Centre for Hydrogeology
and Geothermics (CHYN)
University of Neuchâtel
www.unine.ch/chyn

Advances in characterizing surface water - groundwater interactions: combining unconventional data with complex, fully-integrated models

Ph.D. thesis presented to the Faculty of Science of the University of Neuchâtel
to satisfy the requirements of the degree of Doctor of Philosophy in Science

by

Oliver S. Schilling

Thesis defense date: 01.02.2017
Public defense date: 07.04.2017

Ph.D. Thesis evaluation committee:
Prof. Dr. Philip Brunner, University of Neuchâtel (Director of thesis)
Prof. Dr. Daniel Hunkeler, University of Neuchâtel (Co-director)
Prof. Dr. Harrie-Jan Hendricks Franssen, Forschungszentrum Jülich, Germany
Dr. John Doherty, Watermark Numerical Computing, Australia

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.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 Oliver S. SCHILLING

Titre:

"Advances in characterizing surface water - groundwater interactions: combining unconventional data with complex, fully-integrated models"

sur le rapport des membres du jury composé comme suit:

- Prof. Philip Brunner, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Daniel Hunkeler, co-directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Harrie-Jan Hendricks Franssen, Forschungzentrum Jülich, Allemagne
- Dr. John Doherty, Watermark Numerical Computing, Brisbane, Australie

Neuchâtel, le 16 mars 2017

Le Doyen, Prof. R. Bshary

I have yet to see any problem, however complicated, which when you looked at it in the right way did not become more complicated.

Poul Anderson

Summary

The characterization and simulation of the interactions between surface water and groundwater require observations of hydrological state variables and flow processes. While the latest generation of physically-based flow models allows the integrated simulation of all relevant hydrological processes, the current modelling practice is not adequate to provide reliable predictions. Numerous studies suggest that the main reason for this limited predictive capability is that the complex nature of surface water - groundwater systems cannot be sufficiently described and constrained by only considering the ‘classical’ hydrogeological observations of surface water discharge and hydraulic head. An extensive literature review on this topic is provided in Chapter 2. To overcome the problem of inadequate surface water - groundwater flow model calibration, alternative, unconventional observations should be considered, for example observations of solute concentrations or exchange fluxes. With the appropriate modelling and calibration tools, unconventional observations can not only be successfully included in flow model calibration, but by choosing the right tools it is also possible to quantify the information content of unconventional observations towards reducing the predictive uncertainty of flow models. This was the focus of this thesis and is illustrated in multiple studies: In Chapter 4, a new method that uses tree ring growth records to infer the historic transpiration rates of riparian desert trees of the Tarim River was developed. These new and unconventional observations were successfully used for the calibration of an integrated surface water - groundwater - vegetation flow model built with HydroGeoSphere. A post-calibration uncertainty analysis allowed quantifying the high worth of these unconventional observations in reducing the predictive uncertainty of the flow model. In the study presented in Chapter 5, the established tracer methods using Radon-222 and Helium-Tritium were for the first time complemented by a novel tracer method based on Argon-37, which allowed closing a previously existing gap in residence times characterization. A multi-tracer study carried out on an important drinking water wellfield in Switzerland provided an ideal framework to test the new method alongside other tracer methods. Following the successful characterization of residence times of groundwater in the drinking water wellfield, a quantification of mixing of different types of groundwater, i.e. of recently infiltrated river water and of old groundwater, was achieved through noble gas end-member analysis. The information on mixing of different types of groundwater was subsequently used to inform the parametrization of an integrated surface water - groundwater flow model of the drinking water wellfield. It could be shown that the use of mixing information contains information about exchange fluxes and can successfully inform a flow model parametrization beyond the capabilities of classical observations of groundwater heads and surface water discharge.

One important aspect of surface water - groundwater interactions is the state of connection between the two water bodies. If unsaturated flow processes are expected in a given surface water - groundwater system, it is important that the numerical model which is used to simulate these systems is capable of simulating unsaturated flow. The study in Chapter 3 is dedicated to the development of a Monte-Carlo-based method which allows the rapid quantification of the potential for unsaturated flow processes underneath heterogeneous riverbeds overlying heterogeneous aquifers. This in turn allows a preliminary assessment of the conceptual model of a give surface water-groundwater system.

And finally, as the integrated simulation of surface water - groundwater interactions requires a lot of computational resources, especially when unsaturated flow processes dominate, a modelling framework using HydroGeoSphere, Ensemble Kalman Filter and cloud resources was developed in order to leverage typically limited computational resources. This framework is presented in Chapter 6.

Keywords: surface water-groundwater interactions, tracers, tree rings, Argon-37, noble gases, groundwater model calibration, ensemble Kalman filter, HydroGeoSphere, cloud computing

Résumé

Caractériser et simuler les interactions entre les eaux de surface et souterraines s'avère être un enjeu de plus en plus important afin de garantir une eau de qualité pour des puits de pompage situés à proximité d'une rivière. Grâce à la dernière génération de modèles numériques physiques intégrant les écoulements de surface et souterrains, il est maintenant possible de simuler tous les processus physiques gouvernant ces interactions. Cependant, les résultats de ces modèles sont souvent peu satisfaisants. Plusieurs études suggèrent que l'utilisation des charges hydrauliques et des débits de la rivière n'est pas suffisante pour décrire et contraindre ces processus complexes. Une revue bibliographique de ces différentes études est présentée dans le chapitre 2. Afin de parvenir à un meilleur calage de ces modèles, l'utilisation d'autres types d'observations que l'on peut qualifier de non-conventionnelles doit être envisagée. Ces observations non-conventionnelles peuvent être par exemple les flux caractérisant les interactions ou la concentration de solutés. Le but principal de cette thèse de doctorat est de montrer que l'utilisation de ces observations non conventionnelles permet non seulement d'améliorer significativement le calage de ces modèles, mais également de réduire grandement les incertitudes de leurs projections.

Le chapitre 4 présente le développement d'une nouvelle méthode utilisant l'accroissement de cernes des arbres afin de quantifier l'historique de transpiration des arbres riverains. Les observations non-conventionnelles obtenues grâce à cette méthode ont ensuite été utilisées pour caler un modèle HydroGeoSphere couplant les écoulements souterrains et de surface ainsi que la transpiration des végétaux. Une analyse d'incertitude a permis de quantifier la réduction de l'incertitude induite par l'utilisation de ces observations non-conventionnelles.

Le chapitre 5 présente l'utilisation de traceurs naturels afin de caler un modèle d'écoulements souterrains et de surface. Cette étude fut réalisée dans le cadre d'un essai de traçage réalisé dans un champ de captage situé en Suisse. Après avoir évalué le temps de résidence de l'eau en utilisant les concentrations en argon-37, une analyse du mélange entre les eaux récemment infiltrées et les eaux plus vieilles a été réalisée en analysant les températures d'infiltration des gaz rares. Ces deux informations qui permettent de décrire les interactions entre la rivière et l'aquifère furent utilisées en plus des observations conventionnelles pour caler et contraindre le modèle de ce champ de captage. Les résultats obtenus montrent clairement une amélioration des capacités prédictives du modèle ainsi qu'une diminution de ses incertitudes.

Le chapitre 3 décrit une nouvelle méthode utilisant la méthode de Monte-Carlos afin d'identifier des zones non saturées entre le lit d'une rivière et d'un aquifère tous deux aux propriétés hétérogènes. Finalement, le chapitre 6 présente une nouvelle approche de simulation utilisant HydroGeoSphere, le filtre de Kalman d'ensemble (EnKF) et du nuage informatique (cloud computing) afin d'augmenter les ressources de calcul qui généralement sont importantes dans le cadre de modélisations numériques complexes.

Preface

Water is our most important resource, and in surface water-groundwater systems water interacts with the soil, our second most important resource. We have an obligation to protect these ecosystems and our two most important resources, especially to guarantee a safe and sustainable drinking water supply. After a little more than four years of working on my Ph.D. at the CHYN, with a lot of effort, countless field-trips, numerous inspiring discussions, hundreds of papers read, multiple conference visits and interesting presentations, many new colleagues met and friends won, this thesis in a way represents the culmination of all this. To me the main goal in the beginning of this Ph.D. was to achieve a better understanding of the incredible complexity of hydrogeological modelling, because I came to realize that surface water-groundwater modelling requires expertise not only in hydrology and hydrogeology, but also in algebra, numerics, inverse mathematics, statistics, biology, geology, climate processes, fluid dynamics, isotope chemistry and ecology. Only if one has a sufficient understanding of all these different disciplines one can produce reasonable flow models and predictions with them. I realized that there is still much that I needed to learn, so that the simulations I make can better support the decision-making process. Now that I finished my Ph.D. at the CHYN, I can say that I have learned a tremendous amount about almost all the topics above. But I also came to accept that being a real expert in all of the individual fields simultaneously is pretty much impossible. However, I'm convinced that surface water-groundwater systems research requires interdisciplinary knowledge, and this in turn requires the collaboration of experts from many different fields and groups. Luckily I had the chance to work with some of the best in the field, and this really helped taking the research presented in this thesis to a whole other level. With this very interdisciplinary Ph.D. thesis I hope that I contribute to supporting a more scientifically sound and more sustainable decision-making process when it comes to protecting our two most important resources water and soil.

Acknowledgements

This thesis would not have been possible without the support and contributions of many excellent colleagues and friends. I want to thank all the people who worked with me, and made this thesis possible. Working with you was great, and I've learned a tremendous amount from everyone's unique expertise. First and foremost, I would like to thank Roberto, Laurent, Anne-Marie, Christoph and Roland. Your support and expertise in the field and with measurement techniques was indispensable and crucial to the success of my Ph.D. Philipp, Alice, Claire, Léa, Celia, Julien, Fabien and James it was a pleasure working with you at the CHYN. Thank you for your support both in the field and elsewhere. Dear officemates Christian, Guillaume, Dan Partington and Daniel Kaeser, it was great sharing E119 with you. The discussions about work but also about (PhD) life were inspiring. Daniel Kaeser, thank you very much for all the great work you did in the Emmental prior to my arrival, and your introduction and support in the beginning of my Ph.D. Dan Partington, what can I say? Thank you for the inspiration, the help, and above all your friendship. I'm looking forward to future collaborations and happenings together. Wolfgang, Tang Qi and Andrei, thank you for our great collaboration on our HGS-EnKF-cloud project. What an effort! It's incredible that we managed to pull this off between Jülich and Neuchâtel. Simon, Matthias, Lina, Yama, Elahe, Yvo, Wendelin and Andrea, thank you for your help in the lab and with the noble gas and miniRuedi measurements. Dylan and Sarah, you inspired me and became good friends in Australia. With your great scientific minds and your academic progression, you continue to inspire me. Jürg and the whole Weber-Bann group, thank you for providing me asylum at ETH Hönggerberg on Fridays and doing sports with me. Eric and Tom, it was great to regularly present the latest progress of my Ph.D. at the SFWMD headquarters in West Palm, and to do the occasional field work with you guys at the LILA site. Thank you for supporting me. Eric, you are an inspiration to me in many ways and I hope that we can keep up our almost annual meetings. Lynn and Lloyd, doing two internships with you and visiting you in Australia essentially shaped my ideas about where I want to go and what I need to know to get there. The time spent and discussions held with you during our time together at CSIRO and NTEC motivated me to do this Ph.D. in the first place, and provided me with an outside-of-the-box, let's-first-take-a-step-back-and-think way of thinking that proved to be essential during my studies. I would like to highlight that nothing would have been possible without the endless support of my supervisors. Philip, Daniel, Harrie-Jan and RoKi, your inputs and support was extremely appreciated and I learned a tremendous amount from you. Thank you for the many inputs and ideas for the Emmental project, Daniel. Roland, it was great to have you and Christoph aboard the Emmental project and do the multi-tracer study. I think this turned out to be an exceptionally good project. Dear RoKi, you have supported me throughout my BSc, MSc and PhD, and we are even going to collaborate during my postdoc. I am sure that you understand how much I appreciate this. Harrie-Jan, thank you for your inputs and the many discussions with Wolfgang, Tang Qi and Philip. The meetings always provided new insights and ideas, and helped to improve my work. And last but not least, Philip. You are my science mentor, your inspiration makes me want to keep on doing research, and you are the best supervisor I could imagine. You were always on top of things, and your inputs always helped me to progress my work. I think we achieved more than we hoped for in the beginning. I cannot thank you enough. I hope that we can continue our fruitful collaborations, and develop and investigate many more ideas and topics together. Finally, of course, I especially want to thank my girlfriend Tanja, my family Ursula, Sigurd and Iris, and my friends. Thanks to you I was able to focus on work when I wanted to, but also forget about work when I needed to. You're the best family and friends one can have.

Table of contents

Summary .... **VII**

Résumé .... **IX**

Preface .... **XI**

Acknowledgements .... **XIII**

1. Introduction **2**
- 1.1. Context and motivation 2
  - *1.1.1. Surface water-groundwater systems* *2*
  - *1.1.2. Current state of research* *3*
- 1.2. Research objectives & structure of this thesis 5
  - *1.2.1. Field sites* *5*
  - *1.2.2. Review of unconventional data in SW-GW model calibration* *5*
  - *1.2.3. Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems* *6*
  - *1.2.4. Using tree rings as a proxy for transpiration in SW-GW model calibration* *6*
  - *1.2.5. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel ^{37}Ar tracer method* *7*
  - *1.2.6. EnKF-HGS: A cloud-based data assimilation framework using the Ensemble Kalman Filter and HydroGeoSphere* *7*
- 1.3. References 7

2. Review on the worth of unconventional observations in groundwater model calibration **12**
- 2.1. Introduction 12
- 2.2. Review of unconventional observation types 16
  - *2.2.1. Review strategy* *16*
  - *2.2.2. Temperature* *18*
  - *2.2.3. Exchange fluxes* *19*
  - *2.2.4. Solute concentrations* *21*
  - *2.2.5. Evapotranspiration* *27*
  - *2.2.6. Soil moisture* *28*
- 2.3. Summary and conclusions 29
- 2.4. References 34

3. Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems **44**
- 3.1. Introduction 45
- 3.2. Materials & methods 48
  - *3.2.1. Overview of proposed approaches* *48*
  - *3.2.2. The stochastic 1-D approach* *48*
  - *3.2.3. Three experiments to assess the reliability of the stochastic 1-D approach* *49*
- 3.3. Terminology and general considerations 49
- 3.4. Numerical simulator 51
- 3.5. Experimental setups 52
  - *3.5.1. Experiment 1: Comparing the unsaturated area estimated with the stochastic 1-D approach to fully-heterogeneous 3-D reference simulations* *53*
  - *3.5.2. Experiment 2: Influence of horizontal redistribution on the development of unsaturated zones* *55*
  - *3.5.3. Experiment 3: Influence of heterogeneity on the development of unsaturated zones* *56*
- 3.6. Results 58
  - *3.6.1. Results of Experiment 1: Comparing the unsaturated area estimated with the stochastic 1-D approach to fully-heterogeneous 3-D reference simulations* *58*

3.6.2. The influence of soil type and background saturation on the development of unsaturated zones 59
3.6.3. The influence of geologic structures on the development of unsaturated zones 62
3.7. Discussion 64
3.8. Conclusions 65
3.9. Acknowledgements 66
3.10. References 66

4. Using tree rings as a proxy for transpiration in SW-GW model calibration 72
4.1. Introduction 73
4.2. Materials & methods 75
4.2.1. Study site 75
4.2.2. Estimating transpiration from tree ring growth 77
4.2.3. Modelling 79
4.3. Results 87
4.3.1. Riparian water use efficiency 87
4.3.2. Predictive uncertainty and data worth 88
4.3.3. Flow of information 89
4.4. Discussion & conclusions 91
4.5. Appendix 92
4.6. Acknowlegements 94
4.7. References 94

5. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel ^{37}Ar tracer method 100
5.1. Introduction 101
5.2. Material & methods 104
5.2.1. Theory of natural environmental tracer methods 104
5.2.2. Study site 107
5.2.3. Pumping experiment & sampling strategy 109
5.2.4. Surface water-groundwater modelling 110
5.3. Results 114
5.3.1. Classical observations: GW levels, temperature and SW discharge 114
5.3.2. Results of the SW-GW simulations 116
5.4. Environmental tracer results 119
5.4.3. Identifying the appropriate model structure: comparing measured and simulated pumped water mix and travel times 126
5.5. Discussion 126
5.6. Conclusions 128
5.7. Acknowledgements 128
5.8. References 129

6. Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources 136
6.1. Introduction 137
6.2. Conceptual framework 139
6.2.1. Data acquisition and management 140
6.2.2. HydroGeoSphere 141
6.2.3. Data assimilation with HydroGeoSphere 141
6.2.4. Cloud computing with CLAUDE 144
6.3. Validation example 146
6.3.1. Hydrological model and data assimilation setup 147
6.3.2. Simulation results 149
6.3.3. Performance analysis of the EnKF-HGS modelling and data assimilation platform in the cloud environment 152
6.3.4. Setup of performance comparison 153

6.4. Discussion 158
6.5. Conclusions 160
6.6. Acknowlegements 161
6.7. References 161

7. Conclusions 168

Chapter 1

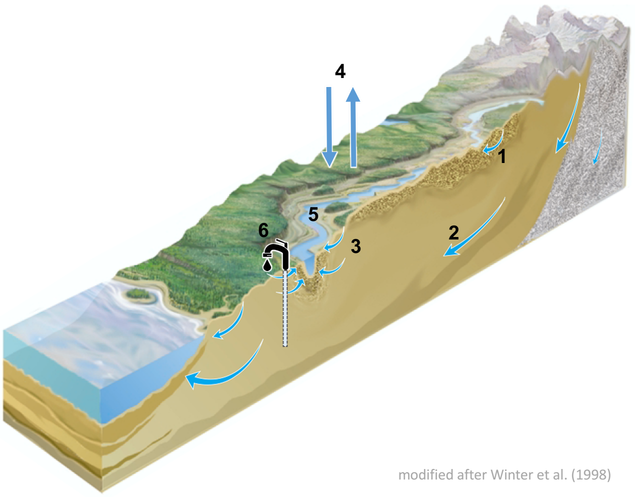
1. Introduction

1.1. Context and motivation

1.1.1. Surface water-groundwater systems

Surface water-groundwater (SW-GW) systems are the focus of a lot of ongoing research (Boano et al., 2014; Boulton et al., 2010; Brunner et al., 2009a; Fleckenstein et al., 2010; Harvey and Gooseff, 2015). Of particular interest are alluvial valleys, where the interactions between groundwater (GW), surface water (SW), and surrounding ecosystems, are highly dynamic on various spatial and temporal scales (**Fig. 1.1**). Such alluvial valleys are immensely important ecosystems and act as hotspots of biological and chemical processes (Boano et al., 2014; Boulton et al., 2010). Moreover, the typically high permeability of alluvial aquifers provides ideal conditions for drinking water production. In Switzerland, for example, 36% of the total drinking water supply is GW pumped in the vicinity of rivers (Diem, 2013; Sinreich et al., 2012; Spreafico and Weingartner, 2005); the study of alluvial systems is therefore of particular importance to Switzerland.

Flow paths in alluvial systems

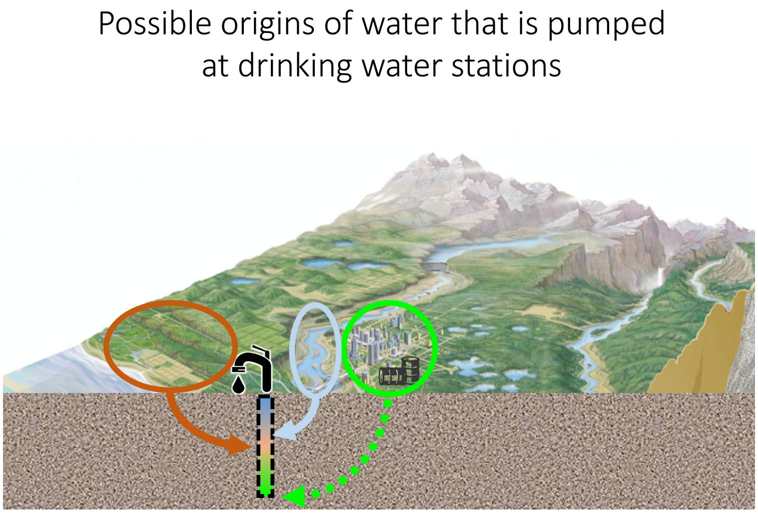


modified after Winter et al. (1998)

Figure 1.1: Conceptual illustration of the flow paths through typical alluvial systems, modified after Winter et al. (1998). The different flow paths include: (1) shallow groundwater flow, (2) deep groundwater flow, (3) hyporheic flow, (4) precipitation and evapotranspiration, (5) surface water flow, and (6) artificial flow paths, such as groundwater pumping.

Alluvial water resources are valuable not only due to the typically high supply, but also due to the high quality (Diem, 2013); their management is therefore necessary to simultaneously guarantee the quality and sustainability of drinking water supplies, as well as, the protection riparian ecosystems from over-exploitation (Delottier et al., 2016).

Many of the river corridors in the world have undergone substantial changes due to anthropogenic influences such as channelization, agriculture, urban development, flood protection, and transport infrastructure. Besides the negative effects of ecosystem degradation due to altered channel morphology, sediment transport, and flow patterns, these anthropogenic developments have also increased the potential for contamination of drinking water resources (**Fig. 1.2**). With a changing climate and a growing human population, the pressure on these resources is expected to further increase in the near future (Aeschbach-Hertig and Gleeson, 2012; Hunkeler et al., 2014).


Possible origins of water that is pumped at drinking water stations

modified after Winter et al. (1998)

Figure 1.2: Conceptual illustration of potential origins of the water that is pumped at alluvial drinking water stations in developed areas, modified after Winter et al. (1998). While the majority of the pumped water most likely originates from recharged river water (blue), both urban runoff (green) and agricultural runoff (brown) could eventually end up in the pumped water mix, and in a worst-case scenario, contaminate the drinking water resources.

To counteract these developments and increase the overall quality of Switzerland’s alluvial ecosystems, improve flood protection measures, and protect the alluvial drinking water resources, the Swiss government has implemented a new legislation, which requires 4’000 km of river restoration and channel-widening until 2084 (Göggel, 2012; Kurth and Schirmer, 2013). This distance is equivalent to the total distance of the Rhine (1’232 km) and the Danube (2’857 km) rivers combined.

Despite this major pending financial investment in river restoration, our understanding and prediction tools for the complex and dynamic alluvial systems have been inadequate to inform how alluvial river systems should ideally be restored and managed. How optimal conditions for sustainable drinking water, ecology health and flood protection can be simultaneously provided is thus still an open question, and is mirrored in the fact that no single, unified quality assessment or post-audit system of river restorations exists in Switzerland; past restorations in Switzerland have been post-audited according to strongly differing, individual measures (Kurth and Schirmer, 2013).

1.1.2. Current state of research

One of the reasons why our understanding of SW-GW systems was not adequate for the definition of a unified quality assessment system for river restorations is the fact that for a long time, SW and GW bodies were considered and studied as separate entities. We now understand that these systems are intimately linked and need to be studied in an integrated way (Winter et al., 1998). Many recent efforts to better characterize and predict connected SW-GW systems have focused on more integrative technologies, and many excellent tools are now available. These necessary tools to further improve our understanding and our predictions of SW-GW systems can be grouped into two main components: (1) integrated measurement techniques that allow the investigation of coupled SW-GW systems, and (2) integrated numerical modelling for SW-GW systems.

Recent advances in measurement techniques now allow observing and quantifying the interactions between SW, GW, and flora and fauna, as well as, drinking water stations, and include:

- a greater ability to measure parameters of SW-GW systems at a larger spatial and temporal extent for example through distributed temperature sensing (e.g., Kurth (2014)), drone-supported hyper spectral temperature sensing (e.g., (Dugdale, 2016)), or photogrammetry to measure the topography of immersed riverbeds (e.g., Feurer et al. (2008)).
- reconstruction of historic discharge regimes through the study of tree rings (dendrochronology, e.g., Gangopadhyay et al. (2009) or Meko et al. (2007))
- new approaches to investigate the subsurface structure of SW-GW systems (e.g., improved geohydraulic characterization of riverbeds through freeze-coring (Strasser et al., 2015)).
- improved detection of extremely rare isotopes, for example through ultra-precise purification techniques (e.g., for ^{37}Ar and ^{39}Ar (Riedmann and Purtschert, 2016)) or through atom trapping (e.g., Ritterbusch et al. (2016)). These developments now allow measuring extremely rare isotopes such as ^{37}Ar, which can be used as a tracer for residence times and mixing in the subsurface on intermediate timescales.

Recent advances in modelling techniques now allow the simulation of the relevant riparian processes and enable predictions of the future behavior of riparian systems, and encompass:

- the fully-integrated, physically-based simulation of surface water, groundwater and vegetation processes in one simulator (e.g., HydroGeoSphere (Therrien et al., 2010), ParFlow (Kollet and Maxwell, 2006)).
- tracking the movement of water through the modelled domain without much extra computational costs (e.g., the Hydraulic Mixing Cell method (Partington et al., 2011)). This enables us to quantify mixing of water from different sources at any point within the model.
- highly parametrized regularized inversion, which allows the simultaneous and systematic calibration of hundreds of parameters according to mathematical procedures that optimally distribute the information content of observations to calibrated parameters (e.g., using PEST (Doherty et al., 2010)).

Substantial potential lies in combining the two different components, novel measurements of SW-GW systems and fully-integrated flow modelling, are combined (Simmons et al., 2012). However, while the recent scientific and technological advances have finally generated the necessary tools to overcome previous limitations in characterizing alluvial SW-GW interactions, Simmons et al. (2012) noted that there is a substantial gap between science and general practice: Many of the advance measurement techniques and modelling codes are rarely used outside of the scientific community.

The fact that this gap exists, and that the potential of unconventional observations in the context of flow modelling remains largely untapped, forms the major motivation for this Ph.D. thesis: The overarching aim is to provide better tools for the sustainable management of SW-GW systems. The following section provides an overview of the research objectives and the different chapters of this Ph.D. thesis.

1.2. Research objectives & structure of this thesis

This Ph.D. thesis is an article-based thesis. The different articles form the different chapters of this thesis, and are self-containing, which means that in every chapter, an introduction to the topic and to the applied methods is given. A brief overview of the main field sites and research articles is provided below.

1.2.1. Field sites

1.2.1.1. Upper Emmental

One of the sites that will soon undergo restoration measures is the River Emme in the Upper Emmental. The Upper Emmental also hosts one of the most important drinking water stations for the Swiss capital Bern: on the Ramsei plain in Aeschau, GW is abstracted in the vicinity of the Emme River to provide roughly 40% of the total drinking water supply to the capital and its surroundings. Due to this interesting setting, the site recently started to be systematically investigated by the Centre of Hydrogeology and Geothermics (CHYN) of the University of Neuchâtel within the Swiss National Science Foundation National Research Program (NRP) 61 "Sustainable water management". Throughout my Ph.D. studies I spent a considerable amount on this field site, and continued and managed the ongoing monitoring and instrumentation.

1.2.1.2. Tarim Basin

The other important field site in this Ph.D. thesis is also characterized by restoration efforts, however, in a different environmental setting and context: The lower reaches of the Tarim River. The Tarim River is located on the border of the Taklamakan Desert, an extremely arid region in Xinjiang, China. The unique riparian forests of the lower reaches of the Tarim River have suffered severe degradation due to an absence of SW flow since the 1970's, when increased agricultural water use upstream began. Around the year 2000, the Chinese government initiated controlled freshwater releases in order to save the strongly endangered riparian forests. As part of my Ph.D., I studied the interactions between the riparian vegetation, the GW and the SW, and investigated the efficiency of these freshwater releases in restoring the riparian forests.

1.2.2. Review of unconventional data in SW-GW model calibration

Addressing the concern raised by Simmons et al. (2012) that in general practice not all available modelling tools and observations are used to assess SW-GW systems, one of the question that remained to be answered was whether, and to what extent, novel observations may improve the characterization of SW-GW systems. While the more different data are available on a system, the better the system can be characterized, how these data should be used to improve our predictions of the future behavior of SW-GW systems has not been systematically assessed. In numerical flow models used for the prediction of SW-GW systems, mainly the classical observations of hydraulic heads and SW discharge are used to calibrate the flow model. How novel, unconventional observations should be included into the flow model calibration process, on the other hand, and how much information these data contain in reducing the predictive uncertainty of the model, is largely unknown. An extensive and urgently needed review of the existing literature on this topic constitutes **chapter 2**. How classical and unconventional data were included in the calibration of numerical flow models, and to what degree these data allowed informing model parameters and reducing predictive uncertainty, is summarized in this chapter.

1.2.3. Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems

An important step in numerical flow model generation lies in the preliminary assessment of the conceptual model and the appropriate model structure. In terms of the conceptual model, it is important to understand and decide, which processes need to be included in the numerical model, e.g., only saturated or also unsaturated flow processes, fully-coupled and physically-based or strongly simplified SW-GW interactions. In terms of the appropriate model structure, it is important to choose the right amount of complexity that is included in the numerical model: Is it, for example, important to represent fully-heterogeneous and spatially-distributed subsurface structures, or can they be simplified to homogeneous structures? Or is it necessary to include unsaturated flow processes, or can the system be adequately represented with saturated flow processes. Multiple studies have investigated the state of connection between rivers and aquifers (e.g., Brunner et al. (2009a) and (2009b), Fleckenstein et al. (2006), Lamontagne et al. (2014), Wang et al. (2016)), and whether simplification of heterogeneous to homogeneous subsurface structures in the numerical model lead to a bias in predictions of the future connection states of the river-aquifer system (Irvine et al., 2012). None of these studies, however, systematically investigated how both heterogeneity of the riverbed and heterogeneity of the aquifer influence the development of unsaturated areas underneath the riverbed; even though an understanding of the potential spatial extent of unsaturated zones is crucial in choosing the appropriate conceptual model and an appropriate complexity for numerical flow simulations. To bridge this gap, we developed a simple stochastic 1-D approach to predict the upper bound of the spatial extent of unsaturated zones for a given heterogeneous riverbed-heterogeneous aquifer system. We furthermore investigated how horizontal capillary forces in the subsurface influence the development of unsaturated zones, and how heterogeneous structures control the distribution of unsaturated areas. This study is in submission as an original research article to Water Resources Research and constitutes **chapter 3**.

1.2.4. Using tree rings as a proxy for transpiration in SW-GW model calibration

How unconventional data can be successfully combined with fully-integrated numerical flow modelling using state-of-the-art inverse methods is presented in **chapter 4**: a novel approach to use tree ring growth records as a proxy for the estimation of historic transpiration rates is used to calibrate a physically-based numerical model. In this study, the worth of the tree-ring-based transpiration estimates in improving predictions made with a numerical flow model was systematically assessed using PEST, a sophisticated inverse tool with integrated uncertainty assessment. The study focused on the interactions between a river, GW and desert vegetation along the Tarim River. The Tarim River is the sole source of water for a vast area in the extremely arid Taklimakan Desert in Northwest China. Through transpiration of water, the trees along the Tarim River form a major component of the water balance of this SW-GW system. Since the 1970's, due to massive divergence of water to a water reservoir in the upstream, the trees in the downstream, which provide the sole protection against the progressing desertification in the area, did not receive water anymore. Through artificial ecological freshwater releases in the Tarim River, starting after the year 2000, the Chinese authorities tried to save these trees by providing desperately needed groundwater recharge. The study summarizes the impact of the freshwater releases on the *Populous euphratica* trees along a stretch of the Tarim River, and quantifies the amount of transpired water. The tree ring growth records, an unconventional data source, served as proxy for plant transpiration. These transpiration estimates were subsequently included in the multi-variable objective function for the calibration of a SW-GW model. This study is an extension of my Master of Science thesis from the Institute of Environmental Engineering (IfU) at the Swiss Federal Institute of Technology in Zürich (ETHZ), Switzerland, and was published as an original research article in the Journal of Hydrology in 2014.

1.2.5. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel ^{37}Ar tracer method

The overarching goal of the studies carried out in the Upper Emmental was to develop an ideal multi-tracer and modelling framework to quantify the interactions between river, groundwater and drinking water abstraction, and to identify the residence times and sources of the pumped GW. For this purpose, (i) a physically-based and fully-integrated numerical flow model of the drinking water wellfield was constructed, (ii) a multi-tracer study using state-of-the-art natural environmental tracers was carried out, and (iii) simulations and measurements were subsequently compared, and an appropriate predictive model for the field site identified. Moreover, in this multi-tracer study, a novel natural tracer method for SW-GW interactions based on ^{37}Ar was employed and compared to established tracer methods. The study was carried out in collaboration with the Climate and Environmental Physics & Oeschger Centre for Climate Research of the University of Bern, and with the Water Resources and Drinking Water Department of eawag. The study is presented in **chapter 5** and in submission as an original research article to Water Resources Research.

1.2.6. EnKF-HGS: A cloud-based data assimilation framework using the Ensemble Kalman Filter and HydroGeoSphere

In a collaborative effort between the IBG-6 of the Forschungszentrum Jülich and the CHYN at the Université de Neuchâtel, we developed a coupling of the physically-based flow model HydroGeoSphere (HGS) to the Ensemble Kalman Filter (EnKF) data assimilation framework, which we called EnKF-HGS. This combination allows continuously updating the system variables of a numerical flow model whenever new observations become available, and simultaneously calibrating flow parameters. This so-called data assimilation framework was then “cloudified” by the IIUN at the Université de Neuchâtel, in order to directly integrate the entire EnKF-HGS data assimilation framework on commercial compute clouds. With the integration of physically-based flow modelling with HGS, EnKF-based data assimilation and cloud-based computing possibilities, we combined the extremely powerful state-of-the-art technologies, which harbor an immense potential as simulation and prediction tools for the real-time management of drinking water wellfields. A similar albeit less powerful framework is already used at the biggest drinking water wellfield in Zurich, Switzerland (Hendricks Franssen and Kinzelbach, 2008). The EnKF-HGS-cloud framework is presented in **chapter 6**, and the original research article is under review at Environmental Modelling and Software.

1.3. References

Aeschbach-Hertig, W., Gleeson, T., 2012. Regional strategies for the accelerating global problem of groundwater depletion. Nat. Geosci., 5: 853-861.

Boano, F. et al., 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Rev. Geophys., 52: 603-679.

Boulton, A.J., Datry, T., Kasahara, T., Mutz, M., Stanford, J.A., 2010. Ecology and management of the hyporheic zone: stream– groundwater interactions of running waters and their floodplains. J. N. Am. Benthol. Soc., 29(1): 26-40.

Brunner, P., Cook, P.G., Simmons, C.T., 2009a. Hydrogeologic controls on disconnection between surface water and groundwater. Water Resour. Res., 45.

Brunner, P., Simmons, C.T., Cook, P.G., 2009b. Spatial and temporal aspects of the transition from connection to disconnection between rivers, lakes and groundwater. J. Hydrol., 376(1-2): 159-169.

Delottier, H., Pryet, A., Dupuy, A., 2016. Why Should Practitioners be Concerned about Predictive Uncertainty of Groundwater Management Models? Water Resour. Manage.

Diem, S., 2013. Riverbank filtration within the context of river restoration and climate change, University of Neuchâtel, Neuchâtel, 144 pp.

Doherty, J.E., Hunt, R.J., Tonkin, M.J., 2010. Approaches to Highly Parameterized Inversion: A Guide to Using PEST for Model-Parameter and Predictive-Uncertainty Analysis, U.S. Geological Survey.

Dugdale, S.J., 2016. A practitioner's guide to thermal infrared remote sensing of rivers and streams: recent advances, precautions and considerations. WIREs Water, 3: 251-268.

Feurer, D., Bailly, J.-S., Puech, C., Le Coarer, Y., Viau, A.A., 2008. Very high-resolution mapping of river-immersed topography by remote sensing. Prog. Phys. Geogr., 32(4): 403-419.

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

Fleckenstein, J.H., Niswonger, R.G., Fogg, G.E., 2006. River-aquifer interactions, geologic heterogeneity, and low-flow management. Ground Water, 44(6): 837-52.

Gangopadhyay, S., Harding, B.L., Rajagopalan, B., Lukas, J.J., Fulp, T.J., 2009. A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. Water Resour. Res., 45.

Göggel, W., 2012. Revitalisierung Fliessgewässer. Strategische Planung. Ein Modul der Vollzugshilfe Renaturierung der Gewässer., Bundesamt für Umwelt, Bern, Switzerland.

Harvey, J.W., Gooseff, M., 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resour. Res., 51.

Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res., 44.

Hunkeler, D., Moeck, C., Kaeser, D., Brunner, P., 2014. Klimaeinflüsse auf Grundwassermengen. Aqua & Gas, 11: 42-49.

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

Kollet, S., Maxwell, R., 2006. Integrated surface-groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. Adv. Water Resour., 29(7): 954-998.

Kurth, A.-M., 2014. Investigation of Groundwater-Surface Water Interactions with Distributed Temperature Sensing (DTS), Université de Neuchâtel, Neuchâtel, Switzerland.

Kurth, A.-M., Schirmer, M., 2013. Thirty years of river restoration in Switzerland: implemented measures and lessons learned. Environ. Earth. Sci., 72: 2065-2079.

Lamontagne, S. et al., 2014. Field assessment of surface water-groundwater connectivity in a semi-arid river basin (Murray-Darling, Australia). Hydrol. Process., 28: 1561-1572.

Meko, D.M. et al., 2007. Medieval drought in the upper Colorado River Basin. Water Resour. Res., 34.

Partington, D. et al., 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface wateregroundwater flow models. Environ. Model. Softw., 26: 886-898.

Riedmann, R.A., Purtschert, R., 2016. Separation of argon from environmental samples for Ar-37 and Ar-39 analyses. Sep. Purif. Technol., 170: 217-223.

Ritterbusch, F. et al., 2016. Groundwater dating with Atom Trap Trace Analysis of 39Ar. Geophys. Res. Lett., 41: 6758-6764.

Simmons, C.T., Hunt, R.J., Cook, P.G., 2012. Using Every Tool in the Toolbox Ground Water, 50(3): 323.

Sinreich, M. et al., 2012. Grundwasserressourcen der Schweiz - Abschätzung von Kennwerten, BAFU, Bern.

Spreafico, M., Weingartner, R., 2005. Hydrologie der Schweiz, Ausgewählte Aspekte und Resultate. Berichte des BWG, Serie Wasser. Bundesamt für Wasser und Geologie (BWG), Switzerland, Bern, Switzerland.

Strasser, D. et al., 2015. Improved geohydraulic characterization of river bed sediments based on freeze-core sampling -development and evaluation of a new measurement approach. J. Hydrol., 527: 133-141.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Hydrogeosphere Manual.

Wang, W. et al., 2016. A quantitative analysis of hydraulic interaction processes in stream-aquifer systems. Sci. Rep., 6(19876).

Winter, T.C., Harvey, J.W., Franke, O.L., Alley, W.M., 1998. Ground water and surface water: a single ressource, USGS, Denver, USA.

Chapter 2

2. Review on the worth of unconventional observations in groundwater model calibration

2.1. Introduction

Integrated and physically-based groundwater flow models (IFMs) are a common tool used for the characterization and management of coupled groundwater (GW) – surface water (SW) resources (Anderson et al., 2015). Simulations with IFMs typically require a significant number of parameters to be defined, but in reality these parameters are rarely known. Due to the limited ability of knowing the true parameter values, especially of strongly heterogeneous subsurface properties such as the hydraulic conductivity (K) or the porosity (*f*), IFMs always need to be calibrated against observations of system states. In general practice, only a limited number of different, 'classical' observations of system states are used to constrain the parameters of IFMs: measurements of hydraulic head (H) and measurements of SW discharge (Q_{SW}). Hydraulic heads are the observation type that is used in the calibration of virtually every regional groundwater flow model – and for a good reason: the information contained in observations of H is of pivotal importance for the reproduction of groundwater heads. Unfortunately, while observations of H contain substantial information about the diffusion or propagation of pressure and the related changes in groundwater heads, they do not contain sufficient information about fluxes, that is, the volume of flow (see for example Anderson et al. (2015), Delottier et al. (2016), Haitjema (2006), Hill and Tiedeman (2007), Townley (2012)). While observations of Q_{SW}, on the other hand, do allow to constrain the water balance, unless measured in a dense measurement network they only provide an integrated information about the catchment response; this information does not suffice in informing about the spatial distribution of the hydrologically relevant parameters, and is thus limited in its capability to inform about the volume of flow in the subsurface. Consequently, the large number of spatially distributed and often statistically correlated parameters required for IFMs cannot sufficiently be constrained by these two classical observation types alone. The high dimensionality and complexity IFMs can thus pose a substantial challenge for robust model calibration and for providing predictions with a reasonably small predictive uncertainty (e.g., Everitt and Hothorn (2011), Doherty (2015), Hill and Tiedeman (2007)). The problem is further amplified when predictions other than those related to purely advective flow are required, for example mass transport or heat transport: the simulation of these additional processes introduces even more unknown parameters to the model, on which neither observations of H or Q_{SW} contain much information. To provide an overview on the different processes and the minimally required parameters and boundary conditions in IFMs, a summary is given in **Table 2.1**.

Tab. 2.1 highlights that it is unlikely that one single observation type provides enough information to sufficiently constrain all unknown parameters as well as boundary and initial conditions at once. The high number of degrees of freedom combined with the limited information content of one single observation type inevitably results in correlation, where the increase in one parameter can be eliminated by the decrease in another parameter (in terms of the reproduction of fluxes, for example, an increase in the hydraulic conductivity K can be compensated by an increase in porosity *f*). In the case of incomplete knowledge of the necessary boundary conditions, boundary conditions might even require calibration as well (e.g., Erdal and Cirpka (2016), Hendricks Franssen et al. (2003)).

Table 2.1: Summary over the different parameters of a groundwater flow problem that need to be calibrated, separated for each type of process. For GW age both the simplified (advective flow only) as well as the explicit (advective-dispersive flow, in brackets) forms are given. Simplifying assumptions: isotropy, homogeneity, single solute with one internal source and one internal sink reaction pathway.

	process	**parameters that require calibration**	**boundary conditions**
GW	Pressure propagation	K	H and/or Q
	Advective flow	K, n	H and/or Q
	Mass transport	K, n, α_{sol}, D_{sol}	H and/or Q, C
	Reactive mass transport	K, n, α_{sol}, D_{sol}, λ_{source}, λ_{sink}	H and/or Q, C
	Heat transport	K, n, α_{heat}, c_{bulk}, ρ_{bulk}, κ_{bulk},	H and/or Q, T
	GW age simulations	K, n, (α_{sol}, D_{sol}, λ_{source}, λ_{sink})	H and/or Q, (C)
SW	Advective flow	n, h_d, h_o	H and/or Q
	Mass transport	n, h_d, h_o, $\alpha_{sol,SW}$, $D_{sol,SW}$	H and/or Q, C_{in}
	Reactive mass transport	n, h_d, h_o, $\alpha_{sol,SW}$, $D_{sol,SW,}$ $\lambda_{source,SW}$, $\lambda_{sink,SW}$	H and/or Q, C_{in}
	Heat transport	n, h_d, h_o, $\alpha_{heat,SW}$, c_{bulk}, ρ_{bulk}, κ_{bulk}	H and/or Q, T_{in}

K	hydraulic conductivity [LT^{-1}]
n	bulk porosity [-]
α_{sol}	(longitudinal / transverse) dispersivity in porous medium[L]
$\alpha_{sol,SW}$	(longitudinal / transverse) dispersivity in surface water [L]
D_{sol}	molecular diffusion in the porous medium [L^2T^{-1}]
$D_{sol,SW}$	molecular diffusion in surface water [L^2T^{-1}]
λ_{source}	mass production coefficient in the porous medium [T^{-1}]
$\lambda_{source,SW}$	mass production coefficient in surface water [T^{-1}]
λ_{sink}	decay/adsorption coefficient in the porous medium [T^{-1}]
$\lambda_{sink,SW}$	decay/adsorption coefficient in surface water [T^{-1}]
α_{heat}	thermal dispersivity of the porous medium [L]
$\alpha_{heat, SW}$	thermal dispersivity in surface water [L]
κ_{bulk}	bulk thermal conductivity [L^2T^{-1}]
ρ_{bulk}	density of the bulk material [ML^{-3}]
c_{bulk}	specific heat capacity of the bulk material [$L^2M^{-2}\Theta^{-1}$]
n	friction coefficients (e.g., Manning roughness coefficients [$L^{-1/3}T$])
h_d	rill/depression storage height
h_o	obstruction storage height
H	First order, Dirichlet boundary condition: fixed hydraulic head [L]
Q	Second order, Neumann boundary condition: fixed flow [L^3T^{-1}]
C_{in}	Fixed concentration boundary condition [ML^{-3}]
T_{in}	Fixed temperature boundary condition [Θ]

The complications that arise due to the limited information content of observations of H and Q_{SW} have been identified in a number of different contexts. Schilling et al. ((under review, Chapter 5 of this thesis)), for example, showed that the calibration of a model of an alluvial drinking water wellfield against observations of H only resulted in non-unique combinations of hydraulic conductivity of the aquifer (K_{aq}), hydraulic conductivity of the riverbed (K_{rb}) and aquifer porosity (f_{aq}). However, the various – in terms of reproducing H – equally likely parameter sets differed dramatically in the simulated exchange fluxes between SW and GW, and, consequently, in the mix between older GW and recently infiltrated SW pumped by the drinking water wells. Only measurements of natural tracers that informed about the source of the pumped water allowed identifying an appropriate parameter set. Similarly, Erdal and Cirpka (2016) investigated how spatially distributed K_{aq} and recharge (R) could be estimated simultaneously based on observations of H alone. The study showed that only if the underlying subsurface structure was already sufficiently well-known and included in the model as prior information on K_{aq}, thus constraining the inverse problem and reducing the degrees of freedom, observations of H were sufficient to reproduce the correct spatial distributions of R and K_{aq} simultaneously. Without this substantial prior information on K_{aq}, the predictive capability of the model was strongly impaired. Kurtz et al. (2012) also found that the information content of observations of H is insufficient to adequately characterize the different parameters of IFMs unless prior information on the subsurface structure is high: In their study, transient changes in the riverbed conductance of a river adjacent to an alluvial drinking water station could only be reproduced through to the inclusion of prior information in the form of a well-calibrated K_{aq}. Also Kerrou et al. (2008) showed the inability of observations of H to calibrate a complex transmissivity field beyond improving the simulation of H. Only the inclusion of direct observations of the transmissivity field as conditioning data for the inverse problem alongside observations of H allowed substantially improving both the simulation of H as well as the representation of the transmissivity field. These studies illustrate that the information content of H is substantial, but that complementary information which either allows constraining the water balance or directly pertains to K_{aq} is necessary to calibrate IFMs.

However, as already indicated previously, the inclusion of observations of Q_{SW} alongside observations of H often is not enough to sufficiently constrain the parameters of IFMs: Ehtiat et al. (2016), for example, compared three different calibration procedures in terms of the ability of a regional SW-GW model to reproduce spatially distributed recharge. The model that best reproduced observations of H during the calibration phase was a model calibrated against H only; however, that parametrization resulted in the worst reproduction of H during a subsequent verification phase due to the inability to adequately simulate R. A better performance was only achieved once R was estimated with a separate recharge model that was individually calibrated against Q_{SW}, showing that the addition of Q_{SW} to the observation dataset is a minimum requirement if GW and SW are simulated in a coupled way and when direct observations of the magnitude and distribution of R are non-existent. However, in a less successful attempt Doppler et al. (2014) calibrated an IFM of a regional SW-GW system against observations of H and Q_{SW} simultaneously, but the resulting model failed to accurately reproduce the dynamics and spatial distribution of H as well as soil moisture (S). The authors suggested that the model should also be calibrated against observations of S. He et al. (2013), too, showed that observations of H and Q_{SW} are insufficient to constrain model outputs other than H or Q_{SW}. They calibrated a flow model against observations of H and Q_{SW} with the aim to reproduce travel times (TT) in a complex geological setting. Without substantially constraining K_{aq} with prior information, calibration of the model against the combined classical observations of H and Q_{SW} only reduced the uncertainty of the simulations of H; the uncertainty of simulations of TT, on the other hand, increased. The problem of inappropriate calibration not only pertains to the simulation and prediction of R, S or TT, neither evapotranspiration (ET) nor spatially distributed Q_{SW} can be estimated with observations of Q_{SW} obtained at the catchment outlet and observations of H: Stisen et al. (2011b) showed this for the Ringkøbing Fjord catchment in Denmark: Calibration of a flow model against H and Q_{SW} resulted in multiple equally-likely parametrizations in terms of their ability to simulate H and Q_{SW}. These parameters significantly differed, however, in the simulation of R, ET and Q_{SW} at discrete locations within the catchment. Stisen et al. (2011b) called for more and new observational data to further constrain the inverse problem and improve the water balance characterization of regional watersheds.

The general practice of limiting the calibration of IFMs to the classical observations of H and Q_{SW} is a stark contrast to the findings as discussed above, and more importantly, to the ever increasing availability of other, unconventional observations of GW-SW systems (e.g., Healy and Scanlon (2010), Boano et al. (2014), Harvey and Gooseff (2015), Jakeman et al. (2016), Maliva (2016)). Such unconventional observation types include (but are not limited to): temperature (T), exchange fluxes (Ex) (such as SW infiltration (I), spatially distributed recharge (R), baseflow (Q_{base}), and GW discharge (Q_{GW})), solute concentrations (C), residence times (RT) and travel times (TT), evapotranspiration (ET), and soil moisture (S). Multiple recent publications suggested that including unconventional observation types in the calibration of IFMs would help to better constrain the parameters of IFMs (e.g., Anderson et al. (2015), Brunner et al. (2012), Hill and Tiedeman (2007), Simmons et al. (2012), Townley

(2012)). However, there are several challenges, potential pitfalls and unresolved issues associated with the use of unconventional observation types for the calibration of IFMs:

(1) **Integration of unconventional observations into IFMs:** In contrast to the classical observation types that are directly simulated by IFMs, some unconventional observations relate to processes that are often not directly simulated by default in all IFMs (e.g., heat and mass transport). The integration of unconventional observations into the IFM calibration process can be roughly separated into three approaches: (i) through an explicit simulation of the respective, additional process, (ii) through a simplified simulation procedure, or (iii) through a pre-processing/transformation of the unconventional observations into an observation type that directly relates to the advective flow simulation. Explicit simulation would provide the physically most appropriate representation of additional processes, but, as has been shown in **Tab. 2.1**, results in more parameters that need to be calibrated, as well as additional initial and boundary conditions that need to be known. Moreover, the explicit simulation of additional processes might result in substantially longer runtimes and harbors potential for numerical instabilities. Simplified simulation routines might provide a more efficient alternative of including additional observation types without introducing additional parameters, but could result in oversimplified physics: Flow tracking (e.g., Partington et al. (2011)) or particle tracking (e.g., Anderson et al. (2015)) schemes that track the movement of water parcels throughout the modelling domain and that allow extracting travel time information of these water parcels might, for example, be based purely on advection and ignore diffusion and dispersion. Neglecting diffusion and dispersion, however, might not always be an appropriate simplification. Similarly, transformation of unconventional observation types to more easily implementable observations prior to their integration into IFM model calibration could facilitate the use of unconventional observations, but the transformation might rely on conceptual models or simplifications that are not always justified.

(2) **Calibration strategy:** Which technique should be used to compare model outputs to unconventional observations, and how should the model parameters be calibrated? Should model parameters be calibrated (i) step-wise with multiple single-variable objective functions, where one parameter is first calibrated against one type of observation and then another parameter against another type of observation and so on, or should model parameters be calibrated (ii) simultaneously with one single weighted multi-variable objective function? Step-wise calibration with multiple single-variable objective functions is likely to result in a local minimum solution of the first parameter that is calibrated, impairing an appropriate calibration of the following parameters (Townley, 2012). Simultaneous calibration of all model parameters against all available observations with s single weighted multi-variable objective function, on the other hand, protects against this; finding the most appropriate weighting scheme for a simultaneous calibration of all parameters against all available observations, however, is often not straightforward (e.g., Anderson et al. (2015), Doherty and Welter (2010), Hill and Tiedeman (2007), Townley (2012), Voss (2011b)).

(3) **Weighting of different observations**: Whether model parameters are calibrated step-wise or simultaneously, one either needs to decide on which observation type to use first (for a step-wise calibration) or on how to weight the different observation types in a multi-variable objective function (for a simultaneous calibration). Both parameters and observations need to be weighted according to their importance as well as

according to their uncertainty. Weighting different observation types, however, is not straightforward, as the different observation types might strongly vary in number, accuracy as well as on their coverage of spatial and temporal scales (Doherty and Welter, 2010). Depending on the calibration order in case of step-wise calibration, or on the weighting scheme in case of simultaneous calibration, very different sets of calibrated parameters can result.

Following from the outlined limitations using classical observations alone, and potential pitfalls associated with the use of unconventional observations, the aim of this review is to analyze the current state of using unconventional observation types in the calibration of IFMs. Moreover, the specific goals of this review are (i) to review how different types of unconventional data were used and to what extent they could improve the model (ii) to identify the methods that allowed to successfully include unconventional observations in the calibration of a flow model, and, if the available literature allows it, (iii) to investigate the worth of different unconventional observation types in calibrating flow models, and (iv) to identify current research needs.

2.2. Review of unconventional observation types

2.2.1. Review strategy

The review was limited to studies that used unconventional observations for the calibration of IFMs of a comparably large scale, with the smallest models having the spatial extent of a contaminant plume. One additional, synthetic study at the scale of an aquifer column was also included in this review, as a very extensive data worth and uncertainty analysis was applied. An overview over the reviewed studies is provided in **Table 2.2**: The table provides information about the study type, the employed observation types, the simulated processes, the modeling codes, the horizontal scale, as well as, the calibration and uncertainty estimation method. The review is grouped by how an unconventional observation is used in the calibration procedure of an IFM, rather than reflecting the method of how the observation was obtained. Residence times, for example, cannot be measured directly but are inferred from tracer concentrations. The use of residence times in flow model calibration nonetheless is discussed in a section separate from the section on the use of solute concentration observations.

Table 2.2: Overview of all reviewed studies, the models used, the calibration technique applied and on which, as well as how, unconventional observation types were included.

study	type	scale (km^2)	additional processes	integration	model code	calibration technique	observation types									
							H	K	Q_{SW}	T	Ex	C	RT & TT	ET	S	other / comments
Ala-Aho et al (2015)	3-D real-world	200			HGS	MLE (PEST)	x				Q_{GW}, Q_{base}					
Alaghmand et al (2014)	3-D real-world	1	mass	explicit	HGS	*manual*	x					EC/salinity				
Alaghmand et al (2016)	3-D real-world	0.5	mass	explicit	HGS	*manual*	x					EC/salinity				(EC from remote sensing)
Bauer et al (2001)	2-D real-world	10	mass	explicit	MODFLOW + MT3DMS	*manual*	x					multiple				(^{3}H, ^{85}Kr, CFC-113, SF_6)
Bonton et al (2012)	3-D real-world	15	mass	explicit	WNGM (Agriflux + HGS)	*manual*	x					nitrate				
Boronina et al (2005a)	2-D real-world	300	mass	explicit	MODFLOW + MT3DMS	*manual*	x				R					
Boronina et al (2005b)	2-D real-world	300	mass	simplified (tracking)	MODFLOW + PMPATH	*manual*	x		x			^{3}H				
Brunner et al (2012)	1-D synthetic				HGS	MLE (PEST)	x							x	x	
Camporese et al (2014)	3-D real-world	0.3			CATHY	*manual*	x		x						x	
Castro and Goblet (2003)	2-D real-world	60	mass	explicit	METIS	?	x					^{4}He				
Castro et al (1998)	2-D real-world	700	mass	explicit	METIS	?	x					^{4}He				
Delsman et al (2016)	3-D real-world	0.1	mass + heat	explicit	SEAWAT + MT3DMS	GLUE	x			x	Q_{GW}	EC/salinity				+ plume depth
Doppler et al (2014)	3-D real-world	1			*original*	MLE	x		x							
Doyle et al (2015)	3-D real-world	25		simplified (tracking)	MODFLOW + MODPATH	*manual*	x									+ recharge elevation from noble gas recharge temperatures
Ehtiat et al (2016)	2-D real-world	360			MODFLOW + SWAT	MLE (PEST)	x		x							
Erdal & Cirpka (2016)	2-D synthetic	22.5			*original*	data assimilation (EnKF)	x									
Gannett et al (2012)	3-D real-world	14000			MODFLOW	*manual*	x				Q_{GW}					
Glaser et al (2016)	3-D real-world	0.002			HGS	*manual*	x	x	x						x	
Gusyev et al (2013)	3-D real-world	1072	mass	explicit	MODFLOW + MT3DMS	*manual*	x		x			^{3}H				
He et al (2013)	3-D real-world	200	age	simplified (tracking)	MODFLOW + MODPATH	MLE (PEST)	x		x							
Heilweil et al (2012)	2-D real-world	15	heat	explicit	VS2DH	*manual*	x			x						+ noble gas recharge temperatures
Hendricks Franssen et al (2003)	2-D synthetic	0.25	mass	explicit	*original*	data assimilation (SSC)	x	x				synthetic				
Hendricks Franssen et al (2008)	2-D real-world	26000			*original*	data assimilation (SSC)	x				R					(R from remote sensing)
Hosseini et al (2011)	2-D synthetic	0.04	mass	explicit	*original*	data assimilation (SSC)	x	x				synthetic				
Hunt et al (2006)	3-D real-world	400	age	simplified (tracking)	MODFLOW + MODPATH	MLE (UCODE)	x				Q_{GW}, Q_{base}		x			+ plume depth
Kerrou et al (2008)	2-D synthetic	0.4			*original*	data assimilation (SSC)	x	x								
Knowling and Werner (2016)	2-D real-world	130			MODFLOW	MLE (PEST)	x	prior			Q_{GW}					
Kurtz et al (2012)	3-D real-world	10			SPRING	data assimilation (EnKF)	x									
Kurtz et al (2014)	3-D real-world	10	heat	explicit	SPRING	data assimilation (EnKF)	x			x						
La Vigna et al (2016)	3-D real-world	55			MODFLOW	MLE (UCODE)	x				Q_{GW}					
Li et al (2009)	3-D real-world	15500			MODFLOW	MLE (PEST)	x							x		(ET from remote sensing)
Ma et al (2010)	3-D real-world	0.002	mass + heat	explicit	MODFLOW + MT3DMS/SEAWAT	MLE (PEST)	x			x		bromide				
Manning and Solomon (2005)	3-D real-world	170	mass+heat+age	explicit	FEMWATER	*manual*	x		x	x	Q_{GW}		x			
Michael and Voss (2009b)	3-D real-world	375	age	simplified (tracking)	MODFLOW + MODPATH	*manual*	x						x			
Oehlman et al (2015)	3-D real-world	150	mass	explicit	COMSOL	automatic	x				Q_{GW}		x			
Rasa et al (2013)	3-D real-world	0.001	mass	explicit	MODFLOW + MT3DMS	MLE (UCODE)	x					bromide	x			
Sanford and Buapeng (1996)	3-D real-world	700	age	simplified (tracking)	MODFLOW + MODPATH	*manual*	x						x			
Sanford et al (2004)	3-D real-world	7900	age	simplified (tracking)	MODFLOW + MODPATH	MLE (UCODE)	x						x			+ hydrochemical zones
Schilling et al (2014 / Chapter 4)	3-D real-world	1			HGS	MLE (PEST)	x							x		(ET from tree ring data)
Schilling et al (under review, Chapter 5)	3-D real-world	1	age	simplified (tracking)	HGS	*manual*	x						x			+ mixing from noble gas recharge temperatures and ^{3}He/^{4}He
Stisen et al (2011a)	3-D real-world	2500			MIKE-SHE / MIKE11	MLE (PEST)	x		x							
Stisen et al (2011b)	3-D real-world	3500	heat	explicit	MIKE-SHE / MIKE11	MLE (PEST)	x		x	x						
van Loon and Troch (2009)	2-D real-world	0.45			*original*	data assimilation (Thikonov)	(x)								x	
Wood et al (2016)	3-D real world	3700	mass	explicit	FEFLOW	MLE (PEST)	x					^{14}C				
Xu and Gomez-Hernandez (2016)	2-D synthetic	0.003	mass + heat	explicit	MODFLOW + MT3DMS	data assimilation (EnKF)	x			x		synthetic				
Zuber et al (2005)	3-D real-world	240	mass	explicit	MODFLOW + MODPATH + MT3DMS	?	x					SF_6				

observation types: H = hydraulic heads, K = hydraulic conductivity/transmissivity, Q_{SW} = SW discharge, Ex = Exchange Fluxes, Q_{GW} = GW discharge, Q_{base} = baseflow, R = Recharge, T = Temperature, Ex = Exchange fluxes, C = Concentration, RT/TT = Residence/Travel Times, ET = Evapotranspiration, S = Soil moisture

calibration methods: EnKF = Ensemble Kalman Filter, MLE = Maximum Likelihood Estimation (Least-squares regression), SSC = Sequential Self Calibration

flow models: CATHY (Camporese et al. 2010), COMSOL (Li et al. 2009), FEFLOW (Dirsch 2014), FEMWATER (Yeh 1987), HGS (Therrien et al. 2010), METIS (Goblet 1999), MIKE-SHE/MIKE11 (Abbott et al. 1986; Graham and Butts 2005), MODFLOW (Harbaugh 2005), SEAWAT (Langevin 2009), SPRING (Delta h 2006), VS2DH (Hsieh et al. 1999)

model packages: AgriFlux (Banton and Larocque 1997), MODPATH (Pollock 2016), MT3DMS (Zheng 2010), PMPATH (Chiang and Kinzelbach 2001), SWAT (Bailey et al. 2016)

2.2.2. Temperature

Measurements of water temperature are comparably easy to obtain, can be made in almost every environment, and are normally available at a high precision (Shanafield and Cook, 2014). Recent advances in airborne and fiber optic techniques have enabled distributed temperature sensing (DTS). In the case of fiber optics-based distributed temperature sensing (FO-DTS), for example, temperature changes can be assessed continuously in time (e.g. Vogt et al. (2010), Kurth (2014)). Temperature measured with airborne hyperspectral and near infrared cameras, on the other hand, allows generating detailed maps of surface temperature, albeit at fewer temporal intervals compared to FO-DTS. Temperature contains information on flow and conduction, as well as on boundary conditions such as temperature sources and sinks. For this reason, temperature has received a lot of attention as a GW tracer and as an indicator for exchange fluxes and mixing processes (Anderson, 2005; Shanafield and Cook, 2014). Unfortunately, the information contained in observations of T is often confounding due to the multitude of different processes that control it, and thus temporally and spatially strongly limited to the time and location of the measurement (Shanafield and Cook, 2014). Moreover, due to the many different processes that influence temperature, the explicit simulation of heat transport requires a relatively large number of additional parameters and forcing functions to be defined, which complicates the application of temperature in flow model calibration. Irvine et al. (2015) compared the suitability of observations of T to observations of C as tracers in heterogeneous media, and concluded that observations of T are more influenced by mixing than C, which makes T a better tracer for the average flow velocity but less suited for an analysis of the spatial distribution of flow. Similarly, Xie et al. (2015) estimated the limits at which SW temperature measurements can be used to quantify SW-GW exchange fluxes, and found that the measurements are extremely limited for quantitative purposes.

In their flow model calibration study, Ma et al. (2012) found that, as a tracer of flow, observations of T are inferior in quality compared to observations of C. The study also showed that only under favorable conditions, that is, mainly advection controlled heat transport with well-known sources, T can be used as a tracer and observation target in locations where measurements of C are too difficult or costly to obtain. In a very extensive study, Delsmann et al. (2016) investigated the value of multiple different unconventional observation types in reducing the ill-position of the inverse problem of flow simulations. Data worth was estimated using the GLUE method (Beven and Binley, 1992; Stedinger et al., 2008) which allowed to compare the influence of the different data types on model calibration and predictions. Overall, Delsmann et al. (2016) showed that the inclusion of unconventional observations allowed improving model predictions and parameters, given that the respective observations were of a high enough quality. The unconventional observation types were especially informative for the simulation of solute transport. In terms of observations of T, mainly two parameters could be improved: f_{aq} and the thermal conductivity of the porous medium. However, Delsmann et al. (2016) also found that: (1) all other observation types employed (i.e., H, Q_{GW} and C) constrained the hydraulic gradients and flow field better than observations of T, and that (2) by including T, all model outcomes other than the simulation of heat transport worsened. Albeit, only the inclusion of observations of C (in their study total salinity load) improved all model outcomes simultaneously. In a study by Stisen et al. (2011a), the calibration of a coupled SW-GW-atmosphere model against observations of H and Q_{SW} was compared to a calibration which included remotely sensed, spatially distributed observations of surface T. The authors could show that only through the inclusion of spatially distributed observations of T the model could reproduce spatially distributed hydrological responses: using only H and Q_{SW} did not create a spatially distributed parametrization that would have been better than a homogeneous parametrization. Similarly, Heilweil et al. (2012), who manually calibrated recharge and hydraulic conductivity of a volcanic island cross-section model against observations of H, T and recharge temperatures inferred from noble gas analysis, could show that only the inclusion of observations of T allowed to constrain recharge locations and K_{aq} to match all observation types simultaneously. By considering observations of T, both obtained directly as well as derived from noble gas analysis, alongside observations of GW discharge and residence times, Manning and Solomon (2005) could distinguish adequate flow regimes from inadequate ones with a model simulating flow as well as heat and mass transport. The identification of inadequate flow regimes would not have been possible to the same degree as brought about through the inclusion of both direct and indirect observations of T. In a sophisticated study, Kurtz et al. (2014) used an Ensemble Kalman Filter (EnKF) approach to assimilated observations of H and T in a GW flow model. They found that simultaneous assimilation of both observation types improved the quality of the simulations compared to an assimilation of H only. However, in their real-world example, the assimilation of T only improved the characterization of the temperature field, but not of the flow field. Xu and Gómez-Hernándenz (2016) used a null-space EnKF approach to simultaneously assimilate H, C and T for the calibration of K_{aq} and n_{aq} in a synthetic model. In their modelling study, the authors could show that the best result was obtained by using all three different observation types in combination. However, in their study the thermal properties of the subsurface were already perfectly known prior to calibration, and only K_{aq} and n_{aq} were uncertain; major confounding effects on temperature were therefore not present.

