Analytical and Stochastic Numerical Methods for the Simulation of Subsurface Flow in Floodplains

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Jonas Timon Allgeier

aus Reutlingen

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	22.11.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr.-Ing. Olaf A. Cirpka
2. Berichterstatter:	PD Dr. habil. Thomas Wöhling
3. Berichterstatter:	Felipe P. J. de Barros, PhD

Acknowledgments

“ Hang on to the good days. I can lean on my friends. They help me going through hard times. ”

— PHOENIX

This dissertation would not exist without the help, advice, guidance and support of various people. Thank you, Prof. Dr.-Ing. Olaf A. Cirpka, for covering all these aspects so well. You were an excellent primary supervisor. Your brilliant ideas, attention to detail, ugly questions, and continuous efforts towards improvement have been key to the existence and quality of this document. I am also thankful that you brought the (I)RTG on Integrated Hydrosystem Modelling to life in the first place.

Another big “thank you” goes to you, Dr. Michael Finkel. You were an excellent second supervisor and I always knew I could rely on you (or your inner doctor/virologist). Your composure and your great sense of humor really helped me not to lose my nerves whenever things did not work out right away.

Albeit not being my official supervisors, I have learned a lot from Prof. Dr.-Ing. Wolfgang Nowak, Dr. Ana González-Nicolás, Dr.-Ing. Daniel Erdal, Simon Martin and Philipp Selzer while working on joint publications. The content of this dissertation has benefited a lot from these collaborations, and I am grateful for your contributions.

The hydrogeology group (including Olaf, Michael, Carsten, Daniel, Emilio, Anna, Philipp, Simon, Stefan, Ruth, Carolin, Adrian, Cora and Marie-Madeleine) was a perfect work environment for me. I always looked forward to the interesting, helpful, and fun weekly group meetings. It is really unfortunate that the cake tradition has been affected so much by the remote working, but I am confident that you can bring it back to its original glory!

The research training group (including Anna, Julia, Philipp, Elena, Michelle, Hemanti, Ishani, Luciana and Ran) was another awesome institution I am glad to have been a part of. Our weekly meetings helped me a lot with respect to going through those truly scary times. Of course, they were beneficial for our science, too. Thank you, Monika, for all the related organization, including the wonderful and informative spring and fall schools.

I am really glad to have had the opportunity to (co-)supervise several master students (including Jonathan, Sary, Matthias and Janek). I believe all these projects have led to interesting results and have been mutually beneficial.

Finally, I am indebted to my beloved friends and family who have accompanied me during the last years. In particular, I thank my parents, my brother, and Natalie for their continuous support.

I also acknowledge and appreciate the financial support through the German Research Foundation (DFG) in the framework of the Research Training Group GRK 1829 “Integrated Hydrosystem Modelling” (Grant Agreement GRK 1829/2).

Jonas

Abstract

Floodplain aquifers are important hydraulic connectors between hillslopes and surface-water bodies. The flow field in floodplain aquifers comprises different flow components governed by various geometric and hydrogeologic parameters. In this work, (semi-)analytical and numerical stochastic simulations are used to address three classical problems associated with investigations of floodplain aquifers. To this end, the Ammer floodplain west of Tübingen serves as an exemplary study site.

The first aspect of this dissertation focuses on valley-scale lateral hyporheic exchange in floodplain aquifers driven by widening and subsequent narrowing of the aquifer geometry. By means of a new semi-analytical solution, simple analytical proxy-models can be derived that allow a trivial and quick assessment, whether this type of exchange is relevant in a given setting. The application of these tools to the Ammer floodplain shows that the site has the geometric potential for notable valley-scale hyporheic exchange, but small hydraulic conductivities and lateral influxes from the hillslopes restrict the exchange zone to a negligible extent.

The second topic is concerned with identifying promising points in space, where hydraulic-head information would help to locate groundwater divides separating the catchment area of floodplain aquifers from other catchments. A respective uncertainty-reduction optimization problem is formulated and solved by the application of a stochastic framework based on pre-filtered steady-state flow models. In the context of the Ammer floodplain, this analysis confirms that a presumed shift between groundwater and surface water divide is likely to exist. Three observation points identified by the procedure are predicted to help in reducing the related uncertainty by more than fifty percent.

The third and final subject deals with calibrating steady-state floodplain models to hydraulic-head data. A modified, proxy-model-based, global calibration routine is able to find well-performing parameter sets that bring a steady-state Ammer floodplain model in agreement with measured field data. Neural Posterior Estimation, a technique from the field of Simulation-Based Inference, confirms these parameter sets and sheds light on the related uncertainties and correlations. A key result of this analysis is the confirmed inter-basin flow from the Ammer hillslopes to the Neckar valley, which takes place in the Erfurt formation beneath the Spitzberg ridge and the Wurmlingen saddle.

Zusammenfassung

Auengrundwasserleiter sind wichtige hydraulische Verbindungen zwischen Hängen und Oberflächengewässern. Das Strömungsfeld in Auengrundwasserleitern umfasst verschiedene Strömungskomponenten, die von unterschiedlichen geometrischen und hydrogeologischen Parametern bestimmt werden. In dieser Arbeit werden (semi-)analytische und numerische stochastische Simulationen eingesetzt, um drei klassische Probleme im Zusammenhang mit Auengrundwasserleitern zu behandeln. Dazu dient die Ammeraue westlich von Tübingen als exemplarisches Untersuchungsgebiet.

Der erste Aspekt dieser Dissertation befasst sich mit groß-skaligem, lateralen hyporheischen Austausch in Auengrundwasserleitern, der durch die Aufweitung und anschließende Verengung der Grundwasserleitergeometrie angetrieben wird. Mit Hilfe einer neuen semi-analytischen Lösung können einfache analytische Proxy-Modelle abgeleitet werden, die eine triviale und schnelle Einschätzung erlauben, ob diese Art des Austausches in einem gegebenen Umfeld relevant ist. Die Anwendung dieser Werkzeuge auf die Ammeraue zeigt, dass der Standort das geometrische Potenzial für einen nennenswerten hyporheischen Austausch besitzt, aber geringe hydraulische Leitfähigkeiten und seitliche Zuflüsse von den Hängen die Austauschzone auf ein vernachlässigbares Niveau beschränken.

Das zweite Thema befasst sich mit der Identifizierung vielversprechender Landschafts-Punkte, an denen Grundwassermessungen helfen würden, Grundwasserscheiden zu lokalisieren, die das Einzugsgebiet von Auengrundwasserleitern von anderen Einzugsgebieten trennen. Ein entsprechendes Optimierungsproblem zur Unsicherheitsreduzierung wird formuliert und durch die Anwendung eines stochastischen Verfahrens auf der Grundlage vorgefilterter stationärer Abflussmodelle gelöst. Im Zusammenhang mit der Ammeraue bestätigt diese Analyse, dass eine vermutete Verschiebung zwischen Grundwasser- und Oberflächenwasserscheide wahrscheinlich ist. Drei durch das Verfahren identifizierte Beobachtungspunkte könnten voraussichtlich dazu beitragen, die damit verbundene Unsicherheit um mehr als die Hälfte zu reduzieren.

Das dritte und letzte Thema befasst sich mit der Kalibrierung von stationären Auenmodellen auf Basis von Daten zur hydraulischen Spiegelhöhe. Eine modifizierte, auf Proxy-Modellen basierende, globale Kalibrierungsroutine ist in der Lage, gute Parametersätze zu finden, die ein stationäres Ammerauen-Modell in Übereinstimmung mit gemessenen Felddaten bringen. Die neuronale Posterior-Schätzung, eine Technik aus dem Bereich der simulationsbasierten Inferenz, bestätigt diese Parametersätze und gibt Aufschluss über die damit verbundenen Unsicherheiten und Korrelationen. Ein zentrales Ergebnis dieser Analyse ist der bestätigte unterirdische Grundwasserabstrom von den Ammerhängen zum Neckartal, der in der Erfurt-Formation unterhalb des Spitzbergs und des Wurmlinger Sattels stattfindet.

General Abbreviations

AESS	Averaged Effective Sample Size
FDM	Finite Difference Method
FEM	Finite Element Method
FVM	Finite Volume Method
GPE	Gaussian Process Emulator
GPR	Gaussian Process Regression
HGS	HydroGeoSphere
MAF	Masked Autoregressive Flow
MCMC	Markov Chain Monte Carlo
NPE	Neural Posterior Estimation
PDE	Partial Differential Equation
PreDIA	Preposterior Data Impact Assessor
RMSE	Root Mean Square Error
SBI	Simulation-Based Inference
SNPE	Sequential Neural Posterior Estimation

Geological Units

moM	Meissner formation
kuE	Erfurt formation
kmGr	Grabfeld formation
kmSt	Stuttgart formation
kmSw	Steigerwald formation
kmHb	Hassberge formation
kmMh	Mainhardt formation
kmLw	Löwenstein formation
kmTr	Trossingen formation
km2345	Lumped sandstone formations

Local Institutions

ASG	Ammertal-Schönbuchgruppe
Campos	Catchments as Reactors: Metabolism of Pollutants on the Landscape Scale (Collaborative Research Center)
LGRB	Landesamt für Geologie, Rohstoffe und Bergbau
LTZ	Landwirtschaftliches Technologiezentrum Augustenberg
LUBW	Landesanstalt für Umwelt Baden-Württemberg
SWT	Stadtwerke Tübingen

Contents

Chapter I

Overall Background

1 Introduction

More than 4.5 % of Germany's land area morphologically belongs to floodplains (Koenzen and Günther-Diringer, 2021) – those areas around rivers that are frequently inundated during high-discharge events if no flood protection measures are taken (Brunotte et al., 2009). As landscape elements, floodplains connect surface-water bodies, aquifers and hillslopes (Jung et al., 2004; Ó Dochartaigh et al., 2019). This interface-role makes them interesting for various scientific disciplines, including ecology, biology, hydro(geo)logy, geomorphology and biogeochemistry (Hauer et al., 2016). The corresponding research deals with flow of water, sediment transport and (reactive) transport of compounds including nutrients, contaminants and organic matter (Bridge, 2009). Wohl (2021) provides an extensive review on how these processes interact with each other and how the corresponding storages are integrated in floodplains.

Floodplains emerge and evolve through different sedimentary mechanisms involving deposition of material within the river-channel and outside of it (Wolman and Leopold, 1957; Nanson and Croke, 1992; Leopold et al., 2020). With processes like channel migration, as well as the associated erosion and deposition, floodplains form comparably dynamic landscapes from a geomorphological perspective (Bridge, 2009). Floodplain geomorphology can also be linked to riparian ecology (Huggenberger et al., 1998), as floodplains (in this context also referred to as *riparian zones*) are ecological habitats representing hotspots of biological diversity (Meyer and Edwards, 1990; Smock et al., 1992; Brunotte et al., 2009). For instance, Hauer et al. (2016) provide an interdisciplinary review on how riparian biodiversity benefits from ecosystem disturbances associated with flooding. Such considerations are especially relevant in the context of hydrogeomorphical river management: many floodplains have been anthropogenically modified in the past, for example through river regulation and flood protection measures (Follner et al., 2010; Brunotte et al., 2009; Koenzen and Günther-Diringer, 2021). It has been recognized recently that such interference has adverse effects on the proper functioning of the river/floodplain ecosystem and that retaining the natural states of floodplains is desirable (Beechie and Roni, 2012; Biron et al., 2014; Buffin-Bélanger et al., 2015; Dezső et al., 2019; Karpack et al., 2020).

Another aspect of anthropogenic interference with riparian areas is the frequent usage of floodplains and their surroundings for agricultural purposes (Tockner and Stanford, 2002). Crop cultivation and the corresponding fertilization can lead to subsurface contamination with pollutants, most prominently nitrate (Baillieux et al., 2014; Schilling et al., 2015). On the other hand, many floodplain aquifers also provide hydraulic and biogeochemical conditions that are exceptionally suitable

for contaminant degradation (Hill, 1990, 1996; Woessner, 2000; Bates et al., 2000; Cloke et al., 2003; Clilverd et al., 2013), for example due to elevated contents of organic carbon in floodplain sediments (Sutfin et al., 2016). Hill (2019) reviewed the recent state of research regarding nitrate removal in floodplain aquifers, concluding that the denitrification pathway is more important than other microbial degradation processes or uptake through vegetation. There are also cases where the reduction of nitrate concentrations in floodplains is attributed to dilution with river water (Pinay et al., 1998). Biogeochemical processes in floodplains are not limited to nitrogen species, as illustrated by the review of Vidon et al. (2019), which also considers phosphorous, greenhouse gases and emerging contaminants (e.g., fire retardants). How complex the interplay of biogeochemical reactions in floodplains can be, becomes clear with the case study of Yabusaki et al. (2017), which considered chemical species of carbon, nitrogen, oxygen, iron, sulfur and uranium.

Any qualitative and quantitative investigation of transport within a given floodplain, may it concern reactive contaminants, conservative compounds, or sedimentary material, first requires a good understanding of the relevant water flow processes (Büttner et al., 2006; Wohl, 2021).

1.1 Hydrogeology of Floodplain Aquifers

The first aspect that comes to mind with respect to water flow in floodplains is their role as flood wave retention areas from the hydrological perspective (e.g., Rashid and Chaudhry, 1995; Bedient et al., 2008; Valentová et al., 2010; Bornschein and Pohl, 2018). As recognized in the last decades, many groundwater and surface-water bodies form a continuum that requires an integrated consideration (Winter et al., 1998; Cook, 2015). Floodplains are a prime example of this phenomenon: surface-water flood waves can also exert notable effects on groundwater levels in the underlying floodplain aquifers. This process is known as *flood wave propagation* (Sophocleous, 1991). The associated celerity and amplitude of the corresponding groundwater pressure waves, as well as the propagation direction have been studied, for example, by Jung et al. (2004), Cloutier et al. (2014), and Buffin-Bélanger et al. (2016).

The overall flow field in floodplain aquifers, however, is often times not (only) governed by flood wave propagation, but other (regional) flow processes that are less dynamic in nature (Larkin and Sharp Jr., 1992; Woessner, 2000). In this dissertation, I am interested in such long-term or quasi-steady-state regional flow behavior of floodplain aquifers. Here, the conditions are assumed to be stable considering time-scales large enough such that the effects of event-triggered dynamics can be averaged out or neglected, but small enough to still consider the floodplain a static entity from the geomorphological point of view. Real floodplain systems are of course never truly at an equilibrium and local short-term dynamic effects might be superimposed on the general regional flow field. Nonetheless, the analysis and investigation of floodplain aquifers in (assumed) quasi-steady-state conditions can reveal insights into the hydraulic functioning of floodplain systems. It might also be a valuable foundation for follow-up studies that target the transient effects, as for example done by Grapes et al. (2006), Folch et al. (2010), Ostendorf et al. (2012), and Helton et al. (2014).

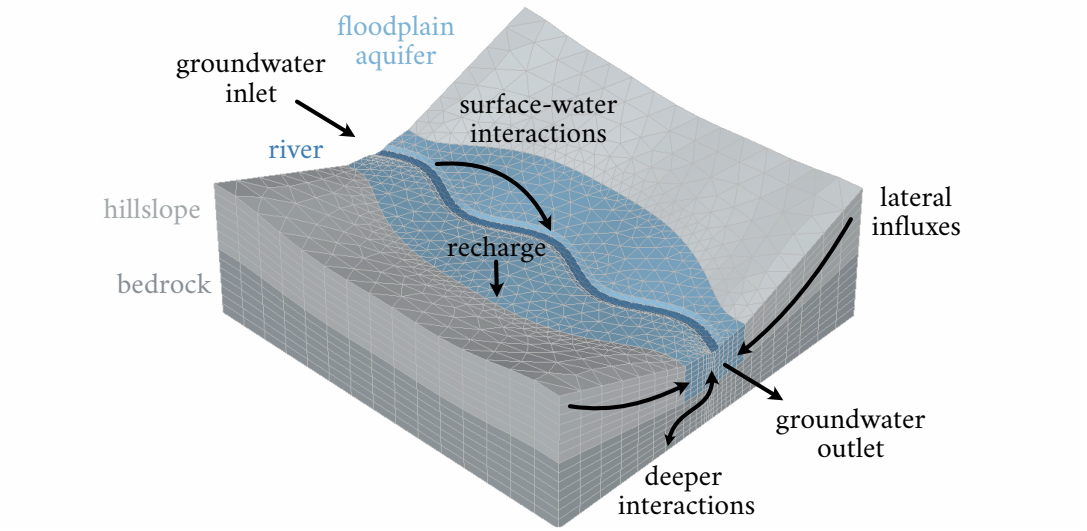

Figure 1: Three-dimensional sketch of potentially relevant flow processes in floodplain aquifers.

An initial conceptual understanding of the subsurface-flow behavior of a floodplain aquifer system can be obtained by a water balance consideration, namely a description of potential sources or sinks of water (Wohl, 2021). Figure 1 provides a schematic overview in this regard, highlighting several flow controls that are potentially relevant in floodplain aquifers (focusing on natural processes, and thereby neglecting extraction wells for instance). In the following, I will further characterize the individual contributions.

First of all, there is the longitudinal slope of the surface elevation which corresponds to the mean effective slope of the river. Typically, this slope also leads to an along-valley flow component in the subsurface (e.g., MacDonald et al., 2014), termed *underflow* component in the review of Larkin and Sharp Jr. (1992). Depending on the connectivity of the floodplain to the preceding and following subsurface-water bodies, this can result in groundwater fluxes entering or leaving the floodplain at its upstream and downstream ends (as for example in the case of Peyrard et al., 2008). Secondly, there is often another topographic gradient superimposed, pointing from the outer rims of the floodplain towards the river. This can lead to a cross-valley component of flow, termed *baseflow* component in the review of Larkin and Sharp Jr. (1992), caused by lateral input of water entering the floodplain from adjacent hillslopes on the sides (e.g., Sun et al., 2017). Similarly, the floodplain might be connected vertically to deeper groundwater resources (e.g., Macpherson and Sophocleous, 2004). Depending on the given pressure situation, the floodplain aquifer might gain water from deeper aquifers, or lose it towards the subsurface instead. The net effect of precipitation and evapotranspiration can result in groundwater recharge within the floodplain. Finally, there is the interaction of the floodplain aquifer and the river itself. Such surface-water/groundwater interactions attract considerable scientific attention (see for example the reviews of Flipo et al., 2014; Ward, 2016), because the interface of these usually very different water bodies (e.g., in terms of physical and chemical properties) is of relevance for the hydro(geo)logical functioning of the floodplain and for the turnover of major biogeochemical constituents and anthropogenic pollutants.

Cook (2015) classified surface-water/groundwater interactions into *river gain*: exfiltration/discharge of groundwater into the river (e.g., as a result of collecting the hillslope-influxes), *river loss*: infiltration/recharge of river water into groundwater (e.g., if there is a significant loss towards deeper groundwater systems), *bank storage*: bidirectional exchange between river and groundwater due to dynamic changes in river-water stage, and *hyporheic exchange*: water originating from rivers, taking a detour through groundwater and coming back to the river without a net impact on the water balance of either the aquifer or the river. Hyporheic exchange is important for various aspects of water quality and ecology, including microbial activity, solute turnover, nutrient fate, and redox conditions (Triska et al., 1993; Hayashi and Rosenberry, 2002; Boano et al., 2009; Fabian et al., 2011; Ward, 2016; Lewandowski et al., 2019).

As the subsurface is basically inaccessible, it is difficult to investigate which of the described flow processes are actually taking place at a specific site, and how relevant they are. Most often, there is only little or sparse information available, for example in form of (1) point-like data obtained from groundwater observation wells, (2) general geometric descriptions (e.g., approximate length, width, thickness and shape of the aquifer), and (3) estimated hydraulic properties obtained from field measurements or as expert knowledge/guess-work. With respect to analyzing the flow behavior in (floodplain) aquifers, groundwater observation wells provide the most valuable data, but unfortunately their installation is costly. As a result, the locations of such wells should be chosen wisely to gain the most information from as few wells as possible. In some cases, direct observational data may be sufficient to answer scientific questions (e.g., the approximation of hillslope contribution fluxes by Martin et al., 2020). More often, a quantitative and physically consistent analysis of the flow field is necessary (e.g., when flow paths need to be identified). The field of hydrogeological modeling provides adequate and versatile tools for this purpose.

1.2 Hydrogeological Modeling

Models are abstract and simplified representations of real, complex systems that are developed to answer problem-specific questions (Asbrand et al., 2002; Anderson et al., 2015). These questions can be related to system-understanding, physically-consistent interpolation, parameter inference or prediction of states. In hydrogeology, modeling represents an especially important research area, because the direct investigation of subsurface-flow systems in all aspects and in the desired degree of detail is often not feasible, too expensive or even impossible. This is due to the fact that physical access to the subsurface is limited and that subsurface-flow systems can cover extensive areas, while often reacting only very slowly to changes (Hölting and Coldewey, 2013). The field of hydrogeological modeling is vast and covers many aspects (e.g., model definition, model calibration, modeling methods, sensitivity analysis, etc.). Extensive reviews on early modeling techniques are given by Prickett (1975) and Bredehoeft (2012). Anderson et al. (2015) provide a detailed overview of the modern state of hydrogeological modeling. In the following, I will focus mostly on the different types of models in the context of hydrogeology and how they relate to this dissertation.

At the most general level, (hydrogeological) models can be differentiated into *conceptual models* and *quantitative models*. A conceptual model is a characterization of a (flow) problem that describes all processes and properties that are relevant for the investigated system and modeling purpose. For instance, this can include descriptions of the domain, the hydrostratigraphy, and assumed or known flow conditions. While conceptual models can stand on their own, a conceptual model is often the precursor of another model that attempts to give quantitative answers to modeling questions. Such a quantitative model always requires a conceptual model (sometimes the conceptual model is presented implicitly) and in many cases the modeling process consists of several iterations of refining both, the conceptual and the quantitative model after gaining insights from the intermediate model results.

Quantitative models can be grouped into *physical models* and *mathematical models*. Physical models use experimental physical systems to reproduce those aspects of a real system that were deemed relevant in the conceptual model. These physical systems can take the form of miniature versions of larger-scale flow problems. For instance, the work of Darcy (1856) can be seen as a very early example of using a small-scale laboratory model to simulate flow on a much larger scale. More recent examples for lab-scale physical models are the works of Rodhe (2012), who developed several physical models of classical hydrogeological problems for class-room teaching, and Boyraz and Kazezyılmaz-Alhan (2017), who investigated surface-water/groundwater interactions through lab-scale experiments. Physical models of hydrogeological systems can also be based on the flow of quantities that are different from water, but subject to equivalent physical laws. For instance, currents in electrical analogue studies can be related to groundwater discharges (e.g., Tóth, 1968; Vaux, 1968; Kiràly, 1971). Similarly, there also exist studies on thermal analogues (Piggott and Elsworth, 1989) and viscous fluids (*Hele-Shaw models*; e.g., Dvoracek and Scott, 1963; Awan and O'Donnell, 1972). Physical models are obviously associated with comparably high labor intensity and measurement errors, adverse properties that mathematical models do not have.

Mathematical models are quantitative representations of the conceptual model by means of mathematical relationships. Generally, mathematical models comprise *data-driven models* and *process-based models*. According to Anderson et al. (2015), data-driven models produce output from input by applying empirical or statistical relationships that are inferred from a given data set, often without resolving, distinguishing or even considering physical processes. Examples for data-driven models are the studies of Amaranto et al. (2018), who used data-driven models for the prediction of groundwater levels from meteorological data, and Kirchner (2019), who inferred catchment transit-time distributions from concentration data in surface-water streams. In contrast to such data-driven approaches, process-based models are based on a mathematical representation of the relevant physical processes.

Process-based models are often formulated as boundary value or initial value problems of governing Partial Differential Equations (PDEs). Depending on the model complexity and the PDEs, analytical solutions may be found for such problems. For instance, Theis (1935) formulated the analytical

expression for transient radial flow towards a pumping well in a homogeneous two-dimensional aquifer of infinite extent. Another famous example of a fully analytical model is the work of Tóth (1963) targeting regional and local flow in a drainage basin. Fully analytical solutions are mathematically elegant, precise, scale- and discretization-independent, and quick to evaluate. Unfortunately, they are typically restricted to mostly inflexible, basic problems with predefined geometries (e.g., rectangular domains) or simplistic boundary conditions (like the sinusoidal head fluctuations of Tóth, 1963). For models targeting domains with complicated boundaries, internal heterogeneities or other non-trivial properties, analytical solutions often cannot be derived. In such cases, numerical methods can be used to (approximately) solve the respective problems.

Numerical solutions allow for arbitrary complexity and therefore full flexibility, but come at the cost of larger evaluation times and accuracy limitations often related to spatial (or temporal) discretization. The ever-increasing availability and power of computational resources (in terms of hardware and software) in the last decades has lead to a rise in popularity of such numerical models that are often based on the Finite Difference Method (FDM; e.g., Trescott et al., 1976), the Finite Volume Method (FVM; e.g., Rees et al., 2004) or the Finite Element Method (FEM; e.g., Huyakorn et al., 1984). As a result, the domain of fully numerical hydrogeologic modeling tools offers a number of modern software codes nowadays. For instance, Modflow (McDonald and Harbaugh, 1988) is a widely used open-source modular groundwater flow modeling suite (Langevin et al., 2017, 2022). In this dissertation I rely on HydroGeoSphere (HGS), a FEM-based fully-integrated hydrogeological model environment, that can simulate surface and variably saturated subsurface flow (Therrien et al., 2010; Brunner and Simmons, 2012).

The distinction between analytical and numerical solutions is not strictly binary. In fact, *semi-analytical methods* serve as an intermediate member on the spectrum of process-based mathematical models. Semi-analytical methods are typically based on a set of simplifying assumptions to make the problem tractable, while maintaining the desired flexibility, for example by allowing arbitrary geometries or boundary condition values for some relevant parts of the domain. This can result in exact analytical expressions that require simple numerical methods for the evaluation of integrals or the determination of infinite-series coefficients (e.g., Zlotnik et al., 2011). Once the set of approximate coefficients has been evaluated for a specific model setup, the semi-analytical expression can be used to evaluate the solution at the same convenience as a fully analytical solution. Semi-analytical techniques have been used in past hydrogeological research, in particular for vertical cross-sections of hillslopes connected to rivers, drainage ditches or groundwater bodies (e.g., Powers, 1966; Li et al., 1996; Read, 2007; Craig, 2008), but also for lateral two-dimensional problems (e.g., Suribhatla et al., 2004; Boano et al., 2006; Samani and Sedghi, 2015; Gomez-Velez et al., 2017). The Analytical Element Method (e.g., Strack, 1989; Bakker, 1999; Bakker and Strack, 2003; Strack, 2003; Bakker, 2006; Fitts, 2010; Strack and Nevison, 2015; Strack, 2018) represents a large class of semi-analytical models. It allows the construction of solutions for a given problem with a modular superposition approach, where each boundary condition represents a so-called *analytical element*.

During the rise of computational power, models have also become more and more complex, due to the availability of new numerical methods and higher affordability of computational complexity (Venkataraman and Haftka, 2004; Zhou and Li, 2011; Jakob, 2014). This increased complexity can be observed from the presence of a comparably large number of tunable parameters in modern models. The process of adjusting model parameter values such that the model output agrees (more) with measured data is known as *model calibration*, *inverse modeling* or *parameter inference*. It has a long history and is one of the classical and essential problems in the field of modeling (Carrera et al., 2005; Hill and Tiedeman, 2006; Zhou et al., 2014). As a result, many calibration philosophies and methods exist, ranging from classical *trial-and-error* approaches (*manual calibration*) to automated calibration schemes like those implemented in the PEST suite (Doherty et al., 1994; Doherty and Hunt, 2010; Doherty, 2015), a general-purpose calibration toolbox based on the Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963) with extensions and variants, that is especially popular within the community of hydrogeology (e.g., Selle et al., 2013). Nevertheless, calibration still remains a crucial challenge in modern modeling, especially in cases where models are nonlinear, high-dimensional, and/or computationally expensive. Moreover, off-the-shelf calibration routines are usually insufficient when parameter uncertainties or interdependencies are of interest.

1.3 Overarching Questions and Outline of This Dissertation

In the following, I want to elaborate on how hydrogeological modeling can help to address three typical questions arising in the investigation of floodplain aquifer systems.

Domain Delineation As the physico-chemical composition of a water parcel strongly depends on its origin, it is of crucial importance to identify where the floodplain aquifer draws its water from (which is also relevant for water budget analyses). This involves a delineation of the corresponding catchment area towards the lateral hillslopes. A common assumption when performing such a delineation is that the groundwater table on these hillslopes essentially follows the surface topography (Tóth, 1963; Haitjema and Mitchell-Bruker, 2005), which simplifies the delineation to finding the surface water divides (a comparably simple task that only requires a digital elevation model and a geographic information system; Tarboton et al., 1991).

However, the topography of a phreatic groundwater surface may substantially differ from that of the land surface so that the groundwater and surface water divides may not coincide (Haitjema and Mitchell-Bruker, 2005; Bloxom and Burbey, 2015; Han et al., 2019). In fact, Haitjema and Mitchell-Bruker (2005) reported on a whole class of aquifers naturally exhibiting such shifts between surface and subsurface water divides, namely cases with relatively high hydraulic conductivities in conjunction with a difference between the elevation of drainage points in neighboring valleys. Additional factors contributing to shifts in groundwater divides include tilted aquifer strata, spatial heterogeneity in the recharge rate, and hydraulic anisotropy. Of course, anthropogenic influence (e.g., groundwater abstraction) can also result in shifted groundwater divides.

The location of groundwater divides can be constrained by hydraulic-head measurements. In theory, a very dense network of piezometers could be used to accurately interpolate the groundwater-table map, which could subsequently be analyzed by the same tools as used for delineating surface-water divides. In practice, this is not advisable as the number of observation wells is limited by financial costs, labor intensity, and legal restrictions. That is, groundwater divides must be delineated with head measurements from a limited number of piezometers. A classical way of doing this is by calibrating groundwater flow-and-transport models to the head measurements, which explicitly uses all information fed into the model construction (e.g., the geometry and parameter ranges of geological units and boundary conditions) and leads to hydraulic-head fields that are consistent with conservation principles. As only a limited number of observation wells is affordable, their placements should be specifically optimized for delineating a particular groundwater divide. This leads to the first research question of this dissertation:

1. *How can we determine where to measure hydraulic head in order to reduce the uncertainty in delineating a groundwater divide?*

In Chapter III, I introduce an optimal-design analysis to address this problem by means of numerical modeling. It can identify the best spatial configuration of piezometers for groundwater-divide delineation and is based on formal minimization of the expected posterior uncertainty in localizing the groundwater divide.

Valley-Scale River-Aquifer Interaction In addition to influxes to the floodplain aquifer at its lateral boundaries, the river represents another entity relevant for water origin and fate. In this context, hyporheic exchange is especially important for floodplain aquifers. In general, hyporheic exchange occurs on different spatial and temporal scales (Boano et al., 2014; Barthel and Banzhaf, 2016; Magliozzi et al., 2017, 2018; Ward and Packman, 2019; Zachara et al., 2020), ranging from centimeter-scale exchange induced by bedforms, over meter-scale exchange between step-pool sequences, to kilometer-scale hyporheic exchange, which is sometimes referred to as parafluvial flow (Mallard et al., 2014; Cook, 2015). While small-scale exchange typically takes place in the vertical direction, larger-scale hyporheic exchange can also occur laterally (Hayashi and Rosenberry, 2002; Wagner and Bretschko, 2003; Gooseff et al., 2003; Fabian et al., 2011), for example between or within river meanders (Boano et al., 2009; Gomez et al., 2012).

Floodplain aquifers often exhibit a widening and narrowing geometry (illustrated in Figure 1). As shown by Tonina and Buffington (2009), such changes in cross-sectional area are one of three drivers for hyporheic exchange (besides non-uniform hydraulic conductivity and changes in energy head gradients). As a result, valley-scale lateral hyporheic exchange could be driven by these varying geometries, even in cases where the river is straight and its slope is uniform. This phenomenon has already been hypothesized conceptually (Wondzell and Gooseff, 2013), but a thorough analysis has not yet been performed. This leads to the second research question of this dissertation:

2. *Under which conditions can the lateral widening and narrowing of floodplain aquifer(s) cause valley-scale hyporheic exchange?*

I address this problem in Chapter II. Towards this end, I derive a semi-analytical solution which describes the steady-state groundwater flow for two-dimensional floodplain aquifer systems connected to rivers. Using simplified, generalized aquifer geometries allows me to observe general patterns instead of site-specific local phenomena. Since the semi-analytical models are comparably simple conceptually, I decided to present this topic before the chapter on domain delineation with its numerical models.

Parameter Estimation After delineating the contributing area of a floodplain and determining which processes need to be considered (e.g., surface-water interactions), a site-specific model needs to be set up to address questions of groundwater management. Before a model can be used for that, however, the model parameters need to be estimated. Model calibration is the process of finding the set of parameters that makes the model meet measured data best. The analysis of the posterior/conditional parameter uncertainty quantifies to which extent the parameters can be constrained by measurements, and a sensitivity analysis helps to identify the parameters that control the magnitude of model outputs. In this regard, parameters that model predictions are sensitive to, but that are poorly constrained by the data define the worst case. The uncertainty of parameters that model predictions are insensitive to can be tolerated much better.

For simple models requiring small runtimes and little computational resources, various automated calibration schemes and global sensitivity analysis tools exist. For instance, ensemble-based methods like genetic algorithms (Goldberg, 1989; Gen and Cheng, 1999; Das and Suganthan, 2011) and Markov Chain Monte Carlo (MCMC) methods (Gilks et al., 1995; Brooks et al., 2011) are able to find global minima and may provide a good approximation of the parameter distribution conditioned on the measurements. To their disadvantage, these methods require many model runs. This can be prohibitive for modern numerical subsurface-flow models with long run times and considerable computational demands.

For these reasons, a special branch of calibration research is dedicated to developing global calibration schemes that are as efficient as possible (see Haftka et al., 2016, for a detailed review). Such schemes are often based on *proxy-models* (also referred to as surrogate-models or meta-models). A proxy-model might be a coarsened version of the original model or a black-box-type approximation relating input parameters sets to observed model output in a simplified way (e.g., through machine-learning methods or by interpolation in parameter space). The underlying idea is that the proxy-model is considerably quicker to evaluate than the original model (at the cost of accuracy). The problem of such proxy-model-based global calibration schemes is that they only aim for a single best estimate of the parameters without assessing uncertainties or relationships between the parameters. Ideally, a full posterior parameter distribution should be obtained and analyzed.

One idea to obtain such a distribution from computationally expensive subsurface-flow models is to run a global calibration scheme first. During this operation, all intermediate input/output information of the full model can be stored. This allows the generation of an interpolation-based proxy-model afterwards, which is then used to perform posterior estimation methods that would be too costly otherwise (e.g., MCMC sampling). Another possible solution might be provided by the field of Simulation-Based Inference (SBI) (Cranmer et al., 2020; Tejero-Cantero et al., 2020; Lueckmann et al., 2021). The corresponding tools allow the estimation of a full posterior parameter distribution without evaluating expensive likelihoods. One particular method of this field is the Neural Posterior Estimation (NPE), a likelihood-free posterior estimation tool based on machine learning (Papamakarios and Murray, 2016; Lueckmann et al., 2017; Greenberg et al., 2019). This raises the third research question of this dissertation:

3. *How does NPE compare to proxy-model-based MCMC sampling of a posterior parameter distribution after global calibration of a computationally expensive subsurface-flow model?*

In Chapter IV, I address this question by constructing posterior distributions with both methods and comparing the outcomes to each other and the results of different variants of a proxy-model-based global calibration routine.

Outline In each of the described three main chapters related to hydrogeological models of floodplain aquifers, I use the same specific floodplain system as an example, namely the Ammer floodplain close to Tübingen. This site has been subject to a number of hydrogeological investigations within the framework of the Collaborative Research Center 1253 CAMPOS (Catchments as Reactors: Metabolism of Pollutants on the Landscape Scale). After a detailed site description in Section 3, I will pose additional site-specific questions in Section 3.5 that are also addressed in Chapters II to IV. Before that, I introduce the general methods in Section 2, as the main chapters share some common background related to the underlying scientific theory or methodology. At the end of this dissertation, a concise summary of the outcomes related to the presented questions and final conclusions across the main chapters is given in Chapter V.

2 Methods

2.1 Notation

Within all equations and mathematical terms of this dissertation, scalar symbols are printed as non-bold letters (a), while vectors use lowercase letters in bold font ($\boldsymbol{a}$). For matrices, I use bold, uppercase letters ($\boldsymbol{A}$). Special matrices are the identity matrix $\boldsymbol{I}$ and the zero matrix $\boldsymbol{O}$. Similarly, there are vectors of zeros $\mathbf{0}$ and ones $\mathbf{1}$. To make the dimensions of all tensors clear, I use the notation $n \times m$, where n is the number of rows and m is the number of columns. I use the notation $\boldsymbol{A}^{-1}$ to indicate the inverse of $\boldsymbol{A}$ and $\boldsymbol{A}^{\mathsf{T}}$ for its transpose.

The ∇-operator is the vector of partial derivatives in all dimensions of the variable it is applied to:

$$\nabla_{\boldsymbol{x}} = \left[\frac{\partial}{\partial x_1}, \quad \frac{\partial}{\partial x_2}, \quad \cdots, \quad \frac{\partial}{\partial x_n}\right]. \tag{2.1}$$

Depending on the circumstances, this vector might also be defined as a column vector.

For general quantities that have a physical meaning I provide information about the respective dimension with dimensional descriptors: M is a mass, L is a length and T is a time. For specific physical quantities I provide the units (typically according to the *Système International d'Unités* specification). All elevations are given in meters above sea level.

The meaning of all symbols is *not* strictly unique within this dissertation, because the same symbol might be intuitive for different variables in different contexts. However, I keep symbols consistent and unique at least within a section and I redefine symbols as soon as they change their meaning.

For graphic representations using colors to visualize continuous variables, I rely on the perceptually uniform color maps suggested by Kovesi (2015), which allow for unbiased visual comparisons.

2.2 Governing Equations for Variably Saturated Flow

This section introduces the governing equations of subsurface flow through porous media under the assumption that water and air are the only fluids, and the mobility of air is infinite.

2.2.1 Water Content and Water Saturation

The dimensionless volumetric water content Θ_{w} is defined as the ratio of the water-filled pore space in a porous medium to the total volume of the considered sample. For a given porous material, Θ_{w} can range from a minimum value, the *residual water content* Θ_{r}, to a maximum value, the *saturated water content* Θ_{s}. The current water content can also be expressed as a dimensionless water saturation S_{w}, which describes the fraction of the volume occupied by water to the total pore space volume available:

$$S_{\mathrm{w}} = \frac{\Theta_{\mathrm{w}}}{\Theta_{\mathrm{s}}}. \tag{2.2}$$

The dimensionless residual saturation S_r describes the ratio of residual to saturated water content:

$$S_r = \frac{\Theta_r}{\Theta_s}. \tag{2.3}$$

This is the fraction of water that cannot be removed from pore-space considering conventional flow mechanisms. As a result, all water saturations are within the range $S_r \leq S_w \leq 1$, while all water contents are within the range $\Theta_r \leq \Theta_w \leq \Theta_s$. The *effective saturation* is defined as a fraction that becomes zero at the residual water content Θ_r and one for full saturation:

$$S_e = \frac{\Theta_w - \Theta_r}{\Theta_s - \Theta_r} = \frac{S_w - S_r}{1 - S_r}. \tag{2.4}$$

2.2.2 Subsurface Flow Equation

Subsurface water-contents typically vary both in space and time t. The PDEs for describing the respective flow of water in a porous domain can be derived from the continuity equation (e.g., Farthing et al., 2003):

$$\frac{\partial(\rho_w \Theta_w)}{\partial t} + \nabla(\rho_w \boldsymbol{q}) = \rho_w Q, \tag{2.5}$$

where ρ_w represents the density of water in $\mathsf{M}\,\mathsf{L}^{-3}$, $\boldsymbol{q}$ is the specific discharge in $\mathsf{L}\,\mathsf{T}^{-1}$ and Q describes volumetric source and sink terms in T^{-1}. With the chain rule of differentiation and after dividing by the density of water we arrive at:

$$\frac{\partial \Theta_w}{\partial t} + \frac{\Theta_w}{\rho_w}\frac{\partial \rho_w}{\partial t} + \nabla \boldsymbol{q} = Q, \tag{2.6}$$

where spatial gradients of ρ_w have been neglected (as they are typically not large enough to be relevant for subsurface flow systems). Equation 2.6 holds for variably saturated flow, meaning both water-saturated and -unsaturated conditions. For numerical solutions of the flow equation, it makes sense to express the governing equation not in terms of water saturations, but water potentials (or pressures, typically expressed in L through the division of the gravimetric potential and the density of water). The most relevant potentials are:

- h_p in L: The non-negative pressure-head describing hydrostatic pressure.
- z in L: The gravimetric potential expressed as an elevation above a defined reference elevation.
- h_c in L: The non-negative capillary head due to capillary forces exerted by the porous medium.

Other potentials exist, but are neglected here (e.g., implicitly requiring uniform air pressure). The total hydraulic head h in L is composed of the individual potentials:

$$h = h_p + z - h_c. \tag{2.7}$$

Water flows generally in direction of decreasing total potential (i.e., from large h to small h).

The *groundwater table* describes that vertical location, where the total potential h equals the elevation, meaning both h_c and h_p are zero. Beneath the groundwater table, there are no capillary forces and the associated water potential h_c becomes zero. Above the groundwater table, there is no hydrostatic pressure and the associated water potential h_p becomes zero. The parts of the subsurface with saturations smaller than unity are called the *unsaturated zone*, whereas the parts where the saturation is unity are called the *saturated zone*. With this knowledge, Equation 2.6 can be expressed in terms of h, which requires two separate formulations, one for the parts above the groundwater table and one for the parts beneath it.

Beneath the groundwater table, the porous medium is fully saturated and the water content Θ_w equals the saturated water content Θ_s. The storage terms can simply be expressed in terms of h:

$$\frac{\partial \Theta_w}{\partial t} + \frac{\Theta_w}{\rho_w}\frac{\partial \rho_w}{\partial t} = \frac{\partial \Theta_s}{\partial t} + \frac{\Theta_s}{\rho_w}\frac{\partial \rho_w}{\partial t} = \frac{\partial \Theta_s}{\partial h}\frac{\partial h}{\partial t} + \frac{\Theta_s}{\rho_w}\frac{\partial \rho_w}{\partial h}\frac{\partial h}{\partial t}. \tag{2.8}$$

It is common (e.g., Freeze and Cherry, 1979) to combine the two storage terms into a single parameter denoted *specific storage coefficient* S_s in L^{-1}, which describes how much water is stored in the porous medium when h is changing:

$$S_s = \frac{\partial \Theta_s}{\partial h} + \frac{\Theta_s}{\rho_w}\frac{\partial \rho_w}{\partial h}. \tag{2.9}$$

This results in the governing equation in terms of h for flow beneath the groundwater table:

$$S_s \frac{\partial h}{\partial t} + \nabla \boldsymbol{q} = Q. \tag{2.10}$$

For the parts above the groundwater table, the water content does not in general equal Θ_s. However, we can make use of the simplification

$$\frac{\partial \Theta_w}{\partial t} + \frac{\Theta_w}{\rho_w}\frac{\partial \rho_w}{\partial t} \approx \frac{\partial \Theta_w}{\partial t}, \tag{2.11}$$

because typically in the unsaturated zone the changes in water content are much larger than the temporal changes in water density. This results in Equation 2.12 for the governing flow equation above the groundwater table:

$$\frac{\partial \Theta_w}{\partial h}\frac{\partial h}{\partial t} + \nabla \boldsymbol{q} = Q. \tag{2.12}$$

It is possible to combine the two formulations again, to construct a single, approximate, governing equation for variably saturated flow (Cooley, 1971; Huyakorn et al., 1984; Therrien et al., 2010):

$$\eta \frac{\partial h}{\partial t} + \nabla \boldsymbol{q} = Q, \tag{2.13}$$

with an *overall storage coefficient* η in L^{-1} accounting for effects beneath and above the groundwater table:

$$\eta = S_w \cdot S_s + \frac{\partial \Theta_w}{\partial h}. \tag{2.14}$$

Beneath the groundwater table, the pore space is fully water-saturated ($S_w = 1$) and the water content does not depend on the total hydraulic head ($\partial\Theta_w/\partial h = 0$). The overall equation simplifies to Equation 2.10. Above the groundwater table, water saturations start to drop and $\partial\Theta_w/\partial h$ quickly becomes much larger than $S_w \cdot S_s$. Hence, the overall equation approximately simplifies to Equation 2.12. Equation 2.13 is the formulation used in HGS and is sometimes called the *(modified) Richards equation* after the work of Richards (1931). The earliest formulation of a governing equation for flow in the unsaturated zone was given by Richardson (1922).

The specific discharge $\boldsymbol{q}$ in a variably saturated porous medium can be described by a generalized version of Darcy's law (Darcy, 1856):

$$\boldsymbol{q} = -\boldsymbol{K}_{sat} k_{rel} \nabla h, \tag{2.15}$$

where $\boldsymbol{K}_{sat}$ in $\mathsf{L}\,\mathsf{T}^{-1}$ is a tensor of saturated hydraulic conductivity and k_{rel} is a dimensionless relative permeability, here considered to be isotropic. The former is a property of the porous medium, while the latter depends on the current state of saturation. In the saturated zone, the relative permeability equals unity. Above the groundwater table, it depends on the water saturation and can be in the range $0 \leq k_{rel} \leq 1$. In numerical modeling of subsurface flow, it is common to relate the relative permeability to the effective saturation S_e by some parametrization $k_{rel} = k_{rel}(S_e)$ (Tocci et al., 1998; Farthing et al., 2003; Suk and Park, 2019).

2.2.3 Parametrization of the Unsaturated Zone

Different parametrizations of the unsaturated zone have been developed in the past. A particularly common choice is the model of Brooks and Corey (1964):

$$k_{rel}(S_e) = S_e^{3+\frac{2}{\lambda}}, \tag{2.16}$$

where λ is the dimensionless *pore size index*. It is positive ($\lambda > 0$) and is typically larger for porous media that have a narrow pore size distribution (i.e., more uniform pore sizes; Brooks and Corey, 1964). There are also more complex parametrizations in use; the most common one being the one of van Genuchten (1980), which is based on the work of Mualem (1976):

$$k_{rel}(S_e) = \sqrt{S_e}\left(1 - \left(1 - S_e^{\frac{N}{N-1}}\right)^{\frac{N-1}{N}}\right)^2, \tag{2.17}$$

where N is an empirical parameter that has to be larger than one. The parametrization of van Genuchten (1980) converges to the model of Brooks and Corey (1964) for large capillary pressures (i.e., small effective saturations). In this case, the relationship between N and the pore size index is:

$$N = 1 + \lambda. \tag{2.18}$$

This means that a porous medium with uniform pore size can be described with a large N-value, and small N close to one represent more heterogeneous pore size distributions.

So far we have only related the relative permeability to the effective saturation, which is not the primary variable in subsurface flow models. Instead, relationships with respect to the total hydraulic head h are required. Those can be constructed by defining parametrizations relating the (effective) water saturation with h (or h_c, to be more precise, as the saturation is always one where the capillary pressure is zero). As the effective saturation can easily be converted to a water saturation by means of Equation 2.4, the expression $S_e(h)$ can also be used for the storage term $\partial\Theta_w/\partial h$ as a byproduct.

Again, the parametrizations of Brooks and Corey (1964) and van Genuchten (1980) are commonly used. Equation 2.19 is the parametrization of Brooks and Corey (1964):

$$S_e(h_c) = \min\left[\left(\frac{h_{AE}}{h_c}\right)^{\lambda}, 1\right], \tag{2.19}$$

where h_{AE} in L is a fixed potential (commonly referred to as *air entry pressure*) that produces water saturations smaller than unity if it is exceeded by the capillary head. However, there is also a fully water-saturated portion above the groundwater table, where capillary forces are still active ($0 < h_c < h_{AE}$). This is zone is called *capillary fringe*. It is the reason why the expressions "saturated zone" and "beneath the groundwater table" are not fully equivalent.

Equation 2.20 displays the parametrization of van Genuchten (1980):

$$S_e(h_c) = \left(1 + (\alpha h_c)^N\right)^{\frac{1-N}{N}}, \tag{2.20}$$

where α in L^{-1} is another empirical parameter. With this parametrization, the saturation drops below one immediately above the groundwater table. A capillary fringe is not explicitly considered. There is, however, a zone with saturations close to unity that extends up to a certain capillary head which is comparable to the capillary fringe. We can come up with an expression approximating an air entry pressure that would be equivalent to a given van Genuchten model (van Genuchten, 1980; Rawls and Brakensiek, 1985, 1989):

$$h_{AE} \approx \frac{1}{\alpha}. \tag{2.21}$$

Hence, the two presented parametrizations approach each other for a specific set of properties and certain unsaturated conditions. The main difference between the two parametrization from a numerical point of view lies in the fact that the Brooks and Corey model is discontinuous but simpler than the smooth van Genuchten equations. As discontinuities often lead to adverse convergence behavior in numerical models, often the van Genuchten parametrization is preferred. However, for small N (i.e., $N \to 1$) this parametrization results in a quick and drastic decrease of k_{rel} when S_e drops below 1 (very similar to a discontinuity; Vogel and Cislerova, 1988; Vogel et al., 2000; Schaap and Leij, 2000). For this reason, some modelers (e.g., Touma, 2009; Kuang et al., 2021) combine the

two parametrizations by taking the saturation curve $S_e(h_c)$ of van Genuchten (1980) and the relative permeability curve $k_{rel}(S_e)$ of Brooks and Corey (1964) (and using Equation 2.18). This combination leads to a smooth decrease of relative permeability with increasing capillary head, even for small values of N or λ. Of course this comes at the cost of breaking the conceptual consistency with the underlying capillary/pore distribution model.

2.2.4 Boundary Conditions

PDE-based mathematical models require boundary and initial conditions in addition to the underlying PDE to be fully defined. Under steady-state conditions, initial conditions are not necessary. A multitude of surface and subsurface boundary conditions can be applied in subsurface flow models. In the following, I summarize the boundary conditions that are used in this dissertation.

Dirichlet Boundary A Dirichlet boundary fixes the primary unknown (i.e., the total hydraulic head) of the subsurface flow equation:

$$h = h_{fix} \quad \text{on } \Gamma_D, \tag{2.22}$$

where h_{fix} is a known hydraulic head in L and Γ_D is that part of the model-domain boundary Γ that the Dirichlet boundary condition is applied to. Dirichlet boundaries can be used on any outer or inner part of the modeling domain (i.e., at the top/bottom surface, the lateral sides, and internally).

Neumann Boundary A Neumann boundary imposes a constraint on the normal derivative of the primary variable. In case of subsurface flow models, such a boundary condition is typically formulated as a fixed, specified flux boundary:

$$\boldsymbol{n} \cdot \boldsymbol{q} = q_{fix} \quad \text{on } \Gamma_N, \tag{2.23}$$

where $\boldsymbol{n}$ is the dimensionless spatial unit normal vector perpendicular to the boundary pointing outwards, q_{fix} is a prescribed normal flux in $\mathsf{L}\,\mathsf{T}^{-1}$ and Γ_N is the boundary. Neumann boundaries can be applied to all parts of the domains surface.

Seepage Boundary This modified Dirichlet boundary can be used to prevent groundwater tables above surface elevations. Its formal definition can be expressed as in Equation 2.24:

$$h = \min[h_{sim}, z_{surf}] \quad \text{on } \Gamma_S, \tag{2.24}$$

where h_{sim} in L is the head a simulation would produce if the seepage boundary was not active, z_{surf} in L is the surface elevation, and Γ_S is the seepage boundary. This type of boundary condition is usually only applied at the model domain top surface. In numerical flow models, the boundary

condition is typically enforced by triggering an artificial sink term, as soon as the groundwater table hits the surface (e.g., in HGS; Therrien et al., 2010).

Leaky Boundary A leaky boundary can be used to simulate contact to another water body with a prescribed head h_{fix} in L. In contrast to a Dirichlet boundary, this contact is not assumed to be hydraulically perfect, but occurring through a conductive zone. In numerical models, for example HGS, it is usually defined by a source/sink term Q in $\mathsf{L}^3\,\mathsf{T}^{-1}$:

$$Q = C_{\text{L}} \cdot (h - h_{\text{fix}}) \quad \text{on } \Gamma_{\text{L}}, \tag{2.25}$$

where C_{L} in $\mathsf{L}^2\,\mathsf{T}^{-1}$ is a conductance and Γ_{L} is the leaky boundary. Leaky boundaries can be used on all outer surface parts of the domain. At lateral sides of the domain, the boundary condition is sometimes referred to as fluid-transfer boundary. Here, the conductance term is defined by:

$$C_{\text{L}} = \frac{A}{L_{\text{int}}} \cdot K_{\text{int}}, \tag{2.26}$$

where A in L^2 is the area affected by the boundary, L_{int} in L is the separation distance between the boundary and the other water body, and K_{int} in $\mathsf{L}\,\mathsf{T}^{-1}$ is the interjacent hydraulic conductivity. If a leaky boundary is applied to the bottom or top of the domain's surface, for example to simulate contact to a river, the conductance can be defined by Equation 2.27:

$$C_{\text{L}} = \frac{L_{\text{riv}} \cdot w_{\text{riv}}}{L_{\text{sed}}} \cdot K_{\text{sed}}, \tag{2.27}$$

where L_{riv} and w_{riv} in L are the length and width of the river segment and L_{sed} in L and K_{sed} in $\mathsf{L}\,\mathsf{T}^{-1}$ are the thickness and hydraulic conductivity of a sediment layer between river and subsurface.

Drain Boundary A special type of leaky boundary is the drain boundary. It allows exfiltration only when the groundwater table exceeds a threshold Δh in L compared to the surface elevation:

$$Q = \begin{cases} C_{\text{D}} \cdot (h - z_{\text{surf}}) & \text{if } h - z_{\text{surf}} > \Delta h \\ 0 & \text{otherwise} \end{cases} \quad \text{on } \Gamma_{\text{T}}, \tag{2.28}$$

where C_{D} in $\mathsf{L}^2\,\mathsf{T}^{-1}$ is a drain conductance and Γ_{T} is the drain boundary. In this case, the conductance can be inferred from an associated surface area A in L^2, as well as thickness and hydraulic conductivity of an intermediate layer (L_{int} in L and K_{int} in $\mathsf{L}\,\mathsf{T}^{-1}$):

$$C_{\text{D}} = \frac{A}{L_{\text{int}}} \cdot K_{\text{int}}. \tag{2.29}$$

The drain boundary differs from the seepage boundary by the thresholding term and the leaky flux, which allows for some overpressure to occur.

2.3 Gaussian Process Regression

Sometimes, during pre- or post-processing of modeling data, a collection of inputs and outputs is given, but the output is desired for an input value that is not part of the existing set. In such cases, *interpolation* techniques can be used to infer the desired quantity from the given data. Gaussian Process Regression (GPR) can be interpreted as a versatile stochastic interpolation method (that is also suitable for proxy-modeling). It makes use of so-called Gaussian Process Emulators (GPEs). Within this document, I will use the term GPE to refer to single interpolator instances, while GPR corresponds to the technique itself. Formally, GPR is identical to kriging (Krige, 1951; Matheron, 1963; Cressie, 1990). However, the term kriging is mostly associated with geo-spatial interpolation in one, two or three dimensions, while the term GPR is used for any dimensionality in the context of machine learning applications (e.g., Rasmussen, 2003). For a detailed derivation and explanation of GPR I refer the reader of this dissertation to Kitanidis (1997), Jones et al. (1998), Rasmussen and Williams (2006), and Erdal et al. (2020). In the following, I will provide a brief summary of GPR tailored towards my specific applications.

A GPE is defined for a specific interpolation problem using training data. These data consist of several pairs of input points $\boldsymbol{x}^{\text{train}}$ (each of size $1 \times n_{\text{dim}}$) and scalar output values $y = f(\boldsymbol{x}^{\text{train}})$ that were obtained by some external process f (e.g., by sampling a spatial field or by executing a numerical model). It is helpful to summarize the training points by vertical concatenation, which results in a matrix $\boldsymbol{X}^{\text{train}}$ of size $n_{\text{train}} \times n_{\text{dim}}$. Similarly, the outputs can be collected in a vector $\boldsymbol{y}^{\text{train}}$ of size $n_{\text{train}} \times 1$. The idea of GPR is to replace the data-generating process f by an approximation g, which is a multi-Gaussian field. The approximation g can then be used to estimate the output y^{est} for any test (i.e., query) point $\boldsymbol{x}^{\text{test}}$. The approximated output is a Gaussian random variable:

$$g(\boldsymbol{x}^{\text{test}}) = \mu(\boldsymbol{x}^{\text{test}}) + \mathcal{N}(0, \sigma^2_{\text{est}}(\boldsymbol{x}^{\text{test}})), \tag{2.30}$$

consisting of a mean function μ and a stochastic deviation term $\mathcal{N}$, namely a normal distribution with zero mean and a variance of σ^2_{est}. Both, μ and σ^2_{est} depend on the test point $\boldsymbol{x}^{\text{test}}$ in ways that are summarized in the following.

2.3.1 Prediction Mean and Variance

The mean function consists of a trend function β (here assumed to be a constant) and deviations from this trend:

$$\mu(\boldsymbol{x}^{\text{test}}) = \beta + \boldsymbol{r}^{\mathsf{T}}\boldsymbol{\xi}. \tag{2.31}$$

The vector $\boldsymbol{r}$ (size $n_{\text{train}} \times 1$) contains distance-based covariances between the test point and all training points. $\boldsymbol{\xi}$ (size $n_{\text{train}} \times 1$) can be interpreted as a vector of weights that can be determined (together with β) by solving a system of equations.

In other contexts, this equation system is often referred to as *ordinary kriging* system of equations:

$$\begin{bmatrix} \boldsymbol{Q} & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \beta \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}^{\text{train}} \\ 0 \end{bmatrix}. \tag{2.32}$$

The $n_{\text{train}} \times n_{\text{train}}$ matrix $\boldsymbol{Q}$ is the covariance matrix of all training data points. It should be noted that $\boldsymbol{Q}$ and $\boldsymbol{y}_{\text{train}}$, and therefore also β and $\boldsymbol{\xi}$, solely depend on the training data and not on $\boldsymbol{x}^{\text{test}}$. Each entry Q_{ij} is given by evaluating a covariance function C for two training data points:

$$Q_{ij} = C(\boldsymbol{X}_i^{\text{train}}, \boldsymbol{X}_j^{\text{train}}, \boldsymbol{\theta}), \tag{2.33}$$

where $\boldsymbol{X}_{\text{train},i}$ refers to the i-th training data point of shape $1 \times n_{\text{dim}}$. Similarly, each entry r_i of $\boldsymbol{r}$ is:

$$r_i = C(\boldsymbol{x}^{\text{test}}, \boldsymbol{X}_i^{\text{train}}, \boldsymbol{\theta}). \tag{2.34}$$

The vector $\boldsymbol{\theta}$ contains hyper-parameters depending on the training data (explained later).

Throughout this dissertation I use a stationary covariance function for GPR, which means that only the difference $\boldsymbol{x}_a - \boldsymbol{x}_b$ between the two evaluation points matters; not the absolute location of the points themselves (Rasmussen and Williams, 2006):

$$C(\boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{\theta}) = C(\boldsymbol{x}_a - \boldsymbol{x}_b, \boldsymbol{\theta}), \tag{2.35}$$

where $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ are placeholders for any training or test data point. One particular choice of stationary covariance function, that I rely on, is the Matérn covariance function of order three-half (Matérn, 1960; Stein, 1999):

$$C(d, \sigma^2) = \sigma^2 \left(1 + \sqrt{6}d\right) \exp\left(-\sqrt{6}d\right) \tag{2.36}$$

$$d(\boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{l}) = \sqrt{\sum_{i=1}^{n_{\text{dim}}} \left(\frac{x_{a,i} - x_{b,i}}{l_i}\right)^2}, \tag{2.37}$$

where σ^2 is a scalar variance and $\boldsymbol{l}$ is a $n_{\text{dim}} \times 1$ vector of correlation lengths. Note that there exist several versions of the Matérn covariance function differing in the constant (here $\sqrt{6}$). However, as this constant always occurs as a product with d, the differences between these factors are simply compensated by $\boldsymbol{l}$ during the GPE training. The covariance parameters σ^2 and $\boldsymbol{l}$ are typically summarized into a single vector of hyper-parameters $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \begin{bmatrix} \sigma^2 \\ \boldsymbol{l} \end{bmatrix}. \tag{2.38}$$

It is now obvious, that $\boldsymbol{\xi}$ and β can be determined easily from given training data, if the hyper-parameters $\boldsymbol{\theta}$ are known (by solving Equation 2.32). Determining (i.e., fitting or estimating) these

hyper-parameters from a given set of training data is a common procedure that is available in virtually all kriging/GPR toolboxes. In the "small toolbox for kriging" package (Bect, Vazquez, et al., 2022) that I use, this is done by using a restricted maximum likelihood method (Patterson and Thompson, 1971). It should be noted that estimating $\boldsymbol{l}$ can be decoupled from the estimation of σ^2, by maximizing the following likelihood term $\mathcal{L}$ (likelihood of the samples $\boldsymbol{y}_{\text{train}}$ given the input $\boldsymbol{X}_{\text{train}}$ and the length scales $\boldsymbol{l}$), where all constant contributions were omitted (Jones et al., 1998):

$$\log \mathcal{L} \propto -n_{\text{train}} \log\left(\frac{\left(\boldsymbol{y} - \mathbf{1}\frac{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\boldsymbol{y}^{\text{train}}}{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\mathbf{1}}\right)\boldsymbol{G}^{-1}\left(\boldsymbol{y} - \mathbf{1}\frac{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\boldsymbol{y}^{\text{train}}}{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\mathbf{1}}\right)}{n_{\text{train}}} \right) - \log(\det \boldsymbol{G}), \tag{2.39}$$

where $\boldsymbol{G}$ is analogous to $\boldsymbol{Q}$, but uses the correlation function C/σ^2 (the covariance function divided by the variance) for its entries. In the case of the small toolbox for kriging, the optimization problem of $\boldsymbol{l}$ is solved with Matlab's `fmincon` (Bect, Vazquez, et al., 2022). Afterwards, σ^2 (and even the constant trend coefficient β) can be determined with $\boldsymbol{G}$ using the optimized length scales $\boldsymbol{l}$ with the following relationships:

$$\beta = \frac{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\boldsymbol{y}^{\text{train}}}{\mathbf{1}^\mathsf{T}\boldsymbol{G}^{-1}\mathbf{1}} \tag{2.40}$$

$$\sigma^2 = \frac{(\boldsymbol{y} - \mathbf{1}\beta)\,\boldsymbol{G}^{-1}\,(\boldsymbol{y} - \mathbf{1}\beta)}{n_{\text{train}}}. \tag{2.41}$$

The weight vector $\boldsymbol{\xi}$ might then just as well be evaluated by

$$\boldsymbol{\xi} = \boldsymbol{Q}^{-1}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta). \tag{2.42}$$

Predictions of GPEs always come with uncertainty estimates by definition (as every prediction is a Gaussian random variable). The variance of a prediction is determined by

$$\sigma^2_{\text{est}}(\boldsymbol{x}^{\text{test}}) = \sigma^2 - \boldsymbol{r}^\mathsf{T}\boldsymbol{Q}^{-1}\boldsymbol{r}, \tag{2.43}$$

where σ^2 still refers to the variance in the covariance function (i.e., the first hyper-parameter in $\boldsymbol{\theta}$).

2.3.2 Interpolation

It is simple to quickly demonstrate how the described prediction procedure is an interpolation scheme, by testing a training point (as illustrated by Jones et al., 1998). As testing requires a trained GPE, we can assume all hyper-parameters (including β) to be known, which means we can use the simplified relationship of Equation 2.42 to evaluate $\boldsymbol{\xi}$.

We test for interpolation by taking one of the training points ($\boldsymbol{X}_i^{\text{train}}$) as test point $\boldsymbol{x}^{\text{test}}$:

$$\mu(\boldsymbol{x}^{\text{test}} = \boldsymbol{X}_i^{\text{train}}) = \beta + \boldsymbol{r}^\mathsf{T}\boldsymbol{Q}^{-1}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta). \tag{2.44}$$

In this case, $\boldsymbol{r}$ will be identical to the i-th row/column of the matrix $\boldsymbol{Q}$ (i.e., $\boldsymbol{Q}\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is a unit vector in the i-th direction). This allows the following rearrangement:

$$\begin{aligned}
\mu(\boldsymbol{x}^{\text{test}} = \boldsymbol{X}_i^{\text{train}}) &= \beta + \boldsymbol{r}^\mathsf{T}\boldsymbol{Q}^{-1}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta) && (2.45)\\
&= \beta + ((\boldsymbol{Q}^{-1})^\mathsf{T}\boldsymbol{r})^\mathsf{T}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta) && (2.46)\\
&= \beta + (\boldsymbol{Q}^{-1}\boldsymbol{r})^\mathsf{T}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta) && (2.47)\\
&= \beta + \boldsymbol{e}_i^\mathsf{T}(\boldsymbol{y}^{\text{train}} - \mathbf{1}\beta) && (2.48)\\
&= \beta + (y_i^{\text{train}} - \beta) = y_i^{\text{train}}, && (2.49)
\end{aligned}$$

where we exploit that $\boldsymbol{Q}$ is a symmetric matrix, as well as the following identity:

$$\boldsymbol{Q}^{-1}\boldsymbol{r} = \boldsymbol{Q}^{-1}(\boldsymbol{Q}\boldsymbol{e}_i) = (\boldsymbol{Q}^{-1}\boldsymbol{Q})\boldsymbol{e}_i = \boldsymbol{I}\boldsymbol{e}_i = \boldsymbol{e}_i. \tag{2.50}$$

Similarly, we can see that the variance at a training point has to be zero:

$$\begin{aligned}
\sigma^2_{\text{est}}(\boldsymbol{x}^{\text{test}} = \boldsymbol{X}_i^{\text{train}}) &= \sigma^2 - \boldsymbol{r}^\mathsf{T}\boldsymbol{Q}^{-1}\boldsymbol{r} && (2.51)\\
&= \sigma^2 - \boldsymbol{r}^\mathsf{T}\boldsymbol{e}_i && (2.52)\\
&= \sigma^2 - C(\boldsymbol{X}_i^{\text{train}}, \boldsymbol{X}_i^{\text{train}}, \boldsymbol{\theta}) && (2.53)\\
&= \sigma^2 - C(\boldsymbol{X}_i^{\text{train}} - \boldsymbol{X}_i^{\text{train}}, \boldsymbol{\theta}) && (2.54)\\
&= \sigma^2 - C(\mathbf{0}, \boldsymbol{\theta}) = \sigma^2 - \sigma^2 = 0. && (2.55)
\end{aligned}$$

2.3.3 Noisy Data

Depending on the application, it is possible to account for noise in the training data:

$$\begin{bmatrix} \boldsymbol{Q} + \boldsymbol{R} & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \beta \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}^{\text{train}} \\ 0 \end{bmatrix}, \tag{2.56}$$

where $\boldsymbol{R}$ would be the covariance matrix of training data errors. If unknown, it can be parametrized as $\boldsymbol{R} = \sigma^2_{\text{train}}\boldsymbol{I}$, where the standard deviation of the training data σ^2_{train} is an additional optimization parameter. However, since I am using GPR for the interpolation of deterministic data, I refrain from doing that and assume noise-free training data.

2.3.4 Derivative of Prediction Mean

As the GPE prediction is a smooth, multi-Gaussian field, an analytical derivative can be obtained by differentiation. As previously noted, the prediction mean is given by:

$$\mu(\boldsymbol{x}^{\text{test}}) = \beta + \boldsymbol{r}^\mathsf{T}\boldsymbol{\xi}. \tag{2.57}$$

We can take the derivative with respect to one of the input dimensions (here denoted i). The GPE is assumed trained, so all hyper-parameters (including β and σ^2) and the weight vector $\boldsymbol{\xi}$ are constants and independent of $\boldsymbol{x}_{\text{test}}$:

$$\frac{\partial \mu(\boldsymbol{x}^{\text{test}})}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\beta + \boldsymbol{r}^{\mathsf{T}}\boldsymbol{\xi}\right) \tag{2.58}$$

$$= \frac{\partial}{\partial x_i}\left(\boldsymbol{r}^{\mathsf{T}}\boldsymbol{\xi}\right) \tag{2.59}$$

$$= \left(\frac{\partial \boldsymbol{r}}{\partial x_i}\right)^{\mathsf{T}} \boldsymbol{\xi}. \tag{2.60}$$

The entries in the derivative of $\boldsymbol{r}$ with respect to one of the dimensions can be expressed in the following way by the chain rule of differentiation:

$$\frac{\partial r_j}{\partial x_i} = \frac{\partial C(d,\sigma^2)}{\partial x_i} = \frac{\partial C(d,\sigma^2)}{\partial d}\frac{\partial d}{x_i}, \tag{2.61}$$

where d is a shorthand of $d(\boldsymbol{x}^{\text{test}}, \boldsymbol{X}_j^{\text{train}}, \boldsymbol{l})$. With the Matérn covariance function, both components of this product can be differentiated analytically:

$$\frac{\partial C(d,\sigma^2)}{\partial d} = -6\sigma^2 d \exp\left(-\sqrt{6}d\right) \tag{2.62}$$

$$\frac{\partial d}{x_i} = \frac{x_i^{\text{test}} - X_{j,i}^{\text{train}}}{l_i^2 d}. \tag{2.63}$$

This results in:

$$\frac{\partial r_j}{\partial x_i} = -6\sigma^2 \frac{x_i^{\text{test}} - X_{j,i}^{\text{train}}}{l_i^2} \exp\left(-\sqrt{6}d\right). \tag{2.64}$$

3 Site Description

Within the framework of the Collaborative Research Center 1253 CAMPOS (Grathwohl et al., 2013), the Ammer catchment located west of Tübingen in South-Western Germany has been the site of coordinated interdisciplinary research (e.g., Chavez Rodriguez, 2021; Müller et al., 2021; Osenbrück et al., 2022; Petrova et al., 2022). The catchment's main river, the *Ammer*, originates in Herrenberg and stretches for about 22.5 km in south-eastern direction towards the city of Tübingen. The total surface-water catchment of river Ammer has a size of 238 km^2, of which 73 km^2 belong to the Goldersbach watershed, a tributary that only confluences with river Ammer shortly before the Ammer itself flows into river Neckar in Tübingen. Besides the mostly coniferous Goldersbach subcatchment, the Ammer valley is dominated by agricultural land-use with aggregates of urban areas and few small forests and grasslands.

The climate in the Ammer valley is warm and temperate, being located at the transition zone from a temperate oceanic climate to a warm-summer humid continental climate (Beck et al., 2018) as defined in the Köppen climate classification (Köppen, 1884). From January 1, 2014 to December 31, 2021, the measured mean annual precipitation was 608.5 mm a^{-1} and the mean temperature was 11.0 °C (with a 5-th percentile of −0.8 °C and a 95-th percentile of 25.0 °C; LTZ, 2021). Precipitation occurs throughout the year, with stronger, mostly convective, sporadic events in summer and more continuous, front-related, minor events in winter (Martin et al., 2020). Similar to other temperate regions, the majority of groundwater recharge takes place in winter.

