Modelling Surface Water-Groundwater Exchange

Evaluating Model Uncertainty from the Catchment to Bedform-Scale

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Reynold Chow, M.Sc., P.Geo
aus Toronto, Kanada

Tübingen
2019

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

Tag der mündlichen Qualifikation:	21.05.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr.-Ing. Wolfgang Nowak
2. Berichterstatter:	Dr. rer. nat. Thomas Wöhling
3. Berichterstatter:	Prof. Dr.-Ing. Olaf A. Cirpka

Abstract

My dissertation focuses on evaluating model uncertainties when numerically simulating surface water-groundwater (sw-gw) interactions at different scales. To do so, I mainly use HydroGeoSphere, a physically-based distributed finite element model that fully couples variably saturated subsurface flow with surface water flow. I evaluate three predominant model uncertainty types at three different scales of sw-gw interaction. For each of these investigations, I selected a corresponding study site.

Firstly, I evaluate structural (conceptual) uncertainty from delineating baseflow contribution areas to gaining stream reaches, or stream capture zones, at the catchment-scale. I investigate how the delineated stream capture zone (in the Alder Creek watershed) can differ due to the chosen model code and delineation method (Chow et al. 2016). The results indicate that different models can calibrate acceptably well to the same data and produce very similar distributions of hydraulic head, but can produce different capture zones. The stream capture zone is highly sensitive to the post-processing particle tracking algorithm. Reverse transport is an alternative and more reliable approach that accounts for local-scale parameter uncertainty and provides probability intervals for the stream capture zone. The two methods can be combined to enhance the overall confidence in the delineated stream capture zone.

Secondly, I evaluate parameter uncertainty when simulating meander-scale hyporheic exchange by conducting a model-based sensitivity (bathymetry) and uncertainty (subsurface *K*-distribution) analysis. I select the Steinlach River Test Site for demonstration. I conduct a sensitivity analysis to determine the aspects of river bathymetry that have the greatest influence on the predictive biases (Chow et al. 2018). Results indicate that simulating hyporheic exchange with a high-resolution detailed bathymetry using a 3D fully coupled sw-gw model leads to nested multi-scale hyporheic exchange systems. A poorly resolved bathymetry will underestimate the small-scale hyporheic exchange, biasing the simulated hyporheic exchange towards larger scales, thus leading to overestimates of hyporheic exchange residence times. The detailed river slope alone is not enough to accurately simulate the locations and magnitudes of losing and gaining river reaches. Thus, local bedforms in terms of bathymetric highs and lows within the river are required. Incorporating local bedforms will likely capture the nested nature of hyporheic exchange, leading to more physically meaningful simulations of hyporheic exchange fluxes and transit times.

Additionally, I conduct an uncertainty analysis to evaluate the trade-offs between intrinsic (irreducible) and epistemic (reducible) model errors when choosing between homogeneous and highly complex subsurface parameter structures (Chow et al. 2019). Results indicate that, if the parameter structure is too simple, it will be limited by intrinsic model errors. By increasing subsurface complexity through the addition of zones or heterogeneity, we can begin to exchange intrinsic for epistemic errors. Thus, choosing the appropriate level of detail to represent the subsurface parameter distributions depends on the acceptable range of intrinsic structural errors for the given modelling objectives and the available site data.

Thirdly, I evaluate data uncertainty at the bedform-scale in a fractured rock setting by modelling a conservative tracer experiment at the Eramosa Bedrock River Site. I use a stochastic discrete fracture network framework to represent the subsurface fractured bedrock connectivity, and in doing so produce a probabilistic distribution of the potential hyporheic exchange extents and residence times. The results indicate that the coincidence of fractures and hydraulic gradients determine the spatial extents of bedform-scale hyporheic exchange. Furthermore, hyporheic exchange residence times in bedrock rivers at the bedform-scale are potentially orders of magnitude longer when compared to fluvial rivers (i.e., months to years vs. minutes to hours).

In the age of highly-parameterized integrated hydrological models, there is an increasing need to understand whether we are getting the right forecasts for the right reasons. Modelling is only a single tool in the scientific toolbox that works best when combined with others, e.g., field and laboratory experiments. The next generation of integrated hydrological modellers will face evermore challenging objectives, which to achieve will require multi-disciplinary teams. Regardless of the origins of your knowledge-base, it is important to always approach numerical models with a healthy level of skepticism. Ultimately, scientists should approach models with a relentless fearlessness to falsify them, which is an absolute necessity if we wish to continually push the envelope of scientific knowledge.

Zusammenfassung

Meine Dissertation beschäftigt sich mit der Einschätzung von Modellunsicherheiten bei der numerischen Simulation von Oberflächenwasser-Grundwasser-Wechselwirkungen auf unterschiedlichen Skalen. Dafür benutze ich hauptsächlich HydroGeoSphere, ein physikalisch-basiertes finite Elemente Modell, das die veränderlich gesättigte Untergrundströmung mit der Oberflächenströmung vollständig koppelt. Ich erkunde drei vorherrschende Unsicherheitstypen auf drei Skalen, indem ich die Oberflächenwasser-Grundwasser-Wechselwirkungen simuliere. Für jede dieser Untersuchungen wähle ich ein geeignetes Untersuchungsgebiet aus.

Zunächst untersuche ich die strukturbedingte (konzeptionelle) Unsicherheit auf der Einzugsgebietsskala bei der Bestimmung von Basisabflussbeitragsflächen exfiltrierender Flussabschnitte oder Flusseinzugsgebiete. Dafür wähle ich das Einzugsgebiet des “Alder Creek” aus. Ich erforsche, inwieweit die so bestimmten Flusseinzugsgebiete von der Wahl des Modellcodes und der Abgrenzungsmethode abhängen (Chow et al. 2016). Die Ergebnisse deuten darauf hin, dass verschiedene Modelle hinreichend gut auf dieselben Daten kalibriert werden können, wobei sehr ähnliche Verteilungen des hydraulischen Potentials, aber unterschiedliche Geometrien der Einzugsgebiete entstehen. Das Flusseinzugsgebiet ist sehr sensitiv gegenüber der Wahl des Particle-Tracking-Nachverarbeitungsalgorithmus. Umgekehrter Transport stellt einen alternativen und zuverlässigeren Ansatz dar, der lokale Parameterunsicherheit berücksichtigt und Wahrscheinlichkeitsintervalle für das Flusseinzugsgebiet liefert. Beide Methoden können kombiniert werden, um die gesamte Konfidenz des abgeleiteten Flusseinzugsgebiets zu erhöhen.

Als Zweites widme ich mich der Parameterunsicherheit, indem ich eine Sensitivitäts- (Bythymetrie) und Unsicherheitsanalyse (K-Verteilung des Untergrunds) für Simulationen des hyporheischen Austauschs auf der Mäanderskala durchführe. Dies untersuche ich am Beispiel des Steinlach-Testgeländes. Ich vollziehe eine Sensitivitätsanalyse um diejenigen Aspekte der Flussbathymetrie zu identifizieren, die den größten Einfluss auf systematische Vorhersageabweichungen haben (Chow et al. 2018). Die Ergebnisse zeigen, dass die Simulation des hyporheischen Austausches mit einer hochaufgelösten und detaillierten Bathymetrie unter Verwendung eines dreidimensionalen vollständig gekoppelten Oberflächenwasser-Grundwasser-Modells zu verschachtelten Multi-Skalen-Systemen von hyporheischem Austausch führt. Eine schlecht aufgelöste Bathymetrie führt zur

Unterschätzung von kleinskaligem hyporheischen Austausch, wodurch derselbe auf größere Skalen verschoben wird. Dies führt zur Überschätzung von Aufenthaltszeiten in der hyporheischen Zone. Eine detaillierte Beschreibung der Flusssteigung allein reicht dabei nicht aus, um die Lage und den Betrag des Wasseraustauschs von exfiltrierenden und infiltrierenden Flussabschnitten richtig zu simulieren. Folglich sind lokale Flussgeometrien (Sohlformen) in Form von bathymetrischen Hoch- und Tiefpunkten des Flussverlaufs notwendig. Bei Berücksichtigung dieser Sohlform wird die verschachtelte Natur des hyporheischen Austauschs besser erfasst, wodurch physikalisch sinnvollere Simulationen von Austauschflüssen und Verweilzeiten erreicht werden können.

Zusätzlich führe ich eine Unsicherheitsanalyse durch, um den Trade-off zwischen intrinsischen (nicht reduzierbaren) und epistemischen (reduzierbaren) Modellfehlern abzuwägen, der bei der Wahl zwischen homogenen und hoch-komplexen Untergrundparameterstrukturen entsteht (Chow et al. 2019). Die Ergebnisse deuten darauf hin, dass die Aussagekraft eines Modells bei einer zu simplen Parameterstruktur durch intrinsische Fehler eingeschränkt ist. Mit zunehmender Untergrundkomplexität durch die Berücksichtigung von Zonen oder Heterogenität können intrinsische Fehler durch epistemische Fehler ersetzt werden. Daher hängt die Wahl des Detaillierungsgrads im Hinblick auf die Untergrundparameterstruktur vom annehmbaren Grad an intrinsischen Strukturfehlern bei gegebenen Modellzielen, sowie von den verfügbaren Standortdaten ab.

Drittens bewerte ich Datenunsicherheit auf der Sohlformskala in geklüftetem Fels. Als Demonstrator wähle ich die Modellierung eines konservativen Tracerexperiments am Standort "Eramosa Bedrock River". Ich benutze ein stochastisches, diskretes Frakturnetzwerk um die Konnektivität von Klüften des Festgesteins abzubilden. Dabei erzeuge ich eine probabilistische Verteilung von potentiellen Ausmaßen und Aufenthaltszeiten hyporheischer Austauschvorgänge. Die Ergebnisse zeigen, dass das Zusammenspiel von Klüften und hydraulischen Gradienten die räumlichen Ausdehnungen des hyporheischen Austauschs auf Sohlformskala bestimmen. Außerdem sind die hyporheischen Austauschverweilzeiten in Festgesteinsflüssen auf der Sohlformskala im Vergleich zu fluvialen Flüssen potentiell um Größenordnungen länger (d.h. Monate oder Jahre gegenüber Minuten oder Stunden).

Im Zeitalter der hochparametrisierten integrierten hydrologischen Modelle gibt es eine zunehmende Notwendigkeit zu verstehen, ob aus den richtigen Gründen richtige Vorhersagen getroffen werden. Modellierung ist nur ein einzelnes Werkzeug im naturwissenschaftlichen Werkzeugkasten, das am besten in Kombination mit anderen Werkzeugen wie Feld- und

Laborexperimenten funktioniert. Die nächste Generation der Modellierer, die sich mit integrierten hydrologischen Systemen beschäftigt, wird immer anspruchsvolleren Zielen gegenübergestellt sein, deren Umsetzung multidisziplinäre Teams verlangt. Unabhängig von den Ursprüngen der eigenen Wissensbasis ist es wichtig, numerischen Modellen stets mit einem gesunden Maß an Skepsis zu begegnen. Abschließend sollten Naturwissenschaftler mit der unerbittlichen Furchtlosigkeit zur Falsifizierung an Modelle herantreten. Dies ist eine absolute Notwendigkeit, um die wissenschaftliche Erkenntnis ständig zu erweitern.

Acknowledgements

Firstly, I would like to thank Prof. Dr.-Ing. Olaf A. Cirpka for giving me the opportunity to join the Integrated Hydrosystem Modelling Research Training Group at the University of Tübingen (Deutsche Forschungsgemeinschaft Grant GRK 1829/2). Secondly, I would like to thank my primary supervisor Prof. Dr.-Ing. Wolfgang Nowak who has mentored me in my development as a scientist. Thirdly, I would like to thank my secondary supervisor, Dr. rer. nat. Thomas Wöhling, for his continued support and guidance. I would also like to thank Prof. Beth Parker and Dr. Colby Steelman for hosting me at the G^{360} Institute for Groundwater Research at the University of Guelph. Finally, I would like to thank the developers at Aquanty Inc. for technical support and a HydroGeoSphere software license.

I am honoured to have worked and discussed with my friends/colleagues from the Tübingen RTG Hydromod, Stuttgart LS^3, and associated research groups (Abelardo Rodriguez, Alex Haluska, Aline Schäfer, Ana González, Anneli Guthke, Bo Xiao, Carolin Paneru, Daniel Erdel, Diane von Gunten, Elena Petrova, Emilio Sánchez-Léon, Evgenii Kortunov, Felix Bode, Fernando Mazo D'Affonseca, Gaëlle Guillet, Gianna Marschmann, Han-Fang Hsueh, Jannik Haas, Jeremy Bennett, Julia Knapp, Julian Mehne, Jürnjakob Dugge, Karim Moghanjoghi, Marvin Höge, Matthias Loschko, Maximiliane Herberich, Michael Sinsbeck, Ran Wei, Sebastian Most, Shangua Li, Stephan Adam, Yan Liu, and Zhongwen Bao). Special thanks to Jonas Allgeier and Philipp Selzer for reviewing my dissertation.

It was a pleasure to mentor and supervise MSc students from The University of Stuttgart's Water Resources Engineering and Management (WAREM) program and undergraduate co-op (HiWi) students from the University of Waterloo. I would like to thank Diego Motavita and Hao Wu for their hard work on their MSc. I would also like to express my gratitude to Melanie Perreault, Evan McLaughlin, Christine Gomez, and Alex Thoms for their work during their internships.

I am grateful to Monika Jekelius, Alica Abberger, and Ute Coquet who helped me with all the administrative issues.

Last but not least, I would like to thank my wife Junko for her continued support at home.

Author declaration

I declare that I have developed and written the enclosed dissertation completely by myself and that I have not used sources or means without declaration in the text. The contents of the following chapters are from accepted peer-reviewed articles published in academic journals. The contributions of all co-authors are clearly stated here and in the footnotes at the beginning of their respective chapters. Appendix A only presents the article abstract.

Chapter 3. Chow, R., Frind, M.E., Frind, E.O, Jones, J.P., Sousa, M.R., Rudolph, D.L., Molson, J.W., and Nowak, W. (2016). Delineating baseflow contribution areas for streams – A model and methods comparison. Journal of Contaminant Hydrology 195, 11–22.

Chow, R.: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript to address peer-review comments. Finalized and approved published article. Analyzed and evaluated all model results. Conceptualized modelling study. Preliminary results presented at the 2011 GeoHydro conference in Québec City (Chow et al. 2011) and in University of Waterloo Master's thesis (Chow 2012). ***Frind, M.E.***: Developed FEFLOW model and analyzed FEFLOW results. ***Frind, E.O.***: Reviewed preliminary works and original draft manuscript. Helped with revisions to address peer-review comments. Conceptualized modelling study. Provided supervision. ***Jones, J.P.***: Developed preliminary HydroGeoSphere model. Reviewed preliminary works. Conceptualized modelling study. Provided supervision. ***Sousa, M.R.***: Developed FEFLOW model and provided technical modelling support. ***Rudolph, D.L.***: Reviewed preliminary works. ***Molson J.W.***: Provided technical modelling support. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

Chapter 4. Chow, R., Wu, H., Bennett, J.P., Dugge, J., Wöhling, T., and Nowak, W. (2018). Sensitivity of Simulated Hyporheic Exchange to River Bathymetry: The Steinlach River Test Site. Groundwater.

Chow, R.: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript addressing peer-review comments. Finalized and approved published article. Analyzed and evaluated model results. Conceptualized modelling study. ***Wu, H.***: Developed results and figures for original draft manuscript. Preliminary results presented in University of Stuttgart Master's thesis (Wu 2017). ***Bennett, J.P.***: Developed conceptual and Multiple Point Geostatistics (MPS) model. Helped write MPS section in original draft manuscript. ***Dugge, J.***: Developed preliminary Steinlach River Test Site HGS model. Conceptualized modelling study. ***Wöhling, T.***: Reviewed original draft manuscript. Conceptualized modelling study. Provided supervision. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

Chapter 5. Chow, R., Bennett, J.P., Dugge, J., Wöhling, T., and Nowak, W. (2019). Evaluating subsurface parameterization to simulate hyporheic exchange: The Steinlach River Test Site. Groundwater.

Chow, R.: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript addressing peer-review comments. Finalized and approved published article. Analyzed and evaluated model results. Conceptualized modelling study. ***Bennett, J.P.***: Developed conceptual and Multiple Point Geostatistics (MPS) model. Helped write MPS section in original draft manuscript. Helped with article revisions to address peer-review comments. ***Dugge, J.***: Developed preliminary Steinlach River Test Site HGS model. Conceptualized modelling study. ***Wöhling, T.***: Reviewed original draft manuscript. Conceptualized modelling study. Provided supervision. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

Appendix A. Motavita, D.F., Chow, R., Guthke, A., and Nowak, W. (2019). The Comprehensive Differential Split-Sample Test: A stress-test for hydrological model robustness under climate variability. Journal of Hydrology.

Motavita, D.F.: Wrote preliminary works and original draft manuscript. Revised manuscript to address peer-review comments. Reduced and analyzed data. Analyzed and evaluated all model results. Finalized and approved published article. Conceptualized modelling study. Preliminary results presented in University of Stuttgart Master's thesis (Motavita 2016). ***Chow, R.***: Helped write original draft manuscript. Helped revise article to address peer-review comments. Helped finalizing article. Conceptualized modelling study. Provided supervision. ***Guthke, A.***: Helped write original draft manuscript. Reviewed Bayesian data analysis methods. Conceptualized modelling study. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Provided supervision.

Contents

List of Figures

List of Tables

1. Introduction

Groundwater and surface water systems have long been considered separately by geoscientists (hydrogeologists and hydrologists) because they are known to act at very different temporal and spatial scales (Figure 1.1). However, in the last 30 years the science of surface water-groundwater (sw-gw) exchange has advanced tremendously. We now know that the environmental sustainability of some surface water systems depends on groundwater discharge to maintain steady baseflows, temperatures, habitat connectivity, and nutrients to support healthy aquatic eco-systems (Vaux 1968, Beatty et al. 2010, Bätz et al. 2016). Furthermore, the water that is associated with sw-gw exchange has the potential to undergo biogeochemical reactions, which can be attributed to the retention and degradation of pollutants within watersheds, ultimately impacting ecosystem health (Kalbus et al. 2006, Buss et al. 2009, Grathwohl et al. 2013, Trauth et al. 2018). Understanding the natural ecological functions of rivers and identifying riverbed exchange mechanisms are especially important for environmental engineers and freshwater biologists in their efforts to reclaim and enhance sw-gw exchange (Woessner 2000).

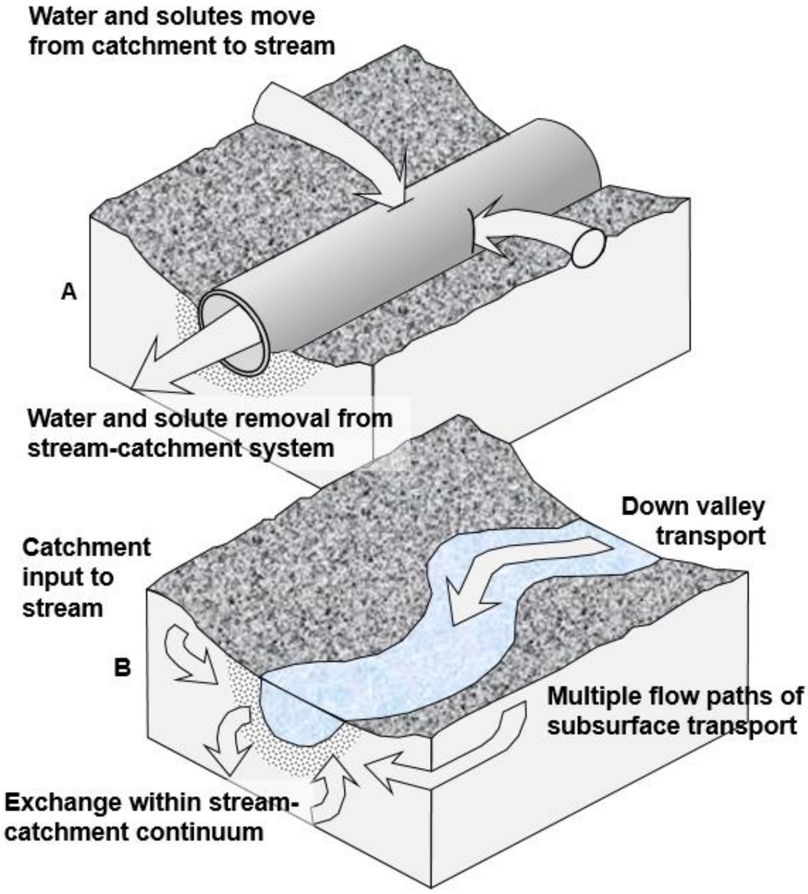


Figure 1.1: Surface water-groundwater exchange early and modern view:

A) Early view of stream function as a closed pipe system, receiving nutrients (or other solutes) from the catchment with negligible amounts of exchange between surface water-groundwater.

B) Contrasting view of the stream's function places the stream as an integral part of the catchment system (adapted from Bencala 1993).

Understanding the biogeochemical nature and physical mechanisms that control sw-gw exchange requires knowledge from a combination of geoscience disciplines. To understand the physical nature of sw-gw exchange alone requires, as a minimum, knowledge of hydrogeology, hydrology and fluid mechanics, as well as geomorphology and sedimentology. My research focuses on studying the physical mechanisms that control sw-gw exchange using a model-based approach.

Model-based approach. Current geoscientific methods available to better understand the physical processes of sw-gw exchange include experimentation in the field and laboratory, as well as numerical modelling. Within each of these methods are vast sub-disciplines of geoscience, each with a wide variety of specialized methods and jargon. For example, hydrogeological (e.g., slug and pumping tests) and geophysical (e.g., electrical resistivity surveys) methods could be employed in the field to characterize sw-gw exchange at a site and to construct a site conceptual model (Doro et al. 2013, Ward et al. 2013, Steelman 2015a, Kennedy 2017). In the laboratory, sw-gw flume experiments could be conducted with the added benefit of controlling boundary conditions and material properties (Salehin et al. 2004, Packman et al. 2004, Tonina and Buffington 2007). Making careful observations in both field and laboratory experiments is critical in producing the measurements that can be used to derive governing equations for system processes. However, field and laboratory experiments can be expensive in both human (e.g., time, skills) and material (e.g., equipment, laboratories) resources. Therefore, an alternative potentially cost-effective form of experimentation is the use of numerical modelling that can incorporate the physical, chemical, and/or biochemical processes relevant to a given situation. Therefore, the research in this dissertation is model-based, using models that were developed in connection with field data collected at select sites.

Due to this focus, it is worthwhile looking at properties, advantages, disadvantages, and challenges of modelling in more detail. Lehr (1979) defines a numerical model as a simplified mathematical representation of a complex physical system and gives the example of a road map as a type of model that depicts a complex network of roads in a simplified manner for navigation. Numerical models can be process-based (i.e., physically-based models), which use processes and principles of physics (e.g., Darcy’s Law) to represent sw-gw flow interactions within a given domain (Anderson et al. 2015). Alternatively, there are also data-driven or empirically-based models, which use statistical equations from available data to derive relationships between a measured variable (e.g., precipitation, temperature) and the desired unknown variable (e.g., discharges at an outlet of a catchment).

Numerical modelling is powerful because it allows one to analyze a wide range of conditions rapidly and inexpensively (not included: integrated numerical models that are computationally expensive and require supercomputing clusters to achieve reasonable run-times) compared to controlled field and laboratory experiments (Frind 2003). For example, one could perform a sensitivity analysis with a numerical model (Section 1.1.4 and Chapter 4) to reveal the parameters that dominate the behaviour of the system under investigation, and thus prioritize the future study of specific parameters and processes. One could also perform a scenario analysis, where a number of “what if” scenarios could be posed to test hypotheses about the system properties and behaviour. Finally, numerical models play an important role when studying potentially hazardous chemicals, which by law cannot be discharged into the environment for experiments (in most developed countries). Thus, validated numerical models (by proxy field and laboratory studies) can be a way to simulate the fate and transport of potentially hazardous chemical experiments without risk to the environment (Sudicky et al. 1983).

Numerical modelling does not come without caveats however, it is important to note that “essentially, all models are wrong, but some are useful” (Box and Draper 1987). The meaning of Box and Draper’s (1987) quote is that every numerical model is wrong because it is a simplification of a complex natural system (i.e., reality), nevertheless numerical models can be useful because they can help us explain and understand the importance of various system components and their interactions.

Relevance of uncertainties. Furthermore, numerical modelling can be used to forecast future systems states. I have chosen to use the word forecast over prediction because it indicates that all estimates of future conditions have uncertainty (Anderson et al. 2015). Uncertainties arise in numerical modelling because all models are wrong (to varying degrees, Box and Draper 1987) and because of uncertainties arising from four broad categories (Gupta et al. 2005, Jones and Mendoza 2012). These four broad uncertainty categories are:

1. Structural/conceptual uncertainty (e.g., missing governing processes, mathematical formulation is incomplete and/or inaccurate, multiple working hypotheses, equifinality).
2. Data uncertainty (e.g., imperfect and/or incomplete measurements)
3. Parameter uncertainty (e.g., nonuniqueness, identifiability)
4. Predictive uncertainty (e.g., out-of-sample simulations)

These four categories of model uncertainty will be further described in Section 1.1.

As a result of the aforementioned model uncertainties, forecasts from numerical models should always be analyzed with a reasonable degree of skepticism and caution (Anderson and Woessner 1992, Konikow and Bredehoeft 1992). Failing to recognize the uncertainties inherent in forecasts from numerical models can lead to disastrous results. For example, overfishing and the eventual demise of the Grand Banks cod fisheries off the coast of Canada was partly due to inaccurate estimates of total allowable catch based on numerically modelled cod stock size (Pilkey and Pilkey-Jarvis 2007).

Anderson et al. (2015) describes a model as being stochastic if any of its parameters, input or boundary conditions have a probabilistic distribution; otherwise, the model is considered to be deterministic. Many geoscientists (e.g., Beven and Binley 1992, Pappenberger and Beven 2006, Cirpka and Valocchi 2016) have argued that the notion of a deterministic calibrated model has no place in environmental models since uniqueness is always achieved at the cost of error-inducing simplifications. Instead, it has debated that a large number of simulations should be used within a stochastic Bayesian framework to explore the range of forecast possibilities rather than making a single best-fit (deterministic) forecast (Pappenberger and Beven 2006, Gómez-Hernández 2006, Cirpka and Valocchi 2016).

Ultimately the usefulness of a numerical model and the amount of uncertainty that one is willing to accept depends on the purpose of the model. The purpose of a forecasting model is to plan future actions based on the likelihood that an event will occur given that some action or inaction will be taken, which is very different from predicting what will happen in the future (Doherty 2011, Anderson et al. 2015). Forecasting models are often used as engineering calculators to plan and design engineering projects (e.g., the dewatering of a proposed open pit mine in a humid region to ensure geotechnical slope stability). Alternatively, interpretive models (Anderson et al. 2015) have the purpose of improving our system understanding, which in this context would be to better understand sw-gw processes and evaluating the properties that govern their interactions. Voss (2011a) adds that there are few insights to be gained from the time-consuming and tedious process of calibrating highly parameterized models to data, rather the goal of a modelling study should be to learn something about the system through a model-based analysis. In other words, a calibrated model should not be the end product of a modelling study, instead the desired result should be the lessons learned from the model development and analysis. Thus, stakeholders concerned with water management issues should buy advice from a competent hydro(geo)logist, not a hydro(geo)logical model (Voss 2011b).

Scales. Sw-gw exchange operate at various spatial and temporal scales. Here, I consider large-scale sw-gw interactions as watershed-scale and river-corridor-scale (1 km to 10 m) processes, which can be mainly topography-driven (Cardenas 2008, Wang et al. 2018). Others (e.g., Chen et al. 2014, Maxwell et al. 2015) classify large-scale hydro(geo)logical modelling as continental-scale (1,000's km) modelling, however continental-scale interactions will not be considered here. I consider medium-scale sw-gw exchange to be meander-scale processes (1 m to 10 cm), which are typically driven by hydrostatic processes (Wondzell and Gooseff 2013). I classify small-scale sw-gw exchange as bedform or riffle-pool-riffle-scale (10 cm to 1 mm), which can be primarily hydrodynamic-driven (Trauth et al. 2013).

An important sw-gw process, known as hyporheic exchange, acts from the topography-driven watershed-scale (km to 100 km) to the hydrodynamic-driven bedform-scale (cm to m) (Cardenas 2008). Hyporheic exchange is defined as the process where river water infiltrates into the subsurface, potentially mixes with groundwater, and returns to the river. Figure 1.2 is a conceptual drawing of the nested hierarchies of hyporheic exchange systems from the river corridor to the bedform-scale.

Boano et al. (2014) have shown that the scale of hyporheic exchange can be associated with different transit times (Figure 1.3). The scales of sw-gw interactions I will focus on in my dissertation will be the catchment-scale (Chapter 3), meander-scale (Chapters 4 and 5), and bedform-scale (Chapter 6). These research chapters (green text) and their corresponding scales (black text) are labelled in Figure 1.3. By examining Figure 1.3, we can get a sense of both the spatial and temporal magnitudes expected at the scales of sw-gw exchange in the upcoming chapters. The distribution of hyporheic transit times (the time taken for water to move through the subsurface) and the percentage of sw-gw exchange per unit length of river have been shown to be critical factors when determining the magnitude of potential biogeochemical reactions (Cardenas 2015, Altenkirch et al. 2016). Anderson et al. (2015) have noted that simulating transit times with groundwater models is more uncertain than the simulation of heads because it requires more detailed knowledge of aquifer heterogeneities.

Scope. This dissertation aims to advance our understanding about the physical processes in sw-gw interactions through modelling-based research. Based on the discussion above, a meaningful definition of the scope of my dissertation and research questions must consider combinations of the various model uncertainty types and different exchange processes at different scales.

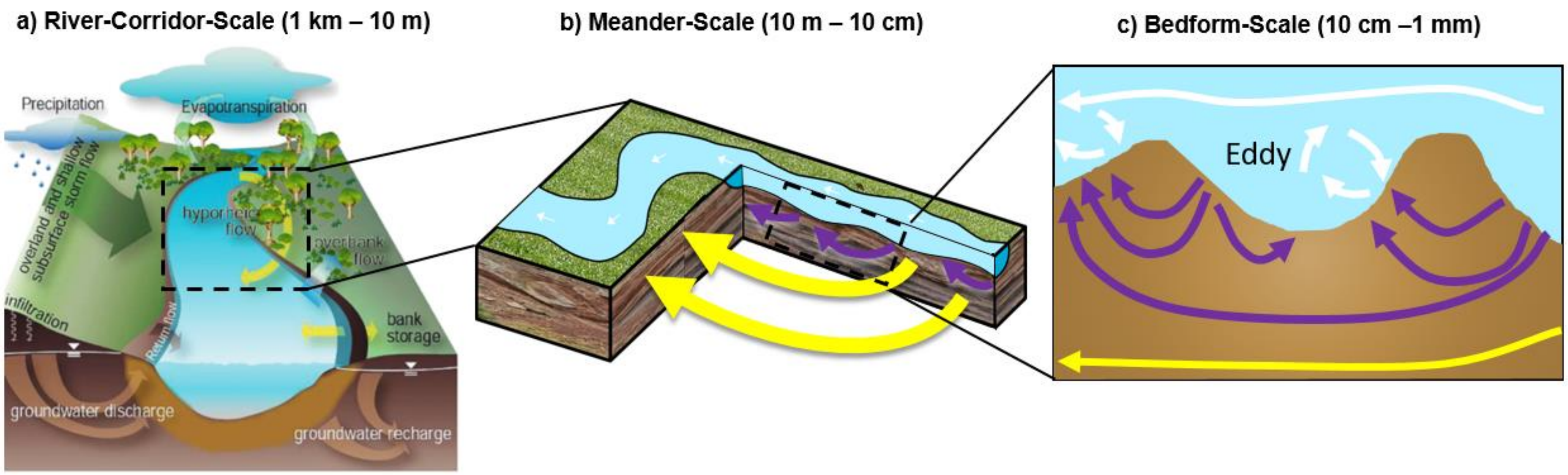
Therefore, the following subsections of Chapter 1 will describe in detail the definitions and categories of uncertainty when modelling sw-gw exchange (Section 1.1), followed by an assessment of knowledge gaps on sw-gw modelling at different scales (Sections 1.2 to 1.4). Finally, a description of which uncertainties being tackled at which scale will be provided, leading to a clear scope and selection of motivating research questions for this dissertation in Section 1.5. The resulting structure of this dissertation will be clarified in Section 1.6.

1.1 Uncertainty in surface water-groundwater modelling

Uncertainty is unavoidable in all efforts of environmental modelling: any model is by definition a simplified and therefore imperfect representation of a system, and this simplified representation is based on measurements (in applied modelling) which can never be error-free. Models used in industry are typically used in a deterministic way without a rigorous quantification of error/uncertainty. It is not necessary (and indeed not possible) to remove the uncertainty from the models in order to use them to design earthwork engineering projects, support decision making, or develop and test hypotheses about natural system processes, as long as the uncertainty is quantified when dealing with model forecasts.

According to Pappenberger and Beven (2006), in order for models to be useful and credible, it is important to have quantifiable measures of uncertainty that are easy to understand for stakeholders who may not have a background in modelling. In cases where the uncertainty of a model prediction is too large for the prediction to be useful for a given purpose, the uncertainty needs to be reduced. This in turn requires knowing at which stage of the modelling process the uncertainty has been introduced and how each individual uncertainty source contributes to the overall forecast uncertainty. Anderson et al. (2015) adds that the modeller is obligated to provide the best forecast possible, report uncertainty bounds, and communicate those limits in a clear and logical way to stakeholders.

By quantifying the contribution of different uncertainty sources, decisions can be made where to focus efforts in reducing the uncertainty. My research focuses on identifying and quantifying different sources of uncertainty associated with the predictions from sw-gw models at different scales in order to gain insights into the relative importance of different uncertainty sources that can be applied to other hydrological modelling problems.



a) River-Corridor-Scale (1 km – 10 m)

b) Meander-Scale (10 m – 10 cm)

c) Bedform-Scale (10 cm –1 mm)

Figure 1.2: Conceptual drawing of nested hierarchical hyporheic exchange circulation cells at different scales: a) River-corridor-scale (1 km -10 m)[adapted from National Research Council 2002]: depicting meander-scale hyporheic exchange nested within regional groundwater discharge and recharge b) Meander-scale (10 m - 10 cm): depicting bedform-scale hyporheic exchange nested within meander-scale hyporheic exchange, c) Bedform-scale (10 cm - 1 mm): depicting the turbulent flow regimes following bedforms that can lead to bedform-scale hyporheic exchange nested within meander-scale hyporheic exchange (Chow et al. 2018).

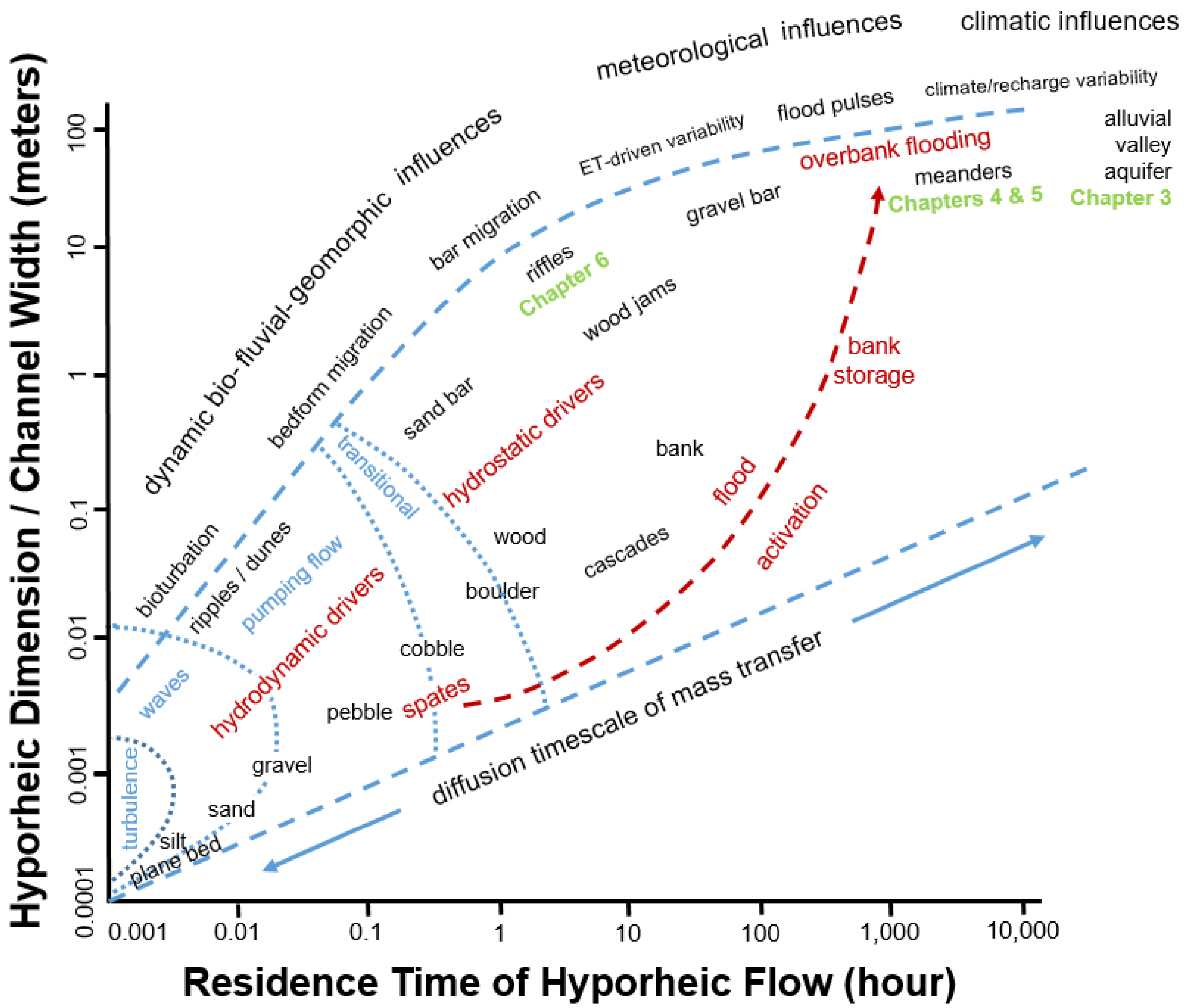


Figure 1.3: Spatial and temporal scaling of hyporheic flow. Important regimes of hyporheic flow are denoted in red with minor regimes in blue and typical channel features influencing hyporheic exchange are shown in black. Chapters addressing hydrological features at a specific scale in green. Approximate spatial and temporal limits of hyporheic flow are illustrated with dashed lines, with external biological, fluvial, and geomorphological influences shown in grey. Hyporheic flow dimension (depth or length) is divided by channel width to account for scaling of size of channel features with river size (adapted from Boano et al. 2014).

1.1.1 States of knowledge (or uncertainty)

The former U.S. Secretary of Defense, Donald Rumsfeld, defined uncertainty as the state of current knowledge, saying that “there are known knowns, there are things that we know we know. We also know there are known unknowns, that is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don’t know we don’t know.” Atherton (2013) organized these states of knowledge in a graphical matrix that I have adapted for sw-gw modelling (Figure 1.4).

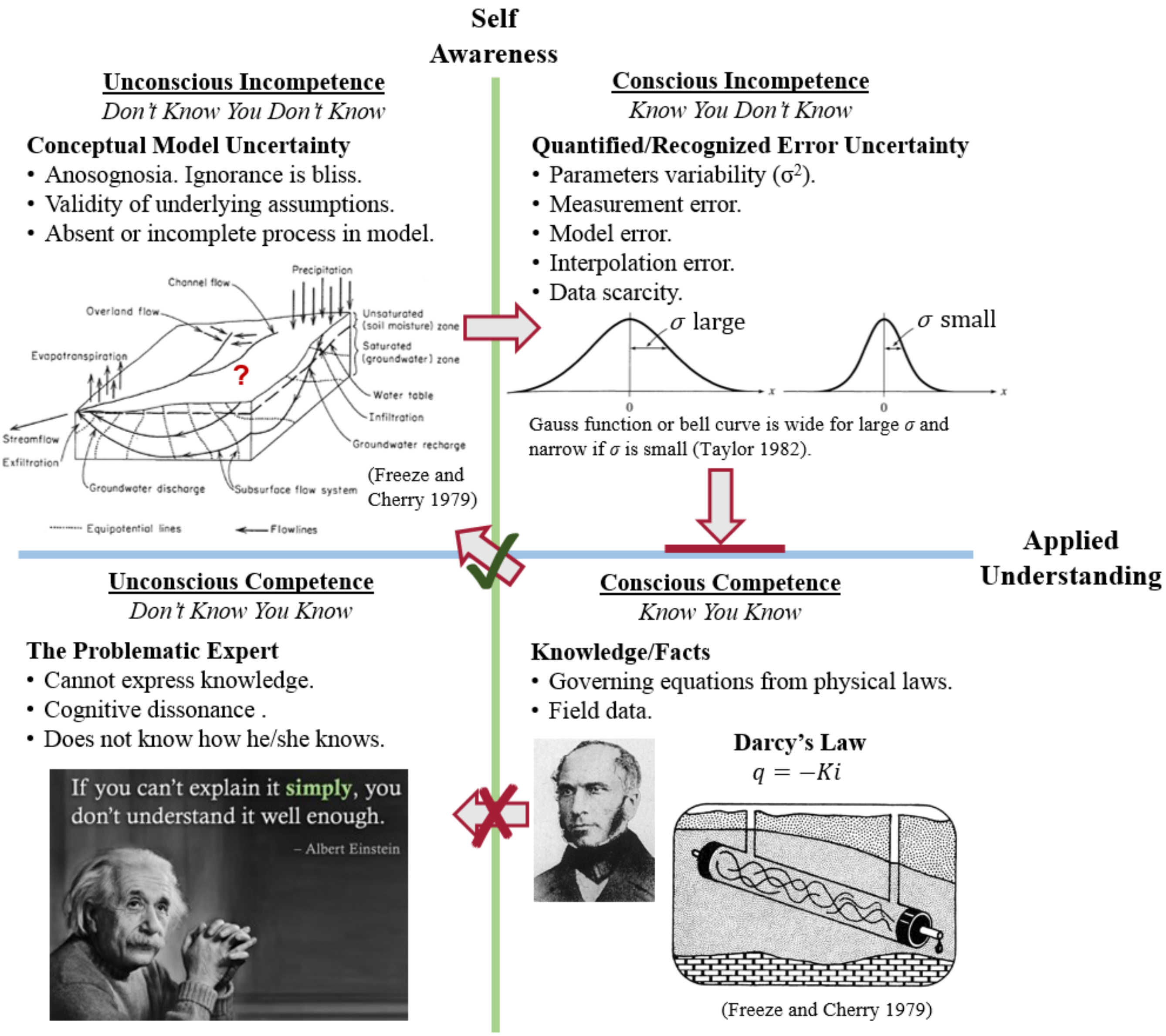


Figure 1.4: Matrix of knowledge states. Four distinct quadrants with varying degrees of self-awareness and applied understanding. The green line separates our unconscious and conscious self-awareness, while the blue line separates our applied understanding. I define applied understanding as a level of mastery where the knowledge can be applied in a straightforward manner to solve engineering-type problems. The arrows indicate the states of knowledge cycle (Adapted from Atherton 2013).

There are four distinct quadrants within the matrix of knowledge states (Figure 1.4):

Unconscious Incompetence (don’t know you don’t know): We are not aware of what we do not know and are therefore unconscious with regards to our incompetence. The “unknown unknowns” are the most elusive by nature and in my opinion the most dangerous type of unknowns because we are unaware of their existence and therefore cannot grasp the consequences of their ignorance. For example, there may be missing or incomplete processes in our sw-gw model that could be important, but we proceed ignorantly.

Conscious Incompetence (know you don’t know): – We are aware that we do not have full knowledge of some particular aspect of the system under investigation, therefore we are consciously incompetent. Stochastic methods play an important role in this knowledge state. For example, unknown parameter and/or boundary conditions of a sw-gw model can be explored by treating them as random variables. With enough realizations and efficient stochastic methods, one could deduce useful insights even if there are a number of variables with uncertain values. In other words, we can quantify (to some extent) the model uncertainties that we are aware of by using stochastic methods.

Conscious Competence (know you know): We have learned through careful experimentation and hypothesis testing fundamental physical laws of nature (e.g., Darcy’s Law in its valid range). These could be useful correlations and/or cause-and-effect relationships. We have reached a level of mastery or conscious competence and can apply this knowledge in useful ways, e.g., engineered designs. Furthermore, this knowledge can be shared through didactic methods.

Unconscious Competence (don’t know you know): We have learned something so well, we have forgotten how we learned it in the first place and are thus operating on auto-pilot (i.e., without deliberate thought). In this case we are unconsciously competent. In other words, we may be so familiar and fluent with something that we now have difficulties explaining it to someone else in a clear and concise way (e.g., driving a car or language use). In the context of sw-gw modelling, one may lack the ability to communicate overly abstract uncertainty quantification concepts to stakeholders.

The first step in the process of understanding sw-gw interactions is to become aware of our ignorance (Figure 1.4: transition from top-left to top-right quadrant). This means going from a state of unconscious incompetence to conscious incompetence. Next, we try to advance our knowledge by going from a state of conscious incompetence to conscious competence. This is often a difficult step (Figure 1.4: barrier between top-right and bottom-right quadrant) that usually requires novel experimentation and can sometimes take years to decades of research by numerous scientists. However, this step can be assisted by prioritizing recognized uncertainties in the upper right quadrant, which is the primary goal of my dissertation.

Finally, once we have gained sufficient understanding, we may end up taking this knowledge for granted, forgetting how we learned it, thus rendering the knowledge difficult to share (Figure 1.4: transition from bottom-right to bottom-left quadrant should be prevented). To avoid transitioning to a state of unconscious competence it is important to properly document our current knowledge, as well as the methods used to obtain it (including our failed attempts, Feynman 1966). Instead, we should seek to use our newfound knowledge to fuel our curiosity to uncover more of natures' "unknown unknowns" (Figure 1.4: transition from bottom-right to top-left quadrant), thus repeating the states of knowledge cycle.

It is important to note that there is a type of observational bias that occurs when focusing on "known unknowns", which is common in stochastic analysis of systems. Watzlawick (1993) popularized the "drunkard's search principle" or the "streetlight effect", which states that there is a tendency to bias one's search (i.e., research resources) where it is easiest to look. This principle originates from an old joke:

> A policeman sees a drunk man searching for something under a streetlight and asks what the drunk has lost. He says he lost his keys and they both look under the streetlight together. After a few minutes the policeman asks if he is sure he lost them here, and the drunk replies, no, and that he lost them in the park. The policeman asks why he is searching here, and the drunk replies, "this is where the light is".

An example of investigating a "known unknown" under the "streetlight" in sw-gw interaction science is research in subsurface heterogeneity for physically-based spatially distributed models, particularly hydraulic conductivity (Cardenas et al. 2004, Salehin et al. 2004, Zhou et al. 2014, Pryshlak et al. 2015, Liu and Chui 2018, Stonedahl et al. 2018). Since, hydraulic conductivity is known to vary orders of magnitude (Freeze and Cherry 1979), the conceptual representation of its heterogeneous distribution in the subsurface continues to be researched, when perhaps scientific resources could be better spent elsewhere. When scientific resources are inefficiently spent on studying similar problems that are obviously uncertain (i.e., drunkard's search), there are often only marginal gains in our understanding of natural system processes. This can lead to stagnation within scientific research, or as Schwartz (2013) calls it “Zombie-Science”. Furthermore, the cost of opportunity losses in research is unknowable.

Schwartz (2013) claims that problem selection is the key to protect against zombie-science. Therefore, I have chosen to investigate different model uncertainty types at different scales of sw-gw exchange modelling, with the aim of achieving novel insights. Although, I dedicate Chapter 5 to investigating subsurface heterogeneity for a sw-gw exchange model, I intend to produce novel insights regarding the potential trade-offs one could encounter when choosing one subsurface structure over another. The model uncertainties I will describe in the following section are the “known unknowns”. However, I will attempt to investigate uncertainties around (not under) the “streetlight” in my dissertation.

1.1.2 Defining types of model uncertainty

A number of different taxonomies have been suggested for classifying types of uncertainty, with sometimes conflicting uses of certain terms. Three broad categories of environmental model uncertainty have been suggested by Gupta et al. (2005) and a fourth by Jones and Mendoza (2012). The four uncertainties are:

1. Structural (Conceptual) Uncertainty
2. Data Uncertainty
3. Parameter Uncertainty
4. Predictive (Model Niche) Uncertainty

Structural (or conceptual) uncertainty has different definitions for hydrologists and hydrogeologists. In the context of lumped conceptual models that are typically used by hydrologists, the definition of this term refers to the types of model components (such as water storage and transfer functions) and connections between these components. In the context of distributed groundwater and surface water models that numerically solve physically-based partial differential equations (PDE), which are typically used by hydrogeologists, the usage is less clear. It can be used to capture the effects of uncertainty about the geometry or limits of the domain and the location of aquitards and facies boundaries, in addition to uncertainties in the mathematical formulation (e.g., missing physical processes, incomplete and/or inaccurate mathematical formulation). Furthermore, multiple structural (conceptual) models can represent processes and hydro(geo)logical responses equally well, this is referred to as equifinality (Anderson et al. 2015).

According to Gupta et al. (2012) model structure encompasses the:

- Conceptual model: the physical and process structures
- Mathematical model: the symbolic representation of the processes in the form of equations and the spatial variability structure
- Computational structure: the discretization and solution of the mathematical formulation. Technically, computational structural choices introduce numerical errors, which are known and differ from model uncertainty.

Data uncertainty refers to observations that are used to build, force, and calibrate the model. An example of data uncertainty are errors in measuring hydraulic heads used for groundwater model calibration or errors in measuring and interpolating precipitation used to estimate recharge for hydro(geo)logical systems.

Parameter uncertainty refers to the situation that there is typically not a single best set of parameter values that can be assigned to a model (Carrera and Neuman 1986), but rather the calibration process finds distinct and different parameter value combinations or parameter value ranges that match the observed data equally well. This is typically referred to as nonuniqueness, where multiple parameter combinations can provide similar fits to the calibration target (Anderson et al. 2015). Furthermore, a model parameter that is both sensitive and uncorrelated is more likely to be estimated than an insensitive and/or correlated parameter, this is known as parameter identifiability. In other words, parameter identifiability reflects the ease with which particular parameter values might be calibrated (Beven 2009). Physically-based hydrogeological models typically have more unknown parameter values than field measurements. Therefore, the inverse problem is underdetermined with respect to the observations (i.e., no. of calibration parameters [unknowns] > no. of observation [knowns]) and is therefore mathematically ill-posed. Parameter nonuniqueness is often reducible by including more and varied calibration data. Different time periods and data types can contain (non-redundant) information related to different aspects of the modelled system (Appendix A).

Predictive (or model niche) uncertainty is described by Jones and Mendoza (2012) as the uncertainty arising from the application of a model outside the system for which it was originally developed and calibrated. Silver (2012) describes model niche uncertainty as the problem of using the past as a guide while recognizing the future may be different. In other words, predictive uncertainty is when a model is used “out of sample”, which means the model is making forecasts with data that only partially matches the prediction trying to be made. This is the case if the calibration data does not cover the entire range of forcings to be accounted for in the prediction, which is typical when a model is being used for forecasting. This is because predictive modelling often requires the model to produce output outside of time frames and stresses in which it was originally developed (e.g., a new well or planned expansion of an open pit mine). Niels Bohr succinctly epitomized predictive model uncertainty saying that “prediction is very difficult, especially about the future.” In light of model predictive uncertainty, Beven and Young (2013) recommends applied modellers to use the term fit-for-purpose (rather than calibrated) to describe a model that has been conditionally optimized for a specific application.

In practice, a calibrated model can only be considered conditionally optimal (Beven 2009) because it depends on data (and their errors) and heuristic criteria for judging optimality that is selected by the modeller (Anderson et al. 2015). Instead, a modeller should acknowledge the aforementioned uncertainties and consider a “family” or ensemble of possible reasonable models because physically-based hydro(geo)logical models are inherently equifinal (i.e., are conceptually uncertain) and nonunique (i.e., have parameter uncertainty). Model evaluation should then be thought of as a form of hypothesis testing, rather than seeking the “best/optimal” model, which leads to the concept of multiple working hypotheses (Beven 2009, Ferré 2017). A skilled modeller should be able to discern dead-end models quickly and focus on a reasonable subset of models (Anderson et al. 2015). Models outside the reasonable realm can be discarded objectively (or assigned a low model weight in a Bayesian framework). I define an unreasonable model as one that fails to achieve a satisfactory match to historical data (i.e., hindcast), does not use defensible parameter values (e.g., within the measured range or ranges reported in the literature for similar properties), and/or is inconsistent with the conceptual model.

1.1.3 Uncertainty types in the modelling process

Another way to understand the different uncertainty types is to label them as they appear in the modelling process (Figure 1.5). The most important step in modelling starts with defining the purpose (Predictive Uncertainty) and building the conceptual model (Structural Uncertainty). In order to build a conceptual model, one needs to study the existing data at a site (Data Uncertainty). Afterwards the appropriate numerical/mathematical model is chosen to represent the conceptual model (Structural Uncertainty). The model choice generally constrains the number of physical processes represented, the discretization scheme, and the representation of boundary conditions. Once the numerical model is built, the model parameters can be calibrated to existing data (Parameter and Data Uncertainty[†]). The model is then used to make predictions outside of the calibrated conditions and thus will be introduced to boundary conditions representing future scenarios (Predictive Uncertainty). Finally, a model can be post audited. This is a form of model validation, where a model is re-visited to judge the accuracy of its predictions (Anderson and Woessner 1992, Konikow and Bredehoeft 1992). Sections 1.5 and 1.6 will describe the uncertainty types addressed by my motivating research questions and the chapters in which they reside, respectively.

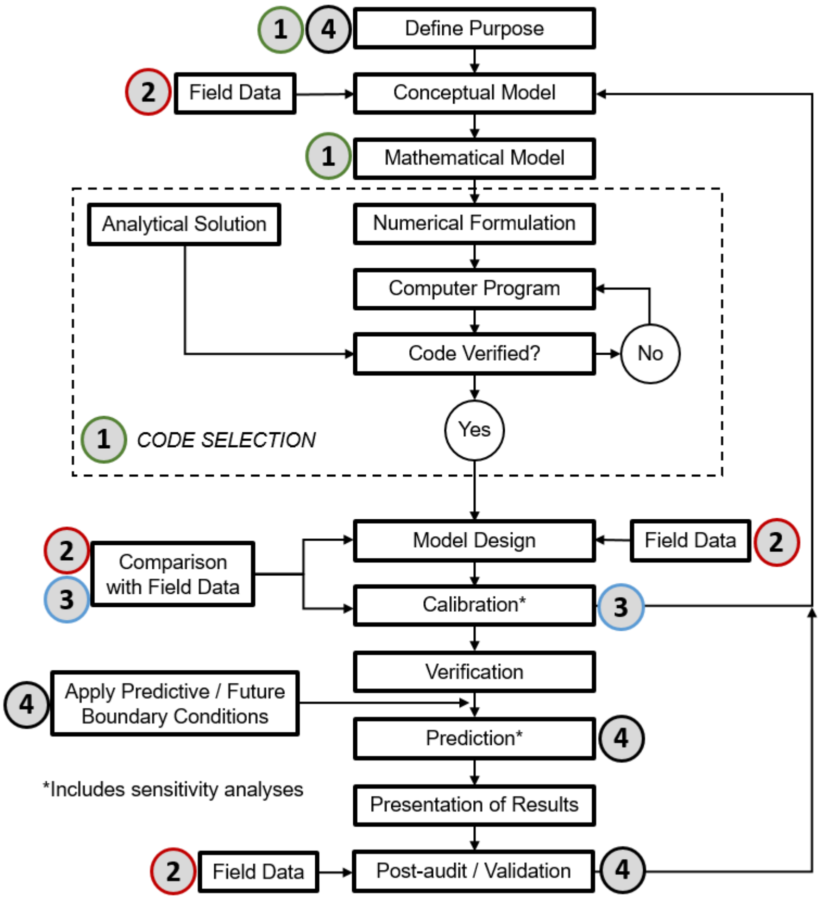


Figure 1.5: Different uncertainty types in hydrogeological model development (adapted from Anderson and Woessner 1992).

Uncertainty types mapped on modelling process:

① *Structural (Conceptual) Uncertainty*

② *Data Uncertainty*

③ *Parameter Uncertainty*

④ *Predictive (Niche) Uncertainty*

1.1.4 Sensitivity analysis

Sensitivity is a measure of how changes in model input values or assumptions impact outputs (Morgan and Henrion 1990). A sensitivity analysis attempts to uncover model input (i.e., parameters, boundary conditions) that may have a significant impact on model output. Depending on the model's purpose, the sensitive input should also significantly impact the

---

[†] A study by Motavita et al. (2019), which I co-authored, investigates the choice of data time series length and hydrological condition on parameter calibration and validation uncertainty using a conceptual rainfall-runoff model (Appendix A).

model output of concern. In other words, a model input could be sensitive, but if it impacts model output that is irrelevant for the design/study, then it is unnecessary to focus resources in constraining that model input.

Therefore, the model input of greatest concern is the one that has little data to constrain it and has a significant impact on the model output of interest. Sensitivity analysis can highlight the existing knowledge gaps once sensitive inputs are identified. It can also help guide the use of future resources in order to constrain model inputs and parameters (i.e., modelling additional scenarios, the design of field investigations).

Sensitive inputs are also compared to how they change the model calibration. A good calibration means that the model honours the observed data, which is interpreted as the model's ability to capture the behaviour of the natural system. Table 1.1 categorizes four different types of sensitivity based on how they change the quality of model calibration (significant vs. insignificant) and how they change the model output (significant vs. insignificant).

Table 1.1: Four sensitivity types (adapted from Wels et al. 2012)

Change of Parameter/ Boundary Condition		Change in Calibration	
		Insignificant	Significant
Change in Model Forecast / Output of Concern	Insignificant	Type I	Type II
	Significant	**Type IV**	Type III

- **Type I**: Input values are varied within a reasonable range but nothing significant happens as a result. This input type does not need further data collection or monitoring.
- **Type II**: There is a significant effect on the quality of model calibration, but an insignificant effect on model output of concern. In other words, changes to the model input or parameters that significantly affect the calibration may not significantly affect the model forecast or output of concern. The input type also does not need further data collection or monitoring. For example, one could calibrate a groundwater model to a flux measurement by changing hydraulic conductivities within a specific hydrogeological unit, but this may not necessarily produce a large change in hydraulic heads (i.e., output of concern) within a differing hydrogeological unit and/or location in the same model domain.

- **Type III**: Has a significant effect on both model calibration and forecast. Having a significant effect on model calibration means that this particular model input or parameter has data that can potentially constrain or identify its value (i.e., parameter identifiability).
- **Type IV** (grey): Has insignificant effect on calibration, but significant effect on model forecasts. This is a cause for concern because nonuniqueness in a model input might allow a range of valid calibrations, but the choice of value significantly impacts the model forecast or output of concern. It is important to determine the value of this input and not rely on model calibration to estimate it. It should be measured with good data, and ideally the data should represent the same stresses as in the predictive model simulations.

1.1.5 Uncertainty analysis

Uncertainty analysis is the quantification of imperfect knowledge regarding model input and how it impacts model output (Jones and Mendoza 2012). Uncertainty analysis and sensitivity analysis are often misinterpreted as the same thing. Sensitivity analysis evaluates the relative impact of a specific input and to model output, and does not consider probabilities. Whereas, uncertainty analysis is a more rigorous attempt to quantify the probability distribution of model output.

Figure 1.6 shows that model uncertainty (grey shaded area) lies at the intersection of three model components: structure (concept), parameter, and data uncertainty. First, by building a conceptual model (green circle) we create a mental picture of how our system behaves. The conceptual model simplifies the system sufficiently to make the solution tractable, while retaining enough of the essential and controlling processes to make the results useful and valid (Frind 2003). From the conceptual model we select our mathematical and numerical model accordingly, thus defining the model structure. Based on our conceptual model we then select boundary conditions and parameters that best represent our system, thus defining our parameters (blue circle). Finally, we compare our model results to observed measurements (red circle) and fine tune our parameter values (blue circle). At the intersection of these three constraints, there may still be large degrees of freedom, thus demonstrating nonuniqueness or uncertainty in the model output.

Hydrologic modelling in practice often takes a deterministic approach, using the “best-fit case” model (a single point in the model space) and perhaps a few scenarios to explore the “upper bounds” (a few points on the outer envelope of the solid grey shaded intersection area). Stochastic modelling processes have been developed to explore this model space in an efficient way (grey area), as well as exploring areas outside of the intersection (grey hatched area), in an attempt to quantify the probability of highly improbable events with severe consequences (red point).

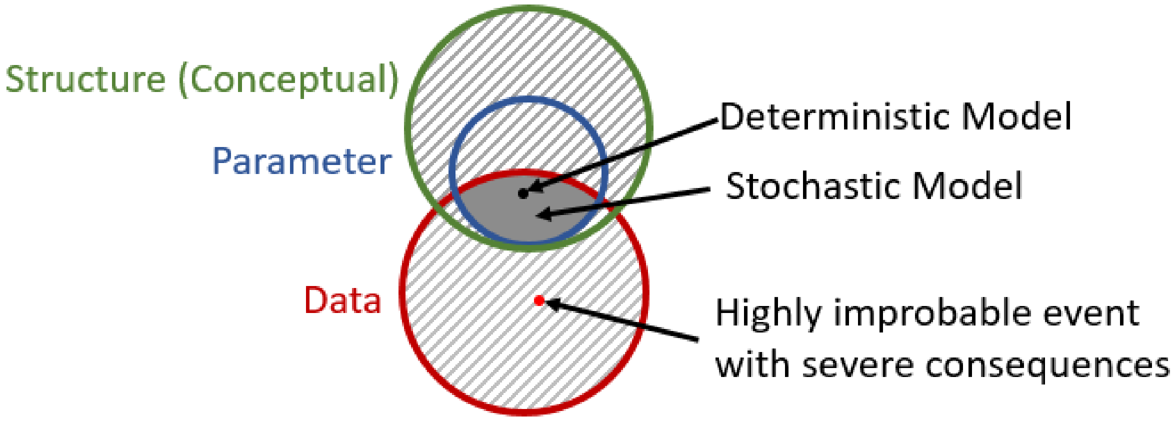


Figure 1.6: Venn diagram of model uncertainties. The grey shaded union of the different sources of model uncertainty (Green: Structural Uncertainty, Red: Data Uncertainty, Blue: Parameter Uncertainty) is the model space we explore in an uncertainty analysis. Stochastic modelling methods can provide efficient ways of exploring the plausible model configurations. A single point within the intersection indicates one realization of the plausible model configurations providing one deterministic model output/forecast. A robust uncertainty analysis (i.e., Bayesian uncertainty quantification) should still include highly improbable events (grey hatched areas) with severe consequences (red dot) even if they have a low likelihood of occurring. This Venn diagram is my attempt to visualize and summarize numerous ideas from the environmental model uncertainty literature.

The upcoming Sections 1.2, 1.3, and 1.4 will describe the sw-gw modelling objectives at different scales and the current research gaps, with a focus on different types of model uncertainties.

