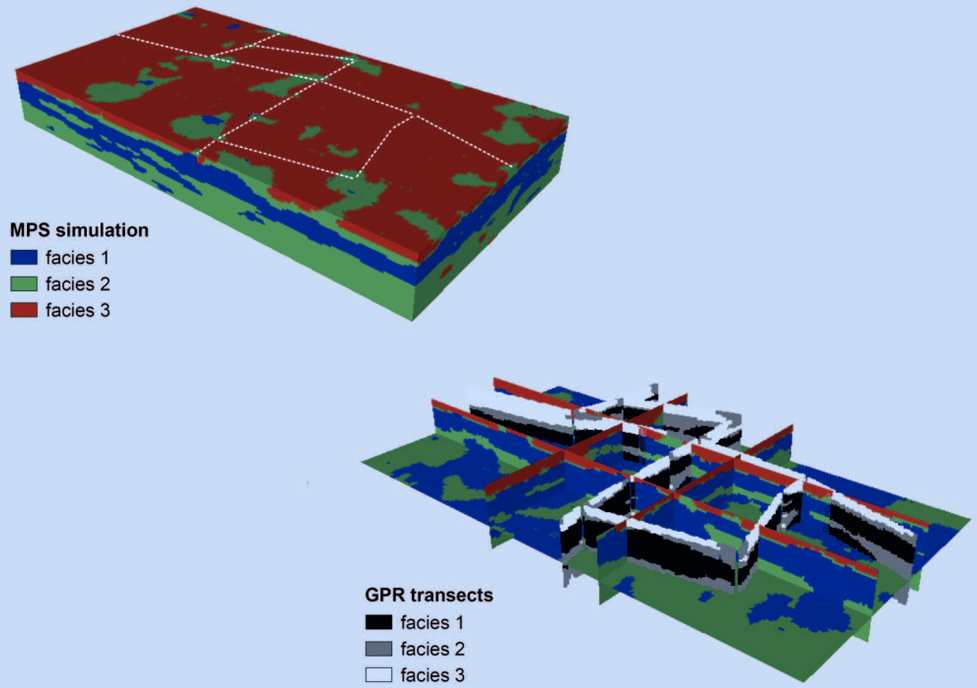
High resolution imaging and modeling of aquifer structure

Nils Güting



Energie & Umwelt / Energy & Environment
Band / Volume 383
ISBN 978-3-95806-253-5

JÜLICH FORSCHUNGSZENTRUM

Energie & Umwelt / Energy & Environment
Band / Volume 383
ISBN 978-3-95806-253-5

JÜLICH FORSCHUNGSZENTRUM

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 383

Forschungszentrum Jülich GmbH
Institute of Bio- and Geosciences
Agrosphere (IBG-3)

High resolution imaging and modeling of aquifer structure

Nils Güting

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 383

ISSN 1866-1793 ISBN 978-3-95806-253-5

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data are available in the
Internet at http://dnb.d-nb.de.

Publisher and Distributor: Forschungszentrum Jülich GmbH
Zentralbibliothek
52425 Jülich
Tel: +49 2461 61-5368
Fax: +49 2461 61-6103
Email: zb-publikation@fz-juelich.de
www.fz-juelich.de/zb

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Grafische Medien, Forschungszentrum Jülich GmbH

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment, Band / Volume 383

D 294 (Diss., Bochum, Univ., 2017)

ISSN 1866-1793
ISBN 978-3-95806-253-5

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JuSER)
at www.fz-juelich.de/zb/openaccess.

CONTENTS

ABSTRACT

Predictive modeling of groundwater flow and solute transport can help to protect groundwater resources and to remediate contaminated sites. It is challenging, however, to develop realistic groundwater models because it is difficult to characterize and model the complex heterogeneity of geologic media. In this work, we propose an approach that combines high resolution geophysical imaging and multiple-point statistical modeling to estimate the 3-D structure of aquifers.

Our study is carried out at the Krauthausen site, Germany, where 15 cross-borehole ground-penetrating radar (GPR) data sets were acquired in an alluvial sand and gravel aquifer. To analyze the GPR data, we apply a recently developed full-waveform inversion approach that is preferable, in terms of spatial resolution, over traditional ray-based inversion approaches. By stitching together the inverted tomograms from adjacent crosshole planes, we are able to image, with a decimeter-scale resolution, the aquifer's electrical properties (dielectric permittivity and electrical conductivity) along vertical cross-sections up to 50 m long and 10 m deep. Comparison of the GPR results with co-located direct-push profiles shows a strong correlation between the porosity derived from GPR dielectric permittivity and the porosity derived from direct-push neutron logs. We can show that the dielectric permittivity obtained from full-waveform inversion more accurately reconstructs sharp contrasts and fine-scale variations in porosity than the dielectric permittivity obtained from traditional ray-based inversion.

One problem with using GPR for hydrogeological site characterization is that GPR yields electrical properties which are only indirectly linked to hydraulic properties. We present two approaches to estimate hydrogeological facies from the GPR results. The first approach, based on k-means cluster analysis, is applied to GPR data from five adjacent crosshole planes. Cluster analysis of the GPR results suggests three facies. Densely spaced cone penetration tests, located along the GPR transect, confirm the number of facies and their spatial distribution in the aquifer cross-section. Grain size distributions and flowmeter data, available from one of the boreholes, show that the derived facies boundaries correlate with changes in grain size and porosity, and to a lesser extent with changes in hydraulic conductivity.

The second approach uses a logistic regression to relate the electrical properties from GPR to different sediments. Compared with the cluster analysis based zonation approach, the logistic regression approach has the advantage that it provides additional information about the reliability of classification. Vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data suggest that the GPR facies obtained from logistic regression are meaningful with regard to porosity and hydraulic conductivity. Comparison of the GPR facies with tracer test data shows a clear and consistent effect of the facies structure on the observed transport, which confirms the hydrogeological relevance of the identified facies. In particular, the GPR results suggest the presence of a thin layer of sand that would explain the observed tracer plume splitting. Because this sand layer is identified by GPR full-waveform inversion

but not by conventional GPR ray-based inversion we conclude that the improvement in spatial resolution due to full-waveform inversion is crucial to detect small-scale aquifer structures that are highly relevant for solute transport.

The GPR derived facies distribution provides valuable information about the aquifer's structure and connectivity. We demonstrate that this information can be used to generate stochastic realizations of the 3-D aquifer structure using multiple-point statistics (MPS). To enable 3-D MPS modeling on the basis of the 2-D GPR transects, we propose a modeling approach that combines sequential 2-D simulations with a 3-D reconstruction scheme. Because the GPR identified subsurface structures are utilized as training and conditioning data, this approach yields 3-D aquifer models that honour the structural information provided by the GPR tomograms. In this way, we are able to model the site-specific heterogeneity of the studied aquifer, which provides an excellent basis to set up realistic groundwater models for numerical flow and transport prediction.

ZUSAMMENFASSUNG

Die Modellierung und Vorhersage von Strömungs- und Transportprozessen im Grundwasser kann dabei helfen, Grundwasserressourcen zu schützen und kontaminierte Standorte zu sanieren. Es ist jedoch schwierig, realistische Grundwassermodelle zu entwickeln, da die komplexe Struktur des Untergrunds nur schwer zu erfassen ist. In dieser Arbeit stellen wir ein Verfahren vor, das geophysikalische Charakterisierung und "multi-punkt-statistische" Modellierung kombiniert, um die 3D-Struktur von Grundwasserleitern abzuleiten.

Wir analysieren 15 Radar-Datensätze, die in einem sandig-kiesigen Grundwasserleiter am Testfeld Krauthausen gemessen wurden. Die Radarmessungen wurden nach tomographischem Messprinzip zwischen jeweils zwei Bohrlöchern durchgeführt. Um die Radardaten auszuwerten, nutzen wir ein kürzlich entwickeltes Verfahren, das auf der Wellenform-Inversion von Bohrlochradardaten basiert. Die invertierten Tomogramme zeigen, mit einer Auflösung im Dezimeterbereich, die elektrischen Bodenparameter (dielektrische Permittivität, elektrische Leitfähigkeit) entlang von vertikalen Querschnittsflächen mit bis zu 50 m Länge und 10 m Tiefe. Die aus der dielektrischen Permittivität abgeleitete Porosität des Grundwasserleiters wird durch vertikale "direct-push"-Porositätsprofile bestätigt. Wir können zeigen, dass die Wellenform-Inversion der Radardaten eine höher auflösende Charakterisierung der Porositätsverteilung ermöglicht als die herkömmliche strahlenbasierte Radaranalyse.

Ein Problem bei der Anwendung von Radar für hydrogeologische Fragestellungen ist, dass die durch Radar ermittelten elektrischen Parameter nur indirekte Informationen über die hydraulischen Parameter des Untergrunds liefern. Wir entwickeln zwei verschiedene Ansätze, um hydrogeologische Einheiten (Fazien) aus den Radarergebnissen abzuleiten. Im ersten Ansatz nutzen wir eine Clusteranalyse für die Klassifizierung. Die Clusteranalyse der Radarergebnisse deutet auf eine Unterteilung des Grundwasserleiters in drei Einheiten hin. Vertikale "direct-push"-Sondierungen bestätigen, dass die aus den Radarergebnissen abgeleiteten Fazien den grundlegenden sedimentologischen Aufbau des Grundwasserleiters widerspiegeln. Korngrößendaten und Flowmetermessungen zeigen, dass die ermittelten Faziengrenzen mit Veränderungen in der Korngröße, der Porosität und, untergeordnet, der hydraulischen Durchlässigkeit korrelieren.

Im zweiten Klassifizierungsansatz nutzen wir ein logistisches Regressionsmodell, um die mit Radar ermittelten elektrischen Bodenparameter unterschiedlichen Sedimenten zuzuordnen. Die logistische Regression hat gegenüber der Clusteranalyse den Vorteil, dass sie zusätzliche Informationen zur Unsicherheit der Klassifizierung der Radar-Ergebnisse liefert. Vertikale "direct-push"-Profile, Flowmetermessungen und Korngrößendaten zeigen, dass die mit Radar abgegrenzten Sedimentfazien durch unterschiedliche Porositäten und hydraulische Durchlässigkeiten charakterisiert sind. Der Vergleich mit den Ergebnissen eines Tracerexperiments zeigt einen klaren Zusammenhang zwischen der auf Basis der Radarergebnisse abgeschätzten Sedimentverteilung und dem beobachteten Transportverhalten des Tracerstoffes. Insbesondere zeigen die

Radarergebnisse eine dünne Sandschicht, die eine Erklärung für die im Experiment beobachtete Aufspaltung der Tracerstoffwolke liefert. Indem wir zeigen, dass diese Sandschicht mit herkömmlicher strahlenbasierter Radaranalyse nicht detektiert werden kann, demonstrieren wir, dass die Wellenform-Inversion von Radardaten entscheidende Vorteile für die Charakterisierung von Grundwasserleitern bietet.

Die mit Radar abgebildeten Sedimentstrukturen liefern wertvolle Informationen über die Struktur und die Konnektivität des Untergrunds. Wir zeigen, dass es auf Basis dieser Informationen möglich ist, die 3D-Struktur des Grundwasserleiters zu modellieren. Dafür entwickeln wir einen "multi-punkt-statistischen" Modellierungsansatz, der sequentielle 2D-Simulationen mit einem 3D-Modellierungskonzept kombiniert. Da die Radartomogramme als Trainings- und Konditionierungsdaten genutzt werden, wird die spezifische Struktur des untersuchten Grundwasserleiters explizit bei der Modellierung berücksichtigt. Das Ergebnis sind 3D-Untergrundmodelle, die eine exzellente Grundlage bieten, um Grundwasserströmung und Stofftransport im Untergrund vorherzusagen.

1 INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

Constituting over 95% of the Earth's unfrozen freshwater, groundwater is the most important freshwater resource on Earth. Particularly in areas with limited or polluted surface waters, groundwater often represents the only supply for the production of drinking water. Ever-increasing water demands and anthropogenic pollution, however, lead to depletion of clean groundwater reserves in many parts of the world [Wada et al., 2010]. Protecting the quality of groundwater, and securing its sustained availability, is thus considered a priority goal for both the present and the future [WHO, 2006].

Achieving this goal requires a precise understanding of flow and transport processes in aquifers as a basis to assist and guide stakeholders in groundwater related questions. For example, water suppliers or farmers may want to know how much water they can pump from an aquifer without overexploiting it. Or, in the case of a contaminated site, the question may arise of where to place and how long to operate remediation wells in order to capture a contaminant plume. One way to answer this kind of questions is to employ groundwater models that numerically simulate flow and transport processes in the subsurface [Kinzelbach, 1992]. The use of such models is appealing because (ideally) they offer the opportunity to evaluate the consequences of an action before actually performing this action. This makes it possible to "try out" several possible strategies for solving some problem in the model, and choose the most effective one for solving the problem in reality.