3.1 Ammer Floodplain and Modeling Domain

The main study region of this dissertation is the Ammer floodplain, which has been the subject of parallel doctoral field research (e.g., Martin et al., 2020; Klingler et al., 2020a,b, 2021). The Ammer floodplain is an elongated portion of the Ammer catchment covering an area of about 8 km^2 from Pfäffingen via Unterjesingen to the west of the city of Tübingen (Klingler et al., 2020b). It exhibits a mostly flat topography around river Ammer in the center of the valley, with mild slopes pointing towards the river, as well as in its flow direction. To the north, the floodplain is bounded by steep hills. Towards the south-east, a ridge ("Kapellenberg" and "Spitzberg") presents the topographic surface-water divide between the Ammer and Neckar valleys. Likewise, on the south-western side, the surface-water divide is on a plateau ("Pfaffenberg"). Between these two steep ridges lies a gentler, undulating hillslope that separates the two valleys ("Wurmlingen saddle").

The Ammer river enters the floodplain on the west and leaves it eastbound. A prominent feature of the main river is a temporary split into two channels, the northern one representing a mill race. These two channels merge again in the floodplain, before another separation further downstream splits the stream into two reaches before passing through the city of Tübingen and discharging into the Neckar.

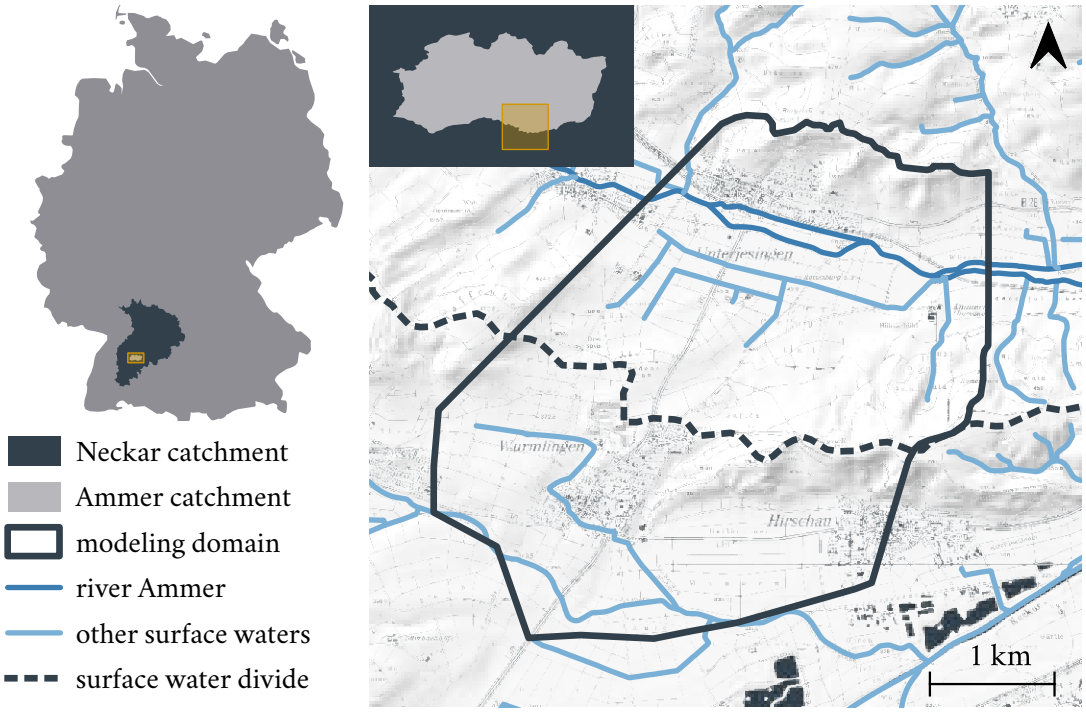

Figure 2: Two-dimensional overview of the model domain as used in Chapters III and IV.

For the site-specific models of Chapters III and IV, I define a model domain that covers most of the Ammer floodplain's areal extent. Figure 2 shows a map of this domain's outline, the surface-water divide between the Ammer and Neckar valleys, as well as streams and drainage channels. In total, the domain covers an area of 13 km^2, of which 7.2 km^2 are located north of the surface water divide (within the Ammer catchment), and 5.8 km^2 belong to the Neckar catchment according to the surface water divide. The surface elevation within the model domain ranges from approximately 330 m to 475 m above sea level.

The geometry of this domain is carefully chosen to approximately have the western and eastern boundaries perpendicular to assumed stream lines in the floodplain, while the boundary on the hillslopes are running along such stream lines (i.e., flow within the hillslopes is mostly assumed to be directed towards the floodplains). The decision on the outline is based on preliminary modeling and expert knowledge. The northern edge of the domain follows a local surface-water divide. At the southern edge, the domain partly covers the Neckar valley, but it does not reach river Neckar. This way, the groundwater divide between the Ammer and Neckar emerges from the model instead of being preset as a boundary condition (see Section 3.2 for further details).

3.2 Hydrogeologic Setting

Regional Geology The regional geology has been subject to many (hydro-)geological investigations (e.g., Quenstedt, 1864; Lang, 1909; Kekeisen, 1913; Köpf, 1926; Kehrer, 1935; Harreß, 1973; Villinger, 1982; D'Affonseca et al., 2020). The bedrock geology in the area is governed by sequences of sandstones and mudstones belonging to the Upper Triassic Keuper formation (Aigner

and Bachmann, 1992; Klingler et al., 2020b). The relevant bedrock formations for my work are briefly characterized in the following, in stratigraphic order from bottom to top:

- *Erfurt formation* (`kuE`): The Erfurt formation is roughly 20 m thick (Hagdorn and Nitsch, 2009; Kirchholtes and Ufrecht, 2015). It is composed of thin layers of sandstones and claystones with dolomite beds and extensive carbonate banks (Geyer and Gwinner, 2011). Towards the top, sulfate rocks (gypsum and anhydrite) can also be found (Hagdorn and Nitsch, 2009; Geyer and Gwinner, 2011). Its lower claystone layers (*Estherienton*) typically act as a regional aquitard, separating the overlying groundwater systems from the underlying regional karstified aquifer of the middle Triassic Muschelkalk formation (including the Meissner formation `moM`; Kleinert, 1976; D'Affonseca et al., 2018, 2020). However, some internal layers (e.g., sandstones and potentially fragmented/weathered carbonate and dolomite banks) of the Erfurt formation can also attain significant transmissivities (e.g., Schollenberger, 1998; Geyer and Gwinner, 2011). Literature information about hydraulic properties of the Erfurt formation is collected in Table 10 (in the appendix).

- *Grabfeld formation* (`kmGr`): The Grabfeld formation is a mudstone unit bearing gypsum, anhydrite and claystones (Geyer and Gwinner, 2011). It can reach thicknesses of about 100 m (Kleinert, 1976; LGRB, 2005). Up to 45 % of this formation can consist of sulfatic rocks (Hagdorn and Nitsch, 2009), which are prone to weathering. Its hydraulic properties vary strongly depending on the degree of this weathering. The unweathered, anhydrite-bearing parts of the Grabfeld formation are considered to be basically watertight, but they may be fractured to allow some water circulation (Geyer and Gwinner, 2011). Water contact within the Grabfeld formation transforms anhydrite to gypsum, which can be dissolved upon further weathering (Ufrecht, 2017). This can increase the hydraulic conductivity by orders of magnitude (Kirchholtes and Ufrecht, 2015; Ufrecht, 2017). Tables 11 and 12 (in the appendix) summarize literature information regarding the hydraulic properties of the unweathered and weathered Grabfeld formation. Table 13 provides a literature overview regarding the maximum weathering depth of the Grabfeld formation beneath the ground surface.

- *Mud- and sandstone formations* (`km2345`): The remaining bedrock formations of the middle Keuper (*Stuttgart formation* `kmSt`, *Steigerwald formation* `kmSw`, *Hassberge formation* `kmHb`, *Mainhardt formation* `kmMh`, *Löwenstein formation* `kmLw`, and *Trossingen formation* `kmTr`), contain interbedded sandstones, claystones, siltstones and dolomite layers (Geyer and Gwinner, 2011). In analogy to Kleinert (1976), Selle et al. (2013), and D'Affonseca et al. (2020), I consider these formations as a single, lumped hydrostratigraphic unit with uniform hydraulic properties. Although the fractured sandstone layers can be well permeable (Geyer and Gwinner, 2011), the hydraulic conductivity of the lumped units is often set to low values in models (e.g., Maier et al., 2013; Selle et al., 2013; D'Affonseca et al., 2018). Harreß (1973) reported a negligible influence of these layers on regional subsurface flow. Literature data on hydraulic properties of these formations are collected in Table 14 (in the appendix).

Three-dimensional spatial information of layer boundaries between these bedrock units is available from a geological model developed by D'Affonseca et al. (2018, 2020). In the domain of interest, the Grabfeld formation is the dominant subcropping unit. The lumped mud- and sandstone strata occur only at the top the hills (Pfaffenberg, Kapellenberg, Spitzberg, as well as at the northern end), where they serve as a cover of the Grabfeld formation. Towards the western end of the domain, two fault lines and a general pinching out of the Grabfeld formation lead to very thin covering of the Erfurt formation, up to a partial subcropping of this unit. This complex setup makes it difficult to assess, which parts of the Ammer floodplain are in (hydraulic) contact with which bedrock formation.

Local Hydrostratigraphy The Ammer and Neckar rivers have carved small, elongated basins into the bedrock (Martin et al., 2020), which are filled with Quaternary sediments forming the floodplains. The Quaternary material of the Ammer floodplain has been investigated by Martin et al. (2020) and Klingler et al. (2020b,a, 2021), who have identified the following four hydrostratigraphic units from bottom to top:

- *Gravel* (Ammer valley): The lowermost floodplain unit in the Ammer valley consists of a Pleistocene clayey gravel body, acting as a local aquifer. Its thickness is in the range of 5 m to 10 m (Klingler et al., 2020a). The estimated hydraulic conductivity of this layer ranges from 10^{-5} m s^{-1} to 10^{-3} m s^{-1} (Klingler et al., 2021). An overview of literature data regarding this unit is given in Table 15 (in the appendix). Klingler et al. (2020b) have hypothesized and detected a deeper paleo-channel structure within this unit, which is not explicitly considered in this dissertation.
- *Clay*: A silty clay unit of approximately 0.5 m to 3 m thickness separates the gravel aquifer from the overlaying formation (Klingler et al., 2020a). Due to lacking direct measurements regarding the hydraulic properties of this layer at the study site, I collected general literature data covering silty clay in Table 16 (in the appendix).
- *Tufa*: Above the clay unit, this Holocene layer consisting mostly of autochthonous calcareous aggregates forms another aquifer (Martin et al., 2020). It has a thickness of several meters (Klingler et al., 2020a). Estimates of hydraulic conductivity based on slug test measurements and single-well pumping test are in the range from 10^{-7} m s^{-1} to 10^{-4} m s^{-1} (Martin et al., 2020). A complete overview of available literature data is given in the appendix (Table 17).
- *Alluvial fines*: Finally, the upper Tufa aquifer is confined by an alluvial clay serving as the top of the Quaternary filling in the Ammer floodplain. This unit is several meters thick and consists of alluvial silty and clayey fines (Klingler et al., 2020a). Table 18 (in the appendix) provides an overview of hydraulic properties of alluvial fines (not limited to the Ammer site).

The three-dimensional layer boundaries of these formations have been inferred from borehole data (Martin et al., 2020). These layers were subsequently implemented into a refined version of the geological model of D'Affonseca et al. (2020).

Finally, I consider the following four additional units:

- *Gravel* (Neckar valley): The floodplain material on the Neckar side consists mostly of Quaternary sandy gravel sediments several meters thick. This unit was already considered as *Quaternary* in the geological model of D'Affonseca et al. (2020). Table 19 (in the appendix) provides an overview of relevant literature data.
- *Hollows*: Colluvial hollow fillings on the southern and northern hillslopes of the Ammer valley cutting into the bedrock formations have been hypothesized and detected (Martin et al., 2020). They reach thicknesses of about 4 m to 13 m and are partially filled with poorly sorted colluvial sediments (including some gravel fraction) deposited by mud-flows or similar processes (Martin et al., 2020). The three-dimensional geometry of these hollows was estimated from mapping out the alluvial fans of the geologic map (LGRB, 2005) and assuming subsurface slopes for their bottom contact. Available data regarding the hydraulic properties of these (or comparable) features are summarized in Table 20 (in the appendix).
- *Soil*: A top soil unit outside of the Ammer floodplain (where alluvial fines serves as top soil) can be considered in the flow model. The layer boundary can simply be constructed by offsetting the top surface elevation. Table 21 in the appendix contains related literature data.
- *River buffer*: Some drilling cores in the Ammer floodplain have been reported to contain comparably clean material deposited by river Ammer (Martin et al., 2020). Still, it is unclear to what extent river Ammer is connected to the subsurface, whether the riverbed generally consists of drastically different material than the alluvial fines, and how deep the riverbed is. Nevertheless, I account for a potential river buffer zone in the model, by introducing another hydrostratigraphic unit surrounding the river.

A three-dimensional rendering providing an overview of all described units considered in this dissertation is given in Figure 3.

Flow Conditions The groundwater recharge and drainage conditions within the Ammer floodplain and the outlined modeling domain are not entirely clear and partly subject of the investigations carried out in this dissertation. There is ambient flow from the upstream end of the Ammer floodplain to its downstream end. Little groundwater recharge is assumed to take place directly within the floodplain, due to the coverage by alluvial fines. However, lateral influxes from the hillslopes might exist (Martin et al., 2020). While such fluxes could be relevant from the perspective of flow within the floodplain aquifer(s), the Ammer Quaternary material is regarded as a barrier to such hillslope inflow, due to comparably small hydraulic conductivities (Martin et al., 2020).

Knowledge about interactions between the floodplain aquifers and surface water bodies is sparse. There is no detailed information about interactions with river Ammer, but the river is generally not regarded as a major net source of groundwater. Within the Ammer floodplain, south of the river, a network of artificial drainage channels has been installed (visible in Figure 2). Those drainage ditches

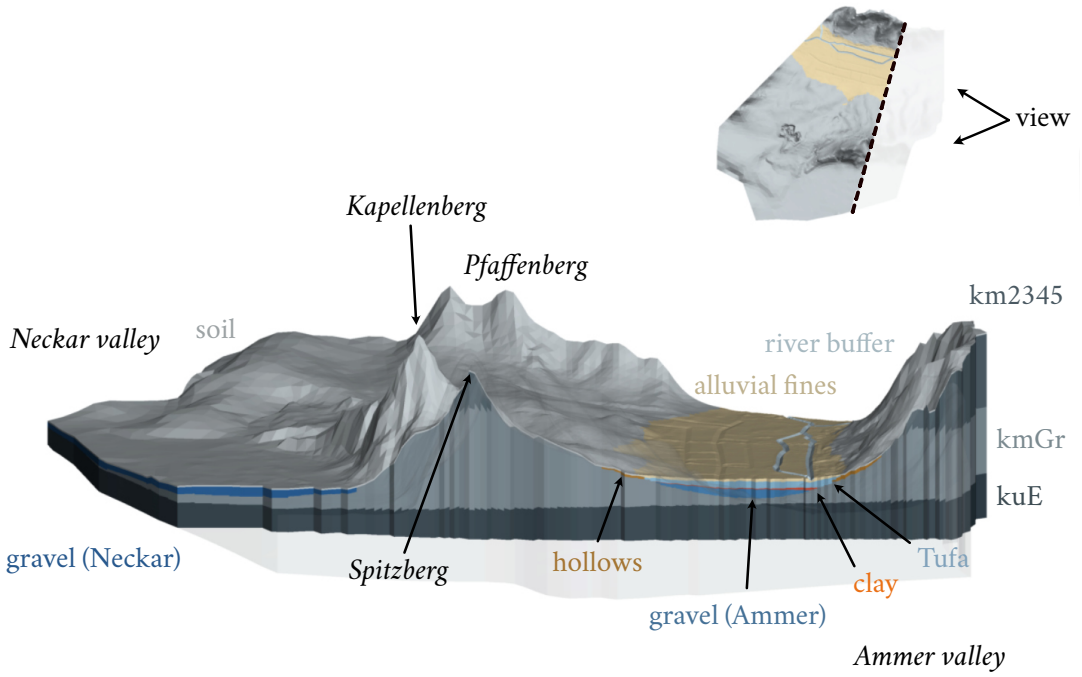

Figure 3: Three-dimensional overview of the geological model with fivefold vertical exaggeration.

running on the south-north axis on the hillslope are dry most of the time, with exception of storm events. The east-west drainages frequently carry water. In addition to these known drainage ditches, there is anecdotal evidence regarding subsurface drainages features (i.e., tile drains) that might have been installed around the beginning of the 20th century (e.g., Kehrer, 1935). Unfortunately, precise information regarding the existence and location of these drainage features is not available. The only surface water body on the Neckar side of the hills is a small creek ("Arbach"), which is considered to be of minor importance with respect to the overall flow field.

The hydraulic connection or separation of the Ammer and Neckar valleys in the vicinity of the Ammer floodplain is also not clear. Previous modeling studies suggested a shift of the groundwater divide in this area towards the Ammer catchment in the north (Kortunov, 2018). This hypothesis is supported by the Neckar valley being about 10 m lower than the Ammer valley elevation-wise, as well as by a partial dipping of the strata towards the south. However, prior to the work of this dissertation there were no groundwater observation wells installed along the southern hillslope of the Ammer floodplain that could have provided data to confirm or falsify this hypothesis.

3.3 Available Data

A large data set of different quantities observed over time is available for the domain of interest. These data comprise meteorological information, groundwater observations, and recordings of river-water stage. Figure 4 provides an overview of the respective time series. The period from October 1, 2018 to November 23, 2018 is emphasized, as it shows extraordinarily stable conditions

with respect to all observations. The steady-state simulations therefore focus on this period and the observations obtained on November 6, 2018 are used as a key-date reference for the calibration in Chapter IV.

Meteorology Meteorological data is available from Landwirtschaftliches Technologiezentrum Augustenberg (LTZ, 2021, Station #146 in Unterjesingen provides precipitation, global radiation, air temperature and relative humidity, Station #055 in Bondorf is the closest station that provides wind speed). Figure 4a shows the precipitation and potential evapotranspiration determined using the semi-empirical estimation equation of Penman (1956).

Groundwater Long-term field investigations targeting the Ammer floodplain have been conducted by Martin et al. (2020) in the framework of the Campos project. Over the course of this project, many groundwater observation wells have been installed in the gravel and Tufa aquifers of the Ammer floodplain. In total, 51 groundwater observation locations exist in the Ammer floodplain, but not all of them recorded hydraulic heads with the same frequency or over the same time frame. As shown in Figure 4e, most of the deep wells targeting the gravel aquifer discontinued measurements in the summer of 2019, because the respective data loggers were relocated (partly to the shallow Tufa aquifer). In the spring of 2020, three additional wells were installed in the Grabfeld formation on the southern hillslope towards the Wurmlingen saddle (partly as an outcome of Chapter III).

I use November 6, 2018 as a key-date for the model calibration in Chapter IV, because it is the date with the most simultaneous (i.e., recorded on the same day) hydraulic head observations during a more or less stable phase of flow conditions.

Additional hydraulic-head data obtained within or close to the study site (not shown in Figure 4) are available from regional and local water suppliers like the Ammertal-Schönbuchgruppe (ASG), the Stadtwerke Tübingen (SWT) and from the former Water & Earth System Science research cluster. In addition, the data of two observation wells in the Grabfeld formation located on top of the Wurmlingen saddle (in the municipality of Wurmlingen) are used in the calibration of Chapter IV (where they are referred to as "hill1c" and "hill2c"). These wells were installed on request of the county office because of subrosion problems in Wurmlingen.

River Water Stage Temporally resolved information about the water stage in river Ammer is available from a gauging station in Pfäffingen (about 700 m upstream of the investigated domain). Time series of water depth at this station are available from May 1, 2016 onward (LUBW, 2021, Landespegel Hydrologie, Messstellennummer 417, *Pfäffingen Ammer*). Up to October 31, 2017, LUBW (2021) also reported the corresponding river discharge. In addition to water-stage and discharge over time, there is a spatial data set (Seitz, 2010) concerning the cross-sectional geometry over the course of the river.

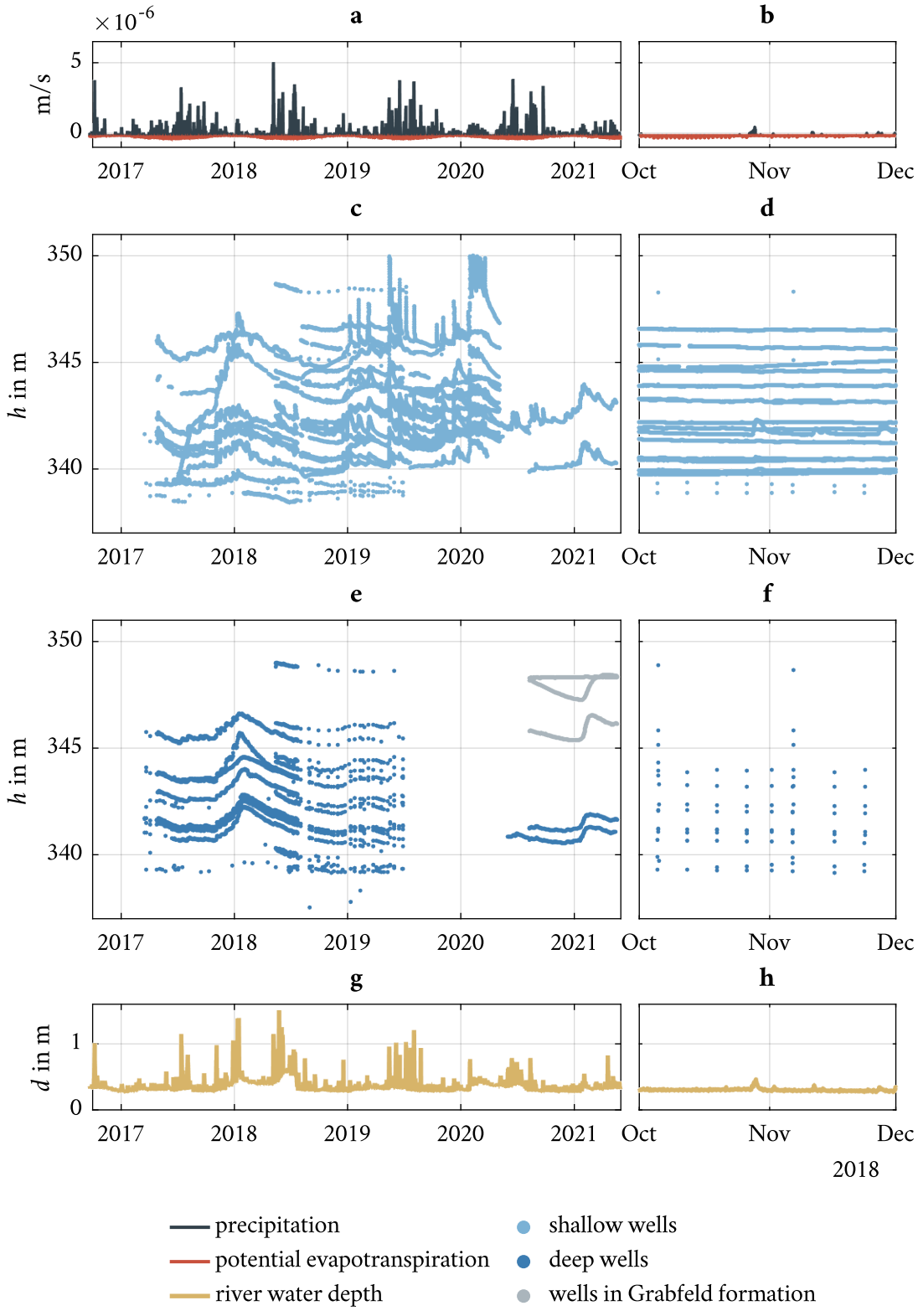

Figure 4: Overview of available data. The metorological data (**a**-**b**) stem from stations in Unterjesingen and Bondorf. The shown hydraulic head data (**c**-**f**) were obtained within the scope of the CAMPOS project. The river water stage data (**g**-**h**) are publicly available from LUBW (*d* represents the water depth at the gauging station).

3.4 Surface-Water Model of River Ammer

In the model of Chapter IV, the connection between subsurface flow and river Ammer is incorporated as a first-type boundary condition. Hence, spatially resolved information about the river-water stage as elevation data is necessary. Unfortunately, there is only a single gauging station in the vicinity of the model domain (see Section 3.3). With the help of the cross-sectional data collected by Seitz (2010), I set up a separate steady-state flow model to (1) convert a single river discharge value to a spatial data set of river-water head (i.e., river-water stage/elevation), and to (2) do this for different discharges, to create a rating curve for each point within the river. This way, all required fixed-head values can be generated from a single discharge value prescribed at the gauging station.

The flow model described in this section is independent of all other models in this dissertation. It was created and used only to infer the relationship between gauging station discharge and river water stage. The model uses the HGS surface-water implementation, which is based on the diffusive wave approximation of the Saint Venant equations (Viessman and Lewis, 1996; Fan and Li, 2006); for the implementation details, see Therrien et al. (2010). It therefore neglects any interactions between surface water and groundwater. This simplifying assumption is necessary, as otherwise a full coupling of surface and subsurface flow would have been needed. This would have required a high spatial resolution to resolve the river morphology and would have also resulted in infeasibly large runtimes. Neglecting the interactions with groundwater is reasonable from the perspective of river-water stage modeling, because the river discharge is much larger than any expected gains or losses from or towards groundwater: even if all rain water within the Ammer side of the modeling domain would infiltrate and eventually drain to the river, the resulting long-term flux of about $0.14\,\mathrm{m^3\,s^{-1}}$ would be small compared to the river discharge in the order of about $1\,\mathrm{m^3\,s^{-1}}$. Accounting for losses towards evapotranspiration and other sinks, the expected water gain of river Ammer within the domain can be safely neglected.

The model geometry is described by a triangular mesh in three spatial dimensions to replicate the real river bathymetry. The mesh consists of 36 670 triangles, whose vertical coordinates interpolate between 97 cross-sections. The interpolation of the one-dimensional cross-section data to a three-dimensional mesh is performed according to the method developed by Caviedes-Voullième et al. (2014). For the two-dimensional meshing, the model relies on the `mesh2D` Matlab toolbox based on the work of Engwirda (2014). The boundary conditions of this surface flow model are trivial: a fixed specified discharge boundary condition at the single upstream end operates as inlet; two critical-depth boundary conditions at both downstream ends serve as water outlets. The model uses a Manning's roughness coefficient of $0.03\,\mathrm{sm^{-1/3}}$ (Manning et al., 1890), as well as rill and obstruction storage heights of 10^{-6} m to allow for undisturbed surface flow. An overview of the model geometry, including locations of cross-sections and boundary conditions is given in Figure 5.

This model is run with six discharges between $0.25\,\mathrm{m^3\,s^{-1}}$ and $1.50\,\mathrm{m^3\,s^{-1}}$ in steps of $0.25\,\mathrm{m^3\,s^{-1}}$ (which covers low to more-than-average discharges). The spatially resolved output of river water

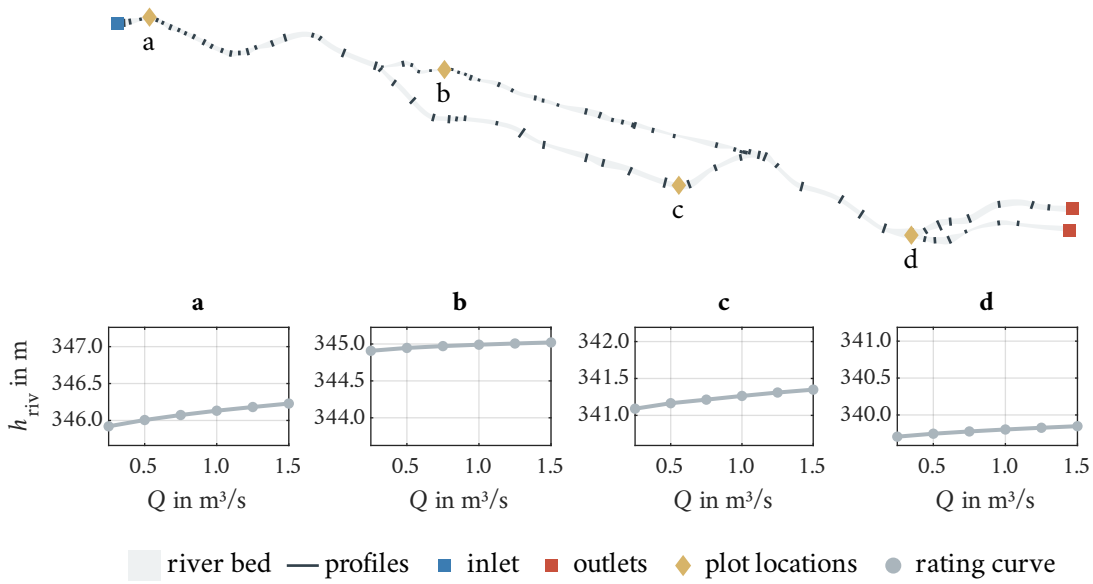

Figure 5: Overview of the surface flow model developed to simulate river Ammer. The four subplots (**a**-**d**) show rating curves (top of water surface h_{riv} in m above sea level versus discharge Q) obtained for four example locations. The bottoms of the vertical axes are aligned with the thalweg elevation at the specific coordinates.

stages is collected and used to define inverse rating curves. These can be used to infer river-water elevations from a given discharge value by linear interpolation. Four examples of such inverse rating curves at different locations are shown in Figure 5. The lower limits of the vertical axes in these plots correspond to the elevation of the thalweg at the specific locations to give a visual impression of river water depth. Multiple conclusions can be drawn from these results:

- The river-water depth is highly heterogeneous. This can be related to very different cross-sectional shapes throughout the river course. Wherever the cross-section resembles a ∨-shape, water depth increases; wherever the cross-section becomes wide and flat, the water depth shrinks. Any fixed-head derivation based on a fixed water depth above the thalweg elevation would therefore be erroneous.
- The slope of the river surface is not uniform (not visible in Figure 5). Instead, there are shorter segments of steeper parts between sections of mild slopes. A simple interpolation of assumed river head elevation at in- and outlets would therefore also be incorrect.
- The rating curves are relatively linear within the range of investigated discharges. However, the slopes differ between the locations. As a result, a change in river discharge might affect the river water head of some locations much more than at others. Therefore, an offsetting of a spatial field of river head (e.g., obtained from a base model) would be oversimplified.

Ultimately, a proper derivation of river-water heads from a fixed discharge requires a model like this, but fewer points in the inverse rating curves might have been sufficient to capture the essential relationship between river discharge and spatial fields of river-water heads.

3.5 Site-Specific Questions

In the following main chapters (Chapters II to IV) I do not only address the general questions raised in Section 1.3, but also the following site-specific questions, some of which were already implicitly raised in Section 3.2:

4. *Is geometry-driven valley-scale hyporheic exchange relevant for the Ammer-floodplain aquifers?*
5. *Is the groundwater divide between the Ammer and the Neckar valley in vicinity of the Ammer floodplain shifted, leading to inter-basin flow?*
6. *Can a steady-state subsurface flow model be calibrated to field data of the Ammer-floodplain aquifers to achieve a decent representation of the observed flow field?*
7. *What is the role of the hydrostratigraphic units in the Ammer-floodplain aquifers, also with respect to the interaction of the floodplain aquifers with the surroundings?*

Question 4 is addressed in Chapter II, Question 5 is discussed in Chapters III and IV, and Questions 6 and 7 are subject of Chapter IV.

Chapter II

Hyporheic Exchange in Idealized Floodplain Aquifers

Context

The contents of this chapter were published as "Systematic Evaluation of Geometry-Driven Lateral River-Groundwater Exchange in Floodplains" in *Water Resources Research* (Allgeier et al., 2021b). The author contributions are: Jonas Allgeier derived the semi-analytical solution, set up all semi-analytical models, performed the computations, created the figures, and wrote the draft manuscript; Simon Martin assisted with the introduction, helped with the interpretation of the results and contributed to manuscript revision; Olaf A. Cirpka conceived the presented idea, supervised the work, provided funding, and revised the manuscript draft.

The model source code used to generate all data of this study is available in form of a repository at https://osf.io/fykr9/ (Allgeier et al., 2021a). It comes in a command-line interface Matlab version and an interactive Python Bokeh application (Bokeh Development Team, 2021). This interactive tool is also available as a free online service (https://jonasallgeier.github.io/fpsimple) that can be accessed and executed in all common web browsers.

4 Introduction

Over the past decades, hydrogeological research has become aware of the necessity of jointly investigating surface and subsurface processes rather than treating them as separate domains (e.g., Winter et al., 1998). As a consequence, a key role in floodplain hydrogeology has been attributed to the interactions between aquifers and the connected rivers (e.g., Ward et al., 2016; Fritz et al., 2018; Ward and Packman, 2019). Especially hyporheic exchange has been identified as a key process of surface-water/groundwater interactions in floodplain aquifers (Triska et al., 1993; Hayashi and Rosenberry, 2002; Boano et al., 2009; Fabian et al., 2011; Ward, 2016; Lewandowski et al., 2019).

4.1 Lateral Hyporheic Exchange

Tonina and Buffington (2009) showed that changes in the cross-sectional area of floodplain aquifers are one of three drivers for hyporheic exchange besides non-uniform hydraulic conductivity and changes in energy head gradients. It is quite typical for floodplain aquifers, to have a varying cross-sectional area in settings of alternating degrees of valley confinements. In geomorphology, the term "confined" typically is used to describe valleys with a narrow floodplain that are laterally bounded

by steep flanks forming a typical ∨-shape (Baxter et al., 1999; Fotherby, 2009; Nagel et al., 2014). In contrast to that, unconfined valleys exhibit extensive floodplains and flat surface topography (Nanson and Croke, 1992; Nagel et al., 2014). Hence, a sequence of confined and unconfined valleys along a stream, as described by Stanford and Ward (1993), Baxter et al. (1999), and Nagel et al. (2014), results in lateral widening and narrowing of floodplain aquifers (Wohl, 2021). Such sequences can form when alternating softer and harder rock layers of a stratigraphic sequence dip into the direction of flow so that the river alternatingly cuts through these softer and harder rocks (Owen and Dahlin, 2005). At the up- and downstream ends of an unconfined basin, the river is hindered to erode the harder bedrock, whereas the soft bedrock in between can be carved out by the river. The widened valley is then filled with fluvial sediments and hillslope material forming the floodplain aquifer. In fact, floodplain aquifers with narrow inlet and outlet cross-sections and a basin-shaped central section occur frequently (e.g., Castro and Hornberger, 1991; Clément et al., 2003; Helton et al., 2012; Ó Dochartaigh et al., 2019; Martin et al., 2020). For example, Ohara et al. (2018) developed a simplified procedure to approximate the boundary of floodplain aquifers solely from digital elevation models. They made use of the fact that the top surfaces of floodplains exhibit small slopes, and that floodplain boundaries are characterized by inflection points of surface elevation (i.e., where the curvature is zero). The mappings of Ohara et al. (2018) show a number of floodplain aquifers along the investigated stream network following the described widening shape between connection points of small width.

As mentioned above, the widening geometry of an unconfined valley in the middle of two confined ones can sustain lateral hyporheic exchange on the valley-scale (Tonina and Buffington, 2009; Buffington and Tonina, 2009; Wondzell and Gooseff, 2013; Nagel et al., 2014). Where the valley widens, river water infiltrates into the aquifer, then flows predominantly down the valley, and is pushed back into the river where the valley narrows again (similar effects on vertical non-uniform cross-sections were already described by Vaux, 1968). This large-scale excursion of river water into the adjacent aquifer defines the riparian hyporheic-exchange zone from a hydrogeological perspective and is the main point of interest of this study. Figure 6 schematically shows that alternating transitions from (partially) confined to unconfined valleys (and vice versa) can be a cause of lateral geometry-driven hyporheic exchange even in straight river reaches, which can often be found in anthropogenically modified, channelized floodplain systems (e.g., Brookes, 1987).

The actual quantitative measurement of hyporheic exchange in field studies (in terms of flux, travel times, or spatial extent) remains a challenge, particularly on larger scales. A number of different experimental quantification techniques exist, including heat tracing (Ren et al., 2019), conservative-tracer tests (Mallard et al., 2014), isotope-data interpretation (Zhang et al., 2017), differential river-discharge measurements (Kalbus et al., 2006), geophysical exploration (Ward et al., 2010), or seepage-meter applications (Langhoff et al., 2006). Each of these methods comes with its own strengths and limitations (see Cook, 2015, for a detailed comparison). Quantitative hydrogeological models represent an attractive addition to field experiments for the quantification of hyporheic

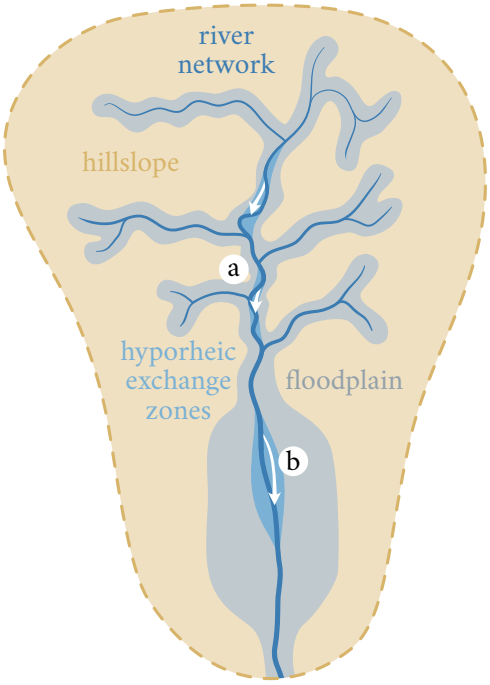

Figure 6: Conceptual drawing of large-scale, lateral, geometry-driven hyporheic exchange zones in a river catchment. **a**: Hyporheic exchange between river meanders (not part of this study). **b**: Hyporheic exchange in widening floodplains with straight (channelized) river section (focus of this study).

exchange, because they are comparably cheap, versatile and allow full control over the respective system. Consequently, there is a long history of surface-water/groundwater exchange modeling studies, of which we want to highlight the ones most significant to our investigations.

4.2 Previous Work and Knowledge Gap

An early two-dimensional model of vertical hyporheic exchange was developed by Vaux (1968), who used electric analogs to visualize his findings, as neither a closed-form solution nor a numerical approximation of his formulation were available/feasible at that time. Lateral hyporheic flow was modeled by Harvey and Bencala (1993), who applied the FDM to analyze the effect of alluvial streambed topography on hyporheic exchange in horizontal, rectangular domains. Revelli et al. (2008) used a FVM model to evaluate lateral hyporheic exchange occurring within a river meander. Similarly, Cardenas (2009a,b) developed FEM models to investigate hyporheic exchange between meanders in horizontal domains, incorporating also ambient river gain or loss. Huang and Chui (2018) derived proxy-equations for pool-riffle systems, serving as simplified empirical models to estimate the spatial scale of a hyporheic zone, as well as the related exchange flux and the median travel time of water flowing through it. Boyraz and Kazezyılmaz-Alhan (2013, 2017) developed (semi-)analytical solutions for two-dimensional flow in closed, horizontal, rectangular domains. As the only source and sink of groundwater in their cases is the simulated river, their studies can be interpreted as hyporheic flow investigations (where all water in the system belongs to the hyporheic

exchange zone by definition). Recently, they expanded their work by deriving an analytical solution for hyporheic exchange in rectangular systems under the influence of groundwater recharge from ponds and wetlands (Boyraz and Kazezyılmaz-Alhan, 2021). In summary, many of these models have targeted hyporheic exchange for various settings, including the case of Figure 6a. However, we see a lack of research studying valley-scale lateral hyporheic exchange driven by the geometry of the floodplain aquifers (i.e., the case in Figure 6b), which we want to address. To this end, we define an idealized two-dimensional plan-view model and solve it semi-analytically.

4.3 Objectives

In this study, we develop a semi-analytical solution for the valley-scale lateral hyporheic-exchange zone driven by the geometry of floodplain aquifers. We aim to answer three questions that are typically of interest when investigating hyporheic exchange (e.g., Kasahara and Wondzell, 2003; Welch et al., 2015; Huang and Chui, 2018). These questions address relevant aspects of biogeochemical reactive processes related to river-borne dissolved compounds:

- *How much water (in terms of the volumetric flux) flows through the valley-scale lateral hyporheic exchange zone?*

Answering this question allows comparing the exchange flux with the total river discharge and with total groundwater fluxes. If, for example, only a small fraction of the total river flow makes it into the aquifer, the river-water composition will not be drastically affected by the hyporheic exchange.

- *What is the spatial extent of the valley-scale lateral hyporheic zone?*

This is important whenever field studies within the aquifer are conducted, because it marks the boundary of "true" groundwater and infiltrated river water, which might carry a different chemical signature (e.g., micropollutants originating from waste-water effluents).

- *How long does the river water stay in the valley-scale lateral hyporheic exchange zone before returning to the river?*

The travel times quantify the contact time between water and aquifer material, which typically determines the degree to which any kinetic reactions or microbial interaction can take place.

After deriving the semi-analytical expression for an idealized two-dimensional floodplain aquifer, we perform a systematic parameter-variation study, to infer how the lateral exchange flux depends on the geometric and hydraulic input parameters. This allows us to construct approximate predictive proxy-models. We demonstrate the applicability of these proxy-models with two examples mimicking field sites close to Tübingen.

A comparable proxy-model for hyporheic exchange fluxes between sinuous river meanders has been developed by Cardenas (2009a,b). Of course, each floodplain site is unique with all its complex three-dimensional geology, morphology and heterogeneity as well as its dynamic processes taking

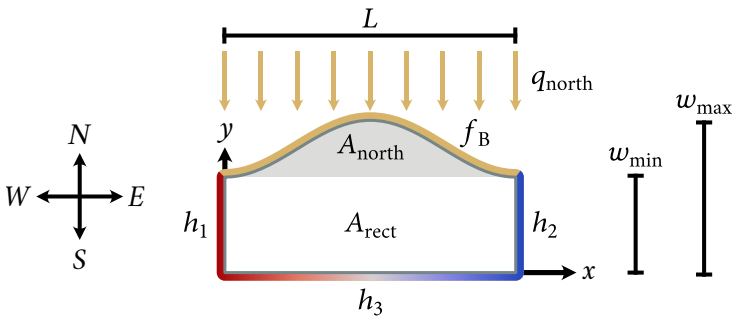

Figure 7: Schematic illustration of the two-dimensional problem definition. The tan line in the north represents a Neumann boundary with constant flux q_{north}. The other three lateral sides reflect Dirichlet boundaries with values h_1, h_2 and $h_3(x)$ (warmer colors indicate higher hydraulic heads). L, w_{min} and w_{max} represent the length of the domain as well as its minimum and maximum width. The shape of the northern boundary is defined by $f_{\text{B}}(x)$, which creates an additional area A_{north} compared to the rectangular area A_{rect}.

place on multiple scales. It is therefore certainly not possible to capture all details of a site with our simplified model, but it can give an easy and a quick order-of-magnitude type of estimation for the hyporheic-zone metrics of interest in cases where only little information about the floodplain is known (this motivation is similar to the one of Huang and Chui, 2018). Finally, our study gives insight into major dependencies, which might get lost in the details of a site-specific modeling study.

5 Methods

5.1 Conceptual Model and Problem Statement

We simulate steady-state horizontal groundwater flow in a simplified two-dimensional aquifer (see schematic illustration in Figure 7) without internal sources or sinks. The overall geometry idealizes the mid-section subcatchment of a typical channelized river in a landscape formed on bedrocks of alternating competence. In such settings, the large-scale along-valley ambient hydraulic gradient and the topography (including aquifer bottom and top) essentially follow the slope of the river, whereas the across-valley slopes are negligible. Laterally connected hillslopes are not part of our modeling domain; their effect on the floodplain aquifer are considered by a boundary condition. The restriction to two dimensions can be justified by the typical large lateral extent of floodplain aquifers compared to their small and mostly uniform thickness (e.g., Clément et al., 2003). For ease of description, we will use the terms "northern" (in the direction of y), "eastern" (in the direction of x), "southern" (against the direction of y) and "western" (against the direction of x) to denote the four directions and boundaries.

The domain extends laterally within $0 \leq x \leq L$ and $0 \leq y \leq f_{\text{B}}(x)$. Here, L in L represents the domain length. The continuous and real-valued function $f_{\text{B}}(x)$ defines the location of the northern boundary and thereby the domain width. At this northern end, a fixed specific flux

$q_{\text{north}}(x)$ in $\mathsf{L}^3\,\mathsf{L}^{-1}\,\mathsf{T}^{-1}$ crosses the boundary in y-direction, which simulates lateral inflow of an adjacent hillslope ($q_{\text{north}} < 0$ implies an influx). In the following derivations, we do not impose any further assumption or restriction with respect to $f_{\text{B}}(x)$, but in the investigated cases we will focus on a "cosinusoidal" curve mimicking the widening shape of a floodplain aquifer:

$$f_{\text{B}}(x) = w_{\text{min}} + \frac{1}{2}(w_{\text{max}} - w_{\text{min}}) \cdot \left(1 - \cos\left(2\pi\frac{x}{L}\right)\right), \tag{5.1}$$

where w_{min} in L is the minimum width of the domain (which applies for the western and eastern end) and w_{max} in L is the maximum width (which occurs at $x = L/2$). In later comparisons, we also investigate two alternate shapes that we denote "composite" and "bump". The latter shape follows a classical bump function defined by the points $(0|w_{\text{min}})$, $(L/2|w_{\text{max}})$ and $(L|w_{\text{min}})$:

$$f_{\text{B}}(x) = w_{\text{min}} + (w_{\text{max}} - w_{\text{min}}) \cdot \zeta\left(\frac{x}{L}\right) \tag{5.2}$$

$$\zeta(x') = \begin{cases} \exp\left(1 - \frac{1}{1-(2x'-1)^2}\right) & ;\ x' \in (0,1) \\ 0 & ;\ \text{otherwise.} \end{cases} \tag{5.3}$$

The "composite" configuration is defined as a piece-wise function by connecting the six points $(0|w_{\text{min}})$, $(x_1|w_{\text{min}})$, $(x_2|w_{\text{max}})$, $(x_3|w_{\text{max}})$, $(x_4|w_{\text{min}})$ and $(L|w_{\text{min}})$ with three constant and two cosinusoidal segments to achieve an elongated shape with continuous first derivative:

$$f_{\text{B}}(x) = \begin{cases} w_{\text{min}} & ;\quad x < x_1 \\ w_{\text{min}} + \frac{1}{2}(w_{\text{max}} - w_{\text{min}})\left(1 - \cos\left(\pi\frac{x-x_1}{x_2-x_1}\right)\right) & ;\ x_1 \le x < x_2 \\ w_{\text{max}} & ;\ x_2 \le x < x_3 \\ w_{\text{min}} + \frac{1}{2}(w_{\text{max}} - w_{\text{min}})\left(1 + \cos\left(\pi\frac{x-x_3}{x_4-x_3}\right)\right) & ;\ x_3 \le x < x_4 \\ w_{\text{min}} & ;\ x_4 \le x, \end{cases} \tag{5.4}$$

with the four parameters $x_1 = 0.025L$, $x_2 = 0.375L$, $x_3 = 0.625L$ and $x_4 = 0.975L$. These parameters create a straight central section of length $L/2$ and straight inlet and outlet sections with a total length of a tenth of that.

Figure 8 shows example geometries for all three domain shapes. We denote the rectangular area of the southern part A_{rect} in L^2 and the additional northern area A_{north} in L^2 (see Figure 7), such that the total surface area of the domain A_{tot} in L^2 is the sum of the two:

$$A_{\text{tot}} = A_{\text{rect}} + A_{\text{north}} = L \cdot w_{\text{min}} + A_{\text{north}}. \tag{5.5}$$

At the western and eastern boundaries, the fixed heads h_1 in L (western boundary) and h_2 in L (eastern boundary) impose an ambient flow field, which we assume to be connected to adjacent aquifers up- and downstream of the investigated catchment segment. The southern boundary is

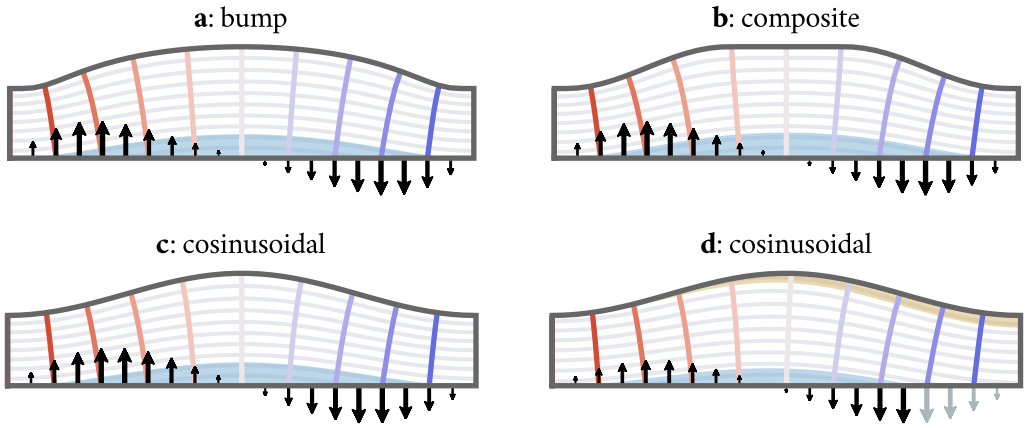

Figure 8: Flow-net examples with colored head contour lines, gray streamlines, and arrow indicators of exchange flux. Black arrows highlight fluxes that occur within the hyporheic exchange zone (shown in blue). The zone of water originating from the northern boundary is shown in tan. The flow nets, flux indicators, and areas visualize actual results obtained with the semi-analytical solution. **a**-**c**: Different shapes without northern influx. **d**: A case with northern influx.

assumed to be in perfect hydraulic contact with a river that provides another fixed-head boundary with $h_3(x)$ linearly varying between h_2 and h_1. We allow the hydraulic conductivity to be anisotropic in the principal directions (x and y), leading to a diagonal transmissivity tensor $\boldsymbol{T}$ in $\mathsf{L}^2\,\mathsf{T}^{-1}$:

$$\boldsymbol{T} = \begin{pmatrix} T_x & 0 \\ 0 & T_y \end{pmatrix}. \tag{5.6}$$

The evaluation of travel times requires a depth-integrated flow-effective porosity, or the aquifer thickness times the mean porosity, which we denote Φ in L.

5.2 Semi-Analytical Solution

The starting point of our derivation is the two-dimensional, steady-state, anisotropic groundwater-flow equation for divergence-free flow. It involves the hydraulic head h, the spatial coordinates x and y and the transmissivities T_x and T_y (e.g., Bear, 1972):

$$\frac{\partial^2 h}{\partial x^2} + \frac{T_y}{T_x}\frac{\partial^2 h}{\partial y^2} = 0. \tag{5.7}$$

The stream function Ψ in $\mathsf{L}^3\,\mathsf{T}^{-1}$ helps in formulating Neumann boundaries and allows for trivial flux evaluations after obtaining its solution. The Cauchy-Riemann equations relate the hydraulic head with the stream function (e.g., Bear, 1972; Strack, 2017):

$$\frac{\partial \Psi}{\partial x} = T_y \frac{\partial h}{\partial y} \tag{5.8}$$

$$\frac{\partial \Psi}{\partial y} = -T_x \frac{\partial h}{\partial x}. \tag{5.9}$$

We want to find the solution $h(x, y)$ of Equation 5.7 meeting the following boundary conditions:

$$h = h_1 \qquad \text{at } x = 0 \tag{5.10}$$

$$h = h_2 \qquad \text{at } x = L \tag{5.11}$$

$$h = h_1 + \frac{h_2 - h_1}{L} x \qquad \text{at } y = 0 \tag{5.12}$$

$$-(\boldsymbol{T}\nabla h) \cdot \boldsymbol{n}(x) = \frac{1}{\sqrt{f'_{\mathrm{B}}(x)^2 + 1}} q_{\mathrm{north}}(x) \quad \text{at } y = f_{\mathrm{B}}(x), \tag{5.13}$$

where $\boldsymbol{n}(x)$ is the unit normal vector along the northern boundary (pointing outwards) and $f'_{\mathrm{B}}(x)$ is the derivative of $f_{\mathrm{B}}(x)$ in the x-direction. Similar to the derivation of Read (2007), it is possible to express the northern Neumann boundary in terms of the stream function. To do that, we start with the definitions of the unit normal vector of a function (in this case $f_{\mathrm{B}}(x)$):

$$\boldsymbol{n}(x) = \frac{1}{\sqrt{f'_{\mathrm{B}}(x)^2 + 1}} \begin{bmatrix} -f'_{\mathrm{B}}(x) \\ 1 \end{bmatrix}. \tag{5.14}$$

Applying the hydraulic transmissivity tensor to the gradient of the hydraulic head and evaluating the scalar product with the unit normal vector converts Equation 5.13 to:

$$T_x \cdot f'_{\mathrm{B}}(x) \cdot \frac{\partial h}{\partial x} - T_y \cdot \frac{\partial h}{\partial y} = q_{\mathrm{north}}(x). \tag{5.15}$$

The application of the Cauchy-Riemann equations (definition of the stream function) yields:

$$-f'_{\mathrm{B}}(x)\frac{\partial \Psi}{\partial y} - \frac{\partial \Psi}{\partial x} = q_{\mathrm{north}}(x) \tag{5.16}$$

$$f'_{\mathrm{B}}(x)\frac{\partial \Psi}{\partial y} + \frac{\partial \Psi}{\partial x} = -q_{\mathrm{north}}(x). \tag{5.17}$$

Since we are operating only along the northern boundary, $f_{\mathrm{B}}(x)$ equals y and $f'_{\mathrm{B}}(x)$ equals $\frac{\mathrm{d}y}{\mathrm{d}x}$:

$$\frac{\mathrm{d}y}{\mathrm{d}x}\frac{\partial \Psi}{\partial y} + \frac{\partial \Psi}{\partial x} = -q_{\mathrm{north}}(x). \tag{5.18}$$

The definition of the total derivative can be used to summarize the terms of the stream function:

$$\frac{\mathrm{d}\Psi}{\mathrm{d}x} = -q_{\mathrm{north}}(x). \tag{5.19}$$

This ordinary differential equation can be integrated to get the stream function formulation of the northern boundary equation:

$$\Psi(x, f_{\mathrm{B}}(x)) = -\int q_{\mathrm{north}}(\xi)\ \mathrm{d}\xi = R(x) \quad \text{at } y = f_{\mathrm{B}}(x). \tag{5.20}$$

Equation 5.13 and Equation 5.20 are equivalent. Note that the stream function is only defined subject to an arbitrary constant offset. Like in most studies, we are only interested in differences between stream-function values so that the offset drops out. To clarify dimensions: $q_{\text{north}}(x)$ is a (potentially location-dependent) specific flux expressed in $\mathsf{L}^2\,\mathsf{T}^{-1}$ and $R(x)$ represents a fixed value for the stream function and is therefore expressed in $\mathsf{L}^3\,\mathsf{T}^{-1}$.

The solution to the boundary value problem is a function of space $h(x, y)$. It can be split into several components that are superimposed, because the problem is linear. Each of these components has to fulfill the groundwater flow equation, and the sum of them has to fulfill the boundary conditions. For the problem at hand, it makes sense to attribute a linear head variation in x, which is given by the western, southern and eastern boundary, to one component of $h(x, y)$ and to attribute all deviations to another component $\varphi(x, y)$:

$$h(x, y) = h_1 + \frac{h_2 - h_1}{L} x + \varphi(x, y). \tag{5.21}$$

Therefore, it is sufficient to find the solution of $\varphi(x, y)$ and to make sure that all boundary conditions are fulfilled. This leads to another anisotropic partial differential equation:

$$\nabla^2 \varphi(x, y) = \frac{\partial^2 \varphi}{\partial x^2} + \frac{T_y}{T_x} \frac{\partial^2 \varphi}{\partial y^2} = 0. \tag{5.22}$$

This problem has three simplified boundary conditions:

$$\varphi(0, y) = \varphi(L, y) = \varphi(x, 0) = 0. \tag{5.23}$$

For now, we ignore the northern boundary. As an ansatz we use a product of two terms that depend only on one of the two spatial variables:

$$\varphi(x, y) = X(x) \cdot Y(y) \tag{5.24}$$

$$\frac{\partial^2 \varphi}{\partial x^2} = \frac{\partial^2 X}{\partial x^2} \cdot Y \tag{5.25}$$

$$\frac{\partial^2 \varphi}{\partial y^2} = \frac{\partial^2 Y}{\partial y^2} \cdot X. \tag{5.26}$$

These expressions can be substituted into Equation 5.22, yielding:

$$\frac{\partial^2 X}{\partial x^2} \cdot Y + \frac{T_y}{T_x} \frac{\partial^2 Y}{\partial y^2} \cdot X = 0. \tag{5.27}$$

Using multiplication and division, we can separate the variables:

$$\frac{1}{X} \frac{\partial^2 X}{\partial x^2} = -\frac{T_y}{T_x} \frac{1}{Y} \frac{\partial^2 Y}{\partial y^2}. \tag{5.28}$$

The two sides have to be equal for all values of x and y, which means that both terms must be constants. We choose a positive constant term k^2 for now and will account for a negative one later:

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} = -\frac{T_y}{T_x}\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} = k^2. \tag{5.29}$$

We can separate the partial differential equation into two decoupled ordinary differential equations:

$$\frac{\partial^2 X}{\partial x^2} = X \cdot k^2 \tag{5.30}$$

$$\frac{\partial^2 Y}{\partial y^2} = -Y \cdot \frac{T_x}{T_y}k^2. \tag{5.31}$$

For these equations, we know the possible general solutions:

$$X(x) = c_1 \exp(kx) + c_2 \exp(-kx) \tag{5.32}$$

$$Y(y) = c_3 \sin\left(k\sqrt{\frac{T_x}{T_y}}y\right) + c_4 \cos\left(k\sqrt{\frac{T_x}{T_y}}y\right). \tag{5.33}$$

We remember that we could have also chosen $-k^2$ as the constant term, which gives us one more set of general solutions:

$$X(x) = c_5 \sin(kx) + c_6 \cos(kx) \tag{5.34}$$

$$Y(y) = c_7 \exp\left(k\sqrt{\frac{T_x}{T_y}}y\right) + c_7 \exp\left(-k\sqrt{\frac{T_x}{T_y}}y\right). \tag{5.35}$$

Iterating over all possible products of the subterms of $X(x)$ and $Y(y)$, we know that the solution must be a linear combination of the following terms:

$$\varphi(x,y) = \begin{Bmatrix} \exp(kx)\sin(k\kappa y), & \exp(-kx)\sin(k\kappa y), \\ \exp(kx)\cos(k\kappa y), & \exp(-kx)\cos(k\kappa y), \\ \exp(k\kappa y)\sin(kx), & \exp(-k\kappa y)\sin(kx), \\ \exp(k\kappa y)\cos(kx), & \exp(-k\kappa y)\cos(kx) \end{Bmatrix}, \tag{5.36}$$

where we use κ as a shorthand for $\kappa = \sqrt{\frac{T_x}{T_y}}$.

At this point, we take care of the boundary conditions. First of all, we know that the northern boundary is the only boundary with a value of $\varphi \neq 0$. Due to the diffusive nature of the Laplace equation, there is no reason for any oscillation in the y-direction implying that we can drop all terms involving a trigonometric function of y:

$$\varphi(x,y) = \Big\{ \exp(k\kappa y)\sin(kx),\ \exp(-k\kappa y)\sin(kx),\ \exp(k\kappa y)\cos(kx),\ \exp(-k\kappa y)\cos(kx) \Big\}. \tag{5.37}$$

The southern boundary ($\varphi(x, 0) = 0$) requires that the coefficients of the linear combination involving the sine functions (and the cosine functions respectively) need to cancel each other out:

$$\varphi(x, y) = \sum_{n=1}^{\infty} \alpha_n \sin(kx) \left(\exp(k\kappa y) - \exp(-k\kappa y)\right) + \sum_{n=1}^{\infty} \beta_n \cos(kx) \left(\exp(k\kappa y) - \exp(-k\kappa y)\right). \tag{5.38}$$

We can simplify this lengthy expression by applying the definition of the hyperbolic sine function:

$$\sinh(t) = \frac{1}{2}\exp(t) - \frac{1}{2}\exp(-t) \tag{5.39}$$

$$\varphi(x, y) = \sum_{n=1}^{\infty} \alpha_n \sin(kx)\frac{1}{2}\sinh(k\kappa y) + \sum_{n=1}^{\infty} \beta_n \cos(kx)\frac{1}{2}\sinh(k\kappa y). \tag{5.40}$$

Without loss of generalization, we can attribute the constant factors of one half to the linear combination coefficients α_n and β_n to keep the equation succinct:

$$\varphi(x, y) = \sum_{n=1}^{\infty} \alpha_n \sin(kx)\sinh(k\kappa y) + \sum_{n=1}^{\infty} \beta_n \cos(kx)\sinh(k\kappa y). \tag{5.41}$$

The western boundary condition ($\varphi(0, y) = 0$) requires dropping terms involving the cosine of x:

$$\varphi(x, y) = \sum_{n=1}^{\infty} \alpha_n \sin(kx)\sinh(k\kappa y). \tag{5.42}$$

The eastern boundary condition ($\varphi(L, y) = 0$) can only be fulfilled, if the following condition holds, which imposes a restriction on k:

$$\sin(k \cdot L) = 0 \tag{5.43}$$

$$k = \frac{l \cdot \pi}{L}; \quad l \in \mathbb{Z} \setminus \{0\}. \tag{5.44}$$

A solution with $l = 0$ does not need to be considered, as the respective contribution to $\varphi(x, y)$ collapses to zero. As the sine function and the hyperbolic sine function are odd, we can assume that all $l < 0$ can be considered by adjusting the coefficients $\alpha_{(-l)}$. Therefore, l can be restricted to $1 \leq l \leq \infty$, and we can use $l = n$:

$$\varphi(x, y) = \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{n \cdot \pi}{L}x\right)\sinh\left(\frac{n \cdot \pi}{L}\kappa y\right). \tag{5.45}$$

The infinite series solution of the original groundwater flow problem thereby becomes:

$$h(x, y) = h_1 + \frac{h_2 - h_1}{L} \cdot x + \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{n \cdot \pi}{L}x\right)\sinh\left(\frac{n \cdot \pi}{L}\kappa y\right), \tag{5.46}$$

where α_n in L is the n-th series coefficient and κ represents the dimensionless square root of the anisotropy ratio $\kappa = \sqrt{T_x/T_y}$.

The associated stream function is:

$$\Psi(x, y) = -T_x \cdot \left(\alpha_0 + \frac{h_2 - h_1}{L} \cdot y + \frac{1}{\kappa} \sum_{n=1}^{\infty} \alpha_n \cos\left(\frac{n \cdot \pi}{L} x\right) \cosh\left(\frac{n \cdot \pi}{L} \kappa y\right) \right), \tag{5.47}$$

in which the coefficient α_0 in L reflects the arbitrary offset of the stream function. The series coefficients $\boldsymbol{\alpha}$ (i.e., α_0 to α_∞) are fully determined by the shape of the northern boundary $f_{\mathrm{B}}(x)$ and the associated boundary condition $R(x)$ (Equation 5.20):

$$-T_x \cdot \left(\alpha_0 + \frac{h_2 - h_1}{L} \cdot f_{\mathrm{B}}(x) + \frac{1}{\kappa} \sum_{n=1}^{\infty} \alpha_n \cos\left(\frac{n \cdot \pi}{L} x\right) \cosh\left(\frac{n \cdot \pi}{L} \kappa f_{\mathrm{B}}(x)\right) \right) \stackrel{!}{=} R(x). \tag{5.48}$$

However, except for very simple cases (e.g., a valley with uniform width and constant normal flux at the northern boundary) it is practically impossible to determine the coefficients analytically. We alleviate this problem by numerical approximations of the series with a finite number of terms N. We use an approach similar to the Analytical Element Method of Strack (1989), that was refined and applied by Barnes and Janković (1999), Janković and Barnes (1999), Read (2007) and Craig (2008). In essence, we choose a finite number of M points along the northern boundary $(x_i | f_{\mathrm{B}}(x_i))$ that are used to evaluate the $N + 1 \leq M$ coefficients $\boldsymbol{\alpha}$ of the truncated infinite series.

The solution of Equation 5.48 can then be obtained through a least-squares formulation minimizing the sum of squared approximation errors ε along the northern boundary:

$$\varepsilon_i(\boldsymbol{\alpha}) = \alpha_0 + \sum_{n=1}^{N} \alpha_n f_n(x_i) - g(x_i), \tag{5.49}$$

with

$$g(x_i) = -\frac{R(x_i)}{T_x} - \frac{h_2 - h_1}{L} \cdot f_{\mathrm{B}}(x_i), \tag{5.50}$$

$$f_n(x_i) = \frac{1}{\kappa} \cos\left(\frac{n \cdot \pi}{L} x_i\right) \cosh\left(\frac{n \cdot \pi}{L} \kappa f_{\mathrm{B}}(x_i)\right). \tag{5.51}$$

The goal is to minimize the (weighted) sum of squared approximation errors across the M points:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{M} w_i \cdot \varepsilon_i(\boldsymbol{\alpha})^2, \tag{5.52}$$

in which w_i represents the weight of the i-th point. This optimization problem is converted to a system of linear equations by setting the derivative to zero (Craig, 2008). This yields $N + 1$ equations with $N + 1$ unknowns, regardless of the number of points M:

$$\alpha_0 \sum_{i=1}^{M} w_i \cdot f_\eta(x_i) + \sum_{n=1}^{N} \alpha_n \cdot \sum_{i=1}^{M} w_i f_\eta(x_i) f_n(x_i) = \sum_{i=1}^{M} w_i f_\eta(x_i) \cdot g(x_i); \quad \eta = 1 \ldots N, \tag{5.53}$$

$$\alpha_0 + \sum_{n=1}^{N} \alpha_n \cdot \sum_{i=1}^{M} w_i f_n(x_i) = \sum_{i=1}^{M} w_i \cdot g(x_i). \tag{5.54}$$

In the case of $M = N + 1$, the system is fully determined and the solution will be met exactly on all M points. Then, these points are referred to as *collocation points* (Barnes and Janković, 1999). Unfortunately, the Gibbs-Wilbraham (Wilbraham, 1848; Gibbs, 1898) and Runge phenomena (Runge, 1901) can lead to significant and strong oscillations between the collocation points. These artifacts cannot be reduced by increasing the approximation order N (for more information see, for example, Hewitt and Hewitt, 1979; Read, 2007; Ray, 2020). Instead, it can be beneficial to choose more points M than coefficients $N + 1$ to reduce the adverse effects of the Gibbs-Wilbraham phenomenon by creating an over-determined system of equations. The corresponding solution will then not be met exactly at all M points, but the error is minimized in an average sense. In such a case of $N + 1 < M$ the points are typically called *control points* (Barnes and Janković, 1999). In our study, we use $N = 10$ and $M = 25$.

As highlighted by Craig (2008), additional Gibbs-Wilbraham effects can be caused wherever Dirichlet and Neumann boundaries meet at angles that lead to inconsistent hydraulic gradients. In our model setup this can happen at the intersection points of the northern boundary and the western and eastern boundaries, depending on the choice of $f_{\mathrm{B}}(x)$ and $q_{\mathrm{north}}(x)$. In problematic cases, according to Craig (2008), non-uniform weights with smaller values close to these points can reduce the influence of these inconsistencies on the remaining parts of the domain. Another way to deal with the problem is to use a non-uniform spacing of the points x_i. Neither of these remedies were necessary in this study, so that we used identical weights for all points x_i spaced equidistantly (tests with parameter combinations outside of the investigated ranges suggest that the most critical parameters with respect to the Gibbs-Wilbraham phenomenon are strong anisotropies and extremely large width-to-length ratios).

The system of equations that we solve includes hyperbolic cosine terms (see Equations 5.51, 5.53 and 5.54). As the hyperbolic cosine function grows rapidly with its argument, it can be useful to reformulate the solution to avoid terms that grow exponentially with n. This is possible by redefining the series coefficients and scaling them with a hyperbolic cosine term:

$$\alpha_n^* = \alpha_n \cdot \cosh\Big(\frac{n\pi}{L}\kappa w_{\max}\Big), \tag{5.55}$$

in which α_n^* in L are the modified series coefficients and $w_{\max}$ is the maximum width. With the two identities of hyperbolic fractions:

$$\frac{\sinh(t)}{\cosh(b)} = \frac{\exp(t-b) - \exp(-t-b)}{1 + \exp(-2b)}, \tag{5.56}$$

$$\frac{\cosh(t)}{\cosh(b)} = \frac{\exp(t-b) + \exp(-t-b)}{1 + \exp(-2b)}, \tag{5.57}$$

the resulting ratios of hyperbolic sine and cosine functions can be reformulated to contain only terms with a negative sign in the exponential (for our case). Such a redefinition of coefficients results in modified versions of the hydraulic head and stream function.

The modified versions of the hydraulic head and stream function solution are:

$$h(x,y) = h_1 + \frac{h_2 - h_1}{L} \cdot x + \sum_{n=1}^{\infty} \alpha_n^* \sin(cx) \frac{\exp(c\kappa(y - w_{\max})) - \exp(-c\kappa(y + w_{\max}))}{1 + \exp(-2c\kappa w_{\max})}, \tag{5.58}$$

and

$$\Psi(x,y) = -T_x \cdot \left(\alpha_0^* + \frac{h_2 - h_1}{L} \cdot y + \frac{1}{\kappa} \sum_{n=1}^{\infty} \alpha_n^* \cos(cx) \frac{\exp(c\kappa(y - w_{\max})) + \exp(-c\kappa(y + w_{\max}))}{1 + \exp(-2c\kappa w_{\max})} \right), \tag{5.59}$$

where we make use of a new shorthand $c = \frac{n\pi}{L}$. A similar procedure has been exploited by Powers (1966) and Powers et al. (1967). If there was an ideal computer with infinite precision the solutions of the two formulations were identical, and the only difference would lie in the magnitude of the series coefficients. Due to round-off errors, however, the modified version yields more precise results when using standard double precision floating-point operations.

5.3 Characterization of the Hyporheic-Exchange Zone

Figure 8 shows semi-analytical flow nets for different model configurations, as well as directions and magnitudes of the flux perpendicular to the southern boundary. This flux is zero at the western end ($x = 0$). Depending on the geometric configuration and the northern influx rate, it typically increases with x until it reaches a maximum. From there on, it decreases again, passes zero and becomes negative until it reaches a minimum. Finally, it increases again to reach zero once more at the eastern end ($x = L$). In summary, we can observe in the western part a net flux from the river to the aquifer and a reversed flux in the eastern part. Without a northern influx the pattern is symmetric and the net exchange between the river and the aquifer is zero. By following the streamlines, we can identify those parts of the domain, where the flow originates from the river and returns to it again. These parts define the valley-scale hyporheic exchange zone (blue areas in Figure 8). In the following we derive how to quantify the flux of water flowing through this zone (the hyporheic exchange flux Q_{ex}), how large the hyporheic exchange zone is (the exchange zone area A_{ex}) and how travel times through this zone are distributed.