Chapters 4 and 5 will evaluate the roles of river bathymetry and subsurface distribution of hydraulic conductivity, respectively, when modelling meander-scale hyporheic exchange. It is far easier to accurately measure channel bathymetry than to accurately measure the distribution of subsurface properties. Therefore, the feasible parameter space for the distribution of subsurface properties is much larger than bathymetry. For this reason only a sensitivity analysis (Chapter 4) will be conducted to identify the aspects of river bathymetry that have the most significant impact on simulated meander-scale hyporheic exchange. Whereas, an uncertainty analysis will be conducted for the distribution of subsurface properties (Chapter 5) to determine the appropriate level of detail necessary to simulate meander-scale hyporheic exchange.

1.2 Catchment-scale surface water-groundwater interactions: stream baseflow capture zone delineation

At the catchment-scale, groundwater discharge to surface water bodies can be important in maintaining steady water levels and baseflows, which keep water temperatures and (bio)chemical constituent concentrations amenable to support healthy aquatic ecosystems (e.g., cold-water fisheries). Water that becomes groundwater discharge originates as precipitation (groundwater recharge) falling on the land surface within the catchment. It is then stored within the aquifer and slowly released into surface water bodies (e.g., lakes, wetlands, rivers, streams) as baseflow (Freeze and Cherry 1979). The dynamics of groundwater discharge (amount, discharge rate, quality) depends on the extent and characteristics of the groundwater recharge and storage areas. For a smaller near-surface water storage area, the transition from precipitation to discharge might take a few days, while for a large watershed it could take centuries or millennia (Winter et al. 1998).

Prevention of actual or potential threats to the quality and quantity of groundwater discharge (baseflow) is critical to ensuring the environmental sustainability of receiving surface water bodies. A major threat is land development for industrial, commercial, or residential purposes, which can increase the risk of groundwater contamination from sources such as chemical spills and road salting (Cleary and Cleary 1991, Howard and Israfilov 2012). Impervious surfaces such as roads, parking lots and roofs can also impact groundwater recharge by promoting storm runoff and reducing aquifer storage.

In order to manage these threats and find a balance between development and the protection of water resources, it is necessary to confidently identify the areas that contribute baseflow to surface water bodies. Appropriate protective measures can then be taken, and the potential economic cost can be assessed. The only practical approach to delineate these areas is by using models that can simulate sw-gw processes. The method to define groundwater recharge areas that potentially provide water to a groundwater sink (e.g., well or surface water body) is known as capture zone delineation (Cleary and Cleary 1991, Frind et al. 2006). The basic relationships for a typical multi-aquifer groundwater flow system is illustrated in Figure 1.7, which contains a hierarchy of scales from local to regional, where the transition from precipitation to discharge can range from days to millennia.

Presently, the industry-standard approach to capture zone delineation (CZD) is by particle tracking, which is usually based on the assumption of homogeneous hydrogeological properties

(AquaResource Inc. 2010). The assumption of homogeneous hydrogeological properties and a uniform flow field can produce thin pencil-like capture zones that can be easily packaged into an envelope under certain conditions (i.e., Grand River Source Protection Area 2015).

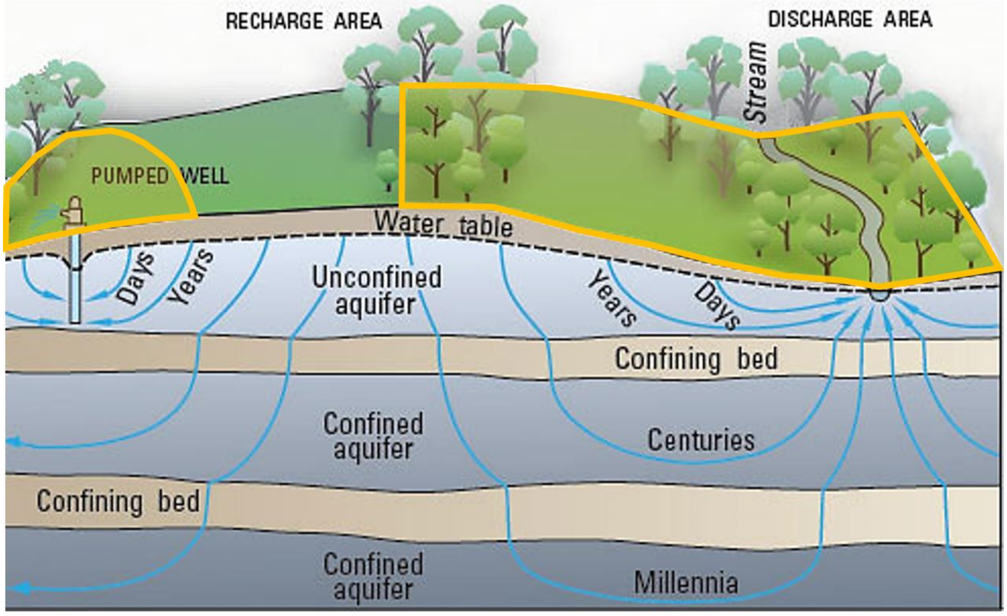


Figure 1.7: Groundwater flow system: well capture zone and stream capture zone are highlighted in orange (Chow et al. 2016).

In contrast, heterogeneous hydrogeological conditions tend to produce erratic (but more realistic) particle tracks that may be subjective and difficult to package into an envelope (Martin and Frind 2005). It is inadvisable to make land use decisions on the basis of such capture zones. Furthermore, once land uses are assigned in a municipality's planning process, it becomes difficult to change them. Therefore, there is an urgent need to update standard practices to account for hydrogeological parameter uncertainty, which challenges decision makers when it comes to interpreting capture zone results and designating land-use (Frind and Molson 2018).

Uncertainties in CZD can also arise from structural or conceptual model choices (Section 1.1), such as the choice between different commercially available hydrosystem modelling software packages, which can differ in their conceptual and numerical approach (e.g., for discretizing the model domain). These differences can introduce different simplifying assumptions that can ultimately impact the delineated capture zone. Thus, choosing between or combining competing capture zones produced by different models can introduce an additional challenge for stakeholders (Sousa et al. 2013).

Thus, I chose to mainly investigate these structural model choice uncertainties because early decisions in the modelling process have been shown to have the potential to cascade and accumulate throughout the modelling process (Beven et al. 2005), which can lead to imprecision and uncertainty in the final model results. Also, I chose to delineate stream baseflow contribution areas at the catchment-scale (10 km to 100 km) because capture zone delineation is typically used for long-term (i.e., years to decades) land-use planning, which is associated with intermediate and regional-scale groundwater flow systems occurring at the catchment-scale.

1.3 Meander-scale hyporheic exchange in fluvial rivers

Hyporheic exchange is driven by hydrostatic and hydrodynamic pressure variations due to variable bedform structures or larger features, such as meanders (e.g., Cardenas 2004). Hyporheic exchange is influenced by the river geomorphology (Kasahara and Wondzell 2003) and structural heterogeneity of the subsurface sediments within the hyporheic zone (e.g., Thibodeaux and Boyle 1987, Stonedahl et al. 2010, Osenbrück et al. 2013).

Numerous studies have evaluated the effects of subsurface heterogeneity on the hyporheic exchange residence times and flux rates (e.g., Cardenas et al. 2004, Salehin et al. 2004, Ward et al. 2013, Zhou et al. 2014, Pryshlak et al. 2015, Tonina et al. 2016, Liu and Chui 2018, Stonedahl et al. 2018). However, the level of detail necessary to simulate hyporheic exchange transit time distributions at the meander-scale and the expected trade-offs in predictive accuracy when choosing one subsurface parameter structure over another is still unclear.

Therefore, I chose to fill this knowledge gap by investigating hyporheic exchange in fluvial rivers at the hydrostatic-driven meander-scale (m to km) because the associated transit times (hours to days) are relevant for anaerobic microbial processes such as denitrification (Zarnetske et al. 2011, Boano et al. 2014). One additional model aspect identified as contributing significantly to the predictive uncertainty of hyporheic exchange is the accurate representation of river bathymetry (Cardenas et al., 2004). However, it remains unclear which aspects of bathymetry are the most critical with respect to the accurate simulation of meander-scale hyporheic exchange. Thus, I also chose to investigate the level of bathymetric detail one would need to simulate meander-scale hyporheic exchange in a physically meaningful way.

1.4 Bedform-scale hyporheic exchange in bedrock rivers

Bedrock rivers represent a unique hydrogeological environment, in which the surface water flows along an exposed bedrock surface, which may include weathered cobbles and sharp angular bedrock fragments. Tinkler and Wohl (1998) define bedrock rivers as reaches where "a substantial proportion of the boundary (≥50 %) is exposed bedrock, or is covered by an alluvial veneer which is largely mobilized during high flows such that underlying bedrock geometry strongly influences patterns of flow hydraulics and sediment movement."

Bedrock rivers are commonly formed by glacial meltwater and are often found in high-elevation headwater regions that are characterized as high-energy erosional environments (Tinkler and Wohl 1998). Bedrock rivers can also be found at lower elevations in regions with no net deposition of fluvial sediments, which are typically high-energy regions dominated by sediment transport where river reaches are stripped of their overburden (Kennedy 2017). Although bedrock rivers are frequently found throughout the world, the number of documented studies is limited (Figure 1.8). Most documented bedrock studies characterize their geomorphological evolution, focusing on channel erosion and transport. Thus, there are few studies that specifically address the nature of sw-gw exchange within bedrock rivers, particularly the mechanisms that could potentially control the presence of hyporheic exchange within bedrock rivers.

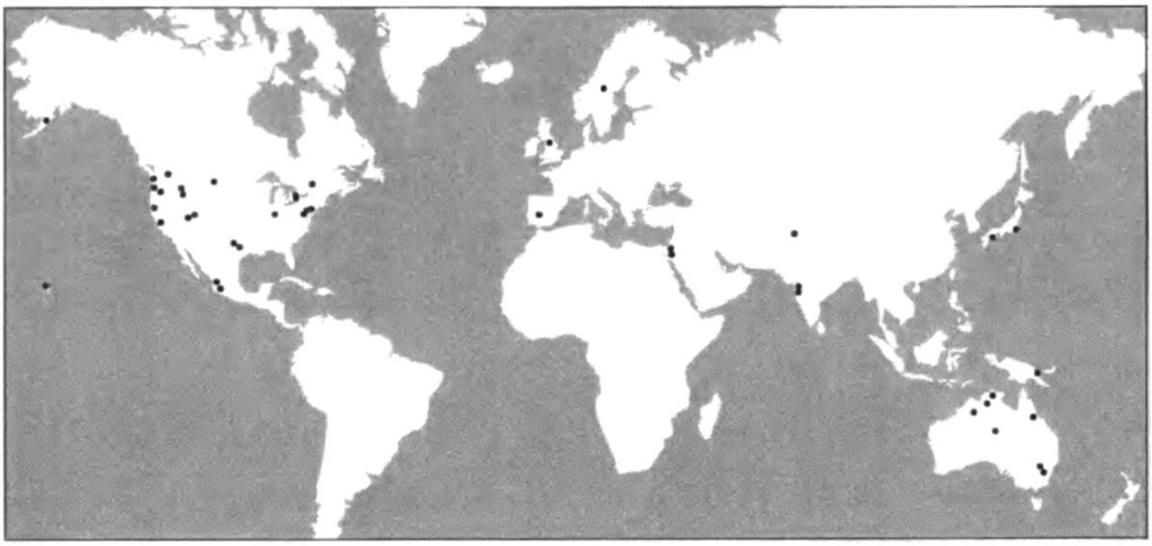


Figure 1.8: World distribution of bedrock river studies (Tinkler and Wohl 1998).

Studies of hyporheic exchange have typically involved rivers composed of unconsolidated fluvial sediments, which leaves a knowledge gap about the behaviour of hyporheic exchange in bedrock rivers (Figure 1.9). It has been shown in river engineering works that it is dangerous to assume that knowledge from alluvial and gravel bed rivers can be transferred directly to bedrock rivers (Vaughn 1990, Tinkler and Parish 1998). This is because the flow regimes of bedrock rivers (even at quite low stages) are associated with greater velocities and shear stresses when compared to fluvial rivers, and usually possess highly aerated and turbulent flow structures (Tinkler and Wohl 1998).

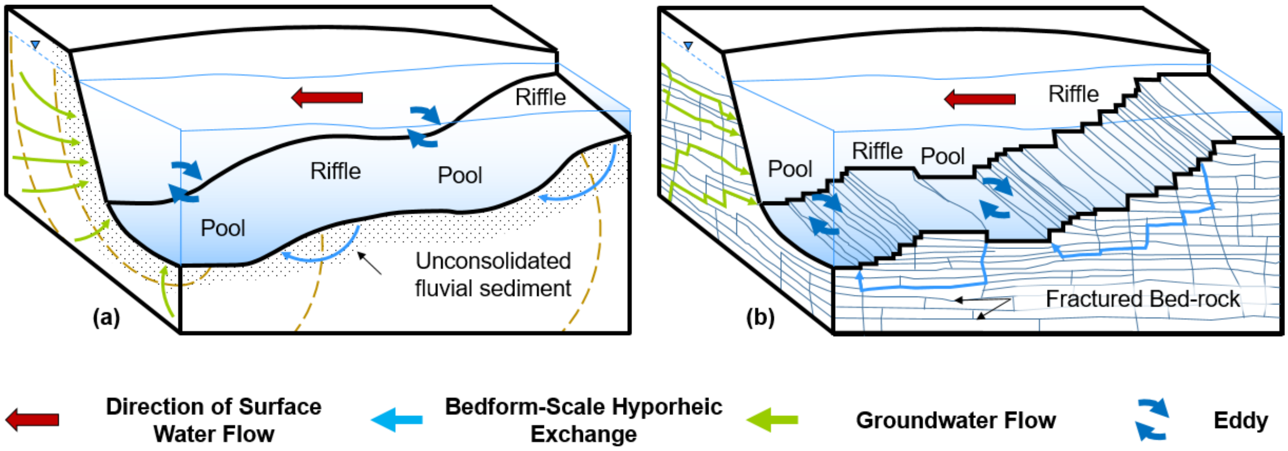


Figure 1.9: Alluvial vs. bedrock rivers (adapted from Kennedy 2017). Bedform-scale hyporheic exchange within (a) alluvial and (b) bedrock river settings. Within a riffle-pool-riffle sequence, turbulent river flow produces eddies on the leeward side of the riffle, causing hydrodynamic pressure variation.

Therefore, I chose to fill this knowledge gap by modelling sw-gw exchange in a bedrock river at the bedform-scale (cm to m) because it is in-line with the scale associated with the discrete fracture network approach, which explicitly represents fractures and matrix as distinct features for subsurface flow and transport (MacQuarrie and Mayer 2005). Furthermore, I was granted access to high-resolution geological and hydrogeological data from a bedrock river at the bedform-scale through a collaboration with the University of Guelph's G^{360} Institute for Groundwater Research (Chapter 6).

1.5 Research goals and motivating questions

Thus, I have derived four main research goals covering three different scales from the aforementioned knowledge gaps on modelling sw-gw exchange. Each research goal focuses on tackling a different type of model uncertainty (Table 1.2).

These four main research goals on modelling sw-gw exchange are:

1. Catchment-scale (Chapter 3): Identify model errors that come from numerical model choice (structural uncertainty) and to provide practical methods that account for parameter uncertainty for delineating areas that provide baseflow to receiving surface water bodies (predictive uncertainty).
2. Meander-scale (Chapter 4): Determine the role river bathymetry plays (data and parameter uncertainty) when simulating hyporheic exchange transit time and fluxes (predictive uncertainty). Also determine which aspects of bathymetry are critical (data uncertainty and sensitivity) so that simplifications of bathymetry do not affect simulated hyporheic exchange too dramatically.
3. Meander-scale (Chapter 5): Evaluate the degree of subsurface detail necessary (conceptual and parameter uncertainty) to accurately simulate hyporheic transit time distributions (predictive uncertainty). Additionally, I wish to determine the subsurface parameter structure that makes the best use of hydrogeological data (data and parameter uncertainty) by reducing the predictive uncertainty (sensitivity).
4. Bedform-scale (Chapter 6): Evaluate factors controlling hyporheic exchange along a fractured sedimentary bedrock river through the use of a discrete fracture network approach to represent the subsurface (data and parameter uncertainty) within a manually coupled numerical sw-gw model. The aim is to understand fracture network heterogeneity and its role in defining the lateral and vertical extents of hyporheic exchange flow paths and residence time distributions (predictive uncertainty) commonly associated with fracture-dominated bedrock river flow systems.

Table 1.2: Uncertainty types addressed modelling SW-GW at different scales

Scale and Site		Chp.	Uncertainty Types*			
			1	2	3	4
Catchment - Alder Creek Watershed		3	✓	✗	✓	✓
Meander - Steinlach River Test Site	Bathymetry	4	✗	✓	✓	✓
	Subsurface Parameterization	5	✓	✓	✓	✓
Bedform - Eramosa Bedrock River Site		6	✗	✓	✓	✓

*Uncertainty Types: (1) Structural, (2) Data, (3) Parameter, (4) Predictive

Four main corresponding research questions (RQ) can be derived from the aforementioned research goals (Figure 1.10):

1. When simulating sw-gw interactions at the catchment-scale, does the choice of numerical model code affect the results of delineating baseflow contribution areas for gaining stream reaches? If yes, then what are the most important criteria to consider when choosing between different numerical model codes? (Chapter 3)
2. When simulating hyporheic exchange at the meander-scale, what are the most important aspects of the river bathymetry that need to be captured in order to capture the scales of exchange appropriately? (Chapter 4)
3. What is the recommended degree of subsurface detail when choosing a subsurface parameterization structure to simulate meander-scale hyporheic exchange? What are the expected trade-offs when choosing between a simple vs. complex (detailed) subsurface parameter structure? (Chapter 5)
4. Can hyporheic exchange occur within bedrock rivers at the bedform-scale? If yes, then how is hyporheic exchange different in bedrock rivers compared to fluvial rivers? Furthermore, what are the main fracture properties that influence the hyporheic exchange residence times and spatial extents? (Chapter 6)

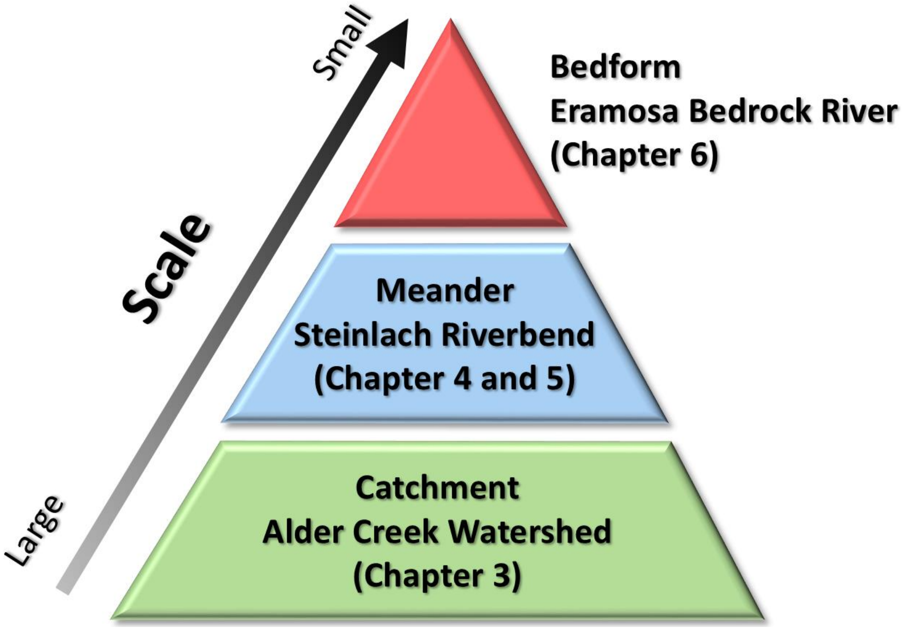


Figure 1.10: Bottom-up approach illustrating the different scales of surface water-groundwater exchange simulated in each Chapter. The colours indicate the predominant uncertainty addressed in each Chapter: Green – Structural Uncertainty, Blue – Parameter Uncertainty, Red – Data Uncertainty.

I will reintroduce Figure 1.10 at the beginning of each research chapter to indicate which scale and which main model uncertainty type is being treated. Each chapter will progressively decrease in system-scale, starting at the catchment-scale and finishing at the bedform-scale. In combination, the stated goals and research questions will address the major questions of uncertainty in modelling the physical process of sw-gw interactions, emphasizing different types and sources of uncertainties that are most relevant at the respective scales and in the respective system.

1.6 Approaches and dissertation structure

Chapter 2 describes the methods used to simulate sw-gw interactions shared between all subsequent chapters. This includes a description of HydroGeoSphere (Aquanty Inc. 2015) and WATRAC (Frind and Molson 2004). Additional chapter-specific methods can be found within the separate chapters.

Chapter 3 starts with catchment-scale sw-gw interactions, addressing the delineation of baseflow contribution areas to stream reaches, also known as stream capture zones (RQ1). This chapter focuses on model-related uncertainties due to model characteristics, discretization schemes, delineation methods, and particle tracking algorithms. To answer these uncertainties, a model and method intercomparison is conducted. The methodology is applied to the Alder Creek watershed in southwestern Ontario. Four different model codes are compared: HydroGeoSphere, WATFLOW, MODFLOW, and FEFLOW. Additionally, two delineation methods are compared: reverse particle tracking and reverse transport, where the latter considers local-scale parameter uncertainty by using a macrodispersion term to produce a capture probability plume. The results from this chapter indicate that different models can calibrate acceptably well to the same data and produce very similar distributions of hydraulic head, but can produce different capture zones. The stream capture zone is found to be highly sensitive to the particle tracking algorithm. It was also found that particle tracking by itself, if applied to complex systems such as the Alder Creek watershed, would require considerable subjective judgement in the delineation of stream capture zones. Reverse transport is an alternative and more reliable approach that provides probability intervals for the baseflow contribution areas, taking uncertainty into account. The two approaches can be used together to enhance the confidence in the final outcome.

Chapter 4 determines the aspects of river bathymetry that have the greatest influence on the predictive biases when simulating meander-scale hyporheic exchange (RQ2). A highly parameterized HydroGeoSphere model of the Steinlach River Test Site in southwest Germany was built as a reference. This model is then modified with simpler bathymetries, evaluating the changes to hyporheic exchange fluxes and transit time distributions. Results indicate that simulating hyporheic exchange with a high-resolution detailed bathymetry using a three-dimensional fully coupled model leads to nested multi-scale hyporheic exchange systems. A poorly resolved bathymetry will underestimate the small-scale hyporheic exchange, biasing the simulated hyporheic exchange towards larger scales, thus leading to overestimates of hyporheic exchange residence times. This can lead to gross biases in the estimation of a catchment's capacity to attenuate pollutants when extrapolated to account for all meanders along an entire river within a watershed. The detailed river slope alone is not enough to accurately simulate the locations and magnitudes of losing and gaining river reaches. Thus, local bedforms in terms of bathymetric highs and lows within the river are required. Bathymetry surveying campaigns can be more effective by prioritizing bathymetry measurements along the thalweg and gegenweg of a meandering channel. The gegenweg is defined as the line that connects the shallowest points in successive cross-sections along the depositional side of a river opposite the thalweg under average flow conditions. Incorporating local bedforms will likely capture the nested nature of hyporheic exchange, leading to more physically meaningful simulations of hyporheic exchange fluxes and transit times.

Chapter 5 evaluates the trade-offs between intrinsic (irreducible) and epistemic (reducible) model errors when choosing between homogeneous and highly complex subsurface parameter structures to simulate meander-scale hyporheic exchange (RQ3). Once more I use the HydroGeoSphere model of the Steinlach River Test Site as a reference to simulate hyporheic exchange and to assess the predictive errors and uncertainties of transit time distributions. Results indicate that if the parameter structure is too simple, it will be limited by intrinsic model errors. By increasing subsurface complexity through the addition of zones or heterogeneity, we can begin to exchange intrinsic for epistemic errors. Thus, choosing the appropriate level of detail to represent the subsurface parameter distributions depends on the acceptable range of intrinsic structural errors for the given modelling objectives and the available site data. Results indicate that a zonated model is capable of reproducing the transit time distributions of a more detailed model, but only if the geological structures are known. An interpolated heterogeneous parameter field (cf. pilot points) showed the best trade-offs between the two errors, indicating fitness for practical applications. Parameter fields generated by multiple-point geostatistics (MPS) produce transit time distributions with the largest uncertainties, however these are reducible by additional hydrogeological data, particularly flux measurements.

Chapter 6 evaluates the conditions that could potentially produce bedform-scale hyporheic exchange to occur within a bedrock river. Furthermore, I examine how hyporheic exchange may be different in bedrock rivers in comparison to fluvial rivers (RQ4). I focus on evaluating data uncertainty at the bedform-scale by modelling a conservative tracer experiment at the Eramosa Bedrock River Site near the City of Guelph, Canada. I use a stochastic discrete fracture network framework to represent the subsurface fractured bedrock connectivity, and in doing so produce a probabilistic distribution of the potential hyporheic exchange extents and residence times. The results indicate that the coincidence of fractures and hydraulic gradients determine the spatial extents of bedform-scale hyporheic exchange. Furthermore, hyporheic exchange residence times in bedrock rivers at the bedform-scale are potentially orders of magnitude longer when compared to fluvial rivers (i.e., months to years vs. minutes to hours).

2. Methods for simulating surface water-groundwater exchange

Modelling sw-gw exchange at relevant scales is important in deepening our understanding of this complex phenomenon. Coupled sw-gw modelling is capable of generating detailed nested hierarchies of hyporheic exchange circulation systems (Buffington and Tonina 2009) at the bedform and meander-scale (Stonedahl et al. 2010). However, at the catchment-scale it becomes difficult to include complex interactions between hyporheic exchange scales (Caruso et al. 2016). Typically, a manually coupled modelling approach is used to simulate hyporheic exchange (Cardenas and Wilson 2007b, Stonedahl et al. 2010, Trauth et al. 2013), where the streambed pressures are derived from a surface water model, which are then assigned as the top boundary condition for a groundwater model. The groundwater model then computes the fluxes at its top surface, representing losing and gaining conditions for the surface water model. Advances in computer technology have made physically based hydrological models that explicitly include all relevant hydrological processes in an integrated approach more tractable. Thus, whenever feasible, fully coupled sw-gw modelling is preferred over manual coupling in my dissertation. Section 2.1 will provide the governing equations for the core model I have used throughout my dissertation.

While the governing equations that are used in coupled sw-gw models are well established, lack of hydrogeological data continues to be a limiting factor in modelling sw-gw interactions (Bredehoeft 2010). For instance, it is rare to have many hydraulic conductivity measurements of the near-subsurface underlying a surface water channel and it is therefore difficult to constrain the model parameters for such a zone. Modelling sw-gw exchange with physically-based spatially distributed models is known to be sensitive to the spatial distribution of subsurface hydrogeological properties (Ward et al. 2013, Pryshlak et al. 2015, Liu and Chui 2018, Stonedahl et al. 2018), as well as surface flow domain properties (Jones et al. 2008, Frei and Fleckenstein 2013, Wang et al. 2018). Thus, it raises the question of whether our current methods for characterizing hydro(geo)logy are sufficient to quantify sw-gw exchange and whether uncertainty in sw-gw exchange can be reduced based on the conceptual hydro(geo)logical model chosen. I will investigate these questions in detail in Chapters 4 and 5 of my dissertation using a fully coupled sw-gw model.

2.1 HydroGeoSphere and governing equations

Freeze and Harlan (1969) were one of the first to propose a fully integrated physically-based hydrologic modelling approach for hydrological forecasting and to assess the adequacy of our hydrological knowledge. With advancements in computing over the last several decades, implementation of Freeze and Harlan's (1969) vision has resulted in several fully-integrated hydrological models, such as the Integrated Hydrology Model (InHM, Vanderwak and Loague 2001), OpenGeoSys (OGS, Kloditz et al. 2012), PARallel FLOW (PARFLOW, Maxwell et al. 2016), and HydroGeoSphere (HGS, Aquanty Inc. 2015). Results from model intercomparison benchmark tests by Maxwell et al. (2014) and Kollett et al. (2017) have shown good agreement and the same qualitative behaviour between different integrated sw-gw models (i.e., CATHY, HGS, OGS, PIHM, ParFlow, PAWS, tRIBS-VEGGIE), which lends confidence in the fundamental physical and numerical implementation of the governing equations in the different integrated sw-gw models tested within their studies. I chose to use HGS because of its flexible discretization scheme, fully coupled treatment of sw-gw exchange, capabilities of representing discrete fracture networks for flow and transport, and parallel computing capabilities.

HGS is a physically-based, 3D finite element numerical model designed to simulate the entire terrestrial portion of the hydrologic cycle (Aquanty Inc. 2015). HGS has the capability of representing numerous hydrological processes (e.g., interception, evapotranspiration, snowmelt, freeze and thaw of pore water), however in my research I only use HGSs' capabilities to fully couple sw-gw flow and to represent discrete fracture networks for flow and transport.

HGS fully couples three-dimensional variably saturated subsurface flow (modified Richards' equation; Richards 1931) with the two-dimensional, depth-integrated diffusive wave approximation of the Saint Venant equations for surface water flow (Viessman and Lewis 1996). HGS is well suited to delineate stream baseflow contribution areas at the catchment-scale (RQ1 in Chapter 3) because it can generate its own water courses without having to assign them as surface water boundary conditions a priori. In other words, the advantage of HGS is that it calculates the spatial distribution of gaining and losing stream reaches (i.e., the exchange flux) within the channel without the need to manually prescribe these areas as a boundary condition to the groundwater flow system beforehand. At the meander-scale, HGS can generate a physically consistent two-dimensional distribution of river stages and the associated distribution of exchange fluxes to simulate hydrostatic-driven hyporheic exchange (RQ2 and 3 in Chapters 4 and 5, respectively). Furthermore, HGS allows flexible unstructured discretization, which permits local refinement of the mesh in areas of interest (i.e., bathymetric details within the river channel, RQ2 in Chapter 4).

HGS uses the standard Galerkin finite element method (Huyakorn and Pinder 1983, Forsyth 1991) to solve Richards' equation with upwinding of the relative permeability and the advective-dispersive transport equation. HGS is capable of using a variety of discretization schemes to solve the 3D porous medium and dual continuum equations (Aquanty Inc. 2015). I used 3D triangular prisms (6-node elements) for my modelling of sw-gw exchange at the catchment and meander-scales (Chapters 3, 4, and 5) because of its flexibility in local refinement in areas of interest (i.e., rivers and wells). For my modelling of sw-gw exchange of a fractured bedrock river at the bedform-scale (Chapter 6), I did not require local refinement. Therefore, I chose to use the 3D finite difference approximation (Panday et al. 1993) for rectangular prisms (8-node elements) to represent the matrix and 2D fractures equations (Tang et al. 1981, Sudicky and McLaren 1992, Therrien and Sudicky 1996) for rectangular (4-node) planes to represent the fractures. HGS solves either linear equations (for saturated flow under combined conditions or solute transport) or non-linear equations (for variably-saturated subsurface flow, surface flow). To solve the non-linear equations, HGS uses the Newton-Raphson iteration scheme (Forsyth and Simpson 1991).

Several assumptions in HGS have been made for subsurface flow:

1. The fluid is incompressible.
2. The porous medium and fractures are non-deformable.
3. The system is under isothermal conditions.
4. The air phase is infinitely mobile.

The governing equation in HGS for subsurface flow is the modified form of Richards' equation used to describe 3D transient subsurface flow in a variably-saturated porous medium (Aquanty Inc. 2015):

$$-\nabla \cdot (w_m q) + \sum_{i=1}^{n} \Gamma_{ex,i} \pm Q = w_m \frac{\partial}{\partial t}(\theta_s S_w)$$

Where:

w_m is a dimensionless weighting factor between a 2nd (or more) interacting porous media (e.g., fractures, macropores);

q is the fluid flux (LT^{-1});

Γ_{ex} is the volumetric fluid exchange rate between n surrounding environmental and hydraulic features (e.g., overland/surface flow domains Γ_o, fractures Γ_f) expressed as a volume-specific source/sink strength ($L^3L^{-3}T^{-1}$);

Q is the volumetric fluid flux per unit volume representing a source (positive) or a sink (negative) to the system (e.g., wells, tile drains), also expressed as a volume-specific source/sink strength ($L^3L^{-3}T^{-1}$);

θ_s is the saturated water content (dimensionless), which is assumed to equal porosity;

S_w is the water saturation (dimensionless).

The product of θ_s and S_w is the volumetric water content, which equals the fraction of water-filled bulk volume.

The fluid flux q (LT^{-1}) is given by:

$$q = -\boldsymbol{K} \cdot k_r \nabla(\Psi + z)$$

Where:

$k_r = k_r(S_w)$ is the relative permeability of the medium (dimensionless), which is a function of the water saturation S_W (dimensionless);

Ψ is the pressure head (L);

z is the elevation head (L);

$\boldsymbol{K}$ is the hydraulic conductivity tensor (LT^{-1}) given by: $\boldsymbol{K} = \frac{\rho g}{\mu}\boldsymbol{k}$

Where:

ρ is the density of water (ML^{-3});

g is gravitational acceleration (LT^{-2});

μ is the dynamic viscosity of water ($ML^{-1}T^{-1}$)

$\boldsymbol{k}$ is the porous medium permeability tensor (L^2)

The governing equation in HGS for overland flow (e.g., surface water and runoff) is the diffusive wave approximation of the Saint Venant equations for unsteady shallow flow. It assumes depth-averaged flow velocities, hydrostatic pressure distribution vertically, mild slope, dominant bottom shear stress, and local and convective acceleration. Furthermore, it assumes that either the Manning, the Chezy, or the Darcy-Weisbach formulae (based on user-choice) are valid to calculate frictional resistance forces (Cirpka and Bürger 2012, Aquanty Inc. 2015):

$$\frac{\partial \phi_o h_o}{\partial t} - \frac{\partial}{\partial x}\left(d_o K_{ox}\frac{\partial h_o}{\partial x}\right) - \frac{\partial}{\partial y}\left(d_o K_{oy}\frac{\partial h_o}{\partial y}\right) + d_o \Gamma_o \ \pm Q_o = 0$$

Where:

ϕ_o is a surface flow domain porosity which is unity for flow over a flat plane, and varies between zero at the land surface and unity at the top of all rills and obstructions, for flow over an uneven surface (dimensionless);

d_o is the depth of flow (L);

z_o is the bathymetry or bed/land surface elevation (L);

$h_o = z_o + \ d_o$ is the surface water head or elevation (L);

K_{ox} and K_{oy} are surface conductances that depend on the equation used to approximate the friction slopes (LT^{-1});

Q_o is a volumetric flow rate per unit area representing external sources and sinks (LT^{-1}); and

t is the time (T).

Two different approaches are available for the surface water (overland) - groundwater (subsurface) flow coupling in HGS: (1) the common-node approach (Therrien and Sudicky 1996), and (2) the dual-node approach. The common-node approach assumes continuity of hydraulic head between the overland and subsurface flow domains, corresponding to an instantaneous equilibrium between them. The dual-node approach differs from the common-node approach in that coupling between the overland and subsurface flow domains is based on a first-order exchange coefficient using the concept of a coupling length (Park et al. 2009). The dual-node approach is more robust compared to the common-node approach because it does not strictly enforce pressure continuities between the overland and subsurface flow domains, which allows the dual-node approach to overcome certain numerical issues. Therefore, the dual-node approach is frequently used for applied HGS studies. The dual-node approach requires additional parameters to exchange flux between the overland and subsurface flow domains, i.e., hydraulic conductivities (K_{zz}, K_{dzz}), permeabilities (k_r, k_{dr}), and a coupling length (l_{ex}). Moreover, as the coupling length used in the dual-node approach becomes a small number (i.e., a few millimeters in physical hydrogeological terms), it is expected to produce results similar to the common-node approach. Nonetheless, the inability to directly estimate the parameters required for the dual-node approach has made the application of the concept problematic in certain situations of sw-gw exchange (Ebel et al. 2009). The governing equation for the dual-node approach in HGS is represented by the exchange term (Aquanty Inc. 2015):

$$d_o \Gamma_o = w_m \frac{k_r K_{zz}}{l_{ex}} (h - h_o) + w_d \frac{k_{dr} K_{dzz}}{l_{ex}} (h_d - h_o)$$

Where:

Γ_o represents flow from the subsurface system to the surface system when positive and vice versa when negative ($L^3L^{-3}T^{-1}$);

h and h_d are the subsurface porous medium and dual medium heads, respectively (L);

k_r and k_{dr} are the relative permeabilities for the porous medium and dual medium exchange fluxes, respectively (dimensionless);

w_m and w_d are the subsurface and dual-continuum volumetric fraction of the total porosity, respectively (dimensionless);

K_{zz} and K_{dzz} are the vertical saturated hydraulic conductivities of the underlying porous and dual media, respectively (LT^{-1});

l_{ex} is the coupling length (L).

Additionally, HGS is capable of representing discrete fractures (RQ4 in Chapter 6) that are represented as idealized space between two-dimensional parallel surfaces, with the assumption that the total head is uniform across the fracture width. HGS uses an extension of the saturated fracture flow equations (Berkowitz et al. 1988, Sudicky and McLaren, 1992) and an analogy of Richards' equation (1931) for a porous medium to derive a two-dimensional variably-saturated fractured flow equation with the form:

$$-\bar{\nabla} \cdot (w_f q_f) - w_f \Gamma_f = w_f \frac{\partial S_{wf}}{\partial t}$$

Where the fluid flux q_f (L/T) is given by:

$$q_f = -K_f \cdot k_{rf} \bar{\nabla} (\Psi_f + z_f)$$

Where:

$\bar{\nabla}$ is the two-dimensional gradient operator defined in the fracture plane;

w_f is fracture width (L);

k_{rf} is the relative permeability of the fracture (dimensionless);

Ψ_f and z_f are the pressure and the elevation heads within the fracture, respectively (L);

S_{wf} is the water saturation for the fracture (dimensionless);

The saturated hydraulic conductivity of a fracture K_f (LT^{-1}), having a uniform aperture w_f, is given by (Bear 1972):

$$K_f = \frac{\rho g w_f^2}{12\mu}$$

Where:

ρ is the density of water (ML^{-3});

g is gravitational acceleration (LT^{-2}); and

μ is the dynamic viscosity of water (ML^{-1}T^{-1})

2.2 WATRAC: Finite element particle tracking

Particle tracking is a post-processing method, following the solution of a groundwater flow model, whereby infinitesimally small water parcels (particles) are released and tracked within the advective flow field to compute flowpaths and the travel times (Anderson et al. 2015). Particles can either be released within the model domain or at its boundary and tracked forward in time through the subsurface, or released at the point of interest (well screen or groundwater discharge area) and allowed to travel backwards until they reach the surface or some other boundary. The two options are referred to as forward or reverse particle tracking. The advantage of particle tracking methods is that they do not introduce numerical dispersion, whereas Eulerian transport schemes, such as the finite element method implemented in HGS, introduce numerical dispersivities that scale linearly with the grid resolution (Daus et al. 1985, Frind 2003).

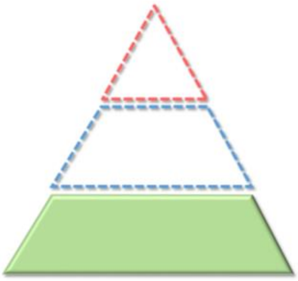
A well-known particle tracking routine today is MODPATH (Pollock 1989), which is available as a post-processing module in MODFLOW (McDonald and Harbaugh 1988). MODPATH uses a semi-analytic solution method by Pollock (1989) to calculate three-dimensional particle tracks from the steady-state flow solution generated by MODFLOW. This method requires the interfacial fluxes between cells and assumes that the velocity varies linearly within a cell in order to calculate the average velocity components. Given the entry point of a particle, the exit face is selected based on the shortest travel time between entry and exit points. After choosing the exit face for the particle, the exit position on the selected face is calculated. This method avoids interpolating velocities between cells, producing physically realistic particle tracks for heterogeneous conditions.

HGS does not have its own particle tracking capabilities. Therefore, I chose WATRAC (Frind and Molson 2004) as a post-processing particle tracker for HGS. WATRAC was developed for unstructured finite element grids and shares fundamental principles with the semi-analytical method by Pollock (1989). Alternatively, particle tracking in finite element models can be approached by using average elemental velocities from nodal heads (e.g., Yang et al. 2018). The problem arising from this approach is that elemental velocities are generally discontinuous across element boundaries. Higher-order interpolation schemes to interpolate velocities smoothly over several elements can be used (Pokrajac and Lazic 2002), but these approaches are inconsistent with the calculated continuous hydraulic head field and may not easily handle the change in flow direction at material interfaces (Frind and Molson 2004). These considerations are important, without them the particle tracks tend to smear through low hydraulic conductivity layers rather than deviating around them when encountering sharply contrasting hydraulic conductivities.

The approach in WATRAC (Frind and Molson 2004) introduces the principle of cell-by-cell velocity interpolation into the finite element method. Thus, combining the advantages of the finite element method in representing complex heterogeneous systems with the advantage of Pollock-type particle tracking (i.e., producing the correct change in flow direction at material interfaces). A drawback from this approach is that fluxes approximated by WATRAC are not locally mass conservative. This is due to the underlying principle of the finite element method, which only guarantees global mass conservation, in combination with an interpolation scheme that does not remedy this problem. Therefore, greater numerical accuracy could be achieved by developing a particle tracking method compatible with HGS (Selzer 2017, Selzer and Cirpka 2019) that uses a mass conservative approach for the determination of fluxes for unstructured finite elements (Odsæter et al. 2017). However, at the time of my dissertation, corresponding advancements in finite element particle tracking (Selzer and Cirpka 2019) was not yet available.

In order to use WATRAC for HGS steady-state hydraulic head distributions, it is necessary to convert HGS output into a format compatible with WATRAC. Afterwards WATRAC can be used for either forward or reverse particle tracking. When placing particles with WATRAC, the user specifies the x-y coordinates of the particle and the elemental layer of the starting position. The particle can be placed anywhere within a cell or element; however, the vertical placement of the particle is always in the center of the chosen layer.

3. Delineating baseflow contribution areas for streams – a model and methods comparison[*]



Chapter highlights

- Different models can calibrate acceptably well to the same data and produce very similar distributions of hydraulic head, but can produce different capture zones.
- The stream capture zone is highly sensitive to the particle tracking algorithm.
- Reverse transport is an alternative approach that provides probability intervals for the baseflow contribution areas, which accounts for local-scale parametric uncertainty.
- The two approaches can be combined to enhance the confidence in the final delineated capture zone.

3.1 Motivation and research questions

An environmentally sustainable stream depends on groundwater discharge for maintaining the steady baseflow and temperature needed to support a healthy aquatic ecosystem. Prevention of actual or potential threats (Cleary and Cleary 1991) to the quality and quantity of stream baseflow is critical to ensuring the environmental sustainability of streams. A major threat is

---

[*] The contents of this chapter are modified from ‘Chow, R., Frind, M.E., Frind, E.O, Jones, J.P., Sousa, M.R., Rudolph, D.L., Molson, J.W., and Nowak, W. (2016). Delineating baseflow contribution areas for streams – A model and methods comparison. Journal of Contaminant Hydrology 195, 11–22. http://dx.doi.org/10.1016/j.jconhyd.2016.11.001’

Author contribution statement: ***Chow, R.***: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript to address peer-review comments. Finalized and approved published article. Analyzed and evaluated all model results. Conceptualized modelling study. Preliminary results presented at the 2011 GeoHydro conference in Québec City (Chow et al. 2011) and in University of Waterloo Master’s thesis (Chow 2012). ***Frind, M.E.***: Developed FEFLOW model and analyzed FEFLOW results. ***Frind, E.O.***: Reviewed preliminary works and original draft manuscript. Helped with revisions to address peer-review comments. Conceptualized modelling study. Provided supervision. ***Jones, J.P.***: Developed preliminary HydroGeoSphere model. Reviewed preliminary works. Conceptualized modelling study. Provided supervision. ***Sousa, M.R.***: Developed FEFLOW model and provided technical modelling support. ***Rudolph, D.L.***: Reviewed preliminary works. ***Molson J.W.***: Provided technical modelling support. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

land development for industrial, commercial, or residential purposes (Howard and Israfilov 2012). Impervious surfaces such as roads, parking lots and roofs can impact groundwater recharge, promote storm runoff, and reduce aquifer storage (Hibbs and Sharp, 2012). Development can also introduce contaminants such as road salt (Bester et al., 2006) and increase the risk of chemical spills from point sources such as gas stations.

In order to manage these threats and find a balance between development and the protection of water resources, it is necessary to identify, with some confidence, the areas that contribute baseflow to sensitive streams or stream reaches. Appropriate protective measures can then be taken, and the potential economic cost can be assessed. The only practical approach to the delineation of these areas is through the use of simulation models.

Since an independent validation of model predictions is rarely possible, critical issues come from numerical model errors and parameter uncertainty (Section 1.1). One important source of numerical model errors when delineating capture zones is from selecting the appropriate model code. The research goals (Section 1.5) in this chapter are to identify numerical model errors that come from model choice and to provide practical methods that account for parameter uncertainty for delineating areas that provide baseflow to streams. The motivating research question I wish to address here is (RQ1, Section 1.5): does numerical model code choice affect the results of delineating baseflow contribution areas for gaining stream reaches? If yes, then what are the most important criteria to consider when choosing between different numerical model codes?

Firstly, Section 3.2 describes previous work that was completed on this topic. Sections 3.3 and 3.4 introduce the stream capture zone concept and describes the model codes considered in this comparison, respectively. Section 3.5 describes a proof-of-concept on the Alder Creek Watershed. Then Sections 3.6 and 3.7 describe two methods for capture zone delineation (CZD) and a discussion of the results, respectively. Other sources of uncertainty when delineating capture zones, as well as possible ways to control them, are briefly discussed in Section 3.8. Finally, Section 3.9 concludes the chapter.

This chapter shows that: (1) methodologies developed for well CZD can be applied to gaining stream reaches, (2) given the same model conceptualization, different groundwater models can produce different delineation results, and (3) results from different delineation methods can be combined to enhance the credibility of delineated capture zones.

3.2 Previous work

Preliminary results were first presented at the 2011 GeoHydro conference in Québec City, Canada. Chow et al. (2011) compared stream baseflow capture zones delineated using two different groundwater models, MODFLOW (McDonald and Harbaugh 1988) and WATFLOW (Molson et al. 2002) using their particle tracking codes MODPATH (Pollock 1988) and WATRAC (Frind and Molson 2004), respectively.

Three conclusions were drawn in Chow et al. (2011):

1. In principle, baseflow capture zones for reaches of gaining streams can be delineated using the same methodology used for well capture zones.
2. Capture zones are highly sensitive to numerical errors and different models can produce different capture zones.
3. Capture zones delineated from natural gradients are more sensitive to numerical errors than those for wells (i.e., larger anthropogenic gradients).

Following Chow et al. (2011), Chow (2012) added a third model, HydroGeoSphere (Aquanty 2015), and an alternative delineation method known as reverse transport (Uffink 1989, Wilson and Liu 1994, Fogg et al. 1999, Frind et al. 2006) to the comparison.

Finally, Chow et al. (2016) added a fourth model, FEFLOW (DHI-WASY 2015), to the comparison, overhauled the entire analysis in order to isolate for factors affecting CZD (i.e., disabled unsaturated zone processes in MODFLOW, WATFLOW, and FEFLOW models discussed in Section 3.5.3), and performed additional quantitative analyses to compare the delineated capture zones (Section 3.7).

3.3 The stream capture zone concept

In principle, the area that provides water to a stream encompasses the watershed or catchment area. However, it may not be practical to protect an entire watershed. The alternative is to identify the portion of the watershed that contributes baseflow for a specific reach of an environmentally sensitive stream. Veale et al. (2014) demonstrated this approach by means of a limited particle tracking analysis.

Conceptually, an area contributing to a stream reach should obey the same principles that govern a well capture zone. Figure 1.7 shows the similarities between well and stream capture zones. On this basis, much of the well-established methodology for delineating capture zones for drinking water wells should apply to streams. In the following sections, we will use the term “stream capture zone” to mean “baseflow contribution area for streams”.

There are several critical conceptual differences between delineating a stream capture zone as opposed to a well capture zone. Three fundamental differences are:

1. The Nature of the Gradients. Natural gradients near a stream will generally be much smaller than gradients induced by pumping at a well and therefore will be subject to greater relative numerical errors (Chow et al. 2011). In this study, the near-subsurface has been finely discretized vertically in order to reduce the numerical errors from natural gradients (Section 3.5.1).
2. The Nature of the Source/Sink Function. A well is a fixed-rate point source/sink, while a stream is a line source/sink variably distributed over the length of the stream. Where and how much groundwater a stream is gaining is generally not known. This study uses a fully integrated state-of-the-art sw-gw model to obtain the spatial distribution and rates of groundwater exchange at streams (Section 3.5.2).
3. The Significance of Transient Flow. Stream discharge is more variable in time than water pumped from a well. For a water supply well, the pumping rate is generally constant for longer periods. Conversely, flow for a stream is strongly influenced by precipitation events and seasonal conditions. The purpose of a capture zone is to designate an area where the planned land use will provide a certain measure of protection. This area cannot change over the seasons or from year to year. Transience can play a role in the delineation (e.g., Rodriguez and Nowak 2018), but in the end, the delineated area must be fixed. Accordingly, this study assumes a steady-state flow system and assumes that transient effects originating at the ground surface generally dampen out over a long period of time. The issue of transience is discussed further in Section 3.8.

Present standard practice in wellhead protection is to calibrate a model to available data and then to use the calibrated model predictively for CZD. A wellhead protection area is the two-dimensional areal extent of a well capture zone, commonly referred to as the contributing area (or area contributing recharge) to the well, which is typically used when there is a regulatory aspect to the area (Barlow et al. 2018). This procedure neglects structural uncertainty and numerical errors due to model-related differences such as the discretization scheme and the computational algorithm.

Pinder and Frind (1972) have shown that, with increased grid refinement, both finite element and finite difference model types converge to the same answer. However, run times increase with the discretization, and in practice, the time available for model runs is often limited. Therefore, groundwater models are often not discretized sufficiently fine. Discretization aspects specific to this study are discussed in Section 3.5.1.

The computational method used for the CZD can also have a major effect on the delineation. Section 3.6 compares two well-known methods, particle tracking and reverse transport. Section 3.7 extends the comparison to different particle tracking algorithms and shows how they can be combined.

In this study, four well-known model codes are used to delineate capture zones for two gaining stream reaches in the Alder Creek watershed in southern Ontario.

3.4 Model codes considered and comparison approach

The following four models were chosen for this study:

HydroGeoSphere (HGS): This physically-based, 3D finite element numerical model (Aquanty Inc. 2015) fully couples variably saturated subsurface flow (modified Richards' equation; Richards 1931) with surface water flow (diffusive wave simplification of the Saint Venant equations). The governing equations for HGS are presented in Section 2.1. As such, it is well suited for a stream capture zone study because it can generate its own water courses. It currently does not have a particle tracking routine, but is compatible with WATRAC (Section 2.2, Frind and Molson 2004) for particle tracking. To generate capture probability plumes I chose the Waterloo Transport Code (WTC) (Molson and Frind 2004), a 3D advective-dispersive mass transport code, because it is both capable of reverse transport and compatible with HGS.

WATFLOW: The finite element flow model WATFLOW (Molson et al. 2002) has been used extensively in previous studies of the Waterloo Moraine (Martin and Frind 1998, Frind et al. 2014). WATFLOW has an integrated automatic calibration algorithm (Beckers and Frind 2001), a particle tracking code, WATRAC (Frind and Molson 2004), and a transport code, WTC (Molson and Frind 2004). WATRAC is based on the Pollock method (Pollock 1988), adapted for triangular prismatic elements.

MODFLOW 2000 (VERSION 1.19.01): The finite difference model MODFLOW (McDonald and Harbaugh 1988) is the most widely used groundwater code worldwide. It is linked to the particle tracking code MODPATH (Pollock 1988). The original FORTRAN version is freely available through the U.S. Geological Survey, and several graphical user interface (GUI) versions are commercially available. This study uses Visual MODFLOW (Waterloo Hydrogeologic 2015) as the GUI for groundwater flow modelling and PEST (Doherty 2005) for calibration.

FEFLOW 7.0: This commercially available finite element flow and transport code is used widely for both industry and academic purposes (DHI-WASY 2015). FEFLOW has a sophisticated GUI and its own particle tracking algorithm, which is based on the 4th order Runge-Kutta method (Diersch 2013). The key difference between that method and the Pollock method is that the former smooths out particle tracks across an interface between different materials, while the latter has particles change direction at such interfaces, as required by theory (Bear 1972). As a result, 4th order Runge-Kutta particles may deviate from Pollock particles at material interfaces. The differences in particle tracking results depend on the degree of heterogeneity and discretization. The transport code also offers the user the choice between the conventional transport formulation and the Lichtner formulation (Lichtner et al. 2002). The latter makes a difference if the system contains significant areas with predominantly vertical velocities (Frind, M. et al. 2015).

An overview of the model and method comparison is described in four steps (Figure 3.1):

1. Develop a fully integrated surface water – groundwater model (HGS) to identify losing and gaining stream reaches (Section 3.5.2).
2. Select saturated groundwater models and apply hydraulic properties, boundary conditions and exchange fluxes (Section 3.5.3).
3. If necessary, recalibrate the selected groundwater models using an optimization algorithm (Section 3.5.3).
4. Delineate stream capture zones with each model (Section 3.6).

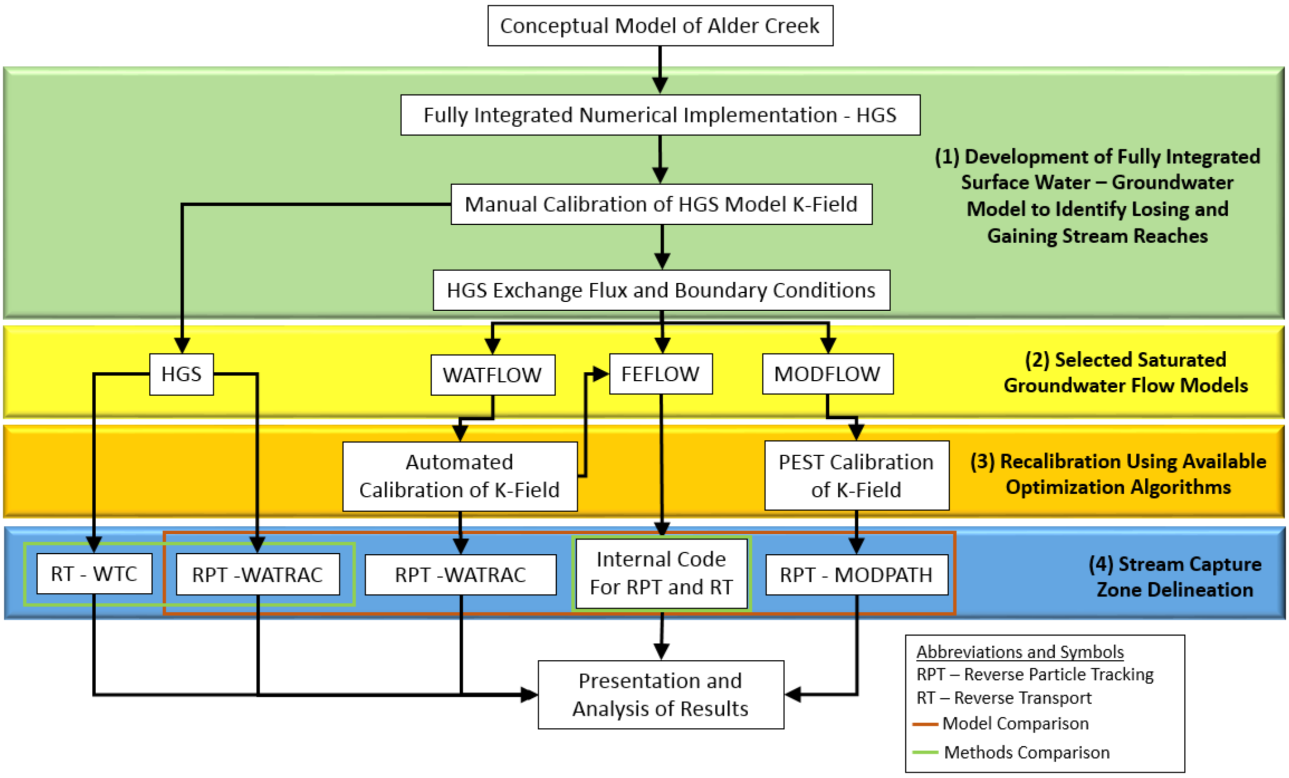


Figure 3.1: Approach to model comparison.

3.5 The Alder Creek watershed

The study area is the Alder Creek watershed, which is part of the Grand River watershed in southern Ontario (Figure 3.2). The watershed has a humid continental climate, with annual precipitation ranging from 780 to 1000 mm/year. It covers approximately 79 km^2 and contains agricultural, urban, forested, and grassland regions (Jones et al. 2009).

The Alder Creek watershed is an important water source for the Region of Waterloo (population 550,000). The region obtains 80 % of its water supply from groundwater sources, with approximately 30 % of that coming from the Mannheim well field which straddles the Alder Creek watershed. Due to its close proximity to the twin cities of Kitchener and Waterloo, the Alder Creek watershed is exposed to considerable development pressure. Understanding the areas that contribute to stream baseflow can help protect the aquatic ecosystems in the watershed.

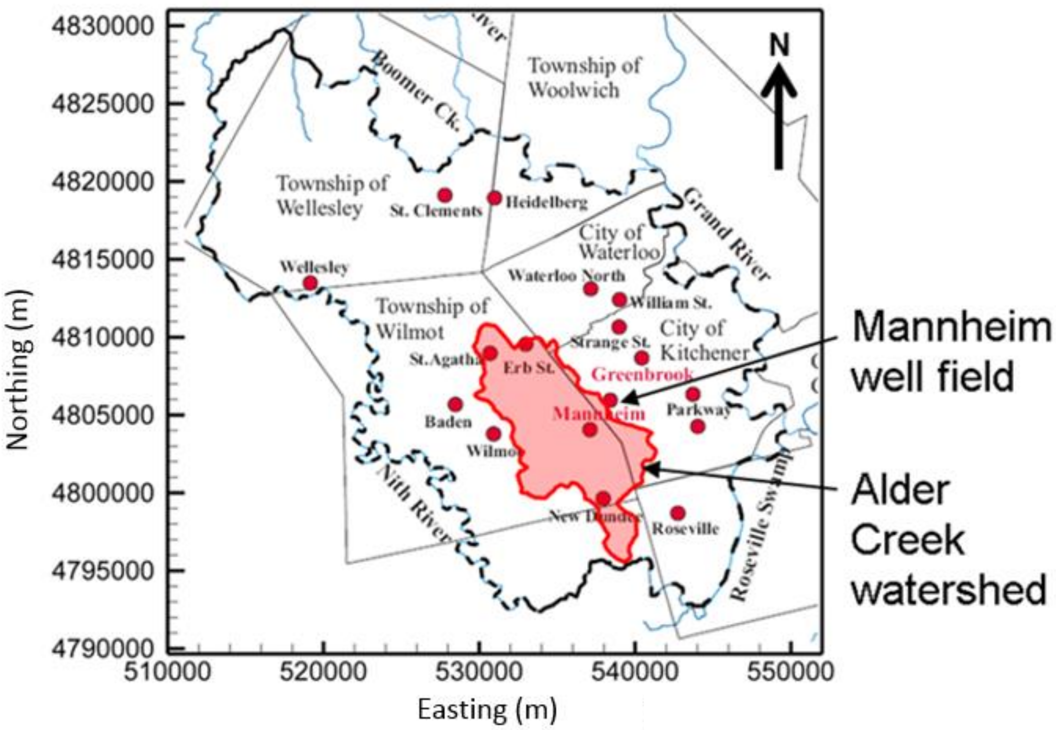


Figure 3.2: Waterloo Moraine model with Alder Creek watershed. Region of Waterloo well fields (red dots) shown (adapted from Frind et al. 2002).

Figure 3.2 shows the well fields (red dots) operated by the Region of Waterloo throughout the Waterloo Moraine area. The Alder Creek watershed itself contains 10 pumping wells and 28 observation wells. The average water table elevation from 1991 to 2000 was mapped by CH2MHILL and S.S. Papadopulos & Associates Inc. (2003). The groundwater flow direction was interpreted to be generally from northwest to southeast within the northern part of the Alder Creek watershed, following the regional gradient. There is also a distinct pattern of convergence toward the creek at the southern tip of the watershed.

3.5.1 Alder Creek model: conceptualization and discretization

The original 3D Waterloo Moraine groundwater flow model (Martin and Frind 1998) was developed from an extensive hydrostratigraphic database maintained by the Ontario Ministry of the Environment. Jones et al. (2009) developed the HGS Alder Creek model as a sub-model of the Waterloo Moraine model to study unsaturated flow and solute transport, which required a high-resolution discretization of the vadose zone. Jones et al. (2009) divided the domain vertically into 87 elemental layers (Figure 3.3a), with the first meter below ground surface discretized with ten 10 cm layers, the next 19 m with fifty-seven 33 cm layers and the final 20 m to bedrock with 20 evenly distributed layers. The present study continues using the fine grid discretization for HGS in order to provide a high degree of accuracy in the calculation of the exchange flux (Section 3.5.2). The same discretization was used for the other two finite element models, WATFLOW and FEFLOW. Because of this fine discretization, inclusion of river bed conductance as a separate parameter was not considered necessary.

For the Alder Creek MODFLOW model, which cannot easily handle vadose zone flow, a finite difference discretization scheme with a coarser lateral and vertical discretization (Figure 3.3b), consistent with industry standards was used (Anderson et al. 2015). The vertical discretization is one model layer per hydrostratigraphic layer and the lateral refinement is localized around the municipal supply wells. This is to avoid long simulation run times and convergence issues due to the rewetting of dry cells. The newest version of MODFLOW, MODFLOW USG (Panday et al. 2015), handles unstructured grids; however, it is not used here because it has yet to gain popular acceptance as the industry standard version of MODFLOW and is missing key post-processing functions such as particle tracking.

For the present study, a no-flow boundary condition around the model periphery is used to represent a groundwater flow divide based on the topographic highs. Around the perimeter of the bottom-most model layer, the no-flow boundary is replaced by a constant-head boundary condition to allow for lateral flow representing regional flow towards the Grand River. Head values for this boundary were obtained from the Waterloo Moraine model (Martin and Frind 1998). An exchange flux, discussed in the following section, is assigned to the top surface of the model. Figure 3.4 shows a conceptual cross section of the boundary conditions.

