



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**

---

Fakultät Umweltwissenschaften

---

## **Reactive transport modeling at hillslope scale with high performance computing methods**

Dissertation zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

**M.Sc. Geoökologie Wenkui He**

geboren am 19. November 1987 in Lijiang, Yunnan

Gutachter:

Herr Prof. Dr.-Ing. Olaf Kolditz, Technische Universität Dresden

Herr Junior-Prof. Dr. Haibing Shao, TU Bergakademie Freiberg

Herr Prof. Dr. Tianfu Xu, Jilin Universität

Tag der Verteidigung: 7. November 2016

## Erklärung des Promovenden

Die Übereinstimmung dieses Exemplars mit dem Original der Dissertation zum Thema:

**„Reactive transport modeling at hillslope scale with high performance computing methods“**

wird hiermit bestätigt.

.....

Ort, Datum

.....

Unterschrift (Vorname Name)

---

## ABSTRACT

---

Reactive transport modeling is an important approach to understand water dynamics, mass transport and biogeochemical processes from the hillslope to the catchment scale. It has a wide range of applications in the fields of e.g. water resource management, contaminated site remediation and geotechnical engineering.

To simulate reactive transport processes at a hillslope or larger scales is a challenging task, which involves interactions of complex physical and biogeochemical processes, huge computational expenses as well as difficulties in numerical precision and stability. The primary goal of the work is to develop a practical, accurate and efficient tool to facilitate the simulation techniques for reactive transport problems towards hillslope or larger scales.

The first part of the work deals with the simulation of water flow in saturated and unsaturated porous media. The capability and accuracy of different numerical approaches were analyzed and compared by using benchmark tests.

The second part of the work introduces the coupling of the scientific software packages OpenGeoSys and IPhreeqc by using a character-string-based interface. The accuracy and computational efficiency of the coupled tool were discussed based on three benchmarks. It shows that OGS#IPhreeqc provides sufficient numerical accuracy to simulate reactive transport problems for both equilibrium and kinetic reactions in variably saturated porous media.

The third part of the work describes the algorithm of a parallelization scheme using MPI (Message Passing Interface) grouping concept, which enables a flexible allocation of computational resources for calculating geochemical reaction and the physical processes such as groundwater flow and transport. The parallel performance of the approach was tested by three examples. It shows that the new approach has more advantages than the conventional ones for the calculation of geochemically-dominated problems, especially when only limited benefit can be obtained through parallelization for solving flow or solute transport. The comparison between the character-string-based and the file-based coupling shows, that the former approach produces less computational overhead in a distributed-memory system such as a computing cluster.

The last part of the work shows the application of OGS#IPhreeqc

for the simulation of the water dynamic and denitrification process in the groundwater aquifer of a study site in Northern Germany. It demonstrates that OGS#IPhreeqc is able to simulate heterogeneous reactive transport problems at a hillslope scale within an acceptable time span. The model results shows the importance of functional zones for natural attenuation process.

---

## KURZFASSUNG

---

Modellierung des reaktiven Stofftransports ist ein wichtiger Ansatz um die Wasserströmung, den Stofftransport und die biogeochemischen Prozesse von der Hang- bis zur Einzugsgebietsskala zu verstehen. Es gibt umfangreiche Anwendungsgebiete, z.B. in der Wasserwirtschaft, Umweltsanierung und Geotechnik.

Die Simulation der reaktiven Stofftransportprozesse auf der Hangskala oder auf größeren Maßstäbe ist eine anspruchsvolle Aufgabe, da es sich um die Wechselwirkungen komplexer physikalischer und biogeochemischer Prozesse handelt, die riesigen Berechnungsaufwand sowie numerischen Schwierigkeiten bezogen auf die Genauigkeit und die Stabilität nach sich ziehen. Das Hauptziel dieser Arbeit besteht darin, ein praktisches, genaues und effizientes Werkzeug zu entwickeln, um die Simulationstechnik für reaktiven Stofftransport auf der Hangskala und auf größeren Skalen zu verbessern.

Der erste Teil der Arbeit behandelt die Simulation der Wasserströmung in gesättigten und ungesättigten porösen Medien. Das Anwendungspotential und die Genauigkeit verschiedener numerischer Ansätze wurden mittels einiger Benchmarks analysiert und miteinander verglichen.

Der zweite Teil der Arbeit stellt die Kopplung der wissenschaftlichen Softwarepakete OpenGeoSys und IPhreeqc mit einer stringbasierten Schnittstelle dar. Die Genauigkeit und die Recheneffizienz des gekoppelten Tools OGS#IPhreeqc wurden basierend auf drei Benchmark-Tests diskutiert. Das Ergebnis zeigt, dass OGS#IPhreeqc die ausreichende numerische Genauigkeit für die Simulation reaktiven Stofftransports liefert, welcher sich sowohl auf die Gleichgewichtsreaktion als auch auf die kinetische Reaktion in variabel gesättigten porösen Medien beziehen.

Der dritte Teil der Arbeit beschreibt zuerst den Algorithmus der Parallelisierung des OGS#IPhreeqc basierend auf dem MPI (Message Passing Interface) Gruppierungskonzept, welcher eine flexible Verteilung der Rechenressourcen für die Berechnung der geochemischen Reaktion und der physikalischen Prozesse wie z.B. Wasserströmung oder Stofftransport ermöglicht. Danach wurde die Leistungsfähigkeit des Algorithmus anhand von drei Beispielen getestet. Es zeigt sich, dass der neue Ansatz Vorteile gegenüber die konventionellen Ansätzen für die Berechnung von geochemisch

dominierten Problemen bringt. Dies ist vor allem dann der Fall, wenn nur eingeschränkter Nutzen aus der Parallelisierung für die Berechnung der Wasserströmung oder des Stofftransportes gezogen werden kann. Der Vergleich zwischen der string- und der dateibasierten Kopplung zeigt, dass die erstere weniger Rechenoverhead in einem verteilten Rechnersystem, wie z.B. Cluster erzeugt.

Der letzte Teil der Arbeit zeigt die Anwendung von OGS#IPhreeqc für die Simulation der Wasserdynamik und der Denitrifikation im Grundwasserleiter eines Untersuchungsgebietes in NordDeutschland. Es beweist, dass OGS#IPhreeqc in der Lage ist, reaktiven Stofftransport auf der Hangskala innerhalb akzeptabler Zeitspanne zu simulieren. Die Simulationsergebnisse zeigen die Bedeutung der funktionalen Zonen für die natürlichen Selbstreinigungsprozesse.

---

## ACKNOWLEDGEMENTS

---

There are so many people to thank during my PhD. Without their helps I could not have achieved my goals in the PhD work, which require a broad spectrum of knowledge and expertise as well as teamwork.

I would like to express my sincere thanks to all my supervisors Prof. Dr.-Ing. Olaf Kolditz (Helmholtz-Centre for Environmental Research-UFZ, department of Environmental Informatics, ENVINF), Dr. Jan H. Fleckenstein (UFZ, Department of Hydrogeology), Prof. Dr. Rudolf Liedl (Technische Universität Dresden, Institute for Groundwater Management, IGW) and Dr. Thomas Kalbacher (UFZ, ENVINF) for their continuous encouragements, trusts and supervisions through out my PhD.

My special thanks goes to Junior Professor Dr. Haibing Shao, who helped me to move from Braunschweig to Leipzig and gave me all kinds of practical suggestions in my work regarding to reactive transport modeling. My deep gratitude also goes to Dr. Christof Beyer (Christian-Albrechts-Universität zu Kiel, Institute of Geosciences), who always gave me constructive and detailed suggestions for both code development and publishing my works. I would like to express my sincere thanks to all of my colleagues from the diverse team of ENVINF: Dr. Wenqing Wang, who provided me numerous helps for solving problems during running simulations and code implementation; Dr. Dmitri Naumov, who continuously helps me to improve my skills in debugging during my works; Dr. Thomas Fischer, who gave me lots of suggestions for parallel programming; Junior Professor Dr. Marc Walther, who always explained me the model setups in OGS clearly and patiently; Leslie Jacobs, who helped me to improve English language for my publications; Dr. Karsten Rink, who aided me with meshing; Lars Bilke, who provided technical supports for software engineering; Agnes Sachse, who gave me detailed suggestions on my PhD thesis. I also thank my fellow PhD students Eunseon Jang, Erik Nixdorf, Haiyang Yi, Yonghui Huang, Tianyuan Zheng, Renchao Lu, Christof Lehmann, Xing-Yuan Miao and Miao Jing, with whom I have not only worked with, but also have had great times during my PhD.

I'm also grateful to the HIGRADE graduate school of the UFZ for providing me a variety of insightful courses and career events

(e.g. CareerCafé). My deep gratitude also goes to Barbara Timmel (UFZ, International Office) for her immense help regarding to my visa issues and organization of many meaningful intercultural events.

Last but not the least, I would like to thank my girlfriend for her trusts and encouragements during my PhD and my family for their love and supports through my study and research in Germany.



---

# CONTENTS

---

Contents	vii
List of Figures	ix
List of Tables	xi
1 INTRODUCTION	1
1.1 Motivation	1
1.2 Previous research	1
1.3 Objective and scope	2
1.4 Dissertation structure	3
2 THEORY AND METHODS	4
2.1 Flow in saturated and unsaturated zone	4
2.1.1 Water flow in aquifer system	4
2.1.2 Water flow in unsaturated soils	5
2.2 Solute transport	6
2.2.1 Fundamental processes	6
2.2.2 Advection-dispersion equation	8
2.3 Biogeochemical modelling	9
2.3.1 Equilibrium reaction	10
2.3.2 Kinetic reaction	11
2.4 Approach for reactive transport modeling	12
3 NUMERICAL SIMULATION OF WATER FLOW IN SATURATED AND UNSATURATED ZONES	15
3.1 Introduction	15
3.2 This problem	15
3.2.1 Benchmark description	17
3.2.2 Results	18
3.3 Numerical simulation of unsaturated water flow	18
3.3.1 Mixed-form of Richards equation revisited	18
3.3.2 Benchmark analysis	21
3.4 Summary	26
4 IMPLEMENTATION AND ANALYSIS OF THE COUPLING INTERFACE OGS#IPHREEQC	27
4.1 Introduction	27
4.2 OpenGeoSys	27
4.3 PHREEQC and IPhreeqc module	28
4.4 Implementation of the coupling interface	30
4.5 Benchmarking	32
4.5.1 Isotope fractionation	32
4.5.2 Engesgaard benchmark	35

## CONTENTS

4.5.3	Uranium leaching	36
4.6	Summary	39
5	PARALLELIZATION AND PERFORMANCE ANALYSIS	42
5.1	Introduction	42
5.2	Methodology	42
5.3	Implementation	44
5.4	Computational platform	46
5.5	Parallel performance tests of the string-based coupling	48
5.5.1	Isotope fractionation, 2-D	48
5.5.2	Isotope fractionation, 3-D	50
5.5.3	Uranium leaching benchmark	53
5.6	Comparison of the string-based and file-based approach	54
5.7	Summary	55
6	APPLICATION OF THE TOOL FOR A CASE STUDY	58
6.1	Study site	58
6.2	Model setup	58
6.3	Results and discussion	61
6.4	Summary	63
7	CONCLUSIONS AND OUTLOOK	64
7.1	Conclusions	64
7.2	Outlook	65
	BIBLIOGRAPHY	67
	LIST OF PUBLICATIONS	74
	Appendices	75
A	OGS#IPHREEQC INPUT FILE DESCRIPTION	76
B	INSTRUCTION FOR OGS#IPHREEQC COMPILATION AND EXECUTION	87

---

## LIST OF FIGURES

---

Figure 1	Hydraulic drawdown in an aquifer as a result of pumping activity.	16
Figure 2	The employed mesh for Theis 3-D.	18
Figure 3	Calculated hydraulic head drawdowns at a distance of 9.639 m from the well.	19
Figure 4	Distribution of volumetric water content in two scenarios by using three different algorithms.	23
Figure 5	Initial and boundary conditions of the benchmark example (from He et al. 2015d).	24
Figure 6	Comparison of the simulation results of different approaches together with experiment data.	25
Figure 7	The general idea of the coupling procedure between OGS and IPhreeqc (from He et al. 2015b).	31
Figure 8	Concentration profiles of the light CHC isotopologues and isotope signatures along the model domain.	34
Figure 9	The concentration profiles of different minerals and aqueous species simulated by OGS#IPhreeqc, PHREEQC and OGS-ChemApp (from He et al. 2015b).	36
Figure 10	Uranium leaching from a mill tailing.	37
Figure 11	Geometry and boundary conditions of the uranium leaching example (from supplementary material of He et al. 2015b).	38
Figure 12	Time evolution of uranium concentration profile after 200 (a), 500 (b) and 1000 (c) days (from supplementary material of He et al. 2015b).	39
Figure 13	Distribution of pH, calcite ( $mol \cdot kgw^{-1}$ ) and gypsum ( $mol \cdot kgw^{-1}$ ) after 1000 days (from supplementary material of He et al. 2015b).	40
Figure 14	comparison of the scalability of the conventional and the new parallelization scheme in different scenarios.	43

## LIST OF FIGURES

Figure 15	The concept of MPI grouping and communication of the parallelization scheme (modified from He et al. 2015b).	45
Figure 16	Collection of the concentration values from the local buffers to a global concentration vector using <i>MPI_Allreduce</i> operation.	46
Figure 17	Pseudo code for the presentation of the tasks in both MPI groups (modified from He et al. 2015b).	47
Figure 18	Comparison of the results using serial and parallel simulation for isotope fractionation 2-D example.	49
Figure 19	Performance of the parallelization scheme for simulation of isotope fractionation 2-D example on ENVINF (from He et al. 2015b).	51
Figure 20	The parallel efficiency of flow/transport and IPhreeqc calculation as a function of sub-domains	52
Figure 21	Performance of the parallelization scheme for the simulation of isotope fractionation 3-D example on EVE (from He et al. 2015b).	52
Figure 22	Relative speedup of the simulation of the uranium leaching example as a function of number of DDCs and compute cores.	54
Figure 23	Breakdown of the total time corresponding to curve AB in Fig. 22 (from He et al. 2015b).	55
Figure 24	Comparison of the parallel performance of the string-based and file-based coupling interface.	56
Figure 25	Mesh generation for the Selke "book" model.	59
Figure 26	The permeability distribution of the model domain.	60
Figure 27	The geometry and boundary condition.	60
Figure 28	The distribution of first-order denitrification rate (larger than $4.2E-8 \text{ s}^{-1}$ ) in the model domain.	61
Figure 29	The velocity distribution of groundwater flow in the model domain at steady state.	62
Figure 30	The concentration contour maps of nitrate after 578 days.	62

---

LIST OF TABLES

---

Table 1	The numerical capabilities of different RTM codes (from Steefel et al. 2015). 14	
Table 2	Parameters and their values applied for Theis problem (from He et al. 2015c). 17	
Table 3	Summary of the model concepts for Theis' problem (from He et al. 2015c). 17	
Table 4	Soil hydraulic parameters of Berino loamy fine sand and Glendale clay loam (from Hills et al. 1989). 22	
Table 5	Model setups and mass balance errors of different algorithms for a 5-day simulation. 23	
Table 6	Soil hydraulic parameters based on the data from Vauclin et al. 1979 and Clement et al. 1994. 24	
Table 7	Mass balance errors of different algorithms for the water table benchmark. 25	
Table 8	Comparison of the key geochemical features of different RTM codes (from Steefel et al. 2015). 29	
Table 9	The first-order degradation rate and enrichment factor for CHCs (from van Breukelen et al. 2005) 33	
Table 10	A comparison of different portions of the simulation time for the van Breukelen benchmark by using OGS#IPhreeqc, KinReact and PHREEQC (in seconds) (from He et al. 2015b). 34	
Table 11	Material properties of the 1-D calcite column (from He et al. 2015b). 35	
Table 12	Initial and boundary conditions for the Engesgaard benchmark (from He et al. 2015b). 35	
Table 13	A comparison of different portions of the simulation time for the Engesgaard benchmark by using OGS#IPhreeqc, PHREEQC and OGS-ChemApp (in seconds) (from He et al. 2015b). 36	

## LIST OF TABLES

Table 14	Soil hydraulic parameters of the hillslope (from Šimůnek et al. 2012).	38
Table 15	Initial and boundary setups for geochemical system of uranium leaching example (from supplementary of He et al. 2015b).	38
Table 16	Input files and descriptions for the isotope fractionation benchmark.	77

---

## ACRONYMS

---

- ADE** Advection-Dispersion Equation
- CFL condition** Courant–Friedrichs–Lewy condition
- DEM** Digital Elevation Model
- DNF** Denitrification
- DDC** Domain Decomposition
- ER** Electrical Resistivity
- IBVP** Initial-Boundary Value Problem
- MPI** Message Passing Interface
- ODE** Ordinary Differential Equation
- OGS** OpenGeoSys
- REV** Representative Elementary Volume
- RTM** Reactive Transport Modeling
- SIA** Sequential Iterative Approach
- SNIA** Sequential Non-Iterative Approach

---

## INTRODUCTION

---

### 1.1 MOTIVATION

The hillslope is the fundamental element of the catchment. To understand water dynamics, mass transport and biogeochemical processes at such a scale is essential for assessing the environmental impacts of anthropogenic activities such as agricultural land use or mining on water resources.

Reactive transport modeling (RTM) is an important approach to get a profound understanding of these processes and make plausible predictions or assessments for various applications. In the fields of contamination remediation or water resource management, RTM has been applied to predict the fate of environmental chemicals e.g. uranium, pesticide in soils or groundwater reservoirs (e.g. Hammond and Lichtner 2010; Henzler et al. 2014; Molins et al. 2010; Yabusaki et al. 2011). Additionally, RTM has also been applied in geotechnical applications such as risk assessment of nuclear waste disposals in geological repositories (e.g. Kosakowski and Watanabe 2014; Shao et al. 2009) or evaluation of geological sequestration of carbon dioxide (e.g. Beyer et al. 2012; Xu et al. 2004; Xu et al. 2006).

The development of reliable and efficient modeling tools is the foundation for the successful application of RTM in various fields, especially when simulations at the hillslope or larger scales have to be performed.

### 1.2 PREVIOUS RESEARCH

In the past decades, many software packages have been developed for RTM, for example PHREEQC (Parkhurst and Appelo 1999; Parkhurst and Appelo 2013), OpenGeoSys (OGS) (Kolditz et al. 2012), HYTEC (van der Lee et al. 2003), ORCHESTRA (Meeussen 2003), TOUGHREACT (Xu et al. 2011), eSTOMP (Yabusaki et al. 2011), HYDROGEOCHEM (Yeh and Tripathi 1990), CrunchFlow (Steeffel et al. 2015), MIN3P (Mayer et al.



2002), PFLOTRAN (Lichtner et al. 2015), and so on.

In the meantime, code coupling has often been applied as a straightforward approach to combine the advantages of different software. Just to name a few: PHREEQC and HYDRUS (Šimůnek et al. 2006; Šimůnek et al. 2012), PHREEQC and COMSOL Multiphysics (Wissmeier and Barry 2011; Nardi et al. 2014; Nasir et al. 2014); PHREEQC and MT3DMS (Morway et al. 2013); OGS and GEMs (Kosakowski and Watanabe 2014); OGS and BRNS (Centler et al. 2010); MODFLOW-UZF and RT3D (Bailey et al. 2013).

RTM at a large scale is often accompanied with big computational expenses, which makes model calibrations and parameter sensitivity analysis unrealistic, especially when these kinds of tasks have to be completed within a limited amount of time. An established approach to treat this challenge is to use parallel computing, which becomes more and more attractive with the continuous innovation of software and hardware infrastructures for high performance computing (HPC). Existing parallelized codes for RTM are for example, PFLOTRAN (Hammond et al. 2014), TOUGH-MP (Hubschwerlen et al. 2012), eSTOMP and OGS-GEMs (Kosakowski and Watanabe 2014). Apart from eSTOMP, which uses a one-sided communication and global shared-memory programming paradigm (Yabusaki et al. 2011), the other codes mentioned above apply domain decomposition (DDC) approach for their code parallelization.

### 1.3 OBJECTIVE AND SCOPE

Despite plenties of efforts, development of practical tools for RTM at the hillslope scale still remains a demanding task. The major challenges are listed below:

- ◆ RTM often involves a variety of hydrological and biogeochemical processes that interact with each other in a complex manner. This requires RTM simulators to have comprehensive modeling capabilities to solve coupled problems;
- ◆ RTM often requires the simulation of non-linear problems such as unsaturated flow, which poses big challenges in numerical accuracy and stability;
- ◆ Computational efficiency is another vital issue for RTM at large scales. Despite the use of HPC platforms, to improve the parallel performance of a code is still a challenging task. For a coupled simulator, the efficiency of the coupling interface can become a crucial factor influencing the scalability of the parallelized code

## 1.4 DISSERTATION STRUCTURE

when large number of compute cores are employed.

The current work tries to tackle these challenges, thus to facilitate the techniques for RTM at a hillslope or even catchment scale.

✧ The capabilities and accuracy of different numerical approaches, which can be used to solve water flow in coupled soil-aquifer system, were analyzed and compared with benchmarking tests;

✧ The scientific software packages OGS and IPhreeqc were coupled at the code level to enable the setup and simulation of a variety of multi-physical problems with a wide range of chemical reactions that are known to influence the water quality in porous media;

✧ A new parallelization scheme was developed for the coupled tool, which realizes a flexible allocation of computational resources for different computing tasks.

## 1.4 DISSERTATION STRUCTURE

The dissertation is organized as follows. Chapter 2 presents the fundamental theories and governing equations for RTM. Chapter 3 analyzes the capability and accuracy of different numerical approaches for simulating water flow in saturated and unsaturated porous media. Chapter 4 introduces the implementation and benchmarking of the coupling interface between OGS and IPhreeqc. Chapter 5 presents the idea, the implementation and performance tests of a new parallelization scheme. In chapter 6, the parallelized tool is applied for a case study at the hillslope scale. The summary and outlook of the work are given at the end.

---

## THEORY AND METHODS

---

Modeling reactive transport in the subsurface environment involves the coupling of multiple processes, which mainly consists of water flow in saturated and unsaturated zones, solute transport and biogeochemical reactions. Each processes can be defined as individual initial-boundary-value problems (IBVPs). In this chapter, the governing equations of each problem will be presented separately, followed by the introduction of the coupling methods.

### 2.1 FLOW IN SATURATED AND UNSATURATED ZONE

Soil and aquifer bodies are consist of porous media. In order to quantify the water motion in the porous media at a macroscopic level, a continuum approach called representative elementary volume (REV) is applied, which is universally applicable for porous media and guarantees the average property values to be measurable and representative for the selected volume (Bear and Verruijt 1987). This approach is employed for the mathematical models simulating water flow in soil and aquifer systems, which are described below.

#### 2.1.1 *Water flow in aquifer system*

Darcy's law (Darcy 1856) can be applied to describe the water motion in the saturated soil medium, which says the flow rate of water through a porous medium is proportional to the piezometric head gradient and the area of cross section, as given by Eq. 1.

$$Q = -KA \frac{\Delta h}{L} \quad (1)$$

where  $Q$  [ $m^3 \cdot s^{-1}$ ] is the flow rate;  $A$  [ $m^2$ ] is the cross-sectional area;  $K$  [ $m \cdot s^{-1}$ ] is the hydraulic conductivity of the porous medium;  $\Delta h$  is the difference of piezometric head;  $L$  [ $m$ ] is the length of soil column. Darcy's law is valid for laminar flow with a Reynolds number between 1 and 10, which is applicable for most of natural

groundwater flows (Bear and Verruijt 1987).

Based on the equation of mass balance of saturated flow and Darcy's law, the general equation for groundwater flow in a confined aquifer is given in Eq. 2

$$S_s \frac{\partial h}{\partial t} = \nabla(K \cdot \nabla h) + q \quad (2)$$

where  $S_s$  [ $m^{-1}$ ] is the specific storage, which is the addition of volume of water to the storage for a unit volume of porous medium as the increase of one unit of piezometric head;  $K$  is the hydraulic conductivity tensor,  $q$  is the source/sink term.

For liquid flow processes, in which pressure is the primary variable, Eq. 2 can be rewritten into Eq. 3.

$$\frac{S_s}{\rho g} \frac{\partial p}{\partial t} = \nabla \left( \frac{k}{\mu} (\nabla p + \rho g \nabla z) \right) + q \quad (3)$$

where  $p$  [ $Pa$ ] is the pressure;  $\rho$  [ $kg \cdot m^3$ ] and  $\mu$  [ $Pa \cdot s$ ] are the fluid density and viscosity, respectively;  $g$  [ $m \cdot s^{-2}$ ] is the constant of gravity acceleration;  $k$  [ $m^2$ ] is the permeability of the porous medium;  $z$  [ $m$ ] is the elevation.