5.3.1 Exchange Flux

As exchange flux, we define the discharge (in $\mathsf{L}^3\,\mathsf{T}^{-1}$) of water originating from the river at the southern boundary and returning to it again. We can quantify the exchange flux by considering the stream function, which is defined such that the net discharge Q_{pq} crossing a line between two points p and q equals the absolute difference of the stream-function values at the two end points:

$$Q_{pq} = |\Psi(x_q, y_q) - \Psi(x_p, y_p)|. \tag{5.60}$$

In Figure 8 we can see different manifestations of the exchange zone. It can span the entire southern boundary (see Figure 8a-c), parts of it (see Figure 8d) or may vanish completely (e.g., if the northern influx is very large). In any case, however, it is bounded by a dividing streamline that starts and/or ends at one of the southern corner points of the domain (i.e., either (0|0) or (L|0)). According to Equation 5.8, an increase in the stream-function value along the southern boundary in the positive x-direction implies an exfiltration flux (groundwater discharge to the river). Thus, in the cases shown in Figure 8, the stream-function first decreases along the southern boundary, reaches a minimum and increases back until it reaches its initial value. In the case of Figure 8d it further increases, because the river gains more water than it loses. In this setting, the exchange flux is given by the stream-function values at (0|0) and at the minimum point. With different geometries (e.g., a valley that first gets narrower and then widens) the southern boundary may start with exfiltrating conditions (river gaining groundwater). In that case the exchange flux is given by the difference of the stream-function values at the end point (L|0) and at the minimum. All cases are covered by:

$$Q_{\mathrm{ex}} = \min[\Psi(0,0), \Psi(L,0)] - \min[\Psi(x,0)]. \tag{5.61}$$

If there is no hyporheic-exchange zone, the end point with the smaller stream-function value is identical with the point of minimal stream-function value, leading to the correct exchange flux of zero. In general, the minimum of the stream function along the southern boundary $\min(\Psi(x,0))$ must be evaluated numerically. If the number M of points to determine the coefficients $\boldsymbol{\alpha}$ was large enough, it is convenient to re-use the same set of x-locations that were also selected for constructing the system of equations.

5.3.2 Area of the Exchange Zone

The volume V_{ex} of the exchange zone can be expressed in terms of a two-dimensional area:

$$A_{\mathrm{ex}} = \frac{V_{\mathrm{ex}}}{\Phi}, \tag{5.62}$$

where we make use of the depth-integrated effective porosity Φ. Just as for the calculation of the exchange flux, a general analytical solution for the area of the exchange zone is not available. Hence, we determine the area numerically by constructing a polygon bounded by the southern domain border and the dividing streamline separating the exchange zone from the northern rest of the domain. The dividing streamline is the contour line of Ψ representing a value of $\min[\Psi(0,0), \Psi(L,0)]$. It can be constructed with standard contouring algorithms from a set of point observations of $\Psi(x, y)$ placed throughout the domain. The evaluation of the polygonal area is trivial; we only need to make sure that we use enough points to approximate the polygon well enough. Again, it might be convenient to reuse the set of x-nodes in combination with a set of y-nodes for the construction of a mesh of points where Ψ is evaluated.

5.3.3 Travel Time Distribution

In order to evaluate the full travel-time distribution of all water parcels flowing through the exchange zone, we construct n_t contour lines of Ψ across the full range of Ψ-values within the exchange zone (i.e., from $\min[\Psi(x,0)]$ to $\min[\Psi(0,0),\Psi(L,0)]$). By choosing equal steps between the contour line values (i.e., a constant $\Delta\Psi$), we construct stream-tubes of identical discharge. For each contour line we determine the respective travel time t in T and the fraction $F(t)$ of discharge that has a travel time smaller than t. This gives one point of the travel-time distribution per contour line.

The travel time t_i of the i-th contour line can be approximated by summation of the travel times t_seg of all its n_seg line segments. A segment's travel time can be determined from its length L_seg in L and its average flow velocity $\boldsymbol{v}_\mathrm{avg}$ in $\mathsf{L}\,\mathsf{T}^{-1}$. We assume that the velocities of the segment endpoints ($\boldsymbol{v}_1$ and $\boldsymbol{v}_2$) apply both each for half of the segment length, which results in an arithmetic average:

$$t_\mathrm{i} = \sum_{j=1}^{n_\mathrm{seg}} t_\mathrm{seg} = \sum_{j=1}^{n_\mathrm{seg}} \frac{L_\mathrm{seg}}{\boldsymbol{v}_\mathrm{avg}} = \sum_{j=1}^{n_\mathrm{seg}} \frac{2 \cdot L_\mathrm{seg}}{|\boldsymbol{v}_1| + |\boldsymbol{v}_2|}. \tag{5.63}$$

The linear velocity $\boldsymbol{v}$ is given by Darcy's law (Darcy, 1856) and the depth-integrated porosity Φ:

$$\boldsymbol{v} = -\frac{1}{\Phi}\boldsymbol{T}\nabla h. \tag{5.64}$$

The corresponding fraction of hyporheic discharge that has a travel time smaller than t_i can be determined with the stream function:

$$F(t_i) = \frac{\Psi_i - \min(\Psi(x,0))}{Q_\mathrm{ex}}, \tag{5.65}$$

where Ψ_i is the stream function value of the i-th contour line. The n_t points $(t_i|F(t_i))$ describe the approximated cumulative distribution function of travel time. For a better comparison between different settings, we create dimensionless travel times by normalizing with the mean travel time:

$$\tilde{t} = \frac{t}{t_\mathrm{mean}}, \tag{5.66}$$

$$t_\mathrm{mean} = \frac{V_\mathrm{ex}}{Q_\mathrm{ex}} = \frac{\Phi A_\mathrm{ex}}{Q_\mathrm{ex}}. \tag{5.67}$$

5.3.4 Summary of the Semi-Analytical Procedure

In short, the semi-analytical determination of hyporheic-exchange-zone properties for a given setting involves the following steps:

1. Define all parameter values (including those describing the geometry).
2. Decide on N, M and x_i.
3. Set up the system of equations (Equations 5.53 and 5.54).

4. Solve the system of equations to obtain the coefficients $\boldsymbol{\alpha}$.

5. Evaluate the stream function on selected points to determine Q_{ex}, A_{ex} or $F(t)$ according to Equations 5.61, 5.62 and 5.65.

Depending on N and M, this scheme can be computationally expensive. Also, it gives no direct evidence of how the various input parameters affect the exchange flux, the area of the exchange zone, or the travel-time distribution. As this is our primary interest, we perform a systematic parameter-variation study setting the base for easy-to-use empirical relationships.

6 Relating Exchange-Zone Metrics to Hydrogeological and Geometric Properties of the Floodplain Aquifer

In this section we investigate how the various hydrogeological and geometric properties of the floodplain aquifer affect the hyporheic-exchange flux, the hyporheic-zone area and the travel-time distribution within the hyporheic exchange zone.

6.1 Sensitivity Analysis

First, we perform a global sensitivity analysis to investigate how strongly the exchange flux Q_{ex} and the exchange zone area A_{ex} depend on the input parameters. As our semi-analytical solution is reasonably fast, we are able to perform a full variance-based sensitivity analysis. It is typically known as the method of Sobol indices, of which we give a brief summary in the following (for detailed explanations we refer to Sobol′, 1993; Sobol′, 2001).

A variance-based sensitivity analysis aims to quantify what fraction of the variance in the observed variable originates from the variance of individual input variables or combinations thereof. Two metrics can be defined: first-order Sobol indices, which express the effects of individual variables by themselves, and total effect Sobol indices, which express the effects of individual variables and all their combinations with other variables. We determine the Sobol indices for Q_{ex} and A_{ex} with a Monte Carlo approach. Figure 9 shows the outcomes of this investigation for the three different boundary shapes “cosinusoidal”, “bump” and “composite”. We can observe the following:

- For the hyporheic exchange flux, the average hydraulic transmissivity is the dominant individual parameter. This is probably related to the fact that the transmissivities vary over orders of magnitude and are a direct linear factor to all flux calculations.
- Considering combined effects, the Q_{ex} Sobol indices of all parameters are more evenly distributed. Still, the hydraulic transmissivity reaches the highest values, but we can also observe a large effect of the northern influx and the ratio of minimum to maximum domain width.

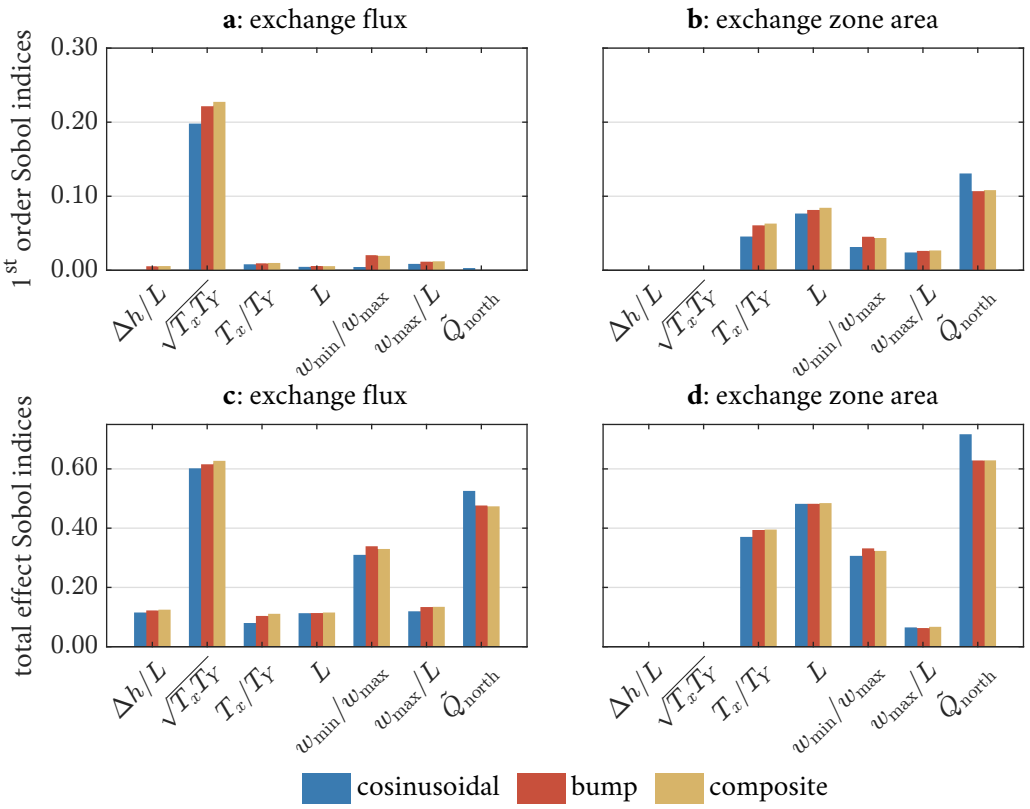

Figure 9: Results of the global variance-based sensitivity analysis for Q_{ex} (**a**, **c**) and A_{ex} (**b**, **d**) in terms of first-order Sobol indices (**a**, **b**) and total effect Sobol indices (**c**, **d**).

- For the area of the exchange zone, the first-order indices and total indices show a very similar pattern. This means combined effects of parameter combinations seem to have a less strong effect compared to the hyporheic exchange flux.
- Two input parameters don't influence the exchange zone at all, as they have Sobol indices of basically zero. These two parameters are the ambient hydraulic gradient and the average hydraulic transmissivity. This observation agrees perfectly with intuition, as these two parameters do not affect the shape of the flow field, but only the magnitudes of h and Ψ.
- The differences between the shape types are not very large.

Overall, however, we believe these results are comparably hard to interpret, because the Sobol indices show average linear effects across the entire parameter space. We therefore continue our investigation by constructing proxy-expressions relating model input to output.

6.2 Exchange Flux

We focus on the "cosinusoidal" shape to infer relationships between the input parameters and Q_{ex} and use the "bump" and "composite" shapes for verifying the derived expressions.

Table 1: Constant hydraulic parameter values for analyzing the effects of geometric properties on hyporheic-exchange metrics.

Parameter	Symbol	Value	Unit
fixed head at inlet	h_1	3	m
fixed head at outlet	h_2	0	m
transmissivity in x	T_x	10^{-3}	$m^2\ s^{-1}$
transmissivity in y	T_y	10^{-3}	$m^2\ s^{-1}$
northern influx	q_{north}	0	$m^2\ s^{-1}$

Cosinusoidal Shape To understand how the hyporheic-exchange flux relates to the domain geometry for a given hydraulic setup, we fix all hydraulic parameters to the values defined in Table 1 and vary the domain length L between 500 m and 3500 m, the minimum width w_{min} between 50 m and 350 m, and the ratio of maximum to minimum width $w_{\text{max}}/w_{\text{min}}$ between 1.0 and 2.5. With this initial sample, we can resemble most of the floodplain examples mentioned previously, which range from a few hundred meters to a few kilometers in length and have approximate w_{max}/L ratios between 10 % and 50 % (e.g., Clément et al., 2003; Ó Dochartaigh et al., 2019) and approximate $w_{\text{max}}/w_{\text{min}}$ ratios of up to 2.5 (e.g., Castro and Hornberger, 1991; Helton et al., 2012).

Figure 10a shows a three-dimensional slice and scatter plot of corresponding exchange fluxes for the $13 \times 7 \times 7 = 637$ parameter combinations that we tested. This plot reveals the following insights: (1) For a width ratio of one, the exchange flux is zero, independent of the other two parameters. This is obvious, because the domain collapses to a rectangle, where the solution of the related system of equations is trivial (α_1 to α_N equal zero) and there is no driving force for any lateral hyporheic exchange. (2) For a fixed length and minimum width, the exchange flux increases monotonically with the ratio of w_{max} to w_{min}. (3) For any fixed width-ratio larger than one, the magnitude of the exchange flux depends on both the domain length and the minimum width. (4) There seems to be a ratio of the latter two that gives a maximum flux. (5) For ratios larger or smaller than that, the exchange flux declines.

This rather complex behavior in three-dimensional parameter space can be simplified and the dimensionality can be reduced. Figure 10b shows that all simulated points can be brought very close to a single curve by plotting the following two transformed quantities against each other: On the horizontal axis we display the ratio of the average width to the domain length. For that we calculate the average domain width by dividing the domain's total area by its length:

$$w_{\text{mean}} = \frac{A_{\text{tot}}}{L}. \tag{6.1}$$

On the vertical axis of the transformed plot, we show the product of the exchange flux and the ratio of domain length to the transmissivity in x-direction and maximum width-difference Δw, which is:

$$\Delta w = w_{\text{max}} - w_{\text{min}}. \tag{6.2}$$

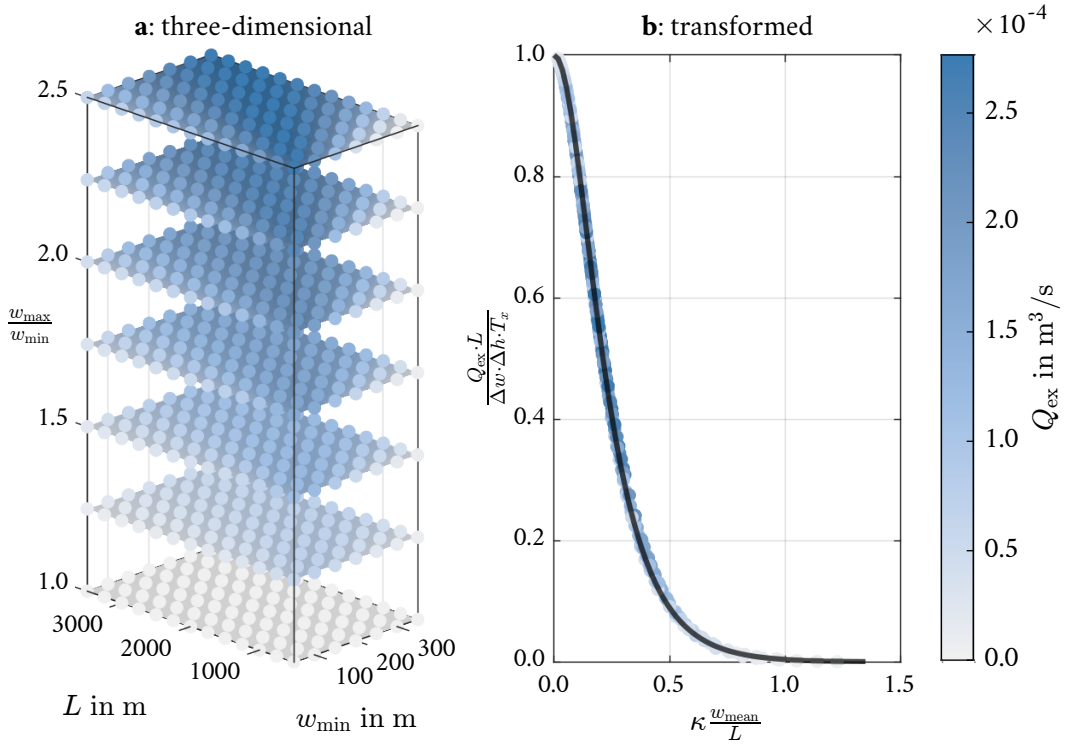

Figure 10: Response of the exchange flux to geometric parameters. **a**: Scatter plot and sliced interpolation involving geometric parameters. **b**: A transformation brings the response of the exchange flux to all three geometric parameters onto a single curve that can be approximated by a scaled hyperbolic secant function (black line).

After subsequent variations of the hydraulic properties h_1, h_2 and κ for fixed geometries (not shown), we found an easy way to incorporate their influence on the exchange flux into the existing transformation as long as there is no northern hillslope influx (i.e., $q_{\text{north}} = 0$): The square root of the anisotropy ratio (κ) affects only the scaling in the direction of the horizontal axis and needs to be multiplied with the ratio of the average width to the domain length. The hydraulic heads h_1 and h_2 only matter as a head difference:

$$\Delta h = h_1 - h_2. \tag{6.3}$$

This difference only affects the vertical axis of the plot and can be accounted for in dividing the existing expression by Δh.

The one-dimensional relationship derived in this way (see Figure 10b) starts at the point (0|1) and is characterized by a monotonic decline that has a small slope in the beginning, a steeper part around 0.25 and a tail asymptotically approaching zero. Such a behavior can be approximated by a hyperbolic secant function scaled in the direction of the horizontal axis (Figure 10b contains such a function fitted to the simulated model results). This relationship can be inverted to a description of the dimensional hyporheic-exchange flux depending on all model parameters (with exception of q_{north}, which has been set to zero so far). As a result, we postulate the approximation:

$$Q_{\text{ex}} \approx Q_0 \cdot \operatorname{sech}\left(a_1 \cdot \kappa \frac{w_{\text{mean}}}{L}\right), \tag{6.4}$$

with a dimensionless fitting parameter a_1 and the reference discharge Q_0 in $\mathsf{L}^3\,\mathsf{T}^{-1}$:

$$Q_0 = I_x \cdot T_x \cdot \Delta w = \frac{h_1 - h_2}{L} \cdot T_x \cdot (w_{\max} - w_{\min}), \tag{6.5}$$

where I_x is the dimensionless ambient hydraulic gradient in x-direction, which is given by the ratio of the head difference Δh to the domain length L.

We can interpret the relationship of Equation 6.4 in the following way: (1) The difference between the maximum and minimum width of the valley exerts a linear control on the hyporheic exchange flux. (2) For a given width difference, there is a maximum potential exchange flux Q_0. (3) The magnitude of this flux linearly depends on the ambient hydraulic gradient and the hydraulic transmissivity in x-direction. (4) The actual exchange flux Q_{ex} is smaller than Q_0 and the ratio of the two is determined by the domain aspect ratio (w_{mean}/L) and the anisotropy ($\kappa = \sqrt{T_x/T_y}$). We explain the details of this relationship in the following.

To understand the equation for Q_0, it makes sense to analyze the illustrative extreme case of $w_{\min}$ approaching zero in a scenario of constant Δw. As seen in Figure 7, in such a case nearly the entire domain belongs to the northern part A_{north}, and A_{rect} becomes small. Basically all water flowing through the domain must come exclusively from the river and return to it again (i.e., $A_{\text{tot}} = A_{\text{north}} = A_{\text{ex}}$). The easiest way to quantify the exchange flux for this specific scenario is to take a look at the widest part of the domain (i.e., $x = L/2$). Here, the discharge Q^* through the domain, is given by Darcy's law with the following approximation:

$$Q^* = \Delta w \left(-T_x \frac{\partial h}{\partial x} \right) \approx \Delta w \cdot T_x \frac{h_1 - h_2}{L}, \tag{6.6}$$

where we assume a uniform hydraulic gradient in x-direction equaling $-\Delta h$ over L. T_x is the only transmissivity that matters here, because the hydraulic gradient in the y-direction is zero. It becomes clear now, that Q^* equals our reference discharge Q_0 and it represents the maximum discharge that could be reached for a given Δw and adjustable $w_{\min}$.

The actual exchange flux can be smaller than Q_0 for two reasons: firstly the curvature of the domain/flow-paths and secondly the distance between northern and southern boundary. Both of these effects are summarized in the term w_{mean}/L within the hyperbolic secant function in the following way: Even in cases with $w_{\min} \to 0$, the actual exchange flux is smaller than Q_0, because of the curvature of the domain's northern boundary. This curvature leads to flow-path lengths larger than L. Consequently, the hydraulic gradient $\partial h/\partial x$ is smaller than I_x at $x = L/2$. For given domain widths, L serves as a control to change the curvature, with $L \to \infty$ resulting in less curved flow-paths yielding $Q_{\text{ex}} \to Q_0$. In addition to that, the deviation of Q_{ex} from Q_0 depends on the average domain width w_{mean}, which is a measure for how far the northern boundary (i.e., the driving force) is separated from the river. As the hyperbolic secant function decreases strictly monotonically, a larger separation distance always leads to a decrease of hyporheic exchange flux.

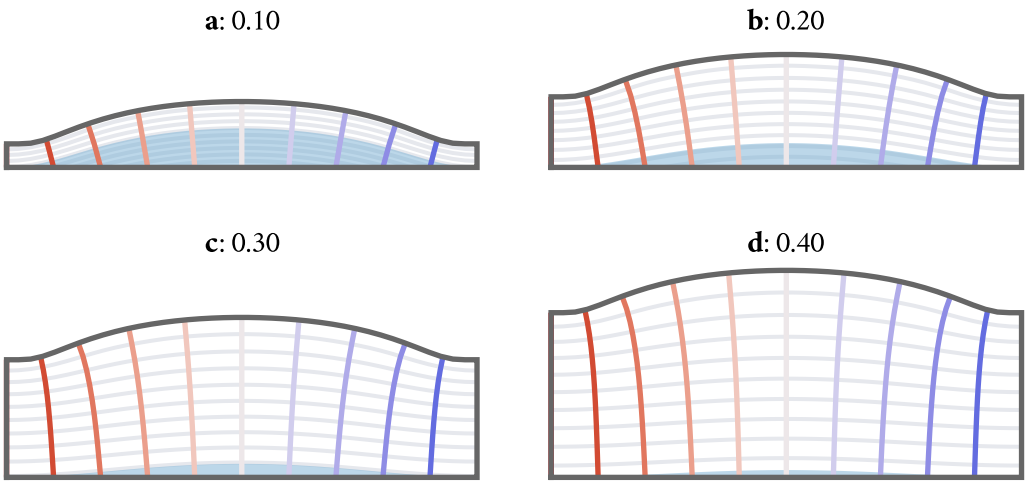

Figure 11: Examples of flow nets showing how increasing separation (from **a** to **d**) between the northern area and the southern boundary shrinks the hyporheic exchange zone. All model parameters with exception of w_{min} and w_{max} are identical in the four cases. The difference of w_{min} and w_{max} was kept constant too. The number in the title reflects the average domain aspect ratio w_{mean}/L.

This is in line with intuition: if the northern valley expansion is far away from the river, the water necessary to "fill" A_{north} can be drawn from the western boundary without affecting the river too much. For illustration of this behavior, Figure 11 contains four model results that show how a larger separation distance of the northern area A_{north} from the southern boundary leads to a reduction of hyporheic exchange.

The hydraulic anisotropy acts as a scaling factor for the average aspect ratio w_{mean}/L of the domain, meaning that if $T_y > T_x$, it becomes easier to draw water from the river, having the same effect as moving the northern boundary closer to it. Vice versa, a case of $T_x > T_y$ can be interpreted as increasing the distance between river and northern boundary. As κ represents the square root of the anisotropy ratio T_x/T_y, we can summarize the anisotropy-corrected aspect ratio to a new dimensionless variable:

$$\tilde{x} = \kappa \frac{w_{\text{mean}}}{L}, \tag{6.7}$$

which can be used to construct a dimensionless formulation for the exchange flux $\tilde{Q}_{\text{ex}}$:

$$\tilde{Q}_{\text{ex}} = \frac{Q_{\text{ex}}}{Q_0} \approx \operatorname{sech}(a_1 \cdot \tilde{x})\,. \tag{6.8}$$

In the following, we want to analyze the dependence of the hyporheic exchange flux on the northern hillslope influx q_{north} to generalize Equation 6.8. Towards this end, we normalize the total northern discharge by Q_0, to obtain the dimensionless quantity $\tilde{Q}_{\text{north}}$:

$$\tilde{Q}_{\text{north}} = \frac{\int_0^L q_{\text{north}}\, \mathrm{d}x}{Q_0} = \frac{q_{\text{north}}}{T_x} \cdot \frac{L^2}{h_1 - h_2} \cdot \frac{1}{w_{\text{max}} - w_{\text{min}}}. \tag{6.9}$$

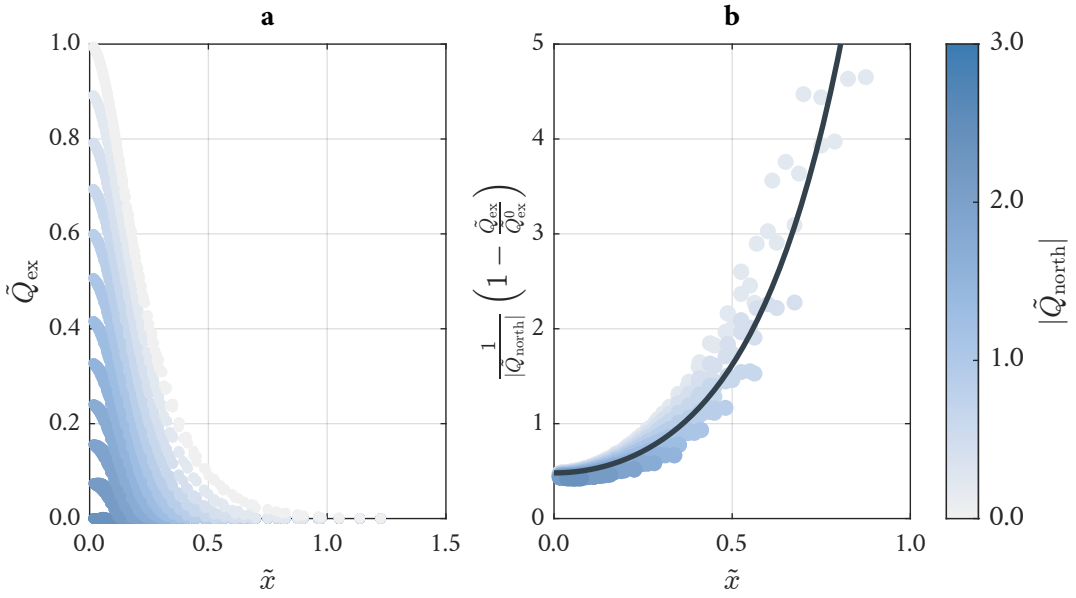

Figure 12: Dependency of $\tilde{Q}_{ex}$ on $\tilde{Q}_{north}$ across different model realizations. $\tilde{Q}^0_{ex}$ represents $\tilde{Q}_{ex}$ for $\tilde{x} = 0$. **a**: Comparison of cases with different $|\tilde{Q}_{north}|$. **b**: Transformation to a common relationship.

An influx from the northern boundary leads to a reduction of the hyporheic exchange flux, because the hyporheic exchange zone is "pushed" towards the southern river boundary. At sufficiently large values of $|\tilde{Q}_{north}|$, the exchange zone may even vanish completely, meaning that no river-water infiltration takes place.

Figure 12a shows the results of a systematic parameter variation study, where $\tilde{Q}_{north}$ varies between -3 and 0 for a revision of the study conducted for Figure 10. It is obvious that $\tilde{Q}_{ex}$ decreases with increasing $\tilde{Q}_{north}$ for a given $\tilde{x}$. The exact relationship between the two is not linear. Figure 12b shows one way of approximating this relationship. The equation of the regression model (shown as a black line) is given by:

$$\tilde{Q}_{ex} \approx \operatorname{sech}(a_1\tilde{x}) \cdot \max\left[0, 1 - a_2 \cdot |\tilde{Q}_{north}| \cosh(a_3\tilde{x})\right]. \tag{6.10}$$

We chose this model because it only needs two additional parameters and has a slope of 0 at $\tilde{x} = 0$ for the transformed quantity, which is close to the simulated data points (see Figure 12). There still exists a spread of the data points after the nonlinear transformation, but especially towards small values of $\tilde{x}$ this model works decently well. As already mentioned, this relationship involves a second empirical function with two coefficients a_2 and a_3, employing the hyperbolic cosine function. The implementation of the maximum-value function constructs a threshold leading to a constant exchange flux of $\tilde{Q}_{ex} = 0$ in cases of large values of $|\tilde{Q}|_{north}$. This threshold can also be used to give a simple approximate logical expression indicating if the exchange zone is present for a given case of $\tilde{Q}_{north}$ and $\tilde{x}$:

$$1 > a_2|\tilde{Q}_{north}| \cosh(a_3\tilde{x}). \tag{6.11}$$

Table 2: Ranges of geometric and hydraulic parameters that were explored in the stochastic simulation to obtain shape-dependent empirical coefficients.

Parameter	Symbol	Minimum	Maximum	Unit
length	L	100.0	3000.0	m
gradient	I_x	0.0	3.0	%
length ratio	$w_{\max}/L$	0.1	0.5	–
width ratio	$w_{\max}/w_{\min}$	0.4	1.0	–
log. transmissivity	$\log_{10}\left(\sqrt{T_x T_y}\right)$	−6.0	−2.3	T in $\mathrm{m^2\,s^{-1}}$
log. anisotropy	$\log_{10}(T_x/T_y)$	−1.0	1.0	–
northern influx	$\tilde{Q}_{\mathrm{north}}$	−3.0	0.0	–

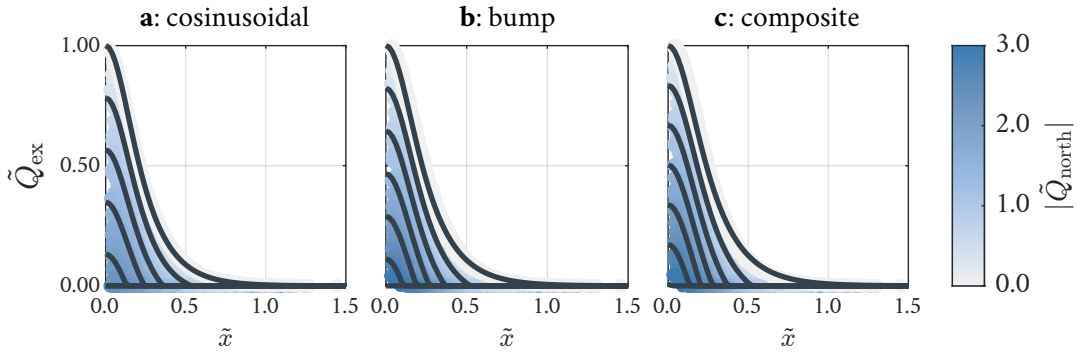

Figure 13: Dimensionless exchange flux results for different values of $|\tilde{Q}_{\mathrm{north}}|$ (0.0, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0) and three domain shapes (**a**-**c**). Fitted proxy-models are shown as black lines.

In summary, our proxy-expression for the estimation of the hyporheic-exchange flux has a single empirical coefficient in case of zero northern influx, and two additional coefficients in cases with non-zero northern influx. We will determine and compare these coefficients in the following for different shapes of the floodplain.

Different Shapes We construct a sample of 1500 quasi-random model realizations to test whether the derived proxy-equation also works for shapes other than the "cosinusoidal" one. The parameter values of all realizations are drawn from uniform probability distributions within the ranges documented in Table 2, in which the sampling of parameter sets is done with a scrambled Halton sequence (Halton, 1960; Kocis and Whiten, 1997; Cheng and Druzdzel, 2013).

For each realization (i.e., parameter combination) we solve the semi-analytical model six times, twice for each of the three shapes ("cosinusoidal", "bump" and "composite"), with and without influx from the northern hillslope. This allows determining the empirical fitting coefficient a_1 independently of a_2 and a_3. Figure 13 shows the resulting normalized hyporheic-exchange flux for this stochastic model sample. This figure could also be used as a tool to determine $\tilde{Q}_{\mathrm{ex}}$ from $\tilde{x}$ and $|\tilde{Q}_{\mathrm{north}}|$ graphically. Table 3 lists the fitted empirical coefficients and metrics indicating the quality of the fits.

Table 3: Fitted proxy-model coefficients and quality of fit metrics for the sample of 1500 model realizations per domain shape. Coefficient values are shown as fit ± uncertainty, where the uncertainty is obtained through linearized uncertainty propagation.

Shape	a_1	a_2	a_3	RMSE_0	$\mathrm{RMSE}_{\neq 0}$
"cosinusoidal"	6.242 ± 0.002	0.434 ± 0.001	4.121 ± 0.024	0.005	0.014
"bump"	5.852 ± 0.009	0.355 ± 0.001	4.607 ± 0.037	0.023	0.025
"composite"	5.515 ± 0.010	0.331 ± 0.001	4.755 ± 0.026	0.027	0.019

The results are similar for all three shapes and closely follow the hyperbolic secant curve in the cases without northern influx. Considerable differences between the shapes occur only for large northern influxes, but are mostly restricted to how quickly the associated exchange flux drops. The "bump" and "composite" shape are very similar to each other and exhibit a slower decrease of $\tilde{Q}_{\mathrm{ex}}$ with increasing $\tilde{Q}_{\mathrm{north}}$ compared to the "cosinusoidal" shape. The qualitative behavior, however, is identical for all domain shapes. This is also reflected in the fitted coefficients a_1, a_2, and a_3, which differ only slightly between the shapes.

We quantify the quality of the proxy-equation with the Root Mean Square Error (RMSE) of $\tilde{Q}_{\mathrm{ex}}$ when comparing the proxy-equation to the semi-analytical solution. We do this independently for the cases in which the northern influx is zero ("RMSE_0"), and those with a non-zero northern influx ("$\mathrm{RMSE}_{\neq 0}$"). Given that $\tilde{Q}_{\mathrm{ex}}$ ranges between zero and one, the RMSE-values of up to 0.027 indicate a very good agreement between the proxy-equation and semi-analytical solution. This reveals that our interpretations are valid across the different domain shapes. Finally, we perform a linearized uncertainty propagation to estimate the uncertainties of the fitted coefficient values. The resulting relative uncertainties (shown in Table 3) are on the order of a few percent or less, indicating a high confidence in the fitted coefficients.

6.3 Area of the Exchange Zone

For each realization of the model sample, we determine the area of the hyporheic-exchange zone. In this section, we construct a simplified proxy-equation relating this area to the model input parameters. To do so, we define a dimensionless area $\tilde{A}$, by normalizing A_{ex} with A_{north} for all realizations. This dimensionless area seems to be approximately proportional to $\tilde{Q}_{\mathrm{ex}}$, and additionally depends on $|\tilde{Q}_{\mathrm{north}}|$ in the following nonlinear way:

$$\tilde{A} = \frac{A_{\mathrm{ex}}}{A_{\mathrm{north}}} \approx \frac{\tilde{Q}_{\mathrm{ex}}}{\sqrt{1 + |\tilde{Q}_{\mathrm{north}}|}}. \tag{6.12}$$

Figure 14 shows that the 3000 simulations decently follow this curve for all three domain shapes, and Table 4 summarizes the respective quality of the fits. Again, the RMSE values ($\mathrm{RMSE}_{\mathrm{area}}$) of up to 0.028 are small compared to the range of observed values (zero to one).

Table 4: Quality of the empirical-fit metrics for the area of the exchange zone. The results are obtained for the sample of 3000 model realizations per domain shape.

Shape	$\mathrm{RMSE}_{\mathrm{area}}$
"cosinusoidal"	0.017
"bump"	0.024
"composite"	0.028

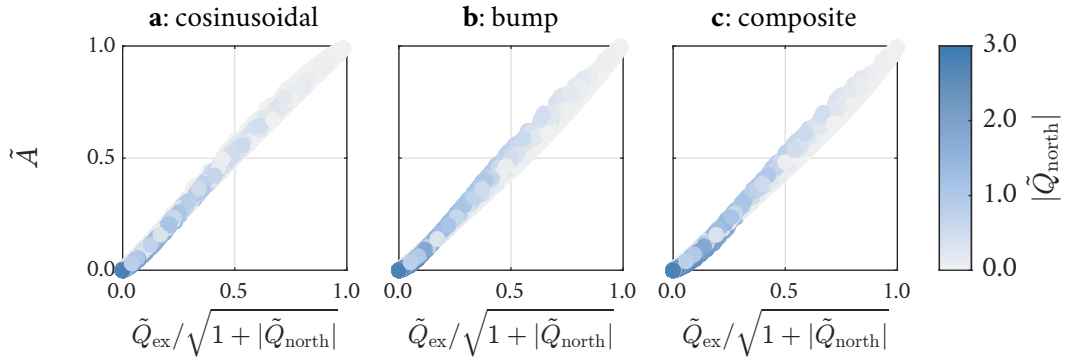

Figure 14: Nearly proportional relationship between the normalized exchange-zone area and the dimensionless discharge for the three domain shapes (**a**-**c**).

The relationship of Equation 6.12 suggests that in cases without a northern influx, the area of the hyporheic exchange zone closely scales with the normalized exchange flux and A_{north}. This implies that pronounced widening (i.e., a large width difference $w_{\mathrm{max}} - w_{\mathrm{min}}$) leads to a larger exchange zone, but a greater ambient hydraulic gradient (I_x) would not have any effect on A_{ex}.

As intuitively expected, an influx on the northern hillslope boundary leads to a reduction of the exchange area. As outlined above, the influx from the northern hillslope leads to a reduction of the exchange flux, but the reduction of the area of the exchange zone is even bigger. One interpretation of that could be that an increase of $|\tilde{Q}_{\mathrm{north}}|$ does not only shrink the exchange zone in the y-direction, but also reduces its extent along the southern boundary. For cases with $\tilde{Q}_{\mathrm{north}} = 0$, on the other hand, the southern boundary of the exchange zone always covers the entire river and the occupied area varies only in the y-direction.

6.4 Hyporheic Travel Times

For all simulated model scenarios, we calculate the distribution of normalized travel times as outlined above. Figure 15a-c show the resulting cumulative distribution functions for the three different domain shapes in the cases without northern influx. Figure 15g-i contain the twin versions of these plots, but show all model runs that had a northern influx ($\tilde{Q}_{\mathrm{north}} \neq 0$). These graphs show similar results and are subject to the same interpretations as done in the following.

We plotted the 5^{th}, 50^{th} and 95^{th} percentiles to highlight the spread across the collection of curves. It becomes clear that this spread is comparably small, and the travel-time distributions are similar, both

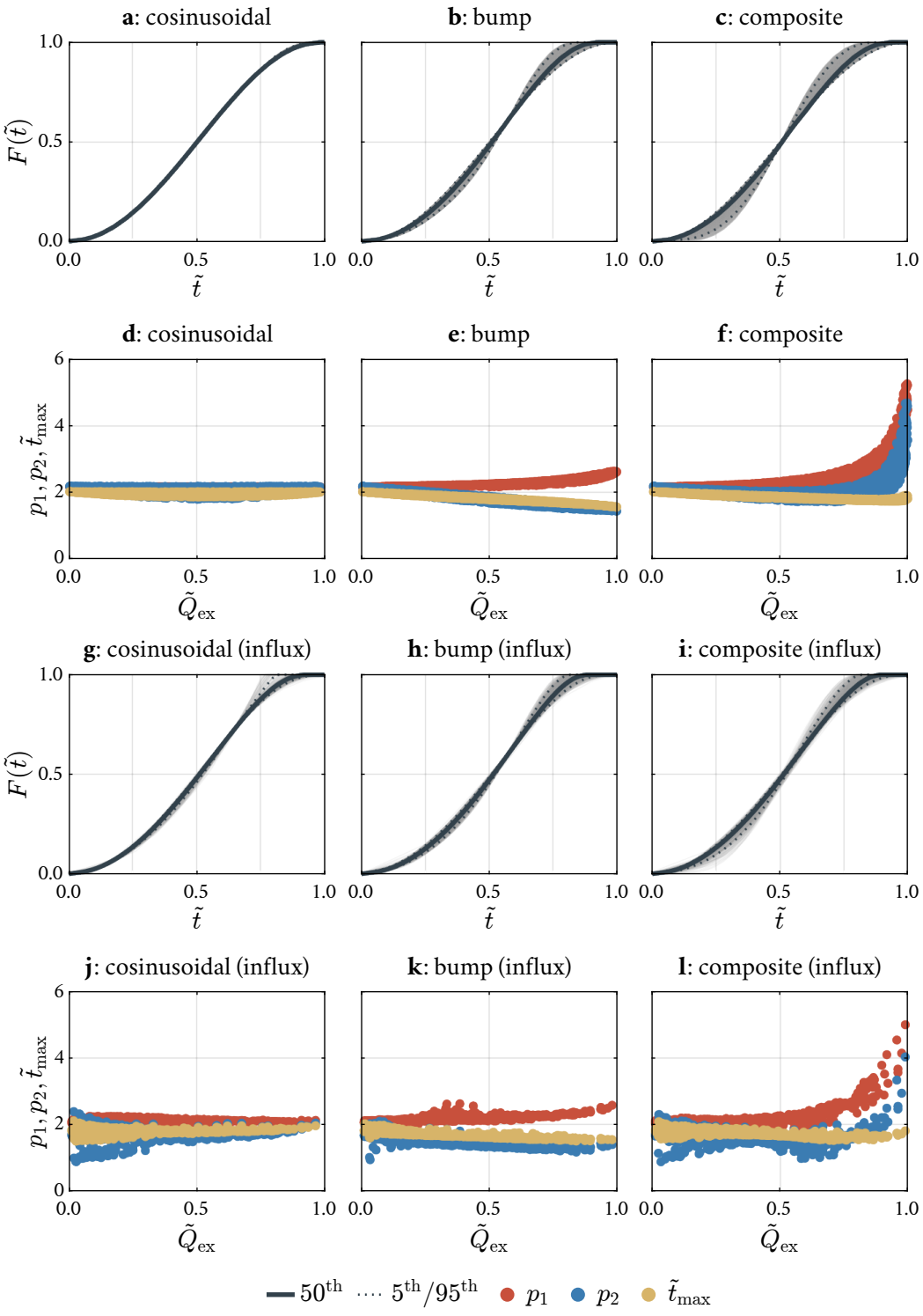

Figure 15: Travel-time distributions (**a**-**c**, **g**-**i**) and coefficients of fitted beta distributions (**d**-**f**, **j**-**l**) for the three shapes. The top figures (**a**-**f**) correspond to cases without northern influx, the bottom figures (**g**-**l**) are the twin versions for cases with northern influx.

for all distributions of a specific shape, and between the three shapes. All curves show a sigmoidal behavior starting at a travel time of zero, approximately passing 50 % cumulative probability at the mean travel time (determined from the area of and discharge through the exchange zone), and finally reaching 100 % probability at about twice the mean travel time.

That the cumulative distribution function of travel times start at the origin is intuitively clear: an infinitesimally small travel time exists at the transition point along the river between losing and gaining conditions. The fact that all distributions nearly pass 50 % at the mean travel time indicates that the median and mean travel times are almost identical. It is obvious that there must be a finite maximum travel time, which corresponds to the time that the water needs to travel along the bounding stream-line of the hyporheic exchange zone, which separates it from the remaining aquifer. This maximum travel time is about twice as large as t_{mean}. Altogether, this results in a symmetric travel-time distribution, which qualitatively differs from skewed travel-time distributions occurring in systems with stagnation points.

All curves shown in Figure 15 closely resemble beta distributions that are scaled in $\tilde{t}$-direction. Hence, we chose to fit the cumulative distribution functions with the beta distribution, in which the dimensionless time $\tilde{t}$ is scaled by the maximum time $\tilde{t}_{\mathrm{max}}$:

$$F(\tilde{t}) = \mathcal{I}_{\tilde{t}/\tilde{t}_{\mathrm{max}}}(p_1, p_2), \tag{6.13}$$

where $\mathcal{I}_{\tilde{t}/\tilde{t}_{\mathrm{max}}}(p_1, p_2)$ is the regularized incomplete beta function and p_1 and p_2 are two shape parameters.

As a result of the fit, we obtain one set of values of p_1, p_2, and $\tilde{t}_{\mathrm{max}}$ for each simulation. Figure 15d-f (and by extension to cases with northern influx Figure 15j-l) shows how these three fitted parameters depend on the dimensionless exchange flux $\tilde{Q}_{\mathrm{ex}}$. Comparing the results of the three different shapes, it is remarkable that all shapes have the same parameter values for small dimensionless exchange fluxes, namely $p_1 \approx p_2 \approx \tilde{t}_{\mathrm{max}} \approx 2$, implying a symmetric distribution with $t_{\mathrm{max}} \approx 2t_{\mathrm{mean}}$ and a standard deviation of $t_{\mathrm{mean}}/\sqrt{5}$. This is probably related to the fact that small values of $\tilde{Q}_{\mathrm{ex}}$ indicate a large separation between the river and the northern boundary, where the specific shape of the boundary loses importance due to the diffusive nature of the groundwater flow equation. With increasing values of $\tilde{Q}_{\mathrm{ex}}$ the fitted parameters begin to differ among the three different shapes. For the "cosinusoidal" case, the parameters barely change across the full range of the hyporheic exchange flux, whereas the other two shapes exhibit distinct trends. With increasing exchange flux, the maximum travel time drops from twice the mean travel time to a value of about $\tilde{t} = 1.6$ for both the "bump" and "composite" shape. This trend is accompanied by a change of the shape parameters p_1 and p_2. For the "bump" shape, p_1 and p_2 start to deviate from each other with increasing exchange flux, creating an asymmetric travel-time distribution. In the "composite" models, asymmetry is also introduced (albeit to a smaller extent, as p_1 and p_2 change in the same direction), but the variance of the travel-time distribution decreases as p_1 and p_2 increase.

Table 5: Geometric and hydraulic parameters used for the two example calculations (I: Ammer floodplain; II: Neckar floodplain).

Parameter	Symbol	Example I	Example II	Unit
shape	–	bump	cosinusoidal	–
domain length	L	$3.00 \cdot 10^3$	$6.50 \cdot 10^3$	m
maximum width	$w_{\max}$	$6.00 \cdot 10^2$	$1.75 \cdot 10^3$	m
minimum width	$w_{\min}$	$1.75 \cdot 10^2$	$5.00 \cdot 10^2$	m
fixed heat at inlet	h_1	$3.49 \cdot 10^2$	$3.45 \cdot 10^2$	m
fixed head at outlet	h_2	$3.41 \cdot 10^2$	$3.24 \cdot 10^2$	m
transmissivity	$T_x = T_y$	$5.00 \cdot 10^{-5}$	$1.25 \cdot 10^{-2}$	$m^2\ s^{-1}$
northern influx	q_{North}	$-2.50 \cdot 10^{-8}$	$-7.50 \cdot 10^{-7}$	$m^2\ s^{-1}$
depth-integrated porosity	Φ	$2.00 \cdot 10^{-1}$	$7.50 \cdot 10^{-1}$	m
average width	w_{mean}	$4.31 \cdot 10^2$	$1.12 \cdot 10^3$	m
northern area	A_{north}	$7.69 \cdot 10^5$	$4.06 \cdot 10^6$	m^2

Most likely, this is caused by the segment of constant width that the "composite" shape exhibits in the mid-section of the floodplain aquifer, as it encourages parallel flow paths in this part of the domain. Overall however, the dominant behavior does neither depend on the domain shape nor on the magnitude of $\tilde{Q}_{\text{ex}}$.

6.5 Application to Study Site

As a result of the previous proxy-model derivations, we propose a simplified estimation of hyporheic-zone properties, for a given setup. For illustrative purposes, we include two examples resembling the floodplains of the rivers Ammer (I) and Neckar (II) close to Tübingen, in South-Western Germany. These two neighboring floodplain aquifers are exposed to similar geomorphological settings (e.g., their aspect ratios, the ambient hydraulic gradients and the degree of river channelization are comparable). However, they differ in their hydraulic properties and absolute size with the larger Neckar aquifer being dominated by sandy gravel resulting in larger transmissivities. The floodplain aquifer located in the adjacent Ammer valley is comparably small and consists of fine-grained material, which results in smaller transmissivities. Figure 16 shows a map of the two locations, as well as superimposed model results in the styling of Figure 8.

Just as stated in the introduction, we are interested in the hyporheic exchange flux, the extent of the associated exchange zone and the travel times of water parcels passing it. For the evaluation, we use the simplified proxy-equations:

1. Determine all geometric and hydraulic parameters. For the examples, we assume the values shown in Table 5.

2. Choose one of the three shape types that resembles the real shape best: we assume "bump" for the first example and "cosinusoidal" for the second one.

3. Read the coefficient values from Table 3. Ammer: $a_1^{\text{I}} = 5.852$, $a_2^{\text{I}} = 0.355$ and $a_3^{\text{I}} = 4.607$.

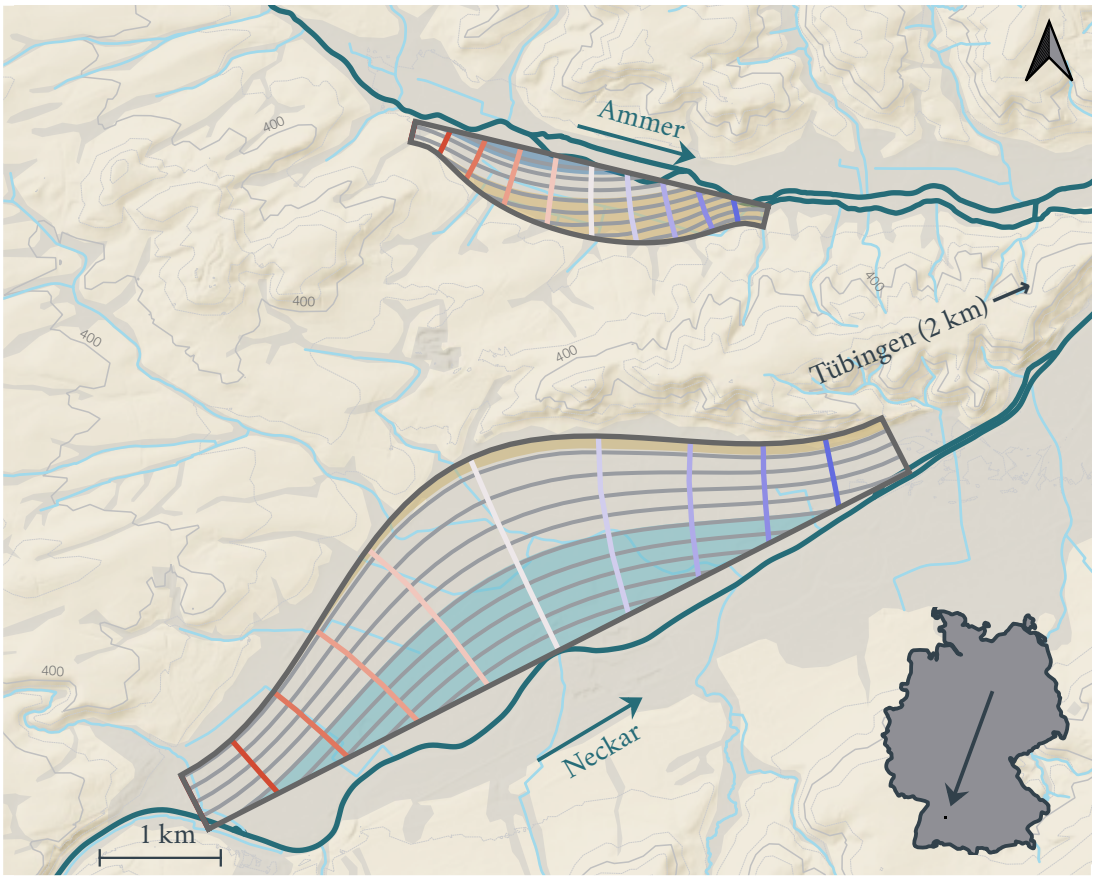

Figure 16: Map of the Ammer (case I, northern part) and Neckar (case II, southern part) floodplains near Tübingen, superimposed with example models. The model results are displayed as flow nets and translucent areas, similar to the styling of Figure 8. In accordance with that, bedrock deposits are shown in tan and floodplain materials are highlighted in light gray.

Neckar: $a_1^{\mathrm{II}} = 6.242$, $a_2^{\mathrm{II}} = 0.434$ and $a_3^{\mathrm{II}} = 4.121$.

4. Evaluate Q_0 with Equation 6.5: $Q_0^{\mathrm{I}} = 5.68 \cdot 10^{-5}\ \mathrm{m^3\ s^{-1}}$ and $Q_0^{\mathrm{II}} = 5.05 \cdot 10^{-2}\ \mathrm{m^3\ s^{-1}}$.
5. Determine $\tilde{x}$ with Equation 6.7: $\tilde{x}^{\mathrm{I}} = 0.144$ and $\tilde{x}^{\mathrm{II}} = 0.173$.
6. Evaluate $\tilde{Q}_{\mathrm{north}}$ with Equation 6.9: $\tilde{Q}_{\mathrm{north}}^{\mathrm{I}} = -1.32$ and $\tilde{Q}_{\mathrm{north}}^{\mathrm{II}} = -0.10$.
7. Determine $\tilde{Q}_{\mathrm{ex}}$ either from Equation 6.10 or graphically from Figure 13: $\tilde{Q}_{\mathrm{ex}}^{\mathrm{I}} = 0.308$ and $\tilde{Q}_{\mathrm{ex}}^{\mathrm{II}} = 0.577$.
8. Find Q_{ex} by multiplying $\tilde{Q}_{\mathrm{ex}}$ with Q_0: $Q_{\mathrm{ex}}^{\mathrm{I}} = 1.74 \cdot 10^{-5}\ \mathrm{m^3\ s^{-1}}$ and $Q_{\mathrm{ex}}^{\mathrm{II}} = 2.91 \cdot 10^{-2}\ \mathrm{m^3\ s^{-1}}$.
9. Determine the dimensionless area of the exchange zone either from Equation 6.12 or graphically from Figure 14: $\tilde{A}^{\mathrm{I}} = 0.202$ and $\tilde{A}^{\mathrm{II}} = 0.551$.
10. Find A_{ex} by multiplying $\tilde{A}$ with A_{north}: $A_{\mathrm{ex}}^{\mathrm{I}} = 1.55 \cdot 10^{5}\ \mathrm{m^2}$ and $A_{\mathrm{ex}}^{\mathrm{II}} = 2.24 \cdot 10^{6}\ \mathrm{m^2}$.
11. Obtain the mean travel time by diving the product of Φ and A_{ex} by Q_{ex}: $t_{\mathrm{mean}}^{\mathrm{I}} = 1.78 \cdot 10^{9}\ \mathrm{s} \approx 56.5\ \mathrm{a}$ and $t_{\mathrm{mean}}^{\mathrm{II}} = 5.76 \cdot 10^{7}\ \mathrm{s} \approx 1.8\ \mathrm{a}$.

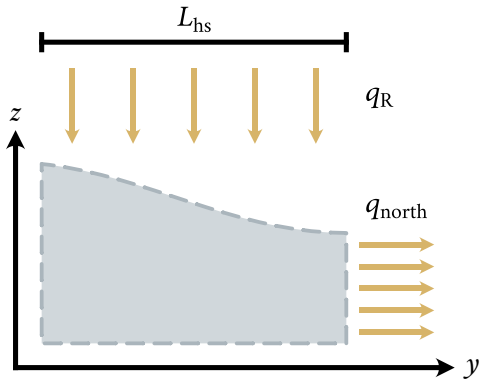

Figure 17: Schematic illustration of how to estimate the incoming flux q_{north} at the northern boundary from the length L_{hs} of the connected hillslope and a hillslope recharge rate q_R.

The results of the simplified estimation compare reasonably well with the results of the semi-analytical solution (first case: $Q_{ex} = 1.74 \cdot 10^{-5}$ m^3 s^{-1}, $A_{ex} = 1.51 \cdot 10^5$ m^2 and 54.0 a; second case: $Q_{ex} = 2.89 \cdot 10^{-2}$ m^3 s^{-1}, $A_{ex} = 2.62 \cdot 10^6$ m^2 and 2.2 a). We see that in both cases the absolute volumetric-exchange flux is small (e.g., compared to the discharge of the associated rivers, which is in the order of 0.5 m^3 s^{-1} and 7 m^3 s^{-1} under base-flow conditions). Comparing the results to the reported collection of surface-water/groundwater interaction fluxes reviewed by Cranswick and Cook (2015), the flux is basically negligible in the first case, while in the second case it operates at the lower end of the exchange-flux spectrum reported for rivers of similar discharge. In the first case (Ammer), the exchange-zone area is also relatively small. For the second case (Neckar), however, a large portion of the aquifer is taken up by the exchange zone. Both cases exhibit average travel times on the order of years, which might be enough time for the infiltrated river water to lose its chemical signature and become very similar or indistinguishable from "true" groundwater (at least on the longer flow-paths).

Our proxy-equations requires a value for the northern influx rate q_{north}, which might not be readily available for a specific site. For such cases, the following substitution might be helpful: q_{north} can be approximated by the length L_{hs} in L of the connected hillslope and an estimated average groundwater recharge rate q_R in $\mathsf{L}\,\mathsf{T}^{-1}$ stemming from a water balance of a vertical two-dimensional hillslope slice:

$$q_{north} \approx L_{hs} \cdot q_R. \tag{6.14}$$

Figure 17 provides a related schematic illustration. The sketch shows a vertical cross-section of a hillslope that is connected with the floodplain on the right-hand side. It should be noted that in the case of the Ammer example, the northern influx was not based on such a calculation, but instead based on fluxes estimated by Martin et al. (2020). A hillslope/recharge-based estimation for the Ammer case would result in much larger fluxes that would deform the flow field to the point of being inconsistent with field data. As a side note, this indicates that at least some water infiltrating on the southern hillslope adjacent to the Ammer floodplain does not enter the floodplain aquifer(s).

7 Conclusions & Outlook

The widening and narrowing of river valleys due to varying bedrock geology, produces large-scale variations in the geometry of floodplain aquifers, which subsequently induce valley-scale lateral hyporheic exchange even for straight river reaches. Estimating the size of the hyporheic exchange zone, the exchange flux, and the hyporheic travel-time distribution is relevant for groundwater management, river-water quality, and ecology. We have computed these properties by a semi-analytical modeling approach for idealized shapes of the floodplain aquifer. We found simple proxy-equations to decently approximate the geometry-driven steady-state exchange flux between floodplain aquifers and connected rivers, as well as the area covered by this hyporheic-exchange zone. The equations involve three empirical coefficients, which we have fitted for three different shapes of the floodplain aquifer. Our semi-analytical solution for the described problem provides the hydraulic-head, specific-discharge, and stream-function values throughout the domain. This information can be used to determine fluxes between different points, to construct dividing streamlines, to highlight zones of different hydrological origin and destination, and to determine travel times.

Our main conclusions from investigating the behavior of valley-scale lateral hyporheic exchange across the geometric and hydraulic parameter space are:

1. The maximum width-difference, the ambient hydraulic gradient, and the longitudinal transmissivity of the floodplain aquifer exert a linear control on the potential maximum exchange flux Q_0 between the river and the floodplain.

2. The ratio of the actual exchange flux Q_{ex} to Q_0 depends non-linearly on the aspect ratio of the domain (w_{mean}/L), which is the ratio of the floodplain width to the channel length. Large aspect ratios lead to less hyporheic exchange.

3. Horizontal hydraulic anisotropy ($T_x \neq T_y$) can act as a scaling factor for the aspect ratio of the domain, controlling the ease of which water can be drawn from the river.

4. Groundwater influx from the hillslope q_{north} exerts a strong control on the size of the exchange zone, where increasing q_{north} effectively pushes it towards the river while also reducing its longitudinal extent.

5. Travel-time distributions of hyporheic exchange water roughly follow beta distributions.

The applicability of the presented model to real case studies is of course limited: Real systems are affected by transient forcings and are subject to three-dimensional heterogeneity. Furthermore, our model only considers two-dimensional divergence-free groundwater flow and thereby assumes that the aquifer and the river are connected across the full aquifer depth, which is often not the case in real systems. Nonetheless, our expressions are useful for quick estimations of the lateral exchange flux in cases with little known information. For example, our results can be used to decide

which of several sites is most promising for targeted measurements of hyporheic exchange if budget restrictions limit field investigations to one site. Actual field data might then be used to calibrate and validate flow models (the presented one, or a more complex model) in order to simulate and quantify the hyporheic exchange more accurately.

Other relevant questions, for example, whether the groundwater sampling point at a given location lies within the hyporheic-exchange zone can be answered by the semi-analytical method, but we have not developed proxy-models for them. Future work may expand on that. Furthermore, it might be interesting to extend the semi-analytical solution to account for river meanders or non-uniform slopes. It might also be worthwhile to analyze the effects of asymmetric northern boundaries/influxes (i.e., $f_{\mathrm{B}}(x)$ and $q_{\mathrm{north}}(x)$) on the hyporheic exchange zone.

Chapter III

Optimal Well Placement for Delineating Groundwater Divides

Context

The contents of this chapter were published as "A Stochastic Framework to Optimize Monitoring Strategies for Delineating Groundwater Divides" in *Frontiers in Earth Science* (Allgeier et al., 2020). The author contributions are: Jonas Allgeier set up the numerical flow and particle-tracking model, implemented the stochastic sampler, performed the computations, created the figures, and wrote the draft manuscript; Ana González-Nicolás performed the optimal experimental design analysis and contributed to manuscript revision; Daniel Erdal developed the stochastic sampler and the pre-selection method; Wolfgang Nowak and Olaf A. Cirpka conceived the presented idea, supervised the work, provided funding, and revised the manuscript draft.

The raw data supporting the conclusions of this study and all Matlab codes used to generate the figures are publicly accessible in form of a repository at https://osf.io/ayb58/ (Allgeier et al., 2022).

8 Introduction

Groundwater divides are curves separating different subsurface catchments. Water entering the subsurface on one side of the groundwater divide ends up in a different receptor than water infiltrating on the other side of the divide. Delineating groundwater divides is therefore important for the analysis of aquifer water budgets, for investigating contaminant fate, and other applications of groundwater management. Groundwater divides also represent attractive geometries for setting second-type boundaries of hydrogeological models, since the water flux across the divide is zero (e.g., Pöschke et al., 2018; Qiu et al., 2019; Erdal and Cirpka, 2019). Obviously, a natural stream network contains many nested surface water and groundwater divides of different order (i.e., a catchment can be subdivided into sub-catchments). It is therefore always important to define the scale of investigation to identify which groundwater divides are relevant and which sub-catchments can be attributed to a higher-order catchment.

The common assumption that groundwater divides and surface-water divides, which are comparably easy to delineate, coincide is not always valid (Haitjema and Mitchell-Bruker, 2005; Bloxom and Burbey, 2015; Han et al., 2019). Especially when hydraulic conductivities are high compared to recharge rates and/or when an elevation difference between the drainage points in neighboring valleys is given a significant shift between the two divides can exist (Haitjema and Mitchell-Bruker,

2005). Drinking water extraction wells, tilted aquifers, heterogeneities and anisotropies can also be contributing factors. In such cases, a proper groundwater divide delineation requires detailed knowledge about the subsurface-flow field from hydraulic-head measurements. Such measurements can be obtained from groundwater observation wells (i.e., piezometers). The installation of such devices involves expensive drillings into the subsurface, which means that typically only few are affordable. As a result, the respective placement should be specifically optimized for delineating a particular groundwater divide. Either, one wants to find the best possible piezometer configuration for a fixed number of wells, in which the optimum is defined by minimizing the uncertainty of the divide's position, or one wants to find the configuration requiring the least number of wells for a fixed target uncertainty of the divide's location. In both cases, the objective is to maximize the information-to-costs ratio, which is a well-known general problem under the name of "optimal design of experiments" (Pukelsheim, 2006; Fedorov, 2013).

In this study, we solve the described optimization problem. We provide a framework to identify the best set of points to delineate a particular groundwater divide. The "goodness" of such a point set is defined by how much the uncertainty in the divide's location is reduced, if hydraulic-head measurements were available at these points. The best set of points might then be implemented as real-world monitoring wells, whose measurements can be used to calibrate a flow model for actually delineating the divide of interest.

Of course, during the stage of identifying promising measurement locations it is unknown which measurement values would be obtained at these locations. To circumvent this problem, we apply a specific optimal experimental design technique called Preposterior Data Impact Assessor (PreDIA, Leube et al., 2012). We feed it with a sample of steady-state groundwater models that is efficiently pre-selected to include only plausible subsurface flow fields (Erdal et al., 2020). By means of delineating the groundwater divide for each individual realization and virtually conducting all possible measurements, we can quantify both, the total uncertainty of the groundwater divide's location across the domain and by how much this scalar quantity can be reduced with a specific measurement configuration.

The main contributions of the present study are the formulation of the problem and the development of a suitable objective function for delineating a groundwater divide, as well as the combination of PreDIA with the pre-selection of plausible model results.

The motivation behind our work originates from a real field site. During the investigation of the Ammer floodplain, it was discovered that the observed lateral groundwater influxes from the hillslopes are too small to drain the water quantities gained by the hillslope's expected recharge. This imbalance of in- and outfluxes has led to the conclusion that the groundwater divide underneath the hillslope is shifted in a way that the contributing area draining towards the floodplain is much smaller than expected, when considering the surface water divide as contributing boundary. The phenomenon of flow crossing surface water divides has been referred to as "inter-basin groundwater

flow". It needs to be quantitatively estimated, before detailed studies focusing on the hillslope or floodplain can be conducted. The information of whether such inter-basin flow occurs in a domain and how pronounced it is can be of great importance, for example if contamination occurs in one basin and a sensitive receptor (e.g., a drinking water supply well) is located in the other one.

We developed our framework for cases, where the (suspected) shift of a groundwater divide is the phenomenon of interest that needs to be quantified. In reality, such a shifted divide might additionally be subject to transient processes (i.e., it might move with time). This is not covered by our methodology, but we believe our analysis might still be useful in such cases (see Section 11.5). We want to emphasize that a shifted divide does not imply its movement over time. A groundwater divide can very well be at a (quasi-)steady state while being shifted due to the geological setting, which does not change significantly over time scales relevant for groundwater management.

Section 9 introduces and explains the underlying framework. Real data from the Ammer site are used in Section 10 to test the method. We want to highlight that we separate our site-specific implementation details (application) from the general approach of our framework (methods). The results of our example study are presented and discussed in Section 11. Finally, we draw conclusions and give an outlook in Section 12.

9 Methods

9.1 Particle Tracking

The optimal experimental design method we use later on (Section 9.4) is based on stochastic runs of a steady-state subsurface flow model. To model saturated and unsaturated parts of the subsurface, we solve the steady-state version of the Richards equation for variably saturated flow in porous media with the van Genuchten parametrization (Richards, 1931; van Genuchten, 1980, see Section 2.2).

After simulating subsurface flow, we use particle tracking to determine the groundwater divide as explained in Section 9.3. Towards this end, we introduce particles at the land surface, track their advective movement according to the advective velocity $\boldsymbol{v}$ in $\mathsf{L}\,\mathsf{T}^{-1}$, and analyze on which side of the groundwater system they end. This approach is a common procedure for delineating subsurface water divides (e.g., Hunt et al., 2001; Han et al., 2019):

$$\frac{\mathrm{d}\boldsymbol{x}_i}{\mathrm{d}t} = \boldsymbol{v}(\boldsymbol{x}_i(t)) \tag{9.1}$$

$$\text{subject to } \boldsymbol{x}_i(t=0) = \boldsymbol{x}_i^{\mathrm{ini}} \tag{9.2}$$

$$\text{with } \boldsymbol{v} = \frac{\boldsymbol{q}}{\Theta_{\mathrm{w}}}, \tag{9.3}$$

where $\boldsymbol{x}_i(t)$ in L is the position of particle i at time t in T, $\boldsymbol{x}_i^{\mathrm{ini}}$ in L is the starting position, $\boldsymbol{q}$ in $\mathsf{L}\,\mathsf{T}^{-1}$ is the specific discharge, and Θ_{w} is the volumetric water content.

The approach of delineating the groundwater divide by particle tracking obviously implies that the divide is located within the modeling domain. This is in contrast to many practical groundwater-modeling studies, where the domain is bounded by the assumed groundwater divides. Under such conditions, these groundwater divides are fixed by the model choice. Since we want to study the uncertainty of the groundwater divide, we require a model domain where the divide is in the interior so that the model has the freedom to shift it.

9.2 Generation of a Plausible Model Sample

In order to capture the uncertainty of the divide's location (prior to any measurements and after hypothetical measurements), our framework makes use of ensemble-modeling. This implies the repeated simulation of the same conceptual model with different numerical representations. These can be formally identical, differing only, for example, in some material property values. They could also differ in more fundamental properties, like the internal structure. We call the final group of model entities a "sample", to avoid confusion with the term "ensemble" referring to such a group of infinite size. Each entity of the sample is termed a realization or sample member.

Formally, a sample member is defined both, by the formulation of the general model itself (common to all members) and by a member-specific set of parameters. In addition to that, the sample member also comprises its deterministic modeling results (after the model was evaluated), which can be reproduced from the general model by using the same parameter set. We denote these parameter sets $\boldsymbol{p}$. They are vectors of all individual properties that differ between realizations. The vector $\boldsymbol{p}$ may include not only material properties, but also boundary conditions or geometric descriptors (for an example, we refer to our application in Section 10.1.3).

In theory, we could create a sample of sufficient size just by drawing random parameter sets from appropriate prior distributions and subsequent numerical modeling of subsurface-flow. These prior distributions could be derived from measurements (e.g., pumping tests for hydraulic conductivities), other models (e.g., recharge rates) or expert knowledge (e.g., anisotropies). Afterwards, particle-tracking would obtain one groundwater divide for each realization. In practice however, we need to exclude parameter sets that lead to implausible model results (e.g., wrong signs of fluxes across boundaries; more examples in context of our application, Section 10.2), because that would ignore obvious insight into the correct system behavior and thus overstretch uncertainty. Conversely, we do not want to restrict the parameter ranges too much because we want to assess the full space of plausible model parameters. Therefore, we keep the prior parameter ranges untouched, but rely on the exclusion of models with obviously unrealistic results (denoted *unbehavioral* or *implausible*).

While excluding unbehavioral realizations is a conditioning step, we would not yet consider it a model calibration, but rather a plausibility check or pre-selection (see Erdal and Cirpka, 2019; Erdal et al., 2020; Erdal and Cirpka, 2020). In a rigorous conditioning step (i.e., "stochastic calibration") that could follow on this pre-selection, we would modify the parameters of sample members to better

meet the exact measurement values. A potential method to do that would be an ensemble Kalman smoother. However, a full stochastic calibration on the existing data would be computationally expensive, but not informative about the quantity of interest, namely the position of the groundwater divide. The lack of hydraulic-head measurements that are informative about the delineation of the groundwater divide is the very reason why we perform the optimal design analysis to begin with.

The decision about the plausibility and acceptance or rejection of a candidate model is based on a set of criteria. Each plausibility criterion compares a scalar model outcome (e.g., the flux across a specific boundary) with a target value that must not be exceeded or fallen below. Only if a model realization fulfills all plausibility criteria, it will be included in the sample for further analysis.

A key problem of the pre-selection is that more than 94 % of randomly drawn parameter sets in our application miss at least one criterion. If we performed full runs of the numerical subsurface-flow model for each model candidate, we would thus waste more than 94 % of the computing time on model runs that must be discarded. To overcome this problem efficiently, we have adopted the pre-selection method of Erdal et al. (2020), which is based on the work of Erdal and Cirpka (2019). This method relates the plausibility criteria with the model parameters $\boldsymbol{p}$ by means of interpolation, to estimate whether a new parameter set is likely to be plausible. To this end, we follow these steps:

1. We create a small initial sample of $\boldsymbol{p}$ by Latin Hypercube sampling (e.g., McKay et al., 1979; Tang, 1993; Lin, Tang, et al., 2015) from appropriate priors and perform numerical subsurface-flow modeling for all sample members. We compute the respective values of the plausibility criteria for each realization.

