Technische Universität Darmstadt
Fachbereich Material- und Geowissenschaften

SUBSURFACE CHARACTERIZATION BY MEANS OF GEOVISUAL ANALYTICS

Zur Erlangung des akademischen Grades eines
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

von

ADRIAN LINSEL, M.SC.

Erstgutachter: Prof. Dr. Matthias Hinderer
Zweitgutachterin: Prof. Dr. Maria-Theresia Schafmeister

Darmstadt 2020

The noblest pleasure is the joy of understanding.

LEONARDO DA VINCI (1452–1519)

The greatest use of practical science is discovery.

SIR HUMPHRY DAVY (1778–1829)

Declaration

I hereby declare that the presented dissertation is based on original research and is the result of my own work. I certify that this dissertation contains no material which has been accepted for the award of any other degree in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

Darmstadt, November 26th, 2020

_____

A. Linsel

_____

place/date

Abstract

This Thesis is concerned with one of the major problems in subsurface characterizations emerging from ever-increasing loads of data in the last decades: What kind of technologies suit well for extracting novel, valid and useful knowledge from persistent data repositories for the characterization of subsurface regions and how can such technologies be implemented in an integrated, community-open software platform?

In order to address those questions, an interactive, open-source software platform for geoscientific knowledge discovery has been developed, which enables domain experts to generate, optimize and validate prognostic models of the subsurface domain. Such a free tool has been missing in the geoscientific community so far. The extensible software platform GeoReVi (Geological Reservoir Virtualization) implements selected aspects of geovisual analytics with special attention being paid to an implementation of the knowledge discovery in databases process. With GeoReVi the human expert can model and visualize static and dynamic systems in the subsurface in a feedback cycle. The created models can be analyzed and parameterized by means of modern approaches from geostatistics and data mining. Hence, knowledge that is useful to both the assessment of subsurface potentials and to support decision-making during the utilization process of the subsurface regions can be extracted and exchanged in a formalized manner.

The modular software application is composed of both integrated and centralized databases, a graphical user interface and a business logic. In order to fulfill the needs of low computing time in accordance with high computational complexity of spatial problems, the software system makes intense use of parallelism and asynchronous programming.

The competitiveness of industry branches, which are aimed at utilizing the subsurface in unknown regions, such as the geothermal energy production or carbon capture and storage, are especially dependent on the quality of spatial forecasts for relevant rock and fluid properties. Thus, the focus of this work has been laid upon the implementation of algorithms, which enhance the predictability of properties in space under consideration of

uncertainty. The software system was therefore evaluated in ample real-world scenarios by solving problems from scientific, educational and industrial projects.

The implemented software system shows an excellent suitability to generically address spatial problems such as interpolation or stochastic simulation under consideration of numerical uncertainty. In this context, GeoReVi served as a tool for discovering new knowledge with special regard to investigating the heterogeneity of rock media on multiple scales of investigation.

Among others, it could be demonstrated that the three-dimensional scalar fields of different petrophysical and geochemical properties in sandstone media may diverge significantly at small-scales. In fact, if the small-scale variability is not considered in field-scale projects, in which the sampling density is usually low, statistical correlations and thus empirical relationships might be feigned.

Furthermore, it could be demonstrated that the simple kriging variance, which is used to simulate the natural variability in sequential simulations, systematically underestimates the intrinsic variability of the investigated sandstone media. If the small-scale variability can be determined by high-resolution sampling, it can be used to enhance conditional simulations at the scale of depositional environments.

Zusammenfassung

Die vorliegende Thesis behandelt eines der zentralen Probleme prognostischer Untergrundmodellierungen in Zeiten stetig wachsender Datenmengen: Wie kann neues, valides und nutzbares Wissen zur Bewertung von Untergrundpotenzialen aus großen Datenbanken mit geowissenschaftlichem Kontext optimal extrahiert und verwertet werden?

Hierzu wurde ein interaktives, quelloffenes visuelles Wissensfindungssystem mit Datenbankanbindung entwickelt, mit dem geowissenschaftliche Datensätze verwaltet und semi-automatisiert prognostische Untergrundmodelle entwickelt, validiert und optimiert werden können. Solch ein quelloffenes System fehlte bislang innerhalb der geowissenschaftlichen Forschungsgemeinschaft. Das System basiert auf dem Prinzip der Wissensfindung in Datenbanken und implementiert ausgewählte Aspekte aus der Disziplin der visuellen Analytik. In einem interaktiven Benutzer-Maschine-Kreislauf können Domänenexperten statische und dynamische Untergrundsysteme modellieren, um neues Wissen zur Beurteilung von Geopotenzialen aus vorhandenen Datensätzen zu extrahieren. Das System stellt hierfür moderne Algorithmen der Geostatistik oder des Data Mining zur Verfügung. Das mehrschichtige, modulare Softwaresystem besteht aus lokalen und zentralisierten Datenbanken, einer graphischen Benutzeroberfläche und einer Geschäftslogik.

Datensätze aus wissenschaftlichen und industriellen Projekten, die vorwiegend aus der Domäne der geologischen Reservoircharakterisierung stammen, wurden in das System importiert und genutzt, um wissenschaftliche Fragestellungen unter Zuhilfenahme des Softwaresystems zu beantworten. Die Wirtschaftlichkeit der Reservoirnutzung ist maßgeblich von der Qualität räumlicher Prognosen relevanter Untergrundparameter abhängig, weswegen der Schwerpunkt dieser Arbeit auf die Implementierung von Algorithmen zur Verbesserung der Vorhersagbarkeit dieser Untergrundparameter gelegt wurde. Das Softwaresystem wurde diesbezüglich in realen Testszenarien evaluiert.

Die Ergebnisse der Fallstudien zeigen, dass sich visuelle Wissensfindungssysteme hervorragend dafür eignen, geowissenschaftliche Fragestellungen unter der Berücksichtigung von Unsicherheiten zu lösen. In den Fallstudien konnte gezeigt werden, dass sich die

räumliche Ausprägung von dreidimensionalen Skalarfeldern physikalisch-chemischer Eigenschaften in Sandsteinmedien auf der sub-Meterskala signifikant unterscheiden kann. Ohne Berücksichtigung der kleinskaligen geologischen Variabilität könnten Rückschlüsse über statistische Zusammenhänge auf der Anwendungsskala, auf der die Beprobungsdichte generell gering ist, fälschlicherweise impliziert werden.

In dieser Fallstudie konnte des Weiteren gezeigt werden, dass die Simple Kriging Varianz, die zur stochastischen Nachbildung der natürlichen Variabilität in sequentiellen Simulationen Verwendung findet, die natürliche Variabilität innerhalb der untersuchten porösen Sandsteinmedien systematisch unterschätzt. Falls die natürliche Variabilität in hochauflösenden Studien ermittelt werden kann, kann diese genutzt werden, um die Simulation der lokalen Variabilität innerhalb sedimentärer Ablagerungsräume zu verbessern.

Acknowledgments

Numerous people and institutions contributed to this Thesis in various ways. Therefore, I would like to express my sincerest thanks to:

Prof. Dr. Matthias Hinderer for the countless invaluable discussions and for the decisive and honest advises, which helped to continuously improve my work and to expand my horizon. You gave me the chance to work self-dependently and to implement my own ideas — yet always letting me know that there is an open ear if any problems or questions arise. Without you, this journey wouldn't have been made the way it has been — thank you Matthias.

Prof. Dr. Maria-Theresia Schafmeister for agreeing to act as a second examiner of my dissertation. Being aware that this task involves a tremendous amount of time, your effort is highly appreciated.

Dr. Kristian Bär and **Dr. Jens Hornung**, who supported me in any kind of problem that arose during my studies. Without their regular reviews, constructive advises and thoughtful suggestions this work couldn't have been accomplished in the resultant quality.

Matthias Greb for hosting my stay as a visiting scholar at the Energy & Geoscience Institute in Salt Lake City, Utah, and for his intense support and invaluable advises.

Reimund Rosmann for so many hours of help in both the field and the lab. I really don't know how to thank you for all of your efforts and helpful ideas. Maybe I can partly compensate for it by hosting some pizza-sessions in future.

All other members of the Department of Applied Sedimentary Geology, and especially, Dr. Laura Stutenbecker, PD Dr. Olaf Lenz, and Jürgen Mutzl. I was really glad to have such great colleagues not only for countless discussions and pleasant cooperation but also for having an overall nice time altogether.

Contents

List of Figures

List of Tables

Symbol Directory

Mathematical symbols

$A \Leftrightarrow B$	A holds if and only if B holds	$\mathbf{h}$	lag vector
α, β, γ	angles	∇	Laplace operator
R	coefficient of correlation	μ	Laplace parameter or mean
R^2	coefficient of determination	$\max$	maximum
c	covariance function	$\min$	minimum
$\varnothing$	diameter	mod	modulo
Δ	difference	n	number of elements
γ	dissimilarity	∂	partial derivation
λ_i	eigenvalue	ζ, ξ, η	projected variables
v	eigenvector	rk	rank
ϵ	error	$\hat{e}$	residual
$i, j,$ or k	index	Ω	spatial domain
∞	infinity	$[\mathbf{a}_1, ... \mathbf{a}_n]$	start and end of a matrix
$\int$	integral	$(a_1, ..., a_n)$	start and end of a vector
$\mathbf{I}$	inverse of a matrix	$\sum$	sum
J	Jacobian matrix	$\mathcal{U}$	universe
$\bowtie$	join operator	V	volume

Material properties

T	ambient temperature	$[°C]$
ρ	density	$[g\,cm^{-3}]$
H	enthalpy	$[J]$
q	heat	$[J]$
c	heat capacity	$[J\,kg^{-1}\,K^{-1}]$
k	hydraulic permeability	$[m^2]$
k	hydraulic conductivity	$[m\,s^-1]$
ϕ	porosity	[-]
κ	magnetic susceptibility	[-]
m	mass	[kg]
SiO$_2$	mass fraction of a molecule in a medium	[kg kg^{-1}]
p	pressure	[kg m^{-1} s^{-2}]
v_p and v_s	P- and S-wave velocity, respectively	$[m\,s^{-1}]$
A	radiogenic heat production	$[W\,m^{-3}]$
H	source term	variable
σ	stress	[kg m^{-1} s^{-2}]
λ	thermal conductivity	$[W\,m^{-1}\,K^{-1}]$
α	thermal diffusivity	$[m^2\,s^{-1}]$
V	volume	$[m^3]$

Abbreviations

1-D, 2-D, 3-D	one-, two-, three-dimensional
BL	business logic
CDF	cumulative density function
DAL	data access layer
DIKW	data-information-knowledge-wisdom
DMS	database management system
DSS	direct sequential simulation
EDM	exploratory data mining
ERD	entity-relationship diagram
FEM	finite element method
FK	foreign key
GUI	graphical user interface
GVA	geovisual analytics
KDD	knowledge discovery in databases
KGMVA	knowledge generation model for visual analytics
LDM	logical data modeling or, respectively, logical data model
LpO CV	leave p out cross validation
MAE	mean absolute error
MSE	mean square error
MVVM	model-view-view model
NoSQL	not-only structured query language
NST	normal score transform

ODE	ordinary differential equation
OK	ordinary kriging
PCA	principal component analysis
PDE	partial differential equation
PDF	probability density function
PDM	predictive data mining
PK	primary key
RDBMS	relational database management system
RDM	relational data mining
REV	representative elementary volume
RM	relational model
RMSE	root mean square error
SDM	spatial data mining
SGS	sequential Gaussian simulation
SK	simple kriging or, respectively, super key
SQL	structured query language
TAP	task-based asynchronous pattern
U3D	usage-driven database design
UK	universal kriging
VA	visual analytics
VDM	visual data mining

1. Introduction

In the introduction, the problems addressed in the Thesis at hand will be outlined. Therefore, investigations will be introduced and rationalized with which this Thesis aims to contribute to solving these problems. Moreover, the structure of the Thesis will be presented and the basic notations applied in this work will be provided.

1.1. Problem Statement

Industrial applications which are making use of the subsurface, such as petroleum engineering, geothermal heat production, groundwater extraction, mining operations, carbon-capture and storage, tunneling, or nuclear waste disposal, require a deep understanding and highly accurate spatial predictions of relevant physical or geochemical properties in order to assess the economic feasibility of a target region (Magnus Bergman 1986; Landa & Strebelle 2002; Heap et al. 2017; Kushnir et al. 2018; Rodrigo-Ilarri et al. 2017). The degree of subsurface exploitation has remained more or less constant over the past decades as indicated by Baker Hughes' international rig count[1] (Baker Hughes 2020) which is shown in Figure (1.1). Prior to drilling, the target region in the subsurface needs to be characterized and well understood in order to plan the borehole correctly and to scale and optimize the development strategy which should lead to a maximization of the economic revenue.

The data, which are required for making a comprehensive subsurface characterization, are usually compiled from numerous geoscientific domains such as petrophysics, stratigraphy, geophysics, petrography or hydrochemistry. Thus the compilation of data – data, which is in most cases incomplete – has always been a strenuous task for geologists, petrophysicists or reservoir engineers. The problem is reinforced by the fact that data must be acquired from multiple sources with alternating semantics (Ge & Chen 2010). The scarcity and

---

[1]https://rigcount.bakerhughes.com/intl-rig-count/, last access 22nd November, 2020
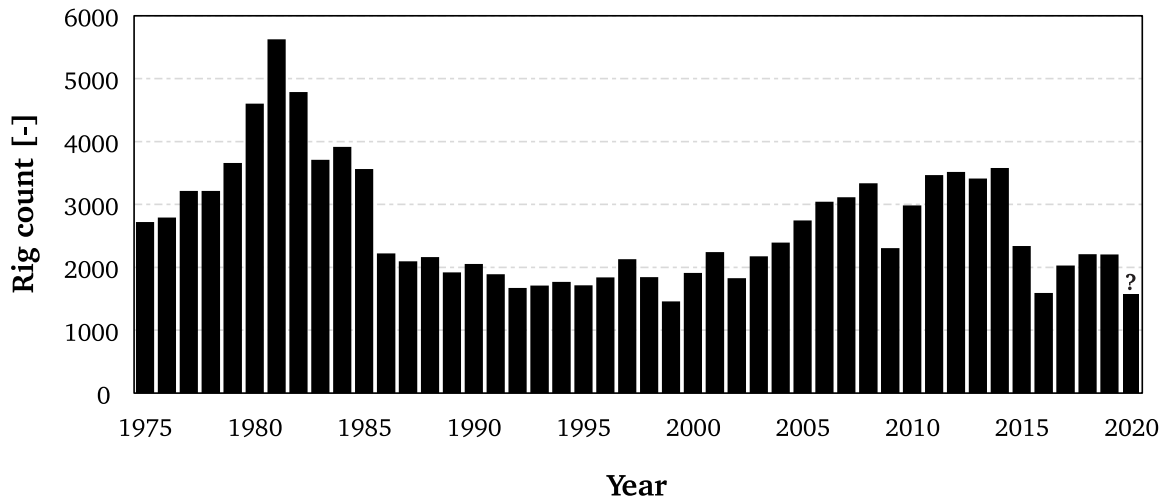
Figure 1.1.: Time distribution of Baker Hughes' international rig count.

complexity of subsurface-related data sets may lead to even higher efforts while integrating the corresponding data sets into existing software systems. Consequently, data incorporation may make up the biggest share of the domain expert's total working hours (Fayyad et al. 1996a).

Although numerous commercial software systems are available providing the capability to organize, model and analyze subsurface data, the geoscientific community is still lacking an open-source system that is capable of supporting the entire knowledge discovery process chain of a domain expert efficiently. By now, open-source, freely available software applications for geoscientific knowledge discovery are either still in their infancy or architecturally outdated.

Generally, knowledge about the geological architecture and the physicochemical properties from subsurface regions is scarce due to cost-intensive exploration methods. Hence, sparsely sampled subsurface information is available for analyses and for the prediction of the subsurface's behavior during production. The competitiveness of those industry branches which are aimed at utilizing the subsurface in unknown regions, such as the geothermal energy production or carbon capture and storage, are especially dependent on the quality of spatial forecasts for relevant rock and fluid properties. These properties are, however, prone to uncertainty induced by measurement errors, interpolation errors and local geological variability. Up until now, geostatistical simulation algorithms were aimed at accounting for measurement and interpolation errors but not for the local geological variability, which is generally not well understood as studies at that scale are time-consuming and usually avoided.

1.2. Contributions of the Thesis

This Thesis aims at addressing the two major major problems, outlined in the previous section, which involve geostatistical simulation algorithms not accounting for local geological variability in subsurface predictions and a missing community-open geological knowledge discovery system. Consequently, the main research objectives of this Thesis result in:

I) Implementing an open-source geological knowledge discovery system to assist domain experts in characterizing, modeling and understanding subsurface systems and their industrial potentials.

II) Validation of the software system by unit testing and performing case studies which aim at modeling the local geological variability of a series of physical and geochemical rock properties in space.

III) Enhancing existing geostatistical simulation algorithms in order to account for local geological variability.

IV) Discovering new, valid and usable knowledge in case studies involving real subsurface systems by means of the developed software system.

Firstly, a new open-source software system is developed that supports geoscientists within typical knowledge discovery process chains. Secondly, data sets from scientific and industrial projects are imported into the system and utilized to investigate real-world problems related to subsurface characterization. Those studies are used to evaluate the software system in ample real-world application scenarios.

The case studies constitute both classical characterization processes of subsurface domains at various scales as well as new methodological approaches that improve the predictability of uncertainty in unknown space.

1.2.1. Journal publications

This cumulative dissertation comprises three journal publications and one book section, which all underwent a single blind peer-review and which are listed below. The full papers are listed in the section of chapter 3.

- **Publication I**:
  Linsel, A., Bär, K., Haas, J., Hornung, J., Greb, M.D. and Hinderer, M. (2020): GeoReVi: A knowledge discovery tool for subsurface characterization. *SoftwareX*. doi:10.1016/j.softx.2020.100597

- **Publication II**:
Hornung, J., **Linsel, A.,** Schröder, D., Gumbert, J., Ölmez, J., Scheid, M., Pöppel-reiter, M. C. (2020): Understanding small-scale petrophysical heterogeneities in sedimentary rocks: The key to understanding pore geometry variations and to predicting lithofacies-dependent reservoir properties. *In:* Grötsch J. & Pöppelreiter, M. C.: Digital Geology – Multi-scale analysis of depositional systems and their subsurface modelling workflows. *EAGE Special Publication* ISBN 9789462823372.

- **Publication III**:
Linsel, A., Wiesler, S., Hornung, J. and Hinderer, M. (2020): High-Resolution Analysis of the Physicochemical Characteristics of 3-D Sandstone Media at the Lithofacies Scale. *Solid Earth*. doi:10.5194/se-2020-13

- **Publication IV**:
Linsel, A., Wiesler, S., Haas, J., Bär, K. and Hinderer, M. (2020): Accounting for Local Geological Variability in Sequential Simulations—Concept and Application. *ISPRS International Journal of Geo-Information*. doi:10.3390/ijgi9060409

The scope and the author's contribution are described in those chapters, which are associated with the specific publication.

1.2.2. Conference publications

Furthermore, the research was presented on conferences and exhibitions in oral presentations and poster sessions:

Linsel, A. , Bär, K. and Hinderer, M. (2017): Linking Siliciclastic Sedimentology to Geothermal Reservoir Characterization with a Relational Database Application. *5th European Geothermal Workshop - Characterization of Deep Geothermal Systems*. (*poster presentation*)

Linsel, A., Bär, K., Hornung, J. and Hinderer, M. (2018): Bridging Information Gaps in Reservoir Studies – Archiving and Retrieving Data With Relational Database Management Systems. *AAPG Annual Convention and Exhibition*, Salt Lake City, 5/20 - 5/23/2018. (*poster presentation*)

Stutenbecker, L., **Linsel, A.**, Caracciolo, L., Vermeesch, P. and Hinderer, M. (2018): A database for compositional data. *Working Group on Sediment Generation*, Dublin, 06/27-06/29/2018. (*poster presentation*)

Linsel, A., Bär, K., Hornung, J. and Hinderer, M. (2019): Visual analytics and information management for comprehensive geothermal reservoir characterization. *NovCare 2019*, Waterloo, ON, 05/28 - 05/31/2019. (*oral presentation*)

Linsel, A., Weinert, S., Bär, K. and Hinderer, M. (2019): Thermo-Hydraulic Heterogeneity Assessment Across First-Order Hiatal Surfaces – A Case Study from the Post-Variscan Nonconformity. *GeoMünster 2019*, Münster, 09/22 - 09/25/2019. (*oral presentation*)

Greb, M. and **Linsel, A.** (2020): Meaningful geoscientific data collection, curation and integration in petroleum exploration and development endeavors. *AAPG Annual Convention and Exhibition*, Houston, TX. (*oral presentation*)

1.3. Structure of the Thesis

This Thesis is structured into five parts (Fig. 1.2) and further sub-parts.

In **Part I** the topic, structure and basic notations of this Thesis will be introduced.

Part II will outline the theoretical fundamentals this Thesis is built upon. These comprise a detailed background of subsurface characterization in general together with aspects from the technical disciplines knowledge discovery in databases (KDD) and geovisual analytics (GVA). At first, we will provide the basics of subsurface characterization by formalizing the concepts of geological compartmentalization, the theory of the regionalized variables and static as well as dynamic data-driven subsurface modeling.

Following up on the discipline of subsurface characterization, fundamentals from the domain of GVA will be discussed. Here, we will focus on the interaction between human and machine and will rationale which methods suit best to fit into the topic of the geological subsurface characterization process chain.

When we come to explaining aspects from KDD, argumentation will split up into fundamentals from exploratory data mining and descriptive data mining. Herein, we will link the computational methods to geoscientific examples, most of which have been produced in the frame of the case studies as a part of this Thesis.

In **Part III** the publications related to this cumulative Thesis will be presented. **Publication I** contains the development of a software system for visual knowledge discovery in the context of subsurface characterization. The architecture, the logical data model and the graphical user interface will be presented. Furthermore, typical process chains as they
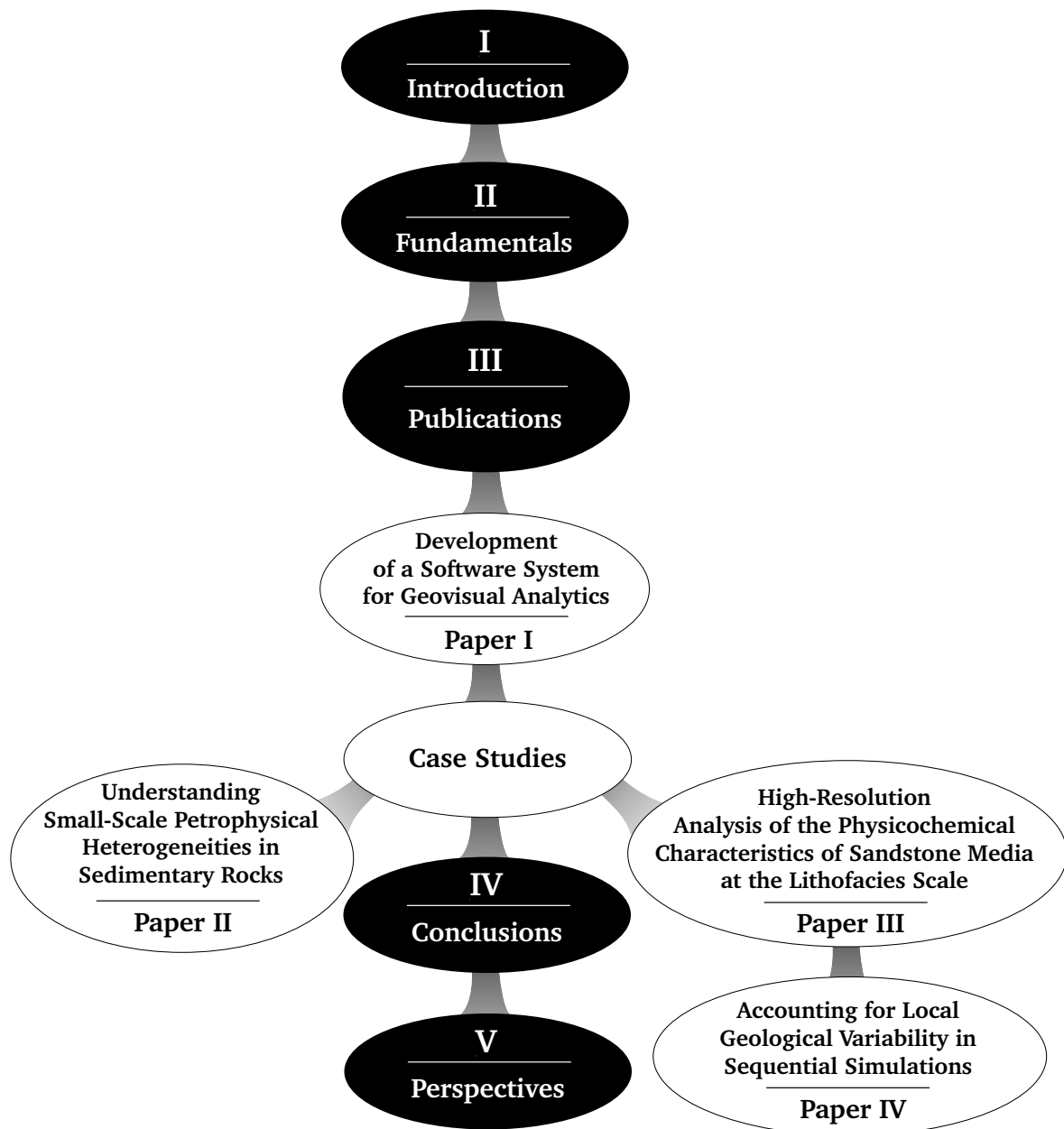
Figure 1.2.: Structure of the dissertation.

are conducted during subsurface characterization will be presented. The software system is the key component of this dissertation.

Moreover, a set of case studies will provide a verification for the implemented software functionality. In each case study, the potential of the framework will be presented to discover new knowledge in different geological and investigative settings around geological subsurface characterization.

The last two parts, namely **Part IV** and **Part V**, will wrap up the findings made in this Thesis and close it with perspectives over further developments in the topic for visual

knowledge discovery for decision making in the context of subsurface utilization.

At this point, I want to mention, that the synopsis presented here contains literal text passages from my first-author publications without giving any reference. I want to clarify, that these self-quotations are present and that they do not denote any intent to deceive.

1.4. Basic Notations

In this Thesis, a set of notation rules regarding mathematical and physical symbols and logical expressions comes into practice:

Scalars are given in lowercase letters. Bold lowercase letters, such as $\mathbf{x}$, stand for vectors while bold capitals, such as $\mathbf{M}$ or $\mathbf{Z}$, stand for a matrix, a tensor, or a (random) variable. In the subject of relational models bold capitals stand for a set of relations. Italic capitals (e.g, R) express a relation, respectively.

For physical and chemical properties internationally standardized symbols are used such as λ representing the thermal conductivity or k representing the hydraulic permeability, respectively. All symbols that are used in the frame of this Thesis are provided in the list of symbols.

Proper names start with capitals and quotes are displayed in italic font style which is framed by quotation marks.

2. Fundamentals

2.1. Introduction

In the following sections, we will describe the theoretical background of subsurface characterization and its integration into the topic of geovisual analytics as it found usage in the publications connected to this Thesis. Thereby, focus will be laid upon both the geological and technical understanding of that domain. A specific example of one form of subsurface usage is given in section 2.5.3 in which aspects from the domain of geothermal reservoir characterization in general are discussed.

2.2. The Subsurface Domain

Lying beneath the atmosphere and the oceans is the subsurface of the Earth, called *geosphere*. By definition, the geosphere comprises that material which is located between the ground surface of the Earth and its center and which is composed of solids, such as minerals or rocks, fluids, including, for instance, water or methane, as well as organic matter. Up until now, utilization of the subsurface takes place in the outermost layer of the geosphere, named the Earth's crust. The Earth's crust measures roughly seven kilometers in vertical direction in the oceanic compartments and averages between 35 and 40 kilometers in the continental compartments (Tarbuck & Lutgens 2008).

2.2.1. Geological Compartmentalization

The crust of the Earth $\mathbf{G}$ is built up by a set of disjoint geological bodies $G_1, G_2, ..., G_n$ and fluids stored in the bodies' pore space below the ground surface, where $1 \leq i \leq n$. Let $\mathbf{G} = G_1 \cup G_2 \cup ... \cup G_n$ where $G_i \neq \emptyset$. Strongly simplified, a geological body consists of three constitutes: rock matrix, liquids and gaseous phases. The degree of internal

compartmentalization in a subsurface domain is dependent on the scale and type of specific problems. In geodynamic problems, for example, major rock units are considered to be continuous geological media (Gerya 2009) under the assumption that there are no macroscopically observable mass-free spaces present. This concept is valid for most parts of the Earth's crusts but is simplified when for instance karst, extensive fracture networks or divergent fault zones must be accounted for in a discretized subsurface model.

G_i can be defined as a specific type of lithology separated by one another through bounding surfaces. Bounding surfaces are 2.5-D features represented by faults, joints, material-, grain size- or structural contrasts (Fig. 2.1). The rock matrix is composed of solid mineral phases such as quartz or feldspar, ductile phases, nonmineral inorganics, or organics. Geological compartmentalization takes place at multiple scales ranging from the nanometer scale, where for instance crystal defects occur, to the kilometer scale, where plate tectonic processes control the crusts architecture. Usually, compartmentalization of the subsurface is hierarchically structured where one geological body G_i is built up by other geological bodies $G_{1a}, G_{1b}, ..., G_{nm}$ so that we can say

$$V_{G_i} = \sum_{k=1}^{m} V_{G_{ik}}, \tag{2.1}$$

where V_{G_i} is the volume of a geological body in the Earth's crust under the assumption that $G_{ik} \subseteq G_i$ and m is the number of geological bodies building up G_i.

Figure 2.1.: Geological compartmentalization produced by erosive bounding surfaces and diagenetic alteration in a siliciclastic environment. The picture was taken in the Valley of Fire State Park in Nevada, USA.

2.2.1.1. The Lithofacies Concept

It is common to categorize rocks observed in the field or drill core into rock types. As a classification scheme, which arouse within the domain of sedimentology, the lithofacies concept has been established over the past decades (Nichols 2009). Summarizing reviews on the definitions, concepts and applications of the lithofacies concept in sedimentary environments can be found in Miall (2000) and Nichols (2009). A comprehensive review is provided by Reading (1996) while the first approaches can be found in Collinson (1969). The lithofacies concept aims to categorize lithological bodies in a hierarchical scheme which can be mathematically expressed as

$$\text{microfacies} \subseteq \text{lithofacies} \subseteq \text{architectural element} \subseteq \text{depositional environment}. \quad (2.2)$$
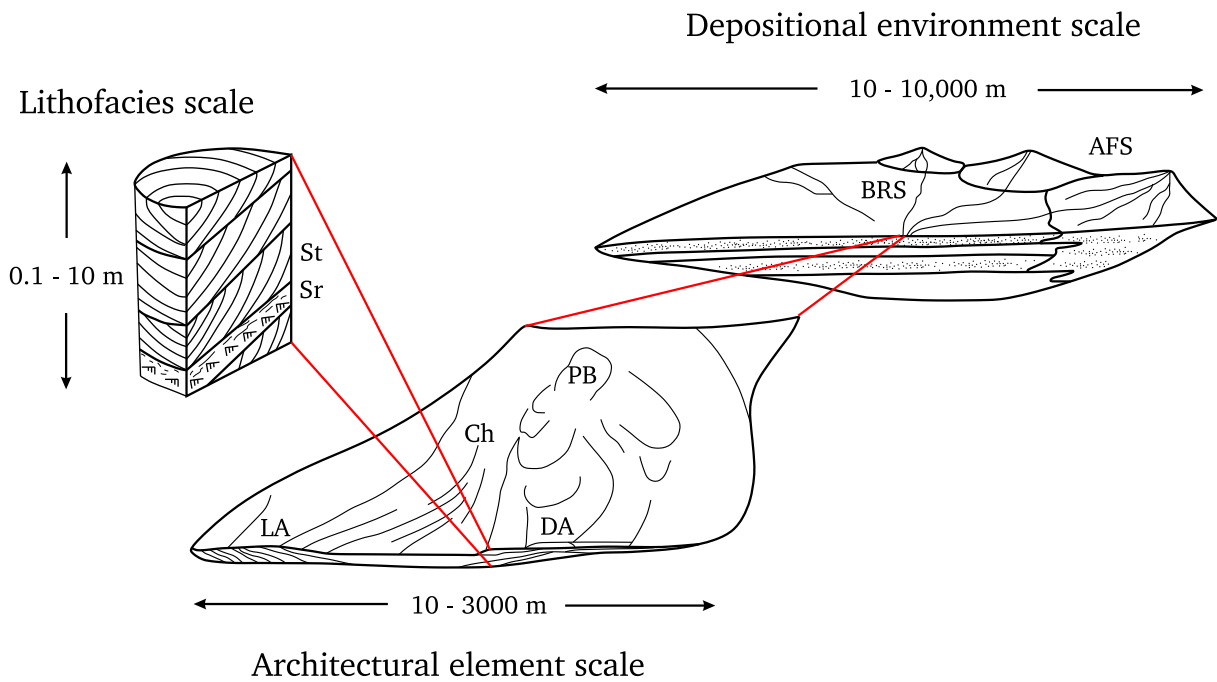
Figure 2.2.: Visual representation of the lithofacies concept (Miall 1985; Miall 1996; Miall 2000). St = sand trough cross-bedded; Sr = sand ripple cross-bedded; PB = point bar; Ch = channel; DA = downstream accretion ; LA = lateral accretion; BRS = braided river system; AFS = alluvial fan system.

Accordingly, each geological body is built up by one or more bodies observed at a smaller scale (see Fig. 2.2). Each component in this scheme is assigned a code representing either a specific lithological category and its macroscopic characteristics (e.g., St = sand trough cross-bedded) or being an acronym for the type of lithological body (e.g., Ch =

channel). For siliciclastic rocks, such a lithofacies code consists of up to three letters; the first of which describing the grain size of the component, the second one the primary sedimentary structure and the third one the type of framework in case of gravel-sized sedimentary rocks (see Fig. 2.2).

Lithofacies types are defined by building distinct classes of rock types based on their macroscopic characteristics (see, e.g., Hornung & Aigner 2002 or Hornung & Hinderer 2011). The lithofacies type can be derived from characteristics from various scales covering the micrometer up to the kilometer scale. Descriptive characteristics comprise, among others: grain/mineral size distribution, color, grade of weathering, internal structure/stratification, width, thickness, texture, mineral composition, porosity, strength and fracture behavior. Rock units, classified into identical classes, should be as similar as possible regarding physical and chemical properties. Hence, the term facies refers to the sum of characteristics of a lithological body.
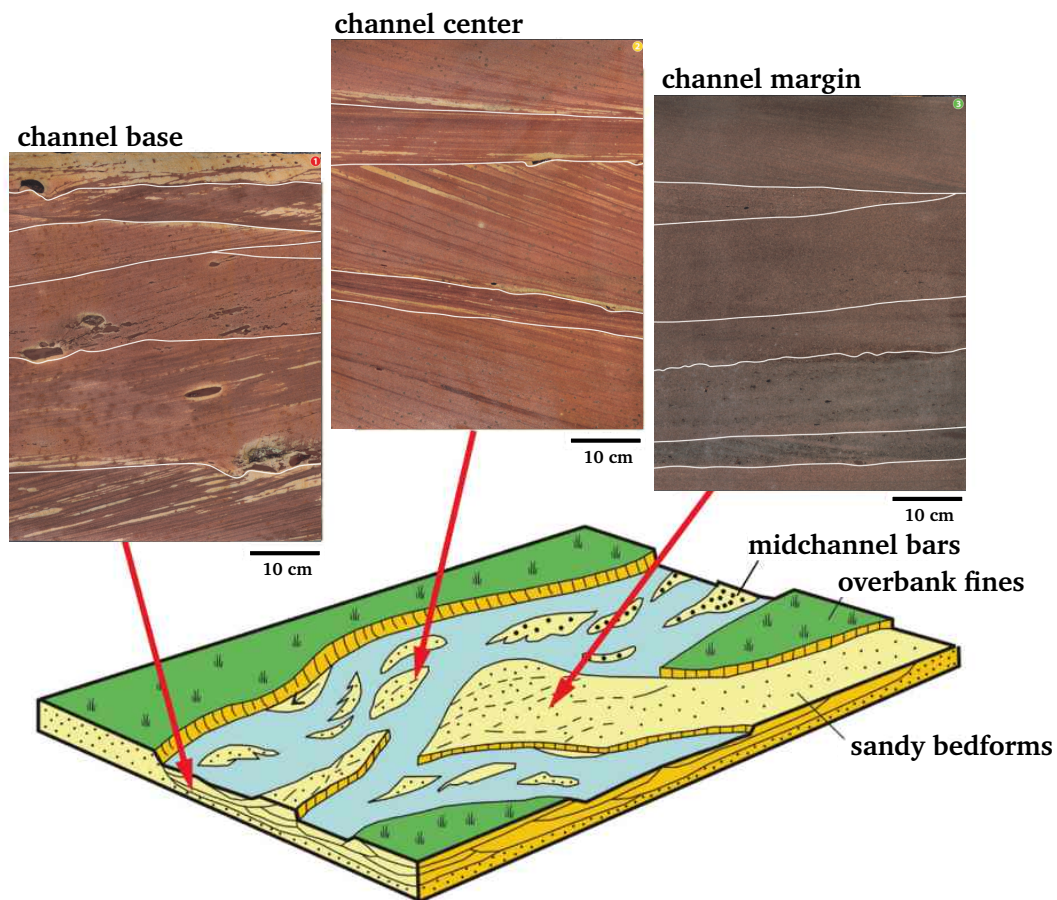
Figure 2.3.: Hierarchical structure of architectural elements in the depositional environment of a siliciclastic braided river system after Einsele (2000). The displayed rock slabs served as objects of investigation for **Publication II** (Hornung et al. 2020, ©EAGE).

In sedimentary environments, the next higher-level scale is represented by architectural elements. Architectural elements are composed of lithofacies types and build distinct, bounded bodies too. Architectural elements, in turn, are the constructive units of depositional environments (Miall 2000) or depositional elements, respectively, as they are termed in the studies of Colombera et al. (2012), Colombera et al. (2016)a, and Colombera et al. (2016)b. Figure 2.3 shows the schematic distribution of architectural elements in a braided river system with real-world examples of architectural elements and lithofacies types.

2.2.2. The Theory of Regionalized Variables

Geological processes can create, interfere with or remove parts of the Earth's crust. Thus, the physical and chemical characteristics of a specific location $\mathbf{x}$ in the subsurface's domain at point t in time is a product of natural processes such as deposition, diagenesis and erosion that had taken place in the time $\leq$ t at point $\mathbf{x}$. It is important to mention that the relative position of $\mathbf{x}$ is dynamically drifting during geological history rather than being a fixed point relative to the Earth's center.
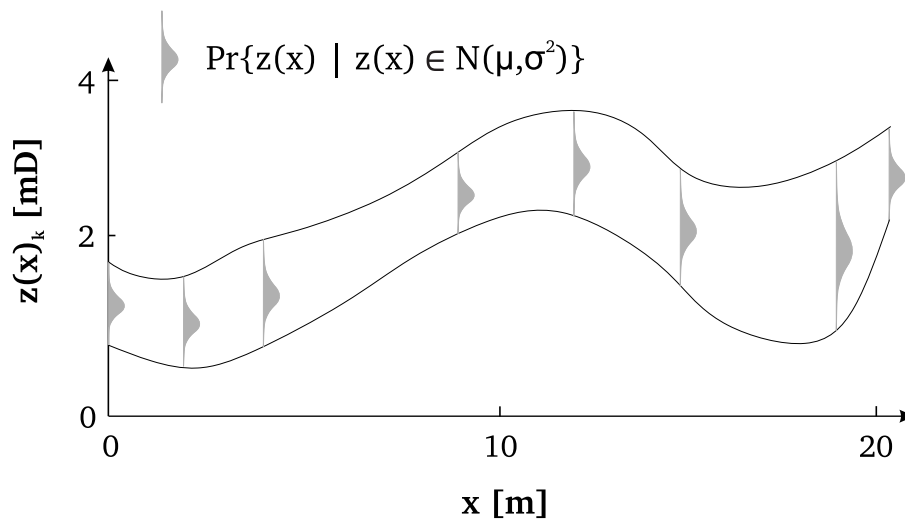
Figure 2.4.: Conceptual schematic of the regionalized variable. The example shows the spatial distribution of a rock property called intrinsic permeability (k) along the x-axis within a subsurface domain accompanied by simplified smaller-scale fluctuations in the form of normal distributions indicating the local variability.

Usually, each location $\mathbf{x}$ can be represented by a set of Cartesian coordinates in a three-dimensional space. In order to reduce the probability of economic failure in mining industries the 1960's the concept of the *regionalized variable* had been developed by Matheron (1963). The regionalized variable is a function that takes a definite value

at each point of space. In geological media the regionalized variable often proved to be too complex to be expressed by mathematical functions. Thus, Matheron (1963) also shaped a term called the *geometrical field* that is inimitably linked to the spatial nature of geological processes. The geometric field was defined as the '*mineralized space (volume of the deposit or of the strata) ... of the regionalization*' (Matheron 1963) making the regionalized variable localized. The geometrical field accordingly corresponds to a geological body G as it had been outlined in section 2.2.1. A regionalized variable is assumed to show a more or less steady continuity in space accompanied by fluctuations or noise (Fig. 2.4).

These local fluctuations result from heterogeneities observed at smaller scales than the considered observation field. We will discuss the term heterogeneity in section 2.2.2.1. The continuity of a regionalized variable is thus dependent on the continuity of the geological media (see Fig 2.5) and may or may not provide continuity in a mathematical sense. Lastly, a regionalized variable may or may not show different kinds of *anisotropy* the theory of which will be discussed in section 2.2.2.2.
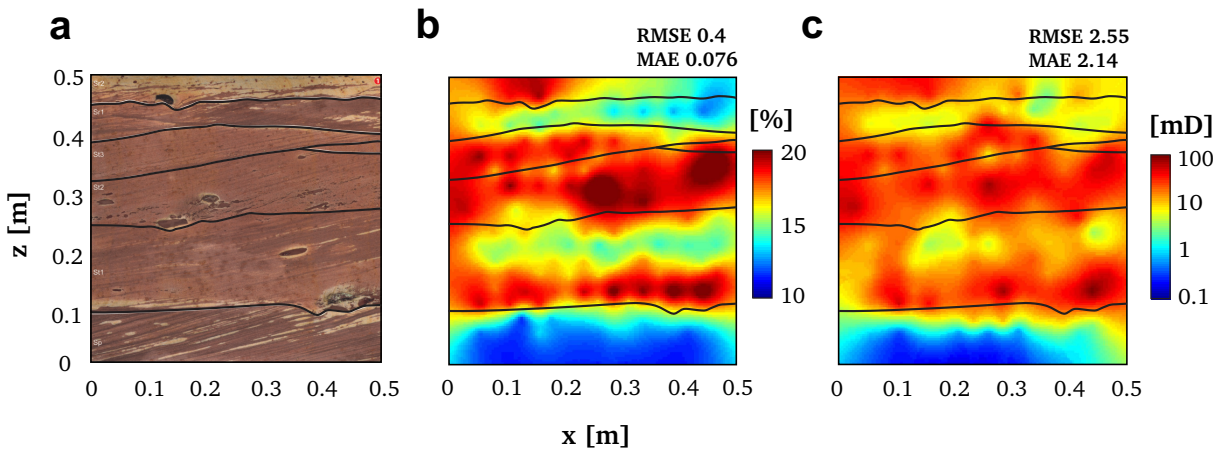
Figure 2.5.: The geometric field of a regionalized variable is expressed by a more or less steady mean in space accompanied by smaller-scale fluctuations. **(b)** and **(c)** show the spatial distribution of the effective porosity ϕ in % and the apparent permeability k in mD in a rock slab measuring $50 \times 50 \times 6$ cm **(a)** which was taken from a Buntsandstein quarry in Central Germany. The interpolation has been performed with a simple kriging interpolation which was constrained by 108 measurements conducted on rock cylinders which were extracted from the rock slab (Hornung et al. 2020, ©EAGE).

In this work, we will use the term *property* for a regionalized variable, the term *field* for the geometric field or the spatial distribution of a property, and the term *spatial domain* for an area of interest.

2.2.2.1. Heterogeneity

Heterogeneity is subject of numerous geoscientific studies (Anyiam et al. 2017; Michie & Haines 2016; Mukerji et al. 1997; De Ros & Scherer 2013). In general, the term is defined as *'Difference or diversity in kind from other things'*, a *'Composition from diverse elements or parts'* or being of *'multifarious composition'* (Oxford Dictionary 2014). In most works, the term heterogeneity is used to describe that an object consists of multiple subsets being different from one another in one or more attributes. Li & Reynolds (1995) restrict the term to be the variability of a system property in three-dimensional space showing that the heterogeneity is linked to the variability of a property.

In the case of uncorrelated properties, the spatial distributions of rock or fluid properties in the subsurface provide different types of patterns (Linsel et al. 2020a). Hence, in order to estimate the heterogeneity of a system, each rock and fluid property must be considered individually or it must be checked for correlation among the properties of interest (Gu et al. 2017).

Jiang (2014) differentiates two ways of thinking regarding geospatial heterogeneity, namely, the Gaussian and Paretian way. The Gaussian way assumes that all things are more or less similar in size and can be characterized by a well-defined mean and standard deviation. The Paretian way, however, assumes things to be highly non-linear and non-stationary. Within this Thesis, we will discuss aspects related to both assumptions.

Fitch et al. (2015) provide a set of methods to quantify heterogeneity within a set of petrophysical observations. The resulting measures are the coefficient of variation (c_v, Eq. 2.3) and the Dykstra-Parsons coefficient (c_{dp}, Eq. 2.4). c_v can be calculated by

$$c_v = \frac{\sqrt{\sigma^2}}{\overline{x}},$$ (2.3)

where σ is the standard deviation and $\overline{x}$ is the arithmetic mean whereas c_{dp} is defined by

$$c_{dp} = \frac{x_{50} - x_{84}}{x_{50}},$$ (2.4)

where x_n is the nth percentile of a set of numeric values.

Another method for heterogeneity quantification is the Lorenz-Curve (Heumann et al. 2016) where an univariate data set $\mathbf{x} = x_1, x_2, ...x_n$ is sorted ascending and the cumulative

value at the observation a is displayed in ratio to the total cumulative value of the data set yielding u_i (Eq. 2.5), like

$$u_i = \sum_{a=1}^{m} x_a \Big/ \sum_{k=1}^{n} x_k \quad \text{with} \quad a = 1, ..., m \quad \text{and} \quad u_0 := 0. \tag{2.5}$$

2.2.2.2. Anisotropy

Experimentally derived readings may show a directional dependence. The orientation and magnitude of directional dependency of a property is called anisotropy. Anisotropy can be caused by any geological feature (grain, fracture, bounding surface, etc.) which has a dominant alignment. If the magnitude of a property of an infinitesimal domain being a subset of a continuous medium is identical in every Cartesian direction, it is quasi-isotropic.

Anisotropy is often simplified or completely neglected in geological 3-D models and physical simulations. On the one hand, this is due to scarce data sets created in time-critical field- or laboratory campaigns. On the other hand, properties would have to be considered as tensors in numerical forward simulations, which would increase the computational effort significantly. For instance the hydraulic permeability (k), regardless of considering a continuous rock medium or a fractured formation, commonly exhibits distinct directional dependencies (Ringrose & Bentley 2015). The mathematical expression of this variable would change from a scalar to a tensor (Eq. 2.6), which would require solving a system of linear equations at each node of the target mesh. In general, tensors may be symmetric as it is in Equation 2.6, or antisymmetric. Visualization of complex tensor fields is still subject to research (Kratz 2013).

$$k = \begin{pmatrix} k_{xx} & k_{xy} & k_{xz} \\ k_{yx} & k_{yy} & k_{yz} \\ k_{zx} & k_{zy} & k_{zz} \end{pmatrix} \tag{2.6}$$

We will differentiate two types of anisotropy, namely, the zonal anisotropy and the polar anisotropy. When the spatial distribution of a property provides a trend such as bedding, we speak of a zonal anisotropy. A zonal anisotropy can be detected with the help of a directional semivariogram. The grade of anisotropy can, for instance, be expressed by the ratio of the ranges between two directional semivariograms, which both consider different spatial directions (Ringrose & Bentley 2015). The directions are not limited to the standard axes of a 3-D Cartesian coordinate system, namely X, Y and Z axis, rather

than representing a generic spatial vector which can be composed of any combination of X, Y and Z respectively. This makes the zonal anisotropy being an infinitesimal problem.

A polar anisotropy is what we understand as the classical direction dependence of a property. Although the polar anisotropy is an infinitesimal problem too, in most cases this problem is reduced to the standard axes of the Cartesian coordinate system. The determination of the full tensor or anisotropy principal axes of properties is particularly important for, e.g., simulating the 3-D mass and heat transport in geological media during geothermal production (Popov et al. 2016; Popov & Mandel 1998; Popov et al. 1999) as well as for forecasting the production rate in hydrocarbon sites (Backeberg et al. 2017; Clavaud et al. 2008).

Following, we will describe the polar anisotropy of elasticity and we will provide measures for the quantification of anisotropy under the simplifying assumption of transverse isotropy. Transverse isotropy assumes that one Cartesian axis is the major symmetry axis in the system that provides a diverging magnitude with regard to the other two Cartesian axes. The elastic modulus tensor can be expressed as a fourth-rank tensor

$$
\mathbf{C} = \begin{pmatrix}
C_{11} & (C_{11} - 2C_{66}) & C_{13} & 0 & 0 & 0 \\
(C_{11} - 2C_{66}) & C_{11} & C_{13} & 0 & 0 & 0 \\
C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\
0 & 0 & 0 & C_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & C_{44} & 0 \\
0 & 0 & 0 & 0 & 0 & C_{66}
\end{pmatrix}
\tag{2.7}
$$

where C_{ij} represents an elasticity modulus and the indices are related to the directional P- and S-wave velocity, under the assumption that z is the symmetry axis, by

$$
v_p^z = \sqrt{\frac{C_{33}}{\rho}}
\tag{2.8}
$$

$$
v_s^z = \sqrt{\frac{C_{44}}{\rho}}
\tag{2.9}
$$

where v_p is the P-wave velocity and v_s is the S-wave velocity parallel to the symmetry axis and ρ is the bulk density. The anisotropy, here expressed for the P-wave polar anisotropy, can be quantified with the Thomsen parameters (Thomsen 1986), which include, among others, ϵ that is calculated by Equation 2.10.

$$\epsilon = \frac{C_{11} - C_{33}}{2C_{33}}. \tag{2.10}$$

If $\epsilon \ll 1$ the material can be classified as weakly anisotropic.

2.2.3. The Representative Elementary Volume

Investigations of subsurface regions are commonly performed at multiple scales. Numerous studies showed that physical variability in geological media must be integrated as a function of measurement volume, also known as the representative elementary volume (REV, Nordahl & Ringrose 2008). The REV denotes a volume where a representative amount of heterogeneity is captured by a measurement device (Nordahl et al. 2014) minimizing the smaller-scale fluctuations.
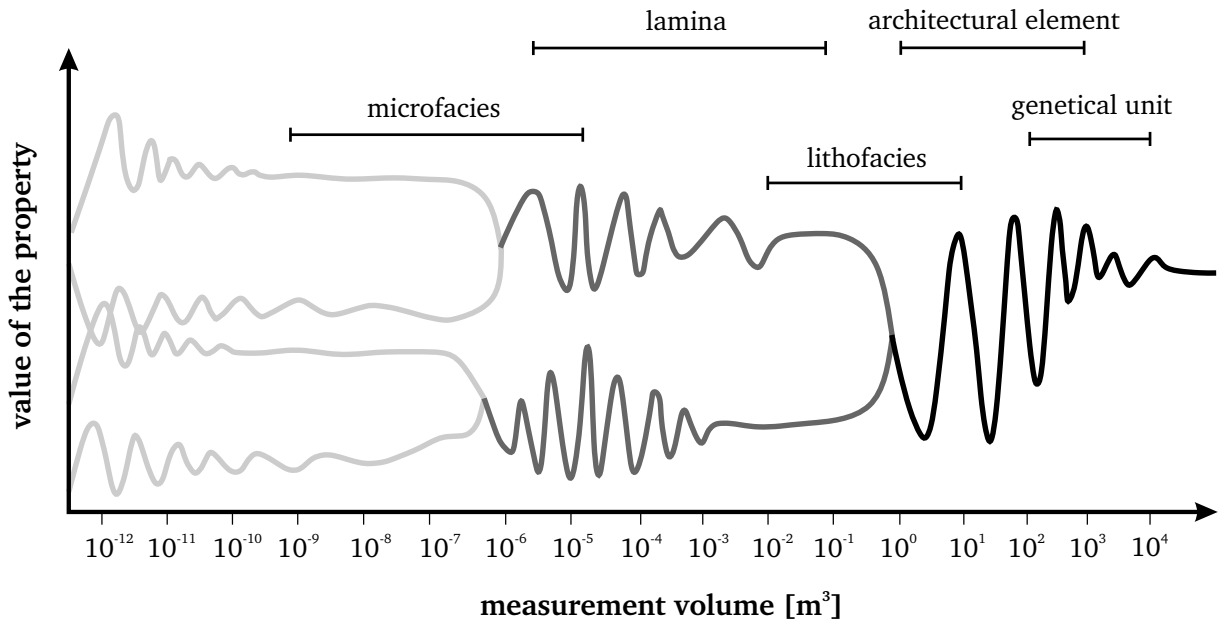
Figure 2.6.: Schematic of the representative elementary volume (REV) concept after Nordahl & Ringrose (2008) with integrated scales of the lithofacies concept according to Miall (1985).

Many approaches exist that try to organize the entire scale shown on the ordinate in Figure 2.6 based on geological processes. Therefore, hierarchical concepts such as the lithofacies concept have provided promising results in the past decades in order to capture the scale-dependence of physical properties with respect to the geological architecture. The concept of the REV emphasizes the fractal nature of geological media which must be

accounted for when taking rock samples from the field, interpolating rock properties in space or making statistical analyses using subsets of the total considered rock volume.

2.3. Subsurface Characterization

Ways of utilizing the subsurface comprise, for instance, groundwater extraction and storage, hydrocarbon production, carbon capture and storage, mineral resource mining, or geothermal heat storage and exploitation (Landa & Strebelle 2002; Heap et al. 2017; Kushnir et al. 2018; Rodrigo-Ilarri et al. 2017; Ito et al. 2017).

An area of interest in this regard is a region within the Earth's crust in which technically exploitable resources such as oil, gas, heat, minerals, or water are stored. Characterization of the subsurface is a crucial procedure before targeting a region of interest by boreholes or mining operations.

In terms of reservoirs, Fanchi (2002) defines reservoir characterization as '*the process of preparing a quantitative representation of a reservoir using data from a variety of sources and disciplines*.' Accordingly, as numerous physical and chemical properties contribute to a reservoir's potential, this problem is of a high dimensionality and complexity with multiple domains contributing towards it. Hence, regardless of whether considering hydrocarbon, geothermal or groundwater systems, carbon capture and storage or final disposal of radioactive material, a comprehensive knowledge on the subsurface architecture and the physical and chemical fields is substantial for making profitable and sustainable decisions in subsurface applications (Ailin 2012).

2.3.1. The Subsurface Characterization Process Chain

2.3.1.1. Data Acquisition and Data Management

Understanding the spatiotemporal distribution of rock and fluid properties in the subsurface is critical for the economic and sustainable usage of the subsurface (Tiab & Donaldson 2012; Gudmundsdottir & Horne 2018). Hence, we need quantitative information about physical and geological characteristics which is often provided as a huge number of heterogeneous, high-dimensional, cross-disciplinary data sets compiled into disparate data integration platforms. These data are produced in well log measurements, during reservoir production, in drill core analyses, outcrop analogue studies or in geophysical surveys and can be both static or dynamic. Information about the lithological architecture and a property's variability can be derived from both, the subsurface itself as well as outcrop analogues (Fischer et al. 2007; Howell et al. 2014; Aretz et al. 2015) that can be easily accessed.

Due to modern sensor systems and low-cost data storage, the number of data produced during exploration and exploitation is rapidly increasing in the last years. Hence, data sets produced in subsurface studies regularly extend the degree of complexity to be efficiently analyzed by domain experts with simple spreadsheet-based tools. Consequently, modern automatized technologies to explore and model subsurface-related data are required to increase the predictability of properties in the subsurface and eventually optimize the efficiency of usage (Das et al. 2015). Some authors such as Cannon (2018) even declare that the process of data management might be the most important entity within the subsurface characterization process chain.

As subsurface-related data is mostly compiled from numerous domains with highly diverging ontologies (Ge & Chen 2010), the terminology among each domain might be inhomogeneously defined. This requires that the terminology being used is homogenized prior to collecting relevant data (Greb & Linsel 2020). Moreover, as data for subsurface characterization is produced in multiple both parallel and sequentially performed process steps, it is crucial that the involved domain experts are allowed to integrate data continuously throughout the entire process. This requires that the group of domain experts must be enabled to perform intermediate analyses in order to optimize the sampling strategy, for planning the exploitation strategy, or for making other business-critical decisions. Figure 2.7 illustrates a summary of the process chain of data acquisition during a subsurface study.

2.3.1.2. Constructing the Rock Model

Rock models are computational representations of subsurface regions which spatially resemble the distribution of distinct rock bodies. Rock models can be useful to more accurately control spatial modeling or simulation of rock and fluid properties which contribute towards the region's economic potential. Furthermore, e.g, rock models can be considered for a better understanding of a reservoir's behavior during production.

A geological rock model consists of building blocks such as lithofacies types, genetic units, architectural elements, depositional environments, or diagenetic units (Fig. 2.8). Which type of building blocks should be used depends on the scale, purpose and economic restrictions of the projects. Rock models are constructed by either object-based approaches which aim to resemble geological objects based on their characteristic geometry (channel-shaped, elliptic, longitudinal, etc.; see Holden et al. (1998)). Alternatively, pixel-based methods including indicator kriging (IK), nearest neighbor (NN) classification,
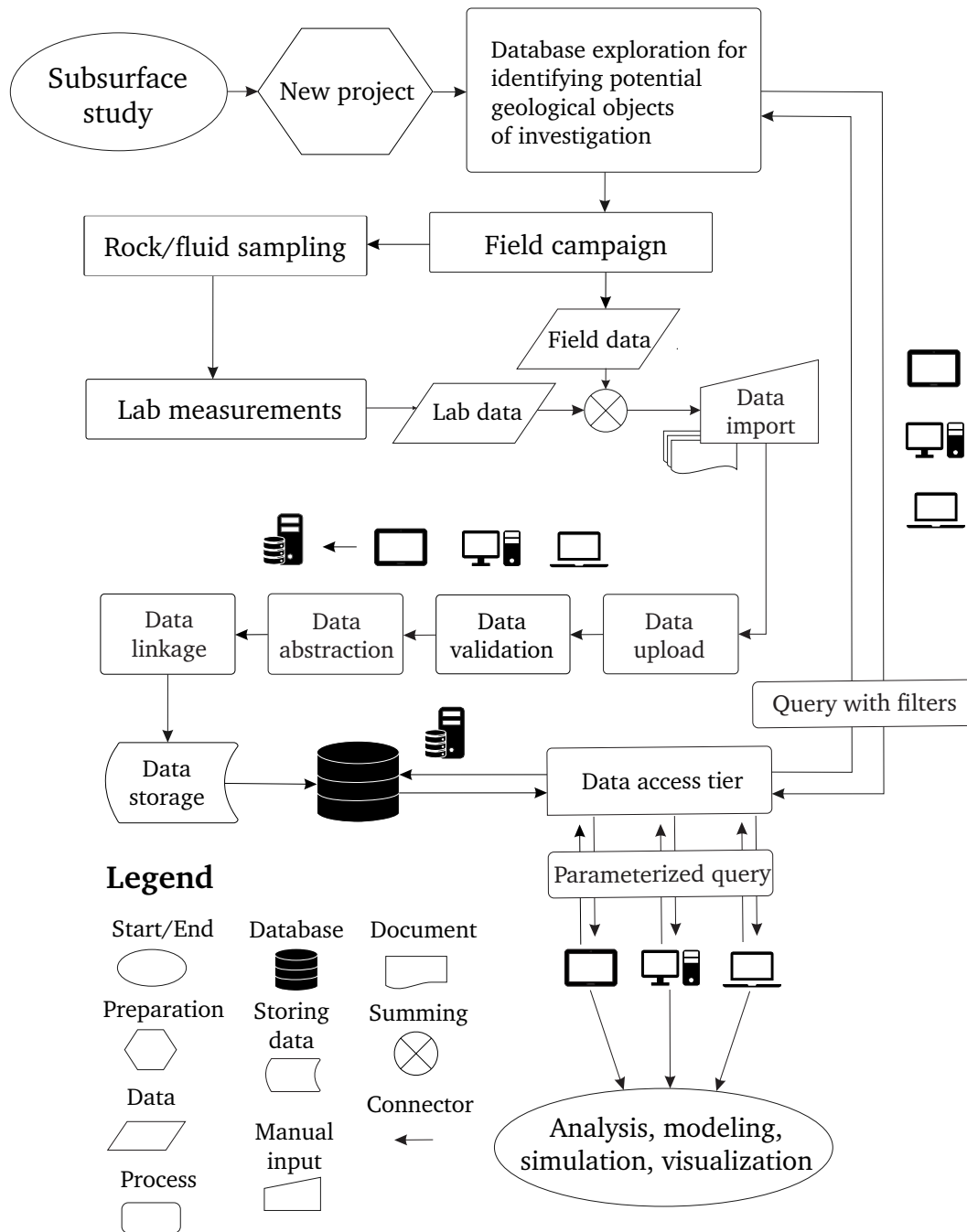
Figure 2.7.: Illustration of the end-to-end process of subsurface characterization.

or sequential indicator simulation (SIS) can be used to model geological bodies on a cell-to-cell basis (Journel & Alabert 1990).

The scale of investigation plays a major role in the investigation of reservoir rocks. Geological heterogeneities – even at small scales – play a key role here as they may constitute undesirable features in the subsurface such as flow-barriers in reservoirs (Landa & Strebelle 2002; Ringrose et al. 1993; Medici et al. 2016; Medici et al. 2019), pathways in
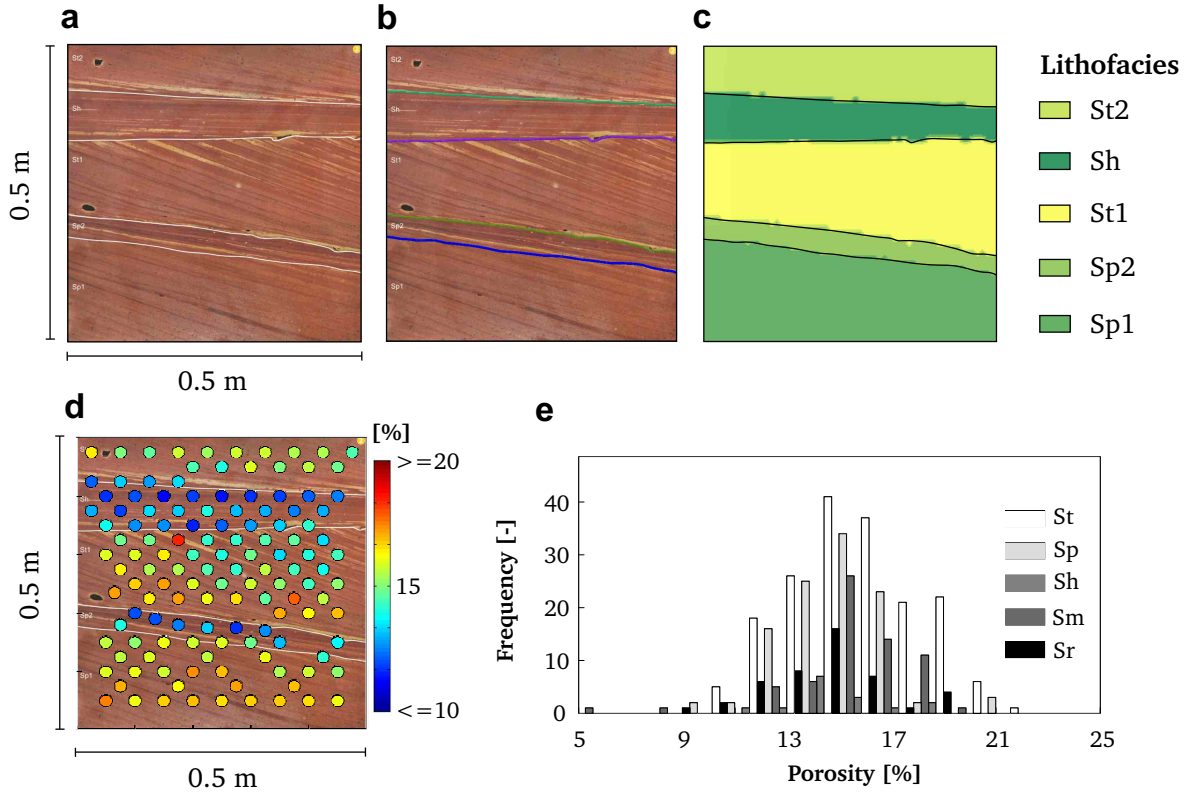
Figure 2.8.: Illustration of the rock modeling process for geological media. The example shows a rock slab of the Buntsandstein which was investigated in **Publication II**. **(a)** Photograph of the rock medium. **(b)** Computational mapping of the bounding surfaces. **(c)** Nearest neighbor modeling result at the scale of lithofacies types. **(d)** Sampling locations at which rock cylinders with $\varnothing = 2$ cm were taken. The locations are colored based on the porosity value that has been measured at the extracted rock sample. **(e)** Empirical histogram of the porosity measured on the rock cylinders from **(d)** (Hornung et al. 2020, ©EAGE).

radionuclide repository sites (Kiryukhin et al. 2008) and in contaminated sites (Tellam & Barker 2006) or geochemical anomalies in mining areas (Wang & Zuo 2018). Hence, the controlling factors of sub-meter variability should be understood and at least roughly quantified before starting the development in the subsurface region.

2.3.1.3. Constructing the Property Model

One of the most important steps in subsurface characterization involves the construction of property models (Ringrose & Bentley 2015, Fig. 2.9) for a target region. In hydrocarbon extraction, for instance, the main purpose of property models is to improve the understanding of fluid distribution in the reservoir itself and to capture the heterogeneity in such a way that the dynamics of fluid flow can be modeled more realistically (Canon 2018).
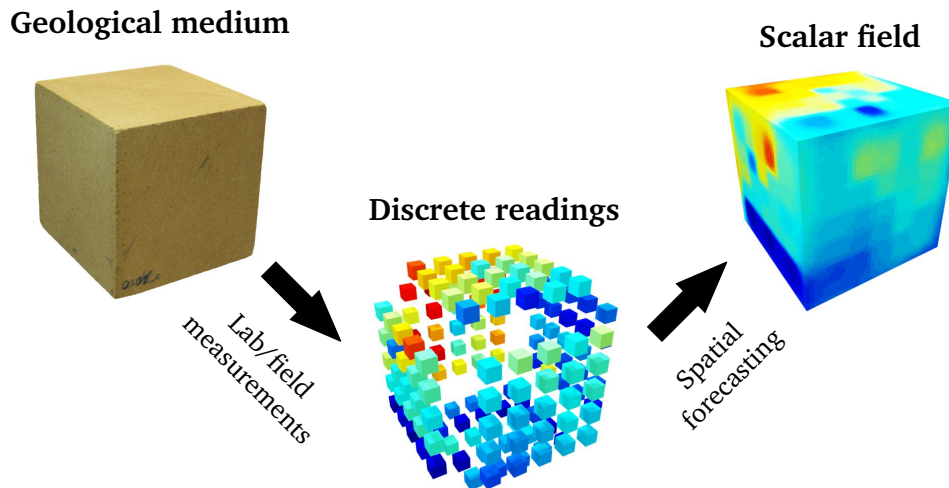
Figure 2.9.: Illustration of the procedure of constructing a subsurface property model. Here, a rock cube being composed of sandstone was extracted from a quarry and measured in order to determine the spatial distribution of the apparent permeability. **Publication III** contains the study this work flow has been performed along.

The cut-offs in petrophysical properties have been identified as a major cause of poor estimation of reserves in place in a reservoir (Landa & Strebelle 2002; Ringrose et al. 1993). The scarceness of adequate petrophysical data (such as porosity and permeability) is widely observed (see, e.g., Landa & Strebelle (2002) and Bär et al. (2020)) but – if available – should be integrated into a study in order to enrich information about the target region. Optimally, the region of interest can be computationally represented by both a rock model and a property model as accurately as possible.

Property models can be used to bridge the gap of missing data in the target-specific subsurface, which cannot be fully surveyed by well logs due to economic and time limitations.

Apart from hydrocarbon or geothermal energy extraction and carbon capture and storage, subsurface property modeling plays a critical role in mining industries (Wang & Zuo 2018) and in the assessment of potential final radionuclide repositories. These models can be used for planning the mining strategy, the pumping rates of production wells and the placement of injection wells in order to maximize reservoir recovery.

Rock property models can be constructed by different techniques. The most widespread group of techniques for determining properties in unknown space are the interpolation and simulation algorithms. Apart from those techniques, properties of geological media can be determined by geophysical measurements, regression, or machine learning techniques such as support vector regression (Abbaszadeh Shahri et al. 2020). The most important algorithms will be described in detail in section 2.4.4.9.

2.4. Geovisual Analytics for Subsurface Characterization

2.4.1. Introduction

The rapid increase of the storage capacity and computing power of personal computers in the last decades placed modern techniques to analyze large data sets into the forefront of current research. Those techniques comprise knowledge discovery, machine learning, data visualization and data mining in multidimensional data sets finding their way not merely into the domain of computer sciences but also far beyond.

Systems for visual knowledge discovery enable a domain expert to deploy data analysis and visualization algorithms onto data sets in order to yield hidden models or patterns. The visual knowledge discovery system GeoReVi, which has been developed in the frame of this Thesis, incorporates elements of three disciplines, namely, visual analytics (VA), geovisual analytics (GVA) and knowledge discovery in databases (KDD), which will be described in detail in the following sections.

2.4.2. Visual Analytics

The term *visual analytics* (VA) came up in 2005 being defined by Thomas & Cook (2005) within the frame of The Research and Development Agenda for Visual Analytics. Here, VA is defined as '*the science of analytical reasoning facilitated by interactive visual interfaces*.'

As the name implies, VA '*combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets*' (Ellis & Mansmann 2010). With the '*human in the loop*', a system for VA aims to enhance the integration of an expert's domain knowledge into the analyses of big data sets. The scope of VA comprises a series of interdisciplinary analysis techniques (Fig. 2.10).

VA is a widespread yet young uprising science which has been more and more integrated into other fields of research in the last decade. Enabling experts to interactively visualizing and understanding complex multidimensional systems (Munzner et al. 2006), computer-aided VA can be used for enhancing the decision-making for very complex problems (Kohlhammer et al. 2011), such as the diagnosis and classification of brain tumors (Glaßer et al. 2013). In recent years, technologies and tools from VA were subject to research in the geosciences too, especially in E&P industries where VA is implemented in various software packages (e.g. RAVA [Brazil et al. 2018]). In geosciences and related

industries, the research of VA and the development of associated VA tools, especially those with an open source structure, are yet still underrepresented.
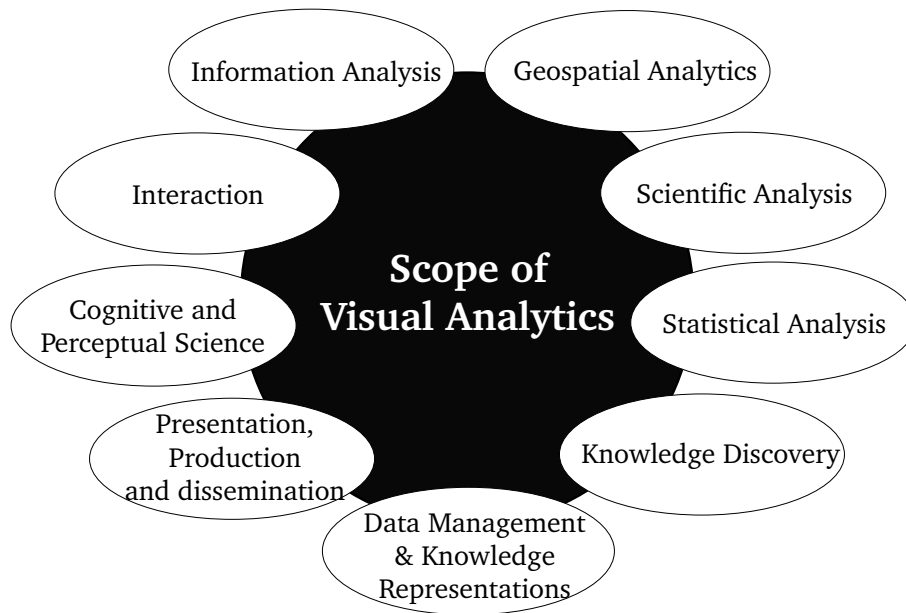
Figure 2.10.: Visual representation of the scope of visual analytics according to Keim et al. (2008).

2.4.2.1. Information Visualization

Visualization is the means through which humans and computers can communicate, cooperate and interact (Andrienko et al. 2010). According to Munzner et al. (2006), '*visualization is fundamental to understanding models of complex phenomena*' and, by definition, this discipline plays a major role in the context of VA.

Human beings are capable of detecting structures, features, patterns, trends, anomalies, and relationships in visual structures. Those visual structures may be agglomerates of simple geometries, numbers, or objects known from real life. Building up onto this capability, visualization aims to support a human being to discovering patterns in complex data sets. Thereby, based on the definition of Fayyad et al. (2002), a perfect visualization system harnesses the perceptual capabilities of the human visual system.

From a computational perspective, a visualization is a visual representation of data. According to Mazza (2009), the process of generating a visual representation of data is composed of three steps, namely, data preprocessing and data transformation, visual mapping as well as view creation (Fig. 2.11). On the way from raw data to the visual end-product the human expert can control how data are preprocessed and transformed before being mapped onto a visual structure.
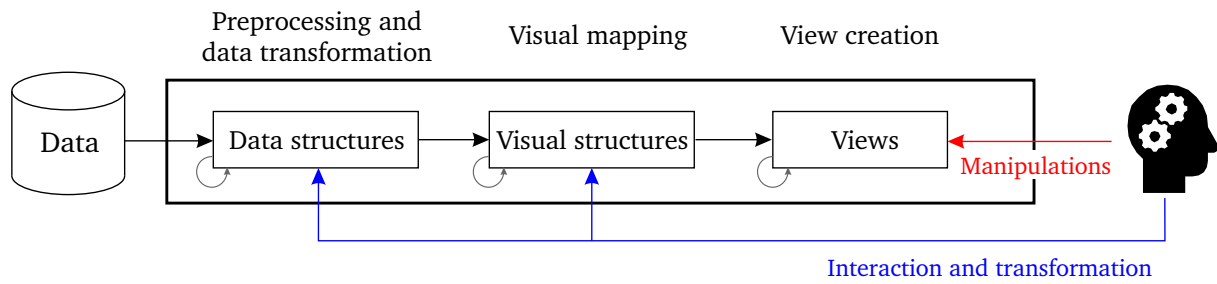
Figure 2.11.: Conceptualization of the process of generating a visual representation of data after Mazza (2009).

Mazza (2009) differentiates three types of visual structures, namely, the spatial substrate, the graphical elements, and the graphical properties, which may be used to map data onto graphical representations.

The spatial substrate defines the dimension of the physical space inside which the visual representation is located. This usually involves zero to four dimensions in geospatial analyses, which are represented by the Cartesian coordinates x, y and z together with the time dimension.

The graphical elements involve those discrete elements, an object can be represented by, including points, lines, surfaces and volumes (Fig. 2.12).

The graphical properties comprise those properties a human being can recognize and interpret by means of the human visual system. These include the size, orientation, color, texture and shape of a visual structures.
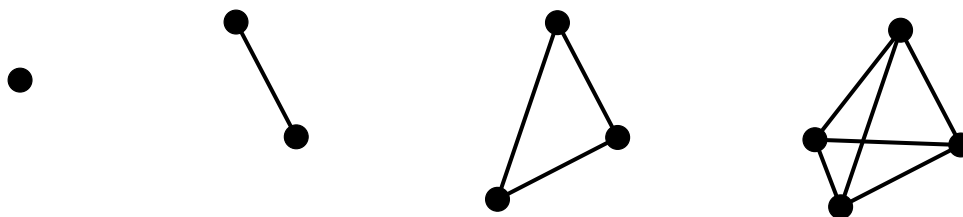
Figure 2.12.: Graphical elements for generating visual structures of objects including a point, a line, a triangle and a tetrahedron (from left to right).

The visual end-product, named view, is a compilation of one or more visual structures. The human expert can perform manipulations on the visual structures of the view or use it to gain or communicate knowledge.