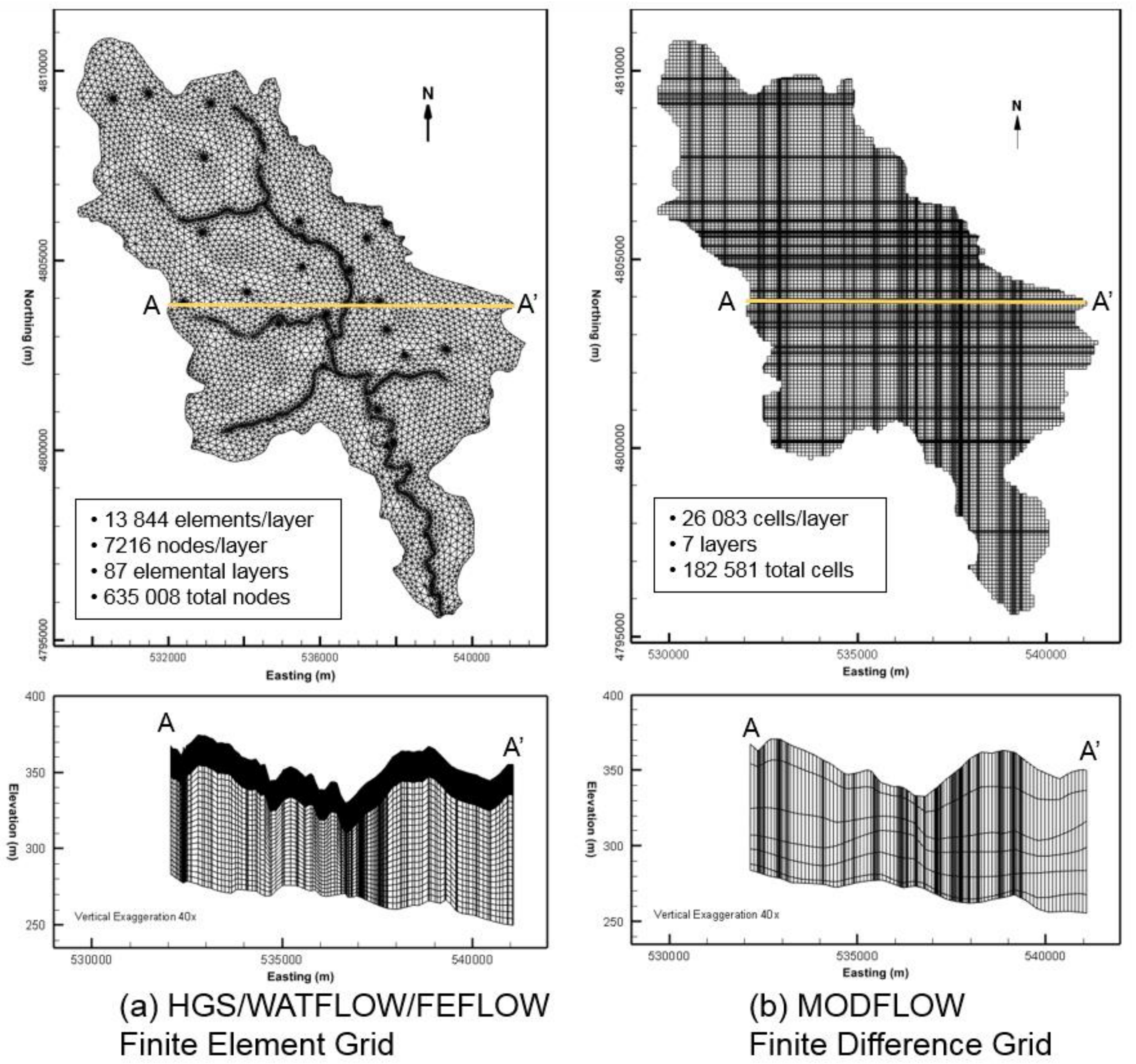
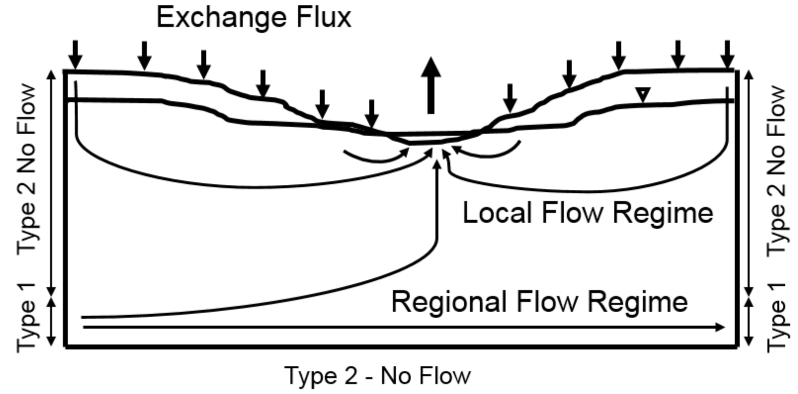


Figure 3.3: Model discretization: (a) Finite element models, (b) Finite difference model.



Type 1 – Constant Head values obtained from regional Waterloo Moraine Model (Martin & Frind 1998)

Figure 3.4: Alder Creek watershed conceptual cross-section of boundary conditions.

3.5.2 The exchange flux: identifying gaining stream reaches

CZD can only be performed on groundwater sinks. A gaining stream is conceptualized as a groundwater sink on the top surface of the model. Therefore, in order to delineate a stream capture zone, we must identify the spatial distribution of groundwater passing through the model top surface. In this study, HGS is used to identify this spatial distribution.

A uniform net rainfall rate of 200 mm/year was applied to the surface of the initially saturated HGS Alder Creek model and run until steady-state flow conditions were reached. Jones et al. (2009) calibrated the hydraulic conductivity of the model manually, with the results summarized in Section 3.5.3.

HGS calculates the distribution of water infiltrating and exfiltrating across the land surface (termed exchange flux), without the need to explicitly specify the surface water boundary condition. The exchange flux distribution over the Alder Creek watershed is shown in Figure 3.5. Groundwater recharge is depicted by the warm colours (positive numbers) with most of the model domain having a positive exchange flux depicting recharge. Groundwater discharge is depicted by the cool colours (negative numbers). The exchange flux along the stream channels depicts areas of gaining and losing streams at steady-state; this can be conceptualized as the long-term average baseflow of the stream. There is a broad accumulation of net discharge around the outlet of stream channel in the south region. Groundwater lost to evapotranspiration is not explicitly accounted for since a net recharge value was applied to the model.

Figure 3.5 also shows two gaining stream reaches, Stream Reaches 1 and 2, that were chosen for CZD. Both are located in the center of the model domain approximately 7.4 km upstream from the discharge outlet, within a well-calibrated area of the watershed (discussed in Section 3.5.3), and are situated sufficiently far from the model boundaries to avoid boundary effects.

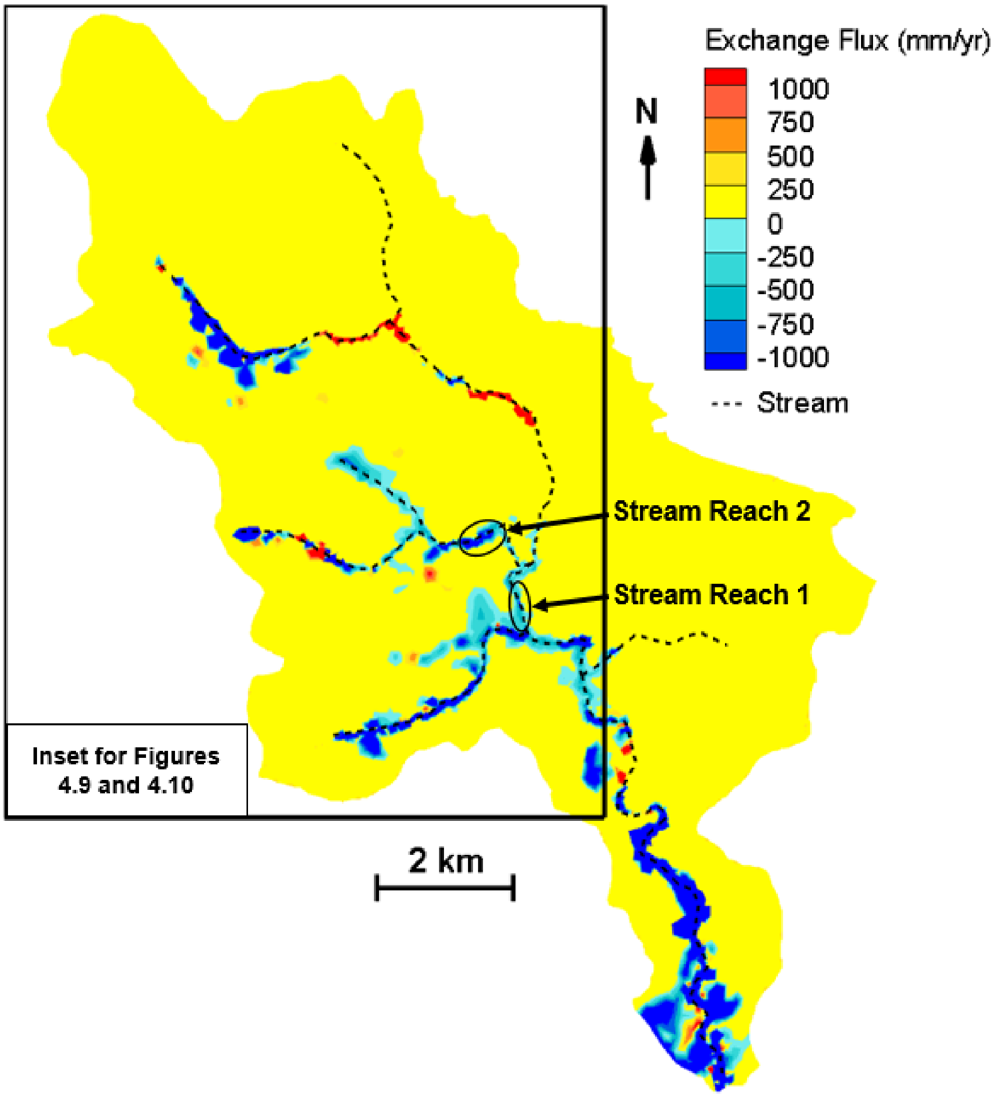


Figure 3.5: Alder Creek watershed exchange flux distribution, showing groundwater recharge, gaining stream reaches, and Stream Reach 1 and 2. Negative exchange flux (cool colours) is groundwater discharge, positive exchange flux (warm colours) is groundwater recharge. Border outlining inset for Figure 3.9 and 3.10.

3.5.3 Model calibration and hydraulic head distribution

The hydraulic properties, boundary conditions and exchange flux from the manually calibrated HGS model were applied to the WATFLOW, MODFLOW, and FEFLOW models. These models were then used to solve for steady-state saturated groundwater flow. The unsaturated zone was not represented in these three models because the overall shape of the steady-state stream capture zone was found to be insensitive to the unsaturated zone. This was tested by comparing the results with a simplified representation of the unsaturated zone in WATFLOW. Furthermore, in this hydrogeologic setting, flow through the unsaturated zone is driven by gravity and is mainly in the vertical direction. Thus the inclusion of unsaturated flow, while potentially affecting travel time, would have no impact on the overall shape of the steady-state capture zones.

Three head observations were discarded from the calibration dataset because of their close proximity to municipal supply wells and model limits, hence only 25 of 28 of the observed head measurements were used in the calibration process. HGS is computationally demanding when overland flow and variably saturated subsurface flow are fully coupled, requiring approximately 18 hours to reach a steady-state solution (using the non-parallelized version). While WATFLOW, MODFLOW, and FEFLOW could reach a steady-state solution (for saturated groundwater flow only) within 1.6 minutes, 18 seconds, and 17 seconds, respectively. The difference in run times played a key role in model calibration.

It is common practice to use an automated calibration routine to reach an optimum solution based on an objective function. WATFLOW and MODFLOW were recalibrated using their respective automated calibration routines (Beckers and Frind 2001, Doherty 2005). The recalibration found one optimum hydraulic conductivity multiplier per hydrostratigraphic layer (within an order of magnitude of the initial calibrated value). The calibrated WATFLOW hydraulic conductivity field was applied to FEFLOW, producing identical calibration results because the same discretization scheme and grid applies in both models. HGS was not recalibrated by automatic means because of its long run times; instead, the original hydraulic conductivity field was used to avoid affecting the exchange flux, which would have required further iterative calibration steps. Slight differences in HGS hydraulic head when compared to the other models (only saturated flow) is due to its full sw-gw coupling and representation of flow in the unsaturated zone.

Figure 3.6 and Table 3.1 show the comparison between observed to simulated heads and calibration statistics for HGS, WATFLOW/FEFLOW, and MODFLOW. All models produced acceptable calibration results when comparing observed to simulated heads, with normalized root mean square errors below 9 % and correlation coefficients greater than 97.5 %. Most importantly, the heads for all models agree well in the central area (heads ranging between 330 and 340 m) where Stream Reaches 1 and 2 are located (Figure 3.5).

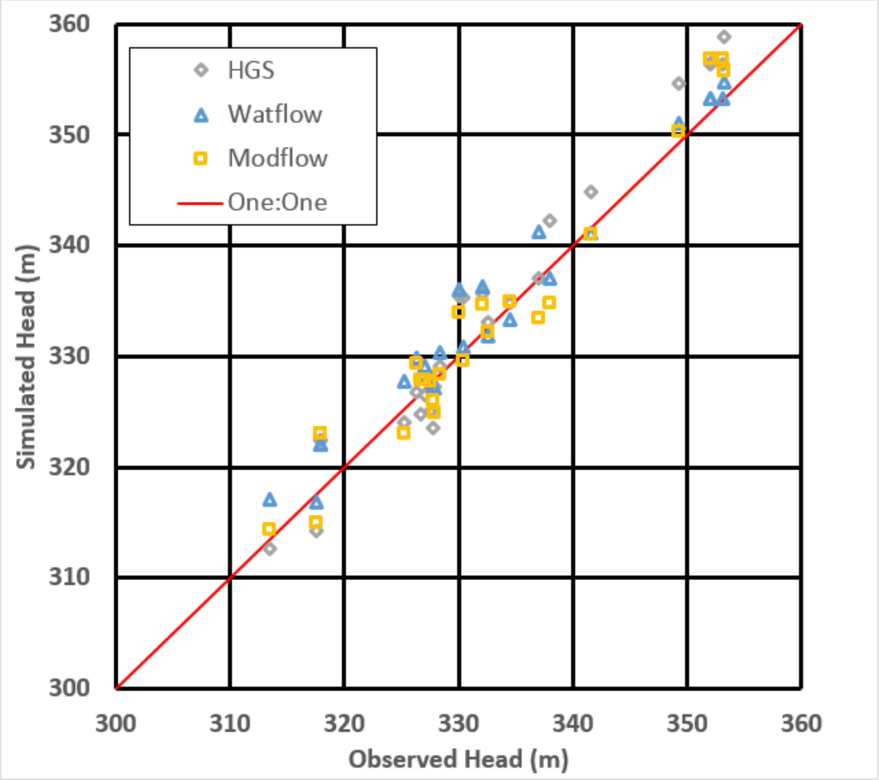


Figure 3.6: Calibration results for HGS, WATFLOW/FEFLOW, and MODFLOW.

Table 3.1: Calibration results for HGS, WATFLOW/FEFLOW, and MODFLOW

Model	**HGS**	**WATFLOW/ FEFLOW**	**MODFLOW**
Number of Observations		25	
Observation Range (m)		39.8	
Average Residual (m)	-1.6	-1.6	-0.9
Absolute Average Residual (m)	2.8	2.0	2.3
Root Mean Square (m)	3.4	2.6	2.8
Normalized Root Mean Square (%)	8.5	6.4	7.1
Correlation Coefficient (%)	98.2	98.5	97.7

Figure 3.7 shows the hydraulic head distribution for the main aquifer (Aquifer 1) for all models. Very similar hydraulic head distributions are produced by all four models, with high hydraulic head levels (approximately 360-370 m) at the northwest edge of the model domain and low hydraulic head levels (approximately 300-310 m) at the southeast edge. Groundwater flow is generally from northwest to southeast, which agrees with interpretations by CH2MHILL and S.S. Papadopulos & Associates Inc. (2003).

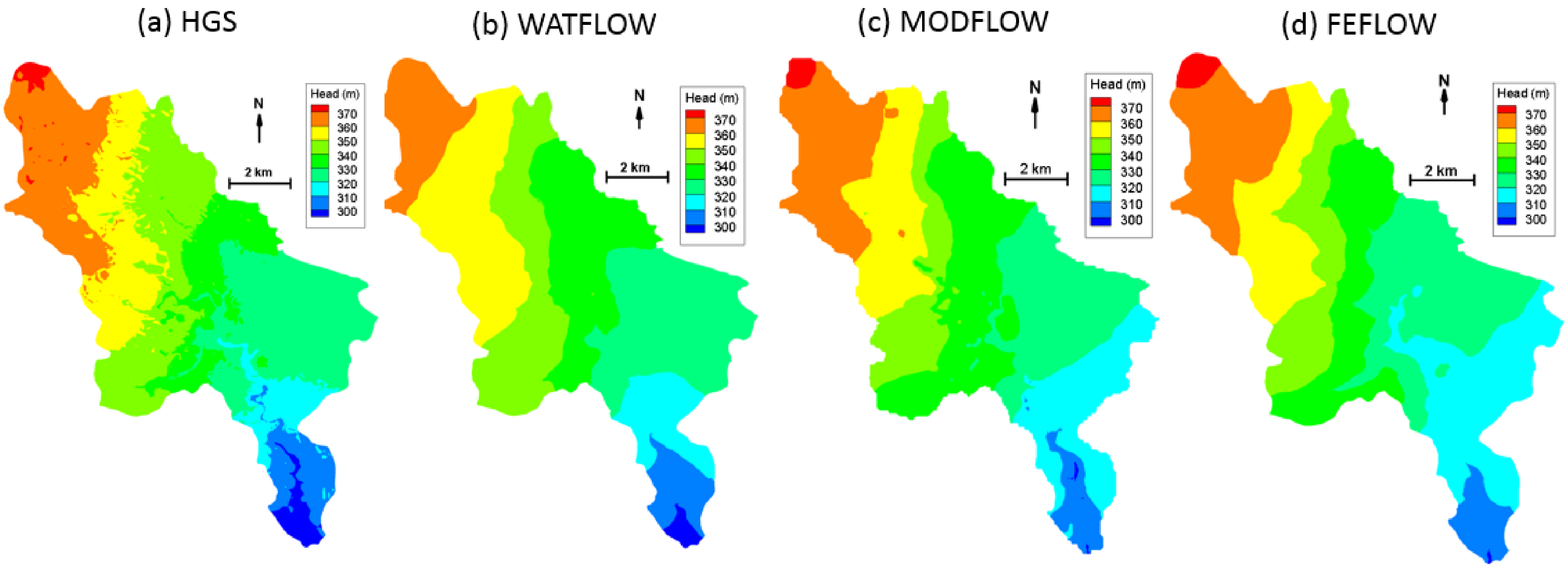


Figure 3.7: Simulated hydraulic heads in main aquifer: (a) HGS, (b) WATFLOW, (c) MODFLOW, (d) FEFLOW.

3.6 Stream capture zone delineation methods

3.6.1 Particle tracking

The usual approach to particle tracking CZD is to place a number of particles around the screen of a well and to track the advective travel path of the particles in the reverse (i.e., upgradient) direction. In this study, the well has been replaced by a gaining stream reach. 150 and 200 particles were placed within the vertical mid-point of the upper-most model layer for Stream Reach 1 and 2, respectively (Figure 3.5). Increasing the number of particles across the stream reaches gave essentially the same results, providing assurance that most preferred flow paths had been identified.

An alternative approach is to apply both forward and reverse particle tracking where the forward particles are placed at or near expected recharge areas. This often helps to identify areas that reverse particles may have missed. In this study, forward particle tracking was not used because particle exit points are highly dispersed over large areas (see Figure 3.9 and 3.10), making this option impractical.

Particle tracks can be sensitive to the particle starting position. Due to the inherent grid differences, starting positions for structured rectilinear and unstructured prismatic grids cannot be exactly the same, so care was taken to minimize the differences. Figure 3.8 shows the particle starting position for the finite difference model (MODFLOW) and the finite element models (HGS, WATFLOW, and FEFLOW). The particles were tracked backwards by advection until they reached the ground surface or no longer advanced, which occurred at approximately 300 years for all models.

In order to focus on model-related differences, particle tracking was performed on only one calibrated realization of the hydraulic conductivity field for each model, thus neglecting parameter uncertainty. Section 3.6.2 describes an alternative methodology that can be used for a more complete analysis including local-scale parameter uncertainty.

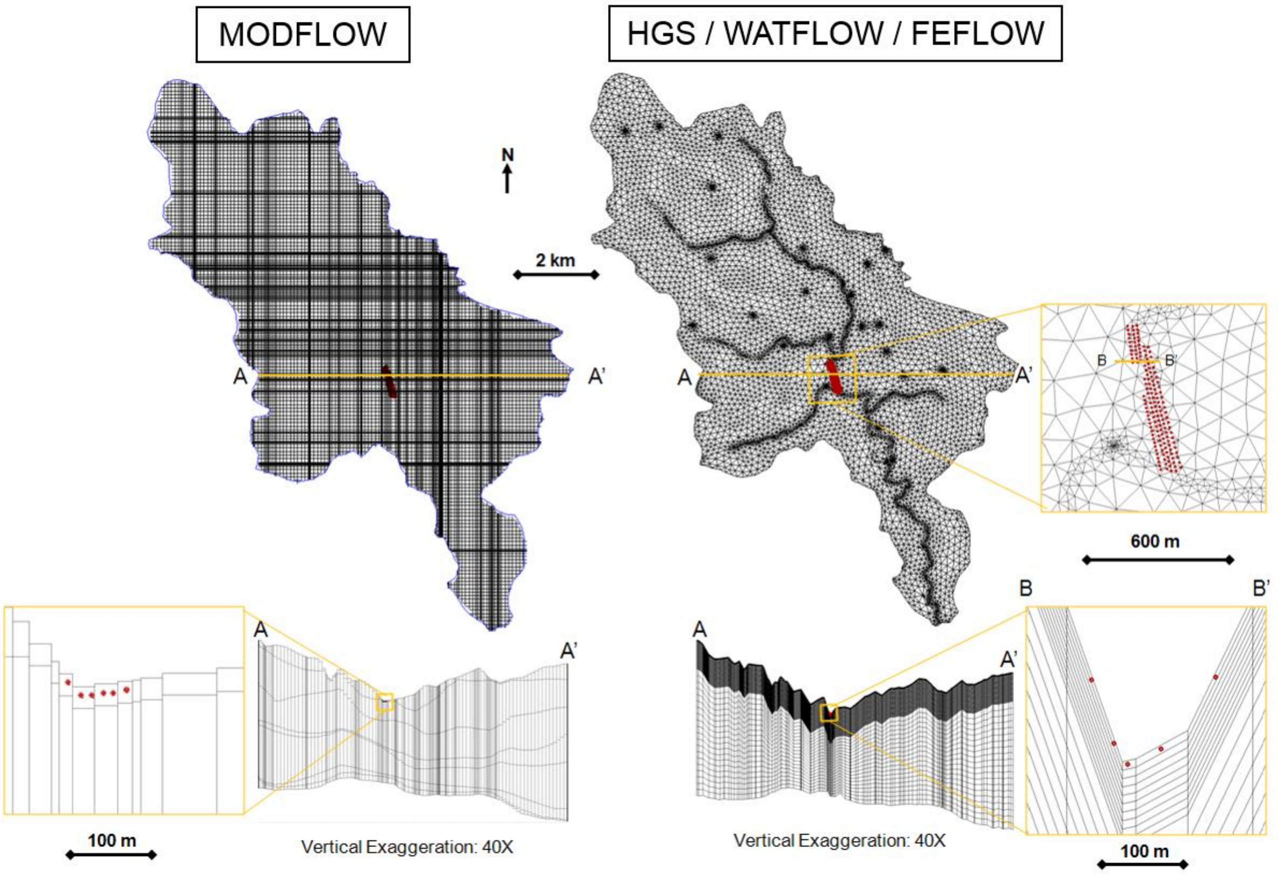


Figure 3.8: Initial particle placement for Stream Reach 1.

3.6.2 Reverse transport

Reverse transport uses the advective-dispersive transport equation and is conceptually equivalent to applying the reverse particle tracking method with multiple realizations of the hydraulic conductivity field (Neupauer and Wilson 2001, Frind et al. 2002, Frind et al. 2006). The macrodispersion term in the advective-dispersion equation can be conceptualized as a term that represents local-scale parameter uncertainty in hydraulic conductivity (Gelhar and Axness 1983). The transport parameters were those used previously by Frind et al. (2002), with an effective molecular diffusion coefficient of 10^{-10} m^2/s, and dispersivities of 20 m, 5 m, and 0.02 m, for longitudinal, transverse horizontal, and transverse vertical dispersion, respectively. These values are within the range of dispersivities given by Gelhar et al. (1992) and Engesgaard et al. (1996) for a model of this scale. These parameter values are equally valid for the forward and reverse forms of the transport equation.

The reverse transport capture zone is expressed in terms of capture probability P. The equation is solved with a Type 1 (specified capture probability) of P = 1 at the stream and P = 0 at distant points, producing a capture probability plume (CPP) with values ranging between $0 \leq P \leq 1$.

Reverse transport was run until a backwards time of 300 years was reached to obtain a capture probability at equilibrium. CPPs were generated with the HGS and FEFLOW steady-state flow field for Stream Reach 1 and 2. These CPPs were then compared to their corresponding reverse particle tracks. The validity and the advantages of the reverse transport method have been discussed, e.g., by Sousa et al. (2013).

3.7 Results and discussion

3.7.1 Models comparison: HGS, WATFLOW, MODFLOW, and FEFLOW

Figure 3.9 shows the reverse particle tracks obtained for Stream Reach 1 for all four models. An examination of these particle tracks reveals that, for all models, the particles first descend to the bottom aquifer, and then turn in the upgradient (northwest) direction.

The particle tracks from HGS and WATFLOW are very similar in distance and direction, both tending west, with those from HGS being shortest. The small difference can be attributed to the differences in hydraulic conductivity distribution resulting from recalibrating WATFLOW. The MODFLOW particle tracks tend to deviate farther to the northwest than HGS and WATFLOW; this difference is expected due to the different discretization scheme. The FEFLOW particle tracks show larger differences, bifurcating into two directions, one towards the northwest along the boundary and the other towards the south, also reaching the model limits. This difference is mainly attributed to the difference in the particle tracking algorithm (4th order Runge-Kutta method for FEFLOW vs. Pollock method for HGS/WATFLOW/-MODFLOW).

Figure 3.10 shows the particle tracks for Stream Reach 2. For all models, particles once again descend to the bottom aquifer before moving west-northwest. Particles from HGS and WATFLOW travel a short distance to the west, while the particles from MODFLOW and FEFLOW travel a greater distance and are directed more towards the northwest. Again, the major causes of the differences in the particle tracks between models are the discretization scheme (for MODFLOW) and the particle tracking algorithm (for FEFLOW). A negligible two out of the 200 particles applied in MODFLOW deviate from the bulk of the particles and travel a considerable distance along the model northwest boundary; these two particles can be ignored.

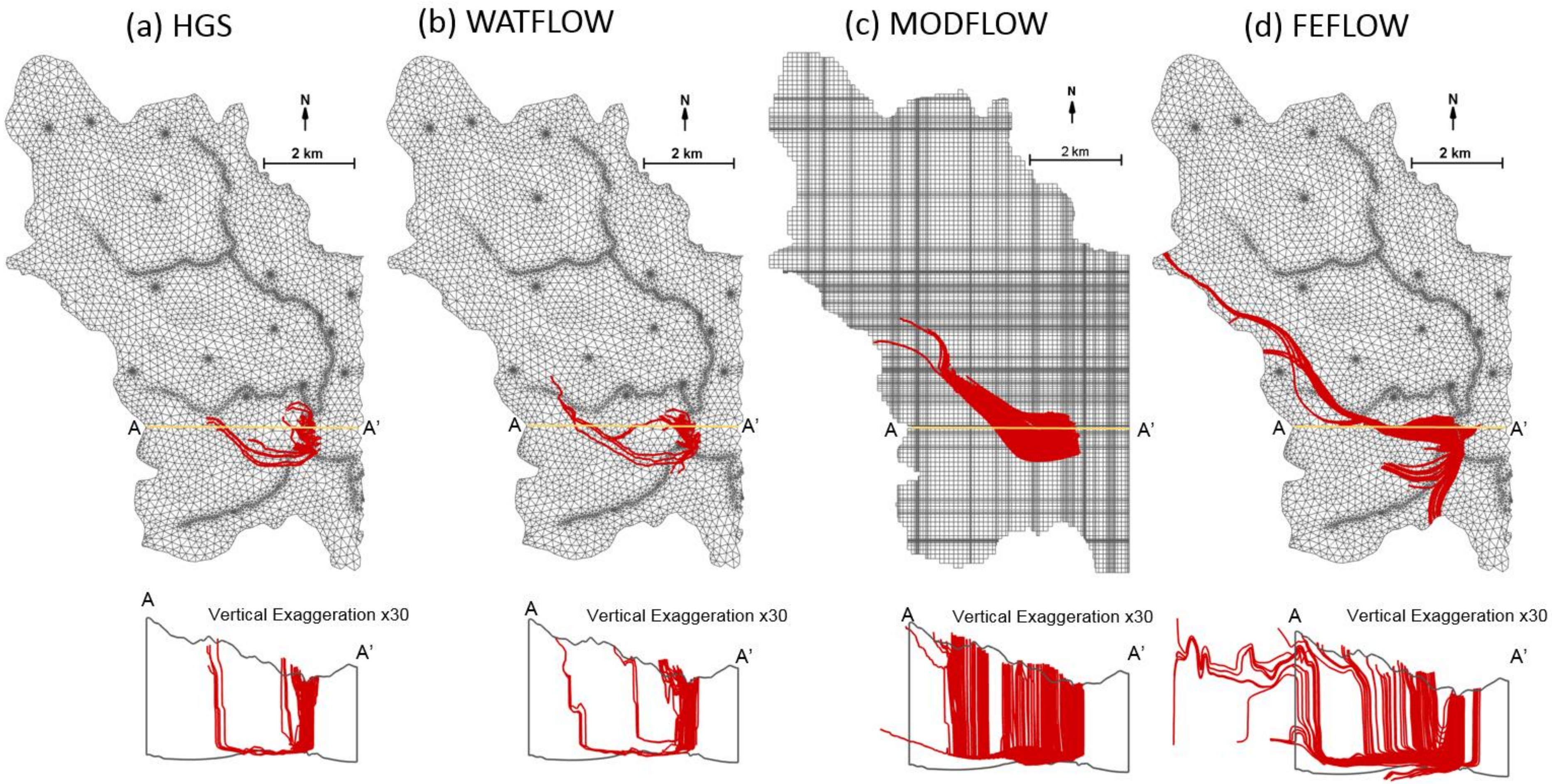


Figure 3.9: Particle tracks for Stream Reach 1: (a) HGS, (b) WATFLOW, (c) MODFLOW, (d) FEFLOW.

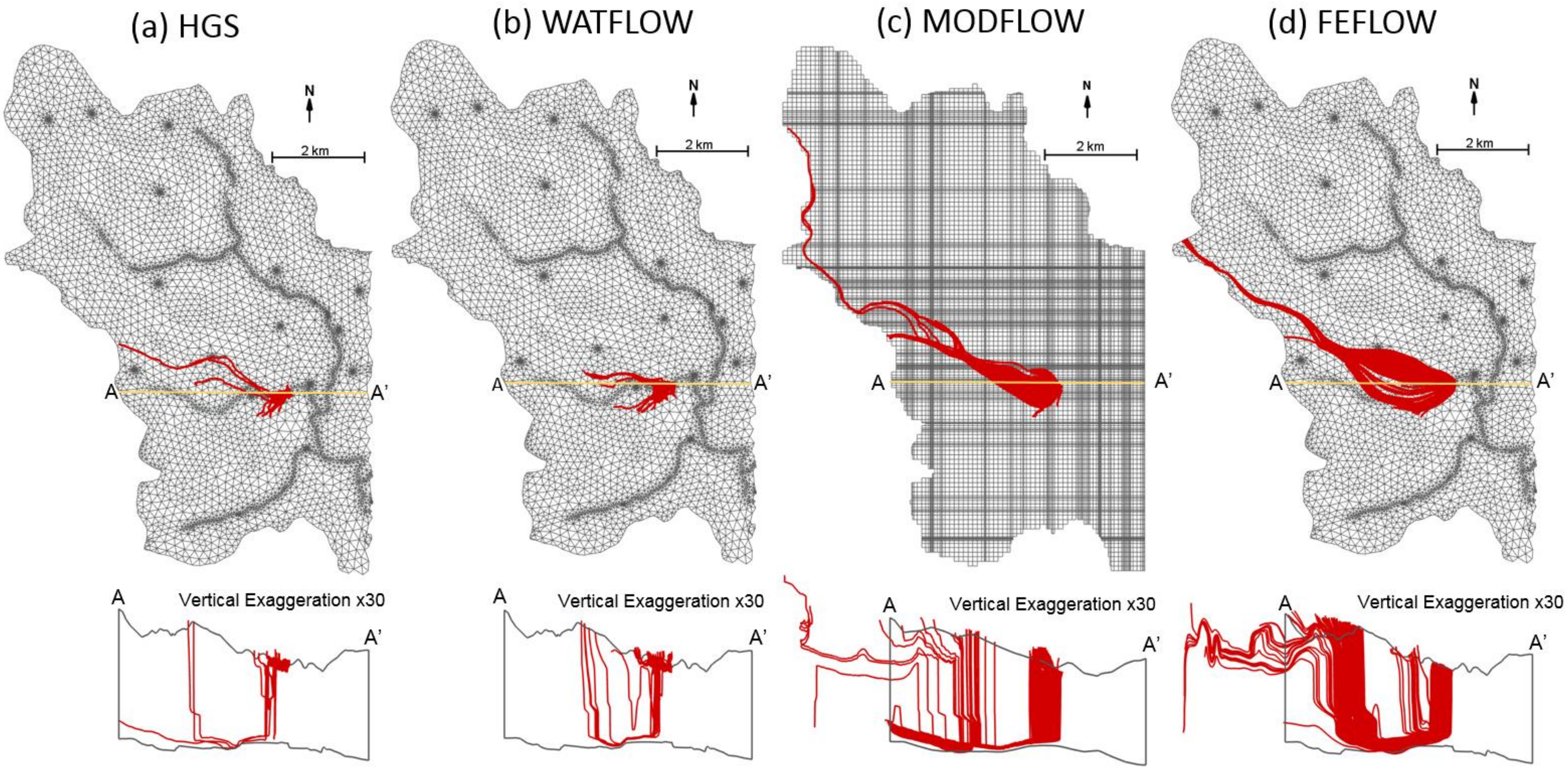


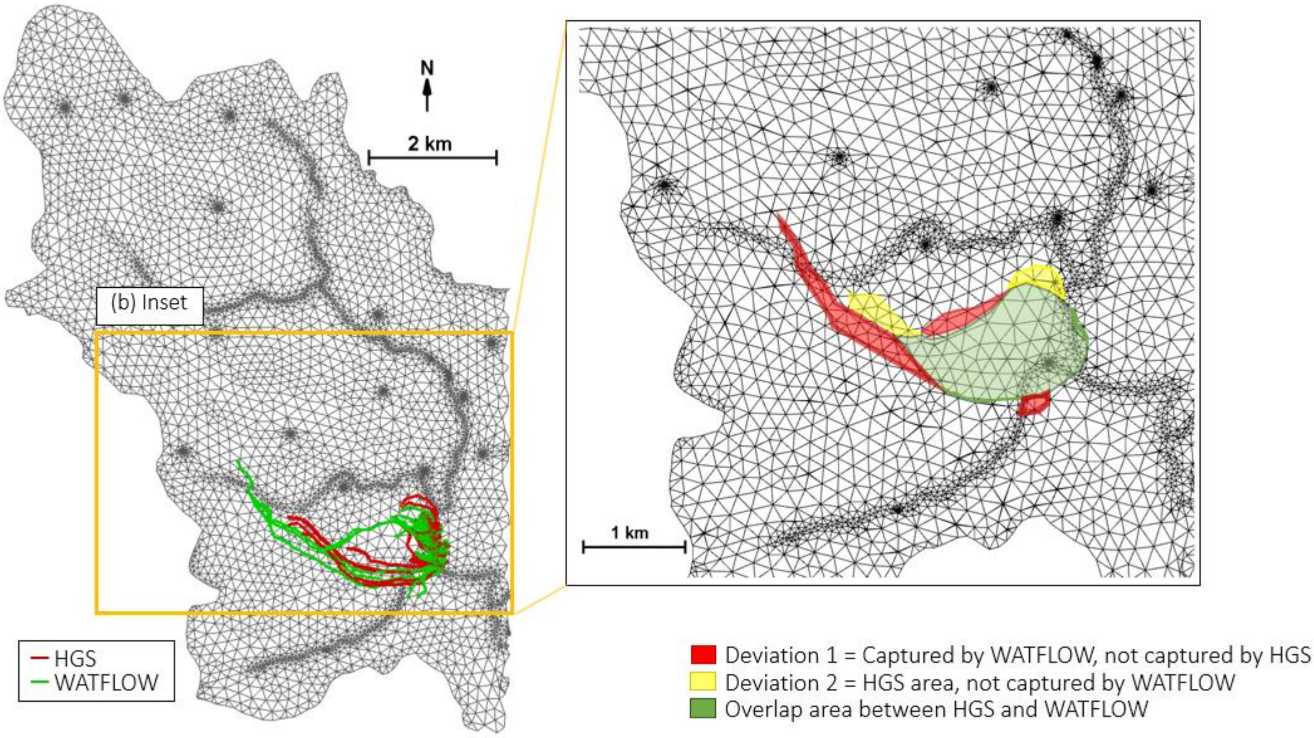
Figure 3.10: Particle tracks for Stream Reach 2: (a) HGS, (b) WATFLOW, (c) MODFLOW, (d) FEFLOW.

For a more quantitative assessment of the particle tracks, we can compare the envelopes drawn around the respective particle bundles under the different models to each other. The comparison will account for differences in the model and the particle tracking algorithm, but it will not account for parameter uncertainty. Drawing the envelope itself involves some subjectivity because of the gaps in the particle tracks; these gaps are mainly due to the bias imparted on the particles due to their starting position in the grid. These gaps will disappear when either the reverse transport method (see Section 3.6.2) or the Monte Carlo approach (Enzenhoefer et al. 2014, Meyer et al. 2014) is used.

The HGS model can be considered as the most rigorous model among those selected, therefore the HGS solution was selected as the benchmark solution for this comparison. However, this selection does not necessarily imply correctness.

Superimposing the capture zones delineated by the other models over the benchmark HGS capture zone, the mismatch between the capture zones can be viewed as deviations relative to the HGS capture zone. Figure 3.11 gives an example for the case of the WATFLOW-HGS comparison. Three different deviations are identified:

1. Deviation Type 1 (highlighted in red): The area captured by the other model but not captured by HGS. This area can be viewed as an area that is mislabeled as being part of the capture zone when compared with HGS.
2. Deviation Type 2 (highlighted in yellow): The area captured by HGS but not captured by the other model. This area can be viewed as an area that should have been part of the capture zone generated by the other model when compared with HGS.
3. Overlap (highlighted in green): The area captured by the HGS particle tracks and the area of the other model. This area can be viewed as an area of model agreement.



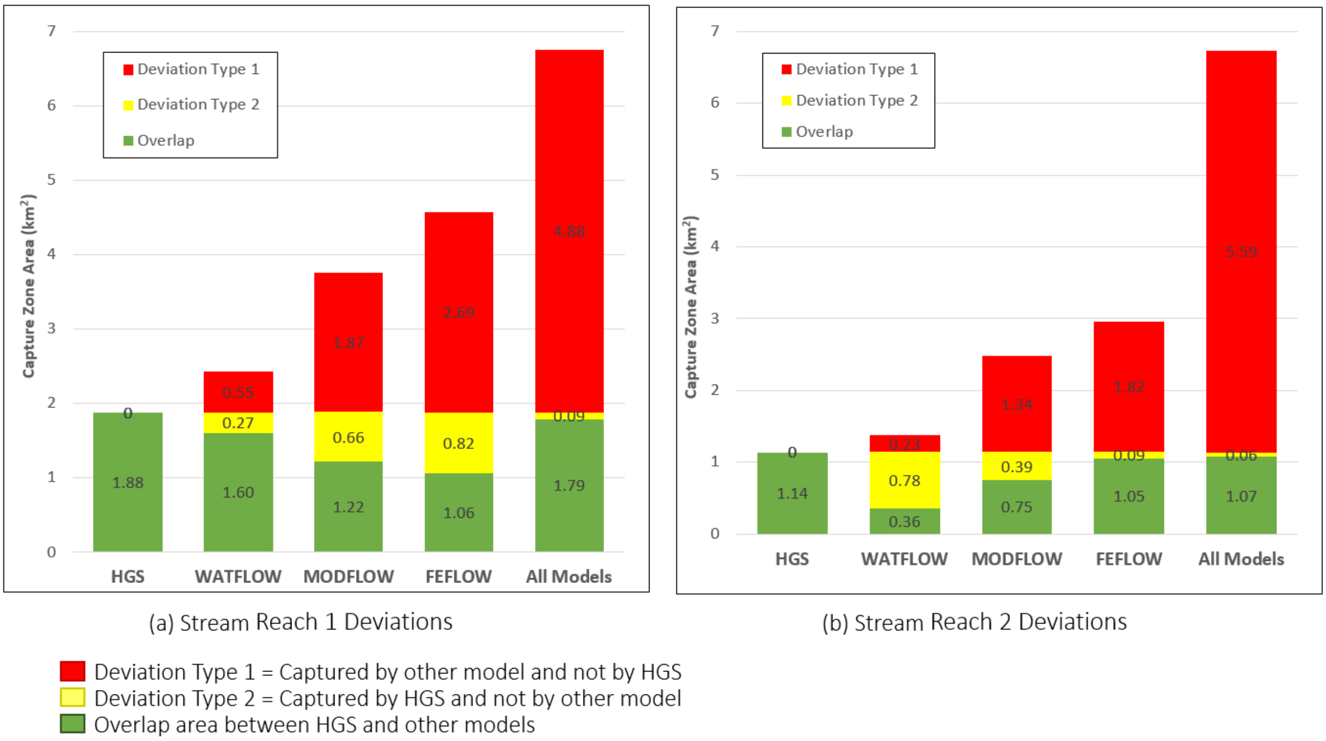
(a) HGS and WATFLOW particles superimposed for Stream Reach 1

(b) Inset: Key capture zone areas benchmarked to HGS

Figure 3.11: Example of key areas when comparing WATFLOW to HGS capture zone for Stream Reach 1: (a) HGS and WATFLOW particles superimposed for Stream Reach 1, (b) Deviations benchmarked to HGS.

These deviations are summarized in a stacked bar graph for each of the selected models (Figure 3.12). Analysis of the bar graph for Stream Reach 1 (Figure 3.12a) reveals that the WATFLOW deviations are much smaller than the deviations from the other models, suggesting that the manually calibrated HGS model and the automatically calibrated WATFLOW model are very similar. FEFLOW deviations are the largest, with FEFLOW Deviation Type 1 being far greater than the deviations from the other models. MODFLOW’s deviations fall between those of WATFLOW and FEFLOW. The causes for the deviations are as follows:

- **WATFLOW**: Differences in the hydraulic conductivity distribution due to the WATFLOW automatic calibration routine (Beckers and Frind 2001).
- **MODFLOW**: Different discretization (finite difference, coarser and rectilinear grid).
- **FEFLOW**: Hydraulic conductivity from WATFLOW recalibration and different particle tracking algorithm (4th order Runge-Kutta method as opposed to Pollock or Pollock-equivalent for other models).



(a) Stream Reach 1 Deviations

(b) Stream Reach 2 Deviations

🟥 Deviation Type 1 = Captured by other model and not by HGS
🟨 Deviation Type 2 = Captured by HGS and not by other model
🟩 Overlap area between HGS and other models

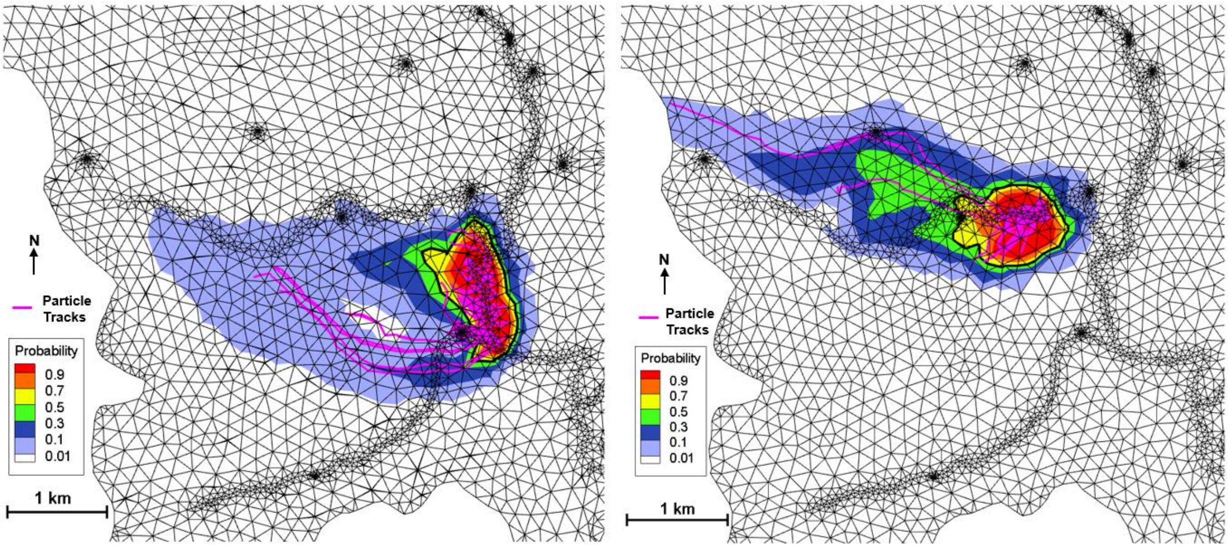
Figure 3.12: Summary of reverse particle track capture zones deviations: (a) Stream Reach 1 Deviations, (b) Stream Reach 2 Deviations.

Repeating the same analysis for Stream Reach 2 (Figure 3.12b), we can see that a similar pattern emerges. HGS and WATFLOW produce the smallest capture zones. In this case, the WATFLOW capture zone is much smaller than the HGS capture zone, leading to a much larger Deviation Type 2. FEFLOW again has the greatest Deviation Type 1, extending much further northwest than the HGS capture zone. MODFLOW again has deviations that fall somewhere in between that of WATFLOW and FEFLOW. This comparison demonstrates that CZD can be extremely sensitive to the post-processing particle tracking algorithm, even more so than the model discretization scheme. (Note that the terminology used here should not be interpreted as meaning correctness or lack thereof.)

The last bar on Figure 3.12, labeled “All Models” shows the total key areas from combining the capture zones from WATFLOW, MODFLOW, and FEFLOW. Compared to the HGS capture area, the total model-related uncertainty can vary approximately 2 to 5 times in size in the form of a Deviation Type 1. This supports the findings of previous researchers that capture zone extent and direction is sensitive to small differences or errors in the hydraulic head due to different conceptualizations or parameter changes (Franke et al. 1998). Model calibration by itself only shows that a model fits the data; it does not allow any conclusion as to which is the “best” model.

3.7.2 Method comparison: reverse particle tracking and reverse transport

Figure 3.13 shows the resulting CPPs obtained for Stream Reach 1 and 2 with the steady-state HGS flow field. The 0.5 probability contour is highlighted; this contour has been recommended by Molson and Frind (2012) as a preferred choice for CZD. The particle tracks are superimposed over the CPP to compare the delineation methods.



(a) Stream Reach 1 (b) Stream Reach 2

Figure 3.13: HGS capture probability plume with 0.5 contour highlighted and particle tracks superimposed (Figure 4.11 Inset): (a) Stream Reach 1, (b) Stream Reach 2.

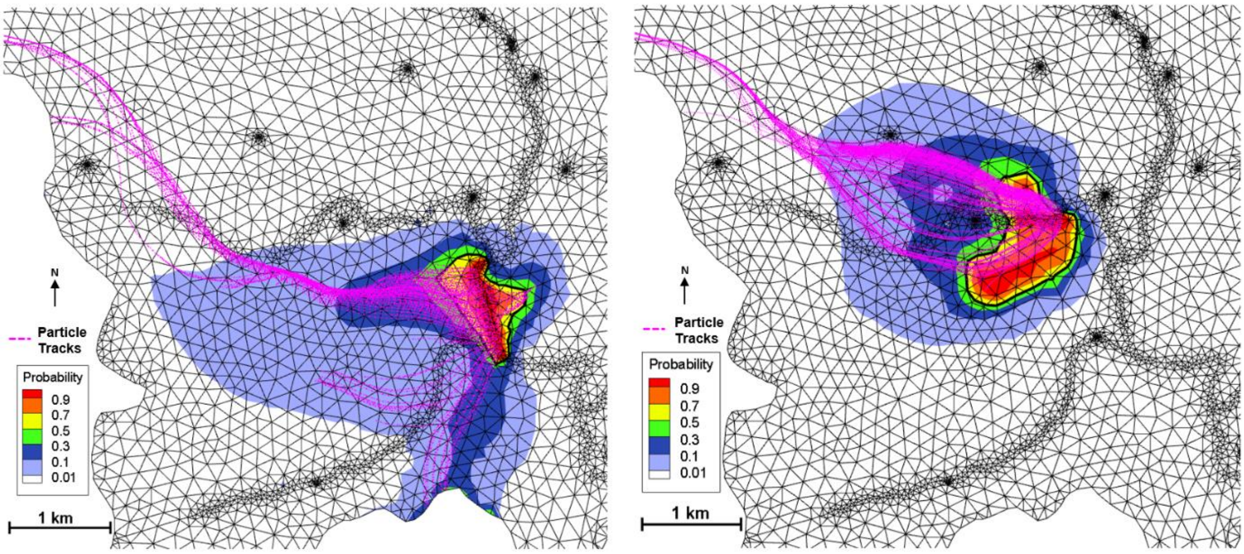
In Figure 3.13a the area outlined by the 0.5 probability contour is approximately 0.67 km^2. The predominant area contributing baseflow to Stream Reach 1 extends to the east and west of the stream. There is a probability of 0.01 that the area extending approximately 2.5 km west also contributes to the base flow.

In Figure 3.13b the area outlined by the 0.5 probability contour is approximately 0.65 km^2. The area contributing baseflow to Stream Reach 2 extends mainly to the north and south of the stream. There is a probability of 0.1 that a contribution to baseflow also comes from the area extending approximately 2.0 km west-northwest. The 0.01 contour reaches the limits of the model domain, but this minor boundary effect has no impact on the core capture zone $P \geq 0.5$.

For Stream Reach 1 (Figure 3.13a), the CPP corresponds well with the superimposed particle tracks. The agreement between the CPP and particle tracks is excellent, with 80 % of the particles terminating within the 0.5 contour. For Stream Reach 2 (Figure 3.13b), the agreement is even better with 98 % of the particles terminating within the 0.5 contour. This suggests that baseflow for this stream reach comes predominantly from local recharge within the area of the 0.5 contour. This reinforces the choice of the 0.5 contour as a reasonable capture zone for the current study.

One important note is that only a small percentage of the total particles released travel beyond the 0.5 contour, these few stray particles travel within the low CPP contour range (0.01 to 0.1). Although these stray particles follow low probability travel paths, they can have a significant impact on the particle envelope and hence on the interpreted capture zone delineated by reverse particle tracking, as shown above. Superposition of the particle tracks on the CPP allows insight into the significance of outlying particles.

Figure 3.14 shows the resulting CPPs obtained for Stream Reaches 1 and 2 with the steady-state FEFLOW flow field, also with the 0.5 contour highlighted. Comparison between the HGS and FEFLOW CPP for Stream Reach 1 (Figure 3.13a and 3.14a) shows that, except for the southern arm of the FEFLOW CPP, the two CPPs are similar. Comparing the HGS and FEFLOW CPP for Stream Reach 2 (Figure 3.13b and 3.14b) shows that the agreement between the two CPPs is not quite as good, but both are plausible. In any case, the CPPs of both models could be merged by selecting the maximum value at each (x,y) point from both plumes, following the procedure developed by Sousa et al. (2013).



(a) Stream Reach 1

(b) Stream Reach 2

Figure 3.14: FEFLOW capture probability plume with 0.5 contour highlighted and particle tracks superimposed (Figure 3.11 Inset): (a) Stream Reach 1, (b) Stream Reach 2.

However, Figure 3.14 also shows that the FEFLOW particle tracks do not agree well with their corresponding CPPs, with the bulk of the FEFLOW particle tracks extending a considerable distance beyond the 0.5 contour. This can be expected due to the particle tracking scheme used by FEFLOW. This method, known as the 4th order Runge-Kutta method, per definition generates smooth particle tracks, which is an advantage in solving the Navier-Stokes equations with inertia terms. However, for groundwater flow in heterogeneous media, it cannot accurately represent the change in the flow direction at internal material interfaces in proportion with the permeability ratio (Bear 1972), unless a very fine discretization is used. With the discretization used in this study, the particle tracks are smoothed and stretched out across natural discontinuities, leading to a mismatch between the FEFLOW particle tracks and the corresponding CPP.

Since neither CZD approach (i.e., reverse particle tracking or reverse transport) can be independently verified, having two approaches producing compatible results supports the credibility of these approaches. It is important to note that for heterogeneous systems, the agreement between the two methods occurs only when particle tracking is based on the Pollock method.

3.8 Future steps

The above results show that relying solely on single-run particle tracking for CZD for complex systems is not recommended. Instead, particle tracking should always be used within the framework of a Monte Carlo procedure (see Enzenhoefer et al. 2014, Meyer et al. 2014). The Monte Carlo approach could be extended through methods such as Bayesian model averaging (Hoeting et al. 1999) to account for the between-model differences encountered in this current study.

In addition, combining Monte Carlo methods (Enzenhoefer et al. 2014) with scenario analysis (Sousa et al. 2013) and applying reverse transport on the resultant flow fields could address both parameter and conceptual uncertainties. In this case, filtering for model realizations would be necessary, only using realizations that provide an acceptable calibration (Beven and Binley 1992, Kavetski et al. 2003). The ensemble of capture zones for each model could then be merged (Sousa et al. 2013) or averaged (Ye et al. 2010). Comparing the resultant capture zones and their sensitivity to different types of uncertainty would be of interest in informing hydro(geo)logic modelling methods.

Another relevant direction would be to study the effects of spatial and temporal scaling in the forms of regional flow and transience on the stream capture zone. Veale et al. (2014) have shown that the source of baseflow can be regional for larger surface water courses, such as rivers. The limits of the current study could be expanded to include the entire Waterloo Moraine in order to investigate impacts of regional baseflow to larger surface water courses. Seasonal transience is somewhat important in the context of well capture zones; however, it could be important in the context of stream capture zones. By combining transient drivers (Rodriguez and Nowak 2018) to stream CZD, probabilistic year-to-year variations may also be considered. This would require flow, particle tracking, and reverse transport to be fully integrated in one model. At the end, a steady capture zone could be delineated on the basis of a transient analysis using the precautionary principle as outlined by Sousa et al. (2013) or based on a frequency analysis as in Rodriguez and Nowak (2018).

In addition to particle tracking and reverse transport, groundwater age simulation as proposed by Molson and Frind (2012) could be applied to the delineation of the stream capture zones. In this approach, a conceptual “age” mass that ages at a rate of one day per day is transported through the system. As the life expectancy of a particle of water (i.e. the time remaining before capture) is zero at the well, increasing with up-gradient distance from the well, the approach can be used to generate a map of life expectancy for the water particles approaching a well (here the stream) under local-scale parameter uncertainty. The result would be a map of groundwater time-of-travel zones with respect to the stream, which would be immediately useful for land-use planning as well as the implementation of protective measures near the stream.

3.9 Conclusions

My main motivating research question (RQ1, Section 1.5) for this study was: does numerical model code choice lead to uncertainties in the delineation of baseflow contribution for gaining stream reaches? My results have shown that areas contributing baseflow to streams can in principle be delineated using wellhead protection methodology. However, the delineation can be affected by both the choice of groundwater model and the CZD method. This multi-model study provides insight into some of the model-choice related uncertainties that arise in the delineation of stream capture zones. Different models can yield different results due to different parameterization, discretization schemes, and particle tracking algorithms.

A critical component for delineating stream capture zones is the exchange flux at ground surface, which can be obtained using integrated sw-gw models such as HGS. The model should be finely discretized near the top boundary to reduce numerical errors when calculating the exchange flux. After obtaining the exchange flux, the same approaches that are used for the delineation of well capture zones can be used for gaining streams. By exploring different approaches, the study provides a first basis for developing protective measures to maintain the quantity and quality of groundwater discharge to sensitive streams.

The results show that one of the largest sources of deviation in CZD is the particle tracking algorithm, followed by the discretization scheme and the parameter distribution. This means that capture zone results for complex heterogeneous systems are highly sensitive to the particle tracking algorithm.

The results also show that in general, for complex systems, the reverse transport approach can be less subjective than particle tracking. Reverse transport can be combined with particle tracking used as a screening tool. Agreement between particle tracking and reverse transport results can be seen as reinforcing the validity of the baseflow contribution area delineated. It is important to note that, for heterogeneous systems, the agreement between the reverse transport and reverse particle tracking (with random walk to account for dispersion) approaches only occurs when particle tracking is based on the Pollock method or other sufficiently exact particle tracking methods, which takes the change in flow direction at internal interfaces into account.

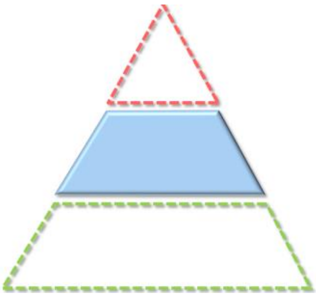
In response to RQ1 formulated in Section 1.5, three recommendations for delineating baseflow contribution areas for streams can be drawn from the conclusions of this study:

1. With finer discretization of the near subsurface and a fully integrated surface water – groundwater model, the same methods of CZD methods for wells can be applied for gaining streams.
2. For heterogeneous systems with internal material interfaces, the Pollock method for particle tracking should be chosen over the 4th order Runge-Kutta method.
3. CZD using particle tracking should be combined with a Monte Carlo approach or reverse transport to address parameter uncertainty.

In practice, multi-model analysis may not always be feasible. In view of this limitation, it is important that practitioners be aware of the uncertainties that can be expected when relying on a single conceptual model or a single software package. Uncertainty in groundwater investigations remains a fact of life, and particularly so in CZD.

Results from this chapter support the use of HGS when simulating sw-gw exchange in the following chapters because of its capability for fully-coupling sw-gw flow. Furthermore, the conclusions from this chapter suggest that if particle tracking is required (Chapters 4 and 5) then it is advisable to use a large number of particles (i.e., 1000's) and that the preference should be to use the Pollock method over the 4th order Runge-Kutta method (however future studies requiring particle tracking should use the method by Selzer and Cirpka 2019). Finally, when possible, advection-dispersive transport (Chapter 6) should be used over particle tracking because it can more explicitly account for local-scale uncertainty.

4. Sensitivity of simulated meander-scale hyporheic exchange to river bathymetry[§]



Chapter highlights

- Simulating hyporheic exchange with a high-resolution detailed bathymetry using a 3D fully coupled model leads to nested multi-scale hyporheic exchange systems.
- A poorly resolved bathymetry will underestimate bedform-scale hyporheic exchange, biasing the simulated hyporheic exchange towards the meander-scale, thus leading to overestimates of hyporheic exchange residence times.
- Incorporating local bedforms will likely capture the nested nature of hyporheic exchange, leading to more physically meaningful simulations of hyporheic exchange fluxes and transit times.

4.1 Motivation and research questions

As discussed in Section 1.3 of the introduction, developing methods to simulate hyporheic exchange at relevant spatial and temporal scales can help us better understand this complex physical phenomenon and provide us with tools to plan ecological restoration projects. I chose to focus on modelling hyporheic exchange at the hydrostatic-driven meander-scale (m to km) because the associated transit times (hours to days) are relevant for anaerobic microbial

---

[§] The contents of this chapter are modified from 'Chow, R., Wu, H., Bennett, J.P., Dugge, J., Wöhling, T., and Nowak, W. (2018). Sensitivity of Simulated Hyporheic Exchange to River Bathymetry: The Steinlach River Test Site. Groundwater, https://doi.org/10.1111/gwat.12816'

Author contribution statement: ***Chow, R.***: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript addressing peer-review comments. Finalized and approved published article. Analyzed and evaluated model results. Conceptualized modelling study. ***Wu, H.***: Developed results and figures for original draft manuscript. Preliminary results presented in University of Stuttgart Master's thesis (Wu 2017). ***Bennett, J.P.***: Developed conceptual and Multiple Point Geostatistics (MPS) model. Helped write MPS section in original draft manuscript. ***Dugge, J.***: Developed preliminary Steinlach River Test Site HGS model. Conceptualized modelling study. ***Wöhling, T.***: Reviewed original draft manuscript. Conceptualized modelling study. Provided supervision. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

processes such as denitrification (Zarnetske et al. 2011, Boano et al. 2014). One recent example in a southwest Ontario watershed, shows meander-scale hyporheic exchange to be the predominant driver of acesulfame (an artificial sweetener acting as a human-specific conservative tracer that comes from wastewater treatment plants) into near-stream shallow groundwater (Hwang et al. 2019).

One model aspect identified as contributing significantly to the predictive uncertainty of hyporheic exchange is the accurate representation of river bathymetry (Cardenas et al. 2004). Pryshlak et al. (2015) compared the hyporheic exchange fluxes between simulations with homogeneous hydraulic conductivity (*K*) and abrupt *K* transitions, concluding that characterizing *K* was more critical than channel morphology for predicting hyporheic exchange. In contrast, Tonina and Buffington (2007) have shown with both laboratory experiments and numerical modelling that the major mechanism for hyporheic exchange was bedform-induced advection. Stonedahl et al. (2013) linearized the effects of different topographic features (i.e., meanders, bars, and dunes) and found that dunes typically contributed more to hyporheic exchange than other topographic features.

Thus, the exact role of bathymetry when simulating meander-scale hyporheic exchange (fluxes and transit times) is unclear because changes in bathymetry simultaneously change water stages in the river and the geometry of the hyporheic aquifer. The goals of this study (Section 1.5) are to determine the role river bathymetry plays when simulating hyporheic exchange at the meander-scale in a physically meaningful way, and to determine which aspects of bathymetry are critical so that simplifications of bathymetry do not affect hyporheic exchange too dramatically.

The principal research question (RQ2, Section 1.5) for this study is: what are the most important aspects of river bathymetry that need to be captured in order to capture the scales of exchange appropriately? To make this principal question more answerable and tractable, I have broken it down to three sub-questions:

1. Is bathymetry important when simulating meander-scale hyporheic exchange?
2. What errors occur when we lack bathymetry data?
3. What aspects of bathymetry matter the most and which can be neglected in order to simulate meander-scale hyporheic exchange in a physically meaningful way?

4.2 Approach

In order to better understand the role bathymetry plays in simulating hyporheic exchange, I first created a highly parameterized, highly resolved, fully coupled sw-gw model of a river meander with detailed river bathymetry and then perform a numerical analysis of various simplified models

I chose the Steinlach River Test Site as a basis for this modelling study because of the extensive site investigations characterizing its hydrogeology and hyporheic exchange (Doro et al. 2013, Osenbrück et al. 2013, Voronina 2013, Wöhling et al. 2016). Informed by previous field investigations and natural field-based tracer experiments (Osenbrück et al. 2013), the model is constrained. This model is then treated as a reference solution or ‘virtual reality’. Next, the model is rerun with simpler bathymetries and compared to the virtual reality model. This lets us identify the river bathymetry features that have the greatest influence on the predictive biases when simulating meander-scale hyporheic exchange using a simplified representation of bathymetry.