### 2.1.2 Water flow in unsaturated soils

The Richards equation is a combination of the mass conservation equation and the Darcy-Buckingham equation. It has been widely used in simulating water flow processes in unsaturated soil or coupled soil-aquifer systems. In the Richards equation there are two primary variables i.e. the volumetric water content and the pressure head. The water content serves as the mass accumulation variable, whereas the pressure head is the driving force variable by introducing capillary pressure gradient and gravity (Krabbenhøft 2007). Additionally, for unsaturated soil the hydraulic conductivity is a function of the soil moisture (or water content). Hence, additional equations are required to quantify the relationship between water content and pressure head (known as soil retention curve) as well as hydraulic conductivity and water content. Empirical equations such as Van Genuchten (Van Genuchten 1980) and Brooks Corey (Brooks and Corey 1966) are often employed for this purpose.

The Richards equation has been expressed in three standard forms: i) the pressure head-based form (h-based form) (see Eq. 4) with pressure head as primary variable, ii) the saturation-based form (see Eq. 5) with saturation as primary variable, and iii) the mixed

## 2.2 SOLUTE TRANSPORT

form (see Eq. 6), in which either pressure or saturation can be chosen as primary variable.

$$C(h_p) \frac{\partial h_p}{\partial t} = \nabla(K(h_p) \cdot \nabla h_p) + \frac{\partial K(h_p)}{\partial z} \quad (4)$$

$$\frac{\partial \theta}{\partial t} = \nabla(D(\theta) \cdot \nabla \theta) + \frac{\partial K(\theta)}{\partial z} \quad (5)$$

$$\frac{\partial \theta}{\partial t} = \nabla(K(h_p) \cdot \nabla h_p) + \frac{\partial K(h_p)}{\partial z} \quad (6)$$

In Eq. 4 to Eq. 6,  $h_p$  [m] is pressure head;  $\theta$  [ $m^3 \cdot m^{-3}$ ] is the volumetric water content;  $K(h_p)$  and  $K(\theta)$  are the capillary conductivity functions;  $C(h_p) = \frac{d\theta}{dh_p}$  is the specific moisture capacity;  $D(\theta) = \frac{K(\theta)}{C(\theta)}$  is the unsaturated diffusivity;  $z$  and  $t$  are vertical coordinate (positive upward) and time, respectively.

Apart from the saturation-based form, the h-based and the mixed form which uses the pressure head as the primary variable (Eq. 6) can be applied in the unsaturated and saturated zone simultaneously (Clement et al. 1994, Kavetski et al. 2001, Krabbenhøft 2007). In the saturated zone, where the volumetric water content in the porous medium stays constant, the specific moisture capacity and the time derivative of water content i.e.  $\frac{\partial \theta}{\partial t}$  converges to zero. In this case, both Eq. 4 and Eq. 6 converge to the general groundwater flow equation.

## 2.2 SOLUTE TRANSPORT

In the subsurface the transport and retardation of different chemical compounds can be described by processes such as advection, dispersion, adsorption, and so on. These processes will be influenced by factors such as flow field, physical-chemical characteristics of the compounds, as well as the properties of the fluid and the porous media through which the flow and solute transport occur. These phenomena can cause the concentration variations of the solutes, thus affect the fluid properties such as density and viscosity, which can become especially significant in certain situations like saline water intrusion in coastal areas.

### 2.2.1 *Fundamental processes*

Advection describes the solute movement due to the motion of fluid. Here, the solute flux is in accordance with the average ve-

locity (including direction) of the fluid, which can be calculated by Darcy's law in the porous media.

Molecular diffusion is the spreading (mixing) of solutes in the fluid as a result of the random walk of molecules, which can produce a solute flux in response to its concentration gradient in the fluid. Based on Fick's law, Bear and Bachmat 1986 derived the equation for diffusion flux at a macroscopic level (see Eq. 7).

$$\mathbf{J}_{dif} = -\mathbf{D}_d^* \cdot \nabla C \quad (7)$$

where  $\mathbf{J}_{dif}$  [ $kg \cdot m^{-2} \cdot s^{-1}$ ] is the flux vector of solute;  $\mathbf{D}_d^*$  [ $m^2 \cdot s^{-1}$ ] is the coefficient of molecular diffusion;  $\nabla c$  is the concentration gradient of solute; the minus sign in the right hand side of the equation means the flux is directed from higher concentrations towards lower ones.

Mechanical dispersion describes the spreading of solute as a result of the variation of fluid velocity at the microscopic level (pore scale). Similar like the diffusion flux, the dispersive flux of the solute can be described by a Fickian type law as Eq. 8.

$$\mathbf{J}_{dis} = -\mathbf{D} \cdot \nabla C \quad (8)$$

$\mathbf{D}$  is the coefficient of mechanical dispersion with

$$\mathbf{D} = [D_{ij}] = \begin{bmatrix} \alpha_L v & 0 & 0 \\ 0 & \alpha_T v & 0 \\ 0 & 0 & \alpha_T v \end{bmatrix} \quad (9)$$

where  $\alpha_L$  and  $\alpha_T$  are longitudinal and transverse dispersivity, respectively.

The term hydrodynamic dispersion is used to denote the spreading phenomenon (both longitudinal and transverse to the stream lines) due to both molecular diffusion and mechanical dispersion. Combining diffusive and dispersive flux gives

$$\mathbf{J}_{dif} + \mathbf{J}_{dis} = -(\mathbf{D}_d^* + \mathbf{D}) \cdot \nabla C = -\mathbf{D}_h \cdot \nabla C \quad (10)$$

where  $\mathbf{D}_h$  is the coefficient of hydrodynamic dispersion.

Besides the mechanisms introduced above, several other processes such as sorption, decay and biogeochemical reaction can also influence the concentration distribution and evolution of chemicals in the aqueous and solid phase of porous media.

According to Appelo and Postma 2005, the term sorption denotes the physical-chemical interactions between soil (aquifer) materials and chemicals, which include processes such as adsorption, absorption and ion exchange. Adsorption involves the adherence of a

chemical species to the solid surface, absorption means the uptake of a chemical into the solid phase, whereas ion exchange refers to the replacement of one ion for another at the surface of the solid. An adsorption isotherm describes the relationship between the quantity of a substance that is dissolved in the fluid and its quantity that is adsorbed on the solid at a fixed temperature. The adsorption isotherms which are often used include the linear (Eq. 11), Freundlich (Eq. 12) and Langmuir (Eq. 13) isotherms.

$$S = k_d C \quad (11)$$

$$S = k_f C^m \quad (12)$$

$$S = \frac{s_{max} C}{k_l + C} \quad (13)$$

In Eq. 11 to Eq. 13,  $S$  [ $kg \cdot kg^{-1}$ ] is the component concentration on the solid;  $k_d$  is called the distribution coefficient;  $m$ ,  $k_f$ ,  $s_{max}$  and  $k_l$  are constant coefficients.

Decay is related with the degradation of radioactive species. The common decay model is the first-order degradation equation given in Eq. 14.

$$\frac{dC}{dt} = -\lambda \cdot C \quad (14)$$

where  $\lambda$  [ $kg \cdot kg^{-1} \cdot s^{-1}$ ] is the decay rate in the liquid phase.

The chemical reaction is another important source/sink for substance in the solute transport process, which will be introduced in details in Sect. 2.3.

### 2.2.2 Advection-dispersion equation

Based on the equations to quantify the fundamental processes for solute transport (introduced in Sect. 2.2.1) and the mass balance equation, the advection-dispersion equation can be derived as below.

$$\frac{\partial(\theta C)}{\partial t} = \nabla(\theta \mathbf{D}_h \cdot \nabla C) - \nabla(\vec{q} \cdot C) - \frac{\partial(\theta_s \rho_s S)}{\partial t} + \theta_s \rho_s \Lambda_s + \theta \rho \Lambda \quad (15)$$

where  $\vec{q}$  is the darcy velocity vector, which can be calculated by Eq. 2 or Eq. 6;  $\rho_s$  [ $kg \cdot m^{-3}$ ] and  $\rho$  [ $kg \cdot m^{-3}$ ] are solid and liquid density, respectively;  $n$  is the porosity, hence  $1 - n$  denotes the volumetric fraction of solid;  $\Lambda_s$  [ $kg \cdot kg^{-1} \cdot s^{-1}$ ] and  $\Lambda$  [ $kg \cdot kg^{-1} \cdot$

### 2.3 BIOGEOCHEMICAL MODELLING

$s^{-1}$ ] are rates of chemical reactions or decay processes for the interested component in liquid and solid phase, respectively.

Both  $\Lambda_s$  and  $\Lambda$  are functions of concentrations of the interested component (i.e.  $C$  and  $S$ ) as well as other relevant components in the liquid and solid phase.  $C$  and  $S$  can be related with each other by using the sorption isotherms introduced in 2.2.1, thus two variables of Eq. 15 can be reduced to only one.

In saturated zone, assuming only adsorption following the linear adsorption isotherm and first-order decay ( $\rho\Lambda = -\lambda C$  and  $\Lambda_s = -\lambda_s S$ ) take place, Eq. 15 can be rewritten into

$$\frac{\partial(nC)}{\partial t} = \nabla(n\mathbf{D}_h \cdot \nabla C) - \nabla(\vec{q} \cdot C) - \frac{\partial((1-n)\rho_s k_d C)}{\partial t} - (1-n)\rho_s \lambda_s k_d C - n\lambda C \quad (16)$$

where  $\lambda_s$  [ $kg \cdot kg^{-1} \cdot s^{-1}$ ] is the decay rate in the solid phase.

If we further assume that decay rates are equal in both phases and porosity is constant and uniformly distributed for the whole domain, Eq. 16 can be reduced into

$$\frac{\partial C}{\partial t} = \nabla\left(\frac{\mathbf{D}_h}{R} \cdot \nabla C\right) - \nabla\left(\frac{\vec{V}}{R} \cdot C\right) - \lambda C \quad (17)$$

where  $\vec{V}$  is the vector for pore water velocity and  $R$  is the retardation factor, which is given in Eq. 18

$$R = 1 + \frac{(1-n)\rho_s k_d}{n} \quad (18)$$

### 2.3 BIOGEOCHEMICAL MODELLING

The biogeochemical reactions in the subsurface includes mineral dissolution and precipitations, homogeneous reactions (having only one phase e.g. aqueous phase reactions), surface complexation, ion exchange, gas-aqueous phase exchange and microbially-mediated reactions (Stefel et al. 2015). These reactions can often be described by two types of reaction mechanisms. Reactions, which take place fast and evolve to equilibrium states in a "neglectable" time (compared to the residence time of groundwater), can be regarded as equilibrium reactions (Appelo and Postma 2005). In contrary, if a reaction is rather "slow", then its reaction kinetics have to been taken into consideration.



2.3.1 *Equilibrium reaction*

Equilibrium reactions can often be generalized as Eq. 19,



where  $A_j$  and  $A_i$  are master and second species, respectively;  $v_{i,j}$  denotes the stoichiometric coefficient;  $N$  is the total number of master species.

At the equilibrium state, the distribution of secondary and master species can be linked by using a coefficient called equilibrium constant  $K_i$ , which can be obtained by using Eq. 20.

$$K_i = a_i^{-1} \cdot \prod_{j=1}^N a_j^{v_{i,j}} \quad (20)$$

where  $a_i$  and  $a_j$  are the activity of secondary and master species, respectively.

Equation 20 is called the law of mass action, which is the fundamental theory for equilibrium reactions.

The molarity of species  $A_i$  i.e.  $C_i$  can be calculated by Eq. 21.

$$C_i = \frac{a_i}{\gamma_i} = K_i^{-1} \cdot \frac{\prod_{j=1}^N a_j^{v_{i,j}}}{\gamma_i} \quad (21)$$

where  $\gamma_i$  is the activity coefficient, which can be calculated by different approaches such as extended Debye-Hückel equation (Apelo and Postma 2005) (Eq. 22) and Davies equation (Davies 1962) (Eq. 23).

$$\log \gamma_i = -\frac{Az_i^2 \sqrt{I}}{1 + Ba_i \sqrt{I}} + b_i I \quad (22)$$

$$\log \gamma_i = -Az_i^2 \left( \frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3I \right) \quad (23)$$

In Eq. 22 and Eq. 23,  $A$  and  $B$  are constants dependent on temperature;  $z_i$  is the ion charge number;  $a_i$  and  $b_i$  are ion-specific fit parameters;  $I$  is the ionic strength.

The equilibrium constant  $K_i$  in Eq. 21 is temperature dependent. In a standard state (at a pressure of 1 bar and temperature of 25°C),  $K_i$  can be calculated based on the standard Gibbs free energy with equation Eq. 24.

$$K_i = e^{-\frac{\Delta_r G_i^\circ}{RT}} \quad (24)$$

where  $R$  [ $J \cdot K_{-1} \cdot mol_{-1}$ ] is the ideal gas constant;  $T$  [ $K$ ] is temperature in Kelvin;  $\Delta_r G_i^\circ$  is the standard Gibbs free energy for the reaction.

Based on the equilibrium constant at the standard state, those for other temperatures i.e.  $K_i(T)$  can be calculated based on Van't Hoff equation (Eq. 25) or a polynomial expression (Eq. 26).

$$\log K_i(T) = \log K_i(T_0) - \frac{\Delta H}{R \ln 10} \left( \frac{1}{T} - \frac{1}{T_0} \right) \quad (25)$$

where  $K_i(T_0)$  is the equilibrium constant at temperature  $T_0$ ;  $\Delta H$  is the reaction enthalpy.

$$\log K_i(T) = A + B \cdot T + \frac{C}{T} + D \cdot \log T + \frac{E}{T^2} \quad (26)$$

where  $A$ ,  $B$ ,  $C$ ,  $D$  and  $E$  are constants.

### 2.3.2 Kinetic reaction

Reaction kinetics are often considered for reactions such as microbial-driven processes (e.g. denitrification) or mineral dissolution/precipitation.

Monod kinetic rate equation (Eq. 27) is often used to model the biodegradation of organic pollutants.

$$\frac{dC}{dt} = -k_m \frac{C}{k_{1/2} + C} \quad (27)$$

where  $C$  [ $mg \cdot L^{-1}$ ] is the concentration of the pollutant;  $k_m$  [ $mg \cdot L^{-1} \cdot s^{-1}$ ] is the maximal degradation rate which is a function of the bacterial mass concentration and growth rate;  $k_{1/2}$  [ $mg \cdot L^{-1}$ ] is the half-saturation constant.

Eq. 27 can be extended to multiplicative monod equation to include the limiting effects of other chemical species (substrates) on the reaction rate.

$$\frac{dC}{dt} = - \left( \frac{k_m C}{k_{1/2} + C} \right) \left( \frac{k_m^1 C^1}{k_{1/2}^1 + C^1} \right) \dots \quad (28)$$

Mineral dissolution and precipitation can be modeled by using different types of rate laws, which were introduced in great details in Steefel et al. 2015. Here, only the Transition State Theory (TST) based rate expression is introduced, by which the dissolution/precipitation of mineral is regarded as reversible process depending on Gibbs energy and the saturation state.

Based on Lasaga et al. 1994, Beyer et al. 2012 and Steefel et al. 2015, the rate law is given as follows:

$$\frac{dM}{dt} = K A \operatorname{sgn}(1 - \Omega^\theta) |1 - \Omega^\theta|^\eta \quad (29)$$

where  $M$  [ $\text{mol} \cdot \text{m}^{-3}$  porous medium] is the mineral concentration,  $A$  [ $\text{m}^2 \cdot \text{m}^{-3}$  porous medium] is the reactive surface area of mineral,  $\theta$  and  $\eta$  are empirical exponents,  $\operatorname{sgn}(1 - \Omega^\theta)$  is the sign function,  $\Omega$  is the ion activity product divided by equilibrium constant,  $K$  is the overall rate constant, which can include several mechanisms depending on different rate constants  $K_i$ , acid, alkaline or activities of inhibiting species. These mechanisms were described in great details in Palandri and Kharaka 2004. The different rate constants can be calculated for temperatures other than  $25^\circ\text{C}$  based on the Arrhenius equation.

$$K_i(T) = K_i(298.15\text{K}) \exp\left(-\frac{E}{R} \left[\frac{1}{T} - \frac{1}{298.15}\right]\right) \quad (30)$$

where  $E$  [ $\text{J} \cdot \text{mol}^{-1}$ ] is the activation energy,  $R$  [ $\text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ ] is the gas constant.

## 2.4 APPROACH FOR REACTIVE TRANSPORT MODELING

If we specifies the reaction terms in the ADE (Eq. 15) based on Steefel et al. 2015, then we can get the reactive transport equation as follows:

$$\begin{aligned} \frac{\partial(\theta C)}{\partial t} = & \nabla(\theta \mathbf{D}_h \cdot \nabla C) - \nabla(\vec{q} \cdot C) - \frac{\partial(\theta_s \rho_s S)}{\partial t} \\ & - \sum_{i=1}^{Na} v_{i,a} R_a - \sum_{i=1}^{Nm} v_{i,m} R_m - \sum_{i=1}^{Ng} v_{i,g} R_g \end{aligned} \quad (31)$$

where  $R_a$ ,  $R_m$  and  $R_g$  are reaction terms for aqueous phase reactions, mineral reactions and gas reactions, respectively.  $v_{i,a}$ ,  $v_{i,m}$  and  $v_{i,g}$  are corresponding stoichiometric coefficients in each reaction types.

There are generally two categories of approach to solve the reactive transport equations numerically. The first one is the global implicit approach (GIA), in which the transport and reaction terms in Eq. 31 are solved simultaneously; whereas the second approach is the operator splitting approach, in which Eq. 31 is decoupled into a transport step (Eq. 32) and a reaction step (Eq. 33), so that transport and reaction are solved sequentially.

$$\frac{(C_i^{\text{transport}} - C_i^n)}{\Delta t} = L(C_i)^n, (i = 1, \dots, N) \quad (32)$$

$$\frac{(C_i^{n+1} - C_i^{transport})}{\Delta t} = R_i^{n+1}, (i = 1, \dots, N) \quad (33)$$

where  $R_i^{n+1}$  denotes the chemical reactions, while  $L$  is the spatial operator with

$$L(C_i)^n = \left[ \frac{\partial(\theta)}{\partial t} + \nabla \cdot (\vec{q} - \theta \mathbf{D}_h \cdot \nabla) \right] C \quad (34)$$

The GIA is generally regarded as the most stable method (Kräutle and Knabner 2007), which is not restricted by the Courant condition during spatial- and temporal discretization (Steeffel et al. 2015). However, it is usually more computational expensive than the operator splitting approach, especially for multi-component (species) systems (Steeffel et al. 2015). Recently, Kräutle and Knabner 2007 has developed a reduction scheme in order to reduce the size of the coupled equation system, so as to cut down the required computational effort.

The operator splitting method, on the other hand, is generally less expensive to solve for one time step, because chemical reaction is solved locally on each node, after the calculation of transport. However, the Courant condition needs to be fulfilled (usually realized by using smaller time step sizes) in order to minimize the operator splitting error. There are generally two different methods i.e. sequential iterative approach (SIA) and sequential non-iterative approach (SNIA). In the former approach, a iteration between solving transport and reaction will be performed until the convergence is obtained, whereas in the later approach transport and reaction are solved only once during each time step. In order to reduce the operator splitting error, adequate small time step sizes are required if the SNIA is applied.

Tab. 1 (Steeffel et al. 2015) summarizes the key numerical capabilities of several RTM codes, in which the available coupled processes and numerical schemes of these codes are compared in details.

Table 1.: The numerical capabilities of different RTM codes (from Steefel et al. 2015).

Capabilities	PHREEQC	HP1	PHT3D	OGS	HYTEC	ORCHESTRA	TOUGHREACT	εSTOMP	HYDROGEOCHEM	CrunchFlow	MIN3P	PFLOTRAN
Coupling												
Reaction-induced porosity-permeability	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Coupled heat of reaction	No	No	No	Yes	No	No	Yes	No	Yes	No	No	Yes
Coupled deformation compaction <sup>1</sup>	No	No	No	Yes	No	No	C-E	No	EGM	C-E	No <sup>2</sup>	Yes
Reaction consumption of phase ( $H_2O, CO_2$ ) <sup>1</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Electrical Double layer transport	Yes	No	No	No	No	Yes	No	No	Yes	Yes	No	No
Numerical scheme												
Operator splitting	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Global implicit	No	No	No	Yes	No	No	No	No	No	Yes	Yes	Yes
Hight Peclet number transport	Yes	No	Yes	Yes <sup>3</sup>	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Spatial discretization <sup>4</sup>	MC	FEM	FVM, MMC	FEM	FVM	MC	FVM	FVM	FEM, MMC	FVM	FVM	FVM
Time discretization <sup>5</sup>	TW	TW	TW	BE	TW	FE/RK	BE/CN	BE	TW <sup>mid</sup>	BE	BE	BE

<sup>1</sup> C-E denotes compaction-erosion, and EGM denotes equilibrium Geo-Mechanics Model<sup>2</sup> Simplified 1-D approach<sup>3</sup> using linearized algebraic flux corrected transport (FCT)<sup>4</sup> MC denotes mixing cell, FEM is Finite Element Method, FVM denotes Finite Volume Method and MMC denotes Modified Method of characteristic or backward particle tracking<sup>5</sup> BE denotes Backward Euler, TW denotes time weighted difference including forward difference, backward difference, Crank-Nicolson Central Difference, mid denotes Mid difference, FE is Finite Element, and RK denotes Runge-Kutta

# 3

---

## NUMERICAL SIMULATION OF WATER FLOW IN SATURATED AND UNSATURATED ZONES

---

### 3.1 INTRODUCTION

The reactive transport of environmental chemicals such as pesticides in the subsurface involves various processes in both unsaturated soil layers and the saturated aquifer. The correct simulation of water flow in the coupled soil-aquifer system is the prerequisite for the plausible simulation of the solute transport and geochemical reaction processes in the subsurface environment.

This chapter is divided into two parts. The first part introduces the numerical simulation of the classic Theis problem by using finite-element-based software OpenGeoSys (OGS) for different dimensions, in order to evaluate the plausibility of OGS to solve groundwater flow problems in saturated porous media under a stress condition. The second part deals with the numerical simulation of unsaturated flow by using the Richards equation. Mainly based on the work of He et al. 2015d, the capability and accuracy of different forms of Richards equation are analyzed.

### 3.2 THEIS PROBLEM

Theis problem describes a transient lowering of the water table in a confined aquifer caused by a pumping well (see Fig. 1). In the current work, two different approaches i.e. groundwater flow (Eq. 2) and liquid flow (Eq. 3) are applied to simulate Theis problem for a homogeneous and isotopic case. Models with different dimensions (2-D, 3-D, radial symmetric 1.5-D and 2.5-D) are built and their simulation results are compared with the analytical solution (Theis 1935). For radial symmetric cases, the governing

### 3.2 THEIS PROBLEM

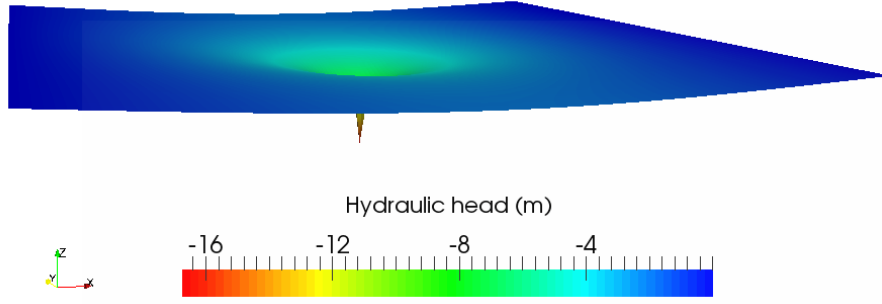


Figure 1.: Hydraulic drawdown in an aquifer as a result of pumping activity.

equations have to be modified to be suitable for polar coordinate system. For example, Eq.2 is changed into Eq.35.

$$S_s \frac{\partial h}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} (K_r r \frac{\partial h}{\partial r}) + \frac{\partial}{\partial z} (K_z \frac{\partial h}{\partial z}) + q \quad (35)$$

where  $r$  denotes the radial distance to the pumping well.