2. We train one GPE per plausibility criterion as a proxy-model, with the initial sample of full model runs. As discussed in Section 2.3, GPEs are kriging interpolators in parameter space that estimate the expected value of the plausibility criterion and quantify its estimation variance, provided that the assumptions of kriging (e.g., statistical stationarity) hold. We want to emphasize here that this is not a spatial interpolation, but an interpolation of the model response to parameter values.

3. We then draw further random samples of $\boldsymbol{p}$. For each of them, we apply the GPEs to compute the compliance probability with each plausibility criterion. If a realization's product of all individual compliance probabilities (i.e., its overall probability) does not exceed a certain threshold value (in our case 50 %), we discard it and draw a new sample. This evaluation is comparably quick (fraction of a second) and saves us modeling time that would be wasted by running a model that would probably need to be rejected due to implausible results.

4. For a model candidate where this product exceeds the threshold probability (a "stage-1-accepted" realization), we perform the simulation of the full subsurface-flow model. A small percentage of sample members (we use 5 %) is run directly without checking against the GPE estimates first.

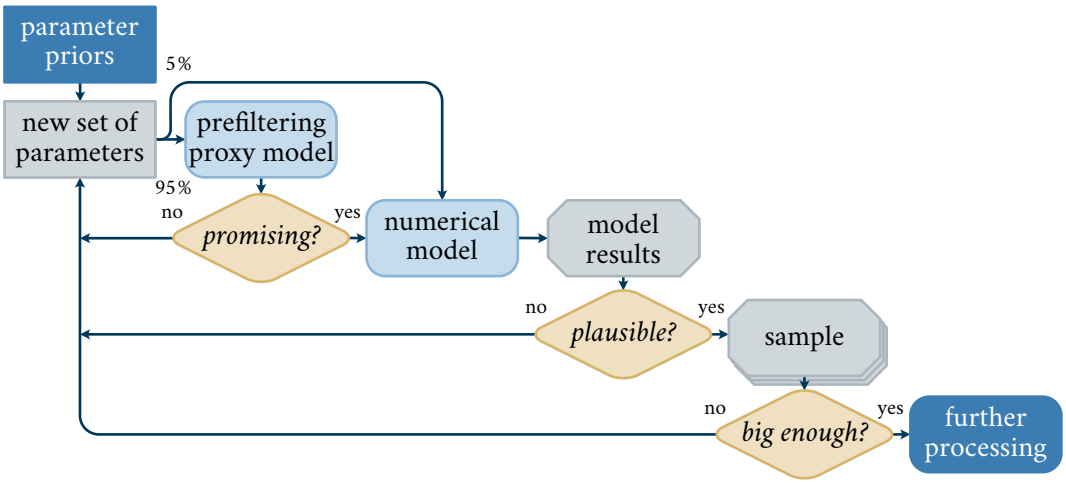

Figure 18: Procedure to generate a sample of physically plausible model realizations.

5. If the model candidate also meets the plausibility criteria after running the full numerical model, it is "stage-2-accepted" (i.e., included in the sample of physically plausible models), and particle-tracking simulations are performed to obtain the groundwater divide. Otherwise, it is discarded.

6. With an increasingly large set of full model runs, the GPE model is regularly retrained to improve its accuracy in predicting the behavioral status of subsequent model candidates.

With this procedure, we were able to increase the overall acceptance ratio, that is, the number of stage-2-accepted full-model runs over the total number of full-model runs. In the initial small sample (full Monte Carlo), only 6 % of the realizations passed the plausibility check (111 out of 2000). With the interpolation method, we were able to achieve an acceptance ratio of 69 % of realizations subject to a full model run (50 000 of 72 481 stage-1-accepted parameter sets; a large number of randomly drawn parameter sets was rejected in stage 1). Figure 18 schematically illustrates the whole sample-generation procedure. It results in n_{sample} stage-2-accepted realizations that will actually be used in the following analysis.

9.3 Uncertainty in Delineating a Groundwater Divide

For each stage-1-accepted parameter realization (see step 4 in Section 9.2), we determine the scalar model outcomes of the plausibility check. Additionally, we simulate virtual measurement values of hydraulic heads at all potential measurement locations, by determining the respective elevations of the groundwater table at these locations. The number and location of such potential measurements is known prior to the analysis and part of the problem statement.

Only for the n_{sample} stage-2-accepted realizations, we compute a vector $\boldsymbol{z}$ of particle fates via particle tracking for a regular map of starting locations: We introduce n_{par} particles at the model domain's surface. These particles are tracked through the domain until they exit the domain through a groundwater outlet. This tracking allows us to classify the particles with two categories summarized

by the classification vector $\boldsymbol{z}$ with $z_i \in \{0, 1\}$ and $i = 1, \ldots, n_{\text{par}}$. A particle i that ends up in one outlet (A) is assigned the value $z_i = 1$, while a particle ending up in the other outlet (B) obtains a value of $z_i = 0$. Since each particle is related to a starting point in two-dimensional space, $\boldsymbol{z}$ represents what we call the *binary particle-fate map*. This binary classification is sufficient to delineate the boundary of a single subdomain, but it cannot be used to delineate all groundwater divides between more than two subdomains (e.g., due to groundwater extraction wells). In Section 9.5 we therefore include a generalization to an arbitrary number of subdomains. In the following, we will focus on binary systems, because this is the most common scenario.

Other approaches than particle tracking for the delineation of groundwater divides exist. They are typically based on locating the "ridge of the groundwater table". However, they have been shown to be less reliable (Han et al., 2019).

The fate of a particle i depends on the parameter vector $\boldsymbol{p}$ (including all variable model decisions). The probability $P(z_i)$ of z_i being one (that is, of the associated starting point to be within the catchment of outlet A) is computed by integrating over the space $\Omega_{\boldsymbol{p}}$ of the parameter vector $\boldsymbol{p}$, weighted with the probability density of $\boldsymbol{p}$:

$$P(z_i) = \int_{\Omega_{\boldsymbol{p}}} z_i(\boldsymbol{p}) \rho(\boldsymbol{p}) \, \mathrm{d}\boldsymbol{p} \approx \sum_{j=1}^{n_{\text{sample}}} z_i(\boldsymbol{p}_j) P(\boldsymbol{p}_j), \tag{9.4}$$

in which $z_i(\boldsymbol{p})$ is the binary fate of particle i for the given parameter vector $\boldsymbol{p}$, and $\rho(\boldsymbol{p})$ is the probability density of $\boldsymbol{p}$. The right-hand side of Equation 9.4 is the Monte-Carlo approximation of $P(z_i)$ by the sample of discrete $\boldsymbol{p}$-values with the probability $P(\boldsymbol{p}_j)$ given to the $\boldsymbol{p}$-value of the j-th realization. In our initial sample, all accepted realizations are equally likely, implying $P(\boldsymbol{p}_j) = 1/n_{\text{sample}} \ \forall j$. Upon conditioning on (virtual) head measurements, $P(\boldsymbol{p}_j)$ will become a Bayesian weight (explained later). Franzetti and Guadagnini (1996) and Hunt et al. (2001) used a similar approach to estimate the uncertainty of capture-zone delineations.

We can now compute the probability $P_{\text{mc}}(z_i)$ of misclassifying the fate of particle i:

$$P_{\text{mc}}(z_i) = 2P(z_i)(1 - P(z_i)). \tag{9.5}$$

This equation expresses the probability that particle i, which actually ends up in outlet A, is estimated to end up in outlet B or vice versa. P_{mc} ranges from 0.0 (full certainty) to 0.5 (maximum uncertainty). The reason for 0.5 being the largest value of P_{mc} is the underlying assumption that the decision threshold for classification is at 50 %. $P(\boldsymbol{z})$ and $P_{\text{mc}}(\boldsymbol{z})$ can be visualized as maps of probability all over the catchment. We integrate the probability of misclassification over all starting locations $\boldsymbol{x}^{\text{ini}}$ of particles to obtain an integral metric U describing the uncertainty of the groundwater divide:

$$U(\boldsymbol{z}) = \frac{1}{A_{\text{2D}}} \int_{A_{\text{2D}}} P_{\text{mc}}(z(\boldsymbol{x}^{\text{ini}})) \, \mathrm{d}\boldsymbol{x}^{\text{ini}} \approx \frac{1}{A_{\text{2D}}} \sum_{i=1}^{n_{\text{par}}} P_{\text{mc}}(z_i) A_i^{\text{ini}}, \tag{9.6}$$

in which A_{2D} is the two-dimensional top surface area of the model domain and A_i^{ini} is the contributing area of particle i, which may be computed by Voronoi tessellation of all starting locations (e.g., Brassel and Reif, 1979). Large values of $U(\boldsymbol{z})$ express that the outlet destination of particles is uncertain on a large fraction of the domain's surface, which is not desirable.

As discussed in the context of Equation 9.4, the probability $P(z_i)$ of starting location $\boldsymbol{x}_i^{ini}$ being in the catchment of outlet A, and thus the associated probability of misclassification $P_{mc}(z_i)$ and ultimately the overall uncertainty $U(\boldsymbol{z})$, depends on the probabilities $P(\boldsymbol{p}_j)$ of individual parameter realizations j. This implies that conditioning the parameter vector $\boldsymbol{p}$ on head observations will change the overall uncertainty U of delineating the groundwater divide. The following optimal design analysis aims at minimizing U.

9.4 Prospective Optimal Experimental Design

To find the optimal placement of piezometers in order to delineate a groundwater divide, we apply the optimal experimental design method PreDIA (the Preposterior Data Impact Assessor; Leube et al., 2012), which we briefly review in the given context.

The scientific question of optimal design is to find the combination of measurements or experiments with the largest information content regarding a target quantity, before the experiment itself is carried out. Formally, the objective is to identify the single design $\boldsymbol{d}_{opt}$ of a set of n_{des} possible designs $\boldsymbol{d}$ in the design space $\boldsymbol{d} \in \boldsymbol{D}$ that maximizes a utility function $\phi(\boldsymbol{d})$ (Leube et al., 2012):

$$\boldsymbol{d}_{opt} = \arg\max_{\boldsymbol{d} \in \boldsymbol{D}} \left[\phi(\boldsymbol{d}) \right]. \tag{9.7}$$

A design in this notation is a vector containing information about how measurements are taken in space (and/or time). The utility function $\phi(\boldsymbol{d})$ is a measure of the usefulness of data obtained with an experiment using design $\boldsymbol{d}$. The evaluation of ϕ requires knowledge about the measurement results of a particular design, which is unknown at the stage of the optimal-experimental-design analysis. PreDIA can circumvent this problem by means of ensemble-based modeling.

As previously described, $\boldsymbol{p}$ denotes the input parameter vector, comprising all uncertain model decisions, such as material properties (e.g., hydraulic conductivity), boundary conditions (e.g., recharge), geometric parameters (e.g., thickness of geological units), or structural modeling parameters (e.g., presence of layers). As outlined above, we create a sample of members with physically plausible behavior. The variability in model input $\boldsymbol{p}$ leads to interdependent variability of model output, both with respect to simulated measurements and simulated target quantities (the particle-fate maps).

For a given realization $\boldsymbol{p}_i$, we can simulate virtual observations $\boldsymbol{f}_y(\boldsymbol{p}_i, \boldsymbol{d})$ for a specific design $\boldsymbol{d}$, in which $\boldsymbol{f}_y$ denotes the simulation outcome of the measured quantities. To account for measurement errors, we add a random error term $\boldsymbol{\varepsilon}_y$ to $\boldsymbol{f}_y(\boldsymbol{p}_i, \boldsymbol{d})$.

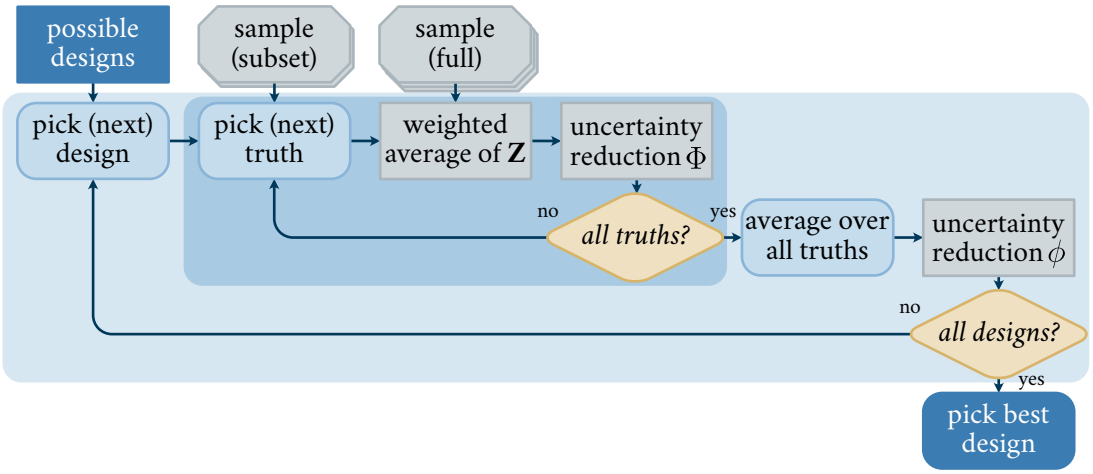

Figure 19: Schematic illustration of PreDIA. Inner loop in dark blue, outer loop in light blue.

This results in virtual measurements $\boldsymbol{y}_i(\boldsymbol{d})$ of a specific design $\boldsymbol{d}$ and parameter realization i:

$$\boldsymbol{y}_i(\boldsymbol{d}) = \boldsymbol{f}_y(\boldsymbol{p}_i, \boldsymbol{d}) + \boldsymbol{\varepsilon}_y. \tag{9.8}$$

To answer the optimal-experimental-design question, we use the stage-2-accepted realizations to compute the $1 \times n_{\text{par}}$ vector of prediction variables $\boldsymbol{z}$ (binary particle-fate map) as discussed above. The prediction solely depends on the input parameter vector $\boldsymbol{p}$ and is independent of the measurement design $\boldsymbol{d}$.

In our particular application, the prediction variable is binary, namely the true/false information whether a particle introduced into the subsurface at a given location belongs to one out of two catchments. The binary nature of $\boldsymbol{z}$ implies that the sample average of it equals the vector of probabilities that the individual elements of $\boldsymbol{z}$ are one.

After acquiring n_{sample} stage-2-accepted sample members with a parameter vector $\boldsymbol{p}$ and after computing the associated virtual measurements and prediction variables, we have $n_{\text{sample}} \cdot n_{\text{des}}$ sets of $\boldsymbol{y}(\boldsymbol{d})$ and n_{sample} sets of $\boldsymbol{z}$ (which can be summarized in a $n_{\text{sample}} \times n_{\text{par}}$ matrix $\boldsymbol{Z}$). As illustrated in Figure 19, PreDIA proceeds in the following way to identify the best design:

1. Compute the unconditional sample mean $P(z_i)$ of all target variables z_i by Equation 9.4 with equal probabilities of all realizations.

2. Compute the vector of unconditional probabilities of misclassification $P_{\text{mc}}(z_i)$ by Equation 9.5 and the associated overall prior uncertainty of groundwater-divide delineation $U(\boldsymbol{z})$ by Equation 9.6.

3. Select a random subset of n_{sub} realizations to define "virtual truths". The distribution of virtually measured values $\boldsymbol{y}$ in this subset should be similar to the corresponding distribution using the full sample (across all designs). When computationally feasible, select all n_{sample} sample members such that $n_{\text{sub}} = n_{\text{sample}}$.

4. For each design $\boldsymbol{d}$, determine the utility function $\phi(\boldsymbol{d})$ by marginalizing an objective function Φ over all n_{sub} realizations:

$$\phi(\boldsymbol{d}) = \frac{1}{n_{\text{sub}}} \sum_{j=1}^{n_{\text{sub}}} \Phi(\boldsymbol{y}_j(\boldsymbol{d})), \tag{9.9}$$

in which we have assumed that all realizations j are equally likely. This step defines the outer loop over all designs $\boldsymbol{d} \in \boldsymbol{D}$, which is illustrated by light blue shading in Figure 19). An inner loop (explained in the following), is used to determine the objective function values Φ.

5. Identify the design $\boldsymbol{d}_{\text{opt}}$ maximizing $\phi(\boldsymbol{d})$ according to Equation 9.7.

In an inner loop (illustrated by dark blue shading in Figure 19), each of the n_{sub} virtual observations for the currently chosen design $\boldsymbol{d}$ are temporarily considered to be the truth. The inner loop results in n_{sub} objective-function values for a given design $\boldsymbol{d}$. It follows this procedure:

1. Declare realization j with the virtual observations $\boldsymbol{y}_j(\boldsymbol{d})$ and the virtual prediction variable $\boldsymbol{z}_j$ temporarily as truth.

2. Assign each realization $i \neq j$ of the full set of n_{sample} realizations a Bayesian weight depending on how close the respective observations $\boldsymbol{y}_i(\boldsymbol{d})$ are to $\boldsymbol{y}_j(\boldsymbol{d})$. The weights are computed by terms describing the likelihoods $\mathcal{L}_i$ of observation $\boldsymbol{y}_i(\boldsymbol{d})$ using the observation $\boldsymbol{y}_j(\boldsymbol{d})$ as temporary truth:

$$w_i = \frac{\mathcal{L}_i}{\sum_i \mathcal{L}_i} \tag{9.10}$$

$$\mathcal{L}_i = \begin{cases} \exp\left(-\frac{1}{2}\left(\boldsymbol{y}_i(\boldsymbol{d}) - \boldsymbol{y}_j(\boldsymbol{d})\right)^{\mathsf{T}} \boldsymbol{R}_\varepsilon^{-1} \left(\boldsymbol{y}_i(\boldsymbol{d}) - \boldsymbol{y}_j(\boldsymbol{d})\right)\right) & \text{if } i \neq j \\ 0 & \text{otherwise,} \end{cases} \tag{9.11}$$

in which n_y is the number of virtual measurements according to the current design $\boldsymbol{d}$, and $\boldsymbol{R}_\varepsilon$ is the $n_y \times n_y$ covariance matrix of measurement errors. We assume this to be a diagonal matrix $\sigma^2_{\text{meas}}\boldsymbol{I}$, which implies that these errors are uncorrelated and normally distributed with a standard deviation of σ_{meas}. The weights are summarized in a $1 \times n_{\text{sample}}$ vector $\boldsymbol{w}$.

3. Compute the weighted mean of all prediction variables in $\boldsymbol{Z}$ ($\boldsymbol{z}$ conditioned on the observations $\boldsymbol{y}_j(\boldsymbol{d})$ obtained with $\boldsymbol{p}_j$ and $\boldsymbol{d}$; this is the conditioned version of Equation 9.4):

$$P(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d})) = \boldsymbol{w}\boldsymbol{Z}. \tag{9.12}$$

The $1 \times n_{\text{par}}$ vector $P(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d}))$ is the vector of probabilities that the individual elements of $\boldsymbol{z}$ are one, conditioned on the vector of observations $\boldsymbol{y}_j(\boldsymbol{d})$ of realization j using the design $\boldsymbol{d}$.

4. Compute the conditional probability of misclassification $P_{\text{mc}}(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d}))$ by substituting the conditional averaged particle fate probabilities $P(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d}))$ rather than the vector of unconditional probabilities $P(\boldsymbol{z})$, into Equation 9.5.

5. From the vectors of conditional misclassification probabilities $P_{\mathrm{mc}}(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d}))$ and unconditional misclassification probabilities $P_{\mathrm{mc}}(\boldsymbol{z})$, compute a scalar metric $\Phi(\boldsymbol{y}_j(\boldsymbol{d}))$ summarizing the relative reduction of uncertainty U in classifying all elements of $\boldsymbol{z}$ by considering the observations $\boldsymbol{y}_j(\boldsymbol{d})$ belonging to design $\boldsymbol{d}$ by using Equation 9.6:

$$\Phi(\boldsymbol{y}_j(\boldsymbol{d})) = 1 - \frac{U(\boldsymbol{z}|\boldsymbol{y}_j(\boldsymbol{d}))}{U(\boldsymbol{z})}. \tag{9.13}$$

The two loops of PreDIA require large sample sizes to make reliable statements about design performances. To estimate whether the chosen sample is large enough for the results to be meaningful, one can use the Averaged Effective Sample Size (AESS; Leube et al., 2012), a concept adapted from Liu, 2008. It is a measure of how many realizations actually contribute to the analysis, where low values indicate filter degeneracy, which needs to be mitigated by increasing the ensemble size.

PreDIA has fundamental advantages over other optimal-experimental-design techniques. It is applicable to inherently non-linear problems without the need of a linearization. It is also very versatile because it imposes few restrictions on the numerical model. Besides the definition and reading of some pre-run input and post-run output quantities, the actual numerical simulation code is independent of PreDIA. This independence makes it trivial to couple any numerical model with PreDIA. It can be seen as a post-processing routine for any modeling sample. PreDIA can capture all kinds of known or estimated uncertainties in boundary conditions, material properties, model structure, or any other model parameters due to its ensemble-based nature.

The disadvantage of PreDIA lies in its computational cost. The analysis requires large sample sizes (i.e., tens of thousands of model runs) and is computationally expensive itself. These difficulties, however, can be overcome with parallel computing techniques (i.e., running multiple realizations at the same time) and simplified models that are comparably quick.

9.5 Generalizations

Non-Binary Systems In cases where one wants to delineate not only a particular (sub-)catchment's boundary, but the (potentially intersecting) groundwater divides between more than two of such catchments, the formulation of our objective function (Equation 9.9) based on binary particle fate maps (Equation 9.5) is insufficient. Here, the particle fates cannot be described with the binary Bernoulli distributions, where the outcome for particle i is $z_i \in \{0, 1\}$.

Instead, one could rely on categorical distributions, which can have more than two outcomes. For example, in a domain with n_{fates} outlets the fate of particle i can be described with $z_i \in \{1, 2, \ldots, n_{\mathrm{fates}}\}$. Each of the outcomes would correspond to one outlet/subcatchment/receptor. To adapt our objective function to these cases, we need to formulate the overall probability of misclassifying the fate of a particle i. We denote the probability that particle i belongs to the receptor k is $P(z_i = k)$.

For such a description, all steps of the method remain as outlined above, with the only exception that the overall probability of misclassification now becomes:

$$P_{\mathrm{mc}}(z_i) = \sum_{k=1}^{n_{\mathrm{fates}}} P(z_i = k) \cdot (1 - P(z_i = k)). \tag{9.14}$$

Transient Systems A potential transient implementation of our framework would require a new formulation of the objective function. In such applications both, the modeled subsurface flow-field and the observations would change over time. This means that also the particle fate maps are transient, since the fate probabilities might change throughout the simulation period. This results in dynamic maps of misclassification probability, that is $P_{\mathrm{mc}}(z)$ becomes $P_{\mathrm{mc}}(z, t)$, which is a function of time t. One potential way to define a metric quantifying the uncertainty of a transient groundwater divide would be to perform an additional integration/averaging over the simulated model duration τ:

$$U(\boldsymbol{z}) = \frac{1}{\tau \cdot A_{2\mathrm{D}}} \int_{\tau} \int_{A_{2\mathrm{D}}} P_{\mathrm{mc}}(z(\boldsymbol{x}^{\mathrm{ini}}), t) \, \mathrm{d}\boldsymbol{x}^{\mathrm{ini}} \, \mathrm{d}t \tag{9.15}$$

$$= \frac{1}{\tau \cdot A_{2\mathrm{D}}} \int_{A_{2\mathrm{D}}} \int_{\tau} P_{\mathrm{mc}}(z(\boldsymbol{x}^{\mathrm{ini}}), t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x}^{\mathrm{ini}}. \tag{9.16}$$

9.6 Numerical Implementation

Our framework does not depend on the choice of any specific software, neither for the flow simulation nor for the optimal-design analysis. In the following application, we use HGS for the former. Because of the Richards equation's nonlinearity, we do not solve for steady-state flow directly. Instead, we use the transient solver of HGS with constant forcings over a simulation time of $3 \cdot 10^{12}$ s $\approx$ 100 000 a using adaptive discretization in time. It is reasonable to assume that steady state is achieved within this time. The velocity field of HGS is then transferred to Tecplot to perform advective particle tracking with Tecplot's streamtracing routine in its command line mode (Tecplot Inc., 2019).

The stochastic engine responsible for the sampling of the parameter space and performing the plausibility check of sample members by the GPE-based proxy-model is written in Matlab and based on the code of Erdal and Cirpka (2019). We execute the stochastic sampler on a mid-size high-performance computing cluster with 24 Intel Xeon L5530 nodes (8 cores per node; 2.4 GHz and 8 MB per chip).

The optimal design analysis using PreDIA is implemented as a separate Matlab code that acts on the full sample of stage-2-accepted realizations after its acquisition.

10 Application to Study Site

The location of the groundwater divide between the Ammer and Neckar catchments north and south of the Wurmlingen Saddle is unclear. A systematic northwards shift of the groundwater divide has been suggested by modeling studies of Kortunov (2018). However, no piezometers currently exist along the decisive hillslope so that the hypothesis of a shifted divide is fairly uncertain.[†] In order to test the hypothesis of a shifted groundwater divide at this location, the installation up to three piezometers is considered. We apply the presented framework to optimize the placement of these observation wells such that they can help to delineate the groundwater divide. Due to legal and logistical reasons, all new groundwater observation points need to be placed on a transect parallel to the road from Unterjesingen to Wurmlingen (twenty proposed locations are shown in Figure 20). We therefore use the described method to determine the best configuration of piezometers along this transect.

10.1 Details of the Subsurface-Flow Model

10.1.1 Discretization

Figure 20 shows a plan view of the model domain, the discretization and the boundary conditions. We discretize the two-dimensional area by 3959 triangles arranged in a conforming unstructured grid. These triangles are extruded in the vertical dimension to generate triangular prisms. Using 35 prism layers from the bottom of the lowermost Erfurt formation to the surface elevation results in a grid of 138 565 three-dimensional elements with 74 412 nodes. The number of prism layers is constant throughout the domain, whereas the prism layer thicknesses vary. The topmost prism layers of the domain are discretized more finely, in order to better resolve the unsaturated zone. The chosen mesh is a compromise between numerical accuracy and computational effort. A comparison between models set up on this grid with models defined on an eightfold refined version revealed some deviations at the coarser parts (mostly on the Neckar side and in the deeper subsurface of the domain). However, we deem these acceptable because they occur where the exact hydraulic heads are of little interest to us anyway and because they are minor compared to the variance between different model realizations. For future applications of the presented method, we suggest performing a grid convergence analysis with a range of different discretizations. The coarsest grid providing adequate accuracy should be selected.

From the geological units outlined in Section 3.2, we define a total of twelve different hydrostratigraphic units that are assumed to have uniform properties. These units are: (1) lower Erfurt formation (`kuE`), (2) upper Erfurt formation (`kuE`), (3) lower Grabfeld formation (`kmGr`), (4) upper Grabfeld formation (`kmGr`), (5) lumped sandstone formations (`km2345`), (6) hillslope hollows, (7) gravel in Neckar floodplain, (8) gravel in Ammer floodplain, (9) clay, (10) Tufa, (11) alluvial

[†]The three Grabfeld formation observation wells on the southern hillslope on the Wurmlingen saddle that were mentioned in Section 3.3 have been installed after the analysis and publication of this chapter.

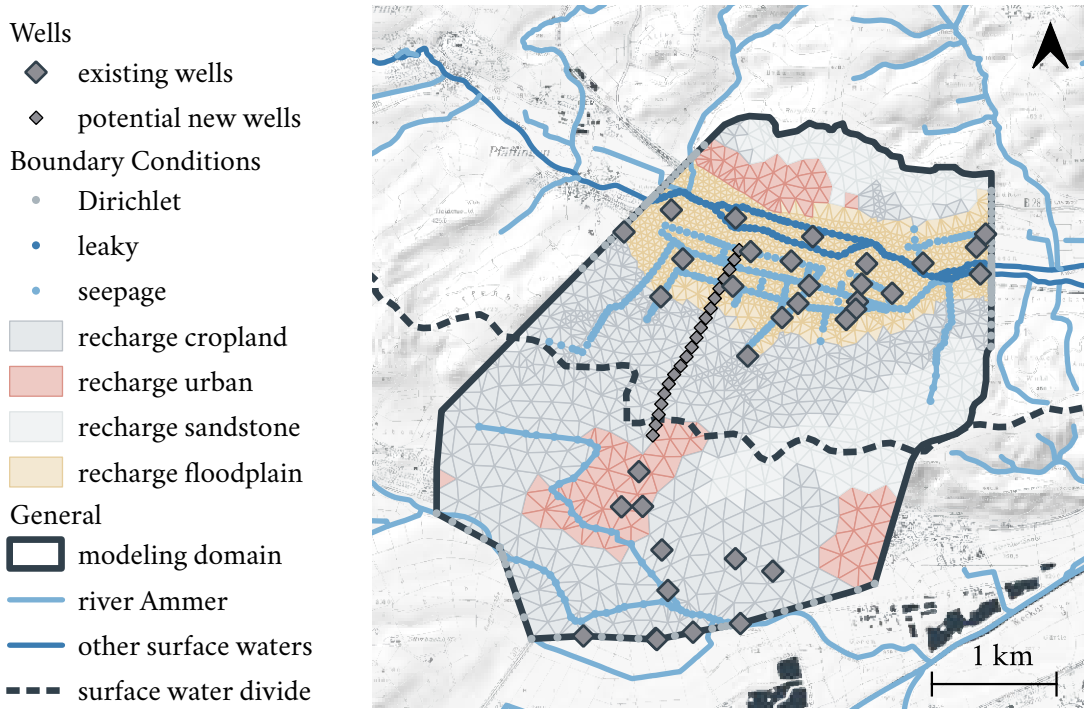

Figure 20: Boundary conditions and two-dimensional discretization of the underlying subsurface flow model.

fines, and (12) a river buffer zone. We decided to split the Erfurt and Grabfeld formation into two separate subunits, to account for heterogeneities in hydraulic conductivity (e.g., due to weathering). Figure 21 illustrates the considered hydrostratigraphic units in three-dimensional renderings.

10.1.2 Boundary Conditions

Along different parts of the boundary, we apply different boundary conditions:

1. If not specified otherwise, all outer mesh faces are assigned a no-flux Neumann boundary condition. These boundaries are either in formations of very low hydraulic conductivity (particularly the bottom) or the boundaries are far away from the area of interest like the northern boundary, which is derived from a secondary surface water divide on the far side of the Ammer valley. The eastern and western boundaries are approximately parallel to the estimated flow field.

2. Three Dirichlet boundary sections are defined at the western, eastern, and southern sides of the domain to allow regional groundwater flow (see Figure 20). To obtain the fixed-head values, we interpolate between observation well data. In the Ammer valley, the Dirichlet boundaries extend over the Quaternary fillings, while on the Neckar side, they extend over the whole depth of the model, where the formation consists of a thin, highly conductive gravel that thins out towards the municipality of Wurmlingen. Because of the high hydraulic conductivity and the absence of significant vertical hydraulic gradients here, we average the interpolated head values over depth for the Dirichlet assignment.

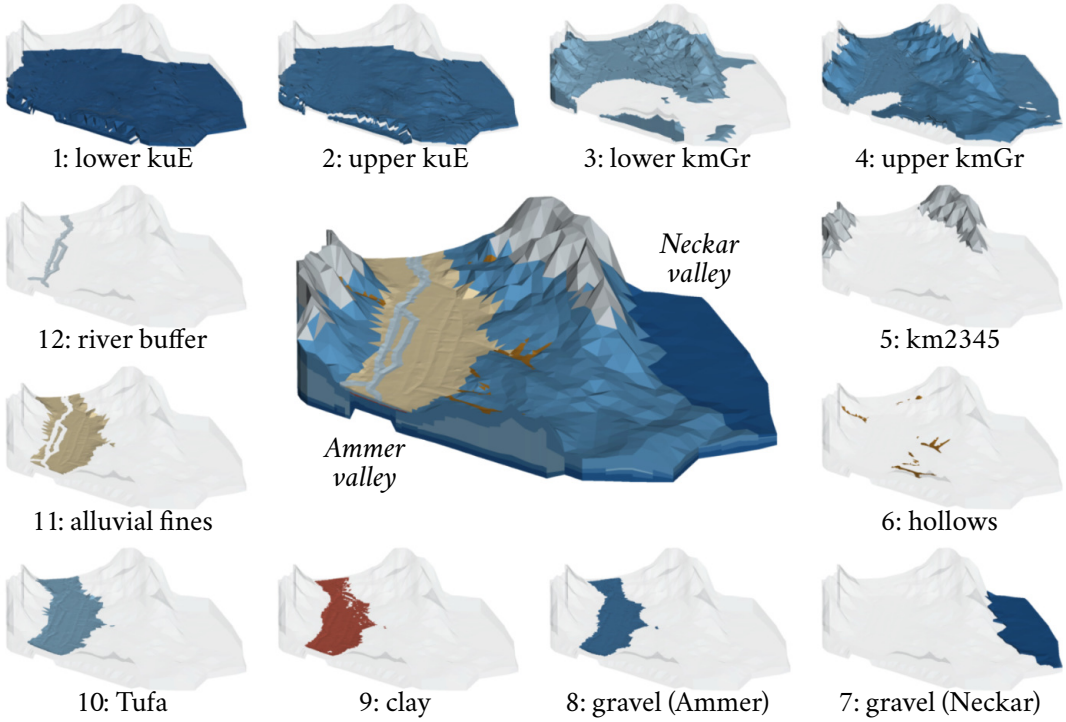

Figure 21: Three-dimensional overview of the subsurface-flow model with a fivefold vertical exaggeration.

3. On the top surface of the domain, we apply recharge as a fixed-flux Neumann boundary condition across element faces. Recharge rates in different zones depend on land use (cropland, floodplain, urban areas, and `km2345`-covered parts). By providing recharge as a model boundary, we lump the dynamic interaction of evaporation, transpiration, precipitation and soil water storage into a single stationary quantity, which is of course a simplification. However, since we are interested in the effective, long-term behavior and not the high-resolution fluctuations, we consider this simplification justified. We base our range of possible recharge rates on previous work conducted in our domain or in comparable aquifers in proximity (Holzwarth, 1980; Wegehenkel and Selg, 2002; Selle et al., 2013).

4. We use a leaky boundary condition to simulate the interaction between groundwater and the Ammer river.

5. For the network of drainage ditches in the Ammer valley and the small surface water creek in the Neckar valley, we apply seepage boundaries.

6. Drain boundary conditions are applied to all other surface nodes, allowing water to drain whenever the groundwater table is above the ground surface. We distinguish between elements that belong to the Ammer floodplain (highlighted in light brown in Figure 20) and the remaining surface.

Note that there are no groundwater abstractions within the model domain so that we do not need to consider corresponding internal boundary conditions.

We tested different initial conditions for the flow solution, but the choice of initial condition affected mostly the run time needed to reach convergence to steady-state, and influenced the steady-state flow field itself only marginally. We settled with initial hydraulic heads equal to the surface elevation. For other applications, we recommend a similar comparison procedure to identify a useful initial condition. Choices that are too far away from a realistic flow field (e.g., a completely dry domain) can lead to convergence problems due to the nonlinearity of Richards' equation.

10.1.3 Uncertain Parameters and Prior Information

Each discretized spatial element (i.e., triangular prism) has a set of parameters defining the hydraulic properties of its material. All elements belonging to the same hydrostratigraphic unit share the same set of parameters. This includes the hydraulic conductivity tensor, which we describe by a horizontal (K_{xy} in $\mathsf{L}\,\mathsf{T}^{-1}$) and a vertical hydraulic conductivity (K_z in $\mathsf{L}\,\mathsf{T}^{-1}$):

$$\boldsymbol{K}_{\text{sat}} = \begin{bmatrix} K_{xy} & 0 & 0 \\ 0 & K_{xy} & 0 \\ 0 & 0 & K_z \end{bmatrix}. \tag{10.1}$$

In doing so, we consider anisotropy in the vertical principal spatial direction. Further parameters that need to be assigned to each unit are the van Genuchten parameters α (in L^{-1}) and N, and the residual water saturation $S_{\text{wr}} = \Theta_{\text{r}}/\Theta_{\text{s}}$. For the transient calculations, we also need storage-related parameters (i.e., porosity or specific storativity), but they do not affect the final steady-state solution.

Table 6 summarizes all material properties considered random. These parameters are the first part of the parameter set $\boldsymbol{p}$, sampled by the stochastic engine. Prior to the pre-selection/conditioning, we assume a uniform distribution of each parameter between a minimum and a maximum value. These distributions reflect unbiased estimates within a range of plausibility based on expert knowledge.

The values in Table 6 are grouped by horizontal saturated hydraulic-conductivity values K_{xy}, anisotropy ratios $K_{\text{z}}/K_{\text{xy}}$, and the van Genuchten parameters α and N. The indices represent the hydrostratigraphic unit using the numbering scheme introduced in Section 10.1.1. In total, we consider 30 variable material properties (named #P1 to #P30), which is less than the number of units times the number of hydraulic properties ($12 \times 4 = 48$) because we chose some parameters to be identical in several geological units. The hydrostratigraphic units 1 to 6 share the same van Genuchten properties, and the units 7 and 8 do not require these unsaturated properties because the gravel aquifers of the Neckar and Ammer valleys are always fully water saturated. We do not treat the residual water saturations as random variables. Instead, we apply the following values in all model runs: $S_{\text{wr},1-8} = 5\,\%$, $S_{\text{wr},9} = 17\,\%$, $S_{\text{wr},10} = 18\,\%$, $S_{\text{wr},11-12} = 25\,\%$.

In total, we use nine random parameters (#B1 to #B9) related to boundary conditions, listed in Table 7. We again assume uniform priors within given bounds. Parameters #B1 to #B4 regulate the groundwater recharge R in $\text{m}\,\text{s}^{-1}$ on the four types of land use. Here we take the random recharge

Table 6: Prior parameter ranges of random material properties for the hydrostratigraphic units considered in the model.

ID	Name	Minimum	Maximum	Unit	Comment
#P1	$\log_{10} K_{xy,1}$	−8.0	−6.0	$\mathrm{m\,s^{-1}}$	
#P2	$K_{\mathrm{xy},2}$	$1/250 \cdot K_{xy,1}$	$1/2 \cdot K_{xy,1}$	$\mathrm{m\,s^{-1}}$	
#P3	$\log_{10} K_{xy,3}$	−9.0	−6.3	$\mathrm{m\,s^{-1}}$	
#P4	$K_{\mathrm{xy},4}$	$K_{\mathrm{xy},3}$	$10^3 \cdot K_{xy,3}$	$\mathrm{m\,s^{-1}}$	
#P5	$\log_{10} K_{xy,5}$	−8.3	−7.0	$\mathrm{m\,s^{-1}}$	
#P6	$\log_{10} K_{xy,6}$	−9.0	−3.0	$\mathrm{m\,s^{-1}}$	
#P7	$\log_{10} K_{xy,7}$	−5.3	−3.0	$\mathrm{m\,s^{-1}}$	
#P8	$\log_{10} K_{xy,8}$	−5.3	−3.0	$\mathrm{m\,s^{-1}}$	
#P9	$\log_{10} K_{xy,9}$	−10.0	−7.0	$\mathrm{m\,s^{-1}}$	
#P10	$\log_{10} K_{xy,10}$	−5.3	−3.0	$\mathrm{m\,s^{-1}}$	
#P11	$\log_{10} K_{xy,11}$	−9.0	−5.3	$\mathrm{m\,s^{-1}}$	
#P12	$\log_{10} K_{xy,12}$	−8.0	−3.0	$\mathrm{m\,s^{-1}}$	
#P13	$K_{z,1}/K_{xy,1}$	1/15	1	–	
	$K_{z,2}/K_{xy,2}$	1/15	1	–	coupled to #P13
#P14	$K_{z,3}/K_{xy,3}$	1/15	1	–	
#P15	$K_{z,4}/K_{xy,4}$	1/15	1	–	
#P16	$K_{z,5}/K_{xy,5}$	1/15	1	–	
#P17	$K_{z,6}/K_{xy,6}$	1/5	1	–	
#P18	$K_{z,7}/K_{xy,7}$	1/5	1	–	
#P19	$K_{z,8}/K_{xy,8}$	1/5	1	–	
#P20	$K_{z,9}/K_{xy,9}$	1/15	1	–	
#P21	$K_{z,10}/K_{xy,10}$	1/15	1	–	
#P22	$K_{z,11}/K_{xy,11}$	1/15	1	–	
#P23	$\alpha_{1\text{-}6}$	0.50	5.00	$\mathrm{m^{-1}}$	
#P24	α_9	0.01	0.10	$\mathrm{m^{-1}}$	
#P25	α_{10}	8.00	12.00	$\mathrm{m^{-1}}$	
#P26	α_{11}	0.50	0.70	$\mathrm{m^{-1}}$	
#P27	$N_{1\text{-}6}$	1.50	6.00	–	
#P28	N_9	1.40	1.70	–	
#P29	N_{10}	1.80	2.20	–	
#P30	N_{11}	1.50	2.10	–	

Table 7: Prior ranges of parameters describing boundary conditions of the model.

ID	Name	Minimum	Maximum	Unit	Comment
#B1	R_{cropland}	$1.5 \cdot 10^{-9}$	$8.0 \cdot 10^{-9}$	$\mathrm{m\,s^{-1}}$	
#B2	$R_{\mathrm{floodplain}}/R_{\mathrm{cropland}}$	0	1	–	coupled to #B1
#B3	$R_{\mathrm{mud/sandstone}}/R_{\mathrm{cropland}}$	0	1	–	coupled to #B1
#B4	$R_{\mathrm{urban}}/R_{\mathrm{cropland}}$	0.25	1	–	coupled to #B1
#B5	$\Delta h_{\mathrm{Neckar}}$	−0.50	0.50	m	
#B6	$\Delta h_{\mathrm{river}}$	−0.25	0.25	m	
#B7	$h_{\mathrm{Ammer,in}}$	346.0	347.0	m	
#B8	$h_{\mathrm{Ammer,out}} - h_{\mathrm{Ammer,in}}$	−8.6	−7.6	m	coupled to #B7
#B9	L_{11}	0.10	1.50	m	

Table 8: Prior ranges of structural parameters.

ID	Name	Minimum	Maximum	Unit	Comment
#S1	L_4	0	50	m	
#S2	size factor hollows	0.5	1.5	–	
#S3	bottom slope hollows	0.0	0.7	%	
#S4	switch hollows	−0.5	0.5	–	no hollows if < 0
#S5	switch riverbed	−0.5	0.5	–	no riverbed if < 0

rate R_{cropland} on undisturbed cropland as reference, which is reduced by random factors for the other land-use types (floodplain material, areas covered by mud-/sandstone, urban areas). The parameters #B5 to #B8 modify the fixed-head values at Dirichlet and leaky boundaries. The base values for the fixed heads used on the southern boundary in the Neckar valley (h_{Neckar}) and the stage of river Ammer (h_{Ammer}) vary in space. In the stochastic setup, we consider random constant shifts of Δh_{Neckar} and Δh_{Ammer} to all nodes belonging to the respective boundaries. The fixed-head values on the groundwater in- and outflow faces in the Ammer floodplain are spatially constant but uncertain, so that the stochastic model directly treats these values, $h_{\text{Ammer,in}}$ and $h_{\text{Ammer,out}}$, as random variables. We have chosen the ranges of these values from time series of hydraulic head measured in existing piezometers close to the boundaries. The last random boundary condition parameter, #B9, represents the uncertain thickness of the drainage boundary in Equation 2.29 for all floodplain elements. The respective hydraulic conductivity is $K_{xy,11}$. For the drainage boundaries outside of the floodplain, we assume a soil layer of 0.2 m thickness and a hydraulic conductivity of 10^{-6} m s^{-1}. The river boundary condition (see Section 2.2.4) uses K_{12} for its conductivity and the geometry parameters $L_{\text{riv}} = 40$ m, $w_{\text{riv}} = 3$ m and $L_{\text{sed}} = 0.5$ m. The river buffer zone shares the van Genuchten parameters of to the alluvial fines layer, and is assumed to be isotropic ($K_{z,12} = K_{xy,12}$).

Finally, we consider a total of five random parameters (#S1 to #S5) describing uncertain geometry of structural units. Table 8 lists the ranges of the parameters. #S1 controls the maximum depth L_4 of the weathered `kmGr` formation: Those parts of the Grabfeld formation that have a distance to the surface elevation smaller than L_4 are considered to be part of the weathered zone (fourth hydrostratigraphic unit). The parameters #S2 and #S3 describe the three-dimensional extent of the hillslope-hollows. #S2 controls the lateral extent of the hollows by expanding or contracting their width by a constant factor. #S3 defines the bottom slope of the hollows, which also controls their maximum depth. The total volume of the sixth hydrostratigraphic unit therefore depends on both, #S2 and #S3. The final two parameters, #S4 and #S5, are converted to binary flags, deciding whether the hillslope hollows (#S4) and explicit river beds (#5) are considered at all. Negative values of #S4 and #5 indicate that the respective features are not considered, whereas positive values lead to realizations including these features. We have introduced these switches because the existence and hydraulic relevance of these hydrogeological elements is uncertain at the real field site. A full parameter set $\boldsymbol{p}$ is the concatenation of all #P, #B and #S values.

10.2 Plausibility Criteria for Model Pre-Selection

We define seven criteria to decide whether the flow solution of a model realization is plausible (i.e., stage-2-accepted). These criteria are listed in the following:

1. To keep the realizations close to data observed in the field, the simulated hydraulic heads are compared to real head measurements obtained in the valleys (see Section 10.3). As the model assumes steady-state flow, we time-average the available series of measured heads at 51 observation wells and compute the RMSE of the corresponding simulated steady-state heads. For a model realization to be stage-2-accepted, its RMSE has to be smaller than 1.5 m. This reflects the order of magnitude of the measured annual fluctuations in hydraulic head, which are in the range of 0.5 m to 2 m.

2. The total groundwater flux Q_{in} crossing the fixed-head boundary at the western inflow end of the Ammer floodplain aquifers must be positive.

3. The total groundwater flux Q_{out} crossing the fixed-head boundary at the eastern outflow end of the Ammer floodplain aquifers must be negative.

4. The magnitude of the two groundwater boundary fluxes Q_{in} and Q_{out} should be similar. It is unclear which of the boundaries exhibits the larger groundwater discharge at the field site. Both scenarios are possible (an increase of discharge from inlet to outlet due to recharge and input from the hillslopes, as well as a decrease of discharge due to drainage towards the rivers and channels). Therefore, we only evaluate the ratio γ of the absolute flux difference over the mean flux:
$$\gamma = 2\frac{||Q_{\text{in}}| - |Q_{\text{out}}||}{|Q_{\text{in}}| + |Q_{\text{out}}|}. \tag{10.2}$$
This ratio can take values between $\gamma = 0$ (both fluxes are identical) and $\gamma = 2$ (one flux is zero). For a stage-2-accepted model realization, we require $\gamma \leq 1$, which is equivalent to requiring $1/3 \leq |Q_{\text{in}}|/|Q_{\text{out}}| \leq 3$.

5. The sum of all exchange fluxes between the subsurface and rivers must be negative (i.e., net groundwater discharge into rivers). Field data on the exchange fluxes are difficult to obtain because the change of river discharge due to surface-water/groundwater exchange is very small along the investigated stretch. Nonetheless, we expect that the rivers are net gaining as there are no groundwater abstractions within the domain. Losing conditions might occur only locally on short stretches of the rivers and channels.

6. A typical behavior shown in many models with randomly drawn parameters is extensive flooding of the model domain. At the real floodplain, by contrast, we do not observe permanent flooding outside of the drainage ditch network. To exclude flooding of the floodplain under steady-state flow conditions, we require that the total flux across all drainage nodes is small (see Section 10.1.2). As plausibility, we set that the total flux leaving at the surface must be smaller than 10 % of the total flux produced by the recharge boundaries.

7. Finally, the water flux leaving at the drainage ditches should not be excessive. In the real floodplain, these ditches carry water mostly seasonally and in small quantities. Since the actual fluxes are unknown and hard to estimate, we require a stage-2-accepted realization to drain less than 50 % of the recharged water through the ditches.

10.3 Tested Experimental Designs

Currently, there are 35 piezometers already installed at the field site, for which a decent-quality data set of hydraulic head in one or multiple depths is available. Figure 20 shows the location of these observation wells. Accounting for different depths in multi-level wells, hydraulic heads are measured at 51 points. However, there are no piezometers located on the hillslope between the two valleys. This lack of observation points results in high uncertainty regarding groundwater flow underneath the hillslope and in the location of the groundwater divide.

In order to fill this gap, the installation of up to three additional piezometers is planned on a transect. We identified twenty potential piezometer locations along this transect, coinciding with edges of the computational grid. These locations are also highlighted in Figure 20. The line of points extends longer on the north than the south, because we expect the divide to be shifted towards the north. This is so, because the northern valley is at a higher elevation than the southern valley, and also the geological units tend to partly dip towards the south-west. Furthermore, a preliminary study conducted by Kortunov (2018) also suggested a shift in this direction.

The optimal experimental design analysis considers designs consisting of one, two, or three new wells, each placed on one of the twenty locations. Our design space $\boldsymbol{D}$ consists of all possible permutations. The total number of possible designs n_{des} for 1, 2 and 3 locations out of a set of n_{pts} can be evaluated by:

$$n_{\text{des}} = n_{\text{pts}} + \frac{1}{2} n_{\text{pts}}(n_{\text{pts}} - 1) + \frac{1}{6} n_{\text{pts}}(n_{\text{pts}} - 1)(n_{\text{pts}} - 2), \tag{10.3}$$

in which n_{pts} is the number of potential observation points. With $n_{\text{pts}} = 20$, Equation 10.3 results in a total of $n_{\text{des}} = 20 + 190 + 1140 = 1350$ individual designs, out of which we need to identify the best one.

While the optimal three-well design will obviously outperform the optimal two- and one-well designs, we want to investigate which information gain (e.g., reduction in uncertainty of delineating the groundwater divide) is achieved by installing more or fewer wells. However, we do not perform a full cost-benefit analysis, as the (financial) costs are difficult to compare to the benefit of reducing the uncertainty in the groundwater-divide delineation.

11 Results & Discussion

Of 72 481 stage-1-accepted realizations, 20 600 needed to be rejected, because they yielded implausible results according to the given criteria. Another 1881 model runs were rejected, because they did not converge within 40 min of wall-clock time, set as limit to use the available computational resources efficiently. The remaining sample consists of $n_{\text{sample}} = 50\,000$ accepted realizations. Among the successful realizations, the model run times roughly followed a log-normal distribution with a mean of 20.7 min, a median of 19.5 min, and a standard deviation of 6.9 min (not shown here). Due to parallelization of up to 57 simultaneous model runs, the total wall-clock time required to obtain all realizations was approximately three weeks. For computational speed-up, we only used $n_{\text{sub}} = 10\,000$ realizations as virtual truths for the optimal design analysis. We checked the validity of this subset size by comparing the average binary fate maps of the whole sample and the subset. There were no significant deviations. For the PreDIA analysis, we assumed a standard deviation of measurement errors of $\sigma_{\text{meas}} = 0.05$ m.

11.1 Uncertainty and Sensitivity of Head Observations to Parameters

Figure 22 shows the distributions of the simulated groundwater-table measurements at the twenty proposed locations. For each suggested observation-well location we show (1) a violin plot of simulated head values of all 50 000 accepted sample members, (2) the median of the simulated head (h_{median}, light-gray dash markers), and (3) the position of the land surface (z_{surf}, black dash markers). The longitudinal distance is evaluated along the line connecting the proposed locations from south to north (i.e., zero corresponds to the first, southernmost investigated point).

At the southern end of the transect (close to the surface-water divide), the distributions of the groundwater table are very wide, whereas at the northern end in the Ammer floodplain they become quite narrow. This behavior can be explained with the plausibility constraints put onto the model selection. As Figure 20 shows, most existing observation wells are within the Ammer floodplain, restricting the variability of hydraulic heads by the first plausibility criterion. Also, excluding realizations showing extended flooding contributes to narrowing the variability of hydraulic heads within the floodplain. By contrast, there are no piezometers to constrain the models along the southern hillslope. Observation wells further away from the hydraulic-head-constraining floodplain show larger uncertainty than those close by, which reflects the uncertainty in groundwater recharge and transmissivity of the weathered part of the Grabfeld formation (`kmGr`). The conditioning by the pre-selection procedure might also explain the gradual change from near-Gaussian distributions for the northern wells to multi-modal wide distributions towards the southern end.

As indicated by the black dashes in Figure 22, the topography along the transect is not strictly monotonic. At about one quarter along the length of the profile, a hillslope hollow oriented in the WSW-ENE direction crosses the transect. Along the transect, the median of the simulated hydraulic head follows the topography to some extent, but with a much smaller range. At the

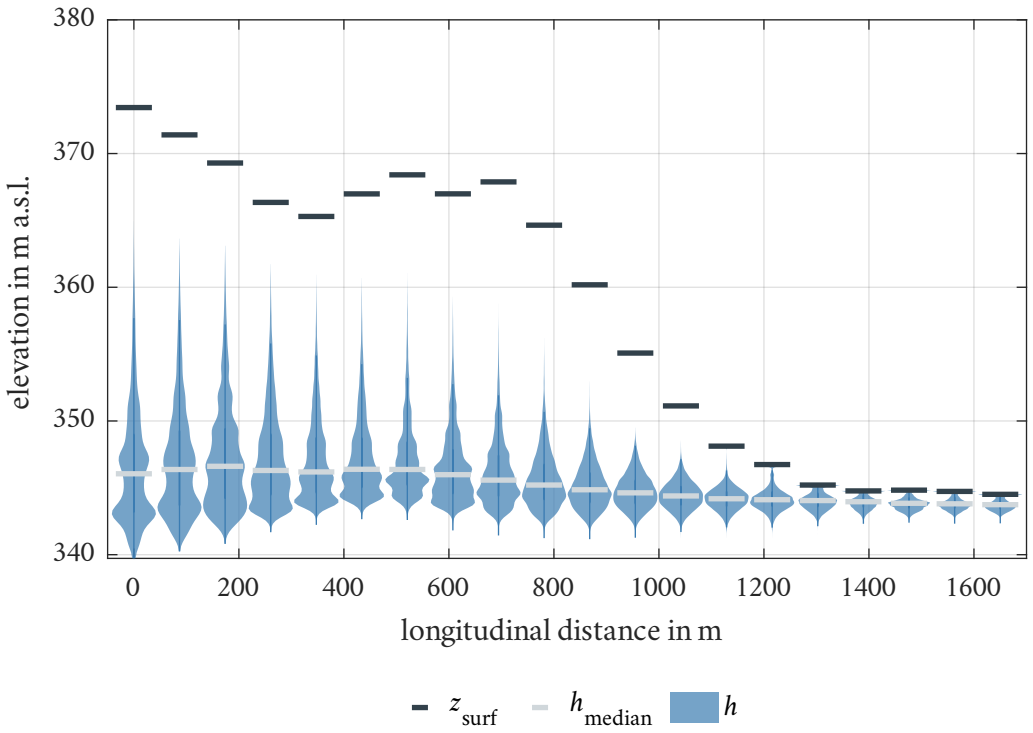

Figure 22: Distributions of virtual hydraulic-head observations using the sample of stage-2-accepted realizations at all twenty potential locations along the transect.

southern end, the median profile of hydraulic head drops towards the south along a distance of 200 m, whereas the surface elevation profile increases. The median groundwater table dipping towards the south of the transect might indicate that the groundwater divide is shifted towards the north, as hypothesized by Kortunov (2018). However, not all individual realizations show the same trend as the median, indicating that the general statement of Kortunov (2018) may be uncertain. This is why we performed the ensemble-based particle-tracking analysis to evaluate the location of the groundwater divide and its uncertainty in the following section.

To gain insights in how the head observations depend on the input parameters, we performed a global sensitivity analysis using the framework developed by Erdal et al. (2020), which applies the method of active subspaces (Constantine et al., 2014; Constantine and Diaz, 2017) supported by GPEs. The active-subspace method results in activity scores, expressing the relative importance of all input parameters for a selected target variable. We performed this analysis for the simulated hydraulic-heads at the 20 potential piezometer locations along the transect. At the 14 southern-most locations along the hillslope in the weathered Grabfeld formation, the activity scores were the highest for the conductivities in the unweathered and weathered Grabfeld formation, the thickness of the weathering layer, and the recharge rate of cropland. At the six northern-most locations, closer to/within the floodplain, we saw a shift towards conductivities of floodplain sediments and recharge in the floodplain. Comparable global sensitivity patterns have been obtained by Erdal and Cirpka (2019) in a study on a neighboring catchment with similar geology.

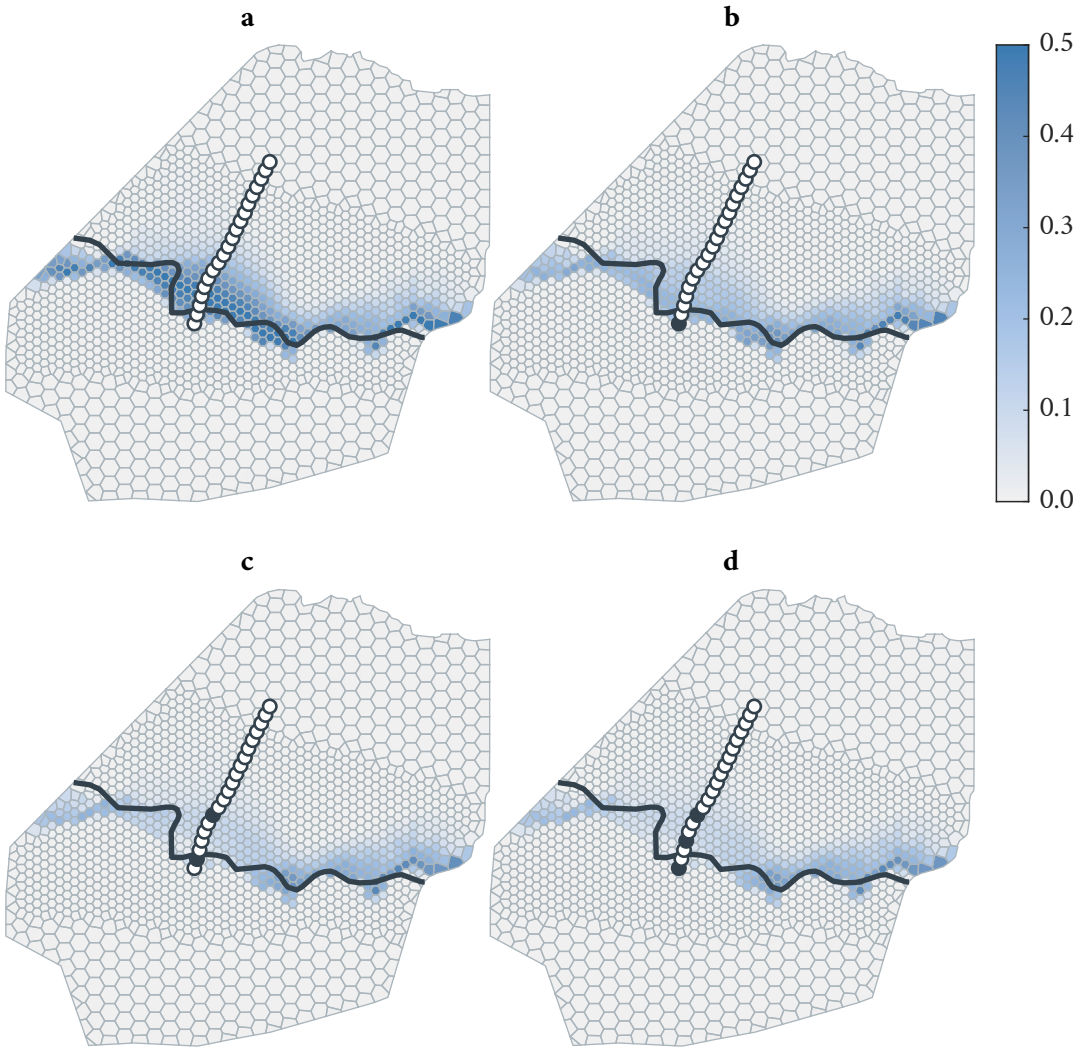

Figure 23: Maps of misclassification probability $P_{mc} = 2P \cdot (1 - P)$. **a**: Prior P_{mc}. **b**-**d**: P_{mc} for the best design with one, two and three additional piezometer(s). The best configuration is marked by black circles (unused locations shown as white circles). Surface water divide shown in black.

11.2 Maps of Misclassification Probability

Figure 23 shows maps of the misclassification probability P_{mc} according to Equation 9.5. It quantifies how likely it is that any point on the map is considered part of one subsurface catchment while belonging to the other one in reality. The 1526 polygons were constructed by Voronoi tessellation based on the set of starting points for particle tracking. The resolution is higher in a stripe within a few hundred meters north and south of the surface water divide, because we suspect the groundwater divide to be within this area. The colors of the polygons reflect the misclassification probability P_{mc} of a particle released in the center of the polygon, which ranges from 0.0 to 0.5.

Figure 23a shows the map prior to installing any new piezometers. The highest values of the misclassification probability occur close to the surface-water divide. On the Neckar (southern) side of the surface-water divide, the misclassification probability drops rapidly. Here, all model

realizations agree that these points belong to the Neckar subsurface catchment. On the Ammer (northern) side of the surface-water divide, by contrast, the misclassification probability decreases gradually, overall resulting in an uncertainty belt of the groundwater divide with a width ranging from 100 m to 800 m. This confirms the hypothesis of Kortunov (2018) that the groundwater divide might be shifted in this direction. At the foot of the hillslope within the Ammer valley, the misclassification probability is again practically zero, because these points belong to the Ammer subsurface catchment in almost all stage-2-accepted model realizations.

The width of the identified uncertainty belt is comparably small at the steeper hillslopes towards the east and at the very western end, where the topmost geological layer is the low conductive lumped sandstone formation (see Figure 21, layer 5). In contrast to that, the width is large on the gentle saddle in the western and middle parts of the domain, where the top subsurface-layer consists of the weathered Grabfeld formation, which has a higher hydraulic conductivity. This observation agrees with the findings of Haitjema and Mitchell-Bruker (2005), stating that groundwater and surface water divides are more likely to differ in aquifers with high transmissivities (for a given recharge rate and geometry). The transect of the proposed piezometer locations crosses the broadest part of the uncertainty zone perpendicular to the course of the belt. This is fortunate for the optimal experimental design, because we can acquire information just within the most uncertain region.

Figure 23b-d shows the maps of the misclassification probability after performing the optimal-experimental-design analysis for one, two, and three additional piezometers, respectively. Figure 23b reveals how the misclassification probability is expected to be reduced by placing a single additional piezometer. The optimal location is the southernmost point along the transect close to the surface-water divide. Unsurprisingly, the location of this piezometer coincides with the location that shows the highest uncertainty of hydraulic heads in Figure 22. A comparison between Figure 23a and 23b shows that the misclassification probability is not only reduced in the direct vicinity of the chosen new piezometer, but essentially over the entire width of the Wurmlingen saddle, whereas the effect at the eastern end of the model domain is negligible. This pattern reflects the smoothness of hydraulic heads, but is strongly affected by the assumption that each lithostratigraphic unit has a uniform set of hydraulic parameters (only the groundwater-recharge values are subdivided by land-use). The latter implies that conditioning the model on a single observation point in a particular unit, here the weathered Grabfeld formation, affects the model outcome at all other points within this unit. However, if we had considered internal variability within the units, individual head measurements would not have reduced the uncertainty at distant points within that unit to the same extent. Consistent to these arguments, the eastern end of the uncertainty belt (where the topmost geological unit is the lumped sandstone formation rather than the weathered Grabfeld formation) is not affected by placing a piezometer along the transect.

Placing a second piezometer at the northern fringe of the uncertainty belt reduces the misclassification probability further (Figure 23c), whereas the uncertainty pattern does not visually change when placing a third additional piezometer between the first and second piezometers (Figure 23d).

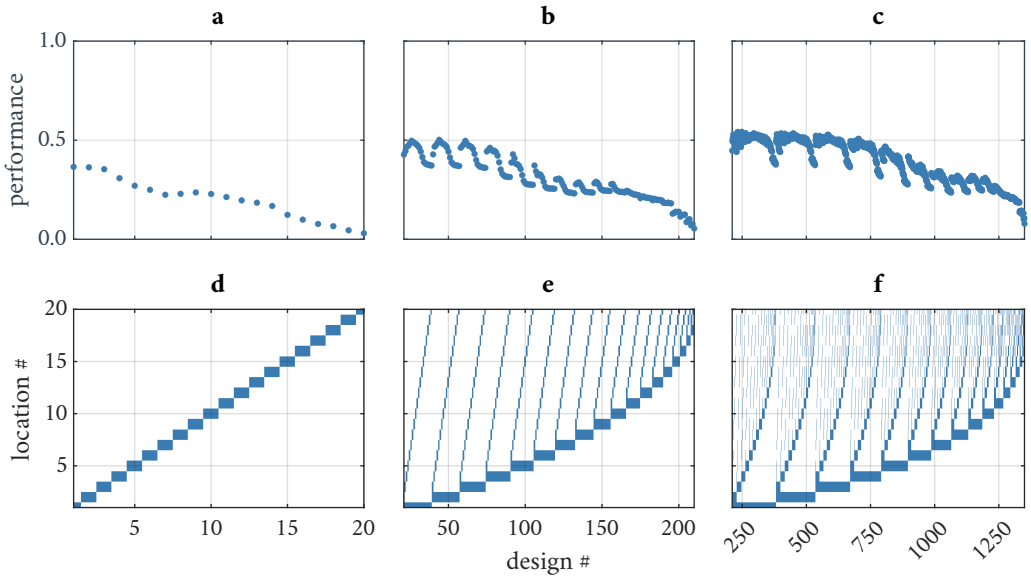

Figure 24: Performance of all 1350 investigated designs. **a**-**c**: Normalized utility function ϕ of the given design according to Equation 9.9. **d**-**f**: Piezometer combination of the given design.

11.3 Performance of Designs

Figure 24 summarizes the performance of all 1350 investigated piezometer configurations (grouped by one-, two- and three-additional-piezometer designs). All subfigures use the design number on the horizontal axis. In the following discussion, we use the notation "(1^{st} piez. | 2^{nd} piez. | 3^{rd} piez.)" to describe a given design, in which the numbers of the piezometer locations are sorted from south to north. Missing piezometers in the one- and two-piezometer designs are marked by a dash. The designs are numbered in the following way: The first twenty designs contain only one additional piezometer, ranging from (1|–|–) to (20|–|–). The designs 21 to 210 are two-piezometer designs, starting with the combination (1|2|–), incrementing the second location in steps of one to (1|20|–), then moving from (2|3|–) to (2|20|–) and so forth, until (19|20|–) is reached. In order to exclude replicates, the index of the second piezometer is always larger than that of the first. Finally, the designs 211 to 1350 start with (1|2|3) and increment the third location first, then the second, and then the first one, until reaching the final design (18|19|20). Again we avoid replicates by requiring that the piezometer indices increase from the first to the third piezometer within all designs. Figure 24d-f visualizes which piezometers the designs use.

Figure 24a-c shows the values of the utility function $\phi(\boldsymbol{d})$ of the given designs $\boldsymbol{d}$ according to Equation 9.9. It quantifies the expected relative reduction of the spatial mean of P_{mc} applying the measurement design $\boldsymbol{d}$. Theoretically, this metric can range between zero (no reduction of uncertainty at all) to one (perfect identification of the groundwater divide with the selected piezometer configuration for all tested truths).

In the single-piezometer designs (Figure 24a), the performance declines with increasing design number (placing the new piezometer further north along the transect). While the first three designs

result in a similar relative uncertainty reduction of about 36 %, $\phi(\boldsymbol{d})$ gradually decreases to a negligible low value of about 3 % at location 20. The optimal design is (1| − |−), resulting in a performance of $\phi = 36.6\,\%$. The best locations for placing a single piezometer coincide with the points at which the prior uncertainty of hydraulic head is the highest (see Figure 22), so that constraining the model by taking a single head measurement at these points yields the highest information gain. As the hydraulic heads at the northern end of the transect are already constrained by the plausibility criteria of the model pre-selection, additional piezometers hardly pay off here.

Figure 24b shows the performances of all two-piezometer designs. Like in the one-piezometer designs, configurations including southern piezometer locations (design numbers 21 to ca. 100) perform better than other designs. For a given first piezometer location, the performance depends on the distance between the two piezometers. At least for the well-performing designs 21 to 100, the optimal distance between the two piezometers is on the order of several hundred meters. Such a configuration performs better than designs in which the two new piezometers are further apart or closer to each other. The best two-piezometer configuration is (2|7|−), leading to an uncertainty reduction of $\phi = 50.2\,\%$. The optimal two-piezometer designs may be explained by the combined effects of having the highest prior uncertainty of hydraulic head at the southern end of the transect (discussed in the context of the one-piezometer designs) and the inherent spatial correlation of hydraulic head caused by the groundwater-flow equation itself: One piezometer should be located at the most informative southern end; placing two piezometers to close to each other would yield redundant information (and observing a small head distance would drown in the measurement error), while placing the second piezometer at the northern end would be of little use because here the hydraulic heads are already constrained by the plausibility criteria.

In the three-piezometer designs (Figure 24c), this pattern is maintained, with the best location of the third piezometer being in the middle of the other two new observation wells. Thus, placing the third well further north, where the head-uncertainty is low, is less beneficial than refining the spatial resolution of head measurements in the southern third of the transect. The best three-piezometer configuration is (1|7|15) with $\phi = 54.2\,\%$, which is not significantly better than the best two-piezometer configuration. We conjecture that adding a fourth piezometer along the transect would yield an even lower increase of performance. Thus, in a practical application, it might be better to invest the money needed to install such a well in other investigations like elaborate well tests, or in entirely different locations (see Section 11.4).

As a quality check, we determined the average effective sample size for the three optimal designs. The values are comparably large ($\mathrm{AESS}_1 = 859.7$, $\mathrm{AESS}_2 = 179.7$ and $\mathrm{AESS}_3 = 68.1$), which means the sample of $n_{\mathrm{sample}} = 50\,000$ was large enough to make reliable statements about the results.

Notably, all three optimal designs use very similar locations. Each larger optimal configuration basically includes the smaller ones as a subset (with the exception of switching between locations 2 and 1 in the two-location design). This means that, in the given application, one could decide

whether and where to install the next observation well after installing the preceding ones, yielding essentially the same optimal designs. Such behavior is beneficial from a practical standpoint of view as, in real-world applications, the decision about extending a measurement network is often made only after realizing that the existing network is not (yet) sufficient. However, we cannot generalize that such a behavior occurs in all cases. In other applications, the optimal designs of many piezometers may not be a superset of the designs with fewer piezometers. Also, the information gained by the actual data value obtained by a first well could change the current state of knowledge, hence leading to (slightly) different later design decisions (Geiges et al., 2015). In such cases, deciding on the number of observation wells would be necessary ahead of the first drilling in order to achieve optimal results.