Based on finding in Chapter 3, to obtain a physically consistent two-dimensional distribution of river stages and the associated distribution of exchange fluxes, the fully coupled sw-gw model HydroGeoSphere (Aquanty Inc. 2015) was used to simulate meander-scale hyporheic exchange as a function of realistic river bathymetry geometries. More detail on this choice is provided in Section 4.3.1.

It is important to note that the purpose of this study is not to calibrate a model of the Steinlach River Test Site. This is not necessary in order to investigate the sensitivity of simulated hyporheic exchange to river bathymetry. Instead, I created a reality-inspired synthetic model as a benchmark for my sensitivity analysis to ensure my conclusions are in line with site observations and the site conceptual model developed by Doro et al. (2013) and Osenbrück et al. (2013). Additionally, I do not determine all factors affecting hyporheic exchange and try to rank them (i.e., by a global parametric uncertainty analysis). Instead, the aim is to determine the level of bathymetry detail necessary to simulate hyporheic exchange in a physically meaningful way. For additional information on the effects of subsurface parameter structure uncertainty when modelling meander-scale hyporheic exchange, see Chapter 5.

As a metric for comparing different scenarios, I consider the hyporheic exchange fluxes in various sub-reaches of the study site and the hyporheic transit time distribution through the river meander. I hypothesize that the key features to be preserved in simplified models are the longitudinal profile of the river bed as well as the relative location of the deepest and highest points within the transverse profile, because the main control on the meander-scale hyporheic exchange should be the longitudinal profile, while the local-scale contrasts should control bedform-scale exchange, which together should produce more physically meaningful proportions of nested multi-scale hyporheic exchange (Figure 1.2).

4.3 Modelling hyporheic exchange at the Steinlach River Test Site

4.3.1 Site description and previous studies

The Steinlach River Test Site is located at the lower reach of the Steinlach River, a tributary of the Neckar River, which belongs to the Rhine basin in southwest Germany (Figure 4.1). The site is approximately 0.01 km^2 in area and consists of a river meander with a relatively steep hydraulic gradient (≈1.2 %) along the valley. Site characterization and monitoring of sw-gw exchange began in 2010. Osenbrück et al. (2013) used hydraulic, chemical, and isotopic monitoring to confirm that meander-scale hyporheic exchange was occurring and estimated the hyporheic exchange fluxes and residence times. Doro et al. (2013) characterized the subsurface heterogeneity of the hyporheic zone and developed a conceptual hydrogeological model of the site based on hydrogeological and geophysical methods. The benefits of geophysical methods to reduce nonuniqueness in sw-gw models have been demonstrated by Ward et al. (2013).

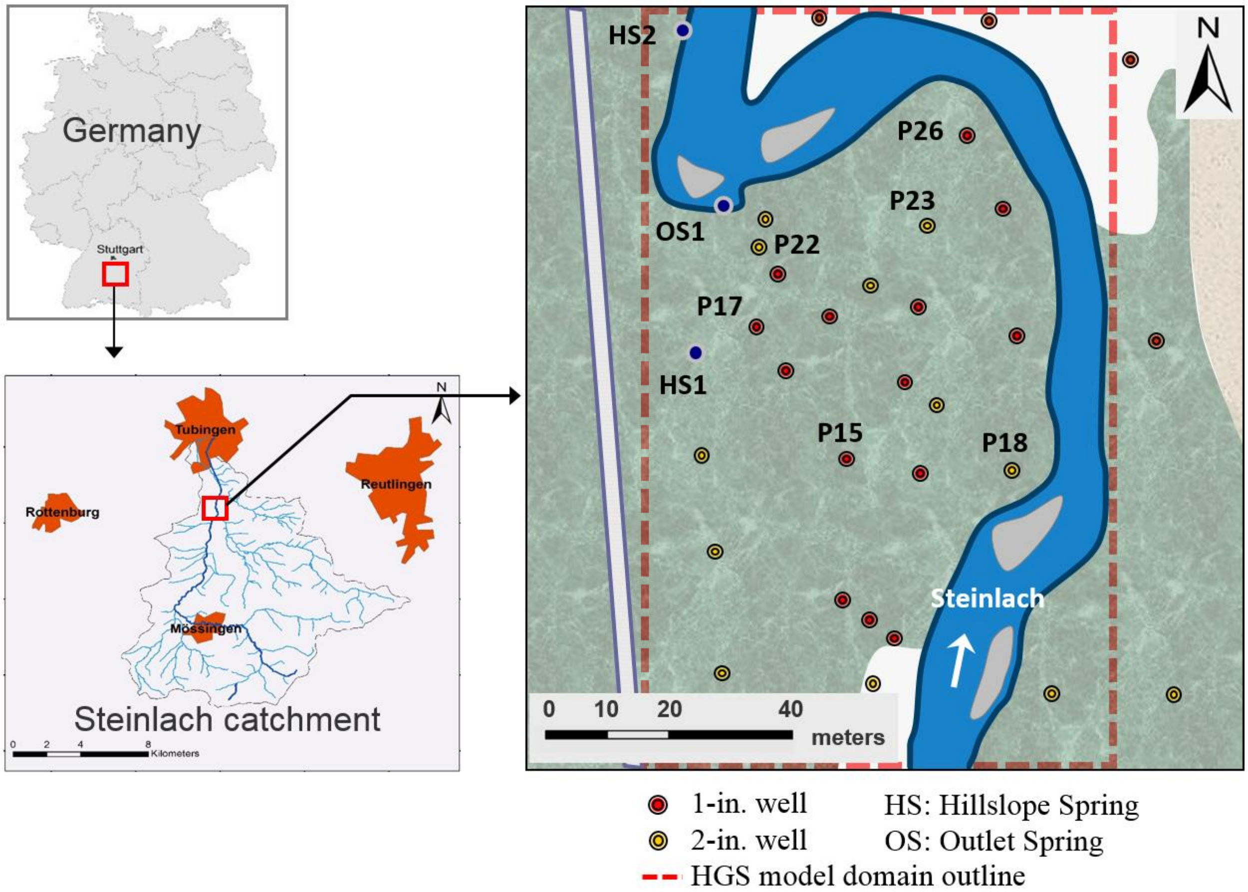


Figure 4.1: Location of Steinlach River Test Site showing installed piezometers, natural springs, and model domain outline (adapted from Osenbrück et al. 2013).

The depositional units present at the site are a topsoil unit and a sandy gravel (or fluvial) deposits, separated by a sharp contact. The topsoil unit consists of fine alluvial material with a high organic-carbon and clay content; these are interpreted as overbank fines. The fluvial deposits are heterogeneous, ranging from fine sand and silt lenses to coarse gravels, which is common for gravel-bed rivers. The fluvial gravel deposits are underlain by sandstone of the Löwenstein formation, belonging to the Upper Triassic. Since the hydraulic conductivity of these sandstones are much lower than the fluvial deposits, they are regarded as an aquitard for the purpose of this study.

Borehole information was classified from Doro et al. (2013) into five hydrofacies: topsoil, silty sand, sandy gravel, gravel, and sandstone (Table 4.1).

Table 4.1: Hydrofacies and associated proportions at the Steinlach River Test Site

	Hydrofacies	Proportion of logged core [%]	Mean K (m/s)
1	Topsoil	23	$1 \cdot 10^{-4}$
2	Silty sand	12	$5 \cdot 10^{-4}$
3	Sandy gravel	46	$1 \cdot 10^{-3}$
4	Gravel	10	$5 \cdot 10^{-3}$
-	Sandstone	8	-

Wöhling et al. (2016) created a MODFLOW model of the Steinlach River Test Site in order to simulate hyporheic exchange, estimate its transit time distributions, and extend a method for optimal sensor placement in groundwater monitoring networks. This MODFLOW model served as a conceptual basis (e.g., model domain limits and boundary conditions) for the modelling conducted in this study.

In the standard streamflow packages of MODFLOW (i.e., RIV, STR1, SFR1, and SFR2) streamflow is computed as a one-dimensional problem, implying a uniform river stage over the width of the river. I deemed this to be inadequate for this river, where bathymetry varies both in the main direction of the river and perpendicular to it. The latter causes two-dimensional variations of river stage that can only be simulated with river-hydraulics models that are at least two-dimensional. The two-dimensional variations of river stage in turn drive hyporheic exchange. For this reason, I decided to transfer the Steinlach River Test Site model to HGS (Section 2.1), a fully coupled sw-gw model. Furthermore, HGS offers a flexible discretization scheme for refinement along the river channel in order to incorporate detailed river bathymetry. Although three-dimensional river models with appropriate closure schemes for turbulence would give even better pressure distributions at the river bed (Trauth et al. 2013), the associated computational effort was assessed to be too high.

4.3.2 Steinlach River Test Site HydroGeoSphere model

As noted in Section 2.1, the advantage of HGS is that it calculates the spatial distribution of gaining and losing river reaches within the channel at high resolution without the need to manually prescribe these areas as a boundary condition to the groundwater flow system a priori.

The domain was discretized using a mesh of triangular prisms (Figure 4.2a). The element edges are about 20 cm long in the river and banks allowing us to incorporate bathymetric details consistent with the resolution of the bathymetry data (Osenbrück et al. 2013). The elements are discretized progressively larger away from the river, and are up to 1.5 m long near the edges of the domain. The model domain area is approximately 0.01 km^2 (85 m x 160 m) with a total of 176,778 nodes and 332,920 elements. Vertically, the domain is split into 20 elemental layers, increasing in thickness with depth. The top layer has a minimum thickness of 0.01 m and the bottom layer has a maximum thickness of 0.5 m. This level of discretization was found to provide a sufficiently precise solution for a meander-scale model, while maintaining reasonable computational costs.

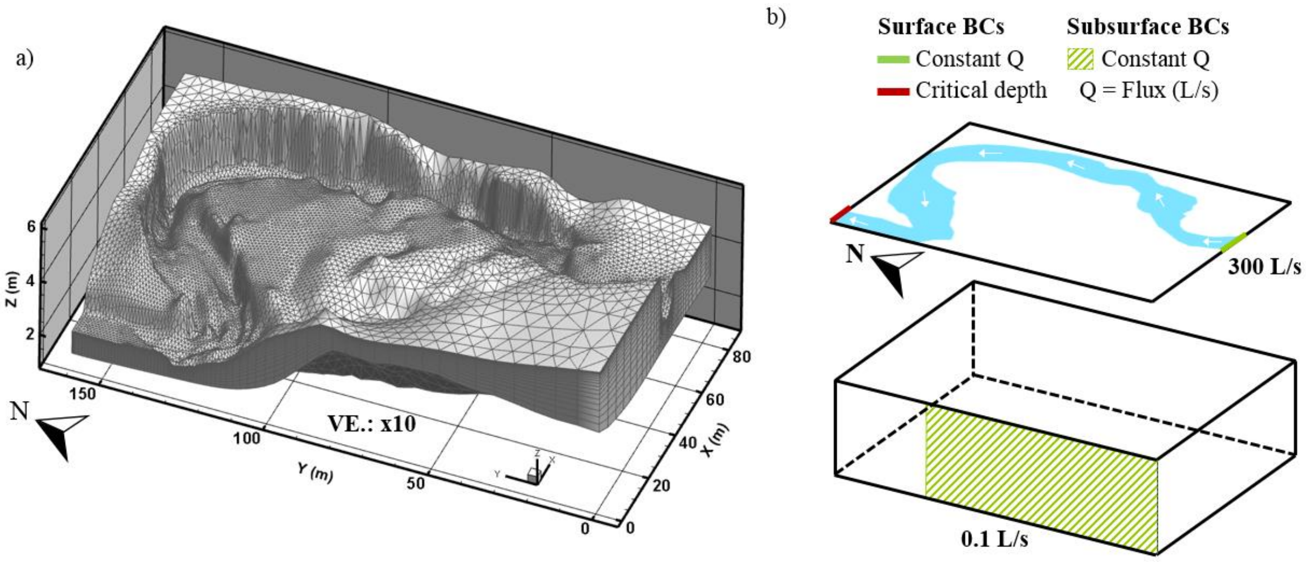


Figure 4.2: Base Case (a) model discretization and (b) boundary conditions (Chow et al. 2018).

Next, the upper boundary of the model was generated from a topographical survey conducted in 2011 and 2012 (Osenbrück et al. 2013). The topographical measurements on land and in the river were then interpolated separately due to the different characteristics and data densities of the two data sets. For the land surface measurements, I used an isotropic bicubic interpolation method. In the river, 540 bathymetric survey points were transformed to a flow-oriented coordinate system, following the approach by Merwade et al. (2005), interpolated them using an inverse-distance-weighting method, and transformed them back to the Cartesian coordinate system. Interpolation was conducted using research codes written in MATLAB.

Finally, the lower boundary of the model was set as a no-flow boundary, representing the contact between the fluvial deposits and the underlying sandstone, which was considered as an aquitard (Doro et al. 2013). I generated the contact surface using the Leapfrog 3D geological modelling software (ARANZ Geo Limited 2017), based on borehole logs and inferred geological trends (Doro et al. 2013).

Figure 4.2b shows the model boundary conditions. At the upstream inlet of the surface domain, a constant inflow rate of 300 L/s was specified based on the observed modal flow rate (Voronina 2013). A subsurface specified flux of 0.1 L/s was defined along the western side of the model domain to account for estimated groundwater fluxes originating from the hillslope (Osenbrück et al. 2013). The downstream outlet in the surface domain was set as a critical depth boundary, which allows water to exit freely without a user-specified rate. A default no-flow boundary was considered at all other sides of the model domain.

4.3.3 Virtual reality model parameterization

The spatial distribution of hydraulic parameters within the Steinlach Test Site was generated in a hierarchical 3-step process, according to the two different depositional environments present at the site. First, the Leapfrog 3D geomodelling software (ARANZ Geo Limited 2017) was used to interpolate unit boundaries (i.e., topsoil and fluvial) based on borehole logs and bank profile information from Doro et al. (2013).

Second, the impala multiple-point geostatistical (MPS) algorithm (Straubhaar et al. 2011) was used to generate the subsurface distribution of hydrofacies, which is a method that can generate hydrofacies consistent with site data and represents the variability, connectivity, and structural properties present in a given training image (Mahmud et al. 2014). The SBED reservoir simulation software (Geomodeling Technology Corp. 1996) was used to generate training images for the MPS simulation. Two training images were generated: one for the topsoil unit, which consisted of horizontal layers, and another for the fluvial unit (Figure 4.3), which consisted of sinusoidal hydrofacies elongated in the assumed direction of paleoflow (congruent with the current surface flow regime). Borehole data (Doro et al. 2013), sequence boundaries, global hydrofacies proportions (Table 4.1), and training images were loaded into impala to generate realizations of hydrofacies distributions for the model. Additional detailed steps in creating the virtual reality model and the impala files used to generate the multiple-point geostatistical realizations are in Appendix B.

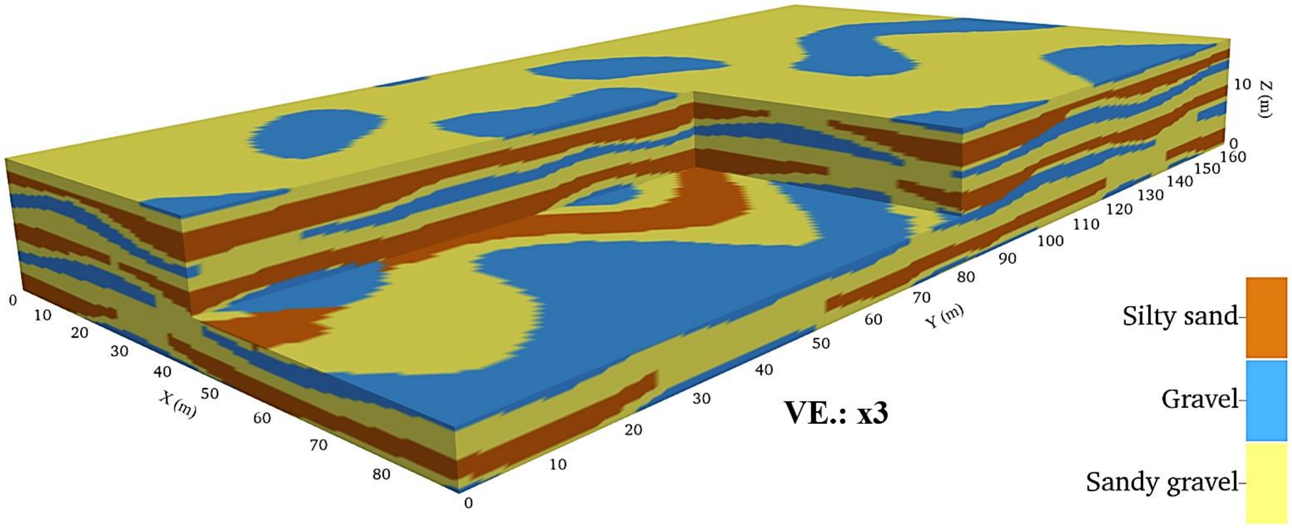


Figure 4.3: Training image of the fluvial unit generated using SBED reservoir simulation software (Geomodeling Technology Corp. 1996). Realizations varying the hydrofacies are based on multivariate distributions inferred from this training image. The sinuosity was derived from analyses of the Steinlach River channel morphology, as well as a site conceptual model informed by geophysical and hydrogeological methods (Doro et al. 2013).

Third, local-scale internal heterogeneity within each hydrofacies was represented by assigning a multi-Gaussian random field for log hydraulic conductivities, and then used a numerical spectral method (Dietrich and Newsam 1993) to generate the corresponding random fields inside the hydrofacies. A Gaussian covariance function was chosen to model the correlation structure of the random fields:

$$C_{ln\,K}(\boldsymbol{r}) = \sigma_{ln\,K}^2 exp(-r'^2)$$

with $r' = \sqrt{\left(\frac{r_x}{\lambda_x}\right)^2 + \left(\frac{r_y}{\lambda_y}\right)^2 + \left(\frac{r_z}{\lambda_z}\right)^2}$, where λ_x, λ_y, and λ_z are the correlation lengths in the x, y and z directions, respectively, $\sigma_{ln\,K}^2$ is the variance of log K, and r_x, r_y, r_z are the separation distances in the x, y, z directions, respectively. The mean K-values for each hydrofacies are from K-ranges derived from a combination of slug and pumping tests (Doro et al. 2013) while accounting for scale effects (Bradbury and Muldoon 1990). Based on estimates of the average extents of the hydrofacies (Doro et al. 2013), the variance σ_{lnK}^2 of log K-values was set to 0.5 and the correlation lengths λ to 1 m and 0.1 m in the horizontal (x,y) and vertical (z) directions for all hydrofacies, respectively. The effective porosity was assumed to be homogeneous at $\eta = 25\,\%$. While the topsoil porosity may have a different porosity than the fluvial river sediments, the hyporheic exchange occurs solely within the saturated river sediments so that the topsoil porosity is of minor relevance for these model simulations.

The riverbed within the Steinlach River Test Site has been described as having minimal in-channel vegetation, steep banks, and a bottom composed predominantly of gravels and cobbles (Doro et al. 2013, Osenbrück et al. 2013). Therefore, a constant Manning's roughness coefficient of $n = 0.02\ \mathrm{s/m}^{1/3}$ was chosen for the river reach according to the classification by Chow (1959). HGS does not require an explicit riverbed conductance value (i.e., MODFLOW streamflow packages), instead HGS uses the hydraulic conductivities assigned to the uppermost elements within the river to compute exchange fluxes. The riverbed sediments were assumed to have the same hydrogeological properties of the fluvial aquifer that were generated with the 3-step process outlined above. This assumption is further discussed in Section 4.4.4.

To obtain the virtual reality model, I selected a single impala realization of the hydrofacies distributions (out of 100 realizations) with local-scale internal heterogeneity (Figure 5.2c) based on its initial fit to 14 head observations (Wöhling et al. 2016), as well as the meander-scale hyporheic exchange fluxes and transit times (Osenbrück et al. 2013). The mean K-values for the hydrofacies were then manually fine-tuned to better agree with this data in an effort to make this virtual reality model more physically meaningful. The final mean K-values for the virtual reality model were $1 \cdot 10^{-2}\ m/s$ for the gravel, $2 \cdot 10^{-3}\ m/s$ for sandy gravel, $4 \cdot 10^{-4}\ m/s$ for silty sand, and $1 \cdot 10^{-4}\ m/s$ for topsoil hydrofacies, respectively.

While this approach has limitations because the information content of the available data is not sufficient for resolving local-scale aquifer heterogeneity (Moore and Doherty 2005), the described method is nonetheless useful as it constrains the model parameters with a dataset that has a significant bearing on predicting hyporheic exchange fluxes and transit times. Furthermore, this approach is based on geologically reasonable assumptions about the larger-scale features, which are simulated by MPS using facies-emulating training images (Figure 4.3). Since I am evaluating the relative impact of different bathymetries and subsurface parameter structures on the predictive uncertainty, it is unnecessary to achieve a perfect calibration of our virtual reality model to the field data.

4.3.4 Simulated vs. observed hyporheic exchange

Similar to Cardenas et al. (2004) and Pryshlak et al. (2015), hyporheic exchange was simulated using simple advective particle tracking, neglecting diffusion and dispersion. This allows us to derive unique hyporheic exchange transit time cumulative distribution functions for different bathymetry scenarios, while simultaneously allowing us to explicitly map the origins of the meander-scale hyporheic exchange. Particle-tracking methods have the advantage that they do not introduce numerical dispersion, whereas Eulerian transport schemes, such as the finite element method implemented in HGS, introduce numerical dispersivities that scale linearly with the grid resolution. WATRAC (Section 2.2, Frind and Molson 2004) was chosen as an offline particle tracker for HGS, which is based on the semi-analytical method by Pollock (1988).

The HGS steady-state hydraulic head distribution was used as the flow field for particle tracking. Initial particle placement was within the river bed upstream of the meander every 0.5 m in both the x- and y-directions for a total of 6,613 particles. Only particles that downwelled into the shallow aquifer and upwelled to the surface downstream of the meander were tracked forward in time with the advective flow field. Figure 4.4a shows the simulated particle trajectories through the Steinlach Test Site hyporheic zone and the corresponding particle transit times.

The hyporheic exchange flux represented by each particle was computed using the initial downward elemental pore water velocities interpolated by WATRAC. The fluxes were the product of the advective velocities, elemental area, and porosity, divided by the number of particles starting within each element. Each particle was assumed to carry a constant amount of flux throughout its travel path and that mass balance errors using interpolated WATRAC velocities were insignificant. These particles were then post-processed using their flux-weighted travel times to compute a cumulative distribution function (CDF) of transit times.

The simulated hyporheic exchange for the calibrated model compares well with observed fluxes and transit times. Osenbrück et al. (2013) estimated hyporheic exchange fluxes at the downstream outlet spring (OS1 in Figure 4.1) to be between 1-2 L/s with an average transit time of approximately 7.8 days. The simulated hyporheic exchange flux was 1.7 L/s and the median transit time was 7.8 days. This HGS model serves as the virtual reality model of the Steinlach River Test Site and is subsequently referred to as the ‘Base Case’.

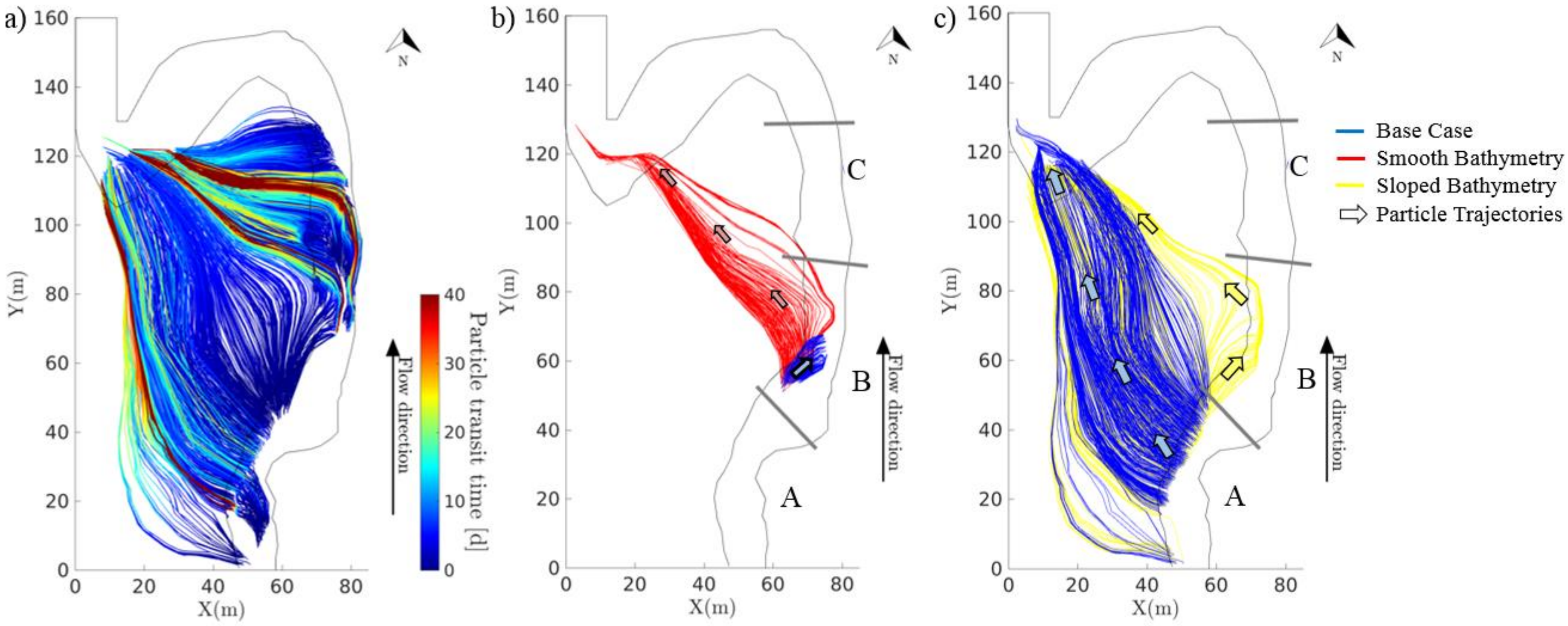


Figure 4.4: Particle trajectories contributing to meander-scale hyporheic exchange. a) Base Case b) difference between Smooth Bathymetry and Base Case with arrows indicating their general directions, Domain B particles that contribute to meander-scale hyporheic exchange in the Smooth Bathymetry contribute to bedform-scale hyporheic exchange in the Base Case; c) difference between Sloped Bathymetry and Base Case with arrows indicating their general directions, Domain A particles take a longer convoluted travel path in the Sloped Bathymetry compared to the Base Case due to a weak-sink effect in Domain B (Chow et al. 2018).

4.4 Effects of simplified bathymetry on simulated hyporheic exchange

To determine the sensitivity of simulated hyporheic exchange to river bathymetry, the detail in bathymetry is systematically degraded by sampling less bathymetric elevations from the Base Case to create a simplified bathymetry. Next, HGS was rerun with the simplified bathymetry to generate the steady-state river stage and hydraulic head distribution, which is then used to recalculate the hyporheic exchange fluxes and transit times. Three simplified bathymetry scenarios are considered:

1. Smooth: Constant slope from river inlet to outlet and devoid of local bedforms.
2. Sloped: Follows river slope but still devoid of local bedforms.
3. Thalweg-Gegenweg: Interpolated from bathymetric minimum and maximum every 9 m.

Figure 4.5 shows the bathymetries for the Smooth, Sloped, and Thalweg-Gegenweg (T-G) cases, including: a plan-view contour plot of the river bathymetries, vertical profiles along the centerline of the river, bathymetric cross-sections, and the bathymetric measurements taken from the Base Case to generate the bathymetry. Figure 4.6 and 4.7 shows the resulting change in the river stage and exchange fluxes, respectively, due to the changes in bathymetry. Figure 4.8 shows the subsequent change in the subsurface head distributions. The results of these three simplified scenarios are discussed, one by one, in the following sections.

4.4.1 Smooth Bathymetry

This case simplifies the river bathymetry by using only two bathymetric measurements, one point at the river inlet and the other at the outlet. I used a slope that was approximately constant in the flow-oriented direction, with all variability from local bedforms in the river cross-section removed. The slope is not perfectly constant due to the transformations from the flow-oriented to Cartesian coordinates and the merging of the river bathymetry to the land surface, however efforts were made to make the bathymetry as smooth and constant as possible. Figure 4.5a shows the resulting ‘Smooth Bathymetry’.

Figure 4.6a shows the elevation contours for the Base Case river stage and land surface, while Figure 4.6b shows the difference in the river stage with the Smooth Bathymetry. Here it is evident that the largest changes occur along the sides of the river, with the exception of a bathymetric high that is removed after the (i.e.,$x = 25\,m, y = 130\,m$) that is located in the middle of the river. Figure 4.7b shows the river exchange flux (exfiltration/infiltration rate) for the Smooth Bathymetry. Figure 4.8b shows the change in the underlying head distribution relative to the Base Case. Here we can see that the largest head increase occurs between 60 to 100 m in the y-direction and the largest head decrease occurs where the bathymetric high was removed (i.e., $x = 30\,m, y = 135\,m$); from this result we would expect a larger head gradient between these two areas.

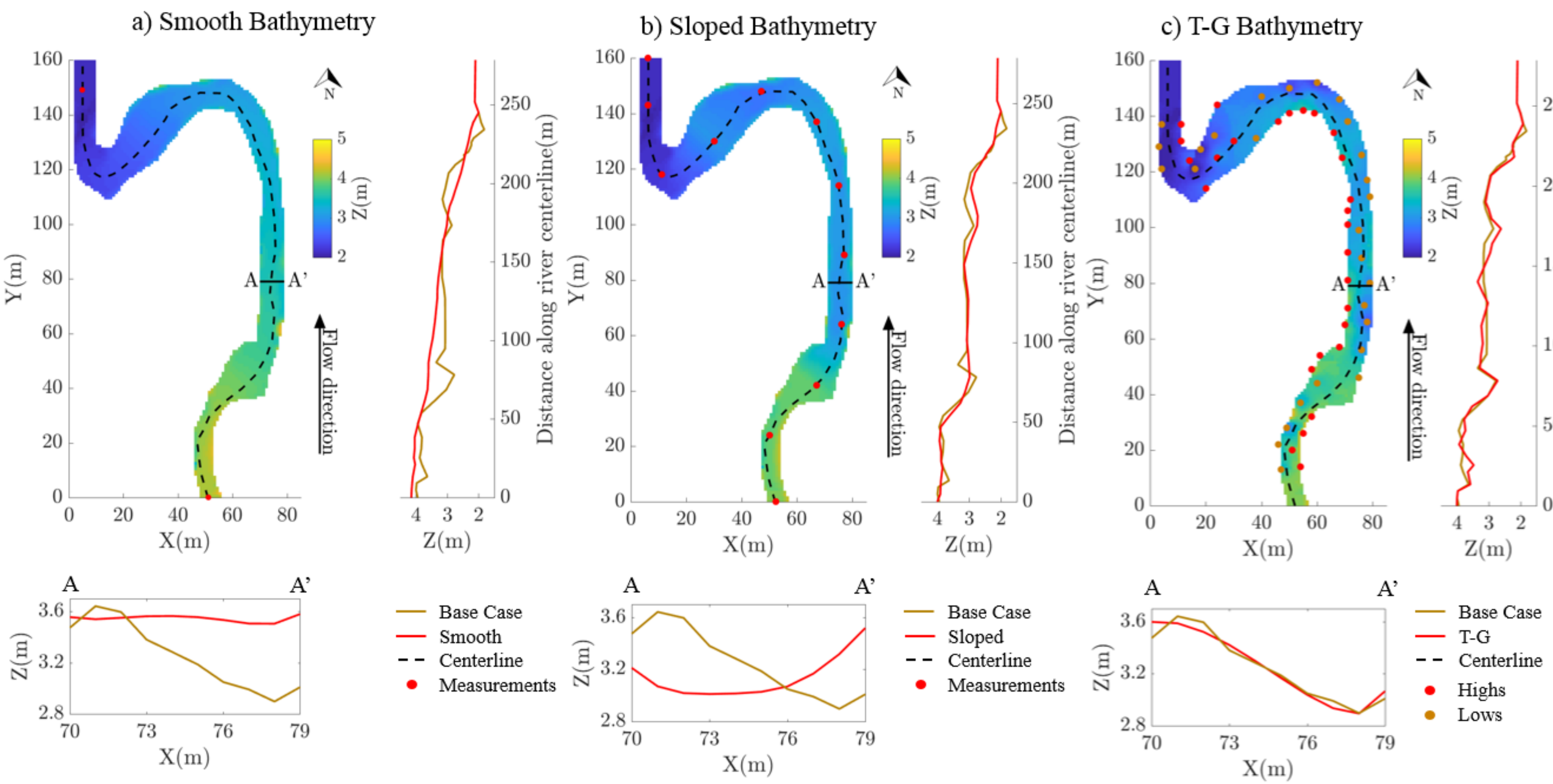


Figure 4.5: Summary of bathymetries: a) Smooth, b) Sloped, c) Thalweg-Gegenweg. Each includes a plan-view contour plot of the river bathymetry, a vertical profile along the centerline of the river (vert. exag.: x20), a cross-section (vert. exag.: x5), and the locations of bathymetric measurements from the Base Case bathymetry used to create the bathymetry (Chow et al. 2018).

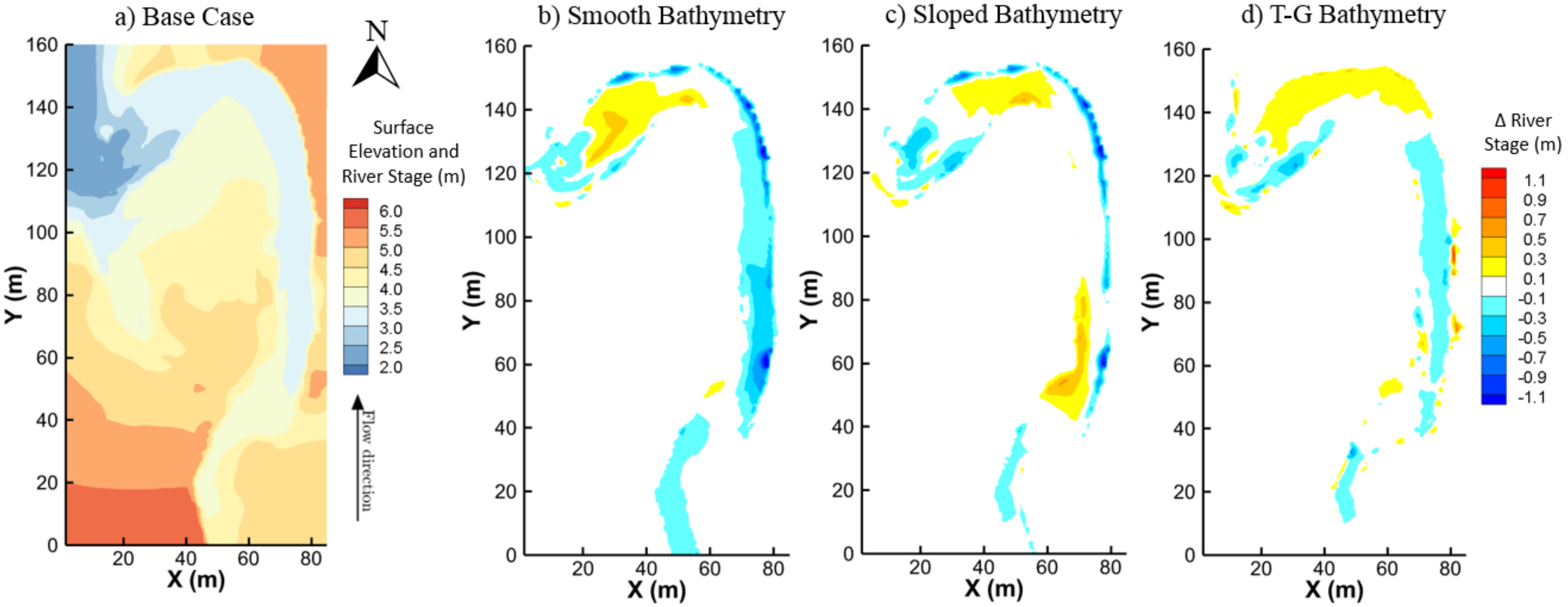


Figure 4.6: River stage differences from bathymetry scenarios: a) Base Case surface and river stage elevation. Difference in river stage elevation: b) Smooth Bathymetry c) Sloped Bathymetry d) Thalweg-Gegenweg (T-G) Bathymetry. River stage differences are calculated relative to the Base Case, warm colours indicate that the Base Case elevation was greater and cool colours indicating the opposite (Chow et al. 2018).

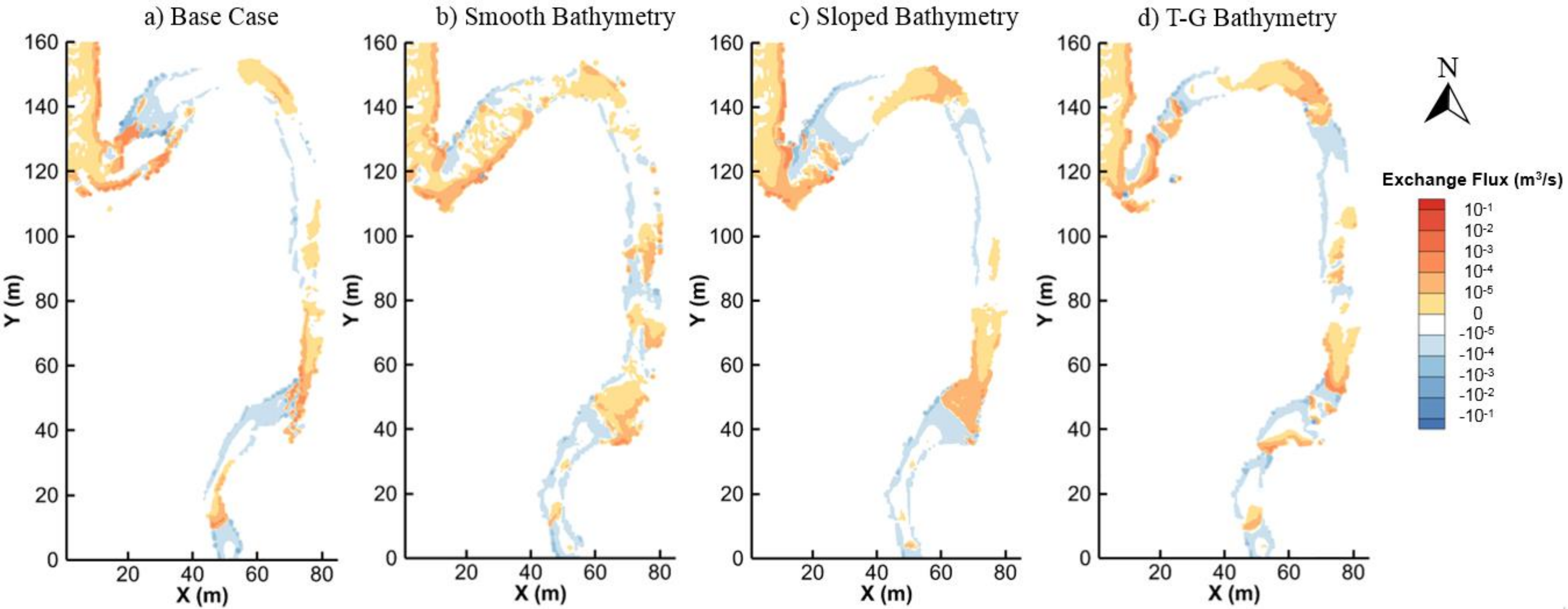
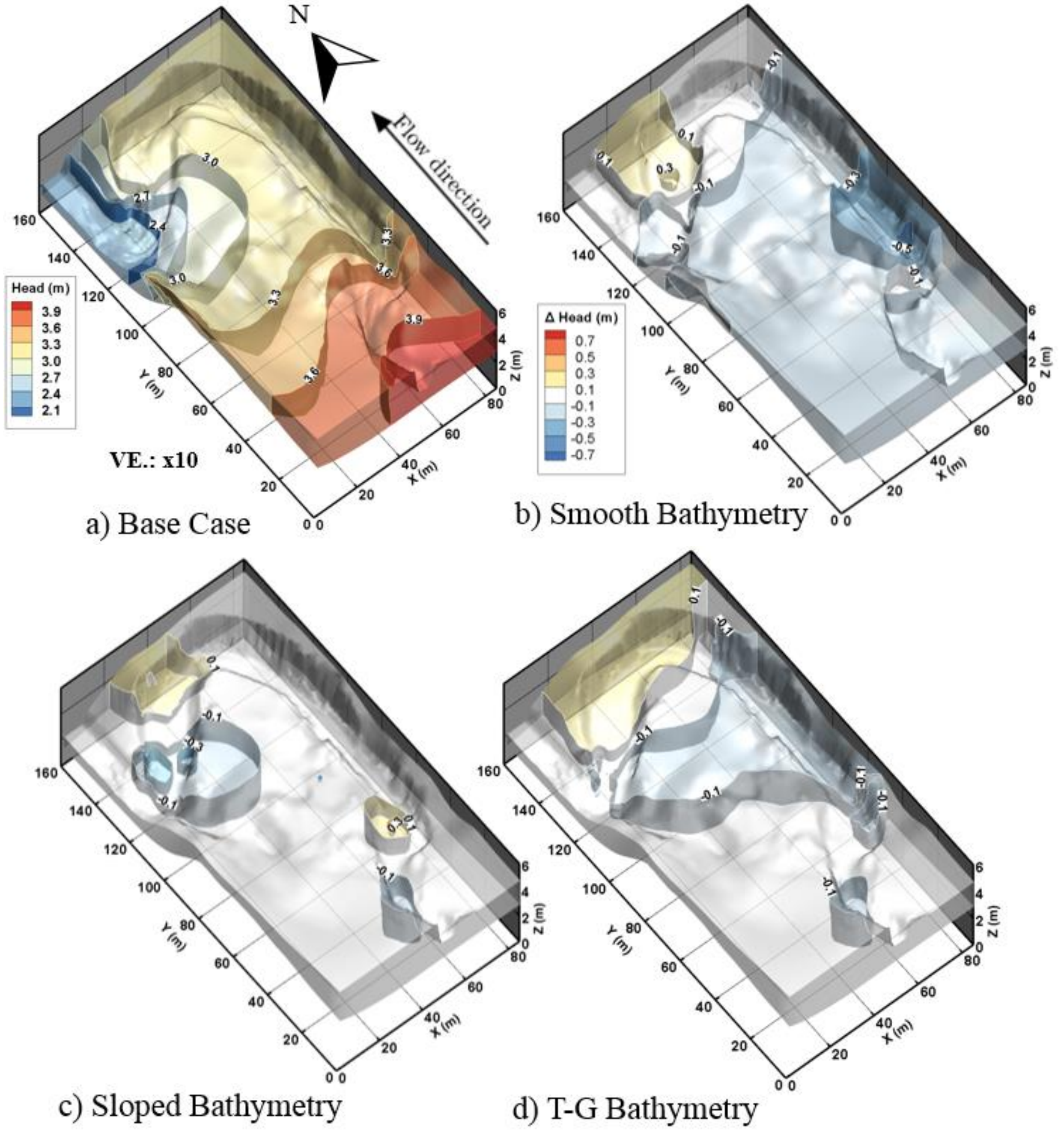


Figure 4.7: Spatially distributed river exchange flux: a) Base Case b) Smooth Bathymetry c) Sloped Bathymetry d) Thalweg-Gegenweg (T-G) Bathymetry (Chow et al. 2018). Exfiltration (warm colours) and infiltration (cool colours)



a) Base Case

b) Smooth Bathymetry

c) Sloped Bathymetry

d) T-G Bathymetry

Figure 4.8: Hydraulic head differences from bathymetry scenarios: a) Base Case 3D subsurface hydraulic head distribution. 3D difference in hydraulic head distributions: b) Smooth Bathymetry c) Sloped Bathymetry d) Thalweg-Gegenweg (T-G) Bathymetry. Hydraulic head differences are calculated relative to the Base Case, warm colours indicate that the Base Case elevation was greater and cool colours indicating the opposite (Chow et al. 2018).

Figure 4.10 summarizes the meander-scale hyporheic exchange fluxes and transit time CDFs for all bathymetry cases, respectively. The total amount of meander-scale hyporheic exchange increased by 27 % from 1.7 L/s to 2.1 L/s from the Base Case to the Smooth Bathymetry, respectively. In order to understand this change in hyporheic exchange flux, the upstream segment was subdivided into three domains of approximately equal area (A, B, and C) to quantify the origins of this flux (Figure 4.9a).

Figure 4.9b - ‘Smooth’ indicates that, for the Smooth Bathymetry, the origins of the exchange flux that re-enters the Steinlach River after the bend is more evenly distributed amongst the upstream domains. Furthermore, the largest change in flux comes from Domain B (0.2 L/s to 0.7 L/s). Comparing the Base Case to the Smooth Bathymetry, the number of particles that originate from Domain B and reach the downstream discharge area increases from 572 to 679, respectively. Figure 4.4b plots the travel paths of the particles that make up this difference. The Base Case particles discharge to a local bathymetric low towards the Northeast within Domain B, thus never reaching the downstream discharge area to become meander-scale hyporheic exchange. This is due to the geometry of the river and the sharp decrease in the Base Case river bathymetry within Domain B. Conversely, in the Smooth Bathymetry these particles are driven by the larger meander-scale gradient towards the Northwest, reaching the downstream discharge area.

This indicates that the presence of local bathymetry contrasts can cause a shift in the scales of simulated hyporheic exchange. This shift in hyporheic exchange scales accounts for approximately 32 % of the increase in meander-driven hyporheic exchange from Domain B. The remaining increase is due to the increase in the hydraulic gradient. The riverbed elevation throughout Domain B in the Smooth Bathymetry is higher compared to the Base Case (Figure 4.5a: river centerline cross-section between 50 to 100 m), which creates a larger hydraulic gradient towards the discharge area downstream of the meander. Thus, in addition to having more particles participating in meander-scale hyporheic exchange in the Smooth Bathymetry, the exchange fluxes are larger and the transit times are shorter in the Smooth Bathymetry as depicted in Figures 4.9 and 4.10, respectively.

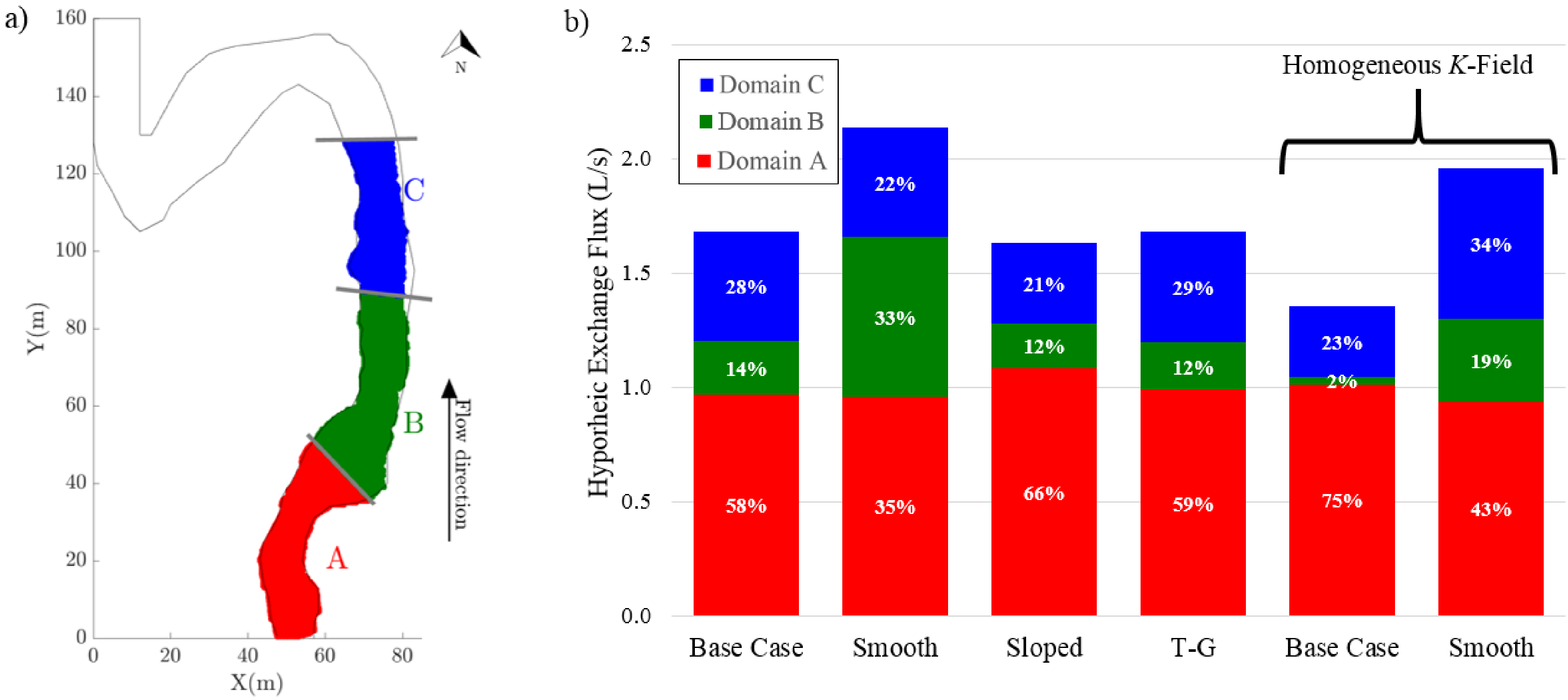
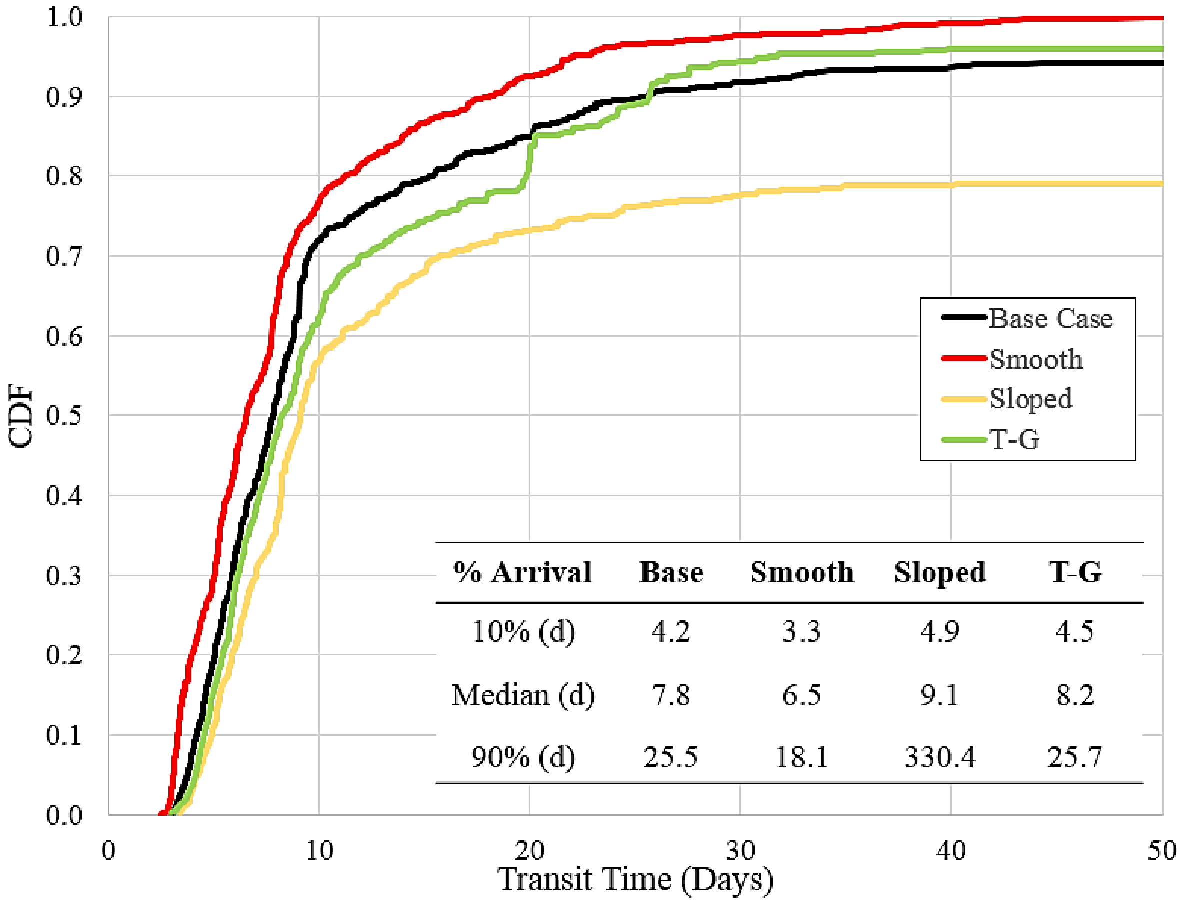


Figure 4.9: Mass balance of meander-scale hyporheic exchange: a) flux originating from upstream domains b) detailed summary from different bathymetry cases (Chow et al. 2018).



% Arrival	Base	Smooth	Sloped	T-G
10% (d)	4.2	3.3	4.9	4.5
Median (d)	7.8	6.5	9.1	8.2
90% (d)	25.5	18.1	330.4	25.7

Figure 4.10: Meander-scale hyporheic exchange transit times CDFs for all bathymetry scenarios. Inset summary table with early-time (10%), median (50%), and late-time (90%) arrivals (Chow et al. 2018).

We can conclude that the Smooth Bathymetry leads to a bias towards larger meander-scale hyporheic exchange and clearly lacks the bathymetry details necessary to accurately simulate meander-scale hyporheic exchange fluxes and transit times; thus more bathymetry data is needed, which leads to the next bathymetry scenario.

4.4.2 Sloped Bathymetry

In this case we go from a bathymetry with a constant slope to one that follows the slope along the centerline of the river. A bathymetric survey point is taken every 25 m along the centerline of the river for a total of twelve points. Within the cross-sections, the river bathymetry remains devoid of variability, similar to the Smooth Bathymetry. Figure 4.5b shows the resulting ‘Sloped Bathymetry’.

Figure 4.6c shows the corresponding change in the Sloped Bathymetry river stage relative to the Base Case. Overall, we can see that the changes in river stage compared to the Base Case are smaller than for the Smooth Bathymetry. Similar to the Smooth Bathymetry, the largest changes in river stage occur at the sides of the river. Overall, the changes to the underlying head depicted in Figure 4.8c are also smaller than those of the Smooth Bathymetry. However, in the Sloped Bathymetry there is a local decrease in head between 55 to 75 m in the Y-direction, which although appearing to be a small change in extent, causes a dramatic change in the hyporheic transit time distribution.

The total meander-scale hyporheic exchange flux for the Sloped Bathymetry is 1.6 L/s, which compares well to the Base Case. Furthermore, Figure 4.9b - ‘Sloped’ shows that the proportions of meander-scale hyporheic exchange from the upstream domains are similar to the Base Case. This evidence supports the Sloped Bathymetry’s ability to capture the bedform and meander-scale hyporheic exchange systems in the correct proportions. However, while the fluxes are comparable, there is a significant delay in the late-time transit times. Figure 4.10 shows increased skewness in the Sloped Bathymetry transit time CDF, taking approximately 330 days for the late-time meander-scale hyporheic exchange to reach the discharge area, compared to 26 days in the Base Case. What is the cause of this delay?

Most of the late-time particles originate from Domain A. Figure 4.4c compares the Base Case and Sloped Bathymetry particles originating from Domain A. It shows that a large number of the Sloped Bathymetry particles deviate northeastwardly towards Domain B, before continuing northwestwardly towards the downstream discharge area. Evidently, the decrease in head within Domain B has become a weak sink in the Sloped Bathymetry case, causing particles originating from Domain A to take a longer and more convoluted travel path through the subsurface, thus increasing their transit times. Abrams et al. (2013) defines a weak sink as a groundwater discharge boundary condition that lets some water pass underneath it, while a strong sink extracts water from the entire aquifer depth (Figure 4.11).

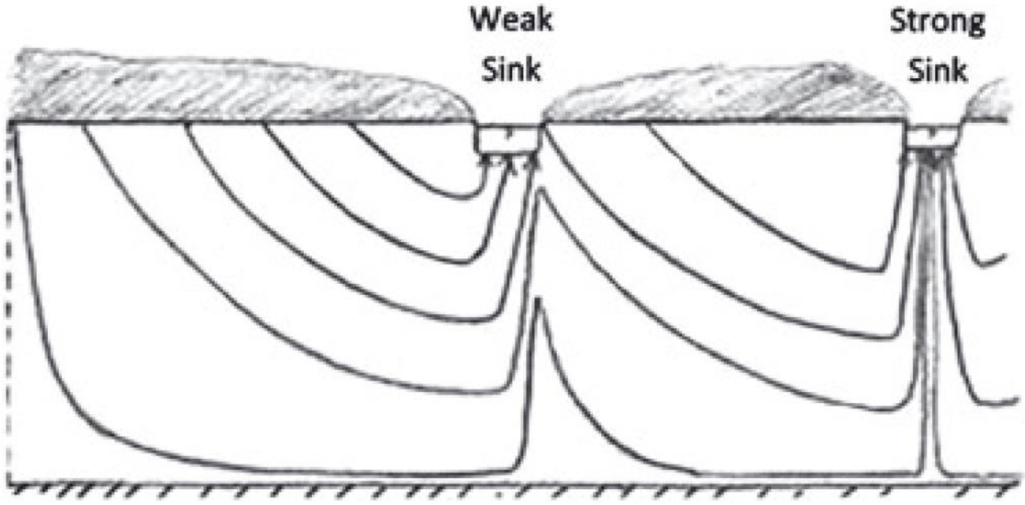


Figure 4.11: Conceptual model of a physical weak sink, which draws water from the upper portion of the aquifer only, and a physical strong sink, which draws water from the entire aquifer depth (Abrams et al. 2013).

Abrams et al. (2013) suggests approaches to account for this weak sink problem for surface water modelling, however the remaining problem is that “there is no way to know whether a specific particle should discharge to the sink or pass through the [weak sink] cell (Pollock 1994)”. In this modelling study, HGS does not require prescribing the spatial distribution of losing or gaining river reaches a priori. Additionally, in places where the river is gaining groundwater, we do not specify the strength of the groundwater sink.

The results indicate that a critical component of the river bathymetry is still missing and that the correct locations and magnitudes of losing (groundwater sources) and gaining (strong and weak groundwater sinks) river reaches are not being represented with the river slope alone. This leads to the final bathymetry scenario.

4.4.3 Thalweg – Gegenweg Bathymetry

In this bathymetry, I hypothesize that the specific locations of the bathymetric highs and lows within the river can produce a result similar to the Base Case. For a meandering river, the bathymetric lows or pools are usually found along the outside of the bend, where the river flow rate and erosion are highest (Hooke 1975). The line connecting the deepest points of successive cross-sections along the riverbed is commonly known as the thalweg (German for way along the valley). Conversely, the bathymetric highs or the shallowest portion of a meandering river are usually found opposite the thalweg in the form of point bars. There is no name given for the line that connects the highest points in a riverbed, thus I use the term ‘gegenweg’ (German for opposite way), which I define as: The line that connects the shallowest points in successive cross-sections along the depositional side of a river opposite the thalweg under average flow conditions. Figure 4.12 summarizes the Thalweg-Gegenweg concept.

To incorporate the positions of the bathymetric highs and lows, a bathymetry was interpolated with the minimum and maximum bathymetry elevations every 9 m in the flow-oriented direction, for a total of 54 points or 27 cross-sections. The resulting bathymetry is shown in Figure 4.5c and is referred to as the 'Thalweg-Gegenweg (T-G) Bathymetry'.

Figure 4.6d shows the corresponding change in Thalweg-Gegenweg Bathymetry river stage relative to the Base Case. Overall, Figure 4.6d shows the smallest magnitudes in river stage change among all simplified bathymetries. This leads to the smallest magnitudes in head difference, depicted in Figure 4.8d.

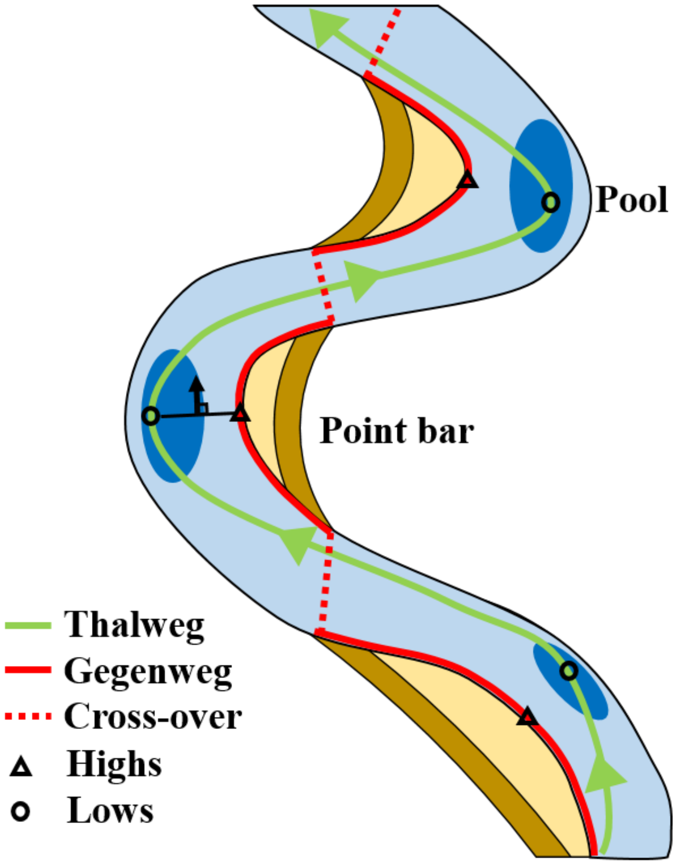


Figure 4.12: Conceptual drawing of the Thalweg – Gegenweg Bathymetry. The bathymetric lows can be surveyed following the thalweg, while the bathymetric highs can be surveyed by taking a measurement of the bank elevation opposite of the thalweg. I define the gegenweg as the line that connects the shallowest points in successive cross-sections along the depositional side of a river opposite the thalweg under average flow conditions. Along the gegenweg is a cross-over point where the shallowest bank will switch to the opposite side of the channel, represented by the red dashed line (Chow et al. 2018).

The resulting meander-scale hyporheic exchange fluxes and transit times from the T-G Bathymetry compare well to those of the Base Case (Figures 4.9b and 4.10). The results indicate that one can simulate hyporheic exchange at the meander scale in a physically meaningful way with a bathymetry that accounts for the bathymetric highs and lows. In this case the number of survey points were effectively reduced by 90 % (540 to 54 points) with the T-G Bathymetry sampling strategy without jeopardizing the simulated meander-scale hyporheic exchange fluxes and transit times. The results suggest that one could design a more effective bathymetry surveying campaign by prioritizing the bathymetric lows and the highs by visually identifying the Thalweg-Gegenweg of a meandering channel, and then strategically taking bathymetry measurements.

It is important to note that taking the minimum and maximum bathymetry elevations every 9 m was sufficient for the Steinlach River Test Site because major bathymetric features existed at that spacing. For other rivers the necessary spacing for an accurate simulation would depend on the level of bathymetry contrasts. Nonetheless, the T-G Bathymetry demonstrates that having the positions of bathymetric highs and lows within the river can lead to more physically meaningful simulations of hyporheic exchange fluxes and transit times than models that neglect the bathymetric variation in the transverse direction altogether.