The corresponding initial-boundary conditions are formulated as follows:

$$\begin{aligned} h(t=0, r) &= h_0 \\ \lim_{r \rightarrow 0} (r \frac{\partial h}{\partial r}) &= \frac{Q}{2\pi T} \\ \lim_{r \rightarrow \infty} h(t, r) &= h_0 \end{aligned} \quad (36)$$

where  $h_0$  is the constant initial hydraulic head [m],  $Q$  is the constant discharge rate [ $m^3 \cdot s^{-1}$ ],  $T$  is the aquifer transmissivity [ $m^2 \cdot s^{-1}$ ] and  $t$  is time [s].

The analytical solution of Theis problem can be obtained based on Eq. 37.

$$h(t, r) = h_0 - \frac{Q}{4\pi T} W(u) \quad (37)$$

where  $W(u)$  is the well function for a confined aquifer which is defined by an infinite series:

$$\begin{aligned} W(u) &= -Ei(-u) = \int_u^\infty \frac{\exp(-\xi)}{\xi} \\ &= -0.5772 - \ln(u) + \sum_{i=1}^{\infty} (-1)^{i-1} \frac{u^i}{i \times i!} \\ \text{with } u &= \frac{S}{4T} \frac{r^2}{t} \end{aligned} \quad (38)$$

where  $S$  is the aquifer storage [-].





Table 2.: Parameters and their values applied for Theis problem (from He et al. 2015c).

Parameter	Symbol	Value	Unit
Pumping rate	$Q$	1.4158E-2	m <sup>3</sup> /s
Hydraulic conductivity	$K$	9.2903E-4	m/s
Intrinsic permeability	$\kappa$	1.2391E-10	m <sup>2</sup>
Specific Storage	$S_s$	1E-3	1/m
Well radius	$r_w$	0.3048	m
Study area length	$r_b$	1000	m
Density of water (10 °C)	$\rho$	999.7026	kg · m <sup>-3</sup>
Viscosity of water (10 °C)	$\mu$	1.308E-03	Pa · s

### 3.2.1 Benchmark description

The properties of the aquifer as well as the parameters for pumping operations are summarized in Tab. 2. Tab. 3 summarizes the geometries and concepts of models with different dimensions. In Theis 1.5-D, the homogeneous aquifer is simplified as a 1-D line segment radiates from a point (the pumping well); whereas in Theis 2.5-D, it becomes a rectangular domain rotating around a line segment as the pumping well. For Theis 2-D, a triangle domain is applied with one of its vertices representing the pumping well (the left vertex). The angle of that vertex is 10°. Theis 3-D is a simple extrusion of Theis 2-D in  $z$  direction. During the spatial discretization for all these model domains, mesh elements for regions around the well are refined. The mesh for Theis 3-D is illustrated in Fig. 2. A detailed description of the model setups in OGS is provided in He et al. 2015c.

Table 3.: Summary of the model concepts for Theis' problem (from He et al. 2015c).

Dimension	1.5-D	2-D	2.5-D	3-D
Geometry				
Concept	Radial symmetry	—	Axisymmetry	—



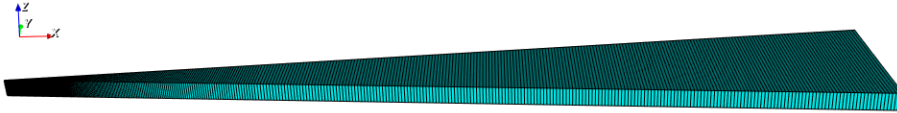


Figure 2.: The employed mesh for Theis 3-D.

### 3.2.2 Results

For all the simulations, a simulation time of 10 days is discretized into 60 time steps with sizes changing from  $1E-5$  (at the starting period of pumping) to 0.9 day (after 1 day). Fig. 3 shows the simulated time series of hydraulic head drawdowns for locations with a distance of 9.639 m from the well. Numerical simulation results of all models are in good agreement with the analytical solution.

## 3.3 NUMERICAL SIMULATION OF UNSATURATED WATER FLOW

Richards equation is widely applied to simulate the coupled saturated-unsaturated flow. As introduced in Sect. 2.1.2, there are several different forms of Richards equation. Several researchers (Celia et al. 1990, Clement et al. 1994, Lehmann and Ackerer 1998) prefers the mixed form (Eq. 6) because it conserves mass more precisely than the h-based form, especially for highly non-linear problems such as infiltration into initially dry soil. However, in the recent past, Sadegh Zadeh 2011 pointed out, that the mixed-form is only applicable in the unsaturated zone. Thus he derived an algorithm to switch between the mixed-form and the h-based form during the simulations.

In order to avoid further confusion and clarify the capability of the mixed form in terms of simulating water flow in unsaturated-saturated system, He et al. 2015d analyzed the capability and accuracy of the mixed, the h-based form as well as the switching algorithm proposed by Sadegh Zadeh 2011 through literature research and benchmarking tests, which is introduced below.

### 3.3.1 Mixed-form of Richards equation revisited

Celia et al. 1990 proposed a modified Picard iteration scheme for the mixed form of Richards equation. This approach was proved to have a better mass balance than the h-based form, the detailed reasons for which was given in Celia et al. 1990. In short, this is

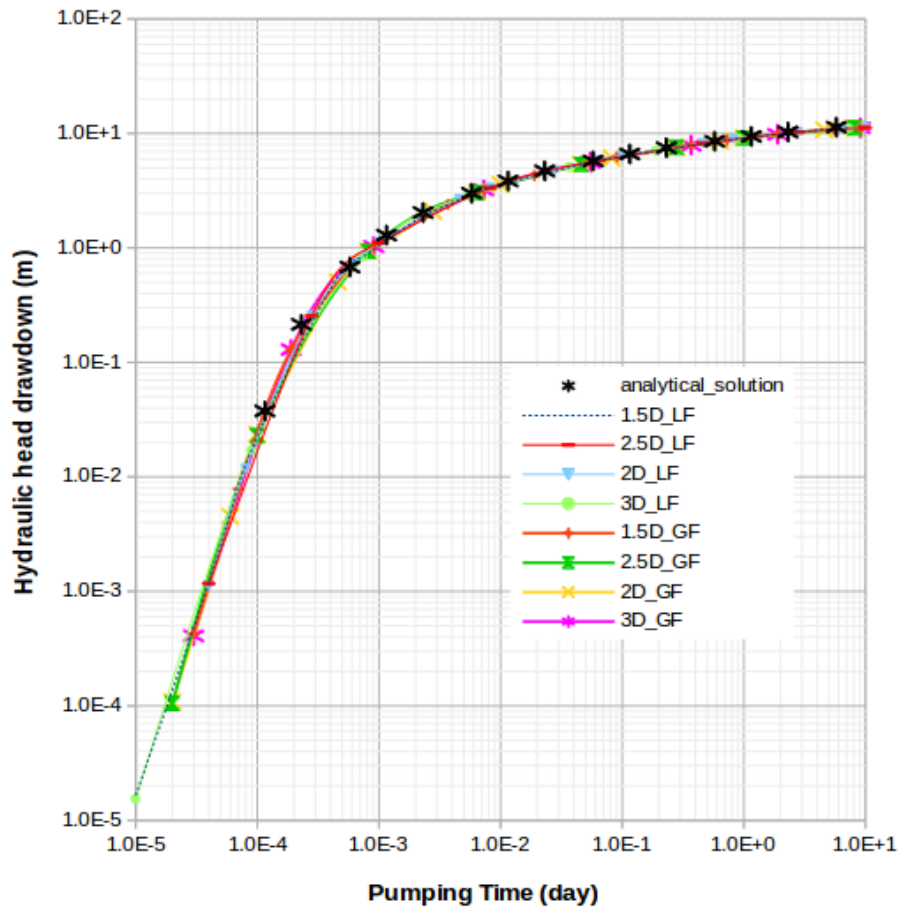


Figure 3.: Calculated hydraulic head drawdowns at a distance of 9.639 m from the well. "GF" denotes groundwater flow, whereas "LF" means liquid flow.

because the two approaches use different ways to approximate the time derivative of water content of Richards equation.

In the h-based form the time derivative term i.e.  $\frac{\partial \theta}{\partial t}$  is approximated by  $C(h_p) \frac{\partial h_p}{\partial t}$ . It can be discretized as follows by using a fully implicit backward Euler scheme:

$$C(h_p) \frac{\partial h_p}{\partial t} \approx C(h_{p\tau}^{n+1}) \frac{h_{p\tau+1}^{n+1} - h_p^n}{\Delta t^{n+1}} \quad (39)$$

where  $n$  is the time step and  $\tau$  the iteration.

Although  $\frac{\partial \theta}{\partial t}$  and  $C(h_p) \frac{\partial h_p}{\partial t}$  are mathematically equivalent in the partial differential equation, their discretized forms are not. The inequality can be intensified dramatically due to the high non-linearity of the term  $C(h_p)$ , which can lead to large mass balance errors when using the h-based form (Celia et al. 1990).

In the mixed form,  $\frac{\partial \theta}{\partial t}$  is directly discretized as  $\frac{\theta_{\tau+1}^{n+1} - \theta^n}{\Delta t^{n+1}}$  and can be further approximated as:

$$\frac{\partial \theta}{\partial t} \approx \frac{\theta_{\tau+1}^{n+1} - \theta^n}{\Delta t^{n+1}} \approx C(h_{p\tau}^{n+1}) \frac{h_{p\tau+1}^{n+1} - h_{p\tau}^{n+1}}{\Delta t^{n+1}} + \frac{\theta_{\tau+1}^{n+1} - \theta^n}{\Delta t^{n+1}} \quad (40)$$

which is obtained by substituting  $\theta_{\tau+1}^{n+1}$  with the first two terms of its truncated Taylor series expansion:

$$\theta_{\tau+1}^{n+1} = \theta_{\tau}^{n+1} + \frac{d\theta}{dh} (h_{p\tau+1}^{n+1} - h_{p\tau}^{n+1}) + O(\delta^2) \quad (41)$$

The finite difference discretization of the mixed form is similar to that of the h-based form apart from the approximation of  $\frac{\partial \theta}{\partial t}$ , and exactly this difference provides an accurate mass balance by the mixed form Celia et al. 1990. In the saturated zone, both  $C(h_{p\tau}^{n+1}) \frac{h_{p\tau+1}^{n+1} - h_{p\tau}^{n+1}}{\Delta t^{n+1}}$  and  $\frac{\theta_{\tau+1}^{n+1} - \theta^n}{\Delta t^{n+1}}$  converge to zero. As a result, there is no difference between the two forms anymore.

Similarly, when using the finite element method the only difference between the two forms is also the approximation of  $\frac{\partial \theta}{\partial t}$ . The detailed finite element approximation of both forms with different iterative schemes such as modified Picard or Newton's method is given in Lehmann and Ackerer 1998. The corresponding modified Picard schemes for both forms in the 1D vertical direction are demonstrated below.

The h-based form:

$$\left[ \mathbf{A}_{\tau}^{n+1} + \frac{\mathbf{B}_{\tau}^{n+1}}{\Delta t^{n+1}} \right] (h_{p\tau+1}^{n+1} - h_p^{n+1}) = \mathbf{F}_{\tau}^{n+1} - \mathbf{A}_{\tau}^{n+1} h_{p\tau}^{n+1} - \frac{\mathbf{B}_{\tau}^{n+1}}{\Delta t^{n+1}} (h_{p\tau}^{n+1} - h_p^n) \quad (42)$$

The mixed form:

$$\left[ \mathbf{A}_\tau^{n+1} + \frac{\mathbf{B}_\tau^{n+1}}{\Delta t^{n+1}} \right] (h_{p\tau+1}^{n+1} - h_p^{n+1}) = \mathbf{F}_\tau^{n+1} - \mathbf{A}_\tau^{n+1} h_{p\tau}^{n+1} - \frac{\mathbf{E}}{\Delta t^{n+1}} (\theta_\tau^{n+1} - \theta^n) \quad (43)$$

where  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{F}$  are the Laplace matrix, the mass matrix and the RHS vector for the h-based form, respectively, with

$$\mathbf{A} = [A_{ij}] = \int_0^L K(h_p) \frac{d\phi_i}{dz} \frac{d\phi_j}{dz} dz \quad (44)$$

$$\mathbf{B} = [B_{ij}] = \int_0^L C(h_p) \phi_i \phi_j dz \quad (45)$$

$$\mathbf{E} = [E_{ij}] = \int_0^L \phi_i \phi_j dz \quad (46)$$

$$\mathbf{F} = (f_i) = \int_0^L K(h_p) \frac{d\phi_i}{dz} dz + \left( K(h_p) \frac{\partial h_p}{\partial z} - K(h_p) \right) \Big|_0^L \quad (47)$$

where  $\phi_i$  and  $\phi_j$  are linear basic functions,  $\left( K(h_p) \frac{\partial h_p}{\partial z} - K(h_p) \right) \Big|_0^L$  is the Neumann boundary condition for upper ( $z = 0$ ) and lower ( $z = L$ ) boundary of the 1D model.

The only difference between the discretized forms of the h-based and the mixed form is thus the third term on the right hand side of Eq. 42 and Eq. 43. In the saturated zone, both  $\frac{\mathbf{B}_\tau^{n+1}}{\Delta t^{n+1}}$  and  $\frac{\mathbf{E}}{\Delta t^{n+1}} (\theta_\tau^{n+1} - \theta^n)$  converge to zero, hence the difference between the two forms vanishes.

### 3.3.2 Benchmark analysis

Sadegh Zadeh 2011 proposed a switching algorithm which uses the mixed form of Richards equation in the unsaturated zone and switches to the h-based form in and near the saturated zone based on a threshold value of pressure head (-2.5 cm).

In this section, the accuracy of the h-based, the mixed form as well as the switching algorithm is tested by using two benchmarks. The first one is a one-dimensional model about the infiltration into layered soils, whereas the second one is a two-dimensional model involving coupled saturated and unsaturated flow.

Table 4.: Soil hydraulic parameters of Berino loamy fine sand and Glendale clay loam (from Hills et al. 1989).

Soil type	$\theta_r[-]$	$\phi[-]$	$\alpha[cm^{-1}]$	$n[-]$	$k_s[m \cdot d^{-1}]$
Berino loamy fine sand	0.0286	0.3658	0.0280	2.2390	541.0
Glendale clay loam	0.1060	0.4686	0.0104	1.3954	13.1

### 3.3.2.1 Infiltration into layered soils Hills et al. 1989

This benchmark is based on Hills et al. 1989. The model domain of 100 cm consists of 5 soil layers (each has a thickness of 20 cm), in which Berino loamy fine sand and Glendale clay loam alternate with each other (with Berino loamy sand layer on the top). The soil hydraulic parameters of both soil types are listed in Tab. 4, where  $\theta_r$ ,  $\phi$ ,  $\alpha$ ,  $n$  and  $k_s$  are the residual saturation, porosity, van Genuchten  $\alpha$ , van Genuchten  $n$  and saturated hydraulic conductivity, respectively. On the upper boundary, a constant Neumann boundary condition with a flux of  $2 \text{ cm} \cdot d^{-1}$  is given. On the lower boundary, the pressure head is fixed as the initial head. The three algorithms are tested by using two scenarios with different initial heads ( $h_0$ ) or vertical discretization ( $\Delta z$ ) (see Tab. 5).

An adaptive time-stepping scheme based on Wang et al. 2011 is applied for all three algorithms. The simulation results are illustrated in Fig. 1, whereas the mass balance errors (calculated based on Eq. 48) by using different algorithms are listed in Tab. 5. With an initial head of  $-1000 \text{ cm}$  (scenario 1), the h-based form brings much larger mass balance errors and has different infiltration front compared with the other two methods (see Fig. 4); under a drier initial condition (scenario 2), the h-based form does not converge with the given spatial discretization. For both scenarios, the mixed form and the switching algorithm produce very similar results and comparable mass balance errors.

$$ERROR = \frac{\left| \sum_e (V_t^e - V_0^e) - \int_0^t \sum_{n\Gamma} Q_n dt \right|}{\max \left( \left| \sum_e (V_t^e - V_0^e) \right|, \left| \int_0^t \sum_{n\Gamma} Q_n dt \right| \right)} 100 \quad (48)$$

where  $V_t^e$  and  $V_0^e$  are the water volumes in element  $e$  at time  $t$  and zero, respectively;  $Q_n$  is the water flux through boundary nodes  $n\Gamma$ .

### 3.3.2.2 Water table experiment

This benchmark is based on the infiltration experiment from Vauclin et al. 1979, which was conducted on a slab of soil 3 m long,

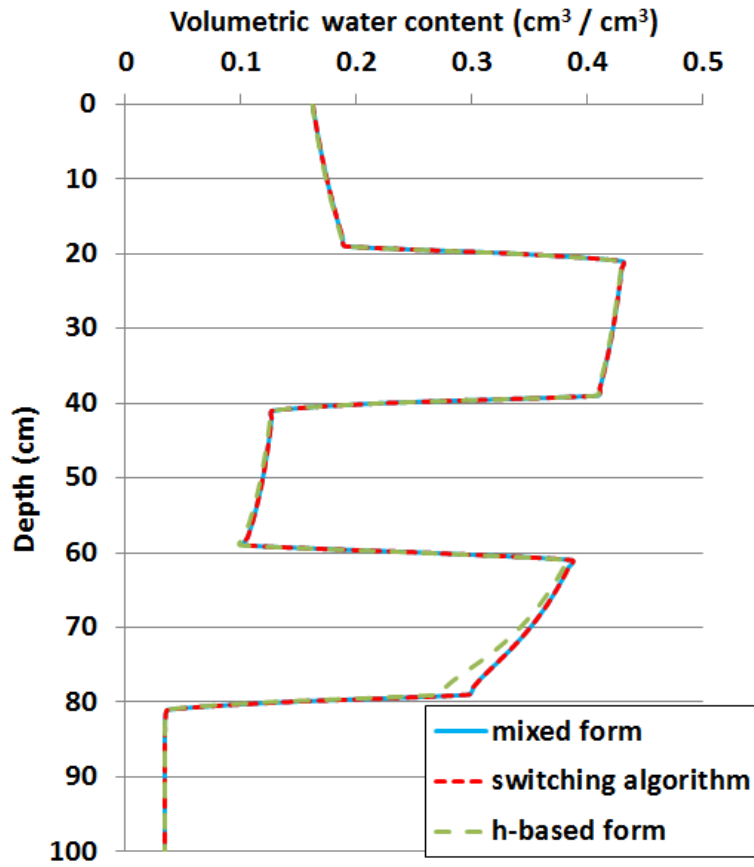


Figure 4.: Distribution of volumetric water content in two scenarios by using three different algorithms.

Table 5.: Model setups and mass balance errors of different algorithms for a 5-day simulation.

Scenario	Algorithm	$h_0[cm]$	$\Delta z[cm]$	Mass balance error [%]
1	h-based form	-1000	1	2.98
	mixed form	-1000	1	4.46E-2
	switching algorithm	-1000	1	7.56E-2
2	h-based form	-10000	0.1	-
	mixed form	-10000	0.1	8.39E-3
	switching algorithm	-10000	0.1	6.94E-3

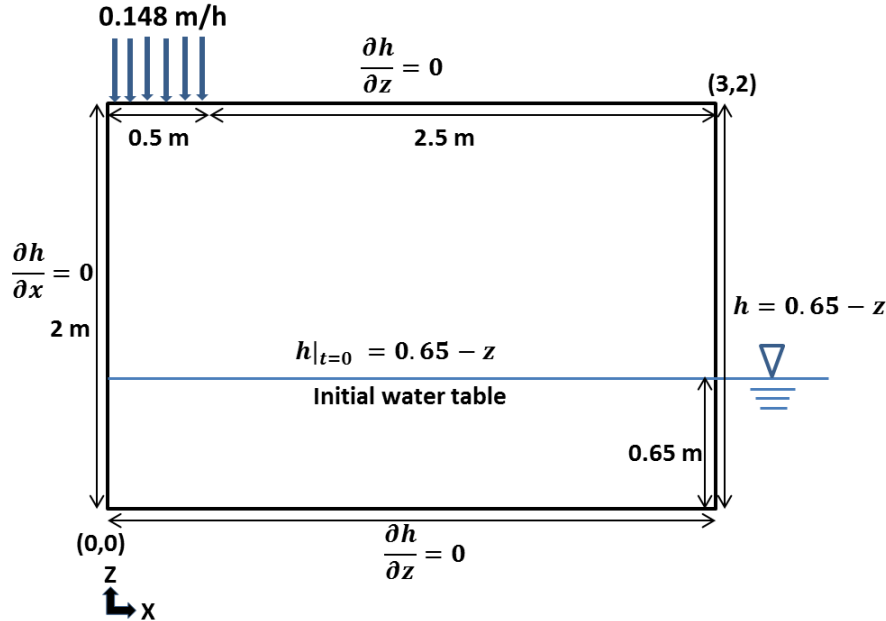


Figure 5.: Initial and boundary conditions of the benchmark example (from He et al. 2015d).  $h(x, z, t)$  is pressure head (in meter) at coordinates  $x$  and  $z$  at time  $t$ .

Table 6.: Soil hydraulic parameters based on the data from Vauclin et al. 1979 and Clement et al. 1994.

Parameter	$\theta_r[-]$	$\phi[-]$	$\alpha[cm^{-1}]$	$n[-]$	$k_s[m \cdot d^{-1}]$
Value	0.01	0.30	0.033	4.1	8.40

2 m high and 5 cm thick in the laboratory. The water table was initially located at an elevation of 0.65 m from the bottom. A constant flux of 0.148 m/h was applied over a width of 0.5 m on the left part of the soil surface (see Fig. 5). There was no flux on the left boundary of the domain, while the pressure head of the right boundary was fixed as constant by combining to a constant head reservoir. The whole experiment lasted 8 hours.

The soil hydraulic parameters for simulations are given in Tab. 6. The initial and boundary conditions are illustrated in Fig. 5. The model domain consists of 1200 rectangular elements, each has a length of 0.10 m and a height of 0.05 m. Same as the first benchmark, the adaptive time-stepping scheme based on Wang et al. 2011 is applied for all the three algorithms during the simulation.

Fig. 6 illustrates a comparison of 12 simulated water table curves at different times together with the according measured data. Generally, the simulation results of the switching algorithm, the mixed

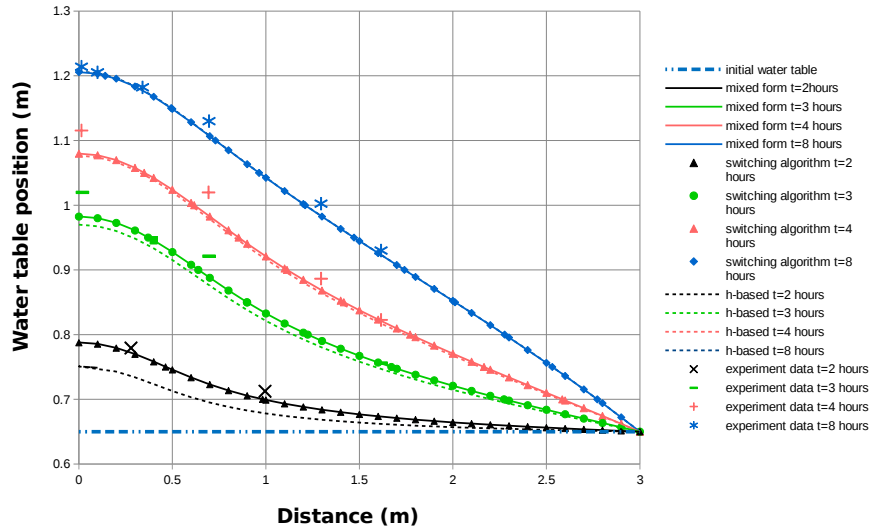


Figure 6.: Comparison of the simulation results of different approaches together with experiment data.

Table 7.: Mass balance errors of different algorithms for the water table benchmark.

Algorithm	Mass balance error (%)		
	h-based form	mixed form	switching algorithm
2 h	7.2	1.85	1.94
3 h	4.0	1.57	1.58

and the h-based form are close to the experimental data. However, at the early stage of the simulation (2 hours and 3 hours), when infiltration into the dry unsaturated zone is the dominating process, the mixed form and the switching algorithm show better correspondences with the experimental data than the h-based form. Additionally, the mixed-form and the switching algorithm have much lower mass balance errors than those of the h-based form (see Tab. 7). A solution to reduce the mass balance error of the h-based form is to use smaller time step sizes. If a fixed time step sizes of 60 s is applied for the simulation, then the h-based form delivers much more accurate results at the early stages (not shown here). Nevertheless, using smaller time steps sizes also leads to a longer execution time.