With the development of versatile groundwater flow and transport modeling codes such as MODFLOW [Harbaugh, 2005] and FEFLOW [Trefry and Muffels, 2007], along with a strong increase in computational power over the recent past, detailed simulation of complex groundwater systems on fine meshes has become feasible. Parametrizing such detailed models, however, poses a problem. The problem lies in the enormous variability of the parameters that control flow and transport in the subsurface, and in the difficulties associated to measuring these parameters and representing their variability properly in a model.

Flow and transport processes in aquifers are mainly controlled by two parameters, the hydraulic conductivity and the porosity. While the porosity shows only relatively mild variations in aquifers, the hydraulic conductivity typically varies over several orders of magnitude and exerts the primary control on groundwater flow and solute migration [Dagan, 1989; Koltermann and Gorelick, 1996]. Classical hydrogeological methods to measure hydraulic conductivity include pumping tests, which yield large-scale averaged effective medium properties; flowmeter tests, which yield one-dimensional profiles of hydraulic conductivity along wells; and permeameter tests, which yield point information. None of these methods is able to combine local resolution with spatial coverage. In principle, if the spatial distribution of hydraulic conductivity (and porosity)

in an aquifer was known at every point in space, simulation of groundwater flow and solute transport in that aquifer would be relatively straight-forward (at least, say, for non-reactive transport, known boundary conditions, etc.). However, perfect knowledge of the parameter field within an aquifer is unrealistic for every practical application, because it is impossible to measure parameters at every point in space. As a consequence, there is a conflict between the scarce information typically available from data, and the exhaustive parametrization required by fine meshed groundwater models.

1.2 THE NEED TO ACCOUNT FOR SPATIAL CONNECTIVITY

Traditionally, this conflict has been approached by measuring the hydraulic conductivity at several locations in the aquifer, characterizing the spatial correlation of the measured data in terms of a variogram, and using variogram-based geostatistical methods such as sequential gaussian simulation or sequential indicator simulation [Deutsch and Journel, 1998] to stochastically generate synthetic parameter fields, which eventually can be used to parametrize the grid nodes of a numerical groundwater model. It was soon realized, however, that the variogram fails to capture spatial connectivity, and thus, may be insufficient to characterize (and model) the heterogeneity of aquifers [Gómez-Hernández and Wen, 1998; Western et al., 2001; Journel and Zhang, 2006]. This is illustrated in Figure 1,

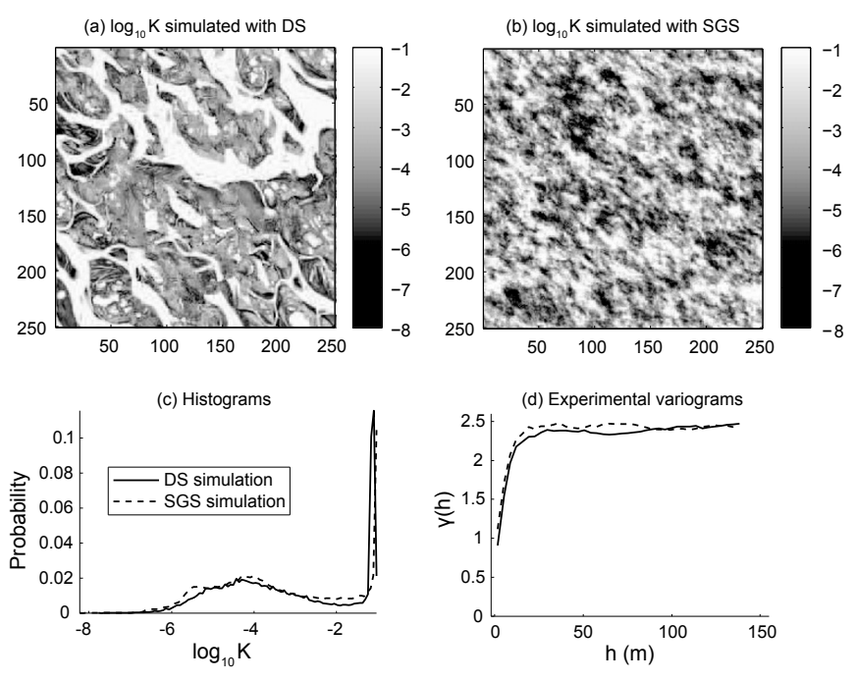


Figure 1: Two heterogeneous fields showing different connectivity patterns but identical histograms and variograms (from Renard and Allard [2013]).

which shows two synthetic hydraulic conductivity fields, one generated using the direct sampling algorithm [Mariethoz et al., 2010], the other generated by sequential gaussian simulation [Deutsch and Journel, 1998]. Although the two fields show quite different connectivity patterns, their histograms and variograms are practically identical.

Spatial connectivity in heterogeneous aquifers has received considerable attention lately and it has been recognized that connected structures explain a range of phenomena that have been observed in natural aquifers. For example, Bianchi et al. [2011] demonstrated that connected structures can explain the early arrival and late-time tailing that is often observed in tracer experiments. Meier et al. [1998] and Knudby and Carrera [2006] analyzed pumping test data and found that differences in the early-time hydraulic response measured in different observation wells could be explained by different degrees of hydraulic connectivity between pumping and observation wells. Generally, connected structures of high permeability act as preferential flow paths and lead to increased water fluxes and increased transport velocities [Fogg et al., 2000; Bianchi et al., 2011; Zheng et al., 2011]; whereas connected structures of low permeability act as flow barriers and lead to decreased water fluxes and decreased transport velocities [Zinn and Harvey, 2003; Huysmans and Dassargues, 2012]. As a consequence, knowledge about the site-specific connectivity structure of an aquifer is crucial to properly predict flow and transport. Zheng et al. [2011] reviewed 25 years of research carried out at the intensively studied MADE site [Boggs et al., 1992] and concluded that the spatial arrangement and the connectivity of heterogeneous sediments is more important to understanding and modeling transport than the variation of hydraulic conductivity alone. Therefore, they argued that novel exploration techniques must be developed that are able to map the detailed distribution and connectivity of subsurface structures. Similarly, Renard and Allard [2013] proposed that identifying connected structures in the field and honoring connectivity in the generation of subsurface models may be the step forward towards more realistic groundwater models.

1.3 MULTIPLE-POINT STATISTICS AND TRAINING IMAGES

Motivated by the inability of variogram-based geostatistics to model the complex connectivity patterns that one finds in nature, alternative modeling approaches based on multiple-point statistics (MPS) have been developed [Guardiano and Srivastava, 1993; Strebelle, 2002; Straubhaar et al., 2011; Mariethoz et al., 2010; Mariethoz and Caers, 2014]. MPS approaches accommodate the stochastic simulation of both categorical and continuous variables and are specifically aimed at representing complex connected structures and curvilinear patterns.

The principle idea of MPS modeling is to generate models by reproducing the multiple-point spatial statistics from a user-defined conceptual model, called training image (TI). The TI is supposed to display the type of geological structures (e.g. channels, bars, layers) that exist at the site being modeled. These structures are then reproduced in the simulation grid, taking conditioning data (if available) into account. In this way, complex models with geologically realistic structures and connectivity patterns can be obtained (Figure 2); but it is crucial to understand that MPS will only produce appro-

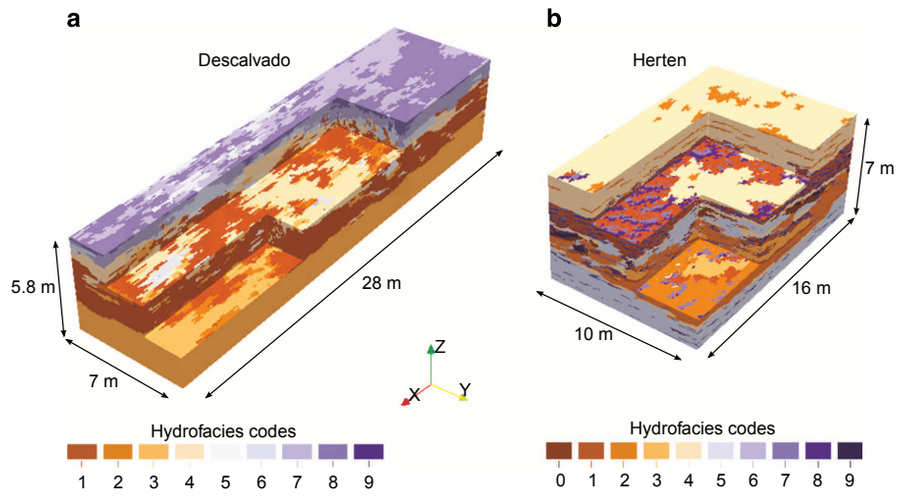


Figure 2: 3-D models of the Descalvado (Brazil) and Herten (Germany) aquifer obtained from multiple-point statistical simulations (from Bayer et al. [2015]). Training images were derived from outcrop photographs.

priate models when an appropriate TI can be provided. The most critical step in MPS modeling thus consists in identifying a TI that properly reflects the specific heterogeneity of the site being modelled. There are different strategies how to obtain such a TI.

One possible approach is to use some external, conceptual TI that reflects the type of spatial structures that are deemed representative for the site being modeled. For example, if the goal is to model an aquifer that is known to represent a former braided river deposit, it might be appropriate to use a TI that includes high permeable gravel channels and low permeable clay lobes [Hermans et al., 2015]. Conceptual TIs for different geological settings can be retrieved from data bases [Comunian and Renard, 2008; Mariethoz and Caers, 2014], or the TI itself can be modeled using process-imitating or object-based methods [Maharaja, 2008; Comunian et al., 2013]. But the problem with using such conceptual TIs for real-case applications is that, obviously, it can be difficult do decide whether a conceptual TI is representative for the (unknown) site to be modeled. Hermans et al. [2015] and Scheidt et al. [2015] therefore proposed to allow several different conceptual TIs, and narrow down the range of these initially plausible TIs by later comparison of model results with acquired data. Although this strategy alleviates the difficulty in deciding for one specific TI it still requires that at least one of the provided TIs properly reflects the actual field.

An alternative strategy was used by Bayer et al. [2015] and Pickel et al. [2015] who used digitized outcrop images as TIs. The rationale behind this approach is that outcrops contain site-specific information, and therefore, it might be advantageous to rely on them rather than on some external, possibly inadequate TI. Outcrops, however, are often not available and, at best, located at the outer boundaries of the domain to be modeled. An interesting alternative, therefore, would be to derive TIs from geophysical

images that could be obtained, non-invasively, directly at the location of interest. Derivation of TIs through geophysical imaging, however, requires that the spatial resolution of the geophysical imaging technique is high enough to resolve the characteristic structure of the subsurface without smoothing out critical features [Hermans, 2014]. Moreover, the derived geophysical images need to be interpreted, for example, by establishing some relationship between geophysical attribute and lithology.

1.4 BENEFITS AND DRAWBACKS OF GEOPHYSICAL IMAGING

Geophysical imaging techniques for the shallow subsurface, such as ground-penetrating radar (GPR) or electrical resistivity tomography (ERT), offer a great potential for characterizing aquifers because they combine local resolution and spatial coverage. Hence, these methods are able to bridge the gap between point measurements and volume-averaged effective properties that are typically obtained from hydrogeological methods. Due to their sensitivity to water content, pore fluid conductivity and clay content, GPR and ERT can be applied, for example, to detect lithological boundaries in aquifers [Bowling et al., 2005] or to quantitatively estimate variations in porosity and clay content [Doetsch et al., 2010]. Particularly cross-borehole tomographic setups are highly effective for characterizing the subsurface because they yield high resolution spatial information on full cross-sections. This makes it possible to assess the lateral continuity of subsurface structures, which is the key to characterize spatial connectivity.

There are two major difficulties, however, that hinder the routine application of geophysical methods for hydrogeological site characterization. The first difficulty is that the hydrogeological interpretation of geophysical data is often challenging because there is a lack of direct and universal relationships between geophysical and hydrogeological properties [Binley et al., 2015]. The second difficulty is that conventional geophysical inversion techniques are often limited in spatial resolution, thus yielding only relatively smooth images of the subsurface. This becomes a problem when the goal is to describe solute transport that is well known to depend on small-scale structures acting as preferential flow paths or flow barriers [Zheng et al., 2011].