The studies presented above confirm that the inclusion of unconventional observations in flow model calibration can improve the characterization of flow fields and parameters, but also highlight the confounding nature of observations of T: the inclusion of observations of T in the model calibration process does not guarantee an overall improvement of flow models – an improvement of the simulation of temperature transport can also result in the worsening of the simulation of flow or solute transport. Typically, substantial prior information on the parameters controlling temperature transport are necessary in order to improve a groundwater model overall. The confounding nature of temperature measurements in terms of reproducing advective flow of GW thus poses a critical limitation to the application of T in the calibration process of flow models.

2.2.3. Exchange fluxes

Exchange fluxes between GW and SW encompass a wide range of different types of fluxes: overland recharge (R), infiltration of SW from a SW body (I), and, in the opposite direction, discharge of GW as springs or into surface water bodies (Q_{GW}). Hyporheic exchange can be considered as an alternation between shallow I and Q_{GW}. And when surface water discharge is measured under low-flow conditions, the measurement represents the integrated SW-GW exchange fluxes under the absence of R for that catchment, which is considered as baseflow (Q_{base}). Observations of R, I and Q_{GW} can be obtained both directly, for example with seepage meters, or indirectly, for example through observations of temperature (Anderson, 2005) or solutes (Cook, 2013). However, seepage meter measurements typically provide extremely local information; the strongly distributed nature of exchange fluxes makes them difficult for upscaling. Observations of exchange fluxes are therefore sometimes obtained based on remote sensing of related quantities, e.g. temperature or soil moisture. But if observations of exchange fluxes are not based on direct measurements, the measurement uncertainty becomes more difficult to assess due to the necessary transformation. This, and the fact that some observations of exchange fluxes represent measurements over an integrated scale (e.g., Q_{base}), makes ranking or weighting of observations of exchange fluxes a difficult task. The accurate simulation of exchange fluxes, on the other hand, is less complicated because it does not require additional processes other than flow to be simulated. Moreover, while traditional flow simulators simulate just one of the two flow domains (i.e., either GW or SW) in a physically-based way, the more recent fully-integrated and physically-based flow simulators allow the simulation of both domains in a physically-based and fully-coupled way. This recent development is well-suited for the simulation of exchange fluxes, as it imposes fewer constraints on the exchange flow dynamics.

Hunt et al. (2006) used a regional SW-GW model to systematically assess the value of diverse observation types in the calibration of a flow model. Besides classical observations of water levels in GW and in a lake, they also included unconventional observations of Q_{base}, Q_{GW} (in and out of a lake), depth of a plume of water with elevated electrical conductivity (EC) and travel times (TT), in the calibration dataset. With a systematic data worth analysis based on influence statistics (i.e., Cook's D, see Hill and Tiedeman (2007)) carried out with UCODE (Poeter et al., 2014), they found that Q_{base} from different locations was the most informative observation type, allowing to allocate exchange fluxes to different sub-basins. Q_{GW} was the most important observation type to quantify the lakebed leakage. Hunt et al. (2006) thus found that observations of Q_{GW} and Q_{base} are of particularly high data worth for flow model calibration if used alongside observations of H. Similarly, Manning and Solomon (2005), who used many different observation types to identify probable flow regimes for a large scale, 3-D catchment model (a study already discussed in relation to T, Section 2.2.2), highlighted the benefit of using observations of Q_{GW} to calibrate model parameters. Also Ala-Aho et al. (2015) included observations of Q_{base} and Q_{GW} into rivers and a lake, alongside observations of H and lake stage, to calibrate a 3-D flow model of a regional SW-GW system. While the model was not calibrated using an automated procedure, choosing parameter values that matched all the different observations types allowed discarding solutions that would have correctly reproduced H and lake stages but failed in reproducing Q_{GW} into lakes. Hendricks Franssen et al. (2008) used remotely sensed soil moisture maps to calculate spatially distributed R. Both the calculated R as well as observations of the spatial pattern of R were then used alongside observations of H and topography to stochastically generate hundreds of equally likely solutions to the inverse problem of a 2-D GW flow model. The authors found that inclusion of information from remote sensing (absolute values as well as the spatial pattern) substantially improved the model outcome compared to a model that was only calibrated against H. This was most clearly visible for the spatial distribution of R and in areas where data of H was scarce. Boronina et al. (2005a) estimated R through measurements of stable water isotopes. They subsequently used the estimated R to manually calibrate a 2-D flow model of the Kouris catchment in Cyprus. Similar to the findings of Hendricks Franssen et al. (2008), compared to a calibration of the flow model against H alone, which only allowed calibrating the relation between transmissivities and R, only the direct inclusion observations of R allowed constraining recharge and thus to calibrate both the absolute values of transmissivities and recharge simultaneously. Gannett et al. (2012) calibrated a regional SW-GW model against observations of H and of Q_{GW} into streams. The calibrated model could simultaneously reproduce multi-decadal, climate- and pumping-influenced fluctuations of H and Q_{GW}. Oehlmann et al. (2015) investigated how the hydraulic parameters of Karst systems can be better estimated with a combination of observations of H, Q_{GW} (spring discharge) and tracer breakthrough curves (i.e., travel time information). Oehlmann et al. (2015) found that only a combined, multi-variable objective function of H, Q_{GW} and TT could properly calibrate the Karst conduit network and hydraulic parameters, and step-wise calibration with single-variable objective functions with either one of the three observation types did not result in adequate models. La Vigna et al. (2016) calibrated a regional SW-GW model against observations of H and Q_{GW} (spring discharge). A systematic data worth analysis carried out with UCODE revealed that the Q_{GW} observations allowed informing not only parameters that are sensitive to observations of H but also parameters that are not sensitive to H; the information content of Q_{GW} was therefore essential in calibrating parameters that remain uninformed by H. This was found despite the fact that only 3 data points of Q_{GW} were used alongside 13 data points of H, and despite the fact that the weighting applied in the weighted multi-variable objective function did not guarantee equal importance both observation groups, but instead reflected the measurement uncertainty. This makes an even stronger case for the information content of Q_{GW} (see (Doherty and Welter, 2010). Knowling and Werner (2016) used a regional SW-GW model to estimate recharge, and could show that the inclusion of Q_{GW} in the form of subsurface GW discharge to the ocean improved the overall estimation of recharge in the catchment. The estimation of recharge without direct observations of recharge proved to be difficult, and the information content of Q_{GW} to inform on the spatial distribution of recharge was limited. In the same study as discussed for temperature (Section 2.2.2), Delsmann et al. (2016) included observations of Q_{GW} into tiles and a ditch alongside observations of H and C. The inclusion of Q_{GW} alongside measurements of H and C substantially improved the reliability of the model compared to using only the classical observations of H. Overall, the studies which applied exchange fluxes showed that the inclusion of observations of exchange fluxes reduced the uncertainty of exchange flux simulations compared to only using classical observations, despite the comparably large uncertainty associated with exchange flux measurements.

The studies that used observations of exchange fluxes, and in particular the studies that applied observations of R and Q_{GW}, make a strong case for using observations of exchange fluxes in flow model calibration: in every study reviewed, the inclusion of exchange flux observations improved the flow model.

2.2.4. Solute concentrations

Observations of solute concentrations (C) can reflect many different types of solutes and measurements: They can, for example, pertain to situations where there is a known, local source of a solute (e.g., a pollution plume, or perhaps a known area of stream infiltration that has an identifiable concentration), they can inform about a source that is spatially invariable but temporally changing (e.g., tritium, or a time marker such as low chloride concentrations due to an increase in recharge following clearing), or the solute could be produced in the subsurface in a spatially and temporally variable manner (e.g., Radon-222 or Argon-37). Moreover, besides reflecting different situations of production and transformation, observations of C can also be used in multiple different ways: either directly for the calibration of explicit mass transport models, or indirectly as derived quantities, for example as travel time or pollution plume depth information. Measurements of C can generally be made at a high precision. This facilitates a ranking or weighting of observations of C. Depending on the solute, however, its implementation into a flow model is more or less straightforward: less parameters are required for observations of conservative solutes compared to observations of non-conservative solutes (see **Tab. 2.1**). Moreover, in the special case where transport of a conservative solute is predominantly driven by advection, a simple flow tracking scheme based on the advection equation, which requires significantly less parameters compared to an explicit simulation of mass transport, might be justified. Unless such simplifications can be made, however, the simulation of solutes typically requires the explicit simulation of mass transport based on an advection-dispersion equation, which in turn requires additional model parameters to be defined (see **Tab. 2.1**). However, in some cases observations of solutes which would require the simulation of additional processes can be transformed prior to their implementation into the model calibration process, for example to observations of fluxes (Q_{GW}) or observations of travel and residence times (TT & RT). Such transformations might allow reducing the number of processes and associated parameters that need to be implemented and calibrated in the flow model. On the other hand, this might introduce inaccuracies due to simplifying assumptions necessary during the transformation step. Not all solutes are therefore equally easy to implement in flow model calibration.

Studies where observations of C were directly applied are reviewed in section 2.2.4.1, and studies that applied transformed observations of C are reviewed in section 2.2.4.2. Section 2.2.4.2 is furthermore separated into a section that discusses transformations to travel and residence times (0) and a section that discusses other transformations of C (0).

2.2.4.1. Direct observations of C

In a synthetic modelling study simulating flow and conservative transport, Hendricks Franssen et al. (2003) systematically assessed the data worth of observations of C of a conservative solute in constraining the inverse problem and improving predictions of transport with the sequential self-calibration method (SSC, (Hendricks Franssen et al., 2009)). C was used alongside observations of H and direct observations of K to calibrate a transmissivity field. The best overall result was achieved by using all three observation types simultaneously, whereas both calibration against H and calibration against C only resulted in poor parametrizations for the simulation of flow and transport. However, combining either H or C with direct observations K both resulted in better parameterizations than using H and C combined. Hendricks Franssen et al. (2003) thus showed that complementing observations of H with observations of C of a truly conservative solute substantially improved the predictive capability of the calibrated model. This study, furthermore, underlines the importance of direct observations of the parameters to be calibrated, and their inclusion in the model calibration procedure. Similarly, in another synthetic flow and mass transport modelling study of Xu and Gómez-Hernándenz (2016) that employed an EnKF (Hendricks Franssen and Kinzelbach, 2008) data assimilation scheme, using observations C (of a conservative solute) alongside H and T substantially reduced the uncertainty of K_{aq} and n_{aq}. However, only using observations of C to constrain the inverse problem resulted in the worst performance of all the possible combinations between H, T and C. Xu and Gómez-Hernándenz (2016) therefore suggested that for the most realistic reproduction of heterogeneous structures in GW flow and mass transport problems, it is best to always assimilate H in combination with at least one other, unconventional type of observation. Alaghmand et al. (2014) and Alaghmand et al. (2016) investigated floodplain salinity in a losing river-floodplain system in response to a lowering of the regional water table and to an artificial river stage management/flooding. The successful calibration of the SW-GW model against a combination of observations of H and C allowed reproducing the propagation of the infiltrating river water into the floodplain and producing recommendations for local floodplain salinity and water management. Bonton et al. (2012) used an extended, physically-based flow and mass transport modelling framework to simulate nitrate transformation and transport in the capture zone of a drinking water well in an agricultural area. The authors concluded that through the calibration of the model against not only observations of H but also against observations of C, changes in nitrate in the drinking water well as a function of agricultural use could be more accurately simulated compared to a calibration against solely one of the two observation types. Castro et al. (1998) simulated GW flow and transport in the Paris Basin. The model was only calibrated against observations of C in the form of measured ^4He concentrations. Both K_{aq} and K of the aquitards, as well as, the sources of ^4He could be successfully calibrated and resulted in a good match between simulated and observed hydraulic heads. In a follow up study, Castro and Goblet (2003) more systematically quantified the degrees of freedom of K if a large-scale GW model is only calibrated against H. They suggested ^{4}He is a well-suited proxy for observations of C, which could constrain the degrees of freedom of K and thus improve model calibration. In their study, calibration of K against H created multiple equally-likely solutions to the inverse problem, whereas the inclusion of observations of C allowed discarding all but one of these solutions. Boronina et al. (2005b) simultaneously simulated flow and the transport of ^{3}H with a flow tracking scheme. In their multi-step calibration approach, the authors found that if K_{aq} was first calibrated only against classical observations of H and Q_{SW}, subsequent calibration of n_{aq} against observations of C in the form of ^{3}H did not allow accurately reproducing measurements of ^{3}H. Gusyev et al. (2013) also simulated the transport of ^{3}H through a regional SW-GW system by explicitly simulating mass transport. In a more successful multi-step calibration attempt than Boronina et al. (2005b), Gusyev et al. (2013) could reproduce observed concentrations of ^{3}H through a step-wise calibration of K and R against observations of H and Q_{base}, and n_{aq} against observations of C. However, ^{3}H in regional SW-GW systems cannot be considered a conservative tracer anymore, and as the two studies above did not simulate the radioactive decay of ^{3}H but instead treated ^{3}H as a conservative tracer, the calibrated models might be biased. Bauer et al. (2001) simulated GW flow plus transport of ^{3}H, ^{85}Kr, CFC-113, and SF_6. In a step-wise calibration of a 2-D watershed scale model, they manually calibrated K against observations of H before calibrating effective porosity against available tracer concentrations. Simulated residence times were subsequently compared to measured residence times (based on ^{3}H). The authors could show that the calibration against H and C allowed reproducing hydraulic head, solute transport as well as residence times simultaneously. In a similar approach, Zuber et al. (2005) calibrated a flow and transport model against observations of H and C by means of SF_6 measurements. Only the calibration of the model against C allowed reproducing transport and travel times sufficiently well – calibration against observations of H alone led to unacceptably large discrepancies between the observed and the simulated tracer concentrations. Hosseini et al. (2011) also applied a multi-step approach. They simultaneously calibrated K and the dissolution and biodegradation rates of a model simulating a non-conservative tracer plume of non-aqueous phase liquids (NAPLs). In their synthetic study, Hosseini et al. (2011) first conditioned a heterogeneous K_{aq}-field to observations of H and K_{aq} using the SSC method (Hendricks Franssen et al., 2009), and subsequently calibrated the dissolution and the biodegradation constants of NAPL against observations of C using the distance-function method (Hosseini et al., 2010). With this approach, the simulated concentration was successfully reproduced. In a sophisticated, recent study, Wood et al. (2017) automatically calibrated a GW model with solute transport against observations of H and C obtained from ^{14}C measurements. They used PEST to simultaneously calibrate K_{aq} together with recharge of 18 different zones using a weighted multi-objective function. Weights of the two observation types were set such that both observation types were accounted for equally (as recommended by Doherty and Welter (2010)). This approach led to the finding that the calibration against observations of C strongly helped in constraining the water balance of the investigated watershed beyond the capabilities of a model calibrated against H only, as the sensitivities of recharge and K_{aq} to measurements of ^{14}C were substantial.

2.2.4.2. Transformed observations of C

2.2.4.2.1. Travel times and residence times

Travel time (TT) is the time that groundwater requires to flow between two discrete locations in the subsurface. Residence times (RT) pertain to the special case of travel time between the recharge or infiltration point and a discrete sampling location. The residence time of GW can therefore also be referred to as ‘groundwater age’. Except for the case where measurements of

TT are obtained through artificial tracer injection experiments and the source of the measured solute is perfectly known, interpretation of measured travel and residence times is not always straightforward, as every groundwater sample represents a mix of water with different travel or residence times, i.e., a travel or a residence time distribution. A measurement of travel or residence time of a given GW sample, therefore, represents the average time that the sampled water has spent in the subsurface between two points. To account for this, the measured residence time of a sample is sometimes referred to as 'apparent groundwater age'. To measure travel times, artificial tracer injection experiments, in which a known quantity of a conservative solute is injected in the upstream and the passage of the solute concentration measured at discrete downstream sampling locations, are often carried out. The RT of GW on the other hand is typically estimated through the measurement of natural, non-conservative tracers that are already present in the environment (e.g., ^{222}Rn, ^{37}Ar, ^{3}H/^{3}He, stable isotopes of H_2O). Subsequently, these measured tracer concentrations are transformed to an observation of residence time based on comparably simple mathematical models (e.g., according to exponential production and decay laws and under the assumption of homogeneity of the subsurface material). Observations of TT and of RT therefore represent transformed observations of C.

The integration of observations of TT between two discrete points within a GW flow system into the calibration of a flow model is relatively straightforward, and under favorable conditions (i.e., under predominantly advection transport) can be done with an advective flow model and a flow tracking scheme, rather than with explicit advection-dispersion based mass transport and residence time simulations. Compared to the direct use of observations of C and of TT, using RT for model calibration, on the other hand, can be problematic (Gardner et al., 2013; McCallum et al., 2015; McCallum et al., 2014): Not only do sampled residence times represent mean residence times, which makes it difficult to assess the residence time distribution within a system. Comparing mean sampled residence times to mean simulated residence times harbors the danger of introducing a bias towards younger groundwater ages, as natural environmental tracers are limited in the coverage of residence times: once the maximum estimable age of a tracer is reached, the additional time spent in the subsurface is not recorded by that tracer anymore and the estimated apparent groundwater age might appear too young. For the particular case of simulations of RT, McCallum et al. (2015) have recommended to directly simulate the transport, production and decay of the respective tracer rather than using simpler approaches. This aspect, together with the fact that observations of RT are based on transformations based on simplified mathematical models, and that observations of RT represent integrated measurements of a large temporal and spatial scale, makes it difficult to accurately rank and weight observations of RT. If observations of RT are used for the calibration of a flow model, the weight that is given to RT observations should therefore represent the large uncertainty that is associated with such measurements (Sanford, 2011).

Travel times

As discussed in the previous section on observations of C (Section 2.2.4.2), Rasa et al. (2013) showed that the most informative observation type in terms of reducing the predictive uncertainty of a GW model which could be extracted from an artificial tracer test was the first order moment of the tracer concentration breakthrough curve, that is, the mean TT. The calibration of K as well as dispersivity were improved when the multi-variable objective function of the flow and transport model was made up of a combination of observations of H and of TT, compared to a combination of observations of H and C. Similarly, Oehlmann et al. (2015) used peak arrival times of tracer breakthrough curves as observations of TT for the calibration of a Karst system. In their study, which was already introduced in Section 2.2.3, calibration against a combination of observations of H and Q_{GW} with TT allowed identifying structural errors in the model that were not visible solely by comparing simulated versus observed H. Only a change of the initial model structure allowed satisfying all three components (H, Q_{GW}, and TT) of the multi-variable objective function simultaneously. The information contained in observations of TT in finding an appropriate model structure and parametrization was of pivotal importance for that Karst system. Hunt et al. (2006) estimated the TT of infiltrating lake water to a given set of wells based on measurements of ^{3}H and chlorofluorocarbons (CFCs). The estimated TT was subsequently included in the calibration of a regional SW-GW flow model with flow tracking alongside many other observation types, which were already discussed in the Sections 2.2.3 and 2.2.4.2. However, the observation of TT did not significantly contribute to the calibration of any parameter of the model, which is in contrast to the findings of Oehlmann et al. (2015). This very clearly illustrates that the data worth of an observation type is dependent on the flow system in question and on the other observation types employed during model calibration.

Residence times

Sanford and Buapeng (1996) used RT inferred from ^{14}C to estimate paleo flow conditions, which then helped to define the structure of a regional GW flow model. In contrast to model structures that were not informed through the RT estimates, the model structure that accounted for paleo flow conditions was the only version to successfully reproduce mean RT in a flow and flow tracking framework. Observations of RT were thus successfully used to inform the conceptual model and the model generation of the GW system in question, rather than directly used in the calibration procedure. In the study by Sanford et al. (2004) already discussed in Section 2.2.4, besides observations of H and observations of C the authors also used measured activity concentrations of ^{14}C to calibrate a flow model with a flow tracking scheme. The authors simulated the ^{14}C activity concentrations by adding a post-processing routine, which calculated the activity concertation of ^{14}C from TT based on an exponential decay law. Spatially distributed ^{14}C activity concentrations at the time of recharge represented one of the parameters that were calibrated. Even though an overly tight fit between the model derived ^{14}C activity concentrations and the measured activity concentrations was avoided, the authors argued that through this calibration the spatially distributed recharge could still be much better estimated than in previous studies which only used observations of H. Also Michael and Voss (2009) attempted to calibrate a regional GW flow model with flow tracking of the huge Bengal Basin to RT derived from ^{14}C measurements. A sensitivity analysis showed that the first attempt using only observations of H in shallow wells allowed calibrating just 3 out of the necessary 32 parameters. In a second attempt, the model was manually calibrated against observations of RT only, which resulted in a much better reproduction of both H and RT. Michael and Voss (2009) pointed out that the simplifying assumptions that needed to be applied to the RT estimation from tracer measurements, as well as, to the RT simulation with the model, potentially resulted in substantial structural inaccuracies. Nonetheless, due to the high sensitivity of K to observations of RT and due to the similarity in the optimized parameter values, the authors concluded that observations of RT should be more often included in the model calibration process. In the multi-observation study by Manning and Solomon (2005) (discussed in Sections 2.2.2 and 2.2.3), the inclusion of observations of RT into the manual calibration of a flow model was of pivotal importance to distinguish between inappropriate and appropriate model parametrizations for a regional flow and transport model of the Salt Lake Valley in Utah.

2.2.4.2.2. Other transformations of C

Rasa et al. (2013) investigated the data worth of observations of a conservative tracer experiment in calibrating K and dispersivity of a GW flow and mass transport model. They simulated a long-term artificial tracer experiment and calibrated the model against observations of H and C. The tracer measurements were used as three different types of observations of C: as transient tracer concentration measurements (i.e., C), as temporal first order moments of the tracer breakthrough curves (i.e., TT), and as tracer mass discharge (C absolute). The measurements of C were thus used (1) directly, (2) to estimate mean travel times, and (3) summed up to cumulative observations of tracer mass discharge. In three calibration experiments, one of the three observation types were paired with observations of H, and the three resulting calibrated models were compared. All three calibration experiments resulted in a significant reduction of the predictive uncertainty of the model. Using direct transient observations of C improved the model the least, whereas calibrating the model against temporal first order moments of the tracer breakthrough curves resulted in the most accurate reproduction of the observed tracer plume. Measurements of C transformed to observations of TT thus contained more valuable information for the calibration of K and dispersivity in comparison to direct observations of C. In the multi-observations flow, heat and mass transport modelling study, already discussed in the context of T and Ex (Sections 2.2.2 and 2.2.3), Delsmann et al. (2016) too found that all C-type observations, i.e., electrical conductivity in drains, total salinity load in drains and the depth of the salinity plume, substantially helped calibrating a SW-GW model, particularly in terms of the reduction of uncertainty of the simulated solute transport. Strikingly, only the observation type C could improve all model outcomes, whereas other observation types (i.e. H, T or Ex) worsened at least some predictions other than the ones directly related to the observation type. Similar to Rasa et al. (2013), Delsmann et al. (2016) found that if the measurements of C were first transformed into a more informative observation, i.e. total salinity load and salinity plume depth, calibration was improved compared to using untransformed measurements. Along similar lines as Delsmann et al. (2016), Hunt et al. (2006) (in the study already discussed in Section 2.2.3) also applied transformed observations of C for improved flow model calibration, but rather than explicitly simulating transport used a flow tracking scheme: Hunt et al. (2006) identified the depth of a lake water plume through the analysis of measurements of stable water isotopes, which allowed differentiating GW from different sources (i.e., lake water versus other sources). The information about the lake water plume depth in the subsurface was then used as transformed observations of C and compared to the source of water obtained through the application of a flow tracking scheme. Despite the small number of plume depth observations compared to other observations, e.g. of H, the data systematic worth analysis with (Poeter et al., 2014) showed that the plume depth observations were very important to reduce the global uncertainty of the model. Sanford et al. (2004) simulated a regional SW-GW flow model with the help of a flow tracking scheme, and besides against observations of H, the model was also calibrated against observations of ^{14}C activity concentrations and observations of C (hydro-chemistry) transformed to water source and mixing information. Similarly to Schilling et al. ((under review, Chapter 5 of this thesis)) discussed in Section 2.1, the hydro-chemical information allowed distinguishing different sources zones of GW, and this information was used to estimate the mix of water at the sampling locations. Incorporating this transformed observation of C into the model calibration dataset allowed calibrating the model such that both the spatially distributed recharge as well as the mix of water could be reproduced. The study of Ala-Aho et al. (2015), which was discussed in Section 2.2.3 on exchange fluxes, also used measurements of stable water isotopes, but used them to infer Q_{GW} into lakes. This also represents an application of transformed observations of C, which substantially helped to improve the quality of the model. Doyle et al. (2015) used observations of H in combination with recharge locations derived from noble gas concentrations to calibrate a 3-D model with backward particle tracking. The inclusion of recharge locations as calibration targets significantly helped constraining K and recharge locations, especially in regions where no observations of H were available. Manning and Solomon (2005) and Heilweil et al. (2012) analysed noble gas concentrations, but they used it to recharge temperatures and were thus already discussed in Section 2.2.2, even though this technically also represents a transformed observation of C.

The above findings show that observations of C, both untransformed as well as transformed, harbour a huge potential for flow model calibration.

2.2.5. Evapotranspiration

Observations of evapotranspiration can be divided into observations of direct evaporation from water bodies and soils, and into transpiration driven by vegetation. In SW-GW systems where evapotranspiration represents a significant portion of the overall water balance, measurements of ET contain substantial information about flow processes within the catchment (Abtew and Melesse, 2013). This situation is typically encountered in arid environments. While classical hydrological observations like H and Q_{SW} in arid environments are often scarce, remote sensing facilitates obtaining spatially distributed information on ET at a considerable resolution in these regions. Observations of ET can not only be obtained on the regional scale through remote sensing, but also at the point scale, for example through local measurements of the evapotranspiration potential with evaporation pans, or through direct measurements of soil evaporation and plant transpiration using evaporation chambers. However, the uncertainty of evapotranspiration measurements is notoriously high, as both point and spatially distributed measurements require models to relate the measurements to relevant different local and regional scales (Abtew and Melesse, 2013). This complicates the assessment of the ranking or weighting of observations of ET for flow model calibration. The fact that observations of ET typically represent observations over an integrated scale further complicates the ranking and weighting. The implementation of observations of ET into flow model calibration, furthermore, requires ET processes to be represented by the flow simulator. In case of unsaturated flow simulations, the simulation of ET processes requires additional parameters to be defined, and due to the strong non-linearity of unsaturated flow processes (Hillel, 1998), could potentially result in a significant additional computational burden and numerical instabilities.

Li et al. (2009a), for example, calibrated a regional SW-GW model of the Yanqi Basin in China against observations of H and ET patterns obtained through remote sensing. They systematically assessed different ways of implementing observations of ET (i.e., different ET pattern characterizations) into the calibration process and different combinations of H and ET in multi-variable objective functions using PEST (Doherty, 2015). Li et al. (2009a) found that if the GW level is close to the surface, i.e., smaller than the ET extinction depth, observations of H can be replaced by observations of ET in the model calibration process if a reproduction of H is desired, no matter which type of ET pattern characterization was used. However, while observations of H could be accurately reproduced using only ET observations in the calibration, reproducing both, observations H and ET was not possible without using a combination of observations of H and ET in the calibration process. Nonetheless, the study showed the huge data worth of observations of ET for regional SW-GW model calibration, and the strong potential to use observations of ET as complementary information alongside observations of H. That observations of ET are of high information content in arid systems was also shown by Schilling et al. ((2014, Chapter 4 of this thesis)). The authors developed a relationship between tree ring growth, the depth to GW and plant transpiration for *Populus euphratica* desert trees in the Tarim Basin in China. Based on the established relationship, Schilling et al. ((2014, Chapter 4 of this thesis)) estimated the transpiration along a losing reach of the Tarim River, and calibrated a regional SW-GW flow model against these ET-type observations alongside observations of H. Through a systematic assessment of different weighting strategies for the weighted multi-variable objective function employed in the model calibration carried out with PEST (Doherty, 2015), the authors could show that observations of ET, even if they were associated with a very high uncertainty, are of pivotal importance to achieve an appropriate fit of both H and ET simultaneously. By only using observations of H without observations of ET it was impossible to close the water balance and simultaneously calibrated K_{aq}, n_{aq}, unsaturated soil and vegetation parameters. Through the inclusion of observations of ET alongside observations of H, all parameters could be significantly constrained and the predictive uncertainty of the model substantially reduced. Moreover, through this successful closure of the water balance also the predictive uncertainty of infiltration of river water could be improved, and therefore the water requirements of the desert trees estimated. The fact that observations of H are insufficient to calibrate parameters of unsaturated flow processes, and that observations of ET allow closing the water balance and thus substantially reducing the ill-position of the inverse problem, was also shown by Brunner et al. (2012) in a systematic calibration experiment of a 1-D soil column. While observations of H allowed calibrating the model to reproduce observations of H, the parameter identifiability was weak, and only improved through the inclusion of observations of ET. The application of observations of ET in the model calibration dataset, however, did not substantially improve the reproduction of H in situations where the water table was close to the surface. Where the water table was far away from the surface, on the other hand, the information content of ET for the calibration of model parameters was substantially elevated.

2.2.6. Soil moisture

As H only contains information about the interface of full subsurface saturation, soil moisture observations can be considered complementary information for the unsaturated zone. Soil moisture observations can be obtained both on the point and on the regional scale. Point measurements of soil moisture are typically influenced by local very confined properties (e.g., macropores, preferential flow paths), and therefore not suited for scaling to larger spatial scales (Hillel, 1998). Remotely sensed observations of S, on the other hand, are often used to infer spatially distributed locations and rates of R (Healy and Scanlon, 2010). The application of such observations of soil moisture in flow model calibration typically requires a scaling of the remotely sensed observations to a model-relevant scale. If observations of S are used to infer R, unsaturated flow processes do not need to be considered. If observations of S are directly used, however, unsaturated flow needs to be explicitly simulated, which introduces additional parameters to the flow model.

van Loon and Troch (2002) calibrated a regional SW-GW model against a combination of structured and objectively weighted observations of H and S. While the inclusion of S in the calibration dataset did not improve the reproduction of H, it improved the reproduction of Q_{SW}: Only through the inclusion of observations of S the moisture content in the upper soil could be properly characterized and dependent runoff generation simulated. Camporese et al. (2014) combined observations of H, Q_{SW} and S to calibrate a regional SW-GW flow model. Calibration of K and specific storage of two different zones (riparian and hillslope) was performed in a trial-and-error procedure. A calibration against these three observation types allowed finding a parametrization that successfully reproduced all three observation types. Also Glaser et al. (2016) applied a manual procedure to calibrate K_{aq}, n_{aq} and various ET parameters of a flow model simulating a shallow, regional SW-GW system. Observations of Q_{SW} and two types of observations of S (i.e., soil moisture time series of the top 10 cm of the soil, and mapped surface saturation patterns) were used. The surface saturation patterns were compared by comparing simulated to observed saturated pixels in a visual output. The authors could show that the parameters that influenced the saturation patterns could be varied to match the patterns without substantially changing the simulation of Q_{SW}. This illustrated that the information contained in observations of S allows informing parameters that otherwise would not have been sufficiently constrained, and that this information is complementary to the information contained in Q_{SW}. In the synthetic study of Brunner et al. (2012) already discussed in Section 2.2.5 remotely sensed observations of S (i.e., observations of S on the surface) were of less data worth compared to observations of H and of ET. As for observations of ET, the worth of observations of S in calibrating a flow model was higher for systems with a low water table, compared to systems with a high water table.

2.3. Summary and conclusions

Many textbooks or modelling guidelines (e.g., Anderson et al. (2015) or Barnett et al. (2012)) suggest that as many different observation types as possible should be used for the calibration of flow models. Including unconventional observation types in flow model calibration is, however, associated with major challenges. Moreover, typically only little – if any – guidance is provided on how unconventional observations should be implemented into the calibration of a flow model alongside classical observations. General conclusions on the worth of different observation types has, so far, not been possible and will clearly depend on the geological settings, climatic forcing functions, and predictions of interest. On the upside, both sophisticated mathematical and modelling tools, which are needed to quantify the worth of different observation types and assess the uncertainty of flow model predictions, are now available (e.g., Doherty (2015), Hill and Tiedeman (2007), Poeter et al. (2014)).

This review fills a void in the discussion of the broad and complex topic of unconventional observations in flow model calibration. In general, this review confirmed that including unconventional observation types of at least one important process of the underlying SW-GW system alongside classical observations strongly reduces the ill-position of the inverse problem, improves parameter identifiability, and increases the predictive accuracy of the model. In the following paragraphs, the findings of this review are summarized per observation type

- **Classical observations:** The classical observations H and Q_{SW} are the most common types of observations used in flow model calibration. Their implementation into flow model calibration is straightforward, their measurement uncertainties are typically well known and small, and ranking or weighting of classical observations is comparably simple. In the studies that only used classical observations and that were presented in the introduction, however, prior information on the magnitude and spatial distribution of K, transmissivity and R were necessary to obtain appropriate model parametrizations through calibration. The review confirmed that observations of H contain substantial and important information about the propagation of pressure, but not sufficient information about fluxes. The addition of observations of Q_{SW} improves constraining the water balance of the respective catchment, but only provides an integrated information, which obscures the spatially distributed nature of the underlying processes. While the simulation and prediction of H with a flow model can be strongly improved by a calibration against observations of H, fluxes within the subsurface and spatially distributed exchange dynamics remained largely uncharacterized when only classical observations were used.

- **Temperature:** Temperature is an unconventional observation type that has been tested numerous times as an unconventional calibration target. However, the fact that temperature is controlled by heat transport, which is much more complicated than flow, requires additional processes to be simulated, and introduces new uncertain parameters and forcing functions to the numerical flow model. The possible gain in the predictive capabilities of the flow model is strongly contrasted by the obscuring nature temperature measurements. Both Delsmann et al. (2016) and Kurtz et al. (2014) showed that because of these obscuring aspects of temperature measurements, observations of T only improved the simulation of heat transport, but did not improve – or even worsened – the simulation of flow. If, however, observations of T are obtained through remote sensing and provide substantial information on the spatial distribution of high and low temperatures, the inclusion of this pattern information allowed improving spatially distributed hydrological responses (shown by Stisen et al. (2011a)). Whether and how observations of T should be used as unconventional observations in model calibration therefore depends on the purpose of the model and the nature of the temperature observations: If the simulation and prediction of heat transport is of interest, observations of T should be included directly. If, on the other hand, the simulation and prediction of flow is of interest, and the simulation of heat transport is not, observations of T should better be included as transformed observations, for example as a pattern constraint that informs about zones of R, I or Q_{GW}. Temperature as a tracer for such exchange fluxes has been discussed by Shanafield and Cook (2014) and demonstrated by Stisen et al. (2011a).

- **Exchange fluxes:** Compared to observations of temperature, the application of observations of the different types of exchange fluxes (i.e., R, I, Q_{GW} and Q_{base}) does not require any additional processes, such as heat or mass transport, to be simulated: Exchange fluxes are controlled by hydraulic parameters and the inclusion of exchange flux observations into the calibration of numerical flow models is comparably straightforward, although a physically-based representation of both the GW and SW flow is desirable. However, exchange fluxes are more difficult to observe than, for example, solutes or temperature. The information content of observations of R, Q_{GW} and Q_{base} have proven to be extremely useful in the studies described above, and in most cases resulted in clearly better parameterizations of the respective models compared to using only observations of H (e.g., Delsmann et al. (2016), Hunt et al. (2006), Oehlmann et al. (2015)). The studies that applied observations of Ex showed that if observations of Ex are used alongside classical observations, the quality of flow predictions and the overall quality of the model, can be greatly improved; more than with any other unconventional observation type. The reason for the high information content of observations of Q_{GW}, Q_{base} and R for the simulation of flow can most probably be attributed to the direct dependence on hydraulic parameters, and most likely also to the ability of informing about local exchange processes, which in turn inform about local gradients and heterogeneity, helping to close the local and overall water balance.

- **Direct use of solute concentrations:** Compared to the integration of observations of exchange fluxes into the model calibration process, direct integration of observations of solute concentrations typically requires the explicit simulation of mass transport in addition to flow. The review of studies that directly applied observations of C to mass transport models clearly showed that including observations of C alongside classical observations improves the quality of models that simultaneously simulate flow and mass transport. One of the advantages of explicit mass transport simulations and the direct use of observations of C is that such models can include effects of mixing processes on the solutes (e.g., dispersion and interaction with aquitards). If observations of solutes are transformed prior to their application to a flow model, or if particle tracking models are used, this effect may be missed. However, if a model uses a simplified geology (as all models do), then unless the dispersion parameter correctly simulates the mixing (which is unlikely), such results could be misleading as they do not indicate the potential structural problems within the model.

- **Transformed use of solute observations:** The explicit simulation of mass transport is normally accompanied by a significant increase of the computational burden and an increase in the number of parameters that need to be calibrated. A prior transformation of observations of C to an observation which does not require the explicit simulation of mass transport can therefore be beneficial. The reviewed publications indeed revealed that using transformed observations of C in model calibration often results in simulations that better reproduce the available observation data set and that these transformed observations typically appear contain a higher data worth than compared to direct observations of C. However, such results could also be misleading, as these model outputs do not indicate the potential structural problems within the model. Especially when using C-based travel times or residence times estimations as observations for flow model calibration, there is a large potential for structural errors. A model that is going to be calibrated against observations of TT or RT should therefore ideally be constructed following the the same criteria as a model which is to be calibrated against observations of C directly. However, a model that is to be calibrated against observations of RT or TT requires, in any case, the inclusion of a flow tracking scheme that provides the possibility to extract temporal information. The most advanced flow tracking schemes do not require any additional processes to be simulated, directly use the flow solution calculated by the flow simulator, and allow hydrograph separation at any given point in the flow model (Partington et al., 2011), therefore not introducing an additional computational burden. As outlined earlier, the reviewed studies confirm that the simulation of RT and the subsequent calibration of the flow model against observations of RT is problematic and, therefore, rarely done. However, observations of RT are typically used as information in the conceptual model construction phase (e.g., Sanford and Buapeng (1996); an approach which has been previously suggested in literature (e.g., Gardner et al. (2013), McCallum et al. (2015), McCallum et al. (2014), Voss (2011a)). Observations of TT, on the other hand, are more often used as calibration targets, as they allow substantially reducing the associated uncertainties when obtained from artificial tracer experiments: first order moments (i.e., mean travel times) (Hunt et al., 2006; Rasa et al., 2013) and peak arrival times (Oehlmann et al., 2015) of tracer breakthrough curves have been successfully used to calibrate both coupled flow and mass transport models and flow models that only use advective flow tracking schemes. These studies revealed that for systems that cannot be characterized through observations of H, Q_{SW} and Ex alone, the information content of observations of TT is very high; this is was most important for complex Karst networks (Oehlmann et al., 2015).

- **Evapotranspiration:** Observations of ET are especially worth considering as additional calibration targets in systems where ET represents a major component of the water balance, for example in arid regions. In order for a flow model to be calibrated against observations of ET, evaporation, and possibly plant transpiration, need to be explicitly simulated in addition to flow, which introduces additional equations and parameters to the numerical flow model. Inclusion of observations of ET alongside observations of H and Q_{SW} then allows better constraining the water balance and model parameters. The fact that spatially distributed measurements of ET are comparably easy to obtain through available satellite/remote sensing resources, particularly for areas where other hydrological measurements such as observations of H are scarce or unavailable, forms a major advantage for the application of observations of ET in flow model calibration. Schilling et al. ((2014, Chapter 4 of this thesis)), Li et al. (2009a) and Brunner et al. (2012) systematically assessed that the worth of observations of ET for flow model calibration is large, particularly for data scarce systems, and that their inclusion is comparably straightforward. The most problematic aspect of observations of ET, however, lies in the fact that weighting of such observations is difficult.

- **Soil moisture:** For observations of soil moisture to be directly applied in model calibration, unsaturated flow processes need to be simulated. However, if the spatial pattern of S is measured, for example through remote sensing, the information can be used to infer observations of recharge, which do not necessarily require unsaturated flow or any additional processes to be simulated by a flow model. This review showed that observations of S are applied significantly less often in flow model calibration compared to observations of H, Q_{SW}, Ex, and C. van Loon and Troch (2002) showed that if observations of Q_{SW} are not available, observations of could S help in finding a parametrization that reproduces Q_{SW}. However, if observations of H and Q_{SW}, plus observations of R or Q_{GW} are available, observations of S do not further improve the model (Brunner et al., 2012). Brunner et al. (2012) systematically showed that observations of S are of less data worth than observations of H and of ET for predictions of H. However, if S is measured through remote sensing, the information often strongly correlates with R and Q_{GW}, and is therefore often transformed to these observation types for model calibration, which typically increases the ability of these observations to constrain the inverse problem (e.g., Hendricks Franssen et al. (2008)).

The findings summarized in this review strongly suggest that observations of the different types of exchange fluxes and of solute concentrations contain the most substantial and complementary information if used alongside classical observations in model development and calibration. Observations of T, while comparably easy to obtain and containing substantial information about the hydrogeological system, are too confounded by processes that are not related to the transport of water for direct use in regional SW-GW model calibration. Through this review, we could also identify that often the best results in terms of reducing the overall predictive uncertainty of a model through unconventional observation types could be achieved by using transformed or derived quantities in the objective function rather than the direct observations. For example, compared to directly using salinity as observations of C in the objective function, using the salinity measurements to delineate interfaces between freshwater and saltwater in case of saltwater intrusion reduces the predictive uncertainty of the model to a stronger degree. Or instead of using the highly confounded observations of T directly, using the pattern of high and low temperatures in the objective function substantially increases the data worth of observations of T. The fact that transformed observations in most cases better constrain the inverse problem of flow models compared to the non-transformed observations highlights the fact that the treatment and transformation of measurements prior to using them to calibrate a model is of great importance and can substantially improve the application of unconventional observation types in flow model calibration. This is particularly evident if the transformation results in observations that do not require additional processes such as heat or mass transport to be simulated and can be represented in a simplified manner, for example with a flow tracking scheme. One should, however, exercise due care when choosing to apply a transformation or a flow tracking scheme: depending on the spatial and temporal scale, and the nature of the original observations, transformations and flow tracking schemes might represent an over-simplification and result in biased models (e.g., Turnadge and Smerdon (2014)).

There lies a huge potential in systematic analyses that allow quantifying the worth of different observation types towards informing different model parameters and towards reducing the predictive uncertainty of model predictions. A systematic quantification of the worth of different observation types in calibrating a flow model was, for example, done by Brunner et al. (2012), Delsmann et al. (2016), Hunt et al. (2006), La Vigna et al. (2016), and Schilling et al. ((2014, Chapter 4 of this thesis)). As the current generation of calibration tools, such as PEST (Doherty, 2015) and UCODE (Poeter et al., 2014) provide the means to analyze data worth in a systematic way using the calculated sensitivities between model parameters, model outputs and observations of the automated calibration with little post-processing efforts. Through the stochastic nature of data assimilation approaches (Hendricks Franssen et al., 2009), these methods also allow a direct quantification of the reduction of uncertainty through inclusion of additional observation points. A systematic separation between the contribution of different observation types in reducing the predictive uncertainty of flow models is, however, not directly provided by data assimilation approaches.

Since only a few of the reviewed studies carried out a systematic analysis of uncertainty and data worth, there still lies a large potential to better characterize ideal unconventional observation types and implementation methods. A clear need for systematic data worth and uncertainty analyses whenever flow models are calibrated can thus be identified, particularly when considering that models can never be regarded as a perfect representation of reality. Predictive uncertainty represents an information of pivotal importance for decision makers, and information on the worth of different observation types is essential in order to optimally plan data acquisition field campaigns (Dausman et al., 2010; Doherty, 2015; Doherty and Welter, 2010; Hill and Tiedeman, 2007).

Moreover, the fact that automated calibration techniques allow a simultaneous calibration of all model parameters against one weighted multi-variable objective function that includes all available observations clearly advocates the use of these tools. While the weighting of different observations in the multi-variable objective function might not be straightforward and harbors potential pitfalls, (1) avoiding a fixation on a local minimum, for example characterized by the first observation type used in a step-wise calibration, (2) the potential to systematically assess the influence of different weighting schemes (as done by Schilling et al. (2014 (Chapter 4 of this thesis), and (3) the ability to analyze the worth of different observations in calibrating different model parameters, outweighs the potential pitfalls and is needed for a systematic investigation of unconventional observation types.

A standard procedure for model calibration and data worth assessment should be the following:

1. Application of weighted multi-variable objective functions, that allows a simultaneous calibration of multiple parameters against multiple observation types.

2. Inclusion of prior parameter information into the calibration process (as, for example, done by Erdal and Cirpka (2016) and Hendricks Franssen et al. (2003)).
3. Testing of multiple weighting schemes, which at least includes both a conservative weighting where the inverse of the measurement uncertainty is used, and a weighting that guarantees that all different observation types are approximately equally important in the objective function (as done by Schilling et al. ((2014, Chapter 4 of this thesis))).
4. Assessment of the prior (before calibration) and posterior (after calibration) predictive uncertainty (as done by Schilling et al. ((2014, Chapter 4 of this thesis))).
5. Assessment of the influence of different observation types on the inverse solution, that is, of data worth, by using either distance methods such as Cook's D or a linear predictive uncertainty assessment, as for example done by La Vigna et al. (2016), Hunt et al. (2006), or Schilling et al. ((2014, Chapter 4 of this thesis)).

Finally, it is important to state that modelling and model calibration efforts should follow standard procedures in order to improve the comparability, reproducibility and falsifiability of hydrogeological models. Future research should be directed towards an improved assessment of the best way of implementation and the worth of unconventional observation types for different flow systems, such that modelers can decide, prior to model development, on which observations to best obtain for the system of interest and goal of the model, and on how to best include these unconventional observation types into the flow modelling process. If this would be achieved, the application of unconventional observations in general modelling practice could hugely improve simulations and predictions of flow systems.

2.4. References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E., Rasmussen, J., 1986. An introduction to the european hydrological system – Systeme Hydrologique Europeen, She. 2. structure of a physically-based, distributed modeling system. J. Hydrol., 87(1-2): 61-77.

Abtew, W., Melesse, A., 2013. Evalporation and Evapotranspiraiton. Springer, Dordrecht,Netherlands.

Ala-Aho, P., Rossi, P.M., Isokangas, E., Kløve, B., 2015. Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging. J. Hydrol., 522: 391-406.

Alaghmand, S. et al., 2014. Modelling the impacts of river stage manipulation on a complex river-floodplain system in a semi-arid region. Environ. Modell. Softw., 59: 109-126.

Alaghmand, S. et al., 2016. Quantifying the impacts of artificial flooding as a salt interception measure on a river-floodplain interaction in a semi-arid saline floodplain. Environ. Modell. Softw., 79: 167-183.

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

Anderson, M.P., Woessner, W.W., Hunt, R.J., 2015. Applied Groundwater Modelling (Second edition). Academic Press, Oxford, United Kingdom.

Bailey, R.T., Wible, T.C., Arabi, M., Records, R.M., Ditty, J., 2016. Assessing regional-scale spatio-temporal patterns of groundwater–surface water interactions using a coupled SWAT-MODFLOW model. Hydrol. Process.

Banton, O., Larocque, M., 1997. Agriflux2.0. User's Manual. Software for the Evaluation of Environmental Losses of Nitrates and Pesticides from Agriculture, INRS-ETE, Université du Québec, Quebec, Canada.

Barnett, B. et al., 2012. Australian groundwater modelling guidelines. Waterlines report. National Water Commission, Canberra, Australia.

Bauer, S., Fulda, C., Schäfer, W., 2001. A multi-tracer study in a shallow aquifer using age dating tracers ^{3}H, ^{85}Kr, CFC-113 and SF6 - Indication for retarded transport of CFC-113. J. Hydrol., 248: 14-34.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process., 6(3): 279-298.

Boano, F. et al., 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Rev. Geophys., 52: 603-679.

Bonton, A., Bouchard, C., A., R., Rodriguez, M.J., Therrien, R., 2012. Calibration and validation of an integrated nitrate transport model within a well capture zone. J. Contam. Hydol., 18: 1-18.

Boronina, A., Balderer, W., Renard, P., Stichler, W., 2005a. Study of stable isotopes in the Kouris catchment (Cyprus) for the description of the regional groundwater flow. J. Hydrol., 308(1-4): 214-226.

Boronina, A., Renard, P., Balderer, W., Stichler, W., 2005b. Application of tritium in precipitation and in groundwater of the Kouris catchment (Cyprus) for description of the regional groundwater flow. Appl. Geochem., 20: 1292-1308.

Brunner, P., Doherty, J., Simmons, C.T., 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. Water Resour. Res., 48.

Camporese, M., Paniconi, C., Putti, M., Orlandini, S., 2010. Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. Water Resour. Res., 46.

Camporese, M., Penna, D., Borga, M., Paniconi, C., 2014. A field and modeling study of nonlinear storage-discharge dynamics for an Alpine headwater catchment. Water Resour. Res., 2014(50): 806-822.

Castro, M.C., Goblet, P., 2003. Calibration of regional groundwater flow models: Working toward a better understanding of site-specific systems. Water Resour. Res., 39(6).

Castro, M.C., Goblet, P., Ledoux, E., Violette, S., de Marsily, G., 1998. Noble gases as natural tracers of water circulation in the Paris Basin: 2. Calibration of a groundwater flow model using noble gas isotope data. Water Resour. Res., 43(10): 2467-2483.

Chiang, W.-H., Kinzelbach, W., 2001. 3D-Groundwater Modeling with PMWIN Springer, Berlin Heideleberg.

Cook, P.G., 2013. Estimating groundwater discharge to rivers from river chemistry surveys. Hydrol. Process., 27: 3694-3707.

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

Delottier, H., Pryet, A., Dupuy, A., 2016. Why Should Practitioners be Concerned about Predictive Uncertainty of Groundwater Management Models? Water Resour. Manage.

Delsmann, J.R., Winters, P., Vandenbohede, A., Oude Essink, G.H.P., Lebbe, L., 2016. Global sampling to assess the value of diverse observations in conditioning a real-world groundwater flow and transport model. Water Resour. Res., 52.

Delta h Ingenieurgesellschaft mbH, 2006. Spring 3.3, Witten, Germany.

Dirsch, H.-J.G., 2014. FEFLOW - Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media. Springer, Heidelberg, Germany.

Doherty, J.E., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J.E., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res., 46.

Doppler, T., Honti, M., Zihlmann, U., Weisskopf, P., Stamm, C., 2014. Validating a spatially distributed hydrological model with soil morphology data. Hydrol. Earth Syst. Sci., 18: 3481-3498.

Doyle, J.M., Gleeson, T., Manning, A.H., Mayer, K.U., 2015. Using noble gas tracers to constrain a groundwater flow model with recharge elevations: A novel approach for mountainous terrain. Water Resour. Res., 51: 8094-8113.

Ehtiat, M., Mousavi, S.J., Vaghefi, S.A., Ghaheri, A., 2016. Analysis of recharge conceptualization in inverse groundwater modelling. Hydrolog. Sci. J., 61(15): 2789-2801.

Erdal, D., Cirpka, O.A., 2016. Joint inference of groundwater-recharge and hydraulic-conductivity
fields from head data using the ensemble Kalman filter. Hydrol. Earth Syst. Sci., 20: 555-569.

Everitt, B., Hothorn, T., 2011. An Introduction to Applied Multivariate Analysis with R. Use R! Springer Science+Business Media, New York, Dordrecht, Heidelberg, London.

Gannett, M.W., Wagner, B.J., Lite, K.E.J., 2012. Groundwater Simulation and Management Models for the Upper Klamath Basin, Oregon and California, USGS, Reston, Virgina, USA.

Gardner, W.P., Hammond, G., Lichtner, P., 2013. High Performance Simulation of Environmental Tracers in Heterogeneous Domains. Ground Water, 53: 71-80.

Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L., 2016. On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. Water Resour. Res., 52.

Goblet, P., 1999. Programme METIS: Simulation d'e´coulement et de transport miscible en milieu poreux et fracture, notice de conception, Mise à jour au1er/11/99,Rapp.CIG/LHM/RD/99/38. Ecole Nationale Supérieure des Mines de Paris, Paris.

Graham, D.N., Butts, M.B., 2005. Flexible, integrated watershed modelling with MIKE SHE. In: Singh, V.P., Frevert, D.K. (Eds.), Watershed Models. Taylor and Francis, Boca Raton, FL, USA, pp. 245–272.

Gusyev, M.A. et al., 2013. Calibration of a transient transport model to tritium data in streams and simulation of groundwater ages in the western Lake Taupo catchment, New Zealand. Hydrol. Earth Syst. Sci., 17: 1217-1227.

Haitjema, H., 2006. The Role of Hand Calculations in Ground Water Flow Modeling. Ground Water, 44(6).

Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—the Ground-Water Flow Process, U.S. Geological Survey Techniques and Methods 6–A16. USGS, Reston, Virgina, USA.

Harvey, J.W., Gooseff, M., 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resour. Res., 51.

He, X. et al., 2013. Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time. Hydrol. Earth Syst. Sci., 17: 3245-3260.

Healy, R.W., Scanlon, B.R., 2010. Estimating groundwater recharge. Camebridge University Press, Camebridge, United Kingdom.

Heilweil, V.M., Healy, R.W., Harris, R.N., 2012. Noble gases and coupled heat/fluid flow modeling for evaluating hydrogeologic conditions of volcanic island aquifers. J. Hydrol., 464-465: 309-327.

Hendricks Franssen, H.J. et al., 2009. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. Adv. Water Resour., 32: 851-872.

Hendricks Franssen, H.J., Brunner, P., Makobo, P., Kinzelbach, W., 2008. Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images. Water Resour. Res., 44(44).

Hendricks Franssen, H.J., Gómez-Hernández, J.J., Sahuquillo, A., 2003. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. J. Hydrol., 281: 281-295.

Hendricks Franssen, H.J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res., 44.

Hill, M.C., Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: with analysis of data, sensitivities, predictions, and uncertainty. John Wiley & Sons, Inc., Hoboken, USA.

Hillel, D., 1998. Environmental soil physics. Academic Press, San Diego, USA.

Hosseini, A.H., Deutsch, C.V., Biggar, K.W., Mendoza, C.A., 2010. Probabilistic Data Integration for characterization of spatial distribution of residual LNAPL. Stoch. Environ. Res. Risk Assess, 24(5): 735-749.

Hosseini, A.H., Deutsch, C.V., Mendoza, C.A., Biggar, K.W., 2011. Inverse modeling for characterization of uncertainty in transport parameters under uncertainty of source geometry in heterogeneous aquifers. J. Hydrol., 405: 402-416.

Hsieh, P.A., Wingle, W., Healy, R.W., 1999. VS2DI - a graphical software package for simulating fluid flow and solute or energy transport in variably saturated porous media., U.S. Geological Survey, .

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

Irvine, D.J., Simmons, C.T., Werner, A.D., Graf, T., 2015. Heat and Solute Tracers: How Do They Compare in Heterogeneous Aquifers? Ground Water, 53: 10-20.

Jakeman, A.J., Barreteau, O., Hunt, R.J., Rinaudo, J.-D., Ross, A. (Eds.), 2016. Integrated Groundwater Management - Concepts, Approaches and Challenges. Springer International Publishing, Switzerland.

Kerrou, J., Renard, P., Hendricks Franssen, H.J., Lunati, I., 2008. Issues in characterizing heterogeneity and connectivity in non-multiGaussian media. Adv. Water Resour., 31: 147-159.

Knowling, M.J., Werner, A.D., 2016. Estimability of recharge through groundwater model calibration: Insights from a field-scale steady-state example. J. Hydrol., 540: 973-987.

Kurth, A.-M., 2014. Investigation of Groundwater-Surface Water Interactions with Distributed Temperature Sensing (DTS), Université de Neuchâtel, Neuchâtel, Switzerland.

Kurtz, W., Hendricks Franssen, H.J., Kaiser, H.P., Vereecken, H., 2014. Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. Water Resour. Res., 50: 1665-1688.

Kurtz, W., Hendricks Franssen, H.J., Vereecken, H., 2012. Identification of time-variant river bed properties with the ensemble Kalman filter. Water Resour. Res., 48.

La Vigna, F., Hill, M.C., Rossetto, R., Mazza, R., 2016. Parameterization, sensitivity analysis, and inversion: an investigation using groundwater modeling of the surface-mined Tivoli-Guidonia basin (Metropolitan City of Rome, Italy). Hydrogeol. J., 24: 1423-1441.

Langevin, C.D., 2009. SEAWAT: A Computer Program for Simulation of Variable-Density Groundwater Flow and Multi-Species Solute and Heat Transport, USGS, Fort Lauderdale.

Li, H.T., Brunner, P., Kinzelbach, W., Li, W.P., Dong, X.G., 2009a. Calibration of a groundwater model using pattern information from remote sensing data. J. Hydrol., 377: 120-130.

Li, Q., Ito, K., Wu, Z., Lowry, C.S., Loheide, S.P., 2009b. COMSOL Multiphysics: A Novel Approach to Ground Water Modeling. Ground Water, 47(4): 480-487.

Ma, R., Zheng, C., Zachara, J.M., Tonkin, M.J., 2012. Utility of bromide and heat tracers for aquifer characterization affected by highly transient flow conditions. Water Resour. Res., 48.

Maliva, R.G., 2016. Aquifer Characterization Techniques. Schlumberger Methods in Water Resources Evaluation Series, 4. Springer International Publishing, Switzerland.

Manning, A.H., Solomon, D.K., 2005. An integrated environmental tracer approach to characterizing groundwater circulation in a mountain block. Water Resour. Res., 41.

McCallum, J.L., Cook, P.G., Simmons, C.T., 2015. Limitations of the Use of Environmental Tracers to Infer Groundwater Age. Ground Water, 53: 56-70.

McCallum, J.L., Cook, P.G., Simmons, C.T., Werner, A.D., 2014. Bias of Apparent Tracer Ages in Heterogeneous Environments. 52(2): 239-250.

Michael, H.A., Voss, C.I., 2009. Estimation of regional-scale groundwater flow properties in the Bengal Basin of India and Bangladesh. Hydrogeol. J., 17: 1329-1346.

Oehlmann, S., Geyer, T., Licha, T., Sauter, M., 2015. Reducing the ambiguity of karst aquifer models by pattern matching of flow and transport on catchment scale. Hydrol. Earth Syst. Sci., 19: 893-912.

Partington, D. et al., 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface wateregroundwater flow models. Environ. Model. Softw., 26: 886-898.

Poeter, E.P., Hill, M.C., Lu, D., Tiedeman, C.R., Mehl, S., 2014. UCODE_2014.

Pollock, D.W., 2016. User Guide for MODPATH Version 7—A Particle-Tracking Model for MODFLOW USGS, Reston, Virginia, USA.

Rasa, E., Foglia, L., Mackay, D.M., Scow, K.M., 2013. Effect of different transport observations on inverse modeling results: case study of a long-term groundwater tracer test monitored at high resolution. Hydrogeol. J., 21: 1539-1554.

Sanford, W.E., 2011. Calibration of models using groundwater age. Hydrogeol. J., 19: 13-16.

Sanford, W.E., Buapeng, S., 1996. Assessment of a groundwater flow model of the Bangkok Basin, Thailand, using carbon-14-based ages and paleohydrology. Hydrogeol. J., 4(4): 26-40.

Sanford, W.E., Plummer, L.N., McAda, D.P., Bexfield, L.M., Anderholm, S.K., 2004. Hydrochemical tracers in the middle Rio Grande Basin, USA: 2. Calibration of a groundwater-flow model. Hydrogeol. J., 12: 389-407.

Schilling, O.S. et al., (2014, Chapter 4 of this thesis). Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water– vegetation interactions. J. Hydrol., 519: 2258-2271.

Schilling, O.S. et al., (under review, Chapter 5 of this thesis). Advancing physically-based flow simulations of alluvial systems through observations of ^{222}Rn, ^{3}H/^{3}He, atmospheric noble gases and the novel ^{37}Ar tracer method
. Water Resour. Res.

Shanafield, M., Cook, P.G., 2014. Transmission losses, infiltration and groundwater recharge through ephemeral and intermittent streambeds: A review of applied methods. J. Hydrol., 511: 518-529.

Simmons, C.T., Hunt, R.J., Cook, P.G., 2012. Using Every Tool in the Toolbox. Ground Water, 50(3): 323.

Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resour. Res., 44.

Stisen, S., McCabe, M.F., Refsgaard, J.C., Lerer, S., Butts, M.B., 2011a. Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. J. Hydrol., 409: 337-349.

Stisen, S., Sonnenborg, T.O., Højberg, A.L., Troldborg, L., Refsgaard, J.C., 2011b. Evaluation of Climate Input Biases and Water Balance Issues Using a Coupled Surface–Subsurface Model. Vadose Zone J., 10: 37-53.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Hydrogeosphere Manual.

Townley, L.R., 2012. Calibration and sensitivity analysis. In: Barnett, B. et al. (Eds.), Australian Groundwater Modelling Guidelines. Waterlines Report 82. National Water Commission, Canberra, AUS, pp. 57-78.

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

van Loon, E.E., Troch, P.A., 2002. Tikhonov regularization as a tool for assimilating soil moisture data in distributed hydrological models. Hydrol. Process., 16: 531-556.

Vogt, T., Schneider, P., Hahn-Woernle, L., Cirpka, O.A., 2010. Estimation of seepage rates in a losing stream by means of fiber-optic high-resolution vertical temperature profiling. J. Hydrol., 380: 154-164.

Voss, C.I., 2011a. Editor's message: Groundwater modeling fantasies —part 1, adrift in the details. Hydrogeol. J., 19: 1281-1284.

Voss, C.I., 2011b. Editor's message: Groundwater modeling fantasies —part 2, down to earth. Hydrogeol. J., 19: 1455-1458.

Wood, C., Cook, P.G., Harrington, G.A., Knapton, A., 2017. Constraining spatial variability in recharge and discharge in an arid environment through modeling carbon-14 with improved boundary conditions. Water Resour. Res., 53.

Xie, Y., Cook, P.G., Simmons, C.T., Zheng, C., 2015. On the limits of heat as a tracer to estimate reach-scale river-aquifer exchange flux. Water Resour. Res., 51: 7401-7416.

Xu, T., Gómez-Hernández, J.J., 2016. Characterization of non-Gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures. Water Resour. Res., 52: 6111-6136.

Yeh, G.T., 1987. 3DFEMWATER: A three-dimensional finite element model of water flow through saturated-unsaturated media, Rep. ORNL-8386. Oak Ridge Natl. Lab., Oak Ridge, Tennessee.

Zheng, C., 2010. MT3DMS v5.3, Supplemental User's Guide, Department of Geological Sciences, University of Alabama.

Zuber, A. et al., 2005. Groundwater dating with ^{3}H and SF_6 in relation to mixing patterns, transport modelling and hydrochemistry. Hydrol. Process., 19: 2247-2275.

Chapter 3

3. Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems

Abstract

The presence of unsaturated zones at the river-aquifer interface has large implications on numerous hydraulic and chemical processes. However, the hydrological and geological controls that influence the development of unsaturated zones have so far only been analyzed with oversimplified conceptualizations of flow processes, or homogeneous conceptualizations of hydraulic conductivity in either the aquifer or the riverbed. To predict the spatial extent of unsaturated zones in river aquifer-systems, a novel stochastic 1-D approach, which takes heterogeneity of both the riverbed and the aquifer into account, is proposed. The simple approach constitutes an efficient Monte Carlo sampling technique and allows the rapid estimation of the upper bound of the spatial extent of unsaturated areas underneath the riverbed. The approach requires statistical descriptions of the hydraulic properties of the riverbed and aquifer. Through a systematic analysis based on numerical modelling experiments, we subsequently investigated (1) how horizontal capillary forces control the spatial extent of unsaturated zones, and (2) the importance of heterogeneity in both the riverbed and the aquifer in the development of unsaturated zones. The findings of this study allow estimating the spatial extent as well as the distribution of unsaturated zones based on a simple criterion and the analysis of the statistical properties of the hydraulic conductivity of the riverbed and the aquifer. These analyses can guide modelers in the choice of an appropriate numerical simulator, as well as, on the conceptualization of heterogeneity in the numerical model.

Keywords:
Surface water-groundwater modelling, unsaturated zones, heterogeneity, horizontal redistribution, disconnection, riverbed

This chapter is under review as an original research article:
Schilling, O.S., Irvine, D.J., Hendricks Franssen, H.-J., and Brunner, P. (under review): Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems. Submitted to Water. Resour. Res.

3.1. Introduction

Riverbeds are hotspots of chemical and biological processes, and control the exchange of water and solutes between the river and the aquifer (e.g. Boano et al. (2014)). The permeability of the riverbed controls exchange fluxes between the surface and the subsurface. Modelling and field studies have shown that unsaturated zones can develop under riverbeds (e.g. Brunner et al. (2009a) & (2009b), Fleckenstein et al. (2006), Frei et al. (2009), Irvine et al. (2012), Kalbus et al. (2006), Lamontagne et al. (2014), Newcomer et al. (2016), Treese et al. (2009), Wang et al. (2016)). This situation can occur if parts or the entire riverbed are clogged through fine sediments or biofilms. Three significantly different states (**Fig. 3.1**) of surface water-groundwater (SW-GW) interactions were identified in the above studies: (a) fully saturated conditions (i.e., fully connected), (b) fully unsaturated conditions (i.e., fully disconnected), and (c) partially saturated – partially unsaturated conditions (i.e., transitional). These three states describe the connection status of a SW-GW system across the entire riverbed, rather than that of an isolated column within the SW-GW system.