## 3.4 SUMMARY

This chapter represents different numerical approaches to simulate water flow problems in variably saturated porous media.

In the first part, This problem is numerically simulated by using OGS. The comparison of simulation results with the analytical solution demonstrates, that OGS is able to model water flow in confined aquifer system with different dimensions.

In the second part, the capability and accuracy of different forms of Richards equation is analyzed with two benchmark examples. The mixed form shows a better accuracy than the h-based form, especially when an automatic time stepping is applied. This is because the mixed form can better conserve mass than the h-based form, especially when coarse temporal discretization is applied. The mixed form can be used in both saturated as well as unsaturated zone, and the switching algorithm of Sadegh Zadeh 2011 does not outperform the mixed form in terms of accuracy. Hence, the switching between h-based and mixed form, as proposed by Sadegh Zadeh 2011, may not be necessary.

# 4

---

## IMPLEMENTATION AND ANALYSIS OF THE COUPLING INTERFACE OGS#IPHREEQC

---

### 4.1 INTRODUCTION

In the current study, the scientific open source software OpenGeoSys (OGS) was coupled to the IPhreeqc module of the geochemical solver PHREEQC. OGS is responsible for the calculation of flow and solute transport processes, whereas IPhreeqc simulates the geochemical reactions locally on each finite element node. A coupling interface was developed to transfer data between these two softwares.

In this chapter, the concepts and implementation of the coupling interface between OGS and PHREEQC will be introduced, its plausibility and accuracy will then be tested with 3 benchmarks.

### 4.2 OPENGEOSYS

Based on finite-element method, OpenGeoSys (OGS) is able to simulate a wide range of thermo-hydro-mechanical/chemical (THM-/C) coupled processes (Kolditz et al. 2012). OGS can be used to simulate fluid flow processes such as groundwater flow, liquid flow, Richards flow, multiphase flow and density-dependent flow. For solving non-linear PDEs (e.g. Richards equation) both Picard and Newton-Raphson schemes can be applied. For the modeling of mass transport, the advection-dispersion equation is applied (see Eq. 15).

There are several possibilities to model reactive transport problems with OGS. On the one hand, its internal KinReact module can be applied for the simulation of kinetically controlled reactions; on the other hand, there exists several couplings between OGS and external geochemical solvers such as PHREEQC (Xie et al. 2006), GEMs (Shao et al. 2009, Kosakowski and Watanabe 2014), BRNS (Centler et al. 2010) and ChemApp (Beyer et al. 2012, Li et al. 2014). Each approach has its own strengths and

drawbacks. For example, OGS-ChemApp is well suitable for "extreme" systems with high temperature and pressure, it does not allow for the consideration of reaction kinetics; the KinReact module includes solvers for stiff ODE systems and thus can be used to simulate kinetic reactions, it cannot be applied for equilibrium reactions.

For more detailed information regarding to OGS developments and benchmarking can be found at <http://www.opengeosys.org/>. The official releases as well as code resources can also be accessed from there.

### 4.3 PHREEQC AND IPHREEQC MODULE

As one of the most widely used open-source geochemical solvers, PHREEQC provides a wide range of geochemical capabilities such as aqueous, mineral, gas, surface, ion exchange, solid-solution equilibria and kinetic reactions (Parkhurst and Appelo 1999, Parkhurst and Appelo 2013). A comparison of the geochemical features of PHREEQC and other reactive transport modeling codes is listed in Tab. 8 (from Steefel et al. 2015). The calculation capabilities of PHREEQC include initial speciation calculations, batch reaction simulations, inverse modeling, one-dimensional advective-dispersive transport calculations, and so on.

Table 8.: Comparison of the key geochemical features of different RTM codes (from Steefel et al. 2015).

	PHREEQC	HP1	PHT3D	OGS	HYTEC	ORCHESTRA	TOUGHREACT	ESTOMP	HYDROGEOCHEM	CrunchFlow	MIN3P	PFLOTRAN
Capabilities	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Extended Debye-Huckel	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Davies activity model	Yes	Yes	No	Yes <sup>1</sup>	No	No	No	Yes	No	No	Yes	No
Pitzer activity model <sup>1</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Non-isothermal geochemistry <sup>1</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Surface complexation <sup>2</sup>	DDL, Non-edl	DDL, Non-edl	DDL, Non-edl	DDL, Non-edl	DDL, CC	DDL	DDL, CC	DDL, Non-edl	DDL, Non-edl	DDL, Non-edl	Non-edl	Non-edl
Ion exchange	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Aqueous-gas exchange	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kinetic mineral precipitation-dissolution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
mineral nucleation	Yes	Yes	Yes	No	No	Yes	No	No	No	Yes	No	No
Mineral solid-solutions	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Equilibrium isotope fractionation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No
Kinetic isotope fractionation	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes <sup>2</sup>	No
Aqueous kinetics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Radioactive decay chains	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

<sup>1</sup> Not available in the native chemistry of OGS, but can be realized through its coupling interfaces to other geochemical or biogeochemical solvers which can use Pitzer activity model

<sup>2</sup> DDL denotes diffuse double layer, non-edl denotes non-electrostatic model, and CC denotes constant capacitance

<sup>3</sup> Not available in MIN3P-THCm, but available in a customized version of MIN3P

There are different PHREEQC versions such as batch version, graphical user interface (for Windows system) as well as modules that allow PHREEQC to be linked with other software (Parkhurst and Appelo 2013). One of these modules is IPhreeqc ("I" stands for interface), which is written in C++. It provides a set of well-defined approaches for data exchange between PHREEQC and client programs (Charlton and Parkhurst 2011). For example, a client program can prepare the PHREEQC input data as either a file or a character string; these data can then be executed by IPhreeqc with methods like *RunFile* (for input file) or *RunString* (for input character string). The simulation results from PHREEQC can also be retrieved as a file or a string buffer depending on the methods applied. Apart from that, the *GetSelectedOutputValue* method can be used to return the individual data item at a specific position of the result array. For more detailed introductions about different IPhreeqc methods please refer to Charlton and Parkhurst 2011.

#### 4.4 IMPLEMENTATION OF THE COUPLING INTERFACE

He et al. 2015b described the coupling interface OGS#IPhreeqc in great details. The SNIA approach (introduced in Sect. 2.4) is applied for the coupling between OGS and IPhreeqc, in which OGS is responsible for solving non-geochemical processes such as water flow and solute transport, whereas the geochemical reactions will be calculated by IPhreeqc. It is worth mentioning, that the Courant–Friedrichs–Lewy (CFL) condition is not embedded in the coupling itself. In order to reduce the operator splitting errors and also to ensure numerical stability for the ADE, it is important to take the CFL condition into account during the spatial- and temporal discretization of the model setup.

The coupling between these two software packages is realized at source code level, which means IPhreeqc functions can be accessed directly in the coupling interface. This is the major difference between OGS#IPhreeqc and the existing coupling between OGS and PHREEQC (Xie et al. 2006), in which PHREEQC is executed externally with a system call.

The interface is highly independent from the code updating from both software packages, which means that it can stay unchanged when a new version is released. When a new release from IPhreeqc is given, it can be integrated efficiently by updating the source code from the IPhreeqc side. When new IPhreeqc files are added (or

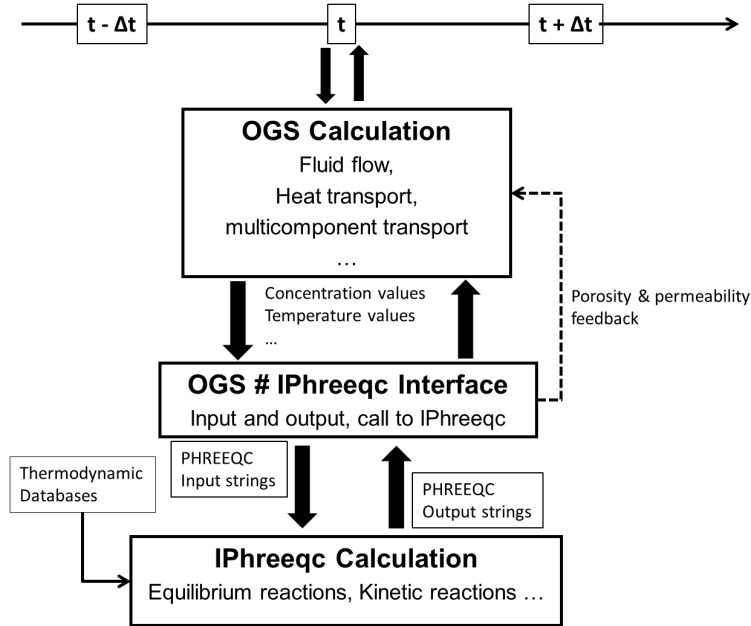


Figure 7.: The general idea of the coupling procedure between OGS and IPhreeqc (from He et al. 2015b).

old files are removed) in the new version, only a reconfiguration of the build system is required (e.g. update of the CMake file).

To setup a model in OGS, several input files need to be prepared. Each of them is responsible for defining certain aspect of the model e.g. geometry, mesh, initial-boundary conditions, numerical schemes, and so on. To setup reactive transport models using OGS#IPhreeqc, one has to provide a thermodynamic database file and prepare another input file to define the geochemical system in IPhreeqc. The content of the input file is very similar to that of a PHREEQC input file. The major difference is, that the transport module is not defined in the input file for OGS#IPhreeqc, since solute transport is solved by OGS. A detailed description of the OGS input files is given in Appendix A.

Based on the previous work of a file-based coupling (see He et al. 2015a), a character-string-based interface was developed, whose structure is illustrated in Fig. 7.

Before entering the time stepping loop, a call to IPhreeqc (more details will be given below) is made to initialize the whole geochemical system. During each time step, fluid flow will be calculated first, followed by the simulation of heat and solute transport processes. After that, the values of state variables (e.g. concentrations of all species and temperatures) on all the finite-element nodes will be processed by the interface to prepare a single input

string for IPhreeqc.

A call to IPhreeqc will then be made, which includes the following procedures and IPhreeqc methods:

- 1) create a instance for IPhreeqc (*CreateIPhreeqc*);
- 2) load a thermodynamic database (*LoadDatabase*);
- 3) run the IPhreeqc input string (*RunString*);
- 4) collect the selected output from IPhreeqc (*GetSelectedOutputString*);
- 5) release the IPhreeqc instance (*DestroyIPhreeqc*).

It should be noted, that in the current implementation a complete call (including all the steps mentioned above) of IPhreeqc has to be performed during each time step, which means IPhreeqc has to be restarted and reinitialized. Step 1, 2 and 5 can thus be regarded as the overhead for calling IPhreeqc, which will be analyzed in Sect. 4.5.

After IPhreeqc calculation is complete, the values of state variables on all nodes will be updated and used for the coming OGS calculation in the new time step. The porosity or permeability change as a consequence of mineral dissolution/precipitation can also be taken into account.

## 4.5 BENCHMARKING

Several benchmarking tests were made to verify the correctness and accuracy of the coupling between OGS and IPhreeqc, which will be introduced below.

### 4.5.1 *Isotope fractionation*

The first benchmark is based on the work of van Breukelen et al. 2005, which simulates the sequential reductive dechlorination of chlorinated hydrocarbons (CHCs) in a aquifer over a period of 20 years. The aquifer is around 876 m long with a hydraulic conductivity of  $10 \text{ m} \cdot \text{d}^{-1}$  and a porosity of 0.25. The velocity of the groundwater flow in the aquifer is  $0.1 \text{ m} \cdot \text{d}^{-1}$ . Initially, there is no CHCs in the aquifer system. The light and heavy isotopes of tetrachloroethene (PCE) are introduced as constant concentrations at one side of the aquifer with values of 0.9892 and  $0.0108 \text{ mol} \cdot \text{m}^3$ , respectively. PCE is transported along the hydraulic gradient of the aquifer and undergoes sequential degradations illustrated below:



where TCE, DCE, VC and ETH denote tri- and dichloroethylene, vinyl chloride and ethane, respectively. The light ( $^{12}\text{C}$ ) and heavy

Table 9.: The first-order degradation rate and enrichment factor for CHCs (from van Breukelen et al. 2005)

Parameter	Unit	PCE	TCE	DCE	VC
First-order rate	$day^{-1}$	$5.5 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$
Enrichment factor	-	-5.2	-8.5	-17.8	-23.2

( $^{13}C$ ) isotopes of each CHC as well chloride (produced during each dechlorination process) are modeled as individual species. First-order kinetics is applied to simulate all degradation reactions. For each CHC the kinetic isotope fractionation factor  $\alpha_k$  is assumed to be constant during the whole degradation process, which can be calculated with Eq. 49:

$$\alpha_k = 1 + \frac{\epsilon}{1000} \quad (49)$$

where  $\epsilon$  denotes the isotope enrichment factor.

For each CHC, the degradation of its light and heavy isotope can be calculated with Eq. 50 and 51, respectively.

$$\frac{dCHC_l}{dt} = -\gamma_l CHC_l \quad (50)$$

$$\frac{dCHC_h}{dt} = -\gamma_l CHC_h \left(1 + \frac{\epsilon}{1000}\right) \quad (51)$$

The first-order degradation rate and enrichment factor for each CHC are listed in Tab. 9.

This benchmark is simulated with PHREEQC (batch version), the KinReact module of OGS and OGS#IPhreeqc over a simulation time of 20 years, which is discretized into 100 equal time steps. The 1-D model domain is evenly divided into 120 line elements. The simulation results of the three software packages are compared in Fig. 8, which shows the distributions of the light CHC isotopologues and their  $\delta^{13}C$  isotope signatures along the model domain at the end of the simulation time. A good agreement of the simulation results is obtained by using all the three codes. A comparison of their computational time is listed in Tab. 10. For this example, OGS#IPhreeqc and PHREEQC are one order of magnitude slower than the KinReact module. Nevertheless, PHREEQC has a broad range of geochemical capabilities, which are not available in the KinReact module. The OGS#IPhreeqc interface takes 14.7% of the total simulation time; whereas the overhead in calling IPhreeqc (introduced in Sect. 4.4) accounts for 2.3%.



Table 10.: A comparison of different portions of the simulation time for the van Breukelen benchmark by using OGS#IPhreeqc, KinReact and PHREEQC (in seconds) (from He et al. 2015b).

Code	Flow & Mass transport	Chemistry & interface	Total
OGS#IPhreeqc	0.453	32.218	32.671
PHREEQC	-	-	14.196
KinReact	0.453	0.969	1.389

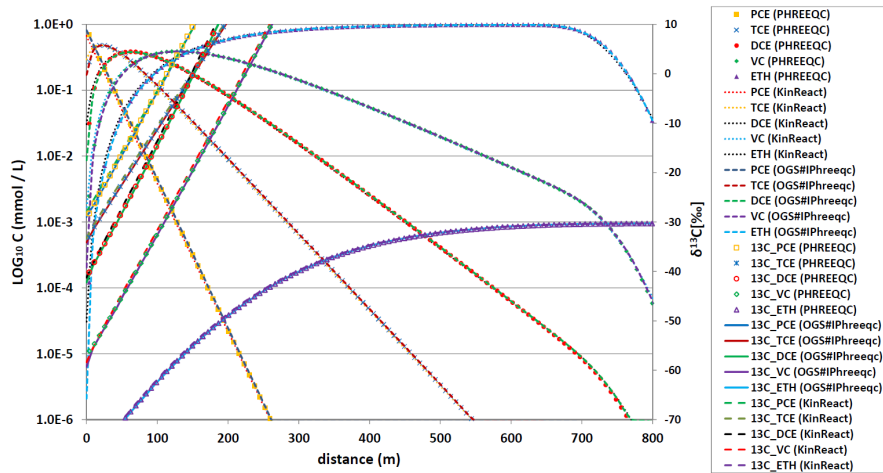


Figure 8.: Concentration profiles of the light CHC isotopologues and  $\delta^{13}C[‰]$  isotope signatures along the model domain simulated by OGS#IPhreeqc, KinReact module of OGS and PHREEQC (symbols) after 20 years.

Table 11.: Material properties of the 1-D calcite column (from He et al. 2015b).

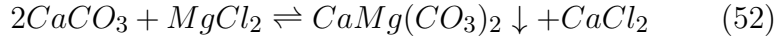
Parameter	Value	Unit
Effective porosity	0.32	–
Bulk density	$1.80 \times 10^3$	$kg \cdot m^{-3}$
Longitudinal dispersivity	$6.70 \times 10^{-2}$	m
Flow rate	$3.00 \times 10^{-6}$	$m \cdot s^{-1}$
Temperature	298.15	K

Table 12.: Initial and boundary conditions for the Engesgaard benchmark (from He et al. 2015b).

Species	Initial conditions	Boundary conditions	Unit
$Ca^{2+}$	$1.23 \times 10^{-1}$	$1.00 \times 10^{-7}$	$mol \cdot m^{-3}$
$Mg^{2+}$	$1.00 \times 10^{-9}$	1.00	$mol \cdot m^{-3}$
C(4)	$1.23 \times 10^{-1}$	$1.00 \times 10^{-7}$	$mol \cdot m^{-3}$
$Cl^-$	$1.00 \times 10^{-9}$	2.00	$mol \cdot m^{-3}$
pH	9.91	7	–
pe	4	4	–
Calcite	$5.7412 \times 10^{-2}$	–	$mol m^{-3}$
Dolomite	0.0	–	$mol m^{-3}$

#### 4.5.2 Engesgaard benchmark

The Engesgaard benchmark (Engesgaard and Kipp 1992) describes the phenomenon occurs when a solution containing magnesium chloride is injected into a calcite column with a length of 0.5 m: calcite dissolves with the movement of the solution; whereas temporary precipitation of dolomite occurs at the dissolution front of calcite. Eq. 52 can be used to describe this phenomenon.



Unlike the original benchmark, in which the dissolution/precipitation of both calcite and dolomite is modeled as equilibrium reaction, here the dissolution/precipitation of the dolomite is modeled as a kinetic reaction using a rate law based on Lasaga et al. 1994 and rate parameters from Palandri and Kharaka 2004.

Tab. 11 and Tab. 12 list the material properties and the initial-boundary conditions of this benchmark, respectively. The model domain consists of 100 line elements with the same length. The total simulation time 21333.32 s is divided into 40 equal time steps. A PHREEQC script is provided in the supplementary material of He et al. 2015b. The simulation results by using OGS#IPhreeqc,

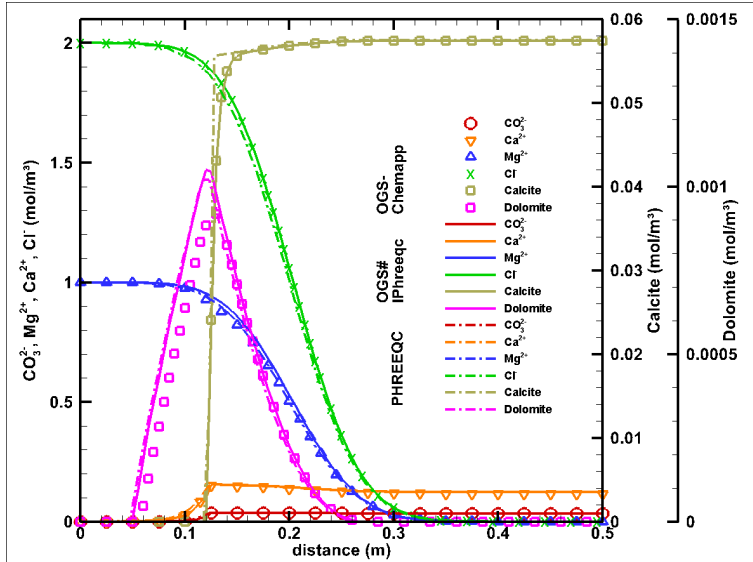


Figure 9.: The concentration profiles of different minerals and aqueous species simulated by OGS#IPhreeqc, PHREEQC and OGS-ChemApp (from He et al. 2015b).

Table 13.: A comparison of different portions of the simulation time for the Engesgaard benchmark by using OGS#IPhreeqc, PHREEQC and OGS-ChemApp (in seconds) (from He et al. 2015b).

Codes	Flow & mass transport	Chemistry & interface	Total
OGS#IPhreeqc	0.047	7.814	7.861
Phreeqc	—	—	5.74
OGS-ChemApp	0.183	23.467	23.65

OGS-ChemApp (combined with the KinReact module) and PHREEQC (batch version) are illustrated in Fig. 9. Generally, good agreements are achieved by using all the three codes except the amount of dolomite. Tab. 13 lists the different portions of the simulation time by using these three codes. For this example, OGS#IPhreeqc is slower than PHREEQC (mainly caused by the interface), but around 2 times faster than OGS-ChemApp. The OGS#IPhreeqc interface and overhead in calling IPhreeqc take 12.7% and 3.8% of the total execution time, respectively.

#### 4.5.3 Uranium leaching

The benchmark presented here is modified from the uranium leaching example of Šimůnek et al. 2012 and Yeh and Tripathi 1991. It simulates uranium leaching as well as the substitution of calcite

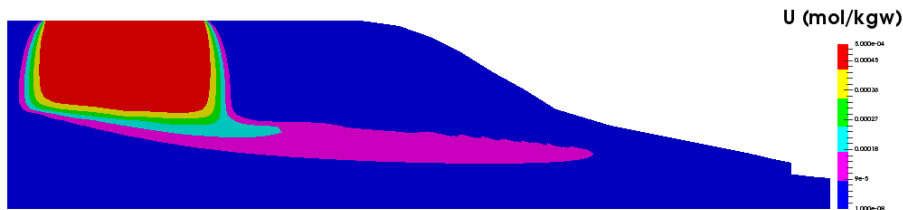


Figure 10.: Uranium leaching from a mill tailing.

by gypsum at a hillslope scale (see Fig. 10), in which water flow in coupled soil-aquifer system is considered. A total of 35 aqueous species and 14 minerals are involved in the reaction network. The model domain is discretized into 14648 triangle elements with a node number of 7522. The total simulation time is 1000 days, which is discretized into 6369 time steps ranging from  $10^{-7}$  to 24000 s.

Fig. 11 shows the geometry, boundary conditions as well as the initial position of the groundwater table of the model domain. Boundary 2 and 3 are set as Neumann boundary condition with infiltration rate of  $1.61\text{E-}7 \text{ m} \cdot \text{s}^{-1}$  and  $1.61\text{E-}8 \text{ m} \cdot \text{s}^{-1}$ , respectively. A constant pressure head of 4.5 m is imposed on boundary 4. A "constrained" pressure head of 12 m is given on boundary 1, which is different from the setup in Šimůnek et al. 2012. This type of boundary condition will switch between the Dirichlet and Neumann boundary condition based on the calculated pressure head. Dirichlet boundary condition will be applied when the pressure head is greater than 0; otherwise zero flux (Neumann boundary condition) will be assigned. The values of the soil hydraulic parameters on the hillslope are given in Tab. 14.

The geochemical reaction networks and thermodynamic database of Šimůnek et al. 2012 are applied here. The concentration of aqueous species and the amount of minerals are set as constant in the mill tailing through out the whole simulation. The initial and boundary conditions for mass transport and geochemical reactions are listed in Tab. 15.

Simulation results are described in great details in the supplementary materials of He et al. 2015b, which includes the distribution of water saturation, pressure head, the concentration evolution of uranium from 250 to 1000 days (Fig. 12), as well as the concentration profiles of pH, calcite and gypsum after 1000 days (see Fig. 13).

The general trends of these results are in line with those presented in Šimůnek et al. 2012. The discrepancies in these two

## 4.5 BENCHMARKING

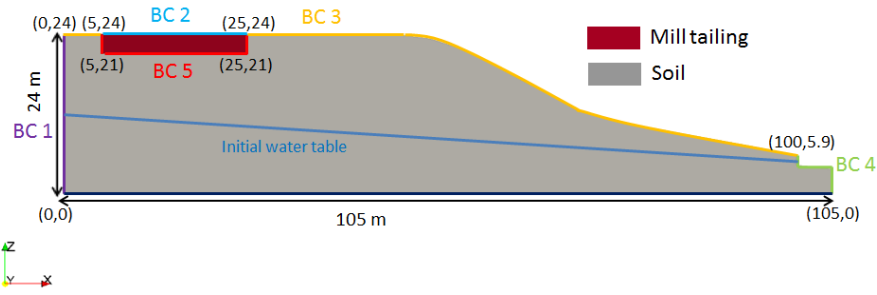


Figure 11.: Geometry and boundary conditions of the uranium leaching example (from supplementary material of He et al. 2015b).