Recent advances in full-waveform inversion of GPR data have shown great potential for characterizing aquifers with higher resolution [Klotzsche et al., 2015]. While waveform approaches have been applied in seismic inversion for decades, their application to crosshole GPR inversion is still relatively new, and there have been just a few studies that used full-waveform inversion to interpret GPR field data [Belina et al., 2012; Klotzsche et al., 2013, 2014; Keskinen et al., 2017]. Full-waveform inversion of crosshole GPR data focuses on the analysis of radar waves traveling from a transmitter placed in a borehole to a receiver placed in another borehole several meters away. Because underground structures alter the radar wave as it travels through the subsurface, the wave picked up by the receiver contains information about structures located between the boreholes. Full-waveform inversion aims at revealing these structures by searching for a dielectric permittivity and electrical conductivity distribution that explains the measured data. In contrast to conventional traveltime inversion approaches that employ approximate ray-based forward simulators to reconstruct first arrival times, full-waveform

inversion employs a numerical solution of Maxwell's equations to reconstruct entire waveforms [Ernst et al., 2007a,b; Meles et al., 2010]. Full-waveform inversion thus uses a significantly larger amount of the measured data and simulates more accurately the propagation of electromagnetic waves in the ground. As a consequence, this approach allows characterizing the subsurface with a considerably improved spatial resolution (Figure 3).

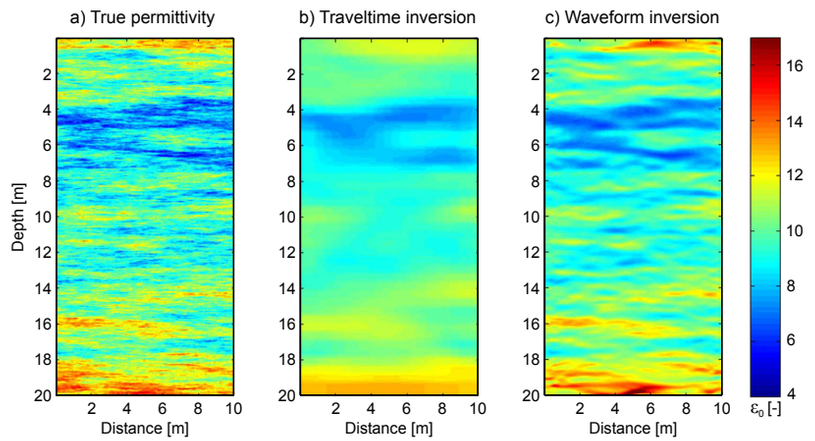


Figure 3: Synthetic experiment illustrating the differences in resolution between traveltime and full-waveform inversion of crosshole GPR data (modified from Belina et al. [2012]). The true permittivity field is used to generate synthetic radar traces. Traveltime and full-waveform inversion are employed to reconstruct the true permittivity field.

This improvement in spatial resolution makes crosshole GPR full-waveform inversion very interesting for hydrogeological site characterization because it might allow for imaging the subsurface with the resolution needed to predict solute transport in aquifers. Moreover, it appears promising to use this high resolution imaging tool to derive TIs for MPS simulations. In this way, GPR full-waveform inversion would be used to collect site-specific information about the detailed spatial structures that exist at a site; and MPS simulations would be used to generate 3-D aquifer models by explicitly considering the spatial structures and connectivity patterns inferred from GPR.

1.5 OBJECTIVES AND THESIS OUTLINE

This thesis has two main objectives. The primary objective is to evaluate the benefit of high-resolution GPR full-waveform inversion for hydrogeological site characterization. The secondary objective is to test if the GPR tomograms can be utilized in a multiple-point statistical modeling framework to derive 3-D aquifer models that honor the site-specific structure and connectivity inferred from GPR.

The two objectives are addressed in three individual Chapters – the primary objective in Chapters 2 and 3, the secondary objective in Chapter 4. Each Chapter is self-contained

and includes a specific introduction, a description of the methods used, and a discussion of the obtained results. Although focusing on different aspects, all three Chapters use data from the Krauthausen test site, partly collected within the framework of this thesis, and partly being available from previous studies carried out at this site. The remaining manuscript is organized as follows:

Chapter 2 investigates if full-waveform inversion of crosshole GPR data followed by cluster analysis is a suitable approach to delineate lithological facies. GPR data from five adjacent crosshole planes are analyzed and partitioned into facies using cluster analysis. The GPR derived zonation is compared to the zonation derived independently from densely spaced cone penetration tests. On the basis of this comparison the reliability of the obtained facies classification is discussed.

Chapter 3 focuses on the hydrogeological interpretation of the information provided by GPR. A logistic regression model is employed to translate the GPR derived electrical properties into facies, whose hydrogeological relevance is assessed based on vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data. The GPR derived facies distribution is compared to the breakthrough of a tracer observed in a field experiment. To evaluate the true benefit of using GPR full-waveform inversion instead of traditional GPR ray-based inversion, the results from both inversion strategies are compared and discussed.

Chapter 4 explores different approaches to utilize the GPR cross-sections as training images to derive 3-D aquifer models using MPS. The main challenge consists in carrying out 3-D MPS simulations on the basis of the 2-D information from GPR. Two existing approaches to this problem are tested and compared in a quantitative manner. In addition, a new combination of the two former approaches is proposed and evaluated.

Finally, Chapter 5 summarizes the overall conclusions of this thesis and indicates possible future research directions.

2 HIGH RESOLUTION IMAGING OF AQUIFER STRUCTURE USING CROSSHOLE GPR FULL-WAVEFORM INVERSION†

ABSTRACT

Spatially highly resolved mapping of aquifer heterogeneities is critical for the accurate prediction of groundwater flow and contaminant transport. Here, we demonstrate the value of using full-waveform inversion of crosshole ground penetrating radar (GPR) data for aquifer characterization. We analyze field data from the Krauthausen test site, where crosshole GPR data were acquired along a transect of 20 m length and 10 m depth. Densely spaced cone penetration tests (CPT), located close to the GPR transect, were used to validate and interpret the tomographic images obtained from GPR. A strong correlation was observed between CPT porosity logs and porosity estimates derived from GPR using the complex refractive index model (CRIM). A less pronounced correlation was observed between electrical conductivity data derived from GPR and CPT. Cluster analysis of the GPR data defined three different subsurface facies, which were found to correspond to sediments with different grain size and porosity. In conclusion, our study suggests that full-waveform inversion of crosshole GPR data followed by cluster analysis is an applicable approach to identify hydrogeological facies in alluvial aquifers and to map their architecture and connectivity. Such facies maps provide valuable information about the subsurface heterogeneity and can be used to construct geologically realistic subsurface models for numerical flow and transport prediction.

2.1 INTRODUCTION

Reliable prediction of groundwater flow and solute transport is needed for environmental engineering tasks such as the definition of protection zones for drinking water wells or the development of efficient remediation strategies at contaminated sites. However, predicting subsurface flow and transport is challenging due to the complex heterogeneity found in most geologic media. Physical and chemical properties in the subsurface vary over orders of magnitude and their spatial variability exerts a primary control on groundwater flow and solute migration [Dagan, 1989]. Sedimentary deposits, which commonly represent soils and aquifers, are often composed of several distinct units or facies, separated by recognizable boundaries at which subsurface properties such as grain size or hydraulic conductivity may abruptly change [Miall, 1985; Fogg, 1986;

† this chapter has been published as: Gueting N., Klotzsche A., van der Kruk J., Vanderborght J., Vereecken H., Englert A. (2015), Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests, *Journal of Hydrology* **524**, 680–695, http://dx.doi.org/10.1016/j.jhydrol.2015.03.030.

Koltermann and Gorelick, 1996]. This has motivated the conceptualization of heterogeneous sedimentary aquifers as assemblages of distinct (hydro-)lithological facies with less variability within than between facies [Anderson, 1989; Webb and Anderson, 1996; Barrash and Clemo, 2002; Riva et al., 2006; Bayer et al., 2011].

Field and modeling studies have highlighted the influence of facies architecture and geometry on flow and transport processes [Zheng and Gorelick, 2003; Feyen and Caers, 2006; Ronayne et al., 2008; Zheng et al., 2011; Haendel and Dietrich, 2012]. In particular, the spatial connectivity of facies appears to be of substantial importance, because connected structures of high permeability act as preferential flow paths and lead to increased water fluxes and increased transport velocities, while connected structures of low permeability act as flow barriers and lead to decreased water fluxes and decreased transport velocities [Zinn and Harvey, 2003; Knudby and Carrera, 2006; Huysmans and Dassargues, 2009; Bianchi et al., 2011; Renard and Allard, 2013]. Hence, reliable prediction of flow and transport in the subsurface critically depends on a detailed characterization of the spatial distribution of subsurface facies.

Unfortunately, a detailed characterization of the subsurface is difficult to obtain in practice because the subsurface is not easily accessible through measurements. Conventional subsurface characterization techniques tend to rely on pumping tests or borehole based geophysical or lithological logs. Pumping tests yield appropriate upscaled properties with regard to flow but suffer from a limited spatial resolution of local properties, which are needed for proper transport prediction. Borehole logs provide detailed information about the vertical distribution of local properties but are inherently one-dimensional in nature. Any inference based on one-dimensional vertical profiles, will be blind to the lateral distribution of subsurface properties, and thus, the interconnectivity of facies. Outcrop analyses can be applied to map the distribution of facies on outcrop walls, which provides valuable information about the lateral connection of facies [Huysmans et al., 2008; Huysmans and Dassargues, 2009; Bayer et al., 2011; Comunian et al., 2011]. However, the characteristics derived from an outcrop are not necessarily valid for the specific conditions at locations some distance away from the outcrop.

In contrast, minimally invasive tomographic geophysical imaging methods such as crosshole ground penetrating radar (GPR) can be used to map the spatial distribution of subsurface properties on full cross-sections directly at the location of interest. The two-dimensionality of the tomographic images provides the opportunity to characterize the lateral distribution of structures and to evaluate their architecture and connectivity. Traditional tomographic inversion of GPR data is carried out using ray-based techniques [Holliger et al., 2001; Maurer and Musil, 2004]. Because ray-based methods consider only the first arrival times and the first-cycle amplitudes they exploit only a limited amount of the information contained in the full recorded waveform. As a consequence, the spatial resolution of such methods is limited, and only relatively smooth images of the subsurface can be obtained [Belina et al., 2009; Klotzsche et al., 2010]. In contrast, full-waveform inversion techniques make use of the full recorded waveform, which includes information beyond the first arrival times and first-cycle amplitudes. As a consequence, full-waveform inversion techniques are capable to resolve subsurface properties with higher spatial resolution and yield tomographic images with a significantly improved level of detail [Ernst et al., 2007a; Meles et al., 2010; Yang et al., 2013; Klotzsche et al., 2013, 2014].

Recent studies showed the potential of GPR full-waveform tomography for hydrogeological application by comparing GPR results with independent measurements of porosity [Yang et al., 2013; Klotzsche et al., 2013] and hydraulic conductivity [Klotzsche et al., 2013]. Within the scope of hydrogeological site characterization the strength of full-waveform GPR tomography lies in its potential to bridge the gap in terms of resolution and coverage that exists among traditional hydrogeological methods such as small-scale core analyses and large-scale pumping tests. However, this advantage comes along with difficulties regarding the hydrogeological interpretation because GPR returns the electrical properties which are only indirectly related to hydrogeological parameters such as porosity or hydraulic conductivity. Without information about the site-specific relationship between electrical and hydrogeological properties, the use of GPR for hydrogeological site characterization can therefore be limited. Another difficulty, in field applications, is to evaluate the reliability of the obtained tomographic images, because in the absence of secondary information it is difficult to validate if the structures seen in the tomographic images represent real subsurface structures and not inversion artifacts.

In the present study we show the value of applying full-waveform GPR tomography in combination with complementary investigation tools to hydrogeologically characterize a site. We analyze field data from the Krauthausen test site, where GPR data were recently acquired along several cross-borehole planes. Because the Krauthausen test site has been intensively investigated in previous studies [Vereecken et al., 2000; Tillmann et al., 2008], we are able to confront the GPR results with densely spaced secondary data from cone penetration tests (CPT), grain size analyses and flowmeter measurements which allow detailed validation and interpretation of the tomographic images obtained from GPR. In detail, our approach consists of the following steps: Firstly, a GPR full-waveform inversion is applied to infer the subsurface permittivity and electrical conductivity distribution along a vertical aquifer cross-section of approximately 20 m length and 10 m depth. Results are validated by comparison with co-located CPT porosity and electrical conductivity logs. Secondly, a cluster analysis is conducted to partition the obtained GPR tomographic images into clusters. In this way, different lithological facies such as sand and gravel that differ in their permittivity and electrical conductivity signatures can be distinguished. The outcome of the cluster analysis is a cross-sectional map showing the spatial distribution of different facies in the aquifer. We discuss the reliability of the facies classification by comparing the distribution of facies obtained from GPR with the distribution of facies obtained independently from CPT. Finally, we use the combined information from GPR, CPT, grain size analyses and flowmeter measurements to develop a hydrogeological interpretation of the obtained facies and to characterize them in terms of grain size, porosity and hydraulic conductivity.