In this Thesis, the knowledge generation model for visual analytics (KGMVA) after Sacha et al. (2014) is applied (Fig. 2.13). This model represents a process, in which a human being extracts knowledge from a big data set through continuous interaction with a computer-

aided cycle of data selection, modeling and visualization. An important characteristic of such a VA system is the continuous interplay of automatic background processes and interactive visualization.
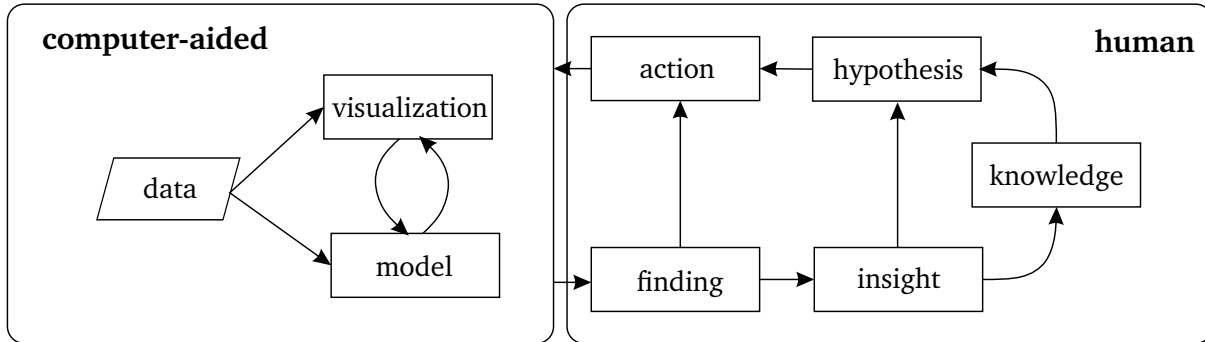
Figure 2.13.: Knowledge generation model for visual analytics (KGMVA) after Sacha et al. (2014).

2.4.3. Geovisual Analytics

Information visualization is fundamental for geoscientists to discovering new knowledge in complex data sets. In the earth sciences, visualization is often referred to as *geovisualization* (DiBiase 1990). Geovisualization is defined as '*a loosely bounded domain that addresses the visual exploration, analysis, synthesis, and presentation of geospatial data by integrating approaches from cartography with those from other information representation and analysis disciplines, including scientific visualization, image analysis, information visualization, exploratory data analysis, and GIScience*' (Dykes et al. 2005).

Ultimately, the domain of geovisual analytics (GVA) has evolved from geovisualization and VA (Kraak 2008). According to Andrienko et al. (2007), GVA is a sub-domain of VA which distinguishes from its mother-domain by (1) dealing with the complex nature of spatiotemporal data, (2) having multiple actors with (3) tacit criteria and knowledge.

Thus GVA aims at supporting space-time related decision making by enhancing the human's capability to analyze, envision, reason and deliberate through computer-techniques (Andrienko et al. 2007). Moreover, an important aspect of GVA is to start data exploration without any hypotheses (Schiewe 2013).

Although the terms of VA and GVA firstly came up in 2005 and 2008, respectively – see section 2.4.2 — the techniques involved in those disciplines, such as, among other, statistical data analyses or data visualization, were discovered and described much earlier.

GVA-based software tools should comply with the following requirements (Andrienko et al. 2010):

1. Seamless integration of visualizations with computational techniques such as spatial statistics, time-series analysis, simulation models, spatiotemporal data mining and so on;
2. Support for documenting the analysis process; keeping provenance of findings, reporting and story-telling;
3. Support for collaboration.

Recently, many new open-source software tools found their way into the geoscientific communities which support researchers in the modeling and analysis of complex subsurface regions (see, e.g., Wellmann et al. (2019)). Yet there is still a technology gap for a software tool which assists researchers along the entire process chain of subsurface characterization. The domain of GVA though provides the means to implement a tool for representing this process chain. Hence, when planning and implementing the software system in the frame of this Thesis, special attention had been paid to comply with the requirements for a GVA-based system.

In the following sections, the spatiotemporal data structures as they are required to computationally depicting subsurface data in the context of GVA will be described.

2.4.3.1. Computational Representation of Subsurface Data

As outlined in section 2.2.1, subsurface regions are hierarchically structured, the elements of which are represented by geological bodies (3-D elements) and bounding surfaces (2-D). Moreover, samples or measurements taken from drill cores or outcrops are commonly represented by 1-D structures. If subsurface properties such as the hydraulic head or subsidence are analyzed, we need to take the time-dimension (t) into account too.

Spatiotemporal Data

Spatiotemporal data are data which are provided with information about space and time. Most entities connected to subsurface analysis require a geographic or local reference as well as an orientation in order to be usable. In geoscientific subsurface analyses spatiotemporal objects are measurements, rock or fluid samples, or geological bounding surfaces.

Those objects are typically provided with a time information together with coordinates from the World Geodetic System 1984 (WGS84) datum in degree-order format (Decker
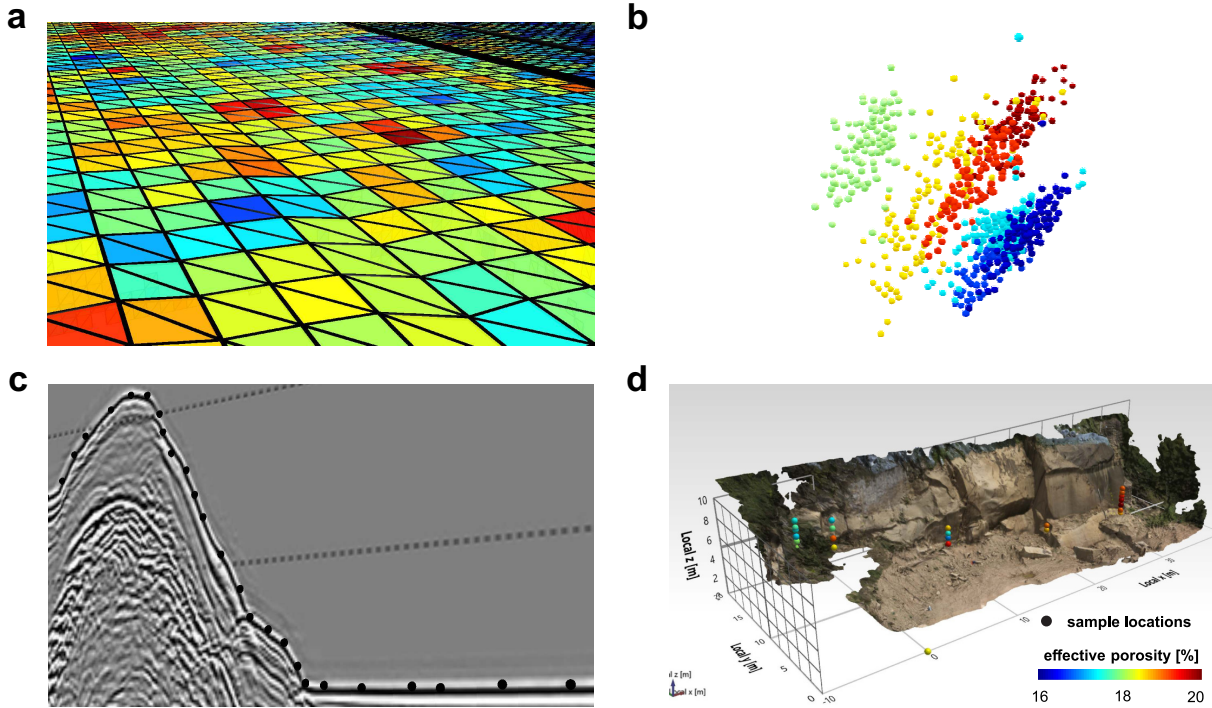
Figure 2.14.: Samples produced with GeoReVi. **(a)** Geocellular 3-D model; **(b)** Result of a principal component analysis projected into 3-D space; **(c)** Mapping a bounding surface from a seismic section; **(d)** Visualizing effective porosity measurements from rock samples taken from a quarry wall in 3-D space. The data for these analyses come from **Publication I** and **Publication II** and the seismic transect is a public data set provided by the Bundesanstalt für Geowissenschaften und Rohstoffe[2].

1986). The WGS84 datum was developed by The Defense Mapping Agency in 1986 as a global basic reference frame for a multitude of mapping tasks. The coordinates of the WGS84 datum consist of a latitude, longitude and elevation. One limitation of the geographic WGS84 is that it is only an approximation to the Earth's geoid and thus provides significant accuracy drawbacks especially at high latitudes.

Geographic coordinates from the WGS84 datum often must be converted into a metric coordinate system in order to apply spatial analyses. As an example, the Mercator projections can be considered which are discussed in detail in Osborne (2013).

$$\mathbf{x} = (x, y, z, t) \tag{2.11}$$

A metric spatiotemporal datum in general is represented by three Cartesian coordinates (x, y, and z) which represent the location at a particular point in time (t) (Eq. 2.11) with respect to an origin which is located at at x=y=z=0.

---

[2]https://www.bgr.bund.de/EN/Themen/GG_Geophysik/Marine_Geophysik/Seismik/Bilder/seism_verfahren_2_reflexionsseism_sektion_p_en.html, last access 23rd November, 2020.

Orientation of Geological Bounding Surfaces

Geological bounding surfaces are responsible for the compartmentalization of a geological region into multiple sub-regions, called compartments. In the case of a planar surface, such bounding surfaces can be characterized by two parameters: dip direction and dip angle. Usually, however, bounding surfaces in the subsurface are not planar which requires to statistically characterize them by a reasonable number of measurements. In sedimentary environments, layers can be characterized by internal stratification surfaces and by external bedding surfaces (Fig. 2.15). The bedding surfaces are those, which can be mapped in order to discretize a domain in the subsurface numerically (see section 2.4.3.2).
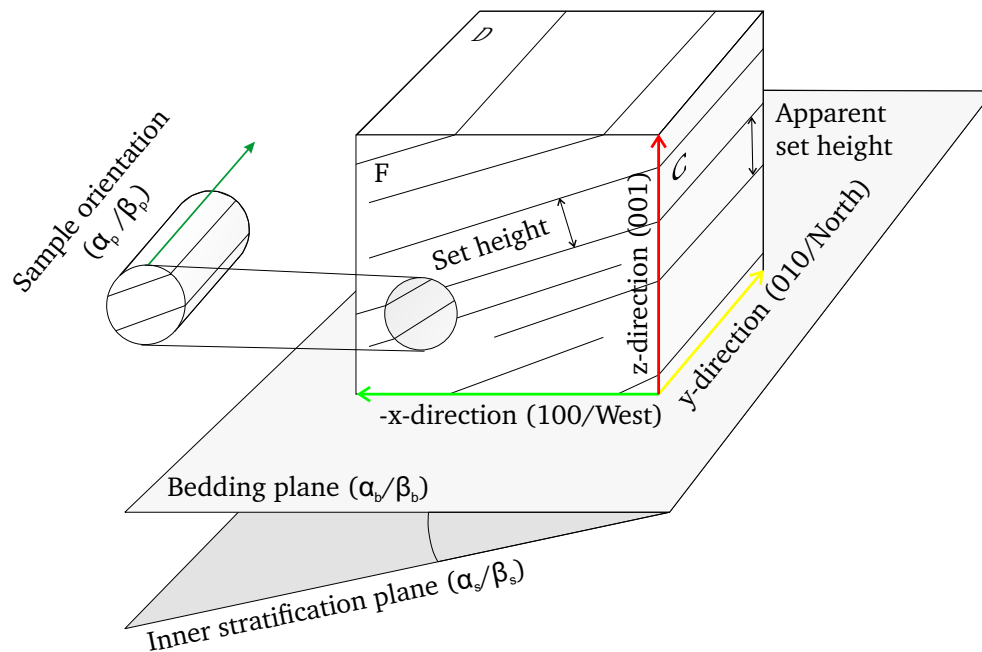
Structural element	Dip direction [°]	Dip angle [°]
Inner stratification	α_s	β_s
Bedding	α_b	β_b
Sample orientation	α_p	β_p

Figure 2.15.: Schematic representation of structures needed to capture the orientation of geological media and their internal bounding surfaces in a siliciclastic depositional environment.

Collection of Geoscientific Data

Geoscientific data can be collected in the field or in the laboratory. In the field, the geoscientist may derive data from objects of investigation such as outcrops, drill cores,

wells or even geophysical transects. During field campaigns samples such as rock cylinders, cuboids, handpieces, thin sections or water samples can be extracted from previously mentioned objects of investigation. Such samples serve as targets for measurements in the laboratory, if measurements are not conducted within the object of investigation itself. The generated field and laboratory readings can be recorded and directly referred to the object of investigation.

Each object of investigation, sample, sub-sample and field/laboratory reading can be spatially localized using either geographic coordinates — e.g, in the WGS84 datum using the decimal degree format — or by a local reference system using metric x, y and z coordinates which represent easting, northing and altitude respectively. Field measurements are directly taken on or within an object of investigation. The spatial reference is documented as a position relative to an origin, which is likewise located within an object of investigation (Fig. 2.16). Laboratory measurements in turn are always connected to a sample. Hence, when retrieving data, the local coordinate vector of the laboratory measurement should be added to the local coordinate of the extracted sample, which in turn should be added to the projected global coordinate of the object of investigation to precisely locate each measurement in space. Equation 2.12 represents the localization vector with o standing for the object of investigation, s for a sample, m for a reading, g/l for global/local, respectively, and p for a projected point.

$$\mathbf{x}_{m,g} = \mathbf{x}_{o,g,p} + \Delta \mathbf{x}_{s,l} + \Delta \mathbf{x}_{m,l} \qquad (2.12)$$

2.4.3.2. Spatial Discretization

Most types of geoscientific problems — and subsurface characterization in special — are limited to a finite spatial domain Ω which can be represented by a finite 1-D, 2-D or 3-D geometry, or also called mesh. For solving a partial differential equation numerically, for instance, Ω must be discretized into finite elements Ω_p. Those finite element meshes build the base for problems which can be solved by the finite element method (FEM, see section 2.5.3). Furthermore, the construction of finite element meshes may be a requisite step for performing interpolations, extrapolations or geostatistical simulations within Ω.

Especially in those subsurface regions, in which multiple irregular boundaries are intensely compartmentalizing the domain, adequate discretization can be a challenging task. Therefore, different discretization methods exist to partition Ω into compartments. Here,
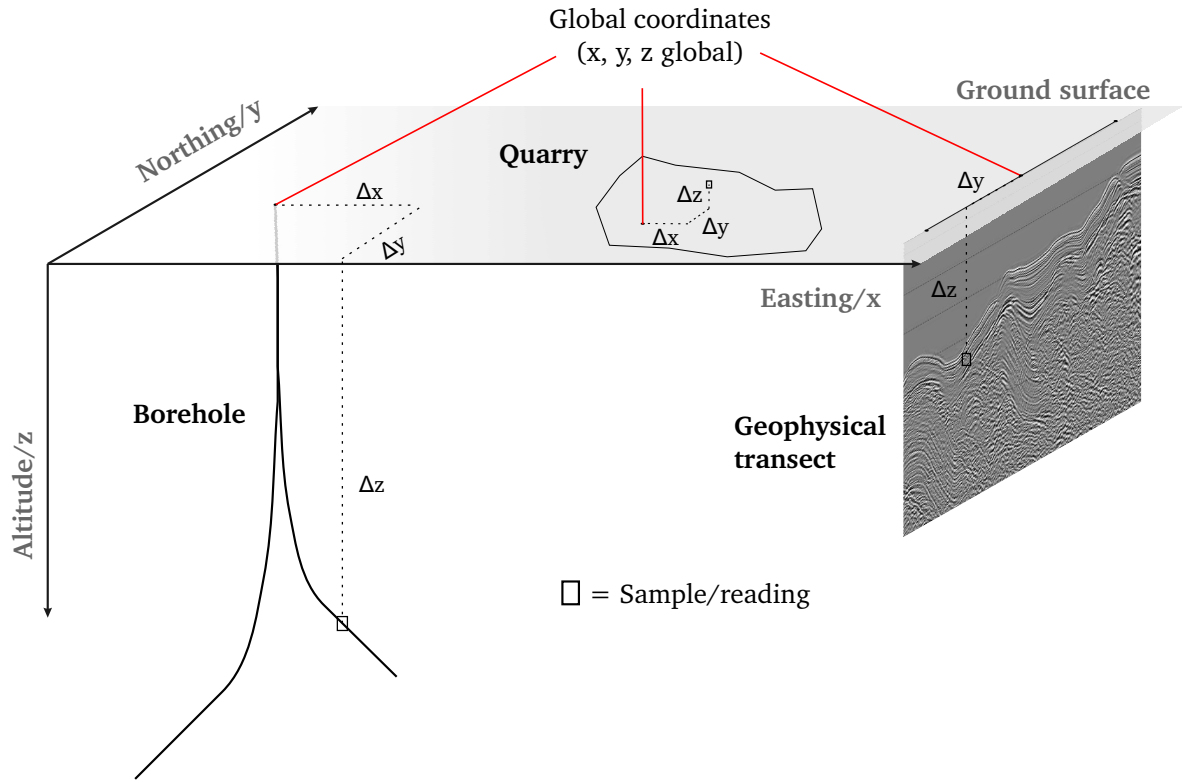
Figure 2.16.: Localization of objects of investigation and from samples and readings which were taken from those (not to scale).

two types of discretization methods are commonly applied: the regular and irregular mesh generation (Liseikin 2010).

The Elements of a Mesh

Geological subsurface domains are usually modeled in the form of meshes. Those meshes consist of simple shapes which aim to represent subsets Ω_p of Ω. Considering a 3-D domain, one of these shapes is, for instance, the hexahedron (Fig. 2.17) which consists of eight nodes and six faces. This 3-D geometry can be constructed by six quadrilaterals, which are 2-D shapes in turn and consist of four nodes. We already introduced some of these shapes in section 2.4.2.1. The relation of elements, faces and nodes within a hexahedral mesh can be determined by

$$6 \cdot n_e = 2 \cdot n_i + n_b, \tag{2.13}$$

where n_e is the number of elements, n_i is the number of internal faces and n_b is the number of boundary faces.
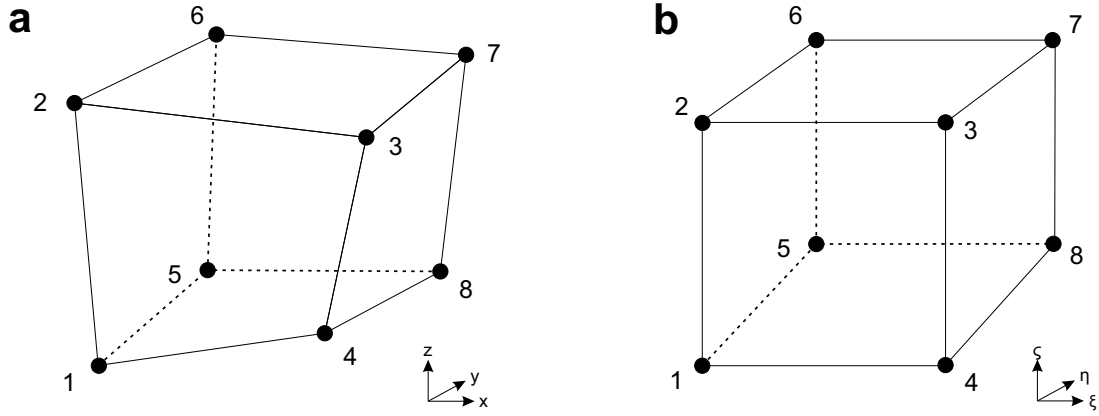
Figure 2.17.: Hexahedral element in physical space **(a)** and local space **(b)**.

Regular Grids

Regular grids are meshes in which each node is allocated to an index. In 2-D space the index is composed of two elements, namely, i and j, each of which is representing one direction. Those directions, however, do not necessarily correspond to the Cartesian axes (Fig. 2.18). In 3-D space, the index is composed of the three elements i, j, and k. In local space, i, j, and k correspond to the local coordinates ζ, ξ, and η. There are different types of regular grids existing: the coordinate grid and the boundary-conforming grid (Liseikin 2010).

The coordinate grid is a regular grid, the nodes of which are all providing the same distance to one another along one direction — e.g. step width of one kilometer along the X-direction. The boundary-conforming grid, however, is — as the name implies — bound to explicit boundaries and does not necessarily have the same step width along one direction.

Having two surfaces — one representing the base and one the top of a rock formation — a 3-D boundary-conforming grid, which is representing the rock formation in the finite domain, can be created. Therefore, the 2-D meshes of the surfaces are projected from a physical to a local coordinate system (Fig. 2.18).

Subsequently, the nodes from the two source data sets serve as boundary nodes for the new mesh. After projection — under the assumption that

$$i_{row} = (i - 1)(n_\eta + 1) + 1 \tag{2.14}$$
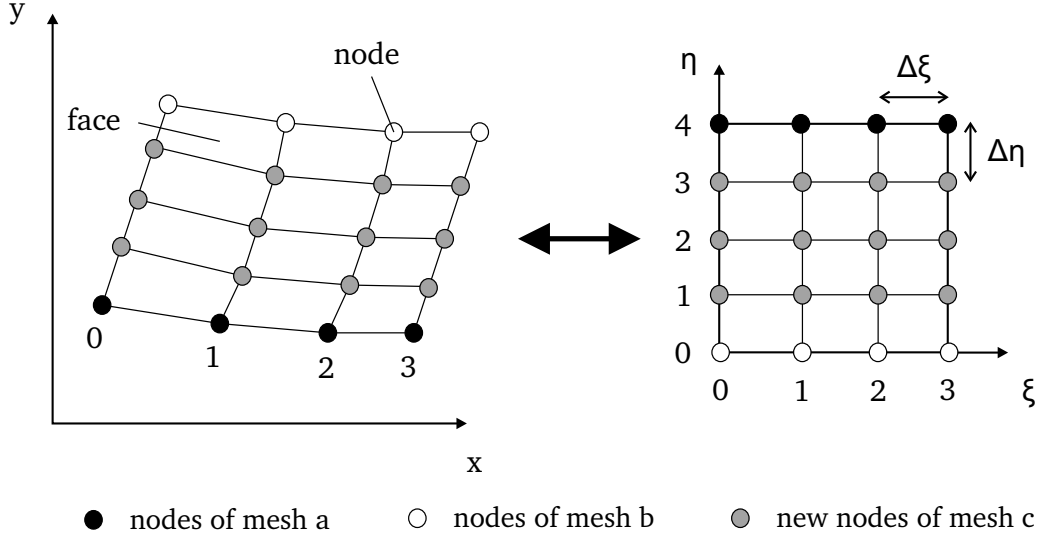
Figure 2.18.: Conceptual illustration of a mesh generation procedure using two constraint meshes (white and black) of dimensionality D = 1 with D = 2 as the target dimension.

and

$$i_{column} = mod((i - 1), (n_\xi + 1)) + 1,$$ (2.15)

with i_{row} as the number of the row, i_{column} as the number of the column and n_η/n_ξ as the number of elements in the specific direction — the points in between the two boundary meshes are created according to

$$x = \sum N_k(\eta, \xi)x_k$$

$$y = \sum N_k(\eta, \xi)y_k$$ (2.16)

where η and ξ are the transformed coordinates of a node and N_k as the nth element in the specific direction. Figure 2.19 shows a final 3-D boundary-conforming grid being constructed by interpolating the coordinates of those points, which are located between two bounding surfaces.

Mesh Statistics

A mesh provides characteristic properties like dimensionality, count of nodes, faces and cells or — dependent on the dimensionality – its bulk volume, area, or length, respectively.
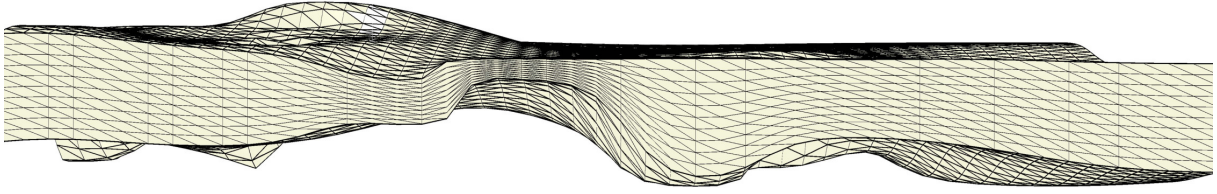
Figure 2.19.: 3-D representation of a curvilinear, regular grid. The model represents the volume between the top and the base boundary of the volcano-sedimentary Rotliegend unit in the Sprendlinger Horst area which is located in Hesse, Germany. The model has a span of approximately 3,000 × 70 m and is 10-fold exaggerated in vertical direction.

In case of a 3-D mesh, the volume is defined as

$$V_m = \sum_{i=1}^{n} V_i,$$ (2.17)

where n is the number of elements and V_i is the volume of the ith element. The volume of a tetrahedral element consisting of the nodes $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ is

$$V_t = \frac{|(\mathbf{a} - \mathbf{d}) \cdot ((\mathbf{b} - \mathbf{d}) \times (\mathbf{c} - \mathbf{d}))|}{6}.$$ (2.18)

In order to calculate the volume of a hexahedral element, it will be partitioned into six tetrahedral elements whose cumulative volume represents the volume of the hexahedral element like

$$V_h = \sum_{i=1}^{6} V_{t_i}.$$ (2.19)

Spatial Visualization of Subsurface Properties

Subsurface properties can be visualized in terms of their spatial location. Most commonly, such visualizations involve 1-D, 2-D and 3-D charts (Keim 2005). Meshes in general can be visualized by constructing each subset of Ω and adding to a chart object. In order to visualize a hexahedron, for instance, each face is constructed as a quadrilateral and added to the mesh geometry.

The color of the quadrilateral can represent either the property to be visualized or does not need to contain information at all. A property is most often visualized by introducing

a color bar that represents a specific value range. Each face of the mesh is then assigned a color from that color bar based on the values held by the constructing nodes. Usually a sort of mean value of the involved nodes' values is used to determine the color of the face.

If the time dimension is included, the color of the element is determined based on its value at a specific time t. This technique is especially useful when visualizing time series or intermediate results of FEM simulations.

Subdivision of Graphical Elements

The subdivision of graphical elements can be regarded as a smoothing process. All polygonal meshes can be partitioned using the Catmull–Clark scheme (Catmull 1974) in order to smooth the visualization. With this scheme, a new point in a quadrilateral is calculated by

$$\mathbf{x}_j^{k+1} = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{x}_i^k, \tag{2.20}$$

with $\mathbf{x}_j^{k+1}$ as the new point at subdivision step $k+1$ in the center of the element j with n nodes at the subdivision step k (Fig. 2.20).
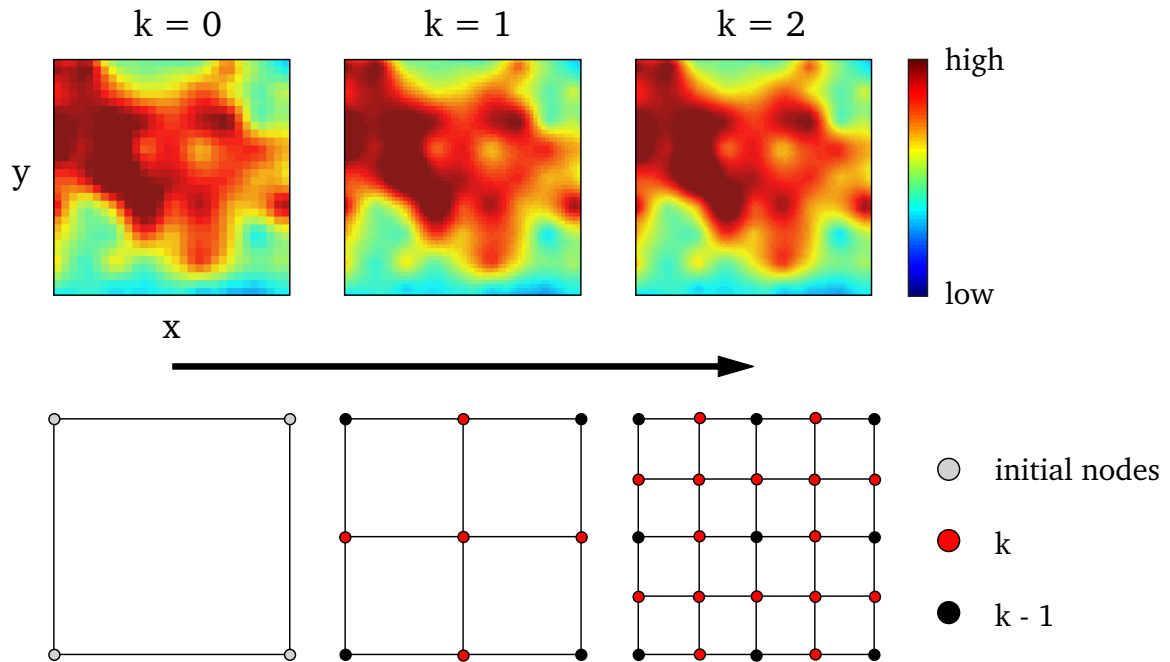
Figure 2.20.: Conceptualization of the Catmull-Clark scheme. The visual shows the spatial distribution of the Fe_2O_3 mass fraction measured on the XY-base face of rock cube OSB1_c (**Publication III**).

2.4.4. Knowledge Discovery in Databases

In the following section, the formalized process of knowledge discovery in databases (KDD) and its sub-disciplines will be described in detail. KDD builds the fundamental process the developed software system within the frame of this Thesis is built upon. KDD is considered a part of VA (Fig. 2.10).

Nowadays, big amounts of geoscientific data are being produced by automated data collection technologies and persisted in mature database management systems (DMS) (Das et al. 2015). Colloquially called *Big Data* (Shyr et al. 2018), these data sets cannot be easily analyzed by means of conventional methods due to their size and representational complexity. Consequently, in the late 1990s the concept of KDD found its way into the forefront of research which is aimed at systematically transforming big amounts of data into useful knowledge.

The interdisciplinary field of KDD comprises a set of semi-automatic, non-trivial methods to extract novel, understandable, valid and useful knowledge from one or more sets of domain-specific data sets stored in databases (Fayyad et al. 1996b, Maimon & Rokach 2010). The domain can be any discipline in which data are produced, managed and analyzed.

2.4.4.1. The KDD Process

Figure 2.21 illustrates the iterative, 8-stage KDD process which was firstly conceptualized by Brachman & Anand (1994) and redesigned and improved by Maimon & Rokach (2010). The process starts with the knowledge of a particular domain owned by the domain expert that is aimed to be maximized within the KDD process.

Core of the KDD process is a set of data mining (DM) algorithms, which are deployed on a processed data set in order to find characteristic patterns or models. Prior to DM, data are selected, projected, cleaned, reduced and transformed by the domain expert. Preprocessing is supported by computer-aided process automation and intelligent preselections.

DM algorithms in general comprise classification, summarization, correlation, regression, prediction and rule-discovery algorithms, whereby each DM algorithm is best suited for a specific problem. Within an iteration the expert can return to a previous task if, for instance, more data is needed to address a certain problem. Before a detailed explanation of the DM algorithms is provided, we will get familiar with a few definitions on data, information and knowledge.
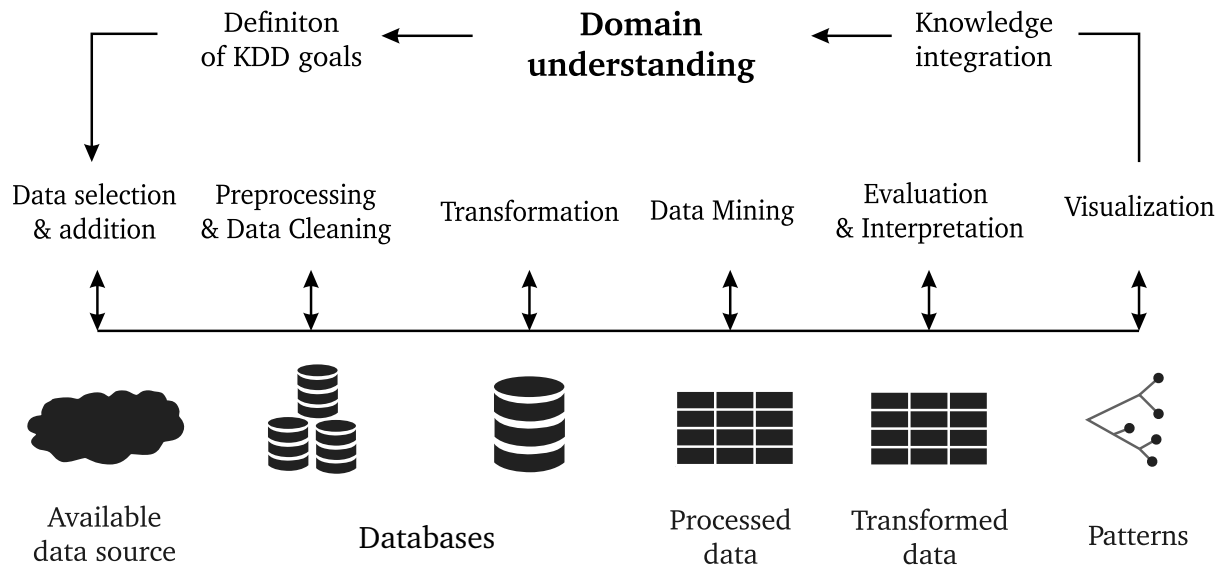
Figure 2.21.: Knowledge discovery in databases (KDD) process according to Fayyad et al. (1996).

2.4.4.2. The Knowledge Hierarchy

In the knowledge hierarchy, which is commonly represented by the data-information-knowledge-wisdom (DIKW) pyramid (Fig. 2.22, Rowley 2007), data is located on the lowermost hierarchy of algorithmic complexity whereas knowledge is located two layers above. The term data is defined very differently in multiple works. Most commonly, data is defined to be a set of "facts" whereas knowledge is a representation of an ordered subset of those facts in one or more symbolisms (Tab. 2.1). The definitions of data and knowledge are, however, still subject to scientific debate.

Information represents the transitioning layer between data and knowledge and is mostly defined as a representation of data in a structured, categorized or classified manner. Rowley (2007) provides definitions for the term wisdom too, which will not be discussed further on in our study.

Data itself can be subdivided into two major groups: structured data and unstructured data. Structured data are all kinds of data that exist in a structured format. A widespread example is a relational table where data is structured in rows and columns or, respectively, into tuples and attributes. Nevertheless, structured data can also be stored in non-relational structures such as images, or streams as long as they are organized in a

well-defined manner. The main difference between structured and unstructured data is that unstructured data are not organized in a well-defined data model (see section 2.4.4.2) which mainly involves, e.g., complex, text-heavy documents, or websites.

Since most parts of this study, are based on structured data, we will have a detailed introduction to those in this section. Structured data consist of instances of real or abstract objects with a set of attributes $\mathbf{A}$. An attribute can be assigned a value from a domain *dom* — not to be confused with a spatial domain — represented by a permissible subset of the universe $\mathcal{U}$ as $dom(\mathbf{A}) = \{a_1, a_2, ..., a_n\}$, where n is the number of elements in the domain of $\mathbf{A}$. Each element $a \in \mathbf{A}$ provides the data type of $\mathbf{A}$ which can be numeric, nominal or ordinal.

To demonstrate a simple example, the nominal attribute *color* of the instance *mineral* can take any symbol of colors like *green, yellow* or *red*. While nominal data types comprise all permissible symbols, numeric data types represent a number which may be an element of $\mathbb{R}$, $\mathbb{N}$, $\mathbb{Q}$, etc. Data can be constrained by rules which reduces the domain to a subset of $dom(\mathbf{A})$.

Nowadays, most data is stored and managed by database management systems (DMS). Those DMSs can be either integrated into distributed systems where a set of users is provided with custom access to data or they can be deployed locally as local data containers such as text or binary files. From those databases, data can be queried and analyzed. As most of the DMSs are still grounded on the relational model after Codd (1970), a brief explanation of this model will be given in section 2.4.4.3.
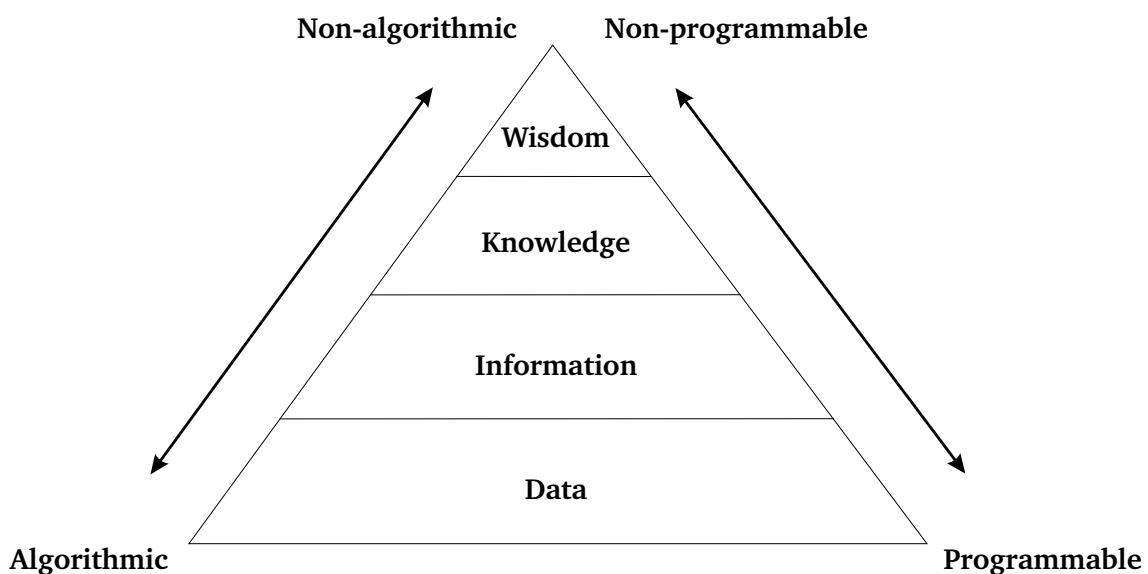
Figure 2.22.: Data-Information-Knowledge-Wisdom (DIKW) pyramid (Rowley 2007).

Authors	Definition		
	Data	Information	Knowledge
Zeleny (1987)	*knowing nothing*	*knowing what*	*knowing how*
Ackoff (1999)	*symbols that represent the properties of objects*	*answers to questions that begin with such words as who, what, when, where, and how many*	*answers to how-to questions*
Thierauf (1999)	*unstructured facts*	*structured data useful for analysis*	*obtained from experts based on experience*
Awad & Ghaziri (2007)	*unorganized, unprocessed, static facts*	*aggregation of data that makes decision-making easier*	*human understanding of a specialized field of interest*
Laudon & Laudon (2012)	*elementary and recorded description of things, events, activities and transactions*	*data that have been shaped into a form that is meaningful and useful to human beings*	*information combined with understanding and capability; it lives in the minds of people*
Jifa & Lingling (2014)	*most basic facts*	*adds context to facts*	*how to use facts*

Table 2.1.: Listing containing definitions from different authors for data, information and knowledge. The list has been adapted from Rowley (2007).

Data can be transformed into information by procedures like sorting, aggregation or contextualization. We refer to knowledge in a geoscientific sense to being patterns/models that are useful to recognizing and substantiating relationships among rock or fluid properties in space and time in order to assist the decision-making before, during and after subsurface utilization or research.

Logical Data Modeling

In order to store and query data that is generated in a subsurface study, databases must be modeled and implemented. Logical data modeling (LDM) represents a set of techniques to model real or abstract objects and their interrelations. In an iterative modeling process (Tillmann 1993) an abstract representation of objects, attributes and relationships is hereby created.

A data model accordingly consists of entities, attributes, relationships and cardinalities. Entities are conceptual representations of a real or abstract object providing a finite set of attributes. Those objects are logically connected through relationships being of a particular type (cardinality). Entities can be subdivided into strong entities, which are independent from other entities, and weak entities which are a composite of a set of attributes from strong entities. For instance, a *user* and a *project* are strong entities whereas the relationship *user in projects* would be a weak entity because it cannot exist without both strong entities (Fig. 2.23).

LDM constitutes a data modeling technique which is feasible to represent most types of databases (Tillmann 2017). Hence, LDM can be applied to conventional data structures like relational or object-oriented but also to modern NoSQL or graph representations. Within this Thesis, we complied with the relational model according to Maier (1983) building up on the pioneering work of Codd (1970) for the data modeling process. Moreover, we applied the entity-relationship technique after Chen (1976) in the *usage-driven database design* (U3D) pattern after Tillmann (2017).

Result of the entity-relationship modeling process is an entity-relationship-diagram (ERD) or entity-relationship-model (ERM). An example of an ERD is provided in Figure 2.23 by using the example from the two paragraphs before.
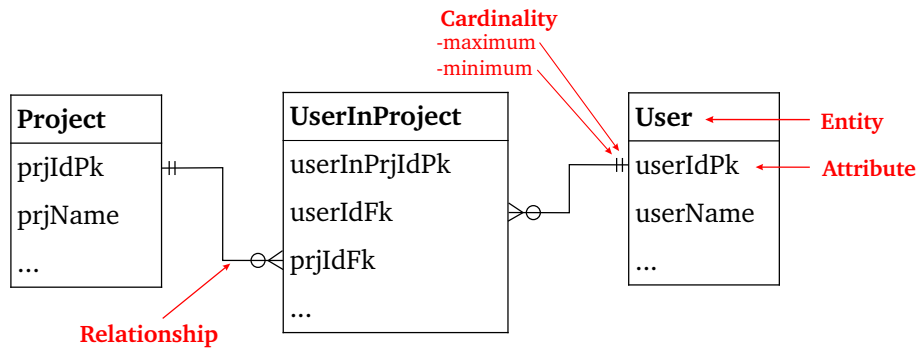
Figure 2.23.: Example of a physical entity-relationship-diagram (ERD) after Chen (1976) using the strong entities *User* and *Project*, which are connected in a weak entity named *UserInProject* via two one-to-many relationships. The cryptic attribute names comply to ERD conventions in which *Pk* stands for primary key and *Fk* for foreign key.

2.4.4.3. Data Management with the Relational Model

This section aims to outline the theory of the relational model (RM, Codd 1970) which builds up the underlying data structure used throughout this Thesis. The RM, besides network or hierarchical models, provides a formal approach to manage data from a domain according to predicate logic.

The RM consists of relations R and relation schemes $\mathbf{R}$ in which R is a finite set of attributes $\{A_1, A_2, ...A_n\}$ and $R \in \mathbf{R}$. D_i represents a finite respectively countably infinite, non-empty, arbitrary set which is corresponding to an attribute A_i where $1 \leq i \leq n$. D_i is also called the domain of A_i $[dom(A_i)]$ (Maier 1983).

A relation R on the relation scheme $\mathbf{R}$ is a finite set of mappings (tuples) $\{t_1, t_2, ..., t_n\}$ from R to $\mathbf{D}$ where $\mathbf{D} = D_1 \cup D_2 \cup ... \cup D_n$. Every tuple $t \in R$ has to be contained in D_i.

A key of a relation R on a relation scheme $\mathbf{R}$ is a subset $K = \{B_1, B_2, ...B_n\}$ of R. For every two distinct tuples t_1 and t_2 there is a $B \in K$ such that $t_1(B) \neq t_2(B)$ and $\exists! t R(t)$. Accordingly, a key helps to identify a designated tuple in $\mathbf{R}$ (superkey [SK]) or R (primary key [PK]). A key consists of at least one attribute A. Keys consisting of more than one attribute are called composite keys. In contrast to SK and PK a foreign key (FK) identifies a relationship between relations.

Relations can be transformed by a set of operations. The add operation adds a tuple t to a relation $R(t)$ the command of which can be expressed by

$$ADD(R;\ A_1 = t_1, A_2 = t_2, ...A_n = t_n). \tag{2.21}$$

Vice-versa, the delete operation removes a tuple from a relation $R(t)$:

$$DEL(R;\ A_1 = t_1, A_2 = t_2, ...A_n = t_n). \tag{2.22}$$

If a tuple is modified, one or more attributes will change their values. This can be expressed by the update operator UPD which is defined as

$$\begin{aligned} UPD(R; A_1 = t_1, A_2 = t_2, ...A_n = t_n; \\ C_1 = e_1, C_2 = e_2, ...C_n = e_n), \end{aligned} \tag{2.23}$$

where C_i is the changed attribute of R with the value e_i.

Once R contains one attribute and one tuple at the least, the relational operators select (SEL), project (PRJ) and join (JN) together with boolean operators can be applied to one or more R from $\mathbf{R}$.

The boolean operators check what tuples of two or more relations belong to the same entity. Thus, the boolean operators are yielding relations $Q(\mathbf{R})$ as a result of a boolean operation on two (or more) relations R and S, which both belong to $\mathbf{R}$. The boolean operations comprise $R \cap S$, $R \cup S$ and $R - S$.

$R \cap S$ yields all tuples which are contained in both R and S. $R \cup S$ contains all tuples which are either in R or in S and $R - S$ contains all tuples in R which are not in S.

The SEL operator is widely applied to select a subset R' of all tuples of R with specific values on one or more specific attributes. A specific example would be the expression

Select all t equal to a in R, with a as an element of the domain of R, which can be expressed as

$$R' = \{\, t \in R \,|\, t(A) = a \,\}. \tag{2.24}$$

The PRJ operator, on the other hand, is applied to select a subset R' of specified attributes of R with unique tuples. Let X be a subset of R in which all undesired attributes are removed. Subsequently, R' can be defined as

$$R' = \{\, t(X) \,|\, t \in R \,\}. \tag{2.25}$$

The join operator ($\bowtie$) is a binary operator which combines two or more relations based on all of their common attributes (Maier 1983). If the relations do not share a non-empty intersection, the join operator yields the Cartesian Product of the set of relations. Thus when we have two relations R and S with the attributes A, B, and C

$$\begin{array}{cc} R\,(A\ B\,) & S\,(B\ C\,) \\ a_1\ b_1 & b_1\ c_1 \\ a_2\ b_2 & b_2\ c_1 \end{array} \tag{2.26}$$

the join operation will lead to the following relation Q

$$R \bowtie S = \begin{array}{c} Q\,(A\ B\ C\,) \\ a_1\ b_1\ c_1 \\ a_2\ b_2\ c_1. \end{array} \tag{2.27}$$

In terms of database management, the join operator is used to select data from two or more tables based on identical cell entries. For geoscientific data sets this might be useful when the user wants to join different types of measurements derived from a sample and enrich the selected data set by meta data from the object of investigation, petrographic group, or the lithostratigraphic unit. The join operation is then performed on the foreign keys for the respective joining target.

2.4.4.4. Data Preprocessing

The Sparsity of Data Sets

It is not uncommon that data sets, which were produced in subsurface studies, provide a sparse structure. This means, that a set of values from one or more features provides no data. The sparsity can be calculated as $s = \sum x_s / \sum x_t$ with $s \in [0, 1]$ where x_s is the count of the sparse data cells and x_t is the count of all cells in the data set.

Sparse values can heavily influence statistical analyses since either the sparse cell needs to be substituted by a dummy value or the affected row needs to be removed from the input data prior to the analysis. Especially geoscientific data sets often provide a sparse structure due to missing measurement values, repetitive measurements, alternating measurement conditions, etc. However, tools exist to find fitting values for these sparse data sets.

In order to predict missing data the value could be estimated by a specific mean or median of the feature space or it could be determined through data mining techniques such as linear or curvilinear regression.

Data Transformation

The process of converting data from one space into another one is known as data transformation. Dependent on the type of empirical distribution, data need to be transformed before, for example, data mining algorithms can be deployed onto them. Many properties in the earth sciences show a skewed, asymmetric distribution caused by, for instance, few very high/low values. For strongly skewed or log-normally distributed data sets (section 2.4.4.7), it might be useful to transform the data set into normal space or by a logarithmic transformation.

In Table A.3 the most relevant transformation methods are listed. There are, however, also more complex transformation methods existing. For instance, prior to sequential Gaussian simulations (SGS, see section 2.4.4.9) it might be necessary to transform a source data set into standard normal space since SGS is computed under the assumption that data follow a multi-Gaussian distribution. We will discuss the normal score transform and its back-transformation here.

The normal score transform (NST) or quantile-quantile transform, respectively, converts an empirical cumulative density function (CDF) into normal score space. The transformation simulates the CDF of the standard normal distribution with $\mu = 0$ and $\sigma^2 = 1$ and each point of the empirical CDF is transformed into the counterpart of the simulated standard normal distribution (Goovaerts 1997).

$F(z)$ and $G(y)$ are stationary CDFs of the original random function $Z(u)$ and the standard normal random function $Y(u)$ (Fig. 2.24). In order to transform any point in the CDF ($F(z)$) of any random variable $Z(u)$ to a random function $Y(u)$ and vice versa the following function must be applied

$$Y(u) = \phi(Z(u)) = G^{-1}[F(Z(u))], \qquad (2.28)$$

where G^{-1} is the inverse Gaussian CDF of $Y(u)$, which is also called quantile function (Goovaerts 1997). Thus, $z_i \in Z$ and $y_i \in Y$ correspond to the same probabilities. This relationship can be expressed by two dictionaries where z and y are listed together with their probability $F(z_i)$ and $G(y_i)$.

In order to transform y_i back into the original variable's space, following equation can be applied

$$z_i = F^{-1}\left(G(y_i)\right). \qquad (2.29)$$
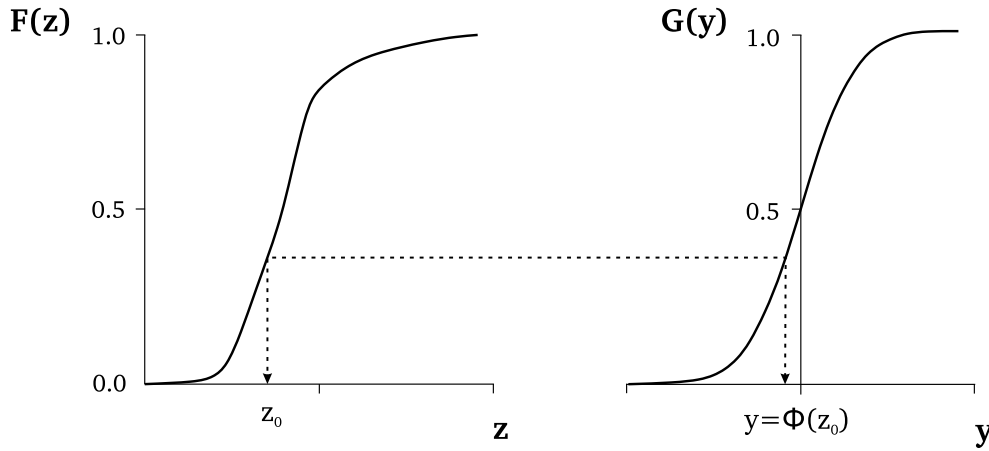
Figure 2.24.: Conceptualization of a normal score transformation and back transformation according to Goovaerts (1997).

The implementation of this equation is based on an interpolation which is aided by a conversion table z_i, y_i with $i = 1, ..., n$ in which each value of y is mapped onto a value of the original variable's space. The corresponding interpolation is performed piece-wise for each interval s_j with $j \in 1, ..., m$ of the original variable's complementary CDF, which corresponds to the tail function.

First, all values of y are ranked and assigned to the corresponding interval of the complementary CDF of $F(z)$ with the upper limit $F(z)_{j,\,upper}$ and $F(z)_{j,\,lower}$ what yields $Y(F(z))$.

Subsequently, the value z_k is derived as

$$z_k = \mathbf{F(z)}_{j, lower} + \frac{\mathbf{F(z)}_{j, upper} - \mathbf{F(z)}_{j, lower}}{n_{\mathbf{Y(F(z))}}} \quad \forall y_k \in \mathbf{Y(F(z))}. \tag{2.30}$$

Outlier Detection

For subsurface characterization, rock and fluid properties are measured in the laboratory or under in-situ conditions in the field. Every method provides individual error ranges and might be prone to failed measurements due to operational errors or other influence factors. Nevertheless, it is not straight-forward to detect those erroneous measurements because in many cases, nature might behave strongly irregularly and in an unexpected manner.

Nonetheless, outliers should be detected and removed prior to data analysis and knowledge extraction. Most outlier detection methods are dependent on the data structure. For univariate data sets a commonly used technique is to declare a set of observations as outliers which are differing from a target distribution $\mathbf{F}$ — with $\mathbf{F}$ being assumed to be normally distributed $N(\mu, \sigma^2)$. Those observations are located in the so-called outlier region.

For normally-distributed and log-normally-distributed random variables, the method of interquartile-range (IQR) as it is described in Tukey (1977) or Heumann et al. (2016) can be used. Here, the difference between the 1st and 3rd quartile is calculated. It is assumed that every point exceeding 1.5-times the interquartile range starting from the arithmetic mean is classified as an outlier. The procedure is recursive which means, that the IQR must be calculated again after outlier exclusion. This method is robust considering normally-distributed data sets.

2.4.4.5. Data Mining

Data mining (DM) in general is a '*mechanized process of identifying or discovering useful structure in data*' (Fayyad et al. 2002) and aims to extract knowledge from data. Therefore, the process of DM allocates algorithms from dispersion analysis, correlation, regression, prediction and classification that can be used within the knowledge discovery process (Fig. 2.25).

In the following passages, techniques and algorithms from visual data mining (VDM), exploratory data mining (EDM) and from predictive data mining (PDM), with special regard to spatial data mining (SDM), are described. Each of these sub-disciplines belongs
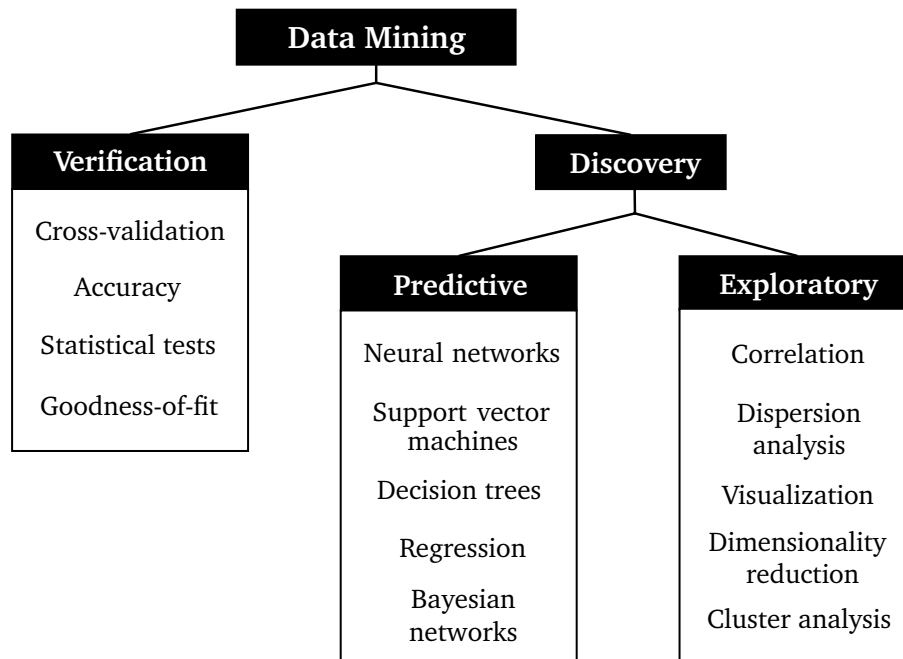
Figure 2.25.: The data mining taxonomy as being used in this work. The scheme is adapted from Maimon & Rokach (2010).

to the discipline of relational data mining (RDM) which is the discipline of finding patterns in data structured in relational format.

2.4.4.6. Visual Data Mining

Visual data mining can be considered being an intersection of KDD and VA. Simoff et al. (2008) define visual data mining (VDM) as being a process of '*interaction and analytical reasoning with one or more visual representations of an abstract data that leads to the visual discovery of robust patterns in these data that form the information and knowledge utilized in informed decision making*.' VDM is human-centered, which means that a domain-expert is continuously interacting with the software system throughout the process. Figure 2.26 illustrates the process of VDM.

It is evident that VDM and KDD provide a series of intersecting sub-processes such as the data preparation, data mining, model visualization, and knowledge/information extraction. Thus, VDM can be seen as a subset of the KDD process with the main difference in being focused on the visualization of models and the interaction with those visualizations (Simoff et al. 2008).
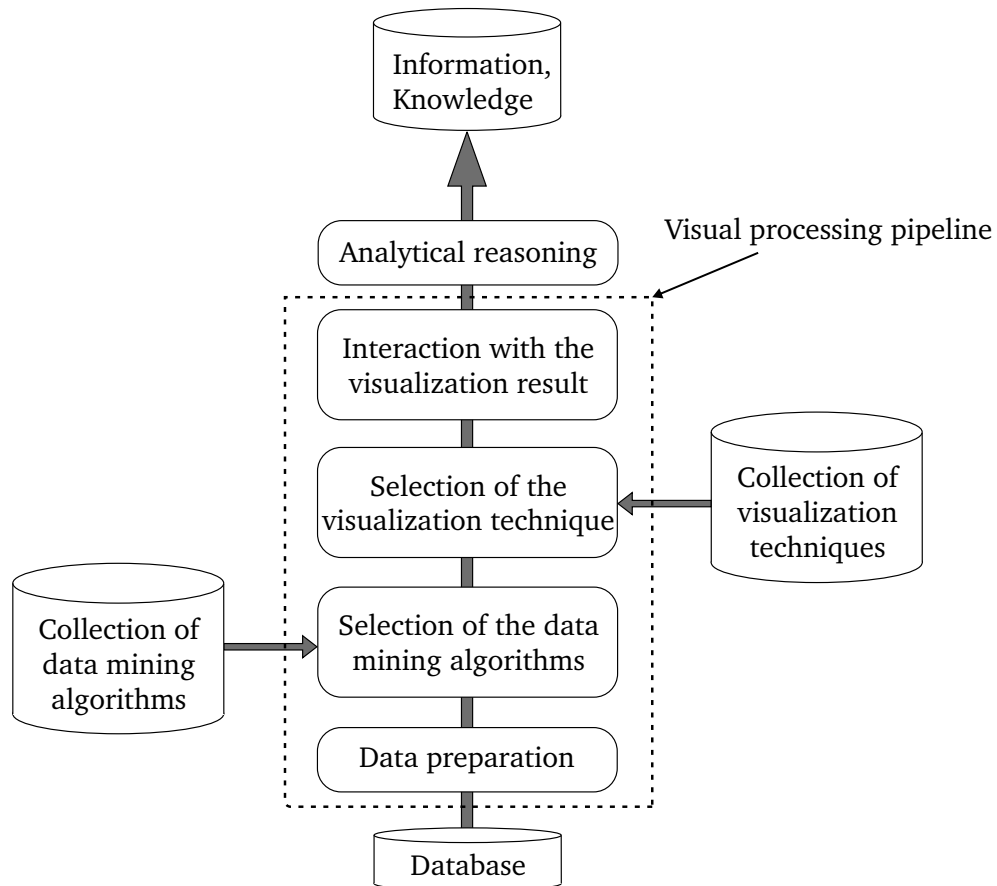
Figure 2.26.: Visual data mining (VDM) process adapted from Simoff et al. (2008).

2.4.4.7. Exploratory Data Mining

Exploratory data mining (EDM) aims to provide means to explore data sets for specific patterns in the form of analyzing the dispersion of a single property and identifying correlation, trends, or clusters in high-dimensional space.

Correlation Analysis

A correlation analysis determines the quantity of correlation of a set of parameters. Correlation analyses can be separated into linear and non-linear correlation analysis. Most commonly, a linear correlation is determined by the covariance matrix or the Pearson correlation coefficient. Non-linear correlation is usually determined by rank-based methods such as Spearman's rank correlation.

For a relational data set $R \subset \mathcal{R}^p$ a covariance matrix $\mathbf{C}$ describes the covariance of the

attributes i, j within R. The calculation of every element $c_{ij} \in \mathbf{C}$ can be expressed as

$$c_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_k^i - \overline{x}^i)(x_k^j - \overline{x}^j) \quad \text{with} \quad i,j = 1, ..., p. \tag{2.31}$$

Higher values of c_{ij} indicate a positive correlation of the attributes whereas low negative values indicate a strong negative correlation. 0 indicates a non-correlation between the attributes i and j.

From the covariance matrix, the Pearson correlation coefficient (R$_P$) can be calculated by dividing c_{ij} by the standard deviation σ of i and j (Eq. 2.32).

$$
\begin{aligned}
R_{P_{ij}} &= \frac{c_{ij}}{\sigma_i \, \sigma_j} \\
&= \frac{\sum\limits_{k=1}^{n} (x_k^i - \overline{x}^i)(x_k^j - \overline{x}^j)}{\sqrt{\left(\sum\limits_{k=1}^{n} \left(x_k^i\right)^2 - n \cdot \left(\overline{x}^i\right)^2 \right) \cdot \left(\sum\limits_{k=1}^{n} \left(x_k^j\right)^2 - n \cdot \left(\overline{x}^j\right)^2 \right)}} \in [-1, 1]
\end{aligned}
\tag{2.32}
$$

$$with \quad i,j = 1, ..., p.$$

Non-linear correlations can be quantified with Spearman's rank correlation (R_{sp}) coefficient that is calculated with the ranks $rk(x_i)$ and $rk(y_i)$ of the independent variables **x** and **y**. The rank is defined as the index i of the value x_i when **x** was sorted ascendingly. Accordingly, if all ranks are distinct integers R_{sp} can be calculated as

$$R_{sp} = 1 - \frac{6 \cdot \sum\limits_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{2.33}$$

with d_i as $rk(x_i) - rk(y_i)$.

Dispersion Analysis

Dispersion analysis provides helpful insights into the variability of an univariate data set in the format $\mathbf{x} = (x_1, x_2, ...x_n)$ (Trauth 2015). The patterns of distributions are dependent on the mathematical nature of the considered variable. Distribution analysis deals on the one hand side with descriptive empirical distribution analysis and predictive theoretical distribution analysis.

Empirical distribution analysis comprises measures of dispersion such as the range, mode, standard deviation, variance, skewness and kurtosis. Detailed derivations and calculations of these simple measures can be found in Trauth (2015) and in numerous other works as well (e.g., Wackernagel (2003) or Heumann et al. (2016)).

Another important measure of distributions is the average. For symmetric distributions the arithmetic mean or weighted arithmetic mean describes the average most robustly. For asymmetric and asymmetric non-linear distributions like exponential growth rates, the harmonic or geometric mean provides the most robust average value. Additionally, a distribution can be characterized by the calculation of quartiles, minimum and maximum value and median. Empirical distributions can be approximated by theoretical models. These models comprise the uniform, binomial, poisson gaussian/normal, log-normal, Student's t, Fishers F and χ^2 distributions.

All corresponding equations from the measures being mentioned in the last two paragraphs are listed in the Tables A.1 and A.2, which are appended to this Thesis.

Dimensionality Reduction

Some scientific problems require a data set X of dimensionality p to be transformed into a lower dimensional data set Y of dimensionality q. The problem in general can be expressed as

$$f : X \subset \mathbb{R}^p \to Y \subset \mathbb{R}^q. \tag{2.34}$$

Dimensionality reduction can be used to compress an observed data matrix, to detect multivariate outliers, to detect intrinsic correlation (Wackernagel 2003) or to proof the quality of a classification scheme which has been applied on a data set.

As dimensionality reduction algorithms the principal component analysis (PCA) and Self-organizing maps (SOM) will be considered in the following sections.

Principal Component Analysis

Main purpose of the PCA is to reduce the dimensionality of the observation space (Maćkiewicz & Ratajczak 1993) in order to derive trends in high-dimensional data sets (Brandsegg et al. 2010). The PCA is a tool commonly applied in the geosciences in the frame of, for instance, reservoir characterization (Hornung et al. 2020, Fig. 2.27), paleoenvironment reconstruction (Lenz et al. 2016), or sedimentary provenance analysis (Schneider et al. 2016; Lewin et al. 2018).

The PCA is a statistical method which finds a linear projection of a data set X in relational format which is representing the data structure insofar as that the variance is maximized in the lower-dimensional projection.
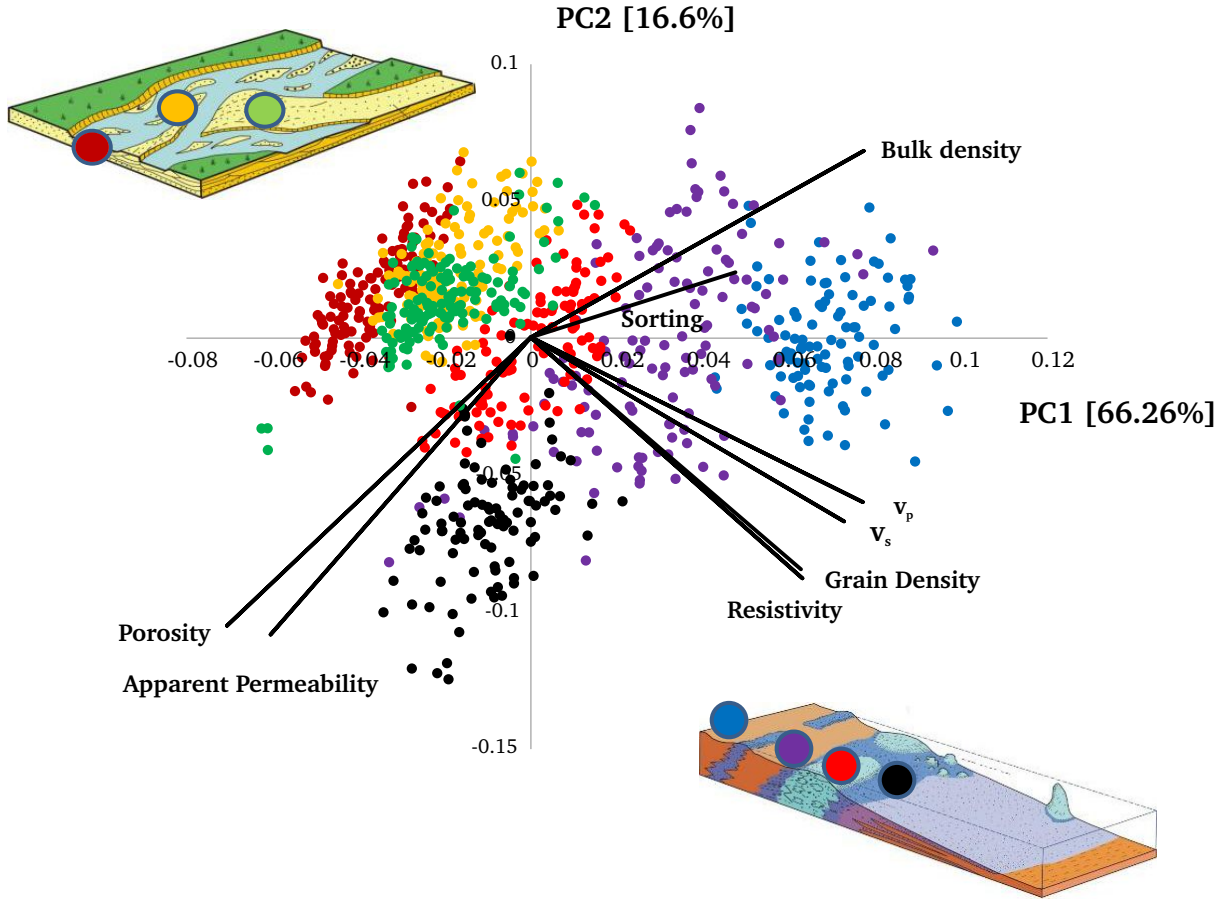
Figure 2.27.: Principal component analysis of the data set produced in **Publication II** (Hornung et al. 2020, ©EAGE) in which $\geq$10,000 measurements of nine properties measured on 887 samples — taken from seven rock slabs — were projected into 2-D space.

Result from a PCA is a linear transformation, represented by a translation and rotation component. This transformation is also called *principal axis transformation*:

$$y_k = x_k - \overline{x} \cdot \mathbf{E} \tag{2.35}$$

with $\mathbf{E}$ being a rotation matrix which is dependent on X. To determine $\mathbf{E}$, the variance

(σ_y^2) in the target data set Y must be maximized (Eq. 2.36):

$$
\begin{aligned}
\sigma_y^2 &= \frac{1}{n-1} \sum_{k=1}^{n} y_k^T y_k \\
&= \frac{1}{n-1} \sum_{k=1}^{n} \left((x_k - \overline{x}) \cdot \mathbf{E} \right)^T \cdot \left((x_k - \overline{x}) \cdot \mathbf{E} \right) \\
&= \frac{1}{n-1} \sum_{k=1}^{n} \mathbf{E}^T \cdot (x_k - \overline{x})^T \cdot (x_k - \overline{x}) \cdot \mathbf{E} \\
&= \mathbf{E}^T \cdot \left(\frac{1}{n-1} \sum_{k=1}^{n} (x_k - \overline{x})^T \cdot (x_k - \overline{x}) \right) \cdot \mathbf{E} \\
&= \mathbf{E}^T \cdot \mathbf{C} \cdot \mathbf{E}
\end{aligned}
\tag{2.36}
$$

where $\mathbf{C}$ is the covariance matrix (Eq. 2.31) of X with the elements c_{ij}. Since the transformation matrix $\mathbf{E}$ should be restricted to rotation, the following applies

$$
\mathbf{E}^T \cdot \mathbf{E} = 1
\tag{2.37}
$$

In order to maximize the variance like outlined in Equation 2.36 under the condition of Equation 2.37 we apply the Lagrange function

$$
\mathbf{L} = \mathbf{E}^T \cdot \mathbf{C} \cdot \mathbf{E} - \lambda \cdot (\mathbf{E}^T \cdot \mathbf{E} - 1).
\tag{2.38}
$$

The necessary condition is

$$
\frac{\partial \mathbf{L}}{\partial \mathbf{E}} = 0
\tag{2.39}
$$

$$
\Leftrightarrow \mathbf{C} \cdot \mathbf{E} + \mathbf{E}^T \cdot \mathbf{C} - 2\lambda \mathbf{E} = 0
\tag{2.40}
$$

$$
\Leftrightarrow \mathbf{C} \cdot \mathbf{E} = \lambda \mathbf{E},
\tag{2.41}
$$

where we describe an eigenvalue problem. By transforming Equation 2.41 into a homogeneous system of equations we can derive

$$
(\mathbf{C} - \lambda \mathbf{I}) \cdot \mathbf{E} = 0.
\tag{2.42}
$$

The columns of $\mathbf{E}$ can be derived from the eigenvectors v_i of $\mathbf{C}$ and the variances correspond to the eigenvalues λ_i of v_i because

$$\lambda = \mathbf{E}^T \cdot \mathbf{C} \cdot \mathbf{E} = \sigma_y^2 \Leftrightarrow \mathbf{C} \cdot \mathbf{E} = \lambda \mathbf{E}. \tag{2.43}$$

Thus, the PCA also delivers the variances of Y which qualifies it to use the dimensionality reduction under the condition of a maximum grade of variance. In order to reduce the dimensionality to q, the first q eigenvectors of $\mathbf{E}$ must be considered. For instance, if the projection should cover $\geq 50\%$ of the total variance, q must be selected so that

$$\sum_{i=1}^{q} \lambda_i / \sum_{j=1}^{p} \lambda_j \geq 0.5. \tag{2.44}$$

Self-Organizing Maps

Self-organizing maps (SOM) aim to non-linearly project a data set X into an n-dimensional representation Y by minimizing the pairwise distances of X and Y so that $d_{ij}^x \approx d_{ij}^y$.

An example of iterative Sammon mapping (SM), which is a form of SOM, is provided in Figure 2.28 in which an eight-dimensional data set resulting from the petrophysical data set from **Publication I** is reduced to two dimensions using 5,000 iterations. The blue and red areas each represent samples from two different lithological units in the investigated quarry.

When applying the SM algorithm, a representation Y is initiated randomly and the initial Sammon's stress (E) is calculated which is expressed as

$$E = \frac{1}{\sum_{i=1}^{n} \sum_{j=i+1}^{n} d_{ij}^x} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{(d_{ij}^y - d_{ij}^x)^2}{d_{ij}^x}, \tag{2.45}$$

where d is the Manhatten distance (Eq. 2.46) between the point pairs of X and Y (Sammon 1969).

$$d(X, Y) = \sum_{i=1}^{n} |x_i - y_i| \tag{2.46}$$

The SM algorithm minimizes the Sammon's stress gradient in a user-defined number of iterations. In order to minimize the projection error, Sammon (1969) applied a steepest-descent iteration. Here, the mapping error is stated as

$$E(m) = \frac{1}{c} \sum_{i<j}^{n} (d_{ij}^* - d_{ij}(m))^2 / d_{ij}^*, \qquad (2.47)$$

with

$$c = \sum_{i<j}^{n} (d_{ij}) \qquad (2.48)$$

and

$$d_{ij}(m) = \sqrt{\sum_{k=1}^{d} \left(y_{ik}(m) - y_{jk}(m) \right)^2}. \qquad (2.49)$$

The projected data set at iteration step $m + 1$ can be derived as

$$y_{pg}(m + 1) = y_{pg}(m) - \alpha \cdot \Delta_{pg}(m) \qquad (2.50)$$

where α is the so called *magic factor* that can be chosen to be between 0.3 and 0.4 and

$$\Delta_{pg}(m) = \frac{\partial E(m)}{\partial y_{pq}} \Big/ \left\| \frac{\partial^2 E(m)}{\partial y_{pg}(m)^2} \right\|. \qquad (2.51)$$

The derivatives can be derived as

$$\frac{\partial E(m)}{\partial y_{pq}} = \frac{-2}{c} \sum_{j=1}^{n} \left(\frac{d_{pj}^* - d_{pj}}{d_{pj}d_{pj}^*} \right) (y_{pq} - y_{iq}) \qquad (2.52)$$

and

$$\frac{\partial^2 E(m)}{\partial y_{pg}(m)^2} = \frac{-2}{c} \sum_{j=1}^{n} \frac{1}{d_{pj}^* d_{pj}} \cdot \left[(d_{pj}^* - d_{pj}) - \frac{(y_{pq} - y_{iq})^2}{d_{pj}} \left(1 + \frac{d_{pj}^* - d_{pj}}{d_{pj}} \right) \right]. \qquad (2.53)$$

Since those formulations provide significant performance draw-backs, we used an implementation of (Kohonen 2001) where the corrections of the target vector $y_{pg}(m + 1)$ can be derived as

$$\Delta y_{pg} = \lambda \cdot \frac{d_{ij} - |r_i - r_j|}{|r_i - r_j|} \cdot (r_i - r_j) \tag{2.54}$$

where

$$\Delta r_j = -\Delta r_i. \tag{2.55}$$
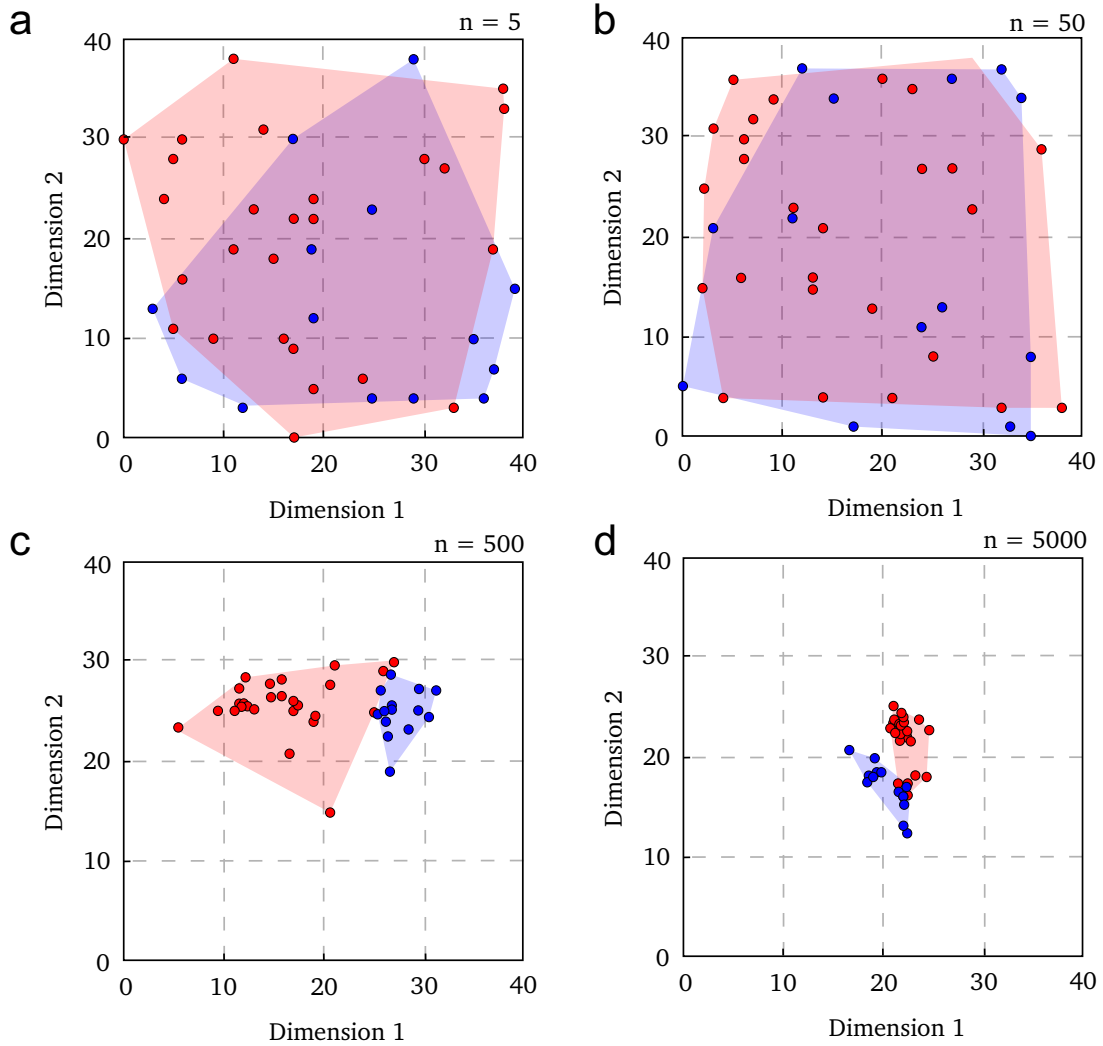
Figure 2.28.: Iterative Sammon's Mapping using two feature classes. Displayed are the petrophysical data sets from the global rock samples which were investigated in **Publication I**. The figures visualize the result of the Sammon's mappings after 5 **(a)**, 50 **(b)**, 500 **(c)** and 5000 **(d)** iterations.