The presence of an unsaturated zone underneath a riverbed has profound implications on the interactions between surface water (SW) and groundwater (GW): While the flow direction in the unsaturated zone is predominantly vertical and the hydraulic conductivity a function of the water content, the physical properties of the saturated zone remain constant and the direction of flow is not predominately vertical. Numerous field- and modelling studies (e.g. Brunner et al. (2009a) & (2009b), Irvine et al. (2012), Newcomer et al. (2016), Osman and Bruen (2002), Rivière et al. (2014), Wang et al. (2011), Wang et al. (2016), Xie et al. (2014)) have shown that the presence of an unsaturated zone results in infiltration rates that are independent of the large-scale hydraulic gradient between SW and the underlying aquifer. Shanafield et al. (2012) used this concept to explain an unexpected and significant rise of the water table close to an ephemeral river with the presence of an unsaturated zone. Fox and Durnford (2003) demonstrated how the clogging of riverbed and the subsequent development of an unsaturated zone increases the catchment area of a well close to a river. Moreover, the hydraulic behavior associated with the presence of an unsaturated zone has implications for numerous processes, such as hyporheic exchange rates, the transport of heat, solutes or dissolved gases, or the filtration efficiency of bacteria and viruses in the riverbed (e.g. Boano et al. (2014), Harvey and Gooseff (2015)). Understanding the propensity of a given river-aquifer system for the development of unsaturated zones is, therefore, of crucial importance for an adequate conceptualization of the relevant processes and the choice of numerical models used for simulation of such systems, and thus constitutes a key aspect in characterizing SW-GW interactions.

Only relatively few studies focused on the geological and hydrological controls of the development of unsaturated zones beneath streams and rivers. Brunner et al. (2009a) demonstrated that the development of an unsaturated zone is dependent on structural riverbed and aquifer properties, as well as the hydraulic heads in the river and the underlying aquifer. In that study, Brunner et al. (2009a) derived a simple 1-D approach that can be used to determine whether an unsaturated zone can develop underneath a riverbed. The relation is a function of the hydraulic conductivities of both the riverbed (K_{rb} [L/T]) and the aquifer (K_{aq} [L/T]), as well as the vertical riverbed extent h_{rb} [L] and the ponded water depth d [L]. For a fully homogeneous river-aquifer system under losing conditions, Brunner et al. (2009a) found that an unsaturated zone can develop underneath the riverbed if the following criterion is met:

$$K_{rb} / K_{aq} \leq h_{rb} / (d + h_{rb}) \qquad (3.1)$$

Eq. (3.1) shows that an unsaturated zone is more likely to develop beneath a riverbed if K_{aq} increases in comparison to K_{rb}, that is, if K_{rb} acts as a clogging layer. Solving for K_{aq} leads to the critical hydraulic conductivity $K_{aq,critical}$ [L/T], which marks the point that separates the two distinct states saturated and unsaturated:

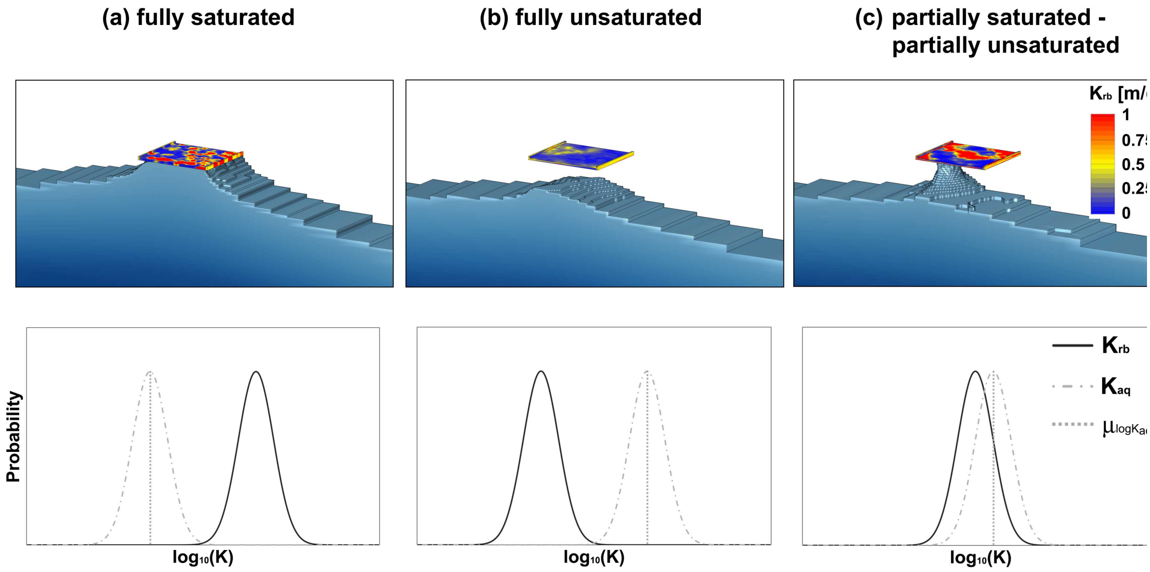
$$K_{aq,critical} = K_{rb} \cdot (d + h_{rb}) / h_{rb} \qquad (3.2)$$

An unsaturated state can thus occur if

$$K_{aq} \geq K_{aq,critical} \qquad (3.3)$$

In a follow up study, Brunner et al. (2009b) also analyzed spatial and temporal controls on the development of an unsaturated zone. The main limitations of these findings are that (i) Eq. 3.1 is a one dimensional (1-D) approach that can only consider vertical flow, and (ii) that in both Brunner et al. (2009a) and Brunner et al. (2009b) the riverbed and the underlying aquifer were assumed to be homogeneous. However, if the spatial distribution of hydraulic conductivities of the riverbed and the underlying aquifer are known, the 1-D approach of Brunner et al. (2009a) could hypothetically be used to estimate if an unsaturated zone underneath the river can develop in a heterogeneous system in response to a lowering of the water table. Whether such a 1-D approach is sufficient to describe the development of unsaturated zones in heterogeneous systems has, however, not yet been analyzed. Horizontal fluxes driven through capillary forces can cause an unsaturated zone under a river section to occur even if no clogging layer is present (Xie et al., 2014). However, Xie et al. (2014) concluded that this is only possible for very narrow rivers. Note that also the study of Xie et al. (2014) was based on homogeneous riverbed and aquifer properties. These capillary effects could nonetheless be important within heterogeneous riverbeds, where small zones without a clogging layer could be located adjacent to clogged areas. In that situation, horizontal fluxes, which are not accounted for in the 1-D approach, could change the state of saturation in isolated pockets underneath the riverbed. It remains unclear how important such horizontal fluxes are, and how they relate to the spatial structure of clogged and un-clogged areas.

Irvine et al. (2012) studied the influence of the heterogeneity of riverbeds on the development of unsaturated zones between a river and an underlying homogeneous aquifer. The authors evaluated under which circumstances the heterogeneity of the riverbed can be simplified by a homogeneous equivalent determined through information on the net exchange flux between river and aquifer, and studied the implications for model predictions based on these homogeneous equivalents. They showed that the state of connection controls how well a heterogeneous riverbed can be represented through a homogeneous equivalent. The three different states of connection of SW-GW systems as presented in Irvine et al. (2012) are illustrated in the top row of **Fig. 3.1**. Corresponding, hypothetical log-normally distributed probability density functions (pdfs) of K_{rb} and K_{aq} that are likely to result in these different states according to Eq. (3.1) are shown in the second column of **Fig. 3.1**.


(a) fully saturated

(b) fully unsaturated

(c) partially saturated - partially unsaturated

Figure 3.1: Conceptual states of connection for heterogeneous riverbeds overlying homogeneous aquifers. Top row: The three different riverbed-aquifer scenarios described in Irvine et al. (2012) that cause the three different states of saturation (a) fully saturated, (b) fully unsaturated, and (c) partially saturated – partially unsaturated. The fully saturated parts of the aquifer are shown in blue. The rectangular, heterogeneous riverbed is centered above a homogeneous aquifer, and colored as a function of the magnitude of K_{rb}. Bottom row: Schematic illustrations of the hypothetical log-normally distributed pdfs of K_{rb} and K_{aq} that could generate the respective states of saturation, according to Brunner et al. (2009a). Note that Irvine et al. (2012) did not consider heterogeneity of the aquifer, but instead based their analysis on a homogeneous value (i.e., μ_{logKaq}).

The main limitation of the study of Irvine et al. (2012) is that heterogeneity was only considered for the riverbed. The influence of aquifer heterogeneity was not evaluated. There are other studies that investigated to what extent heterogeneity should be incorporated into a SW-GW model (e.g. Kurtz et al. (2013), Pryshlak et al. (2015)), or how the heterogeneity should be conceptualized (e.g. Tang et al. (2015)), but none of these studies systematically analyzed how the joint heterogeneity of riverbeds and aquifers influence the distribution and extent of unsaturated zones at the river-aquifer interface.

Given the importance of the degree of saturation at the river-aquifer interface, an appropriate modelling tool as well as an appropriate conceptualization of heterogeneity is of pivotal importance. Knowing the potential of the development of unsaturated zones is an important precondition for choosing an appropriate flow simulator: If, for example, in a river-aquifer system the development of extensive unsaturated zones is expected, a simulator that does not consider unsaturated processes is not appropriate. Knowing the potential of the development of unsaturated zones is also important for the conceptualization of heterogeneity in the subsequent modelling efforts, since simplification of the heterogeneity can, under certain conditions, introduce a bias in predicted exchange fluxes (Irvine et al., 2012). However, the role of heterogeneity in both the riverbed and the underlying aquifer, as well as, the influence of horizontal flow driven by capillary forces on the development of unsaturated zones across the river-aquifer interface, are still poorly understood. In order to advance our understanding on these points, we specifically address the following key questions:

(1) How can prior knowledge of the geological structure allow predicting the potential for unsaturated zones to develop in fully heterogeneous SW-GW systems?

(2) What is the role and importance of horizontal fluxes driven by capillary forces at the interface of saturated and unsaturated areas in heterogeneous systems?

(3) How does structural heterogeneity of the riverbed and the underlying aquifer in combination with horizontal fluxes control the likelihood for unsaturated zones to develop?

The following analyses were carried out: (1) the 1-D approach of Brunner et al. (2009a) was stochastically extended to incorporate structural heterogeneity of both the riverbed and the aquifer. (2) The extent of unsaturated zones estimated with the 1-D approach was compared to stochastic realizations of 3-D reference numerical models. (3) The role of horizontal fluxes in the subsurface driven by capillary forces was analyzed in a systematic way, and (4) the influence of spatial heterogeneity on horizontal fluxes was assessed. Based on these analyses, we ultimately present a method that allows the estimation of the upper limits of the spatial extent of unsaturated zones between a heterogeneous riverbed overlying a heterogeneous aquifer. The proposed approach integrates the influence of heterogeneity in the riverbed and the underlying aquifer.

3.2. Materials & methods

3.2.1. Overview of proposed approaches

In order to answer the questions raised in the introduction, and provide better guidance for modelling of heterogeneous SW-GW systems, a spatial extension to the previously discussed 1-D approach is proposed, and numerous synthetic modelling experiments are carried out. The following paragraphs summarize the approach.

3.2.2. The stochastic 1-D approach

The 1-D approach of Brunner et al. (2009a) does not directly incorporate the heterogeneity of K_{rb} and K_{aq}. A repeated application of the 1-D approach of Brunner et al. (2009a) in a spatially distributed manner could, however, provide the basis to analyze the spatial extent of the unsaturated area across the entire riverbed. For this purpose, we extended Eq. (3.1) to incorporate heterogeneity of both the riverbed and the aquifer by replacing the K_{rb} and K_{aq} with their respective probability density functions (pdf) $P(K_{rb})$ and $P(K_{aq})$. This concept is illustrated on the second row of **Fig. 3.1**, where $P(K_{rb})$ and $P(K_{aq})$ are represented by log-normal pdfs. Eq. (3.2) for $K_{aq,critical}$ then becomes an expression of the pdf of $K_{aq,critical}$ ($P(K_{aq,critical})$):

$$P(K_{aq,critical}) = P(K_{rb}) \cdot (d + h_{rb}) / h_{rb} \qquad (3.4)$$

Eq. (3.4) cannot be solved analytically and therefore a Monte Carlo (MC) procedure is applied: The MC procedure can be used to efficiently sample from $P(K_{rb})$, and insert the value into Eq. (3.4) to solve for $P(K_{aq,critical})$. Also $P(K_{aq})$ is sampled and inserted into Eq. (3.3) to evaluate, together with $P(K_{aq,critical})$, the criterion. Eq. (3.3) is evaluated as many times as MC samples are taken. For the complete set of MC samples obtained in this way, the ratio of riverbed-aquifer pairs that can lead to unsaturated conditions versus the total number of pairs provides an estimate of the spatial extent of the unsaturated area in and below the riverbed. For a number of reasons, however, this approach might not be an accurate predictor: As in the original approach of Brunner et al. (2009a), the stochastic 1-D approach does not take horizontal fluxes driven through capillary forces into account.

3.2.3. Three experiments to assess the reliability of the stochastic 1-D approach

To assess the predictive power of this spatially distributed albeit simplistic predictor, in a first experiment the results were compared to fully-heterogeneous 3-D reference simulations made with a realistic cross-sectional river-aquifer model (Experiment 1). The model setup for the reference simulations for Experiment 1 is described in section 3.5.1, and the results of the two approaches are compared in section 3.6.1. This experiment addresses question 1 raised in the introduction.

Experiment 2 was designed to address question 2 raised in the introduction, and investigate the importance of horizontal redistribution of water driven by capillary forces at the interface between a saturated and an unsaturated zone. The primary goal was to understand whether the horizontal capillary forces are significant enough to cause a column that is predicted to be saturated based on the 1-D criterion to become unsaturated if it is next to an unsaturated column, and vice versa. This influence of horizontal redistribution of water was investigated in three steps, Experiment 2a, 2b and 2c, which consider different soil types (2a & 2b) and the importance of background moisture in the unsaturated column (2c). The experimental setups are described in section 3.5.2, and the respective results are given in section 3.6.2.

In Experiment 3, the effects of structural heterogeneity on the horizontal redistribution, and on the development of unsaturated zones underneath the riverbed, was tested with more complex models than in Experiment 2. This experiment addresses question 3 raised in the introduction. For this purpose, the 2-column systems of Experiment 2 were extended into checkerboard-type multi-column arrangements. The influence of different degrees of variance and spatial correlation were tested in Experiment 3a, and the combined influence of the spatial correlation and the horizontal domain scale were tested in Experiment 3b. The setups of the two experiments are described in section 3.5.3, the corresponding results follow in section 3.6.3.

3.3. Terminology and general considerations

There are several definitions, considerations, as well as, an underlying conceptual model that are common to all upcoming experiments. In the conceptual SW-GW system employed, a river is located on top of an aquifer, and the two compartments are separated by a riverbed. The river itself is conceptualized as ponding water of depth d. The riverbed is characterized by its height h_{rb}, by its hydraulic conductivity K_{rb}, and by its soil water retention characteristics. The aquifer below the riverbed either consists of the same or of different material than the riverbed. Above the water table, in the unsaturated zone, the degree of saturation lies between the residual saturation S_r [-] and full saturation. The state at which the degree of saturation is approximately equal to S_r is denoted as dry throughout the rest of this study. As a prerequisite for unsaturated zones to occur underneath a riverbed, losing conditions are necessary. For losing conditions to occur, the regional water table needs to be sufficiently deep, or in other words, the depth to water table (*DTW*) at a defined distance to the river needs to be sufficiently large. This conceptual river-aquifer model is based on the conceptual models used by Brunner et al. (2009a) and Irvine et al. (2012), among others.

Apart from the river-scale simulations, a range of column experiments consisting of a "riverbed layer" and an underlying "aquifer layer" were set up to explore the importance of horizontal redistribution. Recall that for a 1-D column, saturated conditions can develop in the aquifer layer if $K_{aq,critical} \geq K_{aq}$, and unsaturated conditions if $K_{aq,critical} < K_{aq}$ (Eq. (3.3)). In these column experiments, two columns were put adjacent to each other, and the hydraulic properties of one column were set to fulfill $K_{aq,critical} \geq K_{aq}$. This means that the 1-D criterion will predict saturated conditions for one column. For the second column, on the other hand, the hydraulic conductivity of the aquifer layer were set to fulfill $K_{aq,critical} < K_{aq}$. **Fig. 3.2** shows two possible configurations of such two-column systems. On the left-hand side ("River-River" scenario), a flux $q_{in} > 0$ will infiltrate into the unsaturated column. For the "River-Bank" scenario shown on the right hand side of **Fig 3.2**, vertical infiltration flux into the right column is zero ($q_{in} = 0$). This situation can develop where parts of the riverbed are completely clogged and no infiltration takes place, leading to a sharp interface between saturated and unsaturated conditions. Both the "River-River" and the "River-Bank" two-column scenario correspond to the simplest possible representation of partially saturated – partially unsaturated conditions as presented in **Fig. 3.1c**, and these scenarios are going to be used to study horizontal flow processes.

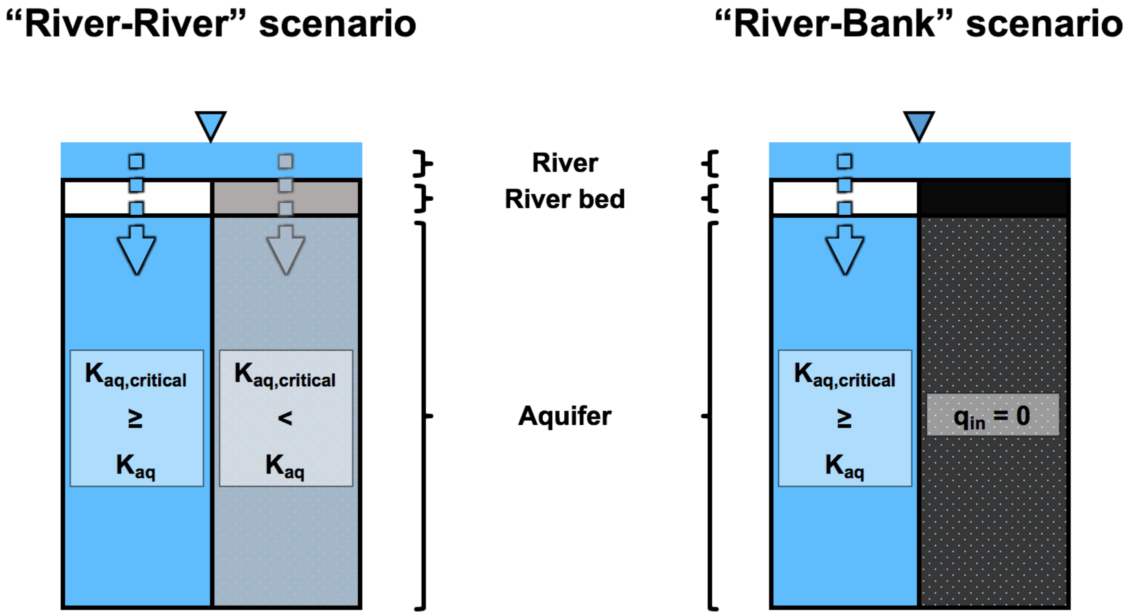


Figure 3.2: Two-column scenario overview: On the left hand side, in the "River-River", scenario two adjacent river-aquifer columns both receive infiltration from above. However, in a purely 1-D scenario, only for the left column the vertical infiltration flux through the riverbed would be sufficient to generate fully saturated conditions ($K_{aq,critical} \geq K_{aq}$), while the right column would not receive sufficient amounts of vertical infiltration due to clogging of the riverbed ($K_{aq,critical} < K_{aq}$). The scenario shown on the right hand side is denoted as the "River-Bank" scenario. This is the special case, in which only the left column receives infiltration from above and would be fully saturated according to the 1-D criterion. Adjacent to that saturated column an unsaturated column is located. In contrast to the "River-River" scenario, the unsaturated column in the "River-Bank" scenario does not receive any infiltration from above, either due to a completely clogged riverbed or due to zero ponded water ($K_{aq,critical} < K_{aq}$, $q_{in} = 0$).

According to Koltermann and Gorelick (1996) and de Marsily et al. (1998), the heterogeneity of K_{rb} and K_{aq} can be approximated with Gaussian statistics. There are more sophisticated, state-of-the-art methods to represent heterogeneity, for example multiple-point geostatistics (Mariethoz and Caers, 2014; Mariethoz et al., 2010). However, the pdfs of K_{rb} and K_{aq} were approximated by log-normal Gaussian distributions for the purpose of this study (see **Fig. 3.1**). The implications of this approach are discussed in section 3.7. Under a log-normal distribution, the three parameters that control the heterogeneity of K in a riverbed or aquifer domain are: the arithmetic mean of $\log_{10}K$ ($\mu_{\log K}$ [L/T]), the variance of $\log_{10}K$ ($\sigma^2_{\log K}$ [L^2/T^2]), and the spatial correlation length of K (τ_K [L]). In particular, the ratio of τ_K : domain length is of importance. For a τ_K : domain length that approaches 0 or 1 the medium becomes effectively homogeneous. In addition to using Gaussian representations of structural heterogeneity, to investigate the underlying physical and hydrological controls for the development of unsaturated zones, also simpler representations of heterogeneity were used. The conceptual reduction of complexity applied for the systematic investigation of effects of structural heterogeneity is illustrated in **Fig. 3.3**. Approximate dimensions of the experimental models used in this study are also indicated.

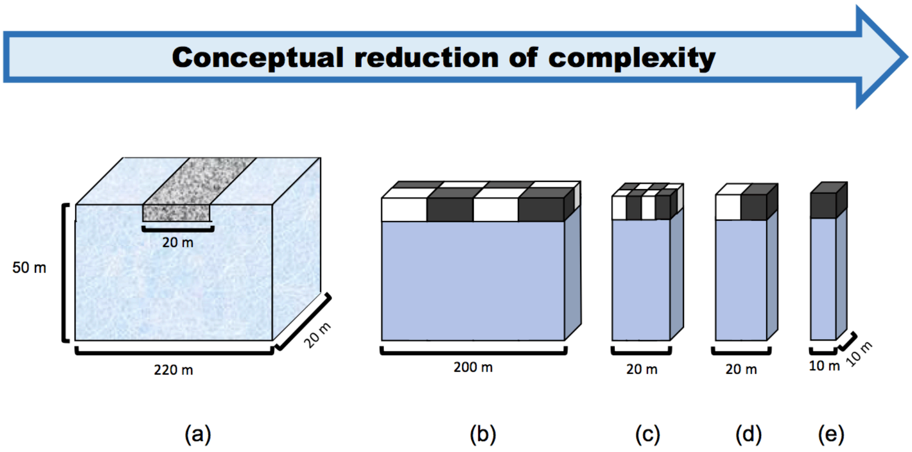


Figure 3.3: Conceptual illustration of the reduction of complexity of structural heterogeneity as done in this study. K_{rb} is colored in grey tones, and K_{aq} is colored in blue tones: the darker the tone, the lower K. The most complex model (a) represents log-normally distributed K_{rb}- and K_{aq}-fields. In the four models to the right, the reduced complexity of the structural heterogeneity is represented by checkerboard-like arrangements of binary K_{rb} zones positioned on top of homogeneous aquifers. In (b), the riverbed is conceptualized by multiple elongated rectangles, which represents an elongation of structures often encountered in natural rivers. In (c), the riverbed is represented by multiple squares. The most basic case of heterogeneity is represented by the two-column system (d). The final model (e) is a one-column system with only one rectangle for the riverbed, which represents a homogeneous K_{rb}. The same reduction of complexity could also be applied to K_{aq}. The indicated dimensions are approximations.

3.4. Numerical simulator

All numerical simulations in this study were carried out using the numerical model HydroGeoSphere (HGS, Therrien et al. (2010), Brunner and Simmons (2012)). The suitability of HGS for the simulation of river-aquifer interactions has been demonstrated in numerous studies (e.g. Banks et al. (2011), Frei et al. (2010), Frei and Fleckenstein (2014), Schilling et al. (2014)). HGS was chosen due to its ability to simulate all the relevant processes of SW-GW flow systems in a fully-integrated, physically-based way. Surface water processes, flow in the saturated and flow in the unsaturated zones are directly-coupled. Moreover, HGS allows the simulation of capillary driven flow in the subsurface, under consideration of heterogeneity. In HGS, unsaturated subsurface flow is based on the Richards equation, and the van Genuchten parametrization is used to define the dependency between k_r [-], pressure ψ [L], and saturation S_w [-]. Under unsaturated conditions, the hydraulic conductivity is a function of the degree of saturation. Fluxes are calculated according to Eq. (3.5), where k_r is the relative permeability, z [L] represents the elevation head and ψ the pressure head:

$$q = -K \cdot k_r(\psi)\, \nabla(\psi + z) \tag{3.5}$$

The relationship between the degree of saturation of the soil S_w and k_r can be expressed by the parametrization of van Genuchten (1980), which uses the two fitting parameters α [1/L] and β [-]:

$$S_w = S_{wr} + (1 - S_{wr}) \left[1 + |\alpha\psi|^{\beta}\right]^{-\upsilon} \quad \text{for } \psi < 0 \tag{3.6}$$

$$S_w = 1 \quad \text{for } \psi \geq 1 \tag{3.7}$$

$$k_r(\psi) = S_e(^{lp}) \left[1 - (1 - S_e^{1/\upsilon})^{\upsilon}\right]^2 \tag{3.8}$$

$$S_e = (S_w - S_{wr}) / (1 - S_{wr}) \tag{3.9}$$

S_{wr}: residual water saturation [–].
υ: given as $1–1/\beta$ with $\beta > 1$.
Se: effective saturation [–].
lp: pore-connectivity parameter (=2)

The fitting parameters α [1/L] and β are different for every soil type. Parameter values of four typical soil types are listed in **Tab. 3.1**.

Table 3.1: Parameters values after Carsel and Parrish (1988) for the 4 typical soil types. Typical hydraulic conductivities are not given, as the influence of K was investigated by systematically varying its magnitude.

Parameter	Loam	Sandy loam	Loamy sand	Sand
porosity (n)	0.43	0.41	0.41	0.43
α	3.6 m^{-1}	7.5 m^{-1}	12.4 m^{-1}	14.5 m^{-1}
β	1.56	1.84	2.28	2.68
S_{wr}	0.078	0.065	0.057	0.045

A detailed description of capillary forces and their potential influence on the movement of water in the unsaturated zone would go beyond the scope of this paper. An excellent overview has been given by Hillel (1998).

3.5. Experimental setups

The conceptual models, the purpose of the different experiments, and the parameters that were systematically varied in each experiment, are summarized in **Tab. 3.2**.

Table 2: Overview of goals and approaches for the different experiments.

Experiment	**Purpose**	**Modelling approach**	**Varied parameters**
Experiment 1	Comparison of the estimated unsaturated area using the stochastic 1-D approach to 3-D reference simulations	Fully-heterogeneous cross-sectional river-section in 3-D with random K_{rb}- and K_{aq}-fields	μ_{logKrb} & μ_{logKaq} σ^2_{logKrb} & σ^2_{logKaq} τ_{Krb} & τ_{Kaq}
Experiment 2a	Role and importance of horizontal redistribution at the interface of saturated and unsaturated areas	1-column and 2-column models	"River-River" vs. "River-Bank"
Experiment 2b	Influence of soil types on horizontal redistribution	"River-Bank" 2-column models	Soil types K_{aq} of the unsaturated column
Experiment 2c	Influence of infiltration on horizontal redistribution	"River-River" 2-column models	Soil types q_{in} of the unsaturated column K_{aq} of the unsaturated column
Experiment 3a	How does structural heterogeneity influence the development of unsaturated zones?	Checkerboard-style arrangement of squares of different K_{rb}, homogeneous K_{aq}	σ^2_{logKrb} τ_{Krb}
Experiment 3b	How do elongated heterogeneous structures influence the development of unsaturated zones?	Checkerboard-style arrangement of elongated rectangles of K_{rb}, homogeneous K_{aq}	τ_{Krb} in one direction

3.5.1. Experiment 1: Comparing the unsaturated area estimated with the stochastic 1-D approach to fully-heterogeneous 3-D reference simulations

In Experiment 1, estimations of the spatial extent of unsaturated areas made with a stochastic 1-D approach were compared to reference unsaturated areas simulated in 3-D for a fully-heterogeneous, cross-sectional river-aquifer model. Both the riverbed and the aquifer included heterogeneous structures, and the same stochastic properties of K were used for both approaches (1-D stochastic and 3-D reference flow simulations). The following steps were carried out to assess the reliability of the new stochastic 1-D approach in estimating the unsaturated area in a heterogeneous aquifer underneath a heterogeneous riverbed:

(1) The unsaturated area that can develop in fully-heterogeneous SW-GW systems was estimated with the stochastic 1-D equation according to the Monte Carlo procedure described in section 3.2.2. The unsaturated area was systematically estimated for 4x7 different degrees of heterogeneity of K_{rb} and K_{aq} (i.e., for different $P(K_{rb})$ and $P(K_{aq})$). **Tab. 3.3** provides a summary over the different K_{rb} - K_{aq} scenarios of this experiment. The different degrees of heterogeneity that were tested are in accordance with typical riverbed properties reported in literature (Calver, 2001; Rubin, 2003; Stewardson et al., 2016). Scenarios rb1 and rb2 are scenarios with a strong heterogeneity (high σ^2_{logKrb} and σ^2_{logKaq}) and scenarios rb3 and rb4 are scenarios with mild heterogeneity (moderate σ^2_{logKrb} and σ^2_{logKaq}). For every heterogeneous scenario, an ensemble of 44'000 pairs of K_{rb} and K_{aq} was sampled from the respective $P(K_{rb})$ and $P(K_{aq})$ to estimate the unsaturated area. The unsaturated area obtained in this way is denoted as *estimated unsaturated area* throughout the rest of this manuscript.

(2) For every degree of heterogeneity of K_{rb} and K_{aq} used in (1), 10 random K-fields of K_{rb} and K_{aq} were generated with geostatistical simulations to be used as input for the 3-D

reference flow simulations. The unconditional geostatistical simulations of the log-normally distributed random-fields of K_{rb} and K_{aq} in the 3-D reference model were based on a spherical variogram and generated with the gstat package (Pebesma, 2004) in R (R Core Team, 2015). For simplification purposes, the hydraulic conductivity was only varied in the two horizontal dimensions (i.e. for the vertical direction an infinite correlation length was used). An ensemble of 10 realizations of random $\log_{10}K_{rb}$- and random $\log_{10}K_{aq}$-fields was generated.

(3) The random K-fields for the riverbed and the aquifer were then incorporated into a 3-D cross-sectional river-aquifer model, and the development of the unsaturated area was simulated in steady state with HGS (the details of the model setup are provided below, in section 3.5.1). The unsaturated area underneath the riverbed was calculated by dividing the number of unsaturated nodes by the total number of nodes in the first layer immediately below the riverbed. The average unsaturated area of the 10 realizations of each scenario will be denoted as *simulated reference unsaturated area* throughout the rest of this manuscript.

(4) In a last step, the average mismatch between the *estimated unsaturated area* from step (1) and *simulated reference unsaturated area* from step (3) was calculated for every pairing of $P(K_{rb})$ and $P(K_{aq})$ tested.

Table 3.3: Parameters of the 4x7 different heterogeneous riverbed-aquifer scenarios tested. $K_{aq,critical}$ was calculated with hrb = 0.5 m and d = 0.5 m, corresponding to the values used in the model setup of the reference 3-D simulations. The correlation lengths correspond to the horizontal directions. For the vertical direction, an infinite correlation structure was chosen.

Scenario	μ_{logKrb} [m/d]	σ^2_{logKrb} [m^2/d^2]	τ_{Krb} [m]	μ_{logKaq} [m/d]	σ^2_{logKaq} [m^2/d^2]	τ_{Kaq} [m]	$\mu_{Kaq} : \mu_{Kaq,critical}$
rb1 - aq1	-1	1	4	0	1	4	5
rb1 - aq5	-1	1	4	1	1	4	50
rb1 - aq9	-1	1	4	3	1	4	5000
rb1 - aq13	-1	1	4	-1	1	4	0.5
rb1 - aq17	-1	1	4	2.3	1	4	1000
rb1 - aq21	-1	1	4	-2	1	4	0.05
rb1 - aq25	-1	1	4	-3	1	4	0.005
rb2 - aq2	-1	1	12	0	1	12	5
rb2 - aq6	-1	1	12	1	1	12	50
rb2 - aq10	-1	1	12	3	1	12	5000
rb2 - aq14	-1	1	12	-1	1	12	0.5
rb2 - aq18	-1	1	12	2.3	1	12	1000
rb2 - aq22	-1	1	12	-2	1	12	0.05
rb2 - aq26	-1	1	12	-3	1	12	0.005
rb3 - aq3	-1	0.1886	4	0	0.1886	4	5
rb3 - aq7	-1	0.1886	4	1	0.1886	4	50
rb3 - aq11	-1	0.1886	4	3	0.1886	4	5000
rb3 - aq15	-1	0.1886	4	-1	0.1886	4	0.5
rb3 - aq19	-1	0.1886	4	2.3	0.1886	4	1000
rb3 - aq23	-1	0.1886	4	-2	0.1886	4	0.05
rb3 - aq27	-1	0.1886	4	-3	0.1886	4	0.005
rb4 - aq4	-1	0.1886	12	0	0.1886	12	5
rb4 - aq8	-1	0.1886	12	1	0.1886	12	50
rb4 - aq12	-1	0.1886	12	3	0.1886	12	5000
rb4 - aq16	-1	0.1886	12	-1	0.1886	12	0.5
rb4 - aq20	-1	0.1886	12	2.3	0.1886	12	1000
rb4 - aq24	-1	0.1886	12	-2	0.1886	12	0.05
rb4 - aq28	-1	0.1886	12	-3	0.1886	12	0.005

3.5.1.1. Detailed model setup of Experiment 1

The reference model of a river-aquifer cross-section was set up in the following way: a 22 m wide river (a 20 m wide riverbed plus a 1 m wide riverbank on each side) was centered over a 200 m wide aquifer. The 20 m wide cross section had a vertical extent of 51 m. The river was conceptualized as a ponded water depth d of 0.5m, acting inside the riverbed at x = 90 - 110 m, y = 0 - 20 m and z = 50.5 m, leaving an effective riverbed height h_{rb} of 0.5 m vertical extent with banks of 1 m height and width at x = 89 - 90 m and x = 110 - 111 m. This geometric design was adapted from Irvine et al. (2012), and corresponds to the most complex model in Fig. 3. The riverbed had typical properties of a loam, and the aquifer typical properties of a sand. In the horizontal direction, the rectangular grid had a resolution of 1 m. In the vertical direction, the resolution varied from 0.03125 m immediately below the riverbed to a resolution of 5 m at the bottom of the aquifer. A regional water table elevation of 31 m was simulated as a fixed head boundary condition at x = 0 m and x = 200 m, creating a *DTW* of 19.5 m below the riverbed and thus a losing system.

3.5.2. Experiment 2: Influence of horizontal redistribution on the development of unsaturated zones

To explain the difference between the *estimated* and the *simulated reference unsaturated area*, the influence of horizontal redistribution between a saturated and an unsaturated column on the development of unsaturated zones was tested in three experiments (Experiment 2a, 2b and 2c). The simplest conceptual case of heterogeneity as shown in **Fig. 3.3d** was used in order to isolate horizontal redistribution from more complex structural effects. The simple two-column test case consisted of two joined columns - one saturated ($K_{aq,critical} \geq K_{aq}$) and one unsaturated ($K_{aq,critical} < K_{aq}$) - either forming a "River-River" case or a "River-Bank" case (as shown in **Fig. 3.2**).

3.5.2.1. Detailed model setup of Experiment 2

Geometrically, each individual column had a horizontal extent of 10 m x 10 m, and a vertical extent of 50 m. The resolution of the models was 0.25 m in the horizontal and 0.2 m in the vertical direction. A fixed hydraulic head boundary condition of 0 m was imposed on the bottom of each column, creating losing conditions. The top meter of each column represented a riverbed with a K_{rb} of 0.001 m/d. For the unsaturated column ($K_{aq,critical} < K_{aq}$), $K_{aq} = 100$ m/d was defined. For the saturated column ($K_{aq,critical} \geq K_{aq}$), K_{aq} was set to 0.001 m/d. For the saturated columns, according to Eq. (3.1), a ponded surface water depth d of 0.5 m, in the form of a constant hydraulic head boundary condition (BC), was imposed on top of the column. In the "River-River" scenarios, the same d was imposed on the unsaturated column, whereas in the "River-Bank" case, no ponded water depth was imposed ($d = 0$) on the unsaturated column. In the base case, both the aquifer as well as the riverbed were conceptualized with properties of a typical loam (see **Tab. 3.1**).

3.5.2.2. Experiment 2a: Desaturation or resaturation?

The aim of Experiment 2a was to test if horizontal redistribution through capillary forces can be of such a degree that sections under clogged areas can become re-saturated, and vice versa if fully saturated zones under unclogged areas could become desaturated. The columns of different degrees of saturation were simulated in steady state as "River-River" or "River-Bank" two-column systems.

3.5.2.3. Experiment 2b: Dependence of horizontal redistribution on the soil type

In Experiment 2b, the effects of different soil types were systematically assessed using the "River-Bank" scenario. This scenario was chosen to isolate the effects of the soil type from potential effects of background moisture in the unsaturated column. The 4 typical soil types that were tested are listed in **Tab. 3.1**. To assess the horizontal redistribution potential of the different soil types, K_{aq} of the unsaturated column was systematically varied: For each soil type tested, the K_{aq} : $K_{aq,critical}$ ratio, was varied between 50000 and 0.005 by gradually decreasing K_{aq} by a factor of 10 from 100 m/d to $1 \cdot 10^{-5}$ m/d, while K_{rb} was kept constant at 0.001 m/d.

3.5.2.4. Experiment 2c: Dependence of horizontal redistribution on the moisture content in the unsaturated zone

How much the horizontal redistribution is influenced by the background moisture content in the unsaturated column was investigated in Experiment 2c. In addition to the variation of K_{aq} of the unsaturated column, as done in the previous experiment that used the "River-Bank" case, in this experiment q_{in} into the unsaturated column was included and systematically varied by using the "River-River" scenario. Different q_{in} into the unsaturated column, which generate different background moisture contents, were simulated by systematically varying K_{rb} of the unsaturated column between 100 and $1 \cdot 10^{-4}$ m/d (encompassing the natural range of K_{rb} (Calver, 2001; Chen, 2011; Lu et al., 2012; Song et al., 2007)). For the current model setup, this resulted in q_{in} between 200 and $2 \cdot 10^{-6}$ m/d. Moreover, for every tested q_{in}, also K_{aq} was systematically varied such that the critical ratio K_{aq} : $K_{aq,critical}$ varied between 0.1 and 5.

3.5.3. Experiment 3: Influence of heterogeneity on the development of unsaturated zones

After the investigation of the horizontal redistribution potential in simple two-column systems, in a next step, a more complex representation of heterogeneity was tested by using a checkerboard-style arrangement of river-aquifer columns with different K_{rb}'s. In the simpler case, the checkerboard of rectangles with different K_{rb} values consists of squares with equal side lengths in the horizontal directions (see **Fig. 3.3c**). In the more complex case, the checkerboard consists of rectangles that are elongated in one horizontal direction (see **Fig. 3.3b**). This more realistic configuration, an elongation of structures in the direction of flow, is observable on the surface of low-gradient systems, for example in the form of tear-shaped islands in wetlands like the Greater Everglades Ecosystem (e.g. McVoy et al. (2011), Mitsch and Gosselink (2007), Lodge (2005), and Schilling et al. (2013)). If the topographic gradient is low enough, similar surface structures can be observed in the form of gravel bank structures in alluvial systems (e.g. Rosgen (1994), Huggenberger et al. (1998)). The influence of checkerboard-style arrangement of square columns on horizontal redistribution was tested in Experiment 3a; the more complex case with elongated rectangles is tested in Experiment 3b. To understand and visualize these effects more easily, only the riverbed was made heterogeneous, and the aquifer was simulated as a homogeneous medium. This allowed isolating the effects of structural heterogeneity on horizontal redistribution to one layer, and allowed visualizing the effects in a layer underneath a heterogeneous layer.

3.5.3.1. Detailed model setup of Experiment 3

Geometrically, the checkerboard-style arranged soil columns had the following properties: the heterogeneous riverbed with a 0.5 m vertical extent (i.e., h_{rb} = 0.5m) was conceptualized as an arrangement of squares (Experiment 3a) or elongated rectangles (Experiment 3b) with different K_{rb} values positioned on top of a homogeneous aquifer of 49.5 m thickness. The riverbed had typical properties of a loam. The aquifer, had typical properties of a sand and a K_{aq} of 1 m/d.

As in the previous experiments, a ponded water depth *d* of 0.5 m was imposed on the top of the riverbed. A constant hydraulic head of 0 m was fixed at the bottom of the column, creating a purely losing system with a *DTW* of 50 m. The complete checkerboard-style arrangement of individual columns had a horizontal extent of 20 x 10 m in Experiment 3a, and was elongated to 2000 x 10 m for Experiment 3b. The horizontal resolution of the rectangular grid was 0.25 m and the vertical resolution 0.5 m; this is slightly less than in Experiment 2, but a test of grid convergence showed that it is sufficient for this experiment.

3.5.3.2. Experiment 3a: Dependence of horizontal redistribution on σ2logK and τK

In Experiment 3a, the checkerboard-style arrangement of squares with different K_{rb} values, the arithmetic mean of K_{rb} of the whole riverbed (μ_{Krb}) was kept constant at 0.1 m/d, while the K_{rb} of the individual squares was varied in 0.02 m/d intervals between 0.0 and 0.2 m/d to simulate different σ^2_{logK}. To test different τ_{Krb} the side lengths of the squares were varied. Three different side lengths were tested: 1 m, 2 m and 5 m. This resulted in 6x3 different models. The tested σ^2_{logK} and τ_{Krb} are summarized in **Tab. 3.4**. Both steady state and transient simulations were carried out. The transient simulations started from a *dry* state, which allowed visualization of the propagation of the wetting front into the aquifer.

Table 3.4: Overview of the different σ^2_{logK}, conceptualized as binary combinations of high K_{rb} and low K_{rb} ($K_{rb,high}$: $K_{rb,low}$), and of the different τ_{Krb}, conceptualized as differing side lengths of the square K_{rb}-fields that were used in Experiment 3a. The different σ^2_{logK} and τ_{Krb} were combined to 3x6 different models.

$K_{rb,high}$: $K_{rb,low}$	τ_{Krb}
0.00 : 0.20 m/d	1 m
0.02 : 0.18 m/d	2 m
0.04 : 0.16 m/d	5 m
0.06 : 0.14 m/d	
0.08 : 0.12 m/d	
0.10 : 0.10 m/d	

3.5.3.3. Experiment 3b: Dependence of horizontal redistribution on τK and the horizontal domain scale

To understand the influence of elongated structures in anisotropic river systems, in Experiment 3b anisotropic correlation structures were implemented by applying a stretch factor of 100 to one side of the domain of the checkerboard-style model. The factor was applied to both the x-domain scale, as well as, to τ_{Krb} in x-direction. This resulted in a horizontal model domain of 2000 x 10 m, and anisotropic horizontal τ_{Krb}'s of (x=100 m, y=1 m), (x=200 m, y=2 m) and (x=500 m, y=5 m). Only the strongest variance case tested in Experiment 3a was used for Experiment 3b (i.e., $K_{rb,high}$ = 0.2 m/d, $K_{rb,low}$ = 0 m/d), and only steady state simulations were carried out.

3.6. Results

3.6.1. Results of Experiment 1: Comparing the unsaturated area estimated with the stochastic 1-D approach to fully-heterogeneous 3-D reference simulations

An overview over the *estimated* and *simulated reference unsaturated areas*, the ensemble statistics of the reference models, as well as the mismatch between the estimation and the reference, is provided in **Tab. 3.5**. The underlying stochastic properties of the different pairings were given in **Tab. 3.3**.

The results presented in **Tab. 3.5** clearly show that the *estimated unsaturated area* obtained using the stochastic 1-D approach based on Eq. (3.4) is always either larger or equal compared to the *simulated reference unsaturated area* obtained with 3-D reference random K-field models. By comparing the *estimated* as well as *simulated reference unsaturated area* to the corresponding ratio of μ_{Kaq} : $\mu_{Kaq,critical}$, it becomes evident that this ratio is a first order control on the development of unsaturated areas, as well as on the mismatch between the *estimated* and *simulation reference unsaturated area*: The closer the ratio is to 1, the more likely partially saturated-partially unsaturated conditions can develop, and the more likely the *estimated unsaturated area* and the *simulated reference unsaturated area* differ. On the other hand, the further the ratio is away from 1, the more likely it is that a fully saturated or a fully unsaturated state develops directly below the riverbed, and the more accurate the predictions by the stochastic 1-D approach. If $\mu_{Kaq} : \mu_{Kaq,critical} >> 1$, fully unsaturated conditions, and if $\mu_{Kaq} : \mu_{Kaq,critical} << 1$, fully saturated conditions become very likely. This can also be directly observed in the magnitude of the *estimated unsaturated area*: The more a very large (≥ 95%) or a very small (≤ 5%) unsaturated area is approached, the more the *estimated unsaturated area* corresponds to the *simulated reference unsaturated area*. Comparing scenarios rb1 and rb2 to scenarios rb3 and rb4 reveals that the higher the variance of K, the more likely it is that partially saturated – partially unsaturated conditions develop.

In summary, the unsaturated area appears to be overestimated with the stochastic 1-D approach, but never underestimated. Moreover, these results indicate that the mismatch between the *estimated unsaturated area* and the *simulated reference unsaturated area* is a result of limiting 1-D assumptions in the stochastic 1-D approach. Neglecting horizontal flow processes seems to result in a significant mismatch if the stochastic 1-D approach predicts partially saturated – partially unsaturated conditions. In these experiments, if the estimated unsaturated areas with the stochastic 1-D approach were up to 45%, full re-saturation was modelled in the 3-D studies. Thus, horizontal processes that are not considered in the stochastic 1-D approach, appear to be strong enough to re-saturate areas of up to 45%, and make an estimated partially saturated – partially unsaturated system a fully saturated system. For this system, a partially saturated-partially unsaturated system was always observed if the unsaturated area was estimated to be between 45% and 95%. Resaturation of up to 45% of the area was only observed for the strongest degree of heterogeneity. In systems with milder degrees of heterogeneity, this mark was lower and approximately 30%. Heterogeneity appears to be a controlling factor for the potential of horizontal redistribution.

Table 3.5: Statistics of the stochastic 1-D *estimated area* and the 3-D *simulated reference unsaturated area* for the different riverbed-aquifer scenarios given in Tab. 3.3.

Scenario	μ_{Kaq} : $\mu_{Kaq,critical}$	*Estimated area* (1D stochastic)	Mean *simulated reference area*	Minimum *simulated reference area*	Maximum *simulated reference area*	Mean mismatch (*Estimated - Simulated*)
rb1 - aq1	5	69.0%	24.9%	8.0%	46.3%	+44.1%
rb1 - aq5	50	88.0%	81.3%	71.8%	88.5%	+6.7%
rb1 - aq9	5000	100.0%	99.7%	99.1%	100.0%	+0.3%
rb1 - aq13	0.5	42.0%	0.5%	0.0%	5.0%	+41.5%
rb1 - aq17	1000	98.0%	97.6%	93.4%	99.3%	+0.4%
rb1 - aq21	0.05	18.0%	0.0%	0.0%	0.3%	+18.0%
rb1 - aq25	0.005	5.0%	0.0%	0.0%	0.0%	+5.0%
rb2 - aq2	5	69.0%	44.5%	4.8%	77.9%	+24.5%
rb2 - aq6	50	88.0%	76.8%	54.0%	100.0%	+11.2%
rb2 - aq10	5000	100.0%	99.8%	98.9%	100.0%	+0.2%
rb2 - aq14	0.5	42.0%	5.0%	0.0%	29.9%	+37.0%
rb2 - aq18	1000	98.0%	97.1%	89.0%	100.0%	+0.9%
rb2 - aq22	0.05	18.0%	0.0%	0.0%	0.5%	+18.0%
rb2 - aq26	0.005	5.0%	0.0%	0.0%	0.1%	+5.0%
rb3 - aq3	5	87.0%	56.1%	49.7%	67.3%	+30.9%
rb3 - aq7	50	100.0%	97.9%	95.1%	100.0%	+2.1%
rb3 - aq11	5000	100.0%	100.0%	100.0%	100.0%	+0.0%
rb3 - aq15	0.5	31.0%	0.3%	0.3%	3.2%	+30.7%
rb3 - aq19	1000	100.0%	100.0%	100.0%	100.0%	+0.0%
rb3 - aq23	0.05	2.0%	0.0%	0.0%	0.0%	+2.0%
rb3 - aq27	0.005	0.0%	0.0%	0.0%	0.0%	+0.0%
rb4 - aq4	5	87.0%	51.0%	30.7%	74.2%	+36.0%
rb4 - aq8	50	100.0%	97.2%	91.6%	100.0%	+2.8%
rb4 - aq12	5000	100.0%	100.0%	100.0%	100.0%	+0.0%
rb4 - aq16	0.5	31.0%	0.1%	0.0%	4.0%	+30.9%
rb4 - aq20	1000	100.0%	100.0%	100.0%	100.0%	+0.0%
rb4 - aq24	0.05	2.0%	0.0%	0.0%	0.0%	+2.0%
rb4 - aq28	0.005	0.0%	0.0%	0.0%	0.0%	+0.0%

3.6.2. The influence of soil type and background saturation on the development of unsaturated zones

3.6.2.1. Results of Experiment 2a: desaturation or resaturation?

The simulation results of Experiment 2a are illustrated in **Fig. 3.4**. First, the results of the one-column experiments are in agreement with Eq. (3.1). On the other hand, the results also clearly indicate that the horizontal redistribution of water can have a significant effect on an initially unsaturated column: the saturation increases once the initially unsaturated column is connected to a saturated column. This is particularly evident in a "River-Bank" setting (**Fig. 3.4b**), where the unsaturated column would be *dry* (q_{in} = 0) if not connected to a saturated column.

In the two-column simulations, a substantial resaturation of the unsaturated column could thus be observed, even directly underneath the riverbed. A substantial desaturating effect of the saturated columns, on the other hand, could not be observed. Only directly at the interface between the two columns deep below the riverbed (more than 20 m below the top of the model domain), a desaturation could be observed. For the SW-GW interactions investigated in this study, however, such a small desaturating effect, far below the riverbed, is not of concern, as its influence on exchange fluxes between the SW and the GW is negligible. For practically relevant SW-GW interactions, therefore, if a column is saturated according to Eq. (3.1), significant desaturation through horizontal redistribution is unlikely to occur even when this column is connected to a *dry* column. Whether a desaturation could occur in a more complex system than a two-column system is investigated in the Experiments 3. However, the findings of Experiment 2a are in agreement with the results from a study by Xie et al. (2014), who found that, even though an unsaturated zone can occur under non-clogging conditions, it could only be observed in extreme scenarios: for very narrow or almost completely dry rivers, and at very high *DTWs*. In Xie et al. (2014), the horizontal desaturation effect already disappeared once the ponded water depth exceeded a critical depth of only 0.001 m in a river of 6 m width. The findings of Experiment 2a shed light on the observations made previously in Experiment 1: if the state of saturation for realistic SW-GW systems is estimated with the new stochastic 1-D approach (Eq. 3.4), an overestimation of the unsaturated area is likely, and an underestimation could not be observed.

"River-River" scenario	**"River-Bank" scenario**

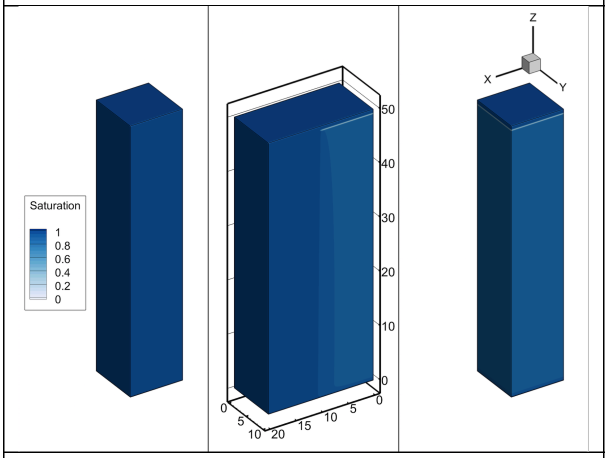
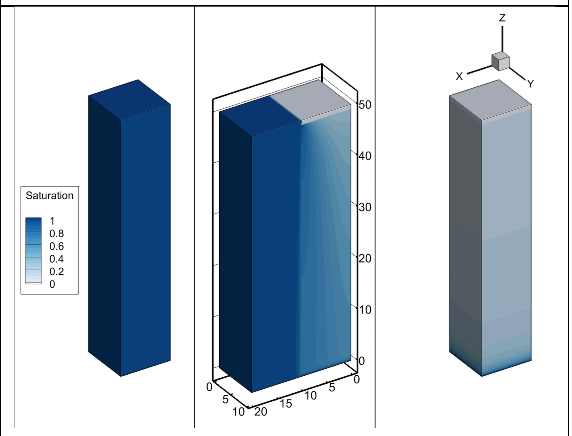

	

Figure 3.4a: "River-River" case for a loamy soil. The individual "River" columns show that a loam can retain a significant amount of moisture. Due to this high background moisture content, connection of a saturated "River" loam column to an unsaturated "River" loam column does not appear to increase the saturation in the unsaturated column far away from the interface between the columns.	**Figure 3.4b:** "River-Bank" case for a loamy soil. If the aquifer that exhibits unsaturated conditions is a *dry* "Bank" case, and therefore doesn't have any vertical water inflow, connection to a saturated "River" column significantly increases the degree of saturation deep into the, initially *dry,* unsaturated column.

3.6.2.2. *Results of Experiment 2b: the influence of the soil type*

The steady state simulation results of Experiment 2b, illustrated in **Fig. 3.5**, show that the loamier the aquifer, the greater its ability to horizontally transport water through the unsaturated zone. However, the results also reveal that there is no significant difference between the soil types if one only considers the advancement of the saturated wetting front into a *dry* column where $q_{in} = 0$. While some water propagated further into loamier soil, the fully re-saturating front propagated approximately the same distance into the unsaturated column for all different soil types. Moreover, the results show that as long as the ratio of $K_{aq} : K_{aq,critical}$ is larger than 0.5, the horizontal redistribution does not alter saturation in the unsaturated zone significantly immediately below the riverbed: The distance (Δx) that water was able to propagate into the unsaturated column in the layer immediately underneath the riverbed was ≈ 0 m. Once $K_{aq} : K_{aq,critical}$ reaches 0.05, horizontal redistribution can significantly alter the unsaturated area below the riverbed ($\Delta x \approx 5$ m). At a K_{aq} to $K_{aq,critical}$ ratio of 0.005, the horizontal redistribution becomes so strong that almost the entire unsaturated column changes to fully saturated conditions (i.e. $\Delta x \approx 20$ m). Thus, if K_{aq} is considerably smaller than the critical value $K_{aq,critical}$, horizontal redistribution can be an important process. Once K_{aq} gets closer to the critical value $K_{aq,critical}$, however, the distance that water can move laterally into the unsaturated column due to capillary forces diminishes rapidly. Nonetheless, depending on the scale and dimensions of the SW-GW system in question, a Δx of 1 or 2 m of resaturation could already be a considerable extent.

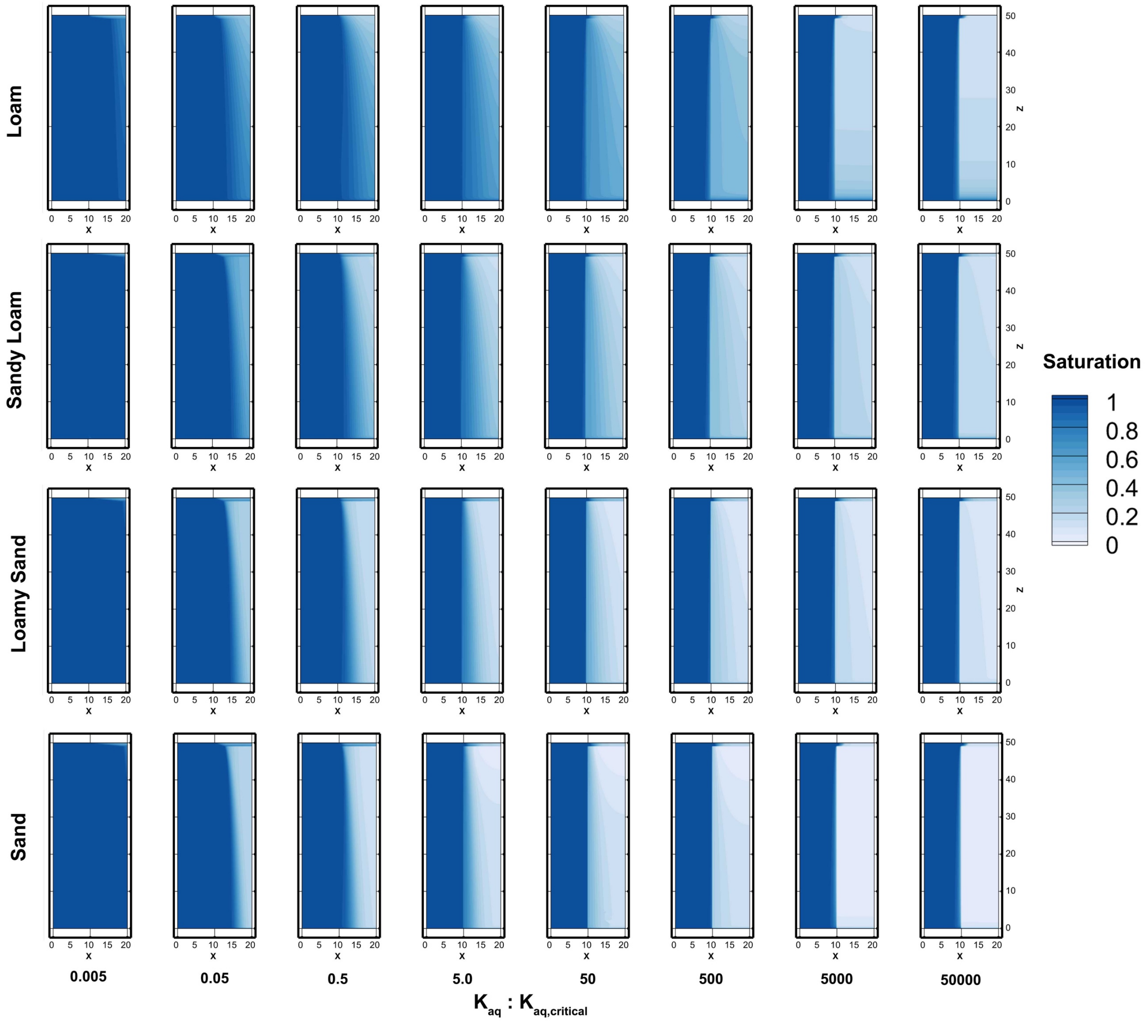


Figure 3.5: Cross-sectional depiction of the steady state saturation for the 4 different soil types (rows) and the 8 different K_{aq} (columns) simulated in Experiment 2b.

3.6.2.3. Results of Experiment 2c: the influence of the moisture content in the unsaturated zone

In Experiment 2c, multiple q_{in} and K_{aq} : $K_{aq,critical}$ were tested for a "River-River" scenario ($q_{in} < K_{aq}$ for the unsaturated column). As opposed to Experiment 2b (see **Fig. 3.5**), where the "River-Bank" scenario was used, the "River-River" scenario creates a constant infiltration of water into the unsaturated zone. While this infiltration cannot fully saturate the unsaturated column ($q_{in} < K_{aq}$), the infiltration nonetheless creates a background moisture content in the unsaturated column. The simulations of Experiment 2c with variable infiltration and K_{aq} : $K_{aq,critical}$ revealed that as soon as there is some infiltration through the clogging layer into the unsaturated column, i.e., as soon as there is some background moisture and the column isn't completely *dry*, the ratio of K_{aq} : $K_{aq,critical}$, at which the system can become *fully saturated* through horizontal redistribution, is shifted. For a "River-Bank" system, where $q_{in} = 0$ prevents any background moisture to build up in the unsaturated column, the critical point, at which the entire two-column system can become *fully saturated* through horizontal redistribution, is K_{aq} : $K_{aq,critical} \approx 0.005$. For a "River-River" system, where also the unsaturated column receives vertical infiltration and background moisture can build up, but not enough to create full saturation through this process alone, the critical point at which full saturation of the entire two-column system can occur as a result of horizontal redistribution is shifted to $K_{aq} : K_{aq,critical} \approx 1.45$ for a sand, and to $K_{aq} : K_{aq,critical} \approx 3.75$ for a loam.

The amount of already present background moisture in the unsaturated column is therefore an important control for the development of unsaturated zones. The importance of this moisture content varies slightly for the different soil types. The results, however, revealed that these ratios are constant for all tested q_{in}, that is, the ratio $K_{aq} : K_{aq,critical}$ for each soil type is independent of the magnitude of q_{in} of the unsaturated column. Moreover, as the ratio of $K_{aq} : K_{aq,critical}$ is a function of K_{rb}, h_{rb}, d, and K_{aq}, this shift can entirely be attributed to the background moisture content and the soil parameters (i.e., porosity, α, β and S_{wr}).

3.6.3. The influence of geologic structures on the development of unsaturated zones

3.6.3.1. Results of Experiment 3a: the influence of σ^2_{logK} and τ_K

The transient simulation results of Experiment 3a of the propagation of the wetting front through a checkerboard-like riverbed into a homogeneous aquifer are illustrated in **Fig. 3.6**. Illustrations represent the wetting front after 6 days of infiltration into a *dry* aquifer. The steady state solutions are presented in **Fig. 3.7**. A visual comparison of **Fig. 3.6** and **3.7** show that the saturated and the unsaturated areas are more distinct for a transient state after 6 days of transient simulation than for the steady state equilibrium. The 1-D vertical flow that is taken into account by Eq. (3.1) is thus a faster process than the horizontal redistribution of water, that is, the gravitational force of water downwards is much stronger than the diffusive, capillary forces that induce horizontal redistribution. The unsaturated area is much smaller once a system has reached steady state. For the homogeneous case ($K_{rb,high} = K_{rb,low} = 0.1$ m/d), shown in the first columns of **Fig. 3.6** and **Fig. 3.7**, all models exhibit the same degree of saturation below the riverbed. On the other hand, for the riverbeds with the highest σ^2_{logK} ($K_{rb,low} = 0$ m/d, $K_{rb,high} = 0.2$ m/d), stark contrasts in the saturation profile can be observed: As shown in the last columns of **Fig. 3.6** and **Fig. 3.7**, there is a significant difference in the degree of saturation below the riverbed of the highest σ^2_{logK} scenario (1) compared to the homogeneous case, and (2) between the models of different τ_K. The larger τ_K, the stronger the distinction between saturated and unsaturated zones. For a small τ_K, unsaturated areas are much less pronounced due to the smaller travel distances and therefore facilitated horizontal re-saturation. Even for the largest σ^2_{logK} tested, where one half of the riverbed is completely impermeable, the shortest τ_K of 1 m was small enough to enable significant re-saturation of unsaturated areas. For $\tau_K = 5$ m and the highest σ^2_{logK}, the horizontal redistribution of water only has a minimal effect on the unsaturated area after $t = 6$ days, and does not significantly change the extent of the unsaturated area immediately below the riverbed even in steady state. In accordance with the findings of Experiment 2b, where full resaturation through horizontal redistribution 1-2 meters into the unsaturated column occurred rapidly, these results underline that horizontal redistribution is a crucial effect for the development of unsaturated zones on scales of a few meters. However, rather than being controlled by σ^2_{logK}, the potential of horizontal redistribution from a saturated column into adjacent unsaturated columns is controlled by the background moisture of the unsaturated columns and the distance that the water needs to travel. The background moisture is controlled by the absolute value of K_{aq} in relation to q_{in} (as shown in Experiment 2a, 2b and 2c), and the distance that water needs to travel is controlled by τ_K.

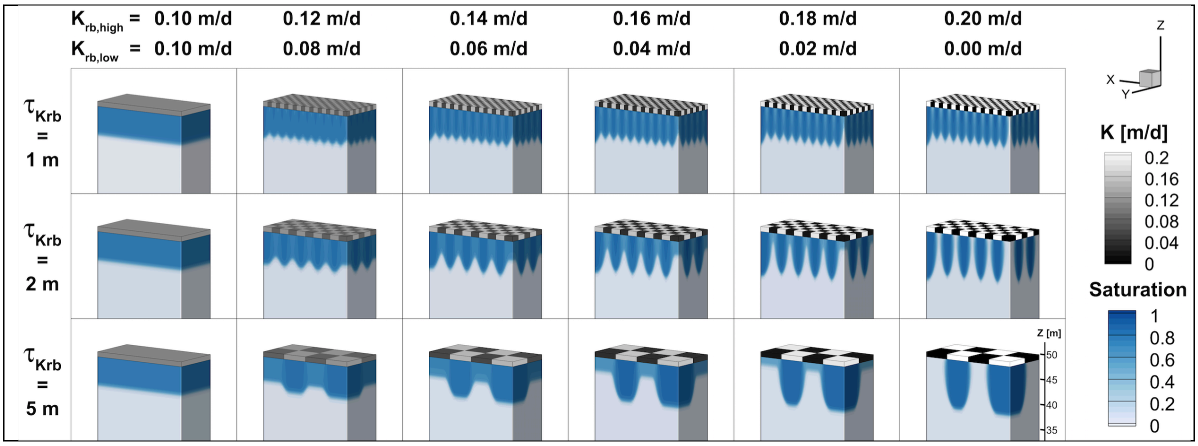


Figure 6: Propagation of a wetting front into a dry aquifer through a riverbed of checkerboard-style heterogeneous structures after 6 days of transient simulation. For better presentation, only the top, unsaturated 15 m of the columns are shown. K_{rb} is displayed in grey scales, the saturation in the aquifer is displayed in blue tones.