We may compare the performance of the optimal designs with those of intuitive choices using the same number of new piezometers. When installing a single piezometer, one might place it on the middle of the transect using the design (10|−|−). The uncertainty reduction of this particular design is $\phi = 22.9$ %, which is considerably smaller than the optimal performance of $\phi = 36.6$ %. When placing installing two piezometers, one could either maximize the distance along the full transect with design (1|20|−) or subdivide the transect into three similarly long sections with the design (7|14|−). The performances of these scenarios are $\phi = 37.2$ % and $\phi = 25.1$ %, respectively, while the best two-piezometer design achieved $\phi = 50.2$ %. Actually, the best single-piezometer design performs almost as good as the intuitive two-piezometer design taken the two end points of the transect, and is considerably better than the intuitive design using identical section lengths. Finally, intuitive choices for the three-piezometer designs would be design (1|10|20), which includes the two end points of the transect, and design (5|10|15), subdividing the transect into sections of similar length. The respective uncertainty reductions are $\phi = 50.5$ % and $\phi = 42.0$ % compared to a reduction of $\phi = 54.2$ % obtained by the optimal design. These calculations exemplify the benefit of an optimal-design-evaluation over intuitive choices.

11.4 Designs With the Third Piezometer Being Placed off the Transect

As shown in Figure 23, installing new piezometers along the suggested transect reduces the misclassification probability $P_{\mathrm{mc}}(\boldsymbol{x})$ on the hillslope parallel to the transect, but hardly affects $P_{\mathrm{mc}}(\boldsymbol{x})$ at the eastern end of the uncertainty belt. This part of the high-uncertainty belt is covered by the lumped sandstone formation. Therefore, this uncertainty depends on the hydraulic properties and groundwater recharge of this hydrostratigraphic unit, and can only be reduced by observations that are sensitive to these properties. Because installing a third piezometer along the transect does not reduce $P_{\mathrm{mc}}(\boldsymbol{x})$ in this zone, the difference between the two- and three-piezometer designs is rather small. We thus hypothesize that placing a third piezometer somewhere else would yield a better performance. We tested this hypothesis by defining an alternative design space: we keep the best two piezometer locations along the transect fixed and then allow the third piezometer to be placed at any node of the two-dimensional computational grid. This results in 2067 additional designs.

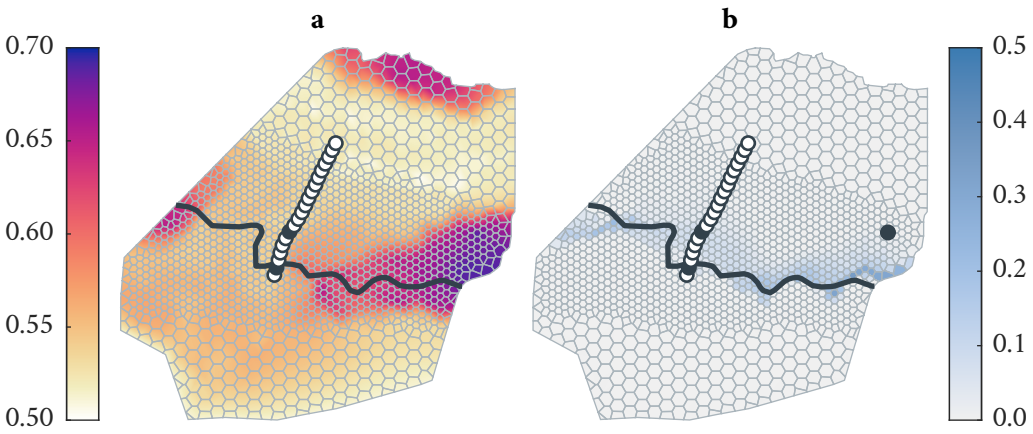

Figure 25: Results with the third piezometer being placed off the transect. **a**: Design performance ϕ as a function of the third observation location. **b**: Misclassification probability P_{mc} for the optimal three-piezometer design with one piezometer not being restricted to the transect.

Figure 25a shows the performance ϕ as a function of the location of the third piezometer. The maximum performance of $\phi = 69.3\,\%$ is obtained by placing it in the eastern part of the domain, roughly 400 m north of the highest-uncertainty region remaining after installing two piezometers (see Figure 23c). This point is located in a hillslope hollow (see Figure 20) that collects groundwater recharged in the lumped sandstone unit. The corresponding hydraulic head is sensitive to the hydraulic properties and groundwater recharge of the lumped sandstone formation, which affects $P_{\text{mc}}(\boldsymbol{x})$ in the eastern section of the uncertainty belt. The latter is confirmed by Figure 25b, which shows the resulting map of misclassification probability $P_{\text{mc}}(\boldsymbol{x})$ for this design. This indicates that the new location of the third piezometer indeed reduces $P_{\text{mc}}(\boldsymbol{x})$ in the eastern section of the uncertainty belt, which was hardly influenced by installing wells exclusively along the transect.

The average effective sample size of the optimal design in this substudy is comparably low ($\text{AESS}_3^* = 4.4$). This drop is caused by the large information gain by the freely moving third well, so that only few realizations achieve significant likelihoods when compared to the hypothetical data values. Given this low number, a larger sample would be necessary to validate the statistical significance of the interpretations. However, given the high computational costs and because this is only a substudy beyond our actual objectives, we refrain from doing so.

Figure 25a includes an interesting and instructive artifact of the model: According to our model, hydraulic-head measurements on the northern hillslope appear to be beneficial for delineating the groundwater divide at the southern boundary of the Ammer valley. Most likely this is caused by the assumed uniformity of hydraulic parameters within each lithostratigraphic unit. In the very north of the model domain, the lumped sandstone formation crops out, implying the same values of hydraulic conductivity and groundwater recharge as in the zone of interest at the southern boundary. Thus, a hydraulic-head measurement within this northern zone constrains model parameters of the lumped sandstone unit, reducing the misclassification probability in the eastern part of the uncertainty belt. However, we are doubtful that this would be confirmed in a real-world application.

11.5 Strengths and Limitations of the Framework

Our framework is easily adaptable to other cases and applications, with the underlying groundwater-flow model being trivially exchangeable. This flexibility makes it convenient to apply the presented technique to other sites. Both interfaces, from the stochastic sampler to the numerical model, and from the numerical model to the optimal experimental design analysis, require only basic input/output operations of parameter values and virtual observations. While we have implemented the stochastic sampler and PreDIA as Matlab scripts, the approach could easily be transferred to other programming environments. However, a particle tracking tool is a necessary requirement for our framework to work.

Among the most labor-intensive parts of the framework is the initial model development, which is needed in quantitative hydrogeological consultancy anyway. Computationally, the creation of the plausible sample is the most costly step, but this can be parallelized effectively. To obtain reasonable uncertainty estimates, several thousand model realizations are needed. This may not be affordable to everybody who might be interested in the uncertainty of groundwater-divide delineation. Such computer-time limitations may be overcome by cloud computing.

In practical applications, the costs related to elaborate modeling in the planning phase of a new observation-well needs to be compared to the other expenses. This includes filing the application for legal approval, advertising for bids, planning of the fieldwork, and the drilling and completion expenses themselves. If the presented optimal-experimental-design method is initiated at the beginning of this process, it becomes an integral part of the decision-making process of how many new piezometers to install and where to place them.

The way we use the chosen optimal-design method PreDIA, we can only rank experimental designs within a given finite set. The number of elements in this set determines the computational costs of the optimal-design part of the analysis. In our application, we confined the design space by restricting the piezometer locations to a transect, reflecting the legal constraints at the given field site. With three piezometers at twenty potential locations, we had to consider 1350 configurations. In the additional study presented in Section 11.4, we removed the constraint to stay on the transect for one piezometer, considering 2067 potential locations. Allowing all three piezometers to be placed at any of these 2067 locations, would have resulted in more than $1.4 \cdot 10^9$ designs (see Equation 10.3), which is computationally prohibitive. Tackling such a problem would need to involve an optimization algorithm around PreDIA to iteratively find a best-performing design without exhaustively testing all of them.

Our application was restricted to steady-state flow. Of course, real flow systems are never fully stationary, since they are always subject to transient forcings. Depending on the investigated site, this can include climatic influences, weather, tides or anthropogenic impacts (e.g., drinking water supply wells). All these processes could affect the position of groundwater divides (e.g., Rodriguez-Pretelin and Nowak, 2018). In cases where the expected movement of the groundwater flow divide

over time is the main research question obviously need to account for this. Characteristics of such systems might be a significant abstraction of groundwater due to pumping wells, a known imbalance of the groundwater flow field or severe temporal fluctuations in groundwater recharge (e.g., Sanz et al., 2009). An interesting extension of our framework would be a transient analysis for such systems, by using transient simulations and time-dependent observations. Consequently, the underlying objective function would need to be redefined. We provided a possible extension towards dynamic systems in Section 9.5. However, the higher uncertainties related to inherently more complex transient models would require a larger sample and would most likely deteriorate the performance of the pre-selection method. In the context of transient data and models, a follow-up project could combine optimal experimental design techniques with data-assimilation methods, but this is beyond the scope of the present study.

For most cases where the divide is suspected to be shifted but not dramatically moving over time, our steady-state framework is applicable (with the interpretation of the steady-state as a "most representative state"). We also want to highlight that the goal of our framework is not to derive the position of groundwater divides themselves. Instead, we want to identify those locations that are best suited to conduct measurements providing insight for this delineation. The actual delineation, for example, can then be carried out by calibrating a groundwater flow model to the obtained measurement data. This second model can be more detailed, more finely discretized and even transient, as probably fewer model runs are necessary. Prior to the calibration, a rigorous grid convergence analysis should be performed to validate the numerical accuracy of the model.

As with every model, the performance of the method depends on the validity of underlying assumptions. In particular, we have assumed that the hydraulic parameters are uniform within each lithostratigraphic unit and that groundwater recharge is spatially uniform in zones defined by the topmost geological layer and land-use. Neglecting spatial variability within these zones expands the spatial ranges over which intended measurements are informative. We may also have missed discrete features altogether, which could affect the position of the groundwater divide but do not influence the existing measurements. The latter would lead to a systematic bias.

The optimal-experimental-design method chosen in this study can accommodate any kind of uncertain parameters or uncertain model choices, provided that a prior uncertainty range is given. Both, identifying the sources of uncertainty and defining the related prior distributions require expert knowledge, thus questioning the objectivity of the analysis. However, as with all Bayesian methods, such choices are at least made transparent. We have made good experience by initially setting fairly wide prior parameter ranges and then constraining the parameter space to behavioral models by the GPE-supported pre-selection method (Erdal et al., 2020).

In the given application, we restricted the observations to hydraulic-head measurements, but this is not a limitation of the method. It is easy to augment the virtual observation vector by other data (e.g., hydraulic tests to be performed using the new observation wells, like borehole dilution or

tracer tests). Like with the extension to transient flow, the consideration of additional data types may also require more (uncertain) parameters. Systematically analyzing which type of data is most informative for which type of question is an ongoing issue of stochastic subsurface hydrology and optimal experimental design beyond the scope of the current study.

12 Conclusions & Outlook

In this work we have presented a framework to identify the best piezometer configuration from a set of possible layouts to delineate local groundwater divides. Through the combination of filtered ensemble-based modeling of steady-state subsurface flow, particle tracking, and the application of the optimal-experimental-design technique PreDIA (Leube et al., 2012), we could identify the piezometer configuration for which we expect the largest reduction in the uncertainty of the groundwater divide. We have applied the method to an appropriate case study, which revealed the following insights:

1. Configurations involving new measurement locations that are far away from existing ones perform better, because then the variability of hydraulic head, consistent with the existing data, is higher.

2. In our application, a medium spacing of a few hundred meters between multiple new piezometers was optimal. Closer points would have led to redundant information due to the spatial auto-correlation of hydraulic head. Larger distances would have pushed observation points into non-informative regions close to existing measurements.

3. The designs, defined as optimal by the presented framework, perform better than intuitive equidistant piezometer placements. In fact, the identified optimal design for a single piezometer provides similar information content as the tested intuitive equidistant placing of two piezometers, implying significant savings in real-world applications.

4. Additional information obtained by adding more piezometers leads to further reduction of uncertainty, but the additional gain of information decreases with each new piezometer.

5. Our procedure may be used to estimate whether the additional information gain is worth the effort of installing an additional observation well. The actual decision depends on the case at hand and involves a tradeoff between desired certainty and available resources. For us, sequential optimization of one piezometer location after the other led to practically the same designs as jointly optimizing multiple piezometer designs, but this observation cannot be generalized.

A worthwhile follow-up study would be the extension of the presented framework to transient flow systems.

Chapter IV

Proxy-Model Assisted Calibration of a Steady-State Subsurface Flow Model

Context

At the time of writing this dissertation, this chapter is currently being prepared for publication. The author contributions are: Jonas Allgeier set up the numerical flow model, implemented the proxy-model assisted calibration routine, performed the computations, created the figures, and wrote the draft manuscript; Olaf A. Cirpka conceived the presented idea, supervised the work, provided funding, and revised the manuscript draft. We acknowledge and appreciate the suggestion of Alexandra Gessner to try Simulation-Based Inference for our problem and the technical help of Michael Deistler in this regard.

The model source files, the calibration code, the raw data supporting the conclusions of this study and all Matlab codes used to generate the figures are publicly accessible in form of a repository at https://osf.io/uptxd/ (Allgeier, 2022b).

13 Introduction

Process-based numerical modeling of subsurface flow is an important tool for hydrogeological research. The computational power (in terms of hardware and software) has improved drastically over the last decades. At the same time, models have increased in complexity, as new numerical methods have become feasible, more processes could be considered and evermore detail could be represented in the models (not only within the hydrogeologic community; Venkataraman and Haftka, 2004; Zhou and Li, 2011; Jakob, 2014). A side effect of this development is that modern models tend to have many tunable parameters. Finding parameter sets that make model outputs match observed measurement data is the goal of model calibration.

In manual calibration, the modeler tests different parameter sets until a satisfactory agreement with the observations is achieved. Albeit still being used regularly (e.g., von Gunten et al., 2014), it is generally regarded as tedious, inefficient, not reproducible, non-transparent, and unable to truly find optimal parameter sets (Carrera et al., 2005; Beckers et al., 2020). Consequently, automated calibration procedures have become more popular with the rise of computational abilities (Yeh, 1986; Solomatine et al., 1999).

Automated calibration schemes aim to minimize a scalar metric quantifying the differences between simulations and observations (i.e., an objective function). Gradient-based methods, like the Gauß-

Newton method and its descendants (e.g., the Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963) or trust-region reflective methods (Powell, 1970a,b; Conn et al., 2000)) are comparably efficient in finding a minimum of the objective function. Unfortunately, they can neither guarantee that this is the global minimum, nor do they provide reliable uncertainty estimates if the functional dependence between model parameters and observables is nonlinear. Ensemble-based methods, like genetic algorithms (Goldberg, 1989; Gen and Cheng, 1999; Das and Suganthan, 2011) and Markov Chain Monte Carlo (MCMC) methods (Gilks et al., 1995; Brooks et al., 2011) are better in finding the global minimum and may provide a good approximation of the parameter distribution conditioned on the measurements, but they require orders of magnitude more model runs than parameters, which can be prohibitive for computationally expensive models with long run times.

For these reasons, a special branch of calibration research tries to develop efficient global calibration schemes based on proxy-models (Haftka et al., 2016). The proxy-models are used to select the most promising point(s) in parameter space for an evaluation with the full model. After evaluation, the proxy-model is retrained, new points are proposed and the next iteration starts. The variants of such proxy-model-assisted calibration schemes differ mostly in the utilized proxy-model, the mechanism of proposing points, and the criteria to select points for model evaluation.

One of the earliest examples making use of this procedure is the Efficient Global Optimization algorithm developed by Jones et al. (1998). In each iteration it identifies a single promising point by optimization of the "expected improvement" metric, which balances the predicted objective function value with the estimated uncertainty. Regis and Shoemaker (2007) presented a more recent example of a proxy-model-assisted global optimization tool, where a single point is selected from a randomly generated set of points by evaluating a metric based on predicted objective function value and distance to all previous points. Regis and Shoemaker (2009) provided an expansion of this method by introducing a parallelized approach. In each iteration, multiple points are iteratively selected from a set of randomly proposed points. All points are then evaluated in parallel using a high-performance computing cluster. Wang and Shoemaker (2014) and Xia et al. (2021) extended this general concept by more sophisticated point proposals, where the number of tweakable parameters is restricted over the course of the calibration. In this study, we propose and apply another variant of proxy-model assisted calibration tailored towards our specific modeling case at hand on the basis of the scheme developed by Regis and Shoemaker (2009). We extend the latter scheme to account for model plausibility, such that the calibration results in a parameter set that produces a plausible model realization.

Global model calibration is associated with finding a single optimal parameter set for a given model and a set of observations. In many cases, especially for models with numerous parameters, there are different points in parameter space that can produce similar or even identical results, which makes the problem of global calibration ill-posed (i.e., a unique solution does not exist; Zhou et al., 2014). This problem is known as equifinality (Beven, 2006). Stochastic calibration methods try to counteract equifinality problems by characterizing the uncertainty of calibrated

parameters, or by even finding a multi-dimensional distribution of points that create acceptable model results. Traditionally, stochastic calibration can be performed by Bayesian methods based on the likelihood approach (e.g., Mohammadi et al., 2018; Beckers et al., 2020). MCMC methods represent a particularly popular example in the Bayesian toolbox. However, these likelihood-based methods also suffer from the fact that many model realizations are necessary, which is often not affordable.

In this study, we circumvent this problem in two different ways to still construct full posterior parameter distributions for our calibration problem. First, we define a proxy-model from those runs of the original model that were conducted during the global calibration. We then apply MCMC sampling not to the original model, but the much faster proxy variant. As an alternative, we generate a full posterior parameter distribution through a likelihood-free method of the field of Simulation-Based Inference (SBI) (Cranmer et al., 2020; Tejero-Cantero et al., 2020; Lueckmann et al., 2021), namely Neural Posterior Estimation (NPE). This tool is based on machine learning and requires only a collection of model inputs and outputs sampled from the prior distribution.

We compare the two resulting posterior distributions with each other and to the results of the optimization-based global model calibration. This allows us (1) to gain insights beyond a best-estimate parameter set, like the assessment of parametric uncertainty and correlation information, and (2) to assess which posterior construction method is more appropriate/reliable for our problem.

Section 14 introduces our methodology and the model we consider. Then we present and discuss the calibration results in Section 15. Finally we summarize our conclusions and give suggestions for further research in Section 16.

14 Methods

We first introduce our notation and terminology. Afterwards we point out challenges that are associated with calibration algorithms concerning models with long run times and how we approach them. Then we present the methods we use for the inference of full posterior parameter distributions (SBI/NPE and MCMC). Finally, we introduce the model that is to be calibrated.

14.1 Calibration Terminology and Notation

Parameters A parameter is a variable model input that is allowed to change between different model realizations. In hydrogeologic models it may internally affect material properties, boundary conditions, or the geometry of geological features. A change in a parameter's value usually also leads to a different model output.

The particular parameter value p of a single realization is a scalar quantity, typically symbolizing a physical or mathematical property. It therefore usually consists of a value and a dimension (with a unit), but it can also be dimensionless. A model typically depends on multiple parameters $p_1, \dots, p_d$,

where d is the *dimension* of the calibration problem (i.e., the total number of parameters). For a more compact notation of all parameters of a model realization, we introduce the parameter vector $\boldsymbol{p} = [p_1, \dots, p_d]$. A parameter set $\boldsymbol{p}$ can also be referred to as a point in *parameter space*.

Working directly with the "physical" parameter values of the model has disadvantages that we will discuss later on. We therefore introduce parameter transformations. The first one converts a physical parameter vector $\boldsymbol{p}$ to a normalized vector $\boldsymbol{p}_{\text{cdf}}$ of the same size with dimensionless entries ranging from 0 to 1. We denote this transformation $\boldsymbol{f}_{\text{cdf}}$ and its inverse $\boldsymbol{f}_{\text{cdf}}^{-1}$, implying that $\boldsymbol{f}_{\text{cdf}}$ can be interpreted as the vector of prior cumulative probabilities of all individual parameters:

$$\boldsymbol{p}_{\text{cdf}} = \boldsymbol{f}_{\text{cdf}}(\boldsymbol{p}) \tag{14.1}$$

$$\boldsymbol{p} = \boldsymbol{f}_{\text{cdf}}^{-1}(\boldsymbol{p}_{\text{cdf}}). \tag{14.2}$$

We introduce a second transformation converting the normalized parameter values $\boldsymbol{p}_{\text{cdf}}$ to re-scaled values $\tilde{\boldsymbol{p}}$, with dimensionless entries ranging from $-\infty$ to $+\infty$. For forward and backward transformation, we again define vector-valued functions:

$$\tilde{\boldsymbol{p}} = \boldsymbol{f}_{\text{scl}}(\boldsymbol{p}_{\text{cdf}}) \tag{14.3}$$

$$\boldsymbol{p}_{\text{cdf}} = \boldsymbol{f}_{\text{scl}}^{-1}(\tilde{\boldsymbol{p}}). \tag{14.4}$$

Note that for the specific case of $\tilde{\boldsymbol{p}}$ following a standard normal distribution, the combination of the two transformations applied here is known as normal-score transformation, or *Gaussian anamorphosis* (Wackernagel, 2003; Everitt and Skrondal, 2010). In summary, we have three equivalent and effortlessly convertible ways to express parameter sets. We use $\tilde{\boldsymbol{p}}$ for calibration-internal calculations and transform these via $\boldsymbol{p}_{\text{cdf}}$ to $\boldsymbol{p}$, whenever a full model run is desired.

Model We denote the full model itself $\mathcal{M}$. A model realization with a specific parameter set $\boldsymbol{p}$ results in the model outcome:

$$\{\boldsymbol{\vartheta}^\bullet, \boldsymbol{\varphi}^\bullet\} = \mathcal{M}(\boldsymbol{p}), \tag{14.5}$$

which is split into two parts. The first one contains all n_{obj} quantities needed for the evaluation of an objective function (explained in the following): $\boldsymbol{\vartheta}^\bullet = [\vartheta_1^\bullet, \dots, \vartheta_{n_{\text{obj}}}^\bullet]$. The second one consists of all n_{plaus} quantities necessary for a plausibility assessment (also explained in the following): $\boldsymbol{\varphi}^\bullet = [\varphi_1^\bullet, \dots, \varphi_{n_{\text{plaus}}}^\bullet]$. These two sets of model observations can intersect and could even be identical. We just use a formal distinction for clarity. We use the symbol "$\bullet$" to indicate the output of a full model run (in contrast to proxy-model outputs introduced in the following).

Proxy-Model We use $\mathcal{P}$ to denote the proxy-model that is used for accelerating the calibration. In our case, the proxy-model is constructed (i.e., trained) from a set of known pairs of model input (in terms of $\tilde{\boldsymbol{p}}$) and model output (in terms of $\{\boldsymbol{\vartheta}^\bullet, \boldsymbol{\varphi}^\bullet\}$). The trained proxy-model can provide a predicted model outcome for a specific re-scaled parameter vector $\tilde{\boldsymbol{p}}$.

This results in a predicted outcome, which ideally approximates the outcome of the full model:

$$\{\boldsymbol{\vartheta}^{\circ}, \boldsymbol{\varphi}^{\circ}\} = \mathcal{P}(\tilde{\boldsymbol{p}}) \approx \mathcal{M}(\boldsymbol{p}) = \{\boldsymbol{\vartheta}^{\bullet}, \boldsymbol{\varphi}^{\bullet}\}, \tag{14.6}$$

where we use $\boldsymbol{p} = \boldsymbol{f}_{\text{cdf}}^{-1}(\boldsymbol{f}_{\text{scl}}^{-1}(\tilde{\boldsymbol{p}}))$. In analogy to the full model, the predicted model outcome consists of objective-function quantities ($\boldsymbol{\vartheta}^{\circ} = [\vartheta_1^{\circ}, \dots, \vartheta_{n_{\text{obj}}}^{\circ}]$) and plausibility-function quantities ($\boldsymbol{\varphi}^{\circ} = [\varphi_1^{\circ}, \dots, \varphi_{n_{\text{plaus}}}^{\circ}]$). The symbol "$\circ$" denotes proxy-model predictions.

Objective Function The objective function f_{obj} compares the $\boldsymbol{\vartheta}$-part of the model output (which could be a predicted model outcome $\boldsymbol{\vartheta}^{\circ}$ or a full model output $\boldsymbol{\vartheta}^{\bullet}$) to a target data set $\boldsymbol{\vartheta}^{*}$ and provides a scalar quantity of agreement y:

$$y_{\boldsymbol{p}} = f_{\text{obj}}(\boldsymbol{\vartheta}^{*}, \boldsymbol{\vartheta}). \tag{14.7}$$

Within the scope of this study, we assume that smaller numbers of y indicate a better fit between model output and target data. The calibration therefore aims to minimize y:

$$\boldsymbol{p}_{\text{best}} = \arg\min_{\boldsymbol{p}} y_{\boldsymbol{p}}, \tag{14.8}$$

where $\boldsymbol{p}_{\text{best}}$ is the best parameter vector found (ideally the global optimum).

Plausibility Function Similar to the objective function, we define a plausibility function f_{plaus} that takes the $\boldsymbol{\varphi}$-part of the model output (which could be $\boldsymbol{\varphi}^{\circ}$ or $\boldsymbol{\varphi}^{\bullet}$) and assigns a scalar quantity of plausibility based on some internal rules:

$$z = f_{\text{plaus}}(\boldsymbol{\varphi}), \tag{14.9}$$

where z is a plausibility score ranging from zero (fully implausible) to one (completely plausible). A detailed description of how we define and use the concept of plausibility is given in Section 14.2.

14.2 Calibration Scheme Challenges

A viable calibration scheme faces several challenges, especially if confronted with models that require a long run time. In the following, we will provide a summary of all problems we encountered when developing our calibration scheme and of the solutions or workarounds that we applied.

Diversity of Model Parameters & Prior Definitions A common, but little talked about problem in model calibration is the diversity within the set of all physical input parameters $\boldsymbol{p}$. For instance, some parameters might describe hydraulic conductivities, others might be related to recharge rates, a third group might define boundary conditions and a last set might describe the geometry of features (e.g., thickness, dip and strike of geological layers, or offsets at faults). Simply summarizing all of these in a single vector $\boldsymbol{p}$ can be problematic for several reasons.

First of all, not all parameters have the same physical dimension. While hydraulic conductivities and recharge rates are in principle compatible (both are expressed in $\mathsf{L}\,\mathsf{T}^{-1}$), the comparison of the respective values is hardly meaningful. Others, like a fixed-head value (expressed in L) and a fixed flux (expressed in $\mathsf{L}^3\,\mathsf{T}^{-1}$) might not even be compatible to each other. The most primitive way to circumvent this problem would be to divide each parameter by its unit, thus creating dimensionless quantities. However, this remedy would arbitrarily depend on the chosen unit.

In fact, there is another problem with directly using $\boldsymbol{p}$ for calibration: the allowable parameter ranges (i.e., support intervals). While some parameters can in theory assume any real value, others might have to be positive (e.g., hydraulic conductivities) and a final group has to stay within a fixed interval (e.g., porosities ranging between zero and unity). Conducting the calibration on $\boldsymbol{p}$ would result in a parameter space with complicated boundaries, where it is not trivial to ensure that newly proposed points are within the valid limits.

A more common remedy to this problem is to normalize each parameter by a relative value between a minimum and a maximum number (e.g., see Chapter III). This way, all parameters can be summarized by a dimensionless number between zero and one and the parameter space becomes a d-dimensional hypercube. We believe this is not the best solution, because it still requires checking validity of points proposed by the calibration scheme with respect to being inside the hypercube. When restricting all parameters to bounded intervals of finite values, it is also sometimes unclear how to set these boundaries, especially for parameters that are in principle unbounded.

An approach that can account for both bounded and unbounded parameter ranges is based on defining a probability density function for each input parameter and operate the calibration in the space of cumulative probabilities ($\boldsymbol{p}_{\mathrm{cdf}}$). For individual model runs, a parameter vector is back-transformed to physical space, where it can assume all values that are considered reasonable, or even physically possible without the necessity of defining bounded intervals (uniform distributions on a bounded interval are possible too, of course). The probability density function and the associated cumulative distribution function do not need to be based on analytical expressions (e.g., the normal, Beta, Gamma or log-normal distribution). They could even be of empirical nature (modern programming/scripting languages allow effortless construction and sampling of such).

We go one step further, by introducing a second transformation step via the inverse logit transform, to re-scale the cumulative distribution values onto the unbounded interval $(-\infty, \infty)$:

$$\tilde{p} = f_{\mathrm{scl}}(p_{\mathrm{cdf}}) = \log\left(\frac{p_{\mathrm{cdf}}}{1 - p_{\mathrm{cdf}}}\right). \tag{14.10}$$

This leads to a homogenized vector $\tilde{\boldsymbol{p}}$ with entries ranging from $-\infty$ to ∞ following a standard logistic distribution. This additional transformation has multiple advantages:

- It opens up the parameter space from a hypercube to the full d-dimensional space, which means that the calibration algorithm does not need to account for any boundaries.

- It avoids a problem of the $\boldsymbol{p}_{\text{cdf}}$-space, which might occur for distributions with pronounced tailing: increasing or decreasing of the parameter value in the direction of tailing means going closer and closer to the boundary of the $\boldsymbol{p}_{\text{cdf}}$-space. As a result, the derivative of the physical parameter value with respect to the $\boldsymbol{p}_{\text{cdf}}$ coordinate becomes very steep towards the respective boundary. This might be problematic for a calibration routine. Obviously, this problem depends very much on the chosen probability density functions, but in general the re-scaled values $\tilde{\boldsymbol{p}}$ do not suffer from this problem to the same extent, as the inverse logit transform contains a double-sided tailing itself.
- By construction, the origin of the re-scaled parameter space is located at the median of all physical parameter value prior distributions. This makes it easy to interpret positive or negative numbers, as they mean "larger than" or "smaller than" the median.
- We note that by using the logit-retransformed space, we make sure that no point can actually reach the maximum or minimum value of the input distributions. As this is typically unwanted in calibration anyway, we consider this a useful feature as long as the input distributions are chosen appropriately.

For the re-scaling, any distribution that maps the space $(0, 1)$ to $(-\infty, \infty)$ (e.g., the inverse Gaussian distribution) is equally applicable. We prefer the unscaled logit transformation, because it shows a slightly more pronounced tailing than a similarly scaled probit transformation. It also has the advantage of a trivial formulation for both forward and backward transformation. A disadvantage is that conditioning techniques that require Gaussian distributions, like Kalman Filtering, are not applicable. As a side-note: If the mapping onto the standard normal distribution is denoted Gaussian anamorphosis, the transformation suggested here might be denoted *logistic anamorphosis*.

Limitations in Computational Resources and Time For spatially distributed subsurface-flow models, like the one we are discussing in this study, a single model run may take several hours or even days. Calibration in the sense of global optimization typically requires many model runs, which means available time and/or computational resources can become a limitation. As outlined previously, one way to alleviate this problem, is the use of proxy-models. The actual calibration algorithm could then operate on the proxy-model, with occasional feedback between the full model and the proxy variant. This feedback might include running the full model at promising points in parameter space that were identified by the proxy-model, as well as re-training the proxy-model after a sufficient number of new input/output pairs was generated by running the full model.

In our approach, we make use of GPR for proxy-modeling (see Section 2.3). This technique is based on the interpolation of model outcomes between evaluated full model runs. The underlying concept and mathematical description is formally identical to kriging, but instead of interpolating in the two- or three-dimensional physical space, it is carried out in the higher-dimensional re-scaled parameter space (i.e., with $\tilde{\boldsymbol{p}}$).

Utilization of Computer-Cluster Resources Another important tool to overcome computational limitations is parallel computing. The presence, availability, and power of high-performance computing clusters has improved drastically in recent years. With such clusters, it is comparably easy to run computations in parallel. This can enhance model calibration in several ways:

- Parallelizing the underlying subsurface-flow model, for example by domain decomposition (e.g., Hwang et al., 2014), can accelerate the individual model runs. However, increasing communication-overhead limits this acceleration (i.e., doubling the number of processors will not lead to a runtime decrease of 50 %). For instance, the HGS manual recommends one processor for each 10^5 model grid nodes for optimal parallel efficiency.
- Parallelization can also be realized on the level of model runs themselves (i.e., running multiple distinct realizations on different processors). As the individual model evaluations should be completely independent of each other, this approach is trivial to implement. Besides the initial model construction, and the final readings of the output, there is also no need for communication between the model run and other processes. Hence, the efficiency does not generally decline with increasing number of processors. However, a direct inversely proportional relationship between core number and runtime is also not guaranteed, because typically not all realizations take the same time and the slowest one might be decisive.
- Finally, the model calibration procedure itself might be able to benefit from parallelization techniques. We outline one possible way to this in Section 14.2.

Schwede et al. (2012) investigated performance gains through the first two approaches.

Early calibration algorithms aiming at optimal use of limited computational resources focus on requiring as few model evaluations as possible (e.g., Jones et al., 1998; Regis and Shoemaker, 2007). However, with parallel computing, the limiting factor is not the absolute number of model evaluations themselves, but the number of iterations that have to be carried out in sequence. We illustrate this with an example: If the available cluster can run 100 models simultaneously, it is insignificant if the calibration algorithm needs 5, 10, 50 or 99 model evaluations as long as they are independent of each other. The wall-clock time will be identical (assuming each model realization takes the same time). Instead, it matters much more if the algorithm requires 500 or 1000 evaluations, because then the wall-clock time for calibration increases by a factor of 5 or 10. Obviously, during typical calibration algorithms, each model run is *not* independent of the previous one. On the contrary, the algorithm suggests new points to test, after evaluating all previous points. Still, the available computing resources should be used as much as possible.

We use a calibration approach that attempts to do this. As many others, it is based on an iterative procedure that cycles between finding promising points and evaluating the objective function at these points with the full model (e.g., Regis and Shoemaker, 2009; Wang and Shoemaker, 2014; Xia et al., 2021). Having access to a computer cluster, we can afford to probe several points within an

iteration, thereby scanning the parameter space and trying to identify the global minimum. To increase the efficiency even further, and to account for the fact that not all model runs take the same time, we even submit more points than required for the next cycle. We then wait until a desired number of newly evaluated points and the respective information is available and continue with the calibration procedure, while the remaining submitted models are still executed. This ensures that the cluster resources are used meaningfully, while the algorithm is busy with finding the next points to be evaluated.

Initial Sample Proxy-model assisted calibration schemes typically require an initial sample of points, where model input and output are known (i.e., pairs of $\tilde{\boldsymbol{p}}$ and $\{\boldsymbol{\vartheta}^\bullet, \boldsymbol{\varphi}^\bullet\}$). It is therefore common to start the calibration procedure with constructing an initial sample of a given size (e.g., five or ten times the number of input parameters d). To sample the parameter space as efficiently as possible, these points should not be drawn fully randomly, because pure random sampling has the tendency to produce clusters of closely-spaced points and sparse regions of low point-density. One way to avoid this, is the use of so-called space-filling designs, which try to space out the points as equally-distributed as possible. A common space-filling design is the Latin hypercube (e.g., McKay et al., 1979; Tang, 1993; Lin, Tang, et al., 2015), which provides a good spacing for a fixed number of points in an arbitrary number of dimensions.

We construct our initial sample by means of the Halton sequence (Halton, 1960; Berblinger and Schlier, 1991), a sequence of low discrepancy that can also be used to define space-filling designs. The main advantage of the Halton sequence over a Latin hypercube design is its expandability: it is not necessary to specify a number of points in advance and each next point of the sequence is close to optimally spaced compared to all previous points. By basing our initial sample on this sequence, we provide an excellent basis not only for the calibration algorithm and the proxy-model, but also for continued global exploration of the parameter space (further discussed in Section 14.2).

Just like Latin hypercube designs, the Halton sequence produces points of arbitrary dimension (in our case d) with coordinates ranging from 0 to 1. We treat these as points in the parameter space of the form $\boldsymbol{p}_{\text{cdf}}$. By means of the inverse cumulative distribution function $\boldsymbol{f}_{\text{cdf}}^{-1}$, we can convert these values to physical parameter sets for the full model. By applying the inverse logit transform $\boldsymbol{f}_{\text{scl}}$, we can convert these values to the homogenized re-scaled parameter space spanning $(-\infty, \infty)$. These definitions result in a denser sampling where the parameter distributions have a higher probability density, which means computational resources are used to explore those regions of the parameter space in more detail that are *a priori* considered to be more likely.

Proxy-Model Accuracy There are several ways of incorporating proxy-models into model calibration. In the most straight-forward approach, a single proxy-model is used to directly relate input parameters to the objective function value of the respective model results:

$$y^\circ = \mathcal{P}(\tilde{\boldsymbol{p}}) \approx f_{\text{obj}}(\boldsymbol{\vartheta}^*, \mathcal{M}(\boldsymbol{p})) = y^\bullet. \tag{14.11}$$

This has the advantages of a trivial implementation and comparably minor computational effort. However, it also comes at the cost of intermingling the behavior of the subsurface-flow model with the behavior of the objective function. For example, two distinct input parameter sets might result in completely different model outputs $\boldsymbol{\vartheta}$, but rather similar objective function values y. This underlying structure is hidden from the proxy-model and the available information (i.e., the raw outputs $\boldsymbol{\vartheta}$) is not utilized to the full extent. Ultimately, this could have adverse effects resulting in a subpar prediction quality of the proxy-model.

One way of avoiding this intermingling is to define, train and use multiple internal proxy-models and summarize them to a single "meta proxy-model". To do that, we can train one GPE for each observation that is used within the objective function. To estimate how a parameter set would perform using the full model, all internal proxy-models are used to predict the individual model outcomes and these predictions are then handed to the objective function:

$$y^{\circ} = f_{\text{obj}}(\boldsymbol{\vartheta}^*, \mathcal{P}(\tilde{\boldsymbol{p}})) \approx f_{\text{obj}}(\boldsymbol{\vartheta}^*, \mathcal{M}(\boldsymbol{p})) = y^{\bullet}. \tag{14.12}$$

In this way, predicting how a specific parameter set would perform with the full model in terms of the objective function becomes more reliable. Obviously, this benefit comes at additional computational costs. These costs are two-fold: First, the construction (i.e., training) cost of the meta-proxy-model increases proportionally with the number of observations n_{obj}. For a limited set of independent observations (e.g., about one hundred or fewer), like in our steady-state subsurface-flow model, these additional costs can easily be alleviated by making use of available cluster resources (as the proxy-models are independent of each other, their training can be trivially parallelized). However, for the calibration of transient models, where the objective function depends on time series, these costs quickly become prohibitive. Second, this approach also leads to an elevated computational effort (that again grows proportionally with the length of $\boldsymbol{\vartheta}$) each time a prediction is requested. We deem this acceptable, as we assume the proxy-modeling is so fast that the difference of evaluating a single or multiple internal proxy-models is small compared to the computing time of a full model run and thereby the overall calibration timeframe.

Balance of Exploitation and Exploration A key issue of calibration with restricted computational resources is to strike a balance between parameter space *exploration* and *exploitation* (Jones et al., 1998; Forrester and Keane, 2009). In this context, exploration refers to drawing and testing points, where the expected model outcome is most uncertain, to learn more about the model behavior. Most of the time, these points tend to be in regions of the parameter space which are far away from all previously evaluated points. On the other side, exploitation tries to make use of the available points in such a way, that new points are drawn and evaluated where the model is expected to show the best performance (i.e., the smallest value of the objective function). An algorithm solely focusing on exploration might take forever to find a reasonably good point to satisfy the convergence criteria, while an algorithm only based on exploitation might quickly get

stuck in a local minimum. As a result, many techniques of balancing these two "drivers" have been developed (e.g., Jones et al., 1998; Regis and Shoemaker, 2007; Wang and Shoemaker, 2014). In his study, we make use of the *surrogate-distance metric* developed by Regis and Shoemaker (2007) and used by Regis and Shoemaker (2009) and Xia et al. (2021) to balance exploitation and exploration. A brief summary of this procedure is given in Section 14.3.

In addition to points selected through the surrogate-distance procedure, we also select and submit some points for full model evaluation based on the Halton sequence. By doing so, we ensure true systematic global exploration. Conceptually, we think of it as following three strategies in parallel:

- local exploitation: track down local minima in the neighborhood of existing good points
- regional exploration: obtain a better impression of the response surface in the somewhat wider region of existing good points
- global exploration: systematically explore the full parameter space without any bias.

Choosing points for all three criteria (with a smooth transition between the first and the second through the weighted surrogate-distance metric), is a promising strategy to approach good points without getting trapped in local minima or completely missing some regions in parameter space.

Plausibility of Calibration Results Ideally, model calibration should result in a parameter set $\boldsymbol{p}_{\text{best}}$ that not only produces a model realization with optimal agreement of measured and modeled observations – this realization should also be *plausible*. We consider (im)plausibility a property of a model realization that would be more or less obvious to a human modeler when looking at the output. An implausible model might, for example, have fluxes across model boundaries that are orders of magnitudes off or even have the wrong sign. To avoid an implausible calibration outcome, the algorithm needs to be made aware of the concept of plausibility.

In principle, this would require a multi-objective optimization, as plausibility and small objective function values do not necessarily have to coincide. However, we think plausibility and low objective function values do not share the same priority. It would therefore not make sense to construct a Pareto front, because plausibility is more important (to us) than the value of the objective function. For example, we would consider even a perfect fit of measured ($\boldsymbol{\vartheta}^*$) and modeled data ($\boldsymbol{\vartheta}^\bullet$) worthless, if the respective realization is implausible. Hence, we can view plausibility as a strict constraint on the calibration. As $\boldsymbol{\vartheta}^*$ is obtained from reality (which should be plausible by definition), it also makes sense to assume that restricting the calibration to plausible points is helpful with respect to finding realistic values of $\boldsymbol{\vartheta}^\bullet$.

Obviously, to inform the calibration algorithm about the concept of plausibility, it needs to be transferred to objective, quantitative metrics. We do this by defining a plausibility function (Section 14.1) based on several plausibility criteria, each one operating on a scale from zero (implausible) to one (plausible). For instance, a wrong sign of a boundary flux would yield zero; a correct sign

would yield one. For some criteria (e.g., the absolute value of a flux or the ratio of two fluxes), the boundary between fully plausible and fully implausible is not as clear. Wherever it is hard to accurately define such a boundary, we use smooth transitions between zero and one to indicate a plausibility fringe. The overall plausibility of a model realization is ultimately defined as the product of all individual plausibility contributions. This ensures, that even if only a single criterion out of many is not fulfilled, the overall assessment is "implausible".

There are several potential ways to incorporate the plausibility function into a calibration algorithm. For example, one might only check after a full model run, whether the results are plausible or not. In the latter case, a penalty could be added when evaluating the objective function. Our implementation considers plausibility already during the stage of creating a set of proposed points. When constructing the set of proposal points, we apply a plausibility-based rejection-sampling approach based on another proxy-model (or meta proxy model) for plausibility assessment. A point with a predicted plausibility…

- …of 0 is immediately rejected and will not be run with the full model.
- …of 1 is immediately accepted and added to the set of proposal points that will be used for selecting new full model runs.
- …between 0 and 1 is compared with a random number drawn from a uniform distribution covering the interval (0, 1). If the plausibility value is larger than the random number, it is added to the set of proposal points.

As a result, the set of proposal points only contains suggestions that are predicted to be at least on the fringe of plausibility. A higher priority is given to points closer to full plausibility.

We do not apply the plausibility-check to points selected from the Halton sequence to ensure full and unbiased global exploration. Such a point, an initial sample point or a point that was erroneously predicted to be plausible, can turn out to be implausible. To avoid that these are treated as a viable solution to the calibration problem, we only consider the subset of points with a plausibility larger than zero for the calibration convergence criteria.

14.3 Applied Calibration Scheme Variants

As outlined above, we use the general parallelized proxy-model assisted calibration procedure developed by Regis and Shoemaker (2009) in an adapted form. We use four different variants differing in the number of internal proxy-models and the way of proposing points for full model evaluations. Nonetheless, they are all based on the following general procedure:

1. **Prior Definitions**: For each model input parameter, a prior probability density function is defined. These distributions allow all physically possible parameter values, but give a higher priority (through probability density) to those values that are considered likely or reasonable.

2. **Initial Sample**: By means of the Halton sequence, an initial set of points (in our case of size 120) is drawn from the prior distributions in a space-filling way. The respective model realizations are generated and submitted to a computing cluster for evaluation.

3. **Main Loop**: This loop is entered when all initial samples are available. It is executed until some convergence criteria are fulfilled. In our case, we run all variations until 3070 model realizations were simulated with the full model. A single evaluation of the full model requires a wall time of about 1 h, which results in a total calibration runtime of a few days per calibration variant (achieved through parallelization). We call each iteration of this loop a *cycle*.

   (a) **Analyze**: The objective and plausibility functions are evaluated for all new model realizations. At this point, the convergence criteria are checked and if they are met, the main loop is terminated.

   (b) **(Re)train Proxy-Model(s)**: We keep track of all model inputs and outputs for all full model realizations. We train a proxy-model for both, the objective function and the plausibility function. Depending on the case, this requires up to one GPE per observation. To accelerate this step, we parallelize the training by submitting training jobs to the cluster.

   (c) **Propose Points**: We generate a set of $n_{\text{pro}} = 10^4$ plausible points scattered around those locations that have proven to be "good" so far (judged by the objective function value). The specific details of how to create those points are outlined below.

   (d) **Select Points**: We iteratively apply the surrogate-distance metric for all (remaining) proposal points to select $n_{\text{set}} = 39$ points. Again, we provide further details of this step in the following.

   (e) **Append Points**: We extend the selected set with 11 points from the Halton sequence, which results in 50 new model realizations that are to be evaluated with the full model.

   (f) **Update**: We import all model realizations that have finished in the meantime. This ensures that the next iteration step does not solely import points from a previous cycle.

   (g) **Follow-up Sample**: We submit the 50 points in the selected set for evaluation with the full model on the cluster.

   (h) **Wait**: The status of all currently running model realizations is continuously checked. When enough models are ready (in our case 39), we continue with the next cycle. The remaining realizations are kept running in the background.

This procedure differs from the one of Regis and Shoemaker (2009) in that (1) we consider plausibility as an additional model constraint, (2) we do not perform intermediate resets to keep all calibration runs comparable, and (3) we use slightly different methods to propose and select points.

Proposing New Points A key step in the outlined algorithm is to propose points for evaluation with the full model. Ideally, those points should result in plausible model realizations that either produce small values of the objective function or help to restrain the proxy-model. The general idea behind our approach to proposing new points is the following:

1. Rank all points that have been evaluated with the full model already by the respective values of the objective function.
2. Apply some random scattering around the good points.
3. Eliminate those points that are predicted to lead to implausible models.
4. Repeat the scattering and elimination until the desired number of points (10^4) is reached.

To select the set of good points among the already performed full model runs objectively, we draft a weighted random subset from the set of all points. The respective weight w_i of each realization is based on its objective function value y_i in comparison to the best $y_{\min}$ and median objective function value $\bar{y}$ of all points:

$$w_i = \frac{\varrho_i}{\sum \varrho} \tag{14.13}$$

$$\varrho_i = \max\left(0, \frac{\bar{y} - y_i}{\bar{y} - y_{\min}}\right)^2. \tag{14.14}$$

By comparing minimum and median objective function value and setting negative weight factors to zero, we ensure that "less than average" performing realizations are not used for generating new points. The squaring leads to a smooth transition at objective function values close to the median. Using the median provides robustness towards extremely bad objective function values.

For the random scattering around the good points, we generalize the approach of Regis and Shoemaker (2009) who added random offsets $\Delta\tilde{\boldsymbol{p}}$ in parameter space. In our case, we split these offsets into a direction vector $\boldsymbol{e}$ (with a length of one) and a magnitude c that are generated independently:

$$\Delta\tilde{\boldsymbol{p}} = c \cdot \boldsymbol{e}. \tag{14.15}$$

We base the magnitude c on a scaling factor σ_{scale}, which takes the role of the standard deviation in Regis and Shoemaker (2009). We start with $\sigma_{\text{scale}} = 1.5$ and increase the scale by 50 % after three consecutive cycles that succeeded in finding a better point. Similarly, we reduce σ_{scale} by 50 % after three consecutive cycles without an improvement. To avoid extreme scales, we restrict σ_{scale} to be in the range from $5 \cdot 10^{-3}$ to 2.5. We sample c from the following probability density function $f(c|\sigma_{\text{scale}}, d)$:

$$f(c|\sigma_{\text{scale}}, d) = \frac{2^{1-d/2}}{\Gamma\left(\frac{d}{2}\right)\sigma_{\text{scale}}} \left(\frac{c}{\sigma_{\text{scale}}}\right)^{d-1} \exp\left(-\frac{1}{2}\left(\frac{c}{\sigma_{\text{scale}}}\right)^2\right), \tag{14.16}$$

which is a scaled and transformed χ^2-distribution describing the length of a vector where each entry is drawn from an independent normal distribution with standard deviation σ_{scale}. This mimics the random magnitudes of Regis and Shoemaker (2009). Our generalization lies in how we produce random offset directions $\boldsymbol{e}$.

Randomized Directions We require that the angle between $\boldsymbol{e}$ and the estimated direction of the steepest descent $\boldsymbol{a}$ of the objective function with respect to $\tilde{\boldsymbol{p}}$ (we elaborate on how to estimate $\boldsymbol{a}$ later) is smaller than some opening angle γ:

$$\gamma \geq \arccos \frac{\boldsymbol{a} \cdot \boldsymbol{e}}{|\boldsymbol{a}||\boldsymbol{e}|} \tag{14.17}$$

$$\boldsymbol{a} = -\nabla_{\tilde{p}} y. \tag{14.18}$$

This restricts the direction vector to be located within the hyperspherical sector of opening angle γ around the direction of the steepest descent. In case of $\gamma = \pi$, no restriction is applied and the approach falls back to pure and unbiased random sampling. In case of $\gamma = 0$, the hyperspherical cap collapses onto a single hyperdimensional line pointing directly towards the direction of the steepest descent. For a value of $\gamma = \pi/2$ we allow all directions whose projection onto that line would point towards $\boldsymbol{a}$ (opposed to $-\boldsymbol{a}$). Within the hyperspherical sector we sample $\boldsymbol{e}$ uniformly, meaning that there is no preference of any direction over another, as long as they are inside the hyperdimensional sector.

Drawing random directions on a full hypersphere is comparably easy (e.g., Muller, 1959; Marsaglia, 1972). Unfortunately, drawing such random directions and rejecting them if they are outside of the hyperspherical cone can become prohibitive for higher dimensions and opening angles $\gamma < 90°$. The reason for that is the rapidly shrinking ratio of the volume of hyperspherical sector V_{sector} to the full hypersphere volume V_{sphere} with increasing number of dimensions (Li, 2011):

$$\frac{V_{\text{sector}}(\gamma, d)}{V_{\text{sphere}}(d)} = \begin{cases} \frac{1}{2}\mathcal{I}_{\sin^2\gamma}\left(\frac{d-1}{2}, \frac{1}{2}\right) & \text{if } 0 \leq \gamma \leq \frac{\pi}{2} \\ 1 - \frac{1}{2}\mathcal{I}_{\sin^2(\pi-\gamma)}\left(\frac{d-1}{2}, \frac{1}{2}\right) & \text{otherwise,} \end{cases} \tag{14.19}$$

where $\mathcal{I}_{\sin^2\gamma}(a, b)$ is the regularized incomplete beta function. A visualization of this curve for different combinations of γ and d is given in Figure 26.

As a result, we present another method of randomly drawing directions within the hyperspherical sector, which relies on hyperspherical coordinates. We are only interested in directions and therefore assume a radial coordinate of one. The remaining coordinates ϕ_1 to ϕ_{d-1} are angles. In the case of a hyperspherical cap around the first coordinate axis, ϕ_1 ranges from 0 to γ, ϕ_2 to ϕ_{d-2} range from 0 to π and ϕ_{d-1} ranges from 0 to 2π. Figure 27 visualizes these angles (and the concept of a hyperspherical sector/cap) for a three-dimensional case. Uniform random sampling of those angles would not lead to uniform directions within the hyperspherical sector.

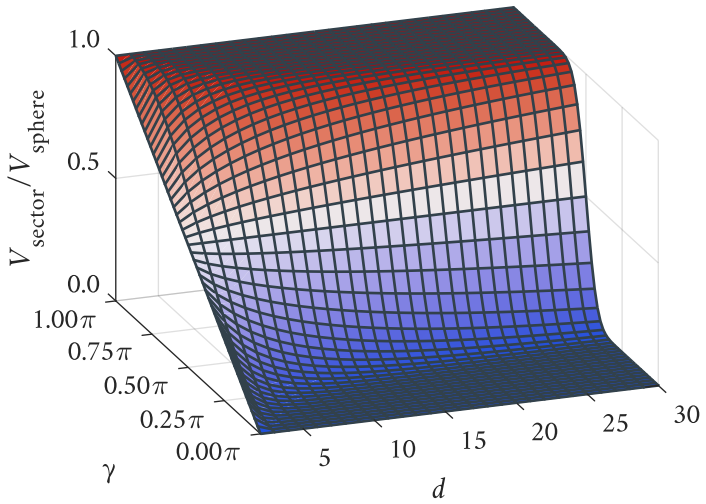

Figure 26: The relative volume of a hyperspherical sector within a hypersphere depends on the number of dimensions d and the opening angle γ.

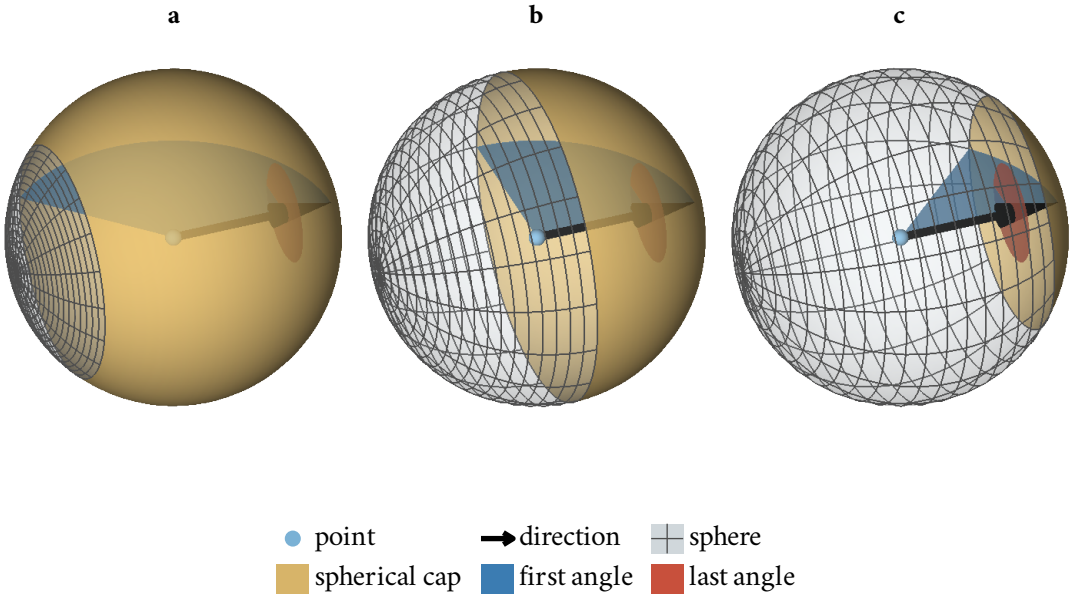

Figure 27: Visualization of a hyperspherical cap in three dimensions. A hypersphere can be created from a point, a direction and a given length. By restricting the first hyperspherical angular coordinate (of a rotated coordinate system) to be smaller than π, the hypersphere reduces to a hyperspherical cap. The first angle is allowed to range from 0 to γ, the last one from 0 to 2π. For higher-dimensional cases, there exist intermediate angles ranging from 0 to π. **a**: $\gamma = 3/4\pi$. **b**: $\gamma = \pi/2$. **c**: $\gamma = \pi/4$.

Instead, the angles have to be sampled according to their contribution to the *hyperspherical surface element* $\mathrm{d}A_{d-1}$ of a hypersphere in d-dimensional space:

$$\mathrm{d}A_{d-1} = \prod_{i=1}^{d-1} \sin^{d-1-i}(\phi_i)\ \mathrm{d}\phi_i \tag{14.20}$$

$$= \sin^{d-2}(\phi_1)\ \mathrm{d}\phi_1 \prod_{i=1}^{d-2} \sin^{d-1-i}(\phi_i)\ \mathrm{d}\phi_i \tag{14.21}$$

$$= \sin^{d-2}(\phi_1)\ \mathrm{d}\phi_1 \mathrm{d}A_{d-2}. \tag{14.22}$$

Hence, for uniform sampling within a hyperspherical sector with opening angle γ around the first coordinate axis, the first angle ϕ_1 has to be sampled from the following probability density function:

$$f(\phi_1) = \frac{\sin^{d-2}(\phi_1)}{\int_0^\gamma \sin^{d-2}(\tau)\ \mathrm{d}\tau}, \tag{14.23}$$

while all other angles can be constructed from a uniform point of a lower-dimensional $(d-2)$-sphere. Obviously, the shape of this probability density function depends on γ and d in a rather complicated way. However, with symbolic math software it is possible to find analytical expressions of $f(\phi_1)$ for a given combination of γ and d. Otherwise, the distribution can also be approximated numerically. In any case, modern programming or scripting languages can use this distribution to generate values of ϕ_1 that approximately satisfy Equation 14.23.

After obtaining valid directions in hyperspherical coordinates that assume the hyperspherical sector to surround the first coordinate axis, a trivial conversion to Cartesian coordinates and a back-rotation to the desired direction $\boldsymbol{a}$ are necessary.

Steepest Descent Estimation Our generalized approach requires an estimated direction of the steepest descent of the objective function value y with respect to the re-scaled input parameters $\tilde{\boldsymbol{p}}$. If the proxy-model is a single GPE that relates $\tilde{\boldsymbol{p}}$ with y, the respective gradient can be directly evaluated using the analytical GPE prediction derivatives outlined in Section 2.3.4. If multiple internal GPEs are used as a proxy-model to relate $\tilde{\boldsymbol{p}}$ with $\boldsymbol{\vartheta}$, the gradient can be decomposed in the following way:

$$\nabla_{\tilde{p}} y = \begin{bmatrix} \frac{\partial y}{\partial \tilde{p}_1} \\ \frac{\partial y}{\partial \tilde{p}_2} \\ \vdots \\ \frac{\partial y}{\partial \tilde{p}_d} \end{bmatrix} = \begin{bmatrix} \frac{\partial y}{\partial \vartheta_1}\frac{\partial \vartheta_1}{\partial \tilde{p}_1} & + & \frac{\partial y}{\partial \vartheta_2}\frac{\partial \vartheta_2}{\partial \tilde{p}_1} & + & \dots & + & \frac{\partial y}{\partial \vartheta_n}\frac{\partial \vartheta_n}{\partial \tilde{p}_1} \\ \frac{\partial y}{\partial \vartheta_1}\frac{\partial \vartheta_1}{\partial \tilde{p}_2} & + & \frac{\partial y}{\partial \vartheta_2}\frac{\partial \vartheta_2}{\partial \tilde{p}_2} & + & \dots & + & \frac{\partial y}{\partial \vartheta_n}\frac{\partial \vartheta_n}{\partial \tilde{p}_2} \\ \vdots & + & \vdots & + & \ddots & + & \vdots \\ \frac{\partial y}{\partial \vartheta_1}\frac{\partial \vartheta_1}{\partial \tilde{p}_d} & + & \frac{\partial y}{\partial \vartheta_2}\frac{\partial \vartheta_2}{\partial \tilde{p}_d} & + & \dots & + & \frac{\partial y}{\partial \vartheta_n}\frac{\partial \vartheta_n}{\partial \tilde{p}_d} \end{bmatrix} = \begin{bmatrix} \frac{\partial \vartheta_1}{\partial \tilde{p}_1} & \frac{\partial \vartheta_2}{\partial \tilde{p}_1} & \dots & \frac{\partial \vartheta_n}{\partial \tilde{p}_1} \\ \frac{\partial \vartheta_1}{\partial \tilde{p}_2} & \frac{\partial \vartheta_2}{\partial \tilde{p}_2} & \dots & \frac{\partial \vartheta_n}{\partial \tilde{p}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \vartheta_1}{\partial \tilde{p}_d} & \frac{\partial \vartheta_2}{\partial \tilde{p}_d} & \dots & \frac{\partial \vartheta_n}{\partial \tilde{p}_d} \end{bmatrix} \begin{bmatrix} \frac{\partial y}{\partial \vartheta_1} \\ \frac{\partial y}{\partial \vartheta_2} \\ \vdots \\ \frac{\partial y}{\partial \vartheta_n} \end{bmatrix}$$

$$= \boldsymbol{J}_{\vartheta,\tilde{p}}^{\mathsf{T}} \nabla_{\vartheta} y, \tag{14.24}$$

where n is n_{obj}. In this case, two things are necessary for the desired gradient: the Jacobian $\boldsymbol{J}_{\vartheta,\tilde{p}}$ of the model outputs $\boldsymbol{\vartheta}$ with respect to the re-scaled model inputs $\tilde{\boldsymbol{p}}$ and the gradient $\nabla_{\vartheta} y$ of the objective function value y with respect to the model outputs $\boldsymbol{\vartheta}$.

Applying the analytical derivatives of Section 2.3.4 for each observation ϑ_i and each parameter $\tilde{p}_j$ yields the entries of $\boldsymbol{J}_{\vartheta,\tilde{p}}$. Whether the gradient of the objective function value y with respect to the model outputs $\boldsymbol{\vartheta}$ can be determined analytically depends on the chosen objective function. We use a common sum of squared residuals evaluation where this is comparably easy:

$$y = f_{\text{obj}}(\boldsymbol{\vartheta}^*, \boldsymbol{\vartheta}) = \sum_{i=1}^{n_{\text{obj}}} (\vartheta_i - \vartheta_i^*)^2 \tag{14.25}$$

$$\frac{\partial y}{\partial \vartheta_i} = 2\,(\vartheta_i - \vartheta^*)\,. \tag{14.26}$$

This results in the following expression for the gradient:

$$\nabla_{\vartheta} y = \begin{bmatrix} \frac{\partial y}{\partial \vartheta_1} \\ \vdots \\ \frac{\partial y}{\partial \vartheta_n} \end{bmatrix} = 2\left(\boldsymbol{\vartheta}^{\mathsf{T}} - \boldsymbol{\vartheta}^{*\mathsf{T}}\right). \tag{14.27}$$

As an overall result, we can obtain estimates of $\nabla_{\tilde{p}} y$ not only for all points that have already been evaluated with the full model, but also (through the help of the proxy-models) for any other point. The gradient $\nabla_{\tilde{p}} y$ points into that direction where the objective function value increases fastest with a change in $\tilde{\boldsymbol{p}}$. As discussed, we use the opposite direction ($\boldsymbol{a} = -\nabla_{\tilde{p}} y$) to propose new points.

Selecting New Points Similar to the approaches of Regis and Shoemaker (2007, 2009) and Xia et al. (2021) we use the surrogate-distance metric to balance exploitation and exploration. This criterion uses a weighted average between two normalized metrics:

- The exploration metric: a normalized measure of the minimal dimensionless distance (in parameter space) between a proposed point and all points that have already been evaluated/selected. The minimal distances of all points in the currently proposed set are linearly normalized by the largest and smallest value that occur in the set:

$$m_i^{\text{explore}} = \frac{\max(\boldsymbol{d}) - d_i}{\max(\boldsymbol{d}) - \min(\boldsymbol{d})}, \tag{14.28}$$

  where m_i^{explore} is the exploration metric of the i-th point in the proposal set and $\boldsymbol{d}$ is the vector of minimum Euclidean distances between the proposed points and all previous points. This metric ranges from 0 to 1, where 0 indicates largest minimal distance and 1 smallest minimal distance. A pure exploration scheme would always strive for the point with the smallest value, as it is the one furthest away from all previous points.

- The exploitation metric: a normalized measure of performance as it is predicted by the proxy-model. The predicted objective function values are linearly normalized by the largest and smallest values of the current set of proposed points:

$$m_i^{\text{exploit}} = \frac{y_i^\circ - \min(\boldsymbol{y}^\circ)}{\max(\boldsymbol{y}^\circ) - \min(\boldsymbol{y}^\circ)}, \tag{14.29}$$

  where m_i^{exploit} is the exploitation metric of the i-th point in the proposal set and $\boldsymbol{y}^\circ$ are the predicted objective function values of the proposed points. This also results in numbers ranging from 0 to 1, where 0 indicates minimal predicted objective function value and 1 indicates the opposite. A pure exploitation scheme would always go for the point with the smallest value, as it is predicted to perform best.

To select n points from a large set based on this metric, we use n weights that linearly scale between zero and one. For each weight ξ we select the evaluation point with the lowest weighted average of the two measures:

$$i_{\text{select}} = \arg\min_i \left[\xi \cdot m_i^{\text{explore}} + (1 - \xi) \cdot m_i^{\text{exploit}}\right], \tag{14.30}$$

where i_{select} is the index of the point that is selected for a full model evaluation. After selecting a point, the exploitation metric is re-evaluated for all remaining points in the set and the next point is selected based on the next weight. We start with a focus on exploitation ($\xi = 0$) and go towards exploration ($\xi = 1$).

Variants In total we apply the presented calibration scheme in four different variants:

- "single (180°)": We use only a single GPE as proxy-model for the objective function, and another single GPE for the plausibility function. We do not restrict the space of point proposals by using an opening angle of $\gamma = 180°$. This variant is conceptually closest to the original implementation of Regis and Shoemaker (2009).
- "many (180°)": We use multiple GPEs as a meta proxy-model for both, the objective function and the plausibility function. We expect the proxy-model predictions to be more accurate, which should enhance the calibration. We still do not restrict the space of point proposals ($\gamma = 180°$).
- "many (90°)": In addition to using multiple GPEs for objective and plausibility functions, we restrict the space of proposed points by narrowing the opening angle of the hyperspherical sector to $\gamma = 90°$. This should avoid stepping into the wrong direction, while still allowing enough randomness to find good new points. As a result, we expect a further improvement of the calibration.
- "uninformed": For comparison, we also apply a naïve global exploration variant of the presented scheme. It does not use the outlined methods for point proposal and selection,

but instead just continues with the next samples of the Halton sequence. As this approach focuses solely on exploration and does not make an effort for exploitation, we expect it to perform poorly compared to the other variants. We also iteratively train multiple GPEs for this case, but the respective prediction information is not used during the calibration itself.

We compare both, the progression of these algorithms over time (over the course of the individual cycles), and the final outcome after more than 3000 full model evaluations.

14.4 Construction of Posterior Distributions

Simulation-Based Inference The calibration schemes outlined above aim at finding a single best-point estimate in the parameter space. Often, it is desirable to also estimate the uncertainty of the individual parameters, ideally in terms of a full posterior distribution. Many classical approaches for inferring this posterior are using likelihood-based Bayesian methods (e.g., MCMC). A common problem of these methods is their slow convergence, especially in higher-dimensional parameter spaces. One possible way to avoid such problems is to use a likelihood-free alternative for estimating the posterior parameter distribution.

The recently developed research on SBI presents such alternatives. One particular example is Sequential Neural Posterior Estimation (SNPE) (Papamakarios and Murray, 2016; Lueckmann et al., 2017; Greenberg et al., 2019), which we apply to our case. The principle of SNPE is based on training a deep neural network. In contrast to the GPE proxy-model, the SNPE method treats model outputs (i.e., observations) as an input. For any queried set of observations, SNPE provides a description of a multi-dimensional distribution of parameter values that are assessed as "likely to produce the queried observations". It is trivial to sample this distribution to infer information about parameter correlations and uncertainties.

We use the single-round version of SNPE (i.e., NPE), because tests with our data have shown that iterative re-training of the neural network does not improve the quality of the posterior distribution. To be precise, we use the SNPE-C implementation (also known as automatic posterior transformation) with atomic loss (Greenberg et al., 2019) and a Masked Autoregressive Flow (MAF) density estimation neural network (Papamakarios et al., 2017). We make use of the flexible interface of the SBI toolbox of Tejero-Cantero et al. (2020) to perform NPE on pre-simulated data. In summary, the procedure works as follows (see Greenberg et al., 2019, for a comprehensive description):

1. Sample n_{set} input parameter sets $\tilde{\boldsymbol{p}}$ from the prior distribution.
2. Obtain the corresponding n_{set} model outputs $\boldsymbol{\vartheta}^{\bullet}$ for these parameters.
3. Formulate or choose an approximate d-dimensional posterior density distribution description $P_{\text{post}}(\boldsymbol{\Psi}, \tilde{\boldsymbol{p}})$ (in our case MAF), which has a set of parameters $\boldsymbol{\Psi}$ and assigns a scalar posterior probability density P_{post} to any parameter set $\tilde{\boldsymbol{p}}$.

4. Define a neural network f_{NN} that relates the density estimator's parameters $\boldsymbol{\Psi}$ to given model outputs $\boldsymbol{\vartheta}^{\bullet}$:

$$\boldsymbol{\Psi} = f_{\mathrm{NN}}(\boldsymbol{\Phi}, \boldsymbol{\vartheta}^{\bullet}), \tag{14.31}$$

where $\boldsymbol{\Phi}$ are the weights (i.e., coefficients) of the neural network.

5. Train the neural network on the sampled data, by adjusting the weights $\boldsymbol{\Phi}$ to minimize the following term:

$$\boldsymbol{\Phi}^{*} = \arg\min_{\boldsymbol{\Phi}} \left[-\sum_{i=1}^{n_{\mathrm{set}}} \log(P_{\mathrm{post}}(f_{\mathrm{NN}}(\boldsymbol{\Phi}, \boldsymbol{\vartheta}_i^{\bullet}), \tilde{\boldsymbol{p}}_i)) \right]. \tag{14.32}$$

6. For a given set of observations $\boldsymbol{\vartheta}^{*}$, the posterior density is then given by:

$$P_{\mathrm{post}}(f_{\mathrm{NN}}(\boldsymbol{\Phi}^{*}, \boldsymbol{\vartheta}^{*}), \tilde{\boldsymbol{p}}). \tag{14.33}$$

There are two reasons, why one should be careful with the interpretation of the NPE results in our application: NPE (1) works with stochastic models (2) under the assumption that the target data can be generated with the model and the right parameters. In our case, however, the subsurface flow model is deterministic and there is little hope that any parameter set would provide a perfect agreement between modeled and measured data. The latter has various reasons, including model structural errors and the comparison of a steady-state model with a snapshot of real-life transient data. As a result, we apply NPE slightly outside of its original purpose, but the results might be interpretable and useful nonetheless.

Markov Chain Monte Carlo Ideally, we would like to compare the SBI results with a posterior distribution obtained from a classical Bayesian inference method based on likelihoods and the full model. We cannot afford this, due to the computational effort of a single model run and the high-dimensional parameter space. However, we can run an MCMC scheme on a quick-to-evaluate GPE-based proxy-model. For that, we train GPEs for all observations on all data generated with the full model (across the four calibration variants) to generate the best possible meta proxy-model that we can afford. We then apply a classical Metropolis-Hastings MCMC algorithm (Metropolis et al., 1953; Hastings, 1970), which is briefly summarized in the following:

1. We randomly sample $n_{\mathrm{chains}} = 12$ points from the prior distribution to initiate different chains. These points are treated as "trial points".

2. We evaluate the proxy-model predictions for the trial points to emulate model runs and to obtain approximate simulated observations $\boldsymbol{\vartheta}^{\circ}$.