Classification Problems

Classification problems address problems where a categorical affiliation, also called class or cluster, must be determined. Classification is useful when the class affiliation of a new sample is not defined. In this work, we will restrict the classification problems to the unsupervised techniques including the nearest neighbor classification and cluster analysis.

Nearest Neighbors Algorithm

One of the easiest and earliest automatic classification methods is the nearest neighbor (NN) algorithm (Shakhnarovich et al. 2006). An unknown point is hereby classified based on the nearest classified point within the feature space. The distance to the nearest neighbor is calculated by a specific measure such as the Euclidean distance or the shortest path distance. An unknown point will be classified as class y_k if

$$|\mathbf{x} - \mathbf{x}_k| = \min_{j=1,\dots,n} |\mathbf{x} - \mathbf{x}_j|, \tag{2.56}$$

where $\mathbf{x}$ is the point with the unknown class, $\mathbf{x}_k$ is the point with the class y_k and $\mathbf{x}_j$ is the jth point in the collection of training samples.

k-Means Clustering

k-means clustering using Lloyd's algorithm (Lloyd 1982) is an automated, multivariate classification method. This clustering method requires a predefined number of clusters $\mathbf{c}$, the data set should be subdivided into, and categorizes the data values by minimizing the sum of squared differences of the data values $\mathbf{x}$ within the clusters' centers μ. Thus, the main objective is to find

$$\arg\min_{j} \sum_{i=1}^{k} \sum_{x \in c_i} ||x_i - \mu_j||^2. \tag{2.57}$$

Initially, the means of the predefined clusters are set either randomly or by means of k randomly selected observations. After initially assigning the observations to the nearest cluster center by applying Equation 2.58 Lloyd's algorithms uses an iterative refinement technique.

$$c := \min_{j} ||x_i - \mu_j||^2 \tag{2.58}$$

Hereby, the clusters' centers are updated based on

$$\mu_i^{(t+1)} = \frac{1}{|c_i^t|} \sum_{x_j \in c_i^t} x_j. \tag{2.59}$$

Subsequently, each observation is assigned to one of the clusters again based on the nearest cluster mean (Eq. 2.58).

2.4.4.8. Predictive Data Mining

Predictive data mining (PDM) aims to predict values in unknown space. In the following sections, some of the involved techniques will be presented and special regard will be laid upon the aspect of spatial data mining (SDM).

Regression Analysis

Regression aims to find a fitting function between samples of two or more random variables. Two types of regression are predominantly applied in the geosciences, namely, the linear and curvilinear regression which are performed in order to find empirical relationships among geological properties (Fig. 2.29).

A linear regression tries to fit a linear function of the form

$$y = b_0 + b_1 x, \tag{2.60}$$

where y and x are the random variables, b_0 is the y-axis interception and b_1 is the gradient. In order to find a function where the squared sum of the Δy deviations is minimized like

$$\sum_{k=1}^{n} (\Delta y_k)^2 = \sum_{k=1}^{n} \left(y_k - (b_0 + b_1 x_k) \right)^2 \tag{2.61}$$

we need to calculate the gradient as

$$b_1 = \frac{\sum_{k=1}^{n} (x_k - \overline{x})(y_k - \overline{y})}{\sum_{k=1}^{n} (x_k - \overline{x})^2}. \tag{2.62}$$

Accordingly, b_0 can be calculated as

$$b_0 = \overline{y} - b_1\overline{x}. \tag{2.63}$$

For curvilinear regression, a function of a degree > 1 will be approximated for a discrete set of values. A second-degree polynomial function for instance would provide the form

$$y = b_0 + b_1 x + b_2 x^2 \tag{2.64}$$

Thus, we would need to find $n + 1$ regression coefficients, where n is the degree of the polynomial function. In general, the regression model yields

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_n x_i^n \quad \text{with} \quad i = 1, 2, ..., n. \tag{2.65}$$

The estimation of the regression coefficients is aimed through solving a system of linear equations as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1^1 & \cdots & x_1^m \\ 1 & x_2^1 & \cdots & x_2^m \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_n^1 & \cdots & x_n^m \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix}, \tag{2.66}$$

where x and y are the samples. Results of a regression analysis can be assessed through jack-knifing, cross-validation or bootstrapping (see section 2.4.4.10).

2.4.4.9. Spatial Data Mining

Introduction

Spatial data mining (SDM), in contrast to PDM, provides a set of methods to extract knowledge specifically from spatial data (Li et al. 2015). Although the methods used for SDM equal the methods involved in PDM, both disciplines are separated from one another as the terminology is slightly different.

Fischer & Getis (2010) provide a comprehensive overview about the methods, applications and software packages related to SDM and spatial data analysis in general. An important member among the methods of SDM in earth sciences is the domain of geostatistics. Past
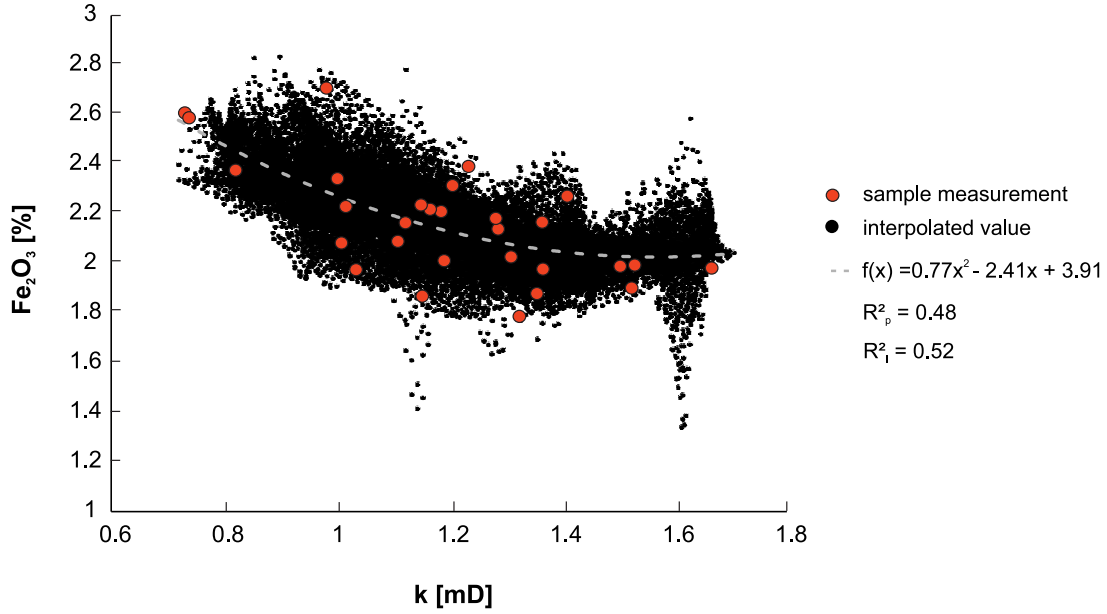
Figure 2.29.: Regression analysis between the two properties Fe_2O_3 and the intrinsic permeability k performed with the data set from **Publication III** in which the small-scale distribution of chemical and physical properties were investigated in terms of statistical relationships and spatial patterns. In the legend, the curvilinear cure-fitting polynomial function together with the coefficients of determination for the original data points (R_p^2) and the interpolated data points (R_i^2) are given.

studies showed that the symbiotic usage of conventional DM algorithms and geostatistical approaches can significantly enhance numerical prognoses of geological properties in space (Kanevski et al. 2004).

Originally intended to optimize ore body mining (Mazzella & Mazzella 2013), nowadays geostatistics plays a major role in various domains which involve spatial data – not only limited to geoscientific applications.

In the following sections we will provide an overview about the statistical and deterministic methods that can be used to analyze and predict rock and fluid properties in space and time.

Spatial Neighborhood

Most geostatistical algorithms require subset-sampling in order to perform reasonably. Commonly, a 3-D search ellipsoid is used to find the neighbors of a point in a mesh. That search ellipsoid can be defined by six properties, namely, azimuth α, dip β, plunge γ, longest r_l, middle r_m and shortest radius r_s of the ellipsoid. α, β and γ define the ellipsoid's rotation along the Z, X and Y axis. Accordingly, the rotation matrix $\mathbf{T}$ can be defined as:

$$\mathbf{T} = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (2.67)$$

First, the neighbors are rotated along the Z axis, secondly along the X axis and last along the Y axis (Fig. 2.30). After rotating the neighboring points, Eq. 2.68 can be used to determine whether a point x with the transformed coordinates x'_x, x'_y and x'_z is located inside (≤ 1) the search ellipsoid or not (> 1).

$$\frac{X'^2}{x'_x} + \frac{Y'^2}{x'_y} + \frac{Z'^2}{x'_z} \leq 1 \qquad (2.68)$$
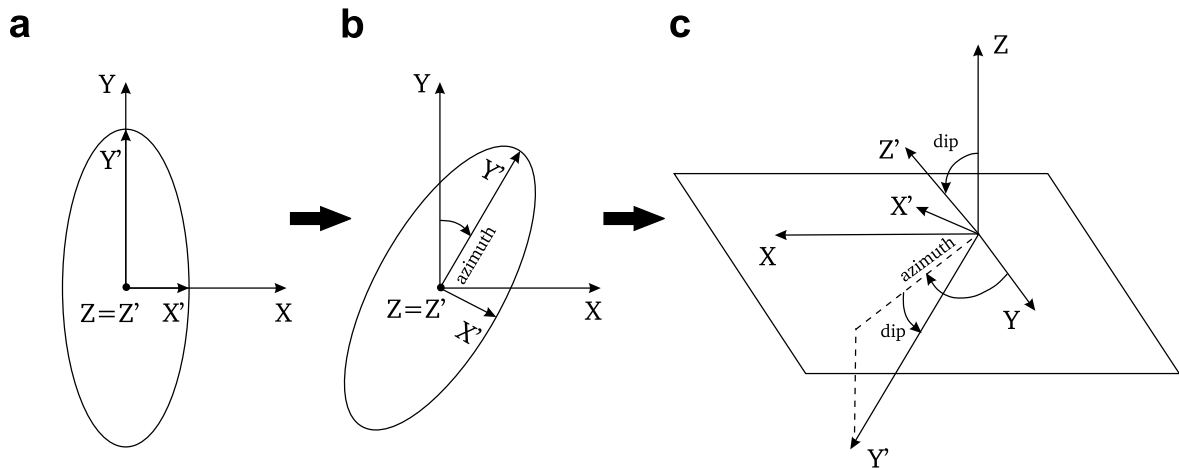
Figure 2.30.: Schematic illustration of the first two steps of a coordinate system rotation when applying an ellipsoid search. **(a)** Starting point; **(b)** Rotation along the Z axis; **(c)** Rotation along the X axis.

Semivariogram

Originally defined by Matheron (1963), the semivariogram describes the spatial dependence of a random field variable. The variability of a regionalized property **Z** is computed at different scales by calculating the dissimilarity between pairs of data values z_α and z_β at the locations $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ in Ω (Fig. 2.31).

As a metric for dissimilarity $\gamma_{\alpha\beta}$ is calculated (Eq. 2.69) as follows

$$\gamma_{\alpha\beta} = \frac{(z_\alpha - z_\beta)^2}{2} \qquad (2.69)$$

Since $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ can be expressed as points separated by a lag vector $\mathbf{h}$ dissimilarity can also be formulated as

$$\gamma *_{\alpha\beta}(\mathbf{h}) = \frac{\left(z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)\right)^2}{2} \tag{2.70}$$

Length of the lag vector $\mathbf{h}$ in n-D Euclidean space with $\mathbf{p}$ respectively $\mathbf{q}$ representing the vectors containing Cartesian coordinates of the points $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ is expressed as the Euclidean norm (Eq. 2.71).

$$\begin{aligned} |\mathbf{p}, \mathbf{q}| &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \end{aligned} \tag{2.71}$$

Variogram calculation results in a set of points representing the cumulative dissimilarity γ of point-pairs with the distance of $|\mathbf{h}|$ in Ω. The experimental semivariogram, however, represents the cumulative dissimilarity of a discrete set of point-pairs x with n_c representing the count of point-pairs within distance classes h_k of identical distance increments (Eq. 2.72).

$$\gamma(\mathbf{h}) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} \left(z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)\right)^2 \tag{2.72}$$

The continuous counterpart, represented by the theoretical semivariogram, is an approximation to the experimental semivariogram assuming $\mathbf{Z}(\mathbf{x})$ to be a stationary random field (Wackernagel 2003). A theoretical variogram γ_{theo} is represented by a covariance function c and with the following relationship

$$\gamma_{theo}(\mathbf{h}) = c(0) - c(\mathbf{h}) \tag{2.73}$$
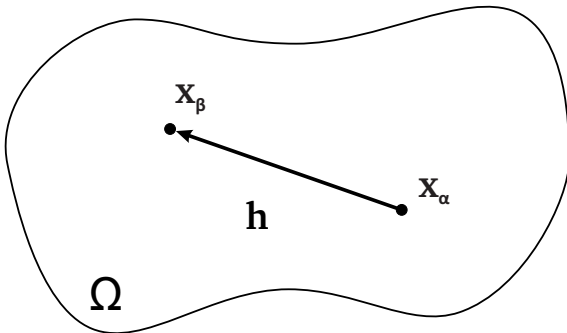
Figure 2.31.: Two points in two-dimensional space linked by a lag vector (h) (Wackernagel 2003).

c is a positive definite, even function with the origin at $\gamma(0) = 0$. If a nugget effect is observable, the theoretical semivariogram is translated by the magnitude of the nugget (n), so that the origin is translated by n on the ordinate.

$$c_{nug}(\mathbf{h}) = \begin{cases} b & \text{for } |\mathbf{h}| = 0 \\ 0 & \text{for } |\mathbf{h}| > 0, \end{cases} \tag{2.74}$$

where $\lim\limits_{|\mathbf{h}| \to 0} \gamma(\mathbf{h}) = b$. Five covariance models are mostly used to fit the experimental semivariogram which are listed in the appendix A.1.3 in Table A.4. Following variables are needed to calculate the covariance functions: nugget (n), range (a) and sill (b) like shown in Figure 2.32.
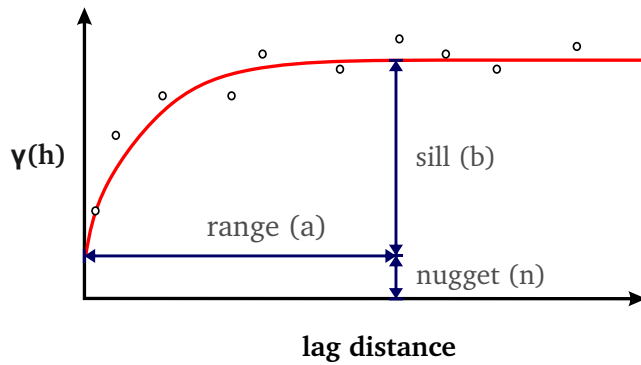
Figure 2.32.: Conceptual visualization of a theoretical semivariogram with a spherical model fitting an experimental semivariogram.

Semivariograms can be used to quantify the spatial or time correlation of a random property (Ringrose & Bentley 2015; Gu et al. 2017; Rühaak et al. 2015c). Further on, the differences in range and sill in dissimilar directional semivariograms can quantify the internal anisotropy of a property (Ringrose & Bentley 2015). The resulting covariance function is an input variable for kriging.

Spatial and Spatiotemporal Interpolation

Spatial and spatiotemporal interpolation is the approximation of the spatial and/or temporal distribution of a property by a function (Mitas & Mitasova 2005). In the following sections, different interpolation methods such as the inverse distance weighting and some varieties of kriging will be outlined.

Inverse Distance Weighting

The inverse distance weighting (IDW), p-value IDW and Shepard's IDW (Shepard 1968) approaches are widely applied methods to deterministically interpolate values in space or time. The IDW interpolation in general calculates an unknown value $z(\mathbf{x}_0)$ at point $\mathbf{x}_0$ by

weighting the distance of that point to each known point ($\mathbf{x}_k$) in space. The underlying formula for IDW is

$$z(\mathbf{x}_0) = \frac{\sum_{k=1}^{n} \left(1/d_k^p\right) \cdot z(\mathbf{x}_k)}{\sum_{k=1}^{n} 1/d_k^p}, \tag{2.75}$$

where d is the Euclidean distance between the point with the known value $\mathbf{x}_k$ and the point with the unknown value $\mathbf{x}_0$ and p is an exponent factor to bias the weights non-linearly. IDW is a reliable and widely applied method to interpolate static properties in 1-D, 2-D and 3-D space (Rühaak 2006.)

Kriging

Kriging is a commonly used stochastic method to interpolate geological properties in space and time (Malvić et al. 2019; Rühaak 2015). The kriging estimator is an optimal estimator as it minimizes the error variance. It incorporates the covariance structure of the global sampled values into the weights for predicting the value $z(\mathbf{x}_0)$ at an unsampled location $\mathbf{x}_0$. Therefore, $z(\mathbf{x}_0)$ is calculated by weighting the neighboring sampled values and building a linear combination of those what yields

$$z(\mathbf{x}_0) = \sum_{k=1}^{n} w_k \cdot z(\mathbf{x}_k), \tag{2.76}$$

where w_k is the weight of the known point $\mathbf{x}_k$ with the value $z(\mathbf{x}_k)$. The idea behind kriging is to find an estimator such that

I) $E(\mathbf{x}_0) = E(\mathbf{x})$ which is satisfied if $\sum w_k = 0$ and if the mean μ is stationary.
II) The prediction variance σ^2 is minimized.

If both conditions are fulfilled the kriging method is the best linear unbiased predictor. In the following paragraphs we will consider the simple kriging (SK), ordinary kriging (OK), regression kriging (RK) and universal kriging (UK) techniques which are the most commonly used varieties.

The types of kriging primarily differ in the derivation of the weight vector, however, RK and UK require data transformation prior to modeling the variogram from the residuals of the regression.

For all systems a set of linear equations must be solved like it is outlined in the following paragraphs. The quality of kriging interpolation is dependent on the theoretical semivariogram and the goodness of fit to the experimental values.

Simple and Ordinary Kriging

In this section, we will consider the simple kriging (SK) technique (Deutsch & Journel 1998) and expand it by the integration of a locally varying mean (Goovaerts 1997). Therefore, we modify Equation 2.76 into

$$z(\mathbf{x}_0)_{SK} = \sum_{k=1}^{n} w_k \cdot z(\mathbf{x}_k) + \left(1 - \sum_{k=1}^{n} w_k\right) \cdot \mu. \tag{2.77}$$

in which the known stationary mean μ has been added (Deutsch & Journel 1998). While SK assumes that μ is globally constant and known, SK with locally varying mean assumes μ to be constant only in the neighborhood of $\mathbf{x}_0$. In order to obtain the SK weights, a system of n linear equations must be solved in which n stands for the number of considered neighbors. This system of equations is defined as

$$\mathbf{A}\mathbf{w} = \mathbf{b}, \tag{2.78}$$

which corresponds to

$$\underbrace{\begin{pmatrix} c(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & c(\mathbf{x}_1 - \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ c(\mathbf{x}_n - \mathbf{x}_1) & \cdots & c(\mathbf{x}_n - \mathbf{x}_n) \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix}}_{\mathbf{w}} = \underbrace{\begin{pmatrix} c(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ c(\mathbf{x}_n - \mathbf{x}_0) \end{pmatrix}}_{\mathbf{b}}, \tag{2.79}$$

with c as covariance function and $\mathbf{x}_n$ as point with known value (Wackernagel 2003). In SK each interpolated point provides a simple kriging variance σ_{SK}^2 (Webster & Oliver 2007), which we can calculate by means of the formula

$$\sigma_{SK}^2 = c(0) - \sum_{k=1}^{n} w_k \, c(\mathbf{x}_k, \mathbf{x}_0). \tag{2.80}$$

In order to obtain the ordinary kriging weights, a set of n+1 equations have to be solved. In matrix notation, this set of equations can be written as

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix} \tag{2.81}$$

with γ as theoretical semivariogram, $\mathbf{x}_n$ as point with known value and μ as Lagrange parameter (Wackernagel 2003).

Regression Kriging and Universal Kriging

Universal kriging splits the random function into a linear combination of a deterministic and a stochastic component such that

$$z(\mathbf{x}_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^{n} w_i \cdot e(\mathbf{x}_i) = \hat{m}(s) + \hat{e}(s_0), \qquad (2.82)$$

where $\hat{m}(s)$ is the deterministic and $\hat{e}(s_0)$ is the stochastic component. $\hat{\beta}_k$ is the deterministic coefficient, $q_k(s_0)$ is the deterministic variable, w_i is the weight of the known residual $\hat{e}$ at point $\mathbf{x}_i$.

RK and UK are based on identical mathematical assumptions, however, in UK the residuals are explicitly calculated as a function of the geographic space. Thus, the deterministic function in UK is expressed as an nth-degree polynomial function of the Cartesian coordinates x, y and z, whereas RK also allows for other drift variables. The mathematical similarity of both methods lead to confusion in the literature.

In order to obtain the weights for RK and UK, a set of n+m equations must be solved. In matrix notation, this set of equations can be written as

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 & f_1(\mathbf{x}_1) & \cdots & f_k(\mathbf{x}_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 & f_1(\mathbf{x}_n) & \cdots & f_k(\mathbf{x}_n) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ f_1(\mathbf{x}_1) & \cdots & f_1(\mathbf{x}_n) & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_k(\mathbf{x}_1) & \cdots & f_k(\mathbf{x}_n) & 0 & 0 & \cdots & 0 \end{pmatrix} \cdot \begin{pmatrix} w_1^{UK} \\ \vdots \\ w_n^{UK} \\ \mu_{UK} \\ \phi_0 \\ \vdots \\ \phi_k \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \\ f_1(\mathbf{x}_0) \\ \vdots \\ f_k(\mathbf{x}_0) \end{pmatrix} \qquad (2.83)$$

with γ as theoretical semivariogram, $\mathbf{x}_n$ as point with known value, μ as Lagrange parameter, and $f_i(\mathbf{x})$ as ith polynomial (Hudson & Wackernagel 1994).

Consideration of Measurement Error Variance

We already saw that kriging induces a local interpolation error by itself, namely, σ_{SK}^2. There are, however, also other components which bias the interpolation result. Besides

σ_{SK}^2, the local and unknown variability of $z(\mathbf{x})$ in Ω_b as well as the measurement error variance σ_m^2 might play an important role (Figure 2.33).
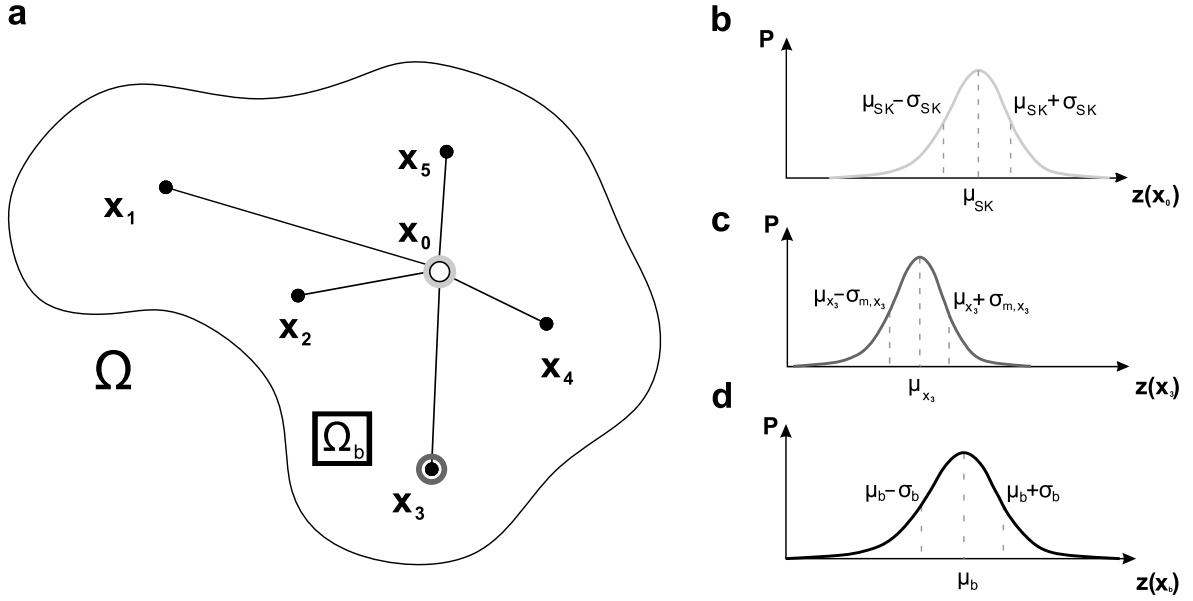
Figure 2.33.: (a) Schematic of the uncertainty components integrated in a predictive model of rock or fluid properties; **(b)** Estimated kriging error variance at $\mathbf{x}_0$; **(c)** Observed measurement error at point $\mathbf{x}_3$; **(d)** Observed block variability in a subset Ω_b of Ω.

Integrating σ_m^2 into an interpolation can be achieved by estimating the measurement error precision σ_m with a variance of σ_m^2 and incorporating it into the kriging system of linear equations giving

$$\begin{pmatrix} c(\mathbf{x}_1 - \mathbf{x}_1) + \sigma_1^2 \cdots & c(\mathbf{x}_1 - \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ c(\mathbf{x}_n - \mathbf{x}_1) & \cdots c(\mathbf{x}_n - \mathbf{x}_n) + \sigma_n^2 \end{pmatrix} \begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix} = \begin{pmatrix} c(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ c(\mathbf{x}_n - \mathbf{x}_0) \end{pmatrix}. \tag{2.84}$$

In contrast to the conventional formula, σ_m^2 with regard to the considered known value at $\mathbf{x}_k$ is added in the diagonal of the matrix (Wackernagel 2003). This accounts for the heteroscedastic nature of geological parameters as they commonly show a higher variability for high values and a lower variability for low values (Deutsch & Journel 1998).

Spatial Bayesian Networks

Another group of algorithms aiming to produce spatial predictions is represented by stochastic simulations which belong to the class of Bayesian networks.

Most interpolation techniques do not represent the original parameter distribution adequately and induce a smoothing effect in the spatial distribution. Furthermore, the original

histogram, semivariogram and variability are not reproduced. Conditional simulations, in contrast, aim to preserve the natural variability of natural phenomena (Remy 2004). Geostatistical simulations in general aim at reproducing both the global histogram and the considered variogram model (Leuangthong et al. 2004).

Some simulation algorithms, such as the sequential Gaussian simulation, are based on the multi-Gaussian approach that assumes data to be normally distributed. Therefore, data needs to be transformed into normal space using a normal score transform before applying the algorithms (Deutsch & Journel 1998).

Sequential Gaussian Simulation

In order to account for the spatial heterogeneity of a property the sequential Gaussian simulation (SGS) algorithm can be utilized. SGS is a popular technique for simulating conditional Gaussian random fields, being based upon the SK interpolation technique.

SGS is based on the multi-Gaussian approach (Goovaerts 1997), which assumes that the kriging error is standard normally distributed with $\mu = 0$ and $\sigma_{SK}^2 = 1$. This requires that each one-point cumulative density function (CDF) of any linear combination of the random vector (RV) is normally distributed, that all subsets of the random field (RF) are multivariate normal, that the two-point distribution is normal and that all conditional distributions of subsets of the RF are normal (Goovaerts 1997).

If the RF fulfills the requirements, then the simple kriging estimate and variance characterize the posterior cumulative CDF under consideration of the normal score variogram model. Thus, we need to transform the original distribution's CDF into standard normal space for SGS (see Eq. 2.28). The algorithm for one SGS realization is provided in Algorithm 1.

Once the global CDF is transformed into standard normal space and the covariance function being defined, each location in the target domain is visited on a random base. At each visited location, the conditional SK estimate and variance is determined, which are conditioned by both the global constraints and the previously simulated values.

For each interpolated point $\mathbf{x}_i$ now a random value of the normal distribution $\mathcal{N}\left(z(\mathbf{x}_i), \sigma_{SK}^2\right)$ is drawn as $z(\mathbf{x}_i)$ using the Box-Muller transform that can be expressed as

$$z(\mathbf{x}_0) = \sqrt{-2 \cdot \log(u_1) \cdot \cos(2\pi \cdot u_2)} \cdot \sigma + \mu \qquad (2.85)$$

with u_1 and u_2 as random numbers $\in [0, 1]$, σ as the standard deviation and μ as arithmetic mean of the original distribution. One problem is obvious here. The more points being

ALGORITHM 1

Sequential Gaussian simulation algorithm

Given: Ω; $\mathbf{x}$; N	$\triangleright$ Target domain; Sampled locations; Neighborhood information;
Initialize: $\mathbf{u_{Sim}}$; $\mathbf{x}'$	$\triangleright$ Simulated locations; Spatial neighbors;
$Y(\mathbf{x}) \leftarrow$ Eq. 2.28	$\triangleright$ Transform to standard normal space
$\gamma(\mathbf{h}) \leftarrow$ Eq. 2.72	$\triangleright$ Estimate the experimental variogram
$\gamma(\mathbf{h})_{sph} \leftarrow$ Eq. A.14	$\triangleright$ Derive the variogram model and the covariance function

for all u_i in Ω **do**

    $\mathbf{x}' \leftarrow$ Eq. 2.67 & Eq. 2.68 $\triangleright$ Determine the neighborhood with N, $\mathbf{x}$ & $\mathbf{u_{Sim}}$

    $\mu_{SK} \leftarrow$ Eq. 2.77 using $\gamma(\mathbf{h})_{sph}$ $\triangleright$ From $\mathbf{x}'$

    $\sigma^2_{SK} \leftarrow$ Eq. 2.80 using $\gamma(\mathbf{h})_{sph}$ $\triangleright$ From $\mathbf{x}'$

    $z(\mathbf{u_i}) \leftarrow$ Eq. 2.85 from $\mathcal{N}\left(\mu_{SK}, \sigma^2_{SK}\right)$ $\triangleright$ Draw a value with σ^2_{SK}

    Add $z(\mathbf{u_i})$ to $\mathbf{u_{Sim}}$

end for

$F(Z(\mathbf{u})) \leftarrow$ Eq. 2.29, $\triangleright$ Back-transform the simulated values into the original space

---

simulated the lower is the simple kriging variance, which is primarily dependent upon the Euclidean distance to the locally neighboring points (see Equation 2.80). This is a problem, which is addressed in **Publication IV** within this Thesis.

Direct Sequential Simulation

The reproduction of the covariance model does not require the multi-Gaussian approach as long as the mean and variance are derived from the SK estimation (Robertson et al. 2006; Soares 2001). Thus, the conditional distribution type chosen in order to simulate the variability at each point does not necessarily need to be Gaussian. With this in mind, it is evident that a normal score transform is not needed before applying the sequential simulation.

When the sequential simulation is performed 'directly' we speak of a direct sequential simulation (DSS). The algorithm for DSS is slightly different to the algorithm for SGS as the transformation into standard normal space is not required.

In the case of a high SK variance, DSS tends to drift towards extremely high or extremely low values. Hence, either a back-transformation into the original variable's space might be necessary or a mapping procedure during simulation must be performed in which the simulated value is projected onto the CDF of the original variable (Soares 2001; Robertson et al. 2006).

2.4.4.10. Validation

Cross-validation

In order to assess the quality of a realization, models, which are constructed by means of interpolation or simulation techniques, should be validated. Commonly, interpolations are validated by cross-validation. This technique is usually performed by using point removal procedures called leave-p-out cross-validation (LpO CV). For the LpO CV, p randomly selected samples are removed from the input data set of size n with $0 < p < n$ and the interpolation or simulation is performed without these samples (Celisse 2014). As measures of goodness of fit, the mean-square error (MSE, Eq. 2.86), the root-mean-square error (RMSE, Eq. 2.87) and the mean-absolute error (MAE, Eq. 2.88) of the realization can be calculated as

$$MSE = \frac{1}{n} \sum_{k=1}^{n} \left(\hat{z}(\mathbf{x}_k) - z(\mathbf{x}_k) \right)^2, \tag{2.86}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(\hat{z}(\mathbf{x}_k) - z(\mathbf{x}_k) \right)^2} \tag{2.87}$$

and

$$MAE = \frac{1}{n} \sum_{k=1}^{n} |\hat{z}(\mathbf{x}_k) - z(\mathbf{x}_k)|, \tag{2.88}$$

where $\hat{z}(\mathbf{x}_k)$ are the estimated points. While Willmott et al. (2009) question the status of the triangle inequality for the RMSE, which is required for a distance function metric, Chai & Draxler (2014) show that the RMSE in fact fulfills this condition. Thus, if the model errors follow a normal distribution, the RMSE is to favor over the MAE (Chai & Draxler 2014).

Geostatistical algorithms usually reduce the number of neighbors that are used to constrain an estimation or simulation. In section 2.4.4.9 it is described, how a subset of the available input data set can be selected by means of a 3-D ellipsoid. By reducing the neighborhood, the covariance model among the selected points slightly deviates from the global covariance model. Accordingly, the estimation of an unknown point is affected by using the subset of the globally available input points which is why the cross-validation errors can be used for optimizing the modeling approach. Figure 2.34 shows that the

dependence of an interpolation is strongly dependent on the selected neighborhood (Linsel et al. 2020b).
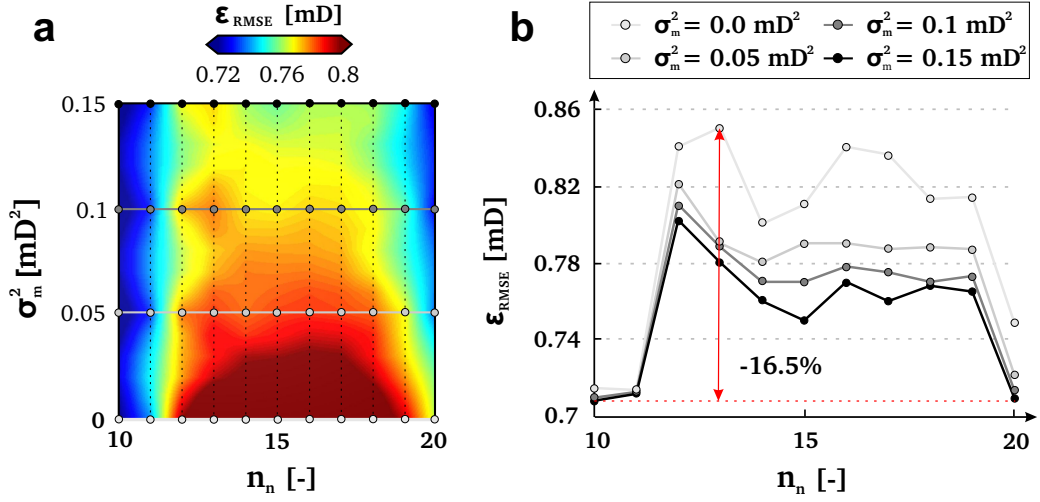
Figure 2.34.: Results of a linear integer programming optimization targeting the minimization of the interpolation error ϵ_{RMSE} for a spatial interpolation. **(a)** RMSE response surface with regard to the incorporated measurement error variance σ_m^2 and the maximum number of neighbors n_n using a leave-one-out cross-validation. **(b)** Cross sections through the response surface of **(a)** (Linsel et al. 2020b).

Curve Fitting

Function approximations like produced in regression analyses are commonly validated using the coefficient of determination (R^2) which is calculated like

$$R^2 = 1 - \frac{s_{res}}{s_{tot}} \in [0, 1], \tag{2.89}$$

where

$$s_{res} = \sum_{k=1}^{n} (y_k - f_k)^2 \tag{2.90}$$

is the explained sum of squares whereas

$$s_{tot} = \sum_{k=1}^{n} (y_i - \overline{y})^2 \tag{2.91}$$

is the total sum of squares.

2.5. Geothermal Reservoir Characterization by Means of Geovisual Analytics

2.5.1. Introduction

The definition of geothermal energy has not been globally standardized yet. In this Thesis, the definition formulated in a directive of The European Parliament and the Council of The European Union (2009) on the promotion of the use of energy from renewable sources will be used further on where '*geothermal energy means energy stored in the form of heat beneath the surface of solid Earth*'. In this directive, geothermal energy is categorized under *energies from renewable sources*.

Figure 2.35 shows a conceptual illustration of the exploitation strategy for hydrothermal electricity production. Hereby, heat is transferred from the geothermal fluid to a secondary working fluid in a heat exchanger. The pressurized working fluid is vaporized and subsequently expanded through a turbine which drives the electrical generator (see, e.g, Mines 2016).
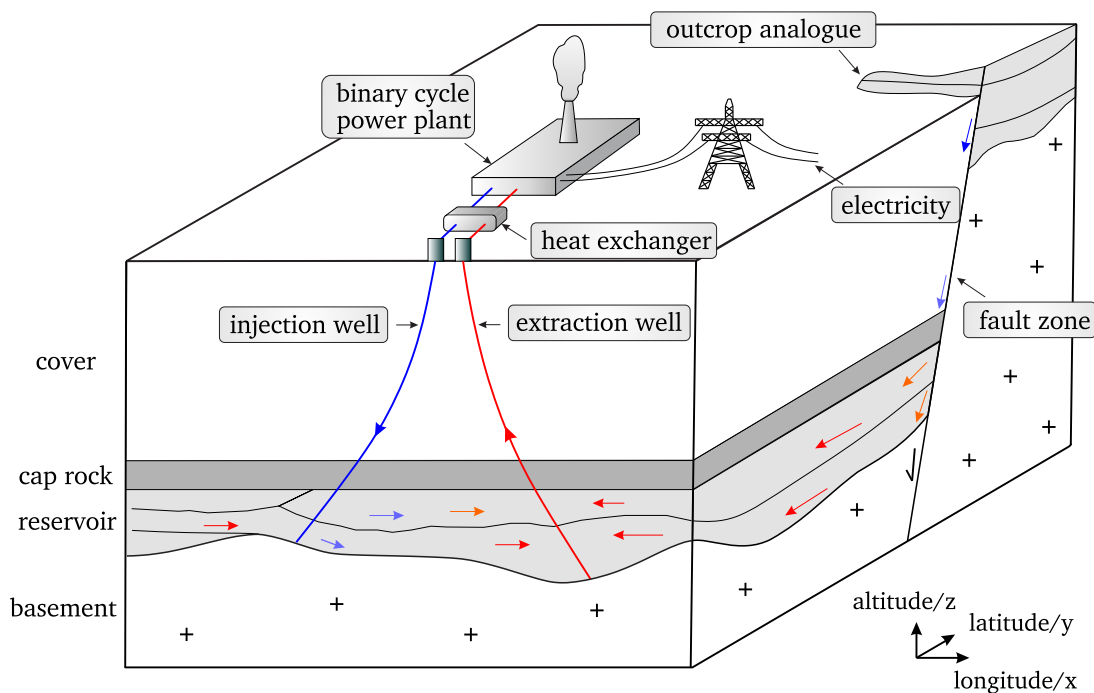
Figure 2.35.: Schematic concept of a binary cycle power plant used to produce electricity with heat extracted by a hydrothermal doublet. Reservoir information can be derived either directly from reservoir depth or in outcrop analogue studies providing insights into the lithological architecture and property distribution within the reservoir unit from surface or near-surface outcrops. Blue (cold) and red (hot) arrows indicate potential fluid trajectories relevant for deep hydraulic circulation.

2.5.2. Controls on a Reservoir's Production Behavior

The performance of a geothermal reservoir is dependent on the internal and adjacent geological architecture, the corresponding rock and fluid properties, together with ambient properties such as the temperature and stress field (Rühaak et al. 2015a; Welsch et al. 2016). Recent studies demonstrated that for each geological scenario, a separate simulation of a geothermal reservoir's performance must be performed in order to reduce the investment risks for geothermal power plants (Schulte et al. 2020). Thus in geothermal reservoir characterization the spatial distribution of all of those characteristics must be taken into account. Figure 2.36 illustrates the spatial characteristics of a temperature field in a rock formation which was modeled by a boundary-conforming grid.

The geological architecture describes the variability and spatial distribution of rock types in association with their compartmentalizing elements in the form of faults, bounding surfaces and fractures in the reservoir. Those characteristics are dependent on geological processes which are taking place during genesis, tectonic activity, burial and exhumation of a geological medium. The processes involved comprise weathering, erosion, transport, deposition, diagenesis and metamorphism. Usually, these characteristics are addressed by subdividing the region of interest into lithological units and a representation of structural elements, based on the purpose that must be served.

The rock and fluid properties are controlled by the lithological, chemical and physical characteristics such as grain size distribution, grain shape, mineralogical composition, grain contacts, and heat content, which may provide (multivariate) functional relationships among one another. Some rock properties differ in their mathematical nature. Permeability, stress, thermal conductivity/diffusivity for instance are direction-dependent parameters and should be considered as tensors (Popov et al. 1999; Clavaud et al. 2008; Dürrast 2004) whereas porosity or density are parameters of scalar type.

The diversity in mathematical formats increases the effort to normalize reservoir-related data models and hampers data analytics and modeling of the properties which is often leading to inaccurate simplifications of heat or mass transfer simulations as depicted in section 2.5.3.
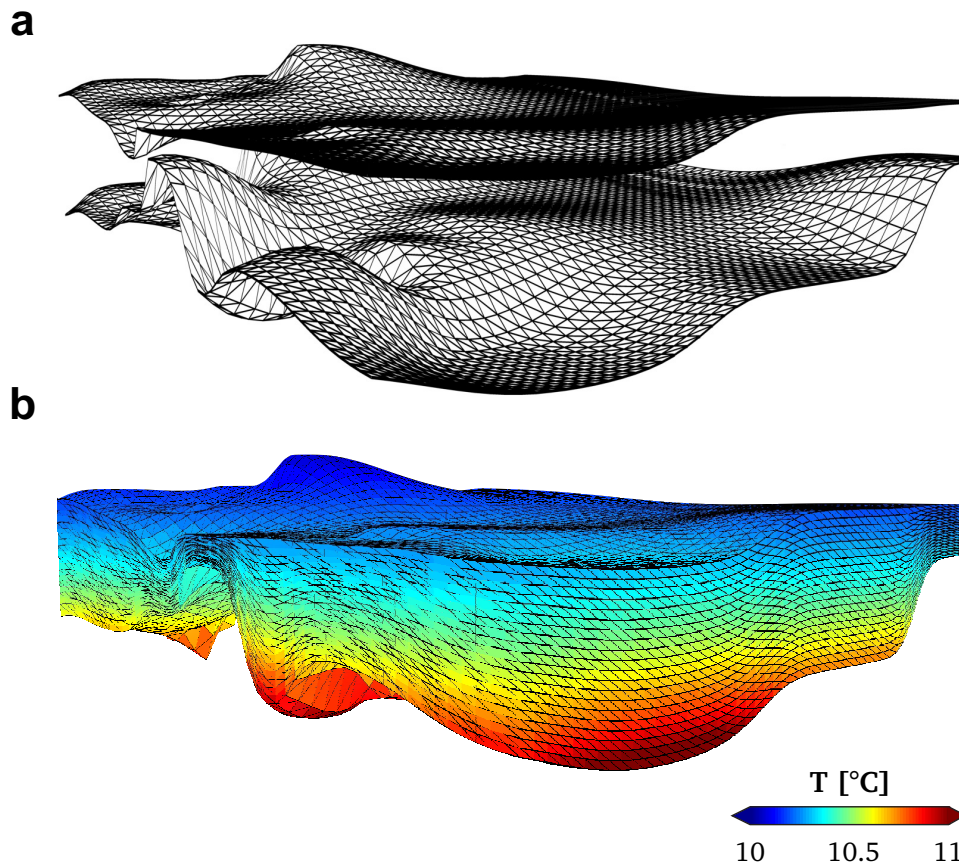
Figure 2.36.: (a) The model represents the top and base of the volcano-sedimentary Rotliegend unit in the Sprendlinger Horst area which is located in Hesse, Germany. **(b)** Exemplary near surface temperature distribution in a 3-D model constructed from both surfaces given in **(a)**. The model has a span of 3,000 × 3,000 × 70 m and is 10-fold exaggerated in vertical direction.

2.5.2.1. Relevant Properties

Recent approaches aiming at the subdivision of properties into property groups came up with a scheme that partitions the properties of reservoir rocks into petrophysical, hydraulic, mechanical and thermophysical properties (Bär et al. 2020). This subdivision applies well if only rock properties are considered as it is implemented in the P^3 database of Bär et al. (2020).

A more generic view would imply that also fluid properties must be considered which brings us to the subdivision of relevant material and ambient properties in geothermal systems as depicted in Table 2.2. Displayed are the general forms of the properties which means that properties such as density can be further subdivided into rock density and fluid density and, when considering rock density, into grain density and bulk density, respectively.

Property	Symbol	Unit	Isotropic
Thermophysical properties			
thermal conductivity	λ	W m^{-1} K^{-1}	no
volumetric heat capacity	c	J m^{-3} K^{-1}	yes
thermal diffusivity	α	m^2 s^{-1}	no
radiogenic heat production	A	W m^{-3}	yes
temperature	T	K	yes
enthalpy	H	J	yes
heat	q	J	yes
Petrophysical properties			
density	ρ	kg m^{-3}	yes
mass	m	kg	yes
Hydraulic properties			
porosity	Φ	[-]	yes
permeability	k	m^2	no
hydraulic conductivity	k	m s^{-1}	no
Mechanical properties			
P-wave velocity	v_p	m s^{-1}	no
S-wave velocity	v_s	m s^{-1}	no
stress	σ	kg m^{-1} s^{-2}	no
pressure	p	kg m^{-1} s^{-2}	yes

Table 2.2.: Thermophysical, petrophysical, hydraulic and mechanical material properties which contribute towards the potential of geothermal systems. Adapted from the data model of the geothermal P^3 database of Bär et al. (2020) and extended by properties which are relevant for heat and fluid flow processes in the subsurface (see section 2.5.3).

The collection of material properties could be extended by ones which are of secondary interest in the process of reservoir characterization. Properties such as the gravitational field strength [N kg^{-1}] or the magnetic susceptibility [-] which aid in understanding the subsurface system do not directly contribute towards the economic potential of a geothermal system.

The ambient properties such as the recent stress field and the temperature distribution mostly depend on regional to global scale processes such as the natural heat flow, the spatial characteristics of fluid convection systems and plate tectonic movements. Just like the material properties of the geological media, the ambient properties are either directly contributing towards a geothermal reservoir's potential or they might proof useful in another way within the reservoir's potential assessment process.

2.5.3. Conductive Heat Transport in the Subsurface

2.5.3.1. Introduction

Many approaches which aim to simulate heat or mass transfer processes in the subsurface are based on simplifications, such as spatial homogeneity or isotropy of rock properties. By the means provided in section 2.3, such simplifications can be overcome by modeling the relevant properties in space which requires that the grade of sample coverage in the region of interest is reasonably high.

In this section, the process of conductive heat transport in the subsurface will be outlined. This process provides the base for subsurface heat storage systems whose quality is highly dependent on the rock properties and their spatial distribution in the subsurface (see, e.g, Welsch (2019) or Rühaak et al. (2015)).

2.5.3.2. Physical Fundamentals

Heat q is a part of the inner energy U that thermodynamic systems can consume or emit. Together with the physical quantity work W it controls the difference of the inner energy in a system (Eq. 2.92).

$$\Delta U = q + W \tag{2.92}$$

The temperature field in the subsurface is controlled by processes that either transfer or consume heat. The temperature T and heat are related as

$$q = m \cdot c \cdot \Delta T, \tag{2.93}$$

where m is the mass of a medium and c is the specific heat capacity.

The total heat in place in a subsurface region can be calculated as

$$
\begin{aligned}
q &= q_s + q_l \\
&= V \,\overline{\phi}\,\overline{c}\,(T_R - T_r) \\
&= \int \left[(1 - \phi)\, c\, \rho_b\, (T - T_{ref}) + \phi\, \rho_w\, (H_w(T) - H_w(T_{ref})) \right] dV
\end{aligned} \tag{2.94}
$$

where q_s is the heat in the solid phase, q_l the heat in the liquid phase, ϕ is the effective porosity of the solid subsurface medium and ρ_b is its bulk density. $H_w(T)$ is the enthalpy of the fluid filling the medium's pore space at a specific temperature T and ρ_w is the fluid's density. V is the volume of the considered subsurface region (Garg & Combs 2015; Muffler & Cataldi 1978). It is particularly important to select T_{ref} appropriately since it determines the fraction of recoverable heat in the medium. Commonly, the ambient temperature (≈ 15 °C) or the condenser temperature (≈ 40 °C) are used (Williams 2014).

2.5.3.3. Governing Equations for Heat Transport

Heat can be transported between physical systems by four mechanisms, namely, conduction, convection, radiation and phase changes. For geothermal processes, mainly convection and conduction apply thus we will reduce the further explanation to those. The formal interaction of both processes can be expressed as

$$\frac{\partial T}{\partial t} + \mathbf{r}^T \nabla T = \nabla^T (\boldsymbol{\alpha} \nabla T) + \frac{A}{\rho c}, \tag{2.95}$$

where $\mathbf{r}$ is the fluid flow vector that can be notated as

$$\mathbf{r} = \left(r_x, r_y, r_z \right), \tag{2.96}$$

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right), \tag{2.97}$$

$$\nabla T = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z} \right), \tag{2.98}$$

and α is the thermal diffusivity tensor that is defined as

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix} \tag{2.99}$$

under assumption of a transverse isotropy. When considering conductive heat transport only, Eq. 2.95 simplifies to

$$\frac{\partial T}{\partial t} = \nabla^T \boldsymbol{\alpha} \nabla T + \frac{A}{\rho c}. \tag{2.100}$$

In order to solve the heat transfer problem by means of the finite element method (FEM) an initial temperature field has to be defined (Fig. 2.37). Additionally, ρ, c, A and $\mathbf{K}$ have to be defined in Ω. Moreover, the uniqueness of a PDE is only guaranteed when the boundaries are imposed with specific conditions called boundary conditions. The following four types of boundary conditions can be encountered:

I) Dirichlet boundary conditions are defined, where the solution of an interval of the domain is fixed (Eq. 2.101).

$$f(x) = a \tag{2.101}$$

II) Neumann boundary conditions define the values in which the derivative of the function should be defined on a boundary (Eq. 2.102).

$$f'(x) = a \tag{2.102}$$

III) Robin boundary conditions, which define the boundary condition as a linear combination of the solution and its derivative.

$$u\,f(x) + v\,f'(x) = a \tag{2.103}$$

IV) Periodic boundary conditions, which assume the solution at one end of the domain is equal to the solution at another end.
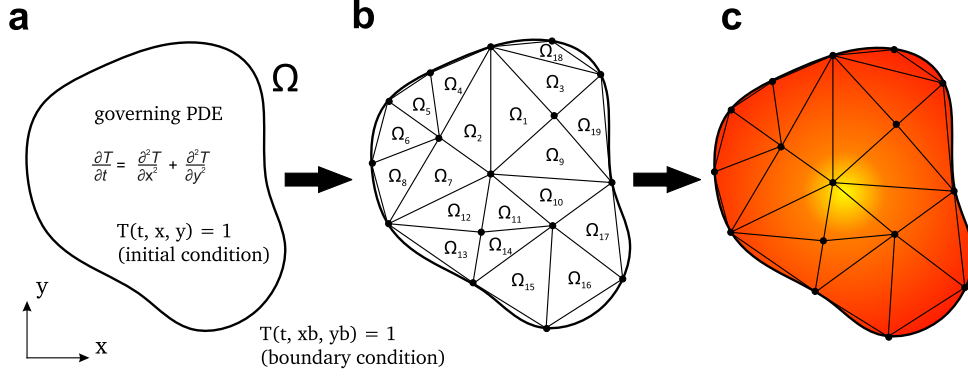
Figure 2.37.: Schematic concept of the finite element method.

If we assume that a domain in the subsurface has been discretized by a set of hexahedral elements, the temperature T at the nodes $\mathbf{x}_i$ in one element can be determined as

$$T \simeq \sum_{i=1}^{8} N_i \cdot T_i \qquad (2.104)$$

with N_i as trilinear shape function at the node $\mathbf{x}_i$. The shape functions for an eight-node hexahedron in terms of local coordinates are listed in Appendix A.1.4. In order to convert the derivatives of the shape function from the local coordinate system to the physical coordinate system, the chain rule can be used as follows

$$\begin{pmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta} \\ \frac{\partial}{\partial \zeta} \end{pmatrix} \mathbf{N} = \begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \mathbf{N} = \mathbf{J} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \mathbf{N}, \qquad (2.105)$$

where $\mathbf{J}$ is the Jacobian matrix which is defined as

$$\mathbf{J} = \begin{pmatrix} \frac{\partial N_1}{\partial \xi} & \frac{\partial N_2}{\partial \xi} & \cdots & \frac{\partial N_8}{\partial \xi} \\ \frac{\partial N_1}{\partial \eta} & \frac{\partial N_2}{\partial \eta} & \cdots & \frac{\partial N_8}{\partial \eta} \\ \frac{\partial N_1}{\partial \zeta} & \frac{\partial N_2}{\partial \zeta} & \cdots & \frac{\partial N_8}{\partial \zeta} \end{pmatrix} \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_8 & y_8 & z_8 \end{pmatrix}. \qquad (2.106)$$

By substituting T in Eq. 2.100 by Eq. 2.104, following discrete element-level system of

equations can be derived:

$$\iiint \mathbf{N}^T \mathbf{N} \, dx \, dy \, dz \, \frac{\partial}{\partial T} + \iiint (\nabla \mathbf{N})^T \mathbf{K} \nabla \mathbf{N} \, dx \, dy \, dz \, \mathbf{T} = \iiint H \mathbf{N}^T \, dx \, dy \, dz.$$

(2.107)

In matrix form, the equation can be expressed as

$$[\mathbf{MM}] \frac{\partial}{\partial t} \mathbf{T} + [\mathbf{KM}] \mathbf{T} = F,$$

(2.108)

which builds the elemental equation for solving the problem by using the FEM. $[\mathbf{MM}]$ thereby represents the element mass matrix while $[\mathbf{KM}]$ constitutes the element stiffness matrix.

The only step left here is the time discretization which is commonly implemented as an implicit finite difference approximation yielding Equation 2.109.

$$\left(\frac{\mathbf{MM}}{\Delta t} + \mathbf{KM} \right) \mathbf{T}^{n+1} = \frac{\mathbf{MM}}{\Delta t} + F$$

(2.109)

3. Publications

3.1. Publication I – Development of an Open-Source Knowledge Discovery System for Subsurface Characterization

3.1.1. Introduction

Publication I constitutes the key publication of this dissertation. In this chapter, an open-source software system called GeoReVi is presented. The acronym GeoReVi stands for *Geological Reservoir Virtualization*. The source code of GeoReVi is available at GitHub[3].

GeoReVi is aimed at addressing a series of tasks associated with subsurface characterization generically. Nowadays, it is still common-practice to make use of multiple software packages – most of which are commercially distributed such as SKUA-GOCAD™ or ANSYS® (ANSYS, Inc. 2020) – in order to perform typical workflows for subsurface modeling in both industries and sciences (Enge et al. 2007; Zehner et al. 2015).

In order to use all of the aforementioned software packages, a substantial amount of money must be spent for licensing. Moreover, each software usually requires a specific file format for the intermediate products which must be transformed from others under a high effort. These obstacles hamper scientific investigations frequently. Although the overall trend – also in the geoscientific community – tends towards the development of open-source software, there is still a lack of open-source geological data management, 3-D modeling and data analysis systems as the development of such software is hard to plan and requires expert knowledge about the computer science, mathematics and geosciences itself.

---

[3]https://github.com/ApirsAL/GeoReVi, last access 23rd November, 2020
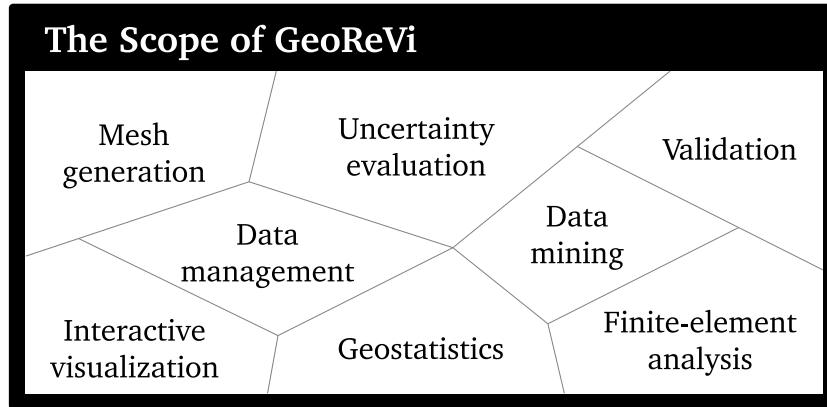
Figure 3.1.: The scope of GeoReVi comprises multiple disciplines that are involved in typical knowledge discovery and decision-making processes during subsurface utilization.

Consequently, there was a potential gap to be filled by providing a generic open-source software platform enabling geoscientists to performing widely applied workflows in one and the same product, which they can adapt to their specific needs. The scope of GeoReVi comprises those entities being illustrated in Figure 3.1.

GeoReVi has been developed with the intention to accompany a domain expert during the entire process of subsurface characterization. Consequently, widely applied workflows commonly performed in subsurface studies in which data can be documented all along the process chain have been implemented in GeoReVi.

Specific focus is hereby laid upon the implementation of a database being connected to a GUI which provides functionality for stratigraphic mesh generation combined with geostatistical and data mining algorithms. These aid the geoscientist to explore the statistical relationships of subsurface properties in space and time which are persistently residing in the database.

Once data are stored in the database, they can be loaded into the buffer memory, where they can be analyzed by means of geovisual analytics as presented in the sections 2.4.2 and 2.4.4. Furthermore, the interactive visualization functionality for modeling results together with validation and uncertainty assessment routines allow for the immediate inspection and evaluation of intermediate products in the process of subsurface characterization. Exemplary outputs of GeoReVi are provided in Figure 2.14.

3.1.2. Author Contributions

- My contribution as the first author of this paper comprises the planning and implementation of the software tool including the development of the databases, the

creation of the graphical user interface and the implementation of the numerical algorithms, the conceptualization and conduct of the case study, the model generation, the preparation of the figures and the statistical data analyses.

- K. Bär, J. Hornung, and M. D. Greb contributed to the creation of the logical data model and to planning the case study.

- J. Haas conducted with me the field work and the laboratory experiments.

- M. Hinderer was the overall supervisor of the research.

- All co-authors contributed to the preparation and revision of the manuscript.

3.1.3. Publication

Published as:
Linsel, A., Bär, K., Haas, J., Hornung, J., Greb, M.D. and Hinderer, M. (2020): GeoReVi: A knowledge discovery tool for subsurface characterization *SoftwareX*.
doi:10.1016/j.softx.2020.100597

Original software publication

GeoReVi: A knowledge discovery and data management tool for subsurface characterization

Adrian Linsel [a],[*], Kristian Bär [a], Joshua Haas [a], Jens Hornung [a], Matthias D. Greb [b], Matthias Hinderer [a]

[a] Technische Universität Darmstadt, Institute of Applied Geosciences, 64287 Darmstadt, Germany
[b] Devon Energy, Oklahoma City, OK, United States

ARTICLE INFO

ABSTRACT

Subsurface characterization is an interdisciplinary and multidimensional problem requiring contribution from numerous geoscientific and technical domains. In order to optimize and automate the process of subsurface characterization and structural modeling we developed a modular, open-source software system called GeoReVi (Geological Reservoir Virtualization). The tool implements the knowledge discovery in databases (KDD) process and utilizes techniques from visual analytics for interactive, interdisciplinary, database-bound knowledge discovery and communication. Multidimensional data sets – produced in subsurface and outcrop analog studies – can be imported, shared, transformed, projected, analyzed, modeled, grouped and visualized interactively in a custom-made graphical user interface. The underlying data model facilitates domain experts to efficiently work in multi-user environments. The knowledge discovery potential is illustrated with an exemplary case study.

Code metadata

Current code version	v1.0.1
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX_2019_352
Legal Code License	GNU General Public License v3.0
Code versioning system used	git
Software code languages, tools, and services used	C#, XAML, WPF
Compilation requirements, operating environments & dependencies	Microsoft Windows, Entity Framework 6, HelixToolkit, Accord.NET, WPF, Caliburn.Micro, FontAwesome.WPF, DotSpatial, GeoAPI, ProjNET, LiteDB, MoreLinq, Extended WPF Toolkit, Math.NET
Support email for questions	contact@georevi.com

Software metadata

Current software version	v1.0.1
Permanent link to executables of this version	https://github.com/ApirsAL/GeoReVi/releases/tag/v1.0.1
Legal Software License	GNU General Public License v3.0
Computing platforms/Operating Systems	Microsoft Windows
Installation requirements & dependencies	Microsoft Windows
If available, link to user manual - if formally published include a reference to the publication in the reference list	https://github.com/ApirsAL/GeoReVi/blob/master/docs/GeoReViUserManual.pdf
Support email for questions	contact@georevi.com

1. Introduction

Subsurface characterization is a crucial step when planning its utilization. A prominent example is the development of natural resources hosted in a subsurface region. Here, with regards to geothermal, groundwater or hydrocarbon reservoirs, a region in the subsurface needs to be characterized in order to assess its petrophysical properties and to predict the performance of the resource exploitation process. Reservoir characterization is defined as '*the process of preparing a quantitative representation of a reservoir using data from a variety of sources and disciplines*' [1].

---

* Corresponding author.
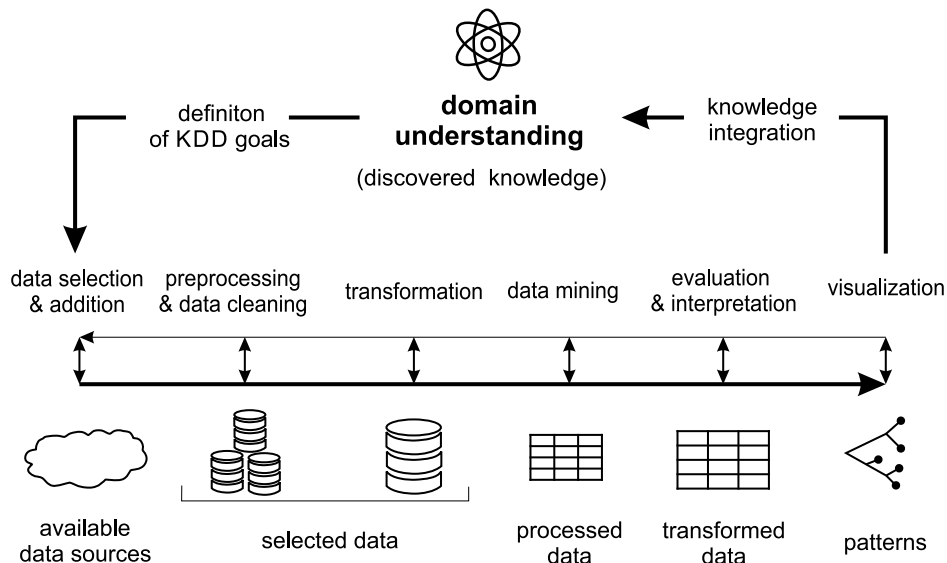 E-mail address: linsel@geo.tu-darmstadt.de (A. Linsel).

Fig. 1. Knowledge discovery in databases process after [2] and [3].

Accordingly, as numerous physicochemical rock and fluid properties contribute to a reservoir's potential, this problem can be considered being multidimensional with numerous contributing domains. However, any type of subsurface utilization requires a comprehensive knowledge of the subsurface architecture and the spatial distribution of relevant properties such as permeability and porosity for making profitable, business-critical yet sustainable decisions [4]. Consequently, modern automatized technologies for integrated exploration and modeling of subsurface-related data are required to increase subsurface predictability and to eventually optimize the subsurface utilization process.

Subsurface-related data are aggregated from multiple scales and numerous domains with contrasting ontologies [5]. These data are produced in well log measurements, in outcrop analog studies, drill core investigations, geophysical surveys or during development and can be both static (e.g. depth of lithological bodies) or dynamic (e.g. production rates). Numerous fundamental rock properties differ in their physical and mathematical nature. Rock properties such as permeability, stress, thermal conductivity/diffusivity are direction-dependent tensors [6–8] whereas porosity or grain density are scalar quantities. This diversity in mathematical and physical formats increases the effort to normalize reservoir-related data models and hampers data analysis and modeling of the properties. Often, these issues lead to inaccurate simplifications during flow and mass transport simulations such as assuming rock and fluid properties to be isotropic and homogeneously distributed in space and/or time.

With this study, we intend to bridge the gap between data collection, data management and integrated data analysis and visualization in the process of subsurface characterization. Therefore, we developed a software system called GeoReVi (*Geological Reservoir Virtualization*) with an internal implementation of the knowledge discovery in databases (KDD) process covering multiple aspects of visual analytics according to [9]. This software system enables domain experts to interactively manage and analyze any kind of spatial and multidimensional data sets through data processing, transformation, selection, import and mining algorithms.

1.1. Knowledge discovery in databases

The interdisciplinary field of KDD comprises a set of semi-automatic, non-trivial methods to extract novel, understandable, valid and useful patterns from domain-specific data sets stored in mature databases [3,10]. Those patterns are evaluated by the domain expert in order to extract '*knowledge*' [3]. The relationship between data and knowledge is commonly illustrated with the Data–Information–Knowledge–Wisdom (DIKW) hierarchy, which has been reviewed by [11]. The first conceptualization of the hierarchical representation of data, information and knowledge is defined in [12] and [13]. [13] defines data as '*symbols*', information as '*data that are processed to be useful*', knowledge as the '*application of data and information*' and wisdom as the '*appreciation of why*'. We will adhere to [13] and refer to knowledge as being '*know-how*' enabling information to be transferred into instructions. An example of this is planning a borehole for which we need information about the spatial distribution of stratigraphic units, structural elements and of physicochemical rock and fluid properties. Fig. 1 illustrates the iterative, 8-stage KDD process, which was conceptualized by [2] and redesigned and improved by [10]. The process starts with the domain expert's knowledge that is aimed to be maximized within the KDD process. The domain can be any discipline in which data are produced, managed and analyzed. Core of the KDD process is a set of data mining (DM) algorithms, which are deployed on a processed data set in order to find characteristic patterns or models. Prior to DM, data are selected, projected, cleaned, reduced and transformed by the domain expert. Pre-processing is supported by computer-aided process automation and intelligent pre-selections. DM algorithms in general comprise classification, summarization, correlation, regression, prediction and rule-discovery algorithms, whereby each DM algorithm is best suited to a specific problem.

2. Software description

2.1. Software architecture

GeoReVi is structured according to a client–server architecture. An illustration of the overall architecture is provided in Fig. 2a. GeoReVi is intended to be used in private, multi-user networks by using an application role authentication together with user credentials, which can be used to store and retrieve user-specific data. The server-side data storage is implemented as a relational database management system. However, GeoReVi can also be used in a local mode, where all data are stored in an integrated NoSQL database. GeoReVi targets the exploration
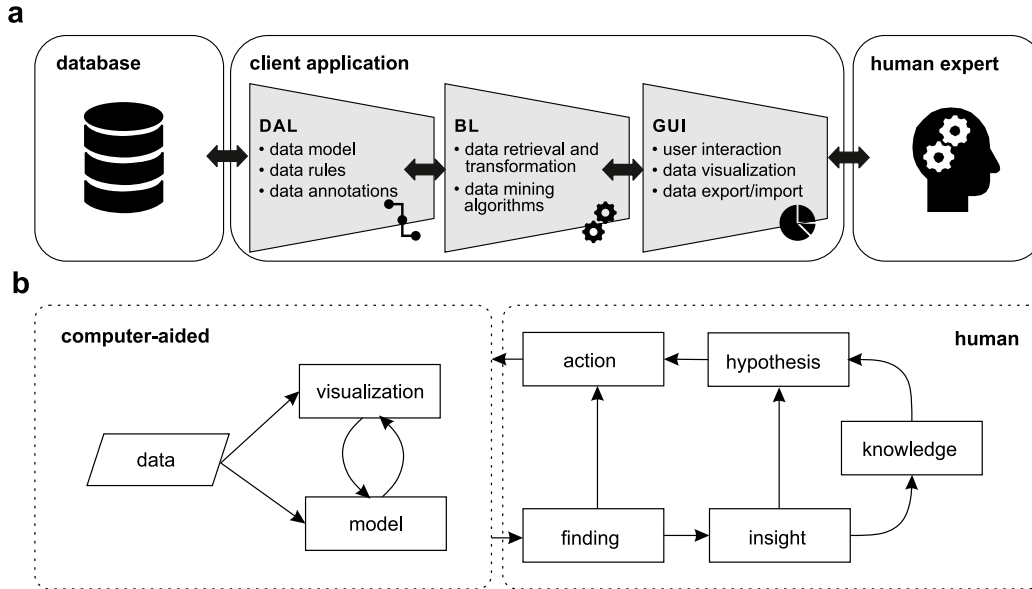
Fig. 2. (a) System architecture of GeoReVi (DAL = data access layer; BL = business logic; GUI = graphical user interface). (b) Knowledge generation model for visual analytics according to [14].
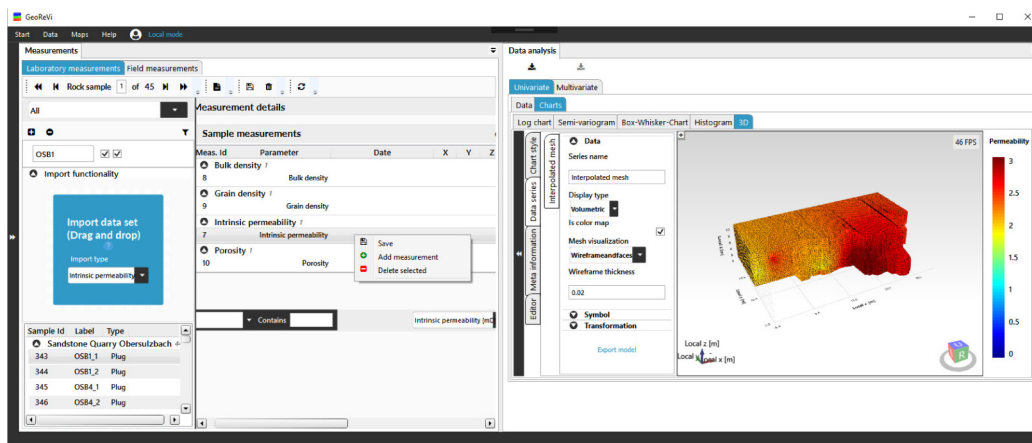
Fig. 3. Graphical user interface of GeoReVi. The measurements view is shown, where readings from laboratory and field measurements can be archived, analyzed and visualized.

loop of the knowledge generation model for visual analytics (KGMVA) after [14] (Fig. 2b). Within this model, a human expert extracts knowledge from huge data sets through interaction with a computer-aided feedback-loop of data selection, modeling and visualization. An important characteristic of a visual analytics system is the continuous interplay of automatic background processes and interactive visualization. With the provided architecture, a domain expert can manage and retrieve subsurface-related data while simultaneously analyzing and modeling a subsurface domain of interest.

2.2. System implementation

GeoReVi's client system has been developed with the *Windows Presentation Foundation* (WPF), which is included in the Microsoft .NET framework. We implemented the Model-View-ViewModel (MVVM) pattern that strictly separates the business logic from the graphical user interface and from the data model by dissociating these components into separated layers. For efficient MVVM-development the *Caliburn.Micro* framework was used. GeoReVi

provides a modular structure using the *Managed Extensible Framework* (MEF), which makes the plug-in-based extension of the system easier for other developers.

In the Data Access Layer (DAL) the implemented data model is represented by a set of Plain Old CLR Objects (POCO) defined in the C# language. For database connectivity, we use the well-established *Entity Framework 6*, an object-relational mapper for the .NET framework, with the Code First approach. In order to provide compatibility to both relational and NoSQL databases, the POCO models were supplemented by data annotations from the *System.ComponentModel* assemblies. The database used for the local version is *LiteDB*, which is open-source and completely written in .NET C# managed code for offline data management. Business logic (BL) was developed with the object-oriented programming language C#. The business logic consists of view models and helper classes. The presentation layer or graphical user interface (GUI) displayed in Fig. 3 has been developed using the XML accent XAML (Extensible Application Markup Language).
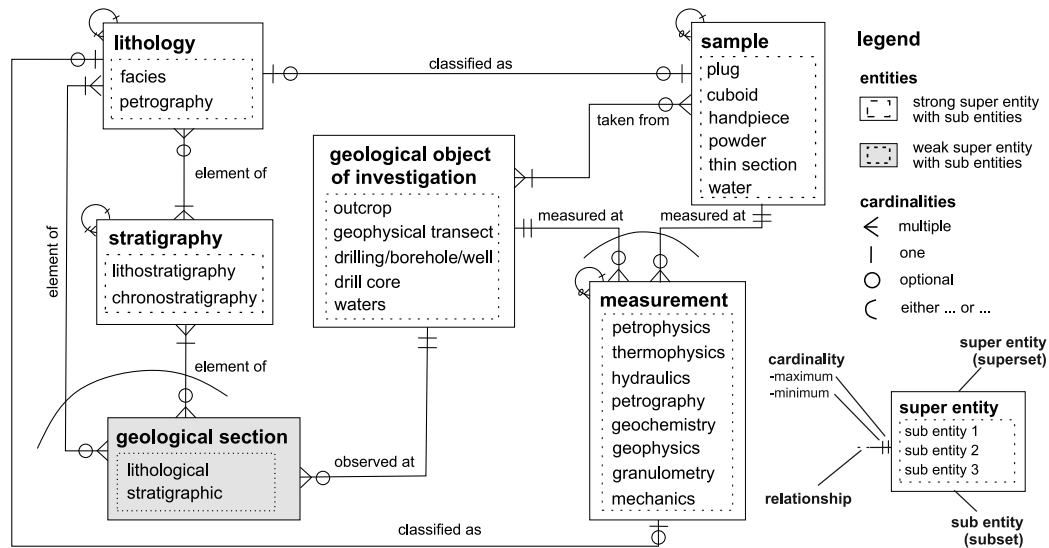
Fig. 4. Core data scheme of GeoReVi represented by an Entity-Relationship model. To provide a better overview, attributes were excluded from the model.

2.3. Software functionality

For a comprehensive overview of the functionality of GeoReVi, we refer to the user manual. However, in the following sections it will be briefly outlined how GeoReVi can be used to store, analyze, model and visualize geoscientific data.

2.3.1. Data storage

GeoReVi's logical data model (LDM, see Fig. 4) comprises the most common entities needed to assess the potential of subsurface geological media to store and extract heat or fluids sustainably. GeoReVi is specialized on geothermal rock and fluid properties, although the most important properties for oil and gas reservoir characterization are included as well. The basis for GeoReVi's LDM is built by the data collection of [15] that was remodeled into a relational data model according to the relational theory after [16]. Additionally, selected parts of the LDM for rock mechanics from [17] and the global geochemical database of [18] were used to extend GeoReVi's LDM. However, most data models are developed within a specific environment, wherefore each LDM had to be adapted to provide compatibility with the core scheme of GeoReVi.

For convenient data storage, custom drop-down menus, list boxes, data grids, color and date picker controls were developed that provide predefined domains. For manual data input, custom text boxes are provided. In order to reduce the number of controls in a view, sub-navigation is implemented via expandable menus or tab controls. To load bulk data sets into the database, a generic import procedure was developed that can semi-automatically assign data from .CSV, .XLS or .XLSX files to the entities in the database.

2.4. Statistical data analysis and subsurface modeling

For statistical data analysis and data visualization the user can load a set of univariate or multivariate measurement values of one or more samples or objects of investigation into the buffer memory, which are displayed in tabular format. Each data set loaded into the buffer memory is by default a mesh. A mesh consists of nodes, faces (quadrilateral or triangular) and cells (hexahedral or tetrahedral). GeoReVi covers individual functionality for 1-D, 2-D and 3-D mesh generation and for 3-D mapping. The user can select the meshes and create the visualizations and

analyses with few commands. The meshes can be parameterized using both stochastic and deterministic algorithms such as inverse distance weighting (IDW), multiple varieties of kriging or conditional simulations. The quality of the models can be assessed with different types of cross-validation. Multivariate analyses in GeoReVi comprise k-Means cluster analysis, principal component analysis, self-organizing maps (Sammon mapping), bivariate regression and correlation analysis. The theory of the most important property modeling and statistical methods is explained in detail in the user manual, which also contains a detailed tutorial.