2.2 MATERIAL AND METHODS

2.2.1 Study site

The Krauthausen test site, set up by the research center Jülich in 1993, is located approximately 10 km northwest of the city of Düren, Germany, in the southern part of the

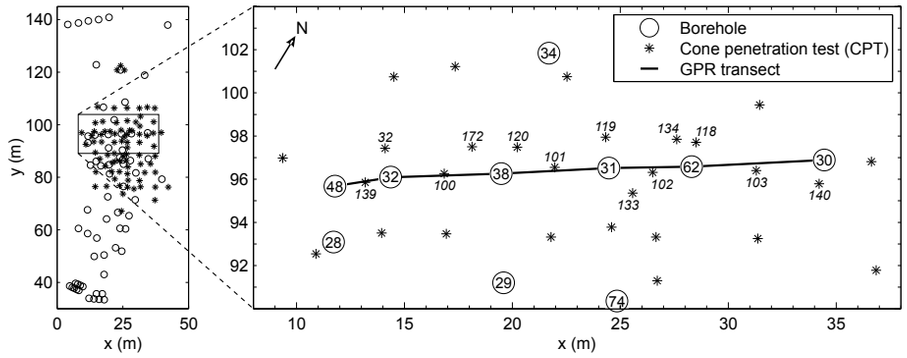


Figure 4: Map of the Krauthausen test site with the location of boreholes (open circles) and cone penetration tests (asterixes). The black solid line in the close-up (right) shows the location of five adjacent cross-borehole GPR planes acquired in the central part of the test site. Selected cone penetration tests located close to the GPR transect are labeled with italic numbers.

Lower Rhine Embayment. A detailed description of the test site is given by Vereecken et al. [2000]. In the last decades, several laboratory and field techniques have been applied at Krauthausen with the goal to study the spatial distribution of aquifer parameters and its effect on solute migration. These techniques include laboratory characterization of sediment samples [Döring, 1997], field pumping and flowmeter tests [Li et al., 2007, 2008], tracer experiments [Vereecken et al., 2000; Vanderborght and Vereecken, 2001], geophysical imaging methods [Kemna et al., 2002; Müller et al., 2010; Oberröhrmann et al., 2013], and cone penetration tests [Tillmann et al., 2008]. Due to the extensive set of subsurface information retrieved in previous studies, Krauthausen provides excellent opportunities to test and validate novel aquifer exploration techniques.

The present study focuses on the uppermost aquifer in the central part of the test site where a series of closely spaced boreholes and cone penetration tests were available (Figure 4). The unconfined aquifer is approximately 10 m thick and composed of alluvial terrace sediments, which were deposited by a local braided river system of the river Rur on top of older Rhine and Maas sediments [Englert, 2003]. According to Döring [1997] and Tillmann et al. [2008] the uppermost aquifer can be broadly divided into three layers (Figure 5a): A bottom layer composed of sandy to gravely grain size, which extends from 6 to 11.5 m depth and is characterized by varying sand to gravel ratio; a well sorted sand layer extending from 4 to 6 m depth; and a poorly sorted gravel layer extending from 1 to 4 m depth. The base of the aquifer is formed by thin layers of clay and sand, at approximately 12 m depth. On top of the aquifer, a loamy soil layer has developed. The groundwater level shows seasonal variations from 1 to 3 m depth. The aquifer sediments are characterized by an average clay content of 2% [Döring, 1997] and by a mean total porosity of 26% [Vereecken et al., 2000]. The mean hydraulic conductivity based on the results of a large scale pumping test is 3.8×10^{-3} m/s [Englert, 2003].

It is important to note that the generalized aquifer cross-section shown in Figure 5a represents a simplified conceptual model. Treating the aquifer as a perfectly stratified medium is an oversimplification of the aquifer's true structure (at Krauthausen and

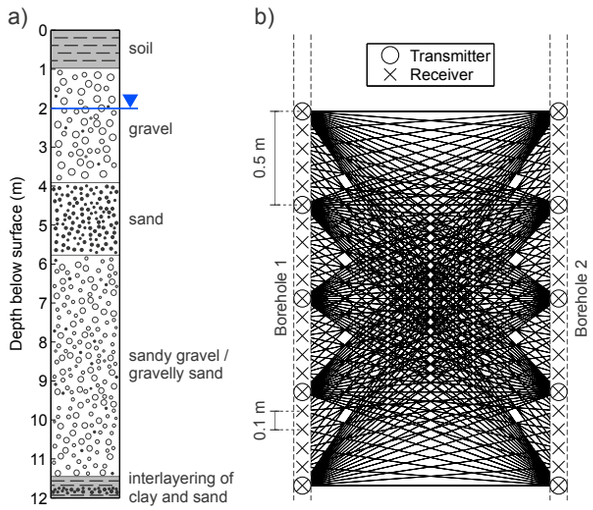


Figure 5: a) Generalized cross-section of the uppermost aquifer after Döring [1997] and Tillmann et al. [2008]. b) Schematic sketch of the crosshole GPR acquisition setup used in the present study. For clarity, the setup is shown for a vertical extension of 2 m only.

elsewhere). For the true aquifer architecture we expect significant lateral variations in layer thickness and properties, as well as the existence of discrete non-layered structures. Characterization thereof is what the present study focuses on.

2.2.2 Crosshole ground penetrating radar tomography

Data acquisition

GPR data were acquired along a transect in the central part of the test site, where six closely spaced boreholes were available (Figure 4). In total, the transect covers a lateral distance of 22.7 m and consists of five adjacent crosshole tomographic planes with individual borehole separations ranging from 2.6 to 6.2 m. We used the Sensors and Software pulseEKKO Pro system with 200 MHz antennas and a semi-reciprocal acquisition setup, with transmitter and receiver spacing of 0.5 m and 0.1 m, respectively (Figure5b). This setup was recently found to represent a compromise between sufficient ray coverage for high resolution inversion results, on the one hand, and affordable acquisition time and computational costs, on the other hand [Klotzsche et al., 2010; Oberröhrmann et al., 2013]. In order to avoid refracted/reflected radar waves associated to the groundwater table [Klotzsche et al., 2010, 2012], GPR data were acquired only within the saturated zone. Measurements were conducted from 3 m depth, which was around 1 m below the groundwater table during the measurements, down to approximately 8-10 m depth, depending on the depth of the boreholes. Note that the transmitter and receiver positions within the boreholes were accurately determined by using a log-

ging tool that measures the lateral deviation of a borehole from a straight vertical line. This was necessary to avoid inversion artifacts.

Full-waveform inversion methodology

To infer the subsurface relative dielectric permittivity (ϵ) and electrical conductivity (σ) from the measured GPR data, we applied a full-waveform inversion. We used an inversion scheme introduced by Ernst et al. [2007a,b] and further developed by Meles et al. [2010] and Klotzsche et al. [2010]. Here, we describe only briefly the major steps in the inversion scheme. For a more detailed description of the inversion procedure and its implementation, the reader is referred to Meles et al. [2010] and Klotzsche et al. [2010].

The overall goal of the inversion is to find the spatial distributions of ϵ and σ whose simulated radar response best matches the observed radar data. This is achieved in an iterative approach using a gradient method to minimize the misfit function

$$C = 0.5 \left\| E^{mod} - E^{obs} \right\|^2, \tag{1}$$

where E^{obs} is the observed radar data, for all transmitter-receiver combinations over the total observation time window, and E^{mod} is the corresponding data predicted by a forward model. Eq. (1) is a sample-by-sample comparison in which all data points are weighted equally. In order to enable the use of the forward model, which is based on a 2-D finite difference time domain solution of Maxwell's equations, the observed data was converted from 3-D to 2-D using the approach by Bleistein [1986]. Initial estimates for the distributions of ϵ and σ were derived using standard ray-based travel time and amplitude inversion [Holliger et al., 2001; Maurer and Musil, 2004]. The full-waveform inversion was performed with the ray-based inversion results as starting models and with an effective source wavelet obtained through a deconvolution method [Ernst et al., 2007a]. For each iteration of the full-waveform inversion, the wavefield predicted by the forward model is subtracted from the observed wavefield to calculate the residual wavefield. Local gradients for the ϵ and σ models are computed through a cross correlation of the back-propagated residual wavefield and the model-predicted wavefield. The gradient indicates if the model parameter values should be increased or decreased in order to reduce the misfit function. In addition, individual step lengths for ϵ and σ are determined which indicate the magnitude of the model updates. The ϵ and σ models are then simultaneously updated using the derived gradients and step lengths [Meles et al., 2010]. Except from a smoothing of the gradient, no regularization is applied. Note that in order to avoid inversion artifacts close to the boreholes, model updates are suppressed within the domain closer than 30 cm to the boreholes. The updated ϵ and σ models are used to parametrize the forward model in the subsequent iteration step until the iteration is stopped when the root mean squared (RMS) misfit between model-predicted and observed data changes less than 1%.

It should be noted that the full-waveform inversion critically depends on adequate starting models for the ϵ and σ distribution. More precisely, for each transmitter-receiver combination, the observed data and the model-predicted data (obtained initially with the starting models) need to show an overlap within half the dominant wavelength in order to ensure proper convergence of the full-waveform inversion. If the offset

between observed and modeled data is greater than half the dominant wavelength, the full-waveform inversion will converge to a local minimum and give unreliable results [Meles et al., 2011; Klotzsche et al., 2014]. To ensure proper convergence of the full-waveform inversion, we carefully checked the match between observed and initially model-predicted data, prior to running the full-waveform inversion.

Another important step in the full-waveform inversion is the definition of perturbation factors for ϵ and σ, which are required for the computation of the model update step lengths [Meles et al., 2010]. We determined appropriate values for the perturbation factors by testing the convergence of the full-waveform inversion with a range of values prescribed for the perturbation factors. With appropriate perturbation factors, the full-waveform inversion converges smoothly towards the optimal solution (maximal reduction of the RMS misfit), whereas too high or too low perturbation factors can be identified by non-smooth convergence and/or minor reduction of the RMS misfit. Note that the full-waveform inversion typically converges to the optimal solution for quite a range of perturbation factors (i.e. within that range the inversion results are robust with regard to changes in perturbation factors), which makes it relatively easy to find appropriate values for the perturbation factors, in practice.

2.2.3 Complex Refractive Index Model (CRIM)

Petrophysical relationships can be used to translate electrical properties obtained from GPR into hydrogeologically more relevant properties such as soil moisture or porosity. Several empirical, semi-empirical and theoretical relationships have been proposed in the literature [Topp et al., 1980; Dobson et al., 1985; Linde et al., 2006; Brovelli and Cassiani, 2008; Steelman and Endres, 2011]. Commonly used approaches are so called mixing models, which are based on the intuitive idea that the bulk permittivity of a multi-phase medium can be related to the permittivities of the individual phases. In its general form, the mixing formula for an n-phase medium is

$$\epsilon_b^\alpha = \sum_{i=1}^{n} V_i \epsilon_i^\alpha, \tag{2}$$

where ϵ_b is the bulk permittivity of the medium, V_i and ϵ_i are the volume fraction and the permittivity of the i-th phase, and α describes the geometrical arrangement of the individual phases within the medium. For fully saturated porous media, Eq. (2) can be written as

$$\epsilon_b^\alpha = (1-\phi)\epsilon_s^\alpha + \phi\epsilon_f^\alpha, \tag{3}$$

where ϕ is the porosity, and ϵ_s and ϵ_f are the permittivities of the solid grain matrix and of the pore fluid, respectively. The exponent α has theoretical upper and lower boundaries of 1 and -1, corresponding to the cases where individual phases are arranged in layers parallel and perpendicular to the applied electric field, respectively [Brown, 1956]. Birchak et al. [1974] argued that for an isotropic phase arrangement, α should take a value of 0.5, which is equivalent to a volumetric average of the complex index of refraction. With $\alpha = 0.5$, Eq. (2) becomes the Complex Refractive Index Model (CRIM) [Birchak et al., 1974]. Although the CRIM relationship is based on an oversimplified

description of the complexity of natural porous media [Dobson et al., 1985; Brovelli and Cassiani, 2008], the CRIM has generally been found to agree well with experimental data and to produce adequate results for various types of soils [Roth et al., 1990; Robinson et al., 2003]. In the present study, we apply the CRIM to convert bulk permittivity, obtained from GPR full-waveform inversion, into porosity, by rearranging Eq. (3) to

$$\phi = \frac{\sqrt{\epsilon_b} - \sqrt{\epsilon_s}}{\sqrt{\epsilon_f} - \sqrt{\epsilon_s}}. \tag{4}$$

2.2.4 Cone penetration tests

A total number of 78 cone penetration tests (CPT) were conducted at Krauthausen and are described in detail by Tillmann et al. [2008]. The CPT survey focused on the central area of the site, where CPT were performed with a horizontal sampling interval of around 1-3 m (Figure 4). At each CPT location, vertical profiles of mechanic cone resistance, natural gamma, bulk density and water content were obtained down to an average depth of approximately 13 m, covering the entire thickness of the uppermost aquifer. For a limited number of CPT, additional measurements of electrical conductivity and pore water pressure were conducted. Note that the sleeve friction, a commonly measured CPT parameter, was not measured in this study. The vertical sampling interval for all measurements was 10 cm. In the present study, we focus on a subset of 13 CPT, which were selected due to their close proximity to the GPR transect. The exact locations and ID numbers of the selected CPT are shown in Figure 4.