Table 14.: Soil hydraulic parameters of the hillslope (from Šimůnek et al. 2012).

Parameter	Residual saturation	Porosity	van Genuchten $\alpha$	van Genuchten $n$	Saturated hydraulic conductivity
Symbol[unit]	$\theta_r[-]$	$\phi[-]$	$\alpha[cm^{-1}]$	$n[-]$	$k_s[m \cdot d^{-1}]$
Value	0.078	0.43	0.036	1.56	3.78

Table 15.: Initial and boundary setups for geochemical system of uranium leaching example (from supplementary of He et al. 2015b).

Variables	Values	Boundary condition (mol/kgw)		Initial condition (mol/kgw)	
		BC3	BC5	Mill tailing	Soil
C(4)		1.5E-3	1.0E-2	1.0E-2	1.5E-3
Ca		1.0E-3	1.2E-2	1.2E-2	1.55E-2
U		1.0E-8	5.0E-4	5.0E-4	1.0E-7
P		1.0E-6	1.0E-6	1.0E-6	1.0E-6
S		1.0E-4	5.81E-2	5.81E-2	1.48E-2
Fe		1.0E-7	3.5E-2	3.5E-2	1.0E-7
O(0)		4.5E-4	4.2E-4	4.2E-4	4.5E-4
pH		9.4	2.0	2.0	7.1
pe		4.0	18.5	18.5	13.5
Cacite		0	0	0	4.7E-4
Gypsum		0	0	1.4E-1	0

## 4.6 SUMMARY

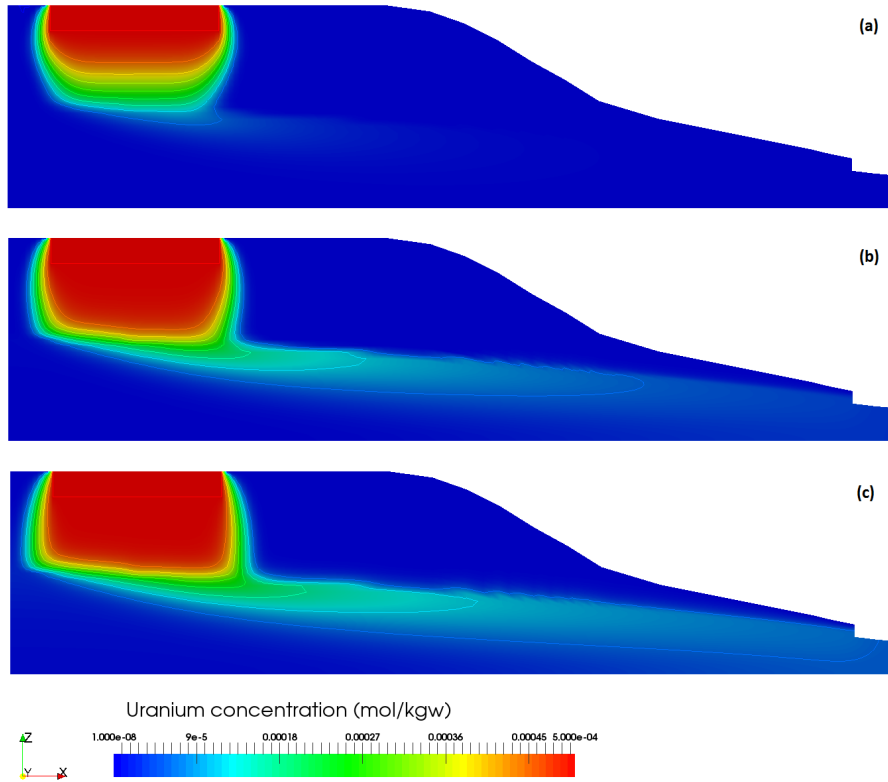


Figure 12.: Time evolution of uranium concentration profile after 200 (a), 500 (b) and 1000 (c) days (from supplementary material of He et al. 2015b).

works are mainly the range of the spread of acid solution and the precipitation of gypsum. In Šimůnek et al. 2012, acid solution presents in a wider region of the vadose zone. Additionally, the precipitation of gypsum also occurs in the lower left boundary of the model domain. These can be caused by using different boundary conditions for simulating saturated-unsaturated flow in these two examples. Among the total time of around 19 hours, the interface takes 3.7% of the total time. Flow/transport and IPhreeqc calculation account for 29.6% and 66.6% of the total time, respectively.

## 4.6 SUMMARY

This chapter introduces the implementation of a character-string-based coupling interface between OGS and IPhreeqc. Three benchmarks involving different flow and geochemical processes are presented to test the accuracy and computational efficiency of the coupled tool. Based on the presented examples, OGS#IPhreeqc

## 4.6 SUMMARY

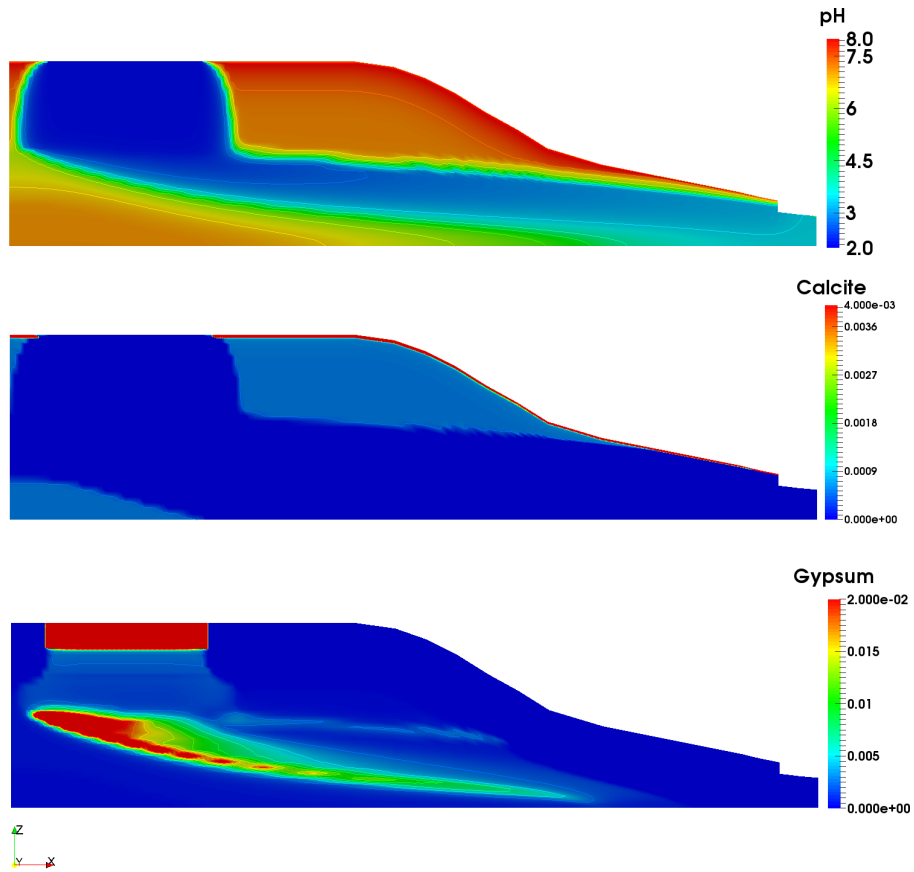


Figure 13.: Distribution of pH, calcite ( $\text{mol} \cdot \text{kgw}^{-1}$ ) and gypsum ( $\text{mol} \cdot \text{kgw}^{-1}$ ) after 1000 days (from supplementary material of He et al. 2015b).

## 4.6 SUMMARY

provides sufficient numerical accuracy to simulate reactive transport problems involving both equilibrium and kinetic reactions in variably saturated porous media. The computational expense of the coupling interface takes less than 15 % of the total execution time for the presented examples.



---

## PARALLELIZATION AND PERFORMANCE ANALYSIS

---

### 5.1 INTRODUCTION

With limited computational resources, reactive transport simulations are often time consuming for large scale problems. Parallel computing is a practical way to improve the computational efficiency for this kinds of problems. Nevertheless, the benefit obtained from adding additional compute cores is limited by the portion of the sequential fraction (the part of computation that can not be parallelized) as well as the portion for inter-processor communication.

A new parallelization scheme is developed in the current work to enable a flexible allocation of compute cores for different types of calculations, thus to achieve better parallel performances than the conventional approach for certain kinds of reactive transport simulations.

In this chapter, the ideas of the new approach will be introduced first, followed by the description of its implementation, performance tests, and a summary.

### 5.2 METHODOLOGY

Based on Amdahl's law, a new term for communication overhead is added to describe the relationship between the "real" parallel performance and the number of employed computer cores (see Eq. 53).

$$S(n) = \frac{T_p + T_s}{\frac{T_p}{n} + T_s + T_c \cdot (n - 1)} \quad (53)$$

where  $S(n)$  is the speedup,  $n$  is the number of cores,  $T_p$ ,  $T_s$  and  $T_c$  are the time consumed for parallel fraction, sequential fraction and the factor for communication, respectively. Here, the time spent for inter-processor communication is assumed to be a linear function of number of compute cores, which increases as the growth of

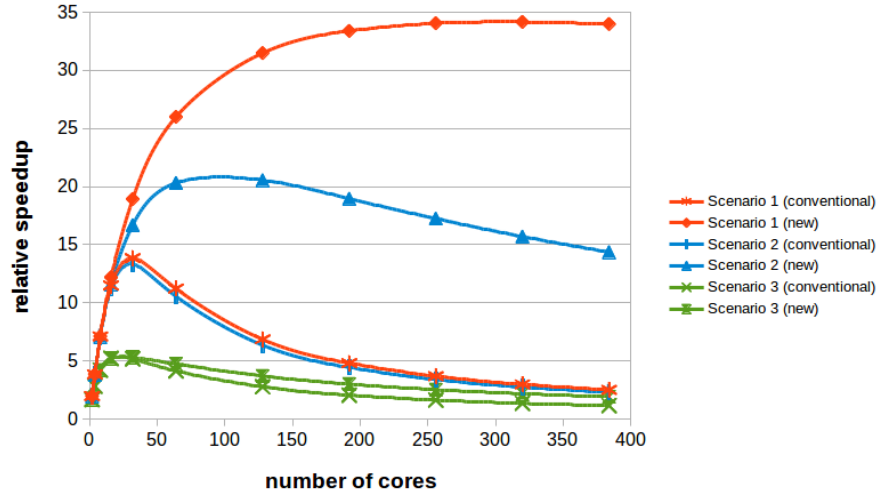


Figure 14.: comparison of the scalability of the conventional and the new parallelization scheme in different scenarios.

number of cores.

When a coupled tool (e.g. OGS#IPhreeqc) is applied for reactive transport modeling, communication overhead comes mainly from two parts: the first part is related to the domain decomposition (DDC) based calculation (for processes such as flow and transport); whereas the second part is the overhead introduced from the coupling interface. In a conventional approach, the same number of cores is employed for DDC related calculations and geochemical reactions (i.e.  $n = DDC$ ), so that the time consumption for both parts will increase as the growth of computer cores. In the current study, a new approach is implemented, in which different number of compute cores can be assigned for the DDC related calculations and geochemical reactions. Because in the SNIA, geochemical reaction (calculated locally on each node) has a much better parallel performance than flow and transport (DDC based calculation), all the cores will take part in the calculation of geochemical reactions, whereas only certain number of cores will be allocated for flow and transport. The number of cores for flow and transport can be fixed at its optimum, so as to avoid the degradation of parallel performance from the DDC part.

The benefit and scalability of the new approach depend mainly on two factors: i) the proportion of time required for chemical reactions; ii) the efficiency of the coupling interface. Three "synthetic" scenarios are presented in Fig. 14.

First scenario ("best case") refers to geochemically dominated problems, in which the communication overhead from the interface is

neglectable when compared with that from DDC part. In this scenario the new approach can bring much better parallel performance and also be scalable for large number of compute cores. In the second scenario, geochemical reaction is still the most time-consuming part and the communication overhead of the interface is much smaller than that of the DDC part. Nevertheless, it is not neglectable anymore. In the third scenario ("worst case"), the overhead of the interface is in the similar magnitude of that of the DDC part. Additionally, the time consumption for flow/transport is comparable with that for reactions. As shown in Fig. 14, if the coupling interface is less efficient, the new approach gradually becomes less scalable, especially for flow/transport dominated problems.

### 5.3 IMPLEMENTATION

In the current study, the message passing interface (MPI) is applied to parallelize the computational tasks of the coupled code. the parallelization of flow/transport is realized through the internal DDC scheme of OGS (Wang et al. 2009), with which the computational tasks of global assembly and linear solver is parallelized; whereas the calculation of geochemical reaction is parallelized by using a loop, in which the computational task of reactions will be distributed into different compute cores.

The general concepts of the parallelization scheme are demonstrated in Fig. 15. In order to allocate different number of compute cores for flow/transport and reactions, two MPI groups (i.e. *Group\_DDC* and *Group\_IPQC*) are created (using MPI function *MPI\_Group\_incl*) which have different members of compute cores. Two relevant intra-communicators (i.e. *Comm\_DDC* and *Comm\_IPQC*) are created as well (by using MPI function *MPI\_Comm\_create*) for the communication among compute cores that belong to the same MPI group. The default communicator *MPI\_COMM\_WORLD* (created after the initialization of MPI environment) is used for communications among the cores from both MPI groups. It is possible to create only one MPI group i.e. *Group\_DDC*, thus all the cores will take part in solving flow/transport and reactions. In this case, the current scheme is identical to the conventional approach, no inter-communication among different MPI groups is required.

Fig. 17 illustrates the tasks that are performed within each MPI group during the entire simulation. Selection statements are used to constrain the processors of a certain MPI group to run a speci-

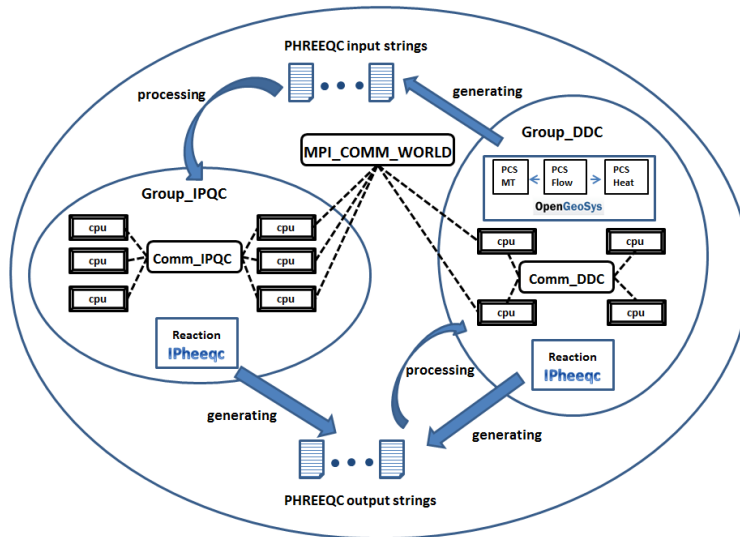


Figure 15.: The concept of MPI grouping and communication of the parallelization scheme (modified from He et al. 2015b).

fied piece of code. Whereas the cores of *Group\_DDC* will perform the execution of the most part of the code, those of *Group\_IPQC* will be constrained in a reaction loop which mainly deals with IPheeq calculations. During each time step, the processors of *Group\_IPQC* will first calculate the flow and transport processes (e.g. heat, mass), followed by the preparation of input strings for IPheeq. A big difference between the serial and the parallel simulation occurs here. Whereas only one input string is generated during each time step in the serial simulation; several input strings are prepared during the parallel run, each of which includes the geochemical information for those mesh nodes being solved on a certain compute core. It is worth mentioning, that flexible grouping of nodes from the whole mesh can be realized here, which is completely independent from the grouping of nodes during domain decomposition. This has a big advantage to ensure an optimum load balance, especially for the simulation of reactive transport problems with sharp fronts.

Communication among the two MPI groups is required when all the input strings are prepared and *Group\_IPQC* do exist. Each compute core from *Group\_IPQC* will receive one input string from *Group\_DDC* to start the calculation of chemical reactions. In the meantime, each compute core from *Group\_IPQC* will execute one IPheeq input string as well.

After the IPheeq calculations are done in both MPI groups, compute cores of *Group\_IPQC* will collect the results (as output

## 5.4 COMPUTATIONAL PLATFORM

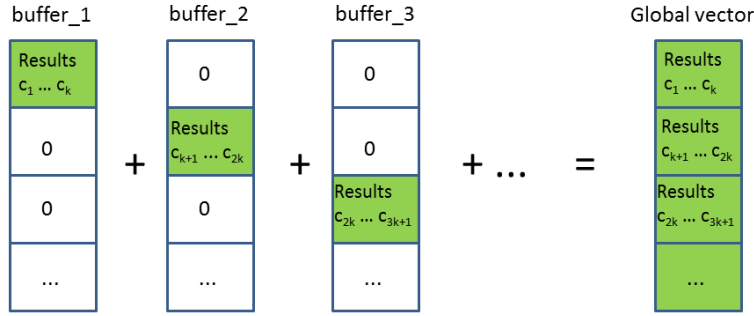


Figure 16.: Collection of the concentration values from the local buffers to a global concentration vector using *MPI\_Allreduce* operation.

string) from both groups and store them in several local buffers. Then values of state variables (such as concentration or temperature) from all local buffers will be gathered into a global concentration vector to update concentrations of different components for the mass transport by using the MPI operation *MPI\_Allreduce*, as is illustrated in Fig. 16. With this method, all the compute cores in *Group\_DDC* will participate in the gather operation. Nevertheless, a more sophisticated approach should be to reduce the memory usage and inter-processor communication. The calculation of flow and transport processes will restart in a new time step, while compute cores of *Group\_IPQC* will stay "idle" until they receive new IPhreeqc input strings to invoke the calls to IPhreeqc again (enabled by using a blocking communication with *MPI\_Receive*).

After simulations for all time steps are gone through, compute cores of *Group\_IPQC* will get a killing signal from *Group\_DDC* to quit the reaction loop. In the end, MPI environment will be terminated by using MPI function *MPI\_Finalize*.

## 5.4 COMPUTATIONAL PLATFORM

Two different platforms i.e. "ENVINF" and "EVE" were applied for the performance tests of the parallelization scheme, which are located at the Helmholtz Centre for Environment (UFZ), Leipzig. "ENVINF" is a multi-core Linux machine with 40 Intel(R) Xeon(R) CPU cores (E5-2680 v2 @ 2.80 GHz) and a shared memory of around 500GB. "EVE" is a cluster with 1008 Intel XEON CPU cores (X5650 @ 2.6 GHz) and a RAM of 5.5 TB. Computer nodes are connected with a 40 Gbits<sup>-1</sup> QDR InfiniBand network. The peak performance of "EVE" is 10 TFLOPs<sup>-1</sup>.

## 5.4 COMPUTATIONAL PLATFORM

```
if (Group_DDC != MPI_UNDEFINED)//ranks of Group_DDC will run following code
{
  read OGS input data;
  while (the current time is smaller than the end time) //time stepping loop
  {
    compute flow processes;
    compute heat and mass transport process;
    prepare the input strings for PHREEQC;
    if Group_IPQC exists
    {
      send start signal to Group_IPQC;
      //inform Group_IPQC that the input strings for IPhreeqc are prepared;
      send input strings to Group_IPQC;
    }
    calculate chemical reactions with IPhreeqc;
    if Group_IPQC exists
      receiving result strings from Group_IPQC;
    handle result strings and update data for mass transport;
  }
  print out results; //if needed
  send Group_IPQC a kill signal;
  //inform Group_IPQC the time stepping loop is over
  terminate MPI environment;
}

if (Group_IPQC != MPI_UNDEFINED)//if ranks of Group_IPQC exist
{
  for () //reaction loop
  {
    waiting for signal from Group_DDC;
    if the signal is a kill signal
      jump out of the reaction loop;
    else
      receive input strings from Group_DDC;
      calculate chemical reactions with IPhreeqc;
      send result strings to Group_DDC;
      //inform Group_DDC that calculation of reaction is done
  }
  terminate MPI environment;
}
```

Figure 17.: Pseudo code for the presentation of the tasks in both MPI groups (modified from He et al. 2015b).

In order to minimize the influence of other submitted jobs on the performance tests, job submission control is applied to guarantee the use of fully free compute nodes (20 free slots). Job submissions can of course be performed without this control. Nevertheless, the parallel performance can become less efficient and varying in this case. Because on the one hand, a job may be distributed to more compute nodes to allow an earlier execution, which would require more inter-compute-nodes communication; on the other hand, the performance can also be interfered by other running jobs on the same compute nodes.

## 5.5 PARALLEL PERFORMANCE TESTS OF THE STRING-BASED COUPLING

The parallel scheme of the character-string based approach was tested by three examples differing with size, complexity of flow and chemical system. These results are presented in He et al. 2015b in great details.

### 5.5.1 *Isotope fractionation, 2-D*

This example is a simple extension of the 1-D benchmark introduced in Sect. 4.5.1 into 2-D. The model domain has a length of 876  $m$  and width of 100  $m$  (illustrated in Fig. 18a), which is discretized into 1200 uniform rectangular elements with a length of 7.3  $m$  and a width of 10  $m$ . Unlike the time discretization in the benchmark example, here 200 time steps with the same size are applied in total. A serial simulation on the ENVINF takes 578  $s$ , among which the chemical reaction accounts for 92.2 %.

This example is tested on the ENVINF with number of compute cores ranging from 4 to 20. A very good agreement of the results between the parallel and serial simulation was obtained, as demonstrated in Fig. 18b.

Fig. 19 shows parallel performance of the approach for simulation of this test example. Although the problem size is small, relative good parallel performance is achieved. The 3-D plot (Fig. 19 a) demonstrates the relative speedup (in comparison to 4 cores with 4 DDCs) as a function of DDCs and the number of cores. Curve *AB* represents the relative speedups for those combinations with the same number of cores and DDCs, which can be regarded as the results for conventional DDC approach (see Sect. 5.2). For a fixed number of DDCs from 4 to 18, a continuous growth of speedup can be observed with the adding of compute cores. For a fixed

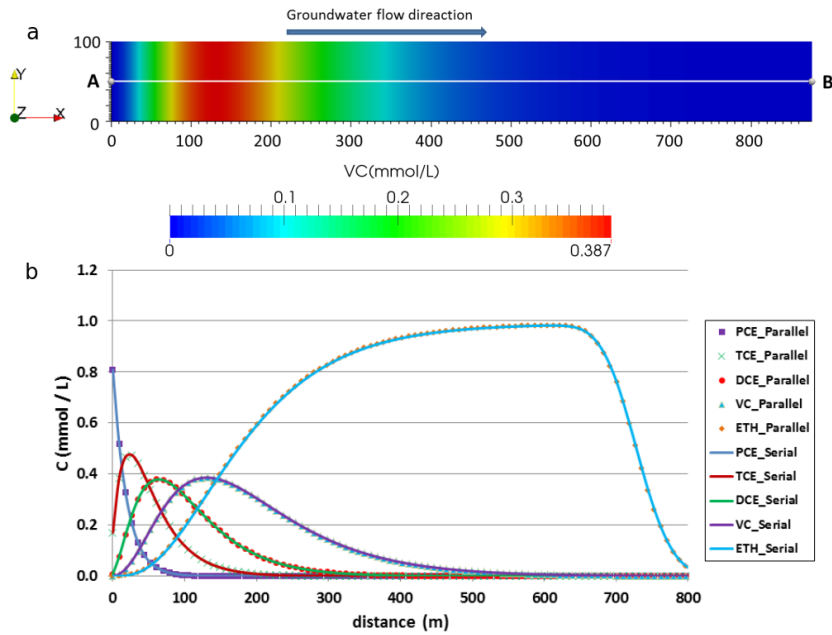


Figure 18.: Comparison of the results using serial and parallel simulation for isotope fractionation 2-D example. **(a)** Geometry of the test example and the concentration profile of light isotope VC by using parallel simulation; **(b)** Comparison of the concentration profiles of CHC light isotopes along the horizontal line in (a) by using serial and parallel simulation



number of cores, however, the increase of the speedup by using more DDCs is not significant anymore after applying 8 DDCs. In Fig. 19 b, the speedup curve "AB" in Fig. 19 a is presented together with the ideal speedup, and the speedup of IPhreeqc calculation and flow/transport against the number of DDCs. As one can see, the IPhreeqc calculation has generally rather good speedup from 4 to 20 cores; whereas the parallel performance for solving flow/transport degrades slightly after 18 cores. In Fig. 20, the parallel efficiency for solving flow/transport and IPhreeqc are plotted together with the proportion of the border nodes among the total nodes as the increase of DDCs. As one can see, the parallel efficiency for calculating flow/transport degrades dramatically to 0.5 when 20 DDCs are applied; whereas that for IPhreeqc calculation decreases only slightly and stays above 0.8. The proportion of sub-domain border nodes grows from around 3 % to around 17 % as the number of DDCs increases from 4 to 20.