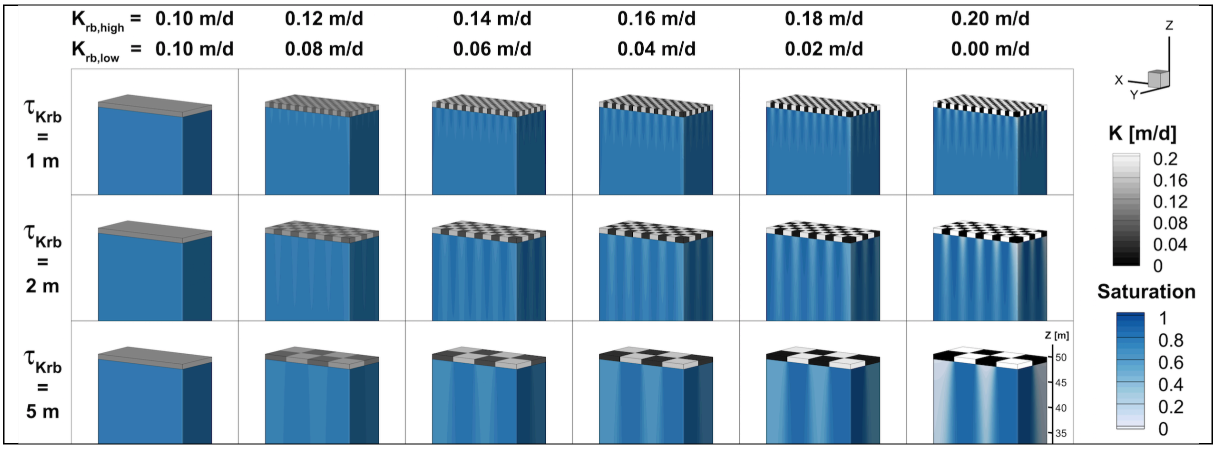


Figure 3.7: Steady state solutions of the checkerboard-style column models. For better presentation only the top 15 m of the columns are shown. K_{rb} is displayed in grey scales, the saturation in the aquifer is displayed in blue tones. The horizontal redistribution of water significantly increased the saturation of the columns that do not receive sufficient amounts of water from direct vertical infiltration in order for them to become saturated. This effect is more pronounced for the steady state solution than for the transient simulations in **Figure 3.6**.

3.6.3.2. Results of Experiment 3b: the influence of τ_K and the horizontal domain scale

The results of the stretched river-like model are illustrated in **Fig. 3.8**. For riverbed structures that are elongated in the direction of SW flow, the simulated saturation profiles illustrated in **Fig. 3.8** show that the horizontal redistribution of water perpendicular to the river flow direction can be an important mechanism. Along the river in the direction of SW flow the horizontal redistribution potential becomes smaller in relative terms due to the interplay between the magnitude of K_{aq}, τ_K, as well as the horizontal domain scale: If structures are elongated in the direction of river flow, the distance that water must travel in the subsurface in order to re-saturate potentially unsaturated areas becomes larger. For this effect to become important, a K_{aq} that is relatively small compared to the given travel distances is necessary. The results of Experiment 3b show that this effect becomes even more important if elongated structures are present, which are quite likely to occur in river systems.

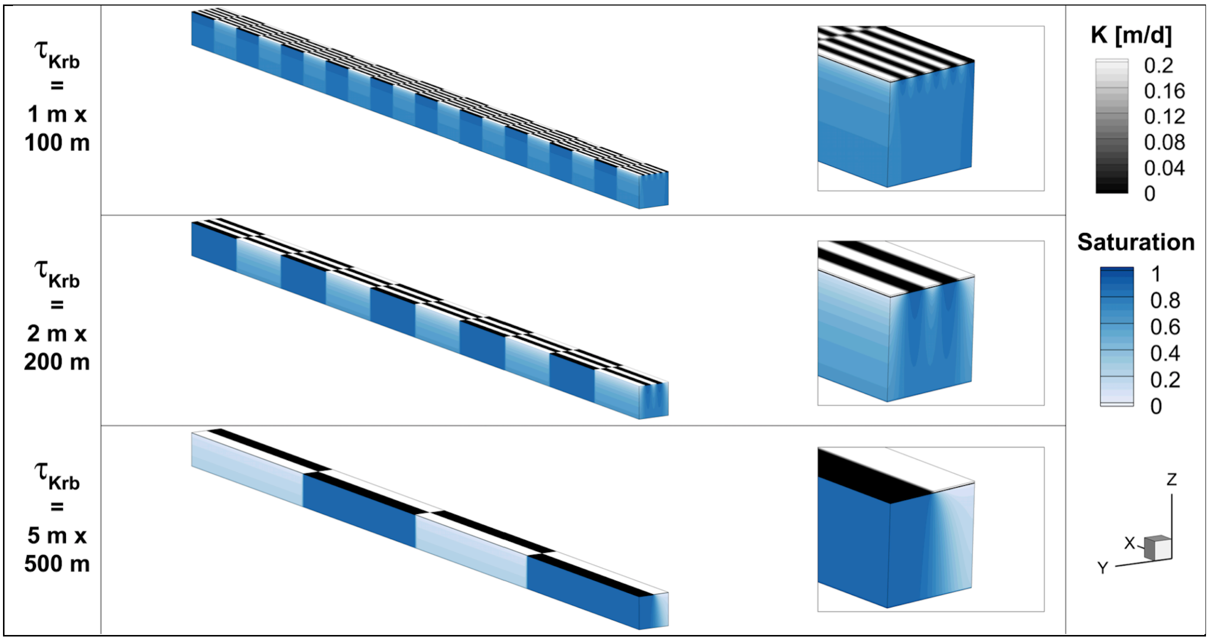


Figure 3.8: Steady state saturation profiles for the elongated river-aquifer models. With this experiment, the influence of the horizontal domain scale on the horizontal redistribution was tested. Horizontal redistribution of water can still significantly influence the saturation in unsaturated zones perpendicular to the river, but for such an elongated riverbed structures, horizontal redistribution along the river is virtually eliminated.

If τ_K perpendicular to the river is small compared to the width of the river, the effect of the increased longitudinal distance can be compensated by horizontal redistribution perpendicular to the river (see the top row of **Fig. 3.8**). However, when τ_K perpendicular to the river approaches the width of the river, the overall horizontal redistribution is significantly reduced, and re-saturation of unsaturated areas becomes much less likely (see the bottom row of **Fig 3.8**). A primary variable to know if one wants to estimate the horizontal redistribution potential is, therefore, τ_K perpendicular to the river. The simulation results moreover confirm that for real-world SW-GW systems it is not only important to know the heterogeneity of K, but also to understand the relation between the magnitude of K and the horizontal domain scale.

3.7. Discussion

The overarching goal of this paper was to provide a comprehensive understanding on the geological and hydrogeological controls of the development of unsaturated zones at the interface of surface water and groundwater. Previous studies have either only considered riverbed heterogeneity (Irvine et al., 2012) or aquifer heterogeneity (Fleckenstein et al., 2006), or focused on systems without clogging riverbed (Pryshlak et al., 2015; Xie et al., 2014). Our study is the first to fully integrate all aspects: We jointly considered heterogeneity of the riverbed as well as the aquifer, and the importance of capillary forces in affecting the degree of saturation through horizontal redistribution in adjacent clogged- and non-clogged areas. We tested the relative importance of horizontal redistribution for different soil types and variable structural heterogeneity. The experiments revealed that the ratio of K_{aq} : $K_{aq,critical}$ and the variance of K_{rb} are first order controls of horizontal redistribution, and to a lesser degree the soil type and the background moisture content in the unsaturated column.

Through the stochastic 1-D approach, our study can provide guidance on the upper limit of the extent of unsaturated zones under losing conditions. This approach (described in section 3.2.2) allows estimating the upper bound of the unsaturated area that can develop in a fully heterogeneous riverbed-aquifer system. Apart from informing the important choice of the numerical model, such an analysis can also be used to provide guidance on how to simplify heterogeneity of the riverbed. Irvine et al. (2012) showed that if the estimated unsaturated area approaches 0% or 100%, the heterogeneity of the riverbed can be replaced by a homogeneous equivalent obtained through calibration, without an increase of uncertainty for predictions of SW-GW exchange fluxes.

The proposed stochastic 1-D approach is conceptually very easy to use. It requires knowledge of the probability distribution of hydraulic conductivity of the riverbed and of the underlying aquifer. In practice, the structural heterogeneity of the riverbed and the aquifer are still difficult to assess. However, there are techniques that allow estimating the heterogeneous nature of the riverbed, for example by fractal analysis of riverbed photographs (Butler et al., 2001). Non-intrusive, geophysical techniques allow assessing the heterogeneous nature of the subsurface better than ever before (an overview over these techniques is provided by Rubin and Hubbard (2005)). Nevertheless, there is an urgent need to develop more efficient approaches to generate highly realistic geological models of riverbeds and their underlying aquifer.

Our chosen approach to characterize heterogeneity was based on Gaussian statistics. Through Gaussian statistics, structural features such as connectivity cannot be retained. Other approaches such as Multiple Point Statistics (Mariethoz and Caers, 2014) could in principle be employed to retain such structural features. However, the conclusions of this study would not be affected: The stochastic 1-D analysis will provide identical results, independent of the spatial arrangement of the different hydraulic conductivities. Moreover, a study by Tang et al. (2015) found that riverbed properties which exhibit non-Gaussian statistics and connectivity can be approximated with Gaussian statistics, without deteriorating the characterization of river-aquifer exchange fluxes. The estimation of the upper limit of the spatial extent of unsaturated zones is therefore not affected by using simplified Gaussian approaches instead of more advanced techniques. What will certainly change, however, is the relative importance of horizontal redistribution. Horizontal redistribution is driven through differences in saturation (and thus suction). If in a riverbed, for example, the unclogged areas are forming large connected zones next to clogged areas, horizontal redistribution would be less important compared to a river where many small unclogged areas are surrounded by clogged zones.

3.8. Conclusions

This paper was motivated by three fundamental questions. Above all, the goal was to explore to what extent knowledge of the geological structure of the riverbed and the underlying aquifer allows predicting the potential for unsaturated zones in and below the riverbed to develop. It is the first study to jointly consider heterogeneity of the riverbed and the underlying aquifer. Our study shows that the upper limit of the unsaturated zones under losing conditions can be predicted through a simple stochastic analysis. This has implications for SW-GW modelling: before a numerical SW-GW model is set up, the stochastic 1-D approach presented in this study should be applied prior to making decisions on the processes and structures that are going to be included in the numerical simulations. However, the bottleneck for a direct application of this approach is the often limited knowledge about the structure and associated hydraulic properties of the riverbed and the underlying aquifer.

The second question addressed in this paper explored the role of horizontal redistribution of water at the interface of saturated and unsaturated zones in the riverbed. Apart from explaining why the stochastic 1-D approach provides an upper bound of the unsaturated areas, this analysis provided new knowledge that helps to understand unsaturated zone processes at the interface of surface water and groundwater. In the final question raised in this paper, the relative importance of horizontal redistribution was explored for different types of structural heterogeneity. This final step allowed integrating not only heterogeneity of the riverbed and the aquifer, but also their spatial configurations. In addition to the upper limit of unsaturated areas that can be obtained through the stochastic 1-D approach, relating the effects of horizontal redistribution to structural heterogeneity allows estimating the spatial distribution of unsaturated zones across the river-aquifer interface.

3.9. Acknowledgements

The authors thank Philippe Renard and Craig Simmons for helpful comments and discussions. There is no observation data in the paper, and all results can be reproduced with the information given in the paper.

3.10. References

Banks, E.W., Brunner, P., Simmons, C.T., 2011. Vegetation controls on variably saturated processes between surface water and groundwater and their impact on the state of connection. Water. Resour. Res., 47.

Boano, F. et al., 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Rev. Geophys., 52: 603-679.

Brunner, P., Simmons, C.T., 2012. HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. Ground water, 50(2): 170-176.

Brunner, P., Simmons, C.T., Cook, P.G., 2009a. Hydrogeologic controls on disconnection between surface water and groundwater. Water Resour. Res., 45.

Brunner, P., Simmons, C.T., Cook, P.G., 2009b. Spatial and temporal aspects of the transition from connection to disconnection between rivers, lakes and groundwater. J. Hydrol., 376: 159-169.

Butler, J.B., Lane, S.N., Chandler, J.H., 2001. Characterization of the Structure of River-Bed Gravels Using Two-Dimensional Fractal Analysis. Math. Geol., 33(3): 301-330.

Calver, A., 2001. Riverbed permeabilities: Information from pooled data. Ground water, 39(4): 546-553.

Carsel, R.F., Parrish, R.S., 1988. Developing joint probability distributions of soil water retention characteristics. Water Resour. Res., 24(5): 755-769.

Chen, X., 2011. Depth-dependent hydraulic conductivity distribution patterns of a streambed. Hydrol. Process., 25: 278-287.

de Marsily, G., Delay, F., Teley, V., Schafmeister, M.T., 1998. Some current methods to represent the heterogeneity of natural media in hydrogeology. Hydrolgeology J., 6: 115-130.

Fleckenstein, J.H., Niswonger, R.G., Fogg, G.E., 2006. River-aquifer interactions, geologic heterogeneity, and low-flow management. Ground water, 44(6): 837-52.

Fox, G.A., Durnford, D.S., 2003. Unsaturated hyporheic zone flow in stream/aquifer conjunctive systems. Adv. Water Resour., 26(9): 989-1000.

Frei, S., Fleckenstein, J.H., 2014. Representing effects of micro-topography on runoff generation and subsurface flow patterns by using superficial rill/depression storage height variations. Environ. Model. Softw., 52: 5-18.

Frei, S., Fleckenstein, J.H., Kollet, S.J., Maxwell, R.M., 2009. Patterns and dynamics of river–aquifer exchange with variably-saturated flow using a fully-coupled model. J. Hydrol., 375(3-4): 383-393.

Frei, S., Lischeid, G., Fleckenstein, J.H., 2010. Effects of micro-topography on surface–subsurface exchange and runoff generation in a virtual riparian wetland — A modeling study. Adv. Water Resour., 33(11): 1388-1401.

Harvey, J.W., Gooseff, M., 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resour. Res., 51.

Hillel, D., 1998. Environmental soil physics Academic Press, London, UK.

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

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

Kalbus, E., Reinstorf, F., Schirmer, M., 2006. Influence of aquifer and streambed heterogeneity on the distribution of groundwater discharge. Hydrol. Earth Syst. Sci., 10: 873-887.

Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in Sedimentary Deposits: A Review of Structure-Imitating, Process-Imitating, and Descriptive Approaches. Water Resour. Res., 32(9): 2617-2658.

Kurtz, W., Hendricks Franssen, H.-J., Brunner, P., Vereecken, H., 2013. Is high-resolution inverse characterization of heterogeneous river bed hydraulic conductivities needed and possible? Hydrol. Earth Syst. Sci., 17: 3795-3813.

Lamontagne, S. et al., 2014. Field assessment of surface water-groundwater connectivity in a semi-arid river basin (Murray-Darling, Australia). Hydrol. Process., 28: 1561-1572.

Lodge, T.E., 2005. The Everglades Handbook: Understanding the Ecosystem. CRC Press, Boca Raton, FL, USA.

Lu, C., Chen, X., Cheng, C., Ou, G., Shu, L., 2012. Horizontal hydraulic conductivity of shallow streambed sediments and comparison with the grain-size analysis results. Hydrol. Process., 26: 454-466.

Mariethoz, G., Caers, J., 2014. Multiple-Point Geostatistics: stochastic modeling with training images. Wiley-Blackwell.

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulation. Water Resour. Res., 46.

McVoy, C.W., Said, W.P., Obeysekera, J., Vanarman, J.A., Dreschel, T.W., 2011. Landscapes and Hydrology of the Predrainage Everglades. University Press of Florida, Gainsville, FL, USA.

Mitsch, W.J., Gosselink, J.G., 2007. Wetlands. John Wiley & Sons, Inc., Hoboken, NJ, USA, 582 pp.

Newcomer, M.E. et al., 2016. Simulating bioclogging effects on dynamic riverbed permeability and infiltration. Water Resour. Res., 52.

Osman, Y.Z., Bruen, M.P., 2002. Modelling stream-aquifer seepage in an alluvial aquifer: an improved loosing-stream package for MODFLOW. J. Hydrol., 264: 69-86.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci., 30(7): 683-691.

Pryshlak, T.T., Sawyer, A.H., Stonedahl, S.H., Soltanian, M.R., 2015. Multiscale hyporheic exchange through strongly heterogeneous sediments. Water Resour. Res., 51(11): 9127-9140.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rivière, A., Gonçalves, J., Jost, A., Font, M., 2014. Experimental and numerical assessment of transient stream-aquifer exchange during disconnection. J.Hydrol, 517: 574-583.

Rosgen, D.L., 1994. A classification of natural rivers. Catena, 22(3): 169-199.

Rubin, Y., 2003. Applied stochastic hydrogeology. Oxford University Press, New York, NY, USA, 416 pp.

Rubin, Y., Hubbard, S.S., 2005. Hydrogeophysics. Water Science and Technology Library, 50. Springer, Dordrecht.

Schilling, O.S., Cline, E., Dreschel, T.W., 2013. Critical flow patterns for an Everglades macro-scale physical model. Fla. Sci., 76(3-4): 381–390.

Schilling, O.S. et al., 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water– vegetation interactions. J. Hydrol., 519: 2258-2271.

Shanafield, M., Cook, P.G., Brunner, P., McCallum, J., Simmons, C.T., 2012. Aquifer response to surface water transience in disconnected streams. Water Resour. Res., 48(11).

Song, J., Chen, X., Cheng, C., Summerside, S., Wen, F., 2007. Effects of hyporheic processes on streambed vertical hydraulic conductivity in three rivers of Nebraska. Geophys. Res. Lett., 34.

Stewardson, M.J. et al., 2016. Variation in reach-scale hydraulic conductivity of streambeds. Geomorphology, 259: 70-80.

Tang, Q., Kurtz, W., Brunner, P., Vereecken, H., Hendricks Franssen, H.-J., 2015. Characterisation of river–aquifer exchange fluxes: The role of spatial patterns of riverbed hydraulic conductivities. J. Hydrol., 531: 111-123.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Manual, Draft October 2010. Groundwater Simulations Group, University of Waterloo, Waterloo, Canada.

Treese, S., Meixner, T., Hogan, J.F., 2009. Clogging of an effluent dominated semiarid river-a conceptual model of stream-aquifer interactions. J. Am. Water Resour. Assoc., 45(4): 1047-1062.

van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Sci. Soc. Am. J., 44: 892-898.

Wang, W. et al., 2016. A quantitative analysis of hydraulic interaction processes in stream-aquifer systems. Sci. Rep., 6(19876).

Wang, W., Li, J., Feng, X., Chen, X., Yao, K., 2011. Evolution of stream-aquifer hydrologic connectedness during pumping-experiment. J. Hydrol., 402(3-4): 401-414.

Xie, Y., Cook, P.G., Brunner, P., Irvine, D.J., Simmons, C.T., 2014. When can inverted water tables occur beneath streams? Ground water, 52(5): 769-74.

Chapter 4

4. Using tree rings as a proxy for transpiration in SW-GW model calibration

Abstract

The interactions between surface water, the vadose zone, groundwater, and vegetation are governed by complex feedback mechanisms. Numerical models simulating these interactions are essential in quantifying these complex processes. However, the notorious lack of field-observations results in highly uncertain parameterizations. We suggest a new type of observation data to be included in the calibration data set for hydrological models simulating interactions with vegetation: Tree rings as a proxy for transpiration. We use the lower Tarim River as an example site for our approach. In order to forestall the loss of riparian ecosystems from reduced flow over a 300 km reach of the lower Tarim River, the Chinese government initiated periodical, ecological freshwater releases. The water exchange processes in this region were simulated for a cross-section on the lower reaches of the Tarim River using a numerical model (HydroGeoSphere) calibrated against observations of water tables, as well as tree ring growth as a proxy for transpiration. A predictive uncertainty analysis quantifying the worth of different components of the observation dataset in reducing the uncertainty of model predictions was then employed. The flow of information from elements of the calibration dataset to the different parameters employed by the model was also evaluated. The uncertainty and the flow of information analyses demonstrated that tree ring records used as a proxy for transpiration can significantly improve confidence in modeling ecosystem dynamics, even if these transpiration estimates are uncertain. To use of the full potential of the historical information encapsulated in the Tarim River tree ring records, however, effort to better define the relationship between tree ring growth and transpiration rates is required.

Keywords:
model predictive uncertainty, ecohydrology, surface water-groundwater interactions, vadose zone, tree ring growth, transpiration

This chapter was published as an original research article:
Schilling, O.S., Doherty, J., Kinzelbach, W., Wang, H., Yang, P.N., and Brunner, P. (2014): Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions. J. Hydrol. 519, 2258-2271, doi: 10.1016/j.jhydrol.2014.08.063.

4.1. Introduction

The Tarim River in Xinjiang in northwest China (**Fig 4.1**) is the most important source of water in the Tarim Basin (Cui and Shao, 2005). Over the last few decades, the rapid development of irrigation in the Tarim Basin has led to an increased demand for water. This threatens the riparian *Populus euphratica* forests in the lower reaches of the river, where water stopped flowing after the construction of Daxihaizi Reservoir (**Fig. 4.1**) in 1972 (Hou et al., 2007b). In response to the severe degradation of these unique ecosystems, the Chinese government implemented the *Ecological Water Conveyance Project* (EWCP) in 2000. Its purpose is to conserve and restore the riparian ecosystems, and to prevent further desertification (Xu et al., 2007). These goals are to be achieved by periodically providing water from both Daxihaizi Reservoir and Lake Bostan (situated northwest of Daxihaizi Reservoir, **Fig. 4.1**) for flow downstream. Between 2000 and 2007, 11 ecological water releases were realized. These releases led to a rise of the water table in the vicinity of the river, and had a positive impact on the riparian vegetation (Chen et al., 2008): The tree ring growth record of *Populus euphratica* indicates significant growth periods in response to the flow releases, as opposed to the pre-release tree ring record (Yang and Li, 2011b, personal communication; Yu et al., 2011).

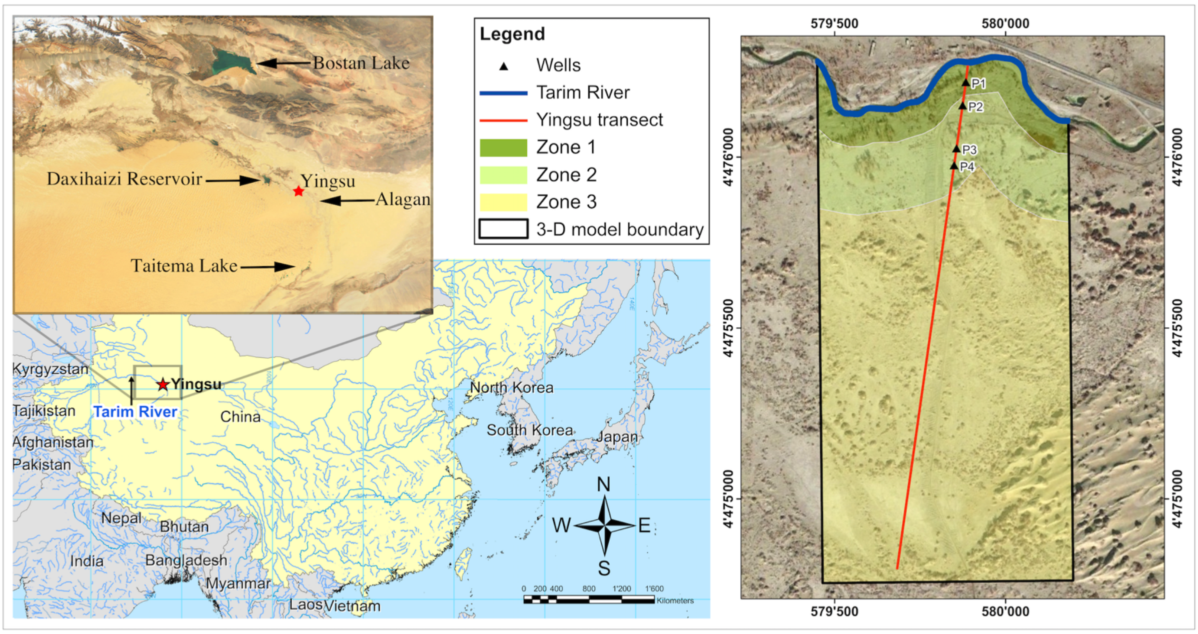


Figure 4.1: The study site on the lower Tarim River. The model outline as well as the transect and the tree distinct zones of vegetation density are outlined in the satellite photo on the right hand side (available through GoogleEarth (2003)). The locations of the wells used in the model (P1-4) are indicated along the Yingsu cross-section. The satellite image on the top left hand side shows the middle and lower reaches of the Tarim River in the Tarim Basin in Xinjiang province, China (satellite image available through NASA (2001)). Daxihaizi Reservoir, the Yingsu measurement site, Bostan Lake, as well as two other important sites, Alagan and Taitema Lake, are indicated. The total length of the lower reach of the Tarim River from Daxihaizi Reservoir to Taitema Lake is approximately 320 km. The location of the Tarim River in China is shown on the map on the bottom left corner. Coordinate system: WGS 1984 UTM Zone 45N.

Unfortunately, these ecological water releases come at high economic cost, as the water made available for the lower Tarim River could otherwise be used to support upstream agricultural production. In order to preserve the downstream ecosystems, while simultaneously minimizing the costs for the agriculture upstream, the flow releases have to be optimized. Optimal in this context means that only the amount of water required by the plants is made available to them; water that is not used by plants replenishes the vast regional groundwater systems and has minimal economical value. Understanding and quantifying the dynamics between infiltration along the river and transpiration of water through the riparian trees is therefore essential for an efficient allocation of the scarce water resources. To quantify and optimize the riparian water use efficiency (defined herein as the ratio of transpiration to infiltration from the river) of past and future water releases, a numerical model simulating all relevant processes is required. Parameters that govern hydraulic and vegetation processes need to be calibrated based on measurements of the system state. Unfortunately, as is the case for many remote field sites, apart from hydraulic head measurements taken at widely separated transects and some isolated discharge measurements, no other hydraulic data are available for the 321 km length of the lower Tarim River. This is problematic, as observations of hydraulic heads cannot fully inform parameters governing plant water uptake processes (Brunner et al., 2012). Given the limited availability and diversity of observations, the assessment of the overall efficiency, and eventually of the economical benefits, of past and future ecological water releases is highly uncertain.

As indicated above, however, the previous flow releases have left a very clear signature in the tree ring growth record of the riparian vegetation (Yu et al., 2011). The calibration dataset could be significantly expanded, and the uncertainty of predictions of the riparian water use efficiency reduced, if it were possible to integrate the historic information encapsulated in the tree ring growth record. But before a significant amount of tree ring growth data is acquired through expensive and time-consuming field campaigns along the Tarim River, a range of important questions have to be addressed:

- How can tree ring growth data be integrated into the model calibration process? Or how can the historic information encapsulated in tree rings inform models simulating surface water, groundwater, and vegetation interactions?

- What is the data worth? In other words, to what extent can tree ring growth data reduce the uncertainty of the predicted riparian water use efficiency beyond that afforded by observations of hydraulic heads?

- Which specific model parameters can be estimated with the new type of data?

This paper aims at answering these three questions, using the Tarim River site as an example. For this purpose, a numerical model that simulates the infiltration along the river, the flow through the unsaturated and saturated zone, as well as plant water uptake equivalent to transpiration, is set up. The model is calibrated using hydraulic head observations, as well as transpiration estimates calculated from tree ring data. The worth of the different types of observation data in reducing the uncertainty of predicted infiltration and transpiration rates is quantified under consideration of their measurement accuracy. Finally, an analysis on the flow of information from the calibration dataset to the calibrated parameters is carried out providing fundamental insights into how the specific observations inform model parameters.

The results of this study are critical for modeling, data usage, and for the planning of future data acquisition and water release strategies in the Tarim Basin. Additionally, this work has significant repercussions to the rapidly growing field of modeling the interactions and feedback mechanisms between hydrological and ecological systems. To the best of the author's knowledge this is the first study that integrates tree ring data (as a proxy for transpiration) into the calibration of a hydrological model and, for the first time, it is shown how such data can inform critical model parameters.

The paper is organized as follows: First the study site and the available data are described. Then the approach of estimating transpiration from tree ring growth records for the use in the calibration process is explained. Descriptions of the conceptual and numerical model, as well as of the model calibration and validation, follow. Subsequently, the methods applied to quantify data worth, predictive uncertainty reduction, and flow of information, are described. Simulation results, together with data worth and the flow of information analyses are then presented. The discussion of conclusions completes this study.

4.2. Materials & methods

4.2.1. Study site

In order to document the response of the water table during the ecological water releases, the local authorities installed observation wells at 9 cross-sections along the lower Tarim River (Chen et al., 2009). For the subsequent analyses, data from the cross-section in Yingsu, located 61 km downstream of Daxihaizi Reservoir (**Fig. 4.1**) are used. This location was chosen due to the availability of local data of the channel geometry, the water level in the Tarim River during the releases, and the subsequent response in groundwater and tree ring growth. Climatic data are also available.

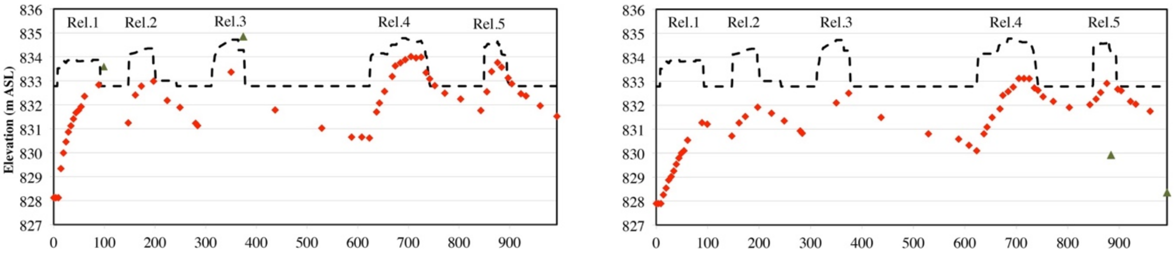
The channel section in Yingsu has a symmetric, trapezoidal shape, with a 10 m wide bed, a depth of 1.77 m and a bed-to-bank slope of 2.2. The bed has an elevation of 832.4 m ASL and the bank of 834.11 m ASL The river flows over bare sand; no streambed or clogging layer has developed in the channel. The aquifer in Yingsu is unconfined, more than 40 meters deep and composed of homogeneous windblown sand (Hou et al., 2007b; Yang et al., 2004). In total, 11 ecological water releases took place, but the first release was not properly documented in Yingsu, and after the 6th release water was routed through a new channel in the proximity of Yingsu. Therefore, only data of the first 5 documented releases are used in this study. These releases were named 2nd, 3rd(I), 3rd(II), 4th, and 5th(I). For better comprehension, the first documented release is hereby defined in this study as release 1 and incrementing numbers are assigned to the subsequent releases.

Discharge of releases 1-4 was measured at three measurement stations (Hou et al., 2007b): at Daxihaizi Reservoir (0 km), in Yingsu (61 km) and in Alagan (143 km) (**Fig. 4.1**). The average infiltration between the stations was estimated by calculating the reduction of flow between Yingsu and the other stations (**Tab. 4.1**). With such limited data, however, a reliable estimation of the infiltration is not possible for a particular location.

Table 4.1: An overview over the duration of stream-flow in Yingsu as well as estimated infiltration volumes. Infiltration volumes were calculated using the discharge volumes from Hou et al. (2007b). No information is available on release 5. The infiltration was separately calculated for the section between Daxihaizi Reservoir and Yingsu (Upstream) as well as for the section between Yingsu and Alagan (Downstream).

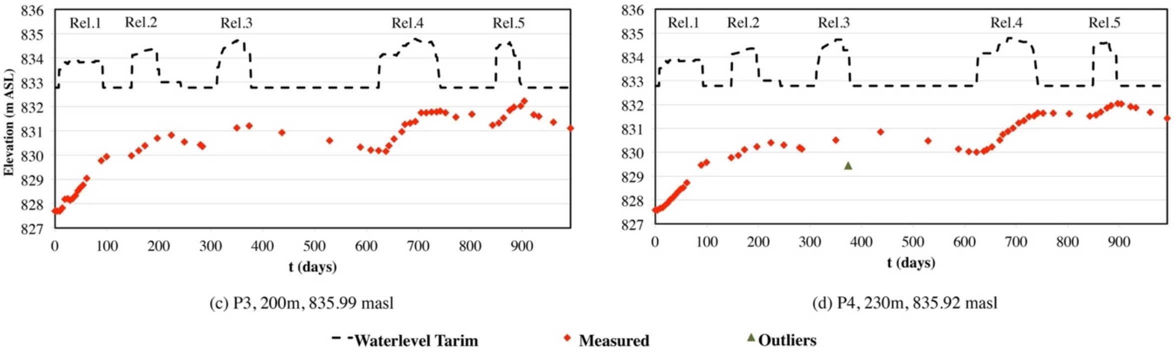
	Dates	Duration [days]	Infiltration [m³/m per riverside]	
			Upstream	**Downstream**
Release 1	17.11.2000 - 07.02.2001	83	106	542
Release 2	04.04.2001 - 07.07.2001	95	82	152
Release 3	15.09.2001 - 19.11.2001	66	62	144
Release 4	24.07.2002 - 19.11.2002	119	88	136
Release 5	06.03.2003 - 29.04.2003	55	-	-
Total		**418**	**338**	**974**

In Yingsu, the response of the groundwater table to the ecological water releases was measured in 8 observation wells along a transect. However, only for four wells reliable and complete data are available. These wells were labeled C3, C4, C5 and C1; for ease of reading they are here referred to as wells P1, P2, P3, and P4, respectively, and are located at distances of 60 m, 130 m, 200 m, and 230 m perpendicular to the channel. The locations of P1-4 are shown in **Fig. 4.1**. Well P4 lies on the north side of the channel. Assuming that the flow of water at the transect is approximately symmetric and that the subsurface is homogeneous, in this study well P4 was mirrored onto the south side of the Yingsu channel (this assumption is validated in Section 2.3.5.). Measurements of the depth to groundwater are available from the beginning of release 1. The measured hydraulic heads in wells P1-4 during the extended 993-days period of the ecological water releases 1-5 as well as the surface water table in the Tarim River are illustrated in **Fig. 4.2**. From the available 96 groundwater level observations, 5 were removed from the subsequent analysis (as indicated in **Fig. 4.2**). These observations were obvious outliers that could only be explained by measurement errors.



(a) P1, 60m, 834.86 masl

(b) P2, 130m, 835.84 masl



(c) P3, 200m, 835.99 masl

(d) P4, 230m, 835.92 masl

Figure 4.2: The diamond markers represent the GW level measurements available for the period of 07.11.2000 – 28.07.2003, comprising releases 1-5, in wells P1-4. The dotted black line represents the surface water level in the Tarim River. Triangular markers indicate the four measurements that were removed from the analysis. Note that there is an extensive unsaturated zone with a thickness between 4 to 7 meters; the thickness increases with increasing distance to the channel.

Potential evapotranspiration (E_P) in the Tarim Basin amounts to 2750 mm per year. Annual rainfall is less than 50 mm, and due to the high E_P a negligible source of groundwater recharge for the region. Apart from water availability and E_P, the density of the vegetation cover controls transpiration. The vegetation cover is expressed by the dimensionless leaf area index (LAI). Ma et al. (2009) found an average LAI of 1.85 for individual *Populus euphratica* trees. The density of *Populus euphratica* trees decreases with increasing distance to the channel. An analysis of satellite images (GoogleEarth, 2003) allowed identifying three distinct zones of vegetation densities in Yingsu (indicated in **Fig. 4.1**). The trees were counted in 100 m wide corridor along the transect. The highest density of *Populus euphratica* is found up to 100 m distance from the channel (Zone 1, 0.007 trees/m^2), at a distance between 100-300 m from the channel the density is less (Zone 2, 0.003 trees/m^2), and at a distance between 300-1500 m, the density is very small (Zone 3, 0.00057 trees/m^2).

Yu et al. (2011) who studied tree ring growth in Yingsu subdivided the growth of *Populus euphratica* into two periods, a low growth period prior to 2001 and a high growth period post 2001, indicating that water is the limiting factor for growth. Tree ring growth of the *Populus euphratica* trees has been recorded along the transect in Yingsu since 1995 (Yang and Li (2011b, personal communication), pers. communication). The tree ring growth data available to this study (**Tab. 4.2**) were measured in 5 plots at different distances to the channel and measurements were taken on branches of similar circumference. A comparison of tree ring growth prior to 2000 to growth in 2003 reveals an increase of growth by up to a factor of 5.

Table 4.2: Annual tree ring growth of *Populus euphratica* measured at different distances from the Tarim River (measured along the Yingsu cross-section). This data has been obtained through personal communication with Yang and Li (2011b, personal communication).

Distance to the river	Annual tree ring growth [mm]					
	2000	**2001**	**2002**	**2003**	**2004**	**2005**
50m	1.411	2.812	3.151	3.425		
150m	1.907	2.106	2.371	2.552	2.620	2.296
350m	1.191	2.306	2.268	2.288		
550m	0.853	1.064	1.532	1.850		
1000m	0.812	1.070	1.175	1.306		

In order to quantify the water balance for this system, and to understand the uptake of water through vegetation in response to the flow releases, information about transpiration is needed. As transpiration was not measured in Yingsu during the water releases an approach to relate tree ring growth to transpiration is required.

4.2.2. Estimating transpiration from tree ring growth

Literature on transpiration rates of *Populus euphratica* is scarce. Only three studies report transpiration of *Populus euphratica* for the larger Xinjiang region (Chen et al., 2011; Fu et al., 2011; Si et al., 2007a). For a number of reasons the application of these results to Yingsu is not straightforward. Chen et al. (2011) reported transpiration rates that exceed the E_P for Yingsu, and the data could therefore not be applied to this study. In the study of Fu et al. (2011) transpiration rates were reported in the form of standardized impact factors without further specifications. As insufficient information was provided to reconstruct the actual transpiration rates, the results of Fu et al. (2011) could not be integrated here. On the other hand, the study of Si et al. (2007a) on *Populus euphratica* proved to be crucial to the analyses of this study. Their study was carried out in the Ejina Basin, a neighboring basin with identical environmental conditions as in Yingsu. In their detailed literature review, the authors demonstrate that their measurements of transpiration are plausible for the given climatic conditions, and in line with transpiration estimates of other species in similar environments. Si et al. (2007a) measured transpiration of *Populus euphratica* together with depth to groundwater; tree ring growth was, however, not reported. The depth to groundwater was measured in Yingsu and in the Ejina Basin and it forms the key link in the estimation of transpiration based on three ring growth data.

Due to the limited data availability and remote nature of the field site a number of key assumptions had to be made to make use of the data provided by Si et al. (2007a):

- That the relations between depth to groundwater, tree ring growth, and transpiration in the Ejina Basin are the same as in Yingsu.

- That the relation between tree ring growth and transpiration is linear for *Populus euphratica*; In the only relevant publication known to the authors, Breda and Granier (1996) also found a linear relationship between transpiration and tree ring growth, in their case for *Quercus petrea*.

- And that tree ring growth is zero if transpiration is zero. This assumption is plausible because biomass production is inevitably linked to photosynthesis, which requires a certain degree of transpiration.

Based on the above-mentioned assumptions, the relationship between tree ring growth and transpiration was established, following the approach described below and illustrated in **Fig. 4.3**:

- In **Fig. 4.3a** tree ring growth in Yingsu (Yang and Li, 2011b, personal communication) is plotted as a function of the depth to groundwater. The data show that an increasing depth to groundwater is associated with decreasing growth, and that this relation is approximately linear.

- The average depth to groundwater and the corresponding transpiration of Si et al. (2007a) are 2.9 m and 4.05 m^3/a, respectively. The two circle markers on the x-axis of **Fig. 4.3a** and **4.3b** indicate these two values.
- The linear relationship based on the data between depth to groundwater and tree ring growth in Yingsu was used to calculate a corresponding value of tree ring growth for the measurement of Si et al. (2007a) (Point A, **Fig. 4.3a**).

- This provides the first point for establishing the relation between transpiration rates and tree ring growth (Point B, **Fig. 4.3b**).

- The second point of the relationship between transpiration and tree ring growth is the coordinate origin (**Fig. 4.3b**).

- Under the plausible assumption that tree ring growth is zero for zero transpiration, and following the findings of Breda and Granier (1996) that the relation between tree ring growth and transpiration is linear, a linear relation defined through the coordinate origin and point B was established (**Fig. 4.3b**). This resulting linear function was used to convert each annual tree ring growth observation in Yingsu to a corresponding transpiration rate.

- Clearly, these estimates of transpiration are highly uncertain. An error of +/- 20 % for the linking point B in **Fig. 4.3b** is assumed. This provides a sufficiently large margin to consider measurement errors of Si et al. (2007a) but also to account for uncertainty related to the assumption of linearity. The assumed error translates into +/- 0.8 m^3/a for point B. The dashed lines in **Fig. 4.3b** represent this uncertainty in the transpiration estimates. The uncertainty is subsequently considered both in the calibration process (see section 4.2.3.4) as well as in in the predictive uncertainty analysis (see section 4.3.2). For the predictive uncertainty analysis, even larger uncertainties are explored.

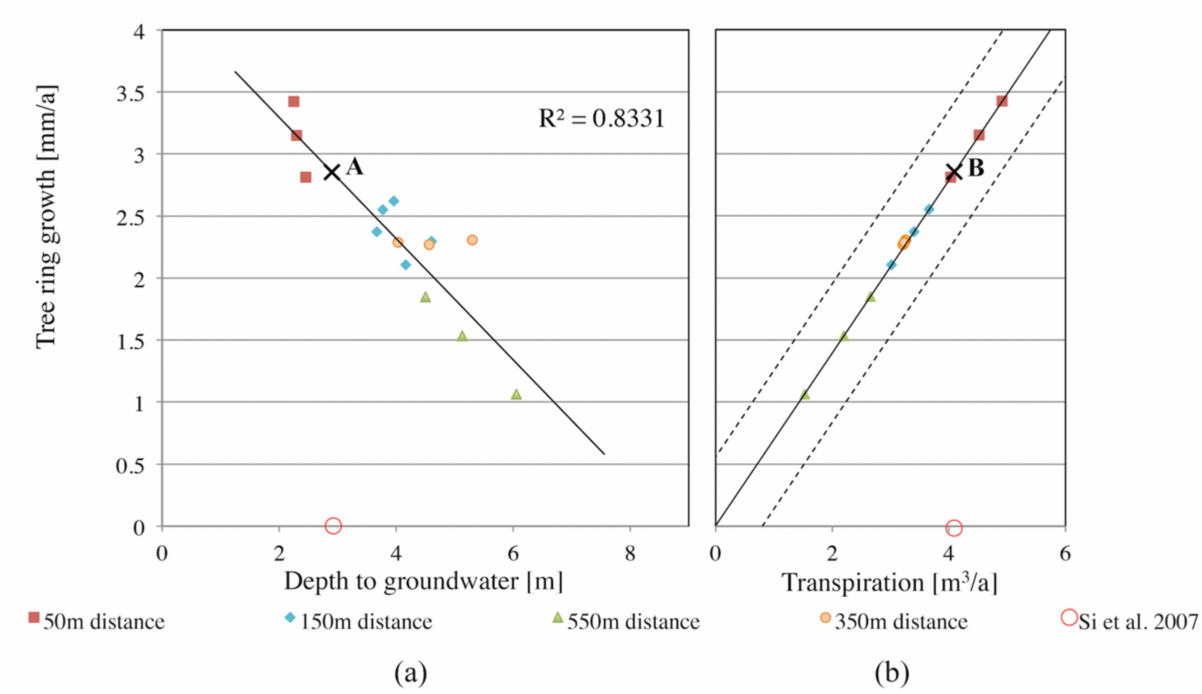


Figure 4.3: (a) Tree ring growth of Populus euphratica as a function of the depth to groundwater in Yingsu (Yang and Li, 2011a, personal communication). The relation is approximately linear (Tree ring growth = -0.488 * Depth to groundwater + 4.273). The different symbols correspond to different distances from the channel in Yingsu. (b) The estimated linear relationship between tree ring growth and transpiration (Tree ring growth = 1.434 * Transpiration). The circles on the x-axes represent the depth to groundwater/transpiration measurement of Si et al. (2007a). Point A and B represent the calculated tree ring growth for the value of Si et al. (2007a). The transpiration uncertainty range of +/- 0.8 m^3/a is indicated in Figure (b) by two dotted lines.

The estimated annual transpiration rates were then multiplied by the number of trees found in the three vegetation zones (**Fig. 4.1**) that correspond to the distances of tree ring measurements. The resulting transpiration volumes are summarized in **Tab. 4.3**.

Table 4.3: Estimated transpiration rates for a cross-section of 1 m width in m^3/a. The total transpiration is the sum of the three different zones of vegetation.

	Annual transpiration volume per vegetation zone [m^3]			
	Zone 1	**Zone 2**	**Zone 3**	**Total**
2001	2.8	1.9	1.1	5.7
2002	3.2	2.0	1.5	6.7
2003	3.4	2.1	1.9	7.4

4.2.3. Modelling

4.2.3.1. Modelling strategy

A 2-D cross-sectional model is used to simulate the dynamics along the Yingsu transect (the red line in **Fig. 4.1**). This modeling strategy was chosen because the computational demand for calibration and uncertainty analysis of a 2-D model is significantly smaller compared to a 3-D model. Modeling Yingsu using a 2-D approach implicitly assumes that the meander does not have a significant three-dimensional influence on groundwater flow and the resulting riparian water use efficiency. The 2-D approach is validated, by comparing the 2-D to a 3-D model.

The 3-D model simulates a bigger area surrounding the transect, as outlined in **Fig. 4.1**. The same parameter values are used in both models. The observations, the simulation results of the 2-D model, and the simulation results of the 3-D model are compared by looking at:

- Groundwater levels using absolute values,
- Groundwater levels using the root mean squared error (RMSE),
- Groundwater levels using the Nash Sutcliffe model efficiency (NSE) coefficient (Moriasi et al., 2007), and
- Transpiration and infiltration.

After demonstrating that a 2-D model is a valid approximation for the flow field in Yingsu, the data worth of the observations in reducing the predictive uncertainty of the model is quantified. This allows calculating to what extent the historic information encapsulated in the tree rings complement observations of hydraulic heads. Finally, the flow of information from observations to model parameters, which are calibrated, is explored.

4.2.3.2. Modelling code

For a number of reasons, the fully-coupled, physically-based HydroGeoSphere (HGS) modeling code (Therrien et al., 2010) is used in this study. Above all HGS was chosen for its ability to simulate all the relevant processes, which are: the flow in the unsaturated zone, groundwater flow, infiltration, transpiration, and evaporation. Neglecting the unsaturated zone under a river can lead to significant errors in simulating the dynamics between surface water and groundwater (Brunner et al., 2010); this must be avoided in modeling the Tarim River site, where a substantial amount of water can be stored in the extensive unsaturated zone underneath the riverbed. Furthermore, the *Populus euphratica* extract water from both the saturated and unsaturated zones; HGS allows vegetation to extract water from both zones. HGS also provides a detailed model output on both transpiration and evaporation separately. The suitability of HGS in similar modeling contexts has been demonstrated in a number of studies, e.g., Doble et al. (2011) who simulated infiltration in response to surface water waves, or Banks et al. (2011) who simulated the interaction between surface water and vegetation.

The governing equations of HGS are described in detail in the Appendix. Subsurface flow in HGS is simulated based on the Richards equation. Saturation is related to the relative permeability (k_r) and to pressure using the approach of Van Genuchten (Therrien et al., 2010). The calculation of evapotranspiration is based on the approach presented by Kristensen and Jensen (1975) (as employed by Panday and Huyakorn (2004)). In the approach of Kristensen and Jensen (1975), plant transpiration is a mechanistic function of E_P and three scaling functions:

- The first relation describes the linear dependency of transpiration on the LAI and is parameterized using the two constants C_1 and C_2. Transpiration increases with increasing LAI (see f_1 in the Appendix (section 4.5)).

- The second relation describes the dependency of transpiration on the soil moisture content: the full transpiration potential is reached at a soil moisture content between the field capacity (θ_{fc}) and the oxic limit (θ_{ox}). Transpiration approaches zero above θ_{ox} and below the wilting point (θ_{wp}). This relation is parameterized using the constant C_3 (see f_2 in the Appendix (section 4.5)).

- The third relation describes the dependency of transpiration on the root depth (rd): The rd defines the maximum depth of transpiration, and the amount of water transpired per depth depends on the root mass distribution (see the Root Distribution Function (RDF) in the Appendix (section 4.5)).

If the E_P is not reached through transpiration alone water is removed by direct soil evaporation, which reduces with increasing depth and is limited by an extinction depth (ET depth). Below the ET depth water cannot be mobilized for direct soil evaporation on the surface; only roots can access deeper water for evapotranspiration.

4.2.3.3. Numerical model

4.2.3.3.1. Geometry

The transect is modeled in 2-D as an unconfined aquifer of 1500 m width and a minimum of 30 m saturated plus 5-9 m unsaturated thickness. The base elevation of the model is 797 m ASL. As the model is a 2-D cross-sectional model, the model thickness is one unit (1 m). The topography along the transect was surveyed at 7 points to a distance of 1500 m from the river (Yang et al., 2004). The topography of the cross-section was interpolated in a linear fashion from these known points of elevation. The Tarim River in Yingsu is approximately symmetric with respect to the flow direction, and therefore, only half space, i.e. the southern side of the river, is modeled to reduce the computational demand by a factor of two. The observation wells are simulated at the distance that corresponds to their perpendicular distance to the channel.

The 3-D model only simulates the southern side of the Yingsu channel section as well. The same channel and aquifer geometries are used in both models. The total length of the channel in the 3-D model is approximately 1000 m and the floodplain area has an extent of approximately 1 million m^2. For the 3-D model the topography of the Yingsu region was obtained as a 90 m resolution DEM dataset from the NASA Shuttle Radar Topographic Mission (SRTM) (Jarvis et al., 2008). The DEM was adjusted by -7 meters in order to match the unknown reference system of Yang et al. (2004).

4.2.3.3.2. Temporal and spatial discretization

Horizontally, the 2-D cross-sectional model consists of a rectangular grid with 222 cells that become narrower the closer they lie to the Tarim River. In the channel the resolution is 0.15 m. The high resolution within the channel and its bank ensures that the relation between water depth and wetted perimeter is following the real bathymetry. Vertically, the model is divided into 120 layers. The vertical node spacing in the lower part of the model domain is 0.875 m. The unsaturated zone is discretized with a maximum vertical resolution of 0.175 m. The spatial discretization of the 2-D cross-sectional model is shown in **Fig. 4.4**.

The 3-D model is spatially discretized with a lower resolution using a triangular grid (in contrast to the 2-D model). This slightly different approach was chosen to better represent the outline of the channel. Horizontally, the resolution decreases stepwise at distances of 6 m, 60 m and 250 m from the centerline of the channel. In the vertical direction, the model is subdivided into 31 layers. The top four layers have an approximate height of 0.125 m, the following 23 layers of 0.5 m and the bottom 4 layers of 7 m each. The spatial discretization of the 3-D model is illustrated in **Fig. 4.6**.

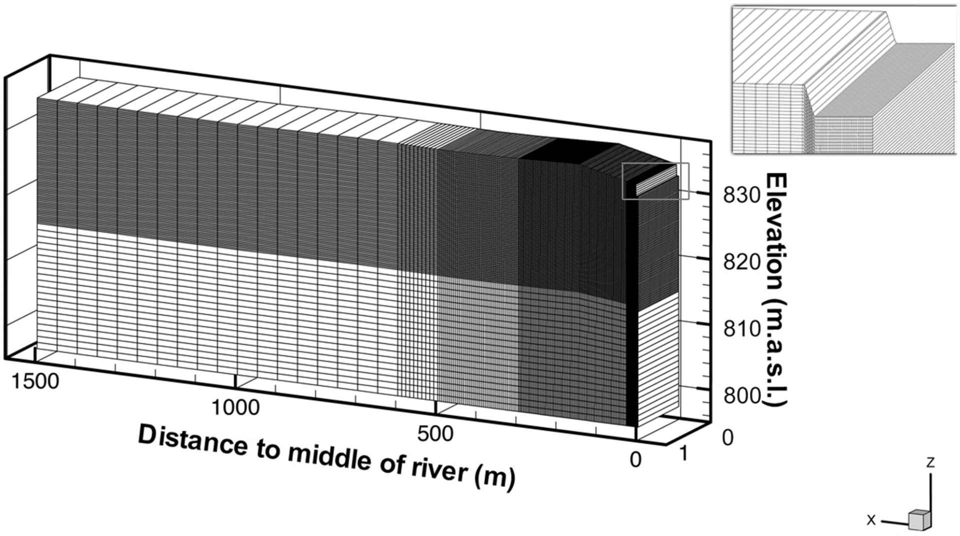


Figure 4.4: Topography and spatial discretization of the Yingsu model. The closer to the channel in the top right corner, the finer becomes the mesh. The vertical coordinates have been exaggerated for better graphical presentation.

An automatic time stepping scheme with a maximum time step size of 1 day and a minimum of 0.1 days is used for both models. The initial time step is 0.1 days.

4.2.3.3.3. Boundary and initial conditions

The base of the models, as well as the symmetry axis, which is represented by the centerline of the stream, is impervious. Groundwater flow along the river is assumed to be perpendicular to the stream and therefore the lateral boundaries of the 2-D and the 3-D models are also impervious. A steady water table 1500 meters away from the channel has established during the decades without water flowing in the Tarim River. This is conceptualized through a constant head boundary condition with a hydraulic head of 827 m ASL. The stage in the Tarim River (represented by the dashed line in **Fig. 4.2**) is modeled as a time dependent head boundary condition.

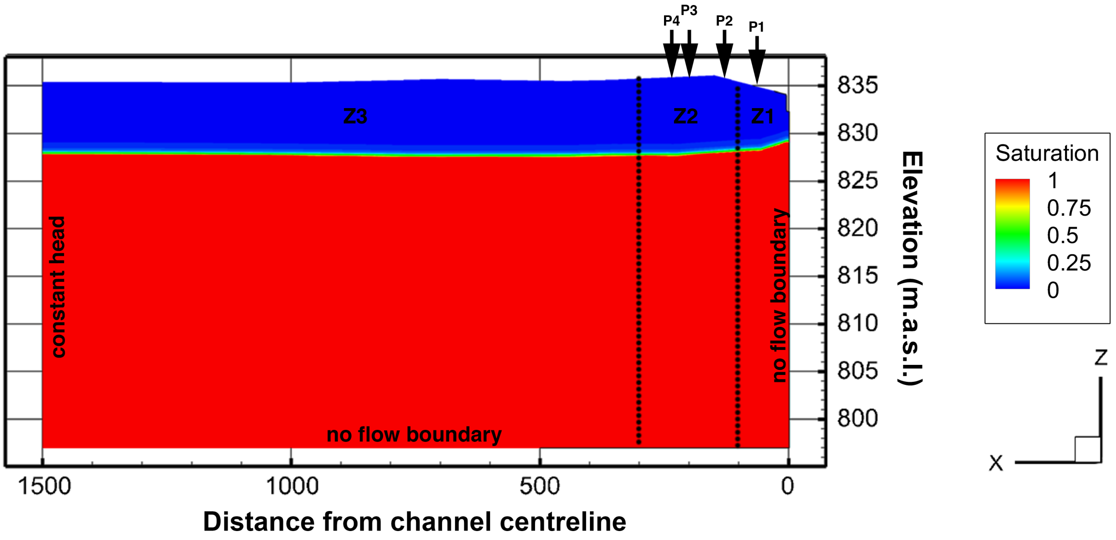


Figure 4.5: The initial and boundary conditions of the 2-D model. The three zones of different vegetation densities (Z1, Z2 and Z3) are indicated by the dotted lines. The vertical coordinates have been exaggerated for better graphical presentation.

The initial groundwater table was obtained from the measured groundwater levels at the beginning of the first water release. No data were available for the depth to groundwater directly below the river and the water table was extrapolated from the measurements at P2 and P1. **Fig. 4.5** illustrates the initial water table of the 2-D model of Yingsu, the different boundary conditions, the locations of the observation wells, as well as the three different vegetation zones.

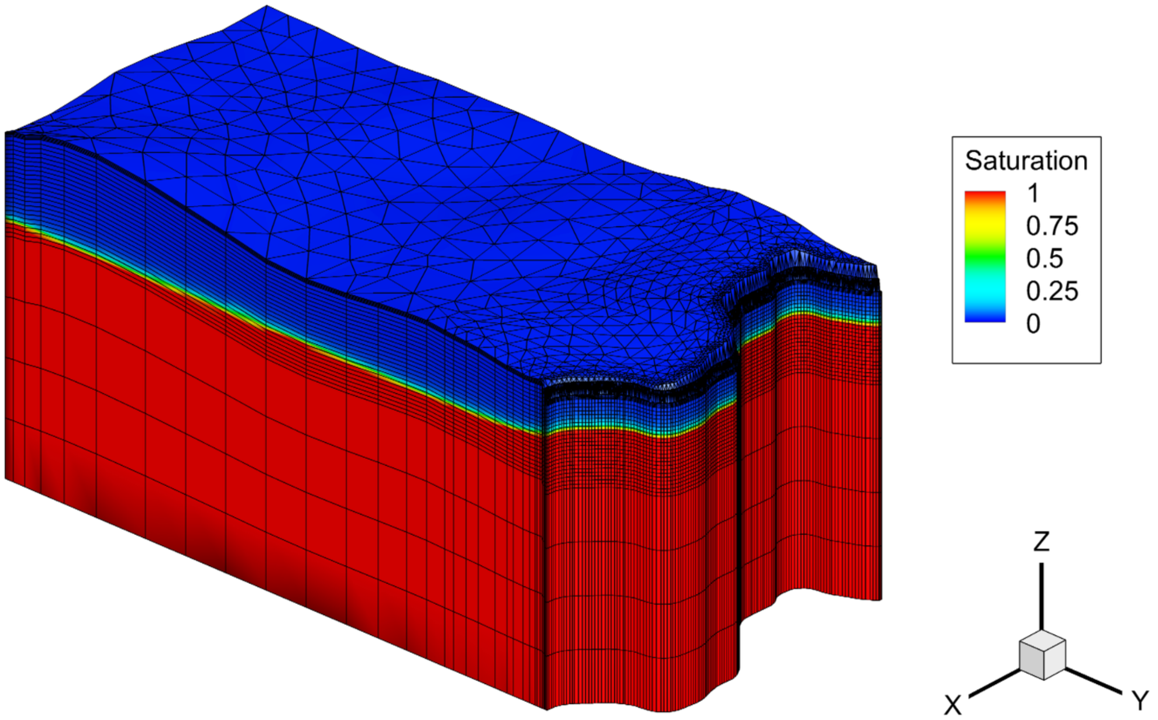


Figure 4.6: Initial condition and topography of the 3-D model of Yingsu. The initial condition is the same as in the 2-D model. The smaller the distance to the channel, the higher becomes the resolution of the grid.

The initial water table of the 3-D model is shown in **Fig. 4.6**. The water table observed along the transect is assumed to be representative for the entire model area and used as initial condition for the entire 3-D model.

4.2.3.4. Model parameters and model calibration

The transient 2-D model was calibrated using PEST (Doherty, 2015). PEST optimizes model parameters by minimizing a least squares objective function of differences between weighted observations and model outputs. A weight must be assigned to every observation and is ideally representing the inverse of the measurement error. The 2-D model was calibrated against data from releases 1 and 2; the subsequent releases (releases 3-5, t=285 until t=993) were used for validation. The observation data sets used in the calibration were:

- Groundwater water levels measured between 07 November 2000 (t=0) and 19 August 2001 (t=285), i.e. 23 observations for each of the four wells P1-4.

- A transpiration estimate representing the first 230 days of 2001, which are included in the calibration period (1 January 2001 (t=55) -19 August 2001 (t=285)).

A weight of 2, representing a very conservative measurement error of 0.5 m, was assigned to all groundwater level measurements. With a weight of 5 assigned to the transpiration estimate, visibility of transpiration in the calibration objective function was guaranteed; this being in accordance with the strategy recommended by Doherty and Welter (2010). The weight for transpiration was chosen in the following way: the inverse of the assumed measurement error of 0.8 m^3/a that was introduced in section 4.2.2 is 1.25. Multiplying 1.25 by the number of observation wells (4) equals a weight of 5.

The parameters to calibrate are shown in bold in **Tab. 4.4**. The upper and lower limits given to each parameter during calibration are shown in brackets. Uncalibrated parameter values were chosen for the following reasons: LAI values represent the reference LAI multiplied by the vegetation density per zone (see section 4.2.2). The coupling length (l_{exch}) for the exchange between the river and the subsurface was set to a low value to reflect the fact that no clogging layer is present in the stream. By setting C_2 to 0 transpiration is inhibited when no trees are present (taking into account that no grass is growing there). C_3 was set to a high value to ensure rapid uptake of water by the trees, as these trees must be able to react immediately to the availability of water (see Appendix for more details on C_2 and C_3). Trees that grow in arid environments, such as in the Tarim Basin, but that are periodically inundated, have developed root systems that allow taking up water even if their roots are fully submerged (Mitsch and Gosselink, 2007). This behavior is conceptualized by setting θ_{ox} to a relative saturation of 1.0 and the anoxic limit (θ_{an}) to 1.001. Other parameters listed in **Tab. 4.4** represent standard values for sandy soils.

Table 4.4: List of model parameters. Parameters subjected to calibration and their optimized values are bold. For the calibrated parameters, the lower and upper boundaries used in calibration are given in brackets. See the Appendix (section 4.5) for detailed descriptions of the parameters.

Aquifer parameters		**Vegetation parameters**	
K [m/d]	**1.84** (1-25)	**C_1** [-]	**0.7** (0.01-200)
n [-]	**0.16** (0.1-0.42)	C_2 [-]	0
α [1/m]	**5.04** (4-14)	C_3 [-]	10
β [-]	**2.5** (1.01-2.9)	**rd1** (zone 1) [m]	**3.5** (0.5-7)
S_{wr} [-]	0.02	**rd2** (zone 2) [m]	**4.3** (1-9)
E_P [m/d]	0.0077	**rd3** (zone 3) [m]	**6** (1-10)
ET depth [m]	0.5	LAI1 (zone 1) [-]	0.0777
l_{exch} [m]	0.001	LAI2 (zone 2) [-]	0.0333
		LAI3 (zone 3) [-]	0.006475
		θ_{wp} [-]	0.05
		θ_{fc} [-]	0.15
		θ_{ox} [-]	1
		θ_{an} [-]	1.001

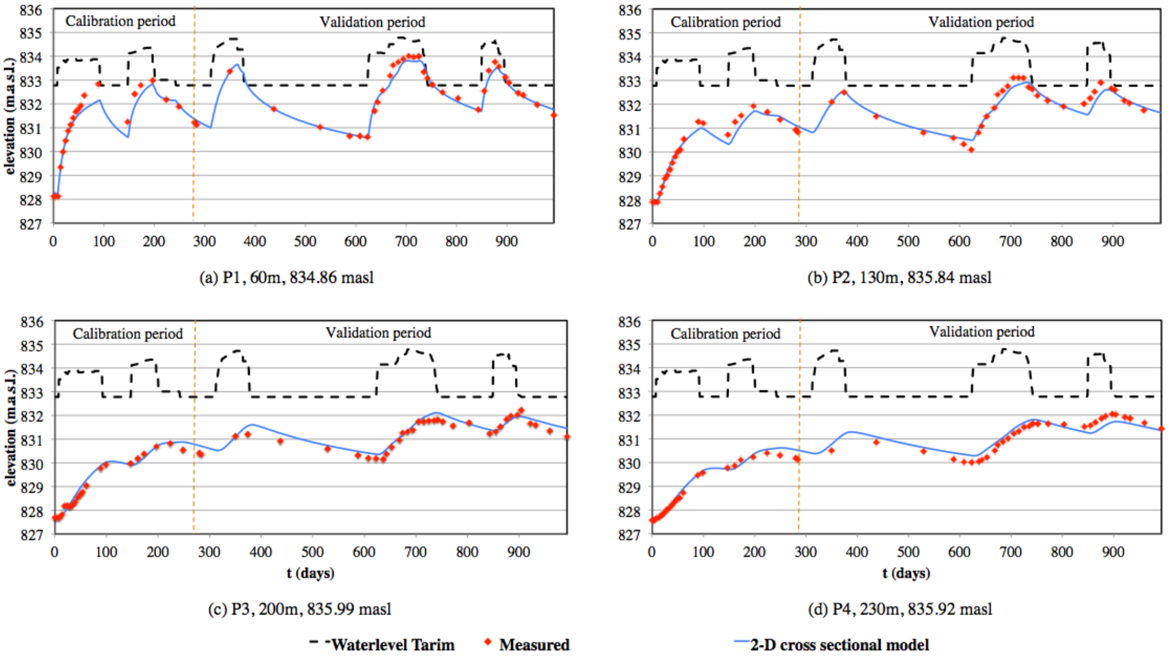
4.2.3.5. 2-D model validation

4.2.3.5.1. Calibrated parameters

Calibrated parameter values are highlighted as the bold values in **Tab. 4.4**. Values for the aquifer parameters hydraulic conductivity (K) and porosity (n) obtained in the calibration process compare well to the values found by Yang et al. (2004), and match the literature values for sandy soil types (Carsel and Parrish, 1988). The vegetation parameters are also reasonable: Root depths of the three distinguishable zones of vegetation density (rd1-rd3) are large enough to reach the saturated zone and match the depths that were found by (Yang and Li, 2011a, personal communication). No reported values of the fitting parameter C_1 for arid environments could be found in literature.

4.2.3.5.2. Comparison between the simulations and the measurements

In **Fig. 4.7** the simulated and the observed groundwater levels are compared for the calibration and the validation period. For the 2-D model, the largest discrepancy exists for release 1 in observation well P1 (**Fig. 7a**). However, all of the following releases are reproduced with a significantly higher accuracy. The mismatch for release 1 therefore seems to be related to the missing information of the initial water table directly below the stream. The RMSE for the 2-D model are 0.39 m, 0.35 m, 0.32 m and 0.35 m for wells P1, P2, P3 and P4, respectively. The RMSE for all four observation wells together is 0.35 m. This mismatch lies within the estimated uncertainty range of groundwater measurements (0.5 m, see section 4.2.3.4). Also, the groundwater levels simulated with the 3-D model, which wasn't directly calibrated but based on parameter values obtained through calibration of the 2-D model, fit the measurements sufficiently well (total RSME of 0.69 m).