2.5. Data visualization

Custom controls were developed to visualize the data sets. Therefore, the dependency injection pattern has been implemented, which ensures that the properties of a control are loosely coupled with the properties of the instantiating class. This pattern simplifies the data binding of a view to its view model as well as the navigation between views. Chart controls comprise scatter charts, matrix charts, bubble charts, bar charts, line charts, box-whisker charts, ternary charts and combined line-bar charts. For 3-D visualization, the base functionality from the well-established *HelixToolkit.WPF* was used. In 3-D space, meshes can be visualized as point clouds, volume or surface meshes or as vector fields. Figures and charts can be exported in raster format (.PNG, .JPEG or .BMP) as well as in vector format (.EMF, .XAML or .PDF). 3-D objects can be exported as .OBJ, .X3D or .XAML files or in a custom, XML-based serialization with the extension .GMSH. GeoReVi enables to import and transform 3-D objects in .OBJ format and images in .PNG and .JPG format. Those may serve as target objects for mapping structures that can be used for surface interpolation. Moreover, this function includes ground-penetrating radar (GPR) and seismic datasets if they are provided in one of the aforementioned file formats.

3. Illustrative example

3.1. Spatial heterogeneity of a compartmentalized sandstone formation

In order to demonstrate the capability of GeoReVi to assist the knowledge discovery process connected to geoscientific problems, the intrinsic 3-D heterogeneity of fundamental petrophysical properties within a potential geothermal target formation was
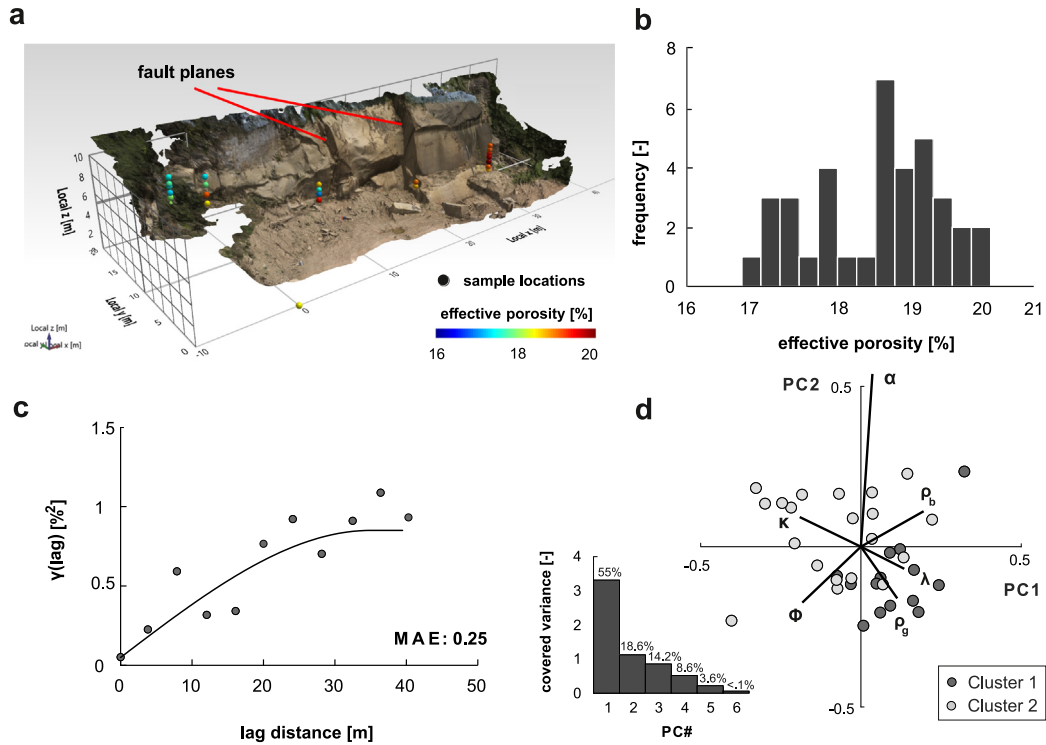
Fig. 5. (a) Photogrammetric model of the investigated sandstone quarry with marked sample locations and structural elements. The faults have a displacement of several meters. (b) Empirical distribution of the effective porosity in the Disibodenberg Formation. (c) Experimental semivariogram and variogram model of the effective porosity used for kriging and sequential Gaussian simulation (MAE = mean-absolute-error) (d) Principal component analysis in the form of a biplot and scree plot combined with a k-Means cluster analysis of the rock properties grain density (ρ_g), bulk density (ρ_b), effective porosity (ϕ), intrinsic permeability (κ), thermal conductivity (λ) and thermal diffusivity (α).

Table 1
Arithmetic mean ($\bar{x}$), minimum (*min*), maximum (*max*), variance (σ^2), coefficient of variation (c_v) and Dykstra–Parson coefficient (c_{dp}) of the measured rock properties: grain density (ρ_g), bulk density (ρ_b), effective porosity (ϕ_e), intrinsic permeability (κ), thermal conductivity (λ) and thermal diffusivity (α).

	ρ_g	ρ_b	ϕ_e	κ	λ	α
Unit	g/cm^3	g/cm^3	%	mD	W/(m · K)	10^{-6} m^2/s
$\bar{x}$	2.66	2.17	18.5	2.64	2.31	1.53
min	2.64	2.12	16.98	0.7	1.99	1.28
max	2.67	2.31	20.23	4.6	2.57	1.75
σ^2	2.1e−5	1e−4	0.6	0.6	0.01	0.01
c_v	0.003	0.015	0.04	0.31	0.07	0.11
c_{dp}	0.002	0.01	0.06	0.31	0.30	0.07

investigated in a quarry that is influenced by tectonic compartmentalization. The case study comprises numerous tasks that are involved in typical subsurface characterization workflows, including data integration, non-orthogonal mesh generation, statistical data analysis, spatial estimation and data visualization.

The investigated outcrop is located in Obersulzbach in southwestern Germany and contains sedimentary rocks from the Disibodenberg Formation (Glan Subgroup) belonging to the Permian Rotliegend Group in the Saar–Nahe-Basin. The Disibodenberg Formation in the quarry is an outcrop analog for the deeply buried formation in the northern Upper Rhine Graben [19]. Here, the deltaic sandstone bodies of the Disibodenberg Formation can be considered as potential hydrothermal reservoirs for power and heat production due to suitable permeability and porosity and sufficient thicknesses [20]. A low-offset strike-slip fault zone, which ranges from x = 14 m to x = 20 m, separates the outcrop into two major parts (Fig. 5a).

Measuring 50 × 15 × 10 meters, the extent of the outcrop is comparable to typical cell sizes of reservoir models built for industrial and scientific applications [21,22]. 36 cylindrical, oriented

rock samples were taken from the outcrop wall and investigated in the laboratory determining the intrinsic permeability, grain and bulk density, effective porosity, thermal conductivity as well as thermal diffusivity (Table 1) of the rock matrix. Those properties can be considered key properties controlling heatflow in porous aquifers with regard to hydrothermal systems [23,24]. All samples and readings are documented in the local database of GeoReVi. The sampling strategy aims to simulate pseudo-wells in 3-D space in order to demonstrate the capability of GeoReVi to operate in 3-D environments.

The laboratory results were analyzed by descriptive and multivariate statistics for initial exploratory data analysis. Subsequently, the data sets were interpolated and simulated in 3-D space using inverse distance weighting (IDW), simple kriging (SK), ordinary kriging (OK) and sequential Gaussian simulation (SGS). For the parameter prediction, a 3-D hexahedral mesh with 80,000 cells was generated using an IDW interpolation of a photogrammetric model of the outcrop wall (Fig. 6a). The results were validated through leave-p-out cross-validation (LPO) providing the root-mean-square error (RMSE) and mean-absolute error (MAE). Each of those steps was performed with GeoReVi using the incorporated data mining and geostatistical algorithms and the interactive visualization capabilities.

3.1.1. Results
According to the classification provided by [21], the coefficient of variation c_v and the Dykstra–Parson coefficient c_{dp} indicate a very low heterogeneity of the sandstone formation (Table 1). However, the grade of heterogeneity varies among the considered rock properties. Intrinsic permeability owns a 100-times higher heterogeneity than grain density, which represents the most homogeneous rock property in turn. This is due to a homogeneous mineralogy of the sandstone. At the same time, secondary porosity, produced by feldspar dissolution [25], and non-pervasive
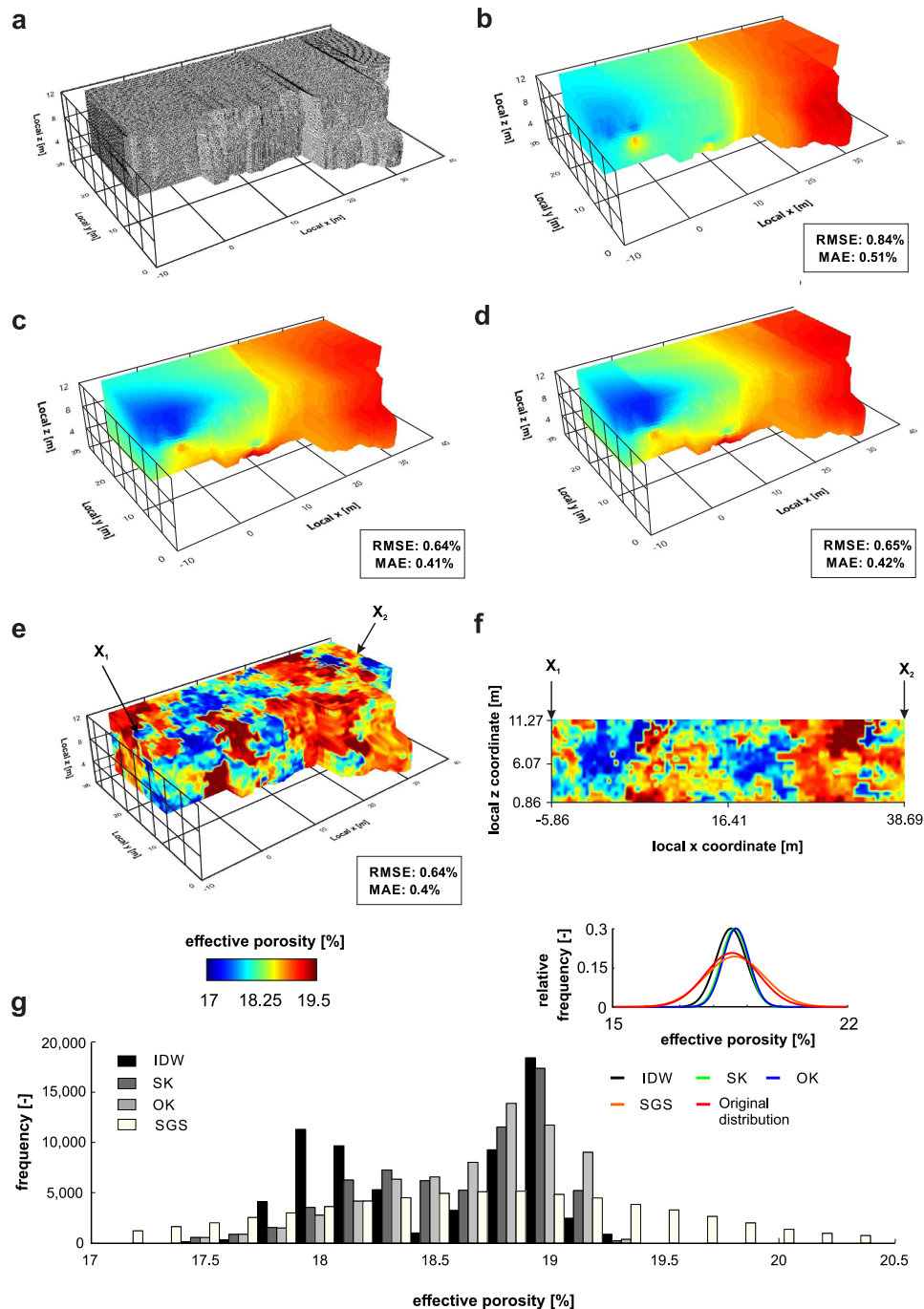
Fig. 6. (a) Hexahedral mesh generated by mapping the photogrammetric outcrop model. The model consists of 80,000 cells and has a volume of 8999 m³. The mesh is used as a target for the interpolation and simulation processes. Comparison of IDW (b), SK (c) and OK (d) interpolations and SGS (e) simulation on the porosity data set of the quarry Obersulzbach. (f) Vertical cross section through the SGS realization (e) from point X_1 to X_2. (g) Empirical and theoretical distributions of the realizations in comparison to the original theoretical distribution.

cementation lead to a heterogeneous network of both closed and enlarged pore throats, strongly increasing the spatial variability of the permeability.

The cluster analysis (Fig. 5d) highlights the differences in the rock properties across the footwall and the hanging wall. Cluster one represents the samples east of the fault zone and cluster two the samples west of it. Five samples were categorized incorrectly with regard to their positions relative to the fault zone. Those samples, however, are located close to it and hence could be affected by tectonic overprint. Moreover, results from cluster analysis correspond well with the principal component analysis

(PCA) where the categories derived from the cluster analysis are projected onto different regions in the biplot.

It can be seen that effective porosity shows a bimodal distribution between 17 and 20% (Fig. 5b). The effective porosity in the eastern part of the quarry is slightly higher than in the western part (Fig. 5a). The SK, OK and SGS realizations provide a low RMSE of 0.64, 0.65 and 0.64% respectively. The RMSE of the IDW realization (0.84%) is higher. Compared to the original histogram (Fig. 5b), IDW, SK and OK underestimate the original range whereas SGS reproduces the range appropriately (Fig. 6g). Contrary to the cross-validation results, IDW reflects the bimodal

porosity distribution and hence the two-fold compartmentalization of the rock volume, which is exposed in the outcrop, more accurately than the other algorithms (Fig. 6b-e). SGS reproduces the observed variability best, which is also indicated by the sharp physical contrasts observable in the 2-D cross section which is taken from the SGS realization (Fig. 6f).

4. Impact

To the best of our knowledge, GeoReVi is the first open-source software system that incorporates the entire KDD process for subsurface characterization into one extensible application. The software system can be applied to address a wide variety of geoscientific research questions related to subsurface characterization — however, also general geoscientific problems can be addressed. This distinguishes GeoReVi from existing software solutions which are often tailored to meet the needs of specific disciplines such as hydrocarbon extraction or heat production. The modular architecture makes GeoReVi easily extendable for other researchers and the broad range of data mining algorithms and conventional geostatistics opens up new paths to go.

The system was tested in a series of outcrop analog studies [26] from both petroleum and geothermal research projects. GeoReVi allows researchers to produce optimized spatial property models through rigorous cross-validation and visual inspection of the results. Academic researchers can use GeoReVi as an integrated data repository, analytical platform and visualization system in the context of subsurface characterization. Thanks to the ability to handle local spreadsheet files, GeoReVi is not only limited to the data model provided by its database. Various types of spatial problems can be addressed by the generic yet simple spatial representation of geoscientific data sets.

5. Conclusions

GeoReVi constitutes an integrated software system that facilitates reservoir engineers, geoscientists, petrophysicists and other researchers to largely automate the subsurface-related data management and knowledge discovery process. The generic knowledge discovery potential of GeoReVi comprises statistical and spatial relationships among any kind of rock properties, optimized spatial predictions at any scale of subsurface investigations, uncertainty estimations and the discovery of multidimensional patterns in relational data sets. The normalized data scheme of GeoReVi makes the software robust to changes in the domain knowledge of subsurface characterization. Semi-automated preprocessing increases robustness of the data mining algorithms regarding sparse, erroneous or multidimensional data sets.

GeoReVi is able to overcome problems from existing open-source software packages related to geostatistics or geological modeling [27] such as the limited applicability of algorithms to 2-D features or regular lattices, restricted expert interaction, single-user environments and data storage limitations. Moreover, expensive commercial software packages – suitable to address those issues – are usually employed as black-box tools. With our work, we aim to contribute to the ongoing development of open-source, intelligent, automated data analysis systems by providing an intuitive and extensible geoscientific data management and analysis tool. Ongoing research will focus on the development of plug ins for finite element simulation of subsurface heat transfer and fluid flow.

CRediT authorship contribution statement

Adrian Linsel: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper, Developed GeoReVi. **Kristian Bär:** Conceived and designed the analysis. **Joshua Haas:** Collected the data, Contributed data or analysis tools, Performed the analysis. **Jens Hornung:** Conceived and designed the analysis. **Matthias D. Greb:** Conceived and designed the analysis. **Matthias Hinderer:** Conceived and designed the analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Sample availability

Samples are available at the Institute of Applied Geosciences in Darmstadt, Germany. Moreover, they are registered in the System for Earth Sample Registration (SESAR, www.geosamples.org) with the code names provided in Table A.1.

Table A.1
Rock samples taken from the outcrop in Obersulzbach with the associated International Geo Sample Number (IGSN).

Sample	IGSN	Sample	IGSN
OSB1_1	IEDAL0046	OSB25	IEDAL000P
OSB1_2	IEDAL0045	OSB26	IEDAL000O
OSB6	IEDAL0004	OSB28	IEDAL000M
OSB7	IEDAL0003	OSB29	IEDAL000L
OSB8	IEDAL0002	OSB30	IEDAL000K
OSB9	IEDAL0001	OSB31	IEDAL000J
OSB10	IEDAL001W	OSB32	IEDAL000I
OSB11	IEDAL001V	OSB33	IEDAL000H
OSB12	IEDAL001U	OSB34	IEDAL000G
OSB13	IEDAL001T	OSB35	IEDAL000F
OSB14	IEDAL001S	OSB36	IEDAL000E
OSB15	IEDAL001R	OSB37	IEDAL000D
OSB16	IEDAL001Q	OSB38	IEDAL000C
OSB17	IEDAL001P	OSB39	IEDAL000B
OSB19	IEDAL001O	OSB40	IEDAL0008
OSB21	IEDAL000S	OSB41	IEDAL0007
OSB22	IEDAL000R	OSB42	IEDAL0006
OSB23	IEDAL000Q	OSB43	IEDAL0005

Appendix B. Data availability

Data is available in the online repository and provided as a training data set for the GeoReVi tutorial. Also, the data is partly integrated into the local database that is shipped with the executable file.

References

[1] Fanchi JR. Fundamentals of reservoir characterization. In: Fanchi JR, editor. Shared Earth Modeling. Woburn: Butterworth-Heinemann; 2002, p. 170–81. http://dx.doi.org/10.1016/B978-075067522-2/50010-0.

[2] Brachman RJ, Anand T. The process of knowledge discovery in a first sketch. Report, 1994.

[3] Fayyad U, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the second international conference on knowledge discovery and data mining. AAAI Press; 1996, p. 82–8, URL http://dl.acm.org/citation.cfm?id=3001460.3001477.

[4] Jia A, He D, Jia C. Advances and challenges of reservoir characterization: A review of the current state-of-the-art. In: Dar IA, editor. Earth Sciences. Rijeka: IntechOpen; 2012, http://dx.doi.org/10.5772/26404.

[5] Ge JK, Chen Z. Constructing ontology-based petroleum exploration database for knowledge discovery. Appl Mech Mater 2010;20–23:975–80. http://dx.doi.org/10.4028/www.scientific.net/AMM.20-23.975.

[6] Popov YA, Pribnow DFC, Sass JH, Williams CF, Burkhardt H. Characterization of rock thermal conductivity by high-resolution optical scanning. Geothermics 1999;28(2):253–76. http://dx.doi.org/10.1016/S0375-6505(99)00007-3.

[7] Clavaud J-B, Maineult A, Zamora M, Rasolofosaon P, Schlitter C. Permeability anisotropy and its relations with porous medium structure. J Geophys Res: Solid Earth 2008;113:1–10. http://dx.doi.org/10.1029/2007JB005004.

[8] Dürrast H. Ultrasonic laboratory measurements on spherical samples — a tool for the investigation of seismic anisotropy in rocks. In: Arnold W, Hirsekorn S, editors. Acoustical imaging. Dordrecht: Springer Netherlands; 2004, p. 383–90. http://dx.doi.org/10.1007/978-1-4020-2402-3_48.

[9] Kohlhammer J, Keim D, Pohl M, Santucci G, Andrienko G. Solving problems with visual analytics. Procedia Comput Sci 2011;7:117–20. http://dx.doi.org/10.1016/j.procs.2011.12.035.

[10] Maimon O, Rokach L. Introduction to knowledge discovery and data mining. In: Data mining and knowledge discovery handbook. Springer Science + Business Media; 2010, p. 1–15. http://dx.doi.org/10.1007/978-0-387-09823-4_1.

[11] Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci 2007;33(2):163–80. http://dx.doi.org/10.1177/0165551506070706.

[12] Zeleny M. Management support systems: Towards integrated knowledge management. Hum Syst Manage 1987;7(1):59–70. http://dx.doi.org/10.3233/HSM-1987-7108.

[13] Ackoff RL. From data to wisdom. J Appl Syst Anal 1989;16:3–9.

[14] Sacha D, Stoffel A, Stoffel F, Kwon BC, Ellis G, Keim DA. Knowledge generation model for visual analytics. IEEE Trans Vis Comput Graphics 2014;20(12):1604–13. http://dx.doi.org/10.1109/TVCG.2014.2346481.

[15] Bär K, Sippel J, Strom A, Mielke P, Sass I. P^3 - International petrophysical property database. In: Proceedings, 42nd workshop on geothermal reservoir engineering. 2017, p. 1–6, URL https://gfzpublic.gfz-potsdam.de/pubman/item/item_2042667.

[16] Codd EF. A relational model of data for large shared data banks. Commun ACM 1970;13(6):377–87. http://dx.doi.org/10.1145/362384.362685.

[17] Liolios P, Exadaktylos G. A relational rock mechanics database scheme with a hierarchical structure. Comput Geosci 2011;37(8):1192–204. http://dx.doi.org/10.1016/j.cageo.2011.02.014.

[18] Gard M, Hasterok D, Halpin J. Global whole-rock geochemical database compilation. Earth Syst Sci Data Discuss 2019;2019:1–23. http://dx.doi.org/10.5194/essd-2019-50.

[19] Becker A, Schwarz M, Schäfer A. Lithostratigraphische Korrelation des Rotliegend im östlichen Saar-Nahe-Becken. Jber Mitt Oberrhein Geol Ver 2012;94:105–33. http://dx.doi.org/10.1127/jmogv/94/2012/105.

[20] Aretz A, Bär K, Götz AE, Sass I. Outcrop analogue study of permocarboniferous geothermal sandstone reservoir formations (northern upper rhine graben, Germany): impact of mineral content, depositional environment and diagenesis on petrophysical properties. Int J Earth Sci 2015;105(5):1431–52. http://dx.doi.org/10.1007/s00531-015-1263-2.

[21] Ringrose P, Bentley M. Reservoir model design. 1st ed. Springer Netherlands; 2015, p. 249. http://dx.doi.org/10.1007/978-94-007-5497-3.

[22] Farkhutdinov A, Goblet P, de Fouquet C, Cherkasov S. A case study of the modeling of a hydrothermal reservoir: Khankala deposit of geothermal waters. Geothermics 2016;59:56–66. http://dx.doi.org/10.1016/j.geothermics.2015.10.005.

[23] Bär K, Arndt D, Fritsche J-G, Götz AE, Kracht M, Hoppe A, et al. 3D-Modellierung der tiefengeothermischen Potenziale von Hessen – Eingangsdaten und Potenzialausweisung. Z dt Ges Geowiss 2011;162(4):371–88. http://dx.doi.org/10.1127/1860-1804/2011/0162-0371.

[24] Agemar T, Weber J, Schulz R. Deep geothermal energy production in Germany. Energies 2014;7(7):4397–416. http://dx.doi.org/10.3390/en7074397.

[25] Molenaar N, Felder M, Bär K, Götz AE. What classic greywacke (litharenite) can reveal about feldspar diagenesis: An example from permian rotliegend sandstone in hessen, Germany. Sediment Geol 2015;326:79–93. http://dx.doi.org/10.1016/j.sedgeo.2015.07.002.

[26] Hornung J, Linsel A, Schröder D, Gumbert J, Ölmez J, Scheid M, et al. Understanding small-scale petrophysical heterogeneities in sedimentary rocks: the key to unravelling pore geometry variations and to predicting lithofacies-dependent reservoir properties. In: Digital Geology – Multi-scale analysis of depositional systems and their subsurface modelling workflows, EAGE Special Volume. 2020.

[27] Goovaerts P. Geostatistical software. In: Fischer MM, Getis A, editors. Handbook of applied spatial analysis: Software tools, methods and applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010, p. 125–34. http://dx.doi.org/10.1007/978-3-642-03647-7_8.

3.2. Publication II – Understanding Small-Scale Petrophysical Heterogeneities in Sedimentary Rocks

3.2.1. Introduction

Figure 3.2.: Multi-scale reservoir characterization concept as applied in the Shell Digital Geology project (from Kloosterman et al. (2017)).

Publication II presents one sub-project of the Shell Digital Geology project which aims at demonstrating how the ongoing digitization can be used to support experts in the process chain of subsurface characterization. Our task within that project involved a comprehensive petrophysical laboratory campaign conducted on seven rock slabs from seven different depositional environments.

The software application GeoReVi (**Publication I**) was used to manage the data sets with ≥10,000 readings of eight representative rock properties derived from 881 rock plugs. Furthermore, GeoReVi was used to perform the statistical analyses and the geostatistical interpolations.

The results of the study were used by Shell for building up their Digital Geology training center. Our study represents the **Grain Scale** (Fig. 3.2) which constitutes one of the smallest scales of investigation in a multi-scale reservoir characterization process chain.

3.2.2. Author Contributions

- The Shell project *Digital Geology* started in 2014 as an independent laboratory campaign coordinated by J. Hornung and M.C. Pöppelreiter. Within two master theses and two bachelor theses, which were conducted by D. Schröder, J. Gumbert, J. Ölmez and M. Scheid, respectively, over 10,000 readings of petrophysical properties were produced on nearly 900 samples.

- The readings were initially managed in an excel spreadsheet until GeoReVi had been used as a data management tool from 2018 onwards. J. Hornung elaborated the scientific concept and organized the sampling and measurement campaign.

- As a co-author, I imported the data sets into GeoReVi, produced a huge partition of the analyses, contributed most of the visualizations and improved the initial manuscript. In the process of manuscript preparation I was in steady exchange of ideas with J. Hornung and contributed custom-made implementations of GeoReVi in order to extract knowledge from the extensive data sets and to aid at visualizing the results.

- All co-authors contributed to the preparation and revision of the manuscript.

3.2.3. Publication

DIGITAL
GEOLOGY

Multi-scale analysis of
depositional systems and
their subsurface workflows

Jürgen Grötsch and
Michael Pöppelreiter (eds.)

Understanding Small-Scale Petrophysical Heterogeneities in Sedimentary Rocks: The Key to Unravelling Pore Geometry Variations and to Predicting Lithofacies-Dependent Reservoir Properties

AUTHORS

J. Hornung[1] I A. Linsel[1] I D. Schröder[1] I J. Gumbert[1] I J.A. Ölmez[1] I M. Scheid[1]
M.C. Pöppelreiter[2]

AUTHORS DETAILS

1   Technische Universität Darmstadt, Institut für Angewandte Geowissenschaften, Schnittspahnstrasse 9,
    64287 Darmstadt, Germany, Hornung@geo.tu-darmstadt.de
2   Shell Kuwait Exploration and Production BV, P.O. Box 28487, Safat 13145, Kuwait, M.Poppelreiter@shell.com

ABSTRACT

Quantitative lithofacies characterization and prediction of reservoir properties is challenging on the scale of individual grid blocks (voxels) of geocellular models. To better understand variability of petrophysical properties on this scale, this study links geological features and petrophysical properties based on high-resolution characterization and innovative analysis methods on grain and bed scale.

Samples from two commonly occurring clastic depositional systems were investigated: i) a siliciclastic fluvial channel and ii) a carbonate ramp. Of these 2 depositional systems, a total of 7 subenvironments were sampled.
i) The fluvial channel system is stratigraphically part of the Triassic Lower Bunter Formation in SW-Germany. The formation consists of sandstone interbedded with shale. Three rock slabs, each representing a subenvironment, were investigated: 1. channel base, 2. midchannel bar and 3. marginal sand bar.
ii) The carbonate ramp samples belong stratigraphically to the Triassic Upper Muschelkalk Formation, also in SW-Germany. The formation consists of limestone, dolomite, and marl. Carbonate samples represent the following subenvironments: 4. lagoon, 5. tidal flat, 6. shoal and 7. foreshoal.

Investigated were rock slabs measuring 50x50 cm (sandstone) and 100x30 cm (carbonates). Up to 145 core plugs measuring 2,45x4 cm were drilled out of each slab. Geological properties measured in detail include grain size and sorting as well as sedimentological attributes of individual lithofacies as indicator for hydrodynamic flow conditions during time of deposition.

Pore systems of sediment samples were investigated using thin sections and scanning electron microscope (SEM). Petrophysical properties analysed include effective He-porosity, apparent permeability measured in 3D, intrinsic permeability, ultrasonic p-wave and s-wave velocity, as well as resistivity.

The resulting database contains almost 1,000 samples and over 10,000 measurements. The data were used to construct uni- as well as multivariate geostatistics. These include distribution analyses, experimental variogram and principal component analyses. Data were visualized as scatter-,

bar-, bi- and box-whisker plots to investigate relationships between geological and petrophysical properties. Moreover, petrophysical core plug measurements were superimposed as "bubble plots" on to each slab with its interpreted lithofacies to avoid statistical data bias. At least simple kriging is introduced to spatially interpolate the readings.

Results show that petrophysical properties show large variability between slabs (= dm to m scale) that represent distinct subenvironments. However, a large variability is also observed on the scale of individual plug measurements (= cm scale) within each slab composed of distinct lithofacies.

Spatial heterogeneities do not exclusively coincide with depositional surfaces (bedding planes, erosive surfaces) but with the textural framework. Both reflect changes in energy conditions during time of deposition. Hence, the definition of a depositional subenvironment of a rock slab (grid block) or a lithofacies type for similar beds may only partially capture the heterogeneities observed. In general, it is crucial to map major bounding surfaces as well as trends within them to serve as proxy for hydrodynamic energy during deposition. With that information, it is possible to partition rocks into areas of similar petrophysical properties and better understand variability as an input to enhance interpolation at grid block scale.

INTRODUCTION

Heterogeneities and rock properties at millimetre to decimetre scale (grain scale) play a significant role for fluid flow dynamics on a grid block scale (Weber, 1986). Centimetre-scale permeability barriers for example reduce the overall permeability of a reservoir, if fluids pass them towards a well bore for example. The lowest permeability of a reservoir rock represents the overall permeability (Jennings and Lucia, 2003). This effect is less significant when reservoir fluids flow parallel to permeability barriers. Hence, it is crucial to treat permeability variations and their anisotropy as a vector property in geological models (Jennings and Lucia, 2003). In this study, we address grain scale heterogeneities in terms of permeability and porosity and relate those rock properties to specific pore systems. Other petrophysical properties such as resistivity and sonic wave velocity were measured at these rock samples, data are shown in this paper as well, but they are not discussed further.

Investigations show that the distribution of petrophysical properties largely mimics heterogeneities due to depositional processes. Thus, petrophysical properties can best be represented in digital models by modelling them co-located with depositional geometries.

This study was part of Shell's Digital Geology Project (see this book) where it represents "grain scale" heterogeneities. This paper presents the most significant data and results of the project. Geological attributes of the rock slabs investigated and the corresponding petrophysical measurements are shown at the Digital Geology training centre mounted in the Shell Headquarter in The Hague, the Netherlands.

DATABASE

Rock slabs of up to 1 m height are used to capture petrophysical heterogeneities at grain and bed scale and obtain results that are representative for specific facies types with statistical significance. Rock slabs of common depositional environments were investigated in terms of lateral and vertical changes of petrophysical properties and their dependency to pore system and sedimentary features. In total, seven rock slabs were selected to cover two commonly occurring depositional environments that frequently contain hydrocarbons:

i) Fluvial stream depositional system
ii) Carbonate ramp depositional system

i) The clastic fluvial deposits investigated belong to the Lower Bunter Sandstone Formation. The rock slabs are derived from quarries in SW-Germany. The slabs represent typical depositional subenvironments as illustrated in the facies model by Einsele (2000). The three slabs represent different parts of a fluvial channel belt (Figure 1) and thus reflect different levels of depositional energy:
• Slab 1: sandstone with mud clasts representing channel base deposits
• Slab 2: sandstone, cross-bedded representing transverse sand bar deposits
• Slab 3: sandstone, ripple bedded to massive, representing channel margin deposits

The depositional system sampled extends for tens of kilometres, much beyond the size of quarries that are hundreds of metres wide. Hence, rock slabs were cut out in different quarries that are all located in the same stratigraphic unit to obtain rock slabs of genetically related depositional elements. The slabs are 50 cm in height (vertical direction or z-axis) and sample the full range of sedimentary structures commonly observed in this depositional system. The width (lateral direction or y-axis) of all rock slabs is 50 cm as well. The y-axis of the rock slabs is oriented in dip direction of the depositional system, that is, the flow direction of the channel system. This assures the slabs sample the area of expected maximum heterogeneity as foresets cross it. As plugs are drilled
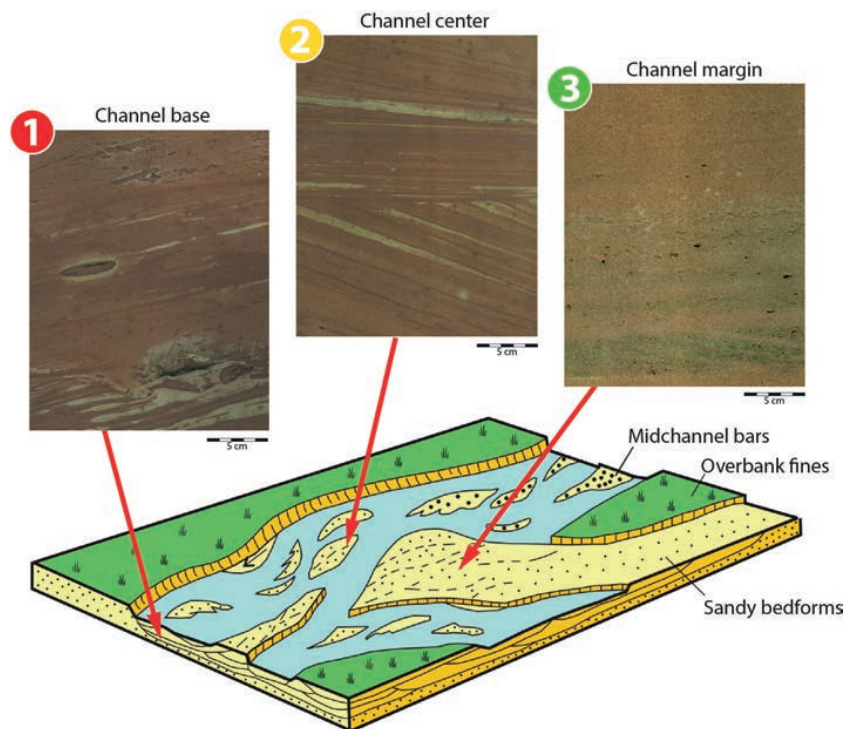
along the x-axis, most of them do show parallel foresets in this direction. Hence, in x- and z-direction, foresets cross the plug. This is related to stream direction as the x-direction represents the strike of the channel system and corresponds to the thickness of the slabs (6 cm). So one can assume a lower degree of heterogeneity in that direction, but this direction cannot be investigated accordingly due to its small extend.

ii) The carbonate samples were selected based on the paleogeographic orientation suggested by the carbonate ramp model proposed by Aigner (1985). The slabs are derived from common and representative ramp subenvironments (Figure 2) representing the facies model of Einsele (2000).

Samples of four typical and commonly occurring subenvironments were selected:
• Slab 4: dolo-mudstone, representing lagoon
• Slab 5: lime-mud- to wackestone, cross-bedded, representing tidal flat
• Slab 6: lime-grainstone, cross-bedded, oolitic representing shoal
• Slab 7: lime-packstone, graded, bioclastic representing foreshoal

The depositional subenvironments extend for several tens of kilometres. Hence, samples were extracted from several quarries requiring also change of the stratigraphic level.

Rock slabs measure 100 cm by 30 cm. The long side (100 cm) represents the vertical direction (z-axis). It samples many different geological layers. The small (30 cm) side (y-axis) was selected along the direction where most heterogeneities can be observed. The thickness (6 cm) of the slab represents the x-axis. All seven rock slabs are cut out in active quarries, from un-weathered surfaces in SW-Germany. The sample locations/formation/quarry owners of all slabs are listed below.

i) Fluvial channel system

All slabs belong to the Lower Bunter Sandstone Formation of Lower Triassic age. Stratigraphically, they belong to the uppermost part of the Calvörde Member and the lowermost Bernburg Member of the Lower Bunter Formation. Locally, the unit is referred to by its trade name as Miltenberger Sandstein Formation. The international stratigraphic code is 'SuM'. In detail, samples come from the following:
• Slab 1: quarry of the company Wassum at 63897 Miltenberg, Germany
  49.7135071°N, 9.2549472°E
• Slab 2: quarry of the company Wassum, 63897 Miltenberg, Germany
  49.7135071°N, 9.2549472°E
• Slab 3: quarry of the company Zeller, 63930 Umpfenbach, Germany
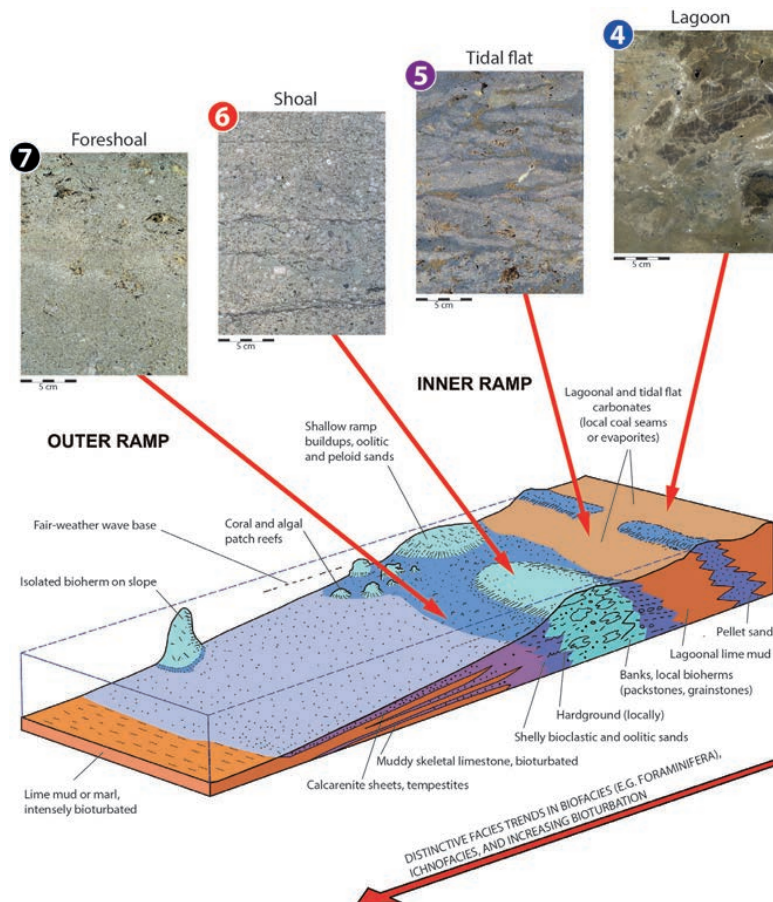 49.689447°N, 9.368433°E

ii) Carbonate ramp system

- Slab 4: Upper Jurassic, Middle Kimmeridge (Malm-Delta), stratigraphic code "ki 2, d, ri+b", trade name Pfraundorfer Dolomit, quarry near Pfraundorf, 49.0034532°N, 11.4284639°E, operating company Zeidler & Wimmel, 97268 Kirchheim/Franken, Germany.
- Slab 5: Upper Muschelkalk-Formation, Trochiten-kalk-Formation, stratigraphic code "moT", trade name Mooser quarry near Geroldshausen-Moos, 49.673615°N, 9.872273°E, operating company Borst, 97268 Kirchheim/Franken, Germany.
- Slab 6: Upper Muschelkalk-Formation, Trochiten-kalk-Formation, stratigraphic code "moT", trade name Crailsheimer Muschelkalk, quarry near Neiden-fels-Kernmühle, 49.1759234°N, 10.0458996°E, operating company Schön & Hippelein, 74589 Crailsheim, Germany.
- Slab 7: Upper Muschelkalk-Formation, Trochiten-kalk-Formation, stratigraphic code "moT"; trade name Kirchheimer Muschelkalk, quarry near Kirchheim/Unterfranken, 49.656497°N, 9.857227°E, operating company Hofmann-Naturstein, 97956 Werbach, Germany.

METHODOLOGY

Nondestructive measurements were performed on both sides of each slab. These measurements included permeability using mini-permeameter technique with air, ultrasonic p- and s-wave velocity, and resistivity. These data are not discussed in this paper. However, they are used for quality control of core plug measurements. Core plugs are drilled out of one side of a slab, while the other side remained untouched as 'display face' for training purposes at the Shell exhibition.

A grid of sample points was laid out covering each slab to indicate where core plugs (diameter 25 mm) were to be retrieved. The grid has a spacing of roughly 50 mm and is aligned to sedimentary structures and other heterogeneities to get different distances for statistical analyses and to avoid having completely different lithofacies types in a plug. Each slab contains up to 145 sampling points. An additional 32 plugs with only 10 mm plug diameter were extracted from small regions of two slabs to check upscaling issues. Core plugs were then drilled out with a

diamond-bearing crown penetrating 40 mm deep into the slab. Penetration time of the drilling crown into the rock was measured under standard conditions mimicking the rate of penetration (ROP) of a drill bit. We do not introduce these data here.

This paper provides a summary of some basic information on the used plug-measurement techniques. Extended descriptions of the project can be found in Schröder (2015), Gumbert (2015), Ölmez (2018) and Scheid (2016). A detailed assessment of data quality, accuracy and reproducibility of gas permeameter readings are provided by Ölmez (2018), Scheid (2016) and Filomena et al. (2014).

Permeability

In principle, two different types of permeabilities exist: the 'intrinsic' permeability and the 'apparent' permeability. The 'intrinsic' permeability is considered a materials property, which is independent of the used fluid and conditions of measurement (Klinkenberg, 1941). It is commonly determined in a cylinder-cased probe (e.g., Hassler cell) using multiple measurements with different pressure steps. The 'apparent' permeability is a single reading where conditions of measurement and fluid properties cannot be cancelled out. Therefore, it is a relative measure that can be compared only to measurements under the same conditions (Filomena et al., 2014). Usually, it is performed by surface injections with an air probe. The method is commonly referred to as "mini-permeability" or "relative permeability".

In this study, an air-driven permeameter (Hornung and Aigner, 2004; Filomena et al., 2014) was used to establish permeability. This permeameter determines the intrinsic permeability as well as apparent permeability. The tool has a measurement range of intrinsic permeability from 0.5 µD to 70 D, if the size of the core is varying. Plugs used in this study had a diameter of 25 mm and are 20 mm to 40 mm in length. Permeability measured at these plugs ranged from 13 µD to 2.5 D. To realize also an exceptional low range of apparent permeability, an injection tip is used, which is adapted to the shape of the plug surface. It is sealed with soft neoprene foam owning a closed cell structure. In tests with an impermeable plug (metal, plastic), leakage was beyond the range of the instrument.

Measurements in 3D are possible using the surface injection probe (apparent permeability) on different sides of a plug sample. Surface injection probe measurements have an injection diameter of 4 mm, integrating a rock volume of just a few cm³. Considering the strong heterogeneity at grain scale, apparent permeability is determined as an arithmetic average of six measurements for each direction (x, y, and z, respectively). This procedure is implemented to minimize variability effects on spatial statistical data

when comparing apparent (surface probe) permeability to intrinsic (cylinder probe) permeability, which integrates over the whole plug volume (Jennings, 2000).

Two other known problems of surface-probe measurements (apparent permeability) had to be addressed:
(1) The different flow directions measured with the surface probe permeameter (apparent permeability) are not fully comparable when using plugs. Measurements in x-direction use the cut-sections of the plug (front and back), which have a flat surface. Measurements with flow directions in y- and z-direction use the curved circumferential surface. Hence, geometry effects of the gas flow may apply such as smaller travel distances and penetrated volume when using the circumferential surface of the plug. This is mainly due to the preferential direction of the sediment structures. Please see details in the anisotropy section later in this paper. To make these analyses, fully comparable drillings in each direction or better cubes should be used in future studies.

Porosity

Porosity was calculated as percentage of pore volume to plug volume. Therefore, the volume of the grains and the envelope volume of the plug are required to establish the difference, which is the pore volume. The grain volume was determined using a helium pycnometer (AccuPyc 1330™) from the company Micromeritics. The envelope volume was calculated based on plug dimensions and by using a powder pycnometer (GeoPyc 1360™), also manufactured by Micromeritics. Caution is required when measuring carbonate samples that contain macroscopically visible-pores. Macroscopic pores (vugs) are not detected by the powder pycnometer. The powder invaded the pores causing erroneous values. As plugs were carefully cut resulting in a regular outer surface, calculating the envelope volume provided better results as compared to the powder-pycnometer method for vuggy carbonates.

Sonic velocity

Ultrasonic velocity (p- and s-wave) was determined along the long axis of the plug. This corresponds to the x-axis (thickness) of the slab. Travel times were picked manually from a time-amplitude diagram. Velocity was calculated using the length of the plug as distance. The device used (USG-40™) was manufactured by the company Geotron.

Resistivity

Resistivity measurements strongly depend on the degree of saturation. Hence, samples for lab measurements require being water saturated as prescribed by the

Archie equation (Archie, 1952). Therefore, plugs were 100% saturated with ion- and gas-free water (lab quality, conductivity <4 µS) under high vacuum. Complete water saturation of plug was established at least for 30 min before measurement took place. Tests reveal that readings showed a high degree of repeatability after this time interval (Boansi, 2015).

With this set-up, neither a qualitative nor a quantitative comparison of measurements with in-situ subsurface samples (e.g., logs) is possible. First, pore fluid properties differ significantly at reservoir conditions. Secondly, saturation is never 100% in reservoirs. In the subsurface, saturation shows relationships between pore size distributions based on the gravity-capillary equilibrium. In turn, in-situ resistivity can be used to estimate porosity and permeability together with an indicator of the rock fabric (Jennings & Lucia, 2003). From the outset, this is excluded with our data. Our results are a proxy for pore networks and petrological composition (rock types) only. In a qualitative way, relative differences among rock types however can be investigated with this experimental set-up.

Statistical analyses

Statistical analyses of data and data visualization were performed with the software tool GeoReVi (Geothermal Reservoir Virtualization) developed at the Technical University of Darmstadt (Linsel et al., 2018). The user-friendly software can generate adaptable stacked histograms, variograms and bubble plots as well as principal component analysis (PCA) and interpolation semi-automatically. It is based on the Knowledge-Discovery in Databases (KDD) process using an in-house database of the Technical University of Darmstadt, Germany. Bulk data sets can be uploaded automatically from the measuring devices or from datasheets in csv format.

RESULTS

Fluvial channel system

Lithofacies description
The three slabs of the fluvial channel system comprise four principal lithofacies types (Figure 3). Petrographically, all are quartz arenites (Pettijohn, 1987) containing approximately 85–95% of quartz, 1–5% clay minerals, 1–3% of mica and 3–7% of feldspar. All lithofacies types show early diagenetic hematite cements coating grains. These confer the reddish colours to the sandstone, a typical name-giving feature of the Bunter Formation. Lithofacies codes were used to differentiate samples based on Miall (2012), Wentworth (1922) and Hornung & Hinderer (2011).

Sandstone, trough cross-bedded (code St)
This lithofacies type is characterized by trough-shaped (tangential) foresets of 5–30 mm thickness and dipping 5–30° in a downstream direction. Sorting is poor to moderate. Each individual foreset shows a fining-up trend towards its top, which makes stratification recognizable. Grain sizes are in the range of 350–500 µm at the base of the foresets and 64–125 µm at their uppermost limit. A lateral fining upward trend exists from the lower to the upper positions of a foreset. A set of stacked individual foresets result in an overall fining upward trend of that lithofacies type (350–500 µm to 177–250 µm).

The base of these bedforms is curved, indicating trough-shaped foresets. Their height must have been upward of 30 cm. Thus, they classify as small dunes (Reineck and Singh, 1980). Using the relationship of the mean grain size, the height and the shape of the dune, it is possible to determine stream velocities during deposition. The deposits were formed in the upper flow regime, with flow velocities between 90 and 110 cm/s (Reineck and Singh, 1980; Harms and Fahnestock, 1965) assuming >100 cm channel depth and normal meteoric water.

Sandstone, planar cross-bedded (code Sp)
This lithofacies type shows the same dimensions and appearance as St with three exceptions: Grain size is from 177 to 350 µm, sorting ranges from well to very well, and the shape of the foresets is planar. This implies that the crest of the ripple is linear and continuous. According to the flume tank experiments (Reineck and Singh, 1980), these megaripples formed under stream velocities around 40–60 cm/s, which is significantly lower than the stream power of facies St.

Sandstone, ripple cross-bedded (code Sr)
The ripples in this facies show the same features as facies Sp and St, but they are much smaller (<10 cm set thickness). Foresets show convex, planar and trough-shaped, nonerosive boundaries. Sorting is poor to moderate and grain sizes range from 177 to 500 µm. These attributes suggest that they were formed in the lower flow regime ranging from 15 to 40 cm/s (Reineck and Singh, 1980).

Sandstone massive (code Sm)
The term 'massive' is used for sediments without visible sedimentary structures. The grain size of such samples is lower than in the other lithofacies types and ranges from 125 to 177 µm. In some cases, samples show weak fining or coarsening upward trends, considering that the grain-size sediment likely formed in the lower flow regime, starting immediately with the transport at 15 cm/s stream power. As there are no sedimentary structures, this means that there was no time to form them. In conclusion, sediment was likely mobilized and deposited suddenly, for example, moving as a grain carpet with a flash flush.

Hence, hydrodynamic conditions can be considered as highly unsteady with variable stream power.

A comparison of the three slabs shows that St and Sp dominate at the channel base (slab 1) or channel centre. Sp dominates the sand bar deposits within a channel (slab 2). Sm and Sr dominate the higher sand bars/channel margins (slab 3). Hence, in terms of stream power, the samples represent a clear trend of decreasing depositional energy and towards intermittent flow. These mimic a typical fluvial cycle in a braided fluvial channel representing an upward shift from channel base to overbanks or margin (Figures 1 and 3).

Multiparametric statistical analysis

Statistical methods lack a spatial connotation. This is a disadvantage when sedimentary rocks, products of di-
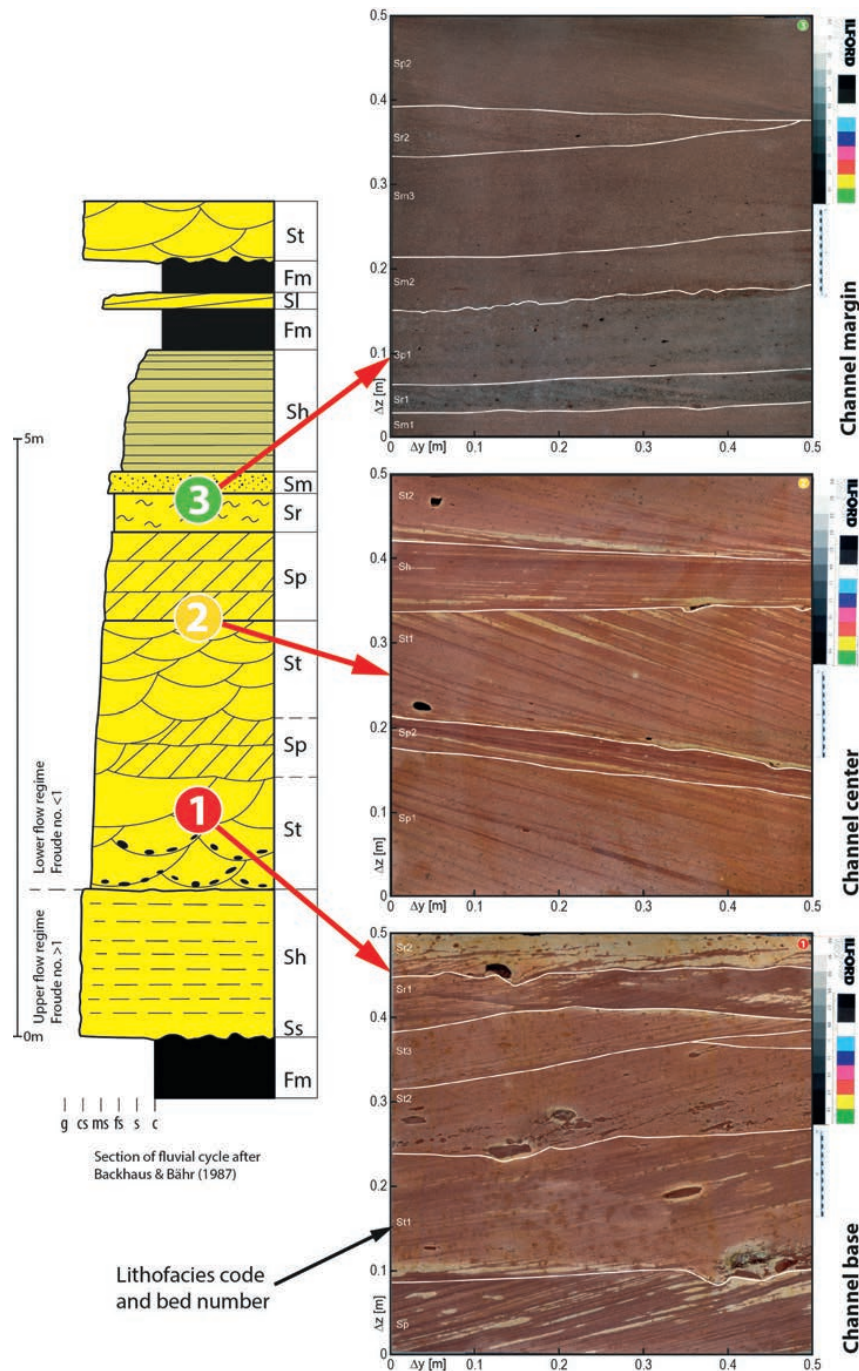
FIGURE 3

Position of the slabs with their lithofacies in a typical braided river succession of the Bunter Sandstone (after Backhaus and Bähr, 1987). Slabs (=subenvironments) show a decrease in depositional energy in higher positions in the fluvial cycle. Sl = Sand low angle cross-bedded, Sp = Sand planar cross-bedded, St = Sand trough cross-bedded, Sr = Sand ripple cross-bedded, Ss = Sand over scoured base (lag deposits), Fm = Fines massive, and Sm = Sand massive.

rectional depositional processes and their petrophysical response, are concerned. This issue is addressed using bubble plots, which relate readings to their original position on the sample (Figure 4a-c). It was realized by using different colours representing the readings. Plots clearly indicated that measurements show a relationship to subenvironments and lithofacies types. In particular, porosity, permeability and resistivity differences correspond well with lithofacies boundaries, boundaries of individual layers or foresets in vertical and horizontal direction. These trends correspond with changes in sediment composition, grain size and sorting. Accordingly, they are related to systematic changes in rock texture due to varying hydrodynamic conditions. In contrast, acoustic velocity (Vp/vs) shows a weak or no dependency to lithofacies. They show a weak dependence to depositional subenvironment.

Principal component analysis (PCA) was used on original readings comprising all measured properties for a comparison of petrophysical and geological relationships of each slab (Figure 5a-c). The two most important variables (components on x- and y-axes) are compared to establish their contribution to variability in an n-dimensional data set. If the lines (biplot) of properties point into one of these 2 axes, this property suggests a strong primary or secondary dependency. Property lines showing the same direction are positively correlated. Property lines pointing in oppo-
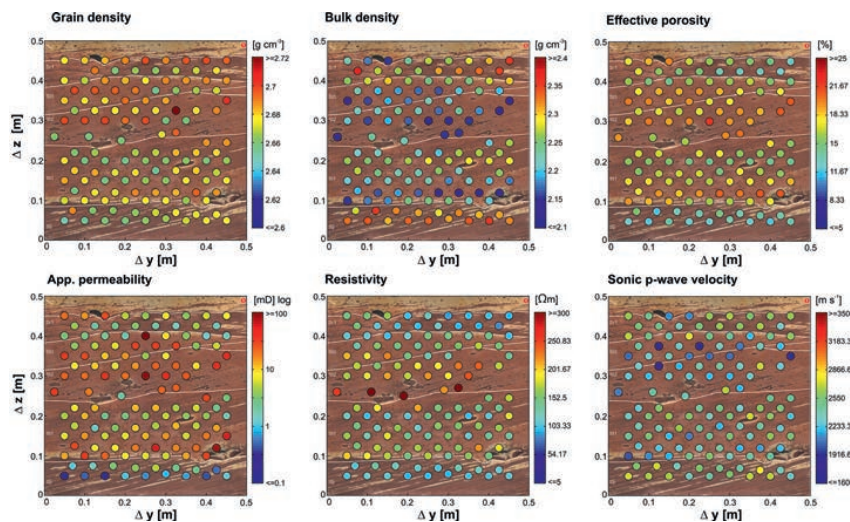
FIGURE 4A

Bubble plots are used to display petrophysical data accurately at their spatial position, in order to relate them to the corresponding lithofacies. Slab 1 (channel base), lithofacies show differences in petrophysical properties and their distribution (except p-wave). Note that specific trends occur within a lithofacies, individual bed and even along a foreset. Systematic trends are beyond the resolution of any statistical method and can be recognized only with these kinds of plots. p-wave velocity readings seem to be independent from lithofacies. For interpretation, see text.
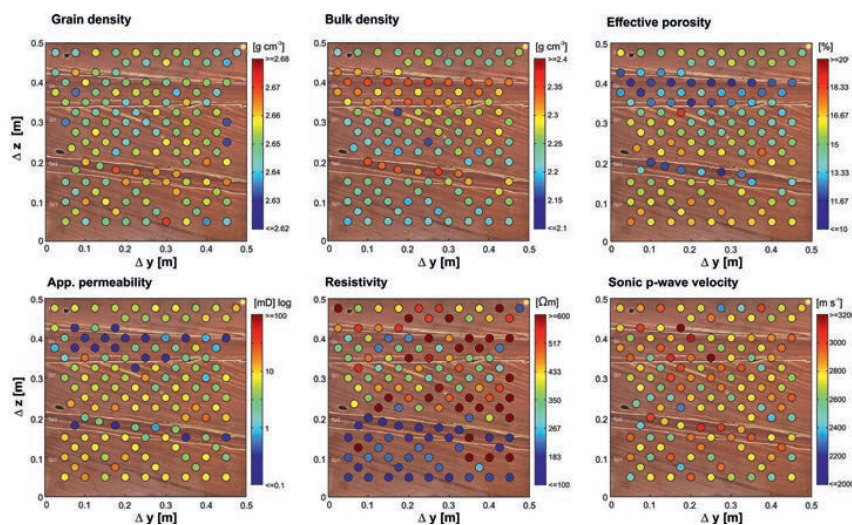
FIGURE 4B

Petrophysical readings of slab 2 (channel centre/midchannel bar sands) reveal systematic trends as described in Figure 4a.

site directions are reversely (negatively) correlated. If they show a 90° angle, they are independent from each other. Property lines explain the spreading of the data set in their specific directions.

All three slabs show that grain size and sorting are negatively correlated. Larger grain sizes are poorly sorted than finer grain sizes. Porosity and permeability show a strong positive correlation. In contrast, some properties show variability throughout subenvironments:

Ultrasonic speed (Vp/Vs) is negatively correlated with porosity and permeability in slab 1 (Figure 5a, channel base) and slab 3 (Figure 5c, channel margin). However, it is uncorrelated to all other properties in slab 2 (Figure 5b, channel centre). All other properties measured at slab 2 are exceptionally well correlated to each other.

FIGURE 5B

PCA of slab 2 (channel centre). Axes display eigenvalues. Note that point clusters plot into distinct areas of the slab, which do not correspond to lithofacies. Dependencies and relationships of the properties are significantly different compared to the other channel subenvironments. Outliers are displayed in red. For details and interpretation, see caption of Figure 5a and text.
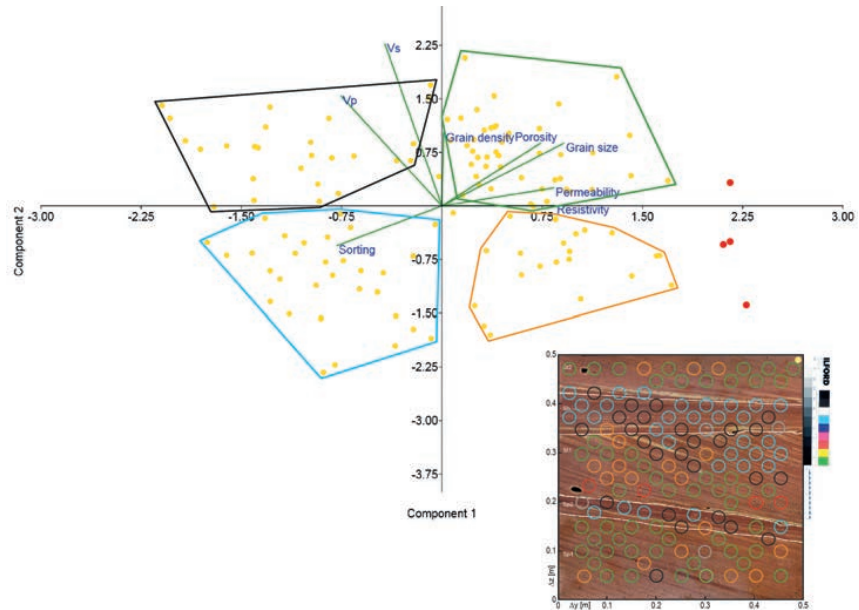
FIGURE 5C

PCA of slab 3 (channel margin). Axes display eigenvalues. Note that clusters of data points dominantly plot into distinct areas of the slab, mostly not corresponding to lithofacies. This is a common feature in all subenvironments. Dependences and relationships of properties are only slightly different compared to channel base subenvironment. Outliers are displayed in red. For details and interpretation, see caption of Figure 5a and text.
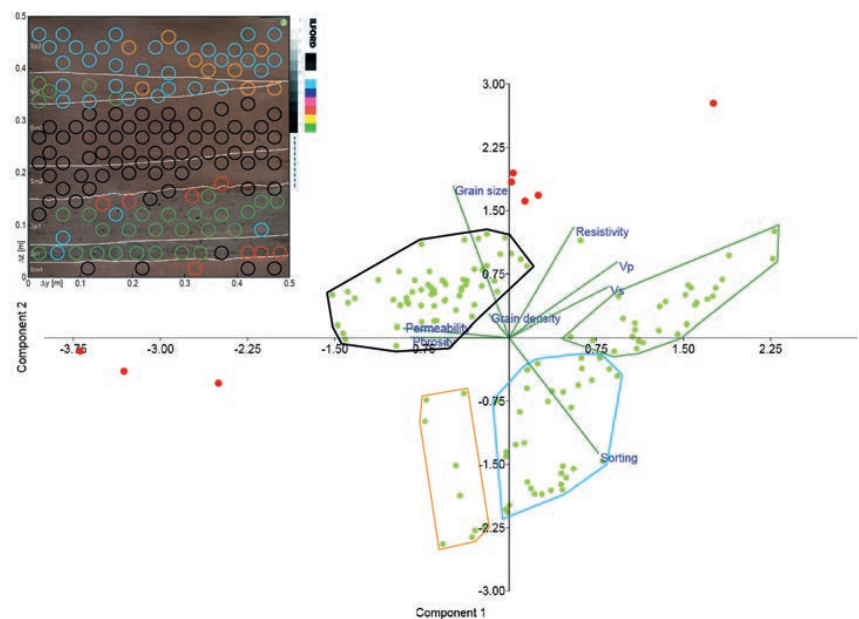
Resistivity changes coincide with porosity and permeability variations in slabs 1 and 2. However, resistivity changes correspond to changes in ultrasonic speed in slab 3.

Data points of all diagrams (Figure 5a-c) show strong clustering despite such inconsistency. Data points are traced back to their positions revealed a correspondence to the lithofacies type of each slab. In slabs 1 and 3, coherent point clusters correspond to successional areas, following stratification and trends within lithofacies types. Each set of

the same lithofacies type however shows slightly different relationships of its properties. In slab 2, property relationships differ significantly between neighbouring areas and do hardly follow lithofacies boundaries. Hence, the tool of lithofacies analysis is not significant enough to characterize different kinds of petrophysical relationships.

Compilation of data from all slabs (Figure 6) reveals more comprehensive relationships between properties and displays significant differences between subenviron-

ments. Here, the log of permeability data was used before performing PCA to better detect linear relationships. Outliers were removed when they were out of the interquartile range.

- Siliciclastic sediments show a relationship of porosity and permeability, which is negatively correlated with sorting.
- Ultrasonic speed is weakly and positively correlated with sorting and negatively correlated with porosity and permeability. This relationship illustrates that fast wave propagation is related to well 'welded' grain contacts.
- Resistivity and grain density are very well negatively correlated, but both are relatively independent from any other property. Surprisingly, this relationship is mainly responsible for the clustering of data points across the different environments. Note that resistivity is on the precondition of 100% saturation. See also discussion in the Methodology section above. This relationship may reflect differences in mineralogical composition rather than in pore size distribution.

Another striking feature in analysing all data points is quite prominent clustering of different slabs. It suggests that properties within a subenvironment are closer correlated than properties across subenvironments.

In conclusion, the variability of petrophysical properties in plug-scale measurements within a slab and between the slabs is significant. Hydrodynamics as interpreted from bedforms and grain size are the most relevant geological properties, governing the distribution of petrophysical properties. This is because hydrodynamical conditions,

in a given type of sediment, strongly change with environment AND each set of a sedimentary structure within a lithofacies type AND even within a sedimentary structure.

Box-whisker plots (Figure 7a) provide absolute values for interpolation and upscaling. Data allow identification and hydrodynamical interpretation of overall trends:

According to these plots, the largest variance (including outliers) in apparent permeability, porosity and sonic wave velocity is observed at the channel top (slab 3). The second largest in slab 1 (channel base). The smallest occurs in slab 2 (channel centre). Resistivity shows roughly the opposite trend.

Three characteristic processes for each environment might be responsible for it:

At the base of a channel (slab 1), a heterogeneous grain size spectrum reflects variable depositional processes such as erosion from turbulent flow with scouring and downcutting. This is also indicated by the frequent occurrence of clay clasts. At the same time, turbulence is indicative for fast-changing stream power at small scale. This is an additional factor creating variance within foresets and across different sets of the same lithofacies type. However, there is always water; extreme conditions, which can create outliers, are rare.

Samples from the channel centre (slab 2) show the smallest variance. Bedforms as expression of stream power are most constant. There is less friction to be expected
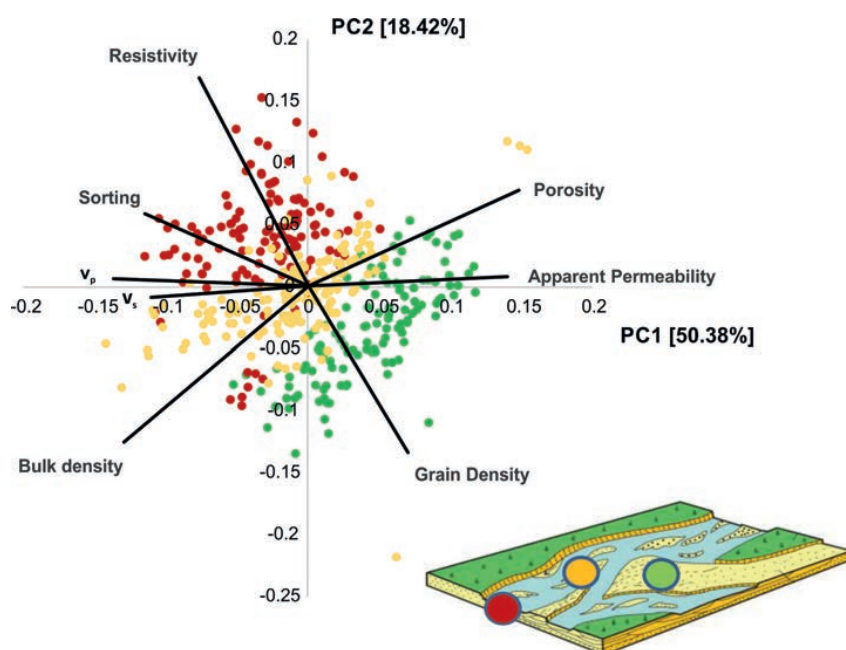
at pre-existent boundaries in this environment. The mid-channel bars are entirely created by the river itself, so that flow can be considered as mostly laminar with only minor turbulences. This results in steady state conditions producing a homogenous type of sediment in rather constant stream power. Hence, less significant variances occur.
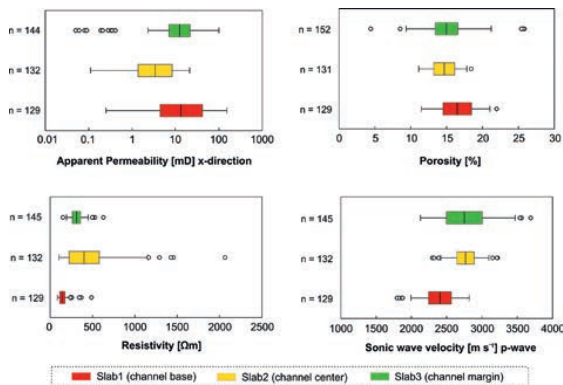
FIGURE 7A

Box-whisker plots are useful to compare values of all siliciclastic subenvironments. Readings of different subenvironments (=slabs) show differences in the arithmetic average as well as in the 25% and 75% confidence interval (box). Circles are considered as outliers (out of innerquartile range). Variance bars and size of the box tell how steady hydrodynamic conditions are during deposition. Channel base shows high stream energy and a lot of turbulence generating higher values with significant variance. Channel centre was most stable, and channel margin experiences variations in depositional energy, probably due to changing water level. Sonic wave propagation and resistivity follow other, but also systematic, rules.

Data variance from samples from the upper bar (slab 3) is the largest, when outliers were considered also. A large number of outliers (circlesdetected by the inter-quartile range method) indicate a large number of extreme conditions. Raising and falling water levels mostly affect this depositional subenvironment as well as spill-over processes at bar surfaces, which are responsible for outliers probably due to unsteady flow conditions and rapid deposition. At the same time, the confidence interval (25% to 75% box) of porosity and permeability is smaller compared to the other slabs. Hence, most of the time, deposition under rather stable flow conditions seems to be responsible for these distinct petrophysical values. At least, diagenesis acts different in an environment characterized by stable, but sometimes extreme conditions. Probably also early diagenetic effects may play a role at bar surfaces. Thin-section analyses and SEM revealed a higher degree of cements in slab 3. As sonic wave velocity is strongly dependent on mechanical properties of grain-grain contacts, a higher variance is likely.

Histograms and Gaussian distributions of porosity and permeability readings (Figure 7b) show that lithofacies do not provide a more detailed tool to relate values to flow conditions. The spread in readings of different lithofacies types are significant. They do not allow for a precise characterization. If readings are sorted after depositional subenvironment (slabs), distributions get more ordered.

Anisotropy of apparent permeability

All measurements discussed earlier are results of individual plug measurements, that is, 1D analyses. They were measured along the direction a plug was drilled. In Figure 8,
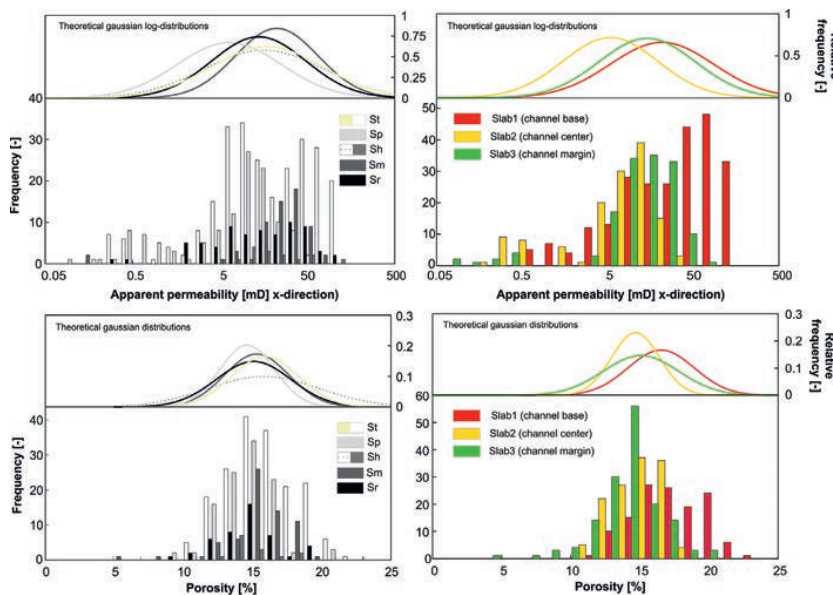
FIGURE 7B

Histograms of porosity and permeability sorted after lithofacies (left) and subenvironment (right). Gaussian distribution of values is shown on top. Distribution after lithofacies shows a large overlap. Arrangement after subenvironment displays distributions that are more distinct.
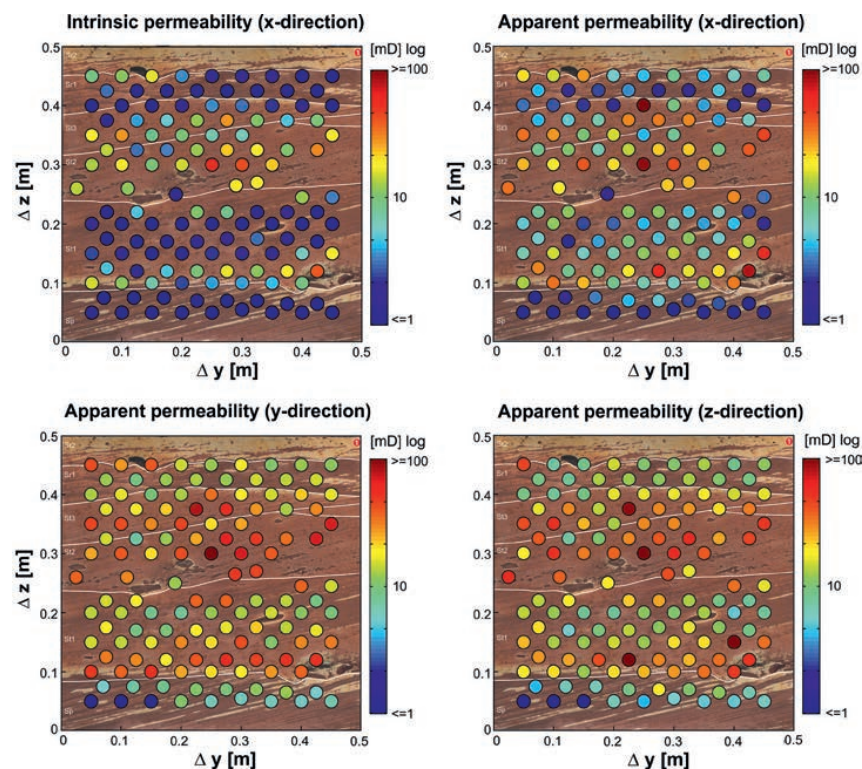
Intrinsic permeability (x-direction) Apparent permeability (x-direction)
Apparent permeability (y-direction) Apparent permeability (z-direction)

this corresponds to the x-direction, which is at the same time the direction of flow of intrinsic permeability measurements (cylinder probe). This is why readings of apparent permeability (surface probe) in x-direction are more similar to intrinsic permeability than they are to y- or z-direction.

To investigate spatial anisotropy of apparent permeability, 3D measurements were performed at slab 1 (channel base, Figure 8) using the surface injection probe permeameter at the sides of the plug (see Methods section). A strong anisotropy is observed in both: the vertical (z) and the horizontal (x and y) direction. A few plugs however show no directional change of flow or only changes in one direction. This mimics that permeability particularly in cross-bedded rocks is quite complex. It often changes at a centimetre scale.

Figure 9 presents the readings as a ternary diagram to better display and calculate anisotropy. The corners of this diagram represent readings of the 3 dimensions normalized to 100%. An isotropic sample would plot in the middle of the triangle. A sample with maximum anisotropy, that is, tight in one direction, would plot at a boundary. Intrinsic (cylinder probe) permeability is represented by the size of a sample point. The diagram reveals that samples with higher intrinsic permeability (larger circles) tend to be more isotropic. Highly anisotropic samples reveal tightness in x-direction only and show an overall small intrinsic permeability (small circles/dots).

The behaviour of anisotropy follows primary depositional features combined with flow geometry effects. Fine-grained foreset boundaries act as permeability barriers. In x-direction, these are mostly oriented parallel to the plug axis, because this is the strike of the river flow direction. Therefore, injected gas of permeability testing is directed through a longer distance along the plug axis. In y- or z-direction, the flow escapes sideward after a short travel distance, because it hits always a foreset boundary hindering further penetration. The shorter flow path in y- and z-directions results in higher apparent permeability. This quite different flow geometry along and crosswise foreset boundaries with injection probe measurements cannot be compensated mathematically, because the exact flow pathways are always unknown. This effect is strongest with small intrinsic permeabilities as flow is forced through the whole plug. This can be recognized in Figure 9 as described above. Hence, interpretation of these results has to be done with respect to this methodological artefact.