4.4.4 Role of subsurface heterogeneity – homogeneous case

To show that the effect of bathymetry on shifting scales of simulated hyporheic exchange was not due to the chosen realization of hydrofacies distributions, I conducted the same analysis with a homogeneous K-field using the arithmetic mean of the heterogeneous K-field ($2.1 \cdot 10^{-3}\ m/s$). The last two columns in Figure 4.9b shows the resulting hyporheic exchange mass balance for the homogenous Base Case and Smooth Bathymetry. Similar to the heterogeneous Base Case, there is an increase in meander-scale hyporheic exchange flux compared to the Smooth Bathymetry (1.4 L/s to 2.0 L/s) with the largest change in flux originating from Domain B and the same change in particle trajectories shown in Figure 4.4b.

The same analysis with a homogeneous K-field was repeated using the geometric mean ($9.4 \cdot 10^{-4}\ m/s$). This resulted in approximately the same distributions of hyporheic exchange (i.e., increase in flux from Base Case to Smooth Bathymetry with the largest change coming from Domain B), only reduced by a factor of 2.5.

Through this simple sensitivity analysis using a homogenous K distribution, we learn two important lessons with respect to the effects of simplifying bathymetry when simulating meander-scale hyporheic exchange:

1. The biases towards larger meander-scale hyporheic exchange when lacking a detailed bathymetry are due to the bathymetry alone (that is, not from spatially varying K, riverbed conductance, and/or Manning's n).
2. The arithmetic mean produces hyporheic exchange fluxes closer to the heterogeneous case compared to the geometric mean, suggesting that the chosen realization of hydrofacies distributions is dominated by well-connected sandy gravel hydrofacies within the hyporheic zone.

From this we can conclude that in our simulations heterogeneity of the subsurface hydraulic properties plays a larger role in determining the connectivity pathways and ultimately hyporheic exchange transit times. Whereas river bathymetry mainly determines the river stage, thus affecting the underlying subsurface hydraulic head distributions, and the proportions of meander-scale to bedform-scale hyporheic exchange flux.

4.4.5 Role of river flow rate

To see the effects of river flow rate on meander-scale hyporheic exchange, I increased the specified constant inflow rate at the upstream inlet, then recalculated the steady-state river stage, hydraulic head distribution, and meander-scale hyporheic exchange flux (MSHEF) for the different bathymetry scenarios. Figure 4.13 summarizes the percent change in MSHEF relative to the MSHEF at the average river flow rate of 300 L/s (Figure 4.9b). Figure 4.13 only extends the river inflow rate (x-axis) to 2000 L/s because the MSHEF was plateauing for all bathymetry scenarios. Even though the T-G bathymetry MSHEF appears to be still rising slightly (green line in Figure 4.13), it had a relative increase of 50 % at a river flow rate of 4000 L/s, suggesting that a plateau for the T-G bathymetry scenario was also reached. River flow rates greater than 4000 L/s produced river stages above the river banks leading to completely different flow patterns. River flow rates less than 300 L/s generally showed little change (<5 %) in MSHEF. The lower limit of river flow (200 L/s) in these simulations coincides with the lower limit of gauged river flows.

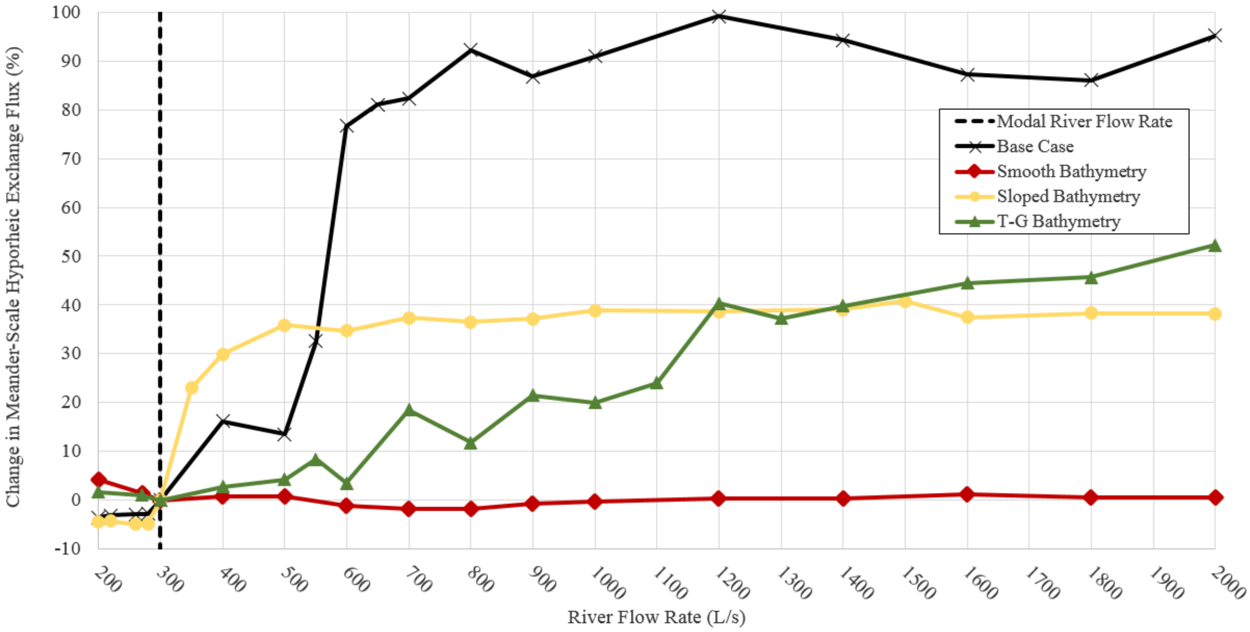


Figure 4.13: Relative change in meander-scale hyporheic exchange flux as a function of river flow rate under different river bathymetry scenarios (Chow et al. 2018).

In the Smooth Bathymetry scenario there is essentially no change in MSHEF with different river flow rates (red line in Figure 4.13). I believe this is because the Smooth Bathymetry scenario lacks small-scale bedforms where the river stage can accumulate behind. Therefore, with an increase in river flow rate, there is nearly a uniform increase in river stage upstream and downstream of the meander, thus the overall head gradient and MSHEF remain virtually constant. The other scenarios show an increase in MSHEF with increase in river flow rate before plateauing. The slopes of this increase for the Base Case, T-G, and Sloped Bathymetries are difficult to predict, however the plateaus are clearly higher in the presence of more bedform variability. The presence of bedform variability causes a non-uniform increase in river stage with increased river flow rate, which produces a larger meander-scale head gradient up to a certain threshold. The MSHEF plateau for the Base Case, T-G, and Sloped Bathymetries are 95 %, 50 %, and 40 %, respectively. In Figure 4.13, we can see a clear relationship between the amount of bathymetry variability and where this plateau is formed.

Figure 4.13 shows that none of the simpler bathymetries can approximate the Base Case MSHEF response to increased river flow. This shows clear limitations to the suggested T-G bathymetry when simulating conditions outside of the average river flow conditions under which it was constructed and illustrates the need for more detailed bathymetry when simulating transient problems. Nonetheless, of the simpler bathymetries, the T-G bathymetry results in a plateau closest to the Base Case.

4.5 Discussion

4.5.1 Nested hyporheic exchange systems

Some of the earliest work investigating the effects of topography on groundwater flow patterns, were performed by Tóth (1963) who recognized that nested local, intermediate, and regional groundwater flow systems can form in low-permeable unconfined aquifers with a topography-controlled water table (Haitjema and Mitchell-Bruker 2005). Numerous scientists have taken the same approach to model hyporheic exchange flow by representing the head fluctuations due to irregularities of the river bed by superimposing sinusoidal functions as a constant head boundary condition (i.e., Stonedahl et al. 2010, Zlotnik et al. 2011, Käser et al. 2013). This is an idealized approach of representing the head fluctuations that drive hyporheic exchange, neglecting the actual effects of bathymetry on the river stage and thus the underlying hydraulic head distribution in the river bed. While the Tóth-flow analysis has been helpful in identifying general key patterns of hyporheic flows, it is hardly applicable to real rivers that typically lack perfect periodicity. Thus, rather than using sinusoidal head boundary conditions, the actual 2D river stage distribution is computed using the fully coupled model HGS without relying on periodicity. Furthermore, the surface-flow simulations are not restricted to 1D calculations as typically done in regional-scale groundwater models, thus accounting for variability in river bathymetry over the width of the river, which is frequently observed in real rivers.

Numerous scientists (Cardenas et al. 2004, Gooseff et al. 2006, Buffington and Tonina 2009) have recognized the role of geomorphology and bathymetry when producing nested hyporheic flow systems. Geomorphology can add a third dimension of hyporheic exchange in the form of river meanders (Stonedahl et al. 2010), while local contrasts in the bathymetry can produce nonhydrostatic pressure variations that are not represented by the Saint Venant equations (Cardenas and Wilson 2007a, Trauth et al. 2013).

The hyporheic exchange scales and transit time distributions generated from 3D fully coupled sw-gw modelling are sensitive to the representation of river bathymetry which affects the river stage, exchange fluxes, and consequently the underlying subsurface hydraulic head distributions (Figure 4.8, respectively). Both bathymetry and river stage contribute to the generation of bedform-scale hyporheic exchange flow systems by creating the flow-driving hydraulic pressure conditions at the river bed, as well as providing pools that act like hydraulic shortcuts for discharging hyporheic exchange (in contrast to the riffles).

In other words, when detailed bathymetry contrasts are present, nested hyporheic exchange systems are produced (i.e., bedform-scale within meander-scale hyporheic exchange). With increased bathymetric contrasts, the frequency of alternating gaining and losing river reaches increases, thus less of the hyporheic zone and river reach participates in larger meander-scale hyporheic exchange. Similarly, Wang et al. (2018) investigated effects of larger-scale topographic mesh refinement on groundwater flow and transport, concluding that lower-resolution/fidelity topography leads to biases towards larger-scale intermediate and regional groundwater flow paths, thus underestimating the quantities of small-scale local groundwater flow paths.

Therefore, the lack of bathymetry detail when simulating hyporheic exchange can lead to biases in the scales of exchange represented. To quantify the scale bias in simulated hyporheic exchange from under-representing river bathymetry, a simple measure of the flushing intensity with depth could be calculated (Zlotnik et al. 2011).

4.5.2 Different river systems and upscaling

The Steinlach River Test Site has been described as a relatively homogeneous and smooth channel with a relatively steep hydraulic gradient, thus it is expected that the overestimation of larger-scale hyporheic exchange processes due to the lack of bathymetry data would be greater in other natural river systems with steeper gradients (i.e., braided and/or mountainous river systems). The greater the bathymetric contrasts, which most natural river systems exhibit, the more likely there will be an overestimation of larger-scale hyporheic exchange when lacking bathymetry data.

Hence, in order to simulate nested hierarchies of hyporheic exchange in a physically meaningful way, three conditions must be met. Firstly, one must be simulating a river that has bathymetric contrasts (i.e., is not a smooth concrete channel). Secondly, data needs to be collected within the river at a high enough resolution to capture these contrasts. Thirdly, the model must be discretized fine enough within the river to incorporate these contrasts. In other words, given that the simulation is discretized fine enough within the river, and that there is survey data that represents the contrasts in the river bathymetry, one should expect to represent the nested scales of hyporheic exchange more realistically.

The implications are that a lack of bathymetry data, which is often the case, can lead to biases in overestimating larger scale hyporheic exchange and residence times (Boano et al. 2014).

When extrapolated to account for all meanders along an entire river or rivers within a watershed, this can lead to gross biases of the catchment’s capacity to act as a reactor to attenuate pollutants (Grathwohl et al. 2013). At the catchment scale it may not be feasible to simulate the detailed contrasts in river bathymetry, however this study indicates the possibility of creating hyporheic exchange scaling factors (Boano et al. 2014) by comparing the simulated hyporheic exchange of a characteristic river segment with varying degrees of detail in bathymetry.

4.5.3 Study limitations and future work

In this study, I was limited to one realization of hydraulic conductivity generated using MPS, which allows for a better representation of laterally extensive preferential flow paths that are important for the distribution of flow through the riverbed and hyporheic zone (Mahmud et al. 2014). However, it was also shown through a homogeneous K-field case that these insights were not due to the chosen realization of hydrofacies distributions. The study presented in the next chapter aims to evaluate the level of detail required in the subsurface parameterization in order to simulate hyporheic exchange at the meander-scale.

Furthermore, I chose to compute the hyporheic transit time distributions using simple particle tracking because it is computationally inexpensive and it allowed us to derive unique hyporheic exchange transit time CDFs, while simultaneously allowing us to explicitly map the origins of the meander-scale hyporheic exchange. Future work could include using the advection-dispersion equation to calculate the hyporheic exchange life expectancy (Molson and Frind 2012). Although this would dramatically increase the computational demand, I would expect to obtain more accurate hyporheic exchange transit time distributions.

Finally, river bathymetries evolve over time, which can have dramatic impacts on hyporheic exchange processes. A river meander is not a static environment but evolves due to erosional and depositional processes. The future of integrated sw-gw models could include feedbacks between the hydraulic processes that drive erosion-deposition (Merritt et al. 2003, Parker et al. 2011, Beckers et al. 2018) and vice versa, in order to simulate the changes in hyporheic exchange within an ever-evolving river.

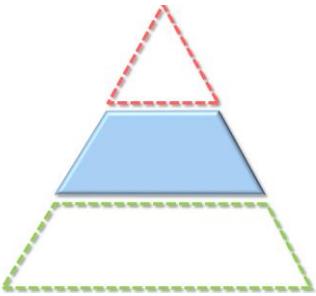
4.7 Conclusions

The goal of this study was to determine the role river bathymetry plays when simulating hyporheic exchange fluxes and transit times and to determine which aspects of bathymetry are the most important to simulate hyporheic exchange in a physically meaningful way. I posed the motivating research question (RQ2, Section 1.5): what are the most important aspects of river bathymetry that need to be captured to simulate meander-scale hyporheic exchange in a physically meaningful way? The results of this study indicate that simulating hyporheic exchange with a 3D fully coupled sw-gw model is sensitive to the representation of river bathymetry. The river bathymetry affects the river stage, and consequently the underlying subsurface hydraulic head distributions. The presence of local-scale contrasts in bathymetry can lead to an increase in smaller bedform-scale hyporheic exchange nested within larger meander-scale hyporheic exchange systems. Conversely, simulations of hyporheic exchange at the meander-scale that lack detailed bathymetry data can lead to the overestimation of larger meander-scale exchange.

In these numerical experiments of the Steinlach River Test Site, meander-scale hyporheic exchange was overestimated by about 27 % when challenged by limited bathymetry data. Overestimating meander-scale hyporheic exchange can lead to biases towards longer residence times. When extrapolated to account for all meanders along an entire river or rivers within a watershed, this can lead to gross biases in the estimated retention and pollutant attenuation capacity of a catchment.

Additional bathymetry information, such as more details of the river slope alone is not enough to accurately simulate the locations and magnitudes of losing and gaining river reaches. Thus it is important to also identify the locations of local bathymetric highs and lows within the river. Bathymetry surveying campaigns can be performed more effectively by prioritizing bathymetry measurements along the thalweg and gegenweg of a meandering channel, which can be identified visually. I define the gegenweg as the line that connects the shallowest points in successive cross-sections along the depositional side of a river opposite the thalweg under average flow conditions. Incorporating local bathymetric contrasts will likely capture the nested hierarchies of hyporheic exchange systems, leading to more physically meaningful simulations of hyporheic exchange fluxes and transit times.

5. Evaluating subsurface parameter structures to simulate meander-scale hyporheic exchange[‡]



Chapter highlights

- If the parameter structure is too simple, it will be limited by intrinsic (irreducible) model errors. By increasing subsurface complexity through the addition of zones or heterogeneity, we can begin to exchange intrinsic for epistemic (potentially reducible with additional data) model errors.
- A zonated model is capable of reproducing the transit time distributions of a more detailed model, but only if the geological structures are known.
- An interpolated heterogeneous parameter field (cf. pilot points) showed the best trade-offs between intrinsic and epistemic model errors, indicating fitness for practical applications.
- Parameter fields generated by multiple-point geostatistics (MPS) produce transit time distributions with the largest uncertainties, however these are reducible by additional hydrogeological data, particularly flux measurements.

5.1 Motivation and research questions

Another model aspect identified as contributing significantly to the predictive uncertainty of hyporheic exchange is the structural heterogeneity of the subsurface sediments within the hyporheic zone (e.g., Thibodeaux and Boyle 1987, Osenbrück et al. 2013, Zhou et al. 2014, Stonedahl et al. 2018). The representation of riverbed and hyporheic zone heterogeneity has

---

[‡] The contents of this chapter are modified from 'Chow, R., Bennett, J.P., Dugge, J., Wöhling, T., and Nowak, W. (2019). Evaluating subsurface parameterization to simulate hyporheic exchange: The Steinlach River Test Site. Groundwater, https://doi.org/10.1111/gwat.12884'

Author contribution statement: ***Chow, R.***: Coordinated work and writing. Wrote preliminary works and original draft manuscript. Revised manuscript addressing peer-review comments. Finalized and approved published article. Analyzed and evaluated model results. Conceptualized modelling study. ***Bennett, J.P.***: Developed conceptual and Multiple Point Geostatistics (MPS) model. Helped write MPS section in original draft manuscript. Helped with article revisions to address peer-review comments. ***Dugge, J.***: Developed preliminary Steinlach River Test Site HGS model. Conceptualized modelling study. ***Wöhling, T.***: Reviewed original draft manuscript. Conceptualized modelling study. Provided supervision. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Conceptualized modelling study. Provided supervision.

long been recognized as a significant contributor to predictive uncertainty when simulating sw-gw exchange (Vaux 1968). Riverbed heterogeneity has been found to have a net effect on compressing the hyporheic zone toward the riverbed and to influence hyporheic exchange transit time distributions at the bedform-scale (Salehin et al. 2004, Tonina et al. 2016, Liu et al. 2018). Pryshlak et al. (2015) found that, for meandering rivers with strongly heterogeneous sediments, characterizing lithological structures or hydrofacies is more critical than characterizing the channel morphology when predicting hyporheic exchange residence times. Since this study focuses on meander-scale hyporheic exchange, I chose to ignore the finer-scale riverbed conductance heterogeneity typical of riverbeds, and instead focus on geological heterogeneities at the 10 m to 100 m scale.

While previous studies show the effects of subsurface heterogeneity on the hyporheic zone and exchange rates, it remains unclear what level of detail is necessary to simulate hyporheic exchange transit time distributions at the meander-scale and what the expected trade-offs are when choosing one subsurface parameter structure over another. Therefore, my research goal (Section 1.5) for this chapter is to evaluate the degree of subsurface detail necessary to accurately simulate hyporheic exchange transit time distributions. In theory, as a model’s complexity increases (e.g., by introducing more processes or adding more parameters), the model will be able to produce output that better fits calibration data. However, with added complexity comes greater variability in the model output because the model has more degrees of freedom. Furthermore, overly complex models can easily lead to over-fitting observations, resulting in a decline in predictive accuracy. This is known as the bias-variance trade-off, where we would expect a lower bias with a more uncertain forecast (larger variance) as we increase the complexity in the subsurface representation of heterogeneity (e.g., Guthke 2017). Note that model complexity is not uniquely defined in the hydrogeological community and is a topic being advanced in the field of model selection (Höge et al. 2018). In this study, I use the term complexity to indicate the number of parameters required to model details in the subsurface and the required knowledge to produce it.

There are several approaches to parameterize the subsurface, which differ in their complexity and their need for data to resolve them. This study investigates three approaches based on their relevance to applied groundwater modelling (Anderson et al. 2015):

1. **Zonation** is the most common and arguably the simplest approach used in practice.
2. **Interpolation** methods such as pilot points relieve the modeller of the responsibility of defining a zonated pattern and are commonly used to represent the heterogeneity inherent in hydrogeological problems (Doherty 2003).
3. **Multiple-point geostatistics** (MPS; Guardiano and Srivastava 1993) are becoming increasingly popular because they do not require the explicit definition of a random function to represent the geological architecture under investigation. Instead, MPS directly utilize empirical multivariate distributions implied by training images and can effectively capture the subsurface geometrical features crucial for characterizing fluid flow (Hu and Chugunova 2008).

Although I have chosen a limited set of subsurface parameter structures to test, I have attempted to represent commonly applied industry-standard (zonated and pilot points) and state-of-the-art (MPS) approaches that cover a wide spectrum of complexity. With the aforementioned approaches, I will test the impacts of assuming homogeneous subsurface parameter distributions in one or more dimensions when simulating meander-scale hyporheic exchange. Specifically, I will use the zonated approach to test the assumption of total and horizontal homogeneity by creating a 1- and 2-layer model. Next, I will use a 2D interpolation approach to test the assumption of vertical homogeneity by varying the subsurface parameters horizontally, while keeping them constant vertically. Finally, I will use the MPS approach to test the assumption of both vertical and lateral homogeneity by varying the parameter distributions in all three dimensions simultaneously.

The principle of parsimony, the idea that a model should be made as simple as possible but not simpler, often guides modellers when defining the optimal level of parameterization in environmental models. Hill (2006) believes that parameter parsimony or the use of a small number of parameters is essential for effective model calibration, and that model complexity should be added gradually and only as necessary. Voss (2011a) adds that there are few insights to be gained from the time-consuming and tedious process of pushing buttons and adjusting knobs of highly parameterized models to fit data, instead the primary objective of a modelling exercise should be learning something about the system through model-based analysis.

However, some researchers argue that the notion of the calibrated model has no place in environmental models since uniqueness is always achieved at the cost of error-inducing simplifications (Beven and Binley 1992, Pappenberger and Beven 2006, Cirpka and Valocchi 2016). Hunt et al. (2007) states that there are diminishing returns whereby some level of parameter complexity improves our simulation capabilities, but too much leads to instability, nonuniqueness, long model runtimes, and an increased potential for predictive error. Instead, it has been argued that a large number of model runs should be used within a probabilistic Bayesian framework to explore the range of predictive possibilities rather than making a single, deterministic (best-fit) prediction (Gómez-Hernández 2006). Bayesian multi-model frameworks can be used to find the data amount required to justify a given level of complexity (Schöniger et al. 2015) and, if appropriately applied, can also enhance predictive performance (Höge et al. 2018).

The principle of parsimony is similar to a literary rule for storytellers by the early 20th Century author Anton Chekhov known as “Chekhov’s Gun”, which states that: “If in Act One you have a pistol hanging on the wall, then it must fire in the last act.” This means that every element in a story must be necessary, and irrelevant elements must be removed; furthermore elements should not appear to make “false promises” by never coming into play (Bit͡sīlli 1983). Chekhov’s Gun suggests to environmental modellers that only relevant model components (i.e., processes that significantly contribute to the desired forecast or modelling objective) should be included in the model, all others should be removed. The concept of Chekhov’s Gun is increasingly pertinent in the age of complex fully integrated hydrological models.

These two problems (error inducing simplifications versus nonuniqueness of calibration) are identical to the bias-variance tradeoff mentioned above. They can likewise be categorized into two broad types of forecast uncertainty: (1) intrinsic and (2) epistemic (Rubin 2003, Anderson et al. 2015). Intrinsic uncertainty, also known as aleatoric uncertainty (Ross et al. 2009, Beven and Young 2013), is the inherent variability of a given problem (i.e., the model forecast from a chosen subsurface parameter structure) that is irreducible by more measurements or new knowledge. Conversely, epistemic uncertainty is the forecast variability that can potentially be reduced with more data.

My motivating research question for this chapter (Q3, Section 1.5) is: what is the degree of subsurface detail necessary to simulate transit time distributions of meander-scale hyporheic exchange in a physically meaningful way (i.e., low intrinsic error, low bias)? To answer this question, it is necessary to quantify the trade-offs from choosing between different subsurface parameterization structures. Additionally, I wish to determine which subsurface parameter structure makes the best use of hydrogeological data by reducing the predictive uncertainty (low epistemic error, low variance).

5.2 Approach

To address these questions, I first build a fully coupled sw-gw model parameterized by a hierarchical MPS method informed by field observations which is treated as the reference solution or "virtual reality" (Section 4.3 and Figure 5.1a). I then systematically test the aforementioned subsurface parameterization structures to determine their relative impact on predictive uncertainty and compare the results to the virtual reality model. The model which forms the basis of this study is a HydroGeoSphere model (Aquanty 2015) of the Steinlach River Test Site developed by Chow et al. (2018).

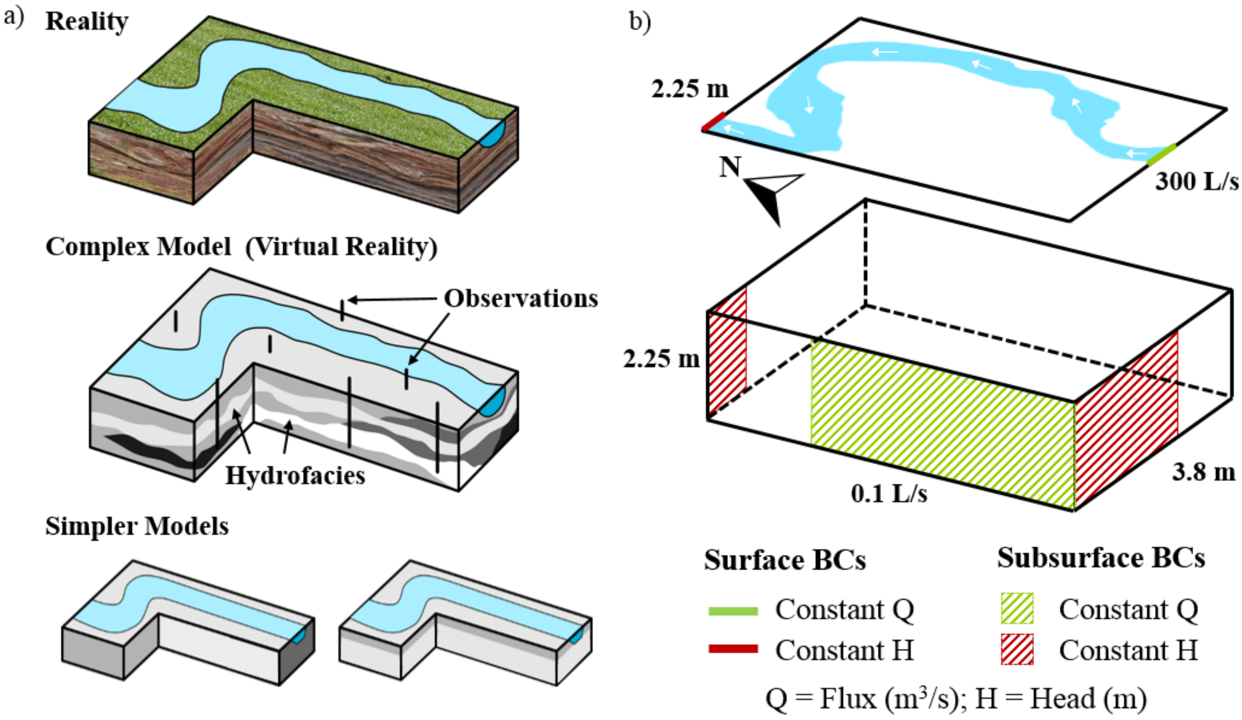


Figure 5.1: Study concept: a) A highly parameterized complex model (virtual reality) is informed by site observations (reality) and is treated as a basis for comparison. This model is in turn modelled with simpler subsurface parameter structures to compare their relative performance; b) Model boundary conditions (adapted from Chow et al. 2018).

I expect that the simplest representations of the hydrogeological structures (i.e., a homogeneous one-zone model with a single effective conductivity value) will be too coarse to produce a transit time distribution similar to the virtual reality model (high bias) and that geostatistical random fields will provide the necessary flexibility (possibly low bias, but high uncertainty). However, the expected trade-off will be that methods that depend on geostatistical random fields will yield much higher variances when predicting transit time distributions, and so they require more calibration data before they can strike a useful trade-off between bias and variance. While similar research questions have been addressed in theoretical works (e.g., Schöniger et al. 2014) or for different systems (e.g., Pham and Tsai 2016), I wish to derive recommendations specific to meander-scale hyporheic exchange.

The following sections in Chapter 5 describe the different methods used to generate alternative subsurface parameter structures, followed by a description of the methods used to evaluate model predictive errors from the different subsurface parameter structures by comparing their results to the results from the virtual reality model (Section 4.3). Finally, I present the predictive error results, followed by a discussion and Chapter specific conclusions.

5.3 Model modifications and alternative subsurface parameter structures

5.3.1 HydroGeoSphere model boundary condition modifications

The same HGS model of the Steinlach River Test Site described in Section 5.2 (Base Case) serves as the reference solution or ‘virtual reality’ for this study. A few slight modifications to the boundary conditions were made to enhance the numerical stability of the model in preparation for randomized subsurface parameter fields used in this study. The model results (e.g., transit time distributions) proved to be insensitive to these changes.

Both surface and subsurface boundary conditions (BC) were specified either as a constant head or prescribed flow (Figure 5.1b). A constant inflow rate of 300 L/s was specified for the south inlet based on the observed modal flow rate of Steinlach River discharge (Voronina 2013). The subsurface BCs were defined along three sides of the model domain. A constant head of 3.80 m and 2.25 m was specified at the upstream and downstream ends, respectively, based on corresponding mean groundwater levels. A constant inflow of 0.1 L/s was specified on the western side of the model domain to account for estimated groundwater fluxes originating from the hillslope (Osenbrück et al. 2013). A default no-flow boundary was applied to all other sides of the model domain.

Section 4.3 provides a detailed description of the model setup and parameterization of the virtual reality model. Again, it should be noted that the riverbed sediments are assumed to have the same hydrogeological properties of the fluvial aquifer because this study mainly focuses on geological heterogeneities at the 10 to 100 m scale. Salehin et al. (2004) and Liu et al. (2018) have investigated the effects of riverbed conductance heterogeneity on hyporheic exchange, concluding that an increase in riverbed heterogeneity can increase both variability in fluxes and ranges in residence times.

5.3.2 Alternative subsurface parameter structures

In this study I consider three categories of subsurface parameter structures: (1) a zonation structure, (2) a two-point geostatistical interpolated structure that emulates pilot point methods, and (3) a hierarchical multiple-point geostatistical structure. An example of the K-distribution for each subsurface structure is presented in Figure 5.2. Each of the three subsurface parameter structures is subdivided into two subcategories that will be evaluated in this study (Table 5.1).

Table 5.1: Summary of subsurface parameter structures to evaluate

Zonated		**Interpolated**		**Multi-Point Geostatistics**	
One-Zone	Two-Zone	66 Pilot Points	128 Pilot Points	Hydrofacies	Hydrofacies + Internal Heterogeneity

5.3.2.1 Zonated model (one-zone and two-zone)

First I test the simplest representation of the subsurface by making it a homogeneous one-zone model. This represents a pessimistic scenario, assuming absolutely no knowledge about major geological structures. Then I create a two-zone model by separating the topsoil and underlying fluvial aquifer. The geometry of the two layers is taken accurately from the virtual reality, which resembles an optimistic case of knowing the contact between hydrogeological units. The K-range used for the one-zone model ranges from $5 \cdot 10^{-3}\ m/s$ to $5 \cdot 10^{-4}\ m/s$, based on hydrogeological tests of the fluvial aquifer (Doro et al. 2013). For the two-zone model (Figure 5.2a), the same K-range was used to represent the fluvial aquifer, with the K-range for the topsoil lowered by an order of magnitude (i.e., $5 \cdot 10^{-4}\ m/s$ to $5 \cdot 10^{-5}\ m/s$).

5.3.2.2 Interpolated model (66 and 128 pilot points)

Next, I test 2D heterogeneous K-fields that emulate pilot point methods, using randomized pilot point values. The goal of the pilot point method is to provide an intermediate approach for characterizing subsurface heterogeneity, striking a balance between direct representation of element-by-element variability and reduction of parameterization to a few homogeneous zones (Doherty et al. 2010). The pilot point method is commonly used to generate smoothly varying heterogeneous parameter fields, where all elements are assigned a parameter value by kriging between a number of parameter estimates available at discrete locations throughout the domain, the “pilot points”. I chose not to include the topsoil layer to simplify the interpolated model in the third dimension, allowing us to test whether a heterogeneous 2D *K*-field is capable of producing transit-time CDFs similar to the virtual reality model.

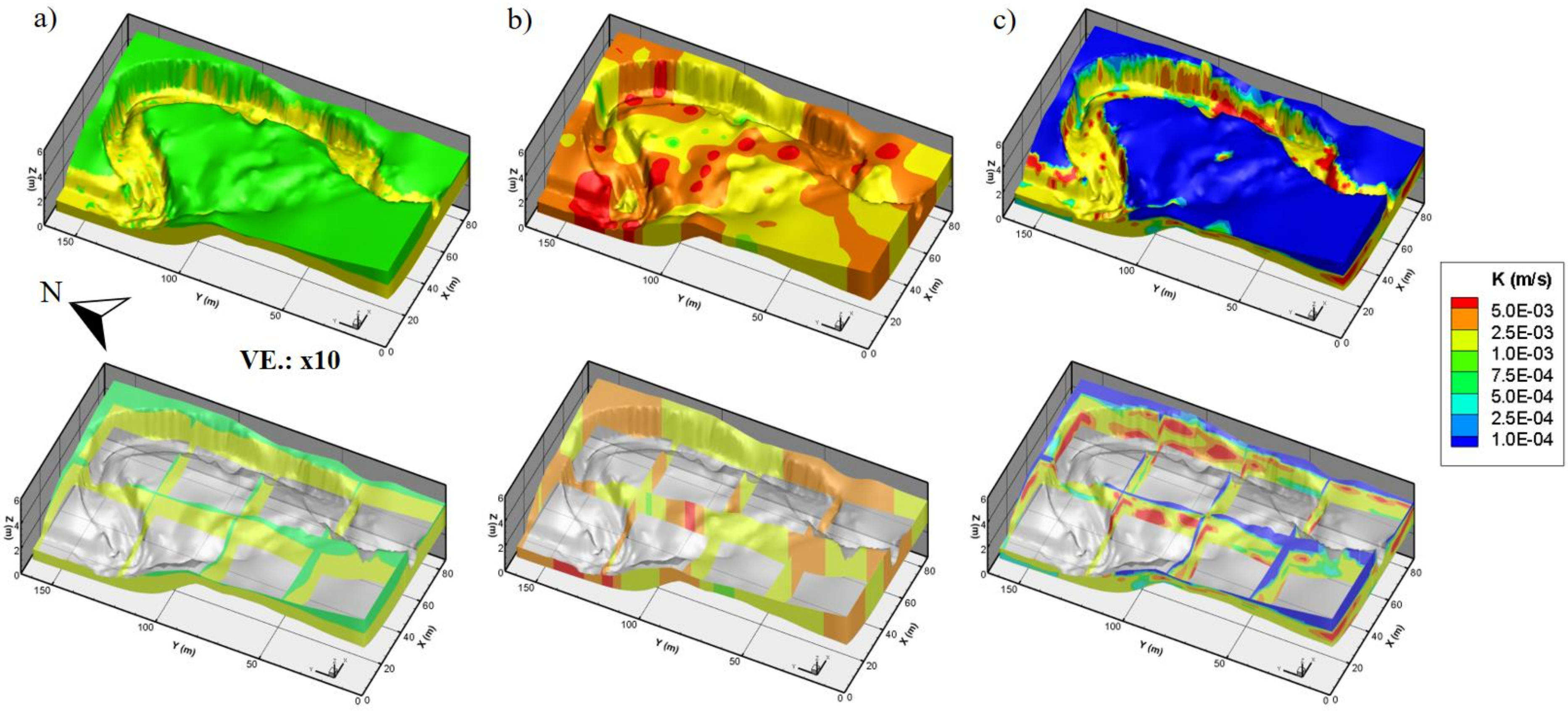


Figure 5.2: Examples of the different subsurface parameter structures (top row) and a fence diagram to show the internal makeup of the structures (bottom row): a) zonated model: two-zones separating overlying topsoil from underlying fluvial aquifer, b) interpolated model: a multi-Gaussian random K-field kriged from 128 equally spaced pilot points, c) Multiple-Point Geostatistics model: overlying topsoil and differing hydrofacies without local-scale internal heterogeneity.

The following steps were executed to generate random heterogeneous 2D K-fields:

1. Sample $log\,K$ from 14 borehole locations in the virtual reality fluvial aquifer.
2. Calculate the 2D depth-averaged arithmetic mean K for each borehole.
3. Create a variogram from the 14 depth-averaged $log\,K$-values.
4. Generate multi-Gaussian random $log\,K$-fields based on the found variogram ($\sigma_{lnK}^2 = 1.5$, $\lambda = 5\ m$, $K_{GEO} = 2.71 \cdot 10^{-3}\ m/s$) using the spectral method described in Section 4.2.3 and condition the realization to the 14 K-values using the method of smallest possible modifications (Kitanidis 1995).
5. Sample K from the multi-Gaussian random log K-field from previous step (4) at uniformly spaced interpolation points (i.e., pilot points) and apply kriging to the field.

Figure 5.2b is an example of a field generated from 128 uniformly spaced interpolation points (8×16 points with a 10 m spacing). The choice of pilot point spacing was based on the Steinlach River Test Site MODFLOW model by Wöhling et al. (2016) and the recommendations from Doherty et al. (2010) on pilot point density. I also tested a scenario with approximately half the interpolation points (66 points total, 6×11 points with a 15 m spacing) to see if acceptable results could be generated with fewer parameters. Both scenarios were run with ensembles of 100 realizations.

5.3.2.3 Multiple-Point Geostatistical model (with and without internal heterogeneity)

Finally, I return to the subsurface parameter structure used to create the virtual reality model. The hierarchical MPS method used to create the virtual reality model is capable of generating multiple realizations; the single realization that is most in-line with the site data and conceptual model is the one used as the virtual reality model. In order to quantify the variability of the forecasts, I generate additional realizations varying the hydrofacies distributions using impala (Straubhaar et al. 2011), based on the multivariate distributions inferred from an SBED training image (Figure 4.3). I run one scenario without local-scale internal heterogeneity, using a homogeneous mean K-value for each hydrofacies. I run a second scenario with local-scale internal heterogeneity. The local-scale internal heterogeneity is generated using the spectral method, variance, and correlation lengths described previously in Section 4.3.3. This will allow us to see the relative impact of local-scale internal heterogeneity on predictive uncertainty of transit time distributions. Again, both scenarios were run with ensembles of 100 realizations.

5.3.3 Simulated vs. observed hyporheic exchange

The simulated hyporheic exchange for the virtual reality model compares well with observed flux and transit times. Osenbrück et al. (2013) estimated meander-scale hyporheic exchange flux at the downstream outlet spring to be between 1 L/s and 2 L/s with an average transit time of approximately 7.8 days. The simulated hyporheic exchange flux was 1.9 L/s and the median transit time was 7.9 days. This model serves as a basis for comparison or virtual reality of the Steinlach River Test Site for evaluating different subsurface parameter structures.

All models (except for the zonated ones) will intrinsically generate various levels of mechanical dispersion as they account for various scales of heterogeneity. I will account for differences in dispersion and scale in the discussion. Yet, the largest contribution for dispersion visible in the hyporheic transit time distributions is morphological dispersion (Rinaldo et al. 1991, Soltani and Cvetkovic 2017), which originates from the geometry of the river meander and the particle pathway length through the hyporheic zone. In other words, even with homogeneous aquifer properties I will produce transit time distributions with dispersion partly because of the relative starting position from the source (initial particle placement within upstream segments of downwelling) to the sink (downstream upwelling at outlet spring).

5.4 Quantifying model predictive error

In this study, I estimate the model predictive error for the different parameterizations (Table 5.1) by evaluating the hyporheic exchange transit time distributions as a function of the subsurface parameter structure. Predictive errors are deviations to the transit time CDF provided by the virtual reality model. My parameter of concern is hydraulic conductivity (K) because it is known to vary by orders of magnitude (Freeze and Cherry 1979). Another subsurface parameter necessary to compute transit times is the effective porosity, which I assume to be constant to help simplify the analysis and because its range is small compared to the range of K. Furthermore, specific storage is neglected because I assume steady-state hydraulic head distributions in this study.

Figure 5.3 summarizes the process used to quantify the total predictive error of hyporheic exchange transit times. In order to quantify the total model predictive error, I change (for each model) the K-values within the range and variances of those measured at the site (Doro et al.

2013). Thus, for a given subsurface parameter structure, I generate a number of realizations differing the K-values with the methods described in Section 5.3.2. The parameter values are either sampled regularly and independently over a specified range (for zonated models) or by Monte Carlo methods that repeatedly generate random K-distributions (ensembles of 100 realizations for interpolated and MPS models). I then rerun the Steinlach River Test Site model for all realizations, keeping boundary conditions and model geometries constant, generating an ensemble of transit time CDFs per model scenario. With an ensemble of transit time CDFs, I then compute statistics representing total model predictive error. I subdivide total model predictive error into two main subcomponents:

a) Bias – is the difference between the average ensemble value and the reference solution (i.e., virtual reality CDF). The difference is called the residual error. A difference is calculated for every percentile of the CDF and is aggregated across the percentiles by calculating the mean square error (MSE).

b) Variance – is the variability or uncertainty of a model prediction calculated as the prediction variance of the realizations. The variance is calculated for every percentile of the CDF and is aggregated by calculating the average variance (σ_{avg}^2).

Thus, the total model predictive error can be plotted as the sum of the bias (MSE) and variance (σ_{avg}^2), allowing us to quantitatively compare the predictive errors from the different subsurface parameter structures. The units for the bias (MSE) and variance (σ_{avg}^2) are squared times (i.e., Days^2), which may feel awkward but assures that the two types of errors are additive for a measure of total error.

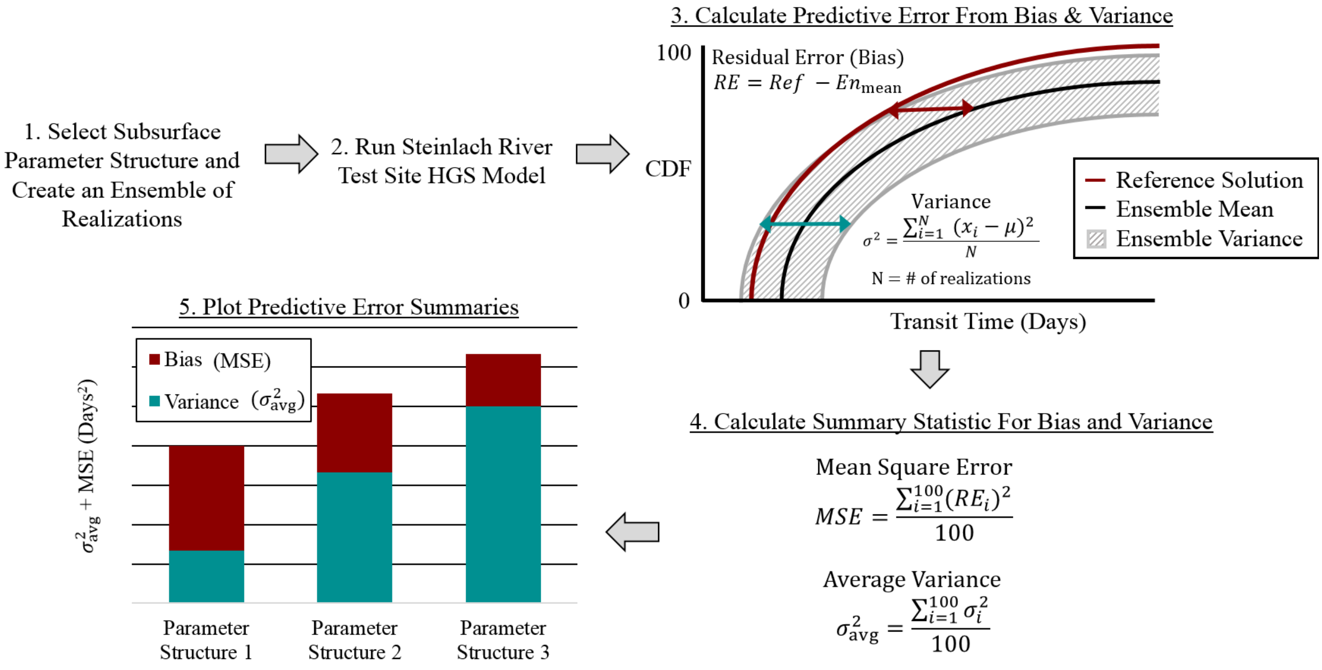


Figure 5.3: Quantifying predictive errors of simulated hyporheic exchange transit time distributions.

5.4.1 Estimating data worth for different parameter structures

In order to evaluate the worth of specific hydrogeological data to a subsurface parameter structure, I condition my model-wise ensembles to data. I achieve conditioning by a procedure known as the smooth bootstrap filter (Pitt 2002). To do so, I compute the likelihood of each realization given the hydrogeological data (i.e., hydraulic head and a hyporheic exchange flux measurement). The closer the realization is to the given hydrogeological data the higher the likelihood. The likelihoods are then summed for all realizations of a given parameter structure, then used as the denominator to divide each likelihood. This turns each likelihood into a model weight that ranges from 0 to 1, with the sum of the weights adding up to 1. This means that the output of each realization is weighted based on how well it fits hydrogeological data. The model bias and variance for a given parameter structure is then recalculated based on these model weights (Figure 5.4). Through this conditioning we can see the relative impact different hydrogeological data types have on error reduction (i.e., data worth).

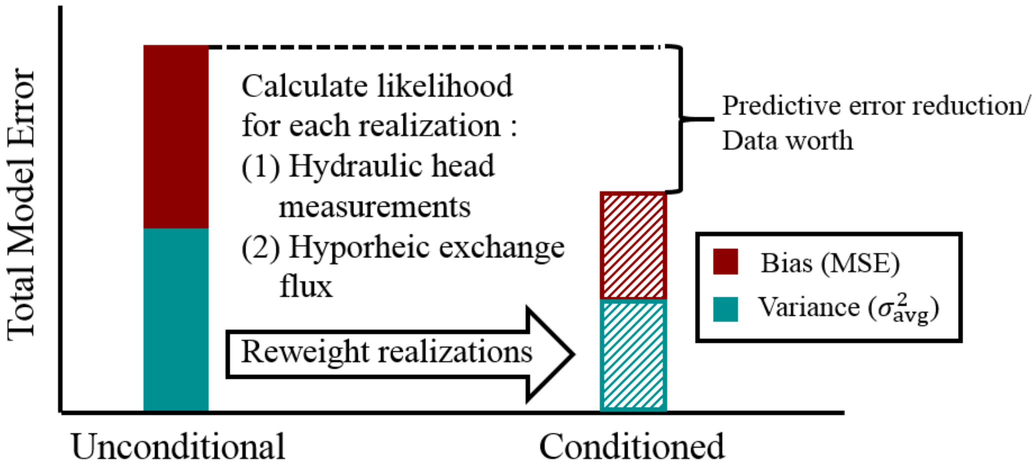


Figure 5.4: Conditioning hyporheic exchange transit time errors to hydrogeological data.

In this study, I condition my realizations to two types of available hydrogeological data: (1) 14 hydraulic head observations and (2) an estimate of the meander-scale hyporheic exchange flux. For the conditioning to hydraulic head data I use a Gaussian likelihood function (Equation 5.1) of the errors between the simulated (h_s) and observed heads from the virtual reality model (h_o) with a measurement error variance of $\sigma_h^2 = 1\ cm^2$ (equating to a $\sigma_h = 1\ cm$, representing the potential measurement error for hydraulic heads), where n is the number of observations:

$$L_h(h_o, \sigma_h^2, h_s) \propto exp\left(-\frac{1}{2\sigma_h^2}\sum_{j=1}^{n}(h_{sj} - h_{oj})^2\right) \tag{5.1}$$

For the conditioning to the hyporheic exchange flux, again I use a Gaussian likelihood function (Equation 5.2) with a mean of $q_o = 1.5\ L/s$ and a variance of $\sigma_q^2 = 0.2\ (L/s)^2$, where q_s is the simulated hyporheic exchange flux:

$$L_q(x|q_o, \sigma_q) \propto exp\left(-\frac{1}{2}\frac{(q_s - q_o)^2}{\sigma_q^2}\right) \tag{5.2}$$

Figure 5.4 summarizes the approach of conditioning total model error of an ensemble. The solid bar is considered “unconditional”, which is made up of equally weighted realizations. The hatched bar is considered “conditioned”, which is the result of re-weighting realizations based on hydrogeological data fit. If there is no change or an increase in total predictive error from conditioning to hydrogeological data, this would indicate that the data cannot constrain the output for this particular parameter structure. In this case, it may be necessary to obtain data at different locations or use a different data type in order to have a meaningful reduction in predictive uncertainty. To assure that 100 realizations are enough for conditioning, I checked that the effective sample size is large enough to robustly estimate averaged variances (Leube et al. 2012).

5.5 Results from differing subsurface parameter structures

Figure 5.5 shows the corresponding transit time CDF error analysis for each subsurface parameter structure. The grey shaded areas depict the 5th and 95th percentile range of the transit time CDF ensemble. The darker grey shading shows the percentile ranges after conditioning to hydrogeological data. The percentile ranges are plotted overlapping one another with the narrowest uncertainty range on top. When the unconditional CDF percentile range is not visible, this indicates that conditioning to this particular data type is uninformative because the CDF percentile range is not reduced within the presented 100 day timespan. The dark red line is the reference CDF from the virtual reality and the black lines are the model ensemble average CDF. The dashed lines are the ensemble average CDF after conditioning to hydrogeological data.

Figure 5.6 is a summary of the CDF error analysis translated into a stacked bar plot in order to compare the total predictive error from different parameter structures (solid bars) and their response when conditioned to hydrogeological data (hatched bars).

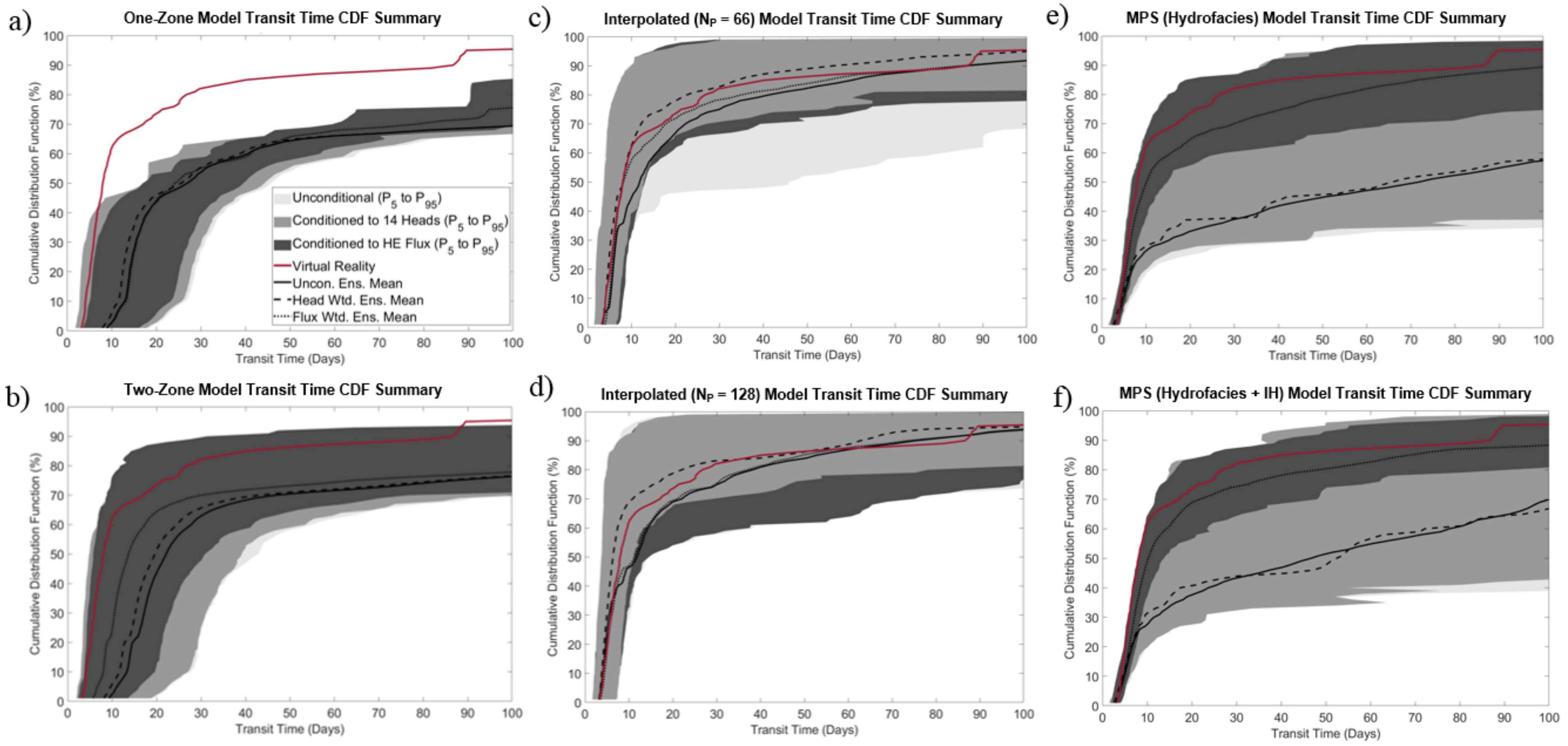


Figure 5.5: Summaries of the corresponding transit time CDF error analysis for each subsurface parameter structure. The grey shaded areas depict the variability or uncertainty of the transit time CDF ensemble by showing the 5th and 95th percentile range and lines represent the ensemble means: a) one-zone model: varying K one order of magnitude, b) two-zone model: varying K one order of magnitude for each zone, c) interpolated model: multi-Gaussian random field kriged to 66 equally spaced pilot points, d) interpolated model: multi-Gaussian random field kriged to 128 equally spaced pilot points, e) MPS model: hydrofacies distributions only, and f) MPS model: hydrofacies distribution with local-scale internal heterogeneity.

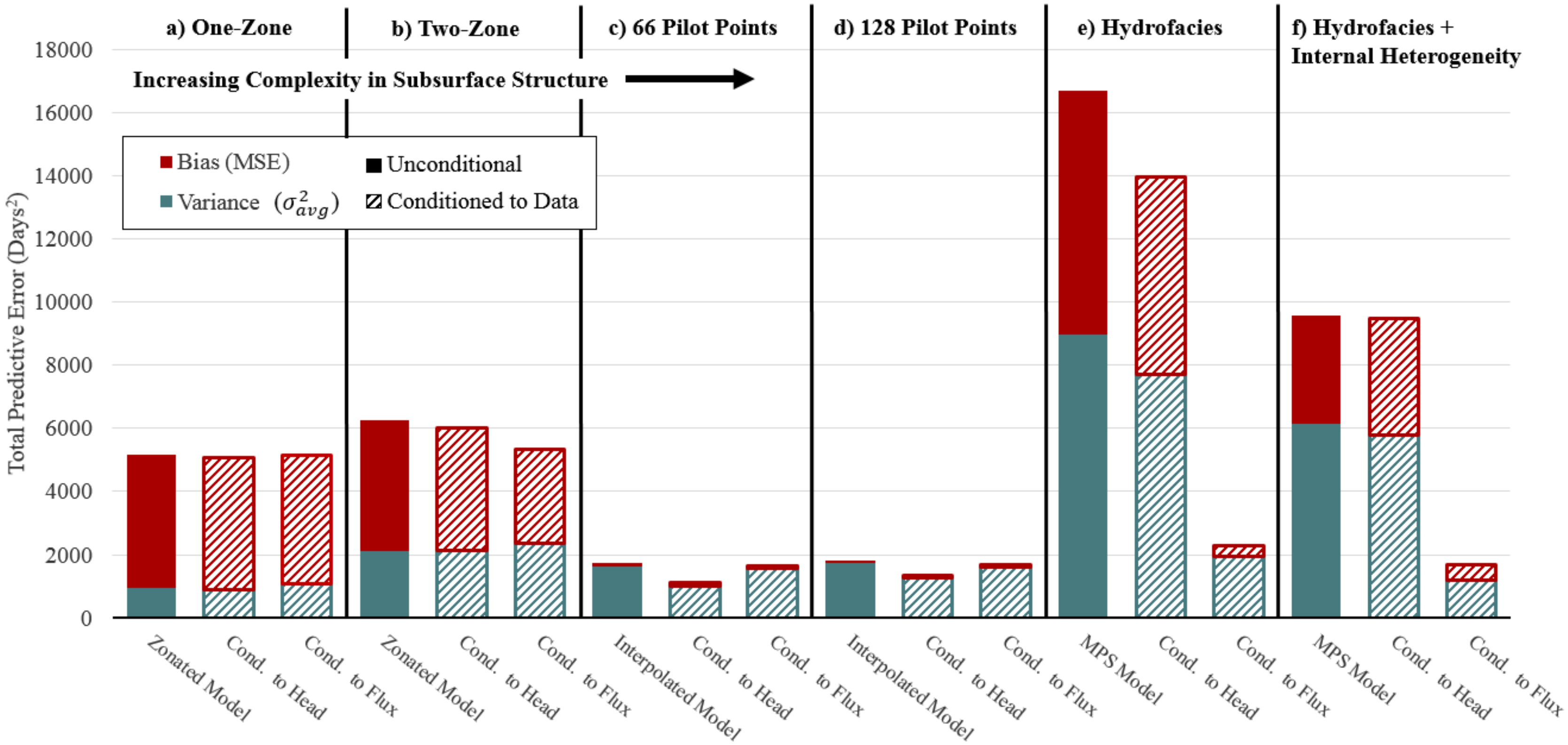


Figure 5.6: Summary of bias and variance predictive errors. Summary of the CDF analysis translated into bias and variance predictive errors presented as a stacked bar plot to compare the total predictive error from different parameter structures. Unconditional errors are shown as solid bars and their response when conditioned to either 14 hydraulic heads or a hyporheic exchange flux measurement are shown as hatched bars. The subsurface parameter structures are arranged by increasing complexity from left to right.

5.5.1 Predictive errors from zonated models

Figure 5.5a shows the transit time distribution summary for the one-zone model, which clearly lacks the ability to recreate the transit time CDF of the virtual reality. From this we can conclude that the one-zone model is too simple to reproduce the virtual reality CDF. Figure 5.6a shows the corresponding predictive error summary for the one-zone model, where we can see a relatively large bias compared to the variance. This indicates that the one-zone model is very inaccurate compared to the virtual reality CDF and that it is over-confident (i.e., erroneous, yet narrow forecast) with a relatively low variance due to its limited K-range. The fact that the one-zone model has essentially no change in its total predictive error when conditioned to data indicates that adding more data to a model with a poor representation of the subsurface structure does not aid in reducing the predictive error. Instead the over-simplified model structure requires revisions through the addition of more subsurface structure, i.e., by adding another zone.

Figure 5.5b shows the transit time distribution summary for the two-zone model, which is able to capture the virtual reality CDF along the outer fringe of its uncertainty range. The uncertainty is quite high, even compared to the later model versions, because K varies globally and has a high sensitivity for all model predictions. The CDF variance after conditioning to head data shows almost no change compared to the unconditional CDF variance. There is some narrowing of the CDF distribution in the early transit times (between 0 % and the 60 % arrival) once conditioned to the hyporheic exchange flux. However, even after the conditioning to a flux measurement, the two-zone model has a relatively high uncertainty with respect to the early transit times. In Figure 5.6b we can see that the variance produced from the two-zone model more than doubles (119 % increase) compared to the one-zone model. This is due to the greater degree of freedom from the additional zone. However, with this added variance we can now see a reduction in bias (17 % decrease) when conditioned to flux data. Therefore, the extra flexibility from the additional zone can now lead to improvements in the model's predictive accuracy.

At first sight it may be disturbing to see how predictive variance does not decrease (or even increases) when conditioning to data (Figure 5.6b, cond. to flux). However, variance is guaranteed to decrease only in a linear inverse problem (Cramér–Rao inequality, Cramér 1946, Rao 1945). In non-linear problems, only the information entropy must decrease, and variance is allowed to increase (Weiss 2005). In the zonated models, the relation between the data, the parameters, the irregularly shaped boundary conditions, and the predictive goal is in fact highly non-linear.

5.5.2 Predictive errors from interpolated models

Clearly there are limitations to simulating hyporheic exchange representing the subsurface as a zonated model. However, representing the subsurface heterogeneity with the detailed characteristics of 3D hydrofacies distribution may be infeasible and/or impractical for most applied hydrogeological studies. Thus, a common intermediary method to generate heterogeneous K-fields is by interpolation using pilot points (Doherty et al. 2010).

Figure 5.5c shows that the interpolated model with 66 pilot points is capable of reproducing the virtual reality CDF. Apparently, by representing heterogeneity, the model is capable of reproducing both the effective mean behaviour of advection (average speed) and of dispersion (general shape of the CDFs). But this improvement goes along with its corresponding uncertainty. While the early transit times ($<$ 40 % arrival) are relatively less uncertain compared to the two-zone model, uncertainty tends to increase after the 40 % arrival, indicating that predicting the late transit times is more uncertain than predicting early transit times. The uncertainty after 40 % arrival appears to be reduced by approximately the same amount by either head or flux data, with slightly greater uncertainty reduction when conditioned to head data.

Figure 5.5d shows the transit time CDF summary when the interpolation points are approximately doubled to 128 pilot points. There is no longer much difference between the unconditional variance and variance conditioned to hyporheic exchange flux, therefore the variance conditioned to hyporheic exchange flux overlaps the unconditional variance which can no longer been seen. This indicates that the 66 pilot point model (Figure 5.5c) has more outlying slower-to-arrive CDFs than the 128 pilot point model.

Figures 5.6c and 5.6d summarize the resulting predictive errors for the 66 pilot point and the 128 pilot point model, respectively. The magnitudes of the total predictive errors between the two models are very similar for both the unconditional errors and errors after conditioning to hydrogeological data. The model error from bias is relatively small compared to the error due to variance. This indicates that the main drawback for this parameter structure is the imprecision in the solution due to nonuniqueness of the heterogeneous two-dimensional K-field and not its ability to produce an accurate transit time CDF.

Figures 5.6c and 5.6d show that the model error reduction is relatively small for both kinds of data, with more variance reduction from conditioning to head measurements. Regardless of the number of interpolation points, there is a greater reduction in variance when conditioned to head data, indicating that hydraulic head data are more informative than flux data for this parameter structure. I expected that flux data would be more informative for all model structures because it is more related to the simulation of transit times, however the results indicate that this is not the case for the interpolated model. I believe this effect is due to two reasons:

1. This parameter structure lacks the confining topsoil layer. Without the topsoil layer, the hyporheic zone acts as an unconfined aquifer and is more sensitive to the horizontal changes in hydraulic conductivity.
2. The method used to generate random heterogeneous K-fields for the interpolated model ensures that the overall average of each realized K-field is close to the assigned geometric mean. Therefore, the hyporheic exchange fluxes produced with the interpolated model are in a relatively narrow range (i.e., approximately 68% of the realizations have fluxes between 1 to 3 L/s) compared to the fluxes simulated by the other model structures. This narrow range of a macroscopic parameter decreases the worth of data that represent a large support volume (like a total flux).