(a) P1, 60m, 834.86 masl

(b) P2, 130m, 835.84 masl

(c) P3, 200m, 835.99 masl

(d) P4, 230m, 835.92 masl

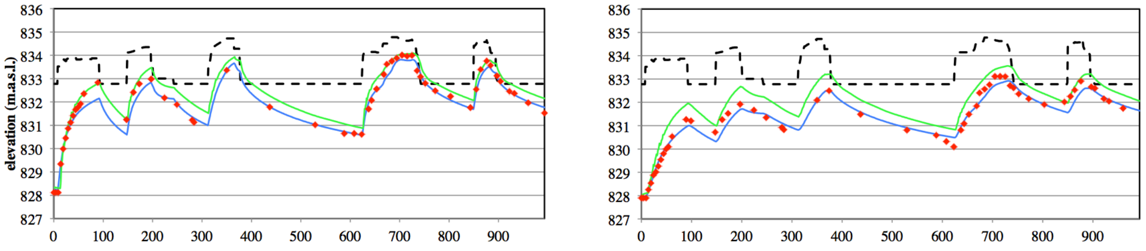
Figure 4.7: Groundwater levels simulated with the calibrated 2-D model in comparison to the measured groundwater levels. Indicated is the calibration period (releases 1 and 2), as well as the validation period (releases 3-5).

As the simulations also match the data of the mirrored well P4 (see section 4.2.2), the results show that the subsurface in Yingsu is approximately homogeneous.

The simulated infiltration volume amounts to a total of 639 m^3/m per bank over all events. As discussed in section 4.2.2, infiltration in Yingsu was not measured during the flood events, but average infiltration rates between Daxihaizi, Yingsu and Alagan could be estimated from discharge measurements (**Tab. 4.1**). The simulated infiltration rates correspond to the estimated average infiltration between the up- and downstream of Yingsu.

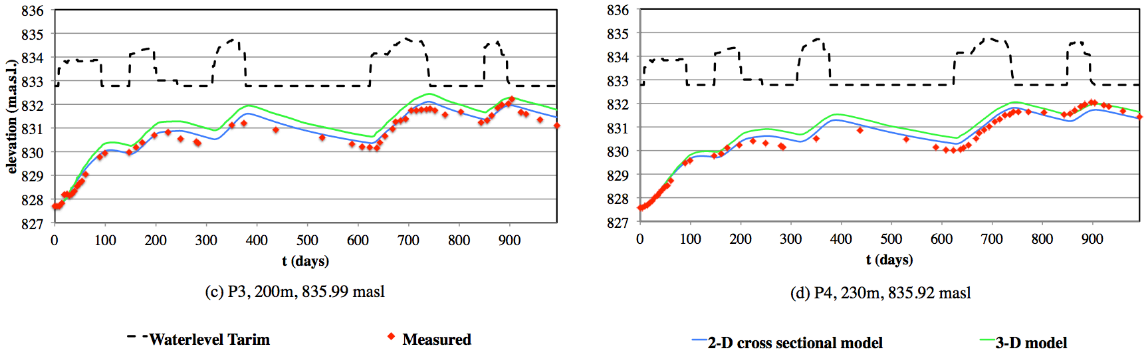
4.2.3.5.3. Comparison between the 2-D and the 3-D model

The 2-D model compares well to the 3-D model of the Yingsu area. A comparison of simulated groundwater levels is shown in **Fig. 4.8a-d**. Differences between the two models are more pronounced in the beginning of the simulation, possibly related to uncertainty in the initial groundwater levels. Considering the NSE coefficients for the groundwater levels of the two models, the models perform almost identically well, with a value of 0.97 for the 2-D and a value of 0.91 for the 3-D model (for the dimensionless NSE coefficient a value of 1 represents the best and a value of -∞ the worst possible fit). Hence, the comparison of absolute groundwater levels, RMSE, and NSE reveal that 3-D effects are minimal and do not significantly affect the questions we seek to answer.



(a) P1, 60m, 834.86 masl

(b) P2, 130m, 835.84 masl



(c) P3, 200m, 835.99 masl

(d) P4, 230m, 835.92 masl

Figure 4.8: Results of the simulations with the 3-D model, obtained with the parameters calibrated in the 2-D model. 3-D model results are compared to the measured groundwater levels and the 2-D simulation results.

In addition to groundwater levels, the riparian water use efficiency was also compared between the models. The comparison revealed that the simulated efficiencies of the two approaches are also nearly identical, providing further evidence that the 2-D approach is a valid approximation of the system. These results are further discussed in section 4.3.1, and shown in **Fig. 4.9**.

4.2.3.6. Predictive uncertainty and data worth analysis

In order to assess the data worth of observations of hydraulic head and of estimates of transpiration, the pre- and post-calibration predictive uncertainties for a prediction of infiltration and for a prediction of transpiration are quantified with a linear analysis using the PEST utility GENLINPRED (Christensen and Doherty, 2008; Doherty, 2010a). Linear uncertainty analysis can provide valuable insights into predictive uncertainty, even if applied to highly non-linear models (Brunner et al., 2012; Dausman et al., 2010). GENLINPRED establishes the worth of a particular observation by calculating how much the uncertainty of a prediction can be reduced through inclusion of this individual observation in the calibration dataset. In this study the data worth of a whole group of elements of the calibration dataset is considered, i.e. either all hydraulic head observations together or the transpiration observation. The reduction in uncertainty of a prediction, achieved through inclusion of different observation groups in the calibration dataset, is a measure of data worth.

The influence of measurement uncertainty on predictive uncertainty can also be explored in this analysis by varying the weights given to the different observation groups during the calibration process. As mentioned above in section 4.2.3.4 and described by Doherty and Welter (2010), the choice of weights can be a subjective process: While weights ideally represent the inverse of measurement uncertainty, both the number and distribution of measurement data points may require alternative weights in order to guarantee visibility of important measurements in the calibration objective function.

4.2.3.7. Flow of information analysis

While linear-type uncertainty analyses, as described above, are very useful to quantify the uncertainty of a given prediction, they do not provide insights into how exactly the observations used in the calibration process inform individual model parameters during calibration. This information can be extracted by the flow of information analysis: The flow of information from a calibration dataset to the parameters undergoing calibration can be tracked by subjecting the calibration sensitivity matrix to a Singular Value Decomposition (SVD) (Dausman et al., 2010; Doherty, 2010b). Quantifying the flow of information is mathematically more challenging than linear uncertainty analysis and requires a solid understanding of the concept of Eigenvectors and Singular Values. The detailed mathematics of this method are described in the PEST manual (Doherty, 2010b) and in studies that applied the method (Doherty and Hunt, 2009; Hill and Østerby, 2003). A qualitative description based on Aster et al. (2005) and Hill and Tiedeman (2007) is given below. Hill and Tiedeman (2007) describe that, in principle, applying SVD to the calibration sensitivity matrix yields the Eigenvalues, i.e. the Singular Values, and the associated Eigenvector pairs, i.e. the Eigenvector of parameters and the Eigenvector of observations. There are as many Eigenvector pairs coupled by Singular Values as there are parameters that are calibrated. Singular Values establish a ranking among the Eigenvectors. Eigenvectors that have Singular Values close to a maximum value of 1 indicate the highest flow of information, i.e. the most sensitive and informative couplings between observations and parameters. Analysis of the members of the Eigenvectors allows simultaneously identifying the sensitive parameters as well as the observations that inform these parameters. In each Eigenvector individual elements are associated with a coefficient between -1 and 1, and the squared sum of all coefficients amounts to 1. The coefficients are a measure of importance of an individual element relative to other elements of the same Eigenvector. Hence, in the Eigenvector of observations the magnitude of a coefficient is a measure of the ability to inform the paired Eigenvector of parameters. And vice versa, the magnitude of the coefficient of each parameter in the Eigenvector of parameters is a measure of its sensitivity to the paired Eigenvector of observations.

4.3. Results

4.3.1. Riparian water use efficiency

A volume of 18.3 m^3 of water is transpired during the 5 ecological water releases (t=0 until t=993) in the 2-D model. This compares well with the volume estimated from tree ring growth (see **Tab. 4.3**). Regarding the riparian water use efficiency, the phreatophytes consume a total of only 2.9% of the water that infiltrates during that time. The remaining 97.1% of the infiltrated water is stored in the aquifer. Part of this infiltrated water remains available to the trees. The ratio of simulated transpiration to infiltration increases with time: Simulation of 1000 additional days without further flooding after release 5 shows that the trees could potentially use up to a total of 3.24% of infiltrated water (**Fig. 4.9**). The riparian water use efficiency therefore improves over time. In order to complement the validation of the 2-D approach (see Section 4.2.3.5), the efficiency calculated with the 3-D approach is also plotted in **Fig. 4.9**. The calculated efficiencies are essentially identical.

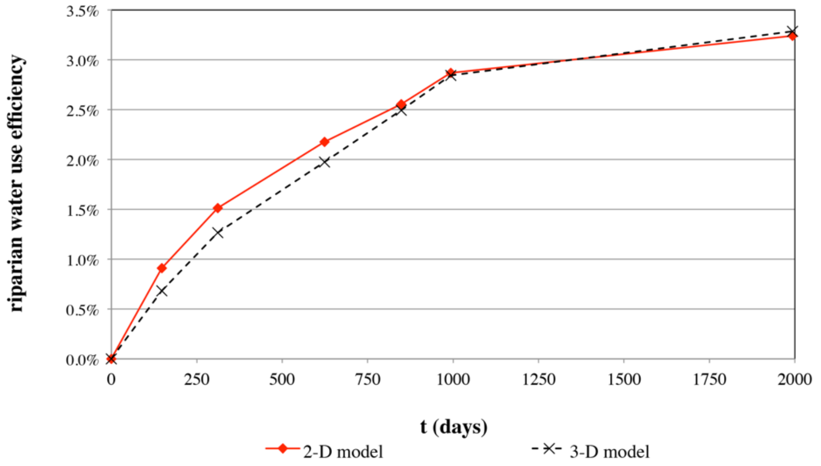


Figure 4.9: Comparison of the riparian water use efficiency simulated with the 2-D and the 3-D models spanning the entire simulation period plus an additional 1000-days relaxation period without additional stream-flow.

4.3.2. Predictive uncertainty and data worth

In order to analyze the data worth of the observation data, the data's potential to reduce the uncertainty for two predictions was quantified: for the transpiration volume of the year 2002 (**Tab. 4.3**) and the infiltration volume of release 3 (**Tab. 4.1**). The uncertainty for both predictions was calculated under consideration of the two groups of data used in calibration: i.e. observations of hydraulic heads (group H) and the estimate of transpiration (group T). The predictive uncertainty reduction was analyzed for the use of both groups of data separately (i.e. only H or T), as well as for their conjunctive use (H + T).

Recall that by systematically varying the weights associated to each observation group in the calibration dataset, the relation between measurement uncertainty and predictive uncertainty can be explored. The standard weights for this analysis are the weights used in the calibration process (see section 4.2.3.4). The weights for H were kept constant, as the measurement accuracy is reasonably well known. As discussed in section 4.2.2, however, the estimates of T are highly uncertain and the uncertainty might exceed the estimated 0.8 m^3/a. To explore this uncertainty, the weights for T were varied, ranging from the highest weight (5), as used during calibration, to a very small weight (0.01), representing a measurement uncertainty of 400 m^3. The resulting predictive uncertainties, expressed as total uncertainty standard deviations, for the transpiration volume of 2002 and the infiltration volume of release 3 are listed in **Tab. 4.5**.

Table 4.5: The predictive uncertainty reduction for the transpiration of 2002 and the infiltration of release 3. H stands for calibration with hydraulic head measurements only, T for calibration with transpiration only, and H+T for combined calibration. Weights associated to individual observations within each observation group are indicated in brackets. The 2-D model was calibrated with H(2) + T(5).

	Predictive uncertainty									
	pre-	**post-calibration total uncertainty standard deviation [m^3]**								
	-	**H(2)**	**T(0.01)**	**T(0.05)**	**T(0.5)**	**T(5)**	**H(2) + T(0.01)**	**H(2) + T(0.05)**	**H(2) + T(0.5)**	**H(2) + T(5)**
Transpiration of 2002	427.8	43.2	161.2	34.7	3.7	1.3	41.9	27.2	3.7	1.2
Infiltration of release 3	211.2	46.8	89.2	46.7	43.6	43.6	46.2	39.7	34.3	34.2

4.3.2.1. Prediction of the transpiration of 2002

If just an observation of T with a high weight of 5 is employed, in the absence of observations of H, the predictive uncertainty of transpiration of 2002 can be reduced by a factor of 330 in comparison to the uncalibrated model. Even a very conservative weight of 0.05, representing a very high uncertainty for the observation of T, still leads to a reduction of the predictive uncertainty of transpiration by a factor of 12.

When just observations of H are employed, a comparison of the pre-calibration with the post-calibration predictive uncertainty of transpiration reveals a reduction by a factor of 10; significantly less compared to using observations of T only.

The combination of both observation groups (H+T) is particularly helpful when the uncertainty of transpiration is very high, as could be the case for the estimates from tree ring growth. But if observations of T are more accurate (e.g. weights of 0.5 and 5), no further reduction of predictive uncertainty for predictions of transpiration can be achieved by complementing observations of T with observations of H.

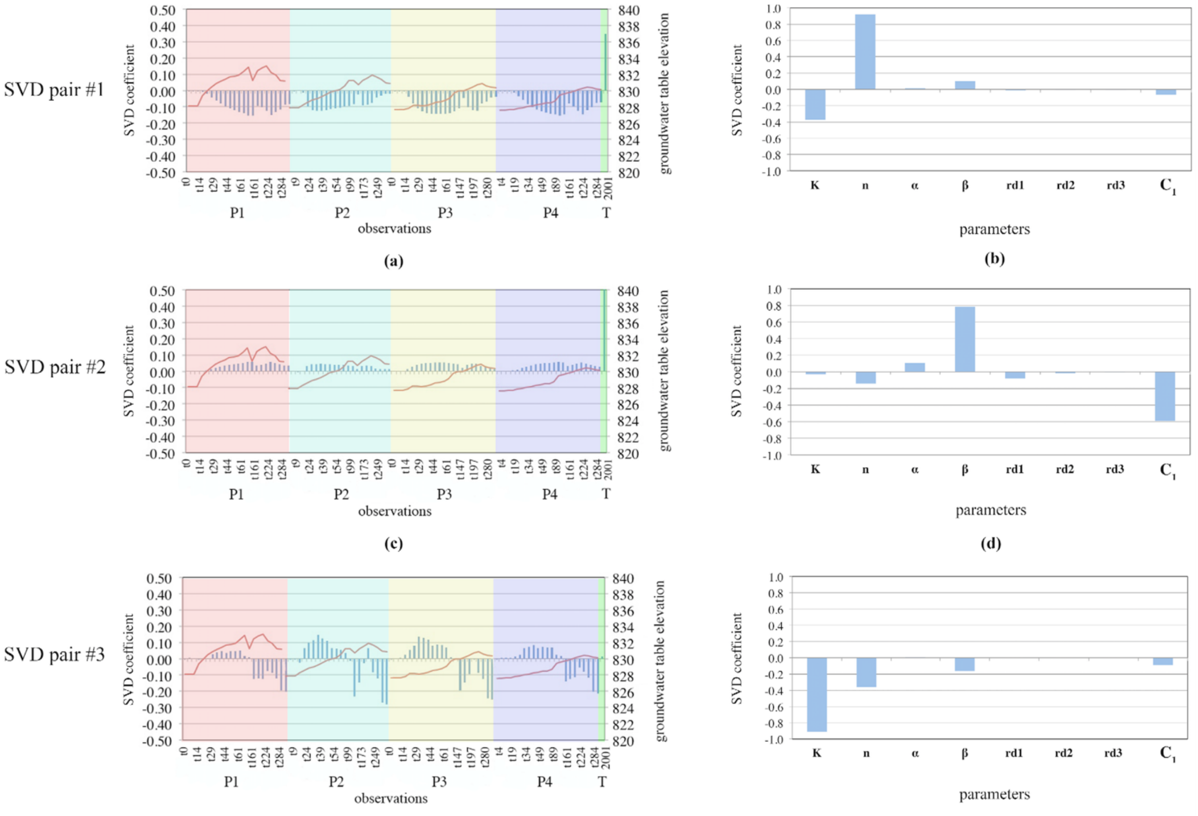
4.3.2.2. Prediction of the infiltration of release 3

Surprising findings result from the prediction of infiltration. Employing the observation of T (weight 5) without observations of H already reduces the predictive uncertainty of infiltration by a factor of 4.8; this is slightly above the reduction gained by employing observations of H only (a factor of 4.5). Even a very conservative weight of 0.01 for the observation of T yields a reduction of the predictive uncertainty of infiltration by a factor of 2.3. The combination of observations of H and T, with a high weight for T (5), results in the highest reduction, by a factor of 6.2.

This demonstrates that observations of T not only provide information about transpiration, but also about infiltration. They are especially valuable when their measurement accuracy is high.

4.3.3. Flow of information

The following results of the SVD analysis, illustrated in **Fig. 4.10**, allow relating the information content of the observations to the parameters subject to calibration. The Eigenvector pairs are ordered by their importance in terms of flow of information, represented by the magnitude of their Singular Values. The pair with the highest Singular Value are plots **4.10a/b**, the second most important pair plots **4.10c/d**, and the third most important pair plots **4.10e/f**. Plots on the left hand side of **Fig. 4.10** (**4.10a**, **4.10c**, and **4.10e**) represent the three Eigenvectors of observations. In each of these plots, the SVD coefficients of groundwater level observations are grouped by piezometer (P1-4) and plotted on the horizontal axis from left to right in order of time. As additional guidance the four solid lines associated with the secondary vertical axis indicate the measured groundwater levels over time for each piezometer separately. The single observation of transpiration (T) forms the last element on the horizontal axis of these plots. Plots on the right hand side of **Fig. 4.10** (**4.10b**, **4.10d**, and **4.10f**) contain the three Eigenvectors of parameters, which are paired with the Eigenvectors of observations on the same row.



(a) (b)

(c) (d)

(e) (f)

Figure 4.10: Eigenvector pairs of the SVD analysis. On the left hand side, Eigenvectors of observations are illustrated. The horizontal axis shows all the groundwater head observations of the four wells that were used in the calibration (P1-P4), ordered by time and grouped by well, as well as the single transpiration observation (T). The primary vertical axis indicates the importance of each observation in informing the paired Eigenvector of parameters; a high absolute SVD coefficient represents high information content. The red line associated with the secondary vertical axis represents the elevation of the simulated groundwater table in each observation well. The Eigenvectors of parameters are shown in the right column of this figure. Every row represents an Eigenvector pair, each pair being associated with one Singular Value. The three pairs are ordered by the magnitude of their Singular Value, with the highest Singular Value and therefore the most important pair in top row.

Recall that observations with high, absolute SVD coefficients in the Eigenvectors of observations have high information content; they inform the parameters that have high, absolute SVD coefficients in the associated Eigenvectors of parameters. As already shown in section 4.3.2 on data worth, plots **4.10a** and **4.10c** indicate high coefficients for transpiration and further support the finding that the information content of transpiration is high. The associated plots of Eigenvectors of parameters show that transpiration informs the inverse relation between the hydraulic conductivity (K) and the porosity (n) (represented by high coefficients in opposing directions in **Fig. 4.10b**) and to a lesser extent the relationship between the vegetation fitting parameter C_1 and the Van Genuchten parameter β (represented by high coefficients in opposing directions in **Fig. 4.10d**). Additionally, plot **4.10b** shows that the relation between K and n is also informed considerably by observations of hydraulic head. This can be intuitively understood, as the relation between K and n is linked to the diffusivity of an aquifer, and this determines the speed of lateral propagation of groundwater movement. In the present case, this determines the time over which the water level is high, and over which water is thus available for transpiration by phreatic vegetation. Plots **4.10c** and **4.10d** also illustrate the high information content of transpiration data with respect to the unsaturated zone, represented by the high coefficients for both T and the parameter β. This might reflect the fact that water must undergo capillary rise through the unsaturated zone to replenish transpiration demands, as the relation between suction pressure and saturation is a function containing β in the exponent.

The combination of observations in plot **4.10e** informs a similar combination of parameters as the combination of observations of plot **4.10a**. The difference between the first (**4.10a**) and the third (**4.10f**) Eigenvector of parameters lies mainly in the algebraic signs of the SVD coefficients of the important parameters. Information in the third Eigenvector of observations (**4.10e**), lies in the interplay between the rise and fall of the groundwater levels, rather than in the interplay between groundwater levels and transpiration, as is the case for the first Eigenvector of observations (**4.10a**). This interplay is represented by the change from positive to negative coefficients for each piezometer after the first release.

4.4. Discussion & conclusions

This study was motivated by an important environmental problem in western China, where *Populus euphratica* forests along the lower Tarim River are endangered due to a dramatically reduced amount of available freshwater. The ecological water releases of the Chinese government that started in 2000 aim at saving these forests, and the efficiency of such measures has to be ensured. Ecologists can provide estimates on the amount of water that is required by the trees. However, numerical modeling is required to quantify the amount of streamflow that is needed to replenish the alluvial aquifers, which provide that water to the trees. Down the line, such a numerical model could provide the basis to develop a release strategy that maximizes the overall efficiency of the ecological water releases.

The high number of parameters that have to be calibrated in such a numerical model simulating surface water, groundwater, and vegetation dynamics is in strong contrast to the limited amount of available observation data. But prior to additional data acquisition in such a remote region, it is necessary to determine the value of additional data. We have explored an approach that allows overcoming the lack of data and significantly expanding the amount of available observations by integrating the historic information encapsulated in tree ring growth. In this paper we have demonstrated how such data can be included in the calibration dataset, we quantified how much these data reduce uncertainty of key predictions, and we highlighted which parameters they inform. The lower Tarim River served as an ideal example site.

In a novel attempt unique to this study, a relationship between depth to groundwater, tree ring growth, and transpiration has been shown to successfully inform the choice of parameters needed in a model developed for the quantification of the riparian water use efficiency. Previous modeling attempts along the Tarim River have focused only on the use of groundwater level measurements. While the information content of head data is considerable, our study suggests that their use alone leaves an information deficit regarding transpiration and infiltration rates, coming from the inability of such data to constrain the water balance. This has also been shown in other studies, e.g., Doble et al. (2006). Our work agrees with these studies, strongly suggesting that measuring transpiration, and including these data into the modeling framework, is of outstanding importance for the quantification and the analysis of surface water, groundwater, and vegetation dynamics.

Our work shows that estimates of transpiration complement observations of groundwater levels in many respects, as only knowledge about transpiration allows closing the water balance and informing parameters describing the unsaturated zone, or the parameters required to simulate transpiration; without transpiration data, the solution for transpiration from shallow groundwater measurements is (within bounds) non-unique. We also show that as a consequence, without observations of transpiration, the high predictive uncertainty of the simulations makes the optimization of ecological water releases very challenging and uncertain. Counter-intuitively, even very low-confidence transpiration data significantly improve the predictive performance of the model. Our uncertainty analysis shows that tree ring growth data provide a possible proxy for transpiration, but also suggests that the current estimation of transpiration through tree ring growth has to be improved and complemented with more accurate, additional information (e.g. sap flow measurements) during future flow releases. This information would allow using the historic data encapsulated in tree rings along to their full potential. Furthermore, observations of infiltration rates would allow constraining the water balance even further, and obtaining an even better understanding of the system.

Specific to the lower Tarim River site, the results of this study indicate that the riparian water use efficiency increased over time with increasing release event number. This is not a surprising result, as the water volume required to infiltrate and bring the water table within range of the root zone is reduced for each subsequent release due to locally retained recharge. Modeling has allowed this effect to be quantified. Such quantification is a vital input to the design of a long-term release strategy, in which, among other things, the relative effectiveness of large occasional releases will be balanced against those of smaller, more frequent releases. Other considerations will of course also play a part in such a design, including the timing of demand for irrigation by upstream users. This will lead to an optimization problem, in which modeling, supported by targeted data acquisition, will play a central role.

4.5. Appendix

The following version of the Richards Eq. (4.1) for subsurface flow is used in HGS:

$$-\nabla \cdot (w_m \mathrm{q}) + \sum \Gamma_{ex} \pm Q = w_m \frac{\partial}{\partial t}(\theta_s S_w) \tag{4.1}$$

S_w: degree of saturation [-]
Γ_{ex}: volumetric fluid exchange rate between the surface and the subsurface zone [$L^3L^{-3}T^{-1}$]
θ_S: saturated water content [-]
w_m: volumetric fraction of the total porosity occupied by the porous medium [-]
Q: fluid exchange with the outside of the simulation domain [$L^3L^{-3}T^{-1}$]

As we only simulate one porous medium and don’t consider fractures, w_m is equal to 1. The fluid flux q [LT^{-1}] is outlined in Eq. (4.2) and contains the relative permeability $k_r=k_r(S_w)$ [-]:

$$\mathrm{q} = -\mathrm{K} \cdot k_r \nabla(\psi + z) \tag{2}$$

K: saturated hydraulic conductivity tensor [LT^{-1}]
ψ: pressure head [L]
z: elevation head [L]

Flow coupling between the surface zone and the subsurface porous medium is calculated with the dual node approach, where exchange is defined by equation (4.3):

$$d_0 \Gamma_0 = \frac{k_r \mathrm{K}_{zz}}{l_{exch}} (H - H_0) \qquad (4.3)$$

K_{zz}: vertical saturated hydraulic conductivity [LT^{-1}]
l_{exch}: coupling length [L]
H_0: surface water head [L]
H: subsurface water head [L]

The subsurface porous medium Eq. (4.4) puts the degree of saturation $S_w$in relation to the matric suction ψ and Eq. (4.5) puts it in relation to the relative permeability k_r.

$$S_w = S_{wr} + (1 - S_{wr})[1 + |\alpha\psi|^{\beta}]^{-\upsilon} \qquad \text{for } \psi < 0$$

$$S_w = 1 \qquad \text{for } \psi \geq 0 \qquad (4.4)$$

$$k_r(\psi) = S_e^{(l_p)}[1 - (1 - S_e^{1/\upsilon})^{\upsilon}]^2 \qquad (4.5)$$

$$S_e = (S_w - S_{wr}) / (1 - S_{wr}) \qquad (4.6)$$

S_{wr}: residual water saturation [-]
α and β: Van Genuchten parameters [-]
υ: given as 1 -1 / β with $\beta > 1$
S_e: effective saturation [-]
l_p: pore-connectivity parameter (=2)

Brunner et al. (2008) describe the equations for the calculation of evapotranspiration (ET) used in HGS as follows: The potential evapotranspiration $E_{potential}$ represents the maximum possible amount of water than can be evaporated with unlimited water supply. Due to the negligible amount of rain, interception and thus canopy evaporation E_{canopy}becomes zero for the water balance in this study. Transpiration from vegetation in contrast, is very important. Key variables of the calculation of transpiration (Eq. (4.7)) are the leaf area index *LAI* [-] (Eq. (4.8)), the soil moisture content θ [-] (Eq. (4.9)), and the quadratic decay root distribution function *RDF* [-].

$$T_p = f_1(LAI)\ f_2(\theta)\ RDF\ [E_{potential} - E_{canopy}] \qquad (4.7)$$

$$f_1(LAI) = \max\{0, \min[1, (C_2 + C_1 LAI)]\} \qquad (4.8)$$

C_1, C_2: transpiration fitting parameters [-]

The function of soil water content $f_2(\theta)$ is a dependence term. Below the wilting point θ_{wp}[-] transpiration is zero. A maximum of transpiration is reached between the field capacity θ_{fc}[-] and the oxic limit θ_{ox}[-].

$$f_2(\theta)=\left|\begin{array}{lll} 0 & & 0\le\theta\le\theta_{wp} \\ f_3 & & \theta_{wp}\le\theta\le\theta_{fc} \\ 1 & for & \theta_{fc}\le\theta\le\theta_{ox} \\ f_4 & & \theta_{ox}\le\theta\le\theta_{an} \\ 0 & & \theta_{an}\le\theta \end{array}\right. \quad (4.9)$$

$$f_3=1-\left[\frac{\theta_{fc}-\theta}{\theta_{fc}-\theta_{wp}}\right]^{C_3} \quad (4.10)$$

$$f_4=1-\left[\frac{\theta_{an}-\theta}{\theta_{an}-\theta_{ox}}\right]^{C_3} \quad (4.11)$$

C_3: transpiration fitting parameter [-]
θ_{wp}: moisture content at wilting point [-]
θ_{fc}: moisture content at field capacity [-]
θ_{ox}: moisture content at oxic limit [-]
θ_{an}: moisture content at anoxic limit [-]

The RDF describes how with increasing depth fewer roots are present. In this study the RDF is a simple quadratic decay function.

For further details on the mathematics of HGS consult the manual (Therrien et al., 2010).

4.6. Acknowlegements

Philip Brunner was funded by the Swiss National Foundation, Ambizione grant PZ00P2_126415. John Doherty is funded by the National Centre for Groundwater Research and Training, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia. We also thank Li X. (Xinjiang Agricultural University, College of Agricultural and Environmental Science; Urumqi 830052) for the data on tree ring growth. We thank Rebecca Doble and the anonymous reviewers for their constructive comments and suggestions.

4.7. References

Aster, R.C., Borchers, B., Thurber, C.H., 2005. Parameter estimation and inverse problems. International Geophysics 90. Academic Press, 296 pp.

Banks, E.W., Brunner, P., Simmons, C.T., 2011. Vegetation controls on variably saturated processes between surface water and groundwater and their impact on the state of connection. Water Resour. Res., 47.

Breda, N., Granier, A., 1996. Intra- and interannual variations of transpiration, leaf area index and radial growth of a sessile oak stand (*Quercus petraea*). Ann. Sci. For., 53: 521-536.

Brunner, P., Doherty, J.E., Simmons, C.T., 2012. Uncertainty assessment and implications for data acquisition in support of integrated hydrologic models. Water Resour. Res., 48.

Brunner, P., Li, H.T., Kinzelbach, W., Li, W.P., Dong, X.G., 2008. Extracting phreatic evaporation from remotely sensed maps of evapotranspiration. Water Resour. Res., 8.

Brunner, P., Simmons, C.T., Cook, P.G., Therrien, R., 2010. Modeling surface water-groundwater interaction with MODFLOW: some considerations. Ground Water, 48(2): 174-180.

Carsel, R.F., Parrish, R.S., 1988. Developing Joint Probability-Distributions of Soil-Water Retention Characteristics. Water Resour. Res., 24(5): 755-769.

Chen, Y.N. et al., 2009. Effects of ecological water conveyance on groundwater dynamics and riparian vegetation in the lower reaches of Tarim River, China. Hydrol. Process., 24(2): 170-177.

Chen, Y.N. et al., 2008. Response of riparian vegetation to water-table changes in the lower reaches of Tarim River, Xinjiang Uygur, China. Hydrogeol. J., 16(7): 1371-1379.

Chen, Y.P., Chen, Y.N., Xu, C.C., Li, W.H., 2011. Groundwater depths affects the daily course of gas exchange parameters of Populus euphratica in arid areas. Environ. Earth Sci.(66): 433-440.

Christensen, S., Doherty, J.E., 2008. Predictive error dependencies when using pilot points and singular value decomposition in groundwater model calibration. Adv. Water. Resour., 31(4): 674-700.

Cui, Y.L., Shao, J.L., 2005. The role of ground water in arid/semiarid ecosystems, Northwest China. Ground Water, 43(4): 471-477.

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

Doble, R., Brunner, P., McCallum, J., Cook, P.G., 2011. An Analysis of River Bank Slope and Unsaturated Flow Effects on Bank Storage. Ground Water, 50(1): 77-86.

Doble, R., Simmons, C., Jolly, I., Walker, G., 2006. Spatial relationships between vegetation cover and irrigation-induced groundwater discharge on a semi-arid floodplain, Australia. J. Hydrol.(329): 75-97.

Doherty, J.E., 2010a. Addendum to the PEST manual, Brisbane, Australia.

Doherty, J.E., 2010b. PEST. Model-independent parameter estimation-User Manual (5th edition, with slight additions), Brisbane, Australia.

Doherty, J.E., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. - PEST: complete theory and what it means for modelling the real world. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J.E., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. J. Hydrol., 366(1-4).

Doherty, J.E., Welter, D., 2010. A short exploration of structural noise. Water Resour. Res., 46.

Fu, A.H., Li, W.H., Chen, Y.N., Zhu, C.G., Ma, J.X., 2011. Analysis of dominant factors influencing moisture change of broad ovate-leaves of Populus euphratica Oliv. in extremely arid region. Photosynthetica, 49(2): 295-308.

GoogleEarth, 2003. Yingsu, Xinjiang, Western China. 40°26'2.75"N, 87°56'30.61"E. GoogleEarth 7.0.

Hill, M.C., Østerby, O., 2003. Determining Extreme Parameter Correlation in Ground Water Models. Ground Water, 41(4): 420-30.

Hill, M.C., Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: with analysis of data, sensitivities, predictions, and uncertainty. John Wiley & Sons, Inc., Hoboken, USA.

Hou, P., Beeton, R.J.S., Dong, X.G., Li, X., 2007a. Response to environmental flows in the lower Tarim River, Xinjiang, China. J Environ Manage, 83(4): 371-382.

Hou, P., Beeton, R.J.S., Dong, X.G., Li, X., 2007b. Response to environmental flows in the lower Tarim River, Xinjiang, China. J. Environ. Manage., 83(4): 371-382.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4.

Kristensen, K.J., Jensen, S.E., 1975. A model for estimating actual evapotranspiration from potential evapotranspiration. Nord. Hydrol., 6(3): 170-188.

Ma, Z.G., Chen, X., Japper, G., Chang, C., 2009. Study on LAI Estimation of Broadleaf Forests in Arid Areas Using Digital Hemispherical Photography - A Case Study in the Tarim River Basin, China, 2009 International Conference on Environmental Science and Information Technology, pp. 375-378.

Mitsch, W.J., Gosselink, J.G., 2007. Wetlands, 4th ed. John Wiley & Sons, Inc., Hoboken, NJ, USA, 582 pp.

Moriasi, D.N. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASAE, 50(3): 885-900.

NASA, 2001. Taklamakan Desert, Western China, Visible Earth.

Panday, S., Huyakorn, P.S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. Adv. Water. Resour., 27(4): 361-382.

Si, J.H. et al., 2007a. Sap flow of Populus euphratica in a desert riparian forest in an extreme arid region during the growing season. J. Integr. Plant Biol., 49(4): 425-436.

Si, J.H. et al., 2007b. Sap flow of Populus euphratica in a desert riparian forest in an extreme arid region during the growing season. J Integr Plant Biol, 49(4): 425-436.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Hydrogeosphere Manual.

Xu, H., Ye, M., Song, Y., Chen, Y., 2007. The natural vegetation responses to the groundwater change resulting from ecological water conveyances to the lower Tarim River. Environ. Monit. Assess., 131(1-3): 37-48.

Yang, P.N., Dong, X.G., Wumaierjiang, 2004. The simulation and analysis on a typical section of groundwater movement after transporting water to the lower reaches of Tarim River (in Chinese). Journal of Xinjiang Agricultural University, 27(3): 57-61.

Yang, P.N., Li, X., 2011a, personal communication. Root depths of *Populus euphratica* in Yingsu, Zürich/Urumqi.

Yang, P.N., Li, X., 2011b, personal communication. Tree ring growth records in Yingsu, Zurich/Urumqi.

Yu, P.J. et al., 2011. Effects of ecological water conveyance on the ring increments of *Populus euphratica* in the lower Tarim River, J. For. Res.

Chapter 5

5. Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel ^{37}Ar tracer method

Abstract

Quantitative and qualitative management of alluvial drinking water wellfields requires a sound understanding of the sources and the residence times of the pumped water. Numerical models are typically used for predicting future states of such systems. Complex, fully-integrated flow models would be ideal tools for wellfield management, but such models are notoriously under-informed, which results in uncertain predictions of system states. To overcome this problem, a combined multi-tracer and modelling study was carried out for an important drinking water wellfield in Switzerland. ^{222}Rn, ^{3}H/^{3}He, noble gases and the novel ^{37}Ar-method were used to quantify the sources and residence times of pumped drinking water. The study showed that to quantify mixing in alluvial drinking water wellfields, atmospheric noble gases and Helium isotopes are optimal tracers, whereas for residence times ^{222}Rn, ^{37}Ar and ^{3}H/^{3}He are ideal. The novel ^{37}Ar-method was for the first time used as a tracer in this context. ^{37}Ar successfully allowed closing a long-existing temporal gap of intermediate residence timescales of days to weeks in the residence time quantification. A comparison of the tracer-based to the simulated pumped water mix revealed that models which are only calibrated against hydraulic heads cannot reliably reproduce mixing patterns or residence times; only the tracer-based estimates of mixing provided the necessary information to identify an appropriate model parametrization. Consequently, multi-tracer studies that cover all relevant residence timescales are required to better inform complex, fully-integrated flow models.

Keywords: surface water-groundwater interactions, physically-based modelling, HydroGeoSphere, ^{222}Rn, ^{37}Ar, ^{3}H/^{3}He

This chapter is under review as an original research article:
Schilling, O.S., Gerber, C., Partington, D.J., Purtschert, R., Brennwald, M.S., Kipfer, R., Hunkeler, D. and Brunner, P. (under review): Advancing physically-based flow simulations of alluvial systems through atmospheric noble gases and the novel ^{37}Ar tracer method. Submitted to Water Resour. Res.

5.1. Introduction

Alluvial surface water-groundwater (SW-GW) systems are often ideal locations to pump groundwater (GW) for drinking water purposes, as the GW is typically of high quality and alluvial aquifers often provide a high yield (Diem, 2013). In Switzerland, for example, 36% of the total drinking water supply is GW pumped from alluvial aquifers, and most of the pumped GW represents river water that infiltrated at various distances upstream of the pumping station (Sinreich et al., 2012; Spreafico and Weingartner, 2005). SW-GW systems are the focus of a lot of ongoing research (e.g., Boano et al. (2014), Boulton et al. (2010), Harvey and Gooseff (2015)). For the management of drinking water resources, it is of particular importance to identify the sources and residence times of the pumped GW, and to identify recharge and infiltration zones, as the knowledge of flow paths, sources and residence times of pumped GW allows delineating optimal protection zones for drinking water wellfields. If, for example, the main source of pumped drinking water is recently infiltrated SW that did not remain in the subsurface for enough time to become naturally cleaned in the subsurface, one must make sure that the water is sufficiently treated prior to being supplied as drinking water. Information about the exchange fluxes, moreover, are essential to determine the availability of water for ecological requirements and manage drinking water abstraction.

The interactions between surface water (SW) and GW in alluvial valleys can be extremely dynamic and complex (Huggenberger et al., 1998; Sophocleus, 2002; Winter et al., 1998). Physically-based flow models such as Modflow (Harbaugh, 2005), HydroGeoSphere (Therrien et al., 2010), and ParFlow (Kollet and Maxwell, 2006), are typically used for the simulation of such SW-GW systems (Anderson et al., 2015), and as prediction tools for GW pumping stations (e.g., Bauser et al. (2010), Hendricks Franssen et al. (2011)). Since the underlying hydraulic parameters of SW-GW systems are often difficult to measure, SW-GW flow models need to be calibrated. In general practice, only hydraulic heads and SW discharge are used for the calibration of SW-GW models (Anderson et al., 2015; Simmons et al., 2012). However, hydraulic heads alone do not contain sufficient information to simultaneously reproduce GW levels, exchange fluxes, mixing and residence times, because the hydraulic conductivity (K) and porosity (n) of the aquifer (K_{aq} & n_{aq}) are strongly correlated towards a reproduction of fluxes, i.e., an increased K_{aq} can be compensated with an increased n_{aq}, and vice versa (e.g., Delottier et al. (2016), Townley (2012)). Consequently, numerical models of SW-GW systems are often under-informed and structurally flawed (Bredehoeft, 2005; Simmons et al., 2012). Moreover, even though the hydraulic conductivity of the riverbed (K_{rb}) is a first order control for exchange fluxes between SW and GW (Boano et al., 2014), K_{rb} is often poorly characterized, rarely considered in numerical flow models, and, in some environments, largely insensitive to hydraulic heads. K_{rb} is often insensitive to hydraulic heads because alluvial valleys are typically characterized by large aquifer cross-sections but small rivers that flow across these aquifers; as a result, the amount of SW-GW exchange fluxes in comparison to the total amount of flow through such valley cross-sections is small. Despite the strongly dampened effect of K_{rb} on hydraulic heads, in these systems, exchange fluxes that are controlled by K_{rb} are a major factor to consider for drinking water stations, as they pose a potential pathway for contamination if the pumped water constitutes a mix of GW and recently infiltrated SW (i.e., GW that has resided in the subsurface only for a few days to a few weeks).

To improve the structure and calibration of SW-GW models and to increase the reliability of predictions, a variety of additional measurements of SW-GW systems should be considered. For example, measurements of the fluxes, information on the sources of the pumped water, and of GW residence times provide important information about alluvial SW-GW systems (Anderson et al., 2015; Simmons et al., 2012). However, residence times are not straightforward to use in the context of flow model calibration (McCallum et al., 2014a; McCallum et al., 2014b). Quantifying the mixing ratio of GW from different sources in the pumped water, for example between older GW and SW that infiltrated only a few days to a few weeks ago, could be a promising alternative to inform complex SW-GW models. For the quantification of mixing ratios and of residence times the flow model needs to be able to track the flow of water and to provide information on travel times of individual water parcels. This, however, is not a standard feature of common flow modelling tools since the explicit simulation of transport is computationally demanding and complex, requiring an advection-dispersion equation to be solved, and potentially geochemical reaction processes to be considered (e.g., Anderson et al. (2015)). However, more efficient but less accurate methods for flow tracking, such as the mixing-cell method, directly use the flow solution and do not require additional parameters to be defined (e.g., Turnadge and Smerdon (2014)). With these efficient flow tracking methods, flow and travel times can be tracked throughout the entire model domain without much additional computational costs.

Two approaches to quantify mixing of water from different sources and to estimate residence times exist: Artificial tracer experiments and measurements of natural environmental tracers (Harvey and Gooseff, 2015). Artificial tracer experiments rely on the injection of, for example, non-reactive dyes at discrete locations into the SW or the GW, and on the subsequent tracking of the movement of the substance through the SW-GW system. They allow quantifying the travel times between the injection point and the sampled locations. The insights gained from such artificial tracer experiments allow the localized quantification of SW discharge, detection of SW-GW interactions, and estimation of the capture zone of a pump (Harvey and Gooseff, 2015). However, the information gained from artificial tracer experiments in GW is limited, as they only allow estimating the sources and travel times of the water that was marked, only allow the coverage of relatively short travel times due to the typically limited experimental duration, and the recovery rate of the injected tracer mass is generally very small (~10%) (Maliva, 2016). Measurements of natural tracers, on the other hand, do not rely on artificially injected substances. They have been widely and successfully used in the last two decades and allow estimating recharge locations, recharge temperatures and the residence times of GW on a large range of temporal and spatial scales, and if used in multi-tracer applications, furthermore allow estimating the mixing of different types of water (Cook and Böhlke, 2000; Harvey and Gooseff, 2015; Kipfer et al., 2002; Purtschert, 2008). Natural tracers that have been successfully applied in SW-GW systems independently or in multi-tracer studies are summarized in **Figure 5.1**.

Even though different natural tracer methods are available, a critical gap in the coverage of residence timescales can be identified: There is a gap between the short-lived radioactive tracer ^{222}Rn (half-life = 3.82 days), which is suited for very short measurement timescales from days to two weeks, and the other available tracer methods, which are suited for timescales of multiple months to millennia. The intermediate timescale from a few days to a few months, however, is often the most relevant for a robust characterization of the sources of pumped drinking water in alluvial SW-GW systems. One particular technological development, however, has the potential to bridge this gap: The very rare radioactive tracer Argon-37 (^{37}Ar) is now measurable at natural levels, and has an ideal half-life of 35.1 days for the measurement of intermediate timescales (Loosli et al., 2000; Riedmann and Purtschert, 2011). **Figure 5.1** provides an overview of tracers used for SW-GW systems with their corresponding dating ranges. The intermediate time scale of weeks to months can potentially be cover by ^{37}Ar. The ^{37}Ar concentration in the atmosphere is very low. However, neutron and muon induced nuclear reactions produce ^{37}Ar in soil gas and GW (Riedmann and Purtschert, 2011), where it accumulates and can potentially be used for dating in a similar way as ^{222}Rn but has not been done before.

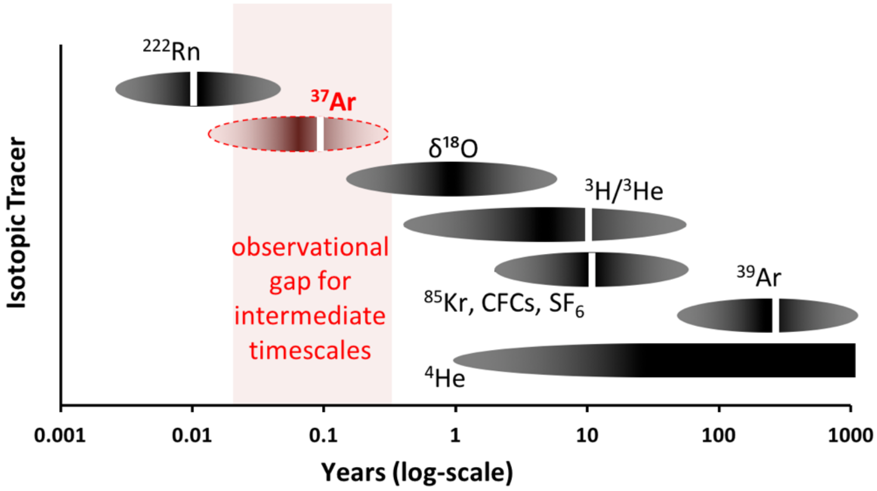


Figure 5.1: Tracer timescales after Purtschert (2008) with the addition of ^{37}Ar. The black ellipses indicate the residence timescale to which each tracer is sensitive. The white lines mark the half-life of the respective radioactive tracer. ^{37}Ar has ideal properties to close the existing gap in residence timescales coverage of intermediate timescales (red area) between the already covered short timescales, which are covered by ^{222}Rn, and long timescales, which are covered by δ^{18}O, ^{3}H/^{3}He, ^{4}He etc. In this study, ^{37}Ar is applied for the first time in the context of SW-GW interactions.

Multi-tracer studies, where tracers of different properties and with different measurement timescales are combined, have proven to be highly suited for the characterization of such complex alluvial SW-GW systems (Åkesson et al., 2015; Althaus et al., 2009; Gardner et al., 2011; Mayer et al., 2014; McCallum et al., 2014a; McCallum et al., 2014b). Except for a few rare cases (e.g., Delsmann et al. (2016), Hunt et al. (2006)), however, solely simple 1-D and lumped parameter models are used in combination with multi-tracer studies. However, such models are strongly limited for the analysis of SW-GW exchange fluxes compared to the complete analyses facilitated by physically-based models (Turnadge and Smerdon, 2014).

The goal of this study is to explore the potential of new and advanced tracer techniques for the application to physically-based flow modelling, and to test ^{37}Ar as a new tracer for intermediate residence timescales. This research is motivated by both practical requirements as well as unresolved scientific questions. From a practical point of view, it is important (1) to characterize the mixing and residence times in alluvial drinking water wellfields, and (2) to improve the simulation of exchange fluxes in flow models used for the prediction of these systems, as exchange fluxes are of pivotal importance for the safety of the pumped drinking water. From a scientific point of view, (1) new tracers are direly needed for intermediate residence timescales from a few days to multiple months, which are the most important timescales for alluvial drinking water stations, and (2) ideal observation types to better inform the choice of appropriate models for the simulation of drinking water wellfields need to be defined.

We want to address both the practical and scientific questions through the combination of a multi-tracer study and fully-coupled, physically-based SW-GW modelling of a highly dynamic alluvial drinking water wellfield under a controlled transient forcing. For this purpose, a numerical flow model of an alluvial drinking water wellfield was first matched against observations of hydraulic heads, and, subsequently, multiple scenarios of K_{rb} and n_{aq}, which all fit the hydraulic head observations equally well, were tested and compared for travel times of infiltrated SW and the mix between SW and GW that is pumped in the drinking water wells. These simulations were then compared to environmental-tracer-based estimates of travel times and mixing. For this purpose, ^{222}Rn, ^{3}H/^{3}He and noble gases, were measured alongside ^{37}Ar – which represents a new tracer in this context – to quantify the interactions in the alluvial drinking water system. The multi-tracer study was carried out during a controlled transient manipulation (40% reduction in GW pumping) to an important alluvial drinking water wellfield in Switzerland. This provided the opportunity of analyzing the transient behavior of the alluvial SW-GW system, and of augmenting the information content of the tracer measurements.

5.2. Material & methods

5.2.1. Theory of natural environmental tracer methods

5.2.1.1. The ^{222}Rn dating method

The radioactive isotope ^{222}Rn with a half-life ($T_{1/2}$) of 3.82 days is produced in the soil via the decay-chain of ^{238}U (Cecil and Green, 2000). The gradual accumulation of ^{222}Rn in GW can be used as a tracer for residence times and was first demonstrated by Hoehn et al. (1992). It is now considered as an established method to estimate groundwater-surface water exchange fluxes (e.g., Bourke et al. (2014), (Cecil and Green, 2000), Vogt et al. (2009), Harvey and Gooseff (2015)). According to Hoehn et al. (1992), the ingrowth of ^{222}Rn in GW can be described by the following equation:

$$AC_{^{222}\mathrm{Rn}}(t) = AC_{^{222}\mathrm{Rn,eq}}\,(1 - e^{-\lambda_{^{222}\mathrm{Rn}} t}) \tag{5.1}$$

with t representing time. $\lambda_{^{222}\mathrm{Rn}}$ is the decay constant (0.182 days^{-1}), $AC_{^{222}\mathrm{Rn,eq}}$ the activity concentration at equilibrium, and $AC_{^{222}\mathrm{Rn}}(t)$ the measured activity concentration of ^{222}Rn. The ^{222}Rn-based residence time (also "apparent age") of groundwater can be estimated by measuring the activity concentration of ^{222}Rn in GW, and solving for t:

$$t_{^{222}\mathrm{Rn}} = \lambda_{^{222}\mathrm{Rn}}{}^{-1} \cdot \ln(1 - \frac{AC_{^{222}\mathrm{Rn}}(t)}{AC_{^{222}\mathrm{Rn,eq}}}) \tag{5.2}$$

After ~3 half-lives of ^{222}Rn, residence times cannot be further differentiated based on ^{222}Rn measurements, as the activity concentration becomes indistinguishable from the equilibrium activity concentration within measurement uncertainty (e.g., Cecil and Green (2000)). The dating range covered by measurements of ^{222}Rn is compared to that of other tracers in **Figure 5.1**.

5.2.1.2. The ^{37}Ar dating method

^{37}Ar ($T_{1/2}$ = 35.1 days) is a very rare radioactive isotope of Argon (Loosli and Purtschert, 2005). However, similarly to ^{222}Rn, ^{37}Ar is only produced in the subsurface in significant quantities. In shallow soil up to 10 meters, activation of Ca by cosmic neutrons is the dominant reaction channel (40Ca(n, α)^{37}Ar (Fabryka-Martin, 1988)). At greater depths, muon capture reactions ^{39}K(μ^-,2n)^{37}Ar and ^{40}K(μ^-,3n)^{37}Ar become increasingly dominant (Fabryka-Martin, 1988; Johnson et al., 2015; Riedmann and Purtschert, 2011; Riedmann and Purtschert, 2016). Both pathways lead to an exponentially decreasing production rate of ^{37}Ar with depth, which is characterized by the attenuation length ℓ. The attenuation length ℓ depends mainly on the bulk density of the soil (including water and soil gas) (Guillon et al., 2016). In case of predominantly horizontal GW flow and full saturation of the porous media, advective and diffusive vertical transport is negligible and the exponential production profile is maintained in the GW ^{37}Ar activity concentration. ^{37}Ar activity concentrations are measured in mBq/L$_{\mathrm{Ar}}$ and converted in mBq/l$_{\mathrm{water}}$ by using the concentration of dissolved Ar in the groundwater (Riedmann and Purtschert, 2011).

At production-decay equilibrium the ^{37}Ar activity concentration as a function of depth d can be expressed by:

$$AC_{^{37}\mathrm{Ar,eq}}(d) = AC_{^{37}\mathrm{Ar,eq}}(0) \cdot e^{-d/\ell} \qquad (5.3)$$

Where $AC_{^{37}Ar,eq}(0)$ is the hypothetical activity concentration at $d = 0$ m. Diffusive loss in the saturated soil column is neglected here.

If this production-decay equilibrium activity concentration profile is known, similarly to ^{222}Rn-dating (Cecil and Green, 2000) ^{37}Ar can be used as a dating tool on timescales of weeks to months for SW that freshly recharged into an aquifer, as the ^{37}Ar activity concentration in GW is approaching the equilibrium activity concentration:

$$AC_{^{37}\mathrm{Ar}}(d,t) = AC_{^{37}\mathrm{Ar,eq}}(d) \cdot \left(1 - e^{-\lambda_{^{37}\mathrm{Ar}} \cdot t}\right) \qquad (5.4)$$

The decay constant $\lambda_{^{37}\mathrm{Ar}}$ is 0.0197 day^{-1}. Similar to ^{222}Rn, the apparent age information is obtained by the ingrowth from zero activity, which is the atmospheric starting point, to the equilibrium profile. This is constant with depth for ^{222}Rn in the case of a homogenous aquifer, but exponentially decreasing with depth for ^{37}Ar. The sensitive timescale of ^{37}Ar is defined by approximately 3 times the half-life of 35.1 days, which makes ^{37}Ar an ideal tracer for intermediate timescales of weeks to months (see **Fig. 5.1**). The ^{37}Ar-residence time is given by:

$$t_{^{37}\mathrm{Ar}} = -\lambda_{^{37}\mathrm{Ar}}{}^{-1} \cdot \ln\left(1 - \frac{AC_{^{37}\mathrm{Ar}}(d,t)}{AC_{^{37}\mathrm{Ar,eq}}(d)}\right) \qquad (5.5)$$

These calculations of ^{37}Ar residence times are based on two simplifying assumptions: First, it is assumed that the GW approx. remains at the same depth, i.e., that flow is approx. horizontal. Second, the aquifer is assumed to be homogeneous with respect to porosity and elemental composition. If the sampling well is screened over several meters, the depth where the equilibrium concentration is equal to the expected equilibrium concentration for the mixed sample is calculated with:

$$z_m = z_u + \ell \cdot (\ln(\ell \cdot (z_l - z_u)) - \ln(1 - e^{-/\ell \cdot (z_l - z_u)})) \qquad (5.6)$$

Where z_u and z_l are upper and lower limits of the screen, respectively.

5.2.1.3. *The ^{3}H/^{3}He dating method*

Tritium (^{3}H) is the radioactive isotope of hydrogen that – with a half-life of ~4500 days – decays to ^{3}He. It is naturally present in the earth's atmosphere and primarily produced through cosmic rays-induced decay of atmospheric nitrogen (Lucas and Unterweger, 2000; Solomon and Cook, 2000). ^{3}H concentrations are expressed in tritium units (TU), and the natural background concentration of ^{3}H in the atmosphere is between 1 and 6 TU (Solomon and Cook, 2000). In the well-established ^{3}H/^{3}He method, the apparent age of GW is estimated by comparing the measured concentration of ^{3}H in GW to the tritiogenic portion of the measured ^{3}He (e.g., Schlosser et al. (1988), Beyerle et al. (1999), Solomon and Cook (2000), Kipfer et al. (2002)). While the tritiogenic portion of ^{3}He (i.e., ^{3}He*) cannot be directly measured, ^{3}He* can be estimated based on noble gas measurements (e.g., Kipfer et al. (2002), Solomon and Cook (2000)). The apparent ^{3}H/^{3}He water age, or ^{3}H/^{3}He-residence time of GW, is defined as:

$$t_{^{3}H/\,^{3}He} = \lambda_{^{3}\mathrm{H}}{}^{-1} \cdot \ln\left(\frac{^{3}He^{*}}{^{3}H} + 1\right) \tag{5.7}$$

$\lambda_{^{3}H}$ is the decay constant of ^{3}H ($1.54 \cdot 10^{-4}$ days^{-1}). The ^{3}H/^{3}He method allows identifying residence times from approx. 1.5 months – 50 years (Harvey and Gooseff, 2015).

5.2.1.4. Analysis of noble gases

5.2.1.4.1. Mixing analysis based on Noble Gas Recharge Temperature

The equilibrium solubility of the stable atmospheric noble gases He, Ne, Ar, Kr and Xe can be used to estimate the water temperature during the moment of recharge (Aeschbach-Hertig and Solomon, 2013; Kipfer et al., 2002). This recharge temperature is commonly known as the noble gas recharge temperature (NGRT). The NGRT in turn allows estimating the time of recharge if the temperature chronicle in the recharge area is known. To calculate the NGRT from stable atmospheric noble gas concentrations, the Closed Equilibrium (CE) model was used in this study (Aeschbach-Hertig et al., 1999; Aeschbach-Hertig and Solomon, 2013; Kipfer et al., 2002).

If, in a given SW-GW system, there are GW components that have residence times of more than one year and, during this long residence time, have become relatively well-mixed, the NGRT of that GW component ($\mathrm{NGRT}_{\mathrm{GW,old}}$) will be close to the average annual air temperature in the recharge area. The NGRT of recently infiltrated SW ($\mathrm{NGRT}_{\mathrm{SW}}$), on the other hand, still maintains the signal of SW temperature at the time of infiltration. In a system where these two components are present, the NGRT can be used to infer mixing if the approximate recharge temperature of the freshly infiltrated SW can be determined through a residence time analysis. While NGRT do not mix linearly (Aeschbach-Hertig and Solomon, 2013; Kipfer et al., 2002), for the small NGRT variation encountered in this study, NGRT mixing can reasonably be approximated by linear mixing:

$$\mathrm{NGRT}_{\mathrm{GW}}(t) = x \cdot \mathrm{NGRT}_{\mathrm{GW,old}} + (1 - x) \cdot \mathrm{NGRT}_{\mathrm{SW}}(t) \tag{5.8}$$

5.2.1.4.2. Mixing analysis based on the 3He/4He isotopic ratio

The isotopic ratio of ^{3}He/^{4}He in GW can also be used as an indicator of two-component-mixture and residence times (Solomon, 2000). The atmospheric ratio of ^{3}He/^{4}He is $1.384 \cdot 10^{-6}$ and $1.36 \cdot 10^{-6}$ in air-equilibrated water (AEW). Multiple processes significantly alter this ^{3}He/^{4}He ratio in shallow GW: (a) terrigenic production of mainly ^{4}He, and (b) tritiogenic production of ^{3}He. If the estimated amount of tritiogenic ^{3}He (^{3}He* estimated as described in section 5.2.1.3) is subtracted from the measured ^{3}He/^{4}He ratio, the resulting $^{3}\mathrm{He}_{\mathrm{corr}}$/^{4}He ratio then only represents the atmospheric and terrigenic components of ^{3}He and ^{4}He. Because neither $^{3}\mathrm{He}_{\mathrm{corr}}$ nor ^{4}He are produced on the short time scales relevant in this study, $^{3}\mathrm{He}_{\mathrm{corr}}$/^{4}He can only change due to mixing of water with different $^{3}\mathrm{He}_{\mathrm{corr}}$/^{4}He ratios. Also the $^{3}\mathrm{He}_{\mathrm{corr}}$/^{4}He ratio can therefore be used to analyze mixing between very young and very old GW components in the system (i.e., between a recently infiltrated SW component with a $^{3}He_{corr}/^{4}He$ ratio still corresponding to the AEW ratio ($1.36\cdot10^{-6}$) ($\frac{^{3}He_{corr}}{^{4}He}(SW)$, and an old component with significantly accumulated ^{4}He resulting in a substantially lower $^{3}He_{corr}/^{4}He$ ratio ($\frac{^{3}He_{corr}}{^{4}He}(GW)$)) (Kipfer et al., 2002):

$$\frac{^{3}He_{corr}}{^{4}He}(t) = x \cdot \frac{^{3}He_{corr}}{^{4}He}(GW) + (1-x) \cdot \frac{^{3}He_{corr}}{^{4}He}(SW) \qquad (5.9)$$

5.2.2. Study site

5.2.2.1. Hydrology, hydrogeology and climate

The Upper Emme valley is a typical pre-alpine, alluvial catchment situated on the northern border of the Swiss Alps (see **Figure 5.2**). Since the last ice age, the valley bottoms in the lower part of the catchment filled up with coarse, quaternary alluvial gravel and sand, forming a highly conductive unconfined aquifer (Blau and Muchenberger, 1997). The catchment is spread over an altitude of 673 - 2221 m ASL, and covers an area of 194 km^2, which is drained by two streams, the Emme River and the Roethebach tributary, with an average discharge of 4.4 m^3/s and 0.7 m^3/s, respectively. These rivers are extremely dynamic and provide the main source of recharge to the alluvial aquifer of the Upper Emme valley (Blau and Muchenberger, 1997; Käser and Hunkeler, 2015). The lowest part of the catchment consists of the main Upper Emme valley with an average topographic gradient of 0.9 % (Käser and Hunkeler, 2015). Approximately 8 km upstream of the outlet, the tributary enters the main valley. The whole aquifer, which extends into the tributary valley, spans an approximate area of 6 km^2 (Käser and Hunkeler, 2015). In the area of the studied aquifer, the valley has a width of 200 - 400 m. The aquifer is limited underneath by impermeable sediments of the freshwater molasses. Due to the high yield and high water quality, the aquifer serves as an important drinking water resource for the capital of Switzerland (45% of the drinking water resources consumed in the region of the Swiss capital Bern (Biaggi et al., 2005)). The wellfield is situated on the Ramsei Plain, towards the outlet of the valley (see close-up in **Fig. 5.2**). The aquifer around the Ramsei Plain wellfield consists of approximately 20% sand and 80% gravel and has an average thickness of 25 m (Würsten, 1991). Two pumping tests at the locations of A25 and A26 (see **Fig. 5.2**) revealed average aquifer hydraulic conductivities (K_{aq}) between 200 and 500 m/d, with maximal values of more than 1350 m/d (Würsten, 1991). Two important parameters that control fluxes and mixing, the hydraulic conductivity of the riverbed (K_{rb}) and the porosity of the aquifer (n_{aq}), have not been measured and thus are unknown.

At the drinking water station the average annual precipitation is 1300 mm, the potential evapotranspiration 550 mm, and the average annual temperature is ~8 °C (Figura et al., 2011; Figura et al., 2013; Figura et al., 2015; Käser and Hunkeler, 2015; Würsten, 1991). In very dry summers and very cold winters, segments of the Emme may run completely dry (Würsten, 1991). With a currently changing climate and the anticipated rise of extreme weather events, the occurrence of floods and droughts is expected to increase (Chaix and Gander-Kunz, 2014; Frei et al., 2007). In such systems the interactions between SW and GW can be extremely dynamic, and it is of utmost importance to minimize the uncertainty of predictions of the systems' behavior in order to optimally manage the abstraction and quality of the drinking water resources. Käser and Hunkeler (2015) measured the water temperature and electrical conductivity in the stream directly above the riverbed near the drinking water wellfield, and identified alternating locations of losing and gaining conditions, indicating that there is a complex pattern of interaction between SW and GW.

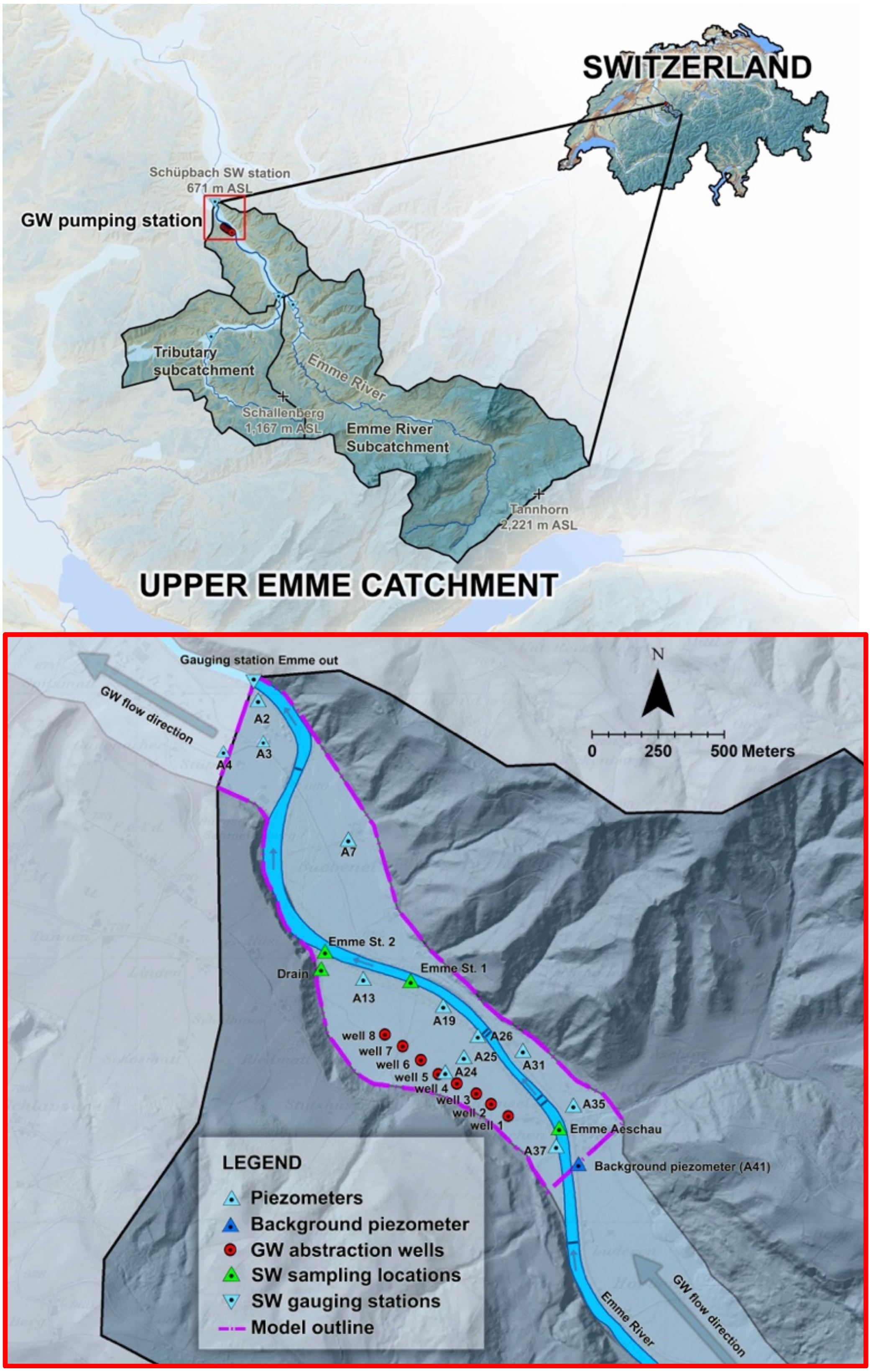


Figure 5.2: Maps of the Upper Emme catchment (upper image) and a close-up of the GW pumping station on the Ramsei Plain in the lower part of the catchment (lower image).