2.2.5 Hydraulic conductivity from flowmeter test and grain size

Flowmeter tests at the Krauthausen site were conducted by Schneider [1995] and are described by Li et al. [2008]. In the field, a constant pumping rate of $Q_p = 1500$ l/h was applied for a sufficiently long time to establish pseudo steady state conditions in the vicinity of the borehole. An electromagnetic flowmeter was used to measure the cumulative vertical flow rate in the borehole as function of depth. Measurements were conducted every 10 cm, starting at the deepest and ending at the highest accessible point within the filtered length of the borehole. If flow towards the borehole is assumed horizontal and there is no ambient flow in the borehole, then the increase of the cumulative flow rate at increment i, ΔQ_i, can be used to calculate the hydraulic conductivity [Javandel and Witherspoon, 1969; Molz and Morin, 1989]

$$K_i = \frac{\Delta Q_i b}{Q_p \Delta z_i} \overline{K}, \tag{5}$$

where K_i is the horizontal hydraulic conductivity in the aquifer at increment i, b is the aquifer thickness, Δz_i is the thickness of the ith measurement increment, and $\overline{K}$ is the aquifer´s depth averaged hydraulic conductivity. Note that we applied Eq. (5) with a constant increment of $\Delta z_i = 20$ cm and with $\overline{K}$ obtained from a large scale pumping test conducted in the central part of the test site [Englert, 2003].

In addition, hydraulic conductivity was estimated from grain size. For a limited number of boreholes at Krauthausen, sediment samples were extracted during drilling and the grain size distribution was determined through sieve analysis [Vereecken et al., 2000]. The grain size data were used to derive hydraulic conductivity estimates based on the empirical relationship given by Bialas and Kleczkowski [1970]

$$K = 0.0036\, d_{20}^{2.3}, \tag{6}$$

where K is the hydraulic conductivity and d_{20} is the 20th percentile of the sediment particle size. According to Bialas and Kleczkowski [1970], Eq. (6) applies for sediments with a coefficient of uniformity, U, in the range of 0.2 <U <60, in which U is defined as the ratio of the 60th and 10th percentile of the particle size. It should be noted that the indirect method of estimating hydraulic conductivity based on grain size can be generally expected to be less reliable compared to methods that directly measure flow properties (such as pumping tests or flowmeter tests) because hydraulic conductivity is related to the pore structure rather than simply to the grain size [Koltermann and Gorelick, 1996]. In the present study we apply Eq. (5) and (6) to obtain hydraulic conductivity profiles for borehole B32, which is one of the boreholes within the GPR transect (Figure 4).

2.2.6 Cluster analysis

The principle of cluster analysis is to partition multivariate data into relatively homogeneous units by grouping data points based on their distances in a multi-dimensional parameter space. Data pre-processing typically involves data normalization (z-score) to assign similar weights to each measured variable and principal component analysis to reduce data dimensions which facilitates cluster analysis and data visualization. Many different clustering approaches exist [Everitt et al., 2001; MacKay, 2002]. Generally, one can distinguish between hierarchical methods that produce nested clusters (i.e. clusters are embedded in others) and partitioning methods that produce a unitary classification into a pre-defined number of clusters. While partitioning techniques are known to be computationally more efficient than hierarchical methods, they require additional user-made specifications such as the number of clusters or the initial cluster centroid positions [Everitt et al., 2001].

Moreover, a distinction can be made between crisp and fuzzy clustering techniques. Crisp methods assign each data point to only exactly one cluster (i.e. to the most likely one), whereas fuzzy methods assign each data point a partial membership to all clusters [Paasche et al., 2010]. A commonly applied crisp partitioning method is the k-means approach [MacQueen, 1967]. The k-means algorithm partitions data points into a predefined number of clusters by minimizing the sum of distances from each data point to its cluster centroid over all clusters. Due to its conceptual simplicity and its computational efficiency, the k-means algorithm allows rapid classification of large multivariate data sets in a straight-forward manner. A notable limitation of k-means clustering is that all clusters are assumed to have the same isotropic variability (i.e. clusters are assumed to be spherical and of same size). This can cause problems when applied to data characterized by clusters with non-spherical shape and/or different size [MacKay, 2002; Rogiers, 2013]. For the interpretation of such data sets, more flexible clustering techniques

such as model-based clustering [Doetsch et al., 2010; Rogiers, 2013] or density-based clustering [Ester et al., 1996] can be favorable. Despite its limitations, however, the k-means algorithm has been proven useful to extract the basic structural information from various types of multivariate data including airborne geophysical data [Anderson-Mayes, 2002], grain size data [Riva et al., 2006] and tomographic geophysical data [Dietrich et al., 1998; Tronicke et al., 2004; Dietrich and Tronicke, 2009].

In the present study, we applied the k-means algorithm to partition the bivariate GPR inversion results (permittivity and electrical conductivity) into clusters. Prior to the cluster analysis, permittivity and electrical conductivity were z-transformed to assign similar weights to the two parameters. A critical step in the cluster analysis is the specification of the number of clusters. To determine the number of natural clusters in a data set, we employed the variance ratio criterion (VRC) [Calinski and Harabasz, 1974; Tronicke et al., 2004; Dietrich and Tronicke, 2009]. Application of the VRC involves running the cluster analysis with different values prescribed for the number of clusters and calculating

$$VRC = \frac{SS_B}{k-1} / \frac{SS_W}{n-k}, \tag{7}$$

where SS_B is the between cluster sum-of-squares, SS_W is the within clusters sum-of-squares, k is the number of clusters and n is the number of samples. The number of clusters that maximizes the VRC value is supposed to optimally represent the data [Calinski and Harabasz, 1974]. Because the final partition can be sensitive to the randomly selected initial cluster distribution, the code was run multiple times to ensure stability of the final solution.

In a similar manner, we applied the k-means algorithm to the multivariate data from 13 CPT profiles located close to the GPR transect (Figure 4). As described in Section 2.2.4, vertical profiles of cone resistance, gamma ray, bulk density and water content were measured at each CPT location, whereas electrical conductivity and pore water pressure were measured only at some CPT locations. To avoid dealing with incomplete data, we conducted the cluster analysis by considering only the measurements available for all CPT (i.e. cone resistance, gamma ray, bulk density, water content). Prior to the cluster analysis, the CPT data were z-transformed and converted into principal components. The first three principal components were used for the cluster analysis, while the fourth principal component was dropped because it explained only 5% of the total variance. The number of clusters was assigned according to the variance ratio criterion.

2.3 RESULTS AND DISCUSSION

2.3.1 GPR full-waveform inversion results and comparison with CPT data

A full-waveform inversion was applied to infer the subsurface permittivity and electrical conductivity along the transect shown in Figure 4. The obtained tomographic images are shown in Figure 6a and 7a, respectively. Black circles and crosses show the exact transmitter and receiver positions within the boreholes, determined by deviation log. Shaded areas depict regions outside of the tomographic cross-section. It is important

to note that each of the five cross-borehole planes was inverted separately, i.e. independent from the adjacent planes. Nevertheless, consistent structures in the permittivity and electrical conductivity distribution can be observed left and right of the boreholes, which indicates robust inversion results. Convergence of the full-waveform inversion for the individual crosshole planes was reached after 16 iterations minimum (B62-B30) and 27 iterations maximum (B31-B62). Compared with the ray-based inversion results, the full-waveform inversion led to a reduction in RMS misfit of 51% minimum (B48-32) and 62% maximum (B32-B38). The stability of the final solution was controlled by running the full-waveform inversion repeatedly with different values prescribed for the perturbation factors (not shown here). Because we found the inversion results to be robust over quite a wide range of perturbation factors, we can expect that the obtained results are reliable.

The tomographic images obtained through full-waveform inversion show a high level of detail and contain spatial structures at the decimeter scale (Figure 6a and 7a). Note that similar to the findings of Klotzsche et al. [2013, 2014] the full-waveform inversion considerably improved the spatial resolution of the initially derived ray-based inversion results (not shown here). A striking feature in the GPR images is the strong increase in both the permittivity and the electrical conductivity between 3 and 4 m depth which

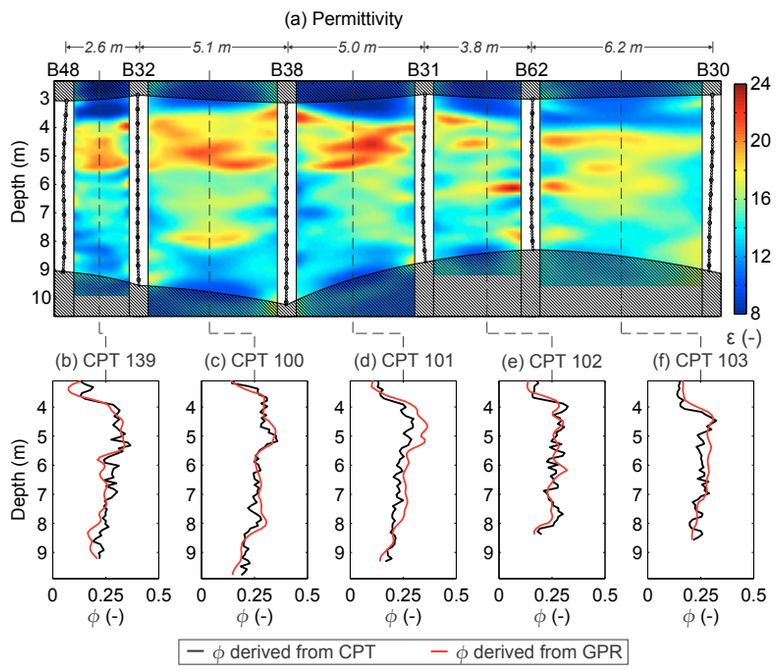


Figure 6: (a) Permittivity full-waveform inversion results along the GPR transect. (b-f) Comparison of vertical porosity profiles derived from CPT (neutron logs) with co-located GPR permittivities converted into porosity using the CRIM (Eq. (4)). The exact locations of the CPT profiles within the GPR transect are indicated by the dashed lines.

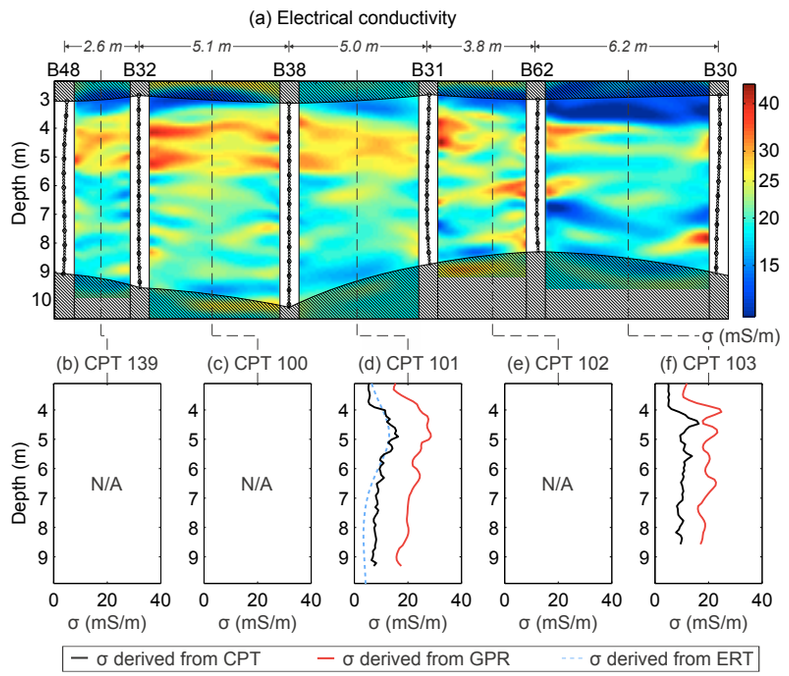


Figure 7: (a) Electrical conductivity full-waveform inversion results along the GPR transect. (b-f) Comparison of vertical profiles of electrical conductivity derived from CPT and GPR. Due to technical problems during the CPT measurements, electrical conductivity logs are available only for CPT 101 and 103. Along CPT 101, the GPR and CPT results can additionally be compared with ERT results [Müller et al., 2010].

indicates a distinct sedimentologic boundary in the aquifer material at that depth. To validate the reliability of the obtained GPR results, we compared them with cone penetration test data. As can be seen in Figure 4, five CPT are located directly within the GPR transect, and thus, can be compared with the GPR results. In the following, we show a comparison of porosity estimates derived from GPR and from CPT; and a comparison of electrical conductivity estimates derived from GPR and from CPT.