When comparing the variance between the zonated and interpolated models (Figure 5.6), we can see that the variance from the interpolated models is greater than from the one-zone model, however it is less than the variance from the two-zone model. The reason for the lower variance in the interpolated model compared to the two-zone model is because the two-zone model regularly samples K-values from the entire K-range (one order of magnitude for the topsoil and fluvial units), while the interpolated model randomly samples multi-Gaussian K-fields that have an effective mean K-value roughly equal to the geometric mean K. This results in hyporheic exchange fluxes and transit time CDFs that vary more for the two-zone model (Figure 5.5b) compared to the interpolated model (Figures 5.5c and 5.5d). In other words, the initial uncertainty from the interpolated model is less than that of the two-zone model because the randomness in K is more local and so has a smaller global impact.

5.5.3 Predictive errors from MPS models

Figure 5.5e shows the CDF summary varying only hydrofacies distributions. The uncertainty of early transit times (< 15 % arrival) is relatively low. After 15 % arrival, the uncertainty begins to increase dramatically, which means that both the bulk arrival time and the degree of dispersion becomes highly uncertain. Figure 5.6e summarizes the predictive errors and shows the effects of conditioning to hydrogeological data. The CDF distributions are relatively insensitive to horizontal changes in simulated head distributions, thus the total predictive error decreases only by 16 % when conditioned to 14 head measurements. This is because the subsurface parameter structure of the MPS model has the confining topsoil layer (similar to the two-zone model), resulting in a smaller reduction in model errors when conditioned to head data. Conditioning to hyporheic exchange flux, on the other hand, proves to be very informative, reducing the total predictive error by 86 % because it is informative globally on the presence or absence of preferential flow paths. The relative impact of conditioning to different data highlights the fact that, although head data are common calibration targets, their ability to constrain predictive errors for travel time distributions is relatively poor, especially in complex parameter structures. The poor ability for head data to constrain forecasts of flow paths and travel times has been documented in works by Moore et al. (2005), Anderson et al. (2015), and Wöhling et al. (2016).

Figure 5.5f shows the CDF error analysis from hydrofacies distributions including local-scale internal heterogeneity. The patterns for the CDF uncertainty are similar to those for the scenario where only hydrofacies vary (Figure 5.5e), except that both bias and variance are smaller. This becomes very clear by comparing the solid bars in Figure 5.6e and 5.6f. Here we can see that introducing internal heterogeneity reduces the total unconditional predictive error by 43 %. This is counterintuitive because, based on the theory of bias-variance trade-off, it would be expected that additional complexity in the form of local-scale internal heterogeneity would add to the total predictive error in the form of additional variance. Instead, we can see that introducing local-scale heterogeneity reduces both bias and variance. This is because introducing local-scale internal heterogeneity as a multi-Gaussian random field increases dispersion within each realization that smooths and counteracts the drastic effects of the MPS model with variable hydrofacies distributions only. This creates an ensemble of transit time CDFs that are more similar to each another. Another important comparison between Figures 5.6e and 5.6f is that, once the model is conditioned to flux, both total predictive errors are similar in magnitude. This suggests that predictive errors from local-scale heterogeneity are negligible relative to the errors from hydrofacies distributions when simulating transit times at the meander-scale. In other words, the predictive uncertainty due to the 3D distribution of hydrofacies is much greater than the uncertainty from local-scale internal heterogeneity at the meander-scale.

Figure 5.6f shows that, when conditioning the realizations of hydrofacies with local-scale internal heterogeneity to 14 head measurements, there is small reduction in variance but at the cost of an increase in bias. Therefore, there is essentially no reduction in the total predictive error. This indicates that head data is not able to resolve the local-scale heterogeneity in the presence of uncertain hydrofacies distributions. Once more, conditioning to hyporheic exchange flux reduces the total predictive error by 82 %, indicating that calibrating to fluxes has a more significant influence on the predictive accuracy and precision of transit times than hydraulic head measurements.

5.5.4 Optimal subsurface parameter structure and best-fit realization

Thus, from these results (Figures 5.5 and 5.6) one may conclude that the optimum level of complexity to simulate meander-scale transit time distributions of hyporheic exchange would be the interpolated model.

The interpolated model may be able to simulate transit time distributions with relatively low total model error. However, if one were to model more complex processes (i.e., reactive transport) that depend on the spatial distribution of connectivity, subsurface mixing, and contact time to specific subsurface materials, then it is likely that the interpolated model would lead to biases, thus requiring a more detailed representation of subsurface heterogeneity. Therefore, I recommend MPS for generating hydrofacies distributions with or without local-scale internal heterogeneity for reactive transport and uncertainty quantification purposes. Although both the interpolated and MPS models have the problem of not being able to identify a unique parameter distribution based on limited data.

Alternatively, if we were to take a deterministic approach by choosing the realization with the best-fit to hydraulic head (lowest root mean squared error [RMSE]) and meander-scale hyporheic exchange flux (MSHE) for the three different model structures, we can see that the simpler zonated model with two zones (Fluvial $K = 2 \cdot 10^{-3}\, m/s$ and Topsoil $K = 5 \cdot 10^{-4}\, m/s$) is capable of producing results similar to the more complex virtual reality model (Figure 5.7 and Table 5.2). Evidently, the topsoil layer is a critical control on the hyporheic exchange transit time distribution and acts as a confining layer increasing the hydraulic gradient and velocities throughout the hyporheic zone. It is important to note that I assume to have perfect knowledge of the location of the surface that divides the topsoil and fluvial aquifer. It would be interesting to investigate in future work the sensitivity of the transit time CDF to the location and three-dimensional geometry of this surface, however the important point here is that a simpler two-zone model is capable of producing results similar to those of a much more detailed model under the specific model settings in this study.

As an additional check, a $K = 9 \cdot 10^{-3}\ m/s$ for the one-zone model was found to produce the lowest CDF RMSE (35.9 days), however it would produce a hyporheic exchange flux of 14.4 L/s, which is far outside the range of the measured flux. Therefore, the best-fit one-zone model was identified to have a hydraulic conductivity of $K = 2 \cdot 10^{-3}\ m/s$, which has a low RMSE with respect to the 14 head measurements and produces a hyporheic exchange flux that is close to the measured flux. Figure 7 shows CDFs produced from the one-zone model with $K = 9 \cdot 10^{-3}\ m/s$ and $K = 2 \cdot 10^{-3}\ m/s$, clearly showing that neither is capable of capturing the dispersion of the virtual reality CDF.

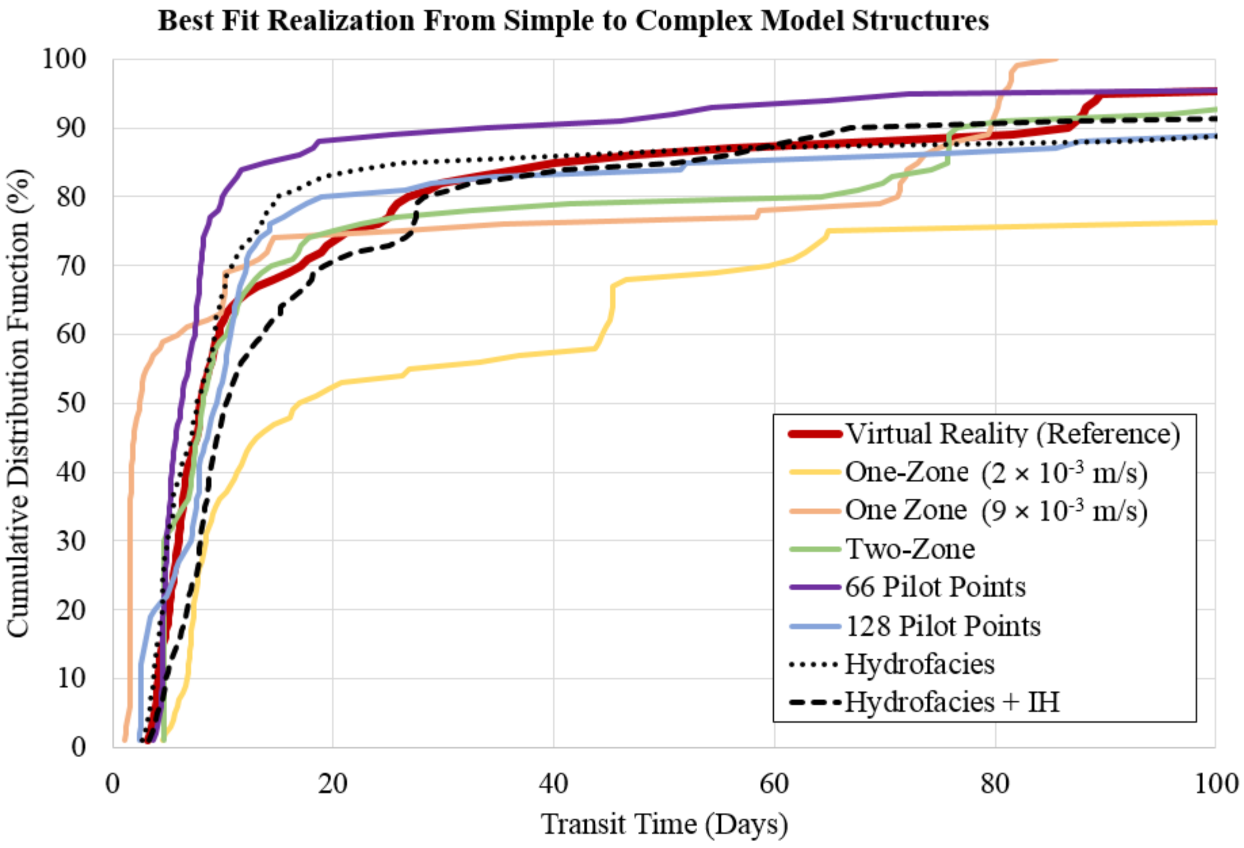


Figure 5.7: Best-fit transit time distribution from individual realization. The model structures are listed from simplest to most detailed, beginning with a model that represents the subsurface hydrostratigraphy as one homogeneous unit to a model with 3D distributions of hydrofacies with internal heterogeneity (IH).

Table 5.2: Output of Best-Fit Realization Differing Model Structures

	Tracer Exp. (Osenbrück et al. 2013)	**Virtual Reality**	**One-Zone ($K = 2 \cdot 10^{-3}$ m/s)**	**Two-Zone**	**Interp. (N_P = 66)**	**Interp. (N_P = 128)**	**MPS (Hydrofacies only)**	**MPS (Hydrofacies + IH)**
10 % Arrival (Days)	**-**	**4.2**	6.8	4.7	4.5	2.6	3.7	4.8
Average Arrival (Days)	**7.8**	**7.9**	17.0	8.2	6.3	9.4	7.7	10.3
90 % Arrival (Days)	**-**	**87.5**	302.0	76.6	34.0	102.5	103.9	66.9
CDF RMSE (Days)	**-**	**-**	99.4	16.5	15.8	17.4	10.0	33.1
MSHE Flux (L/s)	**1.0 to 2.0**	**1.9**	2.1	1.9	1.7	1.4	1.9	2.0
RMSE Head (cm)	**-**	**-**	29	8	6	8	5	8

This additional analysis shows that a simple zonated model of meander-scale hyporheic exchange is capable of simulating most of the variability in transit times when compared to a model with much greater structural complexity, as long as it incorporates the most important physical controls. In this case it appears that the topsoil layer is an important physical control that confines and changes the hydraulic head distributions throughout the hyporheic zone. The zonated model fails to capture all the inflection points or “kinks” in the CDF produced by a model structure with greater complexity because it lacks the preferential flow paths. Therefore, it is likely a zonated model will bias the amount of subsurface mixing and contact time with specific material types, thus rendering it unsuitable for reactive transport modelling. However, if one were to apply the principle of parsimony and were only concerned with estimating the transit time distributions of meander-scale hyporheic exchange it appears to be possible with a zonated model here.

Interestingly, the interpolated model can produce a similar result without the topsoil layer, showing that a heterogeneous 2D K-field has enough parametric freedom to compensate for the lack of 3D hydrogeological knowledge. The best-fit realizations from the interpolated models reveals that kriging with a large number (128) of pilot points can create a CDF with inflections or “kinks” that can be interpreted as preferential flow paths, while decreasing the number of interpolation points by about half results in a CDF that is smoother and more dispersive (kinkless). When we compare it to the best-fit realizations of the MPS models, we see CDFs similar to the virtual reality, however with differing inflection points. Since the best-fit realization from the MPS models are "cut from the same cloth" or produced by the same method, it is not hard to believe that this method can produce CDFs that compare well to the virtual reality CDF. However, in real-world applications we are unable to resolve the subsurface properties in as much detail as they are in real systems. Therefore, results from the MPS model are only valid as an academic benchmark. Furthermore, the number of realizations for the interpolated and MPS models were limited in this study. If I were to run a formal optimization, I would likely encounter a better fit realization that exists in their vast parameter space. Nonetheless, the point here is that the interpolated and MPS models are capable of producing transit time CDFs with qualities consistent with the dispersion from the virtual reality model.

5.6 Discussion and Conclusions

This chapter was motivated by the research question (Q3, Section 1.5): what is the recommended degree of subsurface detail necessary to simulate meander-scale hyporheic exchange in a physically meaningful way? In order to answer this question, it was necessary to evaluate the trade-offs from choosing between different subsurface parameterization structures. Table 5.3 presents a summary of the trade-offs that come with choosing one subsurface parameter structure over the other. The results have clearly shown that the one-zone model cannot recreate the virtual reality transit time CDF and therefore should not be considered. The two-zone model is capable of producing results that fit the available field data and can approximate the transit time CDF of the virtual reality model, while also providing the most unique and parsimonious solution. However, the two-zone model has a relatively large bias and makes little use of hydrogeological data. The interpolated model can recreate the results of the virtual reality model with the lowest total model error. However, the interpolated model does not provide a unique solution. The MPS model also has the problem of a non-unique solution, has relatively high total predictive error with respect to hydrofacies distributions, but makes the greatest use of a hyporheic exchange flux measurement in constraining transit time CDFs.

Table 5.3: Summary of Trade-Offs from Differing Subsurface Parameter Structures

Subsurface Parameter Structure	**Scenarios within structure**	**Fits available field data**	**Predictive coverage of CDF ensemble**	**Most unique and parsimonious**	**Lowest total predictive error**	**Approx. mechanical dispersion / preferential flow paths**	**Makes best use of hydrogeo. data**
Zonated	**One-Zone**	✗	✗	✗	✗	✗	✗
	Two-Zone	✓	✓	✓	✗	✗	✗
Interpolated	**66 Pilot Points**	✓	✓	✗	✓	✓	✗
	128 Pilot Points	✓	✓	✗	✓	✓	✗
Multi-Point Geostatistics	**Hydrofacies**	✓	✓	✗	✗	✓	✓
	Hydrofacies + Internal Heterogeneity	✓	✓	✗	✗	✓	✓

From this analysis we have seen that choosing between different structural representations of subsurface parameter distributions when simulating meander-scale hyporheic exchange transit time CDFs leads to trade-offs between intrinsic and epistemic model errors (Rubin 2003, Anderson et al. 2015). Intrinsic model errors are the structural limitations imposed by the chosen representation of subsurface parameters that cannot be reduced by more knowledge (i.e., new measurements), while epistemic model errors are potentially reducible with new knowledge. Thus, from the range of parameter structures presented in this study, we can see that the simplest model (i.e., the one-zone model) is constrained by intrinsic model errors because it lacks the flexibility to adapt to additional hydrogeological information, therefore its predictive errors are irreducible. Only by increasing complexity through the addition of zones or heterogeneity do we begin to exchange intrinsic model errors for epistemic ones. We see this in the two-zone scenario, where an additional zone leads to a reducible model bias when conditioned to hydrogeological data, at the cost of greater variance from the added flexibility in the model structure. At the other end of the spectrum, we see that the model errors from a hierarchical 3D MPS model are potentially very large but also very reducible from the right choice of additional hydrogeological data.

In general, we have seen evidence of the classic bias-variance trade-off when comparing the different parameter structures, but have also seen that quantifying this trade-off is complex and can lead to counter-intuitive results. Furthermore, this study shows that it is necessary to incorporate measures of data worth when comparing conceptual model structures, because this makes it possible to distinguish model errors caused by the model structure from model errors caused by a lack of data, and thus allows for the quantification of the trade-offs between intrinsic and epistemic model errors.

Another general trend in the predictive uncertainty of hyporheic exchange transit time CDFs, regardless of the parameter structure, is that early-times are relatively less uncertain and that uncertainty tends to increase with increasing transit time. The greatest uncertainty appears to be within the tailing of the transit time distribution or at late-times. This is consistent with reviews of water transit time analysis and modelling by McGuire and McDonnell (2006) and McDonnell et al. (2010), who state that the identification of late transit times is especially uncertain because they are often associated with small proportions of the total mass flux.

Nonetheless, my results indicate that the transit time distributions of meander-scale hyporheic exchange at the Steinlach River Test Site are above the threshold (Kiel et al. 2014) necessary to favour denitrification (and thus nitrogen loss), given that the right biogeochemical conditions are present in the subsurface. Similar to Kiel et al. (2014), I believe that my findings here only serve as a lower limit to the potential metabolic activity because I focus on meander-scale exchange. Therefore, I am likely underestimating the potential reactivity by neglecting the nested vertical bedform-scale exchange (Chow et al. 2018). Gomez-Velez et al. (2015) have consistently observed higher denitrification from vertical bedform-scale exchange in the Mississippi River network, thus a future research direction would be to investigate bedform-scale exchange in more detail to understand its contribution to overall catchment reactivity (Grathwohl et al. 2013).

Taking a deterministic approach and following the principal of parsimony when selecting a subsurface parameter structure reveals that a single best-fit realization of the two-zone model would be capable of reproducing the results of a more complex model that includes the 3D distributions of hydrofacies. However, a zonated model would likely lead to biases if one were to model more complex processes (i.e., reactive transport) that depend on the spatial distribution of connectivity, subsurface mixing, and contact time to specific subsurface materials. Nonetheless, with adequate geological knowledge (i.e., contacts between different zones) it appears that a well-done zonated model is capable of producing a transit time CDF similar to one produced by a more detailed subsurface parameter distribution. Similarly, Wondzell et al. (2009) found that increased model sophistication when simulating hyporheic exchange in a small mountain stream does not necessarily lead to improved results, and that transit time forecasts from a homogeneous model were equal to, or better than the predictions from a heterogeneous model. A caveat is that my result is case-specific, however a zonation-based model has been shown to be a robust choice (given adequate geological knowledge) when challenged by limited data (Schöniger et al. 2015).

The interpolated model provides a tractable solution with relatively low model errors (mostly in the form of variance). However, this parameter structure requires enough data to calculate geostatistical properties, which in this case was 14 boreholes. Furthermore, this parameter structure does not provide a unique solution, but rather works better with an ensemble of plausible realizations. The results can then be averaged and/or provide confidence intervals.

The MPS model provides the greatest flexibility and provides a solution that is consistent with representations of subsurface hydrogeological structures. However, this method requires a defensible training image that represents the site geology and thus requires the highest degree of geological knowledge, which may be uncertain and based on imperfect assumptions. In response to the uncertainty inherent to using a single training image, Khodabakhshi and Jafarpour (2013) have developed a Bayesian mixture-modeling approach that adaptively and efficiently samples conditional facies from multiple uncertain training images. Given there is a defensible training image (or an ensemble of training images), the greatest predictive errors come from the MPS model when defining the distributions of hydrofacies within a heterogeneous meander-scale aquifer. Predictive errors from hydrofacies distributions, although very large, can be reduced by informative data. These results indicate that an estimate of hyporheic exchange flux is much more informative than a small number (14) of head measurements when simulating transit time CDFs with 3D hydrofacies distributions.

Furthermore, these results indicate that the uncertainty from local-scale heterogeneity within the hydrofacies is negligible compared to the uncertainty with respect to the hydrofacies distributions. This is largely due to the meander-scale focus of this study. I suspect that as the scale of the modelling problem increases (Figure 5.8), uncertainty from larger-scale hierarchical elements would dominate (Heinz and Aigner 2003). For instance, hydrofacies distribution could play a lesser role in larger catchment-scale problems than the spatial distribution of geological contacts between aquifer compartments. If my proposition is true, then this could be an important insight for model up/down-scaling, because it would be wrong to assume that results from a sensitivity or uncertainty analysis at a certain scale is transferable to another scale. Thus, a separate sensitivity or uncertainty analysis would be necessary when changing the model scale.

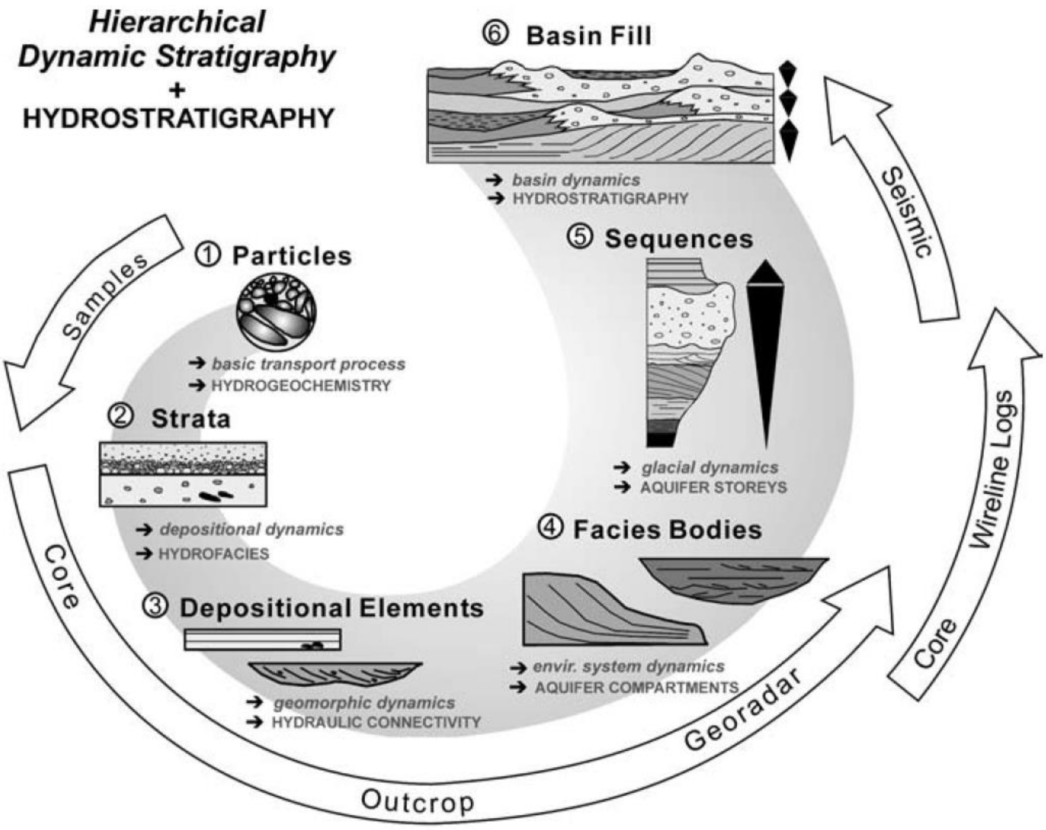


Figure 5.8: Concept of 'dynamic stratigraphy' as applied to the characterization of gravel deposits. The hierarchy of spatial and temporal scales is systematically studied in a process-based analysis; each scale has implications for hydrogeological properties. Note that different investigation tools are necessary for each scale (Heinz and Aigner 2003).

This study provides a clear example of the trade-offs between intrinsic and epistemic model structural errors when evaluating subsurface parameter structures to simulate hyporheic exchange. I have demonstrated that choosing the appropriate level of detail to represent the subsurface parameter distributions will depend on the acceptable range of intrinsic model errors for the given modelling objectives and the available site data.

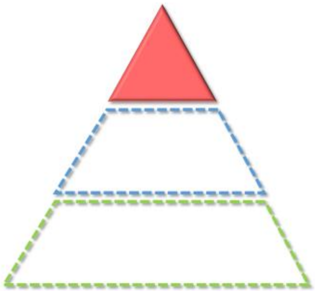
5.7 Bathymetry sensitivity vs. subsurface structure uncertainty

Results from Chapters 4 and 5 indicate that one could achieve results similar to a more detailed model by simply using the thalweg-gegenweg bathymetry (54 instead of 540 bathymetric survey points) and a two-zone (top soil and fluvial aquifer instead of hydrofacies MPS with internal heterogeneity) depiction of the subsurface *K*-distribution. Therefore, by conducting a sensitivity and uncertainty analysis of the bathymetry and subsurface heterogeneity, respectively, the critical few factors that produce the majority of the results (i.e., meander-scale hyporheic exchange flux and transit time distribution) was revealed.

Of course, this is a study specific result and I would expect that the critical factors would likely be different at different sites and scales. Furthermore, there is the underlying assumption that the benchmark 'virtual reality' model is sufficiently representative of a real site. However, the point is that sensitivity and uncertainty analyses are powerful tools that can be applied to a model to potentially identify the critical factors that control the model response for a forecast of interest. With these vital few factors in-hand one can more efficiently and effectively use resources to prioritize investigations into the things that matter most and in doing so save time, money, and perhaps help reveal more "unknown unknowns" (Section 1.1.1).

Conclusions from Chapters 4 and 5 indicate that both bathymetry and subsurface distribution of hydraulic conductivity play an important role when modelling hyporheic exchange at the meander-scale. However, it is far easier to accurately measure channel bathymetry than to accurately measure the distribution of subsurface properties. Furthermore, the feasible parameter space for the distribution of subsurface properties is much vaster than bathymetry. Following the principle of parsimony (Section 5.1) for model analysis, I would recommend conducting a sensitivity analysis first before an uncertainty analysis, because the modelling objectives may be achieved through a sensitivity analysis alone. Afterwards, if the results from a sensitivity analysis are deemed insufficient or inconclusive and an uncertainty analysis is required, the results of the sensitivity analysis could help guide the design and focus of the uncertainty analysis to a vital few tractable modelling processes, parameters, and interactions.

6. Simulating bedform-scale hyporheic exchange in a bedrock river: The Eramosa Bedrock River Site (EBRS)



Chapter highlights

- When challenged by limited data, a stochastic discrete fracture network framework can be used to represent the subsurface fractured bedrock connectivity. Thus, producing probabilistic distributions of the potential hyporheic exchange extents and residence times in a fractured rock setting.
- The coincidence of fractures and hydraulic gradients determine the spatial extents of bedform-scale hyporheic exchange in a fractured rock setting.
- Hyporheic exchange residence times in bedrock rivers at the bedform-scale are potentially orders of magnitude longer when compared to fluvial rivers (i.e., months to years vs. minutes to hours).

6.1 Motivation and research questions

Bedrock rivers represent a unique hydrogeological environment, in which the surface water flows along an exposed bedrock surface (i.e., the riverbed lacks a layer of unconsolidated fluvial sediment). While bedrock rivers are found throughout the world, the number of documented studies is limited (Figure 1.8). This may be due to the difficulties that arise from characterizing the geology and sw-gw exchange within bedrock river systems. Thus, most previous studies of hyporheic exchange are within settings with unconsolidated sediments (e.g., Vaux 1968, Cardenas 2008, Stonedahl et al. 2010). The water that is associated with hyporheic exchange in fluvial river settings (Figure 1.9) has the potential to undergo biogeochemical reactions, which can be attributed to the retention and degradation of pollutants within watersheds (Kalbus et al. 2006, Grathwohl et al. 2013, Trauth et al. 2018). However, it is unclear whether the same can be said for hyporheic exchange (if it occurs at all) within bedrock rivers. Thus, there remains a knowledge gap with respect to the nature of hyporheic exchange in bedrock rivers.

A field investigation between a bedrock river, Twenty Mile Creek, and a local aquifer in

Smithville, Ontario, Canada by Oxtobee and Novakowski (2002) provides some important insights about sw-gw interactions in bedrock rivers. The intent of their investigation was to provide a better understanding of contaminant migration near bedrock rivers, thus hyporheic exchange within bedrock rivers was not their main focus. Nevertheless, their investigation revealed that groundwater discharge was extremely limited within Twenty Mile Creek, with more than 95 % of the groundwater underflowing the creek during baseflow condition. Furthermore, their results indicated that groundwater discharge into Twenty Mile Creek was primarily through discrete fractures and that poor vertical connections in the fracture network limited groundwater exchange. Next, Oxtobee and Novakowski (2003) developed a numerical model based on their field investigation, which indicated that groundwater discharge (or recharge) in bedrock rivers depends on the aperture size of the discharging features and the vertical hydraulic head distribution within the fracture network.

In principle, the same hydrostatic and hydrodynamic drivers for hyporheic exchange at rivers with unconsolidated sediments can drive hyporheic exchange in bedrock rivers. However, it would be incorrect to assume that the same relationship between spatial scales and residence time scales for rivers with unconsolidated sediments (i.e., Figure 1.3) applies to bedrock rivers (Vaughn 1990, Tinkler and Parish 1998), simply because subsurface flow in fractured media behaves differently than flow in unconsolidated porous media (MacQuarrie and Mayer 2005, Singhal and Gupta 2010). The motivating research question is (RQ4, Section 1.5): how does a bedrock river system respond differently to the same hyporheic exchange drivers? Finding ways to simulate hyporheic exchange in bedrock rivers, which incorporate our current knowledge of surface flow hydraulics and groundwater flow through discrete fracture networks, can help us identify the factors that control the spatial extents and residence times of hyporheic exchange within bedrock rivers.

This chapter examines the physical properties that govern hyporheic exchange along a fractured bedrock river through the use of a manually coupled numerical sw-gw model. The aim is to understand the role of fracture network heterogeneity on the formation of hyporheic exchange flowpaths and groundwater residence time distributions associated with fracture-dominated bedrock river flow systems. In order to simulate this in a physically meaningful way, it is important to support my modelling with hydrogeological data that represents a real-life bedrock river flow system, namely the Eramosa Bedrock River Site (EBRS) located within the Barber Memorial Scout Camp (SC) site near the City of Guelph, Ontario, Canada.

6.1.1 Site description, geological setting, and previous studies

The Eramosa River is a major tributary of the Speed River within the Grand River watershed near the City of Guelph, Ontario, Canada (Figure 6.1). The Eramosa River is underlain by two regional bedrock aquifers, known as the Guelph-Amabel Aquifers (Cole et al. 2009). The Guelph-Amabel Aquifers consists of densely fractured dolostone with dissolution-enhanced karst features (Kunert et al. 1998, Kunert and Coniglio 2002) and are the sole drinking water sources for the City of Guelph and surrounding region (i.e., Wellington County). Groundwater presently discharges from these regional bedrock aquifers into the valleys of the Speed and Eramosa Rivers and from pumping wells completed in the aquifers (Cole et al. 2009). The potential effects of increased groundwater pumping on the bedrock rivers and the surrounding ecosystem is poorly understood (Steelman et al 2017b). Understanding the consequences of groundwater withdrawal from these regional aquifers and its potential impact on local surface water ecosystems was the initial motivation for the University of Guelph's G^{360} Institute for Groundwater Research, who have been conducting hydrogeological and geophysical investigations at the Eramosa River since 2010 (Steelman et al. 2015a and 2015b, Kennedy 2017, Steelman et al. 2017b, Capes et al. 2018).

Locally the Eramosa River incises the Eramosa Formation by about 2 to 3 m, exposing vertical and horizontal fractures with little to no fluvial deposits along the riverbed (Steelman 2017b). The Eramosa Formation is up to 11 m thick and is composed predominantly of fractured dolostone, acting regionally as a discontinuous aquitard between the Guelph-Amable Aquifers (Cole et al. 2009). Underlying the Eramosa Formation is approximately 3 m of the cherty, marble-like Goat Island Formation (Steelman et al. 2017b), which is underlain by more than 15 m of Gasport Formation (Figure 6.2). Brunton (2009) provides a full description of these bedrock formations.

The Eramosa Bedrock River Site (EBRS) is a riffle-pool-riffle sequence approximately 60 m in length within a channel meander of the Eramosa River that has an intact dolostone riverbed (Figure 6.1). An important note is that the words "riffle" and "pool" are borrowed from geomorphologist and river morphologists, which are well-defined and depict unconsolidated sedimentary features in fluvial rivers. For convenience I have used the terms to mean riffle-like and pool-like bedform features due to their geometric similarities. However, my use of the words riffle and pool should not be misinterpreted to be a statement about geomorphological provenance. Steelman et al. (2017b) believes that the bulk of the sw-gw exchange occurs along the fracture networks driven by a highly variable head distribution, and that matrix storage could potentially support complex biogeochemical processes.

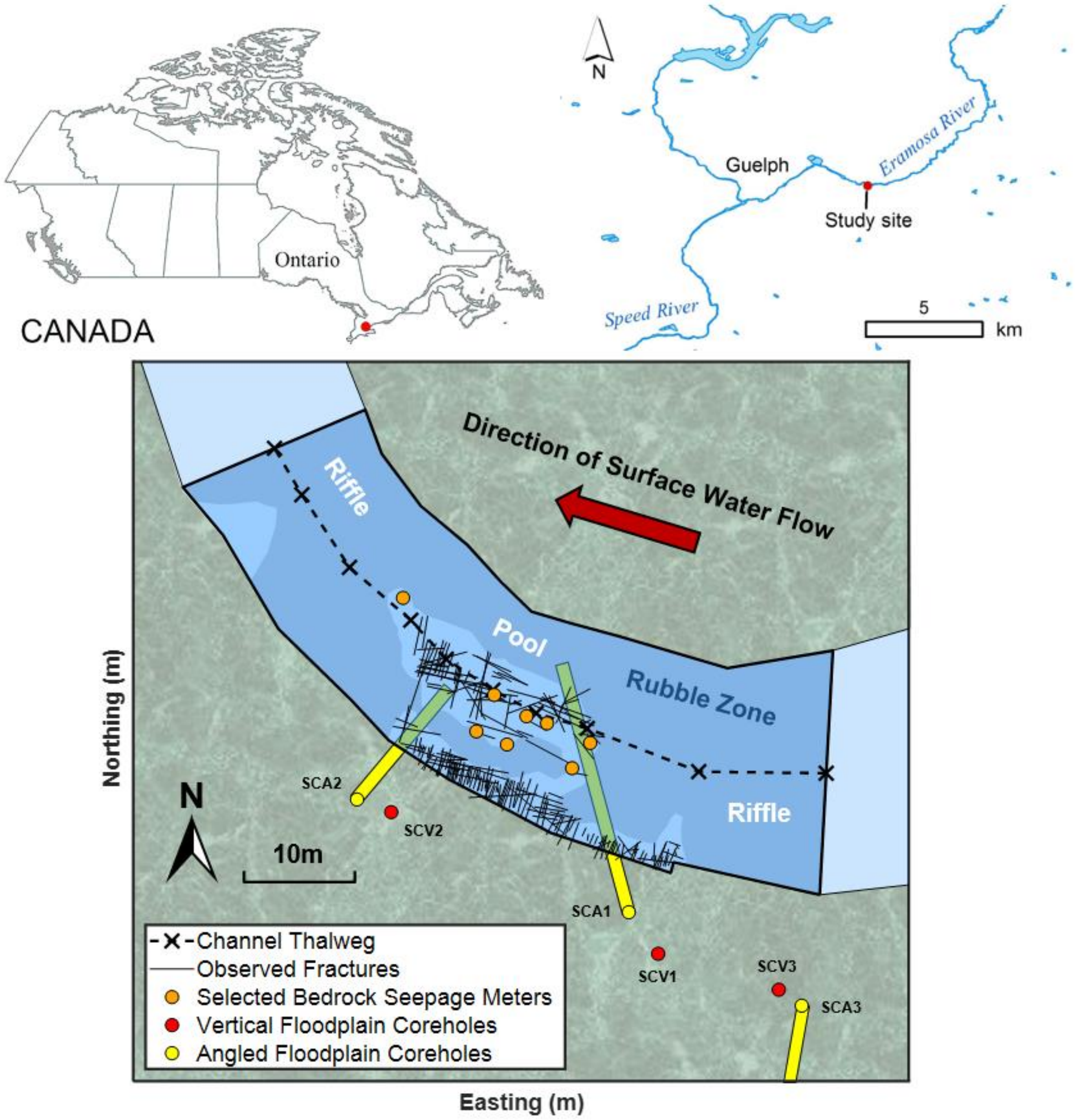


Figure 6.1: Location of Eramosa Bedrock River Site (EBRS), near Guelph, Ontario, Canada.The Eramosa River flows in a southeasterly direction, and encompasses a riffle-pool-riffle sequence (dark blue) within a river meander (adapted from Kennedy 2017 and Steelman et al. 2017b). SCA1 and SCA2 are angled beneath the river (green shading). The light blue areas within the EBRS are rubble-free zones where surficial fracture trace lengths were mapped (Kennedy 2017).

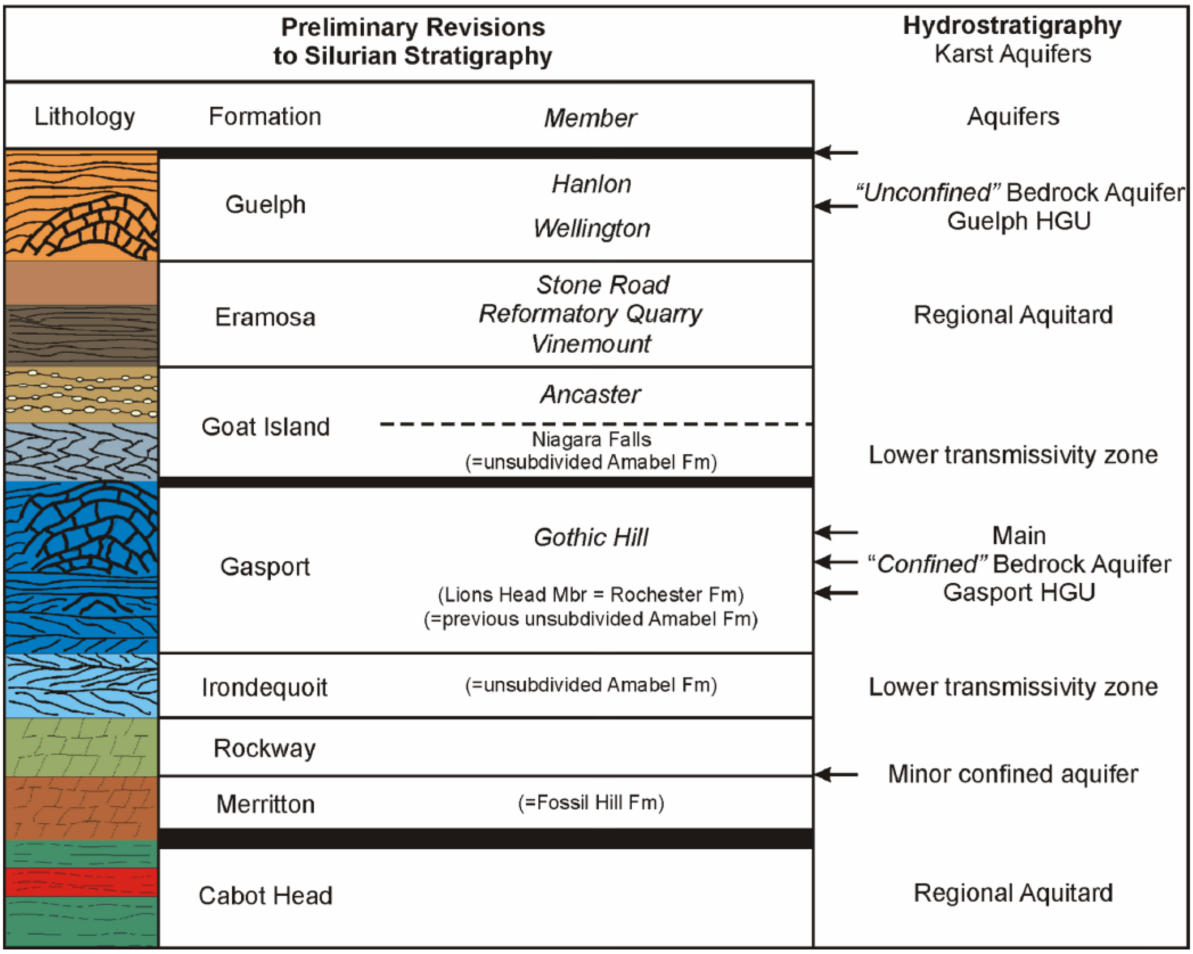


Figure 6.2: Sequence stratigraphic framework for Guelph and surrounding areas of the Niagara Escarpment. Relative thickness of lines separating formations reflects the greater geological time break. The Vinemount Member (Mb) and calcareous shales of the Eramosa Formation (Fm) are a key aquitard. The Niagara Falls Mb of the Goat Island Fm has a relatively low transmissivity. The key regional hydrogeological units include the Gasport Fm and unconfined aquifer known as the Guelph interface aquifer (taken from Brunton 2009).

Kennedy (2017) developed a field-based conceptual model of the EBRS by examining sw-gw exchange using seepage meters especially designed for bedrock rivers, known as bedrock seepage meters (BSM). BSMs are based on seepage meters originally designed for rivers and lakebeds composed of unconsolidated sediments (Lee and Hynes 1978). BSMs (Kennedy 2017) can be used to measure vertical and horizontal flow along horizontal (bedding plane fractures) and vertical (joint fractures) flow paths within a fractured sedimentary bedrock river.

Kennedy (2017) installed 24 BSMs within the riverbed (~25 cm deep) of the EBRS pool (Figure 6.1). Each BSM measured groundwater seepage fluxes and hydraulic heads over a 28-day period beginning on July 7th, 2014 and ending on August 18th, 2014. Corehole (i.e., a borehole with continuous recovered core) data and a survey of exposed fractures along the streambed were also included in the EBRS conceptual model (Figure 6.1).

While monitoring by Kennedy (2017) indicated the existence of sw-gw interactions at the EBRS (i.e., local upward gradients and groundwater discharge within the river), it remains unclear whether bedform-scale hyporheic exchange (i.e., closed-loops of downwelling and upwelling within the EBRS) or merely groundwater discharge was occurring. Therefore, by using a physically-based (as opposed to empirically-based) numerical modelling approach, I intend to simulate potential bedform-scale hyporheic exchange at the EBRS informed by the previous field investigations. Insights gained from the modelling could determine whether hyporheic exchange is possible at the EBRS, and if so, the factors that control hyporheic exchange within bedrock rivers. Furthermore, modelling the EBRS could inform potential future site investigations (e.g., locations for additional piezometers, stream gauging, fracture characterization), including conservative (bromide) and reactive (resazurin) tracer experiments to evaluate the presence (or absence) of hyporheic exchange and potential biogeochemical turnover within the hyporheic zone (Knapp et al. 2017).

6.1.2 Site conceptual model and modelling approach

A site conceptual model of the sw-gw flow system of the EBRS is shown in Figure 6.3. My conceptual model consists of a 2D cross-sectional representation of a 61 m long riffle-pool sequence following the thalweg (see Chapter 4) of the EBRS. The focus of the numerical model is to investigate potential bedform-scale hyporheic exchange between the Vinemont Member (Mb), a subunit of the Eramosa Formation (Brunton 2009, Steelman et al. 2017a), and the surface water of the Eramosa River. Therefore, my conceptual model extends vertically in depth until 303.0 meters above sea level (masl), which is the average elevation of the contact between the Vinemount Mb and underlying Goat Island Formation (Fm).

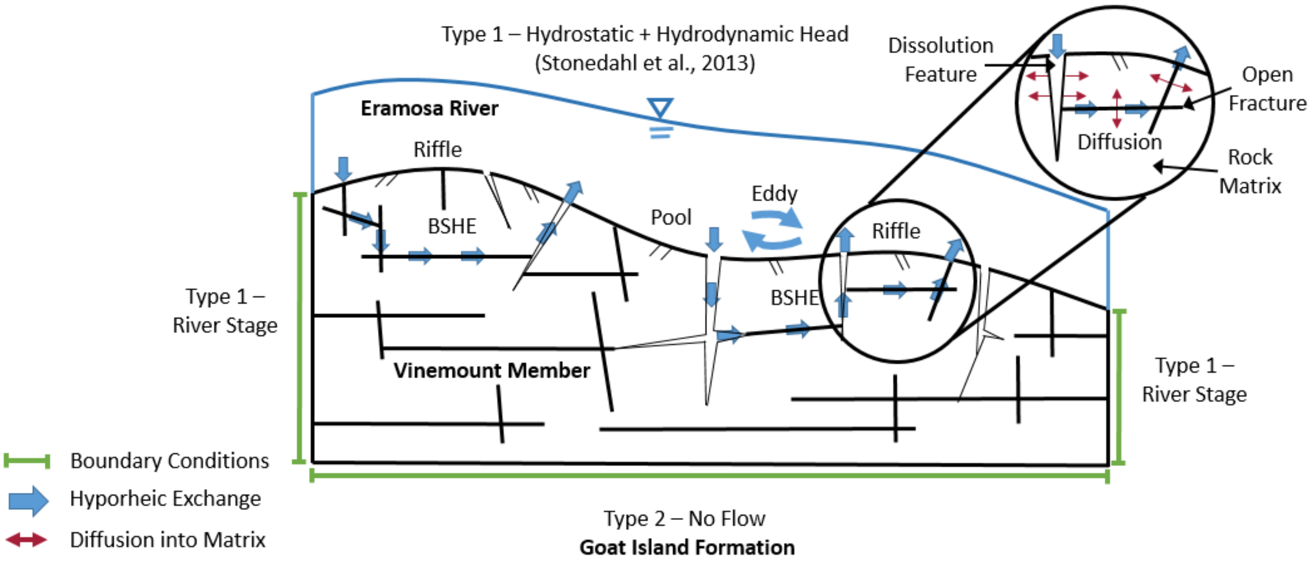


Figure 6.3: Conceptual model of bedform-scale hyporheic exchange (BSHE) at the Eramosa Bedrock River Field Site, a riffle-pool-riffle sequence in a fractured sedimentary bedrock river (approximately 61 m in length). Hyporheic exchange is controlled by the open fractures, which are more hydraulically conductive than the surrounding matrix. Open fractures will be hydraulically connected to the porous matrix, and thus, support potential solute exchange (i.e., diffusion) between the two flow regimes.

Subsurface properties can be represented as an equivalent porous medium (EPM) where a fractured porous rock is characterized by equal fracture and matrix properties, which could be obtained by weighted averaging of fracture and matrix properties such as porosity, tortuosity, and mineral surface areas (MacQuarrie and Mayer 2005). An EPM approach may be reasonable for bulk groundwater flow in settings with ubiquitous fractures that form well-interconnected fracture networks. However, it is inadequate for simulating transport because it would neglect diffusive interaction between the fractures and matrix (Chapman et al. 2014). Furthermore, an EPM approach has been shown to underestimate the variability and average properties associated with transport mechanisms (e.g., matrix diffusion, sorption on fracture walls), which leads to the inability of capturing local-scale measurements or tests at discrete locations within the subsurface (Cacas et al. 1990). MacQuarrie and Mayer (2005) have also noted that a consequence of averaging subsurface properties in fractured rock is that solute travel times (especially non-reactive ones) can be underestimated in the case of fracture-dominated transport.

Due to the aforementioned inadequacies of representing fractured rock systems with an EPM approach, and our current understanding about sw-gw interactions within bedrock rivers (Tinkler and Wohl 1998, Oxtobee and Novakowski 2002 and 2003, Steelman et al. 2017b), I chose to represent fractures and matrix as distinct features and to simulate coupled flow and transport in both the fractures and matrix, simultaneously. This is known as a discrete fracture network (DFN) approach (MacQuarrie and Mayer 2005). A major issue is that 3D DFN simulations at scales relevant to contaminant plumes are computationally expensive. As a possible remedy, Chapman et al. (2014) have proposed a combined EPM-DFN approach to simulate transport in sedimentary bedrock aquifers. The EPM-DFN approach involves using a 3D EPM model that incorporates macro-complexity (e.g., hydrogeological units) to capture bulk groundwater flow, then to identify the predominant plume flow path using forward particle tracking from the source zone. Once the plume flow path has been identified, a 2D DFN cross-sectional approach is used to model transport along the plume flow path, incorporating fracture and matrix properties that influence plume transport.

I can fulfill my study objectives by modelling flow and transport within a 2D DFN cross-section. For this specific site, a full 3D EPM model of the catchment would be unnecessary to identify the predominant groundwater flow direction (i.e., river flow direction), therefore I am only using the latter DFN half of the EPM-DFN approach proposed by Chapman et al. (2014). I will develop a 2D DFN flow and transport model using HGS (Aquanty Inc. 2015, Section 2.1) to simulate sw-gw exchange along the EBRS thalweg.

A major drawback in the application of a DFN approach for contaminant transport is the challenge of obtaining the necessary input parameters (e.g., fracture position, density, angles, apertures, network connectivity, rock matrix parameters) and datasets on solute concentration distributions to calibrate transport parameters (Chapman et al. 2014). Due to these issues with DFN identifiability (see also Section 1.1.2, Sun and Yeh 1990) I have chosen to model the DFN within a stochastic framework. This means that I will represent the subsurface with a suite of statistical DFN realizations, each generated randomly, but in-line with the observed spatial fracture properties from local site measurements. Probability distribution functions of hyporheic exchange residence times and extents can be produced from a suite of statistically valid (with respect to fracture properties) DFN realizations. This could lead to physically meaningful ranges (i.e., within realistic orders-of-magnitudes) of bedform-scale hyporheic exchange fluxes and residence times at the EBRS and similar bedrock rivers. The following sections will describe the field data collected from previous site investigations at the EBRS (Colby et al. 2015 and 2015b, Kennedy 2017, Steelman et al. 2017b, Capes et al. 2018) and the process in which they were reduced to inform the 2D DFN stochastic modelling.

6.2 Fracture properties of the Eramosa Bedrock River Site

The following subsections describe the reduction of data from previous field investigations conducted by the University of Guelph's G^{360} Institute for Groundwater Research (Colby et al. 2015 and 2015b, Kennedy 2017, Steelman et al. 2017b, Capes et al. 2018). I start with describing the geotechnical layers identified by a cumulative fracture intensity with depth analysis (Section 6.2.1), followed by an analysis of fracture dip data (Section 6.2.2). Next, I reduce fracture frequency data from coreholes and surficial riverbed surveys (Section 6.2.3). Then, I calculate fracture aperture statistics from hydraulic packer tests and bedrock seepage meter measurements (Section 6.2.4). Finally, I obtain all other necessary model parameters by reviewing site-relevant literature (Section 6.2.5).

6.2.1 Geotechnical layers of the Vinemount Member

Correctly identifying significant mechanical stratigraphy in rock, their lateral persistence (i.e., spatial distribution), and their potential vertical connectivity is known to greatly improve predictions of subsurface flow and transport for petroleum reservoir engineering, proposed high-level nuclear waste storage facilities, and other types of large subsurface engineering projects in hard rock (La Pointe 2010). This led me to subdivide the Vinemount Mb into mechanical layers (ML), which are intervals of rock that have similar fracture properties (Underwood et al. 2003, Shackleton et al. 2005). Mechanical layers were delineated by identifying inflection points on a cumulative fracture intensity (CFI) vs. depth plot (La Pointe 2010). CFI plots for all six coreholes were generated by ordering all identified fractures by elevation, then assigning them a rank (La Pointe 2010). For instance, the shallowest fracture was given a $\mathrm{rank} = 1$ and the next shallowest fracture was given a $\mathrm{rank} = 2$, with the deepest of n fractures given a $\mathrm{rank} = \mathrm{n}$. Rankings are then normalized by the total number of fractures n for each corehole to create a CFI value, which is plotted against elevation as shown in Figure 6.4.

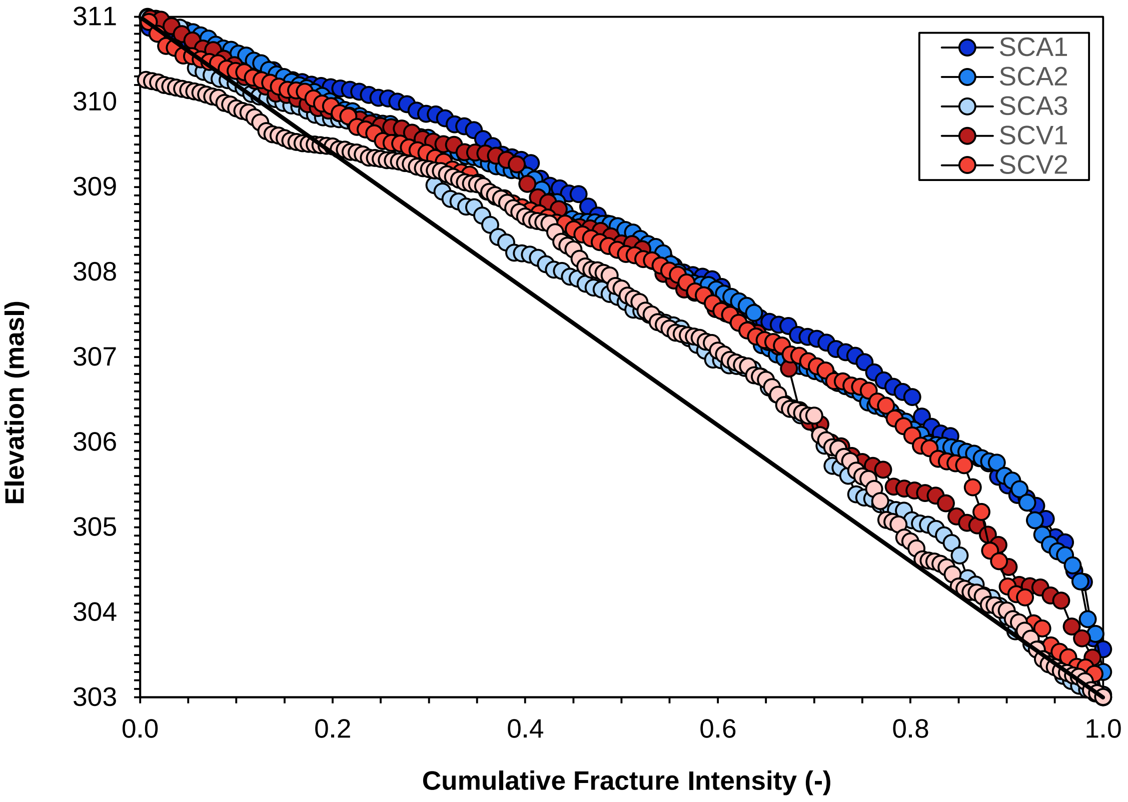


Figure 6.4: Cumulative fracture intensity of total fractures for six coreholes. The black line represents constant fracture intensity. Blue dots (SCA1, 2, and 3) are from the angled coreholes and red dots (SCV1, 2, and 3) are from the vertical coreholes.

To determine the inflection points, the six CFFs were then combined by averaging across fixed depth intervals of 0.1 m along the Vinemount Mb to a single plot representing the average cumulative fracture intensity with elevation (Figure 6.5). I identified four inflection points (red dots in Figure 6.5) from the average CFI plot, which represents the contact between five distinct mechanical layers. The rest of this study will use these five mechanical layers to parameterize my Base Case discrete fracture network (DFN) model.

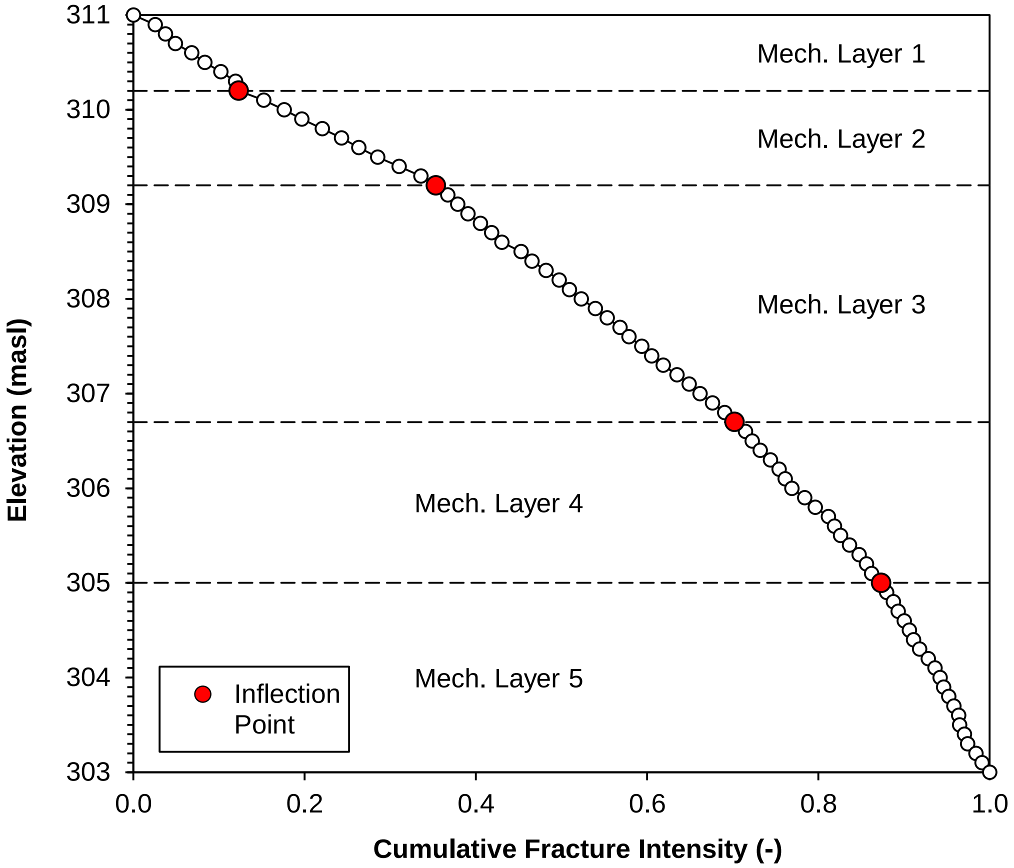


Figure 6.5: Cumulative average fracture intensity of total fractures with depth. Normalized rankings have been averaged across all six coreholes over fixed depth intervals of 0.1 m. Inflection points are marked by red dots indicating a change in fracture intensity and delineating the contact between mechanical layers.

In the absence of close-proximity outcrops, fracture length statistics representative of the site are difficult to attain even with corehole data. Therefore, I used uniform distributions for generating random fracture lengths consistent with methods from Parker et al. (2018) when outcrop data is unavailable. Also, the mechanical layer thickness (MLT) can be a helpful quantity for constraining the uniform distributions for fracture lengths (Underwood et al. 2003, La Pointe 2010). The minimum fracture lengths will be constrained by the model discretization, which will be discussed in Section 6.3.1. Horizontal fracture length l_{h} is randomly selected from $l_{\mathrm{h,min}} = 0.05\ m$ (minimum horizontal grid spacing) to $l_{\mathrm{h,max}} = 10 \cdot \mathrm{MLT}$. Vertical fracture length l_{v} is randomly selected from two uniform distributions because there it is a possibility that some vertical fractures will penetrate mechanical layers: (1) 90% chance from $l_{\mathrm{v,min}} =\ 0.02$ m (maximum vertical grid spacing) to $l_{\mathrm{v,max}} =$ MLT, or (2) 10% chance from $l_{\mathrm{v,min}} =$ MLT to $l_{\mathrm{h,max}} =\ 1.5 \cdot$ MLT (Underwood et al. 2003).

6.2.2 Fracture dips

Fracture dips, the angle between a horizontal plane and a fracture, were normalized by the total fracture count from the angled boreholes (i.e., SCA1, 2, and 3) as shown in Figure 6.6, irrespective of fracture strike. Frequency of fracture dips from vertical coreholes (i.e., SCV1, 2, and 3) were ignored because vertical coreholes are biased against sampling high-angled fractures, and are effectively blind to vertical fractures (Munn 2012).

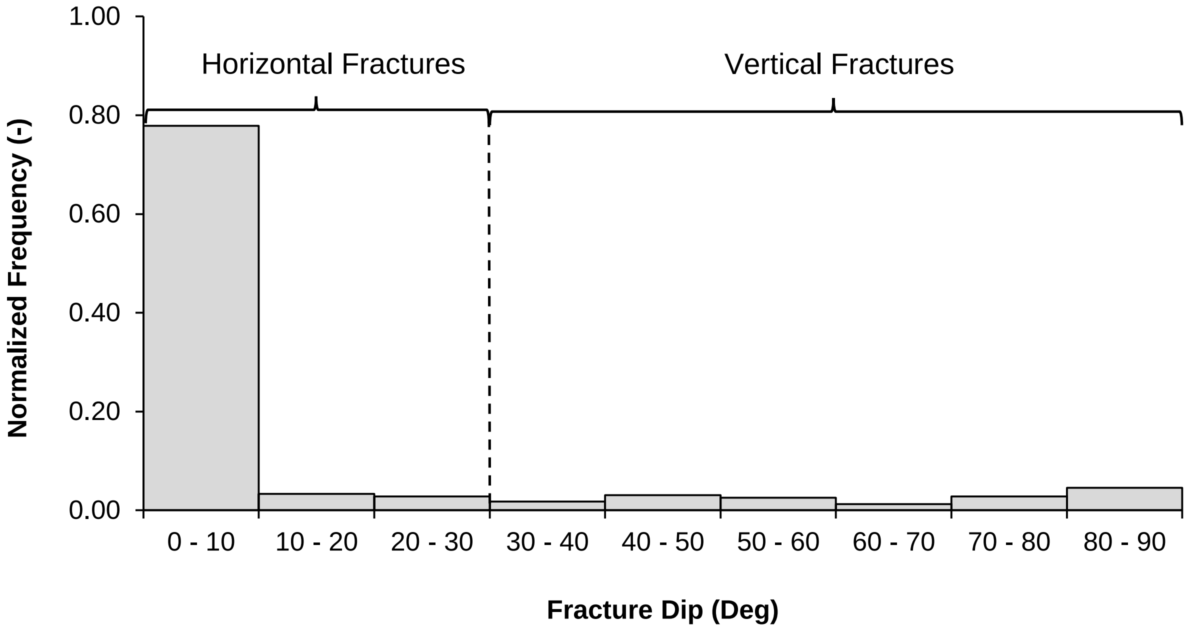


Figure 6.6: Normalized frequency of fracture dip angles from the angled coreholes only. Fractures with dip angles less than or equal to 30° are assumed to be horizontal and fractures with dip angles greater than 30° are assumed to be vertical.

The data indicates that a majority of fractures at the EBRS are subhorizontal (i.e., bedding plane fractures close to being horizontal), with approximately 78 % of fractures with dip angles between 0° and 10°. This suggests that flow may be controlled by subhorizontal flow paths. Similar normalized frequencies (1 % to 5 %) were observed for all other dip angles. Therefore, I chose to represent the EBRS DFN with orthogonal horizontal and vertical fractures only, assuming that all observed fractures with a dip angle equal to 30° or less were horizontal and all fractures with dip angles greater than 30° were vertical.

6.2.3 Fracture frequency

Under the assumption of orthogonal fracture sets, frequency distributions of horizontal and vertical fractures were calculated for each mechanical layer. Horizontal fracture frequency statistics for each mechanical layer were computed from all coreholes and are summarized in Table 6.1. Where the average fracture frequency is $\bar{f}$ and the fracture frequency sample standard deviation is s_f.