5.2.2.2. Drinking water abstraction

At the groundwater abstraction station, the water is pumped in roughly equal parts from 8 wells spaced at approximately 100 m and aligned in parallel to the river. The distance between the wells and the Emme River is 300 m towards the river bend marking the upstream end of the Ramsei plain, and 125 m in parallel to the wells. Water from the three upstream wells is pumped at a depth of 10 m, and at a depth of 15 m in the 5 downstream wells. In total, the groundwater abstraction station pumps 0.4 m^3/s. This GW abstraction is substantial relative to the total water balance of the system: the abstraction can amount to up to 50 % of the total outflow of SW and GW out of the valley (Käser and Hunkeler, 2015; Würsten, 1991).

5.2.2.3. Measurement network

A dense measurement network for the observation of a multitude of hydrological variables covers the entire alluvial valley. SW discharge is continuously monitored at a 10 min interval at 4 river gauging stations, with one station located at the outlet of the catchment, one station 1.5 km downstream of the wellfield, and two stations measuring the inflow from the Emme River and the tributary into the main valley, 5.5 km upstream of the wellfield. GW levels are continuously recorded at 10-15 min intervals in more than 30 piezometers. 14 of these piezometers are located in the immediate proximity of the drinking water wellfield (see Fig. 2). Three of these piezometers are equipped with wireless sensors, which store measurements in a cloud-based infrastructure (Kropf et al., 2014; Kurtz et al., (under review); Lapin et al., 2014). These piezometers are screened in the upper 10-15 m of the soil. A strategic multi-level piezometer (A41) located immediately upstream of the drinking water wellfield, just across the river, was installed to allow sampling of water also at higher depths than the other piezometers and the drinking water wells. The screened depths are 0-10 m, 11.5-13.5 m, 16-18 m and 21.5-23.5 m. Multiple weather stations that record precipitation and air temperature are located throughout the catchment.

5.2.3. Pumping experiment & sampling strategy

A large-scale and controlled transient pumping regime was implemented in collaboration with the water authorities of Bern (WVRB). The experiment was expected to produce (i) altered SW-GW interactions, and (ii) a change in the mix of water from different sources in the pumped water. During the 12 months prior to the experiment, the average abstraction rate Ramsei Plain drinking water station was more than 350 l/s and constant (herein referred to as the *maximum* or *max* pumping regime). This long period of maximal abstraction resulted in a hydraulic quasi-steady-state. A transient manipulation to this hydraulic quasi-steady-state was achieved through a reduction of the pumped GW by 40% for the period of one week. The aim was to create reduced hydraulic gradients and approach a new hydraulic quasi-steady-state. For 7 days, from 26 Jan. 2015 14:00 until 02 Feb. 2015 14:00, the pumping rate was reduced to the smallest technically possible rate of 225 l/s (herein referred to as the *minimum* or *min* pumping regime). The experiment took place during a low flow period at the end of Jan. 2015. During these winter periods, precipitation mainly falls as snow and SW discharge therefore remains approximately constant at a low level, which allows for optimal detection of SW-GW interactions.

The average SW discharge during the 7 days of reduced pumping was 1.4 m^3/s at the closest upstream gauging station and 2.3 m^3/s at the downstream gauging station of the Emme River. The average air temperature was 3.3 °C. Precipitation amounted to a total of 32.9 mm, but due to the cold temperatures was retained as snow on the floodplain; overland flow or infiltration on the floodplain were therefore negligible. The comparably small amount of snow that directly precipitated into river therefore only minimally influenced the SW discharge. To investigate the changes of SW-GW exchange rates, the pumping experiment was accompanied by SW dilution gauging tests at two locations close to the wellfield (upstream: Emme St. 1, downstream: Emme St. 2, see **Fig. 5.2**). The dilution-tests were carried out using fluorescein and the GGUN-FL30 flow-through field fluorometers (Schnegg, 2003).

In addition to continuous hydraulic and meteorological measurements, water samples were taken for analysis of the chemical composition and of different natural environmental tracers (^{222}Rn, ^{37}Ar, ^{3}H/^{3}He, noble gases) at key locations immediately before, during, and after the pumping experiment. The samples were taken at discrete locations in the river, in piezometers A13, A26, A25, A41, as well as in the pumping wells 1, 5 and 7. The GW in the multi-level piezometer A41 is assumed to represent the background end-member, e.g., the oldest possible GW in the studied domain, and was used to obtain depth profiles of ^{37}Ar, and ^{222}Rn.

The solid state alpha detector RAD7 (DURRIDGE, 2014) was used in conjunction with the RAD H2O accessory (DURRIDGE, 2012) to measure the activity concentration of ^{222}Rn in water. ^{37}Ar was measured in the Low Level Counting Laboratory of the University of Bern in Switzerland (Loosli and Purtschert, 2005). To measure the activity of ^{37}Ar 2- 3 tons of water per sample were degassed directly in the field in order to extract sufficient amounts of ^{37}Ar for subsequent laboratory analysis (Purtschert et al., 2013). From the extracted gas, Ar was separated by preparative gas chromatography (Riedmann and Purtschert, 2016) and then measured by low level counting (Riedmann, 2011). Noble gas concentrations, He isotopes and Tritium (^{3}H) were analyzed at the Noble Gas Mass Spectrometry Laboratory of the Swiss Federal Institute of Technology in Zurich, Switzerland, using standard measuring protocols (Beyerle et al., 2000).

5.2.4. Surface water-groundwater modelling

5.2.4.1. Modelling strategy

To estimate the mixing ratios between older GW and recently infiltrated SW (i.e., days to a few weeks ago) in the pumped water, both SW and GW flow were explicitly simulated in a physically-based and fully-coupled way, and the movement of water was tracked throughout the modelling domain. An efficient mixing-cell approach (Partington et al., 2011), which only requires the hydraulic solution of the flow simulator, was used to track the flow of water throughout the model domain, rather than relying on the explicit simulation of transport using the advection-dispersion equation. While potentially less precise, for the relevant temporal and spatial scales the loss in precision introduced through this simplification is negligible.

K_{aq} was calibrated against hydraulic head. To quantify the influence of K_{rb} and n_{aq} on the exchange fluxes and mixing ratios, multiple different scenarios of K_{rb} and n_{aq}, which all fit the hydraulic heads equally well, were tested. Subsequently, the models were analyzed for the sources of the pumped water and for travel times. To quantify the mixing ratio between the different sources of the pumped water, the pumped water was separated into the relative amounts of water that entered the model through the GW boundary conditions versus water that entered the model through the SW boundary conditions. For the analysis of travel times, the mean travel time of the surface water component that is pumped by the abstraction station was used, as this recently infiltrated SW component is the most relevant for the management of drinking water quality. The mean travel time of the pumped SW component between the river and the wells was defined as the time required in the quasi-steady-state models until the SW component has reached 50% of its final value in the pumped water.

The system was simulated both in steady state and transiently. Rather than true steady state simulations, which omit temporal dependency, quasi-steady-state simulations that also allow the analysis of travel times were carried out. Quasi-steady-state simulations constitute simulations with constant forcings for durations that allow a steady state to establish, i.e., a state where hydraulic heads, exchange fluxes, mixing and residence times do not change anymore. Using both steady state and transient models allowed analyzing both the transient short-term, as well as, the long-term reactions of the mixing ratios and travel times to a change in pumping. Both states are relevant for the management of drinking water abstraction. Steady state simulations were first carried out for the maximum pumping regime that prevailed until the beginning of the pumping experiment. Subsequently transient simulations were carried out for the duration of the pumping experiment with the steady state results as initial conditions, which allowed simulating the transient reaction of the system to a change in pumping. And finally, steady state simulations were carried out for the minimum pumping regime and hydraulic states encountered at the end of the pumping experiment, in order to analyze the potential long-term reaction of the system to reduced pumping. The steady state simulations correspond to: (1) the *maximum* (or *max*) pumping regime with measured hydraulic forcings from 21 Jan. 2015 00:00, and (2) the *minimum* (or *min*) pumping regime with measured forcings from 02 Feb. 2015 00:00. Transient simulations covered the duration of the pumping regime and consisted of hourly changing boundary conditions. The transient simulations started off the *maximum* steady state (i.e., 21 Jan. 2015 00:00), and finished four days after the end of pumping experiment, on 06 Feb. 2015 23:59.

Mixing and travel time analyses were carried out for (1) the *maximum* steady state, (2) the transient state immediately before the end of the minimum pumping experiment on 02 Feb. 2015 12:30 (from now on denoted as the state *trans*), and (3) the minimum steady state. The three different states *maximum*, *trans* and *minimum* were subsequently compared to tracer measurements and used to identify an appropriate model among the different scenarios tested.

5.2.4.2. *Model setup*

5.2.4.2.1. *Numerical simulator*

A fully-integrated SW-GW model was developed to simulate the hydraulic heads, SW discharge, exchange flow patterns, drinking water abstraction, mixing, and mean travel times of the alluvial drinking water system, as well as, allowing the direct comparison of the measured isotopic data with numerical simulations. The model HydroGeoSphere (HGS (Brunner and Simmons, 2011; Therrien et al., 2010)) was chosen in conjunction with the Hydraulic Mixing-Cell flow tracking tool (HMC, (Partington et al., 2013; Partington et al., 2011)). The HMC method utilizes the flow solution of the HGS model, tags any inputs to the model domain as specified by the boundary conditions and then tracks the fraction of these inputs in each of the model cells. Previous applications of the HMC method (Partington et al., 2013; Partington et al., 2011; Partington et al., 2012) have been restricted to the surface domain in the context of elucidating streamflow generation mechanisms; however, in this study it was applied to the subsurface for the first time. HGS was chosen as it allows the simulation of all the relevant SW-GW processes in a fully-integrated way, and has been demonstrated to be an ideal tool for the simulation of complex alluvial SW-GW interactions (e.g., Fleckenstein et al. (2006), Banks et al. (2011), Schilling et al. (2014), Schilling et al. (under review)).
HGS uses the Richards equation and the van Genuchten parametrization to simulate unsaturated subsurface flow. The Richards equation links the subsurface flow q to the hydraulic conductivity K [L/T], the relative hydraulic conductivity $k_r(\psi)$ [-], the pressure head (ψ [L]) and the elevation head (z [L]), as well as, the saturation S_w [-] of the porous medium:

$$q = -K \cdot k_r(\psi)\nabla(\psi + z) \quad (5.10)$$

As on the studied field site the river and the aquifer are hydraulically connected, saturated flow dominates and k_r is equal to 1. An important aspect of HGS for the study site is the physically-based simulation of both SW and GW flow and the direct coupling between SW and GW flow processes. The transfer between the two flow domains is defined as:

$$d_0\Gamma_0 = k_r K_{aq,zz} \frac{(H_{GW} - H_{SW})}{l_{exch}} \quad (5.11)$$

d_0: SW depth
Γ_0: volumetric exchange flux [$L^3L^3T^{-1}$] (positive for flow from the subsurface into the surface system)
$K_{aq,zz}$: vertical saturated hydraulic conductivity of the aquifer
H_{GW}: subsurface water head [L]
H_{SW}: surface water head [L]
l_{exch}: coupling length [L]

5.2.4.2.2. Topography of the aquifer and the aquitard

High resolution topographical maps of the floodplain (swisstopo, 2010), the aquifer confining layer (AWA, 2013), and the riverbed were used to define the boundaries of the 3-D model. The riverbed topography of the dry as well as the immersed sections was specifically obtained for this study through photogrammetric analysis (e.g., Feurer et al. (2008)) obtained with a drone on 20 Mar. 2015 (resolution of 0.25 m).

5.2.4.2.3. Numerical setup

Following the guidelines of Käser et al. (2013) for mesh generation of alluvial systems, an approximately equilateral triangular mesh with an average side length of 17.5 m for the floodplain and 8.5 m for the river, resulting in 10983 elements per layer, was generated with GridBuilder (McLaren, 2011). Vertically, the model consisted of 15 layers, covering the entire vertical aquifer extent: the top 5 layers covered 0.61 % each, the next 4 layers 6.1 %, and the last 5 layers 12 %. At the location of the largest vertical extent (46 m) this resulted in layers of 0.28 m, 2.8 m and 5.5 m thickness. The quasi-steady-state solution was obtained by simulating 2500 days with constant forcings. The Newton absolute and residual convergence criteria were set to 10^{-5}.

5.2.4.2.4. Choice of parameters

The aquifer was conceptualized as a homogeneous gravel-sand aquifer with a van Genuchten α of 3.48 m^{-1}, a van Genuchten β of 1.75 and a S_{wr} of 0.05 (Li et al., 2008). Due to the limited number of pumping tests, K_{aq} had to be calibrated against classical observations of hydraulic heads measured in piezometers A4, A7, A19, A24, A25, A26, A31, A35 and A37 in an automated calibration of the *maximum* model using PEST (Doherty, 2015). A K_{aq} of 550 m/d was found to best reproduce the measured hydraulic heads. Values of the porosity of the aquifer (n_{aq}) were not available, and since n_{aq} is strongly correlated to K_{aq} towards reproducing flow (e.g., Anderson et al. (2015), Voss (2011)), K_{aq} and n_{aq} cannot simultaneously be calibrated against changes in hydraulic heads. Therefore, three different n_{aq} scenarios, which span typical values for gravel-sand aquifers documented in literature, were tested in combination with the calibrated K_{aq}: (i) a value of 0.43 (Li et al., 2008), (ii) a value of 0.2 according to (Fetter, 2001), and (iii) a value of 0.1 (Anderson et al., 2015).

Within the river, l_{exch}, which defines the virtual distance between the dual-nodes (**Eq. (5.11)**), was set to a very small value of 0.001 m in order to couple the SW and the GW as directly as possible. A slightly higher value of 0.1 m was defined for the floodplain in order to reduce the computational burden. Due to the very rough surface of the riverbed, with a mix of boulders, rocks, gravel, sand and vegetation, a high flow resistance was incorporated by a high value for Manning's n ($1.7 \cdot 10^{-6}$ d/m$^{1/3}$). The roughness of the floodplain was set to $8.1 \cdot 10^{-5}$ d/m$^{1/3}$. For the riverbed and the floodplain, a rill storage height of 0.1 m was used. For a more detailed explanation of these parameters the reader is referred to the manual of HGS (Therrien et al., 2010).

An assessment of the potential for unsaturated areas to develop underneath the Emme River according to Schilling et al. ((under review)) revealed that unsaturated areas are unlikely. Simplification of the riverbed to a homogeneous layer is therefore justified (Irvine et al., 2012). The riverbed was thus defined for the 4 topmost layers underneath the Emme River, and was conceptualized as homogeneous gravel-sand layer with the same soil properties as the aquifer, but a fixed riverbed porosity (n_{rb}) of 0.41. Automatic calibration of K_{rb} alongside K_{aq} with the *maximum* model revealed that K_{rb}, within a range of approx. 0.24 and 24 m/d, does not significantly influence the simulated hydraulic heads. To quantify the influence of different K_{rb} on mixing and residence times, multiple K_{rb}-scenarios were simulated (i.e., 0.24 m/d, 2.4 m/d, and 24 m/d).

5.2.4.2.5. Boundary conditions

Precipitation was corrected for potential evapotranspiration, which was calculated based on solar radiation and temperature measurements after Spreafico and Weingartner (2005). This resulted in 0 mm/d for the *maximum* regime, and 5.7 mm/d for the *minimum* regime. For the transient simulations, hourly changing values from measurements of a nearby meteorological station were used. Due to the almost negligible amount of precipitation, precipitation was directly simulated as a rain (specified flux) boundary condition (BC) rather than incorporating the computationally more demanding simulation of snow. The upstream BC for GW was split into two first-type (constant head) BCs, one underneath the floodplain that corresponded to the values measured in the background piezometer A41 (*max*: 690.969 m ASL, *min*: 691.167 m ASL), and one underneath the river with the floodplain BC value + 1 m, which corresponds to approx. 0.1 m water depth in the river, ensuring losing conditions in the upstream part of the model. A constant head BC with the measured value of piezometer A3 (*max*: 671.715 m ASL, *min*: 671.728 m ASL) was fixed on the downstream end of the model domain. The SW inflow on the upstream side of the model was conceptualized second-type (constant flux) BC that corresponded to the sum of the inflows of the Emme River and the tributary measured at the closest upstream gauging stations (*max*: 1.702 m^3/s, *min*: 1.425 m^3/s). A small creek entering the Emme River approx. 200m downstream of the upstream BC was implemented as a constant flux BC on the river bank. This little creek is ungauged, and was approximated by a value that corresponds to 25% of the tributary discharge (*max*: 0.086 m^3/s, *min*: 0.0075 m^3/s). The SW outflow was implemented as a critical depth BC. The 8 pumping wells were implemented as constant flux BC's. The pumping rate of the abstraction station is controlled separately for wells 1-4 and wells 5-8. Therefore, during the *maximum* pumping regime, the extracted amount from wells 1-4 was 4230 m^3/d per well and 3430 m^3/d per well for wells 5-8. For the *minimum* regime, the extracted amounts were 2250 m^3/d per well for wells 1-4 and 2630 m^3/d per well for wells 5-8. The values that were used for the BCs of the transient simulations are provided in the results.

5.3. Results

5.3.1. Classical observations: GW levels, temperature and SW discharge

The recorded pumping rates, water temperature, GW levels and SW discharge at the different measurement locations are illustrated in **Figure 5.3**. The SW inflow of the Emme and the tributary was comparably stable during the pumping experiment (26 Jan. 2015 14:00 – 02 Feb. 2015 14:00), and varied slightly around an average of 1.4 m^3/s. The SW outflow of the catchment averaged at 2.29 m^3/s during that period, indicating gaining conditions overall. More specifically, the dilution tests (not shown in the figure) carried out during the *maximum* pumping regime indicated strongly gaining conditions with an average increase in SW discharge between Emme St. 1 and Emme St. 2 of 30%. During the *minimum* pumping regime, the average increase of SW discharge between the two stations rose to 50%, indicating a more exfiltration of GW into the river in reaction to reduced pumping.

The GW levels also changed substantially and rapidly in response to a change of the pumping rate. The change in GW levels relative to the levels of 17 Jan. 2015 00:00 is shown in **Figure 5.3c**. GW levels rose by 1 m and more in the pumping wells, and up to 0.5 m in the sampled piezometers. The GW levels stabilized within a few days, and got lower as quickly as they rose after the end of the minimum pumping experiment. Even the GW levels in the background piezometer A41, approx. 350 m upstream of well 1 and on the far side of the Emme River reacted with a rise of approx. 0.4 m in response to a reduction in the pumping rates. Piezometers as far as 1250 m downstream (A2, A3 and A4) reacted with a rise of approx. 0.1 m.
Water temperature measurements suggest (**Fig. 5.3d**) that the water mix in piezometer A37 throughout the whole experiment, as well as the water in piezometer A26 during the *maximum* pumping regime, was governed by infiltrating SW. The temperatures at all other GW measurement locations correspond to the average annual air temperature of the catchment (~8°C (Käser and Hunkeler, 2015)), which is typical for shallow GW.

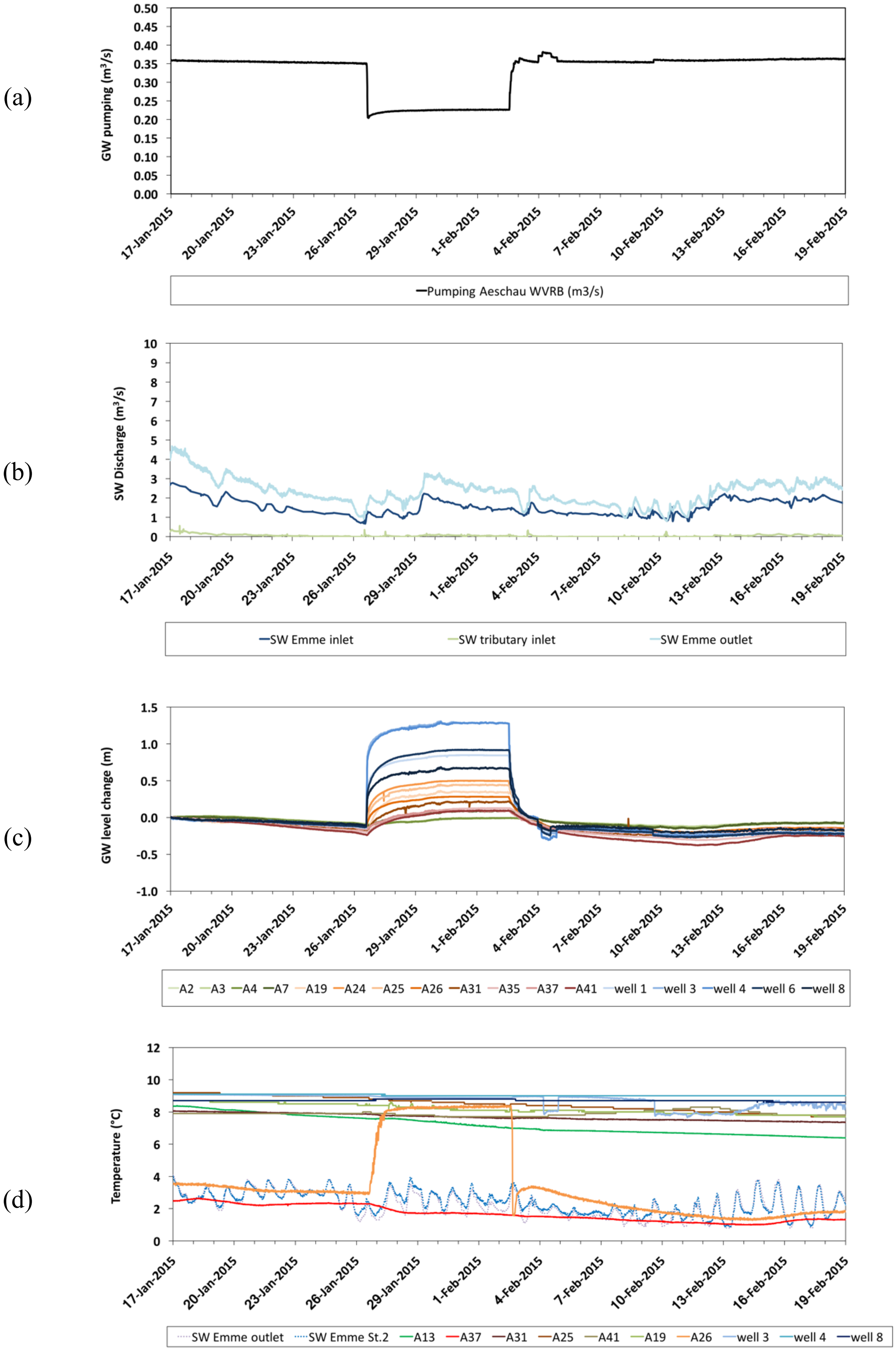


Figure 5.3: Measured classical observations during the period of the pumping experiment: the pumping rate (a), the SW discharge at three different measurement stations (b), and the hydraulic heads throughout the study area (c). GW level changes are given relative to the levels of 17.01.2015 00:00.

5.3.2. Results of the SW-GW simulations

The results of the quasi-stead-state simulations of the *maximum* and *minimum* pumping regimes of the n_{aq} = 0.43 scenario are shown in **Figure 5.4**. Other scenarios are not shown. Overall, these simulations showed that the predominant flow direction in the aquifer is horizontal, with two exceptions: (1) where river water infiltrates into the aquifer, the initial movement is vertical until the water arrives at the average depth to water table, and (2) immediately upstream of the pumps where the movement of water has a considerable vertical component (see **Fig. 5.4**). For (2), the largest vertical movement (~10 m) can be observed for the river water that infiltrates and is subsequently pumped. However, except for the water that is pulled towards the pumps from the river, the general flow direction is horizontal. In all simulations, SW-GW exchange fluxes reveal interchanging sections of gaining and losing conditions. In the upstream part of the model, until the Emme River reaches Y = 195,900 (see **Fig. 5.4**), the exchange fluxes are dominated by two weirs (indicated by the strong black lines crossing the river): directly upstream of the weirs losing conditions dominate, while directly downstream of the weirs gaining conditions prevail. In the river bend that limits the downstream part of the drinking water wellfield, gaining conditions dominate, and are influenced by the strongly variable riverbed topography. The simulated increase in SW discharge matches the observed gaining conditions between Emme St. 1 and Emme St. 2 (see **Fig. 5.2**). The exchange fluxes vary much stronger as a function of K_{rb} than in response to a change of the pumping regime; and while the pattern of exchange fluxes only shifts slightly, the magnitude of exchange is substantially different as a function of K_{rb}. The flow paths in the subsurface, on the other hand, are strongly controlled by the pumping regime: while during *maximum* pumping the subsurface water flow direction tends to strongly bend towards the pumping wells, under *minimum* pumping the flow direction is predominantly parallel to the valley.

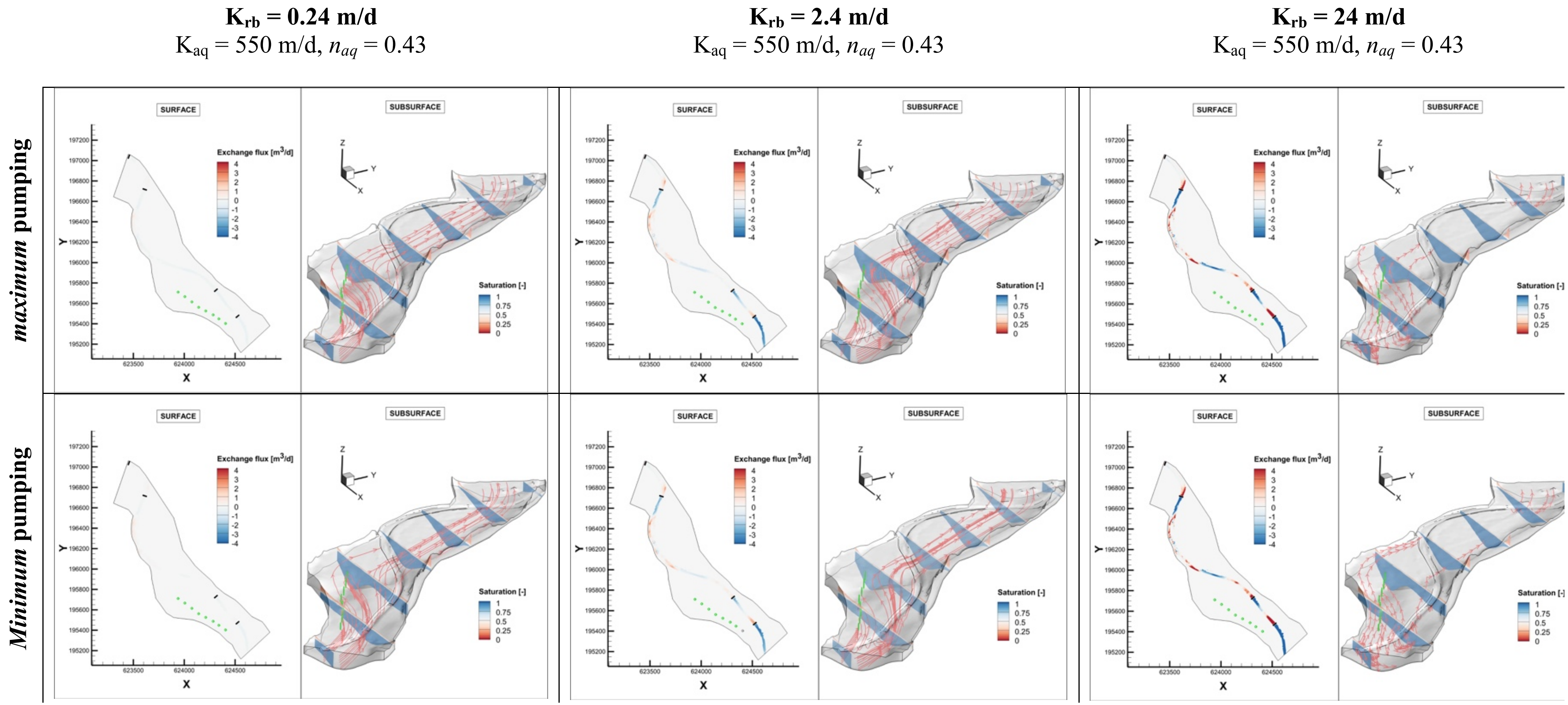


Figure 5.4: Illustrated are the SW-GW exchange fluxes in the river on the surface domain (left hand side of each figure) and the subsurface saturation and flow paths (right hand side of each figure). Exchange flu represent linearly interpolated nodal values, which correspond to the flux per nodal polygon in m^3/d. Positive exchange fluxes (indicated in red) represent gaining conditions, and negative exchange fluxes (indicat in blue) represent losing conditions. The quasi-steady-state solution for the maximum pumping condition is shown in the top row, the steady state solution for the minimum pumping condition in the bottom ro Weirs are indicated by the solid black lines. The pumping wells are indicated in green. North is indicated by the positive y direction. The predominant flow direction for both SW and GW is North-West. Compar the low permeability K_{rb} = 0.24 m/d to the high permeability K_{rb} = 24 m/d scenario reveals that the exchange fluxes are heavily dependent on the permeability of the riverbed, with much larger exchange dynam in a high permeability scenario. Moreover, subsurface flow lines not only react strongly the changes in the pumping rate, but also to changes in the riverbed permeability.

The different simulations also indicate a strong dependency of the pumped water mix on K_{rb} and to a lesser extent on n_{aq} and on the pumping rates. For the three different simulated states *maximum*, *trans* and *minimum*, the contribution of recently infiltrated SW to the pumped water mix in wells 1, 5 and 7 are shown in **Tab. 5.1** for all K_{rb} and n_{aq} scenarios. The recently infiltrated SW water component in the pumped water mix represents water that entered the model domain from the upstream (specified flux) SW boundary condition, and which subsequently infiltrated from the river into the aquifer within the model domain. The remaining part of the pumped water mix is made up by older GW, that is, water that entered the model domain from the upstream (constant head) GW boundary conditions. Additionally, a summary of the mean travel times of the pumped SW component is provided. The Root Mean Square Error (RMSE) between simulated and observed GW levels in the measured piezometers (A3, A4, A7, A19, A24, A25, A26, A31, A35 and A37) is also indicated.

Table 5.1: Overview of the contribution of recently infiltrated SW to the pumped water mix, the RMSE between simulated and observed hydraulic heads (i.e., $RMSE_{GW}$ for wells A3, A4, A7, A19, A24, A25, A26, A31, A35 and A37), and the travel times of the pumped recently infiltrated SW component for the three sampled wells 1, 5 and 7. Results are given for the two quasi-steady-state models (*max* & *min*) and the transient model (*trans*), and separated into the three K_{rb} and three n_{aq} scenarios. The mean travel time of the SW component in the pumped water represents the time at which the SW component has reached 50% of its final contribution to pumped water. Travel time analysis was only carried out for quasi-steady-state models. Since the $RMSE_{GW}$ is independent of the value of n_{aq} in quasi-steady-state, it is therefore represented by one single value. The same is also the case for the SW contribution to the pumped water mix.

K_{rb} scenario	**regime**	**$RMSE_{GW}$**			**Well**	**Mean travel time of the SW component in the pumped water**			**% SW in the pumped water mix**		
		n_{aq} = 0.1	**n_{aq} = 0.2**	**n_{aq} = 0.43**		**n_{aq} = 0.1**	**n_{aq} = 0.2**	**n_{aq} = 0.43**	**n_{aq} = 0.1**	**n_{aq} = 0.2**	**n_{aq} = 0.43**
K_{rb} = 0.24 m/d	*max*	0.5 m			**well 1**	15 d	30 d	40 d	2%		
					well 5	50 d	65 d	80 d	10%		
					well 7	60 d	75 d	100 d	21%		
	trans	0.44 m	0.46 m	0.49 m	**well 1**				1%	1%	2%
					well 5				7%	8%	9%
					well 7				19%	20%	22%
	min	0.42 m			**well 1**	30 d	35 d	45 d	1%		
					well 5	65 d	75 d	95 d	6%		
					well 7	80 d	95 d	140 d	14%		
K_{rb} = 2.4 m/d	*max*	0.30 m			**well 1**	15 d	20 d	35 d	38%		
					well 5	40 d	50 d	75 d	70%		
					well 7	55 d	70 d	95 d	77%		
	trans	0.33 m	0.33 m	0.33 m	**well 1**				29%	31%	34%
					well 5				58%	64%	67%
					well 7				80%	79%	78%
	min	0.32 m			**well 1**	20 d	30 d	40 d	26%		
					well 5	50 d	65 d	85 d	51%		
					well 7	65 d	85 d	110 d	73%		
K_{rb} = 24 m/d	*max*	0.26 m			**well 1**	15 d	20 d	30 d	97%		
					well 5	40 d	50 d	70 d	98%		
					well 7	50 d	70 d	90 d	100%		
	trans	0.31 m	0.31 m	0.31 m	**well 1**				92%	94%	96%
					well 5				95%	96%	97%
					well 7				100%	100%	100%
	min	0.31 m			**well 1**	15 d	25 d	35 d	89%		
					well 5	45 d	60 d	80 d	90%		
					well 7	60 d	80 d	100 d	96%		

In terms of the RMSE of GW levels, the two scenarios with higher riverbed permeability (K_{rb} = 2.4 m/d and K_{rb} = 24 m/d) perform similarly and better than the low permeability scenario

(K_{rb} = 0.24 m/d). The mix of water that is pumped is entirely different between all three K_{rb} scenarios: The pumped water in the K_{rb} = 0.24 m/d scenario is completely dominated by GW, in the K_{rb} = 24 m/d scenario by SW, and in the K_{rb} = 2.4 m/d scenario the pumped water is made up of approx. equal amounts GW and SW (**Tab. 5.1**).

In terms of the pumped SW water component, the simulations generally show that the more upstream the location of the pumping well, the lower the SW component in the pumped mix and the higher the GW component. At first sight, this result might be counter-intuitive, as one would expect that the closer a well is located to the river, the more recently infiltrated SW water is pumped. The simulated results, however, indicate that there is a limited amount of GW available to satisfy the demand of the pumping wells. The results thus show that the more is pumped and the further downstream the well is located, the less water from the GW inflow BC is available, and the more freshly infiltrated SW from the SW inflow BC is drawn towards the pumps to satisfy the demand. Simulations of the *maximum* regime showed a substantially larger SW component in the pumped water mix compared to simulations of the *minimum* regime. The largest change in the pumped water mix in response to a reduction of pumping could be observed between the *maximum* to the *minimum* state of K_{rb} = 2.4 m/d scenario, where shifts of 12 %, 19% and 4% were simulated in wells 1, 5 and 7, respectively (**Tab. 5.1**). For the *trans* scenarios, the smallest porosity scenario (n_{aq} = 0.1) approaches these changes in the pumped water mix after just one week of reduced pumping the most. The two higher porosity scenarios result in significantly smaller changes after 7 days of reduced pumping compared to the *min* quasi stead-state.

In terms of the mean travel time of the SW component, the different simulated scenarios show that if more is pumped the hydraulic gradients are higher and the mean travel times are smaller (*max* vs. *min*). Overall, the simulations also indicate that there is a significant spread of travel times between the different pumping, K_{rb} and n_{aq} scenarios, despite the assumption of homogeneous conditions in the aquifer and the limited extent of the model domain. The mean travel time of the SW component varied between 15 and 45 days in well 1, between 40 and 95 days in well 5, and between 50 and 140 days in well 7 (**Tab. 5.1**). In general, smaller n_{aq} result in smaller the travel times and a faster reaction of the % of SW in the pumped water to a change of the pumping regime.

These modeling results reveal that the water mix, travel times and the dynamic reaction to reduced pumping differ considerably between the different K_{rb} and n_{aq} scenarios. K_{rb} strongly influences the mean travel times and the amount of SW in the pumped water mix, and n_{aq} strongly controls the time it takes for the pumped mix to react to the pumping. Most strikingly, this significant difference in model behavior is nearly non-existent in the standard evaluation criterion for SW-GW models: the $RMSE_{GW}$ of all models are very similar (of the two high-permeability scenarios almost identical). Both the values of K_{rb} and of n_{aq} therefore appear to be almost completely insensitive to a calibration against hydraulic heads. Purely relying on the $RMSE_{GW}$ as the model evaluation criterion would slightly favor the K_{rb} = 24 m/d model over the other models, however, natural tracer measurements are necessary to confirm whether this truly is the appropriate model structure.

5.4. Environmental tracer results

5.4.1. ^{222}Rn

GW samples for the analysis of ^{222}Rn were taken before, during, and after the pumping experiment. Activities of ^{222}Rn measured in the background piezometer (A41) exhibited a high average activity concentration of 16.0 Bq/l_{water}, but varied considerably with a 1σ-standard deviation of 2.3 Bq/l_{water}. Except for one sample, all GW samples were within the uncertainty limits of the background activity. This means that, except for one sample, all sampled GW had a ^{222}Rn-residence time of more than 12 days. The one GW sample that differed from this general observation was taken in well 1 one day after the beginning of the *minimum* pumping regime (8.0 ± 1.4 Bq/l_{water}). This sample appears to represent a pocket of younger water. With an average activity of 3.7 ± 1.3 Bq/l_{water} the SW samples taken at Emme St. 2 show a significant influence of GW, indicating gaining conditions. A slightly elevated ^{222}Rn activity (1.2 ± 1.1 Bq/l_{water}) could also be observed in the drain of the pumping field (see **Fig. 5.2**), which drains parts of the uppermost GW away from the wellfield. The SW samples further upstream correspond to the atmospheric zero activity of ^{222}Rn, which either indicates losing conditions or no SW-GW exchange. In summary, ^{222}Rn indicates GW residence times, or an apparent groundwater age, of at least 12 days at the sampled locations. At the SW measurement stations, ^{222}Rn indicates losing conditions upstream and gaining conditions downstream of the wellfield.

5.4.2. Analysis of noble gases and ^{3}H/^{3}He

The results of the ^{3}H and noble gas measurements, as well as the derived quantities are summarized in **Table 5.2**. The measured quantities were used to derive noble gas recharge temperatures, the different helium components, as well as the ^{3}H/^{3}He age according to the methods described in section 5.2.1. The derived values are also given in **Tab. 5.2** and form the basis of the following discussion. Due to problems during sampling or the gas extraction phase, the samples taken in A19 and in A41,0-10m, could not be properly analyzed.

NGRT varied between 3.8 and 7.4 °C. The apparent ^{3}H/^{3}He GW age varies between 0 and 7 years. Excess air is below 100% except for the samples taken in A13. The excess air, the NGRT, and the ^{3}H/^{3}He analyses all indicate that piezometer A13 is dominated by very recently infiltrated SW (^{3}H/^{3}He age = 0 years during *maximum* abstraction and 1.5 years during minimum *abstraction*, low NGRT, very high excess air). Its proximity to the Emme River apparently leads to a very rapid exchange with the SW. If less is pumped, less SW appears to infiltrate from the river and A13 seems to mix with some older GW, increasing the ^{3}H/^{3}He age from 0 to 1.51 years. A25 with a ^{3}H/^{3}He age of 7.1 years appears to contain the oldest water. However, A25 exhibits a very low recharge temperature despite the high apparent ^{3}H/^{3}He age, which makes the interpretation difficult.

In terms of NGRT, the background well A41 shows a high recharge temperature of 7.4 °C, which is close to the average annual air temperature in the catchment (~8 °C). Well 1 shows a significantly lower NGRT of 5.7 °C, indicating an influence of recently infiltrated SW. Wells 5 and 7 more closely correspond to the NGRT of the background well, revealing a stronger influence of older GW. In reaction to a change of the pumping rate, the NGRT remain relatively constant in wells 1 and 5. The change in NGRT is stronger in well 7, with higher recharge temperatures when less is pumped. The results of the ^{3}H/^{3}He analysis furthermore reveal that the sampled wells 1, 5 and 7 are generally controlled by older GW during the *maximum*, and younger GW during the *minimum* pumping regime. Moreover, during the *maximum* pumping regime, the apparent age of groundwater in the pumping wells is even higher than the water of the background piezometer (A41,22.5-23.5m), probably being a result of the inflow of deeper groundwater than the sampling depth of the background piezometer.

Table 5.2: Summary of the noble gases and ^{3}H measurements, and the derived quantities. NGRT represents the noble gas recharge temperature. Except for ^{3}H with a measurement error of ~3%, all other measurements are associated with a measurement error of less than 1.5%. The error of the ^{3}H/^{3}He water age was calculated according to (Visser et al., 2014). Measurements were made during the *maximum* pumping regime before the beginning of the minimum pumping experiment (*max*), as well as during and immediately before the end of the *minimum* pumping experiment (*min*). The time that has passed between the beginning of the minimum pumping experiment and the sampling is indicated. The samples taken after 7 days of *minimum* pumping serve as a basis for the comparison of the *trans* models.

				measured quantities												**derived quantities**								
location	**sampling**	**pumping regime**	**time into pumping experiment**	**Ar**	**^{36}Ar / ^{40}Ar**	**Kr**	**Xe**	**Ne**	**^{20}Ne / ^{22}Ne**	**He**	**^{3}He**	**^{3}He / ^{4}He**	**^{4}He**	**Ne / ^{4}He**	**^{3}H**	**NGRT**	**err. NGRT**	**excess air**	**^{3}He***	**err. ^{3}He***	**^{3}He$_{corr}$ / ^{4}He**	**err. ^{3}He$_{corr}$ / ^{4}He**	**^{3}H/^{3}He age**	**err. ^{3}H/^{3}He age**
				[ccSTP/g]	[-]	[ccSTP/g]	[ccSTP/g]	[ccSTP/g]	[-]	[ccSTP/g]	[ccSTP/g]	[-]	[ccSTP/g]	[-]	[TU]	[°C]	[°C]	[% Ne]	[ccSTP/g]	[ccSTP/g]	[-]	[-]	[y]	[y]
A41, 21.5-23.5m	22.01.15 12:31	*max*	-	4.02E-04	0.00338	9.33E-08	1.32E-08	2.33E-07	9.791	5.81E-08	8.10E-14	1.39E-06	5.81E-08	4.01	8.90	7.4	*0.3*	59.57	6.17E-15	*1.6E-15*	1.29E-06	*4.4E-08*	4.4	*1.3*
A13	22.01.15 14:10	*max*	-	4.10E-04	0.00338	9.82E-08	1.43E-08	2.08E-07	9.793	4.68E-08	6.27E-14	1.34E-06	4.68E-08	4.44	8.69	5.2	*0.5*	202.66	0.00	*0.00*	1.34E-06	*5.7E-08*	0.0	*1.7*
	02.02.15 10:55	*min*	*7 days*	4.17E-04	0.00338	9.98E-08	1.49E-08	2.10E-07	9.786	4.72E-08	6.62E-14	1.40E-06	4.72E-08	4.45	8.81	4.5	*0.2*	160.67	1.94E-15	*4.8E-16*	1.36E-06	*2.6E-08*	1.5	*0.8*
A25	02.02.15 12:05	*min*	*7 days*	4.19E-04	0.00338	1.01E-07	1.52E-08	2.14E-07	9.802	4.88E-08	7.71E-14	1.58E-06	4.88E-08	4.38	8.76	3.8	*0.1*	8.82	1.07E-14	*6.6E-16*	1.36E-06	*2.8E-08*	7.1	*0.8*
well 1	23.01.15 10:45	*max*	-	4.05E-04	0.00338	9.66E-08	1.41E-08	2.22E-07	9.801	5.30E-08	7.93E-14	1.49E-06	5.30E-08	4.18	8.28	5.7	*0.4*	15.12	8.69E-15	*2.8E-15*	1.33E-06	*7.1E-08*	6.3	*1.6*
	29.01.15 13:20	*min*	*3 days*	4.02E-04	0.00339	9.44E-08	1.40E-08	2.19E-07	9.778	5.27E-08	7.31E-14	1.39E-06	5.27E-08	4.16	8.68	6.5	*0.2*	69.26	4.00E-15	*9.1E-16*	1.31E-06	*3.2E-08*	3.0	*1.1*
	02.02.15 15:15	*min*	*7 days*	4.10E-04	0.00338	9.70E-08	1.41E-08	2.21E-07	9.798	5.27E-08	7.08E-14	1.34E-06	5.27E-08	4.19	8.51	5.7	*0.2*	76.13	1.75E-15	*1.0E-15*	1.31E-06	*3.6E-08*	1.4	*1.2*
well 5	23.01.15 12:30	*max*	-	3.96E-04	0.00338	9.26E-08	1.39E-08	2.22E-07	9.795	5.42E-08	7.79E-14	1.44E-06	5.42E-08	4.09	8.85	7.0	*0.4*	30.95	7.08E-15	*1.7E-15*	1.31E-06	*4.6E-08*	5.0	*1.3*
	29.01.15 16:14	*min*	*3 days*	4.04E-04	0.00339	9.42E-08	1.38E-08	2.24E-07	9.781	5.48E-08	7.73E-14	1.41E-06	5.48E-08	4.08	8.26	6.8	*0.1*	82.90	6.51E-15	*3.4E-16*	1.29E-06	*1.7E-08*	4.9	*0.6*
	02.02.15 14:00	*min*	*7 days*	4.00E-04	0.00338	9.32E-08	1.35E-08	2.23E-07	9.791	5.42E-08	7.39E-14	1.36E-06	5.42E-08	4.12	7.73	7.2	*0.0*	72.59	3.06E-15	*7.7E-17*	1.31E-06	*1.6E-08*	2.6	*0.4*
well 7	23.01.15 14:15	*max*	-	4.02E-04	0.00338	9.50E-08	1.38E-08	2.25E-07	9.795	5.37E-08	8.10E-14	1.51E-06	5.37E-08	4.20	7.85	6.3	*0.2*	17.89	8.62E-15	*1.5E-15*	1.35E-06	*4.9E-08*	6.5	*1.3*
	29.01.15 15:00	*min*	*3 days*	3.92E-04	0.00339	9.16E-08	1.34E-08	2.19E-07	9.789	5.31E-08	7.36E-14	1.39E-06	5.31E-08	4.13	8.21	7.5	*0.1*	46.20	3.64E-15	*4.8E-16*	1.32E-06	*2.5E-08*	2.9	*0.8*
	02.02.15 14:45	*min*	*7 days*	3.95E-04	0.00339	9.23E-08	1.34E-08	2.22E-07	9.787	5.30E-08	7.66E-14	1.45E-06	5.30E-08	4.19	8.35	7.4	*0.0*	47.19	5.88E-15	*1.5E-16*	1.34E-06	*1.7E-08*	4.4	*0.4*

The measured isotopic ratios of Ne/^{4}He are plotted against ^{3}He/^{4}He in **Figure 5.5**. The GW in the background piezometer (A41,21.5-23.5m) appears to have accumulated a significant amount of ^{4}He relative to its initial concentration, thus indicating that this water has resided in the subsurface for the largest amount of time of all sampled GW. The ^{3}H/^{3}He age of the background piezometer suggests that the GW at this location consist of a mix of a very old component with a high amount of ^{4}He and some slightly younger component which pushes the ^{3}H/^{3}He age towards a younger average age compared to the water being pumped in the wells during the maximum regime. The pumping wells generally show lower Ne/^{4}He ratios than the piezometers. Compared to ^{3}He/^{4}He, Ne/^{4}He appears to be largely insensitive to a reduction in pumping: while the Ne/^{4}He ratios remain stable, the wells react strongly to the pumping regime in terms of their ^{3}He/^{4}He ratio, changing from higher ratios during *maximum* to lower ratios during *minimum*. The changes in ^{3}He/^{4}He can therefore only be explained through a change in the ^{3}He* content of the water.

While for the *maximum* regime, the wells seem to have a ratio of ^{3}He/^{4}He that is significantly higher than the AEW ratio, therefore consisting of water that has resided in the subsurface for a longer amount of time, in the *minimum* regime the ratios become smaller and approach the ratio of AEW. This means that the source of the pumped water in the wells appears to be shifting towards more SW during *minimum* compared to during the *maximum* pumping regime. This is in agreement with the ^{3}H/^{3}He residence time estimates. The GW in A25 appears to be most enriched in ^{3}He relative to ^{4}He, but has not accumulated significant amounts of ^{4}He. While this is in agreement with the ^{3}H/^{3}He residence time estimates, the reason for this is unclear. A13 during *maximum* pumping shows the lowest enrichment of ^{3}He relative to ^{4}He and AEW, as well as, the lowest ratio of Ne/^{4}He. This again hints at a strong influence of recently infiltrated SW at this location. In accordance with the ^{3}H/^{3}He and NGRT results, during *maximum* pumping A13 is more controlled by recently infiltrated SW compared to during *minimum* pumping.

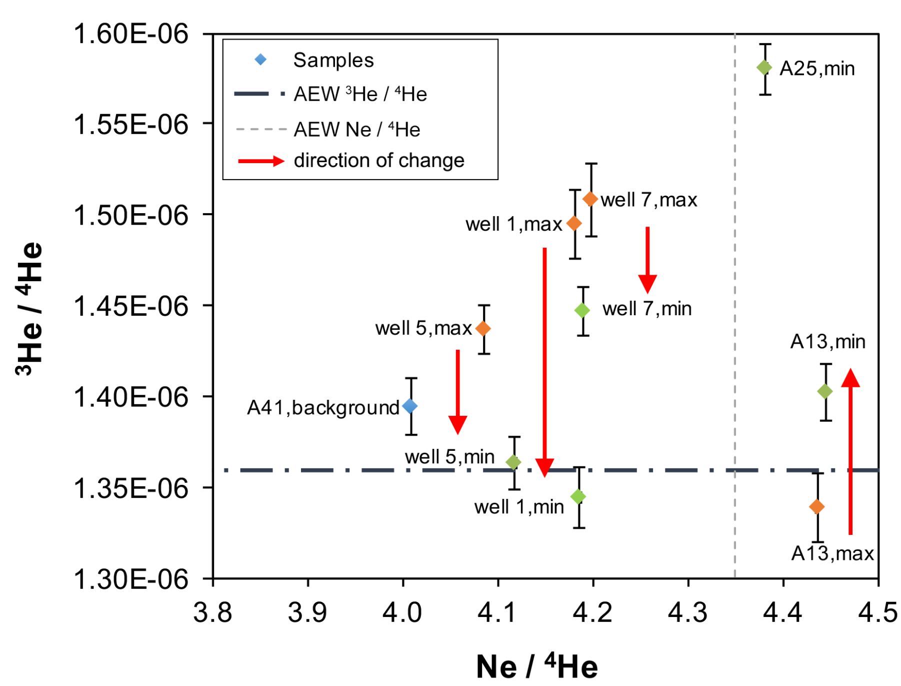


Figure 5.5: Plot of the measured ^{3}He/^{4}He versus the measured Ne/^{4}He ratios. For simplification, only the *maximum* and the measurements immediately before the end of the *minimum* regime (i.e., 7 days passed since the beginning of the pumping experiment) are indicated. The background sample is marked in blue, the *maximum* samples are marked in orange, and the *minimum* samples are marked in green. Reactions to a change in the pumping regime are indicated by arrows. Error bars represent 1σ-standard deviations. The reader is referred to **Tab. 5.2** for the complete list of measurements.

5.4.2.1. ^{37}Ar

The measured activity concentrations of ^{37}Ar are summarized in **Table 5.3,** and illustrated in **Figure 5.6** as a function of their sampling depth. The samples from the depth profile of the background piezometer A41 show an exponentially decreasing activity with increasing depth, which is in agreement with the theory of ^{37}Ar production. Based on the assumption that the water in the background piezometer has reached the production-decay equilibrium of ^{37}Ar , which is indicated by the significantly elevated ^{4}He concentrations, and that flow at that location is mainly horizontal, different residence times isolines (e.g., 1 week and 1 month) could be indicated in **Fig. 5.6**. The two pumping wells 1 and 7, which were sampled before and immediately at the end of the *minimum* pumping experiment, show only a very small, and according to the uncertainty limits not significant, increase in ^{37}Ar activity concentration in reaction to the pumping. The ^{37}Ar activity concentration of the Emme River in the downstream direction of A13 is considerably smaller compared to the GW samples, but not zero. This is in agreement with the findings of the ^{222}Rn measurements and confirms that GW is exfiltrating into the river at that location. The fact that the activity concentration for the samples in wells 1 and 5, as well as, in piezometers A13 and A25, is somewhere between the production-decay equilibrium and zero activity leads to the conclusion that in the sampled GW mix there is a substantial component of freshly infiltrated SW with a residence time between multiple days to 3.5 months.

Table 5.3: Summary of the ^{37}Ar measurement results. The corresponding production-decay equilibrium activity concentrations at the depth of each well are also given.

location	**sampling date**	**Pumping regime**	**Time into pumping experiment**	**$^{37}Ar_{eq}$**	**^{37}Ar**	**err. ^{37}Ar**
				[mBq/l$_{water}$]	[mBq/l$_{water}$]	[%]
A41, 0-10m	17.03.15	*max*	-	3.15E-03	2.95E-03	11%
A41, 11.5-13.5m	17.03.15	*max*	-	2.23E-03	2.23E-03	11%
A41, 16-18m	17.03.15	*max*	-	1.68E-03	1.45E-03	11%
A41, 21.5-23.5m	17.03.15	*max*	-	1.19E-03	1.28E-03	11%
Emme Aeschau	09.04.14	*max*	-	4.89E-03	1.36E-04	0%
A13	09.04.14	*max*	-	3.53E-03	2.36E-03	12%
A25	09.12.14	*max*	-	3.19E-03	8.70E-04	28%
A26	09.12.14	*max*	-	2.25E-03	1.90E-03	11%
A41, 21.5-23.5m	13.10.15	*max*	-	1.19E-03	1.49E-03	11%
well 1	23.01.15 10:45	*max*	-	2.26E-03	1.11E-03	12%
	02.02.15 15:15	*min*	7 days	2.26E-03	1.36E-03	11%
well 5	23.01.15 12:30	*max*	-	1.83E-03	1.04E-03	12%
	02.02.15 14:00	*min*	7 days	-	-	-
well 7	23.01.15 14:15	*max*	-	1.84E-03	1.89E-03	11%
	02.02.15 14:45	*min*	7 days	1.84E-03	1.98E-03	12%

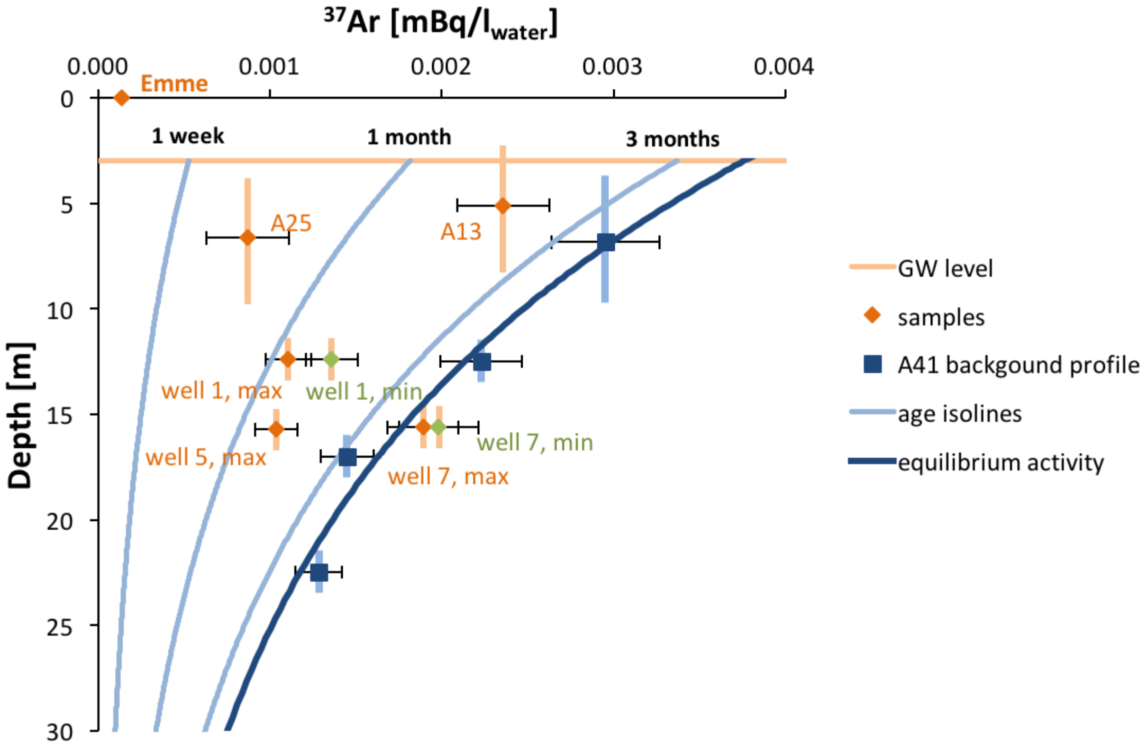


Figure 5.6: ^{37}Ar activity concentrations shown as a function of depth. Age isolines were derived from the ^{37}Ar background profile (A41) and indicate the apparent ^{37}Ar-residence times of 1 week, 1 month and 3 months. The green markers represent samples taken immediately before the end of the minimum pumping regime, and orange markers represent samples taken during the maximum pumping regime. The vertical bars indicate the screen depths of each well.

If translated into residence times for the water pumped by the drinking water station, the ^{37}Ar activities suggest that the sampled water has average residence times of approximately 1.5 month in wells 1 and 5, more than 3.5 months in well 7. If the pumped water were entirely made up of freshly infiltrated SW, the average age of the freshly infiltrated SW and the average age of the water in the respective wells would be identical. It is more likely, however, that the pumped water represents a mix between freshly infiltrated SW and older GW components. If the latter were the case, the residence time of the freshly infiltrated SW would be smaller than the measured average residence time of the samples. The range of possible average residence times for the freshly infiltrated SW in the two pumping wells 1 and 5, therefore, lies between 12 days and the measured ^{37}Ar residence time (~1.5 months). In well 7, ^{37}Ar has reached the production-decay equilibrium and can thus not be used to identify a range of possible residence times. It can only be stated that the minimum residence time in well 7 is more than 110 days.

5.4.2.2. Tracer based analysis of the mixing ratio

As opposed to a standard analysis of mixing, where end-members for the different tracers would be given by true chemically and physically possible end-members (for example a production-decay equilibrium), the analysis in this study is oriented towards an improvement of the SW-GW model of the drinking water wellfield, and end-members therefore are defined in relevance to the Aeschau drinking water wellfield and model: (1) old GW from the background piezometer, and (2) freshly infiltrated SW. The HMC flow tracking component of the flow model allows tracking water that enters the model domain through three different sources: the specified flux inflow BC for SW, the constant head BC for GW inflow, and the precipitation BC. Precipitation was negligible during the pumping experiment, and the contribution of this component is therefore not considered. The two end-members, of which the pumped water mix in the model can thus consist are: (1) recently infiltrated SW entering the model through the SW inflow BC, which corresponds to the Emme River water, and (2) GW entering the model from the GW inflow BC, which corresponds to the GW in the background piezometer (A41). The following tracers can be assumed stable during the travel times (i.e., days to multiple months), which are relevant to the drinking water wellfield, and are therefore suited as indicators for mixing:

- NGRT: The two end-members for NGRT are given by: (1) freshly infiltrated SW with a NGRT corresponding to the SW water temperature during the residence time range of the SW component in the corresponding well, and (2) the GW in A41 (7.4 °C, see **Tab. 5.2**). The end-member for SW were chosen according to the travel time range found in the ^{37}Ar analysis, which suggests that the SW component in well 1 and 5 has a residence time between 12 days to 1.5 months. The average SW temperature during the two months prior to the pumping experiment was 3.7 °C. For A13 the same recharge temperature as for wells 1 and 5 was used, whereas for A25 the average SW temperature during the pumping experiment (2.8 °C). The NGRT-mix was quantified with **Eq. (5.8)**.
- $^{3}\mathrm{He}_{\mathrm{corr}}/^{4}\mathrm{He}$: The two end-member ratios are then given by: (1) water with a ratio corresponding to AEW ($1.36 \cdot 10^{-6}$, see **Tab. 5.2**), and (2) the GW in A41 ($1.288 \cdot 10^{-6}$, see **Tab. 5.2**). The $^{3}\mathrm{He}_{\mathrm{corr}}/^{4}\mathrm{He}$-mix between the two end-members was calculated with **Eq. (5.9)**.

The tracer-based mixing fractions between recently infiltrated SW and GW are summarized in **Tab. 5.4**. The two different tracer-based mixes are comparable: both tracers show that the water mix pumped by the three wells consists of approximately equal parts of freshly infiltrated SW and GW, rather than consisting purely of older GW or purely of freshly infiltrated SW. The water in well 1 appears to be made up of more SW than the water in well 5, whereas well 7 is again more SW influenced than well 5. $^{3}\mathrm{He}_{\mathrm{corr}}/^{4}\mathrm{He}$ suggests a strong reaction to a reduction of the pumping rate in both well 1 and well 7: -28 % and -16 % SW, respectively. NGRT only suggests a strong reaction in A13 (-19%), but not in well 1 or well 5. The $^{3}\mathrm{He}_{\mathrm{corr}}/^{4}\mathrm{He}$-based tracer attributes slightly more SW influence compared to NGRT overall.

Table 5.4: Estimations of contribution of recently infiltrated SW in the water mix based on the analysis of 2 different environmental tracer-based indicators (NGrt and $^{3}\mathrm{He}_{\mathrm{corr}}/^{4}\mathrm{He}$). The end-members for this analysis were chosen according to end-members relevant in the SW-GW model. For this purpose, measurements from A41 were chosen to correspond to the GW end-member, as opposed to the true physically or chemically possible end-members.

location	**regime**	**NGRT - mix**	**$^{3}\mathrm{He}_{\mathrm{corr}}/\,^{4}\mathrm{He}$ - mix**
		% SW in the sampled water mix	**% SW in the sampled water mix**
A13	*max*	61%	71%
	min	78%	100%
A25	*max*	-	-
	min	75%	100%
well 1	*max*	46%	60%
	min	45%	32%
well 5	*max*	10%	26%
	min	8%	26%
well 7	*max*	-	83%
	min	-	67%

5.4.3. Identifying the appropriate model structure: comparing measured and simulated pumped water mix and travel times

Comparison of the measured water mix (see **Tab. 5.4**) to the simulated water mix (see **Tab. 5.1**) of the pumped water in wells 1, 5 and 7 clearly reveals that out of the three tested K_{rb} scenarios, only the K_{rb} = 2.4 m/d model is appropriate for the simulation of mixing. Both the K_{rb} = 0.24 m/d and K_{rb} = 24 m/d model cannot even approximately reproduce the mixing ratio between freshly infiltrated SW and GW as indicated by the tracer data. And in terms of the changes in the sources of the pumped water mix in reaction to 7 days of reduced pumping, the n_{aq} = 0.1 *trans* model best approaches the magnitude of changes observed in the tracer measurements. However, the difference to the K_{rb} =2.4 m/d and *naq* = 0.2 *trans* model is very small.

A comparison of the simulated mean travel times of the SW component in the pumped water to the tracer-based residence times shows that all models are in accordance with the ^{222}Rn-based measurements: the mean travel times exceed the 12 days required for ^{222}Rn to reach equilibrium. However, based on the mean travel times a clear distinction of the most appropriate model parametrization scenario is difficult: In general, the n_{aq} = 0.1 models are best for reproduction of the mean travel times in well 5, whereas the n_{aq} = 0.2 models reproduce the travel times for well 1 the best and the n_{aq} = 0.43 models the travel times for well 7. The best overall reproduction of the tracer-based mean travel times estimates could, however, be attributed to the K_{rb} = 2.4 m/d scenario. However, the comparison between mean travel times of the pumped SW component and tracer based residence times do not allow a very convincing indication of a most appropriate model parametrization. And since the model does not simulate the complete catchment but only the vicinity of the wellfield, the simulated travel times cannot be compared to the residence times measurements based on ^{3}H/^{3}He.

Based on these findings, the most appropriate model parametrizations are the K_{rb} = 2.4 m/d, n_{aq} = 0.1 and the K_{rb} = 2.4 m/d, n_{aq} = 0.2 models. Both models allow simulating similar mixing ratios, particularly in comparison to the K_{rb} = 0.24 m/d and K_{rb} = 24 m/d models, which result in completely different mixing ratios than the measured ones. Moreover, the two most appropriate models result in relatively similar transient reactions of the mixing ratios and similar travel times. However, a yet more appropriate model could potentially be found through a simultaneous calibration of a heterogeneous K_{aq}-field, a K_{rb} and a n_{aq} against combined observations of hydraulic heads, SW discharge, exchange fluxes and mixing ratios. The tracer-based mixes also reveal a structural problem of the model: While the influence of SW in the pumped water increases with increasing distance of the pumping well to the river (i.e., the SW component is highest in well 7), the measured mix suggests that this only holds true for wells 1 and 7. In well 5, on the other hand, much less SW is pumped compared to well 1 (shown by both NGRT and $^{3}He_{corr}/^{4}He$) and compared to well 7 (shown by $^{3}He_{corr}/^{4}He$). This structural problem is most likely due to the simplifying assumption of a homogeneous K_{aq}.