Relation to the pore system

The 3D investigation revealed the importance of permeability barriers. Permeability, as a proxy for pore throat distribution, highlights the significance of the pore system and warrants a visual calibration of the pore system with an SEM and photomicrographs of thin sections. The investigated plugs shown in Figure 10a-d display exclusively
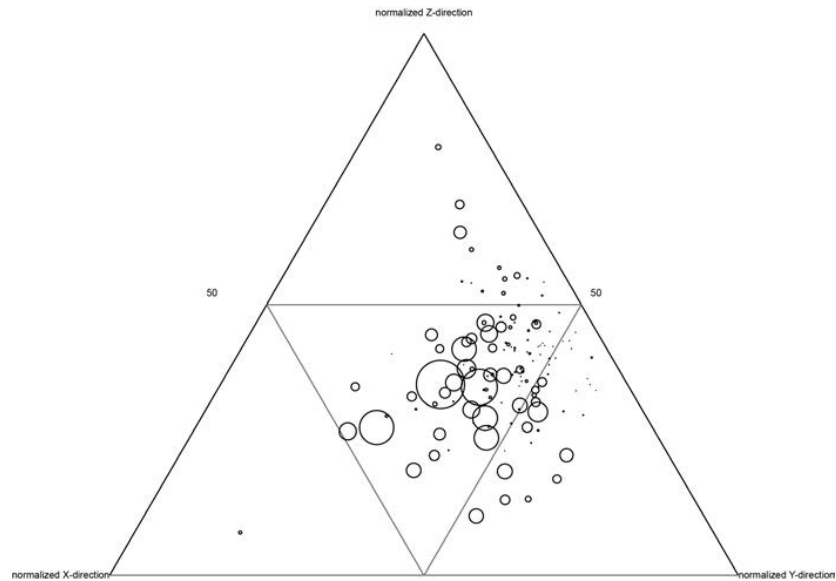
intergranular porosity (Choquette and Pray, 1970; Jennings and Lucia, 2003; Pitman, 1987). A boundary between foresets is shown in the SEM micrograph of Figure 10a and in the thin sections of Figure 10b and c. Bounding laminae are composed of poorly sorted finer grained sediment with grain diameters varying from 32 to 250 μm. The foreset layer itself shows very well sorted grains, with diameters varying between 250 and 350 μm with a large open framework pore geometry. These well-sorted and coarser grains are associated with much bigger and better interconnected pore space. This results in permeability, which is orders of magnitudes higher than that of the bounding laminae. Additionally, small and narrow pore geometries are more prone to cementation as observed in the boundary laminae. Authigenic illite, kaolinite, chlorite as well as hematite cements reduce the remaining pore space even more compared to the much larger pore space of the foreset layer.

In conclusion, boundary layers between individual foresets represent permeability baffles. They deflect or hinder fluid flow, which creates the observed strong anisotropy.

Spatial statistics of permeability and porosity
Variograms are helpful to visualize heterogeneity of petrophysical data as input for interpolation and reservoir models. On the vertical axis of variograms, the calculated curve of the approximation model reaches a 'sill level'. The sill level represents randomly distributed properties where no similarity is observed between values. The distance between data points (bins) is plotted on the horizontal axis. The distance to reach the sill level is the 'range'. To determine the range, a spherical model was used (Wackernagel, 2003). Readings within the range

show stronger similarity to each other the closer they are to the origin of the graph. Some variograms also show 'nugget' effects. A nugget exists when even at zero distance from a reference data point values of a property are not identical. The nugget reflects an even smaller scale of heterogeneity than was measured in a plug sample.
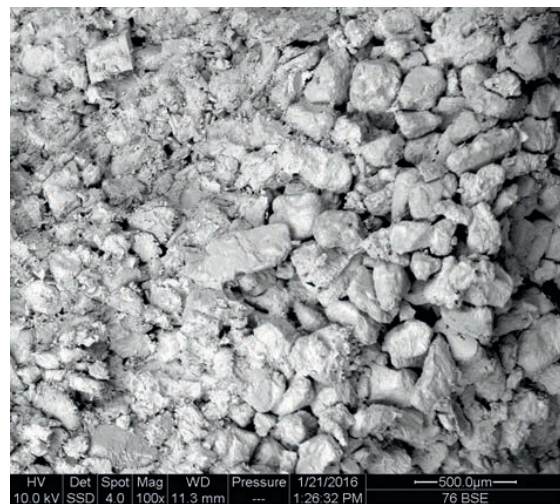
FIGURE 10A

SEM photomicrograph of plug 1-76 (slab 1), lithofacies type St. Lower right part of the image shows open framework pore geometry. The upper left part of the image shows poorly sorted sand with clay cements, both reducing pore space. This photo depicts a foreset boundary of a through cross-bed. Petrophysical data of the plug: porosity 22.0%, apparent permeability 98.7 mD, p-wave speed 2196 m/s, s-wave speed 1362 m/s and resistivity 215 Ωm.
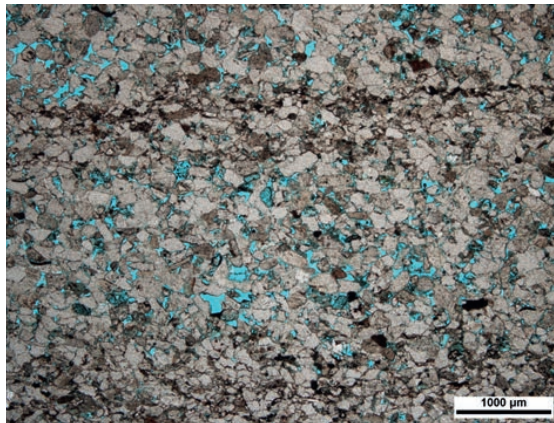
FIGURE 10B

In the thin section of plug 2-38 (slab 2), foreset boundaries of lithofacies type Sp appear as brown bands of opaque minerals (open pore space in blue). These boundaries represent permeability barriers and force fluid flow in a distinct direction, which results in anisotropy patterns. Further explanation is given in the text. Petrophysical data of the plug: porosity 11.8%, apparent permeability 2.4 mD, p-wave speed 2816 m/s, s-wave speed 1568 m/s and resistivity 112 Ωm.

FIGURE 10C

Photomicrograph of a lithofacies without sedimentary structure. Sm of slab 3 in plug 3-98 does show isotropic patterns. Petrophysical data of the plug: porosity 17.1%, apparent permeability 17.7 mD, p-wave speed 2582 m/s, s-wave speed 1486 m/s and resistivity 266 Ωm.

Intrinsic microscale heterogeneity even at the scale of an individual core plug exists.

Variograms of porosity and permeability for slabs 1-3 are shown in Figure 11. For both, porosity and permeability, the range distance of variograms decreases slightly from the rock slab representing channel base (slab 1) to the slab representing channel centre deposits (slab 2) to the rock slab derived from channel margin beds (slab 3). This means that variability of petrophysical values increases from channel base to channel margin in the example investigated. A nugget effect is visible in data from channel base sandstone (slab 1) and channel centre sandstone (slab 3).

Results of the variogram analysis correspond well to the histogram analysis (Figure 7b). Both show patterns for porosity and permeability that are reflective of the rock texture, that is, hydrodynamic regime during time of deposition. Stream power at channel base is strong and turbulences are common. This creates rocks with larger pore throats, higher permeability and porosity and a variance. Variances decrease as stream power reduces and more stable (laminar) flow conditions establish at higher positions of the channel.

Interpolation

Based on the variogram analysis, simple kriging was used to interpolate readings over the entire slab (Figure 12). Quality assessment is provided by calculating the root mean square error (RMSE) and the mean absolute error (MAE), which are displayed for each interpolation, respectively.

Two general observations can be made:
i) Variance of readings is significant within lithofacies but often systematically, for example, a trend from higher values at the base towards lower ones at the top.
ii) Lithofacies boundaries follow often strong contrasts of readings.

By comparing the different subenvironments, strongest and most sudden changes occur at the channel base (slab 1). In the cannel centre (slab 2), contrasts are more calm and do not follow lithofacies boundaries as well as at the channel base. At least at the channel margin (slab 3), large areas are characterized by similar readings ignoring often lithofacies boundaries. However, outliers and sudden strong changes of readings occur within lithofacies.

Summary

At the channel base, turbulences and eddies create strong boundaries and variable hydrodynamic conditions within a lithofacies, for example, creation of cross-bedding. Transported sediment is most different (clasts, grain sizes, etc.). At the channel centre, more calm and continuous flow conditions occur reflected by more homogenous patterns of petrophysical readings. At the channel margin, flow conditions are most calm and quiet, resulting in homogenous patterns. However, this subenvironment is frequently affected by short-term water-level changes, which result in a significant number of outliers. These patterns can be observed in bubble plots, histograms, box plot and variogram analyses and in the interpolation. PCA reveals that grain density plays an important role in explaining variance between the different depositional subenvironments. This reflects depositional energy, which decreases from channel base over centre to margin.

FIGURE 11

Variograms of porosity (left) and permeability data (right). The charts show that the range distance decreases from channel base over channel centre to channel margin (use of spherical model). Note that nugget effects occur also in a systematic manner. For details, see paragraph Spatial statistics of permeability and porosity.
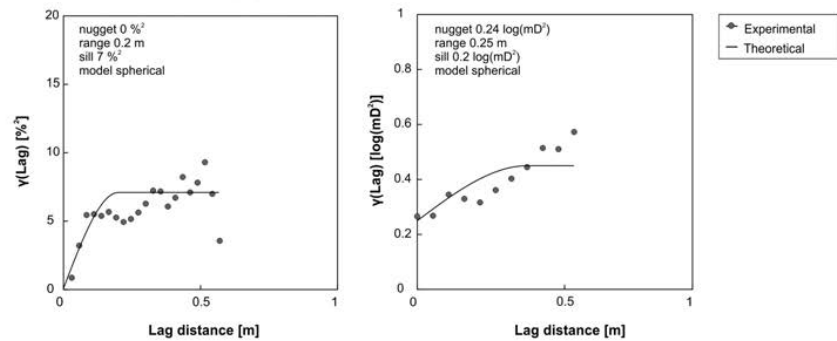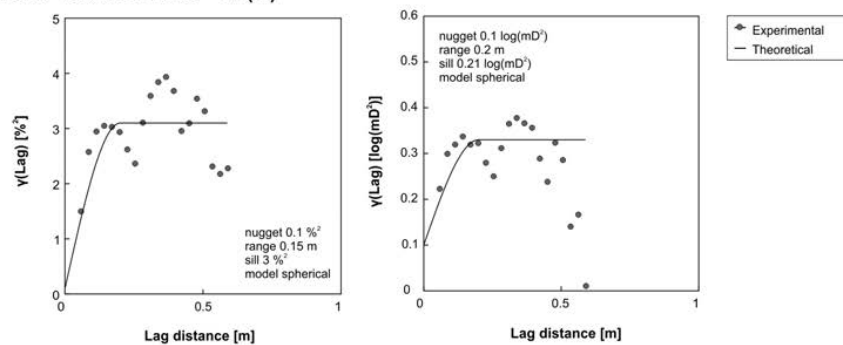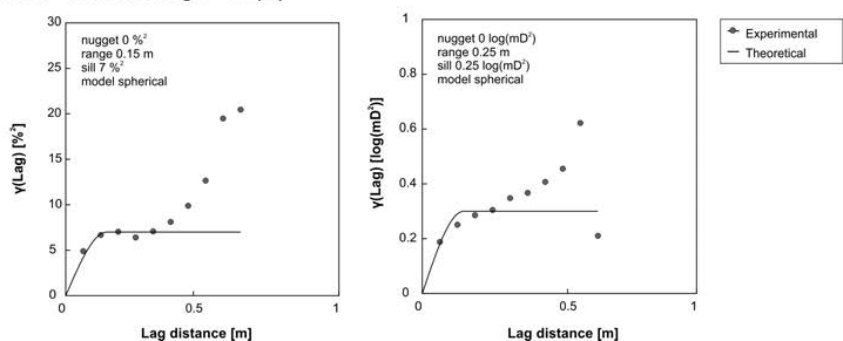
Hence, analysis of petrophysical data from 3 subenvironments of a fluvial channel system suggests that petrophysical properties are strongly governed by sedimentary structures reflecting hydrodynamic conditions in samples during time of deposition. Hydrodynamical condition reflects stream power, flow regime (turbulent or laminar), its direction and changes. Frequency and speed of changes in stream power seem to be a fundamental control on porosity and permeability. Postdepositional diagenetic alteration is of limited importance in the samples analysed.

These processes arrange the sand grains available in the study region and control sedimentary properties like sorting and grain size trends during deposition. Stream power controls grain size, grain density and sedimentary structure. Currently used lithofacies reflect average hydrodynamic conditions mainly based on stability fields as proposed by Reineck and Singh (1980). These hydrodynamic stability fields of sedimentary structures in relation to average stream power and grain size are much too coarse to properly explain or even predict porosity-permeability attributes in the data set investigated.

In addition, lithofacies also reflect variance of depositional energy. As lithofacies requires a subdivision in a range of hydrodynamic conditions, a large scatter and trends are observed within lithofacies types. Much different readings can occur for different sets of the same lithofacies type. This results in permeability baffles, heterogeneity and anisotropy on the scale of individual lithofacies.

However, heterogeneity is scale dependent. On a grid block (voxel) or subenvironment scale, depositional processes exert a first-order control on petrophysical architecture. Hence, subdividing lithofacies based on primary depositional flow characteristics provides a meaningful classification of petrophysical properties on various scales. Depositional flow characteristics are predictable by facies models (Figures 1 and 3) and consequently petrophysical behaviour to some extent.

Carbonate ramp system

Lithofacies description

The Mesozoic carbonate ramp system is represented with 4 rock slabs from 4 distinct depositional subenvironments (Figure 2). The outer ramp is affected by wave-induced traction flow, and the inner ramp is affected by tidal currents.

The study is carried out on samples that suffered limited diagenetic changes only, that is, some dissolution and some cement precipitation altered the pore space as shown in thin sections. Only samples of slab 4 (lagoon) show diagenetic changes altering or obliterating depositional features.

The ramp system sampled is composed of skeletal grains and lime mud in variable proportions (Figure 13). Lithofacies classification used resembles the well-established scheme for siliciclastic deposits (Miall, 2012; Hornung and Hinderer, 2011), which was used in the previous chapter. Grain size and texture (proportion of mud versus components) are described using the classification of Dunham (1960). The facies short code denominates a capital letter for texture and a lower-case letter for sedimentary structures. Stream power is determined based on the chart published by Reineck and Singh (1980) and Harms and Fahnestock (1965). Some 13 lithofacies types are distinguished. A characterization is given below:

Mudstone, dolomitized (code Md)

This lithofacies type is restricted to slab 4 (lagoon). This sample is intensely dolomitized and appears massive or cloudy. Primary sedimentary structures are destroyed. The rock slab is differentiated based on dolomite crystal size, framework and colour. Areas are denoted from 1 to 7:

1) Whitish with clearly bounded, sharp-edged fragments. Borders are formed by white rims of calcite. Apparently, this is a breccia.
2) Dark grey rock fragments disintegrated by fractures, cemented with calcite, which let it appear like an in-situ breccia.
3) Brown-yellowish with minor dark-brown cloudy structures showing gradual colour changes.
4) Brown-yellowish to light-grey colours, which are smeared in a wavy horizontal manner forming distinct shaded areas.
5) Dark grey, dense and homogenous mudstone in which a lot of thin calcite veins occur.
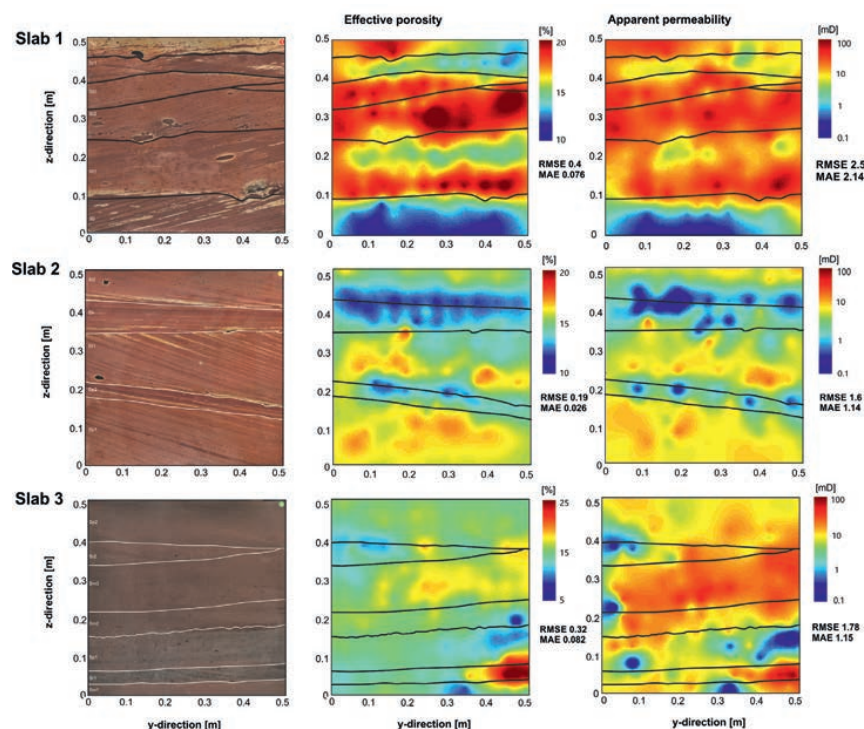
FIGURE 12
Interpolation of porosity (middle) and apparent permeability data (right) using simple kriging. At left, depositional structures and mapped lithofacies boundaries are shown. In general, interpolation algorithms based on the variogram analysis could nicely reproduce and simplify the measured data. Please note significant but systematic changes within a lithofacies type. However, some lithofacies boundaries correspond well to general changes of interpolated readings and some do not. Compare also to Figure 4 a-c. For details, see 'Interpolation' and 'Summary' (fluvial channel system).
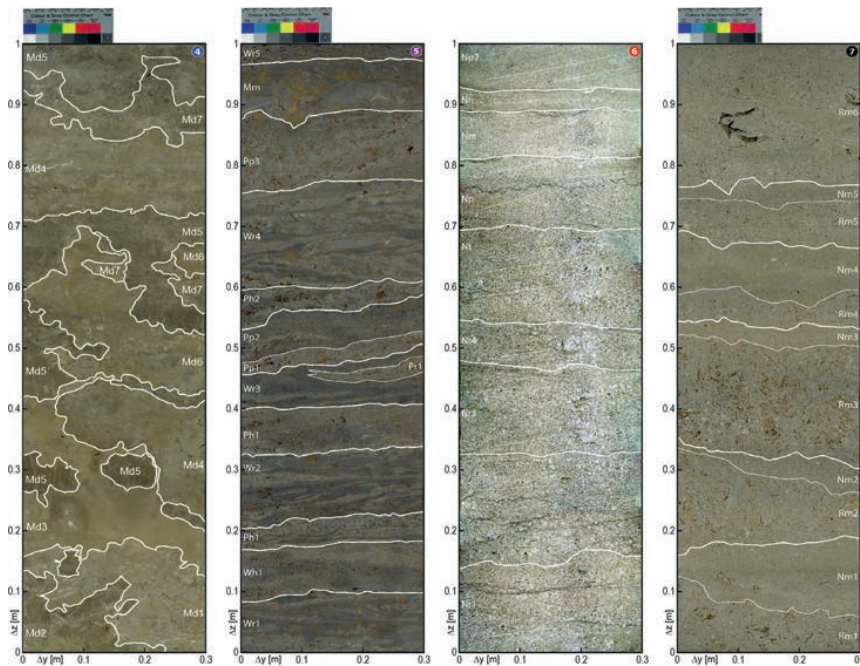
Core slab photographs showing all investigated carbonate slabs with their lithofacies make-up. (4) Lagoon with strong diagenetic overprint (dolomitization, dissolution), (5) tidal flat with herringbone cross-beds, (6) shoal composed of crinoidal grainstones, (7) foreshoal composed of rudstones and some grainstones. Same lithofacies got a number shown at the end of the code in order to trace back petrophysical readings to individual beds. For details, see Lithofacies description (carbonate ramp system).

6) Horizontal bedding, light to yellowish-brown colours occur. Boundaries vary between gradual and sharp. White irregularly shaped spots occur. These represent fenestral pores filled with calcite.

7) Small-scale (1–5 mm) dark mudstone particles supported by a grey matrix.

Mudstone massive (code Mm)

This type is dark-grey in colour and shows no depositional structures except irregular shaped areas with brown ferric oxyhydroxide-siderite grains.

After the hydrodynamic chart of sedimentary structures after Reineck and Singh (1980), this means stream velocities were lower than 10 cm/s. This could be deposits of a shallow lagoonal or marsh area.

Packstone, rippled (code Pr)

On light-grey ripple foresets, thick dark-grey mud drapes are deposited so the rock gets a flaser-like appearance. Dip directions change forming herringbone structures. As bedding planes are planar shaped, ripples must have a straight crest. The average grain size is 500–700 µm, which indicates bidirectional current velocities of 20–30 cm/s. These depositional features are diagnostic for a tidal flat (slab 5).

Packstone, planar cross-bedded (code Pp)

Whitish and light-grey 1–2 cm thick foresets dip with an angle of up to 20°. Set thickness is up to 15 cm. As bedding planes are planar shaped, bedforms must have a straight crest. They always dip in the same direction. Together with the average grain size of 700–1000 µm, this

indicates current velocities of about 60 cm/s. Herringbone stratification indicates bidirectional flow in a tidal channel.

Packstone, horizontally bedded (code Ph)

Intercalating whitish and grey horizontal beds with a thickness of mm to cm form sets from 1 to 10 cm thickness. The average grain size is 700–1000 µm. After Reineck and Singh (1980), this depositional energy belongs to the field of lower plain bed, which is 30–40 cm/s. This bedform is probably found in the tidal flat.

Wackestone horizontally bedded (code Wh)

Same lithology as Ph, but with higher proportions of lime mud at the expense of fragments, which gives the rock a dark grey appearance with smaller whitish or light-grey beds. Depositional energy should be around 20–30 cm/s. This bedform is probably found in the tidal flat.

Wackestone rippled (code Wr)

It is dominated by cm thick mudstone foresets with interbedded millimetre thin packstone foresets; therefore, it shows a lenticular appearance. On average, it is a wackestone. Dip directions do change in co-sets, but not within sets (3–10 cm thick). Dip angles are with 5–10° lower than in lithofacies type Pr, and also grain size is just 350–500 µm. Hence, stream power is also lower, 15–25 cm/s. This bedform is probably found in the tidal flat.

Grainstone, horizontally bedded (code Nh)

The code of grainstone lithofacies types is not consistent to the general scheme, because "G" is already

used for "Gravel" in the siliciclastic lithofacies code scheme. Therefore, the last letter of "grain" has been chosen as the first capital letter for the carbonate clastic system code. In general, all grainstones are of spotty appearance as the bigger grains, calcite precipitations between the grains are whitish, and the smaller grains are light grey. They are most common in slab 6 (shoal subenvironment). Grain size of Nh is 500–700 µm and combined with horizontal beds (1–2 cm thick), and this means a stream power of 20–40 cm/s. This lithofacies type may indicate shoal areas affected by wave-induced traction flows.

Grainstone, planar cross-bedded (code Np)

Foresets are 1–3 cm thick and dip with around 20°. Material change at bedding planes is weak, so layering is hardly visible. Sets are up to 10 cm thick. Grain size is 500–700 µm, and between the grains, diagenetically precipitated calcite occurs frequently. In terms of stream power, this lithofacies represents probably a flow velocity of 45–55 cm/s and commonly indicates shoal areas affected by wave-induced traction flows.

Grainstone, massive (code Nm)

The facies do not show sedimentary structures. Sets are up to 15 cm thick. Grain size is 700–1400 µm and probably there was no time to sort grains and form bedding planes. This lithofacies type indicates rapid sedimentation in shoal areas affected by wave-induced traction flows.

Grainstone, trough cross-bedded (code Nt)

Foresets with components showing grain diameter of 700–2000 µm are interbedded with foresets with grain diameter of 250–1000 µm. Each is 3–5 cm thick and dip with angles up to 20°. Sets are up to 25 cm thick, which is why this is classified as underwater dunes. Mudstone-lithoclasts of up to 15 mm in diameter occur. Hence, this lithofacies type indicates stream power of 90–120 cm/s. This lithofacies type indicates shoal areas strongly affected by wave-induced traction flows.

Grainstone, ripple cross-bedded (code Nr)

Here, only trough-shaped ripples occur. Foresets are 1–3 cm thick and dip with 5–10°. The core of the foresets represent well-sorted crinoidal particles with a grain size of 1000–2000 µm, but pores in-between are often filled with mud, so that the central part of the foreset appear as a packstone or rudstone layer. Bedding planes and the bigger proportion of the foresets show decreased grain sizes (250–1000 µm) with no mud in the pore space, but calcite precipitates. Sets are up to 5 cm thick, co-sets up to 50 cm. In terms of stream power, this lithofacies type points to 40–60 cm/s. This lithofacies type commonly indicates shoal areas affected by wave-induced traction flows.

Rudstone, massive (code Rm)

Rudstone is restricted to slab 7 where it is dominant over grainstone. It exclusively occurs as massive deposits lacking any depositional structure and with chaotic texture. However, some fining-upward trends are observable in individual layers. This trend is even completed when the grainstones on top of the rudstone layers are also considered part of the trend. Beside some mudstone clasts, the majority of components of the rudstone are flat-shaped shell fragments of 2–20 mm in size. Hence, behaviour in currents differs significantly from normal spherical grains. Therefore, no reliable statement can be provided for stream velocities during deposition, also because sedimentary structures are missing. Rudstone layers sometimes show an erosive base but a gradual transition to overlying grainstones. Assuming horizontal bedding as it roughly occurs between grain and rudstone layers in slab 7, stream power can be estimated to be at least more than 100 cm/s, which means that these deposits represent the highest energy of all investigated subenvironments. In conclusion, this lithofacies type can be a tempestite, because of the high depositional energy (decreasing at the end of the deposition), its erosive base and its chaotic texture (Aigner, 1985). Typically, it occurs in foreshoal areas.

Multiparametric statistical analysis

Rock slabs were investigated petrophysically using core plug measurements and probe permeameter measurements as described for the siliciclastic slabs above. Measurements were initially displayed as bubble plots on photographs (Figure 14a-d) of all 4 slabs. These pictures show spatial relationships of measurements and characteristics of the rock slabs such as lithofacies. A first-pass comparison revealed qualitatively a clustering of data, distinct trends within these lithofacies and along individual beds. The visual comparison was an important first pass to detect special dependencies. Statistical methods that are space independent do not reveal such relationships. This visual comparison also revealed larger heterogeneity compared to clastic samples. Readings do not relate to lithofacies alone, and changes in grain shape and mineralogy are superimposed.

A PCA was carried out (Figure 15) to better analyse all readings acquired of a particular rock slab. In all PCA plots of this study, the log of permeability data was used and obvious outliers were removed to be able to better detect linear relationships. Each slab represents a specific depositional subenvironment (slab 4, lagoon; slab 5, tidal flat; slab 6, shoal; slab 7, foreshoal).

- All slabs show a strong negative correlation between porosity and bulk density, as one could expect.
- In general, acoustic wave velocity seems to be independent from all other properties.

Other petrophysical parameters show some spatial correlation but are similarly wide spread among subenvironments:

- Slabs 4 and 6 show little, if any, relationship of porosity and permeability. Slabs 5 and 7 display a fair correlation.

- Resistivity is negatively correlated with grain density and porosity in slab 4. In all other slabs, there is almost no correlation with other parameters.
- Grain density is negatively correlated with bulk density in slabs 4 and 6 and slightly in slabs 5 and 7.
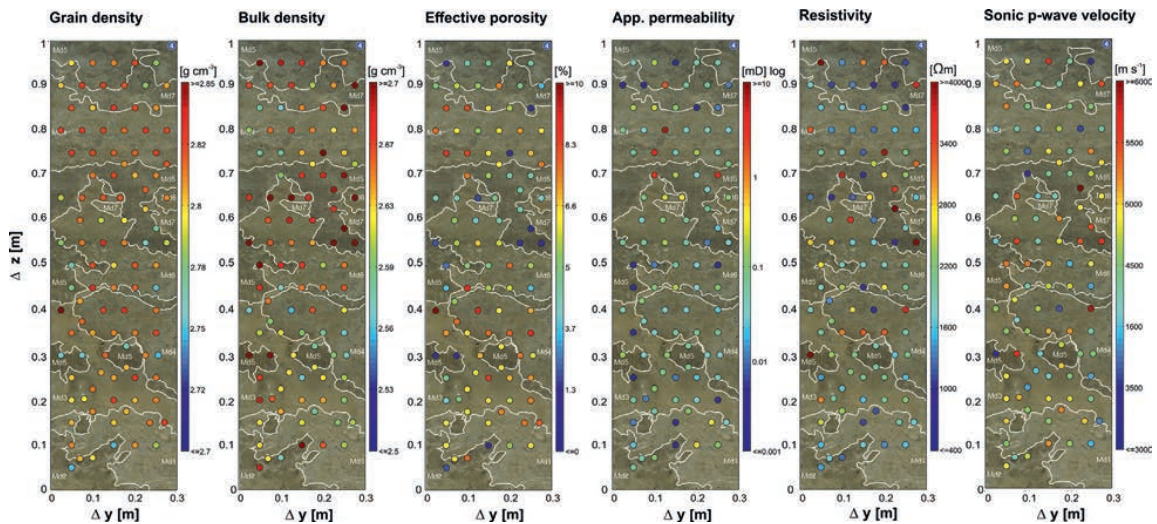
FIGURE 14A

Photograph of rock slab 4 (lagoon) with petrophysical data shown as dots. Colours represent values at their original position. These bubble plots show a heterogeneous, almost chaotic, pattern with a lot of scattering, but still there are trends going along with colour (=type of dolo-mudstone, see text): In terms of bulk density, higher readings occur with dolo-mudstones of darker colours. Grain density seems to have an overall vertical trend to higher values. Porosity follows that roughly, but additionally Md5 seems to have reduced values. Permeability changes often across lithofacies boundaries, but no lithofacies shows a typical distribution. Resistivity and p-wave velocity do not show any patterns, which can be related to the visual appearance of the rock.
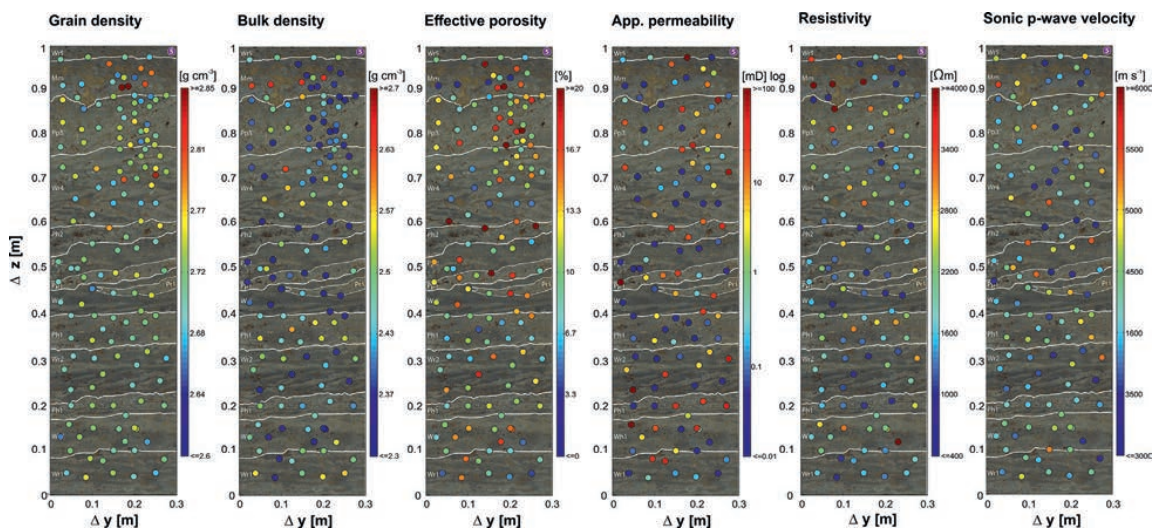
FIGURE 14B

Photograph of rock slab 5 (tidal flat) with petrophysical data shown as dots. Colours represent values at their original position. Readings show a certain dependence on lithofacies (wackestones have lower readings than packstones) and colour variations (as darker as lower the reading). Here, dark colours and wackestones represent higher amounts of lime-mud. In terms of resistivity and p-wave velocity, these trends are not established.

Different carbonate subenvironments showed individual correlations of geological and petrophysical parameters. This is as observed in fluvial channel subenvironments. Specific to the carbonate system is the poor correlation of porosity and permeability in some cases.

To highlight interdependencies, results of a PCA are shown in Figure 16a for all carbonate slabs. The different slabs cluster in different areas of the PCA plot. This sup-

ports that there are specific petrophysical relationships for each subenvironment, respectively, as already suggested by the individual plots:

- According to their position and spreading along the direction, porosity and permeability are the largest contributors to the variance in the data set of slabs 7 (foreshoal) and 5 (tidal flat). Pores in slab 7 showed a large variability. Pores are not equally distributed. This is a consequence of depositional processes and some post-

depositional diagenetic alterations. Slab 7 is a rock type, which is poorly sorted but largely free of mud due to currents washing out fine material. However, components are bioclastic fragments. These form randomly shaped pores. Slab 5 shows mouldic porosity and strong cementation at the same time, which has the same effects.
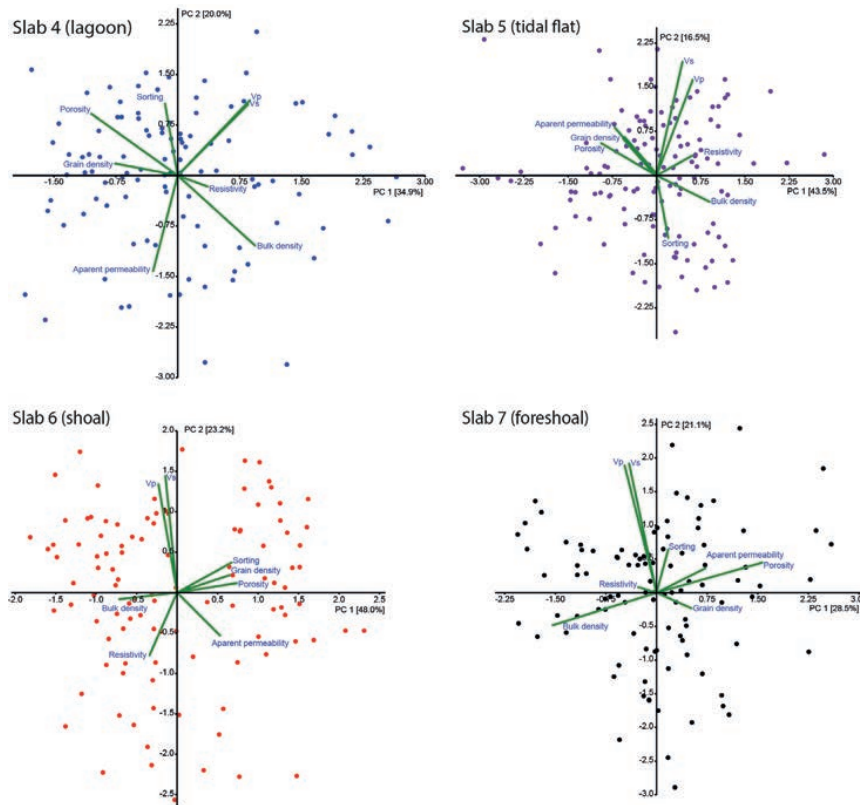
- Acoustic wave velocity, sorting, resistivity and grain density are the largest contributors to the variance in the data set in slab 4 (lagoon) and slab 6 (shoal). Slab 4 is composed of pores that are rather small and similar in shape. However, cement randomly infilled pore space. Slab 4 is a lithofacies made up of mudstone, which was subsequently dolomitized and has no visible relationship to depositional processes. Slab 6 displays the smallest scatter and seems to be dominated by density contrasts. It is composed of variable amounts of crinoid fragments and ooids, which are responsible for this.

A composite PCA of all 7 slabs (Figure 16b) revealed a completely different relationship of petrophysical measurements in siliciclastic and carbonate rocks.
- Siliciclastic data group in the one sector of negative component 1 and positive component 2.
- Bioclastic data group along positive axes of component 1 and negative component 2 (three sectors). They show a significantly larger scatter.

- Grain density, acoustic wave velocity and resistivity are the properties that determine clearly the regions where siliciclastics and carbonates plot. This is as expected and differences in these properties are well known from petrophysical log interpretations.
- Porosity, permeability and bulk density affect all rock types but show significantly more spreading among different types of carbonates (subenvironments).

Absolute values and distribution of readings for porosity, permeability, resistivity and p-wave velocity are shown in box plots (Figure 17a) and in histograms (Figure 17b) for all bioclastic slabs (subenvironments). Data show a large scatter. Readings of individual slabs cluster in an organized fashion reflecting average depositional energy. The median, 25th and 75th percentile, and the variance of all readings reflect this trend towards higher readings from slabs 4 to 7:
- Slab 4 (lagoon) represents the lowest and stable depositional energy but it suffered a strong diagenetic overprint.
- Slab 5 (tidal flat) represents tidal flat deposits with low wave energy but moderate tidal currents.
- Slab 6 (shoal) represents shoal with high wave energy.
- Slab 7 (foreshoal) likely embodies the subenvironment with the highest depositional energy. Storm surges

caused traction with the sea floor in the wave breaker zone. Times without storm action were characterized by moderately high wave energy.

Anisotropy of apparent permeability

To quantify anisotropy of the apparent permeability in carbonates with grain-dominated textures, two subenvironments (slabs 6 and 7) were investigated using 3D measurements of individual core plugs (Figures 18 and 19). The results showed strong anisotropy in many plugs. Anyway, no relationship to lithofacies was found. Permeability was apparently controlled solely by pore size

and pore geometry relative to plug size. It was apparent that small pore systems revealing low permeability are more isotropic on a plug scale. Large pores produced stronger anisotropy often with preferential orientation on this scale. This was a major difference to siliciclastic samples, where sedimentary structures were part of each plug producing continuous permeability barriers at plug scale displaying just the opposite trend.

Relation to the pore system

Analysis of thin sections (Figure 20a-d) revealed increasing amounts of porosity with increasing grain versus mud con-
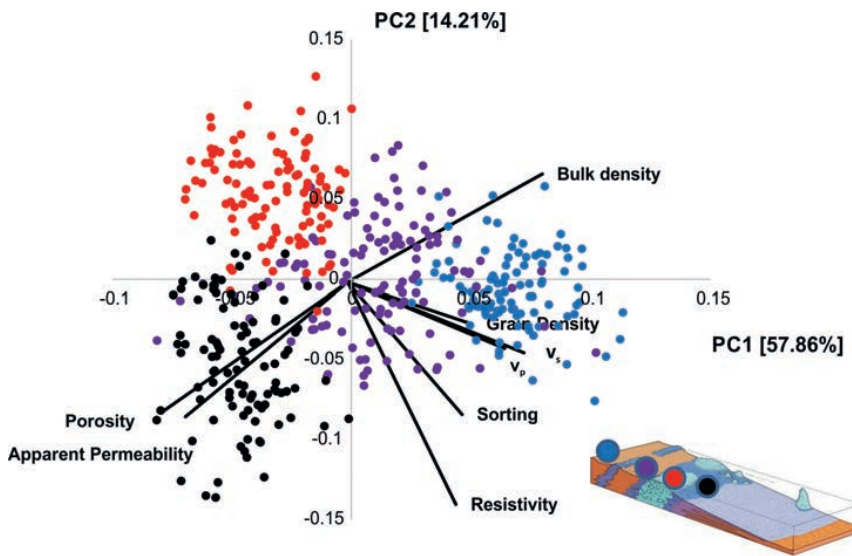
FIGURE 16A

Composite PCA of all carbonate slabs. Colours represent different slabs. The clustering demonstrates that different subenvironments show different controls on deposition and subsequently on petrophysical properties.
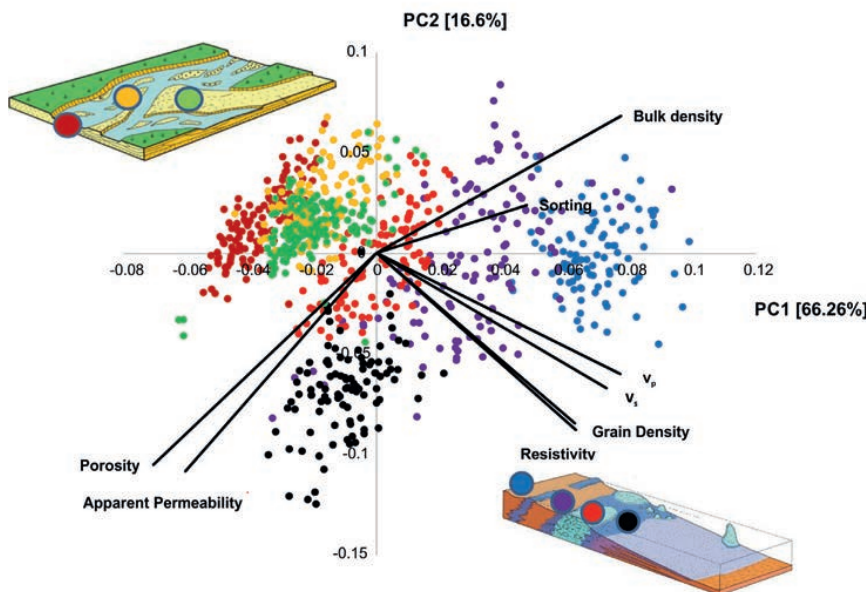
FIGURE 16B

Composite PCA of all investigated carbonate and sandstone slabs. Colours refer to the slab numbers (=subenvironment). The clustering is obvious and demonstrates that in each subenvironment controls on petrophysical properties and deposition are individual. The most significant differences occur between siliciclastic and carbonate clastic systems. More detailed statements are provided in the paragraph 'Multiparametric statistical analysis'.

FIGURE 17A

Box-whisker plot analysis of petrophysical readings of investigated carbonate plugs. Displayed are the 25% and 75% confidence interval (box) and the arithmetic average (division line in box). Circles out of variance bars are considered as outliers (out of innerquartile range). For porosity and permeability, the average, the percentile box and variance bars indicate a clear relationship of subenvironments with depositional energy. The higher the stream power, the higher the values. Sonic speed roughly follows that trend. Resistivity is somehow independent.
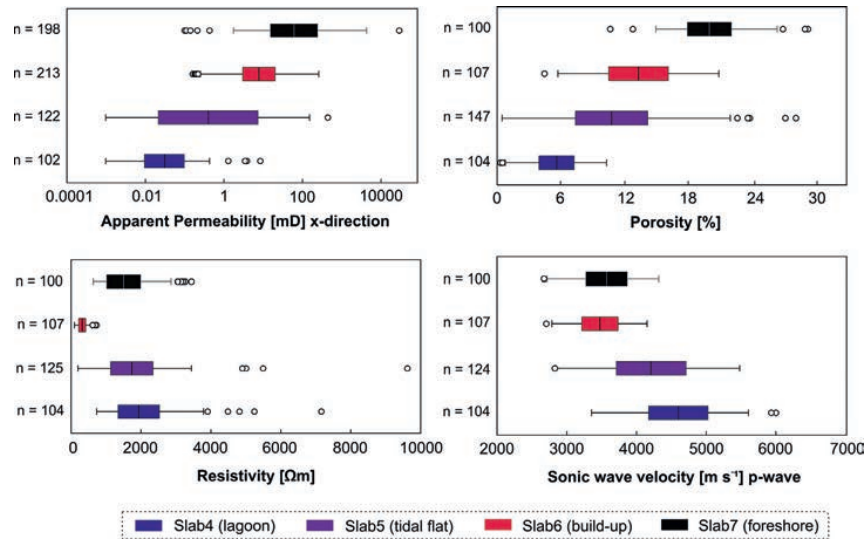
FIGURE 17B

Histogram analysis of porosity and permeability readings of investigated carbonate plugs. Evaluated after lithofacies (left) and subenvironment (right). Gaussian distribution of values is shown on top. Grainstones, wackestones and packstones show similar distributions of values. Characterization based on subenvironment (=slabs) results in a more accurate discrimination.
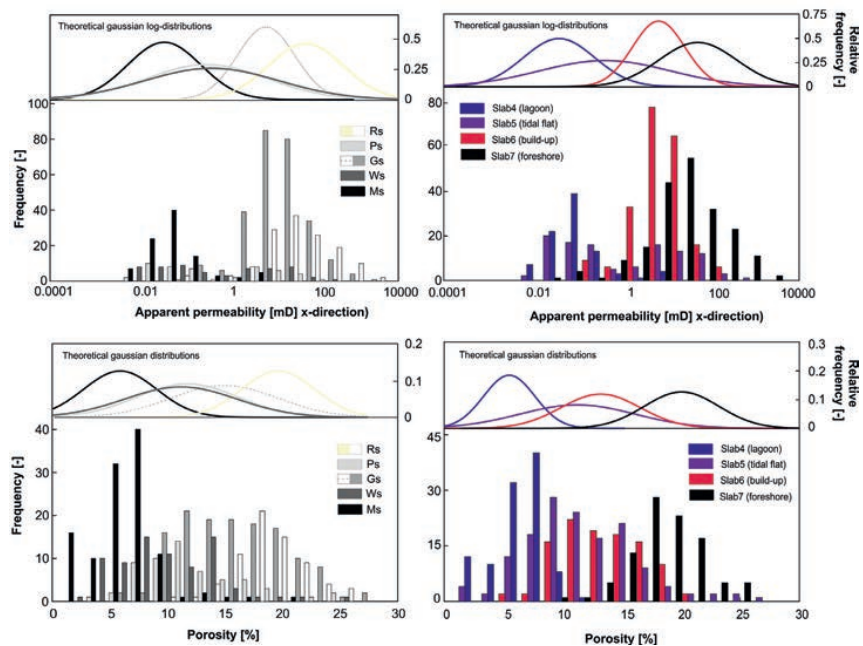
tent (texture) reflecting higher depositional energy. For example, slab 4 (lagoon) showed the lowest, slab 7 (foreshoal) the highest amount of visible porosity. Pores in rudstone (slab 7, Figure 20d) demonstrated the significant influence of the particle shape on pore geometry and direction resulting in a strong anisotropy and variability of permeability. Samples investigated (Figure 20a-c) showed furthermore a poor interconnectedness of the pore space. Pore systems were dominated by mouldic, fenestral, isolated vug, shelter, and intraparticle porosity (Choquette and Pray, 1970; Jennings and Lucia, 2003; Pittman, 1979). This implied narrow and complex pore throats resulting in low permeability, which explains poor correlation of porosity and permeability and their large variance in the PCA analysis.

Spatial statistics of permeability and porosity
Each subenvironment investigated showed distinct porosity and permeability variograms (Figure 21). Slab 4 is highly
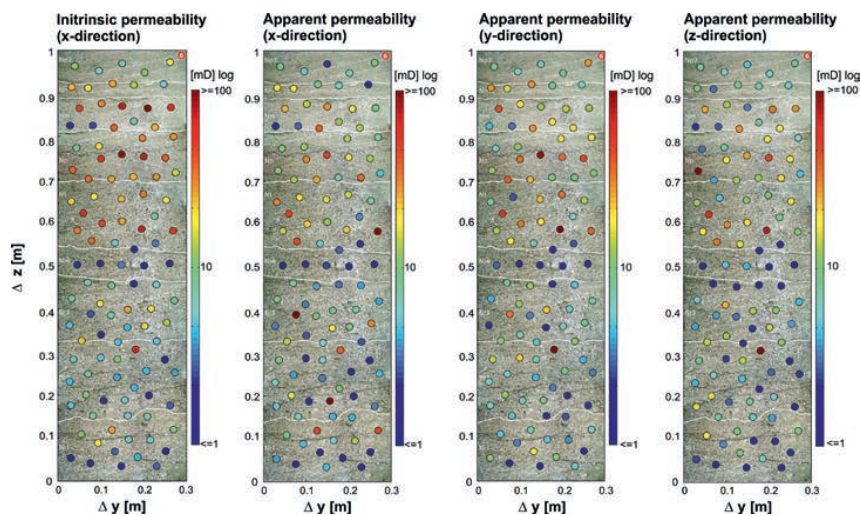
Initrinsic permeability (x-direction) — Apparent permeability (x-direction) — Apparent permeability (y-direction) — Apparent permeability (z-direction)

diagenetically overprinted resulting in complete destruction of primary sedimentary structures. Leaching and dolomitic cementation occurred. These processes seem to act on a dm scale as indicated by the range of porosity and permeability. At the same time, this process shows a small-scale patchy mode (see previous chapter), which results in a nugget effect on sub cm scale.

In contrast, slab 5 (tidal flat) showed strong stratification due to bidirectional current reworking and sorting sediments combined with selective leaching of bioclastic fragments (see previous chapter). This generated significant heterogeneity reflected by a range distance of less than 5 cm (sub-plug scale) and a much higher nugget.

Porosity in slab 6 (shoal) showed an exceptionally large range of approximately 0,6m. This is quite outstanding compared to the range of permeability, which is approximately 10 cm. Spatial patterns of the distribution of permeability and porosity do not correspond to each other. This reflects the decoupling of porosity and permeability due to pervasive patchy dissolution (see previous chapter). However, these heterogeneous pore type distributions seem to be common and are equally throughout the slab.

In slab 7 (foreshoal), permeability shows a higher range as porosity, which means commonly a good interconnectedness of pores, but highly variable in size. This
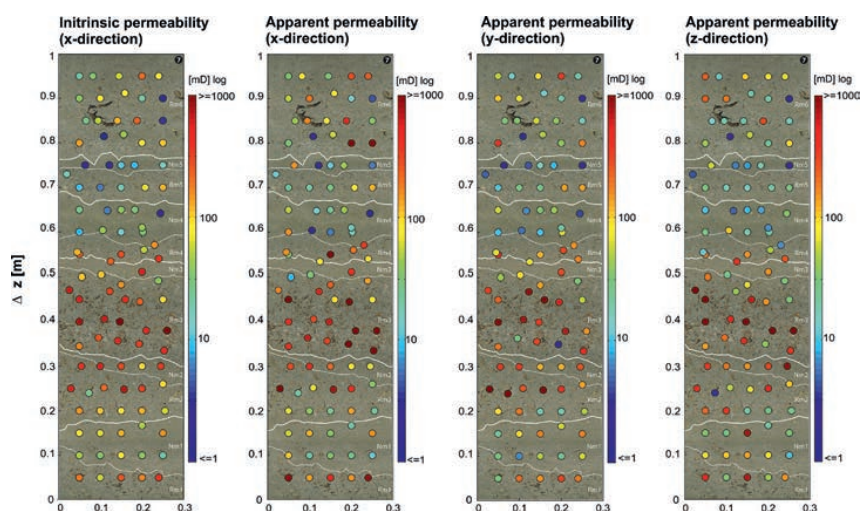
Initrinsic permeability (x-direction) — Apparent permeability (x-direction) — Apparent permeability (y-direction) — Apparent permeability (z-direction)
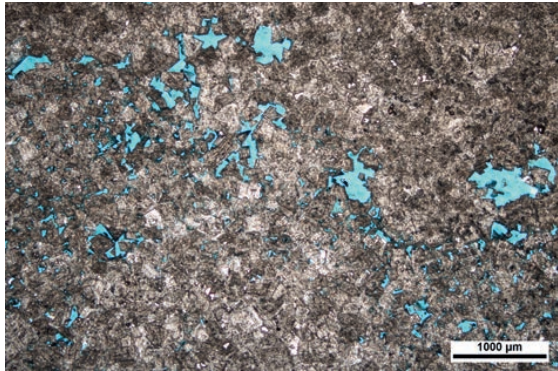
FIGURE 20A

Thin-section photomicrograph of plug 4-77 (slab 4, lagoon). Pores occur mainly at the transition from one lithofacies to another (lower left to upper right). They are poorly connected, which explains the low permeability readings. Petrophysical data of the plug are porosity 6.3%, apparent permeability 0.01 mD, p-wave speed 4973 m/s, s-wave speed 2975 m/s and resistivity 2458 Ωm.
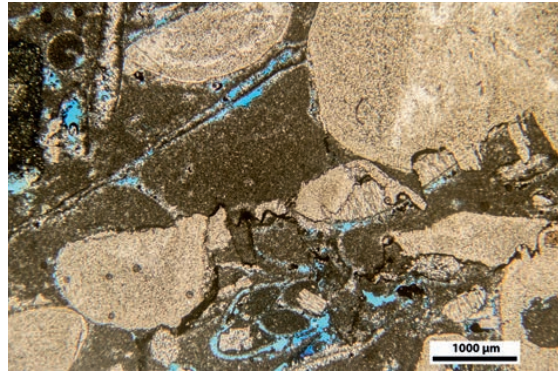
FIGURE 20C

Thin-section photomicrograph of plug 6-27 (slab 6, shoal). Sample is composed dominantly of ooids and shell fragments. Small-scale mouldic, fenestral and vuggy porosity occur frequently. Lime-mud reduces connectivity. Petrophysical data of the plug are porosity 9.9%, intrinsic permeability 3.7 mD, apparent permeability 3.28 mD, p-wave speed 3430 m/s, s-wave speed 2368 m/s and resistivity 292 Ωm.
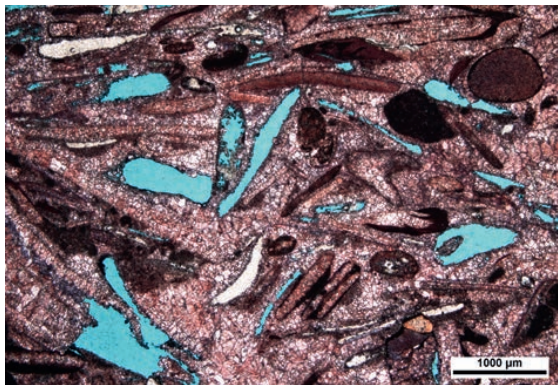
FIGURE 20B

Thin-section photomicrograph of plug 5-61 (slab 5, tidal flat). The sample consists dominantly of shell fragments and ooids. Mouldic porosity is common. Cements hinder connectivity. Permeability is intermediate. Petrophysical data of the plug are porosity 9.6%, apparent permeability 9.5 mD, p-wave speed 4745 m/s, s-wave speed 1352 m/s and resistivity 1227 Ωm.
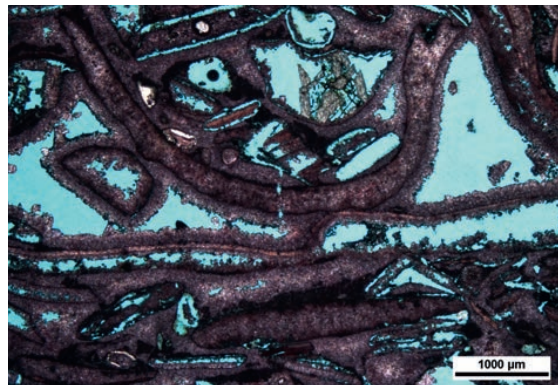
FIGURE 20D

Thin-section photomicrograph of plug 7-3 (slab 7). Sample consists of shell fragments. Shelter and intergranular porosity is common. Intraparticle, mouldic, fenestral and vuggy porosity occur. This lithofacies type (rudstone) has the highest permeability of all investigated slabs. Porosity 29%, intrinsic permeability 261 mD, apparent permeability 392 mD, p-wave speed 3531 m/s, s-wave speed 2124 m/s and resistivity 2200 Ωm.

is due to the large bioclastic fragments and resulting shelter porosity. In addition, this is responsible for the nugget effect observed in permeability data. This means that readings vary significantly already on a centimetre scale. This might be mainly the result of a combination of variable particle shape and depositional processes but also patchily occurring diagenetic leaching. Most of the slab is texturally a rudstone with a chaotic orientation of shell fragments on a centimetre scale (see previous chapter).

Interpolation

Simple kriging (Figure 22) based on the variogram analysis made the distribution of permeability and porosity data better recognizable compared to bubble plots. Slabs 4 and 5 revealed patchy distribution of readings within lithofacies types. Areas with similar patterns cross frequently lithofacies boundaries. However, some lithofacies boundaries are respected. Thin-section analysis comes up with patchy mouldic and vuggy pores suggesting a dominance of diagenetic processes over primary depositional features.

In contrast, primary features are more preserved in slabs 6 and 7, which makes petrophysical features more correspondent to lithofacies boundaries. However, in these two slabs, high amounts of randomly distributed readings occur due to selective leaching and shelter porosity. In conclusion, predictive lithofacies analysis in terms of poroperm readings is almost impossible. However, petrophysical prediction may work well on depositional environment scale.

Summary

The relationships between petrophysical readings and their controls behave significantly different in the carbon-
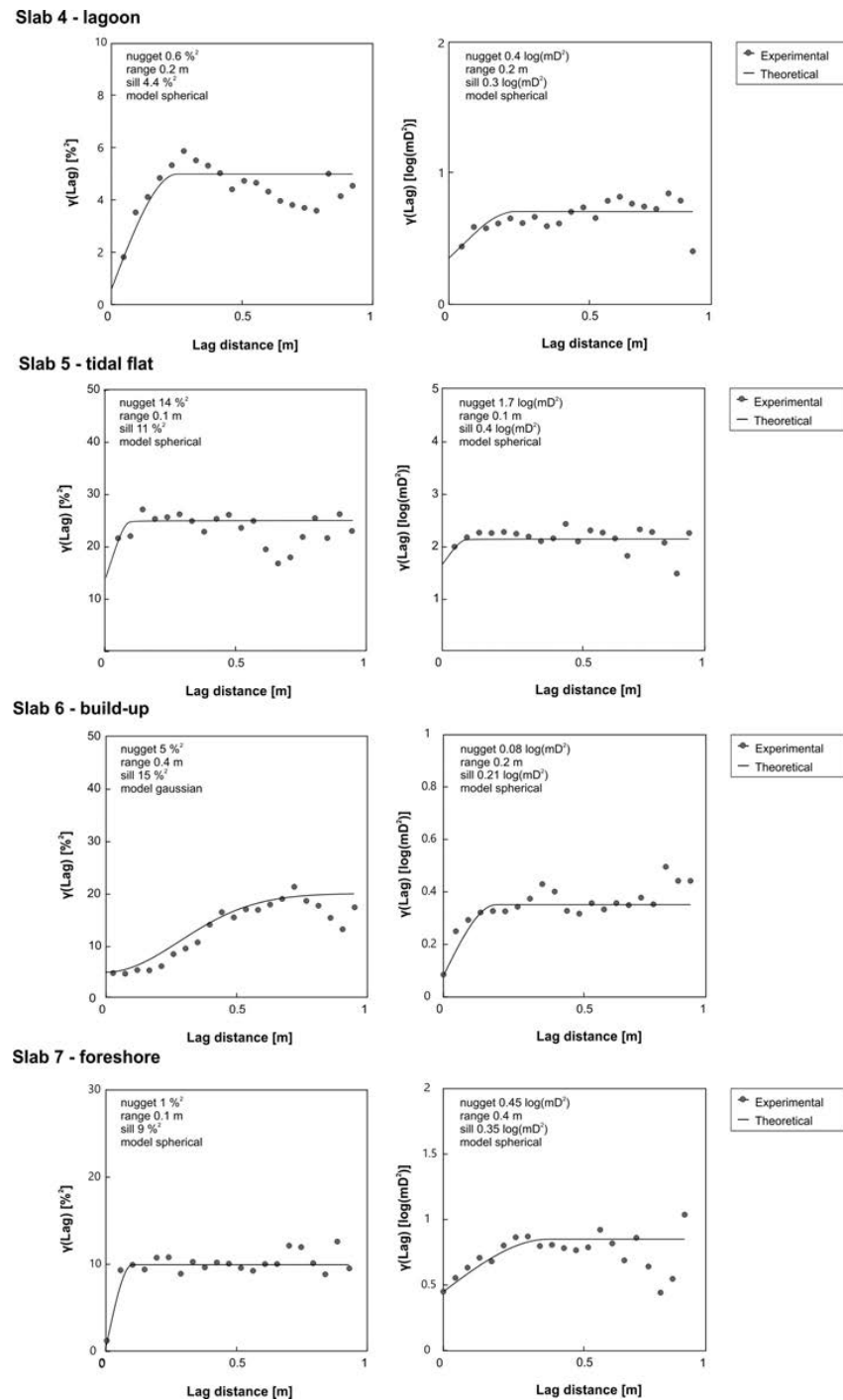
FIGURE 21

Semi-variograms of porosity (left) and permeability (right) for samples from all carbonate subenvironments. They show strong differences. Corresponding lithofacies association appears very different to each other as well. Especially in slab 6 (shoal), porosity and permeability properties are decoupled (largely differing range distance). For details, see 'Spatial statistics of permeability and porosity'.
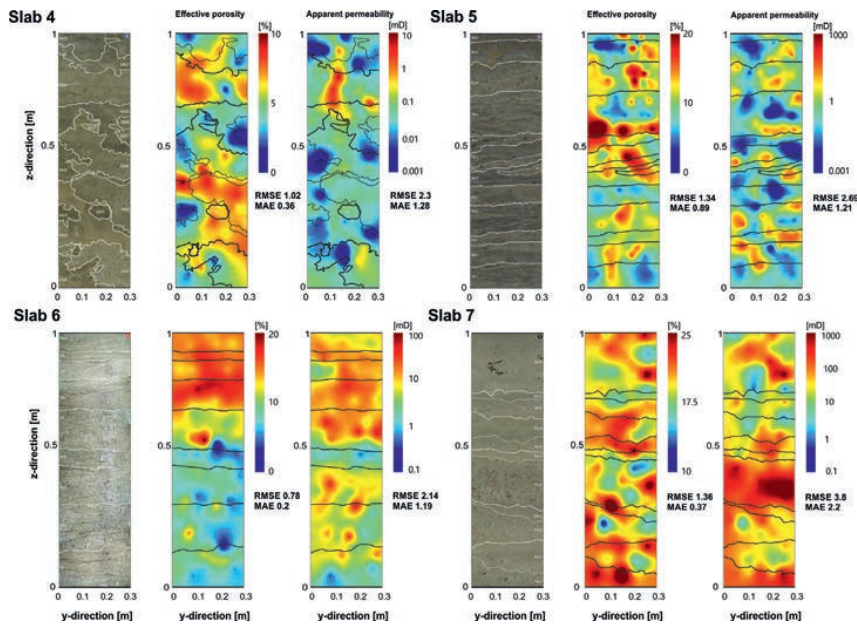
FIGURE 22

Interpolation of porosity (middle) and apparent permeability data (right) using simple kriging for all four slabs, respectively. At left, depositional structures and mapped lithofacies boundaries are shown. In general, interpolation algorithms based on the variogram analysis could simulate the measured data quite well. Interpolated readings show patchy changes within a lithofacies type and across lithofacies boundaries. However, some lithofacies boundaries correspond well to general changes of interpolated readings. Compare also to Figure 14 a-d. For details, see 'Interpolation'.

ate ramp system compared to the siliciclastic system: pore types are much more variable. Their distribution is further complicated by particle shape and size distribution. In addition, diagenetic processes like cementation, dissolution of (aragonitic) skeletal components or dolomitization weaken the dependence of petrophysical parameters to primary depositional structures especially to permeability on a grid block size. Subenvironments however show distinct signatures in petrophysical readings, controlled by hydrodynamics, grain composition and diagenetic alterations.

CONCLUSIONS

Some 1,000 data points and some 10,000 individual measurements from 7 rock slabs of 2 common depositional environments were analysed geologically and petrophysically. Analysis of the measurements revealed fundamental rules and relationships between geological and petrophysical properties:

1) Analysis revealed a hierarchical set of petrophysical heterogeneities spanning from foreset scale to lithofacies scale to subenvironment and to environment scale.
2) Primary depositional textures and the particle size/shape govern petrophysical properties in clastic depositional systems analysed rather than lithofacies.
3) A lithofacies type mimics the average of small-scale interaction of currents and sediment. It does not reflect small-scale processes like rapid variations in current

velocity and direction, turbulent versus laminar flow, and speed of deposition. Lithofacies scale does not allow for a precise characterization of petrophysical properties.
4) Subenvironment scale shows distinct petrophysical dependences and relationships of properties.
5) Petrophysical properties can best be represented as function of petrophysical rock types reflecting a combination of small-scale depositional processes (hydrodynamics), grain types and postdepositional diagenetic changes.
6) Petrophysical properties can be allocated in digital models using a multiscale geological framework to distribute properties three dimensionally.

ACKNOWLEDGEMENTS

REFERENCES

Aigner T. 1985. Storm depositional systems. Lecture Notes in Earth Sciences VIII, 174. DOI: 10.1007/BFb0011411.

Archie G.E. 1952. Classification of carbonate reservoir rocks and petrophysical considerations. AAPG Bulletin 36, 218–298.

Backhaus E. and Bähr R. 1987. Faziesmodelle für den Unteren Buntsandstein Südwestdeutschlands. Facies 17, 1-18.

Boansi L.O. 2015. Influence of porosity on electrical resistivity measurements of rocks at different saturation times: A laboratory method. Scientific Training Report. Technical University of Darmstadt, Dept. of Applied Geosciences, Germany (unpublished).

Choquette P.W. and Pray L.C. 1970. Geologic nomenclature and classification of porosity in sedimentary carbonates. AAPG Bulletin 54, 207–250.

Dunham R.J. 1962. Classification of carbonate rocks according to depositional textures. In: Classifications of carbonate rocks—a symposium (ed. W.E. Ham). AAPG Memoir 1, 108-121.

Einsele G. 2000. Sedimentary Basins: Evolution, Facies and Sediment Budget. Springer Verlag, Berlin, Heidelberg, New York, 795 S.

Filomena C.M., Hornung J. and Stollhofen H. 2014. Assessing accuracy of gas-driven permeability measurements: a comparative study of diverse Hassler-cell and probe permeameter devices. Solid Earth 5.1, 1-11.

Gumbert J. 2015. Zweidimensionale, sedimentologische und petrophysikalische Heterogenität von Lithofaziestypen des Buntsandsteins. Master thesis. Technical University of Darmstadt, Dept. of Applied Geosciences, Germany, unpublished.

Harms J.C. and Fahnestock R.K. 1965. Stratification, bed forms, and flow phenomena (with an example from the Rio Grande). Special Publication (Society of Economic Paleontologists and Mineralogists) 12, 84-115.

Hornung J. and Aigner T. 2002a. Reservoir architecture in a terminal alluvial plain: an outcrop analogue study (Upper triassic, southern Germany). Part 1: Sedimentology and petrophysics. Journal of Petroleum Geology 25, 3-20.

Hornung J. and Hinderer M. 2011. Depositional dynamics and preservation potential in a fluviodeltaic setting: implications for high-resolution sequence stratigraphy (Triassic, China). In: From River to Rock Record: The Preservation of Fluvial Sediments and Their Subsequent Interpretation (eds. Davidson et al.). SEPM Special Publication No. 97, 281-310.

Jennings J.W. Jr. 2000. Spatial Statistics of Permeability Data from Carbonate Outcrops of West Texas and New Mexico: Implications for Improved Reservoir Modeling. Report of Investigations No. 258. The University of Texas at Austin, Bureau of Economic Geology, 50 pp.

Jennings J. and Lucia F. 2003. Predicting permeability from well logs in carbonates with a link to geology for interwell permeability mapping. SPE Reservoir Evaluation & Engineering 6, 215-225. DOI: 10.2118/84942-PA

Klinkenberg L.J. 1941. The Permeability of Porous Media to Liquids and Gases. API-41-200. American Petroleum Institute.

Linsel A., Greb M.D., Bär K., Hornung J. and Hinderer M. 2018. Bridging information gaps in reservoir studies—archiving and providing information for reservoir studies (APIRS). AAPG Annual Convention and Exhibition, Salt Lake City, Utah, May 20-23, 2018. DOI: 10.1306/42274Linsel2018

Miall A.D. 2014. Fluvial depositional systems. Springer Geology. Springer, Cham. ISBN 978-3-319-00665-9. DOI: 10.1007/978-3-319-00666-6

Ölmez J.A. 2018. Kleinskalige 3D-Anisotropie und faziesabhängige Heterogenität der Permeabilität von bioklastischen Kalken (Muschelkalk, Würzburg). Bachelor thesis, Technical University of Darmstadt, Dept. of Applied Geosciences, Germany, unpublished.

Pettijohn F.J., Potter P.E. and Siever R. 1987. Sand and Sandstone. 553 S., 2. Auflage. Springer Verlag, New York.

Pittman E.D. 1979. Porosity, diagenesis and productive capability of sandstone reservoirs. In: Aspects of Diagenesis (eds. Scholle P.A. and P.R. Schluger). SEPM Spec. Publ.. DOI: 10.2110/pec.79.26.0159.

Reineck and Singh. 1980. Depositional Sedimentary Environments. With Reference to Terrigenous Clastics. Springer. ISBN-13: 978-3-540-10189-5. DOI: 10.1007/978-3-642-81498-3

Scheid M. 2016. Kleinskalige 3D-Anisotropie und Faziesabhängige Heterogenität der Sedimentpermeabilität einer fluviatilen Rinne (Buntsandstein, Miltenberg). Bachelor thesis, Technical University of Darmstadt, Dept. of Applied Geosciences, Germany, unpublished.

Schröder D. 2015. Charakterisierung der zweidimensionalen Heterogenität von Lithofaziestypen des Muschelkalks (Trias) in Bezug auf Ultraschall-Wellengeschwindigkeiten, Porosität, Permeabilität und elektrischem Widerstand. Master thesis, Technical University of Darmstadt, Dept. of Applied Geosciences, Germany, unpublished.

Wentworth C.K. 1922. A scale of grade and class terms for clastic sediments. The Journal of Geology 30, 377-392.

Wackernagel H. 2003. Multivariate Geostatistics. 3rd ed. Springer-Verlag Berlin, Heidelberg, 388 pp.

Weber K.J. 1986. How heterogeneity affects oil recovery. In: Reservoir Characterization (eds. L.W. Lake, and H.B. Carroll, Jr.), pp. 487-544. Academic Press Inc., Orlando, Florida, USA.

3.3. Publication III – High-Resolution Analysis of the Physicochemical Characteristics of Sandstone Media at the Lithofacies Scale

3.3.1. Introduction

Publication III provides an innovative methodological approach to characterize the multidimensional physicochemical characteristics of geological media at the lithofacies scale with the help of the software system GeoReVi (**Publication I**). Therefore, the quasi-continuous scalar field for numerous physicochemical rock properties of two rock cubes consisting of sandstone and measuring 0.25 m and 0.2 m were modeled through geostatistical interpolations. The interpolations were constrained by over 1,000 laboratory measurements performed on the cubes' surfaces and on 108 cylinder samples taken from the cubes.

We were able to show that the physicochemical fields show multifarious patterns. Properties such as intrinsic permeability and the Fe-oxide fraction are positively correlated and reflect the primary depositional characteristics expressed through bedding structures.

The spatial characteristics of the K-oxide and Al-oxide fractions, that are positively correlated, are related to meso- to telogenetical processes that have been probably produced by diffusive mass transfer.

Moreover, we were able to show that visible structures may not be indicative for a higher degree of anisotropy in sandstone media. This contribution gives rise to the question whether the lithofacies concept may be too inaccurate for using the lithofacies classes as co-variable for petrophysical property prediction.

3.3.2. Author Contributions

- As first author, I developed the methodological concept, prepared the figures and wrote the manuscript.

- Together with S. Wiesler, I performed the sampling, conducted the field and laboratory measurements and the data analysis.

- J. Hornung contributed to the methodological concept.

- M. Hinderer supervised my research.

- All co-authors contributed to the preparation and revision of the manuscript.

3.3.3. Publication

Published as:

Linsel, A., Wiesler, S., Hornung, J. and Hinderer, M. (2020): High-Resolution Analysis of the Physicochemical Characteristics of 3-D Sandstone Media at the Lithofacies Scale. *Solid Earth*. doi:10.5194/se-2020-13

Solid Earth

High-resolution analysis of the physicochemical characteristics of sandstone media at the lithofacies scale

Adrian Linsel, Sebastian Wiesler, Jens Hornung, and Matthias Hinderer

Institute of Applied Geosciences, Technische Universität Darmstadt, 64287 Darmstadt, Germany

Correspondence: Adrian Linsel (linsel@geo.tu-darmstadt.de)

Abstract. The prediction of physicochemical rock properties in subsurface models regularly suffers from uncertainty observed at the submeter scale. Although at this scale – which is commonly termed the lithofacies scale – the physicochemical variability plays a critical role for various types of subsurface utilization, its dependence on syndepositional and postdepositional processes is still subject to investigation.

The impact of syndepositional and postdepositional geological processes, including depositional dynamics, diagenetic compaction and chemical mass transfer, onto the spatial distribution of physicochemical properties in siliciclastic media at the lithofacies scale is investigated in this study. We propose a new workflow using two cubic rock samples where eight representative geochemical, thermophysical, elastic and hydraulic properties are measured on the cubes' faces and on samples taken from the inside. The scalar fields of the properties are then constructed by means of spatial interpolation. The rock cubes represent the structurally most homogeneous and most heterogeneous lithofacies types observed in a Permian lacustrine-deltaic formation that deposited in an intermontane basin. The spatiotemporal controlling factors are identified by exploratory data analysis and geostatistical modeling in combination with thin section and environmental scanning electron microscopy analyses.

Sedimentary structures are well preserved in the spatial patterns of the negatively correlated permeability and mass fraction of Fe_2O_3. The Fe-rich mud fraction, which builds large amounts of the intergranular rock matrix and of the pseudomatrix, has a degrading effect on the hydraulic properties. This relationship is underlined by a zonal anisotropy that is connected to the observed stratification. Feldspar alteration produced secondary pore space that is filled with authigenic products, including illite, kaolinite and opaque phases. The

local enrichment of clay minerals implies a nonpervasive alteration process that is expressed by the network-like spatial patterns of the positively correlated mass fractions of Al_2O_3 and K_2O. Those patterns are spatially decoupled from primary sedimentary structures. The elastic properties, namely P-wave and S-wave velocity, indicate a weak anisotropy that is not strictly perpendicularly oriented to the sedimentary structures.

The multifarious patterns observed in this study emphasize the importance of high-resolution sampling in order to properly model the variability present in a lithofacies-scale system. Following this, the physicochemical variability observed at the lithofacies scale might nearly cover the global variability in a formation. Hence, if the local variability is not considered in full-field projects – where the sampling density is usually low – statistical correlations and, thus, conclusions about causal relationships among physicochemical properties might be feigned inadvertently.

1 Introduction

The utilization of the subsurface in disciplines such as petroleum reservoir engineering, geothermal heat extraction, mining, carbon capture and storage or nuclear waste disposal requires highly accurate spatial predictions of relevant physical or geochemical properties in order to assess the economic feasibility of a target region (Landa and Strebelle, 2002; Heap et al., 2017; Kushnir et al., 2018; Rodrigo-Ilarri et al., 2017). Although most of these types of utilizations take place at full-field scales, geological variability present at the submeter scale may play an important role during the development process. The scale we are speaking of is commonly

termed the lithofacies scale (Miall, 1985). Geological heterogeneities at the lithofacies scale might constitute undesirable features in the subsurface such as flow barriers in reservoirs (Landa and Strebelle, 2002; Ringrose et al., 1993; Medici et al., 2016, 2019), pathways in radionuclide repository sites (Kiryukhin et al., 2008) and in contaminated sites (Tellam and Barker, 2006) or geochemical anomalies in mining areas (Wang and Zuo, 2018). Hence, the controlling factors of sub-meter variability should be understood and at least roughly quantified before starting the development in the subsurface region.

In sedimentary bodies, the spatial distribution of the properties is mainly controlled by depositional and diagenetic processes (McKinley et al., 2011, 2013). The spatial characteristics of physicochemical properties in sedimentary rock media are complex due to strongly intersecting and interacting processes during sediment transport, deposition and diagenesis (McKinley et al., 2011). Multiple studies aimed to quantify the variability at the lithofacies scale, most of which concentrated on reservoir properties such as permeability and porosity in 2D spaces (McKinley et al., 2011; Hornung et al., 2020). A 2D analysis is well suited to identifying non-visible patterns related to microbedding structures at multiple scales even in very homogeneous sandstones (McKinley et al., 2004). That perspective, however, involves simplifications of the physicochemical variability in 3D spaces since specific rock properties such as permeability are dependent on the Cartesian direction. Also, consideration of geostatistical parameters such as variographic direction, range, sill and nugget revealed differences in 3D compared to 2D spaces (Landa and Strebelle, 2002; Hurst and Rosvoll, 1991).

With proper knowledge of the statistical and causal relationships among physicochemical rock properties at different scales, prognostic property models can be significantly enhanced by the integration of small-scale uncertainty into upscaling or conditional simulation algorithms (Lake and Srinivasan, 2004; Verly, 1993). Especially since multivariate geostatistics can account for interrelationships between rock properties, those relationships can be used as trends or drifts in geostatistical predictions in order to optimize their accuracy in space and time (Hudson and Wackernagel, 1994).

In order to quantify the spatial variability and the multidimensional relationships among physicochemical properties at the 3D lithofacies scale, the quasi-continuous scalar fields of two rock cubes are modeled by means of spatial interpolation, which is constrained by laboratory measurements. The rock cubes have volumes of 0.0156 and 0.008 m³ and have been sampled from a Permian lacustrine-deltaic sandstone formation that deposited in the intermontane Saar–Nahe basin during the Cisuralian series.