In this example, because the portion of time for flow/transport among the total time is rather small when compared with that of reactions, its influence on the total speedup is rather small.

In Fig. 19 c and Fig. 19 d, the total time and its components (IPhreeqc calculation, interface and flow/transport) for  $DDC = 4$  (In Fig. 19 c) and  $DDC = 12$  (In Fig. 19 d) are plotted against the number of cores. In both cases, IPhreeqc calculation accounts for the largest portion of the simulation time. For  $DDC = 4$  it takes from 57.2 % (20 cores) to 86.5 % (4 cores) of the total time; whereas for  $DDC = 12$  the range becomes 73.1 % (20 cores) to 80.5 % (12 cores). The time consumption for the interface mainly includes preparation of input strings, processing of the result strings, as well as the communication among different cores. It takes on average 5.2 % and 10.8 % of the total time for  $DDC = 4$  and  $DDC = 12$ , respectively.

### 5.5.2 Isotope fractionation, 3-D

This example is a further extension of the 2-D test case (Sect. 5.5.1) into 3-D (876 m \* 100 m \* 10 m). The model domain is discretized into 120 000 hexahedral elements (120\*100\*100) with 134431 nodes. Same time discretization was applied as the 2-D example. It takes around 37.5 h for a complete simulation with 2 cores and 2 DDCs on ENVINF.

Further tests were made on the EVE cluster, in which the number of compute cores and DDCs vary from 20 to 80. Results of the parallel performance are illustrated in Fig. 21. Fig. 21a shows

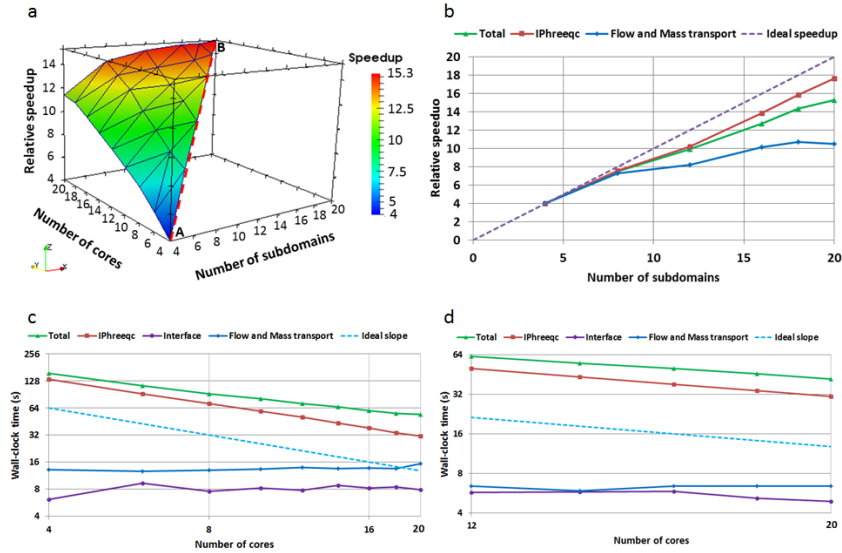


Figure 19.: Performance of the parallelization scheme for simulation of isotope fractionation 2-D example on ENVINF (from He et al. 2015b). **(a)** Relationship between number of DDCs, number of compute cores and relative speedup in comparison to a simulation with 4 cores and 4 DDCs (Color legend shows the value of relative speedup); **(b)** breakdown of the speedup curve AB (marked as dashed line in **a**) into speedup of calculation of chemical reaction i.e. IPhreeqc and flow and mass transport; **(c)** breakdown of the total time for chemical reactions, interface and flow and transport for DDC = 4; **(d)** breakdown of the total time for DDC = 12.

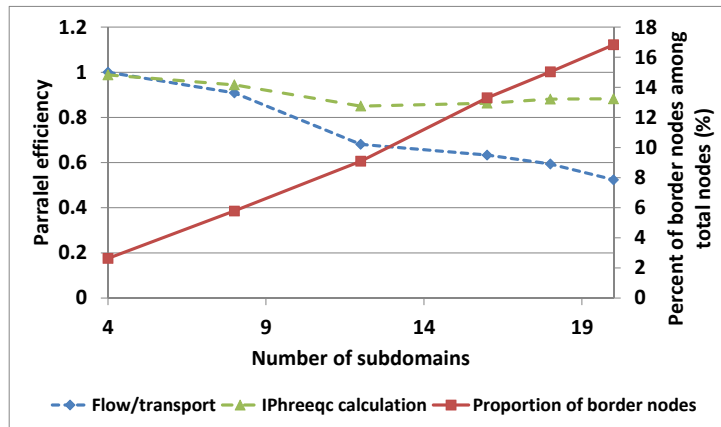


Figure 20.: The parallel efficiency of the conventional approach to calculate flow/transport and IPhreeqc as a function of sub-domains.

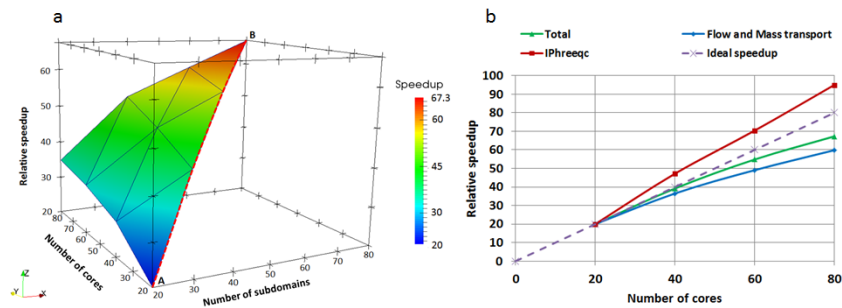


Figure 21.: Performance of the parallelization scheme for simulation of isotope fractionation 3-D example on EVE (from He et al. 2015b). **(a)** Relationship between number of DDCs, number of compute cores and relative speedup in comparison to a simulation with 20 cores and 20 DDCs (Color legend shows the value of relative speedup); **(b)** breakdown of the speedup curve AB (marked as dashed line in **a**) into speedup of calculation of chemical reaction i.e. IPhreeqc and flow and mass transport.

the relative speedup (compared with the simulation with 20 cores and 20 DDCs on EVE) achieved by using different combinations of compute cores and DDCs. Fig. 21 b shows the parallel performance by using the conventional approach (i.e. the number of DDC and cores are the same). The overall speedup (curve "AB" in Fig. 21 a) is presented together with the speedup for IPhreeqc as well as flow/transport calculation. Near linear speedup behaviors are observed for flow/transport and the total speedup. IPhreeqc calculation shows even a super-linear speedup.

For this example, the conventional approach has shown to be rather efficient, which makes the best use of the compute cores employed. By using a 80 cores, the total simulation time was reduced to around 37 min.

The new approach (i.e. using different number of compute cores and DDCs) is less efficient here. As shown in Fig. 21 a), the further increase of the compute cores for a fixed number of DDCs does not bring significant growth of speedup, which is not the case in the 2-D test example.

As problem size becomes larger, time consumption for calculation of flow/transport becomes comparable with that for reactions. By using the combination of 20 cores with 20 DDCs, calculation for flow/transport and IPhreeqc take 36.2 and 54.3% of the total time, respectively. In this case, it is beneficial to use the same number of compute cores and DDCs, in order to accelerate both time consuming calculations.

### 5.5.3 *Uranium leaching benchmark*

This 2-D test example is introduced in Sect. 4.5.3. One simulation of this example on the ENVINF takes around 6.0 h with 2 cores and 2 DDCs. The other tests introduced below were performed on the EVE cluster with combinations of compute cores from 20 to 60 and DDCs from 2 to 60.

Fig. 22 shows the relative speedup for simulating this example against the number of cores and DDCs. Unlike the first two examples, the conventional approach (illustrated by the curve "AB") scales poorly for this example. A strong decrease of the parallel performance occurs when 20 cores and 20 DDCs are employed. There is a slight recovery of the parallel performance when more than 20 DDCs are applied. The new approach shows big advantages over the conventional one in this example. The best performance is obtained when 60 cores and DDCs ranging from 8 to 16 are applied.

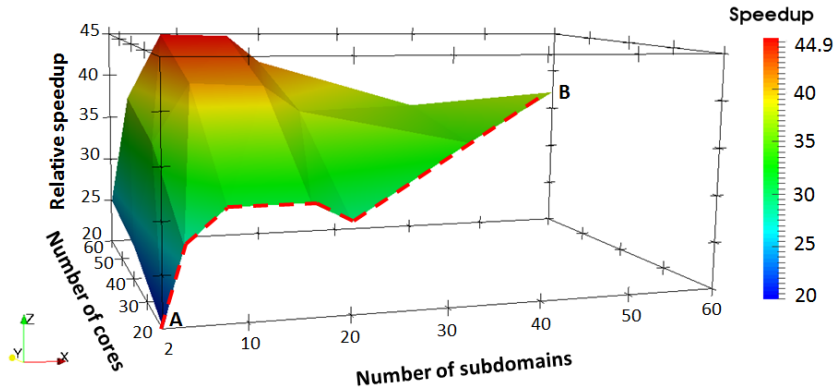


Figure 22.: Relative speedup of the simulation the uranium leaching example as a function of number of DDCs and compute cores in comparison to a simulation with 20 cores and 2 DDCs (from He et al. 2015b).

These speedup behaviors described above can be better explained by breaking down the total simulation time into different portions such as time for linear solver, matrix assembly, IPhreeqc calculation as well as the interface, which is illustrated in Fig. 23. While the time for other portions decrease continuously after 20 cores (DDCs), the time for linear solver increases dramatically after applying 20 DDCs, which leads to the observed overall performance degradation. After 40 DDCs, the increase of time for linear solver slows down, while the time for other portions decreases linearly as the growth of the number of DDCs. As a consequence, the total simulation time goes down slightly.

Because the same number of cores i.e. 20 are applied for DDCs from 2 to 20, the time for IPhreeqc calculation stays almost unchanged before 20 DDCs. One lesson we can learn from this example is, that the interface can become time consuming when a small number of DDCs is combined with a large number of compute cores. In the case, a limited number of cores have to prepare and process a plenty of in- and output strings. For example, when 20 cores are applied with only 2 DDCs, the interface accounts for 23.4 % of the total simulation time.

## 5.6 COMPARISON OF THE STRING-BASED AND FILE-BASED APPROACH

The parallel performance of the file-based approach was tested for test example 1 (5.5.1) and 2 (5.5.2). Detailed results and discussion can be found in He et al. 2015a. Here, the comparison of

## 5.7 SUMMARY

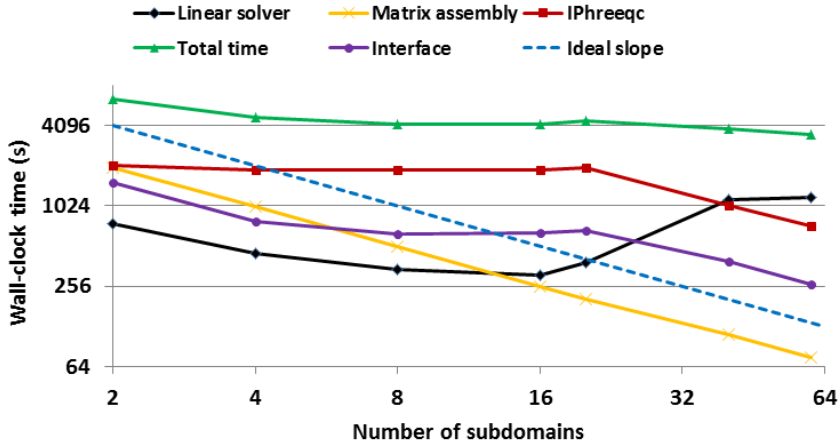


Figure 23.: Breakdown of the total time corresponding to curve AB in Fig. 22 (from He et al. 2015b). 20 cores are always applied for DDCs from 2 to 20; for DDCs over 20, the same number of DDCs and cores are employed.

the parallel performance of file-based and string-based coupling interface for simulating test example 2 will be discussed. The results are illustrated in Fig. 24.

There are no big difference of the time consumption for flow/transport and IPhreeqc calculation by using these two approaches (see Fig. 24 a and b). However, the time consumption for interface are quite different in these two schemes (see Fig. 24 c). The string-based coupling is generally much less time consuming than the file-based coupling. One important reason is that file reading and writing in EVE cluster is realized through the general parallel file system (GPFS), which can be especially inefficient when the GPFS is highly loaded by other running tasks. For a fixed number of compute cores, the time for the interface decreases as more DDCs are given. However, for a fixed number of DDCs, this part of time increases significantly for file-based coupling when more cores are added. When a simulation is performed with 80 cores and 20 DDCs, the interface takes up to 30.9% of the total time with the file-based coupling; whereas that for the string-based coupling is only 10.2%. As a result, the string-based coupling scales much better than the file-based one, which is demonstrated in Fig. 24d.

## 5.7 SUMMARY

This chapter introduces the idea, the implementation as well as the performances of the novel parallelization scheme from He et al.

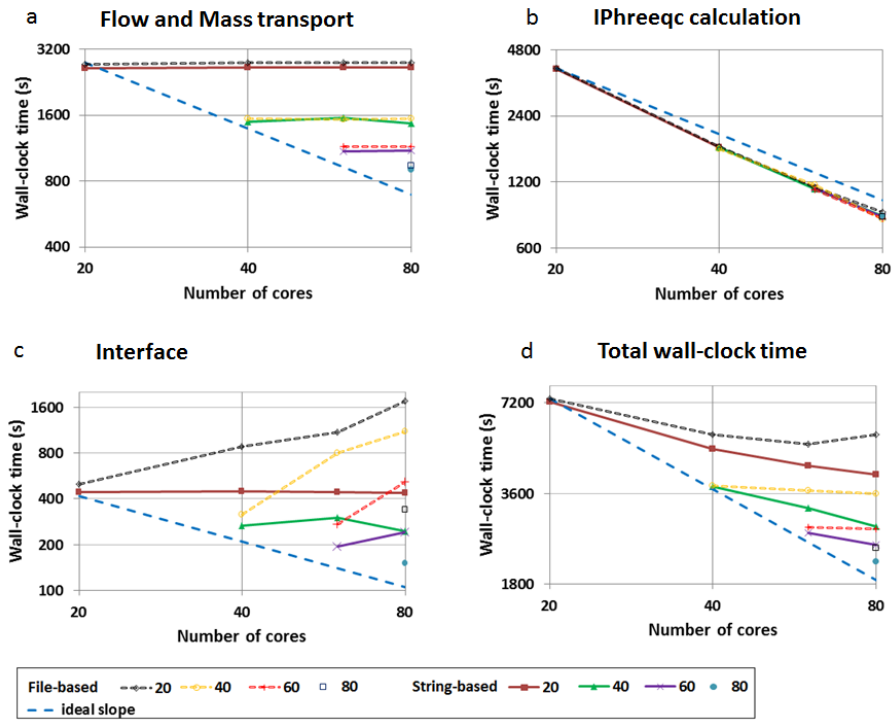


Figure 24.: Comparison of the parallel performance of the string-based and file-based coupling interface to simulate the test example "isotope fractionation 3D" for DDCs from 20 to 80 (from He et al. 2015b). (a) mass transport and flow; (b) geochemical reaction (IPhreeqc); (c) OGS#IPhreeqc interface; (d) total wall-clock time.

2015b. In the SNIA, the optimum number of the required compute cores for DDC-related processes (i.e. flow/transport) can be quite different from that for chemical reactions. Based on this fact, a new approach is developed to enable a flexible allocation of computational resources for solving geochemical and non-geochemical processes.

It is problem dependent, whether the new approach can deliver more advantages over the conventional one, in which the number of cores for flow/transport is the same as that for chemical reactions.

If a problem is dominated by flow/transport, then adding more cores for chemistry alone will not bring significant benefit, if further speedups for flow/transport by using more DDCs still presents. This is especially the case for large scaled problems with simple chemical systems, as demonstrated in Sect. 5.5.2.

If a problem is dominated by chemical reaction, then the new approach can be more advantageous than the conventional one, especially when a severe degradation of parallel performance for flow/transport occurs. In this case, further speedups can still be obtained by adding more cores for chemistry while keep the number of DDCs at the optimum.

The character-string-based coupling interface is shown to be much more efficient than the file-based one, which is shown in Sect. 5.6. Nevertheless, to reduce the computational overhead of the interface still remains an important task. Because it is one of the critical factors, which determine the scalability of the new approach for large amount of compute cores (see Sect. 5.2). One promising solution would be to use an "in-memory" coupling, so that the internal data structures of both OGS and IPhreeqc can be accessed more straightforwardly.



# 6

---

## APPLICATION OF THE TOOL FOR A CASE STUDY

---

In this chapter, the parallelized reactive transport simulator OGS#IPhreeqc is applied to simulate the water dynamics and the natural attenuation of nitrate in the groundwater aquifer of a study site in Northern Germany. The site is introduced first, followed by the description of model setups. Then simulation results as well as parallel performance of OGS#IPhreeqc will be presented and discussed.

### 6.1 STUDY SITE

The site ( $51^{\circ}43'N$   $11^{\circ}18'E$ ) is a section of the Selke River catchment, which is located at the foreland of Harz mountains, Northern Germany. This area is dominated by agricultural activities. Several investigations have been taken place at this site as a part of the TERENO (Terrestrial Environmental Observatories) program. These includes geophysical surveys, tracer tests, and so on. The Selke river, its surrounding flood plains and hill slopes are taken into account for modeling studies described below.

### 6.2 MODEL SETUP

A 3-D model was setup based on the digital elevation model (DEM) and geo-electrical profiles of the site. As the first step, a 2-D surface mesh was generated based on the DEM and the geometry of the Selke river (see Fig. 25 a). Then it was extruded into 3-D with 20 layers in vertical direction (see Fig. 25 b). The elevation of the bottom of the model domain is fixed as 150 m. The mesh consists of 258020 triangular prism elements and 138054 finite element nodes.

In the second step, the permeability field was created for the model domain. Based on the measured profiles of electrical resistivity (ER), the ER values were first extrapolated to the whole model do-

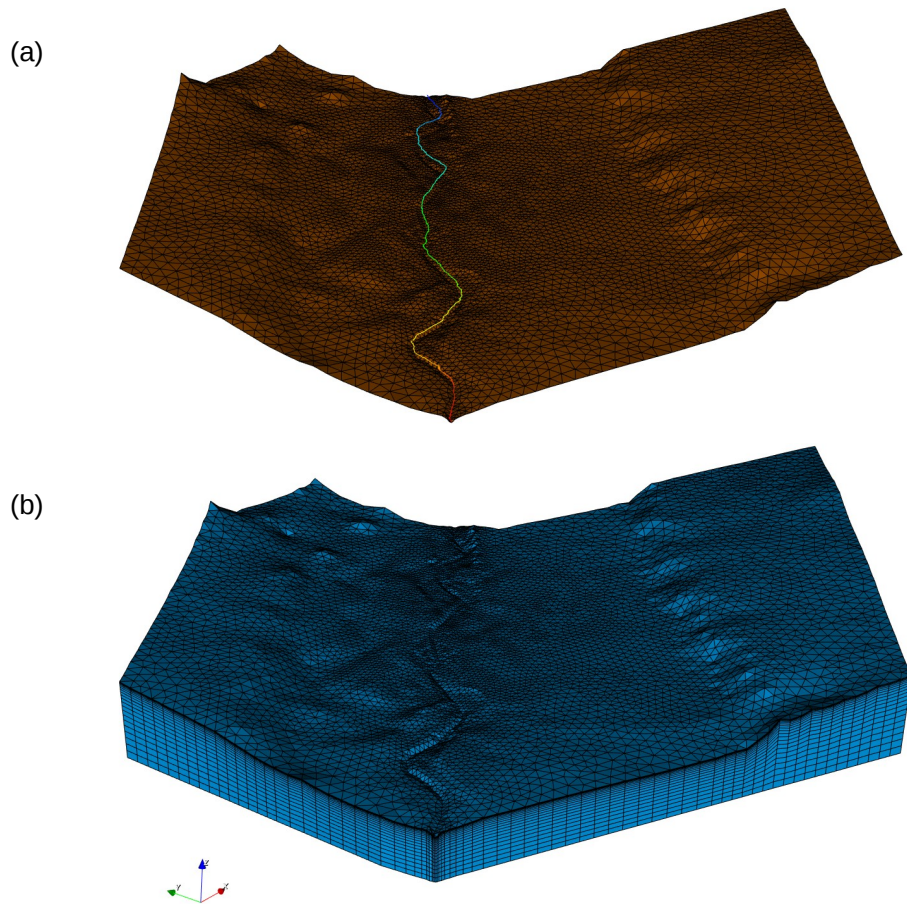


Figure 25.: Mesh generation for the Selke "book" model. **(a)** The surface mesh and the geometry of Selke river; **(b)** The 3-D mesh.

## 6.2 MODEL SETUP

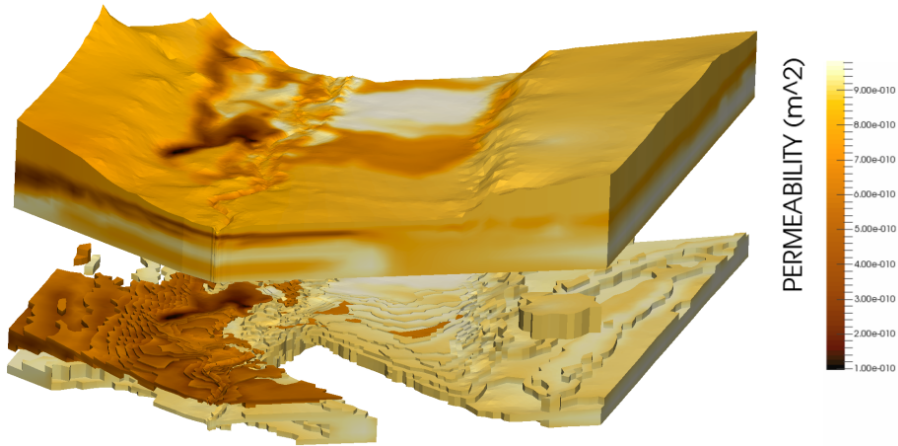


Figure 26.: The permeability distribution of the model domain.

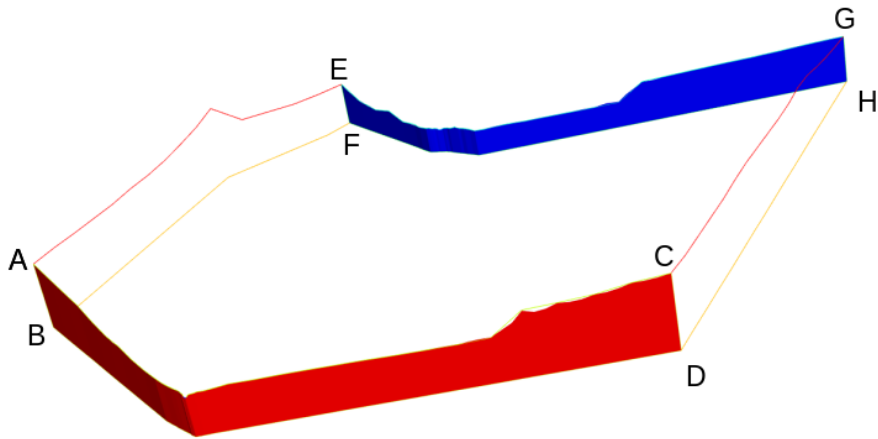


Figure 27.: The geometry and boundary condition.

main. Based on the ER values and the known minimum ( $10^{-10} m^2$ ) as well as the maximum ( $10^{-9} m^2$ ) permeability in the study site, a linear function was deduced which was used to calculate the permeability values by using the ER values. The calculated permeability field of the model domain is shown in Fig. 26, in which less permeable regions are illustrated separately below the complete model domain.

In the third step, a groundwater flow model was built. A hydraulic head of 160 m is given as the initial condition for the entire domain. The Dirichlet boundary condition is applied for both the Southwest (surface ACDB in Fig.27) and Northeast (surface EGHF in Fig.27) boundary of the model. The hydraulic head on these two boundaries are 167.92 and 162 m, respectively.