5.5. Discussion

We combined a multi-tracer study of an important alluvial drinking water station with physically-based flow modelling, and ultimately used the tracer-based mixing analysis of the sources of the pumped water to identify an appropriate model parametrization. The multi-tracer investigation consisted of measurements of ^{222}Rn, ^{3}H/^{3}He, atmospheric noble gases, and ^{37}Ar measurements. ^{37}Ar has never been used as a tracer in this context before. The fact that the ^{37}Ar and ^{3}H/^{3}He dating methods resulted in different residence times revealed that the sampled water consisted of a mix of two different water types: (a) One type constitutes freshly infiltrated

SW that has resided in the subsurface on timescales to which only ^{37}Ar is sensitive (i.e., multiple days to 3.5 months). (b) Another type constitutes older GW that has resided in the subsurface on larger timescales to which ^{37}Ar is not sensitive anymore, but that is detectable by ^{3}H/^{3}He (i.e., multiple months to many years). Hence, through the combined application of the ^{222}Rn, ^{37}Ar and ^{3}H/^{3}He dating methods, residence times and mixing portions of water types (a) and (b) could be constrained.

Through a simultaneous application of these tracers, we could furthermore: (1) demonstrate that the ^{37}Ar-method is well suited for the estimation of intermediate residence times in SW-GW systems, (2) identify suitable tracers for the analysis of mixing of different water components in the pumped water mix, and (3) identify an appropriate flow model parametrization by including tracer-based estimates of mixing and residence times alongside the classical analysis of hydraulic heads.

For the analysis of an appropriate model parametrization, we used multiple models of the alluvial drinking water station that were all well-calibrated against hydraulic heads by means of $RMSE_{GW}$, but that strongly differed in the hydraulic conductivity of the riverbed (K_{rb}) and the porosity of the aquifer (n_{aq}). While all models performed similarly well in reproducing the hydraulic heads, the models strongly differed in the simulation of the pumped water mix (see **Tab. 5.1**). The tracer based results allowed identifying that out of the three different K_{rb} scenarios, only the intermediate permeability scenario allowed reproducing the tracer-based mixing ratios (see **Tab. 5.1 and 5.4**). Moreover, out of the three different n_{aq} scenarios, only the lower porosity scenarios resulted in similarly rapid reactions to a change in the pumping regime as suggested by the tracer-based mixing ratios. Identifying the shortcomings of the inappropriately parametrized models would in this case not have been possible by analyzing hydraulic heads alone. The analyses of the natural tracers also allowed detecting a structural issue in the model: instead of simulating a stronger influence of SW in the wells closer to the river, the models suggested the opposite, i.e., that the wells further away from the river are more strongly influenced by SW. This difference between the tracer measurements and the simulations most likely reveals that the assumption of homogeneity for K_{aq} is not justified, and that preferential flow paths due to heterogeneous structures might exist. These findings highlight the need (1) for an automatic and simultaneous calibration of K_{aq}, K_{rb} and n_{aq} against combined observations of hydraulic heads, SW discharge, exchange fluxes and tracer-based mixing (or tracer concentrations directly), and (2) for inclusion of heterogeneous K_{aq} structures to represent potential preferential flow paths in the model.

Our study highlights (1) that it is absolutely crucial to make tracer-based estimates of mixing and residence times in order to understand the sources of pumped drinking water, (2) that it is necessary to choose tracers with a sensitivity to the expected timescales of the investigated system (which is in accordance with the interpretation of Larocque et al. (2009)), and (3) that multiple tracers need to be applied in order to estimate both mixing and residence times simultaneously. However, it is generally difficult to know the dominant timescales of a SW-GW system beforehand. Nonetheless, with the addition of ^{37}Ar, an important gap in the analysis of residence times could be closed; a residence timescale which is particularly important in alluvial SW-GW systems. Including ^{37}Ar allowed constraining the travel time of freshly infiltrated SW between the Emme River and the GW pumps, and helped identifying the end-members for the NGRT mixing analysis; this would not have been possible with other tracers for residence times such as ^{222}Rn or ^{3}H/^{3}He, which were insensitive to the relevant residence times. Even though sampling for ^{37}Ar requires a relatively large volume of water to be degassed directly in the field, and that the depth dependency of the production of ^{37}Ar needs to be determined locally, the insights gained from ^{37}Ar are were nonetheless extremely useful.

5.6. Conclusions

Fully-integrated, physically-based SW-GW modelling with flow tracking, as provided through HGS and the HMC module, is an indispensable tool to investigate mixing processes and track the sources of pumped water of drinking water wellfields. Without much extra computational costs, the hydraulic mixing-cell method allowed tracking the flow of water without the need to explicitly simulate the advection-dispersion equation, as for the relevant spatial and temporal timescales the simplifying assumptions of the HMC method are justified. The interpretation of physically-based flow models with flow tracking has major advantages: besides quantifying mixing and identifying sources of pumped water, they also allow exploring the validity of preliminary assumptions; assumptions that, for example, are required for the interpretation of the novel ^{37}Ar-method. The results of our study confirmed, however, that complex, physically-based models do not automatically reproduce correct mixing fractions between water from different sources in the subsurface if they are only calibrated in the classical sense, that is, by minimizing the RMSE of hydraulic heads. Ultimately, if predictions are required for the management of the investigated system, tracer-based observations of exchange flows, mixing of water from different sources and of residence times are needed to inform the choice of an appropriate model.

Already a simple scenario analysis, where the outcomes of different equally-likely model parametrizations in terms of RMSE of hydraulic heads were compared to tracer-based observations, allowed improving the choice of the model parametrization beyond an analysis of hydraulic heads alone. The calibration of the model could be further improved through an automatic calibration procedure that includes not only hydraulic heads but all observation types and could introduce heterogeneity in the model parametrization wherever appropriate, for example following the procedures described by Doherty (2015). Through an assessment of the worth of the different observation types in reducing the uncertainty of the model, for example by applying a similar framework as used in Schilling et al. (2014), one could furthermore quantify the value of such a multi-tracer analysis of mixing and residence times.

With respect to the goal raised in the introduction of identifying suitable unconventional observation types to better characterize alluvial SW-GW interactions and inform appropriate model parametrizations for their simulation, we could identify that particularly tracer-based estimates of mixing ratios were very useful. The novel combination of ^{222}Rn, ^{3}H/^{3}He and noble gases with ^{37}Ar provides an optimal multi-tracer framework to simultaneously characterize residence times and mixing ratios, on temporal scales that are also the most relevant for drinking water safety. The combination of complex flow models with such multi-tracer studies so far is an exception, but indicates the direction for future applications, as also pointed out by Turnadge and Smerdon (2014). Our study is a demonstration of the largely unexplored potential of unconventional data in informing SW-GW model construction and calibration. A complex flow model that simultaneously reproduces hydraulic heads, exchange fluxes, travel times and the pumped water mix is an ideal tool to delineate protection zones and manage the amount of GW abstraction for drinking water stations.

5.7. Acknowledgements

The authors would like to thank L. Tyroller, E. Ghadiri, Y. Tomonaga, A. Popp, J. Oser, S. Figura, I. Beck and W. Wild from EAWAG for their support in the noble gas laboratory and during sampling, and A. Badin, C. Carlier, P. Wanner, A. Schomburg, R. Costa and L. Marguet from the CHYN for their assistance during the extensive sampling campaign. The authors are very thankful to B. Burkhalter, B. Wysmüller and R. Hirschi from the WVRB for making the pumping experiment possible.

5.8. References

Aeschbach-Hertig, W., Peeters, F., Beyerle, U., Kipfer, R., 1999. Interpretation of dissolved atmospheric noble gases in natural waters. Water Resour. Res., 35(9): 2779-2792.

Aeschbach-Hertig, W., Solomon, D.K., 2013. Noble Gas Thermometry in Groundwater Hydrology. In: Burnard, P. (Ed.), Advances in Isotope Geochemistry. Springer, Berlin Heidelberg, Germany.

Åkesson, M. et al., 2015. Constraining age distributions of groundwater from public supply wells in diverse hydrogeological settings in Scania, Sweden. J. Hydrol., 528: 217-229.

Althaus, R. et al., 2009. Noble gas tracers for characterisation of flow dynamics and origin of groundwater: A case study in Switzerland. J. Hydrol., 370: 64-72.

Anderson, M.P., Woessner, W.W., Hunt, R.J., 2015. Applied Groundwater Modelling (Second edition). Academic Press, Oxford, United Kingdom.

AWA, 2013. Amt für Wasser und Abfall des Kantons Bern - Digitale Grundwasserkarte des Kantons Bern (GW25), Bern.

Banks, E.W., Brunner, P., Simmons, C.T., 2011. Vegetation controls on variably saturated processes between surface water and groundwater and their impact on the state of connection. Water. Resour. Res., 47.

Bauser, G. et al., 2010. Real-time management of an urban groundwater well field threatened by pollution. Environ. Sci. Technol., 44: 6802-6807.

Beyerle, U. et al., 1999. Infiltration of river water to a shallow aquifer investigated with $^3H/^3He$, noble gases and CFCs. J. Hydrol., 220: 169-185.

Beyerle, U. et al., 2000. A Mass Spectrometric System for the Analysis of Noble Gases and Tritium from Water Samples. Environ. Sci. Technol., 34: 2042-2050.

Biaggi, D., Polack, P., Backman, N., 2005. Grundwasserfassungen Aeschau - Gesuch um Konzessionserneuereung - Fachbericht Hydrologie/Hydrogeologie, Geotechnisches Institut, Bern.

Blau, R.V., 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.

Boano, F. et al., 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Rev. Geophys., 52: 603-679.

Boulton, A.J., Datry, T., Kasahara, T., Mutz, M., Stanford, J.A., 2010. Ecology and management of the hyporheic zone: stream– groundwater interactions of running waters and their floodplains. J. N. Am. Benthol. Soc., 29(1): 26-40.

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

Bredehoeft, J., 2005. The conceptualization model problem—surprise. Hydrogeol. J., 13: 37-46.

Brunner, P., Simmons, C.T., 2011. HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. Ground Water, 50(2): 170-176.

Cecil, L.D., Green, J.R., 2000. Radon-222. In: Cook, P.G., Herczeg, A.L. (Eds.), Environmental tracers in subsurface hydrology. Springer, New York, USA.

Chaix, O., Gander-Kunz, Y., 2014. Grundlagen für die Wasserversorgung 2025 - Risiken, Herausforderungen und Empfehlugen, Bundesamt für Umwelt BAFU, Bern.

Cook, P.G., Böhlke, J.-K., 2000. Determining timescales for groundwater flow and solute transport. In: Cook, P.G., Herczeg, A.L. (Eds.), Environmental tracers in subsurface hydrology. Springer, New York.

Delottier, H., Pryet, A., Dupuy, A., 2016. Why Should Practitioners be Concerned about Predictive Uncertainty of Groundwater Management Models? Water Resour. Manage.

Delsmann, J.R., Winters, P., Vandenbohede, A., Oude Essink, G.H.P., Lebbe, L., 2016. Global sampling to assess the value of diverse observations in conditioning a real-world groundwater flow and transport model. Water Resour. Res., 52.

Diem, S., 2013. Riverbank filtration within the context of river restoration and climate change, University of Neuchâtel, Neuchâtel, 144 pp.

Doherty, J.E., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. - PEST: complete theory and what it means for modelling the real world. Watermark Numerical Computing, Brisbane, Australia.

DURRIDGE, 2012. RAD H_2O User Manual.

DURRIDGE, 2014. RAD7 Radon Detector - User Manual.

Fabryka-Martin, J.T., 1988. Production of radionuclides in the earth and their hydrogeologic significance, with emphasis on chlorine-36 and iodine-129, University of Arizona, Arizona, USA.

Fetter, C.W., 2001. Applied Hydrogeology, 4th edition. Prentice Hall, Upper Saddle Creek, NJ, USA.

Feurer, D., Bailly, J.-S., Puech, C., Le Coarer, Y., Viau, A.A., 2008. Very high-resolution mapping of river-immersed topography by remote sensing. Prog. Phys. Geogr., 32(4): 403-419.

Figura, S., Livingstone, D.M., Hoehn, E., Kipfer, R., 2011. Regime shift in groundwater temperature triggered by the Arctic Oscillation. Geophys. Res. Lett., 38.

Figura, S., Livingstone, D.M., Kipfer, R., 2013. Competing controls on groundwater oxygen concentrations revealed in multidecadal time series from riverbank filtration sites. Water Resour. Res., 49: 7411-7426.

Figura, S., Livingstone, D.M., Kipfer, R., 2015. Forecasting Groundwater Temperature with Linear Regression Models Using Historical Data. Ground Water, 53(6): 943-954.

Fleckenstein, J.H., Niswonger, R.G., Fogg, G.E., 2006. River-aquifer interactions, geologic heterogeneity, and low-flow management. Ground Water, 44(6): 837-52.

Frei, C. et al., 2007. Klimaänderung und die Schweiz 2050 - Erwartete Auswirkungen auf Umwelt, Gesellschaft und Wirtschaft, OoCC / ProClim, Bern.

Gardner, W.P., Harrington, G.A., Solomon, D.K., Cook, P.G., 2011. Using terrigenic 4He to identify and quantify regional groundwater discharge to streams. Water Resour. Res., 47.

Guillon, S. et al., 2016. Alteration of natural ^{37}Ar activity concentration in the subsurface by gas transport and water infiltration. J. Environ. Radioact., 155: 89-96.

Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—the Ground-Water Flow Process, U.S. Geological Survey Techniques and Methods 6–A16. USGS, Reston, Virgina, USA.

Harvey, J.W., Gooseff, M., 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resour. Res., 2015(51).

Hendricks Franssen, H.J. et al., 2011. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. Water Resour. Res., 47.

Hoehn, E., Von Gunten, H.R., Stauffer, F., Dracos, T., 1992. Radon-222 as a Groundwater Tracer. A Laboratory Study. Environ. Sci. Technol., 26: 734-738.

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

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

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

Johnson, C., Armstrong, H., Wilson, W.H., Biegalski, S.R., 2015. Examination of radioargon production by cosmic neutron interactions. J. Environ. Radioact., 140: 123-129.

Käser, D. et al., 2013. Channel Representation in Physically Based Models Coupling Groundwater and Surface Water: Pitfalls and How to Avoid Them. Ground Water, 52(6): 827.836.

Käser, D., Hunkeler, D., 2015. Contribution of alluvial groundwater to the outflow of mountainous catchments. Water Resour. Res., 52.

Kipfer, R., Aeschbach-Hertig, W., Peeters, F., Stute, M., 2002. Noble Gases in Lakes and Ground Waters. In: Porcelli, D., Ballentine, C., Wieler, R. (Eds.), Noble gases in geochemistry and cosmochemistry. Reviews in Mineralogy and Geochemistry, pp. 615-700.

Kollet, S., Maxwell, R., 2006. Integrated surface-groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. Adv. Water Resour., 29(7): 954-998.

Kropf, P. et al., 2014. Wireless Mesh Networks and Cloud Computing for Real Time Environmental Simulations, 10th International Conference on Computing and Information Technology (IC2IT2014), Phuket.

Kurtz, W. et al., (under review). Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. Environ. Model. Softw.

Lapin, A. et al., 2014. Real-time Environmental Monitoring for Cloud-based Hydrogeological Modelling with HydroGeoSphere, High Performance Computing and Communications Conference (IEEE HPCC14), Paris.

Larocque, M., Cook, P.G., Haaken, K., Simmons, C.T., 2009. Estimating Flow Using Tracers and Hydraulics in Synthetic Heterogeneous Aquifers. Ground Water, 47(6): 786-796.

Li, Q. et al., 2008. Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. J. Hydrol., 357: 317-336.

Loosli, H.H., Lehmann, B.E., Smethie, W.M., 2000. Noble Gas Radioisotopes. In: Cook, P.G., Herczeg, A.L. (Eds.), Environmental tracers in subsurface hydrology. Springer, New York, USA.

Loosli, H.H., Purtschert, R., 2005. Rare Gases. In: P. Aggarwal, J.R.G.a.K.F. (Ed.), Isotopes in the Water Cycle: Past, Present and Future of a Developing Science. IAEA, Vienna, pp. 91-95.

Lucas, L.L., Unterweger, M.P., 2000. Comprehensive Review and Critical Evaluation of the Half-Life of Tritium. J. Res. Natl. Inst. Stand. Technol., 105(4): 541-549.

Maliva, R.G., 2016. Aquifer Characterization Techniques. Schlumberger Methods in Water Resources Evaluation Series No. 4. Springer International Publishing, Switzerland.

Mayer, A. et al., 2014. A multi-tracer study of groundwater origin and transit-time in the aquifers of the Venice region (Italy). Appl. Geochem., 50: 177-198.

McCallum, J.L., Cook, P.G., Simmons, C.T., 2014a. Limitations of the Use of Environmental Tracers to Infer Groundwater Age. Ground Water, 53: 56-70.

McCallum, J.L., Cook, P.G., Simmons, C.T., Werner, A.D., 2014b. Bias of Apparent Tracer Ages in Heterogeneous Environments. Ground Water, 52(2): 239-250.

McLaren, R.G., 2011. GridBuilder - A preprocessor for 2-D, triangular element, finite-element programs. Groundwater Simulations Group, University of Waterloo, Waterloo, Ontario, Canada.

Partington, D. et al., 2013. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resour. Res., 49: 5501-5519.

Partington, D. et al., 2011. A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface wateregroundwater flow models. Environ. Model. Softw., 26: 886-898.

Partington, D. et al., 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. J. Hydrol., 458-459: 28-39.

Purtschert, R., 2008. Timescales and Tracers. In: Edmunds, W.M., Shand, P. (Eds.), Natural Groundwater Quality. Blackwell Publishing Ltd., Oxford, UK, pp. 91-108.

Purtschert, R., Yokochi, R., Sturchio, N.C., 2013. ^{81}Kr dating of old groundwater. . In: Suckow, S., Aggarwal, P., Araguas-Araguas, L. (Eds.), Isotope methods for dating old groundwater. IAEA, Vienna.

Riedmann, R.A., 2011. Separation of argon from atmospheric air and measurements of ^{37}Ar for CTBT purposes, University of Bern, Bern, Switzerland.

Riedmann, R.A., Purtschert, R., 2011. Natural ^{37}Ar Concentrations in Soil Air: Implications for Monitoring Underground Nuclear Explosions. Environ. Sci. Technol., 45: 8656-8664.

Riedmann, R.A., Purtschert, R., 2016. Separation of argon from environmental samples for Ar-37 and Ar-39 analyses. Sep. Purif. Technol., 170: 217-223.

Schilling, O.S. et al., 2014. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water– vegetation interactions. J. Hydrol., 519: 2258-2271.

Schilling, O.S., Irvine, D.J., Hendricks Franssen, H.J., Brunner, P., (under review). Estimating the spatial extent of unsaturated zones in heterogeneous river-aquifer systems. Water Resour. Res.

Schlosser, P., Stute, M., Dörr, H., Sonntag, C., Münnich, K.O., 1988. Tritium/^{3}He dating of shallow groundwater. Earth Planet. Sci. Lett., 89: 353-362.

Schnegg, P.-A., 2003. A new field fluorometer for multi-tracer tests and turbidity measurement applied to hydrogeological problems 8th International congress of the Brazilian Geophysical Society. Sociedade Brasileira de Geofisica, Rio de Janeiro, Brazil.

Simmons, C.T., Hunt, R.J., Cook, P.G., 2012. Using Every Tool in the Toolbox Ground Water, 50(3): 323.

Sinreich, M. et al., 2012. Grundwasserressourcen der Schweiz - Abschätzung von Kennwerten, BAFU, Bern.

Solomon, D.K., 2000. ^{4}He in groundwater. In: Cook, P.G., Herczeg, A.L. (Eds.), Environmental tracers in subsurface hydrology. Springer, New York, USA.

Solomon, D.K., Cook, P.G., 2000. ^{3}H and ^{3}He. In: Cook, P.G., Herczeg, A.L. (Eds.), Environmental tracers in subsurface hydrology. Springer, New York, USA.

Sophocleus, M., 2002. Interactions between groundwater and surface water: the state of the science. Hydrogeol. J. , 10: 52-67.

Spreafico, M., Weingartner, R., 2005. Hydrologie der Schweiz, Ausgewählte Aspekte und Resultate. Berichte des BWG, Serie Wasser. Bundesamt für Wasser und Geologie (BWG), Switzerland, Bern, Switzerland.

swisstopo, 2010. Bundesamt für Landestopografie swisstopo - swissALTI3D - Das hoch aufgelöste Terrainmodell der Schweiz, Bern.

Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S., 2010. HydroGeoSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Hydrogeosphere Manual.

Townley, L.R., 2012. Calibration and sensitivity analysis. In: Barnett, B. et al. (Eds.), Australian Groundwater Modelling Guidelines. Waterlines Report 82. National Water Commission, Canberra, AUS, pp. 57-78.

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

Visser, A. et al., 2014. Intercomparison of tritium and noble gases analyses, 3H/3He ages and derived parameters excess air and recharge temperature. Appl. Geochem., 50: 130-141.

Vogt, T. et al., 2009. Fluctuations of electrical conductivity as a natural tracer for bank filtration in a losing stream. Adv. Water Resour., 33: 1296-1308.

Voss, C.I., 2011. Editor's message: Groundwater modeling fantasies —part 1, adrift in the details. Hydrogeol. J., 19: 1281-1284.

Winter, T.C., Harvey, J.W., Franke, O.L., Alley, W.M., 1998. Ground water and surface water: a single resource, Denver, Colorado.

Würsten, M., 1991. GWB - Hydrogeologische Untersuchungen Aeschau: Schlussbericht, Geotechnisches Institut, Zürich.

Chapter 6

6. Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources

Abstract

Online data acquisition, data assimilation and integrated hydrological modelling have become more and more important in hydrological science. The usefulness of these scientific and technological advances for water resources management have been documented in the literature but their joint application is still limited. In this study, we explore cloud computing for integrating field data acquisition and stochastic, physically-based hydrological modelling in a data assimilation and optimization framework as a service to water resources management. For this purpose, we developed an ensemble Kalman filter-based data assimilation system for the integrated hydrological model HydroGeoSphere, which was adapted for the use in a cloud computing environment. The changes in the model codes for the utilization of the cloud infrastructure were minimal and mainly concerned the forward propagation of the model ensemble. A synthetic data assimilation experiment based on the widely used tilted V-catchment problem served as a benchmark for testing the usefulness of the proposed modelling platform and for evaluating the computational efficiency of the cloud-based implementation. Results showed that the computational overhead for the application of the data assimilation platform in a cloud computing environment is minimal, which makes it well suited for practical water management problems. Advantages of the cloud-based implementation comprise the independence from computational infrastructure and the straightforward integration of cloud-based observation databases with the modelling and data assimilation platform.

Keywords: Cloud computing, Integrated hydrological modelling, Data assimilation, Water resources management, HydroGeoSphere, Wireless sensor networks

This chapter has been accepted as an original research article:
Kurtz, W., Lapin, A., Schilling, O.S., Tang, Q., Schiller, E., Braun, T., Hunkeler, D., Vereecken, H., Sudicky, E., Kropf, P., Hendricks Franssen, H.-J., and Brunner, P. (in press): Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. Submitted to Environ. Model. Softw. doi: 10.1016/j.envsoft.2017.03.011

Two collaborations using the EnKF-HGS framework are in preparation:
Tang, Q., Kurtz, W., Schilling, O.S., Brunner, P., Vereecken, H., and Hendricks-Franssen, H.J. (in preparation): Is it important to characterize complex patterns of riverbed hydraulic conductivities for assessing river-aquifer exchange fluxes? An evaluation with an integrated fully coupled hydrological model.

Tang, Q., Schilling, O.S., Kurtz, W., Brunner, P., Vereecken, H., and Hendricks-Franssen, H.J. (in preparation): Reproducing flood induced riverbed transience with physically-based modelling and data assimilation.

6.1. Introduction

Hydrological and hydrogeological systems are highly heterogeneous, and the temporal evolution of their spatially variable states is driven by dynamic forcing functions. Deterministic numerical models are an important tool for understanding and managing such systems. Such models can support the water management decision making process with predictions of the temporal evolution and the spatial distribution of target state variables. Groundwater management often relies on simulations with distributed physically-based hydrological models, e.g., for wellfield operations adjacent to a river. The available numerical models have greatly improved in recent years. For example, there are ongoing efforts towards a better description of the dynamic feedbacks between subsurface and surface water processes (Kollet and Maxwell, 2006; Brunner and Simmons, 2012). One of the advantages of such fully-coupled surface-subsurface models is that the location of surface water features, such as the position of rivers, no longer needs to be predefined through boundary conditions. They are therefore very well suited for simulating changing surface water conditions such as floods or droughts.

Deterministic models need to be calibrated based on existing observations. However, there is a growing awareness of the uncertainty related to such deterministic model predictions (Liu et al., 2012). The uncertainties stem from the limited knowledge about the spatial distribution and magnitude of important model parameters, such as hydraulic conductivity or porosity (Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008). Also, the high spatio-temporal variability of boundary conditions and input variables, such as precipitation, can highly affect model predictions. Moreover, the limited availability of spatial and temporal field data limits the reliability of the calibration process. Finally, the computational requirements of many numerical simulators, especially fully-coupled, physically-based models, exclude in most cases a solid uncertainty analysis. These uncertainties can be substantial and they often undermine the credibility of hydrological and hydrogeological models, especially once it comes to predicting highly dynamic systems.

However, a range of technological and mathematical advances allows overcoming some of the previous limitations. Above all, these advances are related to three key developments: data acquisition techniques, the increasing computational capacities of hydrological models, as well as the integration of measurement data in the modelling process.

Firstly, the acquisition of field data has been greatly facilitated. Traditionally, hydrological field measurements such as piezometer levels, precipitation, soil moisture, discharge or water quality indicators, were acquired either manually in the field at predefined measurement intervals or recorded with data loggers, which have to be read out on a weekly or monthly basis. Ongoing advances in sensor technology and telemetry make it now possible to obtain hydrological data shortly after their acquisition in the field, even for very remote field sites. Wireless sensor networks (WSN) are increasingly applied in environmental studies, e.g., in the context of soil moisture monitoring (Robinson et al., 2008; Ritsema et al., 2009; Bogena et al., 2010) or surface water (Li et al., 2011), studies on wetland dynamics (Watras et al., 2014) or the acquisition of solute transport data for modelling purposes (Loden et al., 2009; Barnhart et al., 2010). Such WSNs consist of distributed sensors, which transmit the measured data through wireless in-built radio modules to a set of router units that manage the communication within the network (Bogena et al., 2010). As a result, the measured data can be accessed by the user on the permanent storage device in near real-time. This can be of great advantage for water management purposes, especially when a system (e.g., pumps for river bank filtration) needs to be controlled and regulated continuously, and the hydraulic forcings (such as the water level in a river) are highly transient.

Secondly, the computational efficiency of hydrological models is continuously increasing. Recent advances in numerical mathematics lead to the development of more efficient solvers and preconditioning techniques (Herbst et al., 2008; Maxwell, 2013). Parallelization of model codes (Ashby and Falgout, 1996; Vereecken et al., 1996; Jones and Woodward, 2001; Mills et al., 2007) makes it possible to solve hydrogeological problems with a high spatio-temporal resolution and on large scales. These advancements in computational efficiency also facilitated the usage of more sophisticated physical process descriptions in the modelling process. For example, state-of-the-art hydrological models now also provide a full 3D solution of the Richards equation, and a physically consistent coupling between surface and subsurface flow equations (Kollet and Maxwell, 2006; Brunner and Simmons, 2012).

Finally, the combination of sequential data assimilation techniques like the ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998) with hydrogeological models now allows integrating real-time data into the modelling process. These methods can be used to effectively merge uncertain model predictions with uncertain observation data in a Bayesian sense. The uncertainty of model predictions is approximated through the forward simulation of an ensemble of model realisations, where each realisation can have a different combination of initial conditions, model forcings and model parameters. The uncertain model predictions are then sequentially updated with measurement data. In this updating step, the uncertainties in the model predictions and the uncertainties of the observations are optimally weighted and the model predictions are effectively adjusted towards the measured data. Besides the correction of state variables, it is also possible to use observation data to update model parameters jointly with the model states (Hendricks Franssen and Kinzelbach, 2008), which makes these methods very effective calibration tools. This methodology has already been applied to a variety of hydrogeological problems including assimilation of hydraulic head data (Chen and Zhang, 2006; Nowak, 2009), transport problems (Liu et al., 2008; Li et al., 2012), river-aquifer interactions (Kurtz et al., 2014; Rasmussen et al., 2015; Tang et al., 2015) or assimilation of discharge data (Camporese et al., 2009). This has, for example, been demonstrated in the context of operational flood forecasting (Seo et al., 2009; Weerts et al., 2010) and integrated hydrological modelling (Shi et al., 2015; Rasmussen et al., 2015; Kurtz et al., 2016). An application in a hydrogeological setting was given by Hendricks Franssen et al. (2011) for groundwater management of the upper Limmat aquifer in Zurich (Switzerland). In this case, a groundwater model is run on a daily basis to support management decisions on groundwater abstraction, and the EnKF methodology is used to continuously correct the model predictions and model parameters with available piezometric head data. These corrected model predictions are then used as input for the real-time optimization of groundwater management activities (Bauser et al., 2010, 2012). In the particular case of the Limmat aquifer, the updated piezometric distribution from the groundwater model is used to optimally control the groundwater abstraction from a well field according to predefined management goals, which include the total abstraction rate and the maintenance of certain hydraulic conditions to prevent the leakage of contaminants to the well field from a close-by disposal site. Other applications of real-time optimization of groundwater resources include the energy efficient operation of well fields (Hansen et al., 2012; Bauer-Gottwein et al., 2016) or the accounting for the thermal regime within an aquifer (Marti, 2014). In Schwanenberg et al. (2011), data assimilation methods are used in conjunction with optimal control algorithms for a large-scale river network.

Such methods, especially in combination with physically-based hydrological models, are usually associated with a high computational burden due to the need to perform hundreds of model simulations in a Monte Carlo framework. This requires the availability of a dedicated computer infrastructure, which is not readily available for every end-user due to the high personal and financial effort for acquiring and maintaining such systems. This can, in part, be overcome by cloud-based services that provide computational resources on demand, and has been suggested as a future platform for hydrological modelling and model calibration (Hunt et al., 2010; Bürger et al., 2012). Such cloud-based solutions are flexible and elastic with respect to the choice of the computing environment (operating system, CPU, main memory, etc.), and can thus host a variety of simulation platforms with different computational requirements. Furthermore, such services are paid according to the actually consumed computation time. Therefore, the costs to the end-user are effectively reduced by avoiding the financial overhead that is required for installing and maintaining an own in-house computer system.

This study presents a fully-operational architecture for a cloud-based real-time prediction and management system in the context of groundwater management. The proposed system allows to perform data assimilation with the integrated hydrological model HydroGeoSphere in a cloud environment, and to use the improved model predictions for the optimization of groundwater management. It also includes a link to the online acquisition of field data, which transfers the measurement data acquired in the field to a cloud-based storage system. To the best of our knowledge, this is the first time that such a complete system for real-time water resources management is described in the context of cloud computing. A second main contribution is that the real-time simulation and management is done with a physically-based, integrated hydrological model, which is challenging in the context of data assimilation due to the large amount of required computational resources.

In the next section, a detailed overview over the different components of this cloud-based real-time monitoring and modelling platform is provided. In section 6.3, we apply this platform to a synthetic groundwater management problem in order to demonstrate the usefulness of this approach and to investigate the performance of the cloud-based solution in comparison to a more conventional (cluster-based) execution environment. Ultimately, we draw conclusions and provide recommendations for using this platform in real-world applications.

6.2. Conceptual framework

The cloud-based real-time monitoring and simulation platform outlined in this paper offers two principal functionalities to groundwater managers: (I) real-time access to field measurement data, and (II) real-time stochastic simulations, which are continuously improved by assimilating the most recent field measurements, with the possibility for additional real-time control of water resource systems. The first functionality includes the full process chain from acquiring the measurement data with online sensors, collecting these data via wireless sensor networks and transferring the data to online storage that can easily be accessed by the end-user. This allows decision makers to monitor the measured key hydrological variables in real-time. The data stored in the cloud also serve as a basis for adjusting hydrological model predictions with data assimilation techniques. The monitoring process chain used in our cloud-based real-time monitoring and simulation platform has been described in detail by Lapin et al. (2014) and is summarized in section 6.2.1. The second functionality is the stochastic real-time prediction of hydrological variables with a physically-based, distributed hydrological model that is corrected by the assimilation of measurement data from the sensor network. This module is the main focus of this paper and aims to support management decisions by providing a probabilistic nowcasting of the system states. This real-time, cloud-based simulation system consists of several components: (I) an interface for the user to orchestrate the cloud-based simulation service and to operate on the input/output, (II) a cloud-based computational infrastructure that hosts and manages the execution of the simulation code, and (III) a software that manages the stochastic forward simulations and performs the assimilation of measurement data. These system components are described in detail in sections 6.2.2, 6.2.3 and 6.2.4.

6.2.1. Data acquisition and management

Wireless sensor networks (WSNs) and Wireless Mesh Networks (WMNs) can be used to acquire and collect hydrological sensor data, especially in remote locations. For example, in a typical alpine or pre-alpine watershed rainfall intensity can vary strongly between sub-catchments, and measuring these variations properly is crucial for a correct estimation of the water budget of the whole watershed. In such environments, cellular network coverage is often not provided. Accessing such sub-catchments, and even more so, installing and connecting all the sensors, however, can be a big challenge. WSNs have proven to be a viable solution in this context due to their portability and relatively low installation costs. A WSN represents a set of smart devices, called sensor nodes, which are equipped with environmental sensors and transmit data from the environment in which they are deployed (Kropf et al., 2014; Lapin et al., 2014). WSN systems work well on smaller scales of up to a few hundred meters. On larger distances, i.e., at the scale of a few kilometers, transporting information from the sensor networks in the field to a fixed Internet network requires an additional WMN infrastructure (Jamakovic et al., 2013). **Figure 6.1** illustrates a typical data-flow from a WSN/WMN to the end-user Data from a WSN is delivered through an intermediate WMN to an Internet Gateway, connecting the WMN with the wired Internet. The data can then be stored in an online storage system and later accessed by an end-user or any kind of processing software over the Internet. Such a data acquisition and management system has been presented by Lapin et al. (2014) and provides the first cornerstone of our cloud-based real-time monitoring and modelling platform.

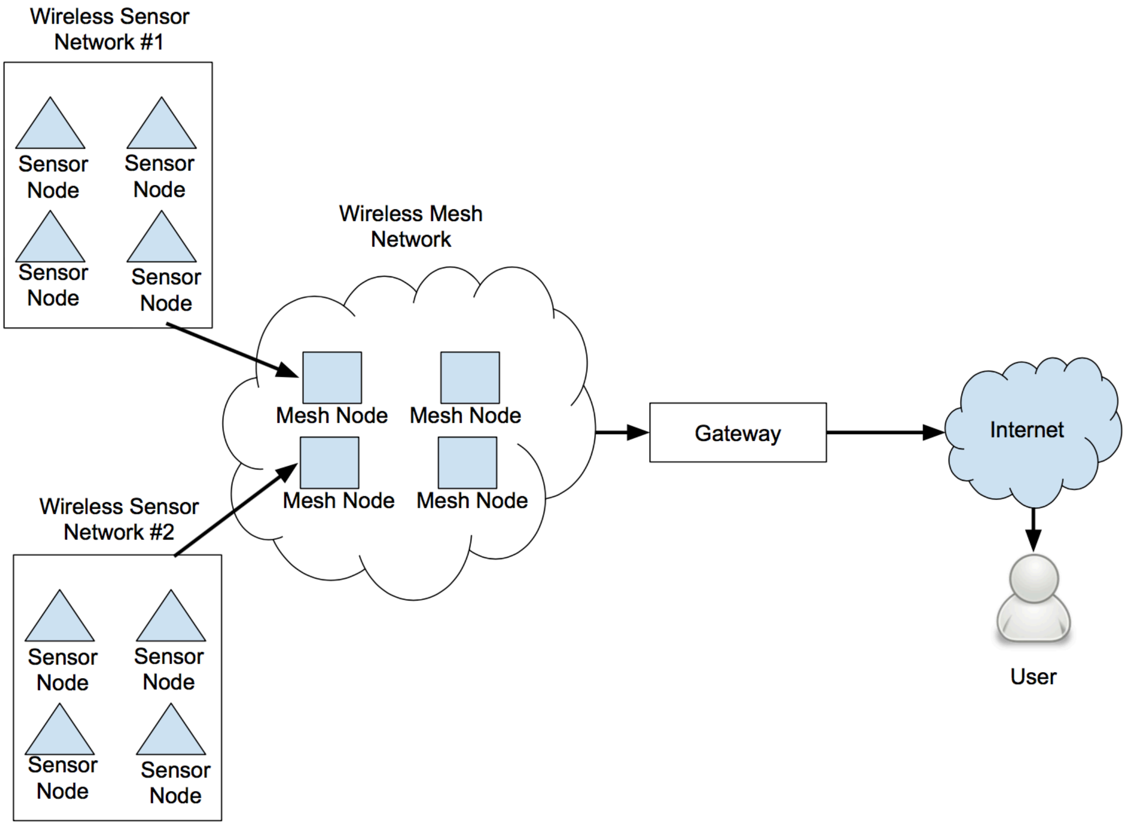


Figure 6.1: Typical data-flow from a Wireless Sensor Network.

6.2.2. HydroGeoSphere

HydroGeoSphere (HGS) is a state-of-the-art numerical code for the fully-integrated simulation of surface water, groundwater, and vegetation processes (Brunner and Simmons, 2012; Aquanty Inc., 2016). It is able to solve variably-saturated subsurface flow in three dimensions and uses a fully-consistent coupling between surface and subsurface flow equations. It has been applied successfully for the simulation of highly complex hydrological systems, such as large-scale solute transport (Blessent et al., 2011), transport in complicated fractured systems (Blessent et al., 2014), systems with natural and anthropogenic structures (De Schepper et al., 2015), as well as interactions between surface water discharge, groundwater recharge and tree ring growth (Schilling et al., 2014 (Chapter 4 of this thesis)). The prediction of surface water discharge, groundwater levels and fluxes demands that the parameters of HGS are well calibrated, and requires an uncertainty analysis. This in turn requires running many instances of HGS in a Monte Carlo framework, which makes HGS perfectly suited for parallelization and implementation in a cloud-based infrastructure.

6.2.3. Data assimilation with HydroGeoSphere

Data assimilation within the cloud-based modelling system is done via the ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998). As already mentioned in the introduction, this method uses a Monte Carlo type approach to approximate the uncertainty of the model predictions. This is done by forward simulations of an ensemble of model realisations, which can differ with respect to the initial conditions, forcing terms and model parameters. The model state vector $\mathbf{x}$ for each model realisation i at time step t (where observations are available) is derived by forward propagation of the dynamical model M using as input the state vector from the previous time step (t−1) and parameters $\mathbf{p}$ and forcings $\mathbf{q}$. Parameters and forcings are different for each model realisation:

$$\mathbf{x}_i^t = M(\mathbf{x}_i^{t-1}, \mathbf{p}_i, \mathbf{q}_i) \tag{6.1}$$

The dynamical model M is the integrated hydrological modelling software HGS in our case. The prognostic variables of HGS can be hydraulic head, temperatures and stream discharge amongst others. In this study, only hydraulic head is simulated and therefore the model state vector $\mathbf{x}$ only consists of this prognostic variable. The measurement equation, which links measured and modelled states is given by:

$$\mathbf{y}_i^t = \mathbf{H}\mathbf{x}_i^t \tag{6.2}$$

where $\mathbf{y}$ is the vector containing the measurements and the matrix $\mathbf{H}$ extracts or interpolates the state vector onto the observation locations. In order to account for measurement uncertainty, the actual observations $\mathbf{y}^0$ for time step t are perturbed with values drawn from a normal distribution N with a mean of zero and a standard deviation corresponding to the estimated measurement error ε:

$$\mathbf{y}_i^0 = \mathbf{y}^0 + N(0, \varepsilon) \tag{6.3}$$

The updated state vector $\mathbf{x}^+$ for each realisation i is then given by:

$$\mathbf{x}_i^+ = \mathbf{x}_{\boldsymbol{i}}^t + \mathbf{K}(\mathbf{y}_i^0 - \mathbf{y}_i^t) \tag{6.4}$$

where the Kalman gain matrix $\mathbf{K}$ is given by:

$$\mathbf{K} = \mathbf{C}_{xy} \left(\mathbf{H}\mathbf{C}_{xy} + \mathbf{R}\right)^{-1} \qquad (6.5)$$

where $\mathbf{C}_{xy}$ is the covariance matrix between the model states $\mathbf{x}$ and the predicted observations $\mathbf{y}$ and $\mathbf{R}$ is the covariance matrix of the measurement errors. If the measurement errors can be considered independent the matrix $\mathbf{R}$ consists of diagonal elements representing the individual measurement errors of the observations. As can be seen from Eq. (6.5), the Kalman gain matrix $\mathbf{K}$ weights the uncertainties in the model predictions (represented by $\mathbf{C}_{xy}$) with the measurement errors (represented by $\mathbf{R}$). These weights are used in the updating equation (Eq. (6.4)) to correct the model prediction $\mathbf{x}^t$ with the perturbed measurements $\mathbf{y}^0$. This state vector can contain several model variables that are then updated with the measurements $\mathbf{y}^0$. This might also include model parameters like hydraulic conductivity or porosity. In this work, hydraulic head is the only model variable that is used to form $\mathbf{x}^t$.

Data assimilation was implemented in combination with the HydroGeoSphere model, in a C program called EnKF-HGS which manages the forward propagation of the ensemble of HGS model realisations (Eq. (6.1)) and performs the update of the simulated state(-parameter) vector with the measurements (Eq. (6.2)-(6.5)). As the forward propagation of a large ensemble of a highly sophisticated hydrological model like HGS is very CPU intensive, the program is parallelized with respect to the ensemble forward propagation and the filtering step in order to speed up calculations. The parallelization is done by the distribution of different realisations among available CPUs so that each CPU handles the forward propagation and updating steps of a specific subset of the whole data assimilation problem. The interfacing between EnKF and HGS is done via the input and output files of HGS. The input of HGS usually consists of a text file that contains all the information to perform the simulation including the definition of the computational grid, the assignment of boundary conditions, the definition of initial model states and model parameters, solver settings, and so forth. Simulations with HGS are then usually performed by first running a pre-processor called 'GROK', which translates the settings specified in the input text file into binary input files which are the actual input used for the HGS simulator. Certain parts of the text input file for the pre-processor GROK (e.g., transient boundary conditions, definition of parameters, etc.) can be sourced out into separate text input files. This feature is used for defining stochastic transient boundary conditions and stochastic model parameters for HGS in combination with EnKF-HGS. The main program EnKF-HGS writes these specific input files for boundary conditions and parameters separately for each model realisation and each time step. This, of course, requires a proper definition of the respective boundary conditions and parameter settings in the text input files for the pre-processor GROK, i.e., all the input information that should be treated as stochastic needs to be sourced out in separate input files. The exchange of prognostic variables (i.e., hydraulic head) between HGS and EnKF-HGS is managed via binary input/output files that are used by HGS. These binary files are written by EnKF-HGS as an input for HGS for the next model integration. After the HGS- simulation has finished, the predicted state variables are extracted by EnKF-HGS from the binary output file of HGS. This information is later used in the ensemble of state vectors of the data assimilation algorithm. The complete program scheme of EnKF-HGS is summarized in **Fig. 6.2**. Note that the ensemble forward propagation as well as the filtering step are calculated in parallel, meaning that each available processor handles a subset of the ensemble members. The program was first built and tested on a standard Linux cluster and then slightly modified to be able to also run in a cloud environment which is described in the following section.

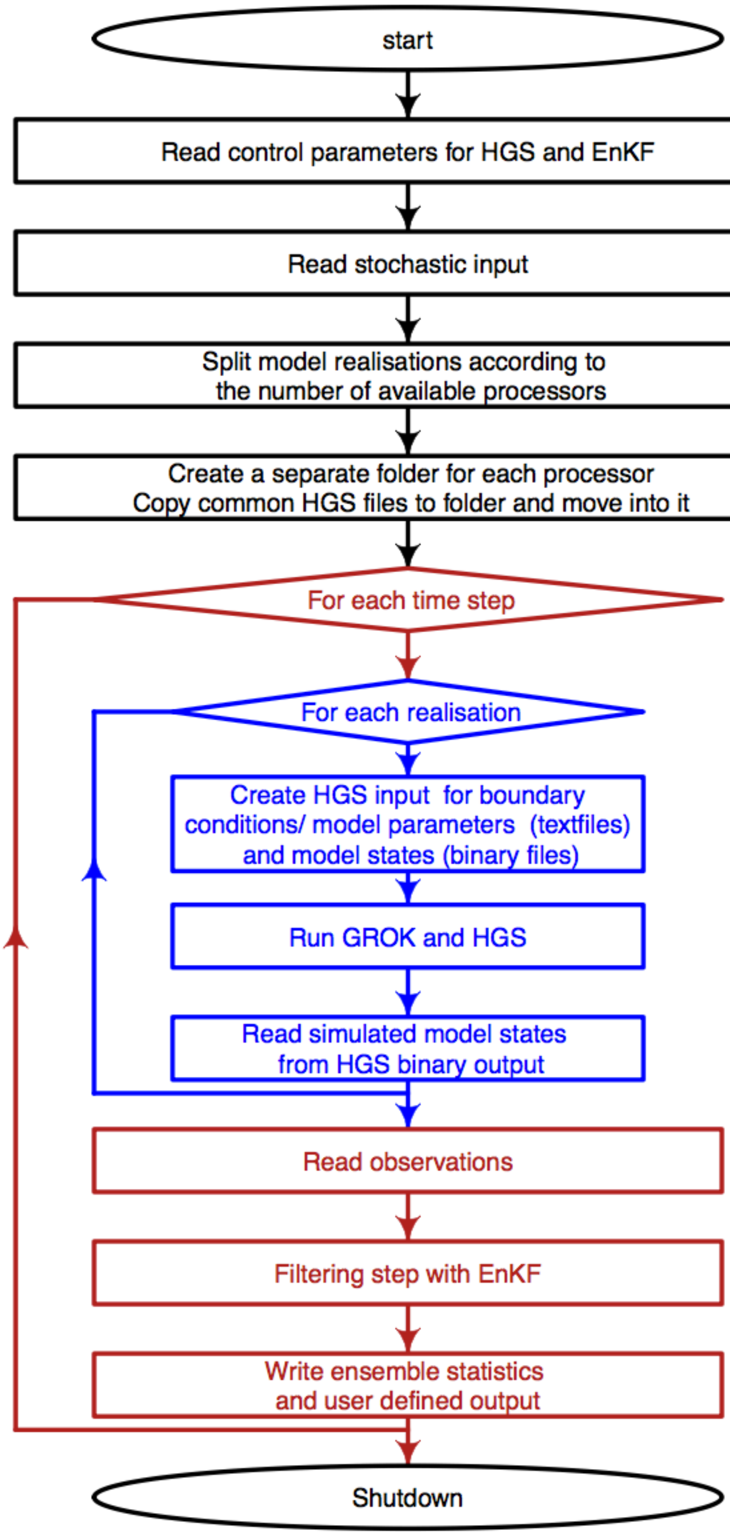


Figure 6.2: Schematic program execution of EnKF-HGS. Black colours mark the initialisation phase of the program, blue colours indicate the ensemble calculations and red colours represent the model forward propagation and filtering step.

6.2.4. Cloud computing with CLAUDE

Cloud computing (Varia, 2008) is a still evolving paradigm, in which various computing resources are delivered to an end-user as services (e.g., Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Software as a Service (SaaS), or XaaS (Everything as a Service)). Virtualization is a fundamental element of cloud computing. In computing, the term *virtualization* refers to provisioning of a virtual resource (e.g., a computer network resource, a storage device, a server) that behaves identically to its physical counterpart. Virtualization allows for physical resource sharing. Virtual resources access the infrastructure through a specialized middleware (i.e., a hypervisor) to improve the infrastructure flexibility and elasticity. While end-users experience cloud computing as an easy-access and easy-to- use technology, cloud computing is a complex and powerful technology with a multitude of important characteristics (Mell and Grance, 2011):

(5) Rapid elasticity and scalability in near real-time via dynamic resource provisioning upon a self-service basis, which allows to on-the-fly adapt system resources to momentary load conditions.

(6) Measured resource consumption: Cloud computing resource usage can be transparently measured, controlled, and reported by both the provider and consumer.

(7) On-demand self-service: The consumer can purchase various services such as applications, networking, or computing using little interaction with the cloud provider.

Broadband access to the cloud infrastructure is provided by the Internet making use of standard network infrastructures and protocols. **Figure 6.3** presents a simplified picture of a) traditional computation and b) typical cloud infrastructures. Traditional computation infrastructure is usually provided by the users themselves, which often results in one of two extremes: (i) Overprovisioning, i.e., a waste of money on the maintenance of non-utilized computational resources, or (ii) Underprovisioning, i.e., limited computing performance due to missing storage or processing infrastructure. Contrarily, cloud users do not need to care about the infrastructure capacity, as they only temporarily purchase computing resources from a cloud provider. The cloud provider is therefore responsible for the maintenance of resources dimensioned for the needs of all the cloud-users. Cloud computing offers SaaS as a major service delivery model. SaaS may provision applications that use a Web browser as their user interface, or predefined Application Program Interfaces (APIs). In the following, we describe our cloud-based simulation service, which we hereinafter refer to as CLAUDE (Lapin et al., 2015), that implements a typical provisioning model for applications delivered as SaaS. CLAUDE provides an easily migratable SaaS supporting a wide range of cloud platforms (e.g., Amazon EC2 (Amazon Web Services, Inc. 2016), OpenNebula (OpenNebula Project 2016), OpenStack (OpenStack Foundation 2016)), while at the same time allowing users to benefit from the features of cloud environments for ordinary, non-interactive, computation-intensive applications commonly used in the high-performance computing domain.

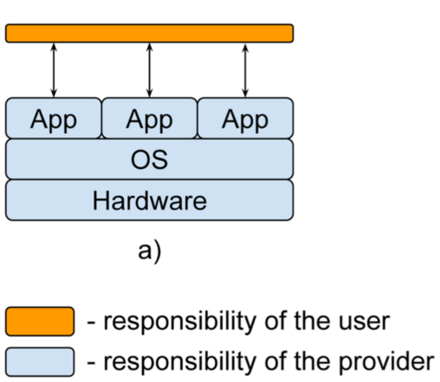


Figure 6.3: (a) Traditional computation infrastructure layers including hardware, operating system (OS) and applications (App); (b) Typical cloud infrastructure layers including virtualization with virtual machines (VM) for three different usage types: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS)

A basic CLAUDE deployment requires a physical or virtual machine for the installation of the CLAUDE core, access to a Cloud Data Storage that supports the Amazon S3 protocol (Amazon Web Services, Inc. 2016) for in- put/output data storage, and access to a cloud resource pool through an Amazon EC2 endpoint. CLAUDE automatically distributes work among available resources, and allows an end-user to balance between minimizing the execution time by exploiting parallel execution (e.g., for Monte Carlo applications) and the final costs by minimizing the amount of consumed resources. CLAUDE allows including data from external data sources (e.g., online databases, WSNs). **Figure 6.4** depicts the integration of the EnKF-HGS real-time modelling using CLAUDE in a cloud environment with a sensor-based live data-flow. The user interacts with the system through a web interface and controls the simulation process (e.g., initialisation, monitoring), the input/output, and the amount of consumed cloud resources. CLAUDE, in turn, orchestrates the running applications on the available computing resources.

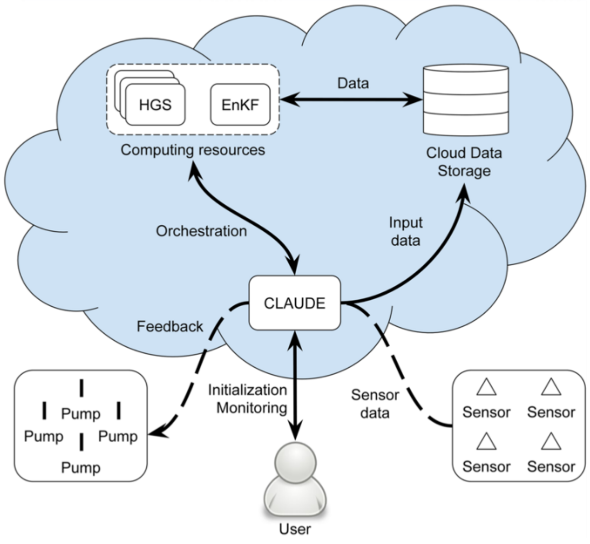


Figure 6.4: CLAUDE-based setup of the EnKF-HGS modelling system.

Accessing and launching EnKF-HGS through CLAUDE has been specifically designed in a very user-friendly way, so that no prior knowledge of running applications on a cloud is required. Furthermore, CLAUDE can easily be generalized for other non-interactive user applications. Once the CLAUDE core is installed and configured, one needs to prepare an Operating System (OS) image for the Virtual Machine (VM) with the required software installed (in our case EnKF-HGS and HGS) and uploading the OS image to the cloud repository. The image is used to create virtual machines, which allow executing the pre-installed software. The uploading process may slightly differ depending on the cloud platform, but the most popular providers (e.g., Amazon EC2, OpenNebula, OpenStack) have step-by-step tutorials explaining the whole process. Additionally, the user is expected to provide the CLAUDE core with a sample Python script that defines execution parameters and input/output locations. After completing all these steps, the system is prepared for execution. As an advantage, CLAUDE provides a set of communication drivers, which implement communication operations in the cloud. This minimizes the necessary changes of the original application source codes. The drivers interact with the CLAUDE core through a standard messaging mechanism provided by a widely used message broker RabbitMQ (Present Pivotal Software, Inc. 2007). CLAUDE therefore allows to easily port an existing application, which was not designed for a cloud environment, to a cloud environment.

In order to port EnKF-HGS (using HGS) to a cloud, the CLAUDE driver was integrated in the program source code, and all the direct execution calls of HGS were replaced with their CLAUDE-based counterparts. This cloudification procedure changes the initial MPI-based execution stack of EnKF-HGS. The cloudified EnKF-HGS separates the main EnKF-HGS simulation loop from the HGS instance, which run on remote computing resources. For this purpose, the CLAUDE driver transmits the relevant input for each HGS model realisation to the Cloud Data Storage, and requests the CLAUDE core to execute corresponding simulations on available cloud computation resources. When all the simulations are completed, the CLAUDE driver retrieves the relevant output from the Cloud Data Storage. The data is then returned to EnKF-HGS, which continues the data assimilation loop. This service oriented approach allows dynamically adjusting the amount of computing resources to momentary workload conditions. Moreover, one can benefit from an external scheduling component for better resource utilization. Additionally, CLAUDE allows minimizing the network Input/Output (I/O) operation using data locality.

6.3. Validation example

For the demonstration of the cloud-based real-time modelling system, the well-known tilted V-catchment problem (e.g., Panday and Huyakorn, 2004; Kollet and Maxwell, 2006) is used to perform a synthetic data assimilation experiment. The tilted V-catchment problem was selected, because it is often used as a benchmark for integrated hydrological models (e.g., Panday and Huyakorn, 2004; Kollet and Maxwell, 2006) and has also been applied in the context of data assimilation with integrated hydrological models before (e.g., Camporese et al., 2009; Bailey and Baù, 2012). The internal flow dynamics of the tilted V-catchment model have furthermore been systematically analysed in original and modified configurations (Gaukroger and Werner, 2011). The original model consists of a V-shaped valley, which is inclined in order to form a sloping stream at the bottom of the valley. The model is forced with one or more precipitation events followed by a recession phase, resulting in distinct discharge peaks at the catchment outlet. The temporal evolution of these peaks also depends on the assigned surface flow properties and the exchange between the surface water and the groundwater domain. Similarly to Gaukroger and Werner (2011), the original V-catchment model was modified to better fit the purpose of this study. First of all, the tilted V-catchment model was extended to allow for more substantial SW-GW interactions: Instead of using the original setup that consisted of a 3 m deep soil-layer to represent the subsurface, a permeable aquifer of 33 m vertical extent was used, allowing significantly more GW dynamics. Secondly, the model was enhanced with eight groundwater wells that simulate typical alluvial groundwater extraction processes for drinking water purposes in the vicinity of the stream. With these modifications to the original model, the modified tilted V-catchment model allows the simulation of substantial SW-GW interactions in a typical alluvial setting with realistic drinking water management. For the synthetic data assimilation experiment, a reference HGS simulation with the modified tilted V-catchment model provides observation data (hydraulic heads) that are subsequently used to correct stochastic model predictions with the EnKF. In addition to the data assimilation aspect, the synthetic experiment used in this study also includes a groundwater management component, where the corrected stochastic model predictions are used to control the groundwater withdrawal of the pumps. This setup represents a simplified groundwater management problem at the catchment scale, where ensemble-based real-time simulations are corrected with observed groundwater levels at regular time intervals and are subsequently used to assist decision makers in the management of groundwater resources (i.e. a synthetic representation of a real world system such as presented by Hendricks Franssen et al. (2011) and Bauser et al. (2010, 2012)). The access to measurement data from the field is treated in a simplified manner in this synthetic experiment by assuming that the observation data are already provided to the cloud infrastructure by the sensor network. For real-world cases, the access to data from the EnKF-HGS modelling and data assimilation platform would involve the data gathering system described in Jamakovic et al. (2013) and Lapin et al. (2014), followed by some steps to treat potentially missing values and to assure the quality of each measurement. Such additional pre-processing of the data is usually quite case-specific and is therefore neglected in our synthetic case.

6.3.1. Hydrological model and data assimilation setup

The modified tilted V-catchment that is used for the data assimilation experiments has a spatial extent of 1620 × 1000 × 33 m and is discretised into 81 × 50 × 11 model nodes. The horizontal discretisation is 20 m in x- and y-direction. The 10 vertical model layers span a total of 33 m and have a variable thickness with increasing resolution towards the top of the model domain (10 m, 10 m, 6 m, 3 m, 2 m, 1.10 m, 0.50 m, 0.25 m, 0.10 m, 0.05 m from bottom to top). The tilted V-catchment topography is formed by assigning a slope of ± 0.05 m/m in x-direction for the western and eastern half of the model. The whole model domain is additionally inclined in y-direction with a slope of -0.02 m/m. The resulting topography for the uppermost model layer is shown in **Fig. 6.5**.

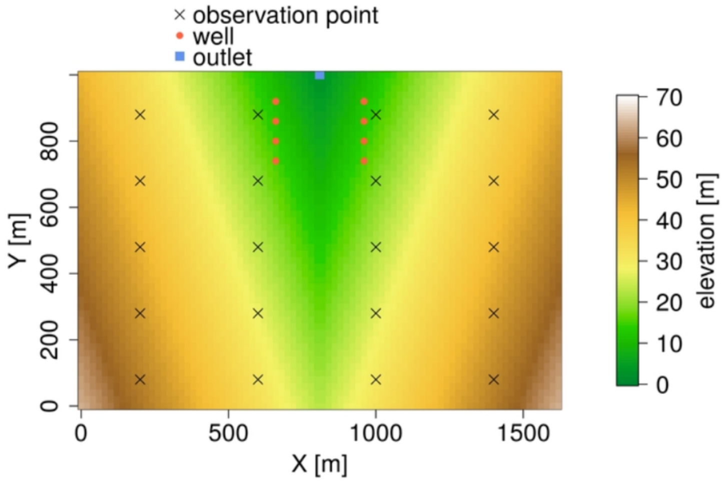


Figure 6.5: Topography and observation network for the modified tilted V-catchment model.

A constant head boundary condition of 23.2 m is assigned to the upstream model face (y = 0 m), which provides a temporally constant base flow to the river. A critical depth boundary condition is assigned to the surface water domain at the downstream model face (y = 1000 m). The lateral model boundaries (x = 0 m and 1620 m) are impermeable. Four wells are placed on either side of the river in the downstream part of the model domain, which operate at a constant withdrawal rate of 0.45 m3/s per well. The hydraulic parameters for surface and subsurface flow are spatially constant and are summarized in **Tab. 6.1**. The total simulation time for the model forward runs is 72 hours, which is discretised in 144 time steps of 1800 s. The reference run is forced with transient time series for rainfall and evapotranspiration (see **Fig. 6.6**). The rainfall time series includes three rainfall events, each with a duration of 6 hours and a precipitation rate of 2.6×10−6 m/s. In-between the precipitation events, evapotranspiration follows a diurnal cycle with an amplitude of 2×10−7 m/s and a duration of 12 hours. **Figure 6.6** shows the temporal sequence of rain and recession phases. During recession, the evapotranspiration follows an ideal diurnal cycle with most of the uncertainty occurring around noon.

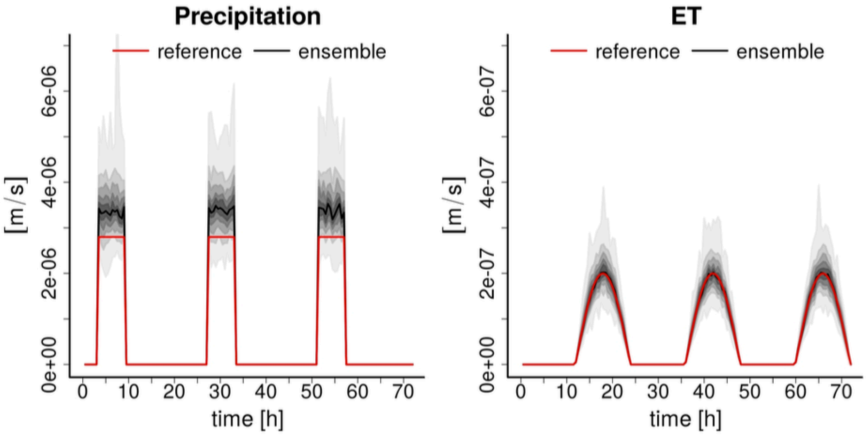


Figure 6.6: Precipitation (left) and evapotranspiration (right) rates for the reference run (in red) and the ensemble simulations (in grey). The different grey areas represent 10%-percentiles of the ensemble distribution and black lines represent the ensemble median value.

The observation data (hydraulic heads) are taken from the reference simulation. In total, 20 observation points are used. The observations are made on the bottom layer of the model (see **Fig. 6.5**), and are assimilated every time step (i.e. every 1800 s) with the EnKF. A measurement error of 0.01 m is assumed for these observation data. The ensemble of model realisations (100 ensemble members) is generated by perturbing the rainfall and evapotranspiration rates with multiplicative noise drawn from a log-normal distribution with a standard deviation of 0.3 (see **Fig. 6.6**). Additionally, a constant bias of 10% is added to the precipitation rates. Precipitation input for hydrological models is typically derived from radar or rain gauge data, or a combination of both. The precipitation rates derived from these measurement techniques are usually associated with a considerable uncertainty range (Liu et al., 2012). The uncertainty in the precipitation input and the additional bias used in this setup intends to resemble the uncertainty of these precipitation measurements. **Figure 6.6** shows the assigned uncertainties of precipitation and evapotranspiration in relation to the reference case. Uncertainties in precipitation are quite large and cover the reference precipitation.

Table 6.1: Model parameters for the modified tilted V-catchment problem.

Variable	Value	Unit
Saturated hydraulic conductivity K	2.9×10^{-3}	m/s
Porosity	0.43	$m^3\ m^{-3}$
van Genuchten α	3.48	m^{-1}
van Genuchten n	1.75	-
Residual saturation (S_{wr})	0.05	$m^3\ m^{-3}$
Manning's coefficient	0.15	$s/m^{1/3}$
Coupling length (l_{exch})	1×10^{-7}	m
Rill storage height	0.1	m

In addition to the assimilation of hydraulic head observations, a simple groundwater management scheme is used in the simulations. This illustrates that the cloud-based EnKF-HGS modelling and data assimilation platform can also be used for real-time adaptation of water resources management, for example to avoid negative impacts of excessive pumping. The example in this paper targets to maintain a minimum 'ecological' river flow by adjusting the well pumping rates based on the simulated ensemble river discharge. The management rule is as follows: If the 30%- percentile of the simulated discharge distribution of the ensemble falls below a threshold value of 1.6 m3/s, the pumping rates are reduced by 5% in the following time step.

6.3.2. Simulation results

Three scenarios were considered with the setup described above, with each scenario including an open-loop run (runs without feedback, i.e., no assimilation of hydraulic head data) and an assimilation run: (I) No well management for the reference and the ensemble simulations. This scenario is intended to first verify the effectiveness of the data assimilation within this setup. (II) Well management for the ensemble simulations but not for the reference. This scenario is intended to test the stochastic well management in conjunction with data assimilation with a large bias between observations and model predictions. (III) Well management for ensemble and the reference simulations. In this case, the measurement data from the reference run are also influenced by the well management which would be typically the case in real-world settings.

Figure 6.7 shows the temporal evolution of simulated discharge for the open-loop simulation and the assimilation experiment without well management (scenario I). The simulated discharge for the open-loop simulation shows a positive bias (compared to the reference), which is related to the assigned bias of 10% in the precipitation input data. The spread in the simulated discharge is more pronounced during the rainfall events. During the recession phase, however, the positive bias in the precipitation input data leads to a mismatch that is increasing with time: in the open-loop simulation the overall groundwater withdrawal through the wells is slightly lower than the precipitation input, compared to the reference simulation. The assimilation of hydraulic head values from the reference run leads to a correction of these discharge values towards the reference discharge, particularly during the recession phases but also during the rainfall events, and especially in the later phase of the simulation period. The correction of simulated groundwater levels towards the observations leads to a correction of the mass balance within the aquifer, which indirectly improves the integrated discharge signal and effectively adjusts it to the true observations.