3. We determine the log-likelihood $\log \mathcal{L}_{\mathrm{post}}$ of each trial point by comparing the virtual observations with the calibration target $\boldsymbol{\vartheta}^{*}$:

$$\log \mathcal{L}_{\mathrm{lik}} = -\frac{1}{2}(\boldsymbol{\vartheta}^{\circ} - \boldsymbol{\vartheta}^{*})\boldsymbol{C}^{-1}(\boldsymbol{\vartheta}^{\circ} - \boldsymbol{\vartheta}^{*})^{\mathsf{T}}, \tag{14.34}$$

where $n_{\text{obj}} \times n_{\text{obj}}$ is a covariance matrix of measurement errors $\boldsymbol{C}$ and all constant terms were omitted, as they are not necessary for the following calculations. We assume independent Gaussian measurement errors with the same variance σ_{obs}^2 for all observations ($\boldsymbol{C} = \sigma_{\text{obs}}^2 \boldsymbol{I}$). If the model was able to meet all observations with the right set of parameters, σ_{obs} would only reflect the uncertainty of the measured observations (in the order of a few centimeters). In this case, however, σ_{obs} should also account for model structural errors (i.e., all reasons why the model can only approximate the observed values). As a result, σ_{obs} needs to be inflated artificially, because otherwise the likelihoods would drop rapidly if not evaluated at the best parameter sets. On the other hand, if the uncertainty is inflated too much, all realizations will be considered approximately equally likely, because even very large deviations between modeled and measured values would be considered acceptable. We obtain our value for σ_{obs} by considering a property of the χ^2-distribution: Assuming all residuals are independent and identically distributed random variables following normal distributions, the sum of squared residuals y follows a χ^2-distribution scaled by the observation variance:

$$y \sim \sigma_{\text{obs}}^2 \chi_{n_{\text{obj}}-d}^2. \tag{14.35}$$

The expected value of this distribution is the number of degrees of freedom ($n_{\text{obj}} - d$):

$$E[y] = \sigma_{\text{obs}}^2 \cdot (n_{\text{obj}} - d). \tag{14.36}$$

By assuming that the sum of squared residuals $y(\boldsymbol{p}_{\text{best}})$ for the best point found in the calibration variants is equal to this expected value, we can derive a reasonable value of σ_{obs}:

$$\sigma_{\text{obs}} = \sqrt{\frac{E[y]}{n_{\text{obj}} - d}} = \sqrt{\frac{y(\boldsymbol{p}_{\text{best}})}{n_{\text{obj}} - d}}. \tag{14.37}$$

4. We determine the prior log-probability $\log \mathcal{L}_{\text{prior}}$ of each trial point via the probability density function of the logistic distribution:

$$\log \mathcal{L}_{\text{prior}} = \sum_{i=1}^{d} -\tilde{p}_i - 2 \cdot \log\left(1 + \exp(-\tilde{p}_i)\right). \tag{14.38}$$

5. The logarithm of the posterior probability density P_{post} is then given by the sum of the two terms:

$$\log P_{\text{post}} = \log \mathcal{L}_{\text{lik}} + \log \mathcal{L}_{\text{prior}}. \tag{14.39}$$

6. A comparison between the posterior densities of the current trial points and the previous points in the chains decides whether the trial points should be accepted or rejected:

$$\delta = \exp\left(\log P_{\text{post}}^{\text{trial}} - \log P_{\text{post}}^{\text{previous}}\right) - \nu, \tag{14.40}$$

where ν is a random number drawn from a uniform distribution between zero and one. Whenever this difference is positive ($\delta > 0$), the trial point is accepted. Whenever it is negative, the trial point is rejected and the previous point is re-accepted (i.e., repeated). As the initial points do not have precursors, they are all accepted.

7. The current list of n_{chains} (re-)accepted points is perturbed to generate new trial points. We use trivial perturbations with offsets generated from scaled standard normal distributions. The associated scaling factor is dynamically tuned over the course of the MCMC procedure to maintain an acceptance rate of about 30 %.

8. With the new set of trial points, the procedure is repeated from the second step until convergence between the chains is achieved. For that we require that the criterion developed by Gelman and Rubin (1992) needs to be smaller than 1.1 for all parameters.

The final list of points represents an MCMC-based sample from the posterior distribution.

14.5 Subsurface-Flow Model

In principle, we apply the calibration to the same subsurface flow model that was described in Section 10.1. However, some insights obtained after publication of the paper reproduced in Chapter III have lead to minor to moderate changes with respect to the model definition. These changes affect the general model formulation (Section 14.5.1), the parametrization including the prior distributions (Section 14.5.2) and the plausibility assessment (Section 14.5.3).

The calibration is carried out by comparing virtual measurements with the hydraulic heads recorded on the key-date November 6, 2018 (see Section 3.3), as it is the largest and most diverse data set available. Unfortunately, this means that the observation wells in the Grabfeld formation, which were installed in the spring of 2020, could not be considered.

14.5.1 Description

First of all, the spatial discretization was changed to create a more appropriate model mesh. This holds for both, the horizontal division into triangles, and the vertical distribution of nodal layers. We paid particular attention to the Quaternary within the Ammer valley. For instance, a modification of the meshing algorithm now ensures that the clay layer between Tufa and gravel is continuous, as suggested by the geological model. Another improvement considers the nodes reflecting river Ammer, which now follow the thalweg as it was inferred in Section 3.4. By optimizing the mesh sizing using the `mesh2D` toolbox of Engwirda (2014), we could keep the number of elements comparably small (4672 horizontal triangles, 102 784 three-dimensional elements and 55 752 three-dimensional vertices).

In contrast to the model of Section 10.1, we do not split the Erfurt formation `kuE` into two parts, because we have little geological information that justifies the division and the split would increase

the number of model parameters. Nevertheless, we still use twelve layers, as we introduce a soil layer of 1.5 m thickness, which covers all surficial parts that are not covered by the Quaternary alluvial fines in the Ammer floodplain. The three-dimensional geometry of the hillslope hollows was refined and does not vary between the realizations anymore.

We modify the boundary conditions, to (1) make the model more realistic, (2) reduce the number of model parameters, and (3) enhance model convergence. The new boundary conditions are the following:

- We apply a single recharge rate at all top faces that are outside of the Ammer Quaternary.
- Instead of Dirichlet boundaries, we use a leaky boundary condition to describe the lateral groundwater inlet and outlet in the Ammer valley.
- We still use a Dirichlet boundary condition to describe the southern groundwater outlet towards the Neckar catchment. However, instead of applying it to the full southern boundary, we restrict it to the eastern end. As the remaining boundary was mostly parallel to streamlines anyway, this change affects the resulting flow field only marginally.
- The interaction with river Ammer is now implemented as a Dirichlet boundary condition, where the respective heads are inferred from the surface water model described in Section 3.4.
- We use a drain boundary for all drainage ditches with an intermediate layer thickness of 0.1 m, an intermediate hydraulic conductivity of $10^{-4}\,\mathrm{m\,s^{-1}}$ and a drainage threshold of 0.02 m. These values lead to a facilitated drainage.
- We use a drain boundary for all top nodes covered by alluvial fines with an intermediate layer thickness of 0.4 m, an intermediate hydraulic conductivity equaling the hydraulic conductivity of the alluvial-fines layer and a drainage threshold of 0.5 m. These values allow some slightly artesian conditions in the Quaternary as observed in field measurements.
- Finally, we use a drain boundary for all remaining top nodes with an intermediate layer thickness of 0.3 m, an intermediate hydraulic conductivity of $10^{-4}\,\mathrm{m\,s^{-1}}$ and a threshold of 0.15 m. This avoids a significant build-up of artesian heads in regions where the recharge rate cannot be discharged towards the subsurface (this occurs mostly where the sandstone formations `km2345` crop out, but also at some coarsely resolved spots in the Neckar valley).

For the parametrization of the unsaturated zone we apply a mixture of the Brooks and Corey model (we use Equation 2.16 for $k_{\mathrm{rel}}(S_e)$ with $\lambda = N - 1$) and the van Genuchten model (we use Equation 2.20 for $S_e(h_c)$). As outlined in Section 2.2.3, this approach has proven to be beneficial for the numerical convergence (especially with randomly drawn parameter sets), while the effect on the resulting steady-state flow field is negligible.

Table 9: Prior distribution definitions for all model parameters.

#	Parameter	Unit	Type	c_1	c_2	c_3	support interval
1	weathering depth	m	$\log_{10}\mathcal{N}$	1.46	0.11		$(0.00, \infty)$
2	anisotropy (bedrock)	–	$\mathcal{B}$	1.52	3.04		$(0.00, 1.00)$
3	anisotropy (Quaternary)	–	$\mathcal{B}$	3.04	1.52		$(0.00, 1.00)$
4	anisotropy (soil)	–	$\mathcal{B}$	3.04	1.52		$(0.00, 1.00)$
5	N (bedrock)	–	$\log_{10}\mathcal{N}_\mathrm{T}$	0.20	0.06		$(1.00, \infty)$
6	N (Quaternary)	–	$\log_{10}\mathcal{N}_\mathrm{T}$	0.20	0.06		$(1.00, \infty)$
7	N (soil)	–	$\log_{10}\mathcal{N}_\mathrm{T}$	0.20	0.06		$(1.00, \infty)$
8	K_{xy} (`kuE`)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−5.27	0.68		$(0.00, \infty)$
9	K_{xy} (lower `kmGr`)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−8.77	1.16		$(0.00, \infty)$
10	K_{xy} (upper `kmGr`)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−5.17	0.73		$(0.00, \infty)$
11	K_{xy} (`km2345`)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−7.51	1.19		$(0.00, \infty)$
12	K_{xy} (hollows)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−6.16	1.00		$(0.00, \infty)$
13	K_{xy} (Neckar gravel)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−3.43	0.65		$(0.00, \infty)$
14	K_{xy} (soil)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−5.19	0.49		$(0.00, \infty)$
15	K_{xy} (gravel)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−4.35	0.84		$(0.00, \infty)$
16	K_{xy} (clay)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−8.11	0.62		$(0.00, \infty)$
17	K_{xy} (Tufa)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−5.09	0.66		$(0.00, \infty)$
18	K_{xy} (alluvial fines)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−7.00	1.02		$(0.00, \infty)$
19	K_{xy} (river buffer)	$\mathrm{m\,s^{-1}}$	$\log_{10}\mathcal{N}$	−5.19	0.49		$(0.00, \infty)$
20	h_leaky (north, inlet)	m	$\mathcal{N}_\mathrm{T}$	350.50	0.75		$(346.69, \infty)$
21	h_leaky (north, outlet)	m	$\mathcal{N}_\mathrm{T}$	335.50	0.75		$(-\infty, 337.26)$
22	h_leaky (south)	m	$\mathcal{N}$	326.25	0.76		$(-\infty, \infty)$
23	Q (river)	$\mathrm{m^3\,s^{-1}}$	$\mathcal{E}_\mathrm{T}$	0.50	0.05	0.13	$(0.34, 1.59)$
24	recharge rate	$\mathrm{m\,s^{-1}}$	$\mathcal{B}_\mathrm{S}$	2.06	6.25		$(0.00, 1.91 \cdot 10^{-8})$

14.5.2 Prior Distributions

The d-dimensional vector of prior distributions $\boldsymbol{f}_\mathrm{cdf}(\boldsymbol{p})$ summarizes all available knowledge about the model input parameters ahead of the calibration. Here, we assume independence between all model parameters (a common choice, because typically little credible information about parameter correlations is available *a priori*). This means we can describe the prior distribution by d individual distributions, each with its own support interval and cumulative density function $f_\mathrm{cdf}(p)$. In the following, we provide an overview of the model parameters and all available data, as well as explanations on how we infer the prior distributions. Table 9 provides a summary of that. We use the following symbols for the different distribution types:

- $\mathcal{N}$: A normal distribution with the mean value c_1 and the standard deviation c_2.
- $\log_{10}\mathcal{N}$: A log-normal distribution where the base-10 logarithm of the parameter has the mean c_1 and the standard deviation c_2.
- $\mathcal{B}$: A beta distribution with the two shape parameters c_1 and c_2.
- $\mathcal{E}$: A generalized extreme value distribution with location parameter c_1, scale parameter c_2 and shape parameter c_3.

We use the symbol "T" to indicate a truncation that was applied to ensure that the support interval does not include unreasonable values. We use the symbol "S" to indicate a linear scaling that was applied to modify the support interval of the beta distribution.

Weathering Depth The first parameter describes the maximum weathering depth of the Grabfeld formation beneath the land surface. Literature estimates for this quantity are collected in Table 13 (in the appendix). For the derivation of a prior distribution, we omit the 1 m estimate of Kehrer, as it was probably not meant to be a reliable estimate of the maximum weathering depth of this formation. We also omit the rather extreme values of Erdal and Cirpka (2019), to avoid a distribution skewed too far towards large values and to limit the variance to a reasonable extent.

The maximum weathering depth beneath the land surface has to be non-negative, as negative values would not have a physical meaning and a fixed physical upper limit does not exist. Hence, we use a $\log_{10}$-normal distribution to describe this parameter, as it is maxentropic for a given mean and log-variance on the interval $(0, \infty)$. The two coefficients can be derived directly from all remaining data by evaluating the mean and standard deviation of the $\log_{10}$-transformed values. The resulting coefficients are listed in Table 9.

Material Properties Parameters #2 to #19 describe various material properties of the hydrostratigraphic units. Relevant literature data are collected in Tables 10 to 12 and 14 to 21 (in the appendix). In principle, each hydrostratigraphic unit requires four parameters:

1. The non-negative saturated horizontal hydraulic conductivity K_{xy} in $\mathrm{m\,s^{-1}}$.
2. The anisotropy expressed as a dimensionless ratio of vertical to horizontal hydraulic conductivity. It typically operates within the range 0.00 to 1.00 (i.e., the vertical permeability is smaller than the horizontal one).
3. The positive van Genuchten parameter α in $\mathrm{m^{-1}}$.
4. The dimensionless van Genuchten parameter N, which has to be larger than 1.

With twelve hydrostratigraphic units we would end up with 48 parameters related to material properties. As each additional parameter increases the dimensionality and therefore the vastness of the parameter space, we decided to use additional regularizations to reduce the number of parameters describing material properties to 18:

- We eliminate the van Genuchten parameter α by directly relating α to the saturated horizontal hydraulic conductivity according to a relationship derived in Section 17 (in the appendix).
- For the anisotropies and N, we group all bedrock (`kuE`, lower/upper `kmGr`, `km2345`) and Quaternary (hollows, Neckar gravel, gravel, clay, Tufa, alluvial fines) hydrostratigraphic units. Together with the soil layer, this results in three groups per parameter.

It is difficult to obtain reliable estimates of anisotropy ratios from the collected prior data as they mostly contain assumptions instead of actual measurements. In some modeling studies it is also not clear whether a value of 1.0 was only chosen because the model could not cope with anisotropy. As a result, we constructed two different beta distributions, one for all bedrock layers, and one for both, all Quaternary layers and the soil layer. The first distribution is skewed towards lower values, with a mean of 1/3 and a standard deviation of 1/5. The second distribution is skewed towards higher values, with a mean of 2/3 and a standard deviation of 1/5. The underlying reasoning is that consolidated bedrock material typically has smaller vertical conductivities due to internal micro-structure (fine layers with hindered flow perpendicular to them), whereas the unconsolidated Quaternary and the soil layer do not show that to the same extent.

The van Genuchten parameter N controls how quickly the relative permeability drops with increasing matric potential. In principle, it can assume values between unity and infinity, but according to Rawls and Brakensiek (1989) the values typically range from 1.0 to 1.4 for unconsolidated sediments. Cases with $N \gg 2.0$ can lead to a strong jump in the $S_e(h_c)$-curve that would result in a sudden loss of saturation (and therefore permeability) even for small matric potentials. To enhance numerical stability, we therefore choose a $\log_{10}$-normal distribution than mostly covers the range 1.0 to 2.5 (see Table 9 for the coefficients). To avoid numbers smaller than one, we truncate the distribution at $N = 1.0$. As we have little data to justify different priors across the three groups, we use the same distribution for all of them.

The saturated horizontal hydraulic conductivities cannot be negative, have no upper physical limit and can easily vary over multiple orders of magnitude. Therefore, we use $\log_{10}$-normal distributions to describe their priors. The coefficients in Table 9 were constructed on a case-by-case basis to meet the data of Tables 10 to 12 and 14 to 21 reasonably well. Where available, experimental data from the study site were considered to be more trustworthy than other data (e.g., data referenced as model assumptions).

Boundary conditions The groundwater inlet and outlet boundary conditions in the Ammer floodplain require head values (i.e., h_{leaky}). We use normal distributions to describe these. For the Ammer valley we use the average longitudinal hydraulic gradient of approximately 8 m/3000 m = 0.27 % (Martin et al., 2020) to estimate the mean hydraulic head at the floodplain ends about 1500 m upstream and downstream of the modeling domain. To account for the uncertainty of these values, we use a standard deviation of 0.75 m.

At the southern boundary towards the Neckar floodplain we use a normal distribution to describe the fixed head. We base the mean value on the modeling results of Keim and Pfäfflin (2006). From data measured by nearby ASG/SWT-wells we know that the hydraulic head in this area can vary by about 2.5 m. Therefore, we use a standard deviation of 0.76 m, which results in a normal distribution where 90 % of the values are within ±1.25 m distance from the mean.

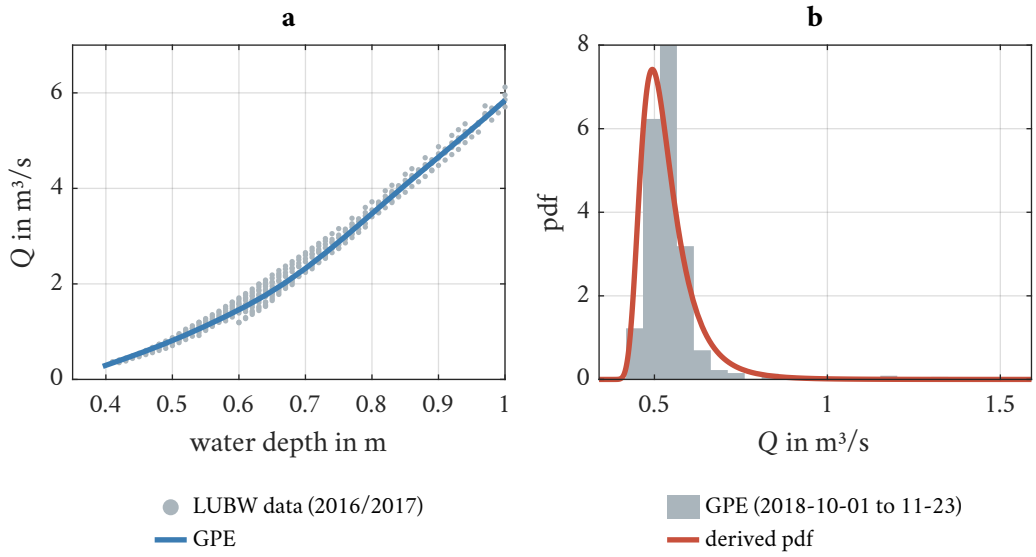

Figure 28: Prior distribution of the Ammer river boundary condition. **a**: River discharge inferred from water depth recordings in the time period of interest. **b**: Derived prior distribution of Q.

The fixed heads of the Ammer river are constructed from the inverse rating curve relationship derived in Section 3.4. Hence, the prior distribution is formulated with respect to the river discharge Q in $\mathsf{L}^3\,\mathsf{T}^{-1}$. We obtain it from the Pfäffingen gauging station data (LUBW, 2021):

1. We train a one-dimensional GPE to infer Q from the water depth. For training, we use the time period in which both, water stage and discharge data are available (May 1, 2016 to October 31, 2017).
2. We use the GPE to infer Q at the time period of interest (October 1, 2018 to November 23, 2018). Fortunately, the respective water stages fall within the range of the training data, which means that extrapolation (in the data space) is not required.
3. We create a histogram of the derived Q values and fit a continuous distribution to it. In this case, a generalized extreme value distribution provided the best agreement.
4. We truncate the distribution at the maximum and minimum values predicted by the GPE.

Figure 28 shows the training data, the GPE interpolation and the resulting prior distribution.

The final parameter is the average groundwater recharge rate, which is a non-negative number, because negative values are only physically possible if there is a subsurface source and if the annual mean evapotranspiration rates exceed precipitation. The physical upper limit of the recharge rate is the average precipitation, which was $1.91 \cdot 10^{-8}\,\mathrm{m\,s^{-1}}$ from January 1, 2014 to December 31, 2020 (LTZ, 2021, measured in Unterjesingen, which is located within the area of interest). As a result, we choose a scaled beta distribution to describe this parameter, as it is defined on a bounded interval and allows for non-uniform probabilities. We derive the respective coefficients (shown in Table 9) from the literature data summarized in Table 22 (in the appendix).

14.5.3 Plausibility Function

We base the plausibility function of a model realization on the individual fluxes across boundaries. These are available as raw output from HGS in form of the water balance summary. In our case, the plausibility function is the product of five contributions (φ_1 to φ_5):

$$f_{\text{plaus}}(\boldsymbol{\varphi}) = \varphi_1 \cdot \varphi_2 \cdot \varphi_3 \cdot \varphi_4 \cdot \varphi_5. \tag{14.41}$$

The first criterion states that the flux Q_{inlet} across the groundwater inlet boundary on the Ammer side must be positive (i.e., the boundary has to be a source), the second one similarly states that the flux Q_{outlet} across the groundwater outlet boundary on the Ammer side must be negative (i.e., the boundary has to be a sink):

$$\varphi_1 = \begin{cases} 1 & \text{if } Q_{\text{inlet}} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{14.42}$$

$$\varphi_2 = \begin{cases} 1 & \text{if } Q_{\text{outlet}} < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{14.43}$$

The third and fourth criteria state that the flux across the outlet or inlet (on the Ammer side) should not be much larger than its counterpart. We accept all realizations where the magnitude of one flux is less than twice the magnitude of the other flux, and we reject all those where one is more than four times larger than the other. Between these two limits we define a gradual transition by the smoothing function $f_{\text{s}}(x) = 3x^2 - 2x^3$:

$$\varphi_3 = \begin{cases} 1 & \text{if } |Q_{\text{outlet}}| < 2|Q_{\text{inlet}}| \\ 0 & \text{if } |Q_{\text{outlet}}| > 4|Q_{\text{inlet}}| \\ f_{\text{s}}\left(4\frac{|Q_{\text{inlet}}|}{|Q_{\text{outlet}}|} - 1\right) & \text{otherwise,} \end{cases} \tag{14.44}$$

$$\varphi_4 = \begin{cases} 1 & \text{if } |Q_{\text{inlet}}| < 2|Q_{\text{outlet}}| \\ 0 & \text{if } |Q_{\text{inlet}}| > 4|Q_{\text{outlet}}| \\ f_{\text{s}}\left(4\frac{|Q_{\text{outlet}}|}{|Q_{\text{inlet}}|} - 1\right) & \text{otherwise.} \end{cases} \tag{14.45}$$

Finally, a criterion states that the Ammer river is not a major source. We require that its net contribution Q_{river} is less than 10 % of the total water balance flux Q_{tot} (the flux across all boundaries):

$$\varphi_5 = \begin{cases} 1 & \text{if } Q_{\text{river}} < 0 \\ 0 & \text{if } Q_{\text{river}} > 0.1 \cdot Q_{\text{tot}} \\ f_{\text{s}}\left(1 - 10\frac{Q_{\text{river}}}{Q_{\text{tot}}}\right) & \text{otherwise.} \end{cases} \tag{14.46}$$

The smoothing between 0 and 1 for the last three criteria is implemented to alleviate the effects of the arbitrarily chosen thresholds. The third-order smoothing function f_{s} ensures that the transition at the thresholds has a continuous first derivative.

15 Results & Discussion

We first describe and discuss the results of running the best-point calibration variants for 3070 full model realizations. Then, we analyze the full posterior distributions (derived with SBI and MCMC). Finally, we discuss noteworthy features of the resulting flow field in the calibrated model.

15.1 Calibration Scheme Variants

Figure 29 shows the development of the objective function over the course of the different calibration variants. We can observe that the uninformed sampling strategy converges very slowly, as expected, even with the space-filling Halton sequence. After nearly sixty cycles it has only reached an objective function value of about 6.1 m^2 (which corresponds to an RMSE of 0.39 m), which was achieved by the other three schemes already within the first two cycles. After more than 3000 realizations (59 cycles), the informed schemes that rely on multiple internal GPEs have found points with objective function values close to each other ("many (180°)": 2.90 m^2, which corresponds to an RMSE of 0.27 m; "many (90°)": 2.80 m^2, RMSE of 0.26 m), while the variant based on a single, lumped GPE produced results between these two and the uninformed scheme (3.32 m^2, RMSE of 0.28 m). In general, this ranking seems to be stable over the cycles: The informed schemes are able to find good regions in parameter space much faster than the uninformed one, and multiple GPEs accelerate the calibration even more. There is also a difference between the two multi-GPE runs with different opening angles, but it is comparably small. The curve corresponding to the narrowed opening angle of 90° is nearly always slightly beneath the line of representing the full opening angle of 180°. Therefore, the reduction of the opening angle seems to help finding better points in parameter space, but the effect is only marginal.

The sampling strategies can be observed in Figure 29b:

- The uninformed variant consistently produces values of the objective function across the range from 6 m^2 to 1000 m^2. As expected, there is no systematic improvement over time and all cycles produce similar results.
- The informed variants show patterns. On the long-term, a systematic improvement can be observed for all three approaches. This can be attributed to finding good regions in parameter space. Within in each cycle, especially in later ones, it can be observed that early realizations yield smaller values of the objective function than later realizations. This is a direct outcome of the linearly decreasing exploitation/exploration weights that initially promote points that are predicted to be favorable (i.e., in the sense of having low objective function values). The increase is the result of shifting the weights towards regional parameter space exploration. The last eleven realizations of each cycle correspond to the unbiased global sampling according to the Halton sequence. Those are recognizable by typically much larger values of the objective function, as they are not aware of the good parameter-space regions.

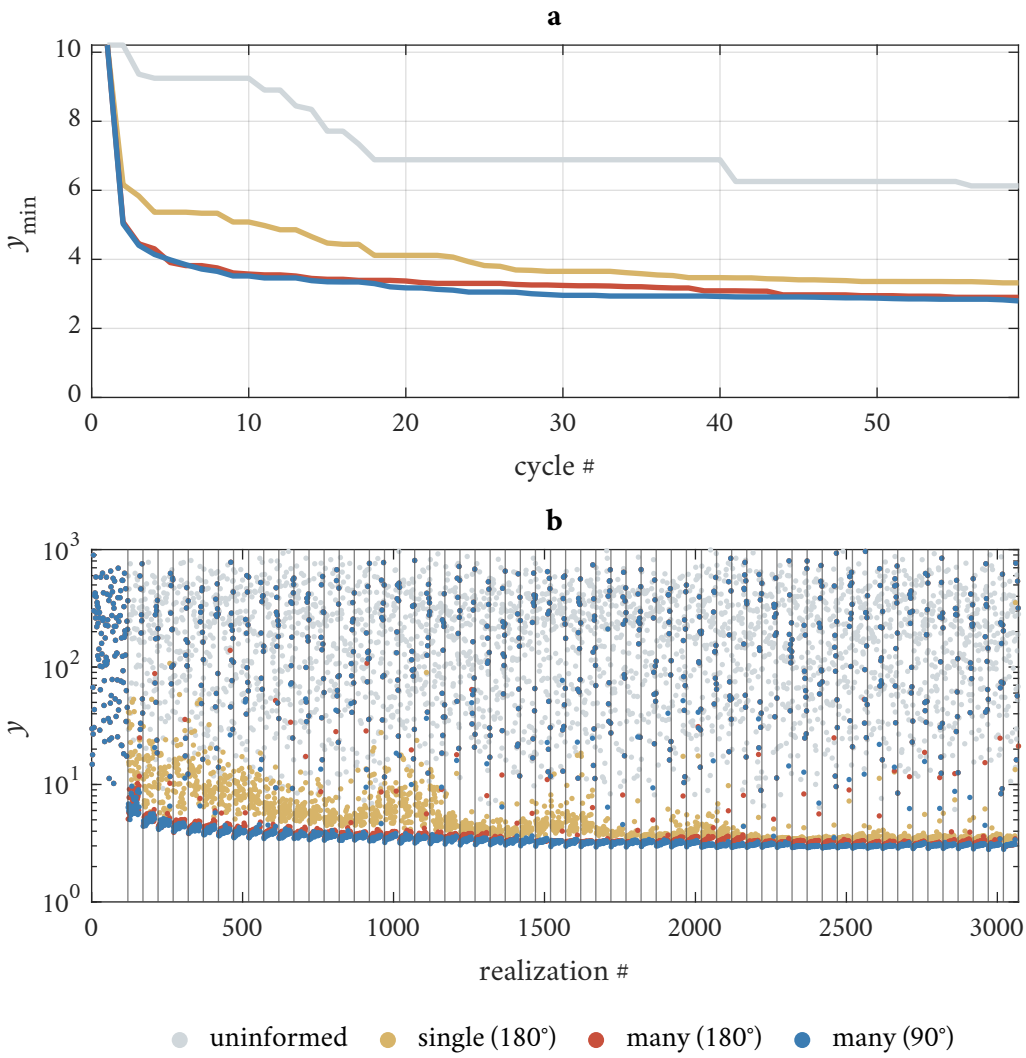

Figure 29: Objective function values y for the different calibration variants. **a**: Best objective function value of all previous iteration cycles. **b**: All individual objective function values on a logarithmic scale. The vertical lines separate the cycles.

- All variants make use of the same Halton sequence. This is visible for the first cycle, where all four variants start with the same set of 120 initial points, and for the last eleven realizations of all subsequent cycles, where the points of the three informed schemes fall on top of each other.

In summary, the performance of all informed schemes is comparable with respect to the final objective function value, but using multiple GPEs helps to find and exploit good regions in parameter space faster.

Figure 30 shows how the four variants differ with respect to the prediction error of the proxy-model over the course of the calibration. The prediction error is defined as the absolute difference between an actually determined objective function value (or plausibility) obtained by running the full model and the value previously predicted by the surrogate model.

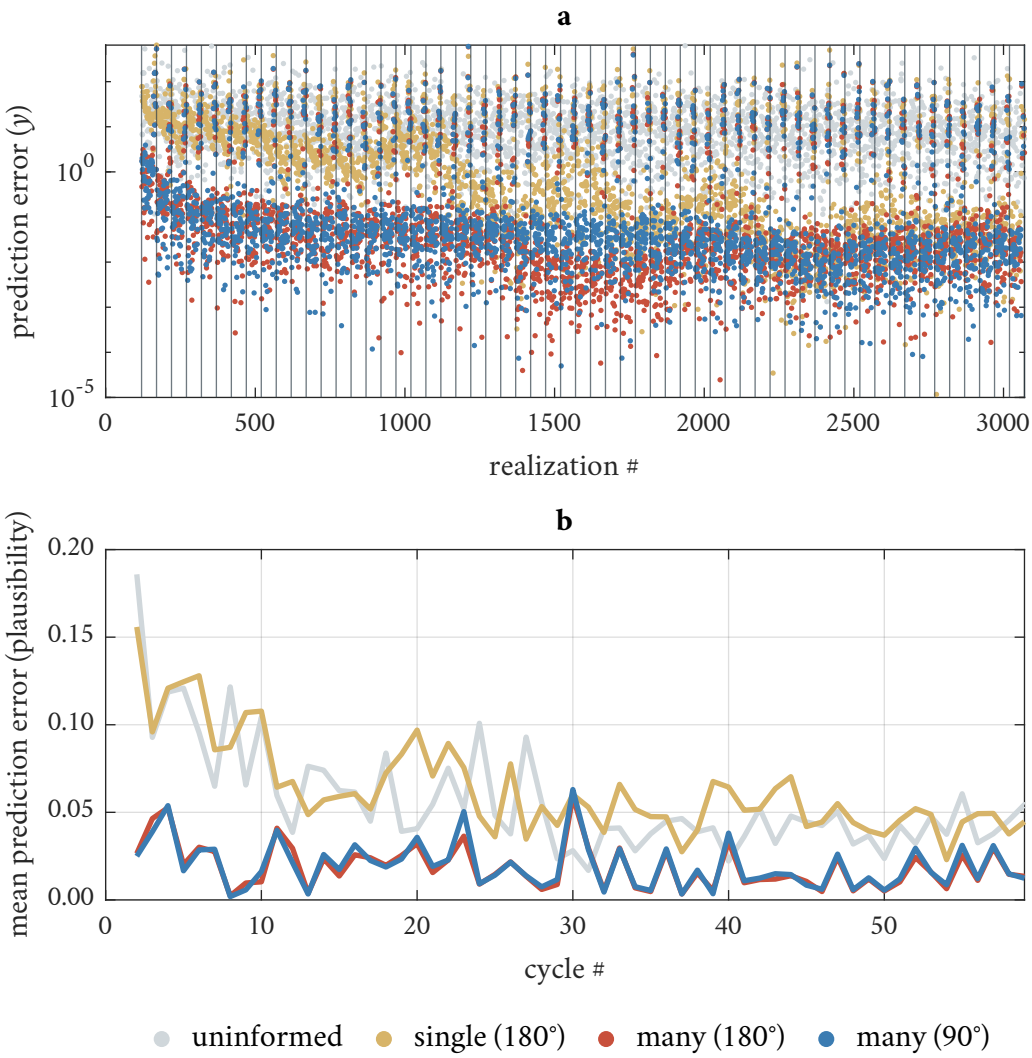

Figure 30: Absolute difference between predictions of objective function value and plausibility score and the actual value obtained from running the full model.

Figure 30a reveals the following:

- Even though the uninformed approach uses multiple GPEs for predicting the value of the objective function, the predictions are of comparably low quality. The prediction error does not drastically decrease over time, which means that the prediction quality increases only marginally even though new information is appended to the GPEs. This illustrates the vastness of the parameter space, as even thousands of space-filling points result in such a low point density that GPE-based interpolation is obviously difficult.
- The informed single-GPE variant roughly starts with similar prediction errors as the uninformed case, but the prediction quality increases over time. This is a result of the scheme approaching better regions in parameter space, where the sampling density increases such that the interpolation works better. In contrast to that, the space-filling Halton design of the uninformed scheme tries to maintain a uniform density throughout the parameter space.

- The calibration scheme variants based on multiple GPEs immediately achieve smaller prediction errors of the objective function, which highlights the drastically improved prediction quality. Over the course of the calibration, the prediction error also decreases until the three informed schemes achieve a similar prediction quality after about 2300 realizations.

The prediction error of the plausibility score behaves similarly, but also shows some differences:

- With values around 0.05 or smaller (from 30 cycles onward), all four variants achieve a decent average prediction quality (a value of 1 would indicate that all plausibility predictions of a cycle were completely wrong; a value of 0 would mean perfect plausibility predictions).
- The uninformed variant and the single-GPE variant perform worse than the multi-GPE cases, but the initially large difference decreases over time. Again, this is probably related to the low point-density, where each new point is very far from all previous points.
- It is interesting to see that the uninformed scheme can also achieve a significant improvement for the prediction of the plausibility score. This indicates that predicting plausibility is easier (i.e., requires a smaller point density) than predicting the objective function value. This might be partly related to the fact that our plausibility function depends on fewer variables than our objective function.
- The informed variants with multiple internal GPEs produce conspicuously similar mean prediction errors. The reason for that can be found in the contribution of the individual realizations to these mean prediction errors. While the two multi-GPE schemes achieve a nearly perfect plausibility prediction of points selected by the surrogate-distance metric (i.e., a prediction error of 0), nearly the entire mean prediction error stems from the Halton sequence points, which are identical in both variants.

So far, the approaches have only been compared with respect to the model outputs in terms of the objective function and the plausibility assessment. We now want to consider the parameter values (i.e., the coordinates in parameter space) of the points found by the individual schemes.

Figure 31 visualizes the estimated logit-score transformed parameters of the final best points of the four calibration variants. To give an impression of the calibration course, the full sequence of all intermediate best points after the first cycle (i.e., all the parameter sets that have contributed to Figure 29a) is shown, too. It is important to note that these point sets do not form an interpretable distribution (at least not a meaningful posterior distribution), as the optimization schemes to not attempt to truly sample the full posterior distribution, but rather aim at finding the single global optimum. The corresponding ranges can therefore also not be related to parameter uncertainties.

In Figure 31 we can see that the best points found with the uninformed scheme obviously are a subset of the prior distribution, but already here some patterns emerge. For example, some parameters exhibit a systematic shift towards positive (e.g., hydraulic conductivities of Tufa and the alluvial fines) or negative (e.g., hydraulic conductivities of Neckar gravel and soil anisotropy) values.

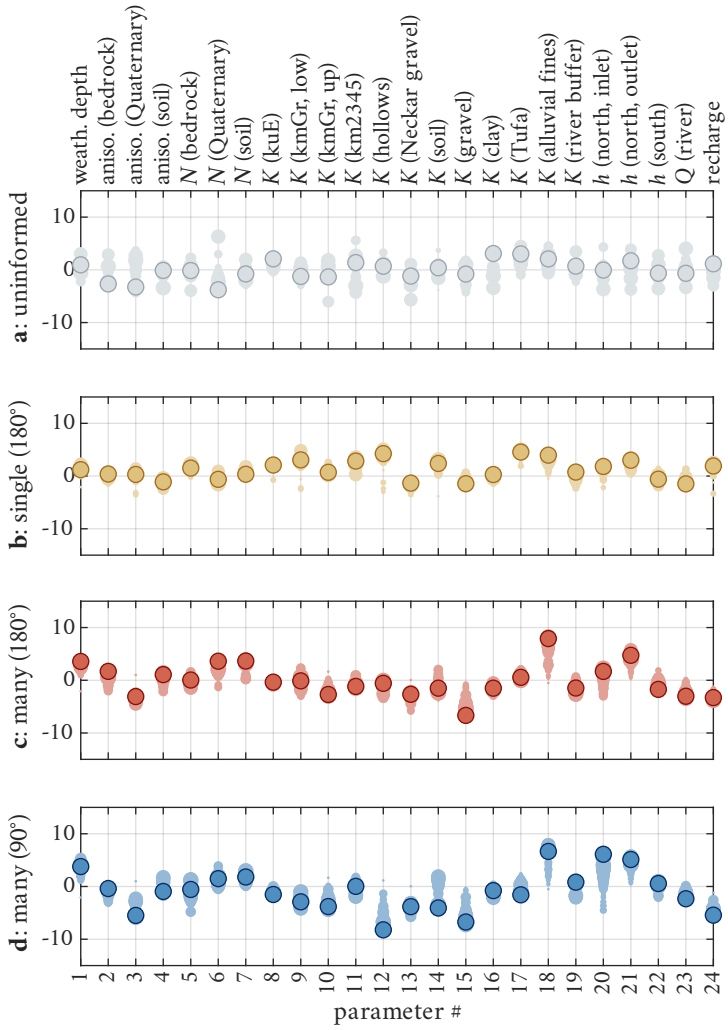

Figure 31: Visualizations of best found parameter sets in the dimensionless $\tilde{p}$-space as large, dark, outlined circles. Intermediate best points are shown as faded smaller circles, where the circle size corresponds to calibration progress (smaller circles occur in earlier cycles).

The best points found with the informed schemes share some common properties. Four commonalities in Figure 31b-d stand out particularly:

- There are many agreements for parameters related to hydraulic conductivities. For example, the best estimate for the Neckar gravel (parameter #13) is negative in all informed variants. Similarly, parameters #15 to #18 (hydraulic conductivity of gravel to hydraulic conductivity of the alluvial fines) form a consistent visual pattern from left to right: The scaled parameter values seem to shift from being more or less strongly negative (gravel) to around zero (clay) to slightly positive (Tufa, with exception of "many (90°)") to large and positive (alluvial fines).
- The scaled parameter values of the hydraulic conductivity of the Erfurt formation (`kuE`) change only slightly over the course of the calibration.
- The parameter values related to hydraulic head offsets in the Ammer valley (#20 and #21) are positive, while the offset in the Neckar valley boundary (#22) scatters around zero.

- Positive parameter values seem to be preferred for the first parameter (maximum weathering zone depth) across all calibration scheme variants.

On the other hand, there are also significant differences between the best points found by the single-GPE and multi-GPE schemes:

- While in both cases the values for the hydraulic conductivity of the Erfurt formation are barely changed during the calibration, the absolute values of this parameter are different between the variants. The single-GPE scheme clearly shifts towards positive values, while the other variants produce negative values.
- Similarly, the values for the recharge rate do not agree. Again, the point found by the single-GPE scheme is a positive normalized value, while the other variants have found good points only with negative shifted values of this parameter.
- Finally, there are large discrepancies for the parameter describing the hydraulic conductivity of the hillslope hollows (parameter #12). Positive, neutral and strongly negative values are all present in the three variants.

While comparing the subfigures with each other, one should obviously not forget that the best points found with the four variants correspond to different values of the objective function. For instance, the uninformed sampling data (i.e., the one with the worst agreement to measured data) are further away from being optimal parameter sets than the other three cases. However, as already stated, the objective-function values of the best points found with the informed variants are reasonably close to each other. The differences between these plots raise the question of which (if any) of the parameter sets should be trusted most and how large the parameter uncertainties really are. The analysis of full posterior distributions allows such investigations.

15.2 Analysis of Posterior Distributions

Sample Construction We apply Neural Posterior Estimation as outlined in Section 14.4. This requires a sample from the prior distribution, for which we use the 3070 realizations of the uninformed sampling scheme. The neural network training step of NPE involves a random split of the input data into a training and a test data set. To make the results robust with respect to that, we apply NPE ten times and average the resulting posterior distributions. We then sample 10 000 points, of which a randomly chosen subset of 250 are used to generate full model realizations.

For the MCMC-based approach, we obtain an inflated standard deviation of measurements of $\sigma_{\text{obs}} = 0.41$ m from the sum of squared residuals at the best found point $y(\boldsymbol{p}_{\text{best}}) = 2.80\,\text{m}^2$. With that, the MCMC scheme outlined in Section 14.4 converges after 27 000 proxy-model realizations. We omit the first 5000 to diminish the influence of the chains' initial states. From the rest we randomly choose 10 000 points that serve as a posterior sample for comparison with its SBI-counterpart. Finally, we run the full model for 250 randomly selected points of this set.

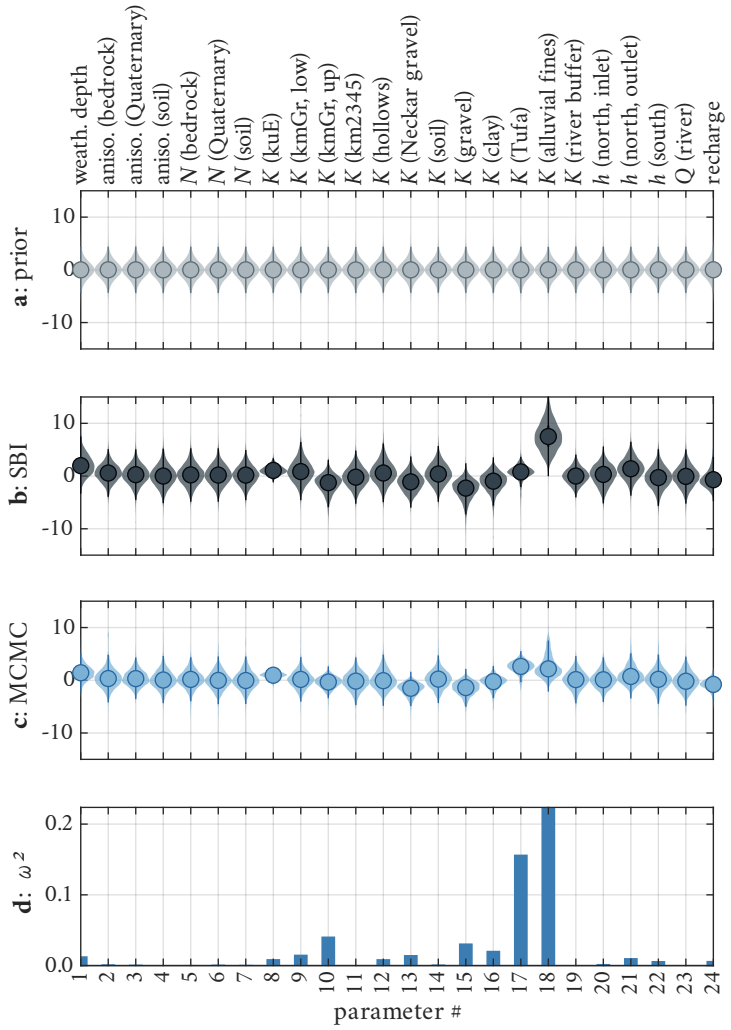

Figure 32: a-c:Comparison of prior and posterior parameter distributions, where circles highlight the median values and the shaded areas show point density (violin plots). All parameter values are displayed in the dimensionless $\tilde{p}$-space. **a**: Prior definitions of parameters. **b**: Posterior parameter distribution as obtained by the SBI procedure. **c**: Posterior parameter distribution as obtained by the MCMC procedure. **d**: Dissimilarity between the SBI and MCMC marginal posterior distributions expressed as ω^2 (Equation 15.1).

Marginal Distributions Both posterior construction methods result in multi-dimensional parameter distributions. In Figure 32, we show the marginal distributions as violin plots for each parameter side by side. Figure 32a uses the full set of points generated by the uninformed calibration variant to give an impression of the prior distribution. We see that the scaled parameter values mostly are within the range −5 to 5, with a higher concentration around 0. This fits perfectly with the definition of the logit transformation. None of the parameters stands out visually, which confirms that the uninformed and unbiased sampling by means of the Halton sequence was successful. Figure 32b shows the marginal distributions obtained with SBI; Figure 32c shows the MCMC-based marginal posterior distributions. To understand consequences for the physical, untransformed parameter values, we also provide the marginal cumulative distribution functions of all unscaled parameter values (i.e., p) for the prior and posterior distributions in Figure 33.

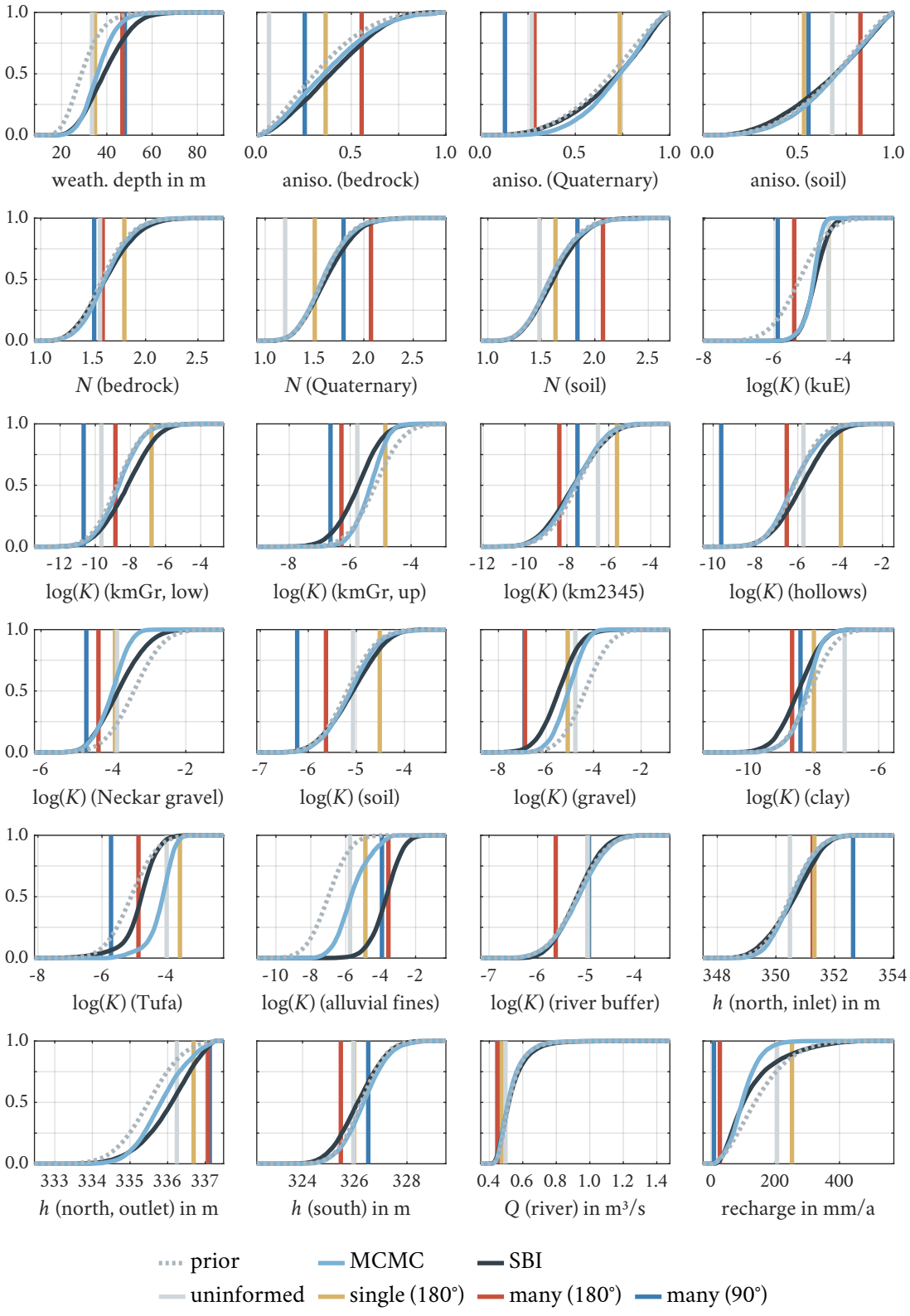

Figure 33: Individual marginal prior and posterior parameter distributions visualized as cumulative distribution functions of unscaled physical parameter values p. Colored lines in the background indicate the best points of the optimization schemes. All hydraulic conductivities are given in m s^{-1} and displayed on a logarithmic axis to the base 10.

In Figures 32 and 33, some remarkable features of the posterior distributions are immediately recognizable and similar to results obtained from the calibration outcomes (Figure 31). For instance, the hydraulic conductivity of the Erfurt formation is the narrowest dimensionless parameter in both cases (Figure 32b-c), with a significant shift of similar magnitude towards larger values. Just as observed in Figure 31, we also see a distinct difference between the prior and posterior values for parameters #15 to #18. In all four cases a parameter shift in the same direction can be observed for both posterior cases and the global calibration outcomes.

In general, the most striking differences between prior and posterior distributions occur for parameters related to hydraulic conductivities. This could indicate, that these parameters exert an important control on the model output (i.e., the hydraulic head observations). This is not surprising, considering that the hydraulic conductivities are defined on a logarithmic scale. In this case, even a minor shift can result in a change by one order of magnitude of material permeability. Not all hydraulic conductivity values are equally important for the observed hydraulic heads. For instance, the conductivities of the lumped sandstone formation, the soil layer and the hillslope hollows are hardly affected by the inference. They obviously exert a low influence on the simulated heads at the measurement locations, which are mostly affected by conductivities in the direct vicinity of the location and by those conductivities that determine the overall flow field. The former is the case for the Ammer gravel and Tufa layers; the latter applies for the hydraulic conductivities in the Erfurt formation and the alluvial fines:

- The alluvial fines form the top of the Quaternary valley-filling in the Ammer floodplain and as such represent the connection between deeper groundwater stories (like the Tufa and gravel aquifers) and the land surface, including the drainage ditches and river Ammer. Increasing the corresponding hydraulic conductivity therefore enhances groundwater discharge towards the drainage network. Apparently, to get the flow system close to measured observations, an intensification of this exfiltration process is required, up to a point where the high-density regions of posterior and prior parameter distribution hardly overlap (at least for the SBI-based distribution). This might be an indication that the parameter is compensating for a structural model error. For instance, a network of tile drains that is assumed to exist in the real domain (mentioned by Kehrer, 1935) could play this exfiltrating role in reality. As it was not implemented in the model due to lacking information about the position of the tile drains, the inference leads to an adjustment of the alluvial fines instead.
- The Erfurt formation is the bottom-most hydrostratigraphic unit of the model. It is also the only one that extends laterally across the entire domain. Depending on its hydraulic conductivity, it might therefore take a leading role in connecting or separating the northern (and elevation-wise upper) Ammer floodplain with/from the southern (lower) Neckar floodplain. The inference suggests a higher hydraulic conductivity of this unit than initially proposed by the prior information. This could mean that a better hydraulic connection between the two valleys is necessary to align the model with the measurements.

It is a bit surprising that the parameters related to the Ammer river (hydraulic conductivity of the river buffer zone and the river discharge responsible for river water stage) are basically not affected by the inference at all. This could indicate that the connection between the groundwater system and the drainage ditch network is more important for the head observations than the connection to the river.

The prior and posterior distributions of the three parameters related to the van Genuchten coefficient N are also nearly identical. This indicates a low sensitivity of the virtual hydraulic head observations to these parameters, which is not surprising, as these parameters have very little effect on the steady-state flow field and the observations.

Another noteworthy point concerns the hydraulic conductivities of Tufa and gravel in the Ammer valley. The posterior distributions suggest a decrease in permeability for gravel and an increase for Tufa, up to a point where the hydraulic conductivity in the Tufa formation exceeds the one in the gravel. This is unexpected, considering the field measurements (see Table 15 and Table 17 in the appendix), which were mostly derived from slug tests. However, the gravel has a clayey matrix and the wells have been developed prior to performing the slug tests. This might have washed out too much clay in the direct vicinity of the deep wells to obtain a reliable estimate of hydraulic conductivity of the formation. Perhaps it would be worthwhile to perform larger-scale pumping tests to confirm or falsify this outcome.

The weathering zone is interesting in that the related depth is shifted towards larger values, but the corresponding hydraulic conductivities are decreased compared to the prior. Again, this agrees well with the best-point estimates found in the global calibration, just like a slight preference for larger values regarding the boundary condition offset parameters in the northern modeling domain (mostly at the outlet), that is missing for the southern boundary condition. Another commonality between the posterior distributions are the lowered values for the conductivity of the Neckar gravel and the recharge rate. The corresponding uncertainties, however, are smaller for the MCMC-posterior, which is noticeable by narrower distributions. There are more also visibly asymmetrical distributions for the MCMC-posterior in general (e.g., in the hydraulic conductivities of Neckar gravel, Tufa and alluvial fines), while in the SBI case they are mostly symmetric. Smaller uncertainties and conspicuous asymmetries might be an indication that the MCMC-based approach is more suitable to constrain the posterior parameter values based on the observation data, compared to NPE. However, at this point we cannot draw any conclusions about the goodness of the posteriors, as this requires an analysis of the drawn full model runs (given in Section 15.4).

One way of objectively comparing the corresponding marginal posterior distributions of a parameter is the ω^2-metric developed by Anderson (1962) based on the work of Cramér (1928) and von Mises (1928):

$$\omega^2 = \int_{-\infty}^{\infty} \left(F(x) - G(x)\right)^2 \, \mathrm{d}H(x), \tag{15.1}$$

where $F(x)$ and $G(x)$ are the individual empirical cumulative distributions functions of the samples and $H(x)$ is the empirical cumulative distribution function of both samples combined. Larger values of ω^2 indicate a greater dissimilarity, and a value of $\omega^2 = 0$ would indicate a perfect agreement.

The resulting ω^2-values are displayed in Figure 32d. It becomes clear, that especially for the hydraulic conductivities of the Tufa and alluvial fines layers, the posterior distributions disagree. In the case of the alluvial fines, the MCMC-based posterior tends to shift much less towards larger values than the SBI-based counterpart. This might be a result of the influence of the prior distribution that condemns points far away from zero as "unlikely" – an information that is neither used, nor directly available in the NPE process. The shifting for the hydraulic conductivity of the Tufa, on the other hand, is even larger in the MCMC case. In both cases, the shift is accompanied by a narrowing, indicating that this parameter can be constrained by the observations to some extent.

Correlation Analysis The visualizations presented so far only looked into marginal information of individual parameters. However, as we have access to full d-dimensional posterior distributions, we can determine correlation coefficients between all 24 parameters. The resulting matrices for both, the SBI- and MCMC-based results are combined in Figure 34.

Visually most striking is a nearly perfect correlation between the hydraulic conductivity of the Erfurt formation and the recharge rate in both posteriors. The former is also correlated to most of the Quaternary hydraulic conductivities (gravel, clay, Tufa, alluvial fines). Obviously, this also implies correlations between recharge rate and these conductivities, which can also be observed in Figure 34. There is also a notable correlation between these hydraulic conductivities themselves.

Apart from these stronger correlations, there is quite some disagreement between the SBI- and MCMC-based distributions. The largest discrepancy might be a sign flip of low to medium correlation between the hydraulic conductivities of Neckar gravel and the Erfurt formation. However, it is difficult to draw further conclusions from Figure 34, as many correlation coefficients are comparably small in magnitude and therefore imply weak relationships that might just be governed by random noise. Such noise is a bit more visible in the MCMC-based posterior distribution, where there are even some minor correlations for the van Genuchten N parameters that have been confirmed to be of low relevance.

Individual correlation coefficients are only scalar summary metrics that work best for linear relationships. As such, they might occlude nonlinear relationships between the two investigated parameters of each pixel in Figure 34. We therefore plot full pairwise two-dimensional posterior distributions in Figure 35. We also show the final best points of the calibration scheme variants. With $24^2 = 576$ individual subfigures, this chart contains a lot of information, so we only focus on the most important aspects.

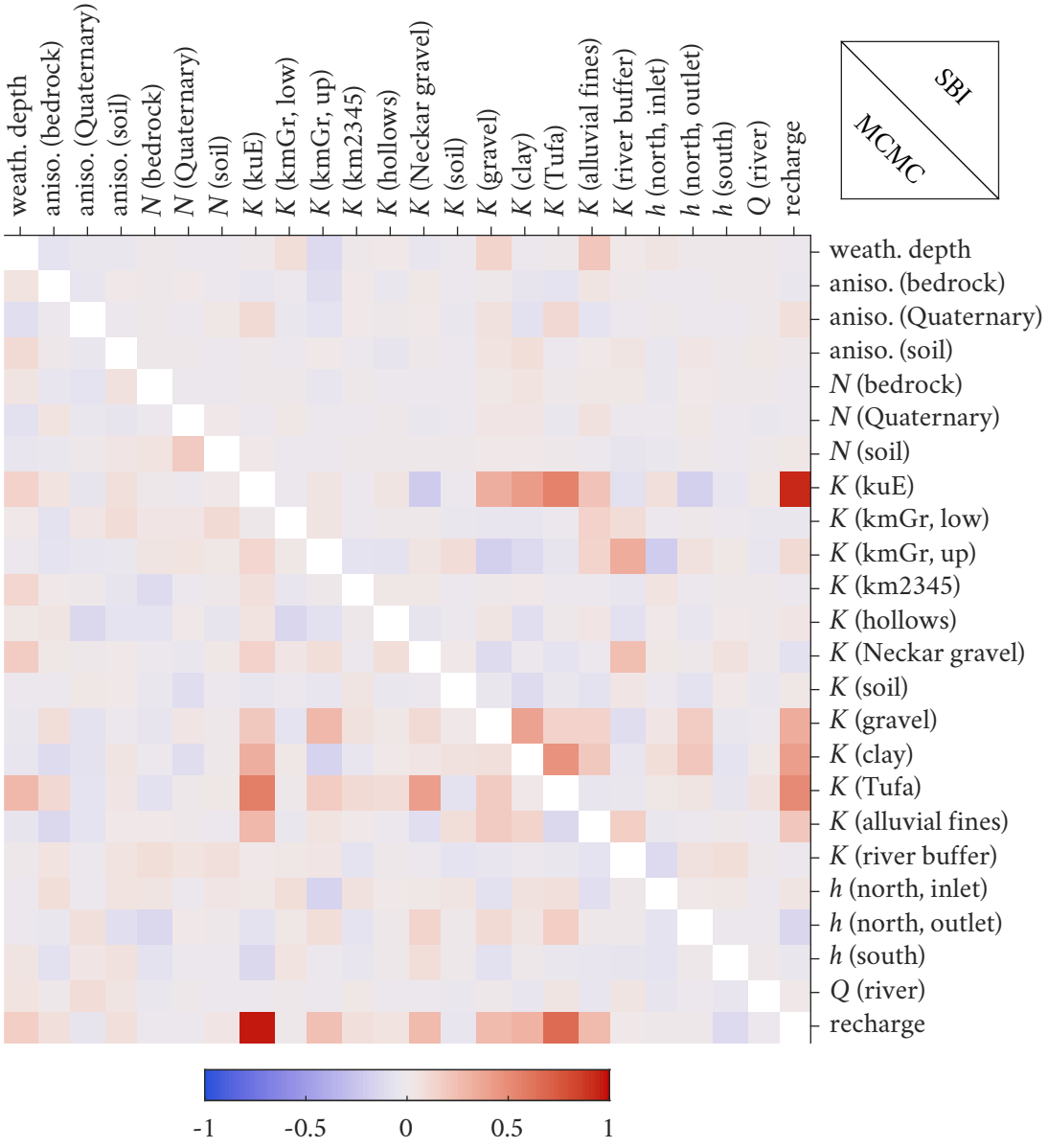

Figure 34: Correlation coefficients between scaled posterior parameters $\tilde{p}$ as determined by the SBI (upper triangular matrix) and MCMC (lower triangular matrix) approaches.

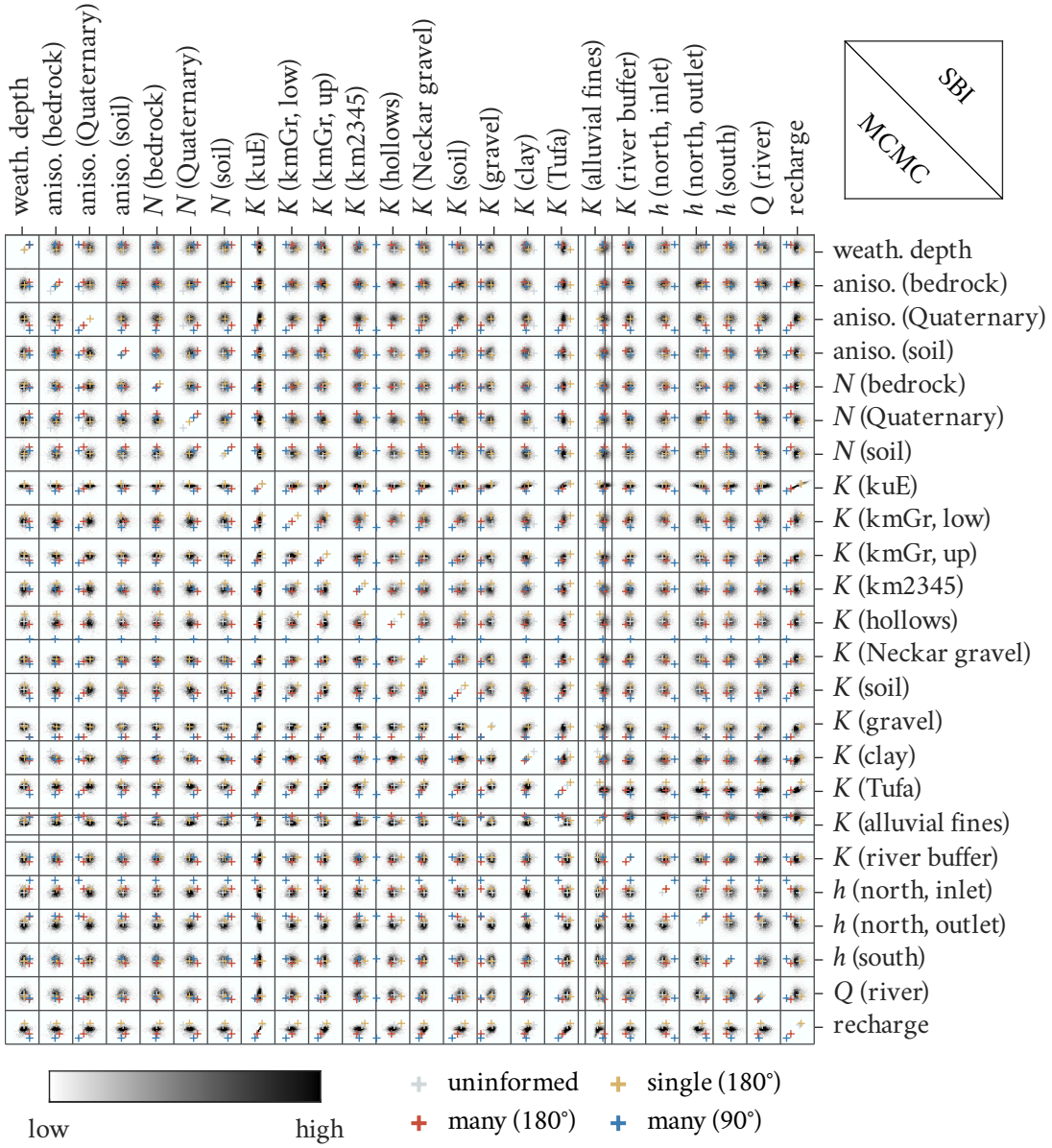

Figure 35: Bivariate marginal density plots of posterior distributions (SBI for upper triangular matrix; MCMC for lower triangular matrix). The axes reflect scaled parameters (i.e., $\tilde{p}$). Overlain are the final best points of the three optimization schemes. All local axes cover the parameter space from −8.5 to 8.5, with exception of the alluvial fines' hydraulic conductivity. Here, additional gray lines indicate these limits, as the SBI-posterior extends beyond them.

We see that the correlation coefficient close to one originates from a nearly perfect one-dimensional, linear relationship between the hydraulic conductivity of the Erfurt formation and the recharge rate (in both posterior distributions). This points out a moderate ill-posedness of the calibration problem in the way the model is currently defined: In steady-state simulations, where all conductivities and the recharge rate are freely allowed to change, only the ratio of conductivities to recharge rate can be inferred from a given set of hydraulic-head observation data. A higher recharge rate can always be balanced by higher hydraulic conductivities. Of course this effect is mitigated by the definition of prior distributions (which means that the parameters are not completely free to change). Still, there seems to be a comparably wide range of allowable recharge rates, where each specific rate requires the hydraulic conductivity of the Erfurt formation to be in a comparably narrow range.

The final best points of the four calibration variants fall onto this linear relationship, and even the intermediate best points follow it closely (not shown). However, the single-GPE variant and multi-GPE variants seem to have found points on opposite ends of the line. This explains why the best point estimates (colored lines in Figure 33) do not agree on the hydraulic conductivity of the Erfurt formation, even though (1) the three informed parameter sets had a comparable performance and (2) this parameter was recognized as important in all cases. Apparently, the three points found through the informed schemes (and to some extent even the best point found by the uninformed scheme) are still part of the same posterior distribution.

No other bivariate distribution shows such a narrow region of high posterior density. However, as already previously noted, similar relationships can be seen for many other combinations of hydraulic conductivity and recharge rate (e.g., Tufa), albeit not as sharply defined.

There do not seem to be any obvious nonlinear relationships that were previously missed by the analysis of the correlation coefficients. However, it cannot be ruled out that there are higher-dimensional relationships that are still masked.

15.3 Sensitivity Analysis

There is one additional tool available to reinforce the drawn conclusions: We can extract derivative information from the GPE proxy-models, to obtain global sensitivity measures. Using the final GPEs (that summarize all full model runs across the optimization schemes), we retrieve the Jacobian (i.e., the derivative of all observations with respect to all parameters) for the MCMC-derived posterior sample of 250 points selected for a full model evaluation. Averaging all entries across these realizations results in a global posterior sensitivity matrix that is visualized in Figure 36.

Figure 36a shows relative sensitivities for each individual observation (i.e., the sensitivity magnitude of the most important parameter was used to normalize the sensitivities of all other parameters within a row). This allows sensitivity comparisons of parameters for any given observation. Figure 36b shows the maximum absolute sensitivity of a specific observation (i.e., row) in m per normalized parameter unit. This allows sensitivity comparisons across the different observations.

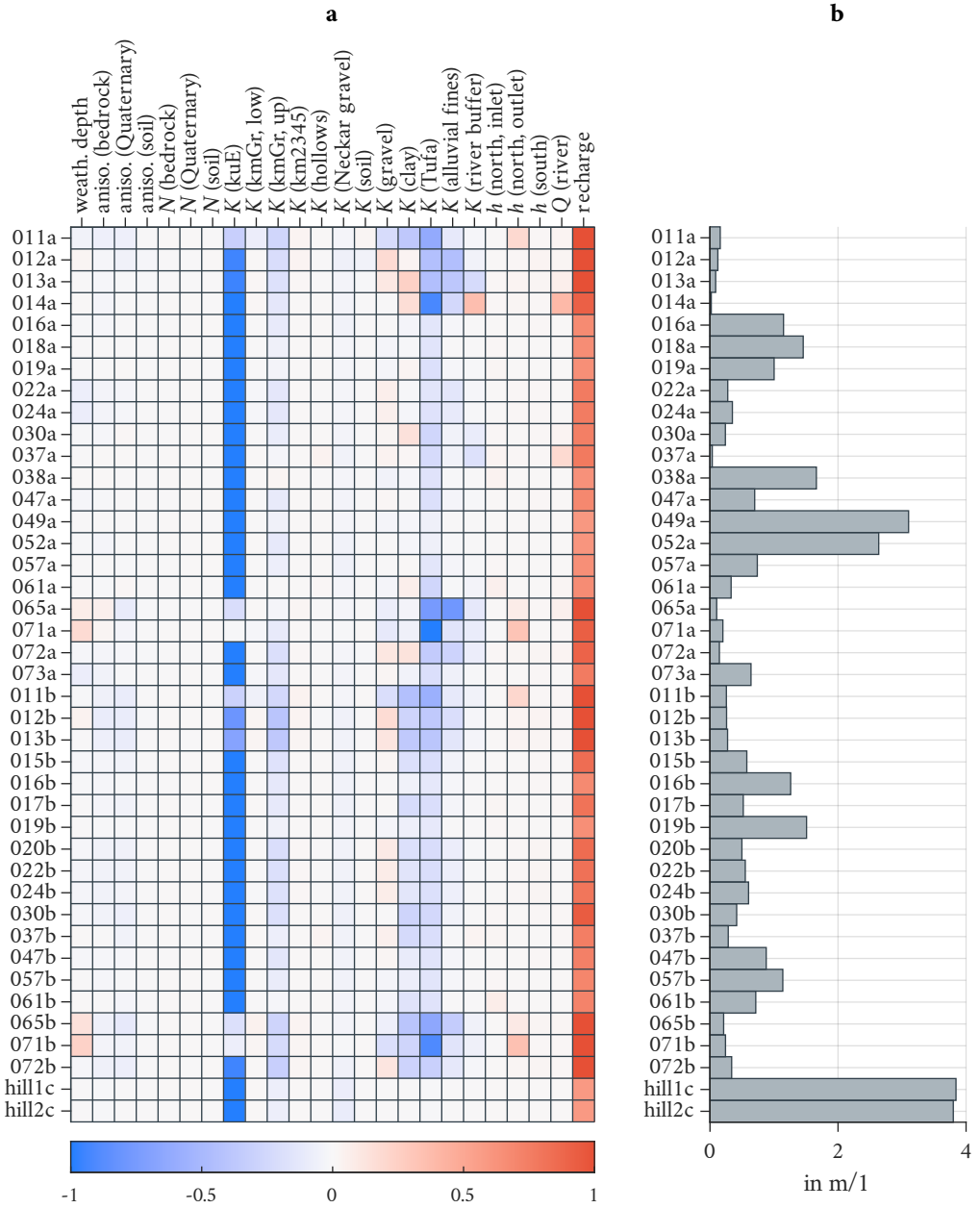

Figure 36: Mean sensitivities of posterior samples as determined by GPE differentiation (in the scaled parameter space $\tilde{p}$). **a**: Sensitivities normalized for each observation well, where the magnitude reveals the relative importance of a parameter for a given observation, and the sign tells whether an increase of the parameter would lead to an increase of observed hydraulic head. **b**: Mean absolute value of sensitivities across all wells and parameters. Larger values indicate a higher overall sensitivity of an observation to the input parameters.