Table 6.1*: Horizontal Fracture Frequency*

Coreholes (Fig. 6.1 for locations)	**ML 1**	**ML 2**	**ML 3**	**ML 4**	**ML 5**
	Horizontal Fracture Count (frac)				
SCA1	15	20	32	15	5
SCA2	20	27	39	27	8
SCA3	8	20	34	15	15
SCV1	11	23	25	17	12
SCV2	14	18	41	11	13
SCV3	na	44	45	18	25
	Thickness of Mechanical Layer (m)				
	0.8	1.0	2.5	1.7	2.0
	Horizontal Fracture Frequency (frac/m)				
$\bar{f}$ (frac/m)	17.0	25.3	14.4	10.1	6.5
s_f (frac/m)	5.6	9.7	2.9	3.2	3.4

na = data not available or incomplete

Vertical fracture frequencies were also computed using only the angled coreholes (SCA1, 2 and, 3) because vertical coreholes are biased against sampling high angled fractures. Table 6.2 summarizes the vertical fracture frequencies for each mechanical layer.

Table 6.2: Vertical Fracture Frequency

	ML 1	ML 2	ML 3	ML 4	ML 5
	Vertical Fracture Count (frac)				
SCA1	3	3	6	2	1
SCA2	0	1	5	0	1
SCA3	5	4	10	6	7
	Horizontal Width of Angled Borehole (m)				
	0.46	0.58	1.44	0.98	1.15
	Vertical Fracture Frequency (frac/m)				
$\bar{f}$ (frac/m)	5.8	4.6	4.8	2.7	2.6
s_f (frac/m)	5.4	2.6	1.8	3.1	3.0

Kennedy (2017) summarizes a surficial fracture mapping survey in Figure D-2 of her PhD dissertation. Surficial vertical fracture (i.e., fractures that intersected the riverbed) trace lengths

(Figure 6.1) were surveyed from four areas within the EBRS (Figure D-1 from Kennedy 2017), providing a measure of total fracture length per unit area that is summarized here in Table 6.3. Where the average fracture density is $\overline{fd}$ and the fracture density sample standard deviation is s_{fd}.One area of exposed trace lengths is a rubble-free zone located within the pool of the EBRS, these mapped fracture traces will be used to determine the locations of vertical fractures within the pool (Section 6.3.2).

Table 6.3: Vertical fracture density along streambed.

Transect	*Vertical Fracture Density (m frac/m^2)*
1	0.44
2	2.45
3	2.03
Floodplain	1.50
Statistics	
$\overline{fd}$	1.6
s_{fd}	0.9

6.2.4 Fracture aperture

The effective hydraulic apertures across the five mechanical layers were approximated from hydraulic conductivity measures derived from packer tests (Capes 2018) and BSMs (Kennedy 2017). These measurements assess the total hydraulic conductivity of both the matrix and fractures. However, the presence of hydraulically connected fractures strongly dominates the flow. Packer testing was conducted across the Vinemount in 1.3 m and 1.5 m intervals within the angled and vertical coreholes, respectively. The average horizontal hydraulic conductivities with depth obtained from the packer tests are shown in Figure 6.7.

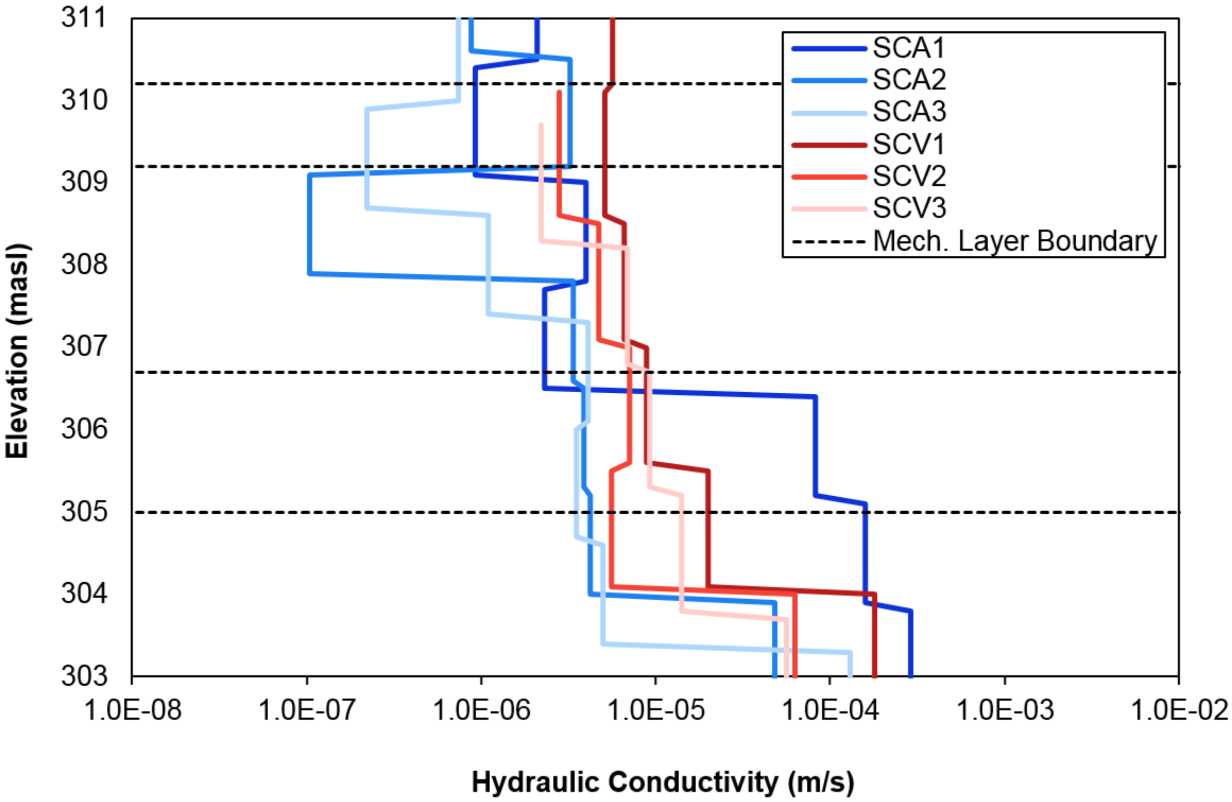


Figure 6.7: Hydraulic conductivity with depth from corehole packer tests.

Hydraulic conductivity measurements where averaged across mechanical layers for each corehole, weighted by interval width. Apertures a [L] were then calculated using the relationship between the hydraulic conductivity of a set parallel fractures $\mathrm{K_s}$ [L/T] and fracture spacing $\Delta\bar{x}$ [L] in an impermeable rock matrix given by Singhal and Guptal (2010):

$$a = \sqrt[3]{\frac{\mathrm{K_s}(12 \cdot \Delta\bar{x} \cdot \mathrm{v})}{\mathrm{g}}}$$

Where $\mathrm{g} = 9.81\ \mathrm{m/s^2}$ is the gravitational constant and $\mathrm{v} = 1.0 \cdot 10^{-6}\ \mathrm{m^2/s}$ is the kinematic viscosity of pure water at 20°C. The average spacing between fractures $\Delta\bar{x}$ was calculated by dividing the mechanical layer thickness by the total horizontal fracture count for each corehole. Table 6.4 summarizes the resulting six apertures for each mechanical layer and their averages (except for ML1 which is based on BSM data and trace length surveys).

Table 6.4: Hydraulic apertures from packer testing hydraulic conductivity

	ML 2	ML 3	ML 4	ML 5
Horizontal Hydraulic Conductivity, K_x (m/s)				
SCA1	9.2E-07	2.9E-06	4.5E-05	2.1E-04
SCA2	3.3E-06	5.5E-07	3.9E-06	1.3E-05
SCA3	3.2E-07	1.0E-06	3.7E-06	8.7E-06
SCV1	5.2E-06	6.4E-06	1.1E-05	6.3E-05
SCV2	2.8E-06	4.3E-06	6.6E-06	2.0E-05
SCV3	2.2E-06	4.4E-06	9.7E-06	2.4E-05
Average Spacing Between Horizontal Fractures, $\Delta\bar{x}$ (m)				
SCA1	0.05	0.08	0.11	0.40
SCA2	0.04	0.06	0.06	0.25
SCA3	0.05	0.07	0.11	0.13
SCV1	0.04	0.10	0.10	0.17
SCV2	0.06	0.06	0.15	0.15
SCV3	0.02	0.06	0.09	0.08
Calculated Hydraulic Aperture, a (μm)				
SCA1	38	65	185	466
SCA2	53	35	67	160
SCA3	27	45	80	112
SCV1	65	92	111	234
SCV2	58	68	108	155
SCV3	39	67	104	132
Hydraulic Aperture Statistics (μm)				
$\bar{a}$ (μm)	45	59	104	185
s_a (μm)	14	20	41	132

Hydraulic apertures for mechanical layer 1 (ML1) were derived from BSM fluxes and hydraulic gradients because BSMs provided measures more representative of the riverbed hydraulics than the floodplain coreholes. Hydraulic apertures were calculated assuming an average ML1 horizontal fracture spacing of 0.06 m from the six coreholes. BSMs that only intersected horizontal fractures were considered in my calculations for hydraulic aperture because some BSMs intersecting angled and vertical fractures were suspected of being leaky due to rock integrity issues.

Table 6.5 summarizes the resulting apertures from the BSMs (including their average and standard deviation) used to represent ML1.

Table 6.5: Hydraulic apertures for mechanical layer 1 from bedrock seepage meters

Bedrock Seepage Meter	K (m/s)	Calculated Hydraulic Aperture, a (μm)
1	3.5E-05	141
3	4.6E-05	154
8	3.6E-06	66
11	5.4E-06	75
12	1.1E-05	96
14	1.7E-06	51
19	1.9E-05	115
23	7.7E-06	85
$\bar{a}$	9.8E-06	92
s_a	1.6E-05	36

6.2.5 Other properties

Here are all other necessary model parameters that were not explicitly acquired through field investigations at the EBRS, but were attained by reviewing site-relevant literature.

- Hydraulic conductivity of the dolostone matrix was assumed to be isotropic and homogeneous with a value of $3.4 \cdot 10^{-9} \text{m/s}$ (Parker et al. 2018).
- Dolostone matrix porosity was set to 4 % (Steelman et al. 2017b).
- Longitudinal dispersivity were set to 0.1 m for both the matrix and fractures, while and transverse dispersivity was set to 0.01 m for the matrix (Parker et al. 2018).
- Free-solution diffusion coefficient was set to $1 \cdot 10^{-5}\ \text{cm}^2/\text{s}$ (Grathwohl 1998, Parker et al. 1994).
- Matrix tortuosity is a scaling factor for molecular diffusion in HGS (Aquanty Inc. 2015, eq. 2.1, p. 34), where an effective diffusion coefficient is the product of matrix tortuosity and the free-solution diffusion coefficient. Matrix tortuosity is known to be related to matrix porosity. Shikaze et al. (1998) used a matrix tortuosity value of 0.1 when matrix porosity was 0.35. Therefore, I expect that the matrix tortuosity at the EBRS should be equal to or $^1/_3$ of the matrix porosity. I chose to use a matrix tortuosity value equal to the matrix porosity (4 %).
- Effective diffusion coefficient is then $4 \cdot 10^{-7}\ \text{cm}^2/\text{s}$, which is in-line with Grathwohl (1998, Table 1.2) for tritium in a Culebra Dolomite and Parker et al. (1994) who reports values from $1.8 \cdot 10^{-6}\ \text{cm}^2/\text{s}$ to $7.5 \cdot\ 10^{-7}\ \text{cm}^2/\text{s}$ for fractured sedimentary rock.

6.3 Numerical model setup

6.3.1 Domain extents, discretization, and boundary conditions

The top of the model domain follows the elevation of the riverbed along the thalweg using 61 bathymetry values (Figure 6.8a) obtained from differential Global Positioning System surveyed by Van Harten Surveying Inc. (Kennedy 2017). The vertical depth of the model domain extends downwards until 303.0 meters above sea level (masl), which is the average elevation of the contact between the Vinemount Mb and underlying Goat Island Formation (an aquitard).

The model domain is discretized with a uniform horizontal grid spacing of 5 cm and a uniform vertical grid spacing of 2 cm from 303.0 masl to 310.2 masl, which coincides with the bottom of Mechanical Layer 1. From 310.2 masl to the top of the model domain (i.e., river bathymetry along the channel thalweg), the domain is discretized into 30 equal layers with a maximum thickness of 2 cm, which yields a total of 475,800 elements and 954,822 nodes. The maximum average horizontal fracture frequency is approximately 25 fractures/m in Mechanical Layer 2. This would compute to an average of about 1 fracture every 4 cm. All other mechanical layers have a lower average horizontal fracture frequency. Therefore, a vertical grid spacing of 2 cm was sufficient to accommodate the average observed horizontal fracture densities. The maximum average vertical fracture frequency is approximately 6 fractures/m in Mechanical Layer 1. This would compute to an average of about 1 fracture every 16 cm. All other Mechanical Layers would have a lower vertical fracture frequency, thus a horizontal grid spacing of 5 cm was sufficient to accommodate the average observed vertical fracture densities. Furthermore, it is important to maintain reasonable computational costs while being able to accommodate observed fracture densities, which in this case was a run time of just under 20 minutes per realization with this level of discretization.

Kennedy (2017) found that the underlying Goat Island Fm had an average $\mathrm{K} = 9.4 \cdot 10^{-10}\ \mathrm{m/s}$. Therefore, I treated the Goat Island Fm as an aquitard and assigned a Type 2 – No Flow boundary condition to the lower limit of the model. The left and right boundaries of the domain were assigned a Type 1 – Constant Head boundary condition of 311.022 m, which represents the average river stage elevation over the entire length of the EBRS at an average flow rate of 1.5 m^3/s (Kennedy 2017).

Using the same hydrostatic head value on the left and right sides of the domain is unrealistic since a downward hydrostatic gradient in the direction of river flow is expected for larger-scale (i.e., meander or catchment) representations of river stage. However, the river stage at the beginning ($x = 0$ m) and end ($x = 61$ m) of the EBRS was measured to be 311.02 masl and 311.09 masl, respectively. These river stages represent the average measurements made in the summer months between June 18, 2014 and June 5, 2015. The higher river stage at the end of the EBRS (downgradient) is caused by the momentum of the river water flowing over a local increase in the riverbed elevation at the end of the riffle-pool-riffle sequence (see Figure 6.8a at $x = 61$ m). Since a hydrostatic gradient in the opposite direction of the river flow would yield physically unrealistic flows, I chose to have a zero slope for the hydrostatic head conditions.

At the bedform-scale (i.e., riffle-pool-riffle), Trauth et al. (2013) have shown that hyporheic exchange is predominantly driven by spatially distributed hydrodynamic pressure variations that are produced by turbulent surface water velocity head. The boundary between surface water and groundwater flow domains would be best represented by a computational fluid dynamics model (i.e., Trauth et al. 2013). However, I used a simplified approach as a first order approximation that has been commonly used in other bedform-scale hyporheic exchange modelling studies (Cardenas 2008, Stonedahl et al. 2010, Käser et al. 2013, Pryshlak et al. 2015). I believe this is sufficient for my modelling objectives and in order to keep computational costs reasonable. I chose to represent the total hydrodynamic pressure head h_{total} along the riverbed as a constant head boundary condition approximated by the modified representation of the basic pumping model for a single bed form (Stonedahl et al. 2013):

$$h_{total} = h'_m \cdot T(x) + h_s$$

h_s is the hydrostatic head component and h'_m is the half-amplitude of the head variation given by:

$$h'_m = \frac{0.28\left(\frac{U(x)^2}{2g}\right)}{\sqrt{2}\sigma_t}\left(\frac{2\sqrt{2}\sigma_t/d_{max}}{0.34}\right)^n, \quad \text{if } \frac{2\sqrt{2}\sigma_t}{d_{max}} < 0.34 \text{ then } n = 3/8, \text{ else } n = 3/2$$

Where $U(x)$ are the average stream velocities observed between June 18 and July 30, 2014 at x m downstream (Table 6.6). The maximum stream depth d_{max} is 0.75 m and the standard deviation of the small topography function σ_t is 0.15 m. The gravitational constant g is 9.81

m/s^2. T(x) is the small topography function, which represents bedform-scale flow-boundary interactions expected to be important (i.e., bars and dunes) at the riffle-pool-riffle scale, and was obtained using a Fourier fitting procedure (Stonedahl et al. 2010) in MATLAB.

Figure 6.8b shows the resulting boundary condition. The hydrodynamic pressure head low is aligned in the area of upwelling within the first pool, which coincides with strong upwards gradients recorded by the bedrock seepage meters (Kennedy 2017).

Table 6.6: Average observed stream velocities at x *m distance along thalweg of EBRS*

Transect	x (m)	Avg. Vel. (m/s)
Tr.0	0.00	0.41
Tr.1	11.51	0.35
Tr.2	22.30	0.29
Tr.2.5	27.11	0.26
Tr.3	31.54	0.28
Tr.3.5	36.61	0.30
Tr.4	41.35	0.29
Tr.5	48.66	0.28
Tr.6	56.51	0.29
Tr.7	61.00	0.42*

*Interpolated value

I assume steady-state hydraulic head conditions, which conceptually represents the long-term average conditions at the EBRS. Therefore, I do not require specific storage parameters.

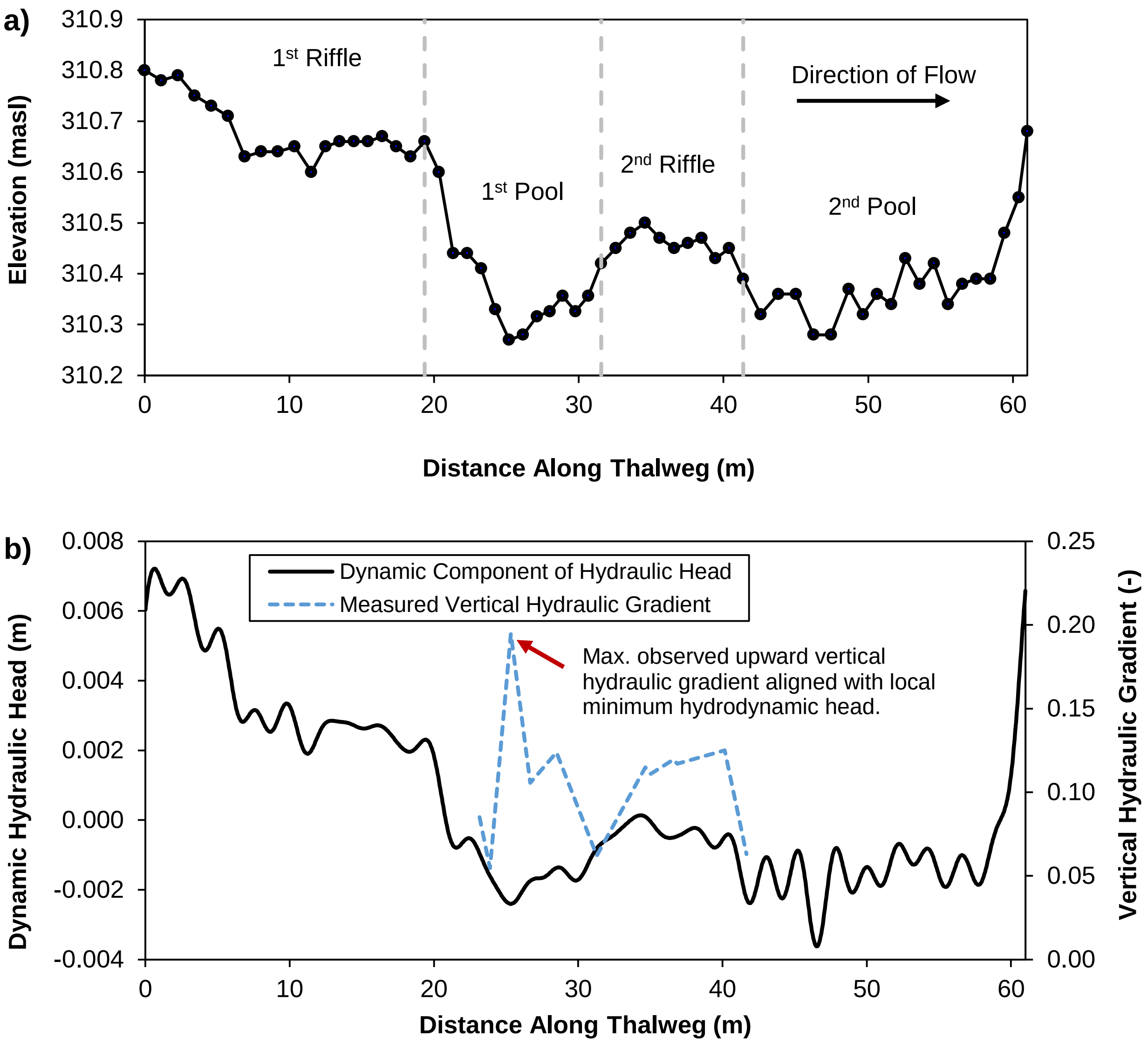


Figure 6.8: Bathymetry profile and dynamic head boundary condition: a) Thalweg bathymetry along riffle-pool-riffle sequence. b) Hydrodynamic head component of the top boundary condition aligned with measured vertical hydraulic gradient measured from bedrock seepage meters under average flow conditions. The hydrodynamic head component is calculated using the basic pumping model for a single bed form (Stonedahl et al. 2013) and added to a hydrostatic head value of 311.022 masl.

6.3.2 Random discrete fracture network generation

A basic assumption of a 2D cross-sectional model along a riffle-pool-riffle sequence is that fractures extend infinitely perpendicular to the flow direction. In other words, horizontal gradients orthogonal to river flow will be neglected in a two-dimensional cross-section. This assumption is suitable for simulating bedform-scale exchange, but may not be suitable for larger bar or meander-scale hyporheic exchange (Stonedahl et al. 2010).

A two-dimensional density is assumed from the one-dimensional fracture frequency (i.e., number of fractures per unit length), by considering all fractures counted within coreholes to be continuous throughout the sample. This assumption results in a 1:1 mapping of fracture frequency (1/L) to fracture density (L/L^2).

A summary of the fracture statistics used to represent the fractures numerically for each mechanical layer is presented in Table 6.7. To better represent vertical fracturing near-surface, I replaced the vertical fracture statistics for ML1 with frequency statistics from the streambed fractures survey (Table 6.3) and the hydraulic aperture values from the BSMs (Table 6.5). I also assume that the same apertures values are used for both horizontal and vertical fractures, which are represented by a log-normal distribution to avoid negative apertures.

Table 6.7: Eramosa Bedrock River Site Mechanical Layer (ML) Fracture Statistics

	ML1	**ML2**	**ML3**	**ML4**	**ML5**
	Elevation (masl)				
Max	311	310.2	309.2	306.7	305
Min	310.2	309.2	306.7	305	303
	Horizontal Fracture Density (m fracture / m^2)				
$\overline{fd}$	17.0	25.3	14.4	10.1	6.5
s_{fd}	6	10	3	3	3
	Vertical Fracture Density (m fracture / m^2)				
$\overline{fd}$	1.6	4.6	4.8	2.7	2.6
s_{fd}	0.9	2.6	1.8	3.1	3.0
	Hydraulic Aperture (μm)				
$\overline{fd}$	92	45	59	104	185
s_{fd}	36	14	20	41	132

The following steps are executed to generate a random DFN consistent with the summarized fracture statistics for the identified mechanical layers (Figure 6.9):

1. Set number of desired DFN realizations, which I set to 100 to maintain reasonable computational costs.
2. Partition model grid by collecting all nodes within defined mechanical layers.
3. Generate random fracture densities for horizontal and vertical fractures. Distribution of fracture density is assumed log-normal for both orientations in order to avoid negative values.

4. Generate fractures until target fracture density is achieved. The routine for fracture generation is as follows:
   a. Horizontal Fractures (HF): Determine point of fracture propagation for HF by randomly selecting any node within the mechanical layer as fracture midpoint based on nodal volumes. Fractures propagate equally on both sides of this midpoint.
   b. Vertical Fractures (VF): For mechanical layer 1, there are a number of mapped fracture traces along the streambed within the rubble-free zone of the pool (Section 6.2.3). I assume these fractures propagate vertically downwards and treat them as known surveyed fractures, thus nodes within this surveyed rubble-free zone do not have random fractures propagating from the ground surface downwards. Furthermore, random VFs chosen to start from the bottom of ML1 and propagate upwards are programed not to intersect the surface ($l_{v,max,ML1}$ = model surface – 0.02 m). For all other mechanical layers (i.e., ML2 to 5), a point of fracture propagation is determined by randomly selecting a node along the mechanical layer boundary with a 50/50 chance of fracture starting from top or bottom. If a node from the top boundary is selected, the fracture will propagate downwards. If a node from the bottom boundary is selected, the fracture will propagate upwards.
5. Determine if newly generated fracture should be kept. A newly generated fracture is kept if it does not coincide with other previously generated fractures within the mechanical unit. Both horizontal and vertical fractures that propagate outside of the model domain are clipped.
6. Calculate fracture density for the stopping criterion in step 4. Horizontal and vertical fracture density is calculated based on the generated fractures within each mechanical layer. The calculation for vertical fracture length density does not consider the vertical fractures generated from other mechanical units that extend upwards or downwards into neighboring mechanical units. I do not account for the vertical fracture lengths that extend beyond their mechanical layer into the adjacent mechanical layer in the calculation of fracture density because their overall effect on fracture length density is negligible.
7. Repeat steps 3 to 6 until the desired number of DFN realizations are generated.

Note that and exception to step 4 is that there are fractures that propagate vertically downwards from the surface in the rubble-free zone of Mechanical Layer 1 are not treated randomly. Instead, their frequency and placement are treated as known values for all realizations. Their frequency and placement values are based on a fracture length trace study conducted by Kennedy (2017). Their lengths, however, are treated stochastically like all other vertical fractures.

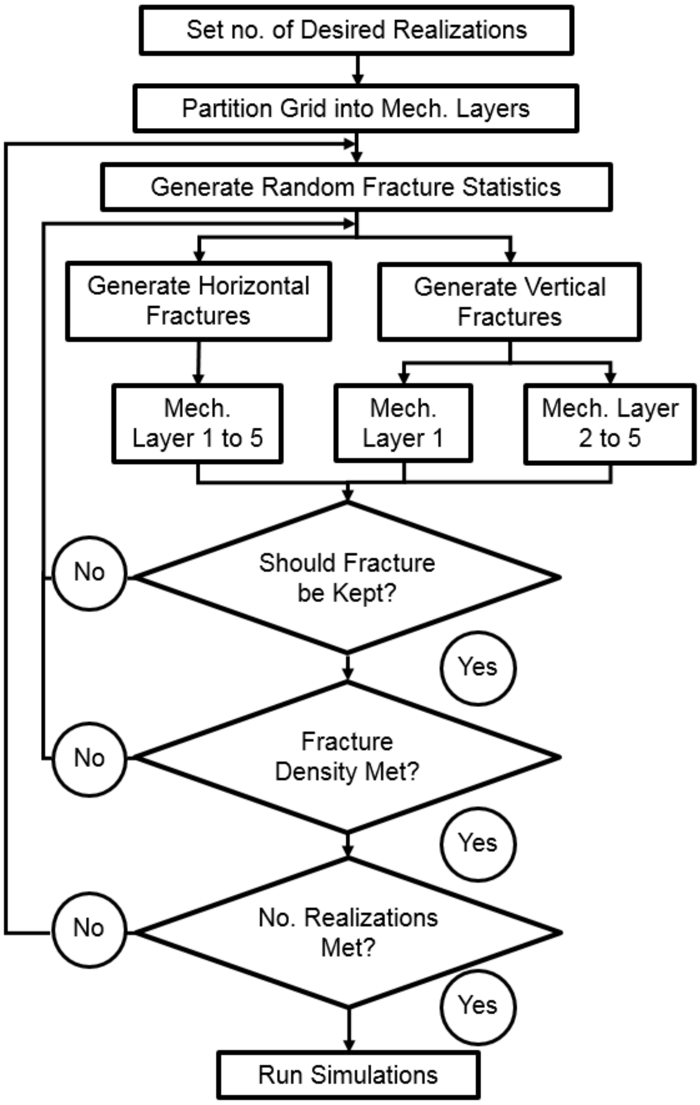


Figure 6.9: Steps for generating 2D discrete fracture network realizations *for the Eramosa Bedrock River Site.*

Figure 6.10 is an example of one DFN realization (no. 39), it is vertically exaggerated 5 times in order to better distinguish the mechanical layers. The observed vertical fractures (blue) represent the mapped fracture traces in the rubble-free zone located within the pool of the EBRS.

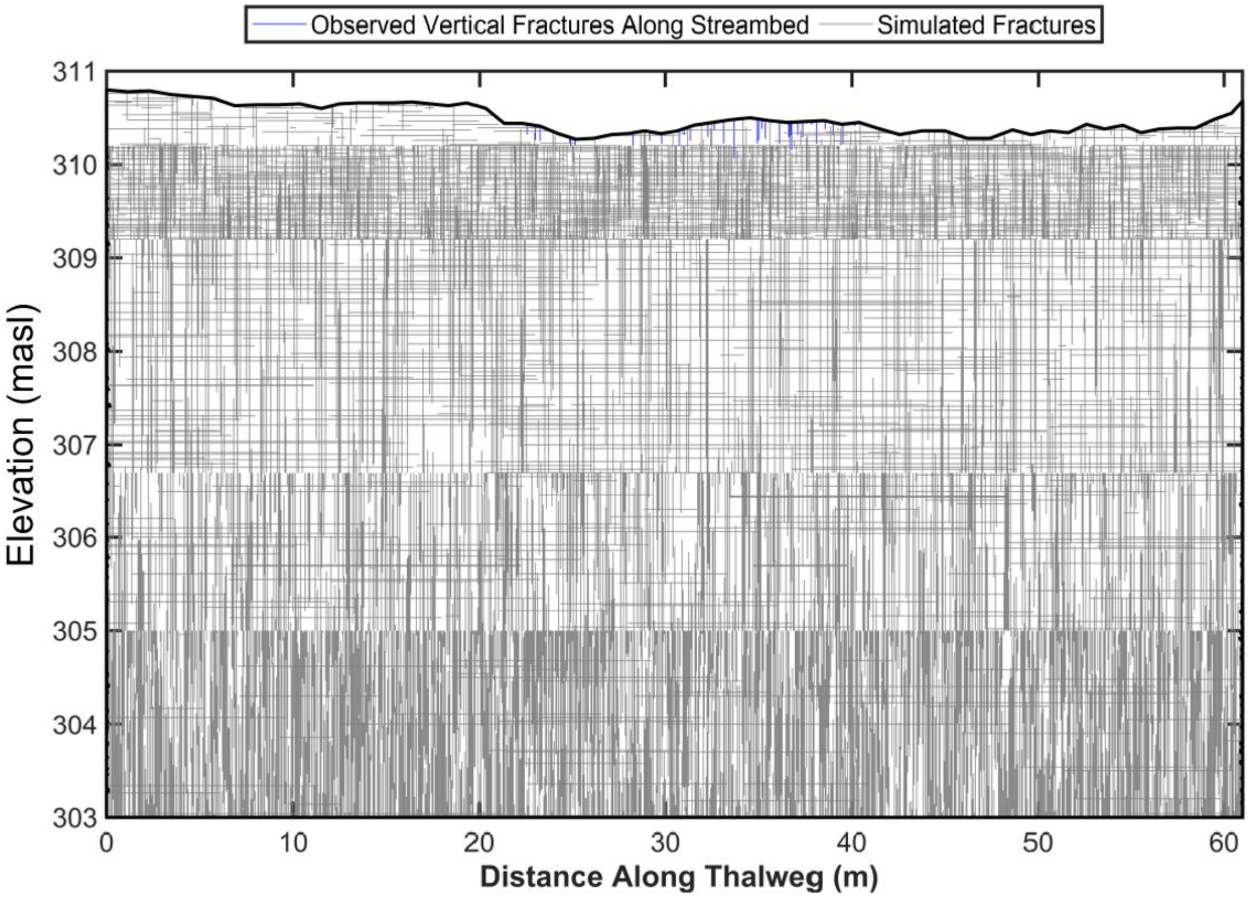


Figure 6.10: Random discrete fracture network (Realization no. 39) based on statistics derived from discrete fracture mapping using acoustic televiewer logs and outcrops. Vertical to horizontal exaggeration = 5:1. Observed fractures in blue represent the mapped fracture traces in the rubble-free zone within the pool of the EBRS.

6.4 Results

The following sections describe the results from running the EBRS HGS model for 100 randomly generated DFN realizations (process described in Section 6.3.2), producing an ensemble of model results. The boundary conditions and grid geometries (Section 6.3.1) are kept constant for all realizations. Sections 6.4.1 and 6.4.2 show the results from a single selected DFN realization (the median with respect to the ensemble of residence time distributions), while Section 6.4.3 shows ensemble results.

6.4.1 Steady-state hydraulic head distribution

Figure 6.11 shows an example of the resulting steady-state head distribution from a single DFN realization (no. 39). Discrete fractures are represented as light grey lines in Figure 6.11. The area of highest hydraulic head within the riverbed is located in the 1st riffle (Figure 6.8a). Areas of lowest hydraulic head are located in the 1st and 2nd pool (Figure 6.8a). Thus, I expect that downwelling would occur in the 1st riffle, with upwelling occurring in the 1st and 2nd pool.

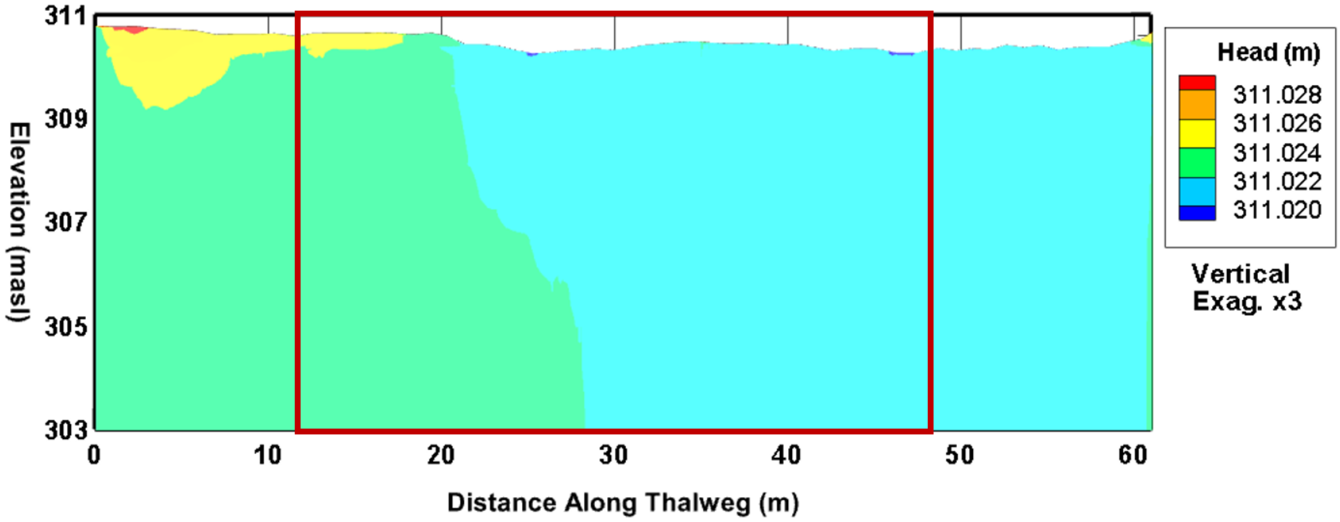


Figure 6.11: Total hydraulic head distribution for DFN under steady-state (Realization no. 39), average flow conditions of 1.5 m^3*/s. Direction of river flow is in the positive x-direction. Red border outlining inset for Figure 6.12. Fractures removed for clarity.*

6.4.2 Transport of conservative tracer

Transport of a conservative tracer was simulated for each DFN realization to determine the residence times and extents of hyporheic exchange at the EBRS. Figure 6.12 is an example of the resulting transport of conservative tracer selected time steps. A constant initial concentration $c_0 = 1000$ mg/L (1 kg/m^3) was applied over a single one-day time-step within the upper-most node of a vertical fracture that intersects the top surface of the model. Conceptually, this represents a one-day pulse of conservative tracer within a surficial vertical fracture. Application of the tracer for each DFN realization is in the left-most surficial vertical fracture between 15.0 m and 19.4 m. 15.0 m (approximately 25 % of the longitudinal domain length) was chosen as the left-most extent for the tracer placement because I wanted to avoid influence from the left-side boundary condition. 19.4 m was chosen as the right-most extent because it is where the elevation begins to decrease (i.e., where the pool begins). Releasing the conservative tracer in the left-most vertical fracture between 15 m and 19.4 m ensures that I am choosing the largest gradient between downwelling in the riffle and upwelling in the pool.

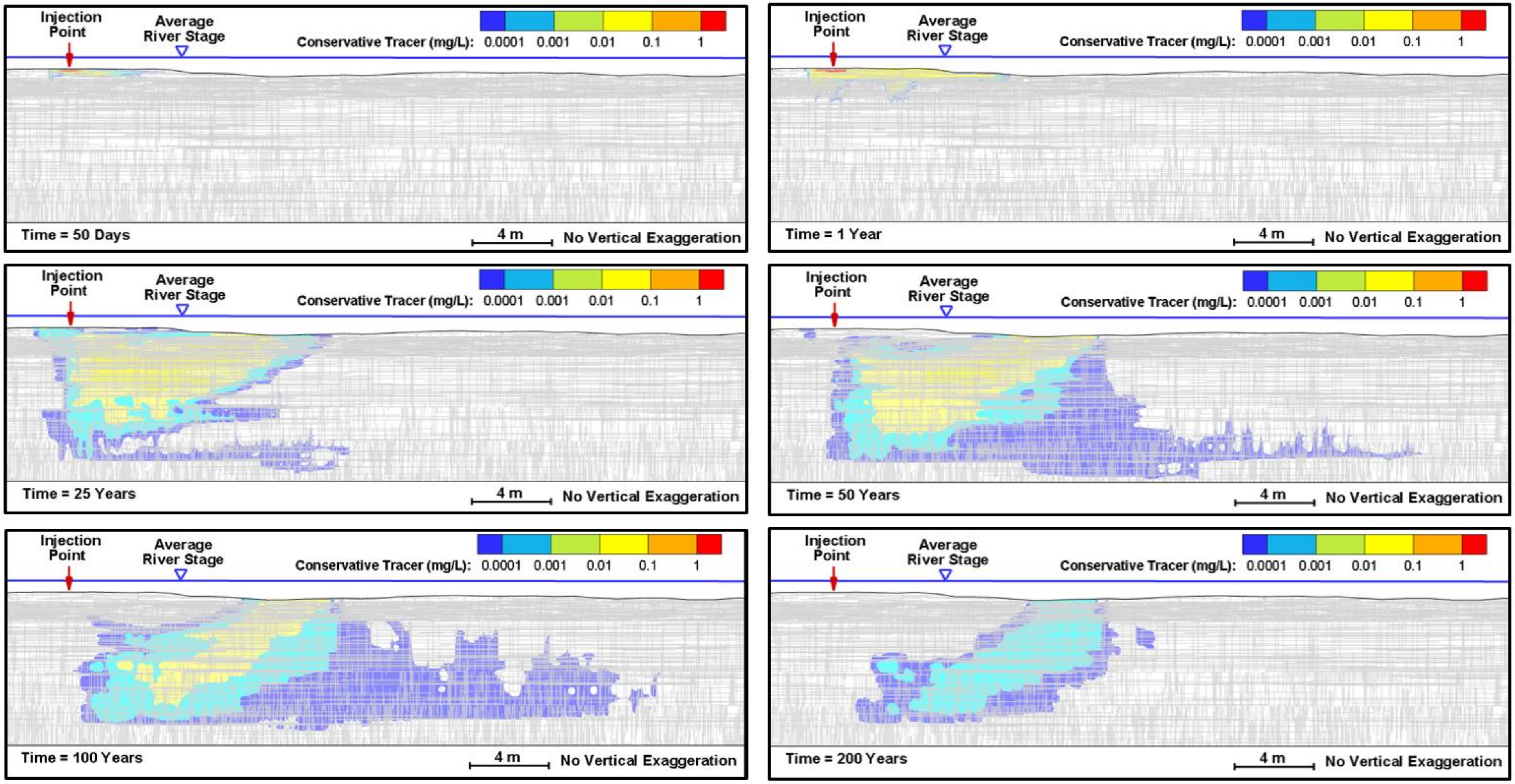


Figure 6.12: Transport of conservative tracer for a single DFN realization (no. 39). Conservative tracer is pulsed over one day into the left-most fracture intersecting the top surface of the model between 15 m and 19.4 m. The tracer migrates from left to right, upwelling in the 1st and 2nd pool. The conservative tracer is below detection (<0.0001 mg/L) after 440 years.

6.4.3 Analysis of hyporheic exchange residence times and extents

In order to determine the residence times of riffle-pool-riffle hyporheic exchange, the total mass balance of the conservative tracer within the model domain was analyzed with respect to time for each DFN realization (Figure 6.13). The tracer mass that enters the system is from an instantaneous one-day (time-step size) pulse injection within the 1st riffle. The total tracer mass within the model domain increases over the injection time t_0, until reaching a maximum mass $M_{\max}$. Afterwards, the mass within the model domain decreases with time as the conservative tracer leaves the model domain through the top or side boundaries of the model.

Solute mass leaving through the top model boundary can be conceptualized in terms of a riffle-pool-scale hyporheic exchange, while the mass leaving through the sides of the model can be conceptualized as larger scale (e.g., meander-scale) hyporheic exchange, which is not explicitly represented in this model. It should be noted however, that conservative tracer mass that leaves through the sides of the model are not necessarily hyporheic exchange.

Finally, a cumulative distribution function (CDF) of the residence times can be created by normalizing the mass $M(t)_{\text{norm}}$ leaving the entire model domain by the maximum total mass (Figure 6.13). Conservative tracer mass is tracked until 95 % of $M_{\max}$ because a majority of the transport simulations could meet this criteria without an excessively long maximum time step (≤ 200 years). Therefore, all realizations still have 5 % of tracer mass left in the domain.

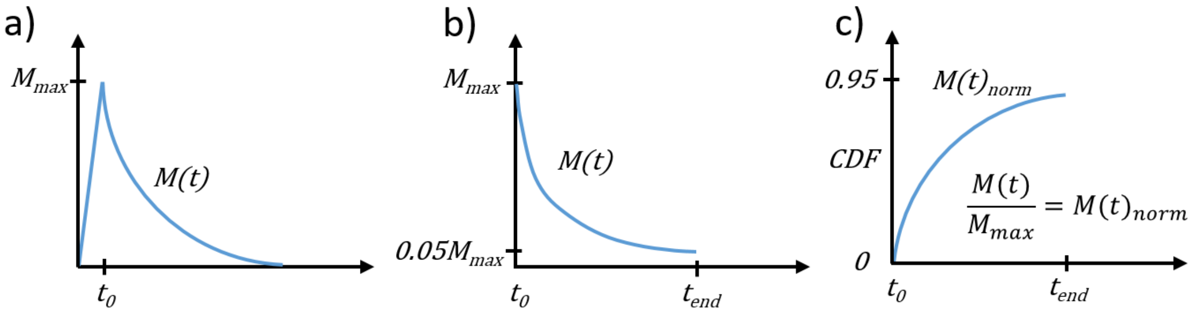


Figure 6.13: Determination of residence time cumulative distribution function (CDF) of conservative tracer mass for each DFN realization: a) initial injection and determination of maximum mass $M_{\max}$ in model domain, b) determination of end time when 5 % of maximum mass remains in model domain, c) normalize cumulative mass leaving model by $M_{\max}$.

Figure 6.14 shows the resulting ensemble of bedform-scale hyporheic exchange residence time CDFs. The first thing to notice is the time units for the x-axis are in years. Right away we can see that the residence times for hyporheic exchange at the EBRS is orders of magnitude longer when compared to similar scales for rivers with unconsolidated sediments. Figure 1.3 from Boano et al. (2014) shows that one should expect residence times in the order of days to weeks for bedform-scale hyporheic exchange in rivers with unconsolidated sediments. The average hyporheic exchange residence time at the EBRS appears to be approximately 13 years. The uncertainty for the average residence time, from the 1st and 3rd quartile, ranges between 4 to 35 years. Also, when looking at the 5th and 95th percentile ranges, one can see that the uncertainty range (especially at CDF values closer to one) is large. The DFN realization (no. 39) that produced the residence time CDF with the smallest sum of square difference (green line in Figure 6.14) relative to the median residence time CDF (red line in Figure 6.14) was chosen as the representative DFN realization, shown previously in Figures 6.10, 6.11, and 6.12.

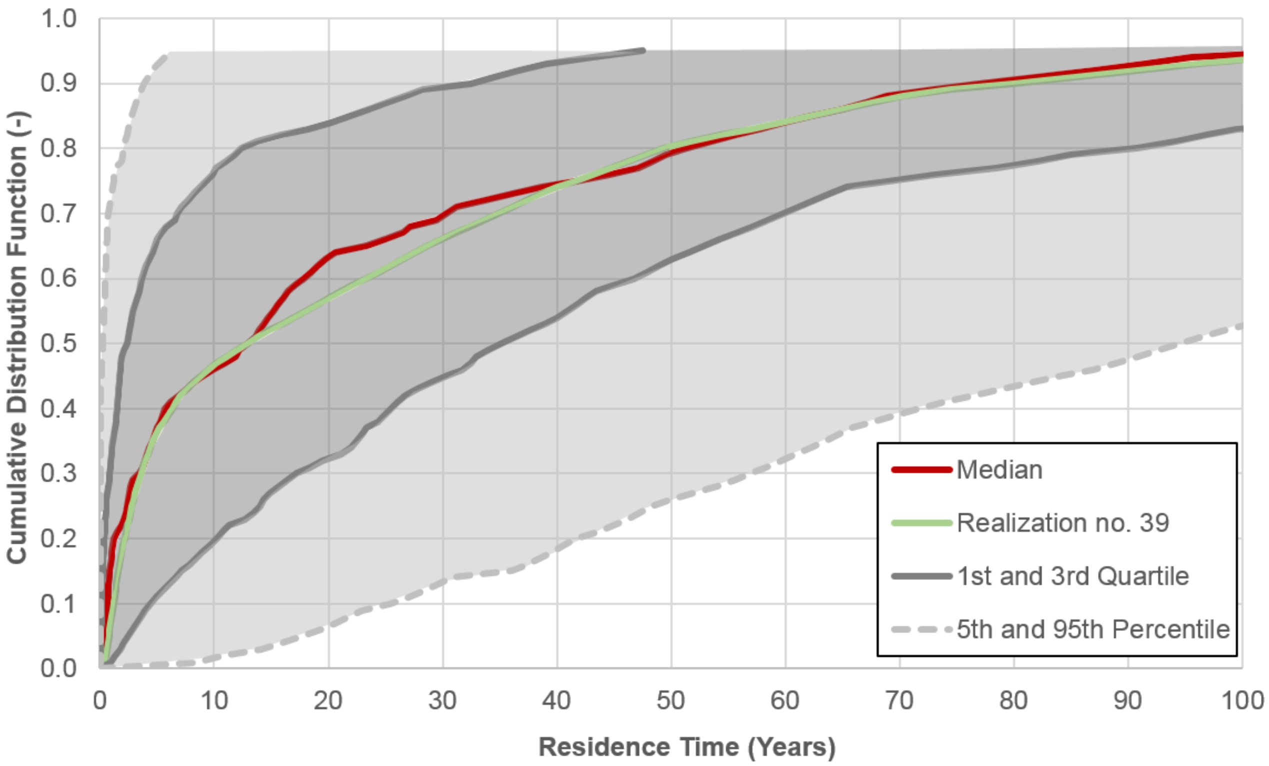


Figure 6.14: Hyporheic exchange residence time cumulative distribution functions (CDF). Solid red line is the median residence time CDF from 100 DFN realizations. Green solid line is the residence time CDF from the realization (no. 39) closest to the median. The solid dark grey line and darker grey shading represent the interquartile range of the 100 residence time CDFs. The dashed grey line and lighter grey shading represent the 5th and 95th percentiles.

To understand the extents of hyporheic exchange, I analyzed where the normalized mass leaves the domain (Figure 6.15). On average, over the 100 DFN realizations, the total mass that entered the model domain after a one-day pulse injection of 1 kg/m^3 was $M_{\max} = 1.1 \cdot 10^{-4}$ kg. The blue line in Figure 6.15 shows the normalized mass leaving through the riverbed (i.e., model domain top surface) averaged over 100 DFN realizations. The area under the blue line sums to 93 %, indicating that on average a majority of the conservative tracer mass leaves through the riverbed. Approximately 2 % of the mass leaves through the right side of the model domain, while the remaining 5 % remains within the model domain after 100 years.

Figure 6.15 shows that most of the mass leaves within the vicinity of the injection interval. The next most common upwelling occurs between 20 m and 30 m, which coincides with the hydraulic head low and the maximum observed upward vertical hydraulic gradient (Figure 6.8). Some minor upwelling also occurs between 40 m and 45 m, which coincides with some small undulations in the bathymetry within the 2nd pool (Figure 6.8). One final note is that upwelling also occurs to the left of the injection interval (i.e., opposite of the river flow direction), which is theoretically possible due to local upstream hydraulic lows caused by turbulent eddies.

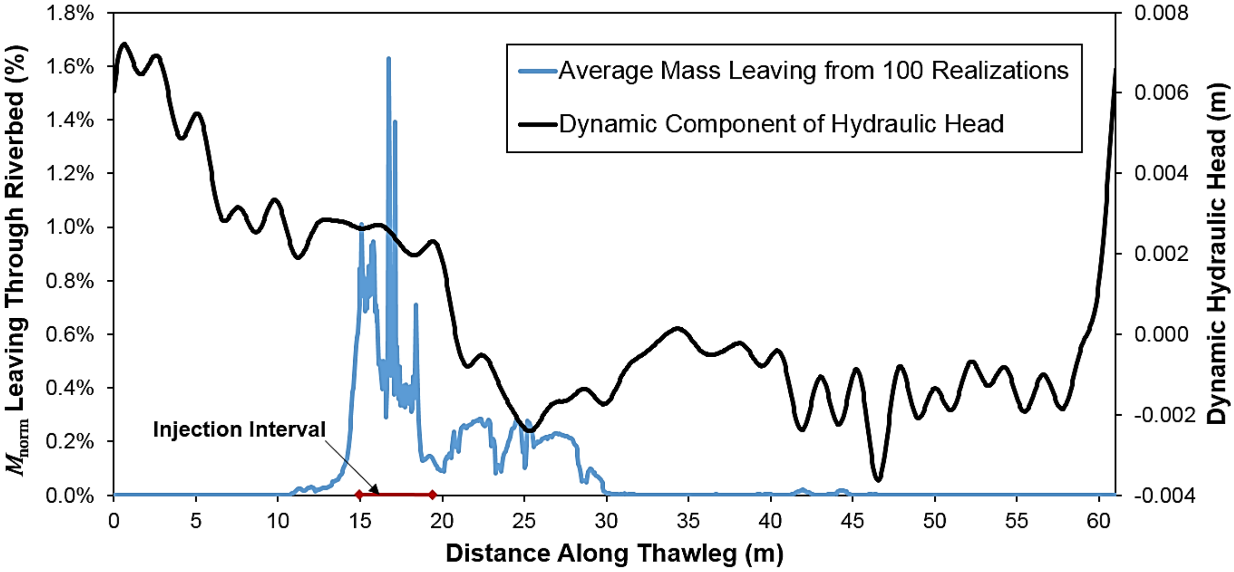


Figure 6.15: Average normalized mass (M_{norm}) leaving through the riverbed (i.e., top surface of the model). Blue line is the M_{norm} leaving averaged over 100 DFN realizations. The dynamic head component (black line) is added to a hydrostatic head component of 311.022 masl. For each individual realization, the first fracture from left to right that intersects the riverbed between 15 to 19.4 m (red line) is selected for the injection of conservative tracer.

6.5 Conclusions and outlook

My goals in this study were to evaluate factors controlling hyporheic exchange along a fractured sedimentary bedrock river by using a discrete fracture network (DFN) approach to represent the subsurface within a manually coupled numerical sw-gw model. The aim was to understand fracture network heterogeneity and its role in defining hyporheic exchange extents and residence time distributions. In order to fulfill these goals, I evaluated data uncertainty and modelled a conservative tracer experiment at the Eramosa Bedrock River Site (EBRS) using a stochastic DFN framework to represent the subsurface fractured bedrock connectivity. In doing so, I produced a probabilistic distribution of the potential hyporheic exchange extents and residence time distributions.

The research questions I aimed to answer in this study were: can hyporheic exchange occur within bedrock rivers at the bedform-scale? If yes, then how is hyporheic exchange different in bedrock rivers when compared to fluvial rivers? Two important conclusions about hyporheic exchange at the EBRS can be drawn from the results. Firstly, by comparing the residence time of hyporheic exchange at the EBRS (Figure 6.14) to a similar scale for fluvial rivers (Figure 1.3 by Boano et al. 2014), we can see that bedform-scale hyporheic exchange in a bedrock river is potentially orders of magnitude slower compared to a fluvial river (i.e., months to years vs. minutes to hours). Secondly, the coincidence of fractures and hydraulic gradients determines the spatial extents of bedform-scale hyporheic exchange.

These conclusions suggest that residence time distributions may not be a good proxy for the potential natural attenuation capacity of bedrock rivers. Since hyporheic exchange residence times are potentially much longer in bedrock rivers than in fluvial rivers, one would expect that diffusion becomes a dominant process causing significant retardation of solutes that are delivered into bedrock riverbeds via hyporheic exchange. The fractures and void spaces would be expected to be the only oxygen-rich environment, while the rock matrix would be expected to be a very reducing environment. The reactive potential of solutes within the rock matrix may not be enough to induce significant degradation. Instead, back-diffusion of potential contaminants from the rock matrix into the water can be a source of long-term trace contaminant concentrations, leading to persistent contamination problems that would be difficult to remediate in-situ. An interesting way to investigate this phenomenon further would be a reactive tracer experiment (Knapp et al. 2017), followed by a reactive transport modelling study.

Numerous directions could be taken to further develop the EBRS modelling study. There are two immediate directions that I think have the potential to deliver critical insights. The first is a direct comparison of the current stochastic DFN model with an equivalent porous medium (EPM) model to potentially find effective dispersion and anisotropy values that could replace the need for an explicit DFN model. The second is a sensitivity analysis of the various stochastic DFN model parameters to identify the vital few that dominate the simulated hyporheic exchange response.

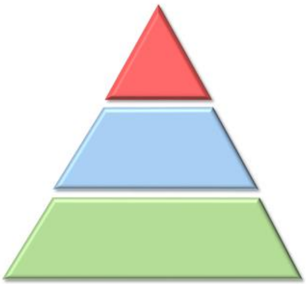
A future direction that would require more effort is to couple the current model to a computational fluid dynamic (CFD) model (e.g., OpenFOAM, Trauth et al. 2013) and include transient surface flows. Although I have only simulated steady-state hydraulic head conditions, I believe interesting insights could be gained by evaluating the transient dynamics of river flow rates on hyporheic exchange systems. I believe that the current computational limitations will be overcome in the near future and thus fully coupled three-dimensional CFD and groundwater flow modelling will be more accessible to sw-gw scientists and engineers.

Nonetheless, I would imagine that during the spring, melting of snow accumulated during the winter would increase river flow rates and thus increase the turbulence at the EBRS. Trauth et al. (2013) has shown that an increase in river flow rate causes turbulence to increase, producing larger hydrodynamic pressure variations within the riverbed hydraulic head distribution (up to a certain threshold). Thus, I believe spring thaws would increase the gradients that drive hyporheic ex-change, leading to an increase in hyporheic exchange fluxes and a decrease in residence times.

I do not believe that the increase in hyporheic exchange fluxes and decrease in residence times will be too dramatic because pressure propagation is generally slow in the subsurface, especially when hydraulic conductivity and connectivity is low. Furthermore, the exact travel paths would probably change due to the changes in the spatial distributions of head gradient under transient conditions. However, once the flow rates slow down after the spring thaw is complete, returning to more average rates (model river flow rate), I would expect to see patterns similar to the steady-state flow model presented here.

Conversely, if river flow rates slowed down (e.g., due to an unusually dry summer), the hydrodynamic gradients would be smaller, and therefore residence times would be longer and hyporheic exchange fluxes would be smaller. If river flow rates were to slow down dramatically (e.g., well below model river flow rates or the river dries up), hyporheic exchange would eventually halt altogether.

7. General conclusions



7.1 Main findings and outlook

My dissertation focused on filling knowledge gaps with regards to model uncertainties from numerical modelling of surface water-groundwater interactions at different scales. It is difficult to separate model uncertainty sources completely. Nonetheless, I have attempted to address three different predominant uncertainty types (Section 1.1.3: structural uncertainty, parameter uncertainty, and data uncertainty) at three distinct scales. To address these uncertainties, I developed three different models described in four chapters within this dissertation. At the catchment-scale (Chapter 3), I developed the Alder Creek watershed model to mainly evaluate the imprecision in model results from choosing between different numerical models, which can be considered a type of structural uncertainty. At the meander-scale (Chapter 4 and 5), I developed the Steinlach River Test Site model to evaluate uncertainty that comes from the parameterization of river bathymetry and subsurface hydrostratigraphy. Finally, at the bedform-scale (Chapter 6), I developed the Eramosa Bedrock River Site model and a stochastic framework in order to deal with data (or the lack thereof) uncertainty inherent to representing groundwater flow and transport with a discrete fracture network approach.

Chapter 3 primarily dealt with structural (conceptual) uncertainty at the catchment-scale by investigating uncertainties from numerical model choice. Technically, choosing between different numerical models is not uncertain because it is known that different model errors will result from different simplifying assumptions in their governing equations and numerical solution methods. Instead, the structural model uncertainties that I investigated in Chapter 3 were uncertainties stemming from imprecision in the model result due to choices in available numerical modelling codes applied to a real-world system.

My main research goal for Chapter 3 was to identify model errors that come from numerical model choice and to provide practical methods that account for parameter uncertainty for delineating areas that provide baseflow to receiving surface water bodies. My motivating research question (RQ1, Section 1.5) for Chapter 3 was: does numerical model code choice lead to errors or deviations in the delineation of baseflow contribution areas for gaining stream reaches? To answer this question, I conducted a model and method intercomparison applied to the Alder Creek watershed in southwestern Ontario, Canada. Four different model codes were compared: HydroGeoSphere, WATFLOW, MODFLOW, and FEFLOW. Additionally, two delineation methods were compared: reverse particle tracking and reverse transport. My results indicated that the delineation of areas contributing baseflow to streams can be affected by both the model choice and the capture zone delineation (CZD) method. Different models can yield different results due to different parameterization, discretization schemes, and particle tracking algorithms.

The results show that one of the largest sources of deviation in CZD is the particle tracking algorithm, followed by the discretization scheme and the parameter distribution. This leads to an important insight regarding uncertainty and its propagation throughout the modelling process. Beven et al. (2005) have proposed, in their model uncertainty cascade concept, that uncertain model input and components from the early stages within the modelling process can cascade and accumulate throughout the modelling process (sometimes non-linearly), which can potentially result in large forecast uncertainties. My results, in contrast, show that, under certain circumstances, the choice of the final post-processing method alone (i.e., particle tracking algorithm) can produce significantly different results. Thus, although the model uncertainty cascade concept is important in highlighting the importance and sensitivity of the choices made in the early stages of modelling, my results indicate that the final post-processing choices could also be significant. Particularly, when it comes to finite element particle tracking, the greatest numerical accuracy can be achieved by choosing a mass conservative approach for the determination of fluxes for unstructured finite elements (Odsæter et al. 2017), which will be compatible with HydroGeoSphere soon (Selzer and Cirpka 2019).

Nonetheless, my findings can be reconciled with the uncertainty cascade concept (Beven et al. 2005) because the importance of post-processing tools and their availability is still a consequence of initial selection between numerical model codes. In other words, it is critical to think through the entire modelling work flow (e.g., Figure 1.5) before choosing a particular model code, because that choice will be followed by model constraints based on its underlying assumptions and available post-processing tools. Obviously, it is imperative to choose a model that can, at a minimum, fulfill the initially stated model objectives. However, the chosen model should also be flexible to allow the modeller the option to represent their conceptual model by different numerical approaches, in case the study objectives change or the initial modelling approach is found to be inadequate. Furthermore, the chosen numerical model code must provide the necessary tools to permit high-quality post-processing (e.g., MODPATH and ZONEBUDGET for MODFLOW). Therefore, choosing the right model must be balanced between having enough complexity and high-quality tools to meet the modelling objectives adaptively, but also as simple as possible to not overburden the modeller with excessive and irrelevant choices. Thinking through the entire modelling work flow can be difficult in research when the ultimate goal may be uncertain, but for most applied hydrological studies with well-defined model objectives it should be a guiding tenet when challenged by model code selection uncertainty.

The issue with numerical model code choice is not only about selecting the appropriate model for the given objective, but once the model is chosen there may be an excessive number of other objective-independent model structural features that the modeller must choose to include (or omit) that can potentially cause imprecision in the model results (equifinality discussed in Section 1.1.2). This is especially true with the advent of fully-integrated physically-based hydro(geo)logical models (e.g., CATHY, HydroGeoSphere, Open-GeoSys, and ParFlow). This type of structural uncertainty stems from the multitude of choices the integrated hydrological modeller must make when simulating multiple environmental processes simultaneously. In other words, structural model choice uncertainty comes from not fully knowing how the numerous integrated modelling choices will combine to produce the final forecast of interest.

Therefore, a clear problem with using fully-integrated spatially distributed physically based models is that there is a higher likelihood of model over-fitting as more processes and their parameters are added to the model (i.e., equifinality and nonuniqueness). Without developing the right hydro(geo)logical theories, it will become increasingly difficult to judge if a fully-integrated hydro(geo)logical model is getting the right answers for the right reasons (Kirchner 2006). Kirchner (2006) believes greater scientific progress can be made by combining theory and data, rather than by increasing the complexity of parameter-rich models to better match data, even if their underlying principles are unrealistic.

One way to remedy these structural model choice uncertainties is by devoting our scientific resources to more benchmark and model comparison studies (e.g., Brunner et al. 2010, Maxwell et al. 2014, Kollet et al. 2017). Model benchmark studies, like the one I conducted in Chapter 3, are especially useful when they provide modellers clear guidance in choosing the appropriate numerical model for common hydro(geo)logical problems, particularly when there is no clear optimum (i.e., lowest resource cost to accuracy ratio) model choice a priori. However, too many additional benchmark studies can quickly lead to dreadful zombie-science (Schwartz 2013). An alternative remedy for these model structural uncertainties is by rigorously applying the principle of parsimony (discussed in Section 5.1) at the conceptual modelling stage. This means to seek simplicity in both defining the modelling problem and in the overall model design.

Chapters 4 and 5 dealt primarily with bathymetry and hydraulic conductivity parameter uncertainty, respectively. Both chapters evaluated these uncertainties with the same meander-scale 3D fully coupled sw-gw model of the Steinlach River Test Site. My research goals were to determine the level of bathymetry and subsurface heterogeneity detail necessary to simulate meander-scale hyporheic exchange in a physically meaningful way. My main research question for Chapter 4 was: what are the most important aspects of the river bathymetry that need to be captured in order to capture the scales of exchange appropriately? And for Chapter 5 it was: what is the recommended degree of subsurface parameterization to simulate meander-scale hyporheic exchange?