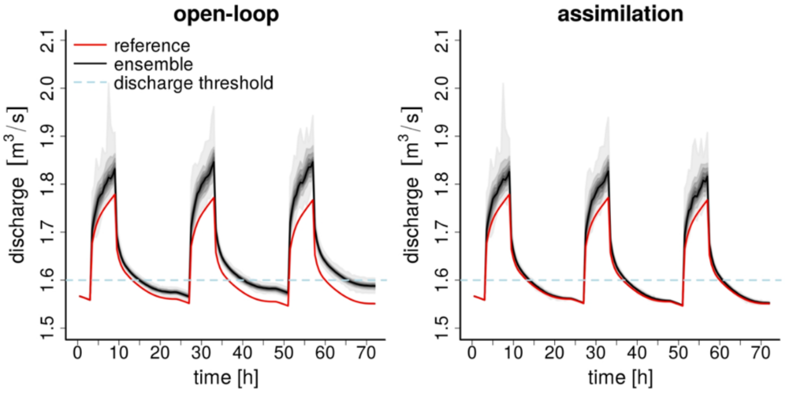


Figure 6.7: Simulated discharge for the open-loop simulations (left) and the assimilation experiment (right) of scenario I. The red lines show the discharge for the reference run and the dashed blue lines mark the discharge threshold for the well management (which, however, was not enforced in this scenario). The different grey areas represent 10%-percentiles of the ensemble distribution, and the black lines represent the ensemble median value.

The improvement of model predictions was also quantitatively assessed by calculating the Mean Absolute Error (MAE) for hydraulic heads and discharge:

$$\mathrm{MAE}(x) = \frac{1}{N_n - N_t} \sum_i^{N_n} \sum_j^{N_t} |\bar{x}_{ij} - x_{ij}^{ref}| \qquad (6.6)$$

here N_n is the number of model nodes (only used in case of hydraulic heads), N_t is the number of model time steps, $\bar{x}_{ij}$ is the simulated ensemble mean of hydraulic head or discharge and x_{ij}^{ref} is the corresponding reference value. For scenario I (no well management), the MAE of hydraulic heads improved from 0.48 to 0.21 m with the assimilation of hydraulic head values. The MAE of discharge reduced from 0.033 to 0.015 m^3/s. **Figure 6.8** shows the spatial distribution of MAE values for hydraulic heads (averaged over the simulation time) for the open-loop simulation and the assimilation run. It can be seen that the assimilation of hydraulic head data in this setup has a significant positive effect on the predicted hydraulic heads, which leads to a significant reduction of errors in hydraulic head and discharge of about 50% compared to the open loop simulations.

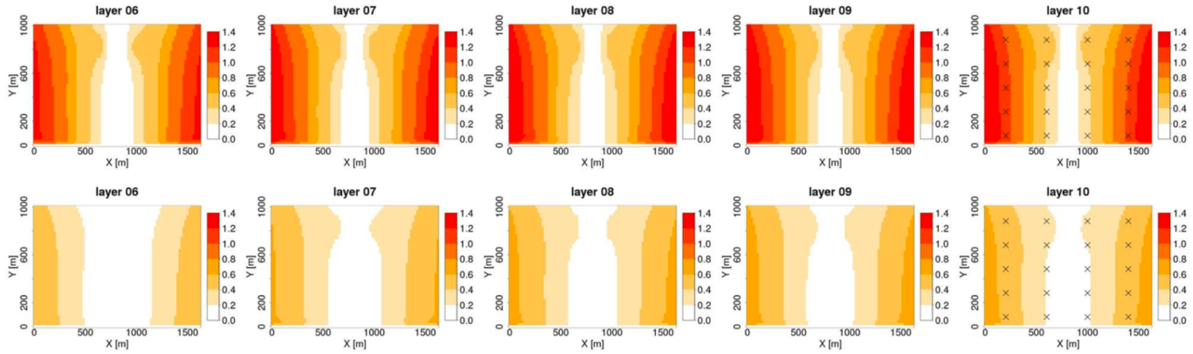


Figure 6.8: Mean absolute error of hydraulic heads for the five uppermost model layers for the open-loop simulation (upper row) and the assimilation run (lower row) of scenario I.

Results for the discharge behaviour of scenario II are shown in **Fig. 6.9**. In both the open-loop and the data assimilation cases, the reduction of pumping rates on the basis of the management criterion effectively keeps the simulated discharge around the predefined discharge threshold value of 1.6 m3/s. In both cases, slight oscillations in discharge occur, which are related to a switching off of the management criterion during the recession phases. When the discharge threshold value of 1.6 m3/s is reached, the pumping wells are shut off until the management criterion is again satisfied, which keeps the discharge of the following time steps above the threshold. The lag time between two consecutive activations of the pump management depends on the response time of the groundwater flow field to the pumping in the wells, which is relatively short in this particular case. At certain times, slightly lower discharge values can be observed in the assimilation experiment during the recession phases, compared to the open loop simulation. This is related to the fact that the correction of simulated groundwater levels in the assimilation run leads to a decrease of the simulated discharge towards the values of the reference run (see **Fig. 6.7**), which also affects the scheduling of the well management. Overall, the management module provides an effective way to keep the discharge around the predefined level (i.e., minimum ecological flow) even if there is a quite large discrepancy between the ensemble simulations and the actual measurements. The MAE values for hydraulic heads and discharge for scenario II are higher than for scenario I due to the management in the ensemble simulations, which is not present in the reference case. MAE of hydraulic head decreases from 0.48 to 0.28 m and MAE of discharge decreases from 0.05 m3/s to 0.04 m3/s in the assimilation run.

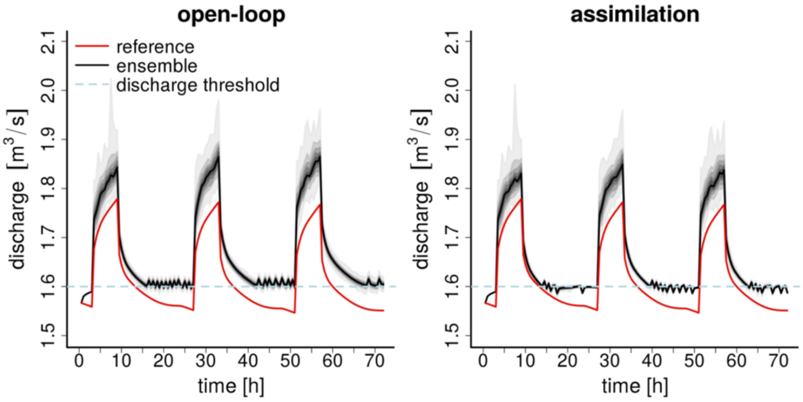


Figure 6.9: Simulated discharge for open-loop simulations (left) and assimilation experiment for scenario II. The red lines show the discharge for the reference run and the dashed blue lines mark the discharge threshold for well management. The different grey areas represent 10%-percentiles of the ensemble distribution, and the black lines represent the ensemble median value.

For scenario III, both the ensemble simulations as well as the reference case include well management. Now, the measurements from the reference case better resemble the aquifer response to the management in the ensemble simulations as compared to scenario II (where measurements are not influenced by the management). However, due to the uncertain and biased precipitation input in the ensemble, there is a discrepancy between the measurements (from the reference run) and the ensemble model predictions, which nonetheless needs to be corrected by EnKF. In **Fig. 6.10**, one can see that the open-loop simulations and the reference show a very similar behaviour during the recession phase where management takes place. For the open-loop simulations the predicted discharge is slightly higher and the amount of correction with the management module is lower than in the reference case, which is due to the positive bias in precipitation. When measurement data are assimilated (right hand side of **Fig. 6.10**), the ensemble discharge prediction in the recession phase is almost identical to the reference, which shows that the precipitation uncertainty can effectively be corrected for by data assimilation in combination with the management module. For scenario III, the MAE values are reduced to a similar magnitude as in scenario I, with a reduction from 0.48 to 0.21 m for hydraulic heads and 0.024 to 0.015 m3/s for discharge.

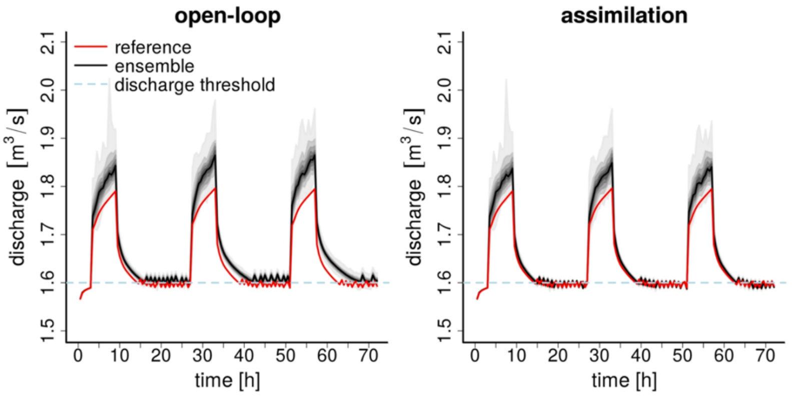


Figure 6.10: Simulated discharge for open-loop simulations (left) and assimilation for scenario III. The red lines show the discharge for the reference run and the dashed blue lines mark the discharge threshold for well management. The different grey areas represent 10%-percentiles of the ensemble distribution, and the black lines represent the ensemble median value.

6.3.3. Performance analysis of the EnKF-HGS modelling and data assimilation platform in the cloud environment

The EnKF-HGS simulator uses a Monte Carlo approach with an iterative workflow. In every iteration, two essential phases can be identified: (i) simulation of an ensemble (i.e., a set) of model realisations (referred to as the forward simulation phase), and (ii) an update, or feedback, phase for state vectors (and optionally parameters) (referred to as the filtering phase). One forward simulation phase computes all individual model realisations of the ensemble. Computing a single model realisation requires executing two applications, i.e., GROK and HGS. GROK is a pre-processor of HGS, which prepares the input files for HGS. The run-time of each HGS realisation strongly depends on input parameters and model complexity, but normally, HGS is a comparably long-running and compute-intensive process due to a high non-linearity of the hydrological processes that are simulated. Moreover, due to a large number of parallel model realisations in the forward simulation phase, the main demand for computing power in EnKF-HGS modelling platform comes from the execution of HGS. In the cloudified version of the EnKF-HGS modelling platform, the for- ward simulation phase is therefore distributed over multiple cloud computing resources. The filtering phase remains centralized.

The focus of this analysis therefore is on the performance of the forward simulation phase (section 6.3.4.1) by measuring the execution time of each implementation. However, execution results of the filtering phase are also presented and discussed (section 6.3.4.2). The execution times of two implementations of the EnKF-HGS platform are directly compared in sections (sections 6.3.4.2 and 6.3.4.3). In particular, CPU utilization, scheduling, and data locality aspects are discussed.

6.3.4. Setup of performance comparison

As discussed in section 6.2.4, the original EnKF-HGS was designed for execution in a traditional cluster environment (i.e., MPI-based), which normally provides: (i) a low latency broadband network (e.g., InfiniBand, Myrinet 10G), (ii) access to a Network File System (e.g., NFS), and (iii) homogeneity of computing resources. None of these properties are guaranteed in a typical cloud environment. This can result in a dramatic performance loss of EnKF-HGS. In order to properly compare the performance of the regular and the cloudified EnKF-HGS, both versions were executed on a local computing infrastructure, rather than the cloudified version being executed on a different infrastructure. Two distinct execution approaches can be identified: Approach 1: an execution of the original MPI-based EnKF-HGS in a virtual cluster environment (i.e., a single VM resource pool), and Approach 2: execution of the cloudified EnKF-HGS with the CLAUDE driver integrated in a cloud environment. Each approach uses a different network file storage. In Approach 1, a VM-based deployment of GlusterFS (Red Hat, Inc. 2016) was used. The storage nodes were organized in a distributed volume with no data replication employed to maximize storage performance. On the computing nodes, the FUSE-based Gluster Native Client (Red Hat, Inc. 2016) was installed, enabling highly concurrent access to the file system. In Approach 2, a VM-based deployment of Riak CS (Basho Technologies, Inc. 2016) was used.

The modified tilted V-catchment model served as a test case. In the experiment, the model performs 144 iterations with an ensemble of 100 model realisations. Since all realisations are independent HGS instances, a set of 4 experiments with varying numbers of CPU cores was conducted: (a) 10, (b) 20, (c) 50, (d) 100. In each experiment, one single CPU core was assigned per HGS instance. In the original MPI-based implementation of EnKF-HGS, this resulted in 10, 5, 2, and 1 HGS executions per CPU core, for (a), (b), (c), and (d) respectively. Whereas for the cloudified version of EnKF-HGS, CLAUDE automatically distributed the execution of the HGS instances over available computing resources for every iteration, using the First-In-First-Out (FIFO) algorithm, which maximized resource utilization. **Tab. 6.2** summarizes the characteristics of selected VM instances for both execution approaches. All VMs were connected via 1GB/s network.

Table 6.2: VM instance selection for the execution of the original MPI-based implementation and the cloudified version of the EnKF-HGS modelling platform. As the memory requirement per HGS instance for the test model is less than 500 MB, the available memory per computing VM is sufficient.

	Approach 1: MPI-based implementation		**Approach 2: Cloudified version**		
VM role	EnKF-HGS platform execution	storage	HGS execution	EnKF-HGS filtering execution	storage
Number of VMs	2/3/7/13	3	2/3/7/13	1	1
CPU cores per VM	8	4	8	1	4
Memory (GB)	7.4	3.5	7.4	2.5	3.5

6.3.4.1. *Performance analysis of the forward simulation phase*

In this first experiment, only the duration of the forward simulation phase is considered. **Figure 6.11** illustrates the computing time ratio between Approach 1 and 2 against the iteration step. In Approach 1, this time is defined as the elapsed time between the beginning of the forward simulation phase and the end of the longest running MPI process. In Approach 2, it is the time interval between the beginning of the forward simulation phase and the end of the last finishing HGS instance. In Approach 2, the overhead of input/output transmissions and cloud scheduling is therefore also accounted for. The execution time of the forward simulation phase for the two approaches is comparable. A small difference of 4-15% in favour of the original MPI-based approach can be observed in experiments (c) and (d).

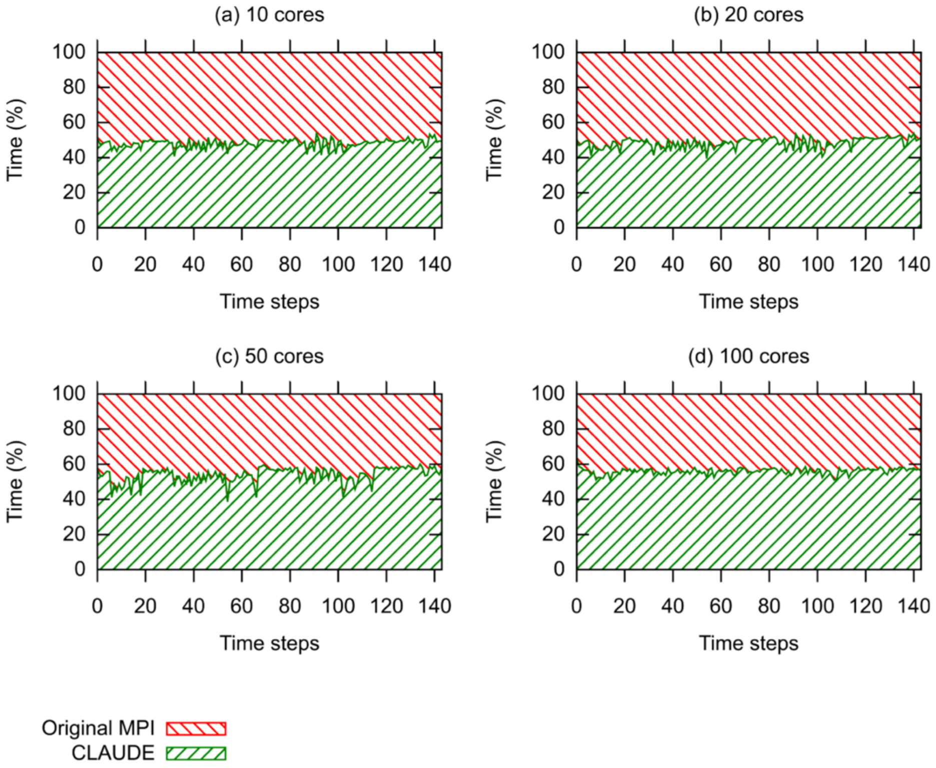


Figure 6.11: Relative computation time of the forward simulation phase for the two different execution approaches.

6.3.4.2. *Performance analysis of the filtering phase*

In contrast to the forward simulation phase, the filtering phase was not distributed, as it requires significantly less computing power and thus much shorter execution times. The EnKF-HGS modelling platform performs filtering on initially allocated system resources. In Approach 1, it includes the entire resource pool ranging from 10 to 100 CPU cores. In Approach 2, only 1 CPU core is initially provided to the filtering step (see **Tab. 6.2**). The relative execution time ratios of the filtering phase of Approaches 1 and 2 are shown in **Fig. 6.12**. Surprisingly, in (d), the filtering phase running on 100 CPU cores is 20% slower than on a single CPU. This difference in performance is a result of the network connection of a virtual cluster in the cloud environment. A regular physical cluster guarantees high throughput and low latency. On a cloud, physical machines can be geographically distributed, and therefore might be connected with links of lower capacity and higher latency. In our experiments, cloud workers were connected with a regular 1 GbE-T Ethernet connection, which is considered slow from the point of view of a regular cluster environment. As a result, the data transmission time between multiple MPI processes surpassed the benefit of parallel computation. The regular, MPI-based

EnKF-HGS setup is therefore not suited for cloud computations with heterogeneous or low capacity resources.

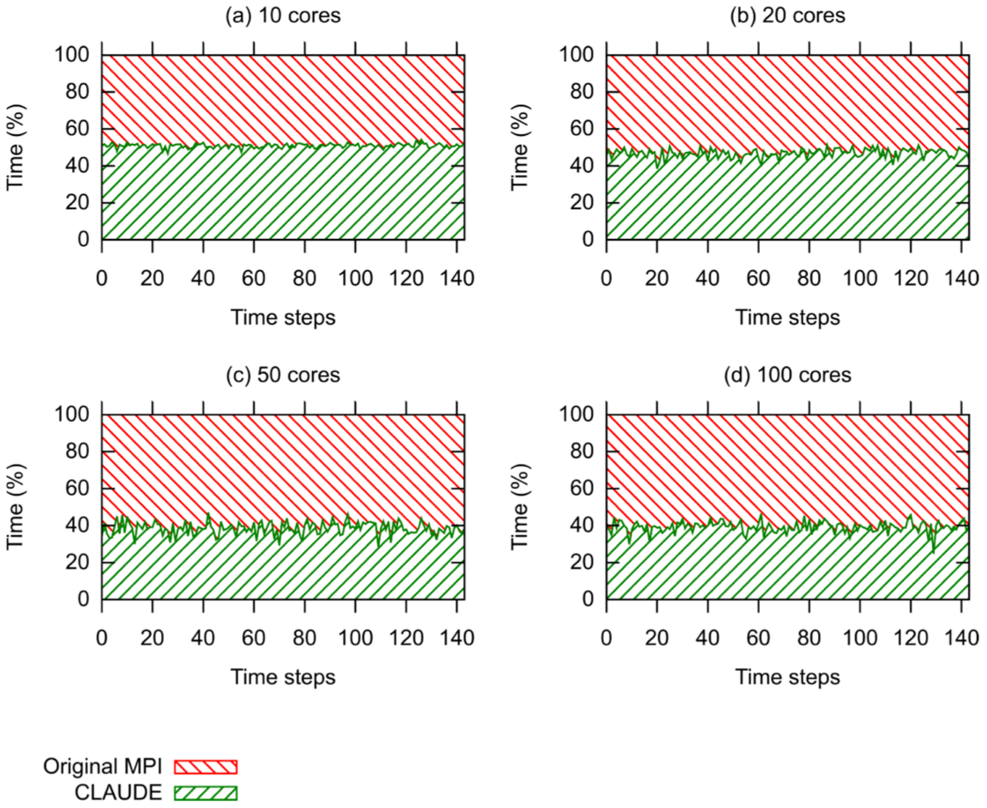


Figure 6.12: Relative computation time of the filtering phase for two execution approaches.

6.3.4.3. Performance analysis of the original implementation

In order to measure the impact of the concurrent access to a network file storage on the execution time of the HGS simulator, it needs to be assured that each HGS instance performs the same amount of I/O operations. We therefore use a similar technique for assessing I/O saturation as Caino-Lores et al. (2016), running a similar I/O stress experiment in a cloud setup, which is described in **Tab. 6.2**. In our case, we use the tilted V-catchment problem as the input hydrogeological model. To this end, a modification of EnKF-HGS to compute the same model *N* times in a row instead of simulating *N* different realisations was made. The modified EnKF-HGS modelling platform was executed using a range of 1 to 64 available CPU cores (see **Fig. 6.13**) to measure the execution time of a HGS instance in every step against the iteration count. The HGS performance plots resemble the results of Caino-Lores et al. (2016). They are, however, skewed towards longer relative execution times for a growing number of concurrent HGS instances. As shown in **Fig. 6.13**, the concurrent access to network file storage significantly affects the performance due to increased I/O latency. This is a result of the saturation of the network link. In this experiment, however, the concurrent I/O operations were artificially equalized. This resulted in a large performance drop as all individual simulations last for the same amount of time and thus highly stress the storage by performing I/O access at the same time. In a realistic ensemble simulation, however, the influence of concurrent execution is less likely to result in such a large number of overlapping I/O operations of parallel HGS instances on input and output.

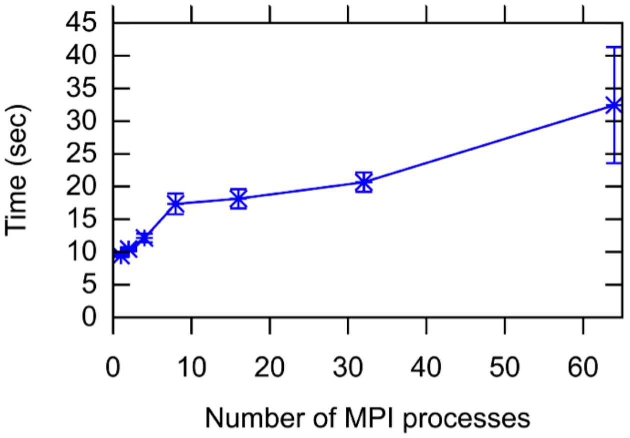


Figure 6.13: Impact of the concurrent access to a network file storage on the execution time of the HGS simulator.

The second important aspect of the MPI implementation is a statically assigned number of model realisations per MPI process. This is a typical design choice of an MPI-based application due to a lack of load balancing, e.g., through a built-in job-stealing mechanism. In an ensemble with individual simulations of heterogeneous duration, EnKF-HGS might result in relatively poor performance, because some MPI processes finish all assigned simulations sooner than others, and EnKF-HGS needs to wait for all individual realisations to finish before the filtering step can be performed. Therefore, the CPU idle time during the simulation was measured. **Figure 6.14** shows the CPU idle time per iteration for the four experiments (a), (b), (c), and (d) described above. The majority of the realisations, or ensemble members, have a comparable simulation time, which results in a relatively short CPU idle time. However, in some cases, especially when a single MPI process executes multiple assigned simulations, the CPU idle time grows up to 150 seconds, which corresponds to 34% of the overall iteration time. For more complex scenarios than the modified tilted V-catchment, with longer simulations runs, the overall execution time of EnKF-HGS could increase substantially due to the possibly long wait for all individual realisations to finish.

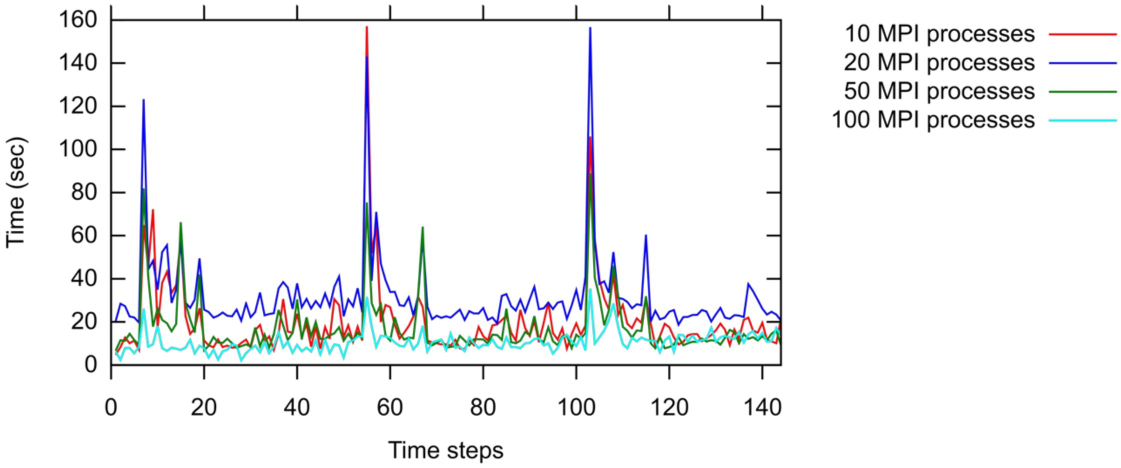


Figure 6.14: CPU idle time duration.

6.3.4.3.1. Performance analysis of the cloudified version

The forward simulation phase in the cloudified EnKF-HGS accommodates three individual, time-consuming pro- cesses: (i) input / output data transmission, (ii) scheduling, and (iii) the simulation phase. In the initial four experiments, the duration of every phase was measured. Moreover, we separated the communication overhead (i.e., worker overhead) of the HGS computing resources with the CLAUDE core from the actual simulation time. **Figure 6.15** shows the relative duration of all these processes. The time spikes in the simulation phase correspond to long-running realisations. **Figure 6.16** shows the absolute overhead. The I/O data transmission time is almost constant for the full ensemble execution, as well as, for the duration of each individual realisation. The total overhead slightly decreases with the growing number of the parallel instances of HGS, but is constant for the duration of one single realisation. The scheduling time grows with the amount of available resources, however, is constant for the duration of each realisation. In practice, this allows parallel execution of complex hydrogeological models with long-running individual realisations in the cloud, while still having comparable constant overhead. While CLAUDE is still in a proof-of- concept phase, it already indicates the benefits of the service-oriented execution approach with external scheduling and enhanced data locality techniques. By using an I/O-optimized implementation and a high-performance scheduler the total overhead could be significantly reduced, as at the moment it corresponds to 40% to 80% of the total overhead.

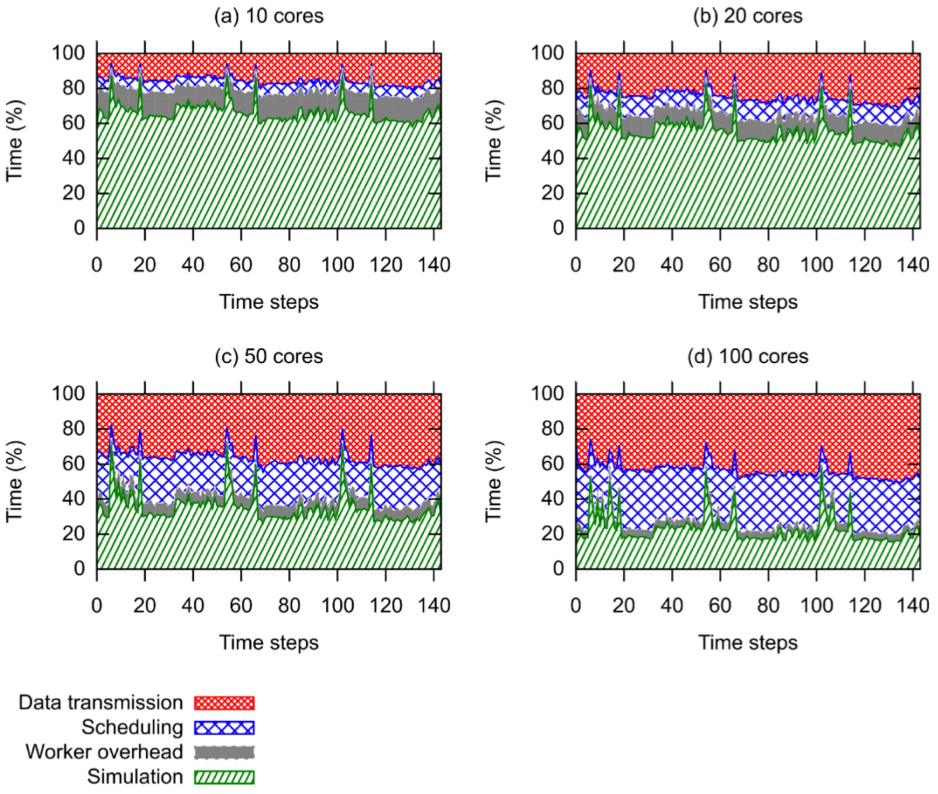


Figure 6.15: Processes involved in the forward simulation phase for the cloudified modification of the simulator.

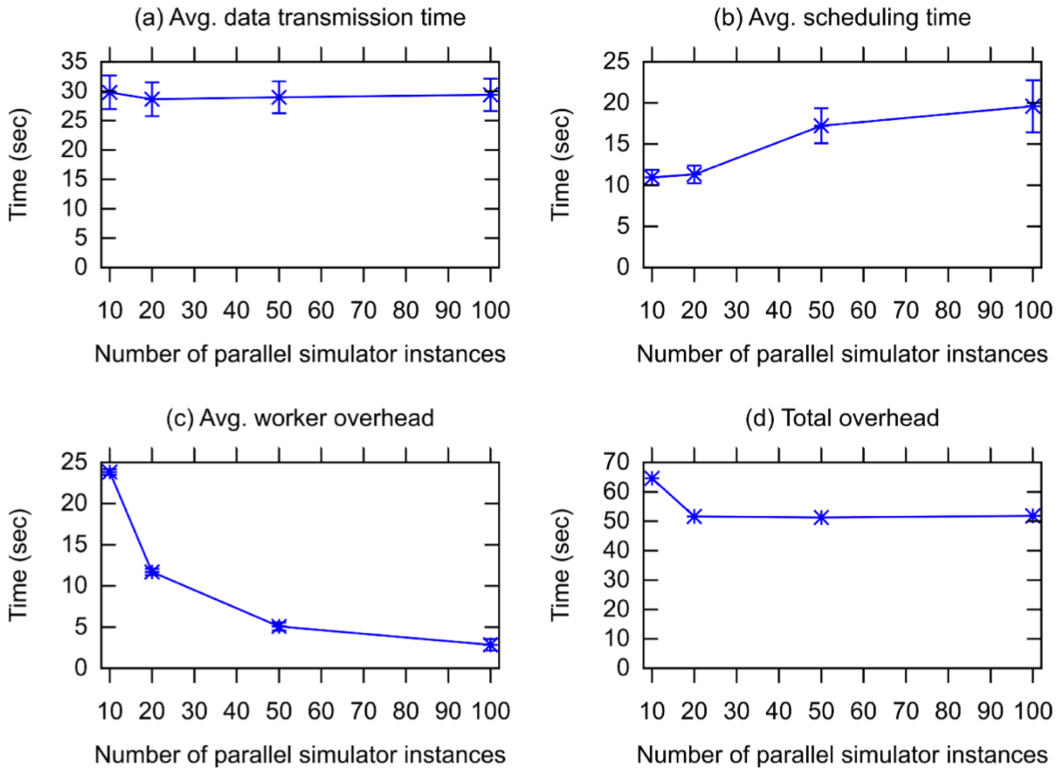


Figure 6.16: Absolute time of the overhead processes.

6.4. Discussion

The data assimilation and modelling platform described in this paper provides a fully-operational architecture for a real-time water management system using the integrated hydrological model HydroGeoSphere in a cloud computing environment. This system is able to provide stochastic predictions of key hydrological variables, which can be continuously improved by assimilating real-time measurement data acquired through wireless sensor networks. The stochastic predictions can then be used to optimize water management in a subsequent step. As individual components of such a real-time modelling system, like online data acquisition, data assimilation, or real-time optimization, have already been described in literature (e.g. Bogena et al., 2010; Hendricks Franssen et al., 2011; Bauser et al., 2012), this paper intends to show the integration of these different methodological advances into an overarching water management framework. An important aspect of this overarching framework is the deployment of such a system in a cloud computing environment. Given the fact that stochastic simulations with integrated hydrological models are usually very CPU-intensive, practitioners in water management may often shy away from the investments in computational infrastructure that is necessary for performing stochastic predictions with such kinds of models. However, integrated hydrological models are useful for many water management problems, especially, for systems with coupled surface water - groundwater flow. In such cases, integrated models have a clear advantage over traditional groundwater models due to their physically more consistent process description of water exchange and surface water routing, which enhances, e.g., the prediction of discharge. In addition, stochastic model predictions in conjunction with data assimilation techniques allow to assess the uncertainty of model predictions and to incorporate field measurements sequentially into the modelling process, which is particularly useful in the context of flood forecasting and contaminant transport.

Cloud technology has the advantage that the computation time for such kind of simulations can be requested on an on-demand basis, meaning that there is no financial and personal overhead for acquiring and maintaining the required computational infrastructure. In this work, the initially cluster-based EnKF-HGS data assimilation system for HydroGeoSphere was translated to a cloud infrastructure with only minimal changes in the original model code and a linkage to pre-existing management tools for the cloud-based computation. These implementation changes were relatively straightforward as EnKF is a Monte Carlo-based data assimilation approach. In the Monte Carlo approach a whole set, or ensemble, of slightly differing model realisations is integrated forward in time to approximate the uncertainties and mutual correlations within the model. This ensemble of simulations can easily be distributed among different CPUs or nodes, and a cloud infrastructure, therefore, provides an ideal environment for such tasks. These findings are also promising for the usage of other, already existing, simulation and data assimilation platforms, since we have been able to show that only a minimal effort is required to port such systems for the use in a cloud infrastructure. Our comparison of the performance of the EnKF-HGS data assimilation platform in a cloud-based and cluster-based computation environment showed that the cloud implementation produced an affordable overhead with respect to data transfer and execution time. Considering the cloud execution, we created the cloud-oriented service CLAUDE that allows distributing an application over available cloud resources. As a result, we obtained a functional, cloud-based version of EnKF-HGS with an execution performance that is comparable to a cluster-like environment, and a predictable service-related overhead. The usage of CLAUDE gives us the possibility to rely on external services for automatic job distribution, efficient job scheduling and optimal utilization of computational resources, which in turn facilitates the performance of the application and eases the application development process. Additionally, CLAUDE offers a well-integrated data storage service, which lowers the demands for the target infrastructure.

The comparison between the cloud and the cluster infrastructure was made with a relatively simple model setup with considerably smaller problem size and execution time than real-world applications. On the one hand, a larger number of model grid cells, as encountered in many practical applications, would increase the size of the data packages that need to be transferred through the network in the cloud-based execution. This would certainly increase the overhead related to data transfer as compared to a cluster-based solution. On the other hand, however, the simulation times for more complex real-world model setups would be considerably higher than the model that was used here in the comparison study. As a consequence, the relative overhead for data exchange in the filtering part of the cloud execution for more complex models should still be in a comparable range to what has been found in this study.

The application of the proposed cloud-based real-time modelling platform to real-world applications, of course, raises some additional issues. One of them is the question of data security/data ownership in cloud-based services. Field measurements as well as simulation results may have to be kept confidential by water managers. This means that the data policy of the cloud provider needs to be thoroughly checked for ownership rights before a data acquisition or simulation system is deployed in such environments. In addition, water managers need to be sure that they have a continuous access to their data and modelling systems, because they usually need these services on an operational basis. If the connection to these services fails, for example, related to a problem with the cloud provider, this could result in considerable problems for water managers, for example, when such a system is used for short-term predictions like flood forecasting. Therefore, when such a system is deployed in a cloud-computing infrastructure, a reliable access and functioning of the services needs to be guaranteed. This could be achieved, for example, by providing a backup management system on a second cloud. However, as cloud computing is a continuously growing business, and more and more IT companies move towards cloud-based infrastructures, water managers have some flexibility with respect to potential cloud providers. Hence, when a cloud-based modelling and monitoring platform needs to be deployed in practice, practitioners can choose between a variety of different cloud service providers in order to meet their specific requirements for data security and operational stability. Another issue in real-world applications is the connection of the modelling system to the online data storage of any kind of wireless sensor networks. In our test-case it was assumed that the measurement data are readily available in the cloud infrastructure. However, in real-world applications it is also necessary to account for sensor failure, non-physical values of the measurements, disconnection from the measurement data base and the quality of the measurement data. This would require an additional pre- processing step which includes an exception handling for such events, which needs to be adjusted to the actually used data type and sensor network. The cloud-based modelling and data assimilation system proposed in this paper is, in general, specifically designed for the assimilation of hydraulic heads with the integrated hydrological model HydroGeoSphere. Most parts of this system are not generic, meaning that the assimilation of another data type (e.g., concentration data) would require additional programming effort in the data assimilation module. Additionally, if another hydrological model with a different model input structure should be used in conjunction with the cloud-based data assimilation, the interface between this model and the data assimilation system would need to be re-programmed and the CLAUDE driver would have to be adjusted to the new model input structure. Our experience from the setting up of the data assimilation system is that the coupling of a new model with the data assimilation software constitutes a large part of the work. Moving the execution from a cluster-based environment to the cloud computation was, on the other hand, relatively straight-forward given the fact that the CLAUDE system took care of the management and scheduling of the ensemble simulations.

6.5. Conclusions

In this paper we established a cloud-based real-time modelling and data assimilation platform for the integrated hydrological model HydroGeoSphere. As a first step, a data assimilation system was created that is able to perform stochastic predictions with HydroGeoSphere, taking into account uncertainties in model parameters as well as bound- ary and initial conditions. Additionally, the system allows to assimilate hydraulic head data sequentially with the ensemble Kalman filter. Updated model predictions may then be subsequently used to perform real-time management of water resources. In a second step, the originally cluster-based data assimilation and modelling platform was adapted to a cloud computing environment. The necessary changes in the code structure were minor and basically affected the forward propagation of the different model realisations: in the originally cluster-based implementation, the ensemble integration was achieved by spawning a certain number of MPI processes on a computing cluster whereas in the cloud-based implementation the calculation of different model realisations was achieved by distributing the workload among different nodes in the cloud environment with CLAUDE. This workload distribution involved the sending of necessary input data to the nodes, the forward simulation on the worker nodes and the subsequent retrieval of simulation results from these nodes. A performance comparison of the cluster-based and cloud-based implementation with a relatively simple tilted V-catchment model showed that the overhead in the cloud-based implementation is within an affordable range, showing that such a system is also well-suited for real-world applications.

Generally, we expect that stochastic real-time simulations using integrated, physically-based hydrological models are going to become more common in practical water resources management applications. The reason is that integrated hydrological models provide major advantages over traditional groundwater models with respect to process description and the coupling between groundwater, surface water, and land surface processes, which are, for example, useful for flood and drought predictions and water balance calculations. Furthermore, there is a growing awareness that uncertainties in model predictions need to be taken into account for successful water management. Along with that, data assimilation offers the opportunity to correct such uncertain model predictions with measurement data, which are increasingly collected by online sensors. Data assimilation can be used in conjunction with real-time optimization methods to improve water resources management. The deployment of such a hydrological data assimilation system in a cloud environment, as shown in this paper, offers several advantages for the application in practical water management problems: (i) Cloud computing resources can be requested on an on-demand basis, which reduces the financial overhead for acquiring and maintaining the computation infrastructure that is required for the computation- ally demanding ensemble predictions with integrated hydrological models. (ii) Moving to cloud-based computations is relatively straightforward for ensemble-based methods like the ensemble Kalman filter, which only require minimal code changes in the forward propagation of the model realisations. (iii) The computational overhead for cloud-based implementations of ensemble predictions is in an affordable range, as shown for the tilted V-catchment problem in this paper (iv) Cloud infrastructure is able to provide a seamless integration of observation data acquired in the field through wireless sensor networks and stochastic real-time simulations with integrated hydrological models for the purpose of data assimilation. Additionally, assessment and control of such systems can be achieved through customized web services, which ease the usability for practitioners.

In this paper, we describe the architecture of a fully operational system for a cloud-based modelling and data assimilation using the integrated hydrological model HydroGeoSphere. Such systems might assist to tackle a wide range of hydrological problems, like the adaptive operation of well fields, flood forecasting, monitoring and management of contaminated sites or optimal irrigation scheduling. In the future, we plan to set up such a system for the Emme catchment in Switzerland, which is equipped with a wireless sensor network for several hydrological variables, and where accurate predictions of discharge and the effect of water withdrawal are highly desired by water managers. This will provide more insights into the practical implementation of cloud-based services in the context of water management.

6.6. Acknowlegements

W. Kurtz gratefully acknowledges the funding provided by the German Research Foundation (DFG) for project SFB-TR32 “Patterns in soil-vegetation-atmosphere systems: monitoring, modelling and data assimilation”. The University of Bern and University of Neuchâtel acknowledge the “AAA/SWITCH – e-Infrastructure for e-Science” programme supporting the development of the acquisition infrastructure via the “Authentication, Authorization, Accounting, and Auditing in Wireless Mesh Networks (A4-Mesh)” and “Easily Deployable A4 Wireless Mesh Networks (eA4-Mesh)” projects. Moreover, the University of Bern and University of Neuchâtel want to express their greatest gratitude to the “Swiss Academic Compute Cloud” supporting the development of the cloud computing infrastructure for hydrogeological use-cases.

6.7. References

Aquanty Inc. *HydroGeoSphere. A three-dimensional numerical model describing fully integrated subsurface and surface flow and solute transport*. Waterloo, ON, Canada, 2016.

S. F. Ashby and R. D. Falgout. A Parallel Multigrid Preconditioned Conjugate Gradient Algorithm for Groundwater Flow Simulations. *Nuclear Science and Engineering*, 124(1):145–159, Sept. 1996.

R. T. Bailey and D. Bau`. Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. *Hydrology and Earth System Sciences*, 16(2):287–304, 2012. ISSN 1607-7938. doi:10.5194/hess-16-287-2012.

K. Barnhart, I. Urteaga, Q. Han, A. Jayasumana, and T. Illangasekare. On Integrating Groundwater Transport Models with Wireless Sensor Networks. *Ground Water*, 48(5):771–780, Sept. 2010. ISSN 1745-6584. doi:10.1111/j.1745-6584.2010.00684.x.

P. Bauer-Gottwein, R. Schneider, and C. Davidsen. Optimizing Wellfield Operation in a Variable Power Price Regime. *Groundwater*, 54(1):92–103, Jan. 2016. ISSN 1745-6584. doi:10.1111/gwat.12341.

G. Bauser, H.-J. H. Franssen, H.-P. Kaiser, U. Kuhlmann, F. Stauffer, and W. Kinzelbach. Real-Time Management of an Urban Groundwater Well Field Threatened by Pollution. *Environmental Science & Technology*, 44(17):6802–6807, Sept. 2010. ISSN 0013-936X. doi:10.1021/es100648j. G. Bauser, H.-J. Hendricks Franssen, F. Stauffer, H.-P. Kaiser, U. Kuhlmann, and W. Kinzelbach. A comparison study of two different control criteria for the real-time management of urban groundwater works. *Journal of Environmental Management*, 105:21–29, Aug. 2012. ISSN 0301-4797. doi:10.1016/j.jenvman.2011.12.024.

D. Blessent, R. Therrien, and C. W. Gable. Large-scale numerical simulation of groundwater flow and solute transport in discretely-fractured crystalline bedrock. *Advances in Water Resources*, 34(12):1539–1552, Dec. 2011. ISSN 0309-1708. doi:10.1016/j.advwatres.2011.09.008.

D. Blessent, P. R. Jrgensen, and R. Therrien. Comparing discrete fracture and continuum models to predict contaminant transport in fractured porous media. *Ground Water*, 52(1):84–95, Feb. 2014. ISSN 1745-6584. doi:10.1111/gwat.12032.

H. R. Bogena, M. Herbst, J. A. Huisman, U. Rosenbaum, A. Weuthen, and H. Vereecken. Potential of Wireless Sensor Networks for Measuring Soil Water Content Variability. *Vadose Zone Journal*, 9(4):1002–1013, Nov. 2010. ISSN 1539-1663. doi:10.2136/vzj2009.0173.

P. Brunner and C. T. Simmons. HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. *Ground Water*, 50(2):170–176, Mar. 2012. ISSN 1745-6584. doi:10.1111/j.1745-6584.2011.00882.x.

C. M. Bürger, S. Kollet, J. Schumacher, and D. Bösel. Introduction of a web service for cloud computing with the integrated hydrologic simulation platform ParFlow. *Computers & Geosciences*, 48:334–336, Nov. 2012. ISSN 0098-3004. doi:10.1016/j.cageo.2012.01.007.

G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126(6):1719–1724, June 1998. ISSN 0027-0644. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

S. Caino-Lores, A. Lapin, P. Kropf, and J. Carretero. Lessons learned from applying big data paradigms to a large scale scientific workflow. In: *Proceedings of the 11th Workshop on*

Workflows in Support of Large-Scale Science, WORKS '16. ACM, 2016. ISBN 978-1-4503-2138-9. doi:10.1145/1235.

M. Camporese, C. Paniconi, M. Putti, and P. Salandin. Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. *Water Resources Research*, 45(10):W10421, Oct. 2009. ISSN 1944-7973. doi:10.1029/2008WR007031.

Y. Chen and D. Zhang. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, 29 (8):1107–1122, Aug. 2006. ISSN 0309-1708. doi:10.1016/j.advwatres.2005.09.007.

G. De Schepper, R. Therrien, J. C. Refsgaard, and A. L. Hansen. Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. *Journal of Hydrology*, 521:374–388, Feb. 2015. ISSN 0022-1694. doi:10.1016/j.jhydrol.2014.12.035.

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, May 1994. ISSN 2156-2202. doi:10.1029/94JC00572.

A. M. Gaukroger and A. D. Werner. On the Panday and Huyakorn surface-subsurface hydrology test case: analysis of internal flow dynamics. *Hydrological Processes*, 25(13):2085–2093, June 2011. ISSN 1099-1085. doi:10.1002/hyp.7959.

A. K. Hansen, H. Madsen, P. Bauer-Gottwein, A. K. V. Falk, and D. Rosbjerg. Multi-objective optimization of the management of a waterworks using an integrated well field model. *Hydrology Research*, 43(4):430–444, Aug. 2012. ISSN 0029-1277, 2224-7955. doi:10.2166/nh.2012.142.

H. J. Hendricks Franssen and W. Kinzelbach. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44(9):W09408, Sept. 2008. ISSN 1944-7973. doi:10.1029/2007WR006505.

H. J. Hendricks Franssen, H. P. Kaiser, U. Kuhlmann, G. Bauser, F. Stauffer, R. Müller, and W. Kinzelbach. Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. *Water Resources Research*, 47(2):W02532, Feb. 2011. ISSN 1944-7973. doi:10.1029/2010WR009480.

M. Herbst, S. Gottschalk, M. Reiel, H. Hardelauf, R. Kasteel, M. Javaux, J. Vanderborght, and H. Vereecken. On preconditioning for a parallel solution of the Richards equation. *Computers & Geosciences*, 34(12):1958–1963, Dec. 2008. ISSN 0098-3004. doi:10.1016/j.cageo.2008.02.020.

R. J. Hunt, J. Luchette, W. A. Schreuder, J. O. Rumbaugh, J. Doherty, M. J. Tonkin, and D. B. Rumbaugh. Using a Cloud to Replenish Parched Groundwater Modeling Efforts. *Ground Water*, 48(3):360–365, Mar. 2010. ISSN 0017467X, 17456584. doi:10.1111/j.1745-6584.2010.00699.x.

A. Jamakovic, D. C. Dimitrova, M. Anwander, T. Macicas, T. Braun, J. Schwanbeck, T. Staub, and B. Nyffenegger. *Real-World Energy Measurements of a Wireless Mesh Network*, pages 218–232. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40517-4. doi:10.1007/978-3-642-40517-4 18.

J. E. Jones and C. S. Woodward. Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24(7):763–774, July 2001. ISSN 0309-1708. doi:10.1016/S0309-1708(00)00075-0.

S. J. Kollet and R. M. Maxwell. Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7):945–958, July 2006. ISSN 0309-1708. doi:10.1016/j.advwatres.2005.08.006.

P. Kropf, E. Schiller, P. Brunner, O. Schilling, D. Hunkeler, and A. Lapin. Wireless mesh networks and cloud computing for real time environmental simulations. In *Recent Advances in Information and Communication Technology. Proceedings of the 10th International Conference on Computing and Information Technology (IC2IT2014)*, Advances in Intelligent Systems and Computing. Springer, 2014. ISBN 978-3-319-06537-3. doi:10.1007/978-3-319-06538-0 1.

W. Kurtz, H.-J. Hendricks Franssen, H.-P. Kaiser, and H. Vereecken. Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. *Water Resources Research*, 50(2):1665–1688, Feb. 2014. ISSN 1944-7973. doi:10.1002/2013WR014823.

W. Kurtz, G. He, S. J. Kollet, R. M. Maxwell, H. Vereecken, and H.-J. Hendricks Franssen. TerrSysMP-PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model. *Geoscientific Model Development*, 9(4):1341–1360, Apr. 2016. ISSN 1991-9603. doi:10.5194/gmd-9-1341-2016.

A. Lapin, E. Schiller, P. Kropf, O. Schilling, P. Brunner, A. J. Kapic, T. Braun, and S. Maffioletti. Real-Time Environmental Monitoring for Cloud-Based Hydrogeological Modeling with HydroGeoSphere. In *2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC,CSS,ICESS)*, pages 959–965, Aug. 2014. doi:10.1109/HPCC.2014.154.

A. Lapin, E. Schiller, and P. Kropf. Claude: Cloud-computing for non-interactive long-running computationally intensive scientific applications. In H. Unger and W. A. Halang, editors, *Autonomous Systems 2015: Proceedings of the 8th GI Conference*, pages 221–232. VDI Verlag, 2015.

L. Li, H. Zhou, J. J. Gmez-Hernndez, and H.-J. Hendricks Franssen. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter. *Journal of Hydrology*, 428429:152–169, Mar. 2012. ISSN 0022-1694. doi:10.1016/j.jhydrol.2012.01.037.

X. Li, X. Cheng, P. Gong, and K. Yan. Design and implementation of a wireless sensor network-based remote water-level monitoring system. *Sensors*, 11(2):1706–1720, 2011. ISSN 1424-8220. doi:10.3390/s110201706.

G. Liu, Y. Chen, and D. Zhang. Investigation of flow and transport processes at the MADE site using ensemble Kalman filter. *Advances in Water Resources*, 31(7):975–986, July 2008. ISSN 0309-1708. doi:10.1016/j.advwatres.2008.03.006.

Y. Liu, A. H. Weerts, M. Clark, H.-J. Hendricks Franssen, S. Kumar, H. Moradkhani, D.-J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec, and P. Restrepo. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.*, 16(10):3863–3887, Oct. 2012. ISSN 1607-7938. doi:10.5194/hess- 16-3863-2012.

P. Loden, Q. Han, L. Porta, T. Illangasekare, and A. P. Jayasumana. A wireless sensor system for validation of real-time automatic calibration of groundwater transport models. *Journal of Systems and Software*, 82(11):1859–1868, Nov. 2009. ISSN 0164-1212. doi:10.1016/j.jss.2009.05.049.

B. S. Marti. *Real-time management and control of groundwater flow field and quality*. PhD thesis, ETH-Zu ̈rich, 2014. R. M. Maxwell. A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, 53:109–117, Mar. 2013. ISSN 0309-1708. doi:10.1016/j.advwatres.2012.10.001.

P. Mell and T. Grance. The NIST definition of cloud computing. Technical Report NIST Special Publication 800-145, National Institute of Standards and Technology, 2011. URL http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf. (last access: 15th October 2016).

R. T. Mills, C. Lu, P. C. Lichtner, and G. E. Hammond. Simulating subsurface flow and transport on ultrascale computers using PFLOTRAN. *Journal of Physics: Conference Series*, 78(1):012051, 2007. ISSN 1742-6596. doi:10.1088/1742-6596/78/1/012051.

W. Nowak. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water Resources Research*, 45(4):W04431, Apr. 2009. ISSN 1944-7973. doi:10.1029/2008WR007328.

S. Panday and P. S. Huyakorn. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Advances in Water Resources*, 27(4):361–382, Apr. 2004. ISSN 0309-1708. doi:10.1016/j.advwatres.2004.02.016.

J. Rasmussen, H. Madsen, K. H. Jensen, and J. C. Refsgaard. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. *Hydrol. Earth Syst. Sci.*, 19(7):2999–3013, July 2015. ISSN 1607-7938. doi:10.5194/hess-19-2999-2015.

C. J. Ritsema, H. Kuipers, L. Kleiboer, E. van den Elsen, K. Oostindie, J. G. Wesseling, J.-W. Wolthuis, and P. Havinga. A new wireless underground network system for continuous monitoring of soil water contents. *Water Resources Research*, 45(4):W00D36, Apr. 2009. ISSN 1944-7973. doi:10.1029/2008WR007071.

D. A. Robinson, C. S. Campbell, J. W. Hopmans, B. K. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zone Journal*, 7(1):358, 2008. ISSN 1539- 1663. doi:10.2136/vzj2007.0143.

O. S. Schilling, J. Doherty, W. Kinzelbach, H. Wang, P. N. Yang, and P. Brunner. Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater-surface water-vegetation interactions. *Journal of Hydrology*, 519, Part B:2258–2271, Nov. 2014. ISSN 0022-1694. doi:10.1016/j.jhydrol.2014.08.063.

D. Schwanenberg, A. v. Breukelen, and S. Hummel. Data assimilation for supporting optimum control in large-scale river networks. In *2011 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pages 98–103, Apr. 2011. doi:10.1109/ICNSC.2011.5874881.

D.-J. Seo, L. Cajina, R. Corby, and T. Howieson. Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, 367(34):255–275, Apr. 2009. ISSN 0022-1694. doi:10.1016/j.jhydrol.2009.01.019.

Y. Shi, K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu. Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment. *Advances in Water Resources*, 83:421–427, Sept. 2015. ISSN 0309-1708. doi:10.1016/j.advwatres.2015.06.009.

Q. Tang, W. Kurtz, P. Brunner, H. Vereecken, and H. J. Hendricks Franssen. Characterisation of river-aquifer exchange fluxes: The role of spatial patterns of riverbed hydraulic conductivities. *Journal of Hydrology*, 531, Part 1:111–123, Dec. 2015. ISSN 0022-1694. doi:10.1016/j.jhydrol.2015.08.019.

J. Varia. *Cloud Architectures-White Paper*. 2008. URL https://aws.amazon.com/about-aws/whatsnew/2008/07/16/ cloud-architectures-white-paper/. (last access: 15th October 2016).

H. Vereecken, O. Neuendorf, G. Lindenmayr, and A. Basermann. A schwarz domain decomposition method for solution of transient unsaturated water flow on parallel computers. *Ecological Modelling*, 93(1):275 – 289, 1996. ISSN 0304-3800. doi:http://dx.doi.org/10.1016/0304- 3800(95)00224-3.

C. J. Watras, M. Morrow, K. Morrison, S. Scannell, S. Yaziciaglu, J. S. Read, Y.-H. Hu, P. C. Hanson, and T. Kratz. Evaluation of wireless sensor networks (WSNs) for remote wetland monitoring: design and initial results. *Environmental Monitoring and Assessment*, 186(2):919–934, Feb. 2014. ISSN 1573-2959. doi:10.1007/s10661-013-3424-8.

A. H. Weerts, G. Y. El Serafy, S. Hummel, J. Dhondia, and H. Gerritsen. Application of generic data assimilation tools (DATools) for flood forecasting purposes. *Computers & Geosciences*, 36(4):453–463, Apr. 2010. ISSN 0098-3004. doi:10.1016/j.cageo.2009.07.009.

Chapter 7

7. Conclusions

Sustainable water resources management must be based on quantitative methods and approaches. Numerical models which simulate hydrological systems under current and future climatic conditions are an important tool in this regard. The current generation of physically-based flow models in principle allow an integrated representation and simulation of all relevant hydrological processes. These include flow processes at the surface, the subsurface and the coupling between them. However, the comprehensive representation of flow processes in such physically-based models is clearly no guarantee for reliable predictions. The reliability of simulations and predictions of a SW-GW systems is strongly dependent on choices of the modeler during the model development phase, particularly on the choice of the parametrization, the observation types that are to be used for model calibration, and on how these observations are implemented into the calibration process. The many parameters that are required for physically-based SW-GW flow simulations cannot be sufficiently constrained with classic observations of hydraulic heads and SW discharge alone.

The overarching goal of this Ph.D. thesis was to improve the conceptualization and simulation of SW-GW systems. Field and modelling methods were developed, employed and combined in this Ph.D. thesis. Field approaches aimed at exploring the worth and applicability of novel and unconventional observations, such as tree ring growth records or the analysis of environmental tracers such as ^{222}Rn, ^{37}Ar, ^{3}H/^{3}He and noble gases. These observations were integrated into physically-based numerical models through calibration approaches. Advanced computational and simulation approaches such as cloud-computing and real-time modelling were employed. Physically-based models were also used to improve our understanding on the underlying physics of the interactions between surface water and groundwater. These findings help to make better informed choices on the conceptualization of streambeds. Below, an overview over the different chapters, their conclusions and future research needs is provided.

- **Chapter 2: Review of scientific literature on the use of unconventional observations in flow modelling:** Such a broad review had not been carried out before, and for the first time provides an overview over the different unconventional observation types used in flow modelling, and their integration into the flow model calibration process. Also, an overview of the worth of different observation types in improving the simulation of flow and transport processes was provided. The review clearly revealed that unconventional observation types can improve flow simulations and, therefore, should be used more frequently. At the same time, the review identified several limitations and challenges in using such data, and revealed major research gaps concerning the implementation of unconventional observations.

- **Chapter 3: Improved assessment of the necessary complexity of riverbed and aquifer heterogeneity in the context of unsaturated subsurface flow:** Through an improved assessment of the complexity of riverbed and aquifer heterogeneity that is required in numerical flow modelling an important preliminary step in SW-GW modelling could be facilitated: With the new stochastic 1-D approach that was presented in chapter 2, one can efficiently estimate the potential for the development of unsaturated areas underneath riverbeds in river-aquifer systems, and decide on the degree of heterogeneity of the hydraulic conductivity that needs to be implemented into the numerical model. This allows a more efficient allocation of the typically limited computational resources towards the simulation of processes that are the most relevant for a given SW-GW scenario; rather than wasting computational resources on the calculation of heterogeneous subsurface flow for systems where heterogeneity could be simplified without introducing a bias in predictions of flow. A future direction of research related to the new stochastic 1-D approach clearly lies in an improved estimation of the probability distribution and spatial correlation of the hydraulic conductivity of both the riverbed and the aquifer. Potential lies in promising developments in geophysics and image analysis, which could provide non-intrusive assessment of the probability distribution and spatial correlation of subsurface properties.

- **Chapter 4: Development of a novel estimation technique for historic transpiration rates based on tree ring growth records, and subsequent integration of these observations into SW-GW modelling through a sophisticated calibration and uncertainty assessment approach:** The notorious lack of observations along a remote stretch of the Tarim River of the arid Taklamakan Desert in China could be successfully counteracted by tapping into the historic data repository of tree ring growth records. The information on historic transpiration derived from the tree ring growth records provided a proxy for transpiration observations. This is the only means of estimating the efficiency of engineered freshwater releases aimed at saving the unique *Populus euphratica* riparian forests along an otherwise dry stretch of the Tarim River. By calibrating a physically-based flow model against this proxy for transpiration as well as observations of hydraulic heads, and by assessing the worth of these observations in reducing the predictive uncertainty of the flow model, allowed a very extensive assessment of the worth of tree ring growth records in eco-hydrological flow simulations. Data worth assessment allows developing targeted and efficient data acquisition schemes. The employed method of predictive uncertainty and data worth assessment could be identified as an ideal method to assess the worth of other unconventional observations in SW-GW model calibration. This is the first time where tree ring growth records have been used to estimate historic transpiration rates, and where such a feedback between vegetation and flow has been systematically used to calibrate a numerical flow model. Moreover, the systematic assessment of the worth of transpiration rates in calibrating the numerical flow model demonstrated the large potential of integrating ecological and hydrological observations into inverse modelling. A huge potential lies in the future research of the relation between tree ring growth and transpiration for other tree species, especially in otherwise data scarce regions.

- **Chapter 5: Combination of observations from a multi-tracer study, including the novel ^{37}Ar tracer method for SW-GW interactions, with physically-based flow modelling:** In this very extensive field and modelling study, it could be shown that unconventional observations, providing information on mixing ratios and residence times, are essential in informing the development of a flow model for alluvial drinking water stations. We could demonstrate that for the system under investigation, riverbed hydraulic conductivities and aquifer porosity, two typically poorly characterized values, are largely insensitive to hydraulic heads compared to the hydraulic conductivity of the aquifer. These two parameters, however, are of pivotal importance to travel times and mixing ratios between water of different origins; and as direct observations of these two parameters were missing, they could only be informed through the observations of mixing and travel times. The employed multi-tracer study, which covers previously unobservable intermediate residence timescales through the inclusion of the novel ^{37}Ar tracer method, can successfully provide the necessary observations. If these observations are combined with a physically-based flow model that includes a flow tracking scheme that tracks the flow of water throughout the entire model domain, an ideal tool for the simulation and prediction of flow in alluvial drinking water stations can be created. Future efforts should be directed towards an automated calibration of the flow model against these tracer-based unconventional observations of mixing between SW and GW in the pumped water mix, in order to further improve the model simulations and systematically quantify the worth of these observations in calibrating the model. Such a follow up study is in preparation.

- **Chapter 6: Coupling of Ensemble Kalman Filter-based data assimilation to the physically-based flow modelling tool HydroGeoSphere, and the integration of this modelling framework into cloud computing environments:** Due to the limited knowledge on the spatial distribution of subsurface properties, state-of-the-art modelling approaches can be further improved by a continuous correction of short-term model predictions whenever new observations become available. For this reason, one of these so-called data assimilation approaches, the powerful Ensemble Kalman Filter (EnKF), was coupled to the physically-based and fully-integrated SW-GW model HydroGeoSphere (HGS), and implemented into cloud-based computing environments. The new EnKF-HGS-cloud modelling framework now, for example, allows considering the transience nature of streambed properties. By moving the entire modelling framework into a cloud computing environment, virtually unlimited computational resources can be employed. To the best of the author's knowledge, this is the first time that the novel computational cloud resources, physically-based and fully-coupled SW-GW modelling, and automated data assimilation, were brought together. Together with the more efficient preliminary assessment of the spatial extent of potential unsaturated zones presented in **chapter 3**, these novel developments strengthen the quality of the SW-GW modelling framework. In order to facilitate the use of the new EnKF-HGS-cloud framework, future efforts should be directed towards the development of web-based interfaces which allow to easily setup, adjust, and oversee the simulations and to facilitate the communication and feedback to between the simulations and the regulation of anthropogenic structures such as pumps or weirs.

As these different studies clearly showed, an improved conceptualization of SW-GW systems, novel observations, and the integration of these observations in the modelling process, provide multiple ways of improving the reliability of simulations and predictions with SW-GW models. However, due to the large number of different measurement techniques and available unconventional observations types (as discussed in **chapter 2)**, this thesis could cover just a few of these different techniques and observations types (e.g., classical observations, plus tree ring growth records and measurements of natural environmental tracers such as ^{222}Rn, ^{37}Ar, ^{3}H/^{3}He, and noble gases), and thus merely scratch the surface of what is still possible. With the calibration and uncertainty method described in **chapter 4**, however, virtually every different observation type can be implemented in the calibration processes and its data worth systematically assessed. This is a strong improvement compared to most studies reviewed in **chapter 2,** where the flow models were not systematically calibrated with the same flexibility to include information from unconventional observations alongside classical observations, and data worth was rarely systematically assessed. Future modelling efforts should therefore focus on a more systematic calibration, uncertainty assessment and data worth analysis, similar to the method presented in **chapter 4**.

In summary, in this Ph.D. thesis, I covered field work, flow modelling and model calibration. In my opinion, this combination of methods represents the highlight of this Ph.D. thesis, as advances could be made in all three fields. And there still lies much more potential in the combination of these different fields. This is especially evident in the light of the findings of the review presented in **chapter 2**, which revealed that the implementation of unconventional observations in the flow model calibration process is not that straightforward, but the implementation harbors huge potential to increase the quality of flow models. Moreover, through the advances that were made in all three fields that this thesis covers, overarching research needs could be identified: Transformation of unconventional observations to observations, which do not require complicated additional processes to be simulated by the numerical flow model, have proven to be of augmented worth for flow model calibration compared to the original observations, for example transforming tree ring growth to historic transpiration rates or transforming observations of solute concentrations to mixing and pumped water sources' information. Future research should be directed towards an easier field assessment of mixing and GW sources, and on how such observations should be implemented into – and potentially transformed for – the flow model calibration process. This would allow better identifying observations that can also efficiently constrain simpler models, which in turn would result in a smaller computational burden, less potential for numerical instabilities, and less additional forcing functions and initial conditions to be defined. At the same time, the ever-increasing availability of computational resources allows more and more complex models to be constructed and simulated. Everything, from inverse methods, to models, and available observations, is being continuously improved. However, the available computational resources, modelling tools, inverse methods and observations are so excessive that their wide-spread use largely remains untapped. But hydrology and hydrogeology must, like all other research fields, embrace these developments and profit from them in order not to get stuck in current practices and problems. The approaches do not need to become more complex and complicated, but instead more appropriate for the different SW-GW systems, and the new technologies more approachable for hydrologists and hydrogeologists. As for the use of unconventional observations in the flow modelling process, more guidance is needed on how to use the newest measurement and computational technologies, and a more systematic assessment of the best tools available should be provided.