Derivation of porosity from GPR and CPT

GPR porosities were derived by converting permittivity inversion results into porosities using the CRIM relation (Eq. (4)). The CRIM was parametrized with $\epsilon_f = 84$, which represents the permittivity of water at 10 °C [Eisenberg and Kauzmann, 1969], and with $\epsilon_s = 4.5$, which represents the solid grain permittivity of quartz [Carmichael, 1982]. Note that the choice of using the permittivity of quartz for ϵ_s is based on the results of picnometer measurements, which gave a mean sediment density of 2.65 g/cm^3 at Krauthausen [Vereecken et al., 2000] which corresponds to the density of quartz.

CPT porosities were derived from neutron log data. Tillmann et al. [2008] used a calibration that was performed by Elgoscar 2000 Ltd. in Budapest, Hungary, to relate

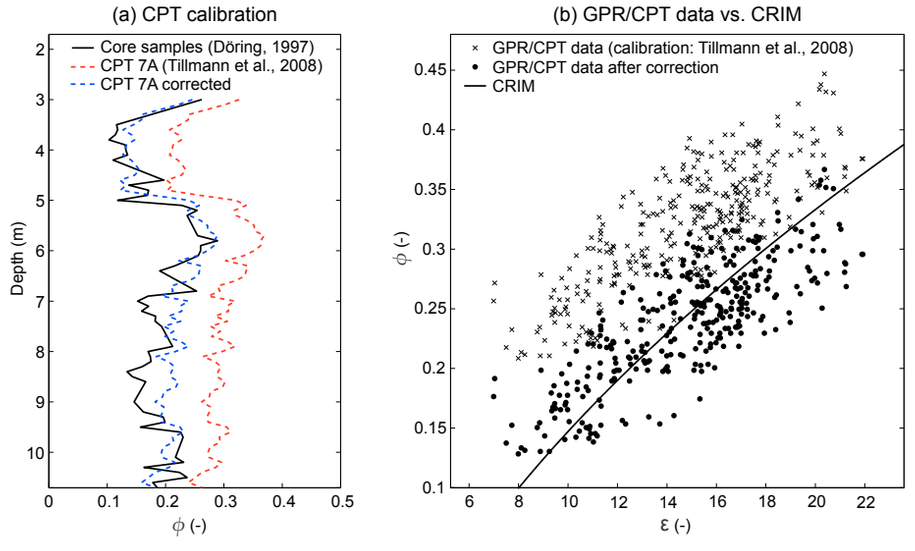


Figure 8: (a) Comparison of CPT porosity estimates with laboratory porosity measurements. (b) Cross-plot of co-located GPR and CPT data. The x-axis denotes the permittivity derived from GPR. The y-axis denotes the porosity derived from CPT. Black solid line shows porosities predicted by the CRIM (Eq. (4)).

neutron log data to water content, which, under saturated conditions and in the absence of significant amount of clay, represents porosity. We critically reviewed the calibration used in Tillmann et al. [2008] by comparing porosity data derived from CPT with laboratory porosity measurements. The black solid line in Figure 8a shows the porosity of sediment samples extracted from borehole B7, which is located in the northern part of the test site. The sediment samples were oven-dried and their porosity was determined by measuring their weight, volume and dry matrix density in the laboratory [Döring, 1997]. The red dashed line shows porosities obtained by applying the calibration used in Tillmann et al. [2008] to neutron log data from a CPT located close to borehole B7. Obviously, the CPT porosities reproduce the trend of the laboratory measurements quite well, but they show consistently higher absolute values. The blue dashed line shows the effect of applying a simple constant-shift correction ($\phi_{\text{corrected}} = \phi - 0.08$) to the CPT porosities. Obviously, this leads to a much better match of the two curves.

Another indication that the calibration used by Tillmann et al. [2008] yields overestimated porosity values is shown in Figure 8b. The cross-plot shows how CPT derived porosities compare with GPR porosities predicted by the CRIM. The calibration according to Tillmann et al. [2008] clearly results in distinctly higher porosities derived from CPT (crosses) compared with the CRIM-predicted porosities (black solid line). Corrected CPT porosities (filled circles), however, scatter closely around the CRIM, and thus agree fairly well with the porosity derived from GPR. Note that changing the value for ϵ_s in Eq. (4), would result in a shift of the CRIM curve along the x-axis in Figure 8b. A satisfactory fit between the CRIM curve and the Tillmann et al. [2008] calibrated data could

be obtained by using an ϵ_s-value of 1.8 instead of 4.5. This must be considered, however, to be an unrealistically low value for ϵ_s at Krauthausen, because it is known from previous studies that the aquifer material at Krauthausen basically consists of quartz [Döring, 1997; Vereecken et al., 2000] and the ϵ_s-value of quartz is 4.5 (Carmichael, 1982). We therefore assume that, for unknown reasons, the calibration used by Tillmann et al. [2008] overestimated the true sediment porosity. A simple constant-shift correction ($\phi_{\text{corrected}} = \phi - 0.08$), however, seems to be applicable to compensate for the overestimation. In the remainder of this paper, thus, corrected CPT porosities are used.

Quantitative comparison of GPR and CPT results

Figure 6b-f and Figure 7d, f show porosity and electrical conductivity profiles derived from GPR and CPT. While the GPR and CPT porosity profiles match quite well, notable deviations exist between the GPR and CPT electrical conductivity profiles. To assess the agreement between GPR and CPT results in a quantitative manner, we calculated correlation coefficients and linear regression parameters. While the correlation coefficient measures how close the observed values come to falling on a straight line in a cross-plot, the slope and intercept of the linear regression allow to assess the deviations from a 1:1 relationship [Isaaks and Srivastava, 1989]. Results for the porosity and for the electrical conductivity data are shown in Figure 9a and b, respectively.

Porosities derived from GPR and CPT show a strong linear relationship with a linear correlation coefficient of $r = 0.8$. Moreover, the slope and intercept of the linear regression are close to 1 and 0, respectively, which means that the data fall close to the 1:1 line.

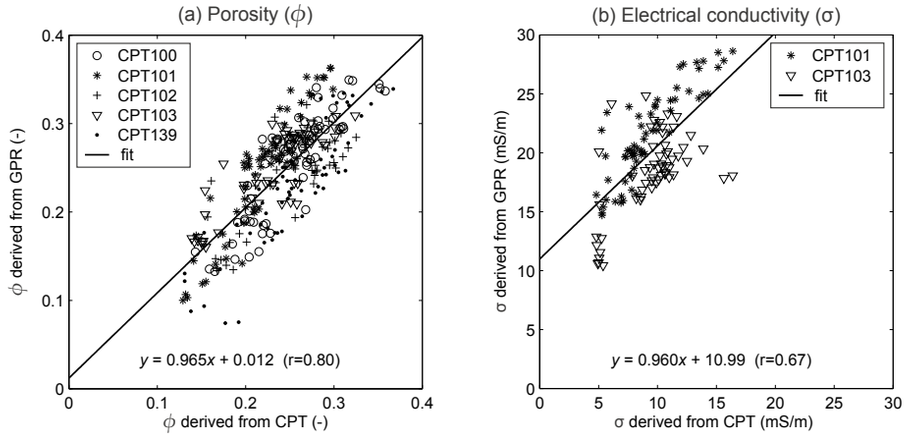


Figure 9: (a) Cross-plot of porosities derived from CPT and GPR for the five profiles for which co-located porosity data exist (Figure 6b-f). (b) Cross-plot of electrical conductivities derived from CPT and GPR for the two profiles for which co-located conductivity data exist (Figure 7d, f). Regression lines through all data points are depicted in black. The corresponding straight-line equations are given at the bottom of the cross plot, r is the correlation coefficient.

The picture is different for the electrical conductivity. First, the correlation between GPR and CPT results is less pronounced ($r = 0.67$). More importantly, while the slope of the linear regression is close to 1, the intercept significantly departs from 0. In other words, the GPR and CPT electrical conductivity profiles agree fairly well regarding the variations in conductivity, but they show a static shift in absolute values. To further investigate the discrepancy in absolute values, we compare the electrical conductivity obtained from GPR and CPT additionally with results from Müller et al. [2010]) who investigated electrical conductivity at Krauthausen based on cross-borehole electrical resistivity tomography (ERT). At CPT location 101, co-located electrical conductivity values are available from CPT, GPR and ERT. The vertical profiles obtained from the three methods are compared in Figure 7d. Obviously, the absolute electrical conductivity values derived from ERT better match the CPT results than the GPR results. Apart from that, it is interesting to note that the three methods show considerable differences in spatial resolution, i.e. ERT basically misses the small-scale variability which is most pronounced in the CPT profile and still clearly visible in the GPR profile.

In the following, we discuss two possible explanations why the absolute conductivity values derived from GPR are higher than the absolute values derived from CPT and ERT. One explanation is that the different absolute values are caused by the inherent differences of the applied methods: ERT and CPT infer electrical conductivity on the basis of direct current resistivity measurements, whereas GPR uses the attenuation of the GPR signal to infer the electrical conductivity at higher frequencies. The attenuation of a GPR signal in the subsurface can be affected by (frequency-dependent) attenuation effects caused by water relaxation or signal scattering [Davis and Annan, 1989]. Because these effects cause an attenuation, which acts in addition to the attenuation caused by electrical conductivity, this could explain the higher absolute conductivities obtained from GPR.

Another possible explanation is that the higher absolute conductivity values derived from GPR are caused by biased starting models used in the full-waveform inversion. As described in Section 2.2.2, we derived starting models for the permittivity and conductivity by applying a ray-based travel time and amplitude inversion, prior to the full-waveform inversion. While ray-based methods generally provide reliable (but smooth) estimates for the permittivity, estimation of electrical conductivity is less straightforward and, in particular the absolute values of the electrical conductivity are typically less reliable [Holliger et al., 2001; Maurer and Musil, 2004]. Unfortunately, inaccuracies in the absolute values of the electrical conductivity starting model are propagated into the amplitude of the estimated effective source wavelet used for the full-waveform inversion because wavelet and electrical conductivity model are coupled [Busch et al., 2012]. As a consequence, absolute electrical conductivities obtained through full-waveform inversion would be biased if the absolute values of the electrical conductivity starting model were biased. As discussed in detail in Busch et al. [2012], a strategy to overcome this issue would be to update the wavelet simultaneously together with the conductivity and permittivity models during the full-waveform inversion.

Irrespective of the discrepancies in the absolute values, however, GPR and CPT results agree fairly well regarding the absolute spatial variations in electrical conductivity (slope of regression line is close to 1 ($a = 0.96$), and the correlation coefficient is 0.67). In the next section, we present the results of the cluster analysis. Because the cluster analysis

interprets the differences in data rather than their absolute values, this allows us to exploit the electrical conductivity inversion results even if absolute values are uncertain.