The lithological characteristics of the sandstones are analyzed, and both isotropic and anisotropic properties, including bulk rock geochemistry, thermophysical, hydraulic and elastic rock properties, are measured on the cubes' faces. In addition, the intrinsic gas permeability under an infinite pressure gradient, the effective porosity, the elemental composition, the thermal conductivity, and the thermal diffusivity together with the P-wave and S-wave velocity are measured on 108 rock cylinders taken from the inside of the cubes representative for each Cartesian direction in order to account for anisotropic patterns.

The measurements are used to interpolate the full 3D field of each property. Prior to interpolation, the discrete measurements are statistically analyzed for correlation and formal relationships. Interpolations are conducted using deterministic and geostatistical methods, including the inverse distance weighting (IDW) and simple kriging (SK) interpolation. The models are evaluated through cross validation, and the observed spatial patterns are categorized. The interpolation results providing the lowest cross validation error are statistically analyzed again and compared with the aforementioned statistical patterns. Eventually, the geological processes, which produced the observed patterns, are interpreted and discussed with the help of qualitative thin section and environmental scanning electron microscope (ESEM) analyses.

The research outputs of this study lie between the scale of a core plug measurement and a wireline log and/or pumping test (Medici et al., 2018). Hence, we aim to contribute to estimating the uncertainty that must be accounted for when performing up- or downscaling between those two scales of investigation (Zheng et al., 2000; Jackson et al., 2003; Corbett et al., 2012; Hamdi et al., 2014).

2 Methodology

2.1 Sedimentological characterization and rock sampling

In order to cover multiple varieties of sedimentary lithofacies types, a quarry in Obersulzbach (Rhineland-Palatinate, Germany) in the Saar–Nahe basin was selected for the investigations (Fig. 1). The quarry belongs to the lacustrine-deltaic Disibodenberg Formation that is assigned to the inner Variscan Rotliegend group and comprises four lithofacies types. This formation is deeply buried (1995 to 2380 m below ground surface) in the northern Upper Rhine Graben in southwestern Germany (Becker et al., 2012) and constitutes a potential target unit for hydrothermal exploitation (Aretz et al., 2015). The maximum past overburden of the field site can be estimated to be between 1950 and 2400 m, as indicated by shale-compaction analyses which were performed by Henk (1992). The outcrop has been chosen in order to estimate the variability of physicochemical properties that could be expected in this formation as an uncertainty factor if it is targeted in a deep geothermal project.

Two rock cubes of $0.2 \times 0.2 \times 0.2$ m (OSB2_c) and $0.25 \times 0.25 \times 0.25$ m (OSB1_c) were extracted from the outcrop wall using a rock chainsaw. According to the outcrop's coordinate system, one edge of the cuboid runs east–west (x),
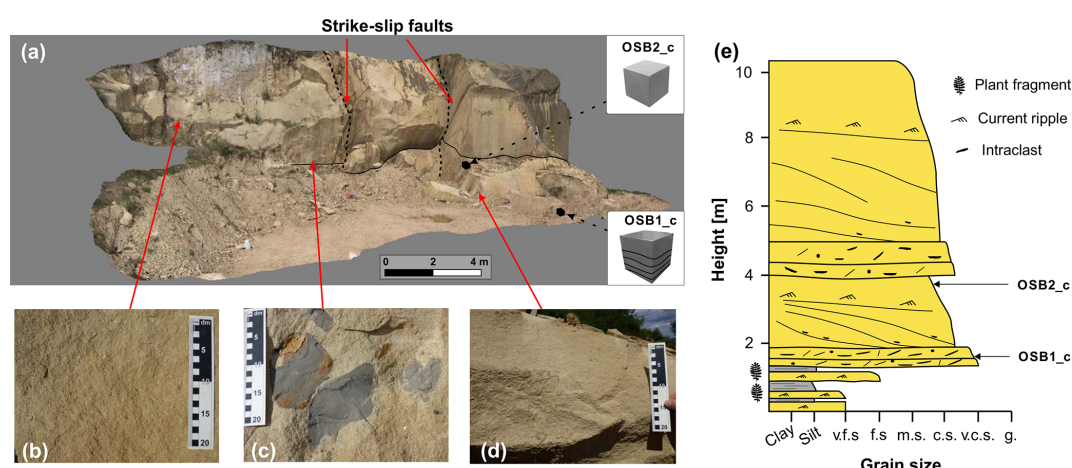
Figure 1. (a) The investigated sandstone quarry in Obersulzbach, Germany. The outcrop is compartmentalized in the central part by two strike-slip faults which belong to the Lauter fault zone (Stollhofen, 1998). The strike-slip faults provide offsets of a few meters. **(b)** Massive sandstone. **(c)** Pelitic rip-up clasts embedded in a massive rock matrix. **(d)** Ripple-cross bedded sandstone. **(e)** Cumulative sedimentary log of the outcrop architecture. The sampling positions of OSB1_c and OSB2_c are marked in the sedimentary log (v.f.s. – very fine sand; f.s. – fine sand; m.s. – medium sand; c.s. – coarse sand; v.c.s. – very coarse sand; and g. – granule).

one north–south (y) and one in an altitudinal (z) direction. The irregular cuboids were reworked to regular cubes with a stationary rock saw. We selected two types of lithofacies (Fig. 1e) – both sandstones – with one representing a heterogeneous, compartmentalized variety (OSB1_c) and the other one being a homogeneous variety (OSB2_c). The cubes were both extracted from a distributary mouth bar that is building a foreset in a fluviatile-dominated lacustrine delta. OSB1_c (Fig. 2) was taken from the high energetic basal part, whereas OSB2_c was taken from the lower energetic top. The sedimentological characteristics, including grain size, sorting, angularity, sedimentary structures and mineral content, were determined by visual inspection, thin section and ESEM analyses. Two different types of zonal anisotropy and spatial patterns were expected to be found with the aforementioned sampling strategy. In other studies, such as McKinley et al. (2011), measurements were directly conducted in the field. This approach, however, often provides a drawback with the accuracy and precision, especially in permeability measurements. In order to address this issue, we performed analyses on the faces of the cubes under laboratory conditions. In the next step, the cubes were cut into rock slabs from which cylinder samples were extracted. In total, 108 rock cylinders – 79 from OSB1_c and 29 from OSB2_c – were extracted from the rock cubes. It was ensured that at least five samples were produced that were representative for each Cartesian direction. By applying the formula for calculating a cylinder's volume V_c with the following:

$$V_c = h \times \pi \times r^2, \tag{1}$$

where h is the height of the cylinder and r the radius, the relative volume covered by the rock cylinders in the rock cubes was calculated to be 25.4 % for OSB1_c and 18.2 % for
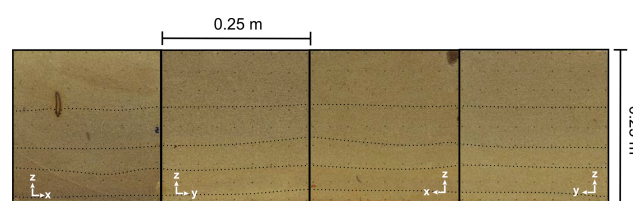
Figure 2. Lateral faces of OSB1_c displayed in the form of an open cube (from left to right: XZ front, YZ front, XZ back and YZ back). The internal bounding surfaces are indicated by the dotted lines.

OSB2_c, respectively. Eventually, target meshes are needed to interpolate the full 3D scalar fields. Therefore, both cubes were modeled in 3D using a regular grid consisting of 27 000 hexahedral orthogonal cells. The elementary cell of OSB1_c has a volume of $5.7 \times 10^{-7} \, \text{m}^3$, whereas OSB2_c's elementary cells have a volume of $3 \times 10^{-7} \, \text{m}^3$.

2.2 Laboratory experiments

First, a local metric coordinate system was defined, where each edge of the cube represents an axis in the Cartesian coordinate system in order to reference each measurement to a point in space. The sampling points were set in a raster of 9×9 points on each face for OSB1_c and 5×5 for each face of OSB2_c. All measurements were conducted in the laboratory of the Institute of Applied Geosciences in Darmstadt, Germany. After drying the rock cubes at 60 °C, noninvasive measurements were conducted on each face of the cube. On the cubes' faces, the P-wave and S-wave velocity and elemental mass fractions were determined (Fig. 3).

(a)   **(b)**
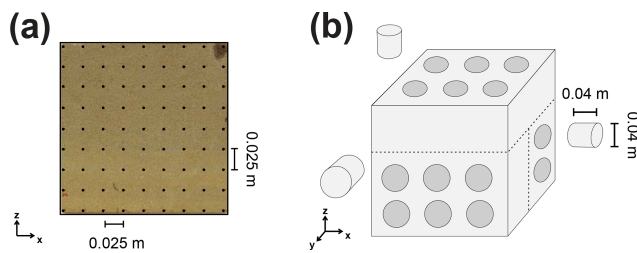
Figure 3. (a) Sampling locations for the noninvasive measurements, including P-wave and S-wave velocity and X-ray fluorescence, displayed for the face XZ back of OSB1_c. **(b)** Schematic of the extraction strategy for sampling the rock cylinders.

After the extraction, the rock cylinders were ovendried at 105 °C and measured in order to determine the intrinsic gas permeability, effective porosity, P-wave and S-wave velocity, elemental mass fractions, thermal conductivity, and the thermal diffusivity in unsaturated conditions. Those properties can be considered key properties of the rock matrix in porous aquifers with regard to hydrothermal systems (Agemar et al., 2014) since they constitute input variables for the governing equations for heat transfer and fluid flow in the subsurface (Carslaw and Jaeger, 1959).

The permeability was measured with the Hassler cell permeameter, which is described in Filomena et al. (2014). The Hassler cell is a gas-driven permeameter which measures the permeability of a cylinder-shaped rock sample under steady-state gas flow. This technique allows for the estimation of the intrinsic gas permeability, which is the permeability at an infinite pressure gradient. The permeameter was set to accept a measurement if 15 consecutive readings did not deviate by more than 5 %. The measurement error, however, can exceed that value, especially in low permeable lithologies. Effective porosity measurements were conducted using an envelope density analyzer (GeoPyc 1360). The accuracy is given by the manufacturer to be within ±0.55 % (Micromeritics, 1998). Thermal properties under unsaturated conditions, namely the thermal conductivity and thermal diffusivity, were determined with a thermal conductivity scanner (TCS) according to the work of Popov et al. (1999). The measurement error is quantified to be ≤ 3 % for thermal conductivity and ≤ 8 % for thermal diffusivity (Popov et al., 1999). The elastic properties of P-wave and S-wave velocity in the rock media were measured with the sonic wave generator UKS-D (Geotron Elektronik) by sending a sonic wave pulse from a pulse-providing test head (UPG-S) to a receiver (UPG-E). The wave velocity is a function of the travel length and time, together with the density of the material. The initial occurrence of the P wave or S wave must be picked manually after visual inspection by the operator. Thus no measurement error can be provided since user bias cannot be assessed quantitatively. Bulk elemental analysis using the Bruker S1 TITAN handheld portable X-ray fluorescent (pXRF) analyzer was used to find correlations between the

elemental composition and the petrophysical properties. The measurement device works on the basis of energy-dispersive X-ray fluorescence (EDXRF) and estimates the elemental mass fractions of a sample. This device produces an ionizing X-ray beam with a diameter of 1.2 cm and quantifies the elemental composition based on the energy emitted by the ionized elements in the targeted area. The portable device can measure the fraction of elements with an ordinal number ≥ 12 and ≤ 235 if the threshold value, defined by the measurement error for the specific element in the sample, is exceeded. For this study, the device was operated in GeoChem, Dual Mining mode, allowing for the detection of the major oxides of SiO_2, Al_2O_3, Fe_2O_3 and K_2O and a wide range of other elements. The device was calibrated with international standards. We used the previously mentioned major oxides for analyses since those can provide insight into the iron oxide and clay mineral distribution, which can significantly impact the petrophysical properties. More details on the measurement devices can be found in the works of Hornung and Aigner (2002), Sass and Götz (2012), Filomena et al. (2014), and Aretz et al. (2015).

2.3 Data analysis and spatial modeling

2.3.1 Variography

The experimental semivariogram represents the cumulative dissimilarity of a discrete set of point pairs x, with n_c as the count of point pairs within the distance classes $\boldsymbol{h}$ of identical distance increments (Eq. 2).

$$\gamma(\boldsymbol{h}) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} \left(z(x_\alpha + \boldsymbol{h}) - z(x_\alpha) \right)^2. \tag{2}$$

The continuous counterpart, represented by the variogram model, is an approximation of the experimental semivariogram that assumes $z(\boldsymbol{x})$ to be a stationary random field (Wackernagel, 2003). A variogram model γ_{theo} is represented by a covariance function c, with the relationship $\gamma_{\text{theo}}(\boldsymbol{h}) = c(0) - c(\boldsymbol{h})$, where c is a positive definite even function. Six covariance models are mostly used to fit the experimental semivariogram, namely the spherical, gaussian, exponential, power, cardinal sine and the linear model (Armstrong, 1998; Ringrose and Bentley, 2015). In this study, we only observe spherical relationships with a nugget effect. This model is calculated as follows:

$$c_{\text{sph}}(\boldsymbol{h}) = \begin{cases} n + b \cdot \left(1 - \frac{3|\boldsymbol{h}|}{2a} + \frac{|\boldsymbol{h}|^3}{2a^3}\right) & \text{for } 0 \leq |\boldsymbol{h}| < a \\ n & \text{for } |\boldsymbol{h}| \geq a, \end{cases} \tag{3}$$

with the variables nugget (n), range (a) and sill (b). Semivariograms can be used to quantify the spatial or time correlation of a random property (Ringrose and Bentley, 2015; Gu et al., 2017; Rühaak et al., 2015). Further on, the differences in range and sill in dissimilar directional semivariograms can quantify the zonal and geometric anisotropy of a property

(Ringrose and Bentley, 2015). The resulting covariance function is an input variable for geostatistical interpolation algorithms.

2.3.2 Rock property interpolation

Spatial inter- and extrapolation can be generated with deterministic and geostatistical techniques. All interpolations are based on the assumption that a point x_k with a known value $z(x_k)$ has a weight on a discrete point x_0 in space with an unknown value $z(x_0)$. The global known points, however, can be reduced to a local neighborhood of x_0.

For deterministic interpolation, the p value inverse distance weighting (IDW; Shepard, 1968) interpolation is used. The IDW interpolation generally calculates an unknown value $z(x_0)$ at point x_0 by weighting the distance of that point to each known value point (x_k) in space. The underlying formula for IDW is as follows:

$$z(x_0) = \frac{\sum_{k=1}^{n} z(x_k)/d_k^p}{\sum_{k=1}^{n} 1/d_k^p}, \tag{4}$$

where d is the Euclidean distance between the point with the known value x_k and the point with the unknown value x_0, and p is an exponent factor to bias the weights nonlinearly. The p value is mostly used for smoothing the results by controlling the distance decay effect (Lu and Wong, 2008). IDW is a reliable and widely applied method to interpolate static rock properties in a 1D to 3D space (Rühaak, 2006).

For geostatistical interpolation, simple kriging (SK) is used. Kriging in general is a popular technique for interpolating geological properties in space (Goovaerts, 1997; Rühaak, 2015; Malvić et al., 2019). Through kriging, the value $z(x_0)$ at an unknown point x_0 is calculated by weighting the neighboring known values and building a linear combination of those via the following formula:

$$z(x_0) = \sum_{k=1}^{n} w_k \cdot z(x_k), \tag{5}$$

where w_k is the weight of the known point x_k with the value $z(x_k)$. SK requires knowledge of the stationary mean μ (Deutsch and Journel, 1998), which modifies Eq. (5) into the following:

$$z(x_0)_{\text{SK}} = \sum_{k=1}^{n} w_k \cdot z(x_k) + \left(1 - \sum_{k=1}^{n} w_k\right) \cdot \mu. \tag{6}$$

To obtain the simple kriging weights, a set of n equations has to be solved. This set of equations can be written as follows:

$$\begin{pmatrix} c(x_1 - x_1) & \cdots & c(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ c(x_n - x_1) & \cdots & c(x_n - x_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$
$$= \begin{pmatrix} c(x_1 - x_0) \\ \vdots \\ c(x_n - x_0) \end{pmatrix}, \tag{7}$$

with c as the covariance function and x_n as the point with a known value (Wackernagel, 2003). The quality of kriging interpolation is dependent on the variogram model, defined neighborhood, sampling density and goodness of fit to the experimental values.

2.4 Cross validation

Cross validation can be used to assess the quality of a model. During cross validation, p randomly selected samples are removed from the input data set of size n with $0 < p < n$, and the interpolation is performed without those samples (Celisse, 2014). The measures of goodness of fit being used in this study include the root mean square error (RMSE) as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(\hat{z}(x_k) - z(x_k)\right)^2}, \tag{8}$$

and the mean absolute error (MAE) as follows:

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^{n} |\hat{z}(x_k) - z(x_k)|, \tag{9}$$

with $\hat{z}(x_k)$ as estimated value at point x_k. Those parameters allow for the quantitative assessment of an interpolation's quality. They might be prone to bias if the sampling density in the target domain is extremely scarce.

2.4.1 Anisotropy

Anisotropy describes the dependence of a physical property on a direction. Rock properties such as stiffness, permeability or thermal conductivity are anisotropic in most cases. Hence, measurements of those properties might show differing magnitudes in different directions if the medium is polar anisotropic. The intrinsic permeability, for example, provides typical ranges for the ratio between the vertical (k_v) and horizontal permeability (k_h) of 10^{-5} to 1 (Ringrose and Bentley, 2015). Anisotropy in geological media is generated by the preferred orientation of mineral grains or cracks and by the intrinsic anisotropy of single crystals (Thomsen, 1986).

In the following, we will provide an exemplary description of the anisotropy of elasticity, and we will provide measures for anisotropy quantification under the simplifying assumption of transverse isotropy. The elastic modulus tensor can be

expressed as a fourth rank tensor as follows:

$$
\mathbf{C} = \begin{pmatrix}
C_{11} & C_{11}-2C_{66} & C_{13} & 0 & 0 & 0 \\
C_{11}-2C_{66} & C_{11} & C_{13} & 0 & 0 & 0 \\
C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\
0 & 0 & 0 & C_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & C_{44} & 0 \\
0 & 0 & 0 & 0 & 0 & C_{66}
\end{pmatrix},
$$

(10)

where C_{ij} represents an elasticity modulus and the indices are related to the directional P-wave and S-wave velocity, under the assumption that z is the symmetry axis. The velocities can be calculated by the following:

$$
v_{\mathrm{p}}^{z} = \sqrt{\frac{C_{33}}{\rho}}
$$

(11)

$$
v_{\mathrm{s}}^{z} = \sqrt{\frac{C_{44}}{\rho}},
$$

(12)

where v_{p} is the P-wave velocity and v_{s} is the S-wave velocity parallel to the symmetry axis and ρ is the bulk density (Yang et al., 2020). The anisotropy, here exemplarily expressed for the P-wave polar anisotropy, can be quantified with the Thomsen parameters (Thomsen, 1986). For example, ϵ can be expressed as follows:

$$
\epsilon = \frac{C_{11}-C_{33}}{2C_{33}}.
$$

(13)

If $\epsilon \ll 1$, the material can be classified as weakly anisotropic.

2.4.2 Correlation and regression analysis

In order to quantify the linear statistical relationship between two independent variables x and y, the Pearson linear product–moment correlation coefficient (R) can be used. R is expressed as follows:

$$
R = \frac{\sum\limits_{k=1}^{n}(x_k-\overline{x})(y_k-\overline{y})}{\left(\sum\limits_{k=1}^{n}x_k^2-n\cdot\overline{x}^2\right)\left(\sum\limits_{k=1}^{n}y_k^2-n\cdot\overline{y}^2\right)},
$$

(14)

with n representing the number of compared point pairs and $\overline{x}$ and $\overline{y}$ standing for the arithmetic mean of x and y.

Regression aims at finding a fitting function between samples of two or more random variables. For curvilinear regression, a function of a degree > 1 will be approximated for a discrete set of values. A second-degree polynomial function $f(x)$, for instance, would be described as follows:

$$
f(x) = b_0 + b_1 x + b_2 x^2.
$$

(15)

Thus, we would need to find $n+1$ regression coefficients, where n is the degree of $f(x)$. In general, the regression model yields the following:

$$
f(x)_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_n x_i^n,
$$

(16)

with $i = 1, 2, \ldots, n$. The regression coefficients b_m are obtained through solving a system of linear equations as follows:

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1^1 & \cdots & x_1^m \\ 1 & x_2^1 & \cdots & x_2^m \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_n^1 & \cdots & x_n^m \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix},
$$

(17)

where x and y are the samples. The function approximations, as produced in regression analyses, are commonly evaluated by the coefficient of determination (R^2), which is calculated through the following:

$$
R^2 = 1 - \frac{s_{\mathrm{res}}}{s_{\mathrm{tot}}} \in [0, 1],
$$

(18)

where

$$
s_{\mathrm{res}} = \sum_{k=1}^{n}(y_k - f(x)_k)^2
$$

(19)

is the explained sum of squares, and

$$
s_{\mathrm{tot}} = \sum_{k=1}^{n}(y_k - \overline{y})^2
$$

(20)

is the total sum of squares.

2.4.3 Spatial modeling and statistical analyses

The spatial dependence of the discrete values is evaluated through experimental semivariograms. The semivariograms are generated for the single rock faces, where measurements are available, and for the plug measurements. The empirical semivariogram is fitted to a variogram model, which is then used for the geostatistical interpolation. Interpolation analyses are performed as IDW and SK realizations (Fig. 4) that are assessed through cross validation. The power parameter for IDW is chosen to be three since this constant provides the lowest RMSE among the realizations. The search radii for each prediction is chosen to be 0.2 m, in x and y direction, and 0.15 m, in z direction, in OSB1_c to account for the sedimentary structures. For OSB2_c, the search radii are chosen to be isotropic with a length of 0.2 m. To make the methods comparable, we selected the maximum number of neighboring points to be 25 to represent between 5 % and 95 % of the measurements.

We decided to waive sequential simulation as large amounts of the cubes' volumes are covered by rock samples. Thus, we do not expect a relevant kriging variance. With this in mind, the simulations are assumed to capture most of the total variance from the measurements themselves. The interpolation results that provide the lowest cross validation error are used for statistical analyses in order to derive correlations
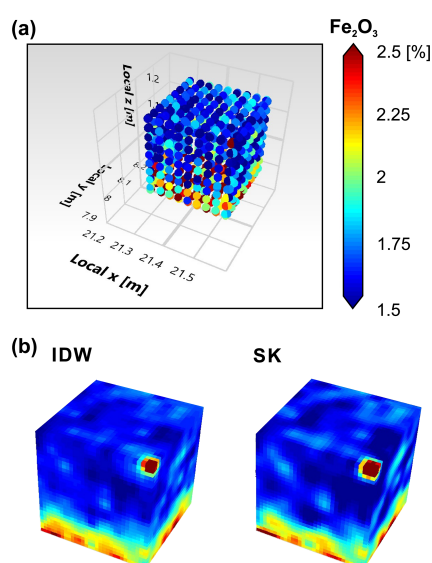
Figure 4. (a) Fe_2O_3 measurement locations on the cube faces of OSB1_c and on the rock samples extracted from the cube. The diameter of one point is $1.2\,cm$, which corresponds to the beam diameter of the pXRF measurement device. **(b)** Visual representation of the inverse distance weighting (IDW) and simple kriging (SK) realization of the 3D scalar field of Fe_2O_3, using the discrete points displayed in **(a)** as known data points.

and regression functions between the scalar fields. Eventually, significant correlations are compared with the noninterpolated data sets. Both the spatial modeling and the statistical analyses are performed with the open source software called Geological Reservoir Virtualization (GeoReVi; Linsel, 2020a). This software tool provides functionality for multidimensional subsurface characterization using the concept of knowledge discovery in databases, which is helpful when handling huge data sets such as those produced in this study.

3 Results

3.1 Sedimentological characteristics

The sandstones belong to a clinothem strata deposited in a fluvial-dominated lacustrine delta. More specifically, the architectural element represents a distributary mouth bar formed by rapid sandstone deposition in sheet-like bodies, as described in Fongngern et al. (2018). The base of those bodies is typically erosive, which is why muddy rip-up clasts commonly occur above the base. Also, the beds, which deposited after the intraclast-rich basal beds, typically show trough or ripple-cross stratification with set heights of 5–15 cm. The vertical orientation of rip-up clasts can be observed in matrix-rich debrites or turbidites deposited under high-energy turbulent hyperpycnal to homopycnal flow conditions (Li et al., 2017). Those are unconformably overlying lacustrine, laminated mud strata from the prodelta environ-
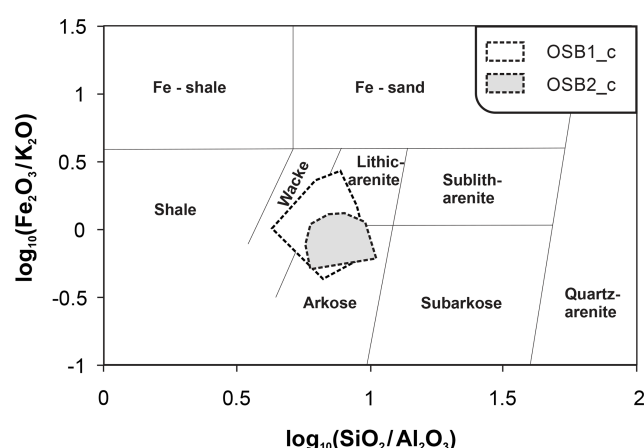
Figure 5. Petrographic classification after Herron (1988), based on the ratio of SiO_2 and Al_2O_3 and Fe_2O_3 and K_2O. The polygons show the convex hull for the measurements derived from the cubes' faces.

ment. Accordingly, Bouma A–E layers (Bouma, 1962; Middleton, 1993) with a prograding trend can be identified in the outcrop. With ongoing sedimentation, the depositional energy in a Bouma sequence typically decreases, which leads to massive sandstones. OSB1_c was taken from a basal bed of the Bouma A interval characterized by a high number of rip-up intraclasts, normal grading and subhorizontal pseudolayering, which may occur in a Bouma A interval if the rip-up clasts experienced buoyancy during transport. OSB2_c was taken from the topmost bed, which corresponds to a Bouma E interval that is characterized by a massive structure.

The average grain size in both cubes ranges from fine to very coarse sand (200–$1400\,\mu m$). While the grain size distribution in OSB2_c does not show a significant variability – mainly characterized by medium to coarse sand – a normal grading is observable in OSB1_c. Here, the grain size gradually decreases from very coarse sand at the base to medium sand at the top. Likewise, sorting increases from poor to moderate. In OSB2_c the sorting is moderate throughout the entire sample volume. The components provide a low to medium sphericity, while the grain shapes vary between subangular and subrounded. Locally, pelitic rip-up clasts occur with diameters of up to 4 cm. The rip-up clasts show a very low textural maturity and are subvertically oriented with respect to bedding.

The original rigid detrital components consist of 50 %–60 % quartz, 20 %–30 % strongly altered feldspar and 10 %–25 % lithic fragments. Mica grains are often bent between more rigid grains. The rock matrix accounts for approximately 10 %–20 % and is built up by detrital grains, coated by iron oxides, ductile, autochthonous pelite grains and fine-grained quartz. According to the geochemical analyses, the rocks can be classified as lithic arenites to arkoses or wackes (Fig. 5), respectively, if the matrix content exceeds 15 % based on the classification of Herron (1988).
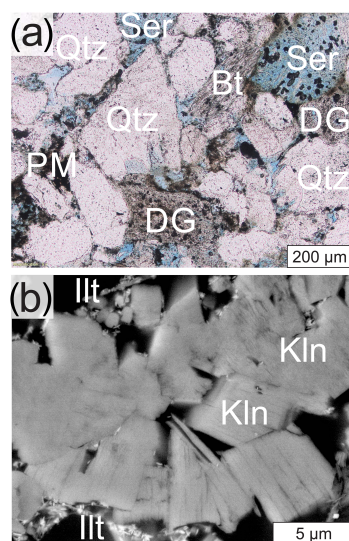
Figure 6. (a) Representative thin section taken from rock cube OSB2_c. The sandstone consists mainly of quartz (Qtz), altered feldspars with residual mineral aggregations (sericite – Ser), altered biotite (Bt) and ductile grains (DG). Feldspar dissolution lead to a high grade of secondary porosity (Molenaar et al., 2015), while the vast majority of the intergranular pore space is filled with primary and pseudomatrix (PM) which is rich in iron oxides. **(b)** Environmental scanning electron microscope (ESEM) image of the authigenic clay minerals, mainly kaolinite (Kln) and illite (Ilt), built in the pore space. Mineral abbreviations were taken from Whitney (2010).

Thin section analysis (Fig. 6a) reveals that most of the pore space is secondary due to grain dissolution. The secondary pores are undeformed, indicating that grain dissolution took place during structural inversion – probably during telogenesis, according to the concept of Worden and Burley (2003). Most of the primary intergranular volume was destroyed during mechanical compaction. ESEM analysis (Fig. 6b) confirms the presence of quartz accompanied by coprecipitated calcite, opaque phases – mainly iron oxides – and authigenic clay minerals including kaolinite and illite in the cement fraction. Thus, chemical compaction had taken place due to iron oxide, quartz and clay mineral precipitation during diagenesis. Here, the earliest cement phase is represented by the opaque phases comprising a high number of iron oxides. Thereafter, kaolinite is formed, mainly in the secondary pore space, and overgrown by illite. Often, the early cement is overgrown syntaxially by quartz. The source of SiO_2 might be internal and related to feldspar dissolution.

3.2 Exploratory data analysis

In order to provide full comparability, the following section will provide an overview of the measurements derived from the rock cylinder analyses. For each property, 79 rock samples from OSB1_c and 29 from OSB2_c were investigated.
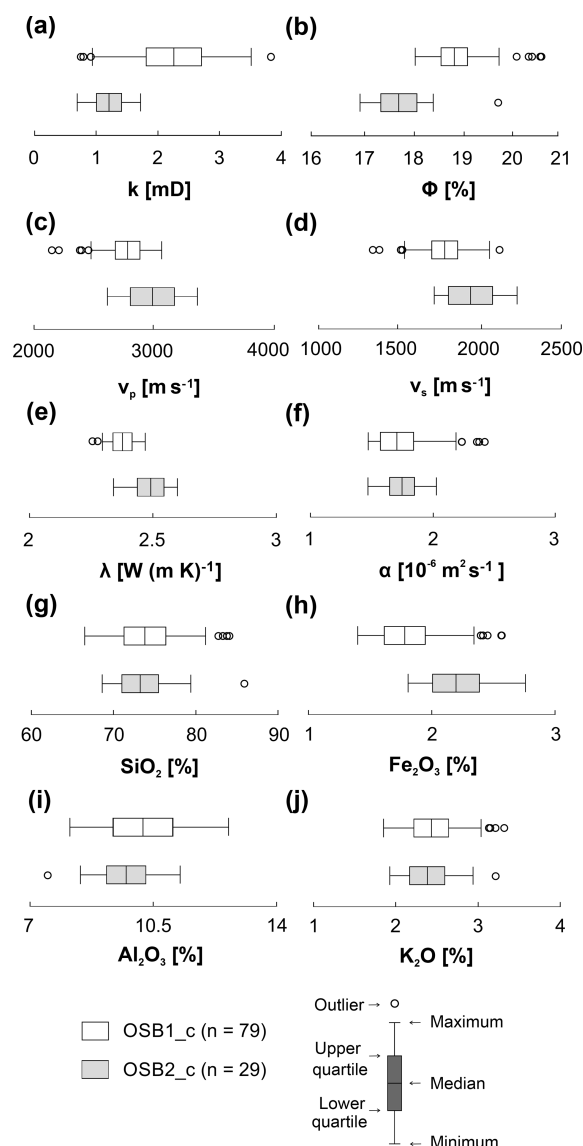
Figure 7. Box and whisker charts showing the empirical distribution of the rock properties measured in the rock cylinders taken from the rock cubes. Outliers were detected according to Tukey's method (Tukey, 1977), where a value is tested to be in the 1.5 times interquartile range of the arithmetic mean. **(a)** Intrinsic permeability – k. **(b)** Effective porosity – ϕ. **(c)** P-wave velocity – v_p. **(d)** S-wave velocity – v_s. **(e)** Thermal conductivity – λ. **(f)** Thermal diffusivity – α; the mass fraction of **(g)** silicon oxide – SiO_2. **(h)** Iron oxide – Fe_2O_3. **(i)** Aluminum oxide – Al_2O_3. **(j)** Potassium oxide – K_2O.

An overview of the rock properties' ranges is provided in the box and whisker charts shown in Fig. 7.

The local variability of OSB1_c is significantly higher than that of OSB2_c. The intrinsic permeability of OSB1_c provides a coefficient of variation of 0.3 and a Dykstra–Parsons coefficient of 0.4, while measurements from OSB2_c show values of 0.2 for the coefficient of variation and 0.18 for the Dykstra–Parsons coefficient, respec-

tively. According to the classification provided by Corbett and Jensen (1992), the intrinsic permeability of both rock cubes can be classified as being very homogeneous. Also, the intrinsic permeability does not show a significant anisotropy.

The range of values in OSB1_c for each property is greater than the range of those in OSB2_c. OSB1_c provides lower values in P-wave and S-wave velocity, thermal conductivity and mass fraction of Fe_2O_3 compared to OSB2_c. Intrinsic permeability and porosity, in turn, are greater. The mass fraction of silicon oxide and thermal diffusivity provides similar statistical parameters in both cubes; however, the ranges are marginally larger in OSB1_c. The measurements of the elastic rock properties revealed a weak anisotropy of the P-wave attenuation, especially in rock cube OSB2_c. The Thomsen parameter ϵ is 0.047 for OSB1_c and 0.096 for OSB2_c. It should be noted that OSB1_c provides visible bedding structures in contrast to OSB2_c; hence, the observed degree of anisotropy is not connectable to the bedding features in this case.

Statistically significant linear correlations (Fig. 8), in the sense of passing a two-tailed significance test at the 0.05 level, were found between porosity and permeability, permeability and Fe_2O_3, v_p and v_s, v_p and SiO_2, v_p and Al_2O_3, v_p and K_2O, Fe_2O_3 and SiO_2, and K_2O and Al_2O_3. The strongest positive linear correlation can be observed between v_p and v_s ($R = 0.88$), K_2O and Al_2O_3 ($R = 0.70$), and porosity and permeability ($R = 0.31$). The strongest negative correlation can be observed between permeability and Fe_2O_3 ($R = -0.56$). Properties not being mentioned do not provide significant statistical correlations to others.

3.3 Submeter-scale spatial correlation

The spatial dependence of the discrete measurements is estimated using experimental semivariograms. Therefore, the geochemical representatives SiO_2 (Fig. 9a) and Fe_2O_3 (Fig. 9b) that were measured on each of the rock faces of OSB1_c are given an exemplary analysis. The experimental semivariograms greatly vary from rock face to rock face in OSB1_c. The nugget effect for each experimental variogram is very low. The range of each semivariogram varies between 0.05 and 0.3 m. In the experimental semivariograms of SiO_2, two types of patterns can be identified. The XY base, XZ back and YZ front of the rock faces show ranges of approximately 0.08 m and a sill between $8\,\%^2$ and $10\,\%^2$, until the semivariance exponentially increases when exceeding a lag distance of 0.2 m. The semivariance on the other rock faces runs similarly, with ranges of 0.2 m and a sill of $4.7\,\%^2$. The semivariogram for Fe_2O_3 shows some similarities. Here, the XY base, YZ front and XZ front of the rock faces show very low ranges between 0.05 and 0.15 m and a sill between $0.1\,\%^2$ and $0.15\,\%^2$ again, with an exponential increase when exceeding a lag distance from 0.2 to 0.25 m. In contrast, the semivariance of YZ back has the highest sill, with $0.21\,\%^2$ and a range of 0.15 m; however, semivariance drops after ex-
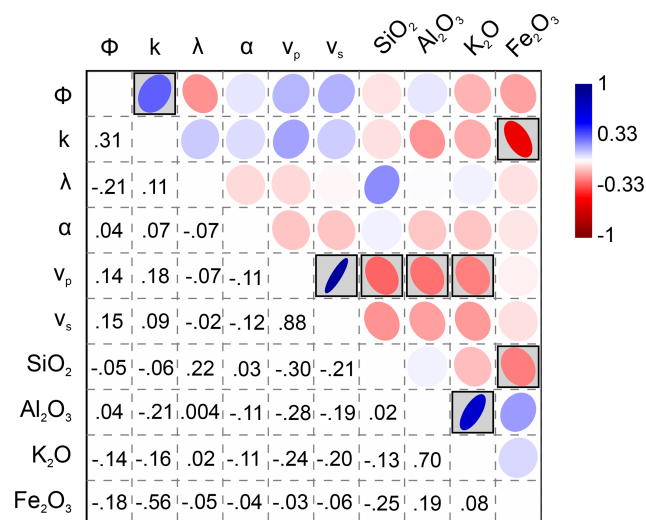
Figure 8. Matrix visualization of the Pearson correlation coefficient derived from the plug measurements. Statistically significant correlations with a p-value ≤ 0.05 are highlighted by gray boxes. The diameter of the ellipses' conjugate axes is dependent on the correlation coefficient. The smaller the length of the axis, the stronger the correlation. The matrix is diagonal, meaning that the Pearson correlation coefficient as numerical expression is located at the diagonal position relative to each ellipsis. Φ – effective porosity; k – permeability; λ – thermal conductivity; α – thermal diffusivity; v_p – P-wave velocity; and v_s – S-wave velocity.
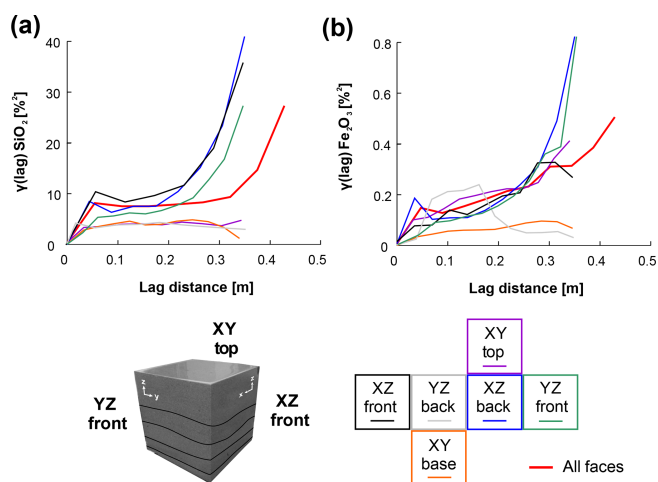
Figure 9. Empirical semivariograms of the mass fraction of SiO_2 **(a)** and Fe_2O_3 **(b)** in rock cube OSB1_c grouped by the investigated rock face.

ceeding a lag distance of 0.2 m. XZ back provides the highest degree of similarity with a range of 0.3 m and a sill of $0.09\,\%^2$, using a spherical approximation. Both geochemical properties show a zonal anisotropy where the sill shows different magnitudes along different directions (Wackernagel, 2003; Allard et al., 2016).

Table 1. RMSE and MAE for the interpolation results of IDW and SK for OSB1_c. k – permeability; Φ – effective porosity; λ – thermal conductivity; α – thermal diffusivity; and v_p – P-wave velocity.

	RMSE IDW	RMSE SK	MAE IDW	MAE SK
k	0.19	0.17	0.15	0.14
Φ	0.54	0.59	0.4	0.42
λ	0.23	0.22	0.18	0.16
α	0.14	0.17	0.1	0.1
v_p	64.19	60.95	52.21	44.74
SiO_2	4.07	3.25	3.05	2.09
Al_2O_3	0.8	0.83	0.66	0.66
K_2O	0.25	0.26	0.19	0.2
Fe_2O_3	0.93	0.32	0.86	0.21

3.4 Spatial pattern analysis

The spatial distributions of the rock properties are interpolated with Shepard's inverse distance weighting (IDW) and simple kriging (SK). Both realizations of a single scalar field provide comparable patterns, which is due to the high sampling density. The interpolation errors are also located in similar ranges; however, IDW seems to be more sensitive to outliers, resulting in much higher interpolation errors with regard to properties like P-wave velocity or mass fraction of SiO_2 (Table 1). IDW tends to underestimate the maximum and minimum values in the scalar fields. Thus, petrophysical and geochemical contrasts are more distinctly reproduced in the geostatistical approach. Also, the IDW realization shows the bull's eye effect, which is a typical artifact of IDW interpolations (Shepard, 1968). Accordingly, the simple kriging realizations are used for further analyses.

The rock properties exhibit a multitude of spatial patterns. Here, discrete, layered and homogeneous patterns, both connected and disconnected to primary sedimentary structures, could be observed in the interpolations.

3.4.1 Patterns connected to sedimentary structures

A bedding-connected pattern is exhibited in the intrinsic permeability and Fe_2O_3 interpolation results of OSB1_c. The mass fraction of Fe_2O_3 varies between 1.25 % and 5 % in OSB1_c. In the histogram displayed in Fig. 11, outliers were removed according to Tukey's outlier-detection method (Tukey, 1977). The local histogram of OSB1_c's intrinsic permeability shows a bimodal distribution ranging from 0.7 to 3.9 mD. The application of Tukey's method revealed no statistical outliers in this scalar field.

The bedding structures in OSB1_c are well reflected by the spatial pattern of the interpolated intrinsic permeability gradually increasing from low values, between 0.7 and 2 mD, in the lower beds to higher values, between 2 and 4 mD, in the upper beds (Fig. 10).
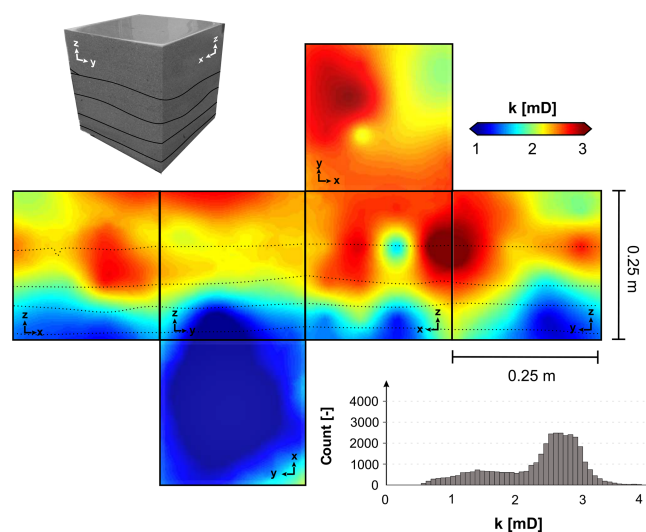
Figure 10. Spatial distribution of the intrinsic permeability modeled with a simple kriging interpolation. The histogram shows a bimodality of the distribution split up into the basal beds and uppermost beds.

The spatial distribution of the mass fraction of Fe_2O_3 in OSB1_c provides a reciprocal trend compared to the permeability. Here, the lowermost bed shows a significantly higher content compared to the upper beds. Both scalar fields show zonal anisotropy. The Fe_2O_3 content is an indicator of the detrital matrix, pseudomatrix and cement content that, in turn, would explain the reciprocal relationship with the permeability measurements. In siliciclastic systems, iron can be contained in clay minerals (up to 30 wt %; Brigatti et al., 2006), mafic components or in iron-rich oxides, hydroxides or carbonates. Local excesses in the Fe_2O_3 content exist in the spatial distribution. Those can be explained by clay-rich intraclasts observed on the rock faces. When comparing the pattern to Fig. 2 at both XZ-oriented cube faces, rip-up clasts can be observed where high Fe_2O_3 mass fractions occur. Those areas provide the maximum values of the Fe_2O_3 distribution.

3.4.2 Patterns decoupled from sedimentary structures

Other scalar fields are decoupled from depositional bounding surfaces. For instance, the geochemical mass fractions of K_2O (Fig. 12) and Al_2O_3 (Fig. 13) provide a significant positive correlation unconnected to visible structural boundaries. Typically, those geochemical properties are indicative of the presence of orthoclase feldspar ($KAlSi_3O_8$) and/or illite ($KAl_3Si_3O_{10}(OH)_2$) in siliciclastic environments. The mass ratio of both components is roughly 1 : 3 to 1 : 4, which is in accordance to the illite fraction that was observed in the thin section and ESEM analyses. Only minor amounts of orthoclase feldspar could be found in the thin sections. Thus,

Figure 11. Spatial distribution of the mass fraction of Fe_2O_3 modeled with a simple kriging interpolation. As in the intrinsic permeability interpolation, a bimodality can be observed in the empirical histogram.

Figure 13. Spatial distribution of the mass fraction of Al_2O_3 modeled with a simple kriging interpolation. The pattern is decoupled from primary sedimentary structures and shows a network-like structure.

4 Discussion

The overall aim of this study was to quantify the 3D interdependencies of thermophysical, hydraulic, elastic and geochemical scalar fields in sandstone media at the lithofacies scale and to identify the controlling factors for the property distributions. With a high-resolution study at the lithofacies scale, statistical and spatial interrelationships between characteristic physicochemical fields could be discovered and traced back to depositional and diagenetic processes.

4.1 Petrophysical and geochemical characteristics

Recent multiscale modeling approaches without the use of local constraints show that the prediction of permeability and porosity in siliciclastic systems is still challenging (Nordahl et al., 2014). Geological sampling almost never includes the entire domain that is investigated. With sampling densities of 25.4 % and 18.2 %, we reached a very high degree of coverage. Studies such as Hurst and Rosvoll (1991) showed that a very high sampling density is necessary to cover the entire variance of permeability at the lithofacies scale. The interpolations performed in this study reproduce the global histogram properly and outliers are also accounted for. This, in fact, implies that the sampling density was selected adequately in order to capture the total variability present in the physical and geochemical scalar fields. This condition is typically only fulfilled in sequential simulations (Robertson et al., 2006) rather than in conventional interpolations.

Although statistically significant correlations may imply a natural relationship between physicochemical properties, this relationship could also be based on random processes

Figure 12. Spatial distribution of the mass fraction of K_2O modeled with a simple kriging interpolation. The pattern is decoupled from primary sedimentary structures and shows a network-like structure.

we assume that the correlation of K_2O and Al_2O_3 can be traced back to the illite phases.

Higher fractions of Al_2O_3 are supposedly due to higher kaolinite ($Al_2Si_2O_5(OH)_4$) fractions in the clay mineral assemblages. The patterns are diffuse, showing autocorrelated areas of slightly enriched and depleted mass fractions. Enriched areas seem to be connected, building network-like patterns, while depleted areas are more isolated.

requiring causality to be verified. Weak correlations were found between the effective porosity and the intrinsic permeability, which are usually positively correlated (Pape et al., 1999). This relationship can be traced back to the Kozeny–Carman equation that connects the permeability with the effective pore throat radius r_{eff}^2 and a formation factor F as follows:

$$k = r_{\text{eff}}^2/(8 \cdot F). \tag{21}$$

The formation factor is defined as the ratio of tortuosity and porosity showing that porosity and permeability provide a positive formal relationship empirically. A high number of secondary pores, produced by feldspar dissolution, did not significantly contribute to the permeability in the investigated sandstones since those pores are often hydraulically isolated. Consequently, secondary porosity did not necessarily lead to increasing radii of the effective pore throats rather than increasing tortuosity. Also, recrystallized quartz cement – blocking a large number of the pore throats – must be taken into account. Both effects, in turn, resulted in a degraded permeability. In addition to the geometrical aspects previously mentioned, the alteration products in the form of clay minerals occupy the pore space, which lead to larger adhesive effects that hinder the ability to transport fluids as well. This observation is in good agreement with observations made by Molenaar et al. (2015) in Rotliegend rocks from the Donnersberg Formation. In addition, these observations are well reflected by the very low values of the intrinsic permeability in both rock cubes. Another reason for the very low intrinsic permeability is the high amount of primary clay and the low maturity of deltaic sheet-like distributary mouth bar deposits (Tye and Hickey, 2001).

The linear correlation analysis revealed a significant negative relationship between hydraulic and geochemical properties that fits to a polynomial regression (Fig. 14). It should be considered that the geochemical measurements cover a very different measurement area – represented by a spot with a 1.2 cm diameter and around 0.5 cm penetration depth compared to the hydraulic measurements performed on an entire rock cylinder with a 40 mm height and diameter. Additionally, instead of using highly precise stationary X-ray fluorescence devices for measurements, a portable, faster device was used to efficiently derive spatial trends in the objects of investigation. This technique weakens the implications for absolute values; however, the trends observed in the measurements from the portable device are in good agreement with trends observed by stationary devices. Also, the observed geochemical characteristics are in accordance with geochemical properties of quartz-rich sandstone varieties that were investigated in Bhatia (1983) or Baiyegunhi et al. (2017).

Geochemical analyses, in contrast to petrographic ones, limit the interpretations of geological processes as mineral phases can only be assumed and not determined for certain. A high mass fraction of Fe_2O_3 may imply that the rock is rich

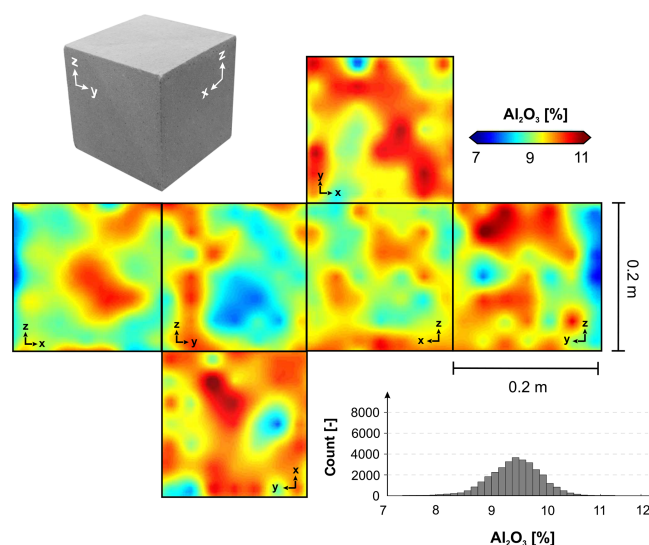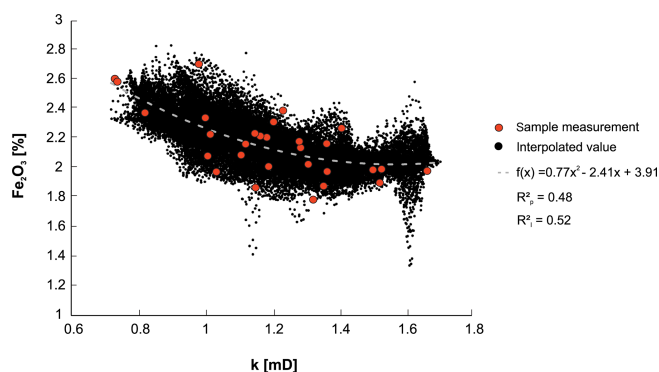**Figure 14.** Regression analysis of the relationship between intrinsic permeability and mass fraction of Fe_2O_3 in the interpolated scalar fields of the rock cube OSB2_c. R_p^2 is the coefficient of determination for the plug measurements, and R_i^2 is the coefficient of determination for the interpolated values.

in iron-bearing minerals like clay minerals, hematite, magnetite, goethite, lepidocrite or ferrihydrite (Costabel et al., 2018), however, a precise classification of the mineral phase is not possible. Iron oxides are more common in secondary precipitates that usually form during eo- and mesodiagenesis (Pettijohn et al., 1987). The degrading impact of iron-oxide-rich coatings on permeability and porosity in unconsolidated sand and gravel has been shown in studies like Costabel et al. (2018). The number of detrital iron-rich phases, such as hematite, which are present in the rock matrix, is typically less (Walker et al., 1981; Turner et al., 1995; Ixer et al., 1979) when compared to the secondary amount. In our case, however, thin section and ESEM analyses revealed that a high degree of the intergranular matrix is still preserved, especially at the base of OSB2_c where high amounts of mud and mud intraclasts are incorporated from basal erosion. The small grain size of the matrix offers a great surface area for iron-oxide-rich precipitates, which might have further enforced degradation of porosity and permeability. The primary matrix typically plugs the pore throats of porous, matrix-rich media. This reduces the ability to conduct fluids compared to matrix-free ones. However, due to progressive compaction, we cannot quantify for certain how large the size of the primary matrix is compared to the pseudomatrix produced by the plastic compaction of ductile, clay-rich grains and by feldspar dissolution.

A significant correlation between K_2O and Al_2O_3 could be detected. The spatial distribution resembles a network-like structure that might be either a product of diffusive mass transport during meso- or telodiagenesis or it might reflect the distribution of feldspar grains and its residues in the sandstone. During feldspar alteration, SiO_2 is dissolved and K remains in the alteration products, which could be an implication for the mesoscale network-like structure into which pore fluids could have had migrated. This relationship is un-

derlined by a negative, yet nonsignificant, correlation of K_2O with SiO_2.

5 Conclusions

Significant nonintuitive relationships between the physical and geochemical scalar fields at the lithofacies scale have been revealed with a deductive approach of spatial field modeling and statistical data analysis. All in all, the following conclusions can be drawn from this study:

1. As specific properties such as the mass fraction of Fe_2O_3 preserve sedimentological textures well in their spatial distribution, other properties seem to be completely decoupled from depositional bounding surfaces. These scalar fields probably reflect processes that might have taken place during the diagenetic overprint of the rocks as a result of burial and exhumation. These processes produce diffuse patterns, as discussed with regard to the correlation of K_2O and Al_2O_3.

2. This study demonstrates that the observation of bedding structures does not necessarily indicate a stronger polar anisotropy compared to macroscopically unstructured lithologies. Here, the microscopic characteristics, like the amount of secondary porosity, might play a more important role in the attenuation of physical waves than the bounding surfaces.

3. It could be shown that hydraulic properties are dependent on the intergranular matrix and cement amount, which are in turn controlled by depositional processes and eogenetic precipitates. Those findings are not new (see Wilson and Pittman, 1977 or Nordahl et al., 2014); however, they have not been evaluated in lithofacies-scale 3D environments yet. We assume that primary matrix and ductile grain content has the most detrimental effect on rock permeability. Ductile grains were mechanically deformed during compaction, leading to plugged pore throats. Feldspar dissolution has a highly productive effect on porosity but not on permeability.

4. We demonstrate that the strength of statistical correlation can be preserved in spatial interpolations as long as the sampling density is sufficient. If the sampling density is too low, a statistical correlation might be inadvertently feigned.

5. As shown in this study, the local geological variability should not be underestimated as an uncertainty factor in spatial predictions and upscaling procedures. In fact, the local geological variability of physicochemical properties might nearly cover the variability being present in an entire formation. Therefore, a high-resolution analysis of physicochemical rock properties can assist in assessing the uncertainty of field-scale property models

which is induced by the local geological variability at the lithofacies scale.

References

Agemar, T., Weber, J., and Schulz, R.: Deep geothermal energy production in Germany, Energies, 7, 4397–4416, https://doi.org/10.3390/en7074397, 2014.

Allard, D., Senoussi, R., and Porcu, E.: Anisotropy models for spatial data, Math. Geosci., 48, 305–328, https://doi.org/10.1007/s11004-015-9594-x, 2016.

Aretz, A., Bär, K., Götz, A. E., and Sass, I.: Outcrop analogue study of permocarboniferous geothermal sandstone reservoir formations (northern Upper Rhine Graben, Germany): impact of mineral content, depositional environment and diagenesis on petrophysical properties, Int. J. Earth Sci., 105, 1431–1452, https://doi.org/10.1007/s00531-015-1263-2, 2015.

Armstrong, M.: Experimental variograms, Springer, Berlin, Heidelberg, Germany, 47–58, https://doi.org/10.1007/978-3-642-58727-6_4, 1998.

Baiyegunhi, C., Liu, K., and Gwavava, O.: Geochemistry of sandstones and shales from the Ecca Group, Karoo Supergroup, in the Eastern Cape Province of South Africa: Implications for provenance, weathering and tectonic setting, Open Geosci., 9, 340–360, https://doi.org/10.1515/geo-2017-0028, 2017.

Becker, A., Schwarz, M., and Schäfer, A.: Lithostratigraphische Korrelation des Rotliegend im östlichen Saar–Nahe–Becken, Jber. Mitt. Oberrhein. Geol. Ver., 94, 105–133, https://doi.org/10.1127/jmogv/94/2012/105, 2012.

Bhatia, M. R.: Plate Tectonics and Geochemical Composition of Sandstones, J. Geol., 91, 611–627, 1983.

Bouma, A. H.: Sedimentology of some Flysch deposits; a graphic approach to facies interpretation, Elsevier, Amsterdam, New York, 1962.

Brigatti, M. F., Galan, E., and Theng, B. K. G.: Chapter 2 Structures and Mineralogy of Clay Minerals, in: Handbook of Clay Science vol. 1, edited by: Bergaya, F., Theng, B. K. G., and Lagaly, G., Elsevier, 19–86, https://doi.org/10.1016/S1572-4352(05)01002-0, 2006.

Carslaw, H. S. and Jaeger, J. C.: Conduction of Heat in Solids, Second Edition, Oxford University Press, Oxford, United Kingdom, 1959.

Celisse, A.: Optimal cross-validation in density estimation with the L2-loss, Ann. Stat., 42, 1879–1910, https://doi.org/10.1214/14-AOS1240, 2014.

Corbett, P. and Jensen, J. L.: Estimating the mean permeability: how many measurements do you need?, First Break, 10, 5, https://doi.org/10.3997/1365-2397.1992006, 1992.

Corbett, P. W. M., Hamdi, H., and Gurav, H.: Layered fluvial reservoirs with internal fluid cross flow: a well-connected family of well test pressure transient responses, Petrol. Geosci., 18, 219–229, https://doi.org/10.1144/1354-079311-008, 2012.

Costabel, S., Weidner, C., Müller-Petke, M., and Houben, G.: Hydraulic characterisation of iron-oxide-coated sand and gravel based on nuclear magnetic resonance relaxation mode analyses, Hydrol. Earth Syst. Sci., 22, 1713–1729, https://doi.org/10.5194/hess-22-1713-2018, 2018.

Deutsch, C. V. and Journel, A.: GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press, Oxford, United Kingdom, https://books.google.de/books?id=CNd6QgAACAAJ (last access: 11 August 2020), 1998.

Filomena, C. M., Hornung, J., and Stollhofen, H.: Assessing accuracy of gas-driven permeability measurements: a comparative study of diverse Hassler-cell and probe permeameter devices, Solid Earth, 5, 1–11, https://doi.org/10.5194/se-5-1-2014, 2014.

Fongngern, R., Olariu, C., Steel, R., Mohrig, D., Krézsek, C., and Hess, T.: Subsurface and outcrop characteristics of fluvial-dominated deep-lacustrine clinoforms, Sedimentology, 65, 1447–1481, https://doi.org/10.1111/sed.12430, 2018.

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Oxford University Press, Oxford, United Kingdom, 1997.

Gu, Y., Rühaak, W., Bär, K., and Sass, I.: Using seismic data to estimate the spatial distribution of rock thermal conductivity at reservoir scale, Geothermics, 66, 61–72, https://doi.org/10.1016/j.geothermics.2016.11.007, 2017.

Hamdi, H., Ruelland, P., Bergey, P., and Corbett, P. W.: Using geological well testing for improving the selection of appropriate reservoir models, Petrol. Geosci., 20, 353–368, https://doi.org/10.1144/petgeo2012-074, 2014.

Heap, M. J., Kushnir, A. R. L., Gilg, H. A., Wadsworth, F. B., Reuschlé, T., and Baud, P.: Microstructural and petrophysical properties of the Permo-Triassic sandstones (Buntsandstein) from the Soultz-sous-Forêts geothermal site (France), Geoth. Energy, 5, 26, https://doi.org/10.1186/s40517-017-0085-9, 2017.

Henk, A.: Mächtigkeit und Alter der erodierten Sedimente im Saar–Nahe–Becken (SW-Deutschland), Geologische Rundschau, 81, 323–331, https://doi.org/10.1007/BF01828601, 1992.

Herron, M. M.: Geochemical classification of terrigenous sands and shales from core or log data, J Sediment. Res., 58, 9, https://doi.org/10.1306/212F8E77-2B24-11D7-8648000102C1865D, 1988.

Hornung, J. and Aigner, T.: Reservoir Architecture in a Terminal Alluvial Plain: An Outcrop Analogue Study (Upper Triassic, Southern Germany) Part 1: Sedimentology And Petrophysics, J Petrol. Geol., 25, 3–30, https://doi.org/10.1111/j.1747-5457.2002.tb00097.x, 2002.

Hornung, J., Linsel, A., Schröder, D., Gumbert, J., Ölmez, J., Scheid, M., and Pöppelreiter, M.: Understanding small-scale petrophysical heterogeneities in sedimentary rocks – the key to understand pore geometry variations and to predict lithofacies-dependent reservoir properties, Digital Geology – Multi-scale analysis of depositional systems and their subsurface modelling workflows, Special Volume, EAGE, Houten, the Netherlands, 71–90, ISBN: 9789462823372, 2020.

Hudson, G. and Wackernagel, H.: Mapping temperature using kriging with external drift: Theory and an example from Scotland, Int. J. Climatol., 14, 77–91, https://doi.org/10.1002/joc.3370140107, 1994.

Hurst, A. and Rosvoll, K. J.: Permeability Variations in Sandstones and their Relationship to Sedimentary Structures, Academic Press, Inc., San Diego, New York, Boston, London, Sydney, Tokyo, Toronto, 166–196, https://doi.org/10.1016/B978-0-12-434066-4.50011-4, 1991.

Ixer, R. A., Turner, P., and Waugh, B.: Authigenic iron and titanium oxides in triassic red beds: St. Bees Sandstone, Cumbria, Northern England, Geol. J., 14, 179–192, https://doi.org/10.1002/gj.3350140214, 1979.

Jackson, M. D., Muggeridge, A. H., Yoshida, S., and Johnson, H. D.: Upscaling Permeability Measurements Within Complex Heterolithic Tidal Sandstones, Math. Geol., 35, 499–520, https://doi.org/10.1023/A:1026236401104, 2003.

Kiryukhin, A. V., Kaymin, E. P., and Zakharova, E. V.: Using TOUGHREACT to Model Laboratory Tests on the Interaction of NaNO$_3$–NaOH Fluids with Sandstone Rock at a Deep Radionuclide Repository Site, Nucl. Technol., 164, 196–206, https://doi.org/10.13182/NT08-A4019, 2008.

Kushnir, A. R. L., Heap, M. J., Baud, P., Gilg, H. A., Reuschlé, T., Lerouge, C., Dezayes, C., and Duringer, P.: Characterizing the

physical properties of rocks from the Paleozoic to Permo-Triassic transition in the Upper Rhine Graben, Geotherm. Energy, 6, 16, https://doi.org/10.1186/s40517-018-0103-6, 2018.

Lake, L. W. and Srinivasan, S.: Statistical scale-up of reservoir properties: concepts and applications, J. Petrol. Sci. Eng., 44, 27–39, https://doi.org/10.1016/j.petrol.2004.02.003, 2004.

Landa, J. L. and Strebelle, S.: Sensitivity Analysis of Petrophysical Properties Spatial Distributions, and Flow Performance Forecasts to Geostatistical Parameters Using Derivative Coefficients, Society of Petroleum Engineers, SPE Annual Technical Conference and Exhibition, 29 September–2 October, San Antonio, Texas, https://doi.org/10.2118/77430-MS, 2002.

Li, S., Li, S., Shan, X., Gong, C., and Yu, X.: Classification, formation, and transport mechanisms of mud clasts, Int. Geol. Rev., 59, 1609–1620, https://doi.org/10.1080/00206814.2017.1287014, 2017.

Linsel, A.: GeoReVi v1.0.1, https://doi.org/10.5281/zenodo.3695815, 2020a.

Linsel, A.: Physicochemical Characteristics of Sandstone Media, figshare, https://doi.org/10.6084/m9.figshare.11791407.v2, 2020b.

Lu, G. Y. and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, Comput. Geosci., 34, 1044–1055, https://doi.org/10.1016/j.cageo.2007.07.010, 2008.

Malvić, T., Ivšinović, J., Velić, J., and Rajić, R.: Kriging with a Small Number of Data Points Supported by Jack-Knifing, a Case Study in the Sava Depression (Northern Croatia), Geosci., 9, 36, https://doi.org/10.3390/geosciences9010036, 2019.

McKinley, J. M., Lloyd, C. D., and Ruffell, A. H.: Use of Variography in Permeability Characterization of Visually Homogeneous Sandstone Reservoirs with Examples from Outcrop Studies, Math. Geol., 36, 761–779, https://doi.org/10.1023/b:Matg.0000041178.73284.88, 2004.

McKinley, J. M., Atkinson, P. M., Lloyd, C. D., Ruffell, A. H., and Worden, R. H.: How Porosity and Permeability Vary Spatially With Grain Size, Sorting, Cement Volume, and Mineral Dissolution In Fluvial Triassic Sandstones: The Value of Geostatistics and Local Regression, J. Sediment. Res., 81, 844–858, https://doi.org/10.2110/jsr.2011.71, 2011.

McKinley, J. M., Ruffell, A. H., and Worden, R. H.: An Integrated Stratigraphic, Petrophysical, Geochemical and Geostatistical Approach to the Understanding of Burial Diagenesis: Triassic Sherwood Sandstone Group, South Yorkshire, UK, John Wiley & Sons, Inc., Chichester, United Kingdom, 231–255, https://doi.org/10.1002/9781118485347.ch10, 2013.

Medici, G., West, L. J., and Mountney, N. P.: Characterizing flow pathways in a sandstone aquifer: Tectonic vs. sedimentary heterogeneities, J. Contam. Hydrol., 194, 36–58, https://doi.org/10.1016/j.jconhyd.2016.09.008, 2016.

Medici, G., West, L. J., and Mountney, N. P.: Characterization of a fluvial aquifer at a range of depths and scales: the Triassic St Bees Sandstone Formation, Cumbria, UK, Hydrogeol. J., 26, 565–591, https://doi.org/10.1007/s10040-017-1676-z, 2018.

Medici, G., West, L. J., and Mountney, N. P.: Sedimentary flow heterogeneities in the Triassic U.K. Sherwood Sandstone Group: Insights for hydrocarbon exploration, Geol. J., 54, 1361–1378, https://doi.org/10.1002/gj.3233, 2019.

Miall, A. D.: Architectural-element analysis: A new method of facies analysis applied to fluvial deposits, Earth-Sci. Rev., 22, 261–308, https://doi.org/10.1016/0012-8252(85)90001-7, 1985.

Micromeritics: GeoPyc 1360 – Envelope Density Analyzer, available at: https://www.micromeritics.com/repository/files/geopyc_1360_reg_and_tap.pdf (last access: 8 August 2020), 1998.

Middleton, G. V.: Sediment Deposition from Turbidity Currents, Annu. Rev. Earth Pl. Sc., 21, 89–114, https://doi.org/10.1146/annurev.ea.21.050193.000513, 1993.

Molenaar, N., Felder, M., Bär, K., and Götz, A. E.: What classic greywacke (litharenite) can reveal about feldspar diagenesis: An example from Permian Rotliegend sandstone in Hessen, Germany, Sediment. Geol., 326, 79–93, https://doi.org/10.1016/j.sedgeo.2015.07.002, 2015.

Nordahl, K., Messina, C., Berland, H., Rustad, A. B., Rimstad, E., Martinius, A. W., Howell, J. A., and Good, T. R.: Impact of multiscale modelling on predicted porosity and permeability distributions in the fluvial deposits of the Upper Lunde Member (Snorre Field, Norwegian Continental Shelf), Geol. Soc. Lon., 387, 25, https://doi.org/10.1144/sp387.10, 2014.

Pape, H., Clauser, C., and Iffland, J.: Permeability prediction based on fractal pore-space geometry, Geophysics, 64, 1447–1460, https://doi.org/10.1190/1.1444649, 1999.

Pettijohn, F. J., Potter, P. E., and Siever, R.: Mineral and Chemical Composition, Springer New York, New York, NY, 25–67, https://doi.org/10.1007/978-1-4612-1066-5_2, 1987.

Popov, Y. A., Pribnow, D. F. C., Sass, J. H., Williams, C. F., and Burkhardt, H.: Characterization of rock thermal conductivity by high-resolution optical scanning, Geothermics, 28, 253–276, https://doi.org/10.1016/S0375-6505(99)00007-3, 1999.

Ringrose, P. and Bentley, M.: Reservoir Model Design, First Edition, Springer, the Netherlands, https://doi.org/10.1007/978-94-007-5497-3, 2015.

Ringrose, P. S., Sorbie, K. S., Corbett, P. W. M., and Jensen, J. L.: Immiscible flow behaviour in laminated and cross-bedded sandstones, J. Petrol. Sci. Eng., 9, 103–124, https://doi.org/10.1016/0920-4105(93)90071-L, 1993.

Robertson, R. K., Mueller, U. A., and Bloom, L. M.: Direct sequential simulation with histogram reproduction: A comparison of algorithms, Comput. Geosci., 32, 382–395, https://doi.org/10.1016/j.cageo.2005.07.002, 2006.

Rodrigo-Ilarri, J., Reisinger, M., and Gómez-Hernández, J. J.: Influence of Heterogeneity on Heat Transport Simulations in Shallow Geothermal Systems, Springer International Publishing, Cham, Germany, 849–862, https://doi.org/10.1007/978-3-319-46819-8_59, 2017.

Rühaak, W.: A Java application for quality weighted 3-d interpolation, Comput. Geosci., 32, 43–51, https://doi.org/10.1016/j.cageo.2005.04.005, 2006.

Rühaak, W.: 3-D interpolation of subsurface temperature data with measurement error using kriging, Environ. Earth Sci., 73, 1893–1900, https://doi.org/10.1007/s12665-014-3554-5, 2015.

Rühaak, W., Guadagnini, A., Geiger, S., Bär, K., Gu, Y., Aretz, A., Homuth, S., and Sass, I.: Upscaling thermal conductivities of sedimentary formations for geothermal exploration, Geothermics, 58, 49–61, https://doi.org/10.1016/j.geothermics.2015.08.004, 2015.

Sass, I. and Götz, A. E.: Geothermal reservoir characterization: a thermofacies concept, Terra Nova, 24, 142–147, https://doi.org/10.1111/j.1365-3121.2011.01048.x, 2012.

Shepard, D.: A Two-Dimensional Interpolation Function for Irregularly-Spaced Data., in: Proceedings of the 1968 ACM National Conference, Association for Computing Machinery, New York, NY, USA, 517–524, https://doi.org/10.1145/800186.810616, 1968.

Stollhofen, H.: Facies architecture variations and seismogenic structures in the Carboniferous–Permian Saar–Nahe Basin (SW Germany): evidence for extension-related transfer fault activity, Sediment. Geol., 119, 47–83, https://doi.org/10.1016/S0037-0738(98)00040-2, 1998.

Tellam, J. H. and Barker, R. D.: Towards prediction of saturated-zone pollutant movement in groundwaters in fractured permeable-matrix aquifers: the case of the UK Permo-Triassic sandstones, Geol. Soc. Lon., Special Publications, 263, 1–48, https://doi.org/10.1144/gsl.Sp.2006.263.01.01, 2006.

Thomsen, L.: Weak elastic anisotropy, Geophysics, 51, 1954–1966, https://doi.org/10.1190/1.1442051, 1986.

Tukey, J.: Exploratory Data Analysis, Pearson, Reading, Massachussetts, USA, 1977.

Turner, P., Burley, S., Rey, D., and Prosser, J.: Burial history of the Penrith Sandstone (Lower Permian) deduced from the combined study of fluid inclusion and palaeomagnetic data, Geol. Soc. Lon., Special Publications, 98, 43–78, https://doi.org/10.1144/GSL.SP.1995.098.01.04, 1995.

Tye, B. and Hickey, J.: Permeability characterization of distributary mouth bar sandstones in Prudhoe Bay field, Alaska: How horizontal cores reduce risk in developing deltaic reservoirs, AAPG Bull., 85, 459–475, https://doi.org/10.1306/8626C91F-173B-11D7-8645000102C1865D, 2001.

Verly, G.: Sequential Gaussian Simulation: A Monte Carlo Method for Generating Models of Porosity and Permeability, in: Generation, Accumulation and Production of Europe's Hydrocarbons III, edited by: Spencer, A. M., Springer, Berlin, Heidelberg, Germany, 345–356, 1993.

Wackernagel, H.: Multivariate Geostatistics, Third Edition, Springer, Berlin, Heidelberg, Germany, https://doi.org/10.1007/978-3-662-05294-5, 2003.

Walker, T. R., Larson, E. E., and Hoblitt, R. P.: Nature and origin of hematite in the Moenkopi Formation (Triassic), Colorado Plateau: A contribution to the origin of magnetism in red beds, J. Geophys. Res.-Sol. Ea., 86, 317–333, https://doi.org/10.1029/JB086iB01p00317, 1981.

Wang, J. and Zuo, R.: Identification of geochemical anomalies through combined sequential Gaussian simulation and grid-based local singularity analysis, Comput. Geosci., 118, 52–64, https://doi.org/10.1016/j.cageo.2018.05.010, 2018.

Whitney, D. L. and Evans, B. W.: Abbreviations for names of rock-forming minerals, Am. Mineral., 95, 185–187, https://doi.org/10.2138/am.2010.3371, 2010.

Wilson, M. D. and Pittman, E. D.: Authigenic clays in sandstones; recognition and influence on reservoir properties and paleoenvironmental analysis, J. Sediment. Res., 47, 3–31, https://doi.org/10.1306/212f70e5-2b24-11d7-8648000102c1865d, 1977.

Worden, R. H. and Burley, S. D.: Sandstone Diagenesis: The Evolution of Sand to Stone, in: Sandstone Diagenesis, edited by: Burley, S. D. and Worden, R. H., 1–44, https://doi.org/10.1002/9781444304459.ch, 2003.

Yang, J., Hua, B., Williamson, P., Zhu, H., McMechan, G., and Huang, J.: Elastic Least-Squares Imaging in Tilted Transversely Isotropic Media for Multicomponent Land and Pressure Marine Data, Surv. Geophys., 805–833, https://doi.org/10.1007/s10712-020-09588-3, 2020.

Zheng, S.-Y., Corbett, P. W. M., Ryseth, A., and Stewart, G.: Uncertainty in Well Test and Core Permeability Analysis: A Case Study in Fluvial Channel Reservoirs, Northern North Sea, Norway, AAPG Bull., 84, 1929–1954, https://doi.org/10.1306/8626c72b-173b-11d7-8645000102c1865d, 2000.

3.4. Publication IV – Accounting for Local Geological Variability in Sequential Simulations—Concept and Application

3.4.1. Introduction

Publication IV involves the modification of two popular geostatistical algorithms aiming at the integration of local geological variability into sequential simulations. The local geological variability is derived from the permeability measurements of the rock samples, which were extracted from the cubes, described in **Publication III**.

Our results indicate that the measure of the local simple kriging variance, which is commonly used to simulate the natural variability in geological media, is systematically underestimating the local geological variability. The modification of the sequential Gaussian simulation (SGS) and the direct sequential simulation (DSS) algorithms involves to simulate the natural variability from a geology-derived and data-driven local variance model (LVM) instead of the local simple kriging variance. The LVM is a global representation of the locally observed variability.

The modified algorithms reproduce the covariance model and the histogram in the range of ergodic fluctuations. Moreover, the spatial distribution of the heterogeneous regions is in accordance with the observable geological architecture, which is not fulfilled in the conventional algorithms.

3.4.2. Author Contributions

- As the first author, I developed the methodological concept, prepared the figures and wrote the manuscript.

- Together with S. Wiesler and J. Haas, I performed the sampling, conducted the field and laboratory measurements and the exploratory data analysis.

- K. Bär contributed towards the methodological concept.

- M. Hinderer supervised my research.

- All co-authors had a hand at the manuscript and contributed to the revision.

3.4.3. Publication

Published as:

Linsel, A., Wiesler, S., Haas, J., Bär, K. and Hinderer, M. (2020): Accounting for Local Geological Variability in Sequential Simulations—Concept and Application. *ISPRS International Journal of Geo-Information*. doi:10.3390/ijgi9060409

Article

Accounting for Local Geological Variability in Sequential Simulations—Concept and Application

Adrian Linsel [1,*] **, Sebastian Wiesler** [1] **, Joshua Haas** [1] **, Kristian Bär** [2] **and Matthias Hinderer** [1]

[1] Department of Applied Sedimentary Geology, Institute of Applied Geosciences,
Technische Universität Darmstadt, 64287 Darmstadt, Germany; Sebastian.Wiesler@hotmail.de (S.W.);
joshuahaas@gmx.de (J.H.); hinderer@geo.tu-darmstadt.de (M.H.)

[2] Department of Applied Geothermal Science and Technology, Institute of Applied Geosciences,
Technische Universität Darmstadt, 64287 Darmstadt, Germany; baer@geo.tu-darmstadt.de

* Correspondence: linsel@geo.tu-darmstadt.de

check for
updates

Abstract: Heterogeneity-preserving property models of subsurface regions are commonly constructed by means of sequential simulations. Sequential Gaussian simulation (SGS) and direct sequential simulation (DSS) draw values from a local probability density function that is described by the simple kriging estimate and the local simple kriging variance at unsampled locations. The local simple kriging variance, however, does not necessarily reflect the geological variability being present at subsets of the target domain. In order to address that issue, we propose a new workflow that implements two modified versions of the popular SGS and DSS algorithms. Both modifications, namely, LVM-DSS and LVM-SGS, aim at simulating values by means of introducing a local variance model (LVM). The LVM is a measurement-constrained and geology-driven global representation of the locally observable variance of a property. The proposed modified algorithms construct the local probability density function with the LVM instead of using the simple kriging variance, while still using the simple kriging estimate as the best linear unbiased estimator. In an outcrop analog study, we can demonstrate that the local simple kriging variance in sequential simulations tends to underestimate the locally observed geological variability in the target domain and certainly does not account for the spatial distribution of the geological heterogeneity. The proposed simulation algorithms reproduce the global histogram, the global heterogeneity, and the considered variogram model in the range of ergodic fluctuations. LVM-SGS outperforms the other algorithms regarding the reproduction of the variogram model. While DSS and SGS generate a randomly distributed heterogeneity, the modified algorithms reproduce a geologically reasonable spatial distribution of heterogeneity instead. The new workflow allows for the integration of continuous geological trends into sequential simulations rather than using class-based approaches such as the indicator simulation technique.