In terms of absolute magnitude, few wells (most prominently "hill1c" and "hill2c", but also GWM049a and GWM052a) dominate the sensitivity matrix with values larger than 2.0 m per normalized parameter unit. Most of the remaining absolute sensitivities are in the range from 0.25 m to 1.0 m. Most likely, this stems from the fact that these two wells are not installed in the valley, but instead at a deep location up on the Wurmlingen saddle ("hill1c" and "hill2c"), or at the outer, southern fringe on the floodplain towards the hillslope (GWM049a and GWM052a). Deviations of hydraulic conductivities can result in more dramatic changes in hydraulic head here, while the observations in the valley are stabilized by head-related boundary conditions. Additionally, the groundwater table on the hill has more vertical space to shift within. For most of the valley wells on the other hand, this vertical freedom is restricted by the comparably shallow surface elevation.

With respect to the relative sensitivities for individual observations, the recharge rate is an important parameter for all of them. Higher recharge rates lead to an increased hydraulic head, which is consistent with basic hydrogeologic knowledge. Most of the other important parameters are hydraulic conductivities. The Erfurt formation evidently is important for all observation wells in this regard. For the shallow groundwater wells in the Ammer valley, the Tufa and the alluvial-fines layer are more important than gravel and clay. For the deeper wells, gravel and clay also play a role. This makes sense considering that gravel and clay are overlain by the Tufa and alluvial-fines layers.

Most of the sensitivities related to hydraulic conductivities are negative, implying a head decrease with increasing parameter value. For some observations, however, the mean sensitivity of the gravel and clay hydraulic conductivities is positive. This implies a counterintuitive head increase with increasing parameter value. The reason for that is unclear. Perhaps, an increased hydraulic conductivity allows more water to enter the aquifer at the groundwater inlet. However, this would also allow a better drainage towards the groundwater outlet and it is unclear why that effect should only be visible for the deeper hydrostratigraphic units in the floodplain.

The weathering depth only plays a minor role for most of the observations, with exception to GWM065a/b and GWM071a/b close to the eastern boundary. Here, an increase of the weathering depth leads to an increased hydraulic head, which is counterintuitive at first. However, at these locations the domain is comparably thin and basically the entire Grabfeld formation is assumed to be weathered anyway. Increasing the weathering depth even further might connect these locations hydraulically to a other regions, resulting in higher observed pressures.

As expected, the observations are not sensitive to the van Genuchten parameters (i.e., N) at all. The relevance of the soil layer and the anisotropies is equally negligible. Only those few observations close to the lateral inlet and outlet boundaries (namely GWM011a/b, GWM038a, GWM61a/b, GWM65a/b and GWM71a/b) show a visible dependence on the respective head offset parameters.

This sensitivity analysis is limited by the quality of the GPE derivatives and over-interpretation should be avoided. Still, the hydraulic conductivity of the Erfurt formation and the recharge rate seem to be the governing parameters, which explains their comparably small posterior uncertainties.

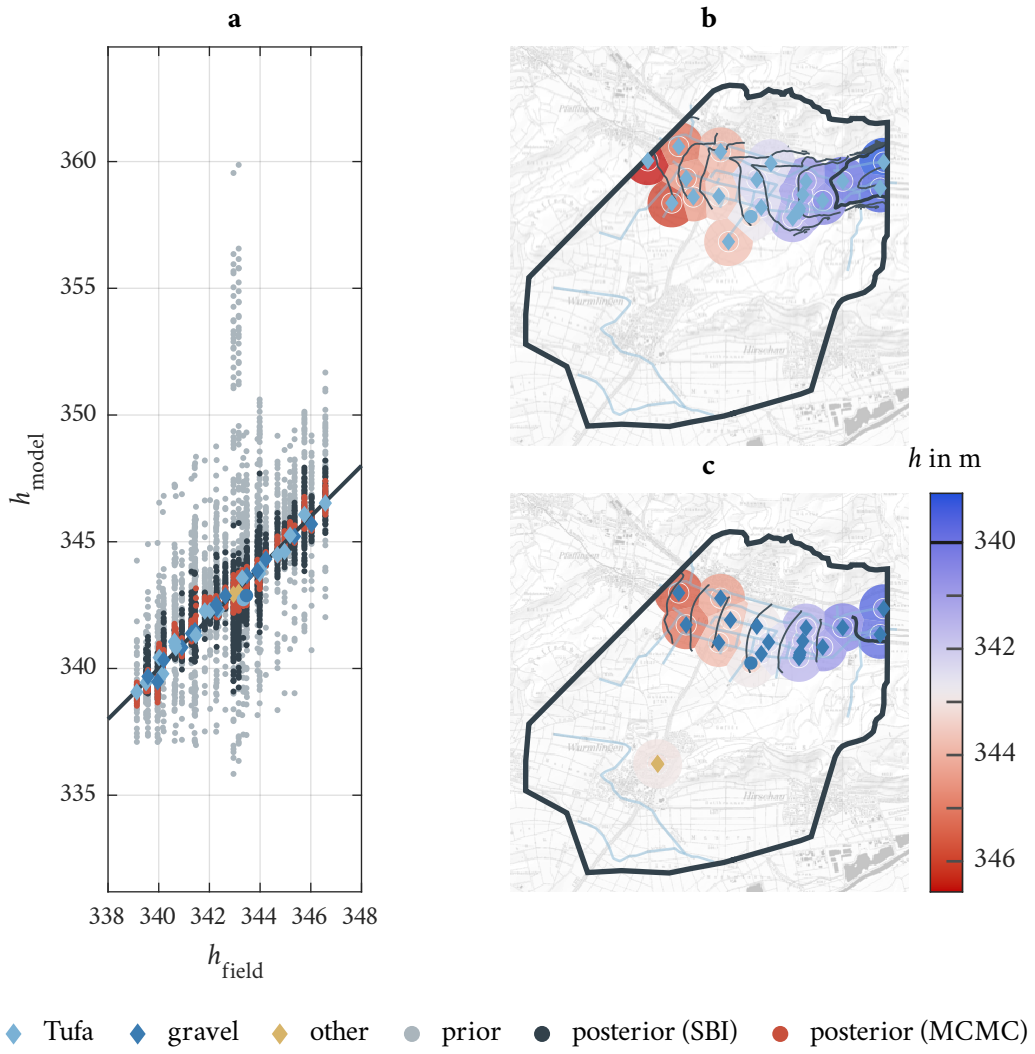

Figure 37: Calibrated flow field. **a**: Comparison of modeled hydraulic heads and the corresponding field observations including the best calibration estimate, and 250 realizations each of the prior, the SBI- and the MCMC-based posterior distribution. **b**: Visualization of the shallow flow field, with contour lines of hydraulic head and concentric rings symbolizing the measured (inner circle) and modeled (outer ring) hydraulic heads. **c**: Same visualization for all deep observations.

15.4 Flow Field of Calibrated Model

After analyzing the system behavior in terms of how the model parameters affect the observations, we now want to focus on the resulting flow field after calibration. Figure 37 provides an overview based on the best parameter set found across all calibration variants. As already stated, this ultimately stems from the variant with multiple internal GPEs and an opening angle of $\gamma = 90°$. Figure 37a shows modeled versus measured observations in comparison to the identity line. For further points of reference, 250 samples of modeled output from the prior and the posteriors (SBI and MCMC; run with the full model) are added.

We observe that the modeled hydraulic head observations scatter around the field data for all three cases (prior, SBI-posterior and MCMC-posterior). The amount of scattering is not uniform

across the different measurement locations Especially the modeled prior observations related to the observations "hill1c" and "hill2c" ("other") show a much larger variability than the others. As already mentioned in the context of the exceptionally large sensitivities of these observations to input parameters, this is probably related to the location of the corresponding wells far away from boundary conditions. The observations generated from SBI-derived posterior samples scatter closer around the identity line than their prior counterparts. This confirms that the NPE was at least partially successful in finding better-performing regions of the parameter space. However, the realizations generated with the MCMC-based verification procedure are even closer to the identity line, which indicates a superior performance.

The performance of the best parameter set identified by the calibration procedure is remarkably good, compared to the scattering of the prior and posterior distributions. The differences between modeled hydraulic heads and the corresponding field measurements (i.e., the residuals) are small and do not show any obvious visual patterns. For instance, the number of points with positive and negative residuals does not drastically differ. There is also no single observation that has a much larger residual than the other ones, and the residuals do not seem to depend on the observed value itself. This it indicates a favorable lack of obvious bias. Obviously, not all measurements are met with the accuracy of typical measurement uncertainties (in the order of few centimeters). However, this is not surprising, considering (1) that we use a steady-state model with homogeneous layers to simulate a snapshot from a transient, heterogeneous, real system, and (2) that the measured elevation of the piezometers themselves might be inaccurate.

Figure 37 also provides a spatial overview of the modeled data, separated by shallow (Figure 37b) and deep (Figure 37c) observation wells. For a facilitated visual comparison between the calibrated flow field and the flow field as it was observed on the key-date (November 6, 2018), we display two concentric rings for each observation. The inner ring represents the measured data, the outer ring the calibrated modeled output. Wherever rings would intersect, we divide the space up by Voronoi tessellation (Brassel and Reif, 1979), leading to a nearest-neighbor visualization in regions with high spatial density of measurement points. On top of that, we provide contour lines of hydraulic head extracted from the model's flow field within the Tufa and Ammer gravel units.

In general, the comparison shows that both, the shallow and the deep flow field are decently represented by the model. In both cases, the dominant hydraulic gradient in the longitudinal direction of the floodplain is clearly visible. The model was also able to reproduce the northwards component of hydraulic gradient in the southern central section of the floodplain, which can be attributed to lateral hillslope contributions (this can be seen more clearly in the shallow Tufa aquifer). Nonetheless, this central part is the region of strongest disagreement between model and reality. One location stands out particularly (GWM016a/b, denoted by a circular symbol), where the model underestimates the hydraulic head by about 0.5 m. A possible reason for this discrepancy might be an insufficient representation of the transition zone between hillslope and floodplain aquifers, as GWM016a/b is located at the southern fringe of the floodplain. Even though hillslope hollows were

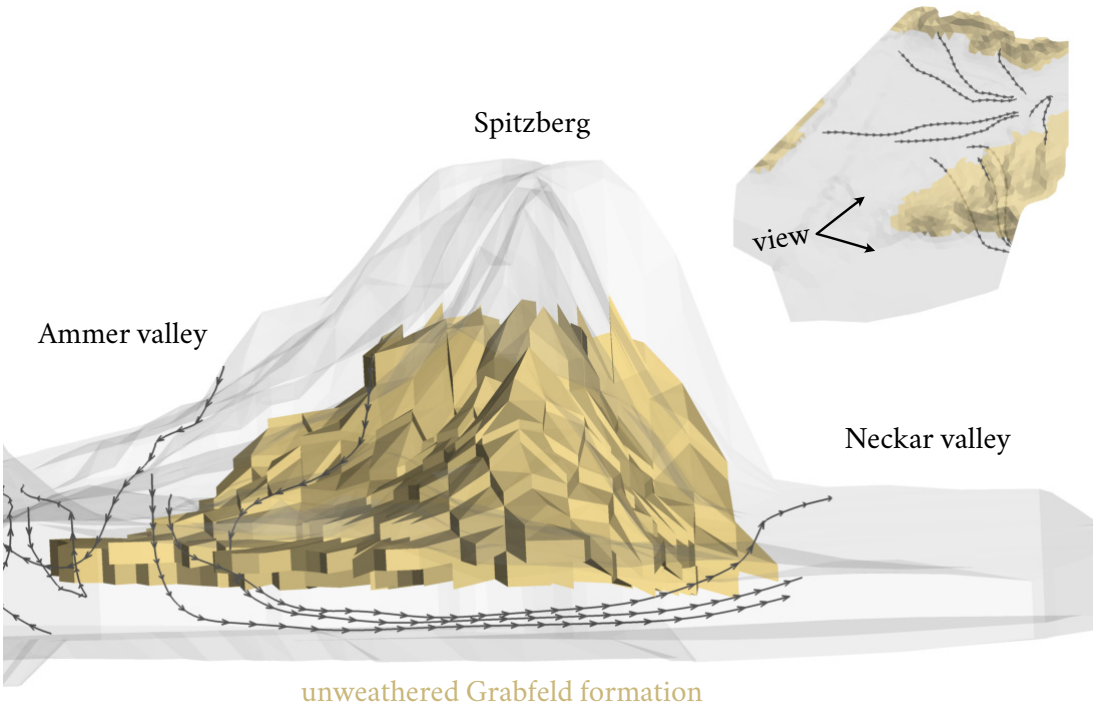

Figure 38: Example streamlines show inter-basin flow diving beneath the unweathered Grabfeld formation (fivefold vertical exaggeration).

explicitly added as separate hydrostratigraphic units (one of them is very close to GWM016a/b), the model might still be oversimplified in this regard. In reality, the transition between hillslope and floodplain is probably smoother than the sharp material property zones of the model allow. The observations of GWM016a/b might be affected by that. This outcome suggests that it might be worthwhile to closely investigate the transitioning zone between the hillslope and the floodplain.

The model was able to capture the head differences between the shallow Tufa and deeper gravel aquifer. This is visible for those observations at the northern inlet that operate in both depths, but also in the contour lines throughout the aquifers. Unsurprisingly, we also see that the shallow Tufa aquifer is much more affected by the Ammer river. The contour lines indicate gaining conditions for most river stretches (this is confirmed by raw nodal model output of HGS), with minor local infiltrations (e.g., in the northern reach right before the confluence of the two main reaches).

Another aspect that is different between gravel and Tufa aquifer is the steepening of hydraulic gradients at the northern and southern floodplain fringes. This is very pronounced for the Tufa aquifer, especially towards the eastern end of the domain. These steep hydraulic gradients indicate lateral inflow from the hillslopes. The gravel aquifer however does not show that, but it also does not have the same lateral extent as the Tufa aquifer. The steepening could simply take place in the adjacent bedrock formations instead.

Figure 38 shows three-dimensional streamlines that illustrate an important aspect of the flow field that was already hypothesized in Section 15.2. We can see pronounced inter-basin flow from the

Ammer side to the Neckar side in the Erfurt formation. The unweathered Grabfeld formation seems to be such a hindrance to flow that streamlines originating on the Ammer side of the Spitzberg first move towards the Ammer floodplain, before they enter the Erfurt formation, shift direction and dive towards the Neckar valley. Of course, this has implications for the location of the groundwater divide. It also implies that not the Grabfeld formation is critical in this regard, but the connection between the two valleys is provided by the Erfurt formation instead.

16 Conclusions & Outlook

We successfully calibrated a steady-state subsurface flow model of the Ammer floodplain and the adjacent section of the Neckar catchment. Using multiple internal GPEs during the calibration helped to improve and accelerate the calibration. Restricting the opening angle of proposed points from 180° to 90° has lead to an additional minor enhancement. We also applied Simulation-Based Inference to infer a full posterior distribution, which helped us to understand the connection between the calibrated parameter sets and also provided estimates on parameter uncertainty. The main properties of this distribution were in accordance with another posterior distribution derived from classical Bayesian MCMC sampling conducted with a GPE-based proxy-model. Considering that the NPE only used 3070 comparably low-performing realizations sampled from the prior distribution, these results are impressive. However, some minor deviations were noticeable, and realizations drawn from the MCMC-posterior outperformed their SBI equivalents in terms of agreement with measured data. Depending on the calibration problem at hand, the application of NPE should therefore be evaluated carefully.

For instance, if the reason for constructing a posterior distribution is to obtain a rough idea about parameter values, uncertainties, and relationships (e.g., as part of a preliminary study) performing NPE on a model sample generated from prior parameter distributions is straightforward and might be sufficient. If details of the posterior distribution are important, or if well-performing samples of the posterior distribution are required, it might be worthwhile to (1) produce additional realizations in promising regions of the parameters space (e.g., by means of a global calibration scheme), (2) train a high-quality proxy-model, and (3) perform MCMC sampling with that proxy-model. Another possible route could be the application of other SBI algorithms (e.g., Papamakarios et al., 2019; Hermans et al., 2020), which were beyond the scope of this study.

The comparison across the different global calibration variants, posteriors and GPE-based sensitivity analysis produces a consistent picture that allows the following conclusions about the hydrogeological system:

- The hydraulic conductivity of the Erfurt formation is the most important material property. This formation seems to take a special role in the modeled system: As a connector between the Ammer and Neckar valleys it acts as a facilitator for inter-basin flow, which is necessary to keep the groundwater levels in the Ammer valley at the observed values.

- The recharge rate is a crucial parameter that affects the measured flow field at all observation wells, but it cannot be inferred independently of the hydraulic conductivity of the Erfurt formation. Instead, there is a close relationship, where a given value of that conductivity allows only a narrow range of recharge rates.
- The hydraulic conductivities of the Quaternary material are also important for most of the observation wells. The calibrated model requires comparably high values for the alluvial fines, which might be the outcome of a parameter compensating for a structural model error, namely the absence of tile drains in the model.

The largest discrepancies between the calibrated flow field and measured data occur towards the southern hillslope in the central region of the Ammer floodplain. A refinement of the model might be necessary in this part of the domain. We also propose to conduct another key-date measurement targeting all available observation wells in the domain, including the wells recently installed in the Grabfeld formation. As these wells are systematically probing the hillslope far from other observation locations, the corresponding data might be especially valuable for model calibration at the hillslope/floodplain interface.

Some aspects of the investigation raised further questions: For instance, the calibrated hydraulic conductivities of Tufa and gravel do not match the prior information in a qualitative sense. Data from multi-well pumping tests could help to assess whether this outcome is true or a mere modeling artifact (potentially caused by the positive global sensitivities associated with the hydraulic conductivity of the Ammer gravel).

The presented multi-GPE calibration variants are not transferable to transient model calibration, because each time point for each well would essentially require a single GPE, as GPEs can only predict scalar quantities. This is unfeasible, as meta proxy-model training time and prediction time increase linearly with the number of GPEs and even with parallel computing there is a limit that is quickly reached already with a very coarse temporal resolution. However, NPE and SBI have already been successfully applied to transient data and high-dimensional model outputs (Lueckmann et al., 2017; Gonçalves et al., 2020). This makes the field of SBI a promising candidate for inferring posterior parameter values of a transient version of the presented model. One key to that issue might be the development of applicable and valid summary metrics from time series data, as they are used for example in Lueckmann et al. (2017). This could potentially re-enable the use of the presented GPE-assisted schemes.

With respect to the presented calibration scheme variants, we propose a rigorous investigation of how the opening angle affects the calibration efficiency. This might include different calibration runs with various opening angles that might even decrease over the course of the calibration. Repeated runs of these scheme variants, also in comparison to the schemes developed by Regis and Shoemaker (2009), Wang and Shoemaker (2014), and Xia et al. (2021) would be ideal. However, we propose to perform this comparison with a simpler model to keep the computational costs tractable.

Chapter V

Overall Conclusion

Main Findings In the general introduction (Sections 1.3 and 3.5) I raised some overarching questions that guided my doctoral research. In this section, I would like address these questions by briefly putting the most important results in context again, before concluding this dissertation with a general outlook.

1. *How can we determine where to measure hydraulic head in order to reduce the uncertainty in delineating a groundwater divide?*

In Chapter III, I formulated an optimization problem whose solution is the piezometer configuration that minimizes the uncertainty in the location of a local groundwater divide. The quantification of this uncertainty is based on particle-tracking information interpreted as (binary) particle fate maps. By means of stochastic, numerical, subsurface-flow modeling, the effect of new hydraulic-head information obtained at planned measurement locations on the uncertainty can be assessed, even though the corresponding data are unknown at the stage of the analysis. A trivial ranking of different measurement configurations with respect to the corresponding uncertainty reduction results in the solution of the optimization problem. To accelerate the stochastic modeling framework, a plausibility-based pre-filtering process can be used.

2. *Under which conditions can the lateral widening and narrowing of floodplain aquifer(s) cause valley-scale hyporheic exchange?*

In Chapter II, I derived a semi-analytical model for two-dimensional flow in idealized widening and narrowing floodplain aquifers. This model allows identifying and quantifying geometry-driven valley-scale lateral hyporheic exchange. I used the semi-analytical solution to perform a systematic evaluation of hyporheic zone characteristics (exchange flux, affected area and travel-time distribution). This has resulted in simplified proxy-models that reveal mechanistic insights in how these characteristics depend on various geometric and hydrogeologic properties. The main conclusion from this study is that the lateral widening and narrowing of floodplain aquifers causes valley-scale hyporheic exchange. This process is driven by the width-difference between the aquifer's widest and narrowest section, and regulated by the separation distance between the floodplain boundary and the river. Influxes from the adjacent hillslopes can shrink the hyporheic exchange zone disproportionately. If such influxes exceed a threshold depending on the domain aspect ratio and the hydraulic anisotropy, this zone even collapses completely such that no valley-scale hyporheic exchange can be observed.

3. *How does NPE compare to proxy-model-based MCMC sampling of a posterior parameter distribution after global calibration of a computationally expensive subsurface-flow model?*

In Chapter IV, I applied different variants of an efficient global calibration scheme to a computationally expensive subsurface-flow model. I compared the resulting best-point estimates with full posterior parameter distributions obtained through NPE, a Simulation-Based Inference method, and MCMC-based sampling of a proxy-model. In general, these posterior distributions aligned well with the results obtained from the calibrations. Furthermore, in the investigated example, they revealed additional information about model parameters in terms of uncertainties and correlations that is in principle unavailable from global calibration outcomes. The two posterior sampling mechanisms produced qualitatively similar results and were able to identify the main correlations. However, realizations drawn from the MCMC-based posterior distribution yielded superior agreement with field observations compared to a similar set generated through NPE.

4. *Is geometry-driven valley-scale hyporheic exchange relevant for the Ammer-floodplain aquifers?*

Applying both, the original semi-analytical solution and the derived proxy-models of Chapter II to the Ammer site, helped to address this question. Even though the Ammer floodplain geometry would have some potential for valley-scale hyporheic exchange (pronounced widening combined with a comparably small average separation distance between river and floodplain boundary), the results indicate that this process is not relevant in terms of hyporheic discharge and affected area. This is the case, because lateral influxes from the southern hillslope push the exchange zone towards the river, which leads to a disproportionate shrinkage of the affected area. The remaining hyporheic zone is characterized by small discharge and large travel times on the order of decades, which are both a direct result of the comparably small hydraulic conductivity in the Ammer floodplain aquifers. Due to the long time that water parcels on the hyporheic flow paths originating from river Ammer spend in the floodplain aquifer, the physical and chemical signature of the surface water would most likely be lost. As a result, water in the hyporheic zone would probably resemble groundwater even at comparably small distances from the river.

5. *Is the groundwater divide between the Ammer and the Neckar valley in vicinity of the Ammer floodplain shifted, leading to inter-basin flow?*

With respect to the Ammer floodplain site, the analysis of Chapter III showed that the uncertainty in the groundwater divide's position is largest on the Wurmlingen saddle. A potential shift of the groundwater divide by several hundred meters towards the Ammer side can be observed in at least a major fraction of the randomly generated model realizations. The sensitivity analysis conducted in Chapter III showed that the hydraulic heads at the Wurmlingen saddle, and therefore the position of the groundwater divide, depend mostly on the Grabfeld formation and its weathering state in particular. In the refined model version of Chapter IV, however, the Erfurt formation takes a leading role with respect to governing the overall flow field. This is an outcome of updated prior parameter distributions based on existing literature, as well as refined model geometries and boundary conditions. Nonetheless, inter-basin flow from the northern Ammer site towards the southern Neckar catchment was observed in the calibrated subsurface-flow model.

6. *Can a steady-state subsurface flow model be calibrated to field data of the Ammer-floodplain aquifers to achieve a decent representation of the observed flow field?*

In Chapter IV, I carried out the calibration of a steady-state model of the Ammer floodplain to hydraulic-head data of a key-date measurement. The application of different variants of a proxy-model-assisted calibration scheme resulted in several similarly-performing parameter sets, that all achieved a decent fit between modeled and observed hydraulic-head data (with exception of an uninformed variant that served as a comparison). The calibrated flow field obtained from running the model with the single best parameter set replicates properties identified through measurements, like a head difference between the Tufa and gravel aquifers. Lateral influxes from the southern hillslopes are visible in the contour lines, too. Interactions with river Ammer in terms of mostly exfiltrating conditions (i.e., river gains water from the subsurface) have been revealed for the Tufa aquifer.

7. *What is the role of the hydrostratigraphic units in the Ammer-floodplain aquifers, also with respect to the interaction of the floodplain aquifers with the surroundings?*

As already discussed in context of Question 3, the posterior parameter distribution obtained through Neural Posterior Estimation and a sensitivity analysis based on Gaussian Process Emulators have helped to understand the function and significance of individual hydrostratigraphic units in Chapter IV. The Erfurt formation, for instance, serves as a hydraulic connector between the Ammer and Neckar valleys, beneath the unweathered Grabfeld formation, which is basically watertight. This also implies a northward shift of the groundwater divide. The top alluvial fines in the Ammer floodplain have also been identified to be of importance for the flow field, most likely as they represent the connection between the floodplain aquifers and the drainage system.

Outlook The semi-analytical and stochastic numerical methods for the simulation of subsurface flow in floodplains presented in this dissertation could be extended in several regards. For example, it would be interesting to investigate how the valley-scale hyporheic exchange in floodplain aquifers of varying width would be affected by sinuosity in the river course and by non-uniformity in the river slope. Such changes are trivial to implement in the conceptual model, but the resulting mathematical problem might become intractable without relying on fully numerical models. With respect to the flow field, these changes would probably result in additional superposition of hyporheic exchange on a smaller scale.

For the methods introduced in Chapters III and IV, an extension to transient systems would be very useful, but also difficult to implement. In the case of optimizing well placement for groundwater divide delineation I already sketched out a possible idea of how to address this issue in Section 9.5. Applying the proxy-model assisted calibration scheme to transient models, however, is not straightforward and might require the development of smart summary metrics that keep the number of internal proxy-models at a feasible level.

I want to end this dissertation with an outlook of how hydrogeologic research in the Ammer floodplain could be continued. Now that steady-state simulations have proven to be successful with respect to replicating patterns observed in the field, the logical consequence would be to extend the model to transient simulations. The overview of available field data (Figure 4) already shows how dynamic the aquifer system can be, and it would be interesting to see if and how these patterns can be matched by modeling. A calibrated transient flow model could give insights and detailed answers on questions regarding the water balance of the Ammer floodplain (e.g., whether the connection between aquifers and the drainage network is active primarily during rain events or throughout the year). However, developing and calibrating a transient model to the Ammer-floodplain data is challenging. Transient models lead to a drastic increase of complexity, as more parameters are needed and relevant (e.g., the unsaturated zone parametrization is much more influential in transient models). The model run times would also increase and the definition of a good objective function judging the fit between measured and modeled time series data is also not trivial.

In my opinion, future modeling studies on the Ammer floodplain should therefore also consider a modification of the model domain, depending on the modeling focus. If the primary interest regards the floodplain Quaternary itself, I would propose an extension of the domain area laterally to both ends of the floodplain (from Pfäffingen or even Poltringen, to the western end of the city of Tübingen). This alleviates problems with the arbitrarily defined boundary geometries and the corresponding boundary conditions. At the same time, it could be worthwhile to restrict the investigation to the Quaternary floodplain aquifer units themselves. Interactions with hillslopes and the deeper subsurface could then be considered by source and sink terms (e.g., by including lateral Neumann boundaries and a leaky boundary towards the bottom). Potentially, this approach might even allow to consider some internal heterogeneity. Alternatively, if the focus lies on the interaction between the Ammer floodplain and the regional hydrogeology, an extension of the domain in all directions accompanied by a decreased resolution (possibly also by lumping all Quaternary units) might be promising. Ultimately, the combination of both approaches might be ideal, for instance by coupling a lower-resolution regional scale model to a high-resolution Quaternary model.

Prior to and during any further modeling activities, I would encourage conducting further key-date measurements of hydraulic head at all available observation locations. This should include all Quaternary wells in the Ammer floodplain aquifers, but also the Grabfeld formation wells that were installed as an outcome of Chapter III, as well as other ones (e.g., Erfurt formation wells recently installed by the Ammertal-Schönbuchgruppe). Such nearly-simultaneous snapshots of the flow field are essential to understand the system.

References

Aigner, T. and G. H. Bachmann (1992). "Sequence-Stratigraphic Framework of the German Triassic." In: *Sedimentary Geology* 80.1, pp. 115–135. ISSN: 0037-0738. DOI: 10.1016/0037-0738(92)90035-P.

Allgeier, J. (2022a). *Code for Inferring Unsaturated Zone Parameters From Saturated Hydraulic Conductivity*. OSF. DOI: 10.17605/OSF.IO/9ZYCB.

Allgeier, J. (2022b). *Raw Data, Code and Plotting Scripts for "Proxy-Model Assisted Calibration of a Steady-State Subsurface Flow Model"*. OSF. DOI: 10.17605/OSF.IO/UPTXD.

Allgeier, J., A. González-Nicolás, D. Erdal, W. Nowak, and O. A. Cirpka (2020). "A Stochastic Framework to Optimize Monitoring Strategies for Delineating Groundwater Divides." In: *Frontiers in Earth Science* 8. ISSN: 2296-6463. DOI: 10.3389/feart.2020.554845.

Allgeier, J., A. González-Nicolás, D. Erdal, W. Nowak, and O. A. Cirpka (2022). *Raw Data, Code and Plotting Scripts for "A Stochastic Framework to Optimize Monitoring Strategies for Delineating Groundwater Divides"*. OSF. DOI: 10.17605/OSF.IO/AYB58.

Allgeier, J., S. Martin, and O. A. Cirpka (2021a). *Code for the Systematic Evaluation of Geometry-Driven Lateral River-Groundwater Exchange in Floodplains*. OSF. DOI: 10.17605/OSF.IO/FYKR9.

Allgeier, J., S. Martin, and O. A. Cirpka (2021b). "Systematic Evaluation of Geometry-Driven Lateral River-Groundwater Exchange in Floodplains." In: *Water Resources Research* 57.8, e2021WR030239. ISSN: 1944-7973. DOI: 10.1029/2021WR030239.

Amaranto, A., F. Munoz-Arriola, G. Corzo, D. P. Solomatine, and G. Meyer (2018). "Semi-Seasonal Groundwater Forecast Using Multiple Data-Driven Models in an Irrigated Cropland." In: *Journal of Hydroinformatics* 20.6, pp. 1227–1246. ISSN: 1464-7141. DOI: 10.2166/hydro.2018.002.

Ammer, U., G. Einsele, W. Arnold, O. Klee, R. Agerer, G. Agster, U. Babel, J. Behringer, W. Böcking, F. .-H. Evers, W. Fleck, H. Günzl, K. .-F. Hofmann, H. Janz, J. Körner, I. Kottke, V. Kracht, A. Krebs, K. Kunzweiler, W. .-D. Langbein, H. .-U. Moosmauer, S. Müller, E. Nickel, S. Pfeiffer, R. Rausch, H. Schmidt-Witte, O. Schwarz, and R. Stegmayer (1983). "Wasserhaushalt, Stoffeintrag, Stoffaustrag und biologische Studien im Naturpark Schönbuch bei Tübingen." In: *Forstwissenschaftliches Centralblatt* 102.1, pp. 282–324. ISSN: 1439-0337. DOI: 10.1007/BF02741862.

Anderson, M. P., W. W. Woessner, and R. J. Hunt (2015). *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. Academic press. ISBN: 0-08-091638-4.

Anderson, T. W. (1962). "On the Distribution of the Two-Sample Cramer-von Mises Criterion." In: *The Annals of Mathematical Statistics* 33.3, pp. 1148–1159. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177704477.

Archer, N., M. Bonell, N. Coles, A. MacDonald, C. Auton, and R. Stevenson (2013). "Soil Characteristics and Landcover Relationships on Soil Hydraulic Conductivity at a Hillslope Scale: A View towards Local Flood Management." In: *Journal of Hydrology* 497, pp. 208–222. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2013.05.043.

Asbrand, M., H. Frisch, B. Hanauer, R. Ludwig, H. Mikat, C. Mikulla, T. Oswald, R. Rausch, J. Riegger, C. Schöpfer, et al. (2002). "Hydrogeologische Modelle: Ein Leitfaden Mit Fallbeispielen." In: *Schriftenreihe Der dt Geol Ges* 24, p. 120.

Awan, N. M. and T. O'Donnell (1972). "Moving Water Tables in Tile-Drained Soils." In: *Journal of the Irrigation and Drainage Division* 98.3, pp. 459–477. DOI: 10.1061/JRCEA4.0000882.

Baillieux, A., D. Campisi, N. Jammet, S. Bucher, and D. Hunkeler (2014). "Regional Water Quality Patterns in an Alluvial Aquifer: Direct and Indirect Influences of Rivers." In: *Journal of Contaminant Hydrology* 169, pp. 123–131. ISSN: 01697722. DOI: 10.1016/j.jconhyd.2014.09.002.

Bakker, M. (1999). "Simulating Groundwater Flow in Multi-Aquifer Systems with Analytical and Numerical Dupuit-models." In: *Journal of Hydrology* 222.1, pp. 55–64. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(99)00089-X.

Bakker, M. (2006). "An Analytic Element Approach for Modeling Polygonal Inhomogeneities in Multi-Aquifer Systems." In: *Advances in Water Resources* 29.10, pp. 1546–1555. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2005.11.005.

Bakker, M. and O. D. L. Strack (2003). "Analytic Elements for Multiaquifer Flow." In: *Journal of Hydrology* 271.1, pp. 119–129. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(02)00319-0.

Barnes, R. and I. Janković (1999). "Two-Dimensional Flow through Large Numbers of Circular Inhomogeneities." In: *Journal of Hydrology* 226.3, pp. 204–210. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(99)00142-0.

Barthel, R. and S. Banzhaf (2016). "Groundwater and Surface Water Interaction at the Regional-scale – A Review with Focus on Regional Integrated Models." In: *Water Resources Management* 30.1, pp. 1–32. ISSN: 0920-4741, 1573-1650. DOI: 10.1007/s11269-015-1163-z.

Bates, P. D., M. D. Stewart, A. Desitter, M. G. Anderson, J.-P. Renaud, and J. A. Smith (2000). "Numerical Simulation of Floodplain Hydrology." In: *Water Resources Research* 36.9, pp. 2517–2529. ISSN: 1944-7973. DOI: 10.1029/2000WR900102.

Baxter, C. V., C. A. Frissell, and F. R. Hauer (1999). "Geomorphology, Logging Roads, and the Distribution of Bull Trout Spawning in a Forested River Basin: Implications for Management and Conservation." In: *Transactions of the American Fisheries Society*, p. 14. DOI: 10.1577/1548-8659(1999)128<0854:GLRATD>2.0.CO;2.

Bear, J. (1972). *Dynamics of Fluids in Porous Media*. https://agris.fao.org/agris-search/search.do?recordID=US201300010992. New York, American Elsevier Pub. Co. [1972]. ISBN: 978-0-444-00114-6.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood (2018). "Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution." In: *Scientific Data* 5.1, p. 180214. ISSN: 2052-4463. DOI: 10.1038/sdata.2018.214.

Beckers, F., A. Heredia, M. Noack, W. Nowak, S. Wieprecht, and S. Oladyshkin (2020). "Bayesian Calibration and Validation of a Large-Scale and Time-Demanding Sediment Transport Model." In: *Water Resources Research* 56.7, e2019WR026966. ISSN: 1944-7973. DOI: 10.1029/2019WR026966.

Bect, J., E. Vazquez, et al. (2022). *STK: A Small (Matlab/Octave) Toolbox for Kriging. Release 2.7.0*. https://github.com/stk-kriging/stk/.

Bedient, P. B., W. C. Huber, B. E. Vieux, et al. (2008). *Hydrology and Floodplain Analysis*. Vol. 816. Prentice Hall Upper Saddle River, NJ.

Beechie, T. and P. Roni (2012). *Stream and Watershed Restoration: A Guide to Restoring Riverine Processes and Habitats*. John Wiley & Sons. ISBN: 978-1-118-40663-2.

Berblinger, M. and C. Schlier (1991). "Monte Carlo Integration with Quasi-Random Numbers: Some Experience." In: *Computer Physics Communications* 66.2-3, pp. 157–166. ISSN: 00104655. DOI: 10.1016/0010-4655(91)90064-R.

Beven, K. (2006). "A Manifesto for the Equifinality Thesis." In: *Journal of Hydrology*. The Model Parameter Estimation Experiment 320.1, pp. 18–36. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2005.07.007.

BfG (2003). *Hydrologischer Atlas Deutschland*. https://geoportal.bafg.de/mapapps/resources/apps/HAD/index.html?lang=de.

Biron, P. M., T. Buffin-Bélanger, M. Larocque, G. Choné, C.-A. Cloutier, M.-A. Ouellet, S. Demers, T. Olsen, C. Desjarlais, and J. Eyquem (2014). "Freedom Space for Rivers: A Sustainable Management Approach to Enhance River Resilience." In: *Environmental Management* 54.5, pp. 1056–1073. ISSN: 0364-152X, 1432-1009. DOI: 10.1007/s00267-014-0366-z.

Bloxom, L. F. and T. J. Burbey (2015). "Determination of the Location of the Groundwater Divide and Nature of Groundwater Flow Paths within a Region of Active Stream Capture; the New River Watershed, Virginia, USA." In: *Environmental Earth Sciences* 74.3, pp. 2687–2699. ISSN: 1866-6299. DOI: 10.1007/s12665-015-4290-1.

Boano, F., J. W. Harvey, A. Marion, A. I. Packman, R. Revelli, L. Ridolfi, and A. Wörman (2014). "Hyporheic Flow and Transport Processes: Mechanisms, Models, and Biogeochemical Implications." In: *Reviews of Geophysics* 52.4, pp. 603–679. ISSN: 1944-9208. DOI: 10.1002/2012RG000417.

Boano, F., C. Camporeale, R. Revelli, and L. Ridolfi (2006). "Sinuosity-Driven Hyporheic Exchange in Meandering Rivers." In: *Geophysical Research Letters* 33.18. ISSN: 1944-8007. DOI: 10.1029/2006GL027630.

Boano, F., R. Revelli, and L. Ridolfi (2009). "Quantifying the Impact of Groundwater Discharge on the Surface–Subsurface Exchange." In: *Hydrological Processes* 23.15, pp. 2108–2116. ISSN: 1099-1085. DOI: 10.1002/hyp.7278.

Bokeh Development Team (2021). *Bokeh: Python Library for Interactive Visualization*. Manual. https://bokeh.org/.

Bornschein, A. and R. Pohl (2018). "Land Use Influence on Flood Routing and Retention from the Viewpoint of Hydromechanics." In: *Journal of Flood Risk Management* 11.1, pp. 6–14. ISSN: 1753-318X. DOI: 10.1111/jfr3.12289.

Boyraz, U. and C. M. Kazezyılmaz-Alhan (2013). "An Investigation on the Effect of Geometric Shape of Streams on Stream/Ground Water Interactions and Ground Water Flow." In: *Hydrology Research* 45.4-5, pp. 575–588. ISSN: 0029-1277. DOI: 10.2166/nh.2013.057.

Boyraz, U. and C. M. Kazezyılmaz-Alhan (2017). "Solutions for Groundwater Flow with Sloping Stream Boundary: Analytical, Numerical and Experimental Models." In: *Hydrology Research* 49.4, pp. 1120–1130. ISSN: 0029-1277. DOI: 10.2166/nh.2017.264.

Boyraz, U. and C. M. Kazezyılmaz-Alhan (2021). "An Analytical Solution for Groundwater Flow Incorporating the Effect of Water Bodies with Sloping Surfaces." In: *Hydrological Sciences Journal*. https://www.tandfonline.com/doi/abs/10.1080/02626667.2021.1925675. ISSN: 0262-6667.

Brassel, K. E. and D. Reif (1979). "A Procedure to Generate Thiessen Polygons." In: *Geographical Analysis* 11.3, pp. 289–303. ISSN: 1538-4632. DOI: 10.1111/j.1538-4632.1979.tb00695.x.

Bredehoeft, J. (2012). "Modeling Groundwater Flow—The Beginnings." In: *Groundwater* 50.3, pp. 325–329. ISSN: 1745-6584. DOI: 10.1111/j.1745-6584.2012.00940.x.

Bridge, J. S. (2009). *Rivers and Floodplains: Forms, Processes, and Sedimentary Record*. John Wiley & Sons. ISBN: 978-1-4443-1126-6.

Brookes, A. (1987). "The Distribution and Management of Channelized Streams in Denmark." In: *Regulated Rivers: Research & Management* 1.1, pp. 3–16. ISSN: 1099-1646. DOI: 10.1002/rrr.3450010103.

Brooks, R. H. and A. T. Corey (1964). *Hydraulic Properties of Porous Media*. Hydrology Papers 3. Fort Collins, Colorado: Colorado State University.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press. ISBN: 978-1-4200-7942-5.

Brunner, P. and C. T. Simmons (2012). "HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model." In: *Ground Water* 50.2, pp. 170–176. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2011.00882.x.

Brunotte, E., E. Dister, D. Günther-Diringer, U. Koenzen, and D. Mehl (2009). *Flussauen in Deutschland Erfassung Und Bewertung Des Auenzustandes*. https://rds-tue.ibs-bw.de/link?kid=1400186277. Bonn-Bad Godesberg: Bundesamt für Naturschutz. ISBN: 978-3-7843-3987-0.

Buffin-Bélanger, T., P. M. Biron, M. Larocque, S. Demers, T. Olsen, G. Choné, M.-A. Ouellet, C.-A. Cloutier, C. Desjarlais, and J. Eyquem (2015). "Freedom Space for Rivers: An Economically Viable River Management Concept in a Changing Climate." In: *Geomorphology* 251, pp. 137–148. ISSN: 0169555X. DOI: 10.1016/j.geomorph.2015.05.013.

Buffin-Bélanger, T., C.-A. Cloutier, C. Tremblay, G. Chaillou, and M. Larocque (2016). "Dynamics of Groundwater Floodwaves and Groundwater Flood Events in an Alluvial Aquifer." In: *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 41.4, pp. 469–483. ISSN: 0701-1784. DOI: 10.1080/07011784.2015.1102651.

Buffington, J. M. and D. Tonina (2009). "Hyporheic Exchange in Mountain Rivers II: Effects of Channel Morphology on Mechanics, Scales, and Rates of Exchange." In: *Geography Compass* 3.3, pp. 1038–1062. ISSN: 1749-8198. DOI: 10.1111/j.1749-8198.2009.00225.x.

Butscher, C., P. Huggenberger, E. Zechner, and H. H. Einstein (2011). "Relation between Hydrogeological Setting and Swelling Potential of Clay-Sulfate Rocks in Tunneling." In: *Engineering Geology* 122.3, pp. 204–214. ISSN: 0013-7952. DOI: 10.1016/j.enggeo.2011.05.009.

Büttner, O., K. Otte-Witte, F. Krüger, G. Meon, and M. Rode (2006). "Numerical Modelling of Floodplain Hydraulics and Suspended Sediment Transport and Deposition at the Event Scale in the Middle River Elbe, Germany." In: *Acta hydrochimica et hydrobiologica* 34.3, pp. 265–278. ISSN: 0323-4320, 1521-401X. DOI: 10.1002/aheh.200500626.

Capuano, R. M. and R. Z. Jan (1996). "In Situ Hydraulic Conductivity of Clay and Silty-Clay Fluvial-Deltaic Sediments, Texas Gulf Coast." In: *Groundwater* 34.3, pp. 545–551. ISSN: 1745-6584. DOI: 10.1111/j.1745-6584.1996.tb02036.x.

Cardenas, M. B. (2009a). "A Model for Lateral Hyporheic Flow Based on Valley Slope and Channel Sinuosity." In: *Water Resources Research* 45.1. ISSN: 1944-7973. DOI: 10.1029/2008WR007442.

Cardenas, M. B. (2009b). "Stream-Aquifer Interactions and Hyporheic Exchange in Gaining and Losing Sinuous Streams." In: *Water Resources Research* 45.6. ISSN: 1944-7973. DOI: 10.1029/2008WR007651.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005). "Inverse Problem in Hydrogeology." In: *Hydrogeology Journal* 13.1, pp. 206–222. ISSN: 1435-0157. DOI: 10.1007/s10040-004-0404-7.

Carsel, R. F. and R. S. Parrish (1988). "Developing Joint Probability Distributions of Soil Water Retention Characteristics." In: *Water Resources Research* 24.5, pp. 755–769. ISSN: 00431397. DOI: 10.1029/WR024i005p00755.

Castro, N. M. and G. M. Hornberger (1991). "Surface-Subsurface Water Interactions in an Alluviated Mountain Stream Channel." In: *Water Resources Research* 27.7, pp. 1613–1621.

Caviedes-Voullième, D., M. Morales-Hernández, I. López-Marijuan, and P. García-Navarro (2014). "Reconstruction of 2D River Beds by Appropriate Interpolation of 1D Cross-Sectional Information for Flood Simulation." In: *Environmental Modelling & Software* 61, pp. 206–228. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2014.07.016.

Chavez Rodriguez, L. (2021). "Optimal Design of Experiments to Improve the Characterization of Atrazine Degradation Pathways in Soil."

Cheng, J. and M. J. Druzdzel (2013). "Computational Investigation of Low-Discrepancy Sequences in Simulation Algorithms for Bayesian Networks." In: *arXiv:1301.3841 [cs]*. http://arxiv.org/abs/1301.3841. arXiv: 1301.3841 [cs].

Clément, J.-C., L. Aquilina, O. Bour, K. Plaine, T. P. Burt, and G. Pinay (2003). "Hydrological Flowpaths and Nitrate Removal Rates within a Riparian Floodplain along a Fourth-Order Stream in Brittany (France)." In: *Hydrological Processes* 17.6, pp. 1177–1195. ISSN: 1099-1085. DOI: 10.1002/hyp.1192.

Clilverd, H., J. Thompson, C. Heppell, C. Sayer, and J. Axmacher (2013). "River–Floodplain Hydrology of an Embanked Lowland Chalk River and Initial Response to Embankment Removal." In: *Hydrological Sciences Journal* 58.3, pp. 627–650. ISSN: 0262-6667, 2150-3435. DOI: 10.1080/02626667.2013.774089.

Cloke, H. L., J. .-P. Renaud, A. J. Claxton, J. J. McDonnell, M. G. Anderson, J. R. Blake, and P. D. Bates (2003). "The Effect of Model Configuration on Modelled Hillslope–Riparian Interactions." In: *Journal of Hydrology* 279.1, pp. 167–181. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(03)00177-X.

Cloutier, C.-A., T. Buffin-Bélanger, and M. Larocque (2014). "Controls of Groundwater Floodwave Propagation in a Gravelly Floodplain." In: *Journal of Hydrology* 511, pp. 423–431. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2014.02.014.

Conn, A. R., N. I. Gould, and P. L. Toint (2000). *Trust Region Methods*. SIAM.

Constantine, P. G. and P. Diaz (2017). "Global Sensitivity Metrics from Active Subspaces." In: *Reliability Engineering & System Safety* 162, pp. 1–13. ISSN: 0951-8320. DOI: 10.1016/j.ress.2017.01.013.

Constantine, P. G., E. Dow, and Q. Wang (2014). "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces." In: *SIAM Journal on Scientific Computing* 36.4, A1500–A1524. ISSN: 1064-8275, 1095-7197. DOI: 10.1137/130916138.

Cook, P. G. (2015). "Quantifying River Gain and Loss at Regional Scales." In: *Journal of Hydrology* 531, pp. 749–758. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2015.10.052.

Cooley, R. L. (1971). "A Finite Difference Method for Unsteady Flow in Variably Saturated Porous Media: Application to a Single Pumping Well." In: *Water Resources Research* 7.6, pp. 1607–1625. ISSN: 00431397. DOI: 10.1029/WR007i006p01607.

Craig, J. R. (2008). "Analytical Solutions for 2D Topography-Driven Flow in Stratified and Syncline Aquifers." In: *Advances in Water Resources* 31.8, pp. 1066–1073. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2008.04.011.

Cramér, H. (1928). "On the Composition of Elementary Errors." In: *Scandinavian Actuarial Journal* 1928.1, pp. 13–74. ISSN: 0346-1238. DOI: 10.1080/03461238.1928.10416862.

Cranmer, K., J. Brehmer, and G. Louppe (2020). "The Frontier of Simulation-Based Inference." In: *Proceedings of the National Academy of Sciences* 117.48, pp. 30055–30062. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1912789117.

Cranswick, R. H. and P. G. Cook (2015). "Scales and Magnitude of Hyporheic, River–Aquifer and Bank Storage Exchange Fluxes." In: *Hydrological Processes* 29.14, pp. 3084–3097. ISSN: 1099-1085. DOI: 10.1002/hyp.10421.

Cressie, N. (1990). "The Origins of Kriging." In: *Mathematical Geology* 22.3, pp. 239–252. ISSN: 1573-8868. DOI: 10.1007/BF00889887.

D'Affonseca, F. M., M. Finkel, and O. A. Cirpka (2020). "Combining Implicit Geological Modeling, Field Surveys, and Hydrogeological Modeling to Describe Groundwater Flow in a Karst Aquifer." In: *Hydrogeology Journal*. ISSN: 1435-0157. DOI: 10.1007/s10040-020-02220-z.

D'Affonseca, F. M., H. Rügner, M. Finkel, K. Osenbrück, C. E. Duffy, and O. A. Cirpka (2018). *Umweltgerechte Gesteinsgewinnung in Wasserschutzgebieten*. Tech. rep. Universität Tübingen.

Darcy, H. P. G. (1856). *Les fontaines publiques de la ville de Dijon. Exposition et application des principes à suivre et des formules à employer dans les questions de distribution d'eau, etc*. V. Dalamont.

Das, S. and P. N. Suganthan (2011). "Differential Evolution: A Survey of the State-of-the-Art." In: *IEEE Transactions on Evolutionary Computation* 15.1, pp. 4–31. ISSN: 1941-0026. DOI: 10.1109/TEVC.2010.2059031.

Dezső, J., D. Lóczy, A. M. Salem, and G. Nagy (2019). "Floodplain Connectivity." In: *The Drava River: Environmental Problems and Solutions*. Ed. by D. Lóczy. Springer Geography. Cham: Springer International Publishing, pp. 215–230. ISBN: 978-3-319-92816-6. DOI: 10.1007/978-3-319-92816-6_14.

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

Doherty, J., L. Brebber, and P. Whyte (1994). "PEST: Model-independent Parameter Estimation." In: *Watermark Computing, Corinda, Australia* 122, p. 336.

Doherty, J. and R. Hunt (2010). *Approaches to Highly Parameterized Inversion: A Guide to Using PEST for Groundwater-Model Calibration*. Scientific Investigations Report. US Department of the Interior, US Geological Survey Reston, VA, USA.

Dvoracek, M. J. and V. H. Scott (1963). "Ground-Water Flow Characteristics Influenced by Recharge Pit Geometry." In: *Transactions of the ASAE* 6.3, pp. 0262–0265. ISSN: 2151-0059. DOI: 10.13031/2013.40885.

Engwirda, D. (2014). "Locally Optimal Delaunay-Refinement and Optimisation-Based Mesh Generation." http://hdl.handle.net/2123/13148. PhD thesis. University of Sydney.

Erdal, D. and O. A. Cirpka (2019). "Global Sensitivity Analysis and Adaptive Stochastic Sampling of a Subsurface-Flow Model Using Active Subspaces." In: *Hydrology and Earth System Sciences* 23.9, pp. 3787–3805. ISSN: 1027-5606. DOI: 10.5194/hess-23-3787-2019.

Erdal, D. and O. A. Cirpka (2020). *Technical Note: Improved Sampling of Behavioral Subsurface Flow Model Parameters Using Active Subspaces*. Preprint. Groundwater hydrology/Stochastic approaches. DOI: 10.5194/hess-2019-629.

Erdal, D., S. Xiao, W. Nowak, and O. A. Cirpka (2020). "Sampling Behavioral Model Parameters for Ensemble-Based Sensitivity Analysis Using Gaussian Process Emulation and Active Subspaces." In: *Stochastic Environmental Research and Risk Assessment* 34.11, pp. 1813–1830. ISSN: 1436-3259. DOI: 10.1007/s00477-020-01867-0.

Everitt, B. S. and A. Skrondal (2010). *The Cambridge Dictionary of Statistics*. ISBN: 978-0-511-78827-7.

Fabian, M. W., T. A. Endreny, A. Bottacin-Busolin, and L. K. Lautz (2011). "Seasonal Variation in Cascade-Driven Hyporheic Exchange, Northern Honduras." In: *Hydrological Processes* 25.10, pp. 1630–1646. ISSN: 1099-1085. DOI: 10.1002/hyp.7924.

Fan, P. and J. Li (2006). "Diffusive Wave Solutions for Open Channel Flows with Uniform and Concentrated Lateral Inflow." In: *Advances in Water Resources* 29.7, pp. 1000–1019. ISSN: 03091708. DOI: 10.1016/j.advwatres.2005.08.008.

Farthing, M. W., C. E. Kees, and C. T. Miller (2003). "Mixed Finite Element Methods and Higher Order Temporal Approximations for Variably Saturated Groundwater Flow." In: *Advances in Water Resources* 26.4, pp. 373–394. ISSN: 03091708. DOI: 10.1016/S0309-1708(02)00187-2.

Fedorov, V. V. (2013). *Theory of Optimal Experiments*. Elsevier.

Fitts, C. R. (2010). "Modeling Aquifer Systems with Analytic Elements and Subdomains." In: *Water Resources Research* 46.7. ISSN: 1944-7973. DOI: 10.1029/2009WR008331.

Flipo, N., A. Mouhri, B. Labarthe, S. Biancamaria, A. Rivière, and P. Weill (2014). "Continental Hydrosystem Modelling: The Concept of Nested Stream–Aquifer Interfaces." In: *Hydrology and Earth System Sciences* 18.8, pp. 3121–3149. ISSN: 1607-7938. DOI: 10.5194/hess-18-3121-2014.

Folch, A., L. Casadellà, O. Astui, A. Menció, J. Massana, G. Vidal-Gavilan, A. Pérez-Paricio, and J. Mas-Pla (2010). "Verifying Conceptual Flow Models in a River-Connected Alluvial Aquifer for Management Purposes Using Numerical Modeling." In: *Proc. XVIII Internat. Conf. Computational Methods in Water Resources (CMWR 2010), Barcelona, Spain*.

Follner, K., T. Ehlert, and B. Neukirchen (2010). "The Status Report on German Floodplains." In: *Large River Basins: Danube Meets Elbe*. Ed. by J. Bloesch. Schweizerbart, p. 6.

Forrester, A. I. J. and A. J. Keane (2009). "Recent Advances in Surrogate-Based Optimization." In: *Progress in Aerospace Sciences* 45.1, pp. 50–79. ISSN: 0376-0421. DOI: 10.1016/j.paerosci.2008.11.001.

Fotherby, L. M. (2009). "Valley Confinement as a Factor of Braided River Pattern for the Platte River." In: *Geomorphology* 103.4, pp. 562–576. ISSN: 0169-555X. DOI: 10.1016/j.geomorph.2008.08.001.

Franzetti, S. and A. Guadagnini (1996). "Probabilistic Estimation of Well Catchments in Heterogeneous Aquifers." In: *Journal of Hydrology* 174.1-2, pp. 149–171. ISSN: 00221694. DOI: 10.1016/0022-1694(95)02750-5.

Franzmeier, D. P. (1991). "Estimation of Hydraulic Conductivity from Effective Porosity Data for Some Indiana Soils." In: *Soil Science Society of America Journal* 55.6, pp. 1801–1803. ISSN: 1435-0661. DOI: 10.2136/sssaj1991.03615995005500060050x.

Freeze, R. A. and J. A. Cherry (1979). *Groundwater*. Englewood Cliffs, N.J: Prentice-Hall. ISBN: 978-0-13-365312-0.

Fritz, K. M., K. A. Schofield, L. C. Alexander, M. G. McManus, H. E. Golden, C. R. Lane, W. G. Kepner, S. D. LeDuc, J. E. DeMeester, and A. I. Pollard (2018). "Physical and Chemical Connectivity of Streams and Riparian Wetlands to Downstream Waters: A Synthesis." In: *JAWRA Journal of the American Water Resources Association* 54.2, pp. 323–345. ISSN: 1093474X. DOI: 10.1111/1752-1688.12632.

Geiges, A., Y. Rubin, and W. Nowak (2015). "Interactive Design of Experiments: A Priori Global versus Sequential Optimization, Revised under Changing States of Knowledge." In: *Water Resources Research* 51.10, pp. 7915–7936. ISSN: 1944-7973. DOI: 10.1002/2015WR017193.

Gelman, A. and D. B. Rubin (1992). "Inference from Iterative Simulation Using Multiple Sequences." In: *Statistical Science* 7.4, pp. 457–472. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177011136.

Gen, M. and R. Cheng (1999). *Genetic Algorithms and Engineering Optimization*. John Wiley & Sons. ISBN: 978-0-471-31531-5.

Geyer, O. F. and M. P. Gwinner (2011). *Geologie von Baden-Württemberg*. Ed. by M. Geyer, E. Nitsch, and T. Simon. https://www.schweizerbart.de/publications/detail/isbn/9783510652679/Geologie_von_Baden_Wurttemberg. Schweizerbart'sche Verlagsbuchhandlung. ISBN: 978-3-510-65267-9.

Gibbs, J. W. (1898). "Fourier's Series." In: *Nature* 59.1522, pp. 200–200.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. CRC Press. ISBN: 978-1-4822-1497-0.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st. USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 978-0-201-15767-3.

Gomez, J. D., J. L. Wilson, and M. B. Cardenas (2012). "Residence Time Distributions in Sinuosity-Driven Hyporheic Zones and Their Biogeochemical Effects." In: *Water Resources Research* 48.9. ISSN: 1944-7973. DOI: 10.1029/2012WR012180.

Gomez-Velez, J. D., J. L. Wilson, M. B. Cardenas, and J. W. Harvey (2017). "Flow and Residence Times of Dynamic River Bank Storage and Sinuosity-Driven Hyporheic Exchange." In: *Water Resources Research* 53.10, pp. 8572–8595. ISSN: 00431397. DOI: 10.1002/2017WR021362.

Gonçalves, P. J., J.-M. Lueckmann, M. Deistler, M. Nonnenmacher, K. Öcal, G. Bassetto, C. Chintaluri, W. F. Podlaski, S. A. Haddad, T. P. Vogels, D. S. Greenberg, and J. H. Macke (2020). "Training Deep Neural Density Estimators to Identify Mechanistic Models of Neural Dynamics." In: *eLife* 9. Ed. by J. R. Huguenard, T. O'Leary, and M. S. Goldman, e56261. ISSN: 2050-084X. DOI: 10.7554/eLife.56261.

Gooseff, M. N., D. M. McKnight, R. L. Runkel, and B. H. Vaughn (2003). "Determining Long Time-Scale Hyporheic Zone Flow Paths in Antarctic Streams." In: *Hydrological Processes* 17.9, pp. 1691–1710. ISSN: 1099-1085. DOI: 10.1002/hyp.1210.

Grapes, T. R., C. Bradley, and G. E. Petts (2006). "Hydrodynamics of Floodplain Wetlands in a Chalk Catchment: The River Lambourn, UK." In: *Journal of Hydrology*. Groundwater - Surface Water Interactions in Wetlands for Integrated Water Resources Management 320.3, pp. 324–341. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2005.07.028.

Grathwohl, P., H. Rügner, T. Wöhling, K. Osenbrück, M. Schwientek, S. Gayler, U. Wollschläger, B. Selle, M. Pause, J.-O. Delfs, M. Grzeschik, U. Weller, M. Ivanov, O. A. Cirpka, U. Maier, B. Kuch, W. Nowak, V. Wulfmeyer, K. Warrach-Sagi, T. Streck, S. Attinger, L. Bilke, P. Dietrich, J. H. Fleckenstein, T. Kalbacher, O. Kolditz, K. Rink, L. Samaniego, H.-J. Vogel, U. Werban, and G. Teutsch (2013). "Catchments as Reactors: A Comprehensive Approach for Water Fluxes and Solute Turnover." In: *Environmental Earth Sciences* 69.2, pp. 317–333. ISSN: 1866-6299. DOI: 10.1007/s12665-013-2281-7.

Greenberg, D. S., M. Nonnenmacher, and J. H. Macke (2019). "Automatic Posterior Transformation for Likelihood-free Inference." In: *Proceedings of Machine Learning Research*, p. 11.

Guarracino, L. (2007). "Estimation of Saturated Hydraulic Conductivity Ks from the van Genuchten Shape Parameter α." In: *Water Resources Research* 43.11. ISSN: 1944-7973. DOI: 10.1029/2006WR005766.

Gudera, T. (2015). "Hoch aufgelöste Modellierung des Bodenwasserhaushalts und der Grundwasserneubildung mit GWN-BW." In: DOI: 10.5675/HYWA_2015,5_1.

Haftka, R. T., D. Villanueva, and A. Chaudhuri (2016). "Parallel Surrogate-Assisted Global Optimization with Expensive Functions – a Survey." In: *Structural and Multidisciplinary Optimization* 54.1, pp. 3–13. ISSN: 1615-1488. DOI: 10.1007/s00158-016-1432-3.

Hagdorn, H. and E. Nitsch (2009). "Triassic of Southwest Germany – 175th Anniversary of the Foundation of the Triassic System by Friedrich von Alberti." In: p. 72.

Haitjema, H. M. and S. Mitchell-Bruker (2005). "Are Water Tables a Subdued Replica of the Topography?" In: *Ground Water* 0.0, p. 050824075421008. ISSN: 0017-467X, 1745-6584. DOI: 10.1111/j.1745-6584.2005.00090.x.

Halton, J. H. (1960). "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals." In: *Numerische Mathematik* 2.1, pp. 84–90. ISSN: 0029-599X, 0945-3245. DOI: 10.1007/BF01386213.

Han, P.-F., X.-S. Wang, L. Wan, X.-W. Jiang, and F.-S. Hu (2019). "The Exact Groundwater Divide on Water Table between Two Rivers: A Fundamental Model Investigation." In: *Water* 11.4, p. 685. DOI: 10.3390/w11040685.

Harreß, H. M. (1973). *Hydrogeologische Untersuchungen Im Oberen Gäu*. https://rds-tue.ibs-bw.de/link?kid=1073957446. Tübingen.

Harvey, J. W. and K. E. Bencala (1993). "The Effect of Streambed Topography on Surface-Subsurface Water Exchange in Mountain Catchments." In: *Water Resources Research* 29.1, pp. 89–98.

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." In: *Biometrika* 57.1, p. 13. DOI: 10.1093/biomet/57.1.97.

Hauer, F. R., H. Locke, V. J. Dreitz, M. Hebblewhite, W. H. Lowe, C. C. Muhlfeld, C. R. Nelson, M. F. Proctor, and S. B. Rood (2016). "Gravel-Bed River Floodplains Are the Ecological Nexus of Glaciated Mountain Landscapes." In: *Science Advances* 2.6, e1600026. DOI: 10.1126/sciadv.1600026.

Hayashi, M. and D. O. Rosenberry (2002). "Effects of Ground Water Exchange on the Hydrology and Ecology of Surface Water." In: *Groundwater* 40.3, pp. 309–316. ISSN: 1745-6584. DOI: 10.1111/j.1745-6584.2002.tb02659.x.

Helton, A. M., G. C. Poole, R. A. Payn, C. Izurieta, and J. A. Stanford (2012). "Scaling Flow Path Processes to Fluvial Landscapes: An Integrated Field and Model Assessment of Temperature and Dissolved Oxygen Dynamics in a River-Floodplain-Aquifer System." In: *Journal of Geophysical Research: Biogeosciences* 117.G4, n/a–n/a. DOI: 10.1029/2012JG002025.

Helton, A. M., G. C. Poole, R. A. Payn, C. Izurieta, and J. A. Stanford (2014). "Relative Influences of the River Channel, Floodplain Surface, and Alluvial Aquifer on Simulated Hydrologic Residence Time in a Montane River Floodplain." In: *Geomorphology*. Discontinuities in Fluvial Systems 205, pp. 17–26. ISSN: 0169-555X. DOI: 10.1016/j.geomorph.2012.01.004.

Hermans, J., V. Begy, and G. Louppe (2020). "Likelihood-Free MCMC with Amortized Approximate Ratio Estimators." In: *Proceedings of Machine Learning Research*, p. 12.

Hewitt, E. and R. E. Hewitt (1979). "The Gibbs-Wilbraham Phenomenon: An Episode in Fourier Analysis." In: *Archive for History of Exact Sciences* 21.2, pp. 129–160. ISSN: 1432-0657. DOI: 10.1007/BF00330404.

Hill, A. R. (1990). "Ground Water Flow Paths in Relation to Nitrogen Chemistry in the Near-Stream Zone." In: *Hydrobiologia* 206.1, pp. 39–52. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/BF00018968.

Hill, A. R. (1996). "Nitrate Removal in Stream Riparian Zones." In: *Journal of Environmental Quality* 25.4, pp. 743–755. ISSN: 0047-2425, 1537-2537. DOI: 10.2134/jeq1996.00472425002500040014x.

Hill, A. R. (2019). "Groundwater Nitrate Removal in Riparian Buffer Zones: A Review of Research Progress in the Past 20 Years." In: *Biogeochemistry* 143.3, pp. 347–369. ISSN: 0168-2563, 1573-515X. DOI: 10.1007/s10533-019-00566-5.

Hill, M. C. and C. R. Tiedeman (2006). *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*. John Wiley & Sons.

Hiller, T., D. Romanov, G. Kaufmann, J. Epting, and P. Huggenberger (2012). "Karstification beneath the Birs Weir in Basel/Switzerland: A 3D Modeling Approach." In: *Journal of Hydrology* 448–449, pp. 181–194. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2012.04.040.

Hölting, B. and W. G. Coldewey (2013). *Hydrogeologie: Einführung in die allgemeine und angewandte Hydrogeologie*. Heidelberg: Spektrum Akademischer Verlag. ISBN: 978-3-8274-2353-5. DOI: 10.1007/978-3-8274-2354-2.

Holzwarth, W. (1980). *Wasserhaushalt Und Stoffumsatz Kleiner Einzugsgebiete Im Keuper Und Jura Bei Reutlingen-Tübingen*. https://rds-tue.ibs-bw.de/link?kid=1078052956. Tübingen.

Huang, P. and T. F. M. Chui (2018). "Empirical Equations to Predict the Characteristics of Hyporheic Exchange in a Pool-Riffle Sequence." In: *Groundwater* 56.6, pp. 947–958. ISSN: 1745-6584. DOI: 10.1111/gwat.12641.

Huch, M., G. f. U. (i. d. G. Gesellschaft (DGG), and H. Geldmacher (2013). *Ressourcen-Umwelt-Management: Wasser · Boden · Sedimente*. Springer-Verlag. ISBN: 978-3-662-09759-5.

Huggenberger, P., E. Hoehn, R. Beschta, and W. Woessner (1998). "Abiotic Aspects of Channels and Floodplains in Riparian Ecology." In: *Freshwater Biology* 40.3, p. 21.

Hunt, R. J., J. J. Steuer, M. T. C. Mansor, and T. D. Bullen (2001). "Delineating a Recharge Area for a Spring Using Numerical Modeling, Monte Carlo Techniques, and Geochemical Investigation." In: *Groundwater* 39.5, pp. 702–712. ISSN: 1745-6584. DOI: 10.1111/j.1745-6584.2001.tb02360.x.

Huyakorn, P. S., S. D. Thomas, and B. M. Thompson (1984). "Techniques for Making Finite Elements Competitve in Modeling Flow in Variably Saturated Porous Media." In: *Water Resources Research* 20.8, pp. 1099–1115. ISSN: 00431397. DOI: 10.1029/WR020i008p01099.

Hwang, H.-T., Y.-J. Park, E. Sudicky, and P. Forsyth (2014). "A Parallel Computational Framework to Solve Flow and Transport in Integrated Surface–Subsurface Hydrologic Systems." In: *Environmental Modelling & Software* 61, pp. 39–58. ISSN: 13648152. DOI: 10.1016/j.envsoft.2014.06.024.

Jakob, C. (2014). "Going Back to Basics." In: *Nature Climate Change* 4.12, pp. 1042–1045. ISSN: 1758-678X, 1758-6798. DOI: 10.1038/nclimate2445.

Janković, I. and R. Barnes (1999). "Three-Dimensional Flow through Large Numbers of Spheroidal Inhomogeneities." In: *Journal of Hy-*

drology 226.3, pp. 224–233. ISSN: 0022-1694. DOI: 10.1016/S0022-1694(99)00141-9.

Jones, D. R., M. Schonlau, and W. J. Welch (1998). "Efficient Global Optimization of Expensive Black-Box Functions." In: *Journal of Global Optimization* 13.4, pp. 455–492. ISSN: 1573-2916. DOI: 10.1023/A:1008306431147.

Jung, M., T. P. Burt, and P. D. Bates (2004). "Toward a Conceptual Model of Floodplain Water Table Response." In: *Water Resources Research* 40.12. ISSN: 1944-7973. DOI: 10.1029/2003WR002619.

Kalbus, E., F. Reinstorf, and M. Schirmer (2006). "Measuring Methods for Groundwater – Surface Water Interactions: A Review." In: *Hydrology and Earth System Sciences* 10.6, pp. 873–887. ISSN: 1027-5606. DOI: 10.5194/hess-10-873-2006.

Karpack, M. N., R. R. Morrison, and R. A. McManamay (2020). "Quantitative Assessment of Floodplain Functionality Using an Index of Integrity." In: *Ecological Indicators* 111, p. 106051. ISSN: 1470160X. DOI: 10.1016/j.ecolind.2019.106051.

Kasahara, T. and S. M. Wondzell (2003). "Geomorphic Controls on Hyporheic Exchange Flow in Mountain Streams." In: *Water Resources Research* 39.1, SBH 3-1-SBH 3–14. ISSN: 1944-7973. DOI: 10.1029/2002WR001386.

Kehrer, W. (1935). *Ein Beitrag Zur Hydrologie Der Umgebung von Tübingen*. https://rds-tue.ibs-bw.de/link?kid=1092730990. Tübingen.

Keim, B. and H. Pfäfflin (2006). *Grundwasserbilanzmodell für die Brunnen der ASG im Neckartal bei Kiebingen*. Tech. rep. A240-1. Ingenieurgesellschaft Prof. Kobus und Partner GmbH.

Kekeisen, F. (1913). *Das Ammertal – Geologische Studie*. https://rds-tue.ibs-bw.de/link?kid=1165633094. Rottenburg a. N.: Pfeffer & Hofmeister.

Kiràly, L. (1971). "Groundwater Flow in Heterogeneous, Anisotropic Fractured Media: A Simple Two-Dimensional Electric Analog." In: *Journal of Hydrology* 12.3, pp. 255–261. ISSN: 0022-1694. DOI: 10.1016/0022-1694(71)90009-6.

Kirchholtes, H. J. and W. Ufrecht, eds. (2015). *Chlorierte Kohlenwasserstoffe im Grundwasser: Untersuchungsmethoden, Modelle und ein Managementplan für Stuttgart*. Wiesbaden: Springer Vieweg. ISBN: 978-3-658-09248-1.

Kirchner, J. W. (2019). "Quantifying New Water Fractions and Transit Time Distributions Using Ensemble Hydrograph Separation: Theory and Benchmark Tests." In: *Hydrology and Earth System Sciences* 23.1, pp. 303–349. ISSN: 1607-7938. DOI: 10.5194/hess-23-303-2019.

Kitanidis, P. K. (1997). "The Minimum Structure Solution to the Inverse Problem." In: *Water Resources Research* 33.10, pp. 2263–2272. ISSN: 1944-7973. DOI: 10.1029/97WR01619.

Kleinert, K. (1976). "Das Grundwasser im Kiesaquifer des oberen Neckartales zwischen Tübingen und Rottenburg." PhD thesis. Eberhard Karls Universität Tübingen.

Klingler, S., O. A. Cirpka, U. Werban, C. Leven, and P. Dietrich (2020a). "Direct-Push Color Logging Images Spatial Heterogeneity of Organic Carbon in Floodplain Sediments." In: *Journal of Geophysical Research: Biogeosciences* 125.12. ISSN: 2169-8953, 2169-8961. DOI: 10.1029/2020JG005887.