2.3.2 Cluster analysis of GPR and CPT data

The cluster analysis was applied to classify the GPR and CPT data into clusters, related to sediment facies with different parameter signatures. Figure 10a and b show the clustered GPR and CPT data. Note that the cluster analysis was conducted independently with the GPR and the CPT data set. The number of clusters, prescribed in accordance to the variance ratio criterion, is three for the GPR data (Figure 10c) and six for the CPT data (Figure 10d). However, in the part of the aquifer cross-section where GPR and CPT data overlap, a consistent number of three clusters appear in both data sets and their spatial distribution in the aquifer cross-section shows a good match (Figure 11b). As a measure for the consistency between the spatial distribution of GPR and CPT clusters, we calculated for each CPT profile the percentage of pixels for which the same cluster ID number was obtained from GPR and from CPT. The results are listed in Table 1. It is important to note that only five of the 13 CPT profiles are located directly within the

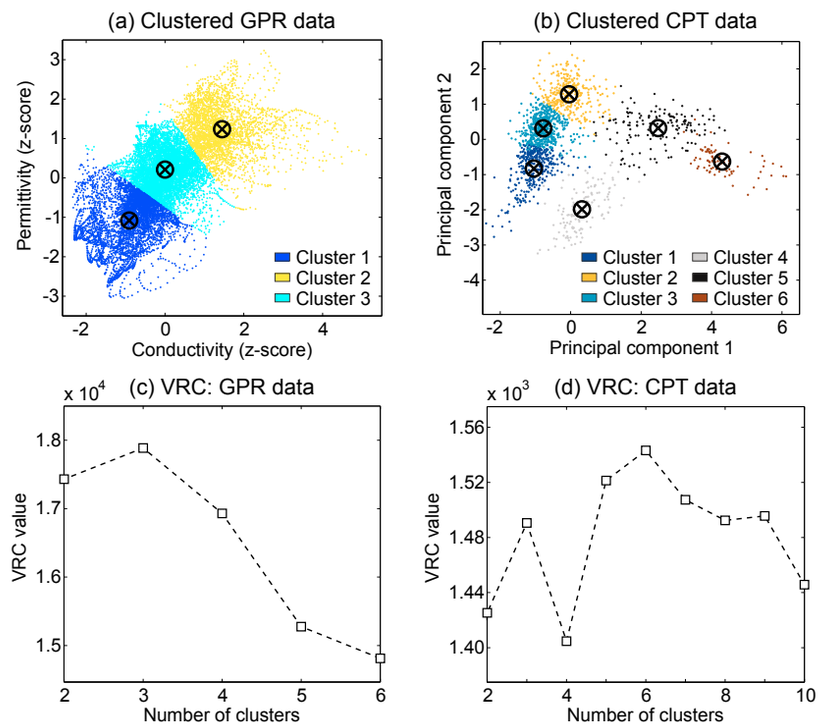


Figure 10: (a, b) K-means clustering of GPR and CPT data. (c, d) The optimal number of clusters according to the variance ratio criterion (Eq. (7)).

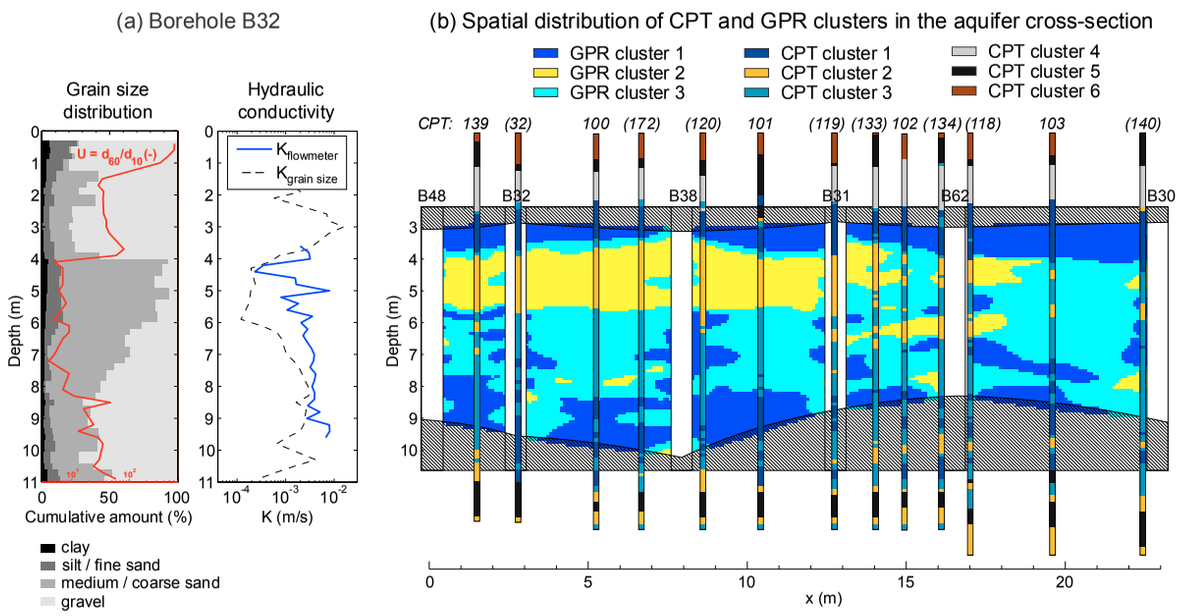


Figure 11: a) Grain size distribution and hydraulic conductivity for borehole B32 (second borehole from left in the GPR transect). Red line represents the coefficient of uniformity, U. Blue and dashed black line represent hydraulic conductivity from flowmeter test and from grain size, respectively. b) Spatial distribution of GPR and CPT clusters. Parentheses indicate CPT that are not located directly within the GPR transect but at a certain lateral distance from it. Exact lateral displacements for the CPT are provided in Table 1.

Table 1: Consistency of GPR and CPT cluster assignment along CPT profiles.

CPT	Distance to GPR transect	Consistent Classification[a] Single profile	Averaged
139	0.1 m	55%	
100	0.1 m	89%	
101	0.2 m	91%	77%
102	0.2 m	74%	
103	0.3 m	73%	
32	1.3 m	-	
172	1.2 m	89%	
120	1.2 m	57%	
119	1.5 m	- %	63%
133	1.2 m	64%	
134	1.3 m	56%	
118	1.1 m	67%	
140	1.1 m	42%	

[a]GPR cluster ID number = CPT cluster ID number

GPR transect. The other eight CPT are located in a lateral distance of 1-1.5 m away from the GPR transect, which can be expected to lead to differences between CPT and GPR results. As can be seen in Table 1, the average agreement between CPT and GPR cluster assignment is indeed larger for the five CPT located close to the GPR transect (77%), than for the CPT located farther away from the transect (63%). Overall, the close agreement of the spatial distribution of GPR and CPT clusters suggests that the cluster analysis, although performed independently with GPR and CPT data, yet defined the same facies in the aquifer. This is a promising result because it suggests that the facies obtained from GPR provide a meaningful classification of the subsurface, not only with regard to electrical properties but also with regard to further sediment properties, such as those measured by CPT.

Reliability of the facies classification obtained through cluster analysis

Despite the overall consistency of the spatial distribution of GPR and CPT clusters, there are some locations in the aquifer cross-section where the cluster assignment based on GPR and CPT is inconsistent. This indicates an uncertainty of the facies classification at these locations. Figure 12 shows the facies classification of co-located GPR and CPT data in a cross-plot of permittivity and electrical conductivity derived from GPR. Only the data along the five CPT profiles which are located directly within the GPR transect are used. For clarity the same color scale was used to represent GPR clusters and CPT clusters. This means, data points that were assigned the same cluster ID number based on GPR and based on CPT, have uniform color; data points for which the cluster assignment based on GPR and CPT is inconsistent, are two-colored. The legend in Figure 12 lists all possible events, sorted by their number of occurrences (n). In total, there are 244 (i.e. 77%) consistent data points, and 74 (i.e. 23%) inconsistent data points. Inter-

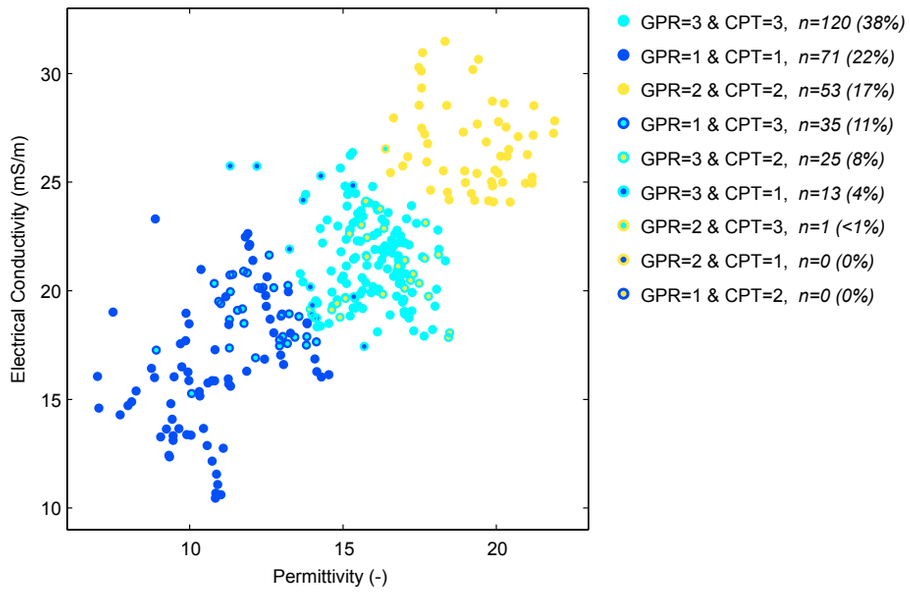


Figure 12: Cluster assignment of co-located GPR and CPT data. Color of circle edges refer to the cluster assignment based on GPR. Color of circle inner area refer to the cluster assignment based on CPT. n is the number of occurrences.

estingly, the inconsistent data points are not evenly distributed in the cross-plot. Only few inconsistent data points plot in the lower left corner and in the upper right corner, the majority of inconsistent data points are scattered around the borders between clusters. This suggests that there are regions in the permittivity-conductivity parameter space, where GPR and CPT facies assignment are more reliable; and there are other regions where GPR and CPT facies assignment are less reliable. Since the classification of facies based on GPR data was completely independent from the classification based on CPT data, inconsistencies in classification can be considered to be related to uncertainty of classification. We used the match/mismatch between the GPR and CPT facies assignment at co-located locations as a basis to approximate the reliability of the facies classification in the entire GPR transect. We defined the ratio R for a certain parameter vector $\vec{X}$ in the GPR derived permittivity-conductivity parameter space as:

$$R(\vec{X}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(\vec{X}-\vec{P}_i)^2}{\frac{1}{m}\sum_{j=1}^{m}(\vec{X}-\vec{Q}_j)^2}, \tag{8}$$

where $\vec{P}_{1,\dots,n}$ and $\vec{Q}_{1,\dots,m}$ represent the parameter vectors of consistently and inconsistently assigned data points for the respective cluster. For example, if $\vec{X}$ was assigned to cluster 1, then $\vec{P}$ would represent the data points that were consistently assigned to cluster 1 based on GPR and CPT; and $\vec{Q}$ would represent the data points that were assigned to cluster 1 only either by GPR or by CPT. Low values for R indicate high reliability of the

facies classification because $\vec{X}$ has a permittivity-conductivity signature similar to those data points where GPR and CPT cluster assignment are consistent; high values for R indicate low reliability of the facies classification because $\vec{X}$ has a permittivity-conductivity signature similar to those data points where GPR and CPT cluster assignment are inconsistent. Note that R does not represent the reliability of the GPR or CPT data themselves. R solely approximates the reliability of the cluster analysis based classification of the data into facies.

Figure 13 shows the distribution of R values in the aquifer cross-section. The R value of a grid cell depends on its relative position, in the data space, to consistently and inconsistently classified data points. Sharp contrasts in R thus reflect the situation when spatially close cells differ in permittivity and/or conductivity and thus plot in data space regions with different densities of consistent/inconsistent points. As expected, high R values occur prevalently along the boundaries between facies (compare with Figure 11b). This reflects our previous finding that the data points for which GPR and CPT cluster assignment are inconsistent generally plot close to the borders between clusters in the parameter space. It is interesting to note, however, that at some locations in the aquifer, facies boundaries are accompanied by only thin bands of high R values, while at other locations, facies boundaries are accompanied by considerably thicker bands of high R values. A thin band of high R values suggests little uncertainty of the exact location of the facies boundary; a thick band of high R values suggests larger uncertainty for the exact location of the facies boundary. We believe that these differences can be attributed to different types of sedimentologic boundaries within the aquifer. For instance, the uppermost facies boundary between 3 and 4 m depth is characterized by a considerably thin band of high R values (Figure 13). At this depth, the grain size distribution in

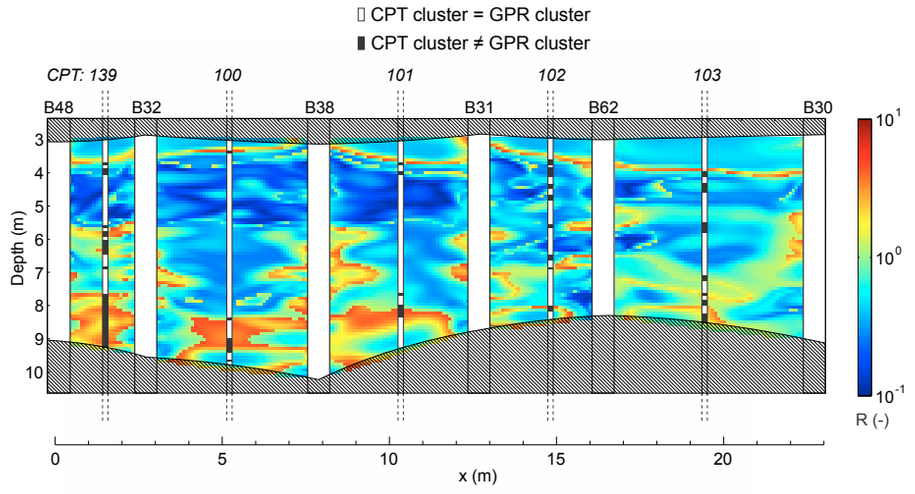


Figure 13: Estimated reliability of the facies classification based on the match/mismatch between GPR and CPT cluster assignment at co-located locations (indicated in white/black). Low *R* values suggest high reliability, high *R* values suggest low reliability (see Eq. (8)).

borehole B32 shows an abrupt change from gravel to underlying sand (Figure 11a). This suggests that this boundary represents an erosional-surface boundary, where sedimentation was interrupted by a phase of erosion. At such a boundary, sediment properties can abruptly change and facies are clearly separated from each other. Note that a distinct boundary at that depth is also indicated by the full-waveform inversion results, which show an abrupt increase in permittivity and electrical conductivity between 3 and 4 m depth (Figure 6a and 7a).