Keywords: sequential simulation; local variance model; geological heterogeneity; uncertainty estimation; subset variability

1. Introduction

Drawing conclusions from uncertain data in Earth sciences is rather usual than unusual. Each measurement in geoscientific studies is affected by measurement errors and represents only a subset of the natural variability of geological media. The natural variability is a substantial business-critical controlling factor of different types of subsurface utilization such as mining, hydrocarbon and geothermal exploitation, carbon capture and storage, or nuclear waste disposal. The physical variability of rocks is defined as the complexity or heterogeneity of a system within time

and space [1]. Even marginal discrepancies from the predicted property distributions in the subsurface can lead to inaccurate simulations of a quarrie's production potential or a reservoir's recovery and life-time [2,3]. Especially the small-scale variability of rock physical properties makes field-sized predictions still challenging.

Natural heterogeneity and the corresponding property distribution in time and space can be modeled through interpolation, statistical regression, machine learning or stochastic simulation [4,5] by using a number of observations or training data. Due to technical, economic or temporal limitations, geoscientific sampling campaigns practically always end up in scarce data sets within a target domain Ω. Accordingly, estimates of properties often do not account for or misfit the observed geological structures in the field and especially conventional interpolation techniques such as kriging produce smooth transitions at sharp geological boundaries. Moreover, they may fail to reproduce the global statistics appropriately. Conventional interpolations tend to underestimate the presence of values in the upper tail of a distribution and likewise in the lower tail, too [5]. Consequently, major geological heterogeneities, such as faults, major bounding surfaces, or physicochemical anomalies, are very likely not to be reproduced appropriately by a continuous random function (RF) [6].

In contrast to conventional interpolation techniques, stochastic simulations aim to reproduce the variance and the histogram observed in the global data [7,8]. Based on either being constrained or not, stochastic simulations split up into unconditional and conditional simulations [9]. Unconditional Monte Carlo-based simulations reproduce the original histogram without spatial constraints. The realizations produced by those methods, however, are regularly far away from representing the true spatial distribution and constitute "most likely" cases at the best. Conditional simulations, in contrast, aim to reproduce the original property distribution by means of discretely sampled points together with spatial characteristics such as the observed variogram model [10].

One type of conventional simulation algorithms is represented by the sequential Gaussian simulation (SGS) in which the local variability is simulated by sampling the local probability density function (PDF) derived from the local simple kriging variance σ_{SK}^2. This parameter results from the previously performed interpolation of the standard normally distributed data set [11]. Early field studies have proven the potential of this method to predict rock properties at unknown locations and to assess the uncertainty that can be expected in the area of interest [12–15]. More recent approaches lead to modifications of the SGS algorithm without the need to transform the original variable into standard normal space. That technique—better known as direct sequential simulations (DSS)—may, for example, sample from the global histogram rather than from the local PDF [9] or perform a quantile-quantile back-transformation into the original variable's space after the simulation. Those approaches can reproduce both the original histogram and experimental semivariogram model as well [10]. The local PDF derived from σ_{SK}^2, however, mainly reflects the degree of uncertainty induced by the interpolation method itself and does not necessarily reflect the local variability observed on a smaller scale than Ω.

In order to enhance the accuracy of sequential simulations, we propose a new workflow, which incorporates the local variability derived from measurements on a subset of Ω into SGS and DSS under the consideration of measurement errors. The modified SGS and DSS algorithms utilize a global representation of the locally observable variance, named local variance model (LVM), in order to draw a value at an unsampled location. Accordingly, the algorithms are called LVM-SGS and LVM-DSS. Before simulation, an integer programming optimization analysis is performed in order to optimize the robustness of the underlying interpolation function. Instead of sampling from the local PDF, which is generated by means of σ_{SK}^2, or by solving a global optimization problem, our parametric approach simulates a local PDF based on a measurement-constrained and geology-driven variance extracted from the LVM. The local PDF hereby is simulated with a Box–Muller transform [16].

The method was tested and validated in a case study, which has been conducted in a potential geothermal reservoir formation in southwestern Germany. Therefore, we measured the intrinsic permeability, representing a key parameter in many types of subsurface utilization, on a set of samples taken from an active quarry. Ω is represented by a 3-D outcrop model, which is constructed by means

of photogrammetric outcrop wall reconstruction. The model covers a volume of 9000 m^3. Small-scale variability is derived from rock samples, which are taken from two representative rock cubes. Those are regarded as Ω_b and cover a volume of 0.0156 m^3 and 0.008 m^3, respectively. The rock cubes are taken from the same outcrop, from which the global samples are taken from. Eventually, our approach is compared to the conventional SGS and DSS algorithms and assessed by its ability to reproduce the global variogram model and the geological heterogeneity.

2. Theoretical Background

2.1. Spatial Variability

In order to reduce the probability of economic failure in mining industries, the concept of the regionalized variable had been developed by Matheron [17] in the 1960s. The regionalized variable is a function that takes a definite value at each point of space. In geological media that regionalized variable often proved to be too complex to be expressed by mathematical functions. A regionalized variable is assumed to show a more or less steady continuity in space accompanied by local fluctuations (Figure 1). In geological media, those fluctuations usually result from the physical variability observed at smaller scales.
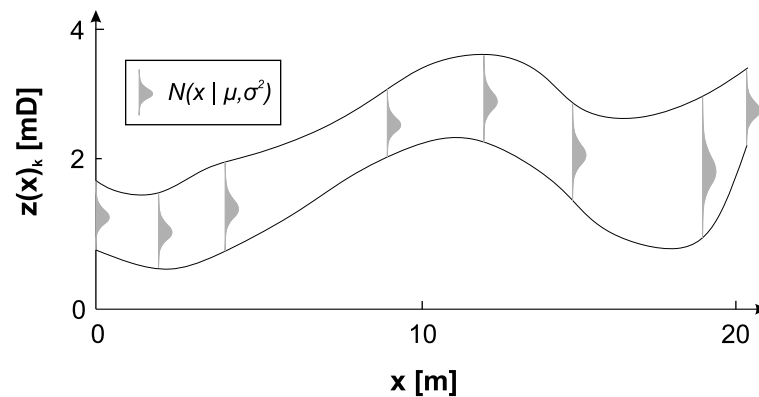
Figure 1. Conceptualization of a regionalized variable after [5] exemplary illustrated for the intrinsic permeability.

Lithological and physical variability is subject of numerous geoscientific studies [18–21] and is commonly termed *heterogeneity*. In the Oxford Dictionary [22] the word heterogeneity is defined as a *Difference or diversity in kind from other things* or a *Composition from diverse elements or parts; multifarious composition*. In most works, this term is used to describe that an object consists of multiple subsets being different to one another in one or more attributes. Li and Reynolds [1] restrict the term to be the variability of a system property in three-dimensional space. Fitch et al. [23] provide a set of methods to quantify heterogeneity within a sample of observations including the coefficient of variation (c_v),

$$c_v = \frac{\sqrt{\sigma^2}}{\mu}, \tag{1}$$

where σ is the standard deviation and μ is the arithmetic mean and the Dykstra–Parsons coefficient (c_{dp})

$$c_{dp} = \frac{x_{50} - x_{84}}{x_{50}}, \tag{2}$$

where x_n is the nth percentile of a distribution.

The continuity of a regionalized variable is thus dependent on the continuity of the geological media and may or may not provide continuity in a mathematical sense. In this work, we will use the term property for a regionalized variable, the term field for the (quasi-)continuous spatial distribution

of a property, and the term target domain Ω for an area of interest. When we mention global and local characteristics, we refer to characteristics of Ω and its subsets Ω_b, respectively.

2.2. Geostatistical Interpolation

Geostatistical interpolation techniques aim to estimate a value at unsampled locations of a property in Ω and build the base for sequential simulations. The most popular geostatistical interpolation technique is kriging. In the following subsections, we will briefly describe the theory behind kriging and focus on its variety simple kriging (SK). Moreover, we will discuss practical computational aspects such as neighborhoods.

2.2.1. Spatial Neighborhood

As the system of linear equations for geostatistical estimations might grow very large, those algorithms require subset-sampling in order to perform reasonably. Therefore, a 3-D search ellipsoid can be used to find the neighbors of a point in a mesh [24]. This ellipsoid can be defined by six properties: azimuth α; dip β; plunge γ together with the radius in X r_x, Y r_y, and Z direction r_z of the ellipsoid. α, β, and γ define the ellipsoid's clockwise rotation around the Z, X, and Y axes in this exact order. Accordingly, the rotation matrix T can be defined as

$$T = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\gamma & 0 & -\sin\gamma \\ 0 & 1 & 0 \\ \sin\gamma & 0 & \cos\gamma \end{pmatrix}. \tag{3}$$

After translating the mesh such that $x_x = x_y = x_z = 0$ and rotating it according to Equation (3), Equation (4) can be used to determine, whether a point x' with the transformed coordinates x'_x, x'_y and x'_z is located inside or on the boundary of the search ellipsoid (≤ 1) or not (>1).

$$\left(\frac{r_x}{x'_x}\right)^2 + \left(\frac{r_y}{x'_y}\right)^2 + \left(\frac{r_z}{x'_z}\right)^2 \leq 1 \tag{4}$$

2.2.2. Variography

The variographic analysis is a crucial prerequisite for numerous geostatistical interpolation techniques. Hereby, the experimental semivariogram represents the cumulative dissimilarity of a discrete set of point-pairs with n_c representing the count of point-pairs within the distance classes **h** of identical distance increments (Equation (5)).

$$\gamma(\mathbf{h}) = \frac{1}{2n_c(\mathbf{h})} \sum_{k=1}^{n_c(\mathbf{h})} \left(z(x_k + \mathbf{h}) - z(x_k)\right)^2 \tag{5}$$

The continuous counterpart, represented by the theoretical semivariogram γ_{theo}, is an approximation of the experimental semivariogram assuming $z(\mathbf{x})$ to be a stationary random field [25]. γ_{theo} is used to fit the experimental variogram. The spherical variogram model γ_{sph} with a nugget effect is a popular nested model used to fit the experimental semivariogram [26,27], which is calculated by

$$\gamma(\mathbf{h})_{sph} = \begin{matrix} n + b \cdot \left(\frac{3|\mathbf{h}|}{2a} - \frac{|\mathbf{h}|^3}{2a^3}\right) & \text{if } 0 \leq |\mathbf{h}| < a \\ n & \text{if } |\mathbf{h}| \geq a, \end{matrix} \tag{6}$$

with n being the nugget, b the sill and a the range [6]. The variogram model represents a covariance function c with the relationship $\gamma(\mathbf{h})_{theo} = c(0) - c(\mathbf{h})$, where c is a positive definite, even function and $c(0) = n + b$ in case of a spherical variogram model with nugget effect. Semivariograms can be used to quantify the spatial or time correlation of a random variable [27–29]. c and γ_{theo} are input variables for geostatistical interpolation algorithms.

2.2.3. Simple Kriging

Kriging is a commonly used stochastic technique to interpolate geological rock properties in space and time [30]. The kriging estimator is the best linear unbiased estimator (BLUE) of a property as it minimizes the error variance. It incorporates the covariance structure of the globally sampled values into the weights for predicting the value $z(x_0)$ at an unsampled location x_0 [31]. Therefore, $z(x_0)$ is calculated by weighting the values of the sampled locations and building a linear combination of those what gives

$$z(x_0) = \sum_{k=1}^{n} w_k \cdot z(x_k), \tag{7}$$

where w_k is the weight of the sampled point x_k with the value $z(x_k)$. The kriging types primarily differ by their derivation of the weight vector. For all kriging systems, a system of linear equations must be solved as it is outlined in the following paragraphs, in which we will consider the simple kriging (SK) technique [32] and expand it by the integration of a locally varying mean [33]. Therefore, we modify Equation (7) into

$$z(x_0)_{SK} = \sum_{k=1}^{n} w_k \cdot z(x_k) + \left(1 - \sum_{k=1}^{n} w_k\right) \cdot \mu. \tag{8}$$

in which the known stationary mean μ has been added [6]. While SK assumes that μ is globally constant and known, SK with locally varying mean assumes μ to be constant only in the neighborhood of x_0. In order to obtain the SK weights, a system of n linear equations must be solved in which n stands for the number of considered neighbors. This system of equations is defined as

$$\mathbf{Aw} = \mathbf{b}, \tag{9}$$

which corresponds to

$$\underbrace{\begin{pmatrix} c(x_1 - x_1) & \cdots & c(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ c(x_n - x_1) & \cdots & c(x_n - x_n) \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix}}_{\mathbf{w}} = \underbrace{\begin{pmatrix} c(x_1 - x_0) \\ \vdots \\ c(x_n - x_0) \end{pmatrix}}_{\mathbf{b}}, \tag{10}$$

with c as covariance function and x_n as point with known value [25]. In SK each interpolated point provides a simple kriging variance σ_{SK}^2 [5], which we can calculate by means of the formula

$$\sigma_{SK}^2 = c(0) - \sum_{k=1}^{n} w_k c(x_k, x_0). \tag{11}$$

The quality of a kriging interpolation is dependent on the variogram model and its goodness of fit to the experimental semivariogram.

2.2.4. Consideration of Measurement Error Variance

We already saw that kriging induces a local interpolation error by itself, namely, σ_{SK}^2. There are, however, also other components which bias the interpolation result. Besides σ_{SK}^2, the local and unknown variability of $z(\mathbf{x})$ in Ω_b as well as the measurement error variance σ_m^2 might play an important role (Figure 2). Integrating σ_m^2 into an interpolation can be achieved by estimating the measurement error precision σ_m with a variance of σ_m^2 and incorporating it into the kriging system of linear equations giving

$$\begin{pmatrix} c(x_1 - x_1) + \sigma_1^2 & \cdots & c(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ c(x_n - x_1) & \cdots & c(x_n - x_n) + \sigma_n^2 \end{pmatrix} \begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix} = \begin{pmatrix} c(x_1 - x_0) \\ \vdots \\ c(x_n - x_0) \end{pmatrix}. \tag{12}$$

In contrast to the conventional formula, σ_m^2 with regard to the considered known value at x_k is added in the diagonal of the matrix [25]. This accounts for the heteroscedastic nature of geological parameters as they commonly show a higher variability for high values and a lower variability for low values.
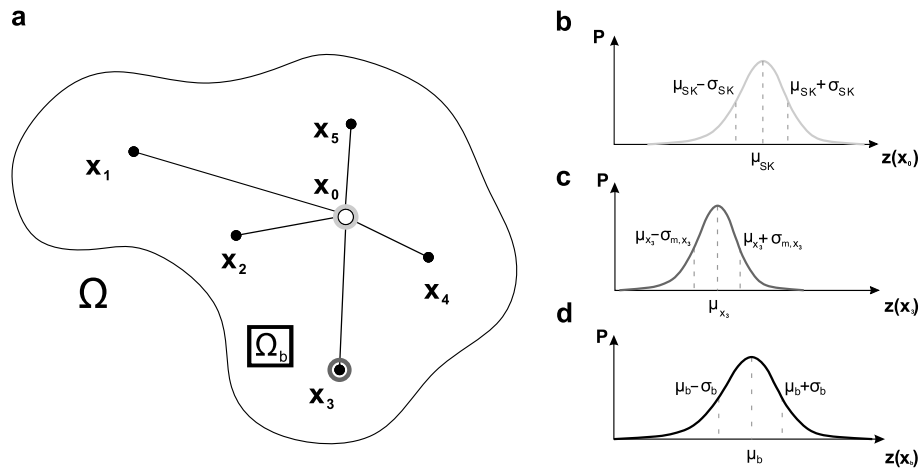
Figure 2. Schematic of the uncertainty components integrated into a predictive model of rock properties. (**a**) Illustration of an interpolation process using neighboring points x_k with known values to predict the unknown value at x_0. (**b**–**d**) Schematic of the local probability density functions (PDFs) in form of a Gaussian distribution defined by σ^2 and μ for the estimated kriging error variance σ_{SK}^2 at x_0 (**b**), the observed measurement error σ_m^2 at the point x_3 (**c**) and the observed variance σ_b^2 in a subset Ω_b of Ω (**d**).

2.3. Sequential Simulation

In contrast to geostatistical interpolation techniques, sequential simulations aim to reproduce the global statistics in form of the considered variogram model and the global histogram. Therefore, in order to account for the spatial heterogeneity of a rock property, the sequential Gaussian simulation (SGS) and the direct sequential simulation (DSS) algorithm can be utilized for univariate simulation. SGS is based on the multi-Gaussian approach [33], which assumes that the kriging error is standard normally distributed with $\mu = 0$ and $\sigma_{SK}^2 = 1$. This requires that each one-point cumulative density function (CDF) of any linear combination of the RV is normally distributed, that all subsets of the RF are multivariate normal, that the two-point distribution is normal and that all conditional distributions of subsets of the RF are normal [33]. If the RF fulfills the requirements, then the simple kriging estimate and variance characterize the posterior cumulative CDF under consideration of the normal score variogram model. Thus, we need to transform the original distribution's CDF into standard normal space for SGS. In order to transform any point in the CDF ($F(Z(\mathbf{u}))$) of any random variable $Z(\mathbf{u})$ to a random function $Y(\mathbf{u})$ and vice versa the following equation can be applied,

$$Y(\mathbf{u}) = \phi(Z(\mathbf{u})) = G^{-1}[F(Z(\mathbf{u}))], \tag{13}$$

where G^{-1} is the inverse Gaussian CDF of $Y(\mathbf{u})$, which is also named quantile function [34], and ϕ is the inverse Gaussian CDF of $F(Z(\mathbf{u}))$. Thus, z and y correspond to the same probabilities. For each previously interpolated point x_j now a random value of the normal distribution $\mathcal{N}(\mu_{SK}, \sigma_{SK}^2)$, whose PDF defines as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{14}$$

is drawn as $z(x_0)$ using the Box–Muller transform [35]. We can perform this transform by applying the equation

$$z(x_0) = \sqrt{-2 \cdot \log(u_1)} \cdot \cos(2\pi \cdot u_2) \cdot \sigma + \mu, \tag{15}$$

with u_1 and u_2 as random numbers $\in [0,1]$, σ as the standard deviation, and μ as the mean of the original distribution. The simulation is eventually back-transformed into the original space using a quantile-quantile back-transformation mapping technique. The reproduction of the covariance model, however, does not require the multi-Gaussian approach as long as the estimate and variance are derived from the SK estimation [9,10]. Thus, the conditional distribution type, which is chosen in order to simulate the variability at each point, does not necessarily need to be Gaussian. With this in mind, it is evident that a normal score transform is not needed before performing a sequential simulation. This results in the DSS approach, which commonly samples from the global PDF by determining the sampling interval from the local PDF [9].

2.4. Model Validation

2.4.1. Cross-Validation

In order to assess the quality of a realization, models, which are constructed by means of interpolation or simulation techniques, should be validated. Commonly, interpolations are validated by cross-validation. This technique is usually performed by using point removal procedures called leave-p-out cross-validation (LpO CV). For the LpO CV, p randomly selected samples are removed from the input data set of size n with $0 < p < n$ and the interpolation or simulation is performed without these samples [36]. As measures of goodness of fit, the mean-square error (MSE, Equation (16)), the root-mean-square error (RMSE, Equation (17)), and the mean-absolute error (MAE, Equation (18)) of the realization can be calculated as

$$MSE = \frac{1}{n} \sum_{k=1}^{n} \left(\hat{z}(x_k) - z(x_k) \right)^2, \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(\hat{z}(x_k) - z(x_k) \right)^2} \tag{17}$$

and

$$MAE = \frac{1}{n} \sum_{k=1}^{n} |\hat{z}(x_k) - z(x_k)|, \tag{18}$$

where $\hat{z}(x_k)$ are the simulated points. While Willmott et al. [37] question the status of the triangle inequality for the RMSE, which is required for a distance function metric, Chai and Draxler [38] show that the RMSE in fact fulfills this condition. Thus, if the model errors follow a normal distribution, the RMSE is to favor over the MAE [38].

2.4.2. Ergodic Fluctuations

The minimum requirement for geostatistical simulations is their ability to reproduce the original data, the global summary statistics and the global variogram model [8,39]. Erdogic fluctuations refer to the discrepancy between the model parameters and the realizations' statistics [6]. In the case of the variogram model, the discrepancy of a realization to the variogram model is related to the limitation of the integrated constraints to a limited neighborhood. This, in fact, leads to higher errors at far ranges within the simulation. In this study, we quantified the ergodic fluctuation of a realization by estimating the average MSE between the experimental semivariogram and the variogram model. If a realization's discrepancy among the experimental variogram and variogram model does not exceed the original values discrepancy, the variogram reproduction is said to be within the range of ergodic fluctuations.

3. Sequential Simulation using a Local Variance Model

In this section, we will describe how the SGS and DSS algorithms need to be modified in order to sample from a local variance model (LVM). The LVM can be described as a global representation of the locally observable variance σ^2_{LVM} in one mesh cell. Thus, the LVM can be referred to as the local geological heterogeneity. The LVM is constructed using a mapping technique in which the value of the mapped variances is constrained by a set of measurements. Those are intended to represent the small-scale variability present at the mapped position. Subsequently, the variance is interpolated onto Ω. The basic concept of interpolating a distribution in space is illustrated in Figure 3c.
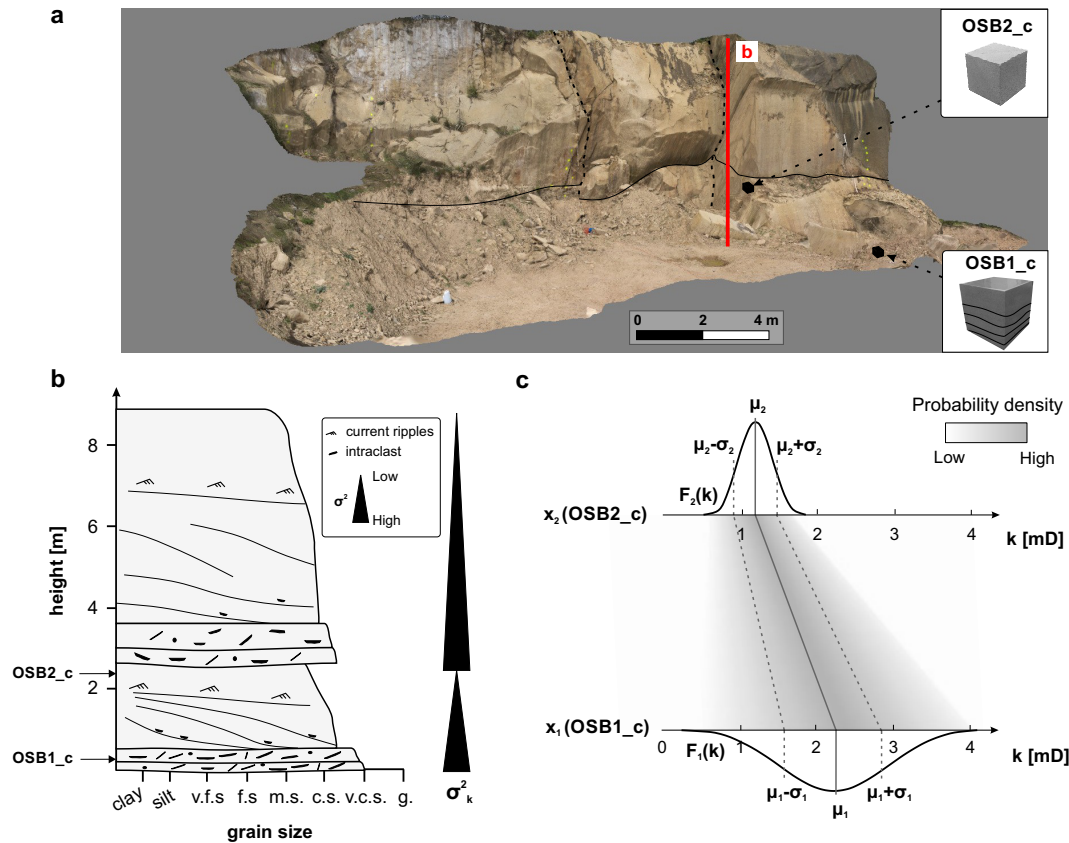
Figure 3. (**a**) Photogrammetric model of the investigated sandstone quarry. The outcrop is compartmentalized by two scissor faults and consists of two lacustrine-deltaic Bouma sequences [40]. (**b**) Sedimentological 1-D section of the sedimentary architecture observed in the outcrop. The Bouma sequence provides an erosive base. One sequence is characterized by a fining-upward trend and consists of intraclasts-rich massive sandstones at the base and trough cross-bedded and ripple cross-bedded sandstones towards top [40]. (**c**) Spatial interpolation of a PDF exemplary illustrated with both theoretical Gaussian distributions derived from the measurements of OSB1_c and OSB2_c.

The sequential simulations are performed on the nodes of Ω using a modification of the SGS and DSS algorithms, namely, the LVM-SGS and LVM-DSS. Our basic idea is that, if and only if the geological heterogeneity is exceeding σ^2_{SK} at x_k, we will sample from the LVM-constructed PDF instead of from the kriging-derived PDF. Otherwise, if the interpolation error is greater than the expectable geological heterogeneity, we will sample from the kriging-derived PDF. The generalized algorithm is displayed in Algorithm 1. All analyses have been conducted with the open-source software GeoReVi [41] in which the new algorithms have been implemented as extensions in the C# programming language (Appendix A).

Algorithm 1 LVM-SGS and LVM-DSS

Given: Ω; $\mathbf{x}$; N ▷ Target domain; Sampled locations; Neighborhood information;

Initialize: $\mathbf{u_{Sim}}$; $\mathbf{x}'$ ▷ Simulated locations; Spatial neighbors;

if GMV-SGS **then**

    $Y(\mathbf{x}) \leftarrow$ Equation (13) ▷ Transform to standard normal space

end if

$\gamma(\mathbf{h}) \leftarrow$ Equation (5) ▷ Estimate the experimental variogram

$\gamma(\mathbf{h})_{sph} \leftarrow$ Equation (6) ▷ Derive the variogram model and the covariance function

for all u_i in Ω **do**

    $\mathbf{x}' \leftarrow$ Equation (3) & Equation (4) ▷ Determine the neighborhood with N applied to $\mathbf{x}$ & $\mathbf{u_{Sim}}$

    $\mu_{SK} \leftarrow$ Equation (8) using $\gamma(\mathbf{h})_{sph}$ ▷ From $\mathbf{x}'$

    $\sigma^2_{SK} \leftarrow$ Equation (11) using $\gamma(\mathbf{h})_{sph}$ ▷ From $\mathbf{x}'$

    Allocate $\sigma^2_{LVM(x'_i)}$

    if $\sigma^2_{SK} \geq \sigma^2_{LVM(x'_i)}$ **then**

        $z(u_i) \leftarrow$ Equation (15) from $\mathcal{N}\left(\mu_{SK}, \sigma^2_{SK}\right)$ ▷ Draw a value with σ^2_{SK}

    else

        $z(u_i) \leftarrow$ Equation (15) from $\mathcal{N}\left(\mu_{SK}, \sigma^2_{LVM}\right)$ ▷ Draw a value with σ^2_{LVM}

    end if

    Add $z(u_i)$ to $\mathbf{u_{Sim}}$

end for

$F(Z(\mathbf{u})) \leftarrow$ Equation (13), ▷ Back-transform the simulated values into the original space

---

3.1. Case Study

In order to test and evaluate the new workflow with the modified algorithms, we conducted an outcrop analogue study in a quarry in Germany. In the following subsections, we will outline the object of investigation, the sampling strategy and the modeling techniques used to implement the LVM-SGS and LVM-DSS algorithms. We decided to use the intrinsic permeability k for the implementation as that property plays a critical role in numerous types of subsurface utilization—especially with regard to subsurface reservoirs.

3.1.1. Object of Investigation

An actively quarried sandstone outcrop (long. 7.647546, lat. 49.523821) in Obersulzbach, which is located in the Saar-Nahe basin in southwestern Germany, has been selected as object of investigation (Figure 3a). The outcrop exposes the Disibodenberg Formation of the innervariscan Rotliegend Group, which constitutes a deeply buried [42] potential hydrothermal reservoir unit [43] in the northern Upper Rhine Graben. The Disibodenberg Formation in the quarry is composed of two Bouma sequences (Figure 3b) from a lacustrine delta, which deposited during Permian times. There were two selection criteria being decisive for selecting the quarry. On the one hand side, the sedimentary beds are ≥ 2 m thick and laterally continuous. Moreover, the outcrop is actively mined, which reduces the impact of recent weathering onto the permeability. Moreover, it was possible to extract both rock samples from the outcrop wall as well as oriented rock cubes from different representative lithofacies types in order to conduct multi-scale three-dimensional investigations. The outcrop measures $50 \times 15 \times 10$ m and thus owns the size of a typical cell in common static and dynamic reservoir models (see, e.g., in [44]).

3.1.2. Sampling Strategy

Numerous studies showed that the physical variability in geological media must be integrated as a function of measurement volume, also known as the representative elementary volume (REV) [45].

The REV denotes a volume, at which a representative amount of heterogeneity is captured by one measurement [46] minimizing the smaller-scale fluctuations. Therefore, a multi-scale approach based on the concept of the REV has been implemented. Accordingly, 39 cylindrical rock samples with diameters and lengths of four centimeters were extracted from the outcrop wall. The samples were taken from six 1-D profiles covering the entire quarry area (Figure 4a). More information regarding the sample positions and orientations can be found in Linsel [41]. Those samples were used for the global field simulations.
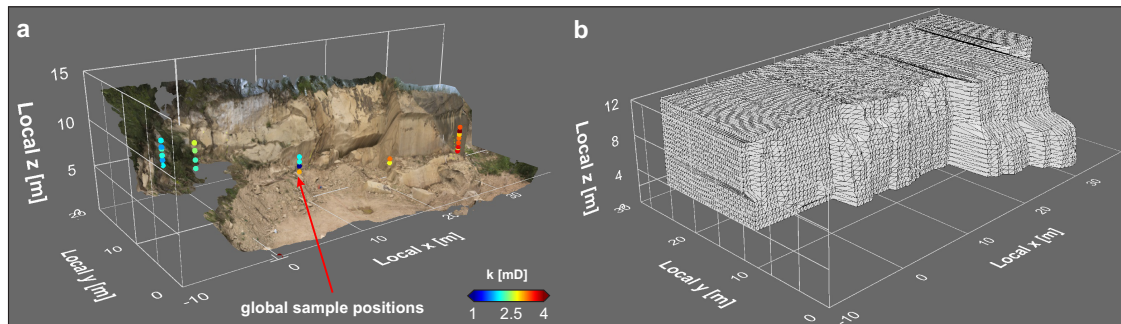
Figure 4. (**a**) Photogrammetric model of the investigated sandstone quarry in Obersulzbach, Germany. Sample locations are displayed as spheres, whose color indicates the observed permeability value at the sample locations. (**b**) Hexahedral non-orthogonal mesh of the investigated outcrop generated by an IDW interpolation using the nodes of the photogrammetric model as constraints.

The quarry contains sequences from a prodelta mouthbar deposited as turbiditic densites. The sequences graduate from a high-energetic depositional environment at the base to a low-energetic environment at the top as the flow velocity is steadily declining. The sequences consist of heterogeneous, intraclast-rich sandstones at the base and of trough cross-bedded, ripple cross-bedded and homogeneous sandstones at the top. Consequently, the sequences can be declared as Bouma sequences containing the Bouma A to Bouma E intervals in a fluvial-dominated lacustrine-deltaic depositional environment [47]. Based on that, we assumed that the variability within one Bouma sequence is highest at the base and lowest at the top (Figure 3b).

Accordingly, two rock cubes of $0.2 \times 0.2 \times 0.2$ m (OSB2_c) and $0.25 \times 0.25 \times 0.25$ m (OSB1_c) were taken—one from the top (Bouma E) and one from the base (Bouma A) of one sequence—in order to capture both the highest and the lowest variability. The locations of the cubes within the quarry and the strata are shown Figure 3a,b.

We selected two types of lithofacies: OSB1_c, a discontinuously cross-bedded, intraclast-rich lithofacies type and OSB2_c, a homogeneous lithofacies type without macroscopically observable internal bounding surfaces. In total, 79 rock cylinders were extracted from rock cube OSB1_c and 29 from OSB2_c. More information regarding the sampling process can be found in Linsel et al. [40]. Those samples were used for constraining the LVM.

3.1.3. Laboratory Measurements

The cylinder samples were cut, oven-dried at 105 °C and measured in the laboratory for determining the intrinsic gas permeability k at unsaturated conditions. k can be considered one of the key parameters of geothermal reservoir rocks with regard to hydrothermal systems in porous aquifers [48]. k was measured with the Hassler cell Darmstadt permeameter. The device's functionality is described in detail in Filomena et al. [49]. The permeability is provided in the industry-standard unit millidarcy (mD), where 1 mD corresponds to 9.869×10^{-16} m^2. The permeability measurement provides an error variance between 0 and 0.15 mD^2 in the range of the observed values [50].

3.1.4. Mesh Generation

In order to construct Ω, the outcrop wall is modeled using a photogrammetric representation that was downsampled into 40×20 faces and subsequently interpolated using Shepard's *p*-value IDW interpolation, which we can write as

$$z(x_0) = \frac{\sum_{k=1}^n \left(1/d_k^p\right) \cdot z(x_k)}{\sum_{k=1}^n 1/d_k^p}, \tag{19}$$

where d is the Euclidean distance between the the point with the known value x_k and the point with the unknown value x_0 and p is an exponent factor to influence the weights non-linearly. IDW has been applied with a short search radius of five meters and a power parameter of four. The interpolation result has an RMSE of 0.024 m, which can be considered low for the surface interpolation. The resulting outcrop surface is used as a bounding surface for a hexahedral mesh, which represents Ω, that is composed of 75,240 cells (Figure 4b, Table 1). The rock cubes, which represent Ω_b, are constructed by an orthogonal, hexahedral mesh containing 25,230 (OSB2_c) and 64,000 cells (OSB1_c), respectively. The volume of an average cell of the outcrop mesh is roughly eight times the volume of OSB1_c and 15 times the volume of OSB2_c (Table 1).

Table 1. Statistical characteristics of the outcrop mesh and both cube meshes (n_n = number of nodes, n_c = number of cells, V = volume of the mesh, $\overline{V}_c$ = average volume of a mesh cell).

Object	n_n [-]	n_c [-]	V [m^3]	$\overline{V}_c$ [m^3]
Outcrop (Ω)	82,000	75,240	9000	0.12
OSB1_c (Ω_b)	68,921	64,000	0.0156	6.19×10^{-7}
OSB2_c (Ω_b)	31,500	25,230	0.008	1.25×10^{-7}

The variance σ_c^2 derived from the measurements conducted on the samples from the rock cubes is assumed to represent the variance $\sigma_{\Omega_b}^2$ that can be expected in one cell of the outcrop mesh so that

$$\sigma_{LVM}^2 \approx \sigma_{\Omega_b}^2, \tag{20}$$

with σ_{LVM}^2 being the local sample variance, which we can calculate by means of the formula

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2, \tag{21}$$

where n is the total number of samples, μ is the mean and x_i is the sample at the ith location.

4. Results

4.1. Spatial Variability

The variogram analysis reveals a range of 0.3 m and 0.2 m for the rock cube samples OSB1_c and OSB2_c, respectively, and a range of 18 m for the outcrop samples (Figure 5a,d,g). The sill is slightly higher in the outcrop region as it is in the rock cubes. Moreover, the outcrop samples show a weak nugget effect. Generally, a scale effect can be observed in which the variance increases with the considered volume. This effect is also present in the descriptive statistics (Figure 5c,f,i).

The measurements from the outcrop region show a c_v of 0.28 and a c_{dp} of 0.31. The histogram indicates a normal distribution of k ranging from 0.7 mD to 4.6 mD (Figure 5b). A two-sided Kolmogorov–Smirnov test [51], which is based on an implementation of Simard and Ecuyer [52], confirmed the hypothesis that all samples come from a normal distribution. The application of Tukey's

outlier detection method [53] reveals no statistical outliers in the sample. By applying the classification scheme of Corbett and Jensen [54], the sample can be classified as being very homogeneous.

The local histogram of *k* from OSB1_c shows a bimodality in the distrubtion ranging from 0.7 to 3.9 mD (Figure 5e). OSB2_c's histogram shows a unimodal range from 0.8 to 1.5 mD (Figure 5h). Again, no statistical outliers can be detected. The local variability of OSB1_c is significantly higher than that of OSB2_c. *k* of OSB1_c provides a c_v of 0.3 and a c_{dp} of 0.4 while measurements from OSB2_c show values of 0.2 for c_v and 0.18 for c_{dp}. c_v and c_{dp} of OSB1_c tend to cover the variability of the global data. This result is in good agreement with the REV theory from Nordahl and Ringrose [45]. Both rock cubes can be classified being very homogeneous as well.
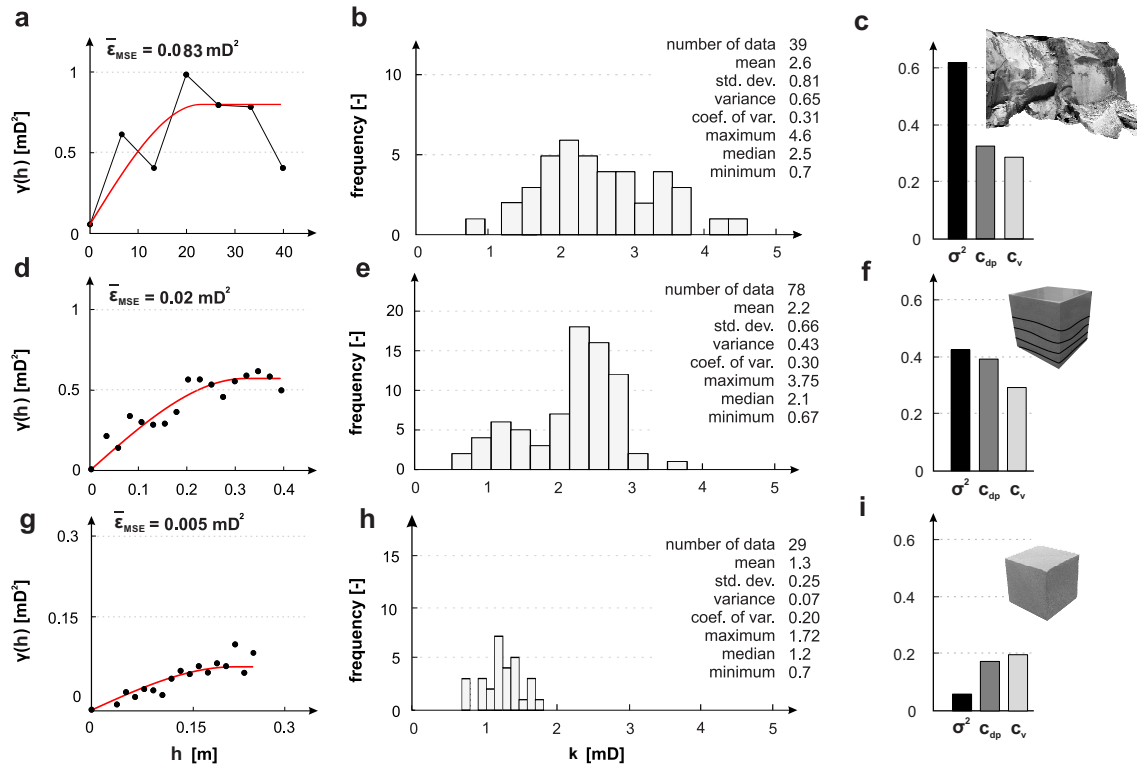
Figure 5. Empirical variogram and variogram model, empirical histogram, and heterogeneity-indexes derived from the *k* measurements for the outcrop (**a**–**c**), and the rock cubes OSB1_c (**d**–**f**) and OSB2_c (**g**–**i**). A scale-effect is observable in the heterogeneity-indicating coefficient of variation, the Dykstra–Parson coefficient and the sample variance. All variogram models are described by a spherical model with nugget effect. The variogram model for (**a**) is described by n = 0.05 mD^2, a = 23 m and b = 0.75 mD^2 with n as nugget, a as range, and b as sill. The model for (**d**) is described by n = 0 mD^2, a = 0.3 m and b = 0.58 mD^2 while the model of (**g**) is described by n = 0.005 mD^2, a = 0.18 m and b = 0.08 mD^2.

Thus, we can observe a significant small-scale variability. The bedding structures in OSB1_c are well preserved in the permeability field of the *k* interpolation, which gradually increases from low values between 0.7 and 2 mD in the lower beds to higher values between 2 and 4 mD in the upper beds (Figure 6a). In OSB2_c the trend is running diagonally through the rock cube (Figure 6b); however, no macroscopic bounding surfaces are visible, which could have had a control on the field of *k* here. It should be noted, however, that the range of *k* is significantly smaller here compared to OSB1_c.
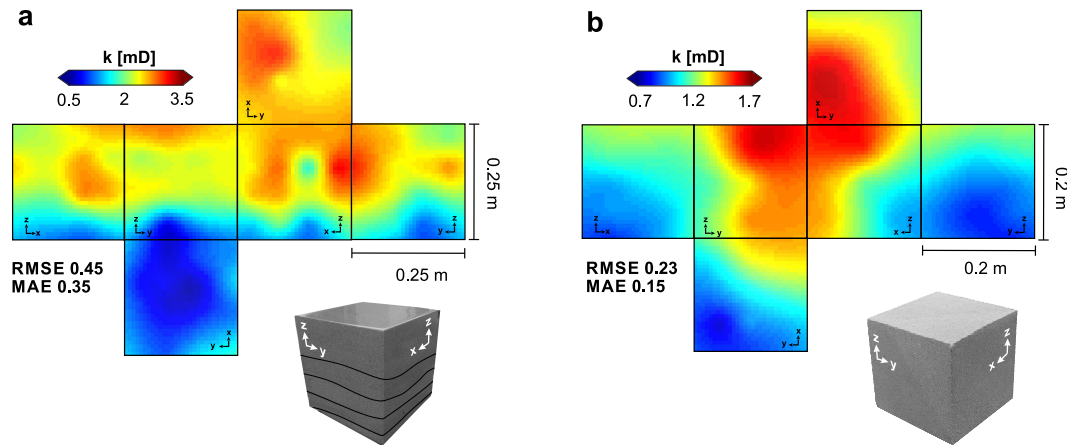
Figure 6. Spatial distribution of the intrinsic permeability in the rock cubes OSB1_c (**a**) [40] and OSB2_c (**b**) interpolated using the SK method.

4.2. Constructing the LVM

The LVM is constructed by means of a 3-D architectural element mapping of both Bouma sequences in the quarry. The base and the top of the sequences are mapped which are being used to constrain the LVM by the locally observable variance σ^2_{LVM}. The exploratory data analysis reveals that the variance of k in OSB1_c is five times larger than that of OSB2_c. This is in accordance with the sedimentological mapping, which indicates a higher heterogeneity at the base of the Bouma sequence.

It is assumed that OSB1_c represents the most heterogeneous and OSB2_c the most homogeneous lithofacies type in the Bouma sequences as it is illustrated in Figure 3b. Accordingly, the positions of those lithofacies types are mapped throughout the quarry area and parameterized with σ^2_{LVM}, which has been determined by the k measurements of OSB1_c and OSB2_c. Thus, we use $\sigma^2_{LVM} = 0.43$ for mapping the base boundaries of the sequences throughout the outcrop area. Likewise, $\sigma^2_{LVM} = 0.07$ is used as a local variance for the topmost boundary of the single sequences. The mapping locations of σ^2_{LVM} are shown in Figure 7a. The points mapped onto the outcrop model are subsequently interpolated onto Ω by using a SK-based interpolation procedure for parametric PDFs (Figure 3c). The interpolation is conducted using 5 neighbors, a range of five meters, a sill of 0.005, a nugget of 0 and a plunge of $10°$ as the strata gently dip towards south. Figure 7b shows the constructed LVM which is being used by the sequential simulation algorithms. It should be noted that we have a decent offset in the LVM in the area of the central fault zones.
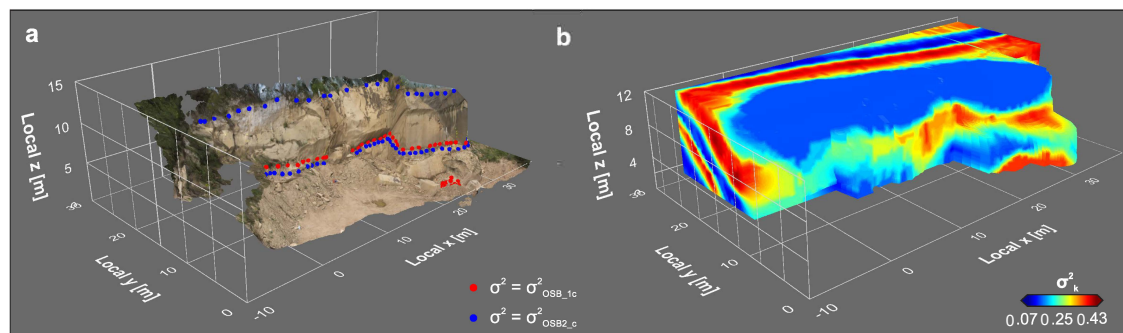
Figure 7. (**a**) Mapping of the local variance with regard to the observed geological structure. The highest variance is indicated by red spheres whereas the lowest variance is indicated by blue ones. The variance is derived from the rock cube measurements of OSB1_c—representing the most heterogeneous lithology at the bottom of the Bouma sequences (red)—and OSB2_c—likewise representing the most homogeneous lithology at the top of the Bouma sequences (blue). (**b**) The 3-D local variance model (LVM) representing the locally observable variance, which is constrained by the mappings shown in (**a**).

4.3. Optimizing the BLUE for Sequential Simulation

Prior to sequential simulation, the optimal SK conditions with regard to the integrated measurement error variance σ_m^2 and the selected neighborhood are determined. Therefore, a simple integer programming optimization is performed using varying measurement error variances ($0.0 \text{ mD}^2 \leq \sigma_m^2 \leq 0.15 \text{ mD}^2$) and a varying number of neighbors ($10 \leq n_n \leq 20$) as inequality constraints. We can express the optimization problem as

$$\min_{\sigma_m^2 \in \mathbb{R}, \, n_n \in \mathbb{N}} \quad \epsilon_{SK}(\sigma_m^2, n_n) \tag{22}$$
$$\text{subject to} \quad 0 \leq \sigma_m^2 \leq 0.15$$
$$10 \leq n_n \leq 20,$$

in which the SK error ϵ_{SK} in form of the RMSE and MAE must be minimized. The response surface of the numerical optimization process indicates that the SK error is generally declining when σ_m^2 is increasing. The lowest errors are produced with an n_n of 10, 11, and 20. This sensitivity of the SK error on the number of neighboring points is not unusual. The numerical optimization reveals that the optimal conditions for SK are met at $n_n = 20$ and $\sigma_m^2 = 0.15$ which yields a RMSE of 0.708 mD (Figure 8). The interpolation error can be reduced by 16.5% for the RMSE and by 18.5% for the MAE. The final SK realization and the spatial distribution of σ_{SK}^2 for that exact model is illustrated in Figure 9. It should be noted that the spatial distribution of σ_{SK}^2 in a sequential simulation is different as previously simulated locations are considered as well.
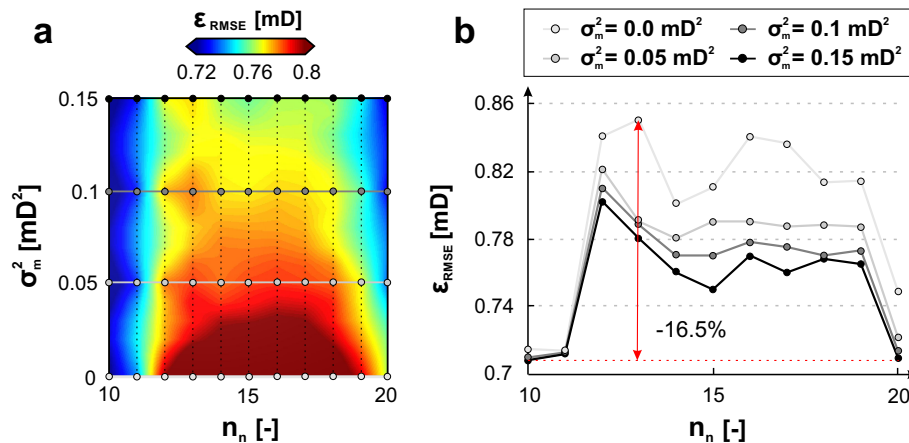
Figure 8. Results of the linear integer programming optimization using the marked sampling points. The interpolation error ϵ_{RMSE} is minimized using the inequality constraints given in Equation (22). (**a**) RMSE response surface with regard to the incorporated measurement error variance σ_m^2 and the maximum number of neighbors n_n using a leave-one-out cross-validation. (**b**) Cross sections through the response surface of (**a**).

The final modeling variables for the sequential simulations are given in Table 2. For SGS and LVM-SGS, the original data are transformed into standard normal space with $\mu = 0$ and $\sigma = 1$. The transformation leads to an adaption of the considered variogram model as the sill is now 1 and not 0.75 with a nugget of 0 instead of 0.05.
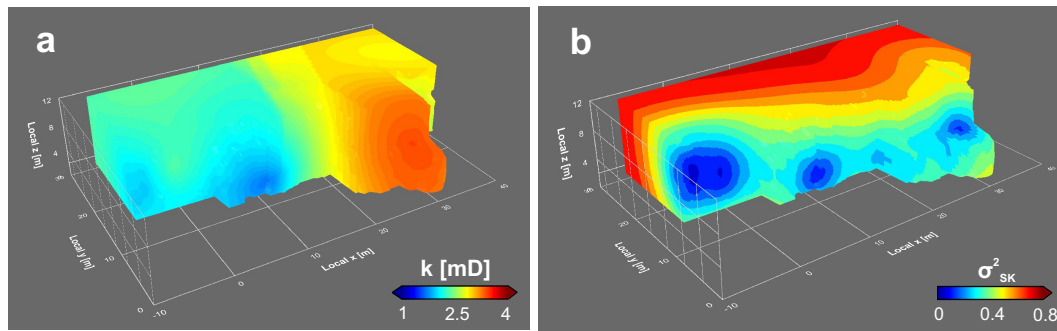
Figure 9. (**a**) Simple kriging estimate (**b**) and the local simple kriging variance for one SK realization.

Table 2. Modeling variables for the sequential simulations.

Variable	SGS & LVM-SGS	DSS & LVM-DSS
BLUE	SK	SK
Normal score transform	yes	no
Quantile-quantile back transform	yes	yes
Range x	50 m	50 m
Range y	50 m	50 m
Range z	15 m	15 m
Nugget	0.05	0
Sill	0.75 mD^2	1 mD^2
Range	23 m	23 m
Max. number of neighbors	20	20
Azimuth	0°	0°
Dip	0°	0°
Plunge	10°	10°
Measurement error variance	0.15 mD^2	0.15 mD^2

4.4. σ^2_{SK} versus σ^2_{LVM}

The statistical and spatial characteristics of σ^2_{SK} and σ^2_{LVM} differ tremendously. σ^2_{SK} is unimodally distributed, whereas σ^2_{LVM} provides a bimodal distribution (Figure 10a). It is evident that σ^2_{SK} covers the total range of the considered covariance model while σ^2_{LVM}'s range is more limited. The probability of simulating variances between 0.2 and 0.43 mD^2 is higher when sampling from the LVM instead of the local SK variance (Figure 10b). The median between σ^2_{LVM} and σ^2_{SK} differs by ≈ 0.08 mD^2, which indicates that the variability simulated in a realization of conventional sequential simulation algorithms is systematically underestimated.

With regard to the variogram model, σ^2_{LVM} has a range of 5 m and a sill of 0.36 mD^2, and σ^2_{SK} has a range of 0.3 m and a sill of 0.1 mD^2. Thus, σ^2_{SK} seems to be spatially uncorrelated and random. However, the grade of variability in the eastern part of the outcrop is slightly higher than in the western part. Therefore, in contrast to σ^2_{LVM}, σ^2_{SK} obviously does not provide the simulation algorithm with a spatial trend when simulating the local variability.
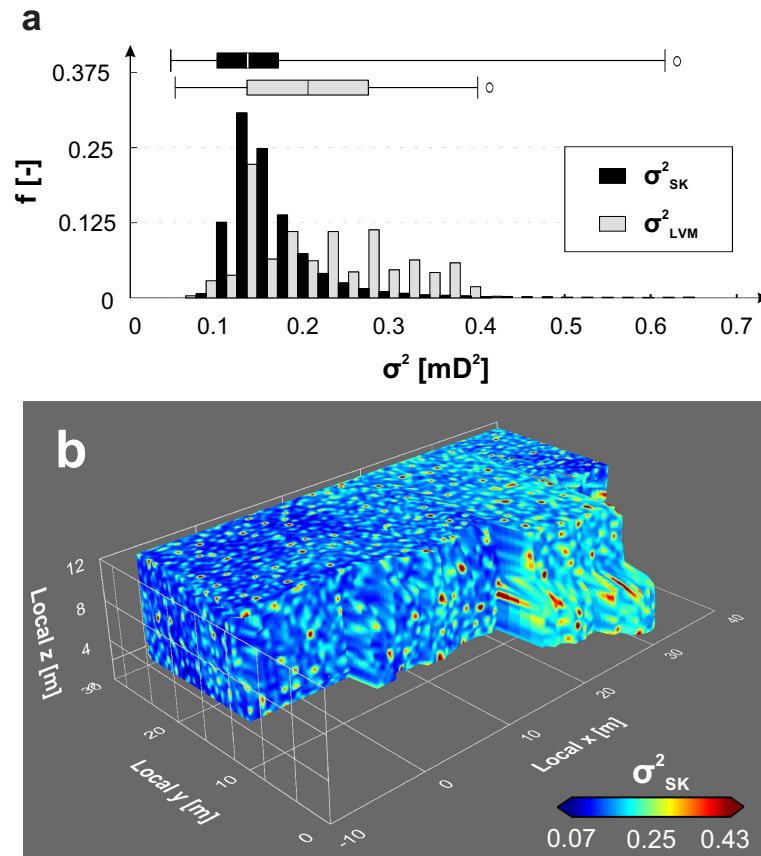
Figure 10. (**a**) Comparison of the empirical histograms of the σ^2_{SK} model produced in a DSS realization with the LVM and (**b**) the empirical distribution of σ^2_{SK} produced in the realization of (**a**).

4.5. Model Validation

All algorithms reproduce the considered variogram model within the range of ergodic fluctuations after back-transformation (Figure 11a–d). The quality of variogram reproduction has been evaluated by calculating the average mean square error $\bar{\epsilon}_{MSE}$ of all realizations between the experimental variogram and the variogram model. The best reproduction is produced by the LVM-DSS and LVM-SGS algorithms, while the latter one provides the lowest degree of ergodic fluctuations with $\bar{\epsilon}_{MSE} = 0.066$ mD^2. All realizations reproduce short-range dissimilarities well but slightly underestimate the dissimilarity at medium ranges. DSS and SGS tend to gentle underestimation at far ranges which is a drawback of limited neighborhoods. This effect, however, is less expressed in the LVM-based algorithms. For both types of sequential simulation, the LVM-based algorithm outperforms the conventional conditional simulation approaches.
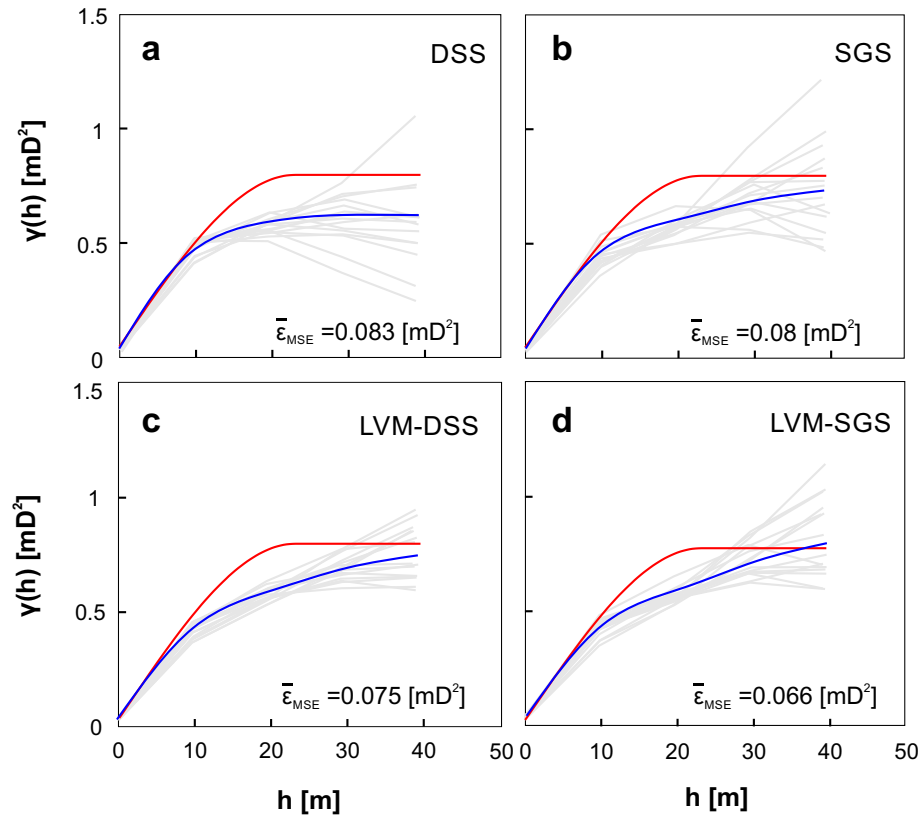
Figure 11. Experimental variograms (gray) for 15 realizations of DSS (**a**), SGS (**b**), LVM-DSS (**c**) and LVM-SGS (**d**) plotted together with the average over all realizations (blue) and the considered variogram model (red), which is described by a nugget of 0.05 mD^2, a range of 23 m and a sill of 0.75 mD^2.

Visual Outputs

It is evident that all simulation algorithms provide visually comparable results (Figure 12). It should be noted that the quadrilaterals of the 3-D models are subdivided using the Catmull–Clark scheme [55] for visualization. Within this scheme, a new point in a quadrilateral is calculated by

$$x_j^{k+1} = \frac{1}{n} \sum_{i=0}^{n-1} x_i^k, \tag{23}$$

with x_j^{k+1} as the new point at subdivision step $k+1$ in the center of the element j with n vertices at the subdivision step k. This technique smooths the observable patterns in the models. There is an obvious trend in all realizations, which indicates that the highest values are located in the eastern part of the quarry and the lowest values in the western part. Having in mind that the applied algorithms are conditional, this trend is in well accordance with the constraints as given by the global measurements, which also provide the highest values in the eastern part of the quarry and the lowest values in the western part (Figure 4a). The trend is most clearly depicted in the DSS and LVM-DSS realizations (Figure 12). SGS and DSS tend to construct homogeneous regions more likely than their LVM equivalents. Thus, those algorithms might indicate a homogeneity, which is likely not present in the strata. Moreover, the heterogeneity of the LVM equivalents is more realistically oriented along the bounding surfaces in the quarry than the models produced by the conventional algorithms.
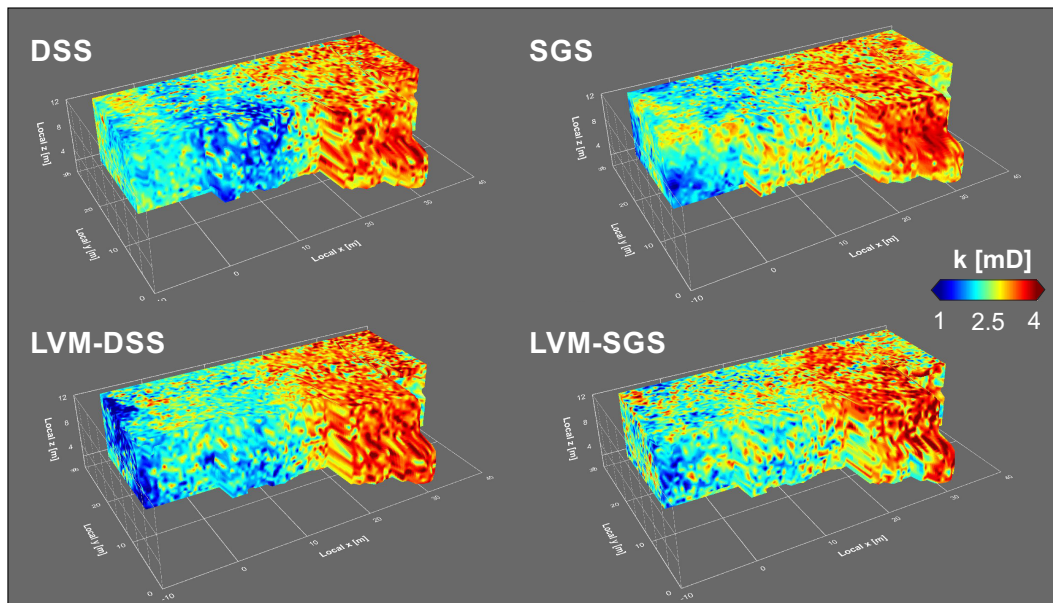
Figure 12. Exemplary model visualizations for the DSS, SGS, LVM-DSS and LVM-SGS realizations.

5. Discussion

In this study, we present a workflow that accounts for the locally observable geological variability in modified versions of conventional sequential simulation algorithms. Our approaches produce similar outputs as the conventional algorithms and reproduce the global variance model together with the global summary statistics, which are important criteria for the validity of a statistical simulation [8,10,39]. Our results are confirming the concept of the REV [45], in which the complexity of a continuous random variable is increasing with reducing the scale of observation. Moreover, we can confirm that σ^2_{SK} constitutes no measure for the local estimation accuracy [56] as it is only reflecting the spatial configuration of the constraining data points being simultaneously independent on the constraints' values [6]. There are, however, two points which must be raised in order to discuss the benefits as well as the drawbacks of our approach.

5.1. Construction of the LVM

The main source of errors in the proposed workflow is based on the construction of the LVM. The LVM has been derived by an integrated approach of measuring the local variability in the most homogeneous and most heterogeneous lithofacies types in the sedimentary succession. The statistical analysis revealed that this assumption proved to be true as the heterogeneity measures in OSB1_c indicate a way higher variability as is present in OSB2_c. This, in fact, is building the basis for this study. The variance has been assumed to be constant at the base and at the top of a Bouma sequence. This assumption is limited by the number of samples taken within this case study. By constructing the LVM with an SK interpolation, we assume that the variance in one sedimentary Bouma sequence is continuous in a mathematical sense. This assumption might be proved to be too simple in future studies. In order to validate those results, more local samples would be necessary to constrain the LVM. This is a drawback in comparison to conventional SGS and DSS algorithms as those are not dependent on estimating a global variability model.

5.2. Comparison of the Spatial Distribution of the Local Variance

Figure 13a,b illustrates the relationship between the LVM and a DSS realization (a) and an LVM-DSS realization, respectively (b). Although the overall trend remains identical among both types of simulation, the spatial distribution of local variability is uncorrelated and inherently different. In the DSS realization, the heterogeneity within the region is randomly distributed. The most heterogeneous

areas in the LVM-DSS realization reside in the light areas—in which σ^2_{LVM} is high—whereas the most homogeneous regions reside in the dark ones—where σ^2_{LVM} is low. As the spatial distribution of σ^2_{SK} is primarily dependent on the distance to the constraining neighbors, the SGS and DSS algorithms, in contrast to their LVM-based modifications, cannot account for a realistic spatial distribution of the local geological variability. This observation is conceptually illustrated in Figure 13c, which shows the spatial relationship between σ^2_{SK} and σ^2_{LVM}, as implied by the results of our study. It is evident that the conventional algorithms underestimate the local geological variability in close ranges to conditional data. It is also evident that σ^2_{SK} systematically underestimates the natural variability present in the geological medium, which is investigated in this study (Figure 10a). Therefore, SGS and DSS might not be able reproduce the total geological variability as shown in this study, which is an advantage of the proposed algorithms instead.
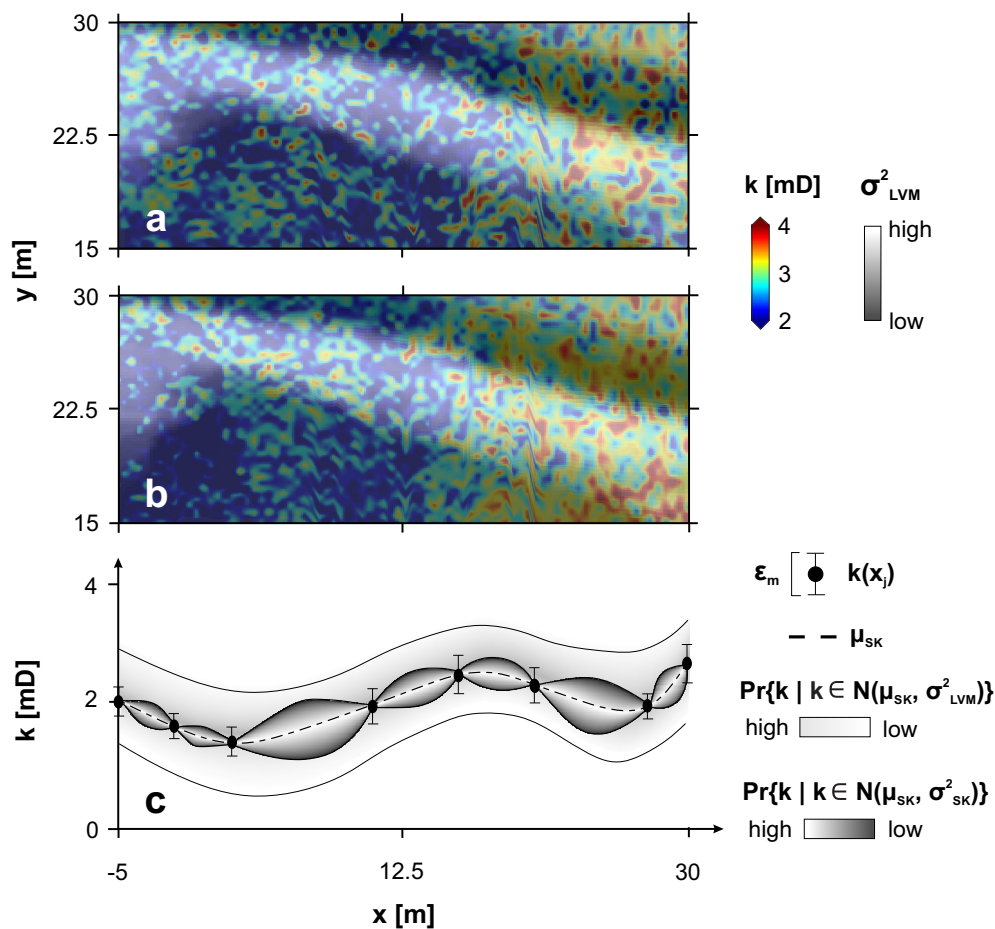
Figure 13. Top-view onto a representative simulation result of DSS (**a**) and LVM-DSS (**b**) superimposed by a gray-scale representation of the LVM with an opacity of 0.6. It is evident that the LVM-based algorithms' heterogeneity is highest in that area of the LVM in which it provides the highest local variance as well. The conventional approach, however, does not reflect the expected variance in space. (**c**) Conceptual illustration showing the spatial distribution of the constraining measurements $k(x_j)$ and the spatial relationship between the simple kriging estimate μ_{SK} with the measurement error ϵ_m and the two parameters used to simulate k in this study namely σ^2_{SK} and σ^2_{LVM}. Pr stands for the probability of k under the condition that k belongs to the Gaussian distribution described by μ_{SK} together with either σ^2_{SK} or σ^2_{LVM}.

6. Conclusions

In this study, we propose a new workflow, which incorporates the locally observed variability σ^2_{LVM} into sequential simulations. We could demonstrate that the local simple kriging variance σ^2_{SK} differs from σ^2_{LVM} in local volumes of the target region. Therefore, the DSS and SGS algorithms have been modified by the replacement of σ^2_{SK} through the measurement-derived σ^2_{LVM} within one mesh cell. This replacement has been done if and only if $\sigma^2_{LVM} \geq \sigma^2_{SK}$. The LVM has been constructed by means of geological mapping and the assumption that the variability is highest in the most heterogeneous lithology and lowest in the least heterogeneous lithology in a Bouma sequence. The proposed approach can be used in any type of spatial property simulation but is especially tailored for geological media.

The LVM-DSS and LVM-SGS approaches reproduce the observed variability in the sedimentary succession adequately yet reproducing the minimum required statistical measures of a valid simulation including the global histogram, the global heterogeneity, and the variogram model. Moreover, in contrast to their conventional representatives, the LVM-based algorithms account for the spatial distribution of the expected local variance adequately. Once the LVM is derived, it may be integrated into other geostatistical simulation algorithms such as the turning bands method [57–60].

From our results we conclude the following.

1. The distance metrics RMSE and MAE in spatial interpolations can be optimized with regard to the measurement error variance and the optimal neighborhood.
2. Geological samples always represent a small subset of the local variability, which should be accounted for by high-resolution sampling at a random basis at the least.
3. The simple kriging variance does not necessarily account for the magnitude of local variability in geological media and definitely does not account for its spatial distribution.
4. The fact that the local simple kriging variance does not reflect a geological trend might lead to unforeseen problems when using sequential simulation-derived models as a basis for subsurface utilization processes because the full geological heterogeneity might not have been taken into account properly.
5. By introducing a measurement-constrained, geology-driven local variance model, the spatial distribution of the variance that is expected in the investigated quarry can be integrated into sequential simulations. This allows to simulate the geological variability, which might be greater than the simulated variability in conventional sequential simulation algorithms.

Future research should focus on comparing σ^2_{SK} and σ^2_{LVM} under the consideration of other physicochemical properties, other geological settings, and other scales. This might require adapting the assumptions on the spatial continuity of the variability which should, however, always be based on reliable geological analyses.

Abbreviations

The following abbreviations are used in this manuscript:

CDF	Cumulative distribution function
DSS	Direct sequential simulation
LVM	Local variance model
LpO CV	Leave-p-out cross-validation
MAE	Mean-absolute-error
MSE	Mean-square error
PDF	probability density function
REV	Representative elementary volume
RMSE	Root-mean-square error
RF	Random function
RV	Random variable
SGS	Sequential Gaussian simulation
SK	Simple kriging

Appendix A. Code and Data Availability

GeoReVi is an open-source software for Windows systems available under https://github.com/ApirsAL/GeoReVi. Data is available under https://www.doi.org/10.6084/m9.figshare.11791407.v2.

References

1. Li, H.; Reynolds, J. On Definition and Quantification of Heterogeneity. *Oikos* **1995**, *73*, 280–284. [CrossRef]
2. Crooijmans, R.A.; Willems, C.J.L.; Nick, H.M.; Bruhn, D.F. The influence of facies heterogeneity on the doublet performance in low-enthalpy geothermal sedimentary reservoirs. *Geothermics* **2016**, *64*, 209–219. [CrossRef]
3. Rodrigo-Ilarri, J.; Reisinger, M.; Gómez-Hernández, J.J. Influence of Heterogeneity on Heat Transport Simulations in Shallow Geothermal Systems. In *Geostatistics Valencia 2016*; Springer: Berlin, Germany, 2017; pp. 849–862. [CrossRef]
4. Shepard, D. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. In Proceedings of the 1968 ACM National Conference, New York, NY, USA, 27–29 August 1968; pp. 517–524. [CrossRef]
5. Webster, R.; Margaret, A.O. *Geostatistics for Environmental Scientists*, 2nd ed.; Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; p. 330.
6. Deutsch, C.V.; Journel, A. *GSLIB: Geostatistical Software Library and User's Guide*; Oxford University Press: Oxford, UK, 1998.
7. Tran, T.T. Improving variogram reproduction on dense simulation grids. *Comput. Geosci.* **1994**, *20*, 1161–1168. [CrossRef]
8. Leuangthong, O.; McLennan, J.A.; Deutsch, C.V. Minimum Acceptance Criteria for Geostatistical Realizations. *Nat. Resour. Res.* **2004**, *13*, 131–141. [CrossRef]
9. Soares, A. Direct Sequential Simulation and Cosimulation. *Math. Geol.* **2001**, *33*, 911–926. [CrossRef]
10. Robertson, R.K.; Mueller, U.A.; Bloom, L.M. Direct sequential simulation with histogram reproduction: A comparison of algorithms. *Comput. Geosci.* **2006**, *32*, 382–395. [CrossRef]
11. Journel, A.; Alabert, F. Non-Gaussian data expansion in the Earth Sciences. *Terra Nova* **1989**, *1*, 123–134. [CrossRef]
12. Verly, G. Sequential Gaussian Simulation: A Monte Carlo Method for Generating Models of Porosity and Permeability. In *Generation, Accumulation and Production of Europe's Hydrocarbons III*; Spencer, A.M., Ed.; Springer: Berlin/Heidelberg, Germany, 1993; pp. 345–356.
13. Ersoy, A.; Yünsel, T.Y. Geostatistical Conditional Simulation for the Assessment of the Quality Characteristics of Cayırhan Lignite Deposits. *Energy Explor. Exploit.* **2006**, *24*, 391–416. [CrossRef]
14. Delbari, M.; Afrasiab, P.; Loiskandl, W. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *CATENA* **2009**, *79*, 163–169. [CrossRef]

15. Pinheiro, M.; Emery, X.; Miranda, T.; Lamas, L.; Espada, M. Modelling Geotechnical Heterogeneities Using Geostatistical Simulation and Finite Differences Analysis. *Minerals* **2018**, *8*, 52. [CrossRef]

16. Box, G.E.P.; Muller, M.E. A Note on the Generation of Random Normal Deviates. *Ann. Math. Statist.* **1958**, *29*, 610–611. [CrossRef]

17. Matheron, G. Principles of Geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [CrossRef]

18. Anyiam, O.A.; Andrew, P.J.; Okwara, I.C. Assessment of the heterogeneity and petrophysical evaluation of reservoirs in the "Akbar Field", Niger Delta, Nigeria. *J. Pet. Explor. Prod. Technol.* **2017**, *7*, 1035–1050. [CrossRef]

19. Michie, E.A.H.; Haines, T.J. Variability and heterogeneity of the petrophysical properties of extensional carbonate fault rocks, Malta. *Pet. Geosci.* **2016**, *22*, 136–152. [CrossRef]

20. Mukerji, T.; Mavko, G.; Rio, P. Scales of Reservoir Heterogeneities and Impact of Seismic Resolution on Geostatistical Integration. *Math. Geol.* **1997**, *29*, 933–950. [CrossRef]

21. De Ros, L.F.; Scherer, C.M.S. Stratigraphic Controls on the Distribution of Diagenetic Processes, Quality and Heterogeneity of Fluvial-Aeolian Reservoirs from the Recôncavo Basin, Brazil. In *Linking Diagenesis to Sequence Stratigraphy*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013; pp. 105–132. [CrossRef]

22. Oxford English Dictionary. *Heterogeneity*; Oxford University Press: Oxford, UK, 2014.

23. Fitch, P.J.R.; Lovell, M.A.; Davies, S.J.; Pritchard, T.; Harvey, P.K. An integrated and quantitative approach to petrophysical heterogeneity. *Mar. Pet. Geol.* **2015**, *63*, 82–96. [CrossRef]

24. Remy, N.; Boucher, A.; Wu, J. *Applied Geostatistics with SGeMS: A User's Guide*; Cambridge University Press: Cambridge, UK, 2009. [CrossRef]

25. Wackernagel, H. *Multivariate Geostatistics*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2003; p. 388. [CrossRef]

26. Armstrong, M. Experimental Variograms. In *Basic Linear Geostatistics*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 47–58. [CrossRef]

27. Ringrose, P.; Bentley, M. *Reservoir Model Design*, 1st ed.; Springer: Dordrecht, The Netherlands, 2015; p. 249. [CrossRef]

28. Gu, Y.; Rühaak, W.; Bär, K.; Sass, I. Using seismic data to estimate the spatial distribution of rock thermal conductivity at reservoir scale. *Geothermics* **2017**, *66*, 61–72. [CrossRef]

29. Rühaak, W.; Guadagnini, A.; Geiger, S.; Bär, K.; Gu, Y.; Aretz, A.; Homuth, S.; Sass, I. Upscaling thermal conductivities of sedimentary formations for geothermal exploration. *Geothermics* **2015**, *58*, 49–61. [CrossRef]

30. Rühaak, W. 3-D interpolation of subsurface temperature data with measurement error using kriging. *Environ. Earth Sci.* **2015**, *73*, 1893–1900. [CrossRef]

31. Bailey, T.; Gatrell, A. *Interactive Spatial Data Analysis*; Longman Group Limited: Harlow, UK, 1995; p. 432.

32. Journel, A.G. Nonparametric estimation of spatial distributions. *J. Int. Assoc. Math. Geol.* **1983**, *15*, 445–468. [CrossRef]

33. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: Oxford, UK, 1997.