In the fourth step, a mass transport and reduction of nitrate by organic carbon was modeled. Constant concentration of ni-

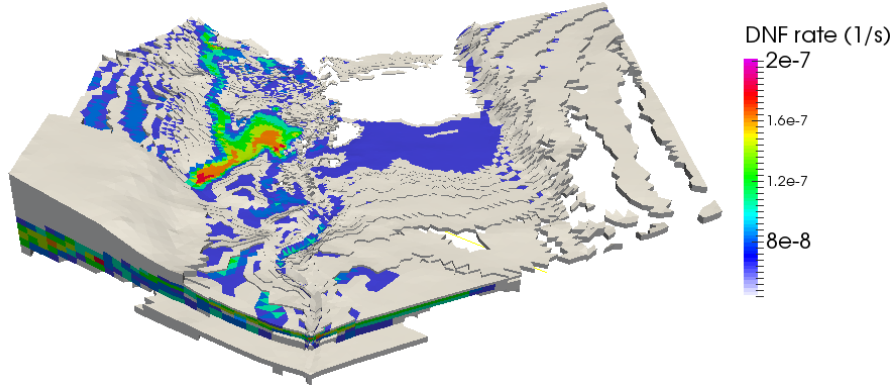


Figure 28.: The distribution of first-order denitrification rate (larger than  $4.2\text{E-}8 \text{ s}^{-1}$ ) in the model domain.

trate ( $4.8\text{E-}4 \text{ mol} \cdot \text{l}^{-1}$ ) is imposed on the Southwest boundary of the model, which is then flushed towards northeast along the flow field. Denitrification (DNF) is simulated as kinetic reaction with a rate law given in Eq. 54.

$$\frac{dC_{Nit}}{dt} = -K_{denit} \cdot f_{TOC} \cdot C_{Nit} \quad (54)$$

where  $C_{Nit}$  is the concentration of nitrate;  $K_{denit}$  is the rate constant for DNF ( $2\text{E-}7 \text{ s}^{-1}$ );  $f_{TOC}$  is a dimensionless factor determining the potential of denitrification by organic carbon, which is related to the availability and activity of total organic carbon (TOC).

It is assumed, that less permeable soil has a higher  $f_{TOC}$  value and thus a higher potential for nitrate reduction. Based on this assumption, the model domain is clustered into 11 material groups with different  $f_{TOC}$  ranging from 0.01 to 1.0. Material group, which has a  $f_{TOC}$  of 1.0, has the highest DNF rate i.e.  $2\text{E-}7 \text{ s}^{-1}$ ; whereas a  $f_{TOC}$  of 0.01 refers to the regions with lowest DNF rate i.e.  $2\text{E-}9 \text{ s}^{-1}$ . Fig. 28 illustrates the model regions with DNF rate larger than  $4.2\text{E-}8 \text{ s}^{-1}$ .

### 6.3 RESULTS AND DISCUSSION

The simulation was performed at the EVE cluster with a simulation time of 578 days. By using 80 cores and 80 DDCs, the simulation took around 1 hour. The simulation results are shown in Fig. 29 and Fig. 30. The velocity field of groundwater in the model domain at steady state is illustrated in Fig. 29, which has a strong correlation with the permeability distribution in the model domain: groundwater flows faster in high permeable regions, and has

### 6.3 RESULTS AND DISCUSSION

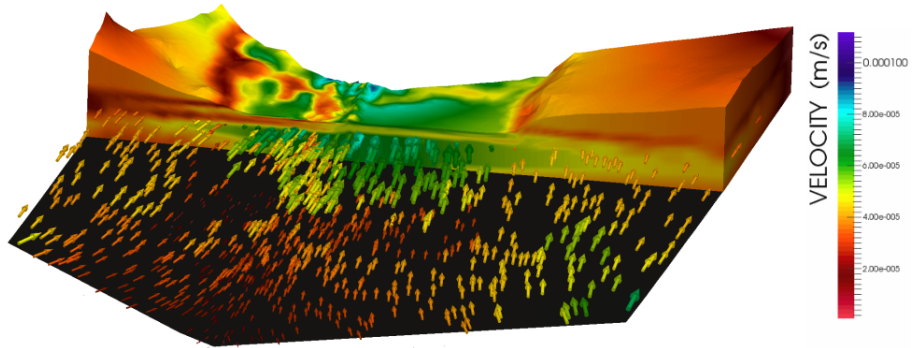


Figure 29.: The velocity distribution of groundwater flow in the model domain at steady state.

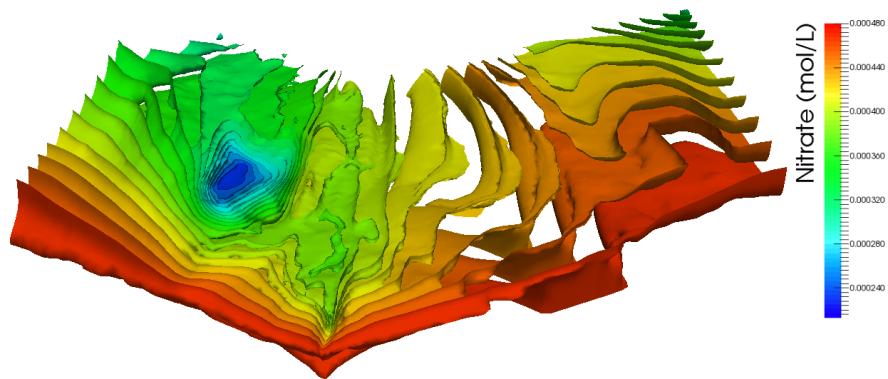


Figure 30.: The concentration contour maps of nitrate after 578 days.

a longer residence time in low permeable materials. The concentration contour maps of nitrate after 578 days is shown in Fig. 30. One can observe a clear sink of nitrate in the model domain, where the nitrate concentration is lower than  $3\text{E-}4 \text{ mol} \cdot \text{l}^{-1}$  (the dark blue region). The low nitrate concentration in this location is mainly caused by two reasons. On the one hand, this region has a higher DNF potential; on the other hand, nitrate has a longer residence time because of the low permeability in this region.

#### 6.4 SUMMARY

This chapter introduces the setup and simulation results of a 3-D model based on a case study. Parallel simulation was performed by using the OGS#IPhreeqc in the EVE cluster. As the available data is limited, only simple flow regime and geochemical system are employed.

Nevertheless, the model considers the real topography and heterogeneity of the site based on the DEM and ER measurements. The model results demonstrates the importance of functional zones for natural attenuation process such as DNF.

The chapter presents the general steps to setup a 3-D reactive transport model at the hillslope scale. Furthermore, it demonstrates that OGS#IPhreeqc is able to simulate such kinds of large scaled problem within acceptable time span, which paves the way for performing uncertainty and sensitivity analysis for such kinds of models in the future work.

---

## CONCLUSIONS AND OUTLOOK

---

### 7.1 CONCLUSIONS

The current work introduces several efforts towards reactive transport modeling at a hillslope scale, which includes:

- ◆ analysis and comparison of different numerical approaches to simulate water flow in saturated and unsaturated flow;
- ◆ development and benchmarking of a character-string based coupling interface between the software packages OpenGeoSys and IPhreeqc;
- ◆ development and analysis of a parallelization scheme using MPI grouping techniques for the coupled tool OGS#IPhreeqc;
- ◆ application of the tool for a case study.

The main findings and achievements of this work are summarized as follows:

- ❖ the mixed-form of Richards equation can conserve mass more accurately than the h-based form, especially when automatic time stepping is applied for the simulation. Additionally, the mixed form can be applied to simulate water flux in both saturated and unsaturated zones;
- ❖ the OGS#IPhreeqc coupling interface can greatly benefit from the broad spectrum of geochemical capabilities and the customizable thermodynamic database from PHREEQC. Moreover, the coupling interface is version-independent (sustainable concept), so that the code developments from both open-source communities can be integrated easily;
- ❖ the character-string-based data exchange between OGS and IPhreeqc is much more efficient than the file-based one for parallel simulations on a distributed-memory system like cluster, in which file reading and writing have to be realized through the GPFS;
- ❖ the parallelization scheme developed in the current work can be applied in different types of HPC platforms such as clusters or shared memory machines. Moreover, it provides more flexi-

## 7.2 OUTLOOK

bility than the conventional approach to allocate computational resources for different computational tasks. The optimal setting of the number of DDCs and total compute cores is problem dependent.

❖ The new approach is advantageous for geochemically dominated problems (e.g. for small- to medium-sized problems with complex chemistry), especially a further increase of the number of DDCs above the optimum will lead to a strong performance degradation for flow/transport. In this case, one can fix the number of cores for flow/transport at its optimum to avoid the performance degradation resulted from this part; while add more cores for calculating chemical reaction.

❖ If the time consumption for flow/transport is of the same magnitude as geochemical reactions (e.g. for large scaled problems where the time for flow/transport becomes more dominant), and the calculation of flow/transport can be further accelerated with the adding of compute cores, then the conventional approach is preferred.

❖ the parallelized OGS#IPhreeqc developed in this work is able to simulate heterogeneous reactive transport problems at a hillslope scale within an acceptable amount of time.

## 7.2 OUTLOOK

In the current study, the test examples as well as the case study are simple regarding to flow regimes and geochemical systems. There are still much work to be done in the future work, in order to simulate "real-world" reactive transport problems at larger scales.

❖ The development and implementation of a more robust and efficient approach is required to reduce or even avoid the non-linearity in solving water flow in unsaturated zone. This can be for example a simplification of Richards equation based on reasonable assumptions, such as the kinematic -wave approximation of Richards equation applied in the UZF package for MODFLOW (Niswonger et al. 2006). Another promising way could be to find alternative approaches, which possess similar capability as Richards equation but have better numerical convergence and stability. A possible good solution is suggested by the work of Ogden et al. 2015, in which a set of ordinary differential equation (ODE) is applied to solve the general unsaturated flow problems.

❖ The SNIA is applied for the coupling between OGS and IPhreeqc.



The presented examples show the sufficient accuracy of this approach. Nevertheless, the implementation of a SIA would be a necessary step in order to deal with problems, in which geochemical reactions can have strong feedbacks on flow and mass transport.

✧ The computational overhead of the coupling interface should be further minimized to enable this approach to have a good scalability for parallel computing with a large number of processors. A promising solution would be to use an in-memory coupling, so that the sharing of the internal data structures of both software packages can be maximized to avoid duplication of mapping of data.

✧ With the presented parallelization scheme, it is possible to avoid the degradation of parallel performance for solving flow and mass transport. However, if a problem is dominated by flow/transport and the maximum speedup for this part of calculation is limited, then the benefit of this approach will be rather small. As a result, it is an important future work to improve the parallel performance for the calculation of flow and mass transport processes.

✧ Due to limited data availability, only simple flow regime and geochemical system are considered in the model for the Selke case study. Furthermore, the assumptions, which are made during the model setup, are needed to be verified or corrected with the help of measurements conducted directly on site. In order to get a better process understanding and a better identification of functional zones in the site, monitoring and modeling activities should be combined and conducted in an iterative way, so that both of them can benefit from each other stepwise. On the one hand, one can use modeling to identify the potential "hot-spots" for sampling and the interesting parameters for performing measurements based on uncertainty and sensitivity analysis. On the other hand, model parameters can be adjusted or optimized with the help of inverse modeling, when more data are available.

---

## BIBLIOGRAPHY

---

- Appelo, C.A.J. and D. Postma (2005). *Geochemistry, Groundwater and Pollution, Second Edition*. CRC Press.
- Bailey, Ryan T., Eric D. Morway, Richard G. Niswonger, and Timothy K. Gates (2013). “Modeling Variably Saturated Multi-species Reactive Groundwater Solute Transport with MODFLOW-UZF and RT3D”. In: *Ground Water* 51.5, pp. 752–761.
- Bear, J. and Y. Bachmat (1986). “Macroscopic modelling of transport phenomena in porous media. 2: Applications to mass, momentum and energy transport”. In: *Transport in Porous Media* 1, pp. 241–269.
- Bear, J. and A. Verruijt (1987). *Modeling Groundwater Flow and Pollution*. Springer Netherlands, p. 414.
- Beyer, Christof, Dedong Li, Marco De Lucia, Michael Kühn, and Sebastian Bauer (2012). “Modelling CO<sub>2</sub>-induced fluid–rock interactions in the Altensalzwedel gas reservoir. Part II: coupled reactive transport simulation”. In: *Environmental Earth Sciences* 67.2, pp. 573–588.
- Brooks, R. H. and A. T. Corey (1966). “Properties of porous media affecting fluid flow”. In: *J. Irrig. Drain. Div. ASCE*, pp. 61–88.
- Celia, Michael A., Efthimios T. Bouloutas, and Rebecca L. Zarba (1990). “A general mass-conservative numerical solution for the unsaturated flow equation”. In: *Water Resources Research* 26.7, pp. 1483–1496.
- Centler, Florian, Haibing Shao, Cecilia De Biase, Chan-Hee Park, Pierre Regnier, Olaf Kolditz, and Martin Thullner (2010). “GeoSys-BRNS—A flexible multidimensional reactive transport model for simulating biogeochemical subsurface processes”. In: *Computers & Geosciences* 36.3, pp. 397–405.
- Charlton, Scott R. and David L. Parkhurst (2011). “Modules based on the geochemical model PHREEQC for use in scripting and programming languages”. In: *Computers & Geosciences* 37.10, pp. 1653–1663.
- Clement, T.P., William R. Wise, and Fred J. Molz (1994). “A physically based, two-dimensional, finite-difference algorithm for modeling variably saturated flow”. In: *Journal of Hydrology* 161.1, pp. 71–90.
- Darcy, H. (1856). *Les Fontaines Publiques de la Ville de Dijon*. Dalmont, Paris.

## Bibliography

- Davies, C.W. (1962). *Ion Association*. Butterworths, Washington DC.
- Engesgaard, Peter and Kenneth L. Kipp (1992). “A geochemical transport model for redox-controlled movement of mineral fronts in groundwater flow systems: A case of nitrate removal by oxidation of pyrite”. In: *Water Resources Research* 28.10, pp. 2829–2843.
- Hammond, G. E., P. C. Lichtner, and R. T. Mills (2014). “Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN”. In: *Water Resources Research* 50.1, pp. 208–228.
- Hammond, Glenn E. and Peter C. Lichtner (2010). “Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing”. In: *Water Resources Research* 46.9.
- He, W. (2015). “Groundwater Flow—Theis’ Revisited”. In: *Thermo-Hydro-Mechanical-Chemical Processes in Fractured Porous Media: Modelling and Benchmarking*. Ed. by Olaf Kolditz, Hua Shao, Wenqing Wang, and Sebastian Bauer. Terrestrial Environmental Sciences. Springer International Publishing, pp. 115–120.
- He, W., C. Beyer, J.H. Fleckenstein, E. Jang, O. Kolditz, D. Naumov, and T. Kalbacher (2015a). “A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IPhreeqc”. In: *Geosci. Model Dev. Discuss.* 8, pp. 2369–2402.
- (2015b). “A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IPhreeqc 5.5.7-3.1.2”. In: *Geosci. Model Dev.* 8, pp. 3333–3348.
- He, W., M. Walther, and O. Kolditz (2015c). “Benchmark: Theis Problem”. In: *OpenGeoSys-Tutorial*. SpringerBriefs in Earth System Sciences. Springer International Publishing, pp. 31–51.
- He, W., H. Shao, O. Kolditz, W. Wang, and T. Kalbacher (2015d). “Comments on “A mass-conservative switching algorithm for modeling fluid flow in variably saturated porous media, K. Sadegh Zadeh, *Journal of Computational Physics*, 230 (2011)””. In: *Journal of Computational Physics* 295, pp. 815–820.
- He, W., C. Beyer, J.H. Fleckenstein, E. Jang, O. Kolditz, D. Naumov, and T. Kalbacher (2015e). “Interactive comment on “A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IPhreeqc” by W. He et al.” In: *Geosci. Model Dev. Discuss.* 8, pp. C1627–C1627.

## Bibliography

- Henzler, Aline F., Janek Greskowiak, and Gudrun Massmann (2014). “Modeling the fate of organic micropollutants during river bank filtration (Berlin, Germany)”. In: *Journal of Contaminant Hydrology* 156, pp. 78–92.
- Hills, R. G., I. Porro, D. B. Hudson, and P. J. Wierenga (1989). “Modeling one-dimensional infiltration into very dry soils: 1. Model development and evaluation”. In: *Water Resources Research* 25.6, pp. 1259–1269.
- Hubschwerlen, Nicolas, Keni Zhang, Gerhard Mayer, Jean Roger, and Bernard Vialay (2012). “Using Tough2-MP on a cluster—optimization methodology and study of scalability”. In: *Computers & Geosciences* 45, pp. 26–35.
- Kavetski, D., P. Binning, and S.W. Sloan (2001). “Adaptive time stepping and error control in a mass conservative numerical solution of the mixed form of Richards equation”. In: *Advances in Water Resources* 24.6, pp. 595–605.
- Kolditz, O., S. Bauer, L. Bilke, N. Böttcher, J.O. Delfs, T. Fischer, U.J. Görke, T. Kalbacher, G. Kosakowski, C.I. McDermott, C.H. Park, F. Radu, K. Rink, H. Shao, H.B. Shao, F. Sun, Y.Y. Sun, A.K. Singh, J. Taron, M. Walther, W. Wang, N. Watanabe, Y. Wu, M. Xie, W. Xu, and B. Zehner (2012). “OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media”. English. In: *Environmental Earth Sciences* 67.2, pp. 589–599.
- Kosakowski, Georg and Norihiro Watanabe (2014). “OpenGeoSys-Gem: A numerical tool for calculating geochemical and porosity changes in saturated and partially saturated media”. In: *Physics and Chemistry of the Earth, Parts A/B/C* 70–71. Mechanisms and Modelling of Waste-Cement and Cement-Host Rock Interactions, pp. 138–149.
- Krabbenhøft, K. (2007). “An alternative to primary variable switching in saturated–unsaturated flow computations”. In: *Advances in Water Resources* 30.3, pp. 483–492.
- Kräutle, S. and P. Knabner (2007). “A reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions”. In: *Water Resources Research* 43.3, pp. 1–15.
- Lasaga, Antonio C., Josep M. Soler, Jiwchar Ganor, Timothy E. Burch, and Kathryn L. Nagy (1994). “Chemical weathering rate laws and global geochemical cycles”. In: *Geochimica et Cosmochimica Acta* 58.10, pp. 2361–2386.
- Lehmann, F. and Ph. Ackerer (1998). “Comparison of Iterative Methods for Improved Solutions of the Fluid Flow Equation

## Bibliography

- in Partially Saturated Porous Media”. In: *Transport in Pours Media* 31, pp. 275–292.
- Li, Dedong, Sebastian Bauer, Katharina Benisch, Bastian Graupner, and Christof Beyer (2014). “OpenGeoSys-ChemApp: a coupled simulator for reactive transport in multiphase systems and application to CO<sub>2</sub> storage formation in Northern Germany”. English. In: *Acta Geotechnica* 9.1, pp. 67–79.
- Lichtner, P. C., G. E. Hammond, C. Lu, S. Karra, G. Bisht, B. Andre, R. T. Mills, and J. Kumar (2015). *PFLOTRAN User Manual: A Massively Parallel Reactive Flow and Transport Model for Describing Surface and Subsurface Processes*. English.
- Mayer, K. Ulrich, Emil O. Frind, and David W. Blowes (2002). “Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions”. In: *Water Resources Research* 38.9. 1174, pp. 13–1–13–21.
- Meeussen, Johannes C. L. (2003). “ORCHESTRA: An Object-Oriented Framework for Implementing Chemical Equilibrium Models”. In: *Environmental Science & Technology* 37.6. PMID: 12680672, pp. 1175–1182.
- Molins, S., K.U. Mayer, R.T. Amos, and B.A. Bekins (2010). “Vadose zone attenuation of organic compounds at a crude oil spill site — Interactions between biogeochemical reactions and multi-component gas transport”. In: *Journal of Contaminant Hydrology* 112.1–4. *Frontiers in Reactive Transport: Microbial Dynamics nad Redox Zonation in the Subsurface*, pp. 15 –29.
- Morway, Eric D., Richard G. Niswonger, Christian D. Langevin, Ryan T. Bailey, and Richard W. Healy (2013). “Modeling Variably Saturated Subsurface Solute Transport with MODFLOW-UZF and MT3DMS”. In: *Ground Water* 51.2, pp. 237–251.
- Nardi, Albert, Andrés Idiart, Paolo Trincherro, Luis Manuel de Vries, and Jorge Molinero (2014). “Interface COMSOL-PHREEQC (iCP), an efficient numerical framework for the solution of coupled multiphysics and geochemistry”. In: *Computers & Geosciences* 69, pp. 10 –21.
- Nasir, Othman, Mamadou Fall, and Erman Evgin (2014). “A simulator for modeling of porosity and permeability changes in near field sedimentary host rocks for nuclear waste under climate change influences”. In: *Tunnelling and Underground Space Technology* 42, pp. 122 –135.
- Niswonger, R.G., D.E. Prudic, and R.S. Regan (2006). *Documentation of the Unsaturated-Zone Flow (UZF1) Package for modeling unsaturated flow between the land surface and the water table*

## Bibliography

- with MODFLOW-2005*. U.S. Geological Survey Techniques and Methods 6-A19. 62 pp.
- Ogden, Fred L., Wencong Lai, Robert C. Steinke, Jianting Zhu, Cary A. Talbot, and John L. Wilson (2015). “A new general 1-D vadose zone flow solution method”. In: *Water Resources Research* 51.6, pp. 4282–4300.
- Palandri, J.L. and Y.K. Kharaka (2004). *A compilation of rate parameters of water-mineral interaction kinetics for application to geochemical modeling*. U.S. GEOLOGICAL SURVEY OPEN FILE REPORT (OF 2004-1068).
- Parkhurst, D. L. and C. A. J. Appelo (1999). *User’s guide to PHREEQC (Version 2) – a computer program for speciation, batch-reaction, one-dimensional transport and inverse geochemical calculations*. US Geological Survey Water-Resources Investigations Report, pp. 99–4259. 312 pp.
- (2013). *Description of input and examples for PHREEQC version 3 – a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations, in: US Geological Survey Techniques and Methods, book 6, chap. A43*. US Geological Survey. 497 pp.
- Sadegh Zadeh, Kouroush (2011). “A mass-conservative switching algorithm for modeling fluid flow in variably saturated porous media”. In: *Journal of Computational Physics* 230.3, pp. 664 – 679.
- Shao, Haibing, Svitlana V. Dmytrieva, Olaf Kolditz, Dmitrii A. Kulik, Wilfried Pfingsten, and Georg Kosakowski (2009). “Modeling reactive transport in non-ideal aqueous–solid solution system”. In: *Applied Geochemistry* 24.7, pp. 1287 –1300.
- Steeffel, C.I., C.A.J. Appelo, B. Arora, D. Jacques, T. Kalbacher, O. Kolditz, V. Lagneau, P.C. Lichtner, K.U. Mayer, J.C.L. Meeussen, S. Molins, D. Moulton, H. Shao, J. Šimůnek, N. Spycher, S.B. Yabusaki, and G.T. Yeh (2015). “Reactive transport codes for subsurface environmental simulation”. In: *Computational Geosciences* 19.3, pp. 445–478.
- Theis, Charles V. (1935). “The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage”. In: *Eos, Transactions American Geophysical Union* 16.2, pp. 519 –524.
- van Breukelen, Boris M., Daniel Hunkeler, and Frank Volkering (2005). “Quantification of Sequential Chlorinated Ethene Degradation by Use of a Reactive Transport Model Incorporating Isotope Fractionation”. In: *Environmental Science & Technology* 39.11. PMID: 15984799, pp. 4189–4197.