Results from Chapter 4 indicate that simulating hyporheic exchange with a 3D fully coupled sw-gw model is sensitive to the representation of river bathymetry. The river bathymetry affects the river stage, and consequently the underlying subsurface hydraulic head distributions. The presence of local-scale contrasts in bathymetry can lead to an increase in smaller bedform-scale hyporheic exchange nested within larger meander-scale hyporheic exchange systems. Conversely, simulations of hyporheic exchange at the meander-scale that lack detailed bathymetry data can lead to underrepresenting smaller bedform-scale exchange. Furthermore, my results indicated that underrepresenting smaller bedform-scale exchange can potentially lead to biases towards overestimating larger meander-scale exchange. Overestimating meander-scale hyporheic exchange can lead to biases towards longer residence times. When extrapolated to account for all meanders along an entire river or rivers within a watershed, this can lead to gross biases in the estimated retention and pollutant attenuation capacity of a catchment.

Furthermore, results from Chapter 4 indicate that it is important to identify the locations of local bathymetric highs and lows within the river. Bathymetry surveying campaigns for meandering channels can be more effective by prioritizing bathymetry measurements along the thalweg and *gegenweg*, which can be identified visually. I define the *gegenweg* as the line that connects the shallowest points in successive cross-sections along the depositional side of a river opposite the thalweg under average flow conditions. Incorporating local bathymetric contrasts will likely capture the nested hierarchies of hyporheic exchange systems, leading to more physically meaningful simulations of hyporheic exchange fluxes and transit times.

In Chapter 5 we saw that choosing between different subsurface parameterization structures leads to trade-offs between intrinsic (aleatoric) and epistemic model errors (Rubin 2003, Anderson et al. 2015). Intrinsic model errors are imposed by the structural limitations of the chosen representation of subsurface parameters that cannot be reduced by more knowledge (i.e., new measurements), while epistemic model errors are potentially reducible with new knowledge. Thus, from the range of parameter structures presented in Chapter 5, we saw that the simplest homogeneous one-zone model was constrained by intrinsic model errors because it lacked the flexibility to adapt to additional hydrogeological information, therefore its predictive errors were irreducible (i.e., intrinsic). We also saw that a two-zone model was capable of reproducing the transit time distributions of a more detailed model, but only if the geological structures are known a priori, which is seldom the case in applied hydrogeological modelling.

Only by increasing complexity through the addition of zones or heterogeneity do we begin to exchange intrinsic model errors for epistemic ones. An interpolated heterogeneous parameter field (cf. pilot points) showed the best trade-offs between the two model errors (i.e., bias and variance), indicating fitness for practical applications. At the other end of the spectrum, we saw that the model errors from a hierarchical 3D multiple-point geostatistics (MPS) model can be potentially large but also reducible with the right choice of additional hydrogeological data, which in this case was an estimate of meander-scale hyporheic exchange flux (Osenbrück et al. 2013).

Results from Chapters 4 and 5 can largely be attributed to the meander-scale focus of the model. I suspect that as the scale of the modelling increases (Figure 5.8), uncertainty from larger-scale hierarchical elements (Heinz and Aigner 2003) would dominate, while rendering the uncertainty from smaller-scale elements less relevant. If my proposition is true, then this could be an important insight for model up/down-scaling, because it would be wrong to assume that results from a sensitivity or uncertainty analysis at a certain scale is transferable to another scale. Thus, a separate sensitivity or uncertainty analysis would be necessary when changing the modelling scale.

Results from Chapters 4 and 5 also indicate that one could achieve results similar to a more detailed model by simply using the thalweg-gegenweg bathymetry (54 instead of 540 bathymetric survey points) and a two-zone (top soil and fluvial aquifer instead of hydrofacies MPS with internal heterogeneity) depiction of the subsurface *K*-distribution. This ratio of a vital few to the trivial many factors (Juran et al. 1974) identified through sensitivity and uncertainty analyses often obeys (qualitatively) the Pareto principle or 80/20 Rule (Sanders 1987), which states that typically 80 % of a desired result (majority of the total effect) comes from 20 % (a vital few members) of the input. In other words, the conducted sensitivity and uncertainty analysis of the bathymetry and subsurface heterogeneity, respectively, revealed the critical few factors that produce the majority of the results (i.e., meander-scale hyporheic exchange flux and transit time distribution).

Although the Pareto principle is a valuable guide when tasked with identifying the most critical model components, it is important to note that it has clear limitations. Firstly, the Pareto principle is based upon empirical observations and was not intended to be used as a rule for action (i.e., to eliminate the 80 % or the many trivial model processes/parameters). Secondly, the Pareto principle only highlights the vital few factors of a particular model in a static snapshot, whereas reality is dynamic and ever-changing (Sanders 1987). In other words, one cannot fully disregard the trivial many (80 %) that may be amongst the vital few (20 %) at different times or under different circumstances.

A caveat to the conclusions from Chapters 4 and 5 is that they are derived from a site-specific result. Therefore, I would expect that the critical few factors would likely differ at different sites, times, and scales. Furthermore, there is the underlying assumption that the benchmark ‘virtual reality’ model is sufficiently representative of a real site. However, the point is that sensitivity and uncertainty analyses are powerful tools that can be applied to a model to potentially identify the critical factors that control the model response or a forecast of interest. With these vital few factors in-hand one can more efficiently and effectively use resources to prioritize investigations into the things that matter most and in doing so save time, money, and perhaps help reveal more “unknown unknowns” (Section 1.1.1).

Conclusions from Chapters 4 and 5 indicate that both bathymetry and subsurface distribution of hydraulic conductivity play an important role when modelling hyporheic exchange at the meander-scale. However, it is far easier to accurately measure channel bathymetry than to accurately measure the distribution of subsurface properties. Furthermore, the feasible parameter space for the distribution of subsurface properties is much vaster than bathymetry. Following the principle of parsimony (Section 5.1) for model analysis, I would recommend conducting a sensitivity analysis first before an uncertainty analysis, because the modelling objectives may be achieved through a sensitivity analysis alone. Afterwards, if the results from a sensitivity analysis are deemed insufficient or inconclusive and that an uncertainty analysis is required, the results of the sensitivity analysis could help guide the design and focus of the uncertainty analysis to a vital few tractable modelling processes, parameters, and interactions.

As an outlook, the challenge of achieving a more identifiable distribution of subsurface properties is a persistent issue for applied hydrogeologists. Ahmed et al. (2016) have developed techniques to combine hydraulic tomography with geophysical measurements which seems promising. Furthermore, advancement in geostatistics, such as the development of Multiple Point Geostatistical methods that do not require a well-defined training image a priori, such as level set (Osher and Fedkiw 2006) and plurigaussian (Armstrong et al. 2011) methods, could lead to more efficient ways of exploring the vast subsurface parameter space while maintaining geological realism.

Chapter 6 dealt with uncertainty predominantly from field data by developing a stochastic DFN framework to represent the subsurface fractured bedrock connectivity. Even though the Eramosa Bedrock River Site (EBRS) was relatively rich in data, compared to the average bedrock river site (Tinkler and Wohl 1998), a number of data gaps still remained. This includes the average length of horizontal fracture, the average number of vertical fractures that pass through a mechanical layer, dispersivities (longitudinal and transverse), and matrix tortuosity.

As an outlook, numerous directions could be taken to further develop the EBRS modelling study. There are two immediate directions that I think have the potential to deliver critical insights. The first is a direct comparison of the current stochastic DFN model with a few well-selected equivalent porous media (EPM) models (e.g., a single homogeneous isotropic vs. anisotropic layer, 5 homogeneous isotropic layers coinciding with the mechanical layers) to investigate the possibility of finding effective dispersion and anisotropy values that could replace the need for an explicit DFN model. The second is a sensitivity analysis of the various stochastic DFN model parameters to identify the vital few that dominate the simulated hyporheic exchange response (i.e., residence time distribution and extents). This can also include stochastifying parameters that were assumed to be constant (e.g., dip angle, dispersivity, matrix tortuosity and porosity).

Two additional future directions that may require more effort are: (1) validate the model, since all available data went into building the current stochastic DFN model. Thus, another data set from additional field investigations (i.e., a reactive tracer test, Knapp et al. 2017) could be used for model validation. (2) Create a three-dimensional model of the EBRS, and (3) couple the current model to a computational fluid dynamic (CFD) model (e.g., OpenFOAM, Trauth et al. 2013) and include transient surface flows.

7.2 Scientists and their role in advancing surface water-groundwater models

A true test for a scientific theory is to see whether it can predict a cause and effect relationship. Moreover, scientists are in constant search for explanations of why these relationships exist and to determine accurately causation from more obscure correlation (Wright 1921). As the old adage goes “even a broken clock is right twice a day” (Addison and Steele 1853). This means that even an unreliable hypothesis can still be right about something every once in a while, even if only by coincidence. When our hypothesis is wrong for the other 99.9 % of the minutes in a day, we may need to revisit and adjust our suggested explanation for the observed phenomenon.

A numerical model is merely a scientific theory expressed as a set of equations arranged in a particular way that represents a simplified real-system. It is well understood that “all models are wrong” (Box and Draper 1987) because they are simplifications of a complex reality. However, models can still be useful given that they incorporate enough of the important processes under the right circumstances. For environmental models, particularly in the age of highly-parameterized integrated hydrological models, there is an increasing need to understand if we are getting the right answers for the right reasons (Kirchner 2006).

Thus, to find out if a hydrological theory is correct, it must be tested to its limits. Therefore, the role of a hydrological modeller in science (excessive in engineering) should be to constantly find ways to falsify his/her model. In other words, the modeller should try to uncover scenarios and circumstances where his/her model no longer works. By seeing where the model “breaks”, we can determine where the limits of our current understanding are and can then challenge our knowledge and intuition. It is easy to find evidence to support a model, a theory, or one’s intuition by looking only for “white swans”, however it takes only a single “black swan” to prove intuition wrong (Pilkey and Pilkey-Jarvis 2007). This can be a way to gain valuable new insights into our system understanding that our intuition may have failed to grasp initially.

Analogously, a scientist to a model is like a test pilot to a prototype aircraft. Test pilots continually refer to what is called the “envelope” for a prototype aircraft (Wolfe 1979). The envelope consists of an upper and lower edge, where the upper edge marks the maximum height and speed that the prototype aircraft has flown before it comes apart, malfunctions, or self-destructs. The goal of the test pilot is to continually push the upper edge of this failure envelope upwards by flying ever faster and higher in order to discover the absolute limits that a particular aircraft is capable of achieving (Tracy 2015). Valuable insights are gained by pushing the prototype aircraft to its physical limits.

The upper edge of the envelope for a hydrological model is where a model is stress-tested (Appendix A) to the point of falsification (i.e., can no longer recreate observed data). An example of a model stress-test is to apply forcing outside of its original design (i.e., outside the calibration period) and seeing whether the model is still capable of adequately forecasting the system response. The scientific modeller should be continuously seeking to define (the limits of our “known knowns”, Section 1.1.1) and advance the upper edge of the hydrological modelling envelope. Furthermore, he/she will hypothesize and study in detail the reasons that cause a model to be falsified (thus revealing “unknown unknowns” and transforming them to “known unknowns and/or knowns”, Section 1.1.1). In doing so, the modeller will simultaneously improve hydrological modelling capabilities and advance the hydrological knowledge base.

Numerical models in hydro(geo)logy will continue to be an important tool in the hydro(geo)logist’s tool box. However, there is a danger when “the only tool you have is a hammer, to treat everything as if it were a nail” (Maslow 1966). This is especially alarming for scientists trained in modelling schools who may erroneously believe that mathematical models can answer all questions. Therefore, it is important to always approach numerical models with a healthy level of skepticism (Pilkey and Pilkey-Jarvis 2007). Furthermore, it is critical to combine the model development process (Figure 1.5) with data obtained by other hydrological investigative tools, such as field and laboratory studies (e.g., Chapters 3, 4, 5, 6, and Appendix A). The idea of combining modelling theory with real-world data should be a tenet modellers adhere closely to, in order to make major advances in our current understanding of hydrological processes (Kirchner 2016).

In today’s era of advanced computing and integrated modelling (Section 2.1) the expectations of models to meet evermore challenging objectives may begin to exceed our current knowledge of environmental systems. Presented with the challenge of increasingly complex environmental problems, there is a growing need for multi-disciplinary knowledge and effective communication between scientists (Hunt and Zheng 2012). Hydro(geo)logists will be expected to defend and explain their models to economists, politicians, as well as other scientists and engineers. To deal with the model uncertainties of tomorrow, it will be necessary for the modeller to be not only an expert in modelling but also to have a firm grasp of the philosophies and practices of other environmental sciences and statisticians (Beven 2009). I hope that my work provides a small step in further uniting the disciplines of surface water and groundwater modelling, providing insights towards more effective solutions for surface water-groundwater interaction problems.

8. References

Abrams, D., Haitjema, H., and Kauffman, L. (2013). On modeling weak sinks in MODPATH. Groundwater, 51(4), pp.597-602.

Addison, J. and Steele, R. (1853). The spectator (Vol. 1). Thomas Bosworth.

Altenkirch, N., Zlatanovic, S., Woodward, K.B., Trauth, N., Mutz, M., & Mokenthin, F. (2016). “Untangling Hyporheic Residence time Distributions and Whole Stream”“Metabolism Using a Hydrological Process Model. Procedia engineering, 154, 1071-1078.

Anderson, M.P. and Woessner, W.W. (1992). The role of the postaudit in model validation. Advances in Water Resources, 15(3), pp.167-173.

Anderson, M.P., Woessner, W.W., and Hunt, R.J. (2015). Applied groundwater modeling: simulation of flow and advective transport, 2nd Edition. Academic press.

Antherton, J.S. (2013). Doceco; Knowing and not knowing [On-line: UK] retrieved January 21, 2016. http://www.doceo.co.uk/tools/knowing.htm

Aquanty Inc., HydroGeoSphere. (2015). A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. Waterloo, Ontario.

AquaResource Inc. (2010). Towns of Orangeville and Mono: Wellhead Protection Area Delineation Report. Submitted March 17, 2010 to R.J. Burnside and Associates. Breslau, Ontario.

ARANZ Geo Limited. (2017). Leapfrog Geo. Retrieved from: http://www.leapfrog3d.com/products/leapfrog-geo

Armstrong, M., Galli, A., Beucher, H., Le Loc'h, G., Renard, D., Doligez, B., Eschard, R. and Geffroy, F. (2011). Plurigaussian simulations in geosciences. Springer Science & Business Media.

Bätz, N., Colombini, P., Cherubini, P., and Lane, S.N. (2016). Groundwater controls on biogeomorphic succession and river channel morphodynamics. Journal of Geophysical Research: Earth Surface, 121(10), 1763-1785.

Barlow, P. M., Leake, S. A., and Fienen, M. N. (2018). Capture Versus Capture Zones: Clarifying Terminology Related to Sources of Water to Wells. Groundwater, 56(5), 694-704.

Bear, J. (1972). Dynamics of Fluids in Porous Media. American Elsevier, New York, N.Y.

Beatty, S.J., Morgan, D.L., McAleer, F.J., and Ramsay, A.R. (2010). Groundwater contribution to baseflow maintains habitat connectivity for Tandanus bostocki (Teleostei: Plotosidae) in a south-western Australian river. Ecology of Freshwater Fish, 19(4), 595-608.

Beckers, F., Noack, M. and Wieprecht, S. (2018). Uncertainty analysis of a 2D sediment transport model: an example of the Lower River Salzach. Journal of Soils and Sediments, pp.1-12.

Beckers, J. and Frind, E.O. (2001). Simulating groundwater flow and runoff for the Oro Moraine aquifer system. Part II. Automated calibration and mass balance calculations. Journal of Hydrology, 243(1-2), 73-90.

Bencala, K.E. (1993). A perspective on stream-catchment connections. Journal of the North American Benthological Society, 12(1), 44-47.

Berkowitz, B., Bear, J., and Braester, C. (1988). Continuum models for contaminant transport in fractured porous formations, Water Resour. Res., 24(8):1225–1236.

Bester, M. L., Frind, E.O., Molson, J.W., and Rudolph, D.L. (2006). Numerical investigation of road salt impact on an urban wellfield. Groundwater, 44(2), 165-175.

Beven, K. (2009). Environmental modelling: An uncertain future? An introduction to techniques for uncertainty estimation in environmental prediction. Routledge, p. 310.

Beven, K. and Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological processes, 6(3), pp.279-298.

Beven, K., Romanowicz, R., Pappenberger, F., Young, P.C. and Werner, M. (2005). The uncertainty cascade in flood forecasting. In Proceedings of the ACTIF meeting on Flood Risk (pp. 1-9).

Beven, K. and Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. Water Resources Research, 49(8), pp.5092-5098.

Bit͡silli, P. M. (1983). Chekhov's art, a stylistic analysis. Ardis.

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

Bradbury, K.R. and Muldoon, M.A. (1990). Hydraulic Conductivity Determinations in Unlithified Glacial and Fluvial Materials. Ground Water and Vadose Zone Monitoring, 1053, p.138.

Bredehoeft, J. (2010). Models and model analysis. Ground Water, 48(3), 328-328.

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

Brunton, F.R. (2009). Update of revisions to the Early Silurian stratigraphy of the Niagara Escarpment: integration of sequence stratigraphy, sedimentology and hydrogeology to delineate hydrogeologic units; in Summary of Field Work and Other Activities 2009. Ontario Geological Survey, Open File Report 6240, p.25-1 to 25-20.

Buffington, J.M. and Tonina, D. (2009). Hyporheic exchange in mountain rivers II: effects of channel morphology on mechanics, scales, and rates of exchange. Geography Compass, 3(3), pp.1038-1062.

Buss, S., Cai, Z., Cardenas, B., Fleckenstein, J., Hannah, D., Heppell, K., Hulme, P., Ibrahim, T., Kaeser, D., Krause, S. and Lawler, D. (2009). The Hyporheic Handbook: a handbook on the groundwater-surface water interface and hyporheic zone for environment managers. Environmental Agency. Almondsbury, Bristol, United Kingdom.

Cacas, M.C., Ledoux, E., Marsily, G.D., Tillie, B., Barbreau, A., Durand, E., Feuga, B. and Peaudecerf, P. (1990). Modeling fracture flow with a stochastic discrete fracture network: calibration and validation: 1. The flow model. Water Resources Research, 26(3), pp.479-489.

Capes, D.C. (2018). Utility of Ambient Groundwater Temperature Profiling within Sealed Bedrock Boreholes for Fracture Flow Characterization in Seasonally Dynamic Environments. Master’s Thesis. University of Guelph. Guelph, Canada.

Capes, D.C., Steelman, C.M., and Parker, B.L. (2018). Hydrologic interpretation of seasonally dynamic ambient temperature profiles in sealed bedrock boreholes. Journal of Hydrology, 567, 133-148.

Cardenas, M.B. (2015). Hyporheic zone hydrologic science: A historical account of its emergence and a prospectus. Water Resources Research, 51(5), 3601-3616.

Cardenas, M.B. (2008). Surface water-groundwater interface geomorphology leads to scaling of residence times. Geophysical Research Letters, 35(8).

Cardenas, M.B. and Wilson, J.L. (2007a). Dunes, turbulent eddies, and interfacial exchange with permeable sediments. Water Resources Research, 43(8).

Cardenas, M.B. and Wilson, J.L. (2007b). Exchange across a sediment–water interface with ambient groundwater discharge. Journal of Hydrology, 346(3), 69-80.

Cardenas, M.B., Wilson, J., and Zlotnik, V.A. (2004). Impact of heterogeneity, bed forms, and stream curvature on subchannel hyporheic exchange. Water Resources Research, 40(8).

Carrera, J. and Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. Water Resources Research, 22(2), 211-227.

Caruso, A., Ridolfi, L., and Boano, F. (2016). Impact of watershed topography on hyporheic exchange. Advances in Water Resources, 94, 400-411.

Chapman, S.W., Parker, B.L., Cherry, J.A., Martin, P., Abbey, D. and McDonald, S.D. (2014). Combined EPM-DFN Modeling Approach for Plumes in Sedimentary Bedrock Aquifers. In DFNE-International Conference on Discrete Fracture Network Engineering.

CH2MHILL and S.S. Papadopulos & Associates, Inc. (2003). Alder creek groundwater study: Final report. Prepared for The Regional Municipality of Waterloo.

Chen, J. (2015). Impact of Climate Change on Canadian Water Resources: A Continental-Scale Hydrologic Modelling Study Using Multiple RCM Projections (Ph. D. thesis). Department of Earth Sciences, University of Waterloo, Waterloo, ON, Canada.

Chow, R. (2012). Delineating Base Flow Contribution Areas for Streams: A Model Comparison. Earth Science Master’s thesis, University of Waterloo.

Chow, R., Frind, E.O., Sousa, M., Jones, J.P., Rudolph, D.L., and Molson, J.W. (2011). Delineating Capture Zones for Environmentally Sensitive Features – A Model Comparison. Geohydro Conference, Québec City, Québec Canada.

Chow, R., Frind, M.E., Frind, E.O., Jones, J.P., Sousa, M.R., Rudolph, D.L., Molson, J.W., and Nowak, W. (2016). Delineating baseflow contribution areas for streams – A model and methods comparison. Journal of Contaminant Hydrology, 195, 11-22. doi:http://dx.doi.org/10.1016/j.jconhyd.2016.11.001

Chow, R., Wu, H., Bennett, J.P., Dugge, J., Wöhling, T., and Nowak, W. (2018). Sensitivity of simulated hyporheic exchange to river bathymetry: The Steinlach River Test Site. Groundwater.

Cirpka, O.A. and Bürger, C.M. (2012). Environmental Modeling 1: Water Cycle and Flow Modeling. Universität Tübginen, Zentrum für Angewandte Geowissenschaften, Hölderlinstr. 12, 72074 Tübingen. WS2011/2012.

Cirpka, O.A. and Valocchi, A.J. (2016). Debates—Stochastic subsurface hydrology from theory to practice: Does stochastic subsurface hydrology help solving practical problems of contaminant hydrogeology?. Water Resources Research, 52(12), pp.9218-9227.

Cleary, T.C., and Cleary, R.W. (1991). Delineation of wellhead protection areas: theory and practice. Water Science and Technology, 24(11), 239-250.

Cole, J., Coniglio, M. and Gautrey, S. (2009). The role of buried bedrock valleys on the development of karstic aquifers in flat-lying carbonate bedrock: insights from Guelph, Ontario, Canada. Hydrogeology Journal, 17(6), pp.1411-1425.

Cramér, H. (1946). Mathematical methods of statistics. Princeton, NJ. Princeton university press.

Daus, A.D., Frind, E.O., & Sudicky, E.A. (1985). Comparative error analysis in finite element formulations of the advection-dispersion equation. Advances in Water Resources, 8(2), 86-95.

DHI-WASY (2015). FEFLOW 7.0: Finite Element Subsurface Flow and Transport simulation system. User Manual. DHI-WASY GmbH, Berlin.

Diersch, H.J.G. (2013). FEFLOW: finite element modeling of flow, mass and heat transport in porous and fractured media. Springer Science & Business Media.

Dietrich, C.R. and Newsam, G.N. (1993). A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resources Research, 29(8), 2861–2869. https://doi.org/10.1029/93WR01070

Doherty, J. (2003). Ground water model calibration using pilot points and regularization. Ground Water, 41(2), 170-177.

Doherty, J. (2005). PEST Version 9 User's Guide. Brisbane, Queensland, Australia: Watermark Numerical Computing.

Doherty, J. (2011). Modeling: picture perfect or abstract art?. Groundwater, 49(4), pp.455-455.

Doherty, J.E., Fienen, M.N., and Hunt, R.J. (2010). Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions. US Geological Survey scientific investigations report, 5168(2010), p.1-36.

Doro, K.O., Leven, C., and Cirpka, O.A. (2013). Delineating subsurface heterogeneity at a loop of River Steinlach using geophysical and hydrogeological methods. Environmental earth sciences, 69(2), 335-348.

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

Engesgaard, P., Jensen, K.H., Molson, J., Frind, E.O., and Olsen, H. (1996). Large-scale dispersion in a sandy aquifer: Simulation of subsurface transport of environmental tritium. Water Resources Research, 32(11), 3253-3266.

Enzenhoefer, R., Bunk, T., and Nowak, W. (2014). Nine steps to risk-informed wellhead protection and management: A case study. Groundwater, 52(S1), 161-174.

Ferré, T.P. (2017). Revisiting the relationship between data, models, and decision-making. Groundwater, 55(5), pp.604-614.

Feynman, R.P. (1966). The development of the space-time view of quantum electrodynamics. Science, 153(3737), pp.699-708.

Findlay, S. (1995). Importance of surface-subsurface exchange in stream ecosystems: The hyporheic zone. Limnology and Oceanography, 40(1), 159-164.

Fogg, G.E., LaBolle, E.M. and Weissmann, G.S. (1999). Groundwater vulnerability assessment: Hydrogeologic perspective and example from Salinas Valley, California. GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION, 108, pp.45-62.

Forsyth, P.A. (1991). A control volume finite element approach to NAPL groundwater contamination, SIAM J. Sci. Stat. Comput., 12(5):1029–1057.

Forsyth, P.A. and Simpson, R.B. (1991). A two phase, two component model for natural convection in a porous medium, Int. J. Num. Meth. Fluids, 12:655–682.

Franke, O.L., Really, T.E., Pollock, D.W., and LaBaugh, J.W. (1998). Estimating areas contributing recharge to wells. US Geological Survey Circular 1174, US Geological Survey, Reston VA.

Freeze, R.A. and Cherry, J. (1979). Groundwater. Prentice-Hall, Ine. Englewood Cliffs, New Jersey, 7632, 604, pp. 4, 26-30.

Freeze, R.A. and Harlan, R.L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model. Journal of Hydrology, 9(3), pp.237-258.

Frei, S. and Fleckenstein, J.H. (2013). Representing effects of micro-topography on runoff generation and sub-surface flow patterns by using superficial rill/depression storage height variations. Environmental modelling & software, 52, pp.5-18.

Frind, E.O. (2003). Fundamentals of Groundwater Modelling. Department of Earth Sciences, University of Waterloo.

Frind, E.O. and Molson, J.W. (2004). A New Particle Tracking Algorithm for Finite Element Grids. FEM-MODFLOW Conference, Carlsbad, Czech Republic.

Frind, E.O. and Molson, J.W. (2018). Issues and Options in the Delineation of Well Capture Zones under Uncertainty. Groundwater, 56(3), pp.366-376.

Frind, E.O., Molson, J.W., and Rudolph, D.L. (2006). Well vulnerability: a quantitative approach for source water protection. Groundwater, 44(5), 732-742.

Frind, E.O., Molson, J.W., Sousa, M.R., and Martin, P.J. (2014). Insights from four decades of model development on the Waterloo Moraine: A review. Canadian Water Resources Journal / Revue canadienne des ressources hydriques 39(2): 149-166.

Frind, E.O., Muhammad, D.S., and Molson, J.W. (2002). Delineation of Three-Dimensional Capture Zones in Complex Multi-Aquifer Systems. Ground Water, Vol. 40, No. 6, pp. 586-589.

Frind, M., Sousa, M., Molson, J., Frind, E., and Rudolph, D. (2015). Predicting the impact of surface sources on an aquifer: the role of dispersion. Proceedings, IAH-CNC Waterloo, 27-30.

Gelhar, L.W. and Axness, C.L. (1983). Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resources Research 19(1): 161-180.

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

Geomodeling Technology Corp. (1996-2016). SBED [Software].

Gómez-Hernández, J.J. (2006). Complexity. Ground Water 4, no.6: 782-785.

Gomez-Velez, J.D., Harvey, J.W., Cardenas, M.B. and Kiel, B. 2015. Denitrification in the Mississippi River network controlled by flow through river bedforms. Nature Geoscience, 8(12), pp.941-945.

Gooseff, M.N., Anderson, J.K., Wondzell, S.M., LaNier, J. and Haggerty, R. (2006). A modeling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA. Hydrological Processes, 20(11), pp.2443-2457.

Gordon, N.J., Salmond, D.J., and Smith, A.F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Paper presented at the IEE Proceedings F-Radar and Signal Processing.

Grand River Source Protection Area (2015). Approved source protection plan. Prepared on behalf of the Lake Erie region source protection committee, under the clean water act, 2006 (Ontario Regulation 287/07).

Grathwohl, P. (1998). Diffusion in Natural Porous Media: Contaminant Transport, Sorption/Desorption and Dissolution Kinetics. Topics in Enviornmental Mechanics. Springer US, Boston, MA.

Grathwohl, P., Rügner, H., Wöhling, T., Osenbrück, K., Schwientek, M., Gayler, S., Wollschläger, U., Selle, B., Pause, M., Delfs, J.O., Grzeschik, M., Weller, U., Ivanov, M., Cirpka, O.A., Maier, U., Kuch, B., Nowak, W., Wulfmeyer, V., Warrach-Sagi, K., Streck, T., Attinger, S., Bilke, L., Dietrich, P., Fleckenstein, J.H., Kalbacher, T., Kolditz, O., Rink, K., Samaniego, L., Vogel, H.J., Werban, U., and Teutsch, G. (2013). Catchments as reactors: a comprehensive approach for water fluxes and solute turnover. Environmental earth sciences, 69(2), pp.317-333.

Guardiano, F.B., and Srivastava, R.M. (1993). Multivariate geostatistics: beyond bivariate moments Geostatistics Troia’92 (pp. 133-144): Springer.

Gupta, H.V., Beven, K.J., and Wagener, T. (2005). Model calibration and uncertainty estimation. Encyclopedia of hydrological sciences.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., and Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. Water Resources Research, 48(8).

Guthke, A. (2017). Defensible model complexity: A call for data-based and goal-oriented model choice. Groundwater, 55(5), pp.646-650.

Haitjema, H.M. and Mitchell-Bruker, S. (2005). Are water tables a subdued replica of the topography?. Groundwater, 43(6), pp.781-786.

Heinz, J. and Aigner, T. (2003). Hierarchical dynamic stratigraphy in various Quaternary gravel deposits, Rhine glacier area (SW Germany): implications for hydrostratigraphy. International Journal of Earth Sciences, 92(6), pp.923-938.

Hibbs, B.J. and Sharp Jr, J.M. (2012). Hydrogeological impacts of urbanization. Environmental & Engineering Geoscience, 18(1), 3-24.

Hill, M.C. (2006). The practical use of simplicity in developing ground water models. Groundwater, 44(6), 775-781.

Höge, M., Wöhling, T., and Nowak, W. (2018). A primer for model selection: The decisive role of model complexity. Water Resources Research, 54(3), pp.1688-1715.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian model averaging: a tutorial. Statistical science, pp.382-401.

Hooke, R.L.B. (1975). Distribution of sediment transport and shear stress in a meander bend. The Journal of geology, 83(5), pp.543-565.

Howard, K.W. and Israfilov, R.G. eds. (2012). Current problems of hydrogeology in urban areas, urban agglomerates and industrial centres (Vol. 8). Springer Science & Business Media.

Hu, L.Y. and Chugunova, T. (2008). Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. Water Resources Research, 44(11).

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

Hunt, R.J. and Zheng, C. (2012). The current state of modeling. Groundwater, 50(3), pp.330-333.

Huyakorn, P.S. and Pinder, G.F. (1983). Computational Methods in Subsurface Flow, Academic Press, New York.

Hwang, H.T., Frey, S.K., Park, Y.J., Pintar, K.D.M., Lapen, D.R., Thomas, J.L., Spoelstra, J., Schiff, S.L., Brown, S.J. and Sudicky, E.A. (2019). Estimating cumulative wastewater treatment plant discharge influences on acesulfame and Escherichia coli in a highly impacted watershed with a fully-integrated modelling approach. Water Research.

Jones, J.P. and Mendoza, C. (2012). Alberta Oil Sands Groundwater Modelling Guidelines. Cumulative Environmental Management Association (CEMA).

Jones, J.P., Sousa, M.R., Frind, E.O., and Rudolph, D.L. (2009). Determining the influence of surface, unsaturated and saturated processes on source water protection strategies: a multi-model study. GeoHalifax. Halifax, Nova Scotia, Canada.

Jones, J.P., Sudicky, E.A. and McLaren, R.G. (2008). Application of a fully-integrated surface-subsurface flow model at the watershed-scale: A case study. Water Resources Research, 44(3).

Juran, J.M., Gryna, F.M. and Bingham, R.S. (1974): Quality Control Handbook. McGraw-Hill, New York. Basic Concepts, pp.2-16 and 2-17.

Käser, D.H., Binley, A., and Heathwaite, A.L. (2013). On the importance of considering channel microforms in groundwater models of hyporheic exchange. River Research and Applications, 29(4), pp.528-535.

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

Kasahara, T. and Wondzell, S.M. (2003). Geomorphic controls on hyporheic exchange flow in mountain streams. Water Resources Research, 39(1), pp.SBH-3.

Kavetski, D., Franks, S. W., and Kuczera, G. (2003). Confronting input uncertainty in environmental modelling. Calibration of watershed models, 6, 49-68.

Kennedy, C.S.C. (2017). Groundwater-Surface Water Interactions in the Discrete Fracture Networks of Bedrock Rivers (Doctoral dissertation). University of Guelph.

Kiel, B.A. and Cardenas, M.B. 2014. Lateral hyporheic exchange throughout the Mississippi River network. Nature Geoscience, 7(6), p.413.

Kirchner, J.W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).

Kitanidis, P.K. (1995). Quasilinear geostatistical theory for inversing. Water Resources Research, 31(10), pp. 2411–2419.

Knapp, J.L., González-Pinzón, R., Drummond, J.D., Larsen, L.G., Cirpka, O.A. and Harvey, J.W. (2017). Tracer-based characterization of hyporheic exchange and benthic biolayers in streams. Water Resources Research, 53(2), pp.1575-1594.

Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.O., Fischer, T., Görke, U.J., Kalbacher, T., Kosakowski, G., McDermott, C.I. and Park, C.H. (2012). OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media. Environmental Earth Sciences, 67(2), pp.589-599.

Kollet, S., Sulis, M., Maxwell, R.M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E.T., Cordano, E., Endrizzi, S., Kikinzon, E. and Mouche, E. (2017). The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks. Water Resources Research, 53(1), pp.867-890.

Konikow, L.F. and Bredehoeft, J.D. (1992). Ground-water models cannot be validated. Advances in water resources, 15(1), pp.75-83.

Kunert, M. and Coniglio, M. (2002). Origin of vertical shafts in bedrock along the Eramosa River valley near Guelph, southern Ontario. Canadian Journal of Earth Sciences, 39(1), pp.43-52.

Kunert, M., Coniglio, M. and Jowett, E.C. (1998). Controls and age of cavernous porosity in Middle Silurian dolomite, southern Ontario. Canadian Journal of Earth Sciences, 35(9), pp.1044-1053.

La Pointe, P.R. (2010). Techniques for Identification and Prediction of Mechanical Stratigraphy in Fractured Rock Masses. In 44th US Rock Mechanics Symposium and 5th US-Canada Rock Mechanics Symposium. American Rock Mechanics Association.

Lee, D.R. and Hynes, H.B.N. (1978). Identification of groundwater discharge zones in a reach of Hillman Creek in southern Ontario. Water Quality Research Journal, 13(1), pp.121-134.

Lehr, J.H. (1979). Mathematical ground-water models may be intellectual toys today, but they should be useful tools tomorrow. Groundwater, 17(5), pp.418-422.

Leube, P.C., Geiges, A., and Nowak, W. (2012). Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. Water Resources Research, 48(2).

Lichtner, P.C., Kelkar, S., and Robinson, B. (2002). New form of dispersion tensor for axisymmetric porous media with implementation in particle tracking. Water Resources Research, 38(8).

Liu, S. and Chui, T.F.M. (2018). Impacts of streambed heterogeneity and anisotropy on residence time of hyporheic zone. Groundwater, 56(3), pp.425-436.

MacQuarrie, K.T. and Mayer, K.U. (2005). Reactive transport modeling in fractured rock: A state-of-the-science review. Earth-Science Reviews, 72(3-4), pp.189-227.

Mahmud, K., Mariethoz, G., Caers, J., Tahmasebi, P. and Baker, A. (2014). Simulation of Earth textures by conditional image quilting. Water Resources Research, 50(4), pp.3088-3107.

Martin, P.J. and Frind, E.O. (1998). Modeling a complex multi-aquifer system: The Waterloo Moraine. Groundwater, 36(4), 679-690.

Maslow, A. H. (1966). The Psychology of Science: A Reconnaissance. Maurice Bassett Publishing. p.15.

MATLAB, MathWorks, Inc. (2015). MATLAB and Statistics Toolbox Release 2015b. Natick, Massachusetts, United States.

Maxwell, R.M., Condon, L.E. and Kollet, S.J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. Geoscientific Model Development, 8(3), p.923.

Maxwell, R.M., Kollet, S.J., Smith, S.G., Woodward, C.S., Falgout, R.D., Ferguson, I.M., Baldwin, C., Bosl, W.J., Hornung, R. and Ashby, S. (2009). ParFlow user's manual. International Ground Water Modeling Center Report GWMI, 1(2009), p.129.

Maxwell, R.M., Putti, M., Meyerhoff, S., Delfs, J.O., Ferguson, I.M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S.J., Kumar, M. and Lopez, S. (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. Water resources research, 50(2), pp.1531-1549.

McDonald, M.G. and Harbaugh, A.W. (1988). A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model: U.S. Geological Survey Techniques of Water-Resource Investigations, book 6, Chapter A1, pp. 586.

McDonnell, J.J., McGuire, K., Aggarwal, P., Beven, K.J., Biondi, D., Destouni, G., Dunn, S., James, A., Kirchner, J., Kraft, P., and Lyon, S. (2010). How old is streamwater? Open questions in catchment transit time conceptualization, modelling and analysis. Hydrological Processes, 24(12), pp.1745-1754.

McGuire, K.J. and McDonnell, J.J. (2006). A review and evaluation of catchment transit time modeling. Journal of Hydrology, 330(3-4), pp.543-563.

Merritt, W.S., Letcher, R.A. and Jakeman, A.J. (2003). A review of erosion and sediment transport models. Environmental modelling & software, 18(8-9), pp.761-799.

Merwade, V.M., Maidment, D.R. and Hodges, B.R. (2005). Geospatial representation of river channels. Journal of Hydrologic Engineering, 10(3), pp.243-251.

Meyer, P. A., Brouwers, M., and Martin, P. (2014). A three-dimensional groundwater flow model of the Waterloo Moraine for water resource management. Canadian Water Resources Journal / Revue canadienne des ressources hydriques 39(2): 167-180.

Molson, J.W., Beckers, J., Frind, E.O., and Martin, P.J. (2002). WATFLOW/3D: A Three-Dimensional Numerical Model for Coupled Groundwater Surface Water Flow, Version 4.0. Department of Earth Sciences, University of Waterloo.

Molson, J.W. and Frind, E.O. (2012). On the use of mean groundwater age, life expectancy and capture probability for defining aquifer vulnerability and time-of-travel zones for source water protection. Journal of contaminant hydrology, 127(1-4), 76-87.

Molson, J.W. and Frind, E.O., (2004). WTC: Waterloo Transport Code, Version 3.0, Advective-Dispersive Mass Transport Model in Three Dimensions, User Guide. Department of Earth Sciences, University of Waterloo.

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

Morgan, G. and Henrion, M. (1990). The nature and sources of uncertainty. In: Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis. Cambridge, U.K.: Cambridge University Press. pp. 47-72.

Motavita, D.F. (2016). Data and parameter uncertainty quantification of hydrological models via Bayesian updating: Comparison between five Bavarian catchments. WAREM Master’s Thesis. University of Stuttgart, Stuttgart, Germany.

Munn, J. (2012). High-resolution discrete fracture network characterization using inclined coreholes in a Silurian dolostone aquifer in Guelph, Ontario. Master’s Thesis. University of Guelph. Guelph, Canada.

National Research Council. 2002. Riparian Areas: Functions and Strategies for Management, Natl. Acad. Press, Washington, D. C.

Neupauer, R.M. and Wilson, J.L. (2001). Adjoint-derived location and travel time probabilities for a multidimensional groundwater system. Water Resources Research, 37(6), 1657-1668.

Odsæter, L.H., Wheeler, M.F., Kvamsdal, T., & Larson, M.G. (2017). Postprocessing of non-conservative flux for compatibility with transport in heterogeneous media. Computer Methods in Applied Mechanics and Engineering, 315, 799-830.

Osenbrück, K., Wöhling, T., Lemke, D., Rohrbach, N., Schwientek, M., Leven, C., and Cirpka, O.A. (2013). Assessing hyporheic exchange and associated travel times by hydraulic, chemical, and isotopic monitoring at the Steinlach Test Site, Germany. Environmental earth sciences, 69(2), 359-372.

Osher, S. and Fedkiw, R. (2006). Level set methods and dynamic implicit surfaces (Vol. 153). Springer Science & Business Media.

Oxtobee, J.P. and Novakowski, K. (2002). A field investigation of groundwater/surface water interaction in a fractured bedrock environment. Journal of Hydrology, 269(3-4), pp.169-193.

Oxtobee, J.P. and Novakowski, K.S. (2003). Ground water/surface water interaction in a fractured rock aquifer. Groundwater, 41(5), pp.667-681.

Packman, A.I., Salehin, M. and Zaramella, M. (2004). Hyporheic exchange with gravel beds: basic hydrodynamic interactions and bedform-induced advective flows. Journal of Hydraulic Engineering, 130(7), pp.647-656.

Panday, S., Huyakorn, P.S., Therrien, R. and Nichols, R.L. (1993). Improved three-dimensional finite-element techniques for field simulation of variably saturated flow and transport. Journal of contaminant hydrology, 12(1-2), pp.3-33.

Panday, S., Langevin, C.D., Niswonger, R.G., Ibaraki, M., and Hughes, J.D. (2015). MODFLOW–USG version 1.3.00: An unstructured grid version of MODFLOW for simulating groundwater flow and tightly coupled processes using a control volume finite-difference formulation: US Geological Survey Software Release, 01 December 2015.

Pappenberger, F. and Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42: 8.

Park, Y.J., Sudicky, E.A., Panday, S. and Matanga, G. (2009). Implicit Subtime Stepping for Solving Nonlinear Flow Equations in an Integrated Surface–Subsurface System. Vadose Zone Journal, 8(4), pp.825-836.

Parker, B.L., Bairos, K., Maldaner, C.H., Chapman, S.W., Turner, C.M., Burns, L.S., Plett, J., Carter, R., and Cherry, J.A. (2018). Metolachlor dense non-aqueous phase liquid source conditions and plume attenuation in a dolostone water supply aquifer. Geological Society, London, Special Publications, 479, SP479-9.

Parker, B.L., Gillham, R.W. and Cherry, J.A. (1994). Diffusive disappearance of immiscible-phase organic liquids in fractured geologic media. Ground Water, 32, pp. 805–820.

Parker, G., Shimizu, Y., Wilkerson, G.V., Eke, E.C., Abad, J.D., Lauer, J.W., Paola, C., Dietrich, W.E. and Voller, V.R. (2011). A new framework for modeling the migration of meandering rivers. Earth Surface Processes and Landforms, 36(1), pp.70-86.

Pham, H.V. and Tsai, F.T.C. (2016). Optimal observation network design for conceptual model discrimination and uncertainty reduction. Water Resources Research, 52(2), 1245-1264.

Pilkey, O.H. and Pilkey-Jarvis, L. (2007). Useless arithmetic: why environmental scientists can't predict the future. Columbia University Press.

Pinder, G.F. and Frind, E.O. (1972). Application of Galerkin's procedure to aquifer analysis. Water Resources Research, 8(1), 108-120.

Pitt, M.K. (2002). Smooth particle filters for likelihood evaluation and maximisation. University of Warwick, Department of Economics, pp.8-10.

Pokrajac, D. and Lazic, R. (2002). An efficient algorithm for high accuracy particle tracking in finite elements. Advances in Water Resources, 25(4), 353-369.

Pollock, D.W. (1988). Semianalytical computation of pathlines for finite-difference models. Ground Water, Vol. 26, No. 6, pp.743-750.

Pollock, D.W. (1994). User's Guide for MODPATH/MODPATH-PLOT, Version 3: A Particle Tracking Post-processing Package for MODFLOW, the US: Geological Survey Finite-difference Ground-water Flow Model: US Geological Survey Open-File Report 94-464, 6 ch. Technical Report, United States Geological Survey.

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

Quinodoz, H., and Valocchi, A. (1990). Macrodispersion in heterogeneous aquifers. Atomic Energy of Canada Ltd., Chalk River, ON (Canada). Chalk River Nuclear Labs; 471 p; ISBN 0-662-18237-5; 1990; p. 455-468; International Conference and Workshop on Transport and Mass Exchange Processes in Sand and Gravel Aquifers; Ottawa, ON (Canada); 1-4 Oct 1990.

Rao, C.R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society. 37:81-89. Mathematical Reviews 0015748.

Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L., and Troldborg, L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Advances in water resources, 36, 36-50.

Richards, L.A. (1931). Capillary conduction of liquids through porous mediums. physics, 1(5), 318-333.

Rinaldo, A., Marani, A., and Rigon, R. (1991). Geomorphological dispersion. Water Resources Research, 27(4), pp.513-525.

Rodriguez-Pretelin, A. and Nowak, W. (2018). Integrating transient behavior as a new dimension to WHPA delineation. Advances in Water Resources, 119, 178-187.

Ross, J.L., Ozbek, M.M., & Pinder, G.F. (2009). Aleatoric and epistemic uncertainty in groundwater flow and transport simulation. Water Resources Research, 45(12).

Rubin, Y. (2003). Applied stochastic hydrogeology. Oxford University Press, New York, NY, p. 4.

Salehin, M., Packman, A.I. and Paradis, M. (2004). Hyporheic exchange with heterogeneous streambeds: Laboratory experiments and modeling. Water Resources Research, 40(11).

Sanders, R., 1987. The Pareto principle: its use and abuse. Journal of Services Marketing, 1(2), pp.37-40.

Schöniger, A., Illman, W.A., Wöhling, T., and Nowak, W. (2015). Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. Journal of Hydrology, 531, pp.96-110.

Schöniger, A., Wöhling, T., Samaniego, L., and Nowak, W. (2014). Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence. Water resources research, 50(12), 9484-9513.

Schwartz, F.W. (2013). Zombie-Science and Beyond. Groundwater, 51(1), pp.1-1.

Selzer, P. (2017). Particle tracking on finite elements for groundwater flow. AEG Master’s Thesis. University of Tübingen. Tübingen, Germany.

Selzer P. and Cirpka O.A. (2019, submitted): Postprocessing of Standard Finite-Element Velocity Fields for Accurate Particle Tracking Applied to Groundwater Flow, Computational Geosciences.

Shackleton, J.R., Cooke, M.L., and Sussman, A.J. (2005). Evidence for temporally changing mechanical stratigraphy and effects on joint-network architecture. Geology, 33(2), 101-104.

Shikaze, S.G., Sudicky, E.A. and Schwartz, F.W. (1998). Density-dependent solute transport in discretely-fractured geologic media: is prediction possible?. Journal of contaminant hydrology, 34(3), pp.273-291.

Singhal, B.B.S., and Gupta, R.P. (2010). Applied hydrogeology of fractured rocks, 2nd Edition. Springer Science & Business Media, pp. 145.

Silver, N. (2012). The signal and the noise: the art and science of prediction. Penguin UK.

Soltani, S.S. and Cvetkovic, V. (2017). Contaminant attenuation by shallow aquifer systems under steady flow. Advances in Water Resources, 108, pp.157-169.

Sousa, M.R., Frind, E.O., and Rudolph, D.L. (2013). An integrated approach for addressing uncertainty in the delineation of groundwater management areas. Journal of Contaminant Hydrology 148: 12-24.

Steelman, C.M., Kennedy, C.S. and Parker, B.L. (2015a). Geophysical conceptualization of a fractured sedimentary bedrock riverbed using ground-penetrating radar and induced electrical conductivity. Journal of Hydrology, 521, pp.433-446.

Steelman, C.M., Parker, B.L. and Kennedy, C.S., (2015b). Evaluating local-scale anisotropy and heterogeneity along a fractured sedimentary bedrock river using EM azimuthal resistivity and ground-penetrating radar. Journal of Applied Geophysics, 116, pp.156-166.

Steelman, C.M., Arnaud, E., Pehme, P. and Parker, B.L. (2017a). Geophysical, geological, and hydrogeological characterization of a tributary buried bedrock valley in southern Ontario. Canadian Journal of Earth Sciences, (999), pp.1-18.

Steelman, C.M., Kennedy, C.S., Capes, D.C. and Parker, B.L. (2017b). Electrical resistivity dynamics beneath a fractured sedimentary bedrock riverbed in response to temperature and groundwater–surface water exchange. Hydrology and Earth System Sciences, 21(6), p.3105.

Stonedahl, S.H., Harvey, J.W. and Packman, A.I. (2013). Interactions between hyporheic flow produced by stream meanders, bars, and dunes. Water Resources Research, 49(9), pp.5450-5461.

Stonedahl, S.H., Harvey, J.W., Wörman, A., Salehin, M., and Packman, A.I. (2010). A multiscale model for integrating hyporheic exchange from ripples to meanders. Water Resources Research, 46(12).

Stonedahl, S.H., Sawyer, A.H., Stonedahl, F., Reiter, C. and Gibson, C. (2018). Effect of Heterogeneous Sediment Distributions on Hyporheic Flow in Physical and Numerical Models. Groundwater, 56(6), pp.934-946.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. Mathematical Geosciences, 43(3), 305-328.

Sudicky, E.A., Cherry, J.A. and Frind, E.O. (1983). Migration of contaminants in groundwater at a landfill: A case study: 4. A natural-gradient dispersion test. Journal of Hydrology, 63(1-2), pp.81-108.

Sudicky, E.A. and McLaren, R.G. (1992). The Laplace transform Galerkin technique for large-scale simulation of mass transport in discretely-fractured porous formations, Water Resour. Res., 28(2):499–514.

Sun, N.Z. and Yeh, W.W.G. (1990). Coupled inverse problems in groundwater modeling: 1. Sensitivity analysis and parameter identification. Water resources research, 26(10), pp.2507-2525.

Tang, D.H., Frind, E.O., and Sudicky, E.A. (1981). Contaminant transport in fractured porous media: Analytical solution for a single fracture, Water Resour. Res., 17(3):555–564.

Taylor, J. R. (1982). An Introduction to Error Analysis, p327, Univ. Sci. Books, Mill Valley, Calif.

Therrien, R. and Sudicky, E.A. (1996). Three-dimensional analysis of variably-saturated flow and solute transport in discretely-fractured porous media. J. Contam. Hydrol., 23(1-2):1–44.

Thibodeaux, L.J. and Boyle, J.D. (1987). Bedform-generated convective transport in bottom sediment. Nature, 325(6102), 341.

Tinkler, K.J. and Parish, J. (1998). Recent adjustments to the long profile of Cooksville Creek, an urbanized bedrock channel in Mississauga, Ontario. GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION, 107, pp.167-188.

Tinkler, K. and Wohl, E. (1998). A primer on bedrock channels. Geophysical Monograph-American Geophysical Union, 107, 1-18.

Tonina, D. and Buffington, J.M. (2007). Hyporheic exchange in gravel bed rivers with pool-riffle morphology: Laboratory experiments and three-dimensional modeling. Water Resources Research, 43(1).

Tonina, D., de Barros, F.P., Marzadri, A., and Bellin, A. (2016). Does streambed heterogeneity matter for hyporheic residence time distribution in sand-bedded streams?. Advances in water resources, 96, pp.120-126.

Tóth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. Journal of geophysical research, 68(16), pp.4795-4812.

Tracy, B. (2015). The power of self-confidence. Wiley and Sons Inc., Hoboken, New Jersey, p.39.

Trauth, N., Musolff, A., Knöller, K., Kaden, U.S., Keller, T., Werban, U. and Fleckenstein, J.H. (2018). River water infiltration enhances denitrification efficiency in riparian groundwater. Water research, 130, pp.185-199.

Trauth, N., Schmidt, C., Maier, U., Vieweg, M., and Fleckenstein, J.H. (2013). Coupled 3-D stream flow and hyporheic flow model under varying stream and ambient groundwater flow conditions in a pool-riffle system. Water Resources Research, 49(9), pp.5834-5850.

Uffink, G.J. (1989). Application of Kolmogorov’s backward equation in random walk simulations of groundwater contaminant transport. Contaminant transport in groundwater, pp.283-289.

Underwood, C.A., Cooke, M.L., Simo, J.A., and Muldoon, M.A. (2003). Stratigraphic controls on vertical fracture patterns in Silurian dolomite, northeastern Wisconsin. AAPG bulletin, 87(1), 121-142.

VanderKwaak, J.E. and Loague, K. (2001). Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model. Water Resources Research 37: 999 – 1013.

Vaughn, D.M. (1990). Flood dynamics of a concrete-lined, urban stream in Kansas City, Missouri. Earth Surface Processes and Landforms, 15(6), 525-537.

Vaux, W.G. (1968). Intragravel flow and interchange of water in a streambed. Fishery Bulletin, 66(3), 479-489.

Veale, B., Cooke, S., Zwiers, G., and Neumann, M. (2014). The Waterloo Moraine: a watershed perspective. Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 39(2), 181-192.

Viessman, W. (Jr.) and Lewis, G.L. (1996). Introduction to Hydrology, 4th Edition, Harper Collins College Publisher, New York.

Voronina, K. (2013). Small-scale river hydraulics modeling study of the Steinlach test site using HEC-RAS. AEG Master's Thesis. University of Tübingen. Tübingen, Germany.

Voss, C.I. (2011a). Editor's message: Groundwater modeling fantasies—part 1, adrift in the details. Hydrogeology Journal, 19(7), pp.1281-1284.

Voss, C.I. (2011b). Editor's message: Groundwater modeling fantasies—part 2, down to earth. Hydrogeology Journal, 19(8), pp.1455-1458.

Watzlawick, P. (1993). The situation is hopeless, but not serious: The pursuit of unhappiness. WW Norton & Company.

Wang, C., Gomez-Velez, J.D. and Wilson, J.L. (2018). The Importance of Capturing Topographic Features for Modeling Groundwater Flow and Transport in Mountainous Watersheds. Water Resources Research.

Ward, A.S., Gooseff, M.N. and Singha, K. (2013). How does subsurface characterization affect simulations of hyporheic exchange?. Groundwater, 51(1), pp.14-28.

Waterloo Hydrogeologic (2015). Visual MODFLOW, Version 2009.1 Pro, Build: 4.4.0.156. Kitchener, Ontario, Canada.

Weiss, N.A. (2005). A Course in Probability. Addison–Wesley, pp.385–386.

Wels, C., Mackie, D. and Scibek, J. (2012). Guidelines for groundwater modelling to assess impacts of proposed natural resource development activities. Ministry of Environment, Water Protection and Sustainability Branch.

Wilson, J.L. and Liu, J. (1994). Backward tracking to find the source of pollution. Water Manage. Risk Remed, 1, pp.181-199.

Winter, T.C., Harvey J.W., Frank, O.L., and Alley, W.M. (1998). Ground water and surface water: a single resource. U.S. Geological Survey Circular (Vol. 1139).

Woessner, W.W. (2000). Stream and fluvial plain ground water interactions: rescaling hydrogeologic thought. Groundwater, 38(3), 423-429.

Wöhling, T., Geiges, A., and Nowak, W. (2016). Optimal Design of Multitype Groundwater Monitoring Networks Using Easily Accessible Tools. Groundwater.

Wolfe, T. (1979). The Right Stuff. New York, NY: Farrar, Straus and Giroux.

Wondzell, S.M., LaNier, J. and Haggerty, R. 2009. Evaluation of alternative groundwater flow models for simulating hyporheic exchange in a small mountain stream. Journal of Hydrology, 364(1-2), pp.142-151.

Wondzell, S.M. and Gooseff, M.N. (2013). 9.13 Geomorphic controls on hyporheic exchange across scales: Watersheds to particles. Treatise on geomorphology, pp.203-218.

Wright, S. (1921). Correlation and causation. Journal of agricultural research, 20(7), pp.557-585.

Wu, H. (2017). Sensitivity of simulating hyporheic exchange to river bathymetry. Environmental Engineering Master’s Thesis. University of Stuttgart, Stuttgart, Germany.

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., and Fleckenstein, J.H. (2018). Exploring the dynamics of transit times and subsurface mixing in a small agricultural catchment. Water Resources Research, 54(3), 2317-2335.

Ye, M., Meyer, P.D., and Neuman, S.P. (2008). On model selection criteria in multimodel analysis. Water Resources Research, 44(3), pp.1-12.

Ye, M., Pohlmann, K.F., Chapman, J.B., Pohll, G.M., and Reeves, D.M. (2010). A model-averaging method for assessing groundwater conceptual model uncertainty. Groundwater, 48(5), 716-728.

Zarnetske, J.P., Haggerty, R., Wondzell, S.M. and Baker, M.A. (2011). Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. Journal of Geophysical Research: Biogeosciences, 116(G1).

Zhou, Y., Ritzi Jr, R.W., Soltanian, M.R. and Dominic, D.F. (2014). The influence of streambed heterogeneity on hyporheic flow in gravelly rivers. Groundwater, 52(2), pp.206-216.

Zlotnik, V.A., Cardenas, M.B., and Toundykov, D. (2011). Effects of multiscale anisotropy on basin and hyporheic groundwater flow. Groundwater, 49(4), pp.576-583.

Appendix A: The Comprehensive Differential Split-Sample Test (Motavita et al. 2019)[§]

Abstract: The choice of data periods for calibrating and evaluating conceptual hydrological models often seems ad-hoc, with no objective guidance on choosing calibration periods that produce the most reliable predictions. We therefore propose to systematically investigate the effects of calibration and validation data choices on parameter identification and predictive performance. We demonstrate our analysis on the Deggendorf/Kollbach catchment in Bavaria, Germany, for its long series of continuous hydrological and meteorological records. After classifying these data into three hydrological conditions (wet, dry and mixed) and combining them into periods of varied data length (2, 4, 8, 15, 20 and 25 years), we repeatedly calibrate a conceptual rainfall runoff hydrological model - Hydrologiska Byråns Vattenbalansavdelning (HBV) to these distinct data sets via Bayesian updating in a Monte Carlo setting. Then, we analyze predictive performance and posterior parameter statistics in various validation periods of distinct hydrological condition and time-series length. We call this the Comprehensive Difference Split-Sample Test (CDSST). Our results suggest that hydrological conditions in calibration tend to have a stronger impact than time-series length, and that calibrating on dry conditions might be a robust choice when aiming at predicting arbitrary future conditions (wet, dry or mixed). Furthermore, we found that posterior parameter estimates converged to a common optimum range with increasing data size under all investigated calibration scenarios, indicating that compensation of model structural errors by parameter fitting is independent of the chosen calibration condition. However, calibrating on time-series 8 years or longer led to overconfident predictions that failed to reliably envelope future data. While these findings are specific to our case study, we recommend using the CDSST to stress-test conceptual hydrological models to identify robust model parameters and/or deficiencies in the model structure. In general, we expect our proposed approach to be a valuable basis for model error

---

[§] The abstract is from 'Motavita, D.F., Chow, R., Guthke, A., and Nowak, W. (2019). The Comprehensive Differential Split-Sample Test: A stress-test for hydrological model robustness under climate variability. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2019.03.054'

Motavita, D.F.: Wrote preliminary works and original draft manuscript. Revised manuscript to address peer-review comments. Reduced and analyzed data. Analyzed and evaluated all model results. Finalized and approved published article. Conceptualized modelling study. Preliminary results presented in University of Stuttgart Master's thesis (Motavita 2016). ***Chow, R.***: Helped write original draft manuscript. Helped revise article to address peer-review comments. Helped finalizing article. Conceptualized modelling study. Provided supervision. ***Guthke, A.***: Helped write original draft manuscript. Reviewed Bayesian data analysis methods. Conceptualized modelling study. ***Nowak, W.***: Reviewed original draft manuscript and revisions. Provided supervision.

diagnosis in any type of dynamic environmental system model, because it answers the following three questions: (1) what is the importance of physical processes not explicitly covered by the model? (2) How much overconfidence is present in the model? And (3), what are case-specific recommendations for appropriate calibration and validation setups?

Appendix B: Creating the Steinlach River Test Site virtual reality (benchmark) model

Creation of the virtual reality for the Steinlach River Test site began with the collation of numerous borehole observations (e.g., Doro et al. 2013). The borehole observations were reclassified into five hydrofacies: topsoil, silty sand, sandy gravel, gravel, and sandstone. The major strata boundaries were then modelled using Leapfrog 3D geomodelling software (ARANZ Geo Limited 2017). Note that it was necessary to constrain the Leapfrog 3D interpolation using bank profiles and inferred geological trends.

The impala multiple-point geostatistical (MPS) algorithm (Straubhaar et al. 2011), as with any MPS method, requires training images for prior parameter distributions. The upper topsoil unit was represented as a uniform layer and therefore did not require a training image. The lower fluvial unit was modelled using the SBED reservoir simulation software (Geomodeling Technology Corp. 1996) to create a parameter field containing sinusoidal hydrofacies with concave-upward lower contact surfaces. The sinuosity was derived from analysis of the Steinlach River channel morphology, as well as a site conceptual model informed by geophysical and hydrogeological methods (Doro et al. 2013). The training image is shown in Figure S1. The impala algorithm was run with the default input parameters.

Once the hydrofacies distributions were simulated using impala, heterogeneity within the hydrofacies was simulated using spectral methods (Dietrich and Newsam 1993) with the parameters outlined in the article. The simulated heterogeneity was then transferred from the regular simulation grid to the irregular FEM model grid using a nearest-neighbor interpolation. Pre- and post-processing of parameter fields was conducted using the Python programming language.

Readers interested in developing parameter fields for their own sites are recommended to use the openly available HyVR simulation algorithm (Bennett et al. 2018) instead of the commercially available SBED reservoir simulation software (Geomodeling Technology Corp. 1996). The HyVR package can simulate training images suitable for MPS methods, as well as horizontal major strata. The HyVR package can also generate unconditional hydraulic parameter fields with multiple scales of heterogeneity that have simulation outputs that are suitable for groundwater flow and solute transport modelling. For more complex geomodelling of major strata contacts, the open-source python module GemPy (de la Varga et al., 2019) may be of interest.

References

Bennett, J. P., Haslauer, C. P., Ross, M., & Cirpka, O. A. (2018). An open, object-based framework for generating anisotropy in sedimentary subsurface models. Groundwater. https://doi.org/10.1111/gwat.12803

de la Varga, M., Schaaf, A., and Wellmann, F. (2019): GemPy 1.0: open-source stochastic geological modeling and inversion, Geosci. Model Dev., 12, 1-32, https://doi.org/10.5194/gmd-12-1-2019.

*Chow_et_al_2019_impala_files_readme.txt

IMPALA (Straubhaar et al. 2011) input data files to create subsurface parameterization of the Steinlach River Test Site virtual reality (benchmark) model presented in Chow et al. (2018 and 2019).

Notes
-----
* Impala Version 2.1 / build number 20160629 was used for the multiple-point geostatistical simulations. Please refer to the impala user guide for instructions on how to run this file from the command line.
* In general, input/output parameter arrays are saved in VTK format. Data files of this type can be visualised with the openly available Paraview software (https://www.paraview.org/).

Files
-----
* st_loop.in - impala input parameter file
* topsoil.vtk - topsoil training image (uniform hydrofacies)
* fluvial12.vtk - fluvial training image (generated using SBED reservoir modelling software)
* zones_noPC.vtk - zones in which each training image is to be used
* mask.vtk - defines in which grid locations properties should be simulated (no simulation at model top/bottom)
* imp_logs.txt - borehole logs used for conditioning of hydrofacies, derived from site investigations
* imp004_xxxxx.vtk - individual impala realisations (MPS outputs)