34. Remy, N. *Algorithmic and Software Methods for a Better Integration of the Geological Information into Numerical Models*; Standford University: Stanford, CA, USA, 2004.

35. Ökten, G.; Göncü, A. Generating low-discrepancy sequences from the normal distribution: Box—Muller or inverse transform? *Math. Comput. Model.* **2011**, *53*, 1268–1281. [CrossRef]

36. Celisse, A. Optimal cross-validation in density estimation with the L2-loss. *Ann. Stat.* **2014**, *42*, 1879–1910. [CrossRef]

37. Willmott, C.J.; Matsuura, K.; Robeson, S.M. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **2009**, *43*, 749–752. [CrossRef]

38. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]

39. Emery, X. Testing the correctness of the sequential algorithm for simulating Gaussian random fields. *Stoch. Environ. Res. Risk Assess.* **2004**, *18*, 401–413. [CrossRef]

40. Linsel, A.; Wiesler, S.; Hornung, J.; Hinderer, M. High-Resolution Analysis of the Physicochemical Characteristics of Sandstone Media at the Lithofacies Scale. *Solid Earth Discuss.* **2020**, *2020*, 1–28. [CrossRef]

41. Linsel, A. ApirsAL/GeoReVi: GeoReVi v1.0.0 Pre-Release. Available online: https://zenodo.org/record/3541136#.XvhBb3ERWUk (accessed on 20 December 2019). [CrossRef]

42. Becker, A.; Schwarz, M.; Schäfer, A. Lithostratigraphische Korrelation des Rotliegend im östlichen Saar-Nahe-Becken. *Jahresberichte Und Mitteilungen Des Oberrheinischen Geologischen Vereins* **2012**, *94*, 105–133. [CrossRef]

43. Aretz, A.; Bär, K.; Götz, A.E.; Sass, I. Outcrop analogue study of Permocarboniferous geothermal sandstone reservoir formations (northern Upper Rhine Graben, Germany): Impact of mineral content, depositional environment and diagenesis on petrophysical properties. *Int. J. Earth Sci.* **2015**, *105*, 1431–1452. [CrossRef]

44. Farkhutdinov, A.; Goblet, P.; de Fouquet, C.; Cherkasov, S. A case study of the modeling of a hydrothermal reservoir: Khankala deposit of geothermal waters. *Geothermics* **2016**, *59*, 56–66. [CrossRef]

45. Nordahl, K.; Ringrose, P.S. Identifying the Representative Elementary Volume for Permeability in Heterolithic Deposits Using Numerical Rock Models. *Math. Geosci.* **2008**, *40*, 753. [CrossRef]

46. Nordahl, K.; Messina, C.; Berland, H.; Rustad, A.B.; Rimstad, E.; Martinius, A.W.; Howell, J.A.; Good, T.R. Impact of multiscale modelling on predicted porosity and permeability distributions in the fluvial deposits of the Upper Lunde Member (Snorre Field, Norwegian Continental Shelf). In *Sediment-Body Geometry and Heterogeneity: Analogue Studies for Modelling the Subsurface*; Geological Society of London: London, UK, 2014; Volume 387, p. 25. [CrossRef]

47. Middleton, G.V. Sediment Deposition from Turbidity Currents. *Annu. Rev. Earth Planet. Sci.* **1993**, *21*, 89–114. [CrossRef]

48. Agemar, T.; Weber, J.; Schulz, R. Deep Geothermal Energy Production in Germany. *Energies* **2014**, *7*, 4397–4416. [CrossRef]

49. Filomena, C.M.; Hornung, J.; Stollhofen, H. Assessing accuracy of gas-driven permeability measurements: A comparative study of diverse Hassler-cell and probe permeameter devices. *Solid Earth* **2014**, *5*, 1–11. [CrossRef]

50. Bär, K. *Untersuchung der tieFengeothermischen Potenziale von Hessen*; Technische Universität Darmstadt: Darmstadt, Germany, 2012; p. 297.

51. Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [CrossRef]

52. Simard, R.; Ecuyer, P. Computing the Two-Sided Kolmogorov-Smirnov Distribution. *J. Stat. Softw.* **2011**, *1*. [CrossRef]

53. Tukey, J. *Exploratory Data Analysis*; Addison-Wesley: Reading, MA, USA, 1977; p. 712.

54. Corbett, P.; Jensen, J.L. Estimating the mean permeability: How many measurements do you need? *First Break* **1992**, *10*, 5. [CrossRef]

55. Catmull, E. *A Subdivision Algorithm for Computer Display of Curved Surfaces*; University of Utah: Salt Lake City, Utah, 1974.

56. Journel, A.G. Geostatistics: Models and tools for the earth sciences. *Math. Geol.* **1986**, *18*, 119–140. [CrossRef]

57. Matheron, G. The intrinsic random functions and their applications. *Adv. Appl. Probab.* **1973**, *5*, 439–468. [CrossRef]

58. Mantoglou, A.; Wilson, J.L. The Turning Bands Method for simulation of random fields using line generation by a spectral method. *Water Resour. Res.* **1982**, *18*, 1379–1394. [CrossRef]

59. Emery, X.; Lantuéjoul, C. TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. *Comput. Geosci.* **2006**, *32*, 1615–1628. [CrossRef]

60. Paravarzar, S.; Emery, X.; Madani, N. Comparing sequential Gaussian and turning bands algorithms for cosimulating grades in multi-element deposits. *Comptes Rendus Geosci.* **2015**, *347*, 84–93. [CrossRef]

4. Conclusions

In this chapter, the contributions of this dissertation to the ongoing development of open-source software solutions with regard to addressing geoscientific problems are recapitulated. Moreover, the most significant findings from the case studies are presented and discussed.

Development of an Open-Source Knowledge Discovery System for Subsurface Characterization

This work was motivated by the uprising demand in geosciences and related industries for open-source software systems, which are aiding the geoscientific knowledge discovery process. Therefore, a software system named GeoReVi has been implemented, which had been validated in three comprehensive real-world applications. The case studies were concerned with scientific investigations of geological media and the evaluation of potential reservoir rocks on multiple scales for both petroleum and geothermal systems.

The desktop-based client has been published by Linsel et al. (2020) (**Publication I**). GeoReVi bridges a significant technology-gap in the geoscientific open-source community by covering numerous competences involved in the subsurface characterization process chain. Hereby, the integrated software system combines a series of technical and scientific disciplines from geovisual analytics including geoscientific data management, geological modeling, spatiotemporal data mining, and uncertainty analysis. The generic data structures of spatial and temporal elements, as they are implemented in GeoReVi, facilitate geoscientists to address a wide range of problems which they might encounter during a subsurface study.

GeoReVi's modular software architecture is scalable from both a technical perspective as well as from an ontological one by providing algorithms which scale flexibly yet providing an easily-extendable geoscientific data model. By complying with standards being considered good practice in software engineering, such as the SOLID principles, object oriented design, unit testing, benchmarking, logical data modeling and clean

code/clean architecture (Martin 2008; Martin 2017), it was aimed to both provide a reasonable quality of the software and to simplify the contribution process for other developers.

The software platform has been designed by using a state-of-the-art software architecture including a graphical user interface (GUI), a business logic (BL) and a data access layer (DAL). Loose coupling among these components make them reusable for other target platforms such as mobile or web. Moreover, the GUI enables software users who are not knowledgeable in any coding language to apply complex analysis on geoscientific data sets. This, in fact, differentiates GeoReVi from other geological modeling tools (e.g., de la Varga et al. 2019) which are currently developed and which require the user to be familiar with a specific programming or scripting language.

Table 4.1 provides a comparison among different desktop-based software systems which are related to subsurface characterization. It is evident that most software packages are specifically suited for a range of specific problems within the domain of subsurface characterization. By developing a suitable software architecture, however, a huge fraction of the entire process chain of subsurface characterization can be performed within one software environment.

Multi-Scale Investigations of Sedimentary Geological Media

In the present work, the software system has been validated in real-world scenarios including multi-scale analyses of sedimentary rock media. Hereby, >1,000 rock samples were extracted from quarries and drill cores and measured in the laboratory resulting in a database, which consists of >20,000 discrete readings. These data have been stored, analyzed and visualized by the means of geovisual analytics, as they are provided by GeoReVi, resulting in three peer-reviewed publications (Hornung et al. 2020; Linsel et al. 2020a; Linsel et al. 2020b).

Within those publications, in which small-scale geosystems have been investigated in terms of physical and geochemical rock properties (**Publications II** and **III**), the following findings could be made:

1. A hierarchical set of petrophysical heterogeneities exists in sedimentary environments, spanning from the bed scale to the depositional environment scale.
2. Primary depositional textures and the particle size/shape, rather than the type of lithofacies, govern petrophysical properties in those clastic depositional systems which had been investigated.

Functionality	Software					
	GeoReVi	SGemS	TetGen	GemPy	JeoStat	Past3
Integrated data management	✓	✗	✗	✗	✗	✗
Non-orthogonal mesh generation	✓	✗	✓	✗	✗	✗
Lithological modeling	✓	✓	✗	✓	✗	✗
Fault modeling	✓	✗	✗	✓	✗	✗
Property modeling/simulation	✓	✓	✗	✓	✓	✓
Cross validation	✓	✗	✗	✗	✓	✗
GUI	✓	✓	✓	✗	✓	✓
Multivariate statistics	✓	✓	✗	✗	✗	✓
FEM	✓	✗	✗	✗	✗	✗
Scripting	✗	✓	✓	✓	✗	✗
Open source	✓	✓	✓	✓	✗	✗
Cross platform	✗	✓	✓	✓	✓	✓
Reference	Linsel et al. (2020)	Remy et al. (2009)	Si (2015)	de la Varga et al. (2019)	Mert & Dag (2017)	Hammer et al. (2001)

Table 4.1.: Comparison between different non-commercial software products which implement at least one entity of the geovisual analytics technology group in the context of subsurface characterization as per the current state of the art. ✗= not supported; ✓= partially supported; ✓= fully supported.

3. In contrast, the investigated samples from the carbonate ramp exhibit a greater heterogeneity than those from the siliciclastic depositional environment. Moreover, spatial distribution of the petrophysical properties of the biochemical sedimentary rocks is more dominantly controlled by diagenetic processes than by depositional processes.

4. The strength of statistical correlation can be preserved in spatial interpolations as long as the sampling density is sufficient. If the sampling density is too low, a statistical correlation might be inadvertently feigned.

5. The local geological variability should not be underestimated as an uncertainty factor in spatial predictions and upscaling procedures. In fact, the local geological variability of physicochemical properties might nearly cover the variability being present in an entire formation. Therefore, a high-resolution analysis of physicochemical rock properties can assist in assessing the uncertainty of field-scale property models which is induced by the local geological variability at the lithofacies scale.

There is an increasing interest within the geoscientific community to investigate small-scale geosystems in order to enhance the understanding of the geological controls onto petrophysical heterogeneity. In the study of Heidsiek et al. (2020), which had been published shortly after our study (Linsel et al. 2020a), Upper Rotliegend sandstones were investigated too, however, with special regard being paid to the diagenetic control factors.

The study revealed that the reservoir properties, namely, porosity and permeability of the investigated medium are primarily controlled by calcite cements, illite/iron oxide coatings and infiltrated clay. These findings are in well accordance with ours, implicating that the small-scale variability of rock medias' properties is a function of depositional conditions and the diagenetic history.

Heidsiek et al. (2020), however, did not provide such a wide range of considered properties as compared to Linsel et al. (2020) nor did they account for the properties' anisotropy or multivariate statistical relationships. In fact, GeoReVi could assist researchers performing such studies to regard these characteristics too.

Accounting for Local Geological Variability in Sequential Simulations

Point five raised in the previously listed enumeration guides to **Publication IV**, which is concerned with accounting for the local geological variability in sequential geostatistical simulation techniques. We demonstrated that the local simple kriging variance does not necessarily reflect the geological variability being present at subsets of the target domain, which led us to the development of two modified versions of the sequential Gaussian

simulation (SGS) and direct sequential simulation (DSS) algorithms to account for the local geological variability in the target domain.

The LVM-DSS and LVM-SGS approaches, which utilize a local variance model (LVM) to simulate the local variability, reproduce the observed variability in the sedimentary succession adequately yet reproducing the minimum required statistical measures of a valid simulation including the global histogram, the global heterogeneity, and the variogram model. Moreover, in contrast to their conventional representatives, the LVM-based algorithms account for the spatial distribution of the expected local variance adequately.

Although the proposed modifications of the SGS and DSS algorithms have yielded promising results, the proposed technique should be implemented with other subsurface properties, in other geological settings and at different scales in order to verify its suitability to generically address the problem of local geological variability. Apart from classical subsurface applications such as mining or reservoir exploitation, the proposed method might prove to be useful for, e.g, remote sensing analyses, GIS analyses, or atmospheric simulations, too.

5. Perspectives

The main outcome of this Thesis is the first extensible, open-source, community-open software system for geological knowledge discovery in the context of geovisual analytics. Although GeoReVi provides a base for solving numerous varieties of geoscientific problems, there is space for further improvements.

One of these improvements, for instance, would involve the development of a scripting tool and a functionality to automatically execute command files as it is implemented in SGemS (Remy et al. 2009) and by default given in GemPy (de la Varga et al. 2019). Moreover, the coverage of code fragments by unit tests (see Appendix A.2.2) should be increased in order to ensure a maximum degree of reliability.

The generic structure of GeoReVi allows for the seamless extension of the system by algorithms which aim at solving numerical simulation problems. In this regard, the first functionality for solving heat transport problems was integrated into GeoReVi, the theoretical background of which is outlined in section 2.5.3. Hereby, GeoReVi yields comparable results as other numerical solvers such as that from Simpson (2017) (Fig. 5.1a and b). Moreover, GeoReVi shows a well fit to the analytical solution of Carslaw & Jaeger (1959) (Fig. 5.1c). The next logical step would involve the extension of GeoReVi to address other physical problems such as consolidation, mass transport and coupled simulations.

In the current form GeoReVi is restricted to modeling continuous geological media in space and cannot account for complex geological structures such as fracture systems, reverse faults, or overturned folds adequately. Geological 3-D modeling can, however, be easily improved by extending the geostatistical methods by potential field-based algorithms (Lajaunie et al. 1997).

The separation of domains as it is provided in Table 4.1 predestines GeoReVi to be implemented in a modern service-oriented architecture. This could be performed in the way of providing granular microservices for each separable business process (see,

e.g., Bucchiarone et al. (2020)), what is recently performed by developing a cloud-hosted version of GeoReVi[4]. With this architecture, GeoReVi may serve the geoscientific community in future by providing an important knowledge repository and database for education, industries and sciences.
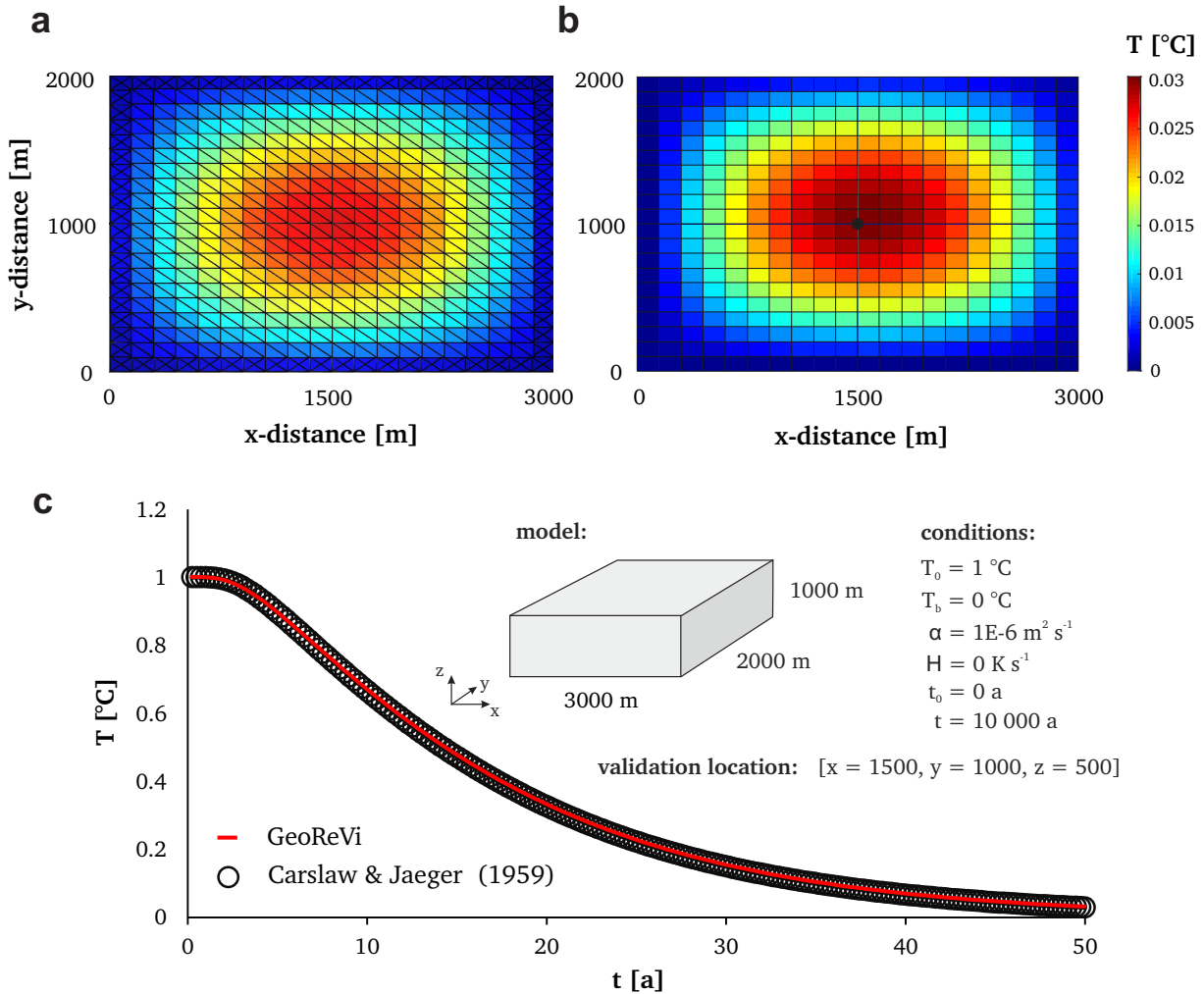
Figure 5.1.: Results of solving a 3-D heat conduction problem. X-Y plane projection of the final cooling model of a rock cuboid measuring $3,000 \times 2,000 \times 1,000 \, m$ in x, y, and z direction from GeoReVi **(a)** and the code provided by Simpson (2017) **(b)**. **(c)** Comparison of the numerical solution of GeoReVi with the analytical solution of Carslaw & Jaeger (1959) calculated for the center of the rock cuboid model at x = 1,500 m, y = 1,000 m, and z = 500 m.

---

[4]https://georevi.com, last access 12th October, 2020

Bibliography

Abbaszadeh Shahri, A., Maghsoudi Moud, F., & Mirfallah Lialestani, S. P. (2020). *A hybrid computing model to predict rock strength index properties using support vector regression*. In: *Engineering with Computers*. ISSN: 1435-5663. DOI: `10.1007/s00366-020-01078-9`. URL: `https://doi.org/10.1007/s00366-020-01078-9`.

Ackoff, R. L. (1999). *Ackoff's Best*. New York: John Wiley & Sons, Inc., p. 356.

Ailin, J. (2012). *Advances and Challenges of Reservoir Characterization: A Review of the Current State-of-the-Art*. In: *Earth Sciences*. DOI: `10.5772/26404`. URL: `https://doi.org/10.5772/26404`.

Andrienko, G. et al. (2007). *Geovisual analytics for spatial decision support: Setting the research agenda*. In: *International Journal of Geographical Information Science* 21.8, pp. 839–857. ISSN: 1365-8816. DOI: `10.1080/13658810701349011`. URL: `https://doi.org/10.1080/13658810701349011`.

Andrienko, G. et al. (2010). *Space, time and visual analytics*. In: *International Journal of Geographical Information Science* 24.10, pp. 1577–1600. ISSN: 1365-8816. DOI: `10.1080/13658816.2010.508043`. URL: `https://doi.org/10.1080/13658816.2010.508043`.

ANSYS, Inc. (2020). *ANSYS® Academic Research Mechanical, Release 18.1, Help System, Coupled Field Analysis Guide*. Computer Program. URL: `www.ansys.com`.

Anyiam, O. A., Andrew, P. J., & Okwara, I. C. (2017). *Assessment of the heterogeneity and petrophysical evaluation of reservoirs in the "Akbar Field", Niger Delta, Nigeria*. In: *Journal of Petroleum Exploration and Production Technology* 7.4, pp. 1035–1050. ISSN: 2190-0566. DOI: `10.1007/s13202-017-0361-z`. URL: `https://doi.org/10.1007/s13202-017-0361-z`.

Aretz, A. et al. (2015). *Outcrop analogue study of Permocarboniferous geothermal sandstone reservoir formations (northern Upper Rhine Graben, Germany): impact of mineral content, depositional environment and diagenesis on petrophysical properties*. In: *International Journal of Earth Sciences* 105.5, pp. 1431–1452. ISSN: 1437-3262. DOI: `10.1007/s00531-015-1263-2`. URL: `http://dx.doi.org/10.1007/s00531-015-1263-2http://link.springer.com/article/10.1007%2Fs00531-015-1263-2`.

Awad, E. & Ghaziri, H. (2007). *Knowledge Management*. Pearson Education India.

Backeberg, N. R. et al. (2017). *Quantifying the anisotropy and tortuosity of permeable pathways in clay-rich mudstones using models based on X-ray tomography*. In: *Scientific Reports* 7.1, p. 14838. ISSN: 2045-2322. DOI: `10.1038/s41598-017-14810-1`. URL: `https://doi.org/10.1038/s41598-017-14810-1`.

Baker Hughes (2020). *World Rig Count*. URL: `https://rigcount.bakerhughes.com/static-files/46fb1985-cd0f-48b4-b2d4-be14347e29e2` (visited on 01/22/2020).

Bär, K., Reinsch, T., & Bott, J. (2020). *The PetroPhysical Property Database (P3) – a global compilation of lab-measured rock properties*. In: *Earth Syst. Sci. Data* 12.4, pp. 2485–2515. ISSN: 1866-3516. DOI: `10.5194/essd-12-2485-2020`. URL: `https://essd.copernicus.org/articles/12/2485/2020/`.

Brachman, R. J. & Anand, T. (1994). *The Process of Knowledge Discovery in A First Sketch*. Report.

Brandsegg, K. B., Hammer, E., & Sinding-Larsen, R. (2010). *A Comparison of Unstructured and Structured Principal Component Analyses and their Interpretation*. In: *Natural Resources Research* 19.1, pp. 45–62. ISSN: 1573-8981. DOI: `10.1007/s11053-010-9110-4`. URL: `http://dx.doi.org/10.1007/s11053-010-9110-4`.

Brazil, E. V. et al. (2018). *Visual Analytics for Reservoir Analogues*. In: *AAPG 2018 Annual Convention & Exhibition*. AAPG, p. 11.

Bucchiarone, A. et al. (2020). *Microservices – Science and Engineering*. Springer. DOI: `10.1007/978-3-030-31646-4`.

Cannon, S. (2018). *Data Collection and Management*. In: *Reservoir modelling: A practical guide*. Ed. by S. Cannon. John Wiley & Sons Ltd, pp. 31–45. DOI: `10.1002/9781119313458.ch2`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119313458.ch2`.

Canon, S. (2018). *Property Model*. In: *Reservoir modelling: A practical guide*. Ed. by S. Cannon, pp. 115–156. DOI: `10.1002/9781119313458.ch6`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119313458.ch6`.

Carslaw, H. S. & Jaeger, J. C. (1959). *Conduction of Heat in Solids*. Second Edition. Oxford: Oxford University Press, p. 517. ISBN: 0198533683.

Catmull, E. (1974). *A Subdivision Algorithm for Computer Display of Curved Surfaces*. University of Utah.

Celisse, A. (2014). *Optimal cross-validation in density estimation with the L2-loss*. In: *Ann. Statist.* 42.5, pp. 1879–1910. ISSN: 0090-5364. DOI: `10.1214/14-AOS1240`. URL: `https://projecteuclid.org:443/euclid.aos/1410440628`.

Chai, T. & Draxler, R. R. (2014). *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature*. In: *Geosci. Model Dev.* 7.3, pp. 1247–1250. ISSN: 1991-9603. DOI: `10.5194/gmd-7-1247-2014`. URL: `https://www.geosci-model-dev.net/7/1247/2014/`.

Chen, P. P.-S. (1976). *The entity-relationship model–toward a unified view of data*. In: *ACM Trans. Database Syst.* 1.1, pp. 9–36. ISSN: 0362-5915. DOI: `10.1145/320434.320440`.

Clavaud, J.-B. et al. (2008). *Permeability anisotropy and its relations with porous medium structure*. In: *Journal of Geophysical Research: Solid Earth* 113.B1. ISSN: 2156-2202. DOI: `10.1029/2007JB005004`. URL: `http://dx.doi.org/10.1029/2007JB005004`.

Codd, E. F. (1970). *A relational model of data for large shared data banks*. In: *Commun. ACM* 13.6, pp. 377–387. ISSN: 0001-0782. DOI: `10.1145/362384.362685`.

Collinson, J. D. (1969). *The sedimentology of the Grindslow shales and the Kinderscout grit; a deltaic complex in the Namurian of northern England*. In: *Journal of Sedimentary Research* 39.1, pp. 194–221. ISSN: 1527-1404. DOI: `10.1306/74d71c17-2b21-11d7-8648000102c1865d`. URL: `https://dx.doi.org/10.1306/74D71C17-2B21-11D7-8648000102C1865D`.

Colombera, L. et al. (2012). *A database approach for constraining stochastic simulations of the sedimentary heterogeneity of fluvial reservoirs*. In: *AAPG Bulletin* 96.11, 2143–2166. DOI: `https://doi.org/10.1306/04211211179`.

Colombera, L. et al. (2016a). *Geological Modeling of Outcrop Successions to Assess Analog-Based Predictions of the Sedimentary Heterogeneity in Fluvial Reservoirs*. In: *AAPG 2015 Annual Convention and Exhibition*.

Colombera, L. et al. (2016b). *The Shallow-Marine Architecture Knowledge Store: a database for the characterization of shallow-marine and paralic depositional systems*. In: *Marine and Petroleum Geology* 75, pp. 83–99. DOI: `https://doi.org/10.1016/j.marpetgeo.2016.03.027`. URL: `http://eprints.whiterose.ac.uk/97154/`.

Das, M., Bin Naharuddin, Z. H., & Al-Khalid, D. (2015). *Optimizing Production and Operation through Automated Well and Network Modeling*. Report. Society of Petroleum Engineers. URL: `https://www.slb.com/-/media/files/premium-content/technical-paper/170/175205.ashx`.

de la Varga, M., Schaaf, A., & Wellmann, F. (2019). *GemPy 1.0: open-source stochastic geological modeling and inversion*. In: *Geosci. Model Dev.* 12.1, pp. 1–32. ISSN: 1991-9603. DOI: `10.5194/gmd-12-1-2019`. URL: `https://gmd.copernicus.org/articles/12/1/2019/`.

De Ros, L. F. & Scherer, C. M. S. (2013). *Stratigraphic Controls on the Distribution of Diagenetic Processes, Quality and Heterogeneity of Fluvial-Aeolian Reservoirs from the Recôncavo Basin, Brazil*. In: *Linking Diagenesis to Sequence Stratigraphy*. John Wiley & Sons, Inc., pp. 105–132. ISBN: 9781118485347. DOI: `10.1002/9781118485347.ch5`. URL: `http://dx.doi.org/10.1002/9781118485347.ch5`.

Decker, L. (1986). *World Geodetic System 1984*, p. 27.

Deutsch, C. V. & Journel, A. (1998). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press. ISBN: 9780195100150. URL: `https://books.google.de/books?id=CNd6QgAACAAJ`.

DiBiase, D. (1990). *Visualization in the earth sciences*. In: *Earth and mineral sciences* 59.2, pp. 13–18.

Dürrast, H. (2004). *Ultrasonic Laboratory Measurements on Spherical Samples — a Tool for the Investigation of Seismic Anisotropy in Rocks*. In: *Acoustical Imaging*. Ed. by W. Arnold & S. Hirsekorn. Dordrecht: Springer Netherlands, pp. 383–390. ISBN: 978-1-4020-2402-3. DOI: `10.1007/978-1-4020-2402-3_48`. URL: `https://doi.org/10.1007/978-1-4020-2402-3_48`.

Dykes, J., MacEachren, A., & Kraak, M. J. (2005). *Exploring Geovisualization*. Elsevier Ltd., p. 730.

Einsele, G. (2000). *Sedimentary Basins: Evolution, Facies and Sediment Budget*. Berlin, Heidelberg, New York: Springer-Verlag, p. 792. DOI: `10.1007/978-3-662-04029-4`.

Ellis, G. & Mansmann, F. (2010). *Mastering the information age: Solving problems with visual analytics*. In: *Proceedings of the Eurographics*. Vol. 2, p. 5.

Enge, H. D. et al. (2007). *From outcrop to reservoir simulation model: Workflow and procedures*. In: *Geosphere* 3.6, pp. 469–490. DOI: `10.1130/ges00099.1`.

Fanchi, J. R. (2002). *Chapter 10 - Fundamentals of Reservoir Characterization*. In: *Shared Earth Modeling*. Ed. by J. R. Fanchi. Woburn: Butterworth-Heinemann, pp. 170–181. ISBN: 978-0-7506-7522-2. DOI: `https://doi.org/10.1016/B978-075067522-2/50010-0`. URL: `http://www.sciencedirect.com/science/article/pii/B9780750675222500100`.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). *The KDD process for extracting useful knowledge from volumes of data*. In: *Commun. ACM* 39.11, 27–34. ISSN: 0001-0782. DOI: `10.1145/240455.240464`. URL: `https://doi.org/10.1145/240455.240464`.

Fayyad, U., Grinstein, G., & Wierse, A. (2002). *Information Visualization in Data Mining and Knowledge Discovery*, p. 391.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. In: *KDD-96 Proceedings*, p. 7.

Fischer, C. et al. (2007). *A 3D High Resolution Model Of Bounding Surfaces In Aeolian-fluvial Deposits: An Outcrop Analogue Study From The Permian Rotliegend, Northern Germany*. In: *Journal of Petroleum Geology* 30.3, pp. 257–273. ISSN: 1747-5457. DOI: `10.1111/j.1747-5457.2007.00257.x`. URL: `http://dx.doi.org/10.1111/j.1747-5457.2007.00257.x`.

Fischer, M. M. & Getis, A. (2010). *Handbook of Applied Spatial Analysis*. DOI: `https://doi.org/10.1007/978-3-642-03647-7`.

Fitch, P. J. R. et al. (2015). *An integrated and quantitative approach to petrophysical heterogeneity*. In: *Marine and Petroleum Geology* 63, pp. 82–96. ISSN: 0264-8172. DOI: `https://doi.org/`

10.1016/j.marpetgeo.2015.02.014. URL: http://www.sciencedirect.com/science/article/pii/S0264817215000434.

Garg, S. K. & Combs, J. (2015). *A reformulation of USGS volumetric "heat in place" resource estimation method*. In: *Geothermics* 55, pp. 150–158. ISSN: 0375-6505. DOI: https://doi.org/10.1016/j.geothermics.2015.02.004. URL: http://www.sciencedirect.com/science/article/pii/S0375650515000280.

Ge, J. K. & Chen, Z. (2010). *Constructing Ontology-Based Petroleum Exploration Database for Knowledge Discovery*. In: *Applied Mechanics and Materials* 20-23, pp. 975–980. ISSN: 1662-7482. DOI: 10.4028/www.scientific.net/AMM.20-23.975. URL: https://www.scientific.net/AMM.20-23.975.

Gerya, T. (2009). *Introduction to Numerical Geodynamic Modelling*. Cambridge: Cambridge University Press. ISBN: 9780521887540. DOI: DOI:10.1017/CBO9780511809101. URL: https://www.cambridge.org/core/books/introduction-to-numerical-geodynamic-modelling/F2CDB0729FE6DE586BCD9B18C059AE07.

Glaßer, S. et al. (2013). *Visual analysis of longitudinal brain tumor perfusion*. Vol. 8670. SPIE Medical Imaging. SPIE. URL: https://doi.org/10.1117/12.2007878.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.

Greb, M. D. & Linsel, A. (2020). *A case for meaningful geoscientific data collection, curation and integration in petroleum exploration and development endeavors*. In: *AAPG ACE 2020*.

Gu, Y. et al. (2017). *Using seismic data to estimate the spatial distribution of rock thermal conductivity at reservoir scale*. In: *Geothermics* 66, pp. 61–72.

Gudmundsdottir, H. & Horne, R. (2018). *Reservoir Characterization and Prediction Modeling Using Statistical Techniques*. In: *43rd Workshop on Geothermal Reservoir Engineering* (Stanford, California). Stanford University.

Hammer, O., Harper, D., & Ryan, P. (2001). *PAST: Paleontological Statistics Software Package for Education and Data Analysis*. In: *Palaeontologia Electronica* 4, pp. 1–9.

Heap, M. J. et al. (2017). *Microstructural and petrophysical properties of the Permo-Triassic sandstones (Buntsandstein) from the Soultz-sous-Forêts geothermal site (France)*. In: *Geothermal Energy* 5.1, p. 26. ISSN: 2195-9706. DOI: 10.1186/s40517-017-0085-9. URL: https://doi.org/10.1186/s40517-017-0085-9.

Heidsiek, M. et al. (2020). *Small-scale diagenetic facies heterogeneity controls porosity and permeability pattern in reservoir sandstones*. In: *Environmental Earth Sciences* 79.18, p. 425. ISSN: 1866-6299. DOI: 10.1007/s12665-020-09168-z. URL: https://doi.org/10.1007/s12665-020-09168-z.

Heumann, C., Schomaker, M., & Shalabh (2016). *Measures of Central Tendency and Dispersion*. In: *Introduction to Statistics and Data Analysis : With Exercises, Solutions and Applications in R*. Cham:

Springer International Publishing, pp. 37–66. ISBN: 978-3-319-46162-5. DOI: `10.1007/978-3-319-46162-5_3`. URL: `https://doi.org/10.1007/978-3-319-46162-5_3`.

Holden, L. et al. (1998). *Modeling of Fluvial Reservoirs with Object Models*. In: *Mathematical Geology* 30.5, pp. 473–496. ISSN: 1573-8868. DOI: `10.1023/A:1021769526425`. URL: `https://doi.org/10.1023/A:1021769526425`.

Hornung, J. & Aigner, T. (2002). *Reservoir Architecture in a Terminal Alluvial Plain: An Outcrop Analogue Study (Upper Triassic, Southern Germany) Part 1: Sedimentology And Petrophysics*. In: *Journal of Petroleum Geology* 25.1, pp. 3–30. ISSN: 1747-5457. DOI: `10.1111/j.1747-5457.2002.tb00097.x`. URL: `http://dx.doi.org/10.1111/j.1747-5457.2002.tb00097.x`.

Hornung, J. et al. (2020). *Understanding small-scale petrophysical heterogeneities in sedimentary rocks: the key to understanding pore geometry variations and to predict lithofacies-dependent reservoir properties*. In: *Digital Geology – Multi-scale analysis of depositional systems and their subsurface workflows*. Ed. by J. Grötsch & M. Pöppelreiter. ISBN: 9789462823372. URL: `https://bookshop.eage.org/product/digital-geology-multi-scale-analysis-of-depositional-systems-and-their-subsurface-workflows/`.

Hornung, J. & Hinderer, M. (2011). *Depositional Dynamics and Preservation Potential in a Progradational Lacustrine Fluvio-Deltaic Setting*. In: *SEPM Special Publications* 97, pp. 281–310. DOI: `10.2110/sepmsp.097.281`.

Howell, J. A., Martinius, A. W., & Good, T. R. (2014). *The application of outcrop analogues in geological modelling: a review, present status and future outlook*. In: *Geological Society, London, Special Publications* 387.1, pp. 1–25. DOI: `10.1144/sp387.12`. URL: `http://sp.lyellcollection.org/content/387/1/1.abstract`.

Hudson, G. & Wackernagel, H. (1994). *Mapping temperature using kriging with external drift: Theory and an example from scotland*. In: *International Journal of Climatology* 14.1, pp. 77–91. ISSN: 0899-8418. DOI: `10.1002/joc.3370140107`. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3370140107`.

Ito, T., Nakajima, T., & Xue, Z. (2017). *Geological Reservoir Characterization and Modelling of a CO2 Storage Aquifer: A Case Study of the Nagaoka Site, Japan*. In: *Energy Procedia* 114, pp. 2792–2798. ISSN: 1876-6102. DOI: `https://doi.org/10.1016/j.egypro.2017.03.1396`. URL: `http://www.sciencedirect.com/science/article/pii/S1876610217315813`.

Jiang, B. (2014). *Geospatial analysis requires a different way of thinking: the problem of spatial heterogeneity*. In: *GeoJournal* 80.1, 1–13. ISSN: 1572-9893. DOI: `10.1007/s10708-014-9537-y`. URL: `http://dx.doi.org/10.1007/s10708-014-9537-y`.

Jifa, G. & Lingling, Z. (2014). *Data, DIKW, Big Data and Data Science*. In: *Procedia Computer Science* 31, pp. 814–821. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.`

2014.05.332. URL: http://www.sciencedirect.com/science/article/pii/S1877050914005092.

Journel, A. & Alabert, F. (1990). *New method for reservoir mapping*. In: *J. Petrol. Technol.* 42, 212–218.

Kanevski, M. et al. (2004). *Environmental data mining and modeling based on machine learning algorithms and geostatistics*. In: *Environmental Modelling & Software* 19.9, pp. 845–855. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2003.03.004. URL: http://www.sciencedirect.com/science/article/pii/S1364815203002032.

Keim, D. A. (2005). *Chapter 2 - Information Visualization: Scope, Techniques and Opportunities for Geovisualization*. In: *Exploring Geovisualization*. Ed. by J. Dykes, A. M. MacEachren, & M.-J. Kraak. Oxford: Elsevier, pp. 21–52. ISBN: 978-0-08-044531-1. DOI: https://doi.org/10.1016/B978-008044531-1/50420-6. URL: http://www.sciencedirect.com/science/article/pii/B9780080445311504206.

Keim, D. A. et al. (2008). *Visual Analytics: Scope and Challenges*. In: *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Ed. by S. J. Simoff, M. H. Böhlen, & A. Mazeika. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 76–90. ISBN: 978-3-540-71080-6. DOI: 10.1007/978-3-540-71080-6_6. URL: https://doi.org/10.1007/978-3-540-71080-6_6.

Kiryukhin, A. V., Kaymin, E. P., & Zakharova, E. V. (2008). *Using TOUGHREACT to Model Laboratory Tests on the Interaction of NaNO3-NaOH Fluids with Sandstone Rock at a Deep Radionuclide Repository Site*. In: *Nuclear Technology* 164.2, pp. 196–206. ISSN: 0029-5450. DOI: 10.13182/NT08-A4019. URL: https://doi.org/10.13182/NT08-A4019.

Kloosterman, H. J. et al. (2017). *Digital geology and the upstream business–Using virtual and augmented reality technologies for subsurface capability building and collaboration*. In: *First Break* 35, 1–4.

Kohlhammer, J. et al. (2011). *Solving Problems with Visual Analytics*. In: *Procedia Computer Science* 7, pp. 117–120. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2011.12.035. URL: http://www.sciencedirect.com/science/article/pii/S1877050911007009.

Kohonen, T. (2001). *Variants of SOM*. In: *Self-Organizing Maps*. Vol. 30. Berlin, Heidelberg: Springer.

Kraak, M. (2008). *Editorial From Geovisualisation Toward Geovisual Analytics*. In: *Cartographic journal* 45.3, pp. 163 –164. ISSN: 1743-2774. DOI: 10.1179/174327708x315183. URL: https://research.utwente.nl/en/publications/from-geovisualization-toward-geovisual-analytics--editorial(a822d25a-cacc-4984-a1ea-2a9bb2dc86ba).html.

Kratz, A. (2013). *Three-Dimensional Second-Order Tensor Fields: Exploratory Visualization and Anisotropic Sampling*. Doctoral thesis.

Kushnir, A. R. L. et al. (2018). *Characterizing the physical properties of rocks from the Paleozoic to Permo-Triassic transition in the Upper Rhine Graben*. In: *Geothermal Energy* 6.1, p. 16. ISSN: 2195-9706. DOI: 10.1186/s40517-018-0103-6. URL: https://doi.org/10.1186/s40517-018-0103-6.

Lajaunie, C., Courrioux, G., & Manuel, L. (1997). *Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation*. In: *Mathematical Geology* 29.4, pp. 571–584. ISSN: 1573-8868. DOI: 10.1007/BF02775087. URL: https://doi.org/10.1007/BF02775087.

Landa, J. L. & Strebelle, S. (2002). *Sensitivity Analysis of Petrophysical Properties Spatial Distributions, and Flow Performance Forecasts to Geostatistical Parameters Using Derivative Coefficients*. Conference Paper. DOI: 10.2118/77430-MS. URL: https://doi.org/10.2118/77430-MS.

Laudon, K. C. & Laudon, J. (2012). *Management Information Systems – Managing The Digital Firm*. Vol. 12. Pearson Prentice Hall, p. 588.

Lenz, O. K., Wilde, V., & Riegel, W. (2016). *ENSO- and solar-driven sub-Milankovitch cyclicity in the Palaeogene greenhouse world; high-resolution pollen records from Eocene Lake Messel, Germany*. In: *Journal of the Geological Society* 174.1, pp. 110–128. ISSN: 0016-7649. DOI: 10.1144/jgs2016-046. URL: https://doi.org/10.1144/jgs2016-046.

Leuangthong, O., McLennan, J. A., & Deutsch, C. V. (2004). *Minimum Acceptance Criteria for Geostatistical Realizations*. In: *Natural Resources Research* 13.3, pp. 131–141. ISSN: 1573-8981. DOI: 10.1023/B:NARR.0000046916.91703.bb. URL: https://doi.org/10.1023/B:NARR.0000046916.91703.bb.

Lewin, A. et al. (2018). *Provenance of sandstones in Ethiopia during Late Ordovician and Carboniferous–Permian Gondwana glaciations: Petrography and geochemistry of the Enticho Sandstone and the Edaga Arbi Glacials*. In: *Sedimentary Geology* 375, pp. 188–202. ISSN: 0037-0738. DOI: https://doi.org/10.1016/j.sedgeo.2017.10.006. URL: http://www.sciencedirect.com/science/article/pii/S0037073817302257.

Li, D., Wang, S., & Li, D. (2015). *Spatial Data Mining*. Berlin, Heidelberg: Springer-Verlag. DOI: 10.1007/978-3-662-48538-5.

Li, H. & Reynolds, J. (1995). *On Definition and Quantification of Heterogeneity*. In: *Oikos* 73, pp. 280–284. DOI: 10.2307/3545921. URL: www.jstor.org/stable/3545921.

Linsel, A. et al. (2020a). *High-Resolution Analysis of the Physicochemical Characteristics of Sandstone Media at the Lithofacies Scale*. In: *Solid Earth Discuss.* 2020, pp. 1–28. ISSN: 1869-9537. DOI: 10.5194/se-2020-13. URL: https://www.solid-earth-discuss.net/se-2020-13/.

Linsel, A. et al. (2020b). *Accounting for Local Geological Variability in Sequential Simulations—Concept and Application*. In: *ISPRS International Journal of Geo-Information* 9.6, p. 409. ISSN: 2220-9964. URL: `https://www.mdpi.com/2220-9964/9/6/409`.

Linsel, A. et al. (2020c). *GeoReVi: A knowledge discovery and data management tool for subsurface characterization*. In: *SoftwareX* 12, p. 100597. ISSN: 2352-7110. DOI: `10.1016/j.softx.2020.100597`. URL: `https://doi.org/10.1016/j.softx.2020.100597`.

Liseikin, V. D. (2010). *General Considerations*. In: *Grid Generation Methods*. Ed. by V. D. Liseikin. Dordrecht: Springer Netherlands, pp. 1–29. ISBN: 978-90-481-2912-6. DOI: `10.1007/978-90-481-2912-6_1`. URL: `https://doi.org/10.1007/978-90-481-2912-6_1`.

Lloyd, S. (1982). *Least squares quantization in PCM*. In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137. ISSN: 0018-9448. DOI: `10.1109/TIT.1982.1056489`. URL: `https://ieeexplore.ieee.org/document/1056489`.

Magnus Bergman, S. (1986). *The development and utilization of subsurface space*. In: *Tunnelling and Underground Space Technology* 1.2, pp. 115–144. ISSN: 0886-7798. DOI: `https://doi.org/10.1016/0886-7798(86)90051-9`. URL: `http://www.sciencedirect.com/science/article/pii/0886779886900519`.

Maier, D. (1983). *Theory of Relational Databases*. Vol. 1. Computer Science Press.

Maimon, O. & Rokach, L. (2010). *Introduction to Knowledge Discovery and Data Mining*. In: *Data Mining and Knowledge Discovery Handbook*. Springer Science +Business Media. Chap. 1, 1–15.

Maćkiewicz, A. & Ratajczak, W. (1993). *Principal components analysis (PCA)*. In: *Computers & Geosciences* 19.3, pp. 303–342. ISSN: 0098-3004. DOI: `https://doi.org/10.1016/0098-3004(93)90090-R`. URL: `http://www.sciencedirect.com/science/article/pii/009830049390090R`.

Malvić, T. et al. (2019). *Kriging with a Small Number of Data Points Supported by Jack-Knifing, a Case Study in the Sava Depression (Northern Croatia)*. In: *Geosciences* 9.1, p. 36. ISSN: 2076-3263. URL: `https://www.mdpi.com/2076-3263/9/1/36`.

Martin, R. C. (2008). *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall PTR, p. 448. ISBN: 0132350882.

– (2017). *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Prentice Hall Press, p. 404. ISBN: 0134494164.

Matheron, G. (1963). *Principles of geostatistics*. In: *Economic Geology* 58.8, 1246–1266. DOI: `10.2113/gsecongeo.58.8.1246`. URL: `https://doi.org/10.2113/gsecongeo.58.8.1246`.

Mazza, R. (2009). *Introduction to Information Visualization*. Springer-Verlag London, p. 149. DOI: `https://doi.org/10.1007/978-1-84800-219-7`.

Mazzella, A. & Mazzella, A. (2013). *The Importance of the Model Choice for Experimental Semivariogram Modeling and Its Consequence in Evaluation Process*. In: *Journal of Engineering* 2013, p. 10. DOI: 10.1155/2013/960105. URL: http://dx.doi.org/10.1155/2013/960105.

Medici, G., West, L. J., & Mountney, N. P. (2016). *Characterizing flow pathways in a sandstone aquifer: Tectonic vs sedimentary heterogeneities*. In: *Journal of Contaminant Hydrology* 194, pp. 36–58. ISSN: 0169-7722. DOI: https://doi.org/10.1016/j.jconhyd.2016.09.008. URL: http://www.sciencedirect.com/science/article/pii/S0169772216302133.

Medici, G., West, L. J., & Mountney, N. P. (2019). *Sedimentary flow heterogeneities in the Triassic U.K. Sherwood Sandstone Group: Insights for hydrocarbon exploration*. In: *Geological Journal* 54.3, pp. 1361–1378. ISSN: 0072-1050. DOI: 10.1002/gj.3233. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/gj.3233.

Mert, B. A. & Dag, A. (2017). *A Computer Program for Practical Semivariogram Modeling and Ordinary Kriging: A Case Study of Porosity Distribution in an Oil Field*. In: *Open Geosciences* 9.1, pp. 663–674. DOI: 10.1515/geo-2017-0050. URL: https://www.degruyter.com/view/j/geo.2017.9.issue-1/geo-2017-0050/geo-2017-0050.xml.

Miall, A. D. (1996). *The Geology of Fluvial Deposits – Sedimentary Facies, Basin Analysis, and Petroleum Geology*. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-08211-5. DOI: 10.1007/978-3-662-03237-4. URL: https://link.springer.com/book/10.1007%2F978-3-662-03237-4.

– (2000). *Facies Analysis*. In: *Principles of Sedimentary Basin Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 141–248. ISBN: 978-3-662-03999-1. DOI: 10.1007/978-3-662-03999-1_4. URL: https://doi.org/10.1007/978-3-662-03999-1_4.

Miall, A. D. (1985). *Architectural-element analysis: A new method of facies analysis applied to fluvial deposits*. In: *Earth-Science Reviews* 22.4, pp. 261–308. ISSN: 0012-8252. DOI: http://dx.doi.org/10.1016/0012-8252(85)90001-7. URL: http://www.sciencedirect.com/science/article/pii/0012825285900017.

Michie, E. A. H. & Haines, T. J. (2016). *Variability and heterogeneity of the petrophysical properties of extensional carbonate fault rocks, Malta*. In: *Petroleum Geoscience* 22.2, pp. 136–152. ISSN: 1354-0793. DOI: 10.1144/petgeo2015-027. URL: http://dx.doi.org/10.1144/petgeo2015-027.

Mines, G. (2016). *Binary geothermal energy conversion systems: Basic Rankine, dual–pressure, and dual–fluid cycles*. In: *Geothermal Power Generation*. Ed. by R. DiPippo. Woodhead Publishing, pp. 353–389. ISBN: 978-0-08-100337-4. DOI: https://doi.org/10.1016/B978-0-08-100337-4.00013-9. URL: http://www.sciencedirect.com/science/article/pii/B9780081003374000139.

Mitas, L. & Mitasova, H. (2005). *Spatial Interpolation*. In: *Geographic Information Systems: Principles, Techniques, Management and Applications*. Ed. by P. Longley et al. 2nd. URL: `www.geos.ed.ac.uk/~gisteac/gis_book_abridged/`.

Muffler, P. & Cataldi, R. (1978). *Methods for regional assessment of geothermal resources*. In: *Geothermics* 7.2, pp. 53–89. ISSN: 0375-6505. DOI: `https://doi.org/10.1016/0375-6505(78)90002-0`. URL: `http://www.sciencedirect.com/science/article/pii/0375650578900020`.

Mukerji, T., Mavko, G., & Rio, P. (1997). *Scales of Reservoir Heterogeneities and Impact of Seismic Resolution on Geostatistical Integration*. In: *Mathematical Geology* 29.7, pp. 933–950. ISSN: 1573-8868. DOI: `10.1023/a:1022307807851`. URL: `https://doi.org/10.1023/A:1022307807851http://dx.doi.org/10.1023/A:1022307807851`.

Munzner, T. et al. (2006). *NIHNSF Visualization Research Challenges Report Summary*. In: *IEEE Comput. Graph. Appl.* 26.2, 20–24.

Nichols, G. (2009). *Sedimentology and Stratigraphy*. 2nd. Wiley-Blackwell.

Nordahl, K. & Ringrose, P. S. (2008). *Identifying the Representative Elementary Volume for Permeability in Heterolithic Deposits Using Numerical Rock Models*. In: *Mathematical Geosciences* 40.7, p. 753. ISSN: 1874-8953. DOI: `10.1007/s11004-008-9182-4`. URL: `https://doi.org/10.1007/s11004-008-9182-4`.

Nordahl, K. et al. (2014). *Impact of multiscale modelling on predicted porosity and permeability distributions in the fluvial deposits of the Upper Lunde Member (Snorre Field, Norwegian Continental Shelf)*. In: *Sediment-Body Geometry and Heterogeneity: Analogue Studies for Modelling the Subsurface*. Vol. 387. Geological Society of London, p. 25. ISBN: 9781862393721. DOI: `10.1144/sp387.10`. URL: `https://doi.org/10.1144/SP387.10`.

Osborne, P. (2013). *The Mercator Projections*. Report.

Oxford Dictionary, E. (2014). *heterogeneity, n.* Oxford University Press. URL: `http://www.oed.com/view/Entry/86452?redirectedFrom=Heterogeneity`.

Popov, Y. A. et al. (1999). *Characterization of rock thermal conductivity by high-resolution optical scanning*. In: *Geothermics* 28.2, pp. 253–276. ISSN: 0375-6505. DOI: `https://doi.org/10.1016/S0375-6505(99)00007-3`. URL: `http://www.sciencedirect.com/science/article/pii/S0375650599000073`.

Popov, Y. & Mandel, A. M. (1998). *Geothermal study of anisotropic rock masses*. Vol. 34, pp. 903–915.

Popov, Y. et al. (2016). *ISRM Suggested Methods for Determining Thermal Properties of Rocks from Laboratory Tests at Atmospheric Pressure*. Vol. 49. DOI: `10.1007/s00603-016-1070-5`.

Reading, H. (1996). *Sedimentary Environments: Processes, Facies and Stratigraphy*. 3rd Edition. Wiley-Blackwell, p. 704.

Remy, N., Boucher, A., & Wu, J. (2009). *Applied Geostatistics with SGeMS: A User's Guide*. Cambridge: Cambridge University Press. ISBN: 9781107403246. DOI: `DOI:10.1017/CB09781139150019`. URL: `https://www.cambridge.org/core/books/applied-geostatistics-with-sgems/35D45D41BF8FD9BB43E8DCD462E30975`.

Remy, N. (2004). *Algorithmic and Software Methods for a Better Integration of the Geological Information into Numerical Models*. Thesis.

Rühaak, W. et al. (2015a). *Prognosefähigkeit numerischer Erdwärmesondenmodelle*. In: *Grundwasser* 20.4, pp. 243–251. DOI: `10.1007/s00767-015-0305-9`.

– (2015b). *Prognosefähigkeit numerischer Erdwärmesondenmodelle*. In: *Grundwasser* 20.4, pp. 243–251. ISSN: 1432-1165. DOI: `10.1007/s00767-015-0305-9`. URL: `https://doi.org/10.1007/s00767-015-0305-9`.

Rühaak, W. et al. (2015c). *Upscaling thermal conductivities of sedimentary formations for geothermal exploration*. In: *Geothermics* 58, pp. 49–61. ISSN: 0375-6505. DOI: `http://dx.doi.org/10.1016/j.geothermics.2015.08.004`. URL: `http://www.sciencedirect.com/science/article/pii/S0375650515001054`.

Ringrose, P. & Bentley, M. (2015). *Reservoir Model Design*. 1st ed. Springer Netherlands, p. 249. ISBN: 978-94-007-5496-6. DOI: `10.1007/978-94-007-5497-3`. URL: `http://www.springer.com/de/book/9789400754966`.

Ringrose, P. S. et al. (1993). *Immiscible flow behaviour in laminated and cross-bedded sandstones*. In: *Journal of Petroleum Science and Engineering* 9.2, pp. 103–124. ISSN: 0920-4105. DOI: `https://doi.org/10.1016/0920-4105(93)90071-L`. URL: `http://www.sciencedirect.com/science/article/pii/092041059390071L`.

Robertson, R. K., Mueller, U. A., & Bloom, L. M. (2006). *Direct sequential simulation with histogram reproduction: A comparison of algorithms*. In: *Comput. Geosci.* 32.3, pp. 382–395. ISSN: 0098-3004. DOI: `10.1016/j.cageo.2005.07.002`.

Rodrigo-Ilarri, J., Reisinger, M., & Gómez-Hernández, J. J. (2017). *Influence of Heterogeneity on Heat Transport Simulations in Shallow Geothermal Systems*. In: *Geostatistics Valencia 2016*. Ed. by J. J. Gómez-Hernández et al. Cham: Springer International Publishing, pp. 849–862. ISBN: 978-3-319-46819-8. DOI: `10.1007/978-3-319-46819-8_59`. URL: `https://doi.org/10.1007/978-3-319-46819-8_59`.

Rowley, J. (2007). *The wisdom hierarchy: representations of the DIKW hierarchy*. In: *Journal of Information Science* 33.2, pp. 163–180. DOI: `10.1177/0165551506070706`. URL: `https://journals.sagepub.com/doi/abs/10.1177/0165551506070706`.

Rühaak, W. (2006). *A Java application for quality weighted 3-d interpolation*. In: *Comput. Geosci.* 32.1, pp. 43–51. ISSN: 0098-3004. DOI: `10.1016/j.cageo.2005.04.005`. URL: `https://doi.org/10.1016/j.cageo.2005.04.005`.

– (2015). *3-D interpolation of subsurface temperature data with measurement error using kriging*. In: *Environmental Earth Sciences* 73.4, pp. 1893–1900. ISSN: 1866-6299. DOI: `10.1007/s12665-014-3554-5`. URL: `https://doi.org/10.1007/s12665-014-3554-5`.

Sacha, D. et al. (2014). *Knowledge Generation Model for Visual Analytics*. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 1604–1613. ISSN: 1077-2626. DOI: `10.1109/TVCG.2014.2346481`.

Sammon, J. W. (1969). *A Nonlinear Mapping for Data Structure Analysis*. In: *IEEE Transactions on Computers* C-18.5, pp. 401–409. DOI: `10.1109/T-C.1969.222678`.

Schiewe, J. (2013). *Geovisualization and Geovisual Analytics: The Interdisciplinary Perspective on Cartography*. In: *KN - Journal of Cartography and Geographic Information* 63.3, pp. 122–126. ISSN: 2524-4965. DOI: `10.1007/BF03546122`. URL: `https://doi.org/10.1007/BF03546122`.

Schneider, S., Hornung, J., & Hinderer, M. (2016). *Evolution of the western East African Rift System reflected in provenance changes of Miocene to Pleistocene synrift sediments (Albertine Rift, Uganda)*. In: *Sedimentary Geology* 343, pp. 190–205. ISSN: 0037-0738. DOI: `https://doi.org/10.1016/j.sedgeo.2016.07.013`. URL: `http://www.sciencedirect.com/science/article/pii/S0037073816302007`.

Schulte, D. O. et al. (2020). *Multi-objective optimization under uncertainty of geothermal reservoirs using experimental design-based proxy models*. In: *Geothermics* 86, p. 101792. ISSN: 0375-6505. DOI: `10.1016/j.geothermics.2019.101792`. URL: `https://doi.org/10.1016/j.geothermics.2019.101792`.

Shakhnarovich, G., Darrell, T., & Indyk, P. (2006). *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)*. The MIT Press. ISBN: 026219547X.

Shepard, D. (1968). *A Two-Dimensional Interpolation Function for Irregularly-Spaced Data*. In: *Proceedings of the 1968 ACM National Conference*, pp. 517–524. URL: `http://dx.doi.org/10.1145/800186.810616`.

Shyr, J., Chu, J., & Woods, M. (2018). *Cognitive Data Analysis for Big Data*. In: *Handbook of Big Data Analytics*. Ed. by W. K. Härdle, H. H.-S. Lu, & X. Shen. Cham: Springer International Publishing, pp. 23–47. ISBN: 978-3-319-18284-1. DOI: `10.1007/978-3-319-18284-1_2`. URL: `https://doi.org/10.1007/978-3-319-18284-1_2`.

Si, H. (2015). *TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator*. In: *ACM Trans. Math. Softw.* 41.2, Article 11. ISSN: 0098-3500. DOI: `10.1145/2629697`. URL: `https://doi.org/10.1145/2629697`.

Simoff, S. J., Böhlen, M. H., & Mazeika, A. (2008). *Visual Data Mining: An Introduction and Overview*. In: *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Ed. by S. J. Simoff, M. H. Böhlen, & A. Mazeika. Berlin, Heidelberg: Springer Berlin Heidelberg,

pp. 1–12. ISBN: 978-3-540-71080-6. DOI: 10.1007/978-3-540-71080-6_1. URL: https://doi.org/10.1007/978-3-540-71080-6_1.

Simpson, G. (2017). *Practical Finite Element Modeling in Earth Science using Matlab*. Wiley-Blackwell, p. 272. ISBN: 978-1-119-24862-0.

Soares, A. (2001). *Direct Sequential Simulation and Cosimulation*. In: *Mathematical Geology* 33.8, pp. 911–926. ISSN: 1573-8868. DOI: 10.1023/A:1012246006212. URL: https://doi.org/10.1023/A:1012246006212.

Tarbuck, E. J. & Lutgens, F. K. (2008). *Earth: An Introduction to Physical Geology*. 9th ed. Pearson Education International, p. 702. ISBN: 0-13-241066-4.

Tellam, J. H. & Barker, R. D. (2006). *Towards prediction of saturated-zone pollutant movement in groundwaters in fractured permeable-matrix aquifers: the case of the UK Permo-Triassic sandstones*. In: *Geological Society, London, Special Publications* 263.1, pp. 1–48. DOI: 10.1144/gsl.Sp.2006.263.01.01. URL: https://sp.lyellcollection.org/content/specpubgsl/263/1/1.full.pdf.

The European Parliament and the Council of The European Union (2009). *Directive 2009/28/EC of The European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC*. In: *Official Journal of the European Union*, p. 47. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009L0028.

Thierauf, R. J. (1999). *Knowlege management systems for business*. Quorum Books, p. 360. ISBN: 1-56720-218-7.

Thomas, J. J. & Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, p. 190.

Thomsen, L. (1986). *Weak elastic anisotropy*. In: *GEOPHYSICS* 51.10, pp. 1954–1966. DOI: 10.1190/1.1442051. URL: https://library.seg.org/doi/abs/10.1190/1.1442051.

Tiab, D. & Donaldson, E. C. (2012). *Chapter 10 - Reservoir Characterization*. In: *Petrophysics (Third Edition)*. Ed. by D. Tiab & E. C. Donaldson. Boston: Gulf Professional Publishing, pp. 667–714. ISBN: 978-0-12-383848-3. DOI: https://doi.org/10.1016/B978-0-12-383848-3.00010-4. URL: http://www.sciencedirect.com/science/article/pii/B9780123838483000104.

Tillmann, G. (2017). *Usage-Driven Database Design. From Logical Data Modeling through Physical Schema Definition*. 1st ed. Apress, p. 374. DOI: 10.1007/978-1-4842-2722-0.

Tillmann, G. (1993). *A practical guide to logical data modeling*. McGraw-Hill, Inc., p. 248.

Trauth, M. H. (2015). *Univariate Statistics*. In: *MATLAB® Recipes for Earth Sciences*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 57–120. ISBN: 978-3-662-46244-7. DOI: `10.1007/978-3-662-46244-7_3`. URL: `https://doi.org/10.1007/978-3-662-46244-7_3`.

Tukey, J. (1977). *Exploratory Data Analysis*. Pearson, p. 712.

Wackernagel, H. (2003). *Multivariate Geostatistics*. Third Edition. Springer-Verlag Berlin Heidelberg GmbH, p. 388. DOI: `10.1007/978-3-662-05294-5`. URL: `https://doi.org/10.1007/978-3-662-05294-5`.

Wang, J. & Zuo, R. (2018). *Identification of geochemical anomalies through combined sequential Gaussian simulation and grid-based local singularity analysis*. In: *Computers & Geosciences* 118, pp. 52–64. ISSN: 0098-3004. DOI: `https://doi.org/10.1016/j.cageo.2018.05.010`. URL: `http://www.sciencedirect.com/science/article/pii/S0098300417313134`.

Webster, R. & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. 2nd ed. Wiley & Sons, Inc, p. 330. URL: `https://www.wiley.com/en-us/Geostatistics+for+Environmental+Scientists%2C+2nd+Edition-p-9780470028582`.

Wellmann, F. et al. (2019). *Chapter 15 - From Google Earth to 3D Geology Problem 2: Seeing Below the Surface of the Digital Earth*. In: *Developments in Structural Geology and Tectonics*. Ed. by A. Billi & k. Fagereng. Vol. 5. Elsevier, pp. 189–204. ISBN: 2542-9000. DOI: `https://doi.org/10.1016/B978-0-12-814048-2.00015-6`. URL: `http://www.sciencedirect.com/science/article/pii/B9780128140482000156`.

Welsch, B. (2019). *Technical, Environmental and Economic Assessment of Medium Deep Borehole Thermal Energy Storage Systems*. Thesis, p. 226.

Welsch, B. et al. (2016). *Characteristics of medium deep borehole thermal energy storage*. In: *International Journal of Energy Research* 40.13, pp. 1855–1868. ISSN: 0363-907X. DOI: `10.1002/er.3570`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/er.3570`.

Williams, C. (2014). *Evaluating the volume method in the assessment of identified geothermal resources*. In: *Transactions - Geothermal Resources Council* 38, pp. 967–974.

Willmott, C. J., Matsuura, K., & Robeson, S. M. (2009). *Ambiguities inherent in sums-of-squares-based error statistics*. In: *Atmospheric Environment* 43.3, pp. 749–752. ISSN: 1352-2310. DOI: `https://doi.org/10.1016/j.atmosenv.2008.10.005`. URL: `http://www.sciencedirect.com/science/article/pii/S1352231008009564`.

Zehner, B. et al. (2015). *Workflows for generating tetrahedral meshes for finite element simulations on complex geological structures*. In: *Computers & Geosciences* 79, pp. 105–117. ISSN: 0098-3004. DOI: `https://doi.org/10.1016/j.cageo.2015.02.009`. URL: `http://www.sciencedirect.com/science/article/pii/S009830041500031X`.

Zeleny, M. (1987). *Management support systems: Towards integrated knowledge management*. In: *Human Systems Management* 7.1, 59–70. DOI: 10.3233/HSM-1987-7108.

A. Appendix

A.1. Mathematical Expressions

A.1.1. Measures for Exploratory Data Analysis

model	function	
arithmetic mean	$\overline{x} = \dfrac{\sum\limits_{i=0}^{n} x_i}{n}$	(A.1)
geometric mean	$\overline{x}_g = \sqrt[n]{\prod\limits_{i=0}^{n} x_i}$	(A.2)
harmonic mean	$\overline{x}_h = \dfrac{n}{\sum\limits_{i=0}^{n} \frac{1}{x_i}}$	(A.3)
sample variance	$\sigma^2 = \dfrac{1}{n-1} \sum\limits_{i=0}^{n} (x_i - \mu)$	(A.4)
sample standard deviation	$\sigma = \sqrt{\sigma^2}$	(A.5)

Table A.1.: Basic statistical measures for exploratory data analysis

model	function	
skewness	$s = \dfrac{n}{(n-1)(n-2)} \dfrac{\sum\limits_{i=0}^{n}(x_i - \overline{x})^3}{\left(\sqrt{\frac{1}{n-1}\sum\limits_{i=0}^{n}(x_i - \overline{x})^2}\right)^3}$	(A.6)
kurtosis	$k = \dfrac{n(n+1)}{(n-1)(n-2)(n-3)} \dfrac{\sum\limits_{i=0}^{n}(x_i - \overline{x})^4}{\left(\sqrt{\frac{1}{n-1}\sum\limits_{i=0}^{n}(x_i - \overline{x})^2}\right)^4} - 3 \cdot \dfrac{(n-1)^2}{(n-2)(n-3)}$	(A.7)

Table A.2.: Basic statistical measures for exploratory data analysis

A.1.2. Transformation Functions

model	function	
logarithmic	$y_i = \log(x_i)$	(A.8)
exponential	$y_i = 10^{x_i}$	(A.9)
z-score	$y_i = \dfrac{x_i - \overline{x}}{\sigma}$	(A.10)
rescaling	$y_i = \dfrac{x_i - x_{min}}{x_{max} - x_{min}}$	(A.11)
mean-rescaling	$y_i = \dfrac{x_i - \overline{x}}{x_{max} - x_{min}}$	(A.12)
subtract-mean	$y_i = x_i - \overline{x}$	(A.13)

Table A.3.: Basic transformation functions during data preprocessing

A.1.3. Covariance Functions

model	function	
spherical	$c_{sph}(\mathbf{h}) = \begin{cases} b \cdot \left(1 - \frac{3\|\mathbf{h}\|}{2a} + \frac{\|\mathbf{h}\|^3}{2a^3}\right) & \text{for } 0 \leq \|\mathbf{h}\| \leq a \\ 0 & \text{for } \|\mathbf{h}\| \geq a. \end{cases}$	(A.14)
gaussian	$c_{gau}(\mathbf{h}) = \begin{cases} b \cdot exp\left(-\frac{\|\mathbf{h}\|^2}{a^2}\right) & \text{for } 0 \leq \|\mathbf{h}\| \leq a \\ 0 & \text{for } \|\mathbf{h}\| \geq a. \end{cases}$	(A.15)
exponential	$c_{exp}(\mathbf{h}) = b \cdot exp\left(-\frac{\|\mathbf{h}\|}{a}\right) \quad \text{with } a, b > 0$	(A.16)
power	$c_{pow}(\mathbf{h}) = b \cdot h^a \quad \text{with } a, b > 0$	(A.17)
linear	$c_{lin}(\mathbf{h}) = \begin{cases} b & \text{for } \|\mathbf{h}\| > 0 \\ 0 & \text{for } \|\mathbf{h}\| = 0 \end{cases}$	(A.18)

Table A.4.: Covariance functions

A.1.4. Shape Functions and Derivatives of an Eight-Node Hexahedron

The shape functions differ for each type of element. A comprehensive list of shape functions is give in Simpson (2017).

$$N_1 = \tfrac{1}{8}(1-\xi)(1-\eta)(1-\zeta)$$

$$N_2 = \tfrac{1}{8}(1-\xi)(1-\eta)(1+\zeta)$$

$$N_3 = \tfrac{1}{8}(1+\xi)(1-\eta)(1+\zeta)$$

$$N_4 = \tfrac{1}{8}(1+\xi)(1-\eta)(1-\zeta)$$

$$N_5 = \tfrac{1}{8}(1-\xi)(1+\eta)(1-\zeta) \tag{A.19}$$

$$N_6 = \tfrac{1}{8}(1-\xi)(1+\eta)(1+\zeta)$$

$$N_7 = \tfrac{1}{8}(1+\xi)(1+\eta)(1+\zeta)$$

$$N_8 = \tfrac{1}{8}(1+\xi)(1+\eta)(1-\zeta).$$

The first derivatives of those shape functions as they are given in Eq. 2.104 are defined as

$$\frac{\partial N_1}{\partial \xi} = -\frac{1}{8}(1-\eta)(1-\zeta) \quad \frac{\partial N_1}{\partial \eta} = -\frac{1}{8}(1-\xi)(1-\zeta) \quad \frac{\partial N_1}{\partial \zeta} = -\frac{1}{8}(1-\xi)(1-\eta)$$

$$\frac{\partial N_2}{\partial \xi} = -\frac{1}{8}(1-\eta)(1+\zeta) \quad \frac{\partial N_2}{\partial \eta} = -\frac{1}{8}(1-\xi)(1+\zeta) \quad \frac{\partial N_2}{\partial \zeta} = \frac{1}{8}(1-\xi)(1-\eta)$$

$$\frac{\partial N_3}{\partial \xi} = \frac{1}{8}(1-\eta)(1+\zeta) \quad \frac{\partial N_3}{\partial \eta} = -\frac{1}{8}(1+\xi)(1+\zeta) \quad \frac{\partial N_3}{\partial \zeta} = \frac{1}{8}(1+\xi)(1-\eta)$$

$$\frac{\partial N_4}{\partial \xi} = \frac{1}{8}(1-\eta)(1+\zeta) \quad \frac{\partial N_4}{\partial \eta} = -\frac{1}{8}(1+\xi)(1-\zeta) \quad \frac{\partial N_4}{\partial \zeta} = -\frac{1}{8}(1+\xi)(1-\eta)$$

$$\frac{\partial N_5}{\partial \xi} = -\frac{1}{8}(1+\eta)(1-\zeta) \quad \frac{\partial N_5}{\partial \eta} = \frac{1}{8}(1-\xi)(1-\zeta) \quad \frac{\partial N_5}{\partial \zeta} = -\frac{1}{8}(1-\xi)(1+\eta)$$

$$\frac{\partial N_6}{\partial \xi} = -\frac{1}{8}(1+\eta)(1+\zeta) \quad \frac{\partial N_6}{\partial \eta} = \frac{1}{8}(1-\xi)(1+\zeta) \quad \frac{\partial N_6}{\partial \zeta} = \frac{1}{8}(1-\xi)(1+\eta)$$

$$\frac{\partial N_7}{\partial \xi} = \frac{1}{8}(1+\eta)(1+\zeta) \quad \frac{\partial N_7}{\partial \eta} = \frac{1}{8}(1+\xi)(1+\zeta) \quad \frac{\partial N_7}{\partial \zeta} = \frac{1}{8}(1+\xi)(1+\eta)$$

$$\frac{\partial N_8}{\partial \xi} = \frac{1}{8}(1+\eta)(1-\zeta) \quad \frac{\partial N_8}{\partial \eta} = \frac{1}{8}(1+\xi)(1-\zeta) \quad \frac{\partial N_8}{\partial \zeta} = -\frac{1}{8}(1+\xi)(1+\eta).$$

$$(A.20)$$

A.2. Code Fragments

A.2.1. Plain Old CLR Object Class

Listing A.1: POCO model for a rock sample entity in GeoReVi.

```
1  // Namespace
2  namespace GeoReVi
3  {
4      // Using directives
5      using LiteDB;
6      ...
7
8      // POCO class for the rock sample entity
9      [Table("tblRockSample")]
10     public partial class tblRockSample
11     {
12         // Sample ID
13         [Key, BsonId]
14         public int sampIdPk { get; set; }
15
16         // Lithofacies Foreign Key
17         public int? sampLithofaciesIdFk { get; set; }
18
19         // Architectural element Foreign Key
20         public int? sampArchitecturalElementIdFk { get; set; }
21
22         // Depositional environment Foreign Key
23         public int? sampDepositionalEnvironmentIdFk { get; set; }
24
25         // Sample label
26         [StringLength(255)]
27         public string sampLabel { get; set; }
28
29         // Sample type
30         [StringLength(255)]
31         public string sampType { get; set; }
32
33         // Sampling method
34         [StringLength(50)]
35         public string sampSamplingMethod { get; set; }
36
37         // Object of investigation Foreign Key
38         [StringLength(255)]
39         public string sampooiNameIdFk { get; set; }
```

```csharp
        // Lithostratigraphic unit
        [StringLength(255)]
        public string sampLithostratigraphyName { get; set; }

        // Chronostratigraphic unit
        [StringLength(255)]
        public string sampChronStratName { get; set; }

        // Petrographic classification
        [StringLength(255)]
        public string sampPetrographicTerm { get; set; }

        // Local x coordinate in meters
        public double? sampLocalXCoordinates { get; set; }

        // Local y coordinate in meters
        public double? sampLocalYCoordinates { get; set; }

        // Local z coordinate in meters
        public double? sampLocalZCoordinates { get; set; }

        // Name of the sampler
        [StringLength(255)]
        public string sampSampler { get; set; }

        // Sampling date and time
        [Column(TypeName = "datetime2")]
        public DateTime? sampDate { get; set; }

        // Project Foreign Key
        public int? sampprjIdFk { get; set; }

        // Latitude in WGS84 decimal degree
        public double? sampLatitude { get; set; }

        // Longitude in WGS84 decimal degree
        public double? sampLongitude { get; set; }

        // Elevation in meters above sea level
        public double? sampElevation { get; set; }

        ...
    }
}
```

A.2.2. Unit Test Class

Listing A.2: Test class for estimating the Euclidean distance in GeoReVi.

```
1  // Namespace
2  namespace GeoReVi.Tests.Helper.Geography
3  {
4      // Using directives
5      using Microsoft.VisualStudio.TestTools.UnitTesting;
6      ...
7
8      /// <summary>
9      /// Summary for GeographyHelperTest
10     /// </summary>
11     [TestClass]
12     public class GeographyHelperTest
13     {
14         public GeographyHelperTest()
15         {
16
17         }
18
19         private TestContext testContextInstance;
20
21         /// <summary>
22         /// Calls the TestContext
23         /// </summary>
24         public TestContext TestContext
25         {
26             get
27             {
28                 return testContextInstance;
29             }
30             set
31             {
32                 testContextInstance = value;
33             }
34         }
35
36         /// <summary>
37         /// Tests the method which estimates the euclidean distance
38         /// between two points
39         /// </summary>
40         [TestMethod]
41         public void EuclideanDistancePoint_Test()
```

```
42          {
43              // Arrange
44              LocationTimeValue loc1 = new LocationTimeValue(0, 0, 0);
45              LocationTimeValue loc2 = new LocationTimeValue(1, 1, 1);
46
47              double expected = 1.732050808;
48
49              // Act
50              double dist = GeographyHelper.EuclideanDistance(loc1, loc2);
51
52
53              // Assert
54              Assert.AreEqual(dist, expected, 0.002, "Calculation error.");
55          }
56
57          ...
58      }
59  }
```

ADRIAN LINSEL

EDUCATION

OCT. 2016
Master of Science in **Applied Geosciences** (TU DARMSTADT)
Specialization: Applied Geology

OCT. 2014
Bachelor of Science in **Applied Geosciences** (TU DARMSTADT)

JUN. 2011
University Entrance Qualification (STARKENBURG-GYMNASIUM HEPPENHEIM)

PROFESSIONAL
EXPERIENCE

MAY 2017 - NOVEMBER 2020
Research assistant (TU DARMSTADT)

MAR. 2018 - JUL. 2018
Short term scholar (UNIVERSITY OF UTAH, USA)

AUG. 2014 - APR. 2017
Tutor and student assistant (TU DARMSTADT)

SCHOLARSHIPS

DEC. 2017 - NOV. 2020
PhD scholarship (FRIEDRICH-EBERT-STIFTUNG)

OCT. 2015 - SEPT. 2016
Deutschland-Stipendium