## Bibliography

- van der Lee, Jan, Laurent De Windt, Vincent Lagneau, and Patrick Goblet (2003). “Module-oriented modeling of reactive transport with HYTEC”. In: *Computers & Geosciences* 29.3. Reactive Transport Modeling in the Geosciences, pp. 265–275.
- Van Genuchten, M. T. (1980). “A closed-form equation for predicting the hydraulic conductivity of unsaturated soils”. In: *Soil Sci. Soc. Am. J.* 44, pp. 892–898.
- Vauclin, M., D. Khanji, and G. Vachaud (1979). “Experimental and numerical study of a transient, two-dimensional unsaturated-saturated water table recharge problem”. In: *Water Resources Research* 15.5, pp. 1089–1101.
- Wang, W., T. Schnicke, and O. Kolditz (2011). “Parallel finite element method and time stepping control for non-isothermal poro-elastic problems”. In: *Computers, Materials and Continua* 21.3, pp. 217–235.
- Wang, Wenqing, Georg Kosakowski, and Olaf Kolditz (2009). “A parallel finite element scheme for thermo-hydro-mechanical (THM) coupled problems in porous media”. In: *Computers & Geosciences* 35.8, pp. 1631–1641.
- Wissmeier, L. and D.A. Barry (2011). “Simulation tool for variably saturated flow with comprehensive geochemical reactions in two- and three-dimensional domains”. In: *Environmental Modelling & Software* 26.2, pp. 210–218.
- Xie, Mingliang, Sebastian Bauer, Olaf Kolditz, Thomas Nowak, and Hua Shao (2006). “Numerical simulation of reactive processes in an experiment with partially saturated bentonite”. In: *Journal of Contaminant Hydrology* 83.1–2, pp. 122–147.
- Xu, Tianfu, John A Apps, and Karsten Pruess (2004). “Numerical simulation of CO<sub>2</sub> disposal by mineral trapping in deep aquifers”. In: *Applied Geochemistry* 19.6, pp. 917–936.
- Xu, Tianfu, Eric Sonnenthal, Nicolas Spycher, and Karsten Pruess (2006). “TOUGHREACT—A simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: Applications to geothermal injectivity and CO<sub>2</sub> geological sequestration”. In: *Computers & Geosciences* 32.2, pp. 145–165.
- Xu, Tianfu, Nicolas Spycher, Eric Sonnenthal, Guoxiang Zhang, Liange Zheng, and Karsten Pruess (2011). “TOUGHREACT Version 2.0: A simulator for subsurface reactive transport under non-isothermal multiphase flow conditions”. In: *Computers & Geosciences* 37.6. 2009 Transport of Unsaturated Groundwater and Heat Symposium 2009 TOUGH Symposium, pp. 763–774.

## Bibliography

- Yabusaki, Steven B., Yilin Fang, Kenneth H. Williams, Christopher J. Murray, Andy L. Ward, Richard D. Dayvault, Scott R. Waichler, Darrell R. Newcomer, Frank A. Spane, and Philip E. Long (2011). “Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment”. In: *Journal of Contaminant Hydrology* 126.3–4, pp. 271–290.
- Yeh, Gour-Tsyh and Vijay S. Tripathi (1991). “A Model for Simulating Transport of Reactive Multispecies Components: Model Development and Demonstration”. In: *Water Resources Research* 27.12, pp. 3075–3094.
- Yeh, G.T. and V.S. Tripathi (1990). *A coupled model of HYDROlogical transport and GEOCHEMical Equilibrium of multi component system, ORNL-6371*. English. Oak Ridge National Laboratory, Oak Ridge.
- Šimůnek, J., D. Jacques, M. Šejna, and M. Th. van Genuchten (2012). *The HP2 program for HYDRUS (2D/3D), A coupled code for simulating two-dimensional variably saturated water flow, heat transport, solute transport and biogeochemistry in porous media (HYDRUS+PHREEQC+2D)*. Version Version 1.0. PC Progress, Prague, Czech Republic. 78 pp.
- Šimůnek, Jirka, Diederik Jacques, Martinus Th. Van Genuchten, and Dirk Mallants (2006). “MULTICOMPONENT GEOCHEMICAL TRANSPORT MODELING USING HYDRUS-1D AND HP11”. In: *JAWRA Journal of the American Water Resources Association* 42.6, pp. 1537–1547.



---

## LIST OF PUBLICATIONS

---

### ISI PUBLICATIONS

- He, W., C. Beyer, J.H. Fleckenstein, E. Jang, O. Kolditz, D. Naumov, and T. Kalbacher (2015b). “A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IPhreeqc 5.5.7-3.1.2”. In: *Geosci. Model Dev.* 8, pp. 3333–3348.
- He, W., H. Shao, O. Kolditz, W. Wang, and T. Kalbacher (2015d). “Comments on “A mass-conservative switching algorithm for modeling fluid flow in variably saturated porous media, K. Sadegh Zadeh, *Journal of Computational Physics*, 230 (2011)””. In: *Journal of Computational Physics* 295, pp. 815–820.

### OTHER PUBLICATIONS

- He, W. (2015). “Groundwater Flow—Theis’ Revisited”. In: *Thermo-Hydro-Mechanical-Chemical Processes in Fractured Porous Media: Modelling and Benchmarking*. Ed. by Olaf Kolditz, Hua Shao, Wenqing Wang, and Sebastian Bauer. Terrestrial Environmental Sciences. Springer International Publishing, pp. 115–120.
- He, W., C. Beyer, J.H. Fleckenstein, E. Jang, O. Kolditz, D. Naumov, and T. Kalbacher (2015a). “A parallelization scheme to simulate reactive transport in the subsurface environment with OGS#IPhreeqc”. In: *Geosci. Model Dev. Discuss.* 8, pp. 2369–2402.
- He, W., M. Walther, and O. Kolditz (2015c). “Benchmark: Theis Problem”. In: *OpenGeoSys-Tutorial*. SpringerBriefs in Earth System Sciences. Springer International Publishing, pp. 31–51.

# Appendices

# A

---

## OGS#IPHREEQC INPUT FILE DESCRIPTION

---

The numerical simulation with OGS#IPhreeqc relies on input-file-based model setups. In this appendix, the structure and contents of each input file for setting up reactive transport models with OGS#IPhreeqc will be introduced. The isotope fractionation benchmark (see Sect. 4.5.1) is used as a demonstration example here. The names of the input files and their short descriptions are given in Tab. 16.

Apart from the PHREEQC database file, all the other files share the same name (but with different file endings) and have the same structure illustrated in Listing. A.1. As one can see, an OGS input file consists of one or several main keywords which are followed by several sub-keywords with corresponding parameter values. An input file ends with the keyword `#STOP`. Everything that is written after the keywords will be ignored when running the program.

Listing A.1: Template for OGS input file

```
#MAIN_KEYWORD1
  $SUB_KEYWORD1
    value value value
  $SUB_KEYWORD2
    value value
#MAIN_KEYWORD2
  $SUB_KEYWORD1
    value value
...
#STOP
```

Listing. A.2 shows the process file in which the corresponding processes, which determine the employed governing equations to be solved, are defined. In this example, two types of processes i.e. *GROUNDWATER\_FLOW* and *MASS\_TRANSPORT* are taken into account. It is worth mentioning, that for every component (such as light isotope of PCE) one *mass transport* module has to

Table 16.: Input files and descriptions for the isotope fractionation benchmark.

Input file	Explanation
<i>isofrac.pcs</i>	process definition
<i>isofrac.gli</i>	system geometry
<i>isofrac.msh</i>	finite element mesh
<i>isofrac.num</i>	numerical properties
<i>isofrac.tim</i>	time discretization
<i>isofrac.ic</i>	initial conditions
<i>isofrac.bc</i>	boundary conditions
<i>isofrac.st</i>	source/sink terms
<i>isofrac.mcp</i>	component properties
<i>isofrac.mfp</i>	fluid properties
<i>isofrac.mmp</i>	medium properties
<i>isofrac.msp</i>	solid properties
<i>isofrac.out</i>	output configuration
<i>isofrac.pqc</i>	PHREEQC input definition
<i>phreeqc.dat</i>	PHREEQC database
<i>isofrac.ddc</i>	domain decomposition*

\* Only necessary for parallel simulation.

be defined.

Listing. A.3 illustrates the geometry file of this example. Each point is defined with an index number,  $x$ ,  $y$  and  $z$  coordinate as well as a name (if necessary); whereas a polyline is defined by a name and indices of points that belong to them. The names of these geometry elements (i.e. point, polyline) can be used when defining the initial-boundary condition, source term as well as model output.

Listing A.2: Process file (*isofrac.pcs*)

```
#PROCESS
$PCS_TYPE
  GROUNDWATER_FLOW
#PROCESS ; Transport Process: Pce_1
$PCS_TYPE
  MASS_TRANSPORT
#PROCESS ; Transport Process: Pce_h
$PCS_TYPE
  MASS_TRANSPORT
...
#STOP
```

Listing A.3: Geometry file (*isofrac.gli*)

```
#POINTS
0 0.0 0.0 0.0 $NAME POINT0
1 876.0 0.0 0.0 $NAME POINT1
#POLYLINE
$NAME
  OUT_LINE
$POINTS
0
1
#STOP
```

The spatial discretization of the model domain is defined in the mesh file (Listing. A.4). The finite element nodes are defined by giving their indices and coordinates, followed by the definition of finite elements, which contains the index, index for material group of the element, element type (e.g. line, triangle) and the indices of nodes that belong to the element.

Listing A.4: Mesh file (*isofrac.msh*)

```
#FEM_MSH
```

```

$NODES
121
0      0      0      0
1      7.3    0      0
2      14.6   0      0
...
120    876    0      0
$ELEMENTS
120
0      0      line      0      1
1      0      line      1      2
...
119    0      line      119    120
#STOP

```

Listing. A.5 to A.7 illustrate the three files dealing with setup of initial-boundary conditions. All of them are related to process (*PCS\_TYPE*) and geometries (*GEO\_TYPE*). The keyword *DIS\_TYPE* specifies the distribution type (e.g. *CONSTANT* and *CONSTANT\_NEUMANN*) and values. In addition, transient boundary conditions or source terms can be defined by adding another keyword *\$TIM\_TYPE*. This is especially useful when time dependent boundary conditions (e.g. seasonal variation of precipitation or temperature) are required.

Listing A.5: Initial condition (*isofrac.ic*)

```

#INITIAL_CONDITION
$PCS_TYPE
GROUNDWATER_FLOW
$PRIMARY_VARIABLE
HEAD
$GEO_TYPE
DOMAIN
$DIS_TYPE
CONSTANT 100.00
#INITIAL_CONDITION
$PCS_TYPE
MASS_TRANSPORT
$PRIMARY_VARIABLE
Pce_1
$GEO_TYPE
DOMAIN
$DIS_TYPE
CONSTANT 1e-12

```

```
...
#STOP
```

Listing A.6: Boundary condition (*isofrac.bc*)

```
#BOUNDARY_CONDITION
$PCS_TYPE
  GROUNDWATER_FLOW
$PRIMARY_VARIABLE
  HEAD
$GEO_TYPE
  POINT POINT0
$DIS_TYPE
  CONSTANT: 100.000
#BOUNDARY_CONDITION
$PCS_TYPE
  GROUNDWATER_FLOW
$PRIMARY_VARIABLE
  HEAD
$GEO_TYPE
  POINT POINT1
$DIS_TYPE
  CONSTANT: 9.781E+01
#BOUNDARY_CONDITION
$PCS_TYPE
  MASS_TRANSPORT
$PRIMARY_VARIABLE
  Pce_1
$GEO_TYPE
  POINT POINT0
$DIS_TYPE
  CONSTANT 0.00098921763658
...
#STOP
```

Listing A.7: Source term (*isofrac.st*)

```
#SOURCE_TERM
$PCS_TYPE
  GROUNDWATER_FLOW
$PRIMARY_VARIABLE
  HEAD
$GEO_TYPE
  POINT POINT1
```

```

$DIS_TYPE
  CONSTANT  0;  -1.15741E-06
#STOP
    
```

The properties of each component (e.g. mobility, molecular diffusivity) for mass transport, fluid (e.g. fluid type, density, viscosity), porous media (e.g. permeability, tortuosity, dispersion coefficient, storativity) and solid (e.g. bulk density) are defined from Listing. A.8 to A.11, respectively. It is worth mentioning, if process *GROUNDWATER\_FLOW* is applied, the value that should be given for the keyword *PERMEABILITY\_TENSOR* is the hydraulic conductivity in  $m \cdot s^{-1}$  (by default); when process *LIQUID\_FLOW* is employed, then the value of permeability in  $m^2$  should be given there.

Listing A.8: Component property (*isofrac.mcp*)

```

#COMPONENT_PROPERTIES  ; 0
$NAME
  Pce_1 ;
$MOBILE
  1          ; MOBIL-Flag: 0=immobile, 1=
  mobile/transported
$DIFFUSION
  1 3.0093E-10          ; diffusion model
  type, diffusion constant
...
    
```

Listing A.9: Fluid property (*isofrac.mfp*)

```

#FLUID_PROPERTIES
$FLUID_TYPE
  LIQUID
$DENSITY
  1 1000.0
$VISCOSITY
  1 1e-3
#STOP
    
```

Listing A.10: Property of porous media (*isofrac.mmp*)

```

#MEDIUM_PROPERTIES
$GEOMETRY_DIMENSION
  3
$GEOMETRY_AREA
  1.0
    
```



```

$POROSITY
  1  0.2500
$TORTUOSITY
  1  1.000000e+000
$STORAGE
  1  0.0
$PERMEABILITY_TENSOR
  ISOTROPIC  1.15741E-04
$MASS_DISPERSION
  1  1.0  1.0E-10
#STOP

```

Listing A.11: Solid property (*isofrac.msp*)

```

#SOLID_PROPERTIES
  $DENSITY
    1  2.65000
#STOP

```

The definition of numerical scheme is also process-specified, as shown in Listing. A.12. For each process, the parameters of linear or non-linear (for solving non-linear equation such as Richards equation) solver e.g. solver type, convergence criteria (error tolerance), number of maximum iterations have to be specified in this file.

Listing A.12: Numerics (*isofrac.num*)

```

#NUMERICS
  $PCS_TYPE
    GROUNDWATER_FLOW
  $LINEAR_SOLVER
; method error_tolerance max_iterations theta
  precondition storage
  2 1.e-014 1000 1.0 1 2
#NUMERICS
  $PCS_TYPE
    MASS_TRANSPORT
  $LINEAR_SOLVER
; method error_tolerance max_iterations theta
  precondition storage
  2 1.e-014 1000 0.50 1
  2
#STOP

```

In *OGS* it is possible to use different time stepping for different processes, as shown in Listing. A.13. In this example, fixed time step number and size ( *\$TIME\_STEPS*) are applied for solving groundwater flow and transport. Nevertheless, automatic time stepping can also be employed, which can be especially useful for non-linear processes.

Listing A.13: Time stepping (*isofrac.tim*)

```
#TIME_STEPPING
$PCS_TYPE
  GROUNDWATER_FLOW
$TIME_STEPS
  100  6307200
$TIME_END
  630720000
$TIME_START
  0.0
#TIME_STEPPING
$PCS_TYPE
  MASS_TRANSPORT
$TIME_STEPS
  100  6307200
$TIME_END
  630720000
$TIME_START
  0.0
#STOP
```

Listing. A.14 illustrates the output controls, which include the values of primary variables (head or component names) for output, geometry (the whole domain, a point or a polyline etc.), data types (e.g. TECPLOT, VTK, PVD) for output and output frequencies or selected output for certain time series (*\$TIM\_TYPE*).

Listing A.14: Output (*isofrac.out*)

```
#OUTPUT
$NOD_VALUES
  HEAD
  Pce_l
  Pce_h
  Tce_l
  . . .
$GEO_TYPE
  DOMAIN
```

```

$DAT_TYPE
TECPLOT
$TIM_TYPE
STEPS 20
#STOP

```

The input file *isofrac.pqc* (see Listing. A.15) is similar to a PHREEQC input file. By reading this file, the OGS#IPhreeqc interface will prepare PHREEQC input that will be executed in IPhreeqc. This file can be prepared based on a PHREEQC input file that is working in PHREEQC. Nevertheless, there are several minor changes have to be made. Firstly, the keyword *# comp* has to be added at the right-hand side of each component that has already been defined in *isofrac.mcp*. Concentration values of components are not important since they will be overwritten based on the setups of initial-boundary conditions as well as the transport calculations during each time step. Secondly, the keyword *# ende* has to be added after each PHREEQC module such as *SOLUTION*, *RATES*, *PRINT*, and so on. Unlike the other *OGS* input files, the terminator in this file is *END* instead of *#STOP*. Additionally, there is no need to include the transport module of PHREEQC, since the transport processes will be calculated by *OGS*.

Listing A.15: PHREEQC input definition (*isofrac.pqc*)

```

SOLUTION 1
units mol/kgw
temp 25.000000
pH 9.91 charge # comp
pe 4.0 02(g) -0.68 # comp
Pce_1 0 # comp
Tce_1 0 # comp
Dce_1 0 # comp
...
Chl 0 # comp
#ende

KINETICS 1
Pce_1_rd
-formula Pce_1 1 Chl -1 Tce_1 -1
-parms 6.366e-8 # first-order degradation
rate per second for PCE
-steps 100 in 1 steps

```

OGS#IPHREEQC INPUT FILE DESCRIPTION

```

...

Vc_h_rd
-formula Vc_h 1 Ch1 -1 Eth_h -1
-parms -23.2 # kinetic isotopic enrichment
      factor (epsilon) for VC

RATES
Pce_l_rd
-start
5 if (tot("Pce_l")+tot("Pce_h")) < 1e-9 then
  goto 60
10 rate = parm(1)*(tot("Pce_l")+tot("Pce_h"))
20 ratio = tot("Pce_l")/(tot("Pce_l")+tot("
  Pce_h"))
30 moles = -ratio * rate * time
40 put(rate, 1)
50 put(ratio, 2)
60 save moles
-end

...

Vc_h_rd
-start
...
-end
#ende

PRINT
  -reset false
  -selected_output true
  -status false
#ende

SELECTED_OUTPUT
-file phout_sel.dat
-high_precision
-reset false
#ende

USER_PUNCH

```

```

20 PUNCH TOT("Pce_1"), TOT("Tce_1"), TOT("
    Dce_1"), TOT("Vc_1") TOT("Pce_h"), TOT("
    Tce_h"), TOT("Dce_h"), TOT("Vc_h"), TOT("
    Eth_h"), TOT("Eth_1"), TOT("Ch1")
30 PUNCH -LA("H+"), -LA("e-")
#ende
END

```

When a simulation is executed parallel, an additional input file i.e. *isofrac.ddc* is needed to include information for domain decomposition (DDC). Listing. A.16 shows a DDC file for parallel simulation of the 2-D example in (Sect. 5.5.1). As one can see, it includes the indices of the elements and inner nodes (non-boarder nodes) that belong to a sub-domain. Mesh partition tools (e.g. Metis) are required to prepare this file, in order to balance the node quantities and minimize the number of boarder nodes among sub-domain efficiently.

Listing A.16: Domain decomposition

```

#DOMAIN 0
$ELEMENTS 62
59
60
...
1144
$NODES_INNER 80
119
118
...
#DOMAIN 1
$ELEMENTS 61
66
67
...
1150
$NODES_INNER 80
133
132
...
#STOP

```

# B

---

## INSTRUCTION FOR OGS#IPHREEQC COMPILATION AND EXECUTION

---

This tutorial introduces the basic steps to compile the source code of OGS#IPhreeqc and perform serial as well as parallel simulation in different operating systems i.e. Windows and Linux.

### B.1 DOWNLOAD THE SOURCE CODE

The source code can be found under the *Github* repository <https://github.com/hobit03/ogs5>. After checking out the source code, one should create a new branch from the remote branch *MPI\_STRING* and switch to that branch. In a Linux environment, one should first install or load *Github*, and then use the *Git* commands shown in Listing. B.1. In a Windows environment, one should first download and install a *Github* client such as *Source Tree* and perform the corresponding operations.

Listing B.1: Github command to get the code

```
git clone https://github.com/hobit03/ogs5
git checkout -b MPI_STRING origin/MPI_STRING
```

### B.2 SERIAL SIMULATION

#### B.2.1 *Source code compiling*

For the linux users, the source code can be compiled by using the following commands (*CMake* should be installed or loaded first):

Listing B.2: Serial compilation under linux environment

```
mkdir Build-Serial
cd Build-Serial
```

```

cmake [path of the source code] -
      DOGS_FEM_IPQC=ON -DCMAKE_BUILD_TYPE=
      Release
make

```

After the successful compilation of the source code, a binary file (OGS executable) will be generated (under *Build-Serial/bin*). For compilation on Windows, build configuration can be done by using either the *CMake – GUI* or command line, followed by build with Visual Studio. The *CMake* option *OGS\_FEM\_IPQC* should be chosen to enable OGS#IPhreeqc interface. More detailed instructions can be found with the following links: <http://devguide.opengeosys.org/win-configure-cmake/> and <http://devguide.opengeosys.org/win-build/>

### B.2.2 Execution

Before running simulations it is necessary to check the line endings of input files and adapt them to the operating system that is employed.

To execute the code in a Linux system, following command lines can be applied. A log file containing the simulation messages (e.g. source code version, data input, iterations during each time step) will be generated.

Listing B.3: Execution of the source code

```

[path of the ogs executable] [name of your
input file without file endings] > [name
of the log file]

```

To run the OGS#IPhreeqc executable under Windows, the simple way would be:

double click the ogs executable => type the path of the input file without file endings => press “enter” on the keyboard

However, this approach can be time-consuming, since all the simulation messages will be printed out on the command line interface.

Another useful approach is to create a batch file which includes the same command for linux shown above. Double click on the batch file will provoke the execution of the program.

## B.3 PARALLEL SIMULATION WITH OPEN MPI

B.3.1 *Source code compiling*

Under linux system, please follow the commands below:

Listing B.4: Parallel compilation under linux environment

```
module load openmpi/gcc/1.8.4-1
mkdir Build-MPI
cd Build-MPI
CC=mpicc CXX=mpic++ cmake [path of the source
    code] -DOGS_FEM_IPQC=ON -
    DPARALLEL_USE_MPI=ON
make
```

B.3.2 *Mesh partition*

For parallel simulation, mesh partition (with e.g. Metis) is required as a preprocessing procedure. A practical tool has been prepared by Wenqing Wang (Helmholtz Centre for Environmental Research-UFZ) for mesh partition. Detailed information on how to get and use this tool can be found at the following page (see parts Prepare mesh partitioning tool and Partition mesh): <https://svn.ufz.de/ogs/wiki/PETScPage> (one has to apply for the assess to the UFZ wiki first). After following these steps, a file containing the information about domain decomposition will be generated (with file endings such as *.20ddc*, the number specifies the number of sub-domains). One more step to do is to modify the file ending into *\*.ddc*.

B.3.3 *Execution*

It is highly recommended to check if the parallel run of flow and mass transport is successful before taking reaction into account (parallel simulation of reactive transport will be a mess if there is already something wrong with flow and mass transport). For multi-core shared memory systems like ENVINF, one can perform the parallel simulation with the following command:

Listing B.5: Command to perform parallel simulation on Open MPI



```
mpirun -np [number of processors] [path of
the executable] -ddc [number of subdomains
] [name of the input file without endings]
> [name of the log file]&
```

Job submission is often required for distributed memory systems. For example, *qsub* is used as the command for submission a job in the Sun Grid Engine queuing system. It would be convenient to prepare a bash script for *qsub*. An example script "paralle.sh" is given below, which works for the Eve cluster at the UFZ:

Listing B.6: Script for job submission on the EVE cluster

```
#!/bin/bash

#$ -N isofrac-2D
# name of the Job

#$ -o /home/hew/isofrac-2D/results.txt
# path to the STDOUT File

#$ -e /home/hew/isofrac-2D/err.txt
# path to the STDERR File

#$ -S /bin/bash
# shell to be used

#$ -l highmem=false
#$ -l h_rt=01:00:00

# Declaration of the runtime of your job
#$ -l h_vmem=2G

#$ -pe openmpi-orte 20
# request your parallel environment and
  number of cores

#$ -m beas -M wenkui.he@ufz.de
# Email address to send you messages

source /etc/profile.d/000-modules.sh

module load openmpi/gcc/1.8.2-1_gcc_4.8.1
# Compiler module, the same as you use for
  compiling the source code
```

```
set -x

APP="/home/hew/Build-MPI-IPQC/bin/ogs"
# The path of the executable file

/bin/echo In directory: `pwd`
# Print the current path

/bin/echo File name: $1
# Print the input file name

cd /home/hew/isofrac-2D
# Change to the directory where the input
  data are.

mpirun -np $NSLOTS $APP -ddc 20 $1
# keyword "-ddc" specifies the number of sub-
  domains. If it is not given, the number of
  sub-domains is the same as the number of
  compute cores by default.
```

After the preparation of this script, the job can be submitted with the following command:

Listing B.7: Command for job submission with qsub

```
qsub parallel.sh [name of the input file
  without endings]
```