Klingler, S., C. Leven, O. A. Cirpka, and P. Dietrich (2020b). "Anomaly Effect-Driven Optimization of Direct-Current Geoelectric Mapping Surveys in Large Areas." In: *Journal of Applied Geophysics* 176, p. 104002. ISSN: 09269851. DOI: 10.1016/j.jappgeo.2020.104002.

Klingler, S., S. Martin, O. A. Cirpka, P. Dietrich, and C. Leven (2021). "Kombination geophysikalischer und hydrogeologischer Methoden zur gezielten Erkundung feinkörniger Talfüllungen." In: *Grundwasser* 26.4, pp. 379–394. ISSN: 1432-1165. DOI: 10.1007/s00767-021-00494-y.

Kocis, L. and W. J. Whiten (1997). "Computational Investigations of Low-Discrepancy Sequences." In: *ACM Transactions on Mathematical Software* 23.2, pp. 266–294. ISSN: 0098-3500. DOI: 10.1145/264029.264064.

Koenzen, U. and D. Günther-Diringer (2021). *Auenzustandsbericht 2021*. First. https://doi.org/10.19217/brs211. DE: Bundesamt für Naturschutz.

Köpf, E. (1926). *Die geologischen Verhältnisse am Spitzberg bei Tübingen*. https://rds-tue.ibs-bw.de/link?kid=1168706254. Tübingen.

Köppen, W. (1884). "Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet." In: *Meteorologische Zeitschrift* 1.21, pp. 5–226.

Kortunov, E. (2018). "Reactive Transport and Long-Term Redox Evolution at the Catchment Scale." PhD thesis. University of Tübingen.

Kovesi, P. (2015). "Good Colour Maps: How to Design Them." In: *arXiv:1509.03700 [cs]*. http://arxiv.org/abs/1509.03700. arXiv: 1509.03700 [cs].

Krige, D. G. (1951). "A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand: By DG Krige." PhD thesis. University of the Witwatersrand.

Kuang, X., J. J. Jiao, J. Shan, and Z. Yang (2021). "A Modification to the van Genuchten Model for Improved Prediction of Relative Hydraulic Conductivity of Unsaturated Soils." In: *European Journal of Soil Science* 72.3, pp. 1354–1372. ISSN: 1365-2389. DOI: 10.1111/ejss.13034.

Lang, R. (1909). *Der mittlere Keuper im südlichen Württemberg*. https://rds-tue.ibs-bw.de/link?kid=133378273X. Stuttgart: Grüninger.

Langevin, C. D., J. D. Hughes, E. Banta, A. Provost, R. Niswonger, and S. Panday (2017). *MODFLOW 6, the U.S. Geological Survey Modular Hydrologic Model*. U.S. Geological Survey. DOI: 10.5066/F76Q1VQV.

Langevin, C. D., A. M. Provost, S. Panday, and J. D. Hughes (2022). *Documentation for the MODFLOW 6 Groundwater Transport Model*. Tech. rep. 6-A61. U.S. Geological Survey. DOI: 10.3133/tm6A61.

Langhoff, J. H., K. R. Rasmussen, and S. Christensen (2006). "Quantification and Regionalization of Groundwater–Surface Water Interaction along an Alluvial Stream." In: *Journal of Hydrology*. Groundwater - Surface Water Interactions in Wetlands for Integrated Water Resources Management 320.3, pp. 342–358. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2005.07.040.

Larkin, R. and J. Sharp Jr. (1992). "On the Relationship between River-Basin Geomorphology, Aquifer Hydraulics, and Ground-Water Flow Direction in Alluvial Aquifers." In: *Geological Society of America Bulletin*. DOI: 10.1130/0016-7606(1992)104<1608:OTRBRB>2.3.CO;2.

Leopold, L. B., M. G. Wolman, J. P. Miller, and E. Wohl (2020). *Fluvial Processes in Geomorphology*. Courier Dover Publications. ISBN: 978-0-486-84552-4.

Leroueil, S., P. Lerat, F. D. W. Hights, and J. J. M. Powells (n.d.). "Hydraulic Conductivity of a Recent Estuarine Silty Clay at Bothkennar." In: *HYDRAULIC CONDUCTIVITY* (), p. 14.

Lessoff, S. C., U. Schneidewind, C. Leven, P. Blum, P. Dietrich, and G. Dagan (2010). "Spatial Characterization of the Hydraulic Conductivity Using Direct-Push Injection Logging." In: *Water Resources Research* 46.12. ISSN: 1944-7973. DOI: 10.1029/2009WR008949.

Leube, P. C., A. Geiges, and W. Nowak (2012). "Bayesian Assessment of the Expected Data Impact on Prediction Confidence in Optimal Sampling Design." In: *Water Resources Research* 48.2. ISSN: 1944-7973. DOI: 10.1029/2010WR010137.

Levenberg, K. (1944). "A Method for the Solution of Certain Non-Linear Problems in Least Squares." In: *Quarterly of Applied Mathematics* 2.2, pp. 164–168. ISSN: 0033-569X, 1552-4485. DOI: 10.1090/qam/10666.

Lewandowski, J., S. Arnon, E. Banks, O. Batelaan, A. Betterle, T. Broecker, C. Coll, J. D. Drummond, J. Gaona Garcia, J. Galloway, J. Gomez-Velez, R. C. Grabowski, S. P. Herzog, R. Hinkelmann, A. Höhne, J. Hollender, M. A. Horn, A. Jaeger, S. Krause, A. Löchner Prats, C. Magliozzi, K. Meinikmann, B. B. Mojarrad, B. M. Mueller, I. Peralta-Maraver, A. L. Popp, M. Posselt, A. Putschew, M. Radke, M. Raza, J. Riml, A. Robertson, C. Rutere, J. L. Schaper, M. Schirmer, H. Schulz, M. Shanafield, T. Singh, A. S. Ward, P. Wolke, A. Wörman, and L. Wu (2019). "Is the Hyporheic Zone Relevant beyond the Scientific Community?" In: *Water* 11.11, p. 2230. DOI: 10.3390/w11112230.

LGRB (2005). *Geologische Karte von Baden-Württemberg 1:25.000 – Erläuterungen Zu Blatt 7420 Tübingen*. Ed. by M. Schmidt, W. Ohmert, A. Schreiner, and E. Villinger. Fifth. Freiburg im Breisgau: Regierungspräsidium Freiburg (Landesamt für Geologie, Rohstoffe und Bergbau).

LGRB (2010). *Geologische Untersuchungen von Baugrundhebungen Im Bereich Des Erdwärmesondenfeldes Beim Rathaus in Der Historischen Altstadt von Staufen i. Br.* Tech. rep. https://produkte.lgrb-bw.de/schriftensuche/sonstige_produkte/1203/. Landesamt für Geologie, Rohstoffe und Bergbau.

LGRB (2012). *Zweiter Sachstandsbericht zu den seit dem 01.03.2010 erfolgten Untersuchungen im Bereich des Erdwärmesondenfeldes beim Rathaus in der historischen Altstadt von Staufen i. Br.* Tech. rep. https://www.lgrb-bw.de/geothermie/staufen/schadensfall_staufen_bericht_2012. Landesamt für Geologie, Rohstoffe und Bergbau.

Li, P., F. Stagnitti, and U. Das (1996). "A New Analytical Solution for Laplacian Porous-Media Flow with Arbitrary Boundary Shapes and Conditions." In: *Mathematical and Computer Modelling* 24.10, pp. 3–19. ISSN: 0895-7177. DOI: 10.1016/S0895-7177(96)00160-4.

Li, S. (2011). "Concise Formulas for the Area and Volume of a Hyperspherical Cap." In: *Asian Journal of Mathematics and Statistics* 4.1, pp. 66–70.

Lin, C. D., B. Tang, et al. (2015). "Latin Hypercubes and Space-Filling Designs." In: *Handbook of design and analysis of experiments*, pp. 593–625.

Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer Science & Business Media. ISBN: 978-0-387-76369-9.

LTZ (2021). *Agrarmeteorologie Baden-Württemberg*. https://www.wetter-bw.de/Agrarmeteorologie-BW/Wetterdaten/Stationen-nach-Region/Tuebingen/BWAM146.

Lu, C., W. Qin, G. Zhao, Y. Zhang, and W. Wang (2017). "Better-Fitted Probability of Hydraulic Conductivity for a Silty Clay Site and Its Effects on Solute Transport." In: *Water* 9.7, p. 466. DOI: 10.3390/w9070466.

LUBW (2021). *Daten aus dem Umweltinformationssystem (UIS) der LUBW Landesanstalt für Umwelt Baden-Württemberg*. https://udo.lubw.baden-wuerttemberg.de/public/q/6dAtdEkpPEGhOGYi1c1PTe.

Lueckmann, J.-M., J. Boelts, D. S. Greenberg, P. J. Gonçalves, and J. H. Macke (2021). "Benchmarking Simulation-Based Inference." In: 130, p. 14.

Lueckmann, J.-M., P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke (2017). "Flexible Statistical Inference for Mechanistic Models of Neural Dynamics." In: p. 11.

MacDonald, A. M., D. J. Lapworth, A. G. Hughes, C. A. Auton, L. Maurice, A. Finlayson, and D. C. Gooddy (2014). "Groundwater, Flooding and Hydrological Functioning in the Findhorn Floodplain, Scotland." In: *Hydrology Research* 45.6, pp. 755–773. ISSN: 0029-1277, 2224-7955. DOI: 10.2166/nh.2014.185.

Macpherson, G. L. and M. Sophocleous (2004). "Fast Ground-Water Mixing and Basal Recharge in an Unconfined, Alluvial Aquifer, Konza LTER Site, Northeastern Kansas." In: *Journal of Hydrology* 286.1, pp. 271–299. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2003.09.016.

Magliozzi, C., R. Grabowski, A. I. Packman, and S. Krause (2017). "Scaling down Hyporheic Exchange Flows: From Catchments to Reaches." In: *Hydrology and Earth System Sciences Discussions*, pp. 1–53. ISSN: 1812-2116. DOI: 10.5194/hess-2016-683.

Magliozzi, C., R. C. Grabowski, A. I. Packman, and S. Krause (2018). "Toward a Conceptual Framework of Hyporheic Exchange across Spatial Scales." In: *Hydrology and Earth System Sciences* 22.12, pp. 6163–6185. ISSN: 1607-7938. DOI: 10.5194/hess-22-6163-2018.

Maier, U., M. Flegr, H. Rügner, and P. Grathwohl (2013). "Long-Term Solute Transport and Geochemical Equilibria in Seepage Water and Groundwater in a Catchment Cross Section." In: *Environmental Earth Sciences* 69.2, pp. 429–441. ISSN: 1866-6299. DOI: 10.1007/s12665-013-2393-0.

Mallard, J., B. McGlynn, and T. Covino (2014). "Lateral Inflows, Stream-Groundwater Exchange, and Network Geometry Influence Stream Water Composition." In: *Water Resources Research* 50.6, pp. 4603–4623. ISSN: 1944-7973. DOI: 10.1002/2013WR014944.

Manning, R., J. P. Griffith, T. Pigot, and L. F. Vernon-Harcourt (1890). *On the Flow of Water in Open Channels and Pipes*.

Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." In: *Journal of the Society for Industrial and Applied Mathematics* 11.2, pp. 431–441. ISSN: 0368-4245. DOI: 10.1137/0111030.

Marsaglia, G. (1972). "Choosing a Point from the Surface of a Sphere." In: *The Annals of Mathematical Statistics* 43.2, pp. 645–646. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177692644.

Martin, S. (2021). *Personal Communication*.

Martin, S., S. Klingler, P. Dietrich, C. Leven, and O. A. Cirpka (2020). "Structural Controls on the Hydrogeological Functioning of a Floodplain." In: *Hydrogeology Journal* 28.8, pp. 2675–2696. ISSN: 1431-2174, 1435-0157. DOI: 10.1007/s10040-020-02225-8.

Matérn, B. (1960). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Vol. 49. Meddelanden från statens Skogsforskningsinstitut.

Matheron, G. (1963). "Principles of Geostatistics." In: *Economic geology* 58.8, pp. 1246–1266.

McDonald, M. G. and A. W. Harbaugh (1988). *A Modular Three-Dimensional Finite-Difference Ground-water Flow Model*. US Geological Survey.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." In: *Technometrics* 21.2, p. 239. ISSN: 00401706. DOI: 10.2307/1268522.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). "Equation of State Calculations by Fast Computing Machines." In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114.

Meyer, J. L. and R. T. Edwards (1990). "Ecosystem Metabolism and Turnover of Organic Carbon along a Blackwater River Continuum." In: *Ecology* 71.2, pp. 668–677. ISSN: 00129658. DOI: 10.2307/1940321.

Minasny, B., J. W. Hopmans, T. Harter, S. O. Eching, A. Tuli, and M. A. Denton (2004). "Neural Networks Prediction of Soil Hydraulic Functions for Alluvial Soils Using Multistep Outflow Data." In: *Soil Science Society of America Journal* 68.2, pp. 417–429. ISSN: 03615995. DOI: 10.2136/sssaj2004.4170.

Mohammadi, F., R. Kopmann, A. Guthke, S. Oladyshkin, and W. Nowak (2018). "Bayesian Selection of Hydro-Morphodynamic Models under Computational Time Constraints." In: *Advances in Water Resources* 117, pp. 53–64. ISSN: 03091708. DOI: 10.1016/j.advwatres.2018.05.007.

Mualem, Y. (1976). "A New Model for Predicting the Hydraulic Conductivity of Unsaturated Porous Media." In: *Water Resources Research* 12.3, pp. 513–522. ISSN: 00431397. DOI: 10.1029/WR012i003p00513.

Muller, M. E. (1959). "A Note on a Method for Generating Points Uniformly on N-Dimensional Spheres." In: *Communications of the ACM* 2.4, pp. 19–20. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/377939.377946.

Müller, M. E., C. Zwiener, and B. I. Escher (2021). "Storm Event–Driven Occurrence and Transport of Dissolved and Sorbed Organic Micropollutants and Associated Effects in the Ammer River, Southwestern Germany." In: *Environmental Toxicology and Chemistry* 40.1, pp. 88–99. ISSN: 1552-8618. DOI: 10.1002/etc.4910.

Nagarajarao, Y. and S. Mallick (1980). "Comparison of Experimentally Determined and Calculated Hydraulic Conductivities for Two Alluvial Sandy Loam Profiles." In: *Zeitschrift für Pflanzenernährung und Bodenkunde* 143.6, pp. 679–683. ISSN: 1522-2624. DOI: 10.1002/jpln.19801430609.

Nagel, D. E., J. M. Buffington, S. L. Parkes, S. Wenger, and J. R. Goode (2014). *A Landscape Scale Valley Confinement Algorithm: Delineating Unconfined Valley Bottoms for Geomorphic, Aquatic, and Riparian Applications*. General Technical Report RMRS-GTR-321. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, RMRS-GTR–321. DOI: 10.2737/RMRS-GTR-321.

Nanson, G. and J. Croke (1992). "A Genetic Classification of Floodplains." In: *Geomorphology* 4.6, pp. 459–486. ISSN: 0169555X. DOI: 10.1016/0169-555X(92)90039-Q.

Ó Dochartaigh, B. É., N. A. L. Archer, L. Peskett, A. M. MacDonald, A. R. Black, C. A. Auton, J. E. Merritt, D. C. Gooddy, and M. Bonell (2019). "Geological Structure as a Control on Floodplain Groundwater Dynamics." In: *Hydrogeology Journal* 27.2, pp. 703–716. ISSN: 1435-0157. DOI: 10.1007/s10040-018-1885-0.

Ohara, N., W. S. Holbrook, K. Yamatani, B. A. Flinchum, and J. T. S. Clair (2018). "Spatial Delineation of Riparian Groundwater within Alluvium Deposit of Mountainous Region Using Laplace Equation." In: *Hydrological Processes* 32.1, pp. 30–38. ISSN: 1099-1085. DOI: 10.1002/hyp.11395.

Osenbrück, K., E. Blendinger, C. Leven, H. Rügner, M. Finkel, N. Jakus, H. Schulz, and P. Grathwohl (2022). "Nitrate Reduction Potential of a Fractured Middle Triassic Carbonate Aquifer in Southwest Germany." In: *Hydrogeology Journal* 30.1, pp. 163–180. ISSN: 1435-0157. DOI: 10.1007/s10040-021-02418-9.

Ostendorf, D. W., E. S. Hinlein, and A. I. Judge (2012). "Floodplain Hydraulics in a Glaciated Bedrock River Valley." In: *Hydrology Research* 43.6, pp. 870–889. ISSN: 0029-1277, 2224-7955. DOI: 10.2166/nh.2012.066.

Owen, R. and T. Dahlin (2005). "Alluvial Aquifers at Geological Boundaries: Geophysical Investigations and Groundwater Resources." In: *Groundwater and Human Development, AA Balkema Publishers, Rotterdam*, pp. 233–246.

Papamakarios, G. and I. Murray (2016). "Fast ϵ-Free Inference of Simulation Models with Bayesian Conditional Density Estimation." In: p. 9.

Papamakarios, G., T. Pavlakou, and I. Murray (2017). "Masked Autoregressive Flow for Density Estimation." In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. https://proceedings.neurips.cc/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf. Curran Associates, Inc.

Papamakarios, G., D. Sterratt, and I. Murray (2019). "Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows." In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. https://proceedings.mlr.press/v89/papamakarios19a.html. PMLR, pp. 837–848.

Patterson, H. D. and R. Thompson (1971). "Recovery of Inter-Block Information When Block Sizes Are Unequal." In: *Biometrika* 58.3, pp. 545–554. ISSN: 0006-3444. DOI: 10.1093/biomet/58.3.545.

Penman, H. (1956). "Estimating Evaporation." In: *Eos, Transactions American Geophysical Union* 37.1, pp. 43–50. ISSN: 2324-9250. DOI: 10.1029/TR037i001p00043.

Petrova, E., E. Kortunov, K. U. Mayer, P. Grathwohl, and M. Finkel (2022). "Travel Time-Based Modelling of Nitrate Reduction in a Fractured Limestone Aquifer by Pyrite and Iron Carbonates under Pore Size Limitation." In: *Journal of Contaminant Hydrology* 248, p. 103983. ISSN: 0169-7722. DOI: 10.1016/j.jconhyd.2022.103983.

Peyrard, D., S. Sauvage, P. Vervier, J. M. Sanchez-Perez, and M. Quintard (2008). "A Coupled Vertically Integrated Model to Describe Lateral Exchanges between Surface and Subsurface in Large Alluvial Floodplains with a Fully Penetrating River." In: *Hydrological Processes* 22.21, pp. 4257–4273. ISSN: 1099-1085. DOI: 10.1002/hyp.7035.

Piggott, A. R. and D. Elsworth (1989). "Physical and Numerical Studies of a Fracture System Model." In: *Water Resources Research* 25.3, pp. 457–462. ISSN: 00431397. DOI: 10.1029/WR025i003p00457.

Pinay, G., C. Ruffinoni, S. Wondzell, and F. Gazelle (1998). "Change in Groundwater Nitrate Concentration in a Large River Floodplain: Denitrification, Uptake, or Mixing?" In: *Journal of the North American Benthological Society* 17.2, pp. 179–189. ISSN: 0887-3593, 1937-237X. DOI: 10.2307/1467961.

Pöschke, F., G. Nützmann, P. Engesgaard, and J. Lewandowski (2018). "How Does the Groundwater Influence the Water Balance of a Lowland Lake? A Field Study from Lake Stechlin, North-Eastern Germany." In: *Limnologica*. Special Issue on Aquatic Interfaces and Linkages: An Emerging Topic of Interdisciplinary Research 68, pp. 17–25. ISSN: 0075-9511. DOI: 10.1016/j.limno.2017.11.005.

Powell, M. J. D. (1970a). "A Hybrid Method for Nonlinear Equations." In: *Numerical Methods for Nonlinear Algebraic Equations*. https://ci.nii.ac.jp/naid/10006528967/.

Powell, M. J. D. (1970b). "A New Algorithm for Unconstrained Optimization." In: *Nonlinear Programming*. Ed. by J. B. Rosen, O. L. Mangasarian, and K. Ritter. Academic Press, pp. 31–65. ISBN: 978-0-12-597050-1. DOI: 10.1016/B978-0-12-597050-1.50006-3.

Powers, W. L. (1966). "Solution of Some Theoretical Soil Drainage Problems by Generalized Orthonormal Functions." In: p. 254.

Powers, W. L., D. Kirkham, and G. Snowden (1967). "Orthonormal Function Tables and the Seepage of Steady Rain through Soil Bedding." In: *Journal of Geophysical Research* 72.24, pp. 6225–6237. ISSN: 01480227. DOI: 10.1029/JZ072i024p06225.

Prickett, T. A. (1975). "Modeling Techniques for Groundwater Evaluation." In: *Advances in Hydroscience*. Vol. 10. Elsevier, pp. 1–143. ISBN: 978-0-12-021810-3. DOI: 10.1016/B978-0-12-021810-3.50006-0.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM.

Qiu, H., P. Blaen, S. Comer-Warner, D. M. Hannah, S. Krause, and M. S. Phanikumar (2019). "Evaluating a Coupled Phenology-Surface Energy Balance Model to Understand Stream-Subsurface Temperature Dynamics in a Mixed-Use Farmland Catchment." In: *Water Resources Research* 55.2, pp. 1675–1697. ISSN: 1944-7973. DOI: 10.1029/2018WR023644.

Quenstedt, F. A. (1864). *Geologische Ausflüge in Schwaben*. Laupp.

Rashid, R. S. M. M. and M. H. Chaudhry (1995). "Flood Routing in Channels with Flood Plains." In: *Journal of Hydrology* 171.1-2, pp. 75–91. ISSN: 00221694. DOI: 10.1016/0022-1694(95)02693-J.

Rasmussen, C. E. (2003). "Gaussian Processes in Machine Learning." In: *Advanced Lectures on Machine Learning*. Springer-Verlag Berlin Heidelberg, pp. 63–71.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press. ISBN: 978-0-262-18253-9.

Rawls, W. J. and D. L. Brakensiek (1985). "Prediction of Soil Water Properties for Hydrologic Modeling." In: *Watershed Management in the Eighties*. ASCE, pp. 293–299.

Rawls, W. J. and D. L. Brakensiek (1989). "Estimation of Soil Water Retention and Hydraulic Properties." In: *Unsaturated Flow in Hydrologic Modeling*. Ed. by H. J. Morel-Seytoux. Dordrecht: Springer Netherlands, pp. 275–300. ISBN: 978-94-010-7559-6. DOI: 10.1007/978-94-009-2352-2_10.

Rawls, W. J., D. L. Brakensiek, and K. Saxtonn (1982). "Estimation of Soil Water Properties." In: *Transactions of the ASAE* 25.5, pp. 1316–1320.

Ray, S. (2020). "A Reconstruction-Based Chebyshev-collocation Method for the Poisson Equation: An Accurate Treatment of the Gibbs-Wilbraham Phenomenon on Irregular Interfaces." In: *Journal of Computational Physics*, p. 20.

Read, W. W. (2007). "An Analytic Series Method for Laplacian Problems with Mixed Boundary Conditions." In: *Journal of Computational and Applied Mathematics* 209.1, pp. 22–32. ISSN: 0377-0427. DOI: 10.1016/j.cam.2006.10.088.

Rees, I., I. Masters, A. G. Malan, and R. W. Lewis (2004). "An Edge-Based Finite Volume Scheme for Saturated–Unsaturated Groundwater Flow." In: *Computer Methods in Applied Mechanics and Engineering* 193.42, pp. 4741–4759. ISSN: 0045-7825. DOI: 10.1016/j.cma.2004.04.003.

Regis, R. G. and C. A. Shoemaker (2007). "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions." In: *INFORMS Journal on Computing* 19.4, pp. 497–509. ISSN: 1091-9856, 1526-5528. DOI: 10.1287/ijoc.1060.0182.

Regis, R. G. and C. A. Shoemaker (2009). "Parallel Stochastic Global Optimization Using Radial Basis Functions." In: *INFORMS Journal on Computing* 21.3, pp. 411–426. ISSN: 1091-9856, 1526-5528. DOI: 10.1287/ijoc.1090.0325.

Ren, J., W. Zhang, J. Yang, and Y. Zhou (2019). "Using Water Temperature Series and Hydraulic Heads to Quantify Hyporheic Exchange in the Riparian Zone." In: *Hydrogeology Journal* 27.4, pp. 1419–1437. ISSN: 1435-0157. DOI: 10.1007/s10040-019-01934-z.

Revelli, R., F. Boano, C. Camporeale, and L. Ridolfi (2008). "Intra-Meander Hyporheic Flow in Alluvial Rivers." In: *Water Resources Research* 44.12. ISSN: 1944-7973. DOI: 10.1029/2008WR007081.

Reynolds, W. D., B. T. Bowman, R. R. Brunke, C. F. Drury, and C. S. Tan (2000). "Comparison of Tension Infiltrometer, Pressure Infiltrometer, and Soil Core Estimates of Saturated Hydraulic Conductivity." In: *Soil Science Society of America Journal* 64.2, pp. 478–484. ISSN: 1435-0661. DOI: 10.2136/sssaj2000.642478x.

Richards, L. A. (1931). "Capillary Conduction of Liquids Through Porous Mediums." In: *Physics* 1.5, pp. 318–333. ISSN: 0148-6349, 2163-5102. DOI: 10.1063/1.1745010.

Richardson, L. F. (1922). *Weather Prediction by Numerical Process*. http://archive.org/details/weatherpredictio00richrich. Cambridge, The University press.

Rodhe, A. (2012). "Physical Models for Classroom Teaching in Hydrology." In: *Hydrology and Earth System Sciences* 16.9, pp. 3075–3082. ISSN: 1607-7938. DOI: 10.5194/hess-16-3075-2012.

Rodriguez-Pretelin, A. and W. Nowak (2018). "Integrating Transient Behavior as a New Dimension to WHPA Delineation." In: *Advances in Water Resources* 119, pp. 178–187. ISSN: 03091708. DOI: 10.1016/j.advwatres.2018.07.005.

Runge, C. (1901). "Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten." In: *Zeitschrift für Mathematik und Physik* 46.224-243, p. 20.

Samani, N. and M. M. Sedghi (2015). "Semi-Analytical Solutions of Groundwater Flow in Multi-Zone (Patchy) Wedge-Shaped Aquifers." In: *Advances in Water Resources* 77, pp. 1–16. ISSN: 03091708. DOI: 10.1016/j.advwatres.2015.01.003.

Sanz, D., J. J. Gómez-Alday, S. Castaño, A. Moratalla, J. De las Heras, and P. E. Martínez-Alfaro (2009). "Hydrostratigraphic Framework and Hydrogeological Behaviour of the Mancha Oriental System (SE Spain)." In: *Hydrogeology Journal* 17.6, pp. 1375–1391. ISSN: 1431-2174, 1435-0157. DOI: 10.1007/s10040-009-0446-y.

Schaap, M. G. and F. J. Leij (2000). "Improved Prediction of Unsaturated Hydraulic Conductivity with the Mualem-van Genuchten Model." In: *Soil Science Society of America Journal* 64.3, pp. 843–851. ISSN: 1435-0661. DOI: 10.2136/sssaj2000.643843x.

Schilling, K. E., P. J. Jacobson, and J. A. Vogelgesang (2015). "Agricultural Conversion of Floodplain Ecosystems: Implications for Groundwater Quality." In: *Journal of Environmental Management* 153, pp. 74–83. ISSN: 0301-4797. DOI: 10.1016/j.jenvman.2015.02.004.

Schlosser, T., M. Schmidt, M. Schneider, and P. Vermeer (2007). "Potenzial der Tunnelbaustrecke des Bahnprojektes Stuttgart 21 zur Wärme- und Kältenutzung." In: *Studie des Zentrums für Energieforschung Stuttgart. Stuttgart*.

Schollenberger, U. (1998). "Beschaffenheit und Dynamik des Kiesgrundwasssers im Neckartal bei Tübingen." PhD thesis. Tübingen: Eberhard Karls Universität Tübingen.

Schwede, R. L., A. Ngo, P. Bastian, O. Ippisch, W. Li, and O. A. Cirpka (2012). "Efficient Parallelization of Geostatistical Inversion Using the Quasi-Linear Approach." In: *Computers & Geosciences* 44, pp. 78–85. ISSN: 00983004. DOI: 10.1016/j.cageo.2012.03.014.

Schweizer, D., H. Prommer, P. Blum, and C. Butscher (2019). "Analyzing the Heave of an Entire City: Modeling of Swelling Processes in Clay-Sulfate Rocks." In: *Engineering Geology* 261, p. 105259. ISSN: 00137952. DOI: 10.1016/j.enggeo.2019.105259.

Schweizer, D., H. Prommer, P. Blum, A. J. Siade, and C. Butscher (2018). "Reactive Transport Modeling of Swelling Processes in Clay-sulfate Rocks." In: *Water Resources Research* 54.9, pp. 6543–6565. ISSN: 00431397. DOI: 10.1029/2018WR023579.

Seitz, T. (2010). *Projektdaten aus der Erstellung der Hochwassergefahrenkarten des Landes Baden-Württemberg*. Tech. rep. Datenausgabe 100928. ISTW Planungsgesellschaft mbH.

Selle, B., K. Rink, and O. Kolditz (2013). "Recharge and Discharge Controls on Groundwater Travel Times and Flow Paths to Production Wells for the Ammer Catchment in Southwestern Germany." In: *Environmental Earth Sciences* 69.2, pp. 443–452. ISSN: 1866-6280, 1866-6299. DOI: 10.1007/s12665-013-2333-z.

Smock, L. A., J. E. Gladden, J. L. Riekenberg, L. C. Smith, and C. R. Black (1992). "Lotic Macroinvertebrate Production in Three Dimensions: Channel Surface, Hyporheic, and Floodplain Environments." In: *Ecology* 73.3, pp. 876–886. ISSN: 00129658. DOI: 10.2307/1940165.

Sobol′, I. M. (2001). "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates." In: *Mathematics and Computers in Simulation*. The Second IMACS Seminar on Monte Carlo Methods 55.1, pp. 271–280. ISSN: 0378-4754. DOI: 10.1016/S0378-4754(00)00270-6.

Sobol′, I. M. (1993). "Sensitivity Analysis for Non-Linear Mathematical Models." In: *Mathematical modelling and computational experiment* 1, pp. 407–414.

Solomatine, D. P., Y. B. Dibike, and N. Kukuric (1999). "Automatic Calibration of Groundwater Models Using Global Optimization Techniques." In: *Hydrological Sciences Journal* 44.6, pp. 879–894. ISSN: 0262-6667, 2150-3435. DOI: 10.1080/02626669909492287.

Sophocleous, M. A. (1991). "Stream-Floodwave Propagation through the Great Bend Alluvial Aquifer, Kansas: Field Measurements and Numerical Simulations." In: *Journal of Hydrology* 124.3-4, pp. 207–228. ISSN: 00221694. DOI: 10.1016/0022-1694(91)90015-A.

Stanford, J. A. and J. V. Ward (1993). "An Ecosystem Perspective of Alluvial Rivers: Connectivity and the Hyporheic Corridor." In: *Journal of the North American Benthological Society* 12.1, pp. 48–60. ISSN: 0887-3593, 1937-237X. DOI: 10.2307/1467685.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media.

Strack, O. D. L. (2003). "Theory and Applications of the Analytic Element Method." In: *Reviews of Geophysics* 41.2. ISSN: 1944-9208. DOI: 10.1029/2002RG000111.

Strack, O. D. L. (2018). "Limitless Analytic Elements." In: *Water Resources Research* 54.2, pp. 1174–1190. ISSN: 1944-7973. DOI: 10.1002/2017WR022117.

Strack, O. D. L. (2017). *Analytical Groundwater Mechanics*. https://doi.org/10.1017/9781316563144. ISBN: 978-1-316-56314-4.

Strack, O. D. L. and P. R. Nevison (2015). "Analytic Elements of Smooth Shapes." In: *Journal of Hydrology* 529, pp. 231–239. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2015.07.001.

Strack, O. D. (1989). *Groundwater Mechanics*. prentice hall.

Suk, H. and E. Park (2019). "Numerical Solution of the Kirchhoff-transformed Richards Equation for Simulating Variably Saturated Flow in Heterogeneous Layered Porous Media." In: *Journal of Hydrology* 579, p. 124213. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2019.124213.

Sun, X., L. Bernard-Jannin, S. Sauvage, C. Garneau, J. G. Arnold, R. Srinivasan, and J. M. Sánchez-Pérez (2017). "Assessment of the Denitrification Process in Alluvial Wetlands at Floodplain Scale Using the SWAT Model." In: *Ecological Engineering*. Wetlands and Buffer Zones in Watershed Management 103, pp. 344–358. ISSN: 0925-8574. DOI: 10.1016/j.ecoleng.2016.06.098.

Suribhatla, R., M. Bakker, K. Bandilla, and I. Janković (2004). "Steady Two-Dimensional Groundwater Flow through Many Elliptical Inhomogeneities." In: *Water Resources Research* 40.4. ISSN: 1944-7973. DOI: 10.1029/2003WR002718.

Sutfin, N. A., E. E. Wohl, and K. A. Dwire (2016). "Banking Carbon: A Review of Organic Carbon Storage and Physical Factors Influencing Retention in Floodplains and Riparian Ecosystems." In: *Earth Surface Processes and Landforms* 41.1, pp. 38–60. ISSN: 1096-9837. DOI: 10.1002/esp.3857.

Szabó, B., M. Weynants, and T. K. D. Weber (2021). "Updated European Hydraulic Pedotransfer Functions with Communicated Uncertainties in the Predicted Variables (Euptfv2)." In: *Geoscientific Model Development* 14.1, pp. 151–175. ISSN: 1991-959X. DOI: 10.5194/gmd-14-151-2021.

Tang, B. (1993). "Orthogonal Array-Based Latin Hypercubes." In: *Journal of the American Statistical Association* 88.424, pp. 1392–1397. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1993.10476423.

Tarboton, D. G., R. L. Bras, and I. Rodriguez-Iturbe (1991). "On the Extraction of Channel Networks from Digital Elevation Data." In: *Hydrological Processes* 5.1, pp. 81–100. ISSN: 1099-1085. DOI: 10.1002/hyp.3360050107.

Tecplot Inc. (2019). *User's Manual Tecplot 360 EX 2019 Release 1*. Bellevue, WA.

Tejero-Cantero, A., J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke (2020). "Sbi: A Toolkit for Simulation-Based Inference." In: *Journal of Open Source Software* 5.52, p. 2505. ISSN: 2475-9066. DOI: 10.21105/joss.02505.

Theis, C. V. (1935). "The Relation between the Lowering of the Piezometric Surface and the Rate and Duration of Discharge of a Well Using Ground-Water Storage." In: *Transactions, American Geophysical Union* 16.2, p. 519. ISSN: 0002-8606. DOI: 10.1029/TR016i002p00519.

Therrien, R., R. McLaren, E. Sudicky, and S. Panday (2010). "HydroGeoSphere: A Three-Dimensional Numerical Model Describing Fully-Integrated Subsurface and Surface Flow and Solute Transport." In: *Groundwater Simulations Group, University of Waterloo, Waterloo, ON*.

Timlin, D. J., L. R. Ahuja, Y. Pachepsky, R. D. Williams, D. Gimenez, and W. Rawls (1999). "Use of Brooks-Corey Parameters to Improve Estimates of Saturated Conductivity from Effective Porosity." In: *Soil Science Society of America Journal* 63.5, pp. 1086–1092. ISSN: 1435-0661. DOI: 10.2136/sssaj1999.6351086x.

Timlin, D. J., R. D. Williams, L. R. Ahuja, and G. C. Heathman (2004). "Simple Parametric Methods to Estimate Soil Water Retention and Hydraulic Conductivity." In: *Developments in Soil Science*. Vol. 30. Development of Pedotransfer Functions in Soil Hydrology. Elsevier, pp. 71–93. DOI: 10.1016/S0166-2481(04)30005-X.

Tocci, M. D., C. Kelley, C. T. Miller, and C. E. Kees (1998). "Inexact Newton Methods and the Method of Lines for Solving Richards' Equation in Two Space Dimensions." In: *Computational Geosciences* 2.4, pp. 291–309. ISSN: 1573-1499. DOI: 10.1023/A:1011562522244.

Tockner, K. and J. A. Stanford (2002). "Riverine Flood Plains: Present State and Future Trends." In: *Environmental Conservation* 29.3, pp. 308–330. ISSN: 1469-4387, 0376-8929. DOI: 10.1017/S037689290200022X.

Tonina, D. and J. M. Buffington (2009). "Hyporheic Exchange in Mountain Rivers I: Mechanics and Environmental Effects: Mechanics of Hyporheic Exchange." In: *Geography Compass* 3.3, pp. 1063–1086. ISSN: 17498198. DOI: 10.1111/j.1749-8198.2009.00226.x.

Tóth, J. (1963). "A Theoretical Analysis of Groundwater Flow in Small Drainage Basins." In: *Journal of Geophysical Research (1896-1977)* 68.16, pp. 4795–4812. ISSN: 2156-2202. DOI: 10.1029/JZ068i016p04795.

Tóth, J. (1968). "Hydrogeological Study of the Three Hills Area, Alberta." In: *Research Council of Alberta*.

Touma, J. (2009). "Comparison of the Soil Hydraulic Conductivity Predicted from Its Water Retention Expressed by the Equation of Van Genuchten and Different Capillary Models." In: *European Journal of Soil Science* 60.4, pp. 671–680. ISSN: 1365-2389. DOI: 10.1111/j.1365-2389.2009.01145.x.

Trescott, P. C., G. F. Pinder, and S. P. Larson (1976). *Finite-Difference Model for Aquifer Simulation in Two Dimensions with Results of Numerical Experiments*. U.S. Department of the Interior, Geological Survey.

Triska, F. J., J. H. Duff, and R. J. Avanzino (1993). "The Role of Water Exchange between a Stream Channel and Its Hyporheic Zone in Nitrogen Cycling at the Terrestrial-Aquatic Interface." In: *Hydrobiologia*, p. 18.

Ufrecht, W. (2017). "Zur Hydrogeologie veränderlich fester Gesteine mit Sulfatgestein, Beispiel Gipskeuper (Trias, Grabfeld-Formation)." In: *Grundwasser* 22.3, pp. 197–208. ISSN: 1430-483X, 1432-1165. DOI: 10.1007/s00767-017-0362-3.

Valentová, J., P. Valenta, and L. Weyskrabová (2010). "Assessing the Retention Capacity of a Floodplain Using a 2D Numerical Model." In: *Journal of Hydrology and Hydromechanics* 58.4, pp. 221–232. ISSN: 0042-790X. DOI: 10.2478/v10098-010-0021-1.

van Genuchten, M. T. (1980). "A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils." In: *Soil Science Society of America Journal* 44.5, p. 892. ISSN: 0361-5995. DOI: 10.2136/sssaj1980.03615995004400050002x.

Van Looy, K., J. Bouma, M. Herbst, J. Koestel, B. Minasny, U. Mishra, C. Montzka, A. Nemes, Y. A. Pachepsky, J. Padarian, M. G. Schaap, B. Tóth, A. Verhoef, J. Vanderborght, M. J. Ploeg, L. Weihermüller, S. Zacharias, Y. Zhang, and H. Vereecken (2017). "Pedotransfer Functions in Earth System Science: Challenges and Perspectives." In: *Reviews of Geophysics* 55.4, pp. 1199–1256. ISSN: 8755-1209, 1944-9208. DOI: 10.1002/2017RG000581.

Vaux, W. G. (1968). "Intragravel Flow and Interchange of Water in a Streambed." In: *Fishery Bulletin of the Fish and Wildlife Service* 66.3, pp. 479–489.

Venkataraman, S. and R. Haftka (2004). "Structural Optimization Complexity: What Has Moore's Law Done for Us?" In: *Structural and Multidisciplinary Optimization* 28.6, pp. 375–387. ISSN: 1615-1488. DOI: 10.1007/s00158-004-0415-y.

Vereecken, H., M. Weynants, M. Javaux, Y. Pachepsky, M. G. Schaap, and M. van Genuchten (2010). "Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: A Review." In: *Vadose Zone Journal* 9.4, pp. 795–820. ISSN: 15391663. DOI: 10.2136/vzj2010.0045.

Vidon, P. G., M. K. Welsh, and Y. T. Hassanzadeh (2019). "Twenty Years of Riparian Zone Research (1997–2017): Where to Next?" In: *Journal of Environmental Quality* 48.2, pp. 248–260. ISSN: 1537-2537. DOI: 10.2134/jeq2018.01.0009.

Viessman, W. and G. L. Lewis (1996). *Introduction to Hydrology*. Harper Collins College.

Villinger, E. (1982). "Grundwasserbilanzen im Karstaquifer des Oberen Muschelkalks im Oberen Gäu (Baden-Württemberg)." In: *Geologisches Jahrbuch, Reihe C* 32, pp. 43–61.

Vogel, T., M. T. van Genuchten, and M. Cislerova (2000). "Effect of the Shape of the Soil Hydraulic Functions near Saturation on Variably-Saturated Flow Predictions." In: *Advances in Water Resources* 24.2, pp. 133–144. ISSN: 0309-1708. DOI: 10.1016/S0309-1708(00)00037-3.

Vogel, T. and M. Cislerova (1988). "On the Reliability of Unsaturated Hydraulic Conductivity Calculated from the Moisture Retention Curve." In: *Transport in Porous Media* 3.1, pp. 1–15. ISSN: 0169-3913, 1573-1634. DOI: 10.1007/BF00222683.

von Gunten, D., T. Wöhling, C. Haslauer, D. Merchán, J. Causapé, and O. A. Cirpka (2014). "Efficient Calibration of a Distributed Pde-Based Hydrological Model Using Grid Coarsening." In: *Journal of Hydrology* 519, pp. 3290–3304. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2014.10.025.

von Mises, R. (1928). *Wahrscheinlichkeit, Statistik Und Wahrheit*. https://www.webofscience.com/wos/alldb/full-record/BCI:BCI19310500016490. Wien: Springer.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer Science & Business Media.

Wagner, F. H. and G. Bretschko (2003). "Riparian Trees and Flow Paths between the Hyporheic Zone and Groundwater in the Oberer Seebach, Austria." In: *International Review of Hydrobiology* 88.2, pp. 129–138. ISSN: 1522-2632. DOI: 10.1002/iroh.200390009.

Wang, Y. and C. A. Shoemaker (2014). "A General Stochastic Algorithmic Framework for Minimizing Expensive Black Box Objective Functions Based on Surrogate Models and Sensitivity Analysis." In: *arXiv:1410.6271*. http://arxiv.org/abs/1410.6271. arXiv: 1410.6271.

Ward, A. S. (2016). "The Evolution and State of Interdisciplinary Hyporheic Research." In: *WIREs Water* 3.1, pp. 83–103. ISSN: 2049-1948. DOI: 10.1002/wat2.1120.

Ward, A. S., M. N. Gooseff, and K. Singha (2010). "Imaging Hyporheic Zone Solute Transport Using Electrical Resistivity." In: *Hydrological Processes* 24.7, pp. 948–953. ISSN: 1099-1085. DOI: 10.1002/hyp.7672.

Ward, A. S. and A. I. Packman (2019). "Advancing Our Predictive Understanding of River Corridor Exchange." In: *Wiley Interdisciplinary Reviews: Water* 6.1, e1327. ISSN: 2049-1948, 2049-1948. DOI: 10.1002/wat2.1327.

Ward, A. S., N. M. Schmadel, S. M. Wondzell, C. Harman, M. N. Gooseff, and K. Singha (2016). "Hydrogeomorphic Controls on Hyporheic and Riparian Transport in Two Headwater Mountain Streams during Base Flow Recession." In: *Water Resources Research* 52.2, pp. 1479–1497. ISSN: 1944-7973. DOI: 10.1002/2015WR018225.

Weber, T. K. D. (2018). *Personal Communication*.

Wegehenkel, M. and M. Selg (2002). "Räumlich hochauflösende Modellierung der Grundwasserneubildung im Neckartal bei Tübingen." In: *Grundwasser* 7.4, pp. 217–223. ISSN: 1430-483X, 1432-1165. DOI: 10.1007/s007670200033.

Welch, C., G. A. Harrington, and P. G. Cook (2015). "Influence of Groundwater Hydraulic Gradient on Bank Storage Metrics." In: *Groundwater* 53.5, pp. 782–793. ISSN: 1745-6584. DOI: 10.1111/gwat.12283.

Wilbraham, H. (1848). "On a Certain Periodic Function." In: *The Cambridge and Dublin Mathematical Journal* 3, pp. 198–201.

Willscher, B., R. Rausch, and M. Selg (2002). "Quantifizierung Des Wasserhaushalts Im Einzugsgebiet Der Brunnen Kiebingen Im Neckartal Bei Rottenburg." In:

Winter, T. C., J. W. Harvey, O. L. Franke, and W. M. Alley (1998). *Ground Water and Surface Water: A Single Resource*. Vol. 1139. US geological Survey.

Wittke, W. (2014). "Swelling Rock." In: *Rock Mechanics Based on an Anisotropic Jointed Rock Model (AJRM)*. John Wiley & Sons, Ltd. Chap. 8, pp. 181–208. ISBN: 978-3-433-60428-1. DOI: 10.1002/9783433604281.ch8.

Woessner, W. W. (2000). "Stream and Fluvial Plain Ground Water Interactions: Rescaling Hydrogeologic Thought." In: *Ground Water* 38.3, pp. 423–429. ISSN: 0017-467X, 1745-6584. DOI: 10.1111/j.1745-6584.2000.tb00228.x.

Wohl, E. (2021). "An Integrative Conceptualization of Floodplain Storage." In: *Reviews of Geophysics* 59.2, e2020RG000724. ISSN: 1944-9208. DOI: 10.1029/2020RG000724.

Wolman, M. G. and L. B. Leopold (1957). *River Flood Plains: Some Observations on Their Formation*. Tech. rep. 282-C. U.S. Government Printing Office, pp. 87–109. DOI: 10.3133/pp282C.

Wondzell, S. and M. Gooseff (2013). "Geomorphic Controls on Hyporheic Exchange Across Scales: Watersheds to Particles." In: *Treatise on Geomorphology*. Elsevier, pp. 203–218. ISBN: 978-0-08-088522-3. DOI: 10.1016/B978-0-12-374739-6.00238-4.

Xia, W., C. Shoemaker, T. Akhtar, and M.-T. Nguyen (2021). "Efficient Parallel Surrogate Optimization Algorithm and Framework with Application to Parameter Calibration of Computationally Expensive Three-Dimensional Hydrodynamic Lake PDE Models." In: *Environmental Modelling & Software* 135, p. 104910. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2020.104910.

Yabusaki, S. B., M. J. Wilkins, Y. Fang, K. H. Williams, B. Arora, J. Bargar, H. R. Beller, N. J. Bouskill, E. L. Brodie, J. N. Christensen, M. E. Conrad, R. E. Danczak, E. King, M. R. Soltanian, N. F. Spycher, C. I. Steefel, T. K. Tokunaga, R. Versteeg, S. R. Waichler, and H. M. Wainwright (2017). "Water Table Dynamics and Biogeochemical Cycling in a Shallow, Variably-Saturated Floodplain." In: *Environmental Science & Technology* 51.6, pp. 3307–3317. ISSN: 0013-936X. DOI: 10.1021/acs.est.6b04873.

Yeh, W. W.-G. (1986). "Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem." In: *Water Resources Research* 22.2, pp. 95–108. ISSN: 00431397. DOI: 10.1029/WR022i002p00095.

Zachara, J. M., X. Chen, X. Song, P. Shuai, C. Murray, and C. T. Resch (2020). "Kilometer-Scale Hydrologic Exchange Flows in a Gravel Bed River Corridor and Their Implications to Solute Migration." In: *Water Resources Research* 56.2. ISSN: 0043-1397, 1944-7973. DOI: 10.1029/2019WR025258.

Zhang, Y., J. Wang, P. Yang, and S. Xie (2017). "Movement of Lateral Hyporheic Flow between Stream and Groundwater." In: *Science China Earth Sciences* 60.11, pp. 2033–2040. ISSN: 1869-1897. DOI: 10.1007/s11430-016-9103-9.

Zhou, H., J. J. Gómez-Hernández, and L. Li (2014). "Inverse Methods in Hydrogeology: Evolution and Recent Trends." In: *Advances in Water Resources* 63, pp. 22–37. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2013.10.014.

Zhou, Y. and W. Li (2011). "A Review of Regional Groundwater Flow Modeling." In: *Geoscience Frontiers* 2.2, pp. 205–214. ISSN: 1674-9871. DOI: 10.1016/j.gsf.2011.03.003.

Zlotnik, V. A., M. B. Cardenas, and D. Toundykov (2011). "Effects of Multiscale Anisotropy on Basin and Hyporheic Groundwater Flow." In: *Groundwater* 49.4, pp. 576–583. ISSN: 1745-6584. DOI: 10.1111/j.1745-6584.2010.00775.x.

Appendix

17 Simplified Parametrization of the Unsaturated Zone

Stochastic numerical simulations are based on repeated simulations of the same flow model with varying values of internal model parameters. In most setups, the parameters are assumed to be independent. This assumption becomes problematic for parameters related to the unsaturated zone, as these are typically correlated (Schaap and Leij, 2000). Especially the parameter related to air entry pressure (α in the van Genuchten model; h_{AE} in the Brooks and Corey model) is often reported to be correlated to the saturated hydraulic conductivity, where higher hydraulic conductivities indicate smaller air entry pressures (Schaap and Leij, 2000; Guarracino, 2007). Other correlations have been identified too (e.g., Franzmeier, 1991; Timlin et al., 1999, 2004). Independent sampling of α and K_{sat} might therefore produce unrealistic parameters sets, even if both values are within realistic ranges by themselves.

An obvious solution to that problem would simply be to measure the parameters in lab studies for any given subsurface material to either specify them as fixed parameters, or to infer the correlations. However, this is difficult and requires a lot of effort (Reynolds et al., 2000; Vereecken et al., 2010). Furthermore, the results will then only be valid for the investigated sample(s), which might not be representative for the full hydrostratigraphic formation.

To avoid these issues, another solution is based on identifying relationships between the unsaturated zone parameters to other material properties that are easier to determine and then exploit the inverse relationship for inference. Such estimation techniques are the key elements of so-called pedotransfer functions (Vereecken et al., 2010; Van Looy et al., 2017; Szabó et al., 2021). A widely popular early study in this regard was conducted by Carsel and Parrish (1988), who worked with artificial data sets generated from the pedotransfer functions of Rawls et al. (1982) and Rawls and Brakensiek (1985, 1989). The outcome of Carsel and Parrish (1988) is a collection of multidimensional distributions, where correlated samples of K_{sat}, Θ_r, α and N can be drawn for any given soil type.

I decided against directly applying the relationships of Carsel and Parrish (1988) in stochastic numerical modeling for the following reasons:

- Accounting for correlations during sampling for stochastic simulations can be difficult, even when the correlations are known.
- The data is quantized into different soil type categories. Often it is unclear, which of the categories to choose, especially for bedrock layers. Such formations are obviously not considered in the collection of Carsel and Parrish (1988), but unsaturated zone parameters are required nonetheless in the modeling, where each hydrostratigraphic unit requires a description of its unsaturated behavior.

- The article of Carsel and Parrish (1988) appears to contain minor errors (e.g., missing zeros in comparison to the original data set of Rawls and Brakensiek (1985) and duplicate entries in the correlation tables). Even though these errors were probably only introduced during typesetting, it means that at least part of the presented data are likely to be incorrect.

To circumvent these problems, I use a direct deterministic, empirical relationship to infer α from a given or (randomly) sampled K_{sat}. For that I use the original relationships developed by Rawls et al. (1982) and Rawls and Brakensiek (1985, 1989), which are based on data sets of Brooks and Corey coefficients and sand, silt and clay content of the respective materials. By sampling within the two-dimensional parameter space of sand and silt content (the clay content is automatically assumed to make up the remaining fraction) I create 1500 virtual data points. To achieve a uniform point density in the ternary sand/silt/clay-fraction diagram, I sample according to a uniform three-dimensional Dirichlet distribution with concentration parameters equal to one. For these 1500 data points, I determine both α and K_{sat} according to the empirical relationships of Rawls and Brakensiek (1989, with some randomness introduced for the bulk density estimation). Finally, I fit a one-dimensional sigmoidal model that treats $\log \alpha$ as a function of $\log K_{\mathrm{sat}}$:

$$\log_{10} \alpha = -0.97 + 5.96 \frac{1}{1 + \exp\Big(-0.34 \cdot (\log_{10} K_{\mathrm{sat}} + 2.74)\Big)}. \tag{17.1}$$

This has the advantages of (1) a direct coupling of α and K_{sat}, which reduces the number of independent model parameters, and (2) providing a relationship that can infer α for any given K_{sat}, which might also be used for bedrock formations. Obviously, this is not a perfect solution, as it is (1) still based only on data from unconsolidated sediments, and (2) neglecting parametric uncertainty by assuming a deterministic relationship. Nonetheless, it is good enough in the context of steady-state modeling, where the parametrization of the unsaturated zone is of minor importance to the overall flow field anyway. The relationship of Equation 17.1 and the data it was fitted to is shown in Figure 39. The Matlab code to re-generate this relationship and figure are available in form of a repository at https://osf.io/9zycb/ (Allgeier, 2022a).

It should be noted that this approach is only used in Chapter IV. For Chapter III, which was developed earlier, I use independent sampling of α, K_{sat} and N.

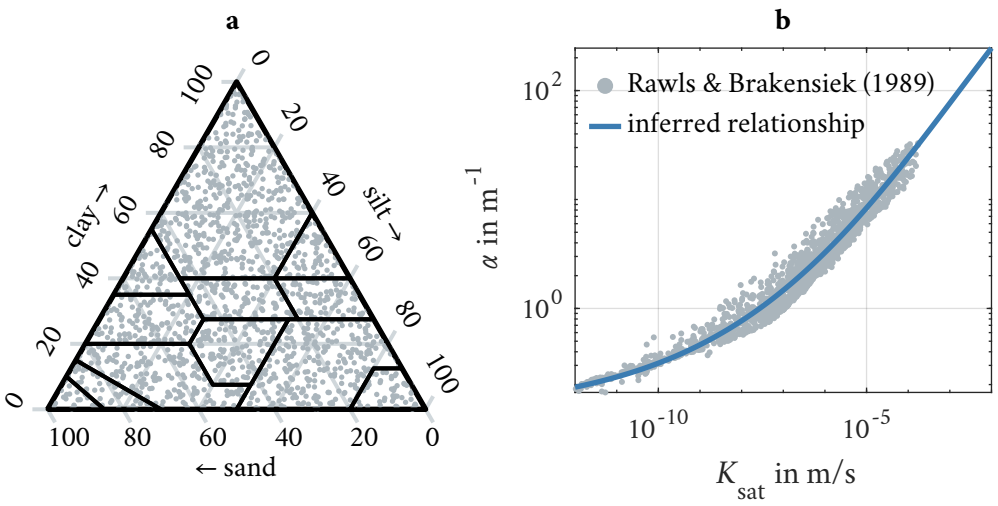

Figure 39: Derivation of a relationship to infer α from K_{sat}. **a**: Set of randomly genereated points. **b**: Corresponding values of α and K_{sat} according to Rawls and Brakensiek (1989), as well as the inferred relationship between those two parameters.

18 Literature Values of Model Parameters

The following tables contain collected literature information regarding the hydrostratigraphic units described in Section 3. With comments I indicate in what context the information was published:

- "modeled": The presented data were used in a modeling study either as input value or as calibrated model output.
- "measured": The presented data were measured either *in situ* or in lab-scale experiments.
- "cited": The presented data was mentioned in text as a reference to another publication, but that (supposedly original) publication is not available to me at the time of writing this dissertation. This might be either due to publisher-related restricted access, insufficient description of the original publication, or because the original publication is not available to the general public (anymore).

As average groundwater recharge rates cannot be measured directly and each quantitative recharge rate estimation is essentially the outcome of some (implicit) model, I use the terms "calculated" and "assumed" in Table 22, where the former indicates outputs of data- or model-driven approaches and the latter to denotes data that were (apparently) based on expert knowledge.

In cases where multiple data were documented, I collected the extreme values (indicated by "min" and "max"), to give an appropriate overview of the value ranges.

Within all tables, K is the hydraulic conductivity, a_r is the anisotropy ratio (vertical to horizontal), ϕ is the porosity and α/N are the van Genuchten parameters. In some cases, K was inferred as the ratio of a given transmissivity to the corresponding layer thickness.

Table 10: Literature values of parameters describing the Erfurt formation.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Butscher et al. (2011)	$1 \cdot 10^{-6}$	0.10				modeled
D'Affonseca et al. (2018)	$1 \cdot 10^{-6}$	0.10	0.06			modeled
D'Affonseca et al. (2020)	$1 \cdot 10^{-6}$	0.10				modeled
Erdal and Cirpka (2019)	$1 \cdot 10^{-8}$	0.02		0.5	1.5	modeled (min)
Erdal and Cirpka (2019)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max)
Erdal et al. (2020)	$1 \cdot 10^{-9}$	0.02		0.5	1.5	modeled (min)
Erdal et al. (2020)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max)
Keim and Pfäfflin (2006)	$6 \cdot 10^{-5}$					measured (min)
Keim and Pfäfflin (2006)	$4 \cdot 10^{-3}$					measured (max)
Kirchholtes and Ufrecht (2015)	$2 \cdot 10^{-7}$					measured (min)
Kirchholtes and Ufrecht (2015)	$5 \cdot 10^{-3}$					measured (max)
LGRB (2010)	$5 \cdot 10^{-6}$					measured (min)
LGRB (2010)	$6 \cdot 10^{-6}$					measured (max)
LGRB (2012)	$5 \cdot 10^{-5}$					measured
Schollenberger (1998)	$8 \cdot 10^{-6}$					measured (min)
Schollenberger (1998)	$4 \cdot 10^{-5}$					measured (max)
Schweizer et al. (2019)	$1 \cdot 10^{-5}$	1.00	0.02			modeled
Selle et al. (2013)	$7 \cdot 10^{-6}$	1.00				modeled (min)
Selle et al. (2013)	$1 \cdot 10^{-4}$	1.00				modeled (max)

Table 11: Literature values of parameters describing the unweathered Grabfeld formation; also comprises effective cases, where only a single formation was considered.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Butscher et al. (2011)	$1 \cdot 10^{-7}$	1.00				modeled
D'Affonseca et al. (2018)	$5 \cdot 10^{-8}$	0.10	0.03			modeled
D'Affonseca et al. (2020)	$5 \cdot 10^{-8}$	0.10				modeled
Erdal and Cirpka (2019)	$1 \cdot 10^{-9}$	0.02		0.5	1.5	modeled (min)
Erdal and Cirpka (2019)	$1 \cdot 10^{-7}$	1.00		5.0	9.0	modeled (max)
Erdal et al. (2020)	$1 \cdot 10^{-9}$	0.02		0.5	1.5	modeled (min)
Erdal et al. (2020)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max)
Hiller et al. (2012)	$1 \cdot 10^{-14}$					cited (min)
Hiller et al. (2012)	$1 \cdot 10^{-7}$					cited (max)
Hiller et al. (2012)	$1 \cdot 10^{-5}$	0.40				modeled
Kirchholtes and Ufrecht (2015)	$2 \cdot 10^{-13}$					measured (min, effective)
Schlosser et al. (2007)	$2 \cdot 10^{-13}$					mentioned (min)
Schlosser et al. (2007)	$3 \cdot 10^{-5}$					mentioned (max)
Schweizer et al. (2018)	$1 \cdot 10^{-10}$		0.03			cited (min)
Schweizer et al. (2018)			0.09			cited (max)
Schweizer et al. (2019)	$1 \cdot 10^{-10}$	1.00	0.02			modeled
Ufrecht (2017)	$1 \cdot 10^{-13}$					mentioned (min)
Ufrecht (2017)	$1 \cdot 10^{-7}$					mentioned (max)
Wittke (2014)	$1 \cdot 10^{-12}$					measured (min)
Wittke (2014)	$1 \cdot 10^{-7}$					measured (max)

Table 12: Literature values of parameters describing the weathered Grabfeld formation.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Butscher et al. (2011)	$1 \cdot 10^{-7}$	1.00				modeled
D'Affonseca et al. (2020)	$8 \cdot 10^{-6}$	1.00				modeled
D'Affonseca et al. (2018)	$8 \cdot 10^{-6}$	1.00	0.06			modeled
Erdal and Cirpka (2019)	$1 \cdot 10^{-7}$	0.02		0.5	1.5	modeled (min)
Erdal and Cirpka (2019)	$5 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max)
Erdal et al. (2020)	$1 \cdot 10^{-9}$	0.02		0.5	1.5	modeled (min)
Erdal et al. (2020)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max)
Hiller et al. (2012)	$1 \cdot 10^{-4}$	0.40				cited
Kirchholtes and Ufrecht (2015)	$1 \cdot 10^{-2}$					measured (max, effective)
Kortunov (2018)	$1 \cdot 10^{-5}$	1.00	0.03	4.0	1.3	modeled, effective
LGRB (2010)	$1 \cdot 10^{-6}$					measured
Martin (2021)	$3 \cdot 10^{-5}$					measured (min)
Martin (2021)	$6 \cdot 10^{-4}$					measured (max)
Schlosser et al. (2007)	$2 \cdot 10^{-9}$					mentioned (min)
Schlosser et al. (2007)	$3 \cdot 10^{-4}$					mentioned (max)
Schweizer et al. (2018)	$4 \cdot 10^{-7}$	0.10	0.08			modeled
Schweizer et al. (2019)	$2 \cdot 10^{-5}$	1.00	0.16			modeled
Selle et al. (2013)	$1 \cdot 10^{-5}$	1.00				modeled (min), effective
Selle et al. (2013)	$2 \cdot 10^{-5}$	1.00				modeled (max), effective
Ufrecht (2017)	$1 \cdot 10^{-6}$					mentioned (min)
Ufrecht (2017)	$9 \cdot 10^{-3}$					mentioned (max)

Table 13: Literature values of weathering zone depth L_w beneath surface (no weathering below) in the Grabfeld formation.

Source	L_w in m	Comment
Butscher et al. (2011)	30	figure
D'Affonseca et al. (2018)	5	modeled
Erdal and Cirpka (2019)	5	modeled (min)
Erdal and Cirpka (2019)	50	modeled (max)
Hiller et al. (2012)	20	modeled (min)
Hiller et al. (2012)	30	modeled (max)
Kehrer (1935)	1	estimated
Martin (2021)	28	measured
Schlosser et al. (2007)	34	mentioned
Schweizer et al. (2019)	35	figure
Ufrecht (2017)	20	mentioned (min)
Ufrecht (2017)	40	mentioned (max)

Table 14: Literature values of parameters describing relevant sandstone formations.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
D'Affonseca et al. (2018)	$5 \cdot 10^{-8}$	0.10	0.06			modeled
D'Affonseca et al. (2020)	$5 \cdot 10^{-8}$	0.10				modeled
Kortunov (2018)	$1 \cdot 10^{-5}$	1.00	0.03	4.0	1.3	modeled
LGRB (2010)	$8 \cdot 10^{-6}$					measured
Maier et al. (2013)	$1 \cdot 10^{-6}$	1.00	0.18	8.7	1.6	modeled (`kmHb`)
Maier et al. (2013)	$3 \cdot 10^{-8}$	1.00	0.41	2.3	1.2	modeled (`kmTr`)
Schlosser et al. (2007)	$1 \cdot 10^{-11}$					mentioned (min; `kmSt`)
Schlosser et al. (2007)	$2 \cdot 10^{-4}$					mentioned (max; `kmSt`)
Schlosser et al. (2007)	$4 \cdot 10^{-13}$					mentioned (min; `kmSw`, `kmHb`, `kmMh`)
Schlosser et al. (2007)	$1 \cdot 10^{-5}$					mentioned (max; `kmSw`, `kmHb`, `kmMh`)
Schlosser et al. (2007)	$6 \cdot 10^{-13}$					mentioned (min; `kmLw`)
Schlosser et al. (2007)	$3 \cdot 10^{-6}$					mentioned (max; `kmLw`)
Schlosser et al. (2007)	$2 \cdot 10^{-13}$					mentioned (min; `kmTr`)
Schlosser et al. (2007)	$2 \cdot 10^{-5}$					mentioned (max; `kmTr`)
Schweizer et al. (2019)	$8 \cdot 10^{-6}$	1.00	0.09			modeled
Selle et al. (2013)	$3 \cdot 10^{-9}$	1.00				modeled (min)
Selle et al. (2013)	$4 \cdot 10^{-9}$	1.00				modeled (max)

Table 15: Literature values of parameters describing clayey gravel (Ammer or comparable).

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
D'Affonseca et al. (2018)	$6 \cdot 10^{-4}$	0.10	0.15			modeled (Quaternary)
D'Affonseca et al. (2020)	$6 \cdot 10^{-4}$	0.10				modeled
Erdal and Cirpka (2019)	$1 \cdot 10^{-7}$	1.00		0.5	1.5	modeled (min; Quaternary)
Erdal and Cirpka (2019)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max; Quaternary)
Erdal et al. (2020)	$1 \cdot 10^{-9}$	1.00		0.5	1.5	modeled (min; Quaternary)
Erdal et al. (2020)	$1 \cdot 10^{-5}$	1.00		5.0	9.0	modeled (max; Quaternary)
Kortunov (2018)	$1 \cdot 10^{-5}$	1.00	0.20	35.0	5.3	modeled
Martin et al. (2020)	$3 \cdot 10^{-9}$					measured (min)
Martin et al. (2020)	$4 \cdot 10^{-4}$					measured (max)
Martin (2021)	$1 \cdot 10^{-6}$					measured (min)
Martin (2021)	$2 \cdot 10^{-3}$					measured (max)

Table 16: Literature values of parameters describing silty clay.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Carsel and Parrish (1988)	$6 \cdot 10^{-8}$		0.36	0.5	1.1	cited
Capuano and Jan (1996)	$3 \cdot 10^{-5}$	0.00				measured (field)
Capuano and Jan (1996)	$1 \cdot 10^{-9}$					measured (lab, min)
Capuano and Jan (1996)	$5 \cdot 10^{-7}$					measured (lab, max)
Kortunov (2018)	$1 \cdot 10^{-8}$	1.00	0.30	0.8	1.2	modeled
Leroueil et al. (n.d.)	$2 \cdot 10^{-10}$	0.30				measured (min)
Leroueil et al. (n.d.)	$3 \cdot 10^{-9}$	1.00				measured (max)
Lu et al. (2017)	$3 \cdot 10^{-11}$					measured (min)
Lu et al. (2017)	$6 \cdot 10^{-9}$					measured (min)
Lu et al. (2017)	$3 \cdot 10^{-6}$					measured (max)

Table 17: Literature values of parameters describing Tufa.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Kortunov (2018)	$1 \cdot 10^{-6}$	1.00	0.40	0.8	1.2	modeled (min)
Kortunov (2018)	$1 \cdot 10^{-5}$					modeled (max)
Martin et al. (2020)	$3 \cdot 10^{-7}$					measured (min)
Martin et al. (2020)	$5 \cdot 10^{-5}$					measured (max)
Martin (2021)	$4 \cdot 10^{-7}$					measured (min)
Martin (2021)	$2 \cdot 10^{-4}$					measured (max)

Table 18: Literature values of parameters describing alluvial fines.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Huch et al. (2013)	$1 \cdot 10^{-10}$					cited
Kortunov (2018)	$1 \cdot 10^{-6}$	1.00	0.30	0.8	1.2	modeled (min)
Kortunov (2018)			0.40			modeled (max)
Maier et al. (2013)	$1 \cdot 10^{-4}$	1.00	0.40	5.5	1.5	modeled
Minasny et al. (2004)	$5 \cdot 10^{-10}$		0.22	0.1	1.1	mentioned (min)
Minasny et al. (2004)	$2 \cdot 10^{-6}$		0.39	1.9	2.0	mentioned
Minasny et al. (2004)	$2 \cdot 10^{-4}$		0.55	12.4	7.6	mentioned (max)
Nagarajarao and Mallick (1980)	$3 \cdot 10^{-5}$		0.39			measured (min)
Nagarajarao and Mallick (1980)	$9 \cdot 10^{-5}$		0.42			measured (max)

Table 19: Literature values of parameters describing gravel (Neckar or comparable).

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
D'Affonseca et al. (2018)	$1 \cdot 10^{-3}$	0.10	0.15			modeled
D'Affonseca et al. (2020)	$1 \cdot 10^{-3}$	0.10				modeled
Keim and Pfäfflin (2006)	$5 \cdot 10^{-5}$					measured (min)
Keim and Pfäfflin (2006)	$4 \cdot 10^{-3}$					measured (max)
Kleinert (1976)	$2 \cdot 10^{-5}$		0.10			measured (min)
Kleinert (1976)	$7 \cdot 10^{-3}$					measured (max)
Kortunov (2018)	$1 \cdot 10^{-3}$	1.00	0.20	35.0	5.3	modeled
Lessoff et al. (2010)	$2 \cdot 10^{-3}$					measured
Willscher et al. (2002)	$1 \cdot 10^{-3}$					modeled

Table 20: Literature values of parameters describing hillslope fillings.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Maier et al. (2013)	$3 \cdot 10^{-4}$	1.00	0.40	2.5	1.3	modeled
Martin et al. (2020)	$8 \cdot 10^{-9}$					measured (min)
Martin et al. (2020)	$6 \cdot 10^{-5}$					measured (max)
Martin (2021)	$1 \cdot 10^{-3}$					measured

Table 21: Literature values of parameters describing generic top soil.

Source	K in m s^{-1}	a_r	ϕ	α in m^{-1}	N	Comment
Kortunov (2018)	$1 \cdot 10^{-6}$	1.00	0.40	0.8	1.2	modeled
Maier et al. (2013)	$1 \cdot 10^{-4}$	1.00	0.38	4.0	2.3	modeled
Archer et al. (2013)	$3 \cdot 10^{-7}$					measured (min)
Archer et al. (2013)	$5 \cdot 10^{-5}$					measured (max)
Weber (2018)	$7 \cdot 10^{-7}$					modeled (min)
Weber (2018)	$3 \cdot 10^{-6}$					modeled
Weber (2018)	$6 \cdot 10^{-5}$					modeled (max)

Table 22: Literature values of groundwater recharge rates r_R in the Ammer floodplain study area or close to it.

Source	r_R in m s^{-1}	Comment
Ammer et al. (1983)	$3.1 \cdot 10^{-9}$	calculated (min, forest)
Ammer et al. (1983)	$3.5 \cdot 10^{-9}$	calculated (max, forest)
BfG (2003)	$7.9 \cdot 10^{-10}$	calculated (min)
BfG (2003)	$4.8 \cdot 10^{-9}$	calculated (max)
Erdal and Cirpka (2019)	$2.5 \cdot 10^{-9}$	assumed (min)
Erdal and Cirpka (2019)	$4.8 \cdot 10^{-9}$	assumed (max)
Harreß (1973)	$3.3 \cdot 10^{-9}$	
Gudera (2015)	$1.6 \cdot 10^{-9}$	cited (min)
Gudera (2015)	$6.3 \cdot 10^{-9}$	cited (max)
Kleinert (1976)	$7.0 \cdot 10^{-9}$	calculated
Kortunov (2018)	$6.3 \cdot 10^{-9}$	cited
Maier et al. (2013)	$1.0 \cdot 10^{-8}$	assumed
Martin et al. (2020)	$6.3 \cdot 10^{-9}$	assumed
Wegehenkel and Selg (2002)	$1.4 \cdot 10^{-9}$	calculated (min)
Wegehenkel and Selg (2002)	$4.8 \cdot 10^{-9}$	calculated (long-term)
Wegehenkel and Selg (2002)	$9.4 \cdot 10^{-9}$	calculated (max)