In contrast, in the lower part of the aquifer cross-section (7-10 m depth), facies boundaries are generally characterized by relatively thick bands of high R values (Figure 13). At this depth, the grain size in B32 shows a gradual increase with increasing depth (Figure 11a) and the permittivity and electrical conductivity distributions show relatively mild variations (Figure 6a and 7a). Gradual changes in subsurface properties can be caused by gradual changes in deposition conditions or by cyclic deposition patterns, which are commonly observed in fluvial systems [Miall, 1996]. If a sediment is characterized by gradual changes in properties, however, it is difficult to define facies and to delineate the boundaries between them.

This highlights an important implication of using crisp clustering methods (such as k-means) to define subsurface facies. Crisp cluster algorithms partition each data point to one of the clusters. More specifically, if a data point plots close to the border between two clusters in the parameter space, small changes in the parameter signature of this data point will result in the assignment of that point to either the one or the other cluster. In the presence of abrupt changes in the subsurface properties, this is not an issue because there are only few or no data points which lie between clusters in the parameter space. Such a boundary seems to exist between 3 and 4 m depth, where the thin band of high R values suggests that the exact location of the corresponding facies boundary is accurately defined. In contrast, if subsurface properties change gradually, more data points are characterized by a parameter signature, which plots close to the borders between clusters in the parameter space. In this case, data points cannot be assigned unambiguously to either the one or the other cluster and facies are not clearly separated from each other. This seems to be the reason for an increased uncertainty of the facies classification in the lower part of the aquifer where thick bands of high R values suggest a larger uncertainty of the exact location of facies boundaries.

In conclusion, our analysis suggests that the simple k-means approach used in the present study is an applicable tool to delineate facies boundaries if the subsurface is characterized by distinct sedimentologic changes. However, if the subsurface is characterized by rather subtle or gradual sedimentologic changes, the exact location of facies boundaries obtained from k-means clustering should be interpreted with caution because the crisp classification of data points that fall between clusters may be misleading. In such environments, fuzzy clustering approaches, which yield additional information about the quality and the internal consistency of the obtained data classification [Paasche et al., 2010, 2012] might be beneficial.

(Hydro-)lithological facies interpretation

In this section, we discuss which type of sediment is related to the different clusters. We first briefly discuss CPT clusters 4-6, which occur in the unsaturated zone and in the

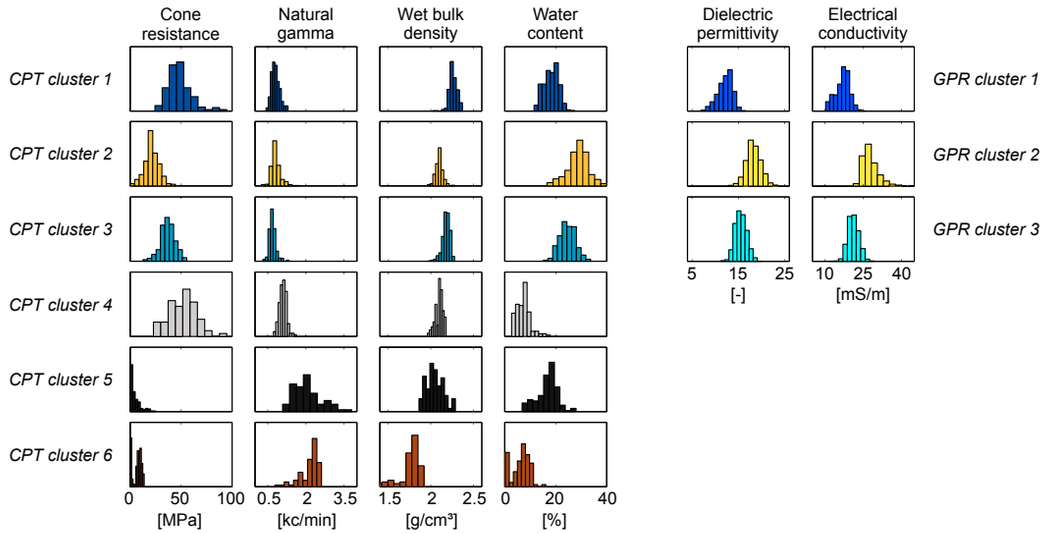


Figure 14: Characteristic histograms for CPT clusters (left) and GPR clusters (right).

deeper part of the aquifer where no co-located GPR data exist. Then we focus on GPR and CPT clusters 1-3, which occur in the central part of the aquifer cross-section where CPT and GPR data overlap.

At the top of each CPT profile, the CPT clusters 5 and 6 occur (Figure 11b). Both clusters are characterized by large natural gamma values and by low cone resistance values (Figure 14). This suggests a soft and clay-rich sediment. Note that CPT cluster 5 occurs again near the bottom of the profiles at approximately 11-12 m depth, where according to Tillmann et al. [2008] the clay layer at the bottom of the uppermost aquifer is located. We interpret CPT clusters 5 and 6 as clay-rich sediments which form the top soil layer and the lower boundary of the uppermost aquifer. The reason why they are not grouped together in just one cluster in the cluster analysis seem to be low bulk density and low water content values in the uppermost dry and less compacted soil decimeters.

In contrast to CPT cluster 5 and 6, the CPT cluster 4 shows large cone resistance values and small natural gamma values (Figure 14). This suggests a hard material such as gravel or gravelly sand with only small amounts of clay. Note that CPT cluster is separated from CPT cluster 1 by a near-horizontal boundary at approximately 2.3 m depth (Figure 11b). The vertical position of that boundary corresponds to the groundwater level during the CPT survey Tillmann et al. [2008]. This indicates that the two clusters could represent the same type of material under saturated and under unsaturated conditions. The CPT data supports this interpretation, because the histograms of cone resistance and gamma activity derived for CPT clusters 1 and 4 look quite similar, while the histograms of water content and wet bulk density deviate for the two clusters (Figure 14).

In the part of the aquifer cross-section where CPT and GPR data overlap, clusters 1-3occur in both data sets. Comparison of the grain size distribution in borehole B32 (Figure 11a) with the spatial distribution of GPR and CPT clusters close to this borehole shows that GPR and CPT cluster 1 occur in those depths where the grain size analysis shows gravel; GPR and CPT cluster 2 occur in those depths where the grain size analysis shows sand; and GPR and CPT cluster 3 occur where the grain size analysis shows an intermediate material such as a gravelly sand or a sandy gravel. This interpretation is supported by the histograms of cone resistance data, which show the largest values for CPT cluster 1, the smallest values for CPT cluster 2, and intermediate values for CPT cluster 3 (Figure 14). The water content histograms, which under saturated conditions can be expected to represent porosity, suggest a relatively small porosity for CPT cluster 1, a relatively high porosity for CPT cluster 2 and an intermediate porosity for CPT cluster 3. As expected, the permittivity histograms for GPR clusters 1-3 show the same trend as the water content histograms for CPT clusters 1-3. Interestingly, also the electrical conductivity histograms show the same trend as the water content histograms. This suggests that the differences in electrical conductivity for GPR clusters 1-3 are mainly caused by differences in porosity. The reason for the correlation between porosity and electrical conductivity is probably that there are only minor changes in clay content in this part of the aquifer, which is supported by the fact that the histograms of gamma activity only slightly deviate for CPT clusters 1-3. With the clay content staying roughly constant, the electrical conductivity derived from GPR is mainly influenced by the electrical conductivity of groundwater, and thus, by changes in sediment porosity.

Note that the uniformity coefficient $U = d_{60}/d_{10}$ (red line in Figure 11a) takes relatively high values in those depths where gravel occurs and relatively low values where sand occurs. This suggests that the higher porosity of the sand facies compared to the gravel facies can be explained by the better sorting of the sand facies compared to the gravel facies.

To evaluate if the facies derived from GPR and CPT provide a meaningful classification with regard to hydraulic conductivity, we compare the facies distribution with vertical profiles of hydraulic conductivity derived from a flowmeter test and from grain size data (Figure 11a). Given the limited amount of hydraulic conductivity data, our goal is not to precisely determine quantitative hydraulic conductivity values for each facies but rather to evaluate if the derived facies show any correspondence to changes in hydraulic conductivity. It should be noted that, in principle, the (direct) flowmeter method can be expected to give more reliable results than the (indirect) grain size method. However, flowmeter results critically depend on proper well construction and well development and can be biased by ambient flow in the borehole [Molz et al., 1994].

Overall, flowmeter and grain size based hydraulic conductivity profiles agree relatively well in the lower part of the aquifer but depart from each other in the upper part of the profile. The grain size based hydraulic conductivity profile shows an obvious correlation with the spatial distribution of GPR and CPT clusters. The lowest values for the hydraulic conductivity are derived between 4 and 6 m depth, which match with the upper and lower boundary of GPR and CPT cluster 2, respectively. The highest values for the hydraulic conductivity are obtained at around 3 m depth where GPR and CPT cluster 1 occur. Intermediate hydraulic conductivity values are obtained from 6 to 9 m depth where GPR and CPT suggest alternating occurrence of clusters 1 and 3.

The flowmeter based hydraulic conductivity profile generally shows a comparable trend but it departs from the grain size based conductivity profile in the upper part of the aquifer. Changes in the flowmeter based hydraulic conductivities are not as clearly associated with the cluster boundaries as it is the case for the grain size based hydraulic conductivities. While the upper boundary of GPR and CPT cluster 2 (at 4 m depth) can be clearly recognized in the flowmeter based conductivity profile, the lower boundary at 6 m depth is only associated with a minor change in hydraulic conductivity. Moreover the profile between those boundaries (4-6 m depth) shows strong inhomogeneities. The distinct peak at 5 m depth represents a particularly high hydraulic conductivity value surrounded by low conductivity values. This suggests a significant heterogeneity within cluster 2. It must be noted, however, that the peak consists of a single value only, and could represent a measurement error. The uppermost layer of GPR and CPT cluster 1 is, unfortunately, only partly covered by the flowmeter data because due to the technical setup, the flowmeter measurements could be conducted only below approximately 3.5 m depth. The hydraulic conductivity of cluster 1 in the uppermost layer can therefore not be evaluated based on the flowmeter data. It can be recognized, however, that the flowmeter based hydraulic conductivity profile shows relatively high values below 6 m depth, where GPR and CPT cluster 1 and 3 occur, and relatively low values between 4 and 6 m depth, where GPR and CPT cluster 2 occur (except for the one value peak at 5 m depth).

Putting all information together, the following interpretation can be derived: GPR and CPT cluster 1 show the distribution of a moderately-poorly sorted gravel facies, which occurs at the top of the GPR cross-section in one continuous layer, and close to the bottom of the GPR cross-section in several isolated patches. Based on the water content values of CPT cluster 1, the gravel facies is characterized by a relatively low average porosity of 18% with a standard deviation of 3%. The CRIM (Eq. (4)) applied to the permittivity data of GPR cluster 1 yields a very similar porosity of $19 \pm 3\%$. Grain size data and flowmeter measurements suggest that the gravel facies is characterized by a relatively high hydraulic conductivity. The gravel facies is underlain by a well-sorted sand facies (GPR and CPT cluster 2) which is characterized by a relatively high porosity of $29 \pm 4\%$ according to CPT water content values and $30 \pm 3\%$ according to GPR permittivity data. Grain size based hydraulic conductivity estimates suggest that the sand facies is characterized by a relatively low hydraulic conductivity. Flowmeter measurements suggest that the hydraulic conductivity of the sand facies is relatively low in average but may include significant inhomogeneities. In the left part of the aquifer cross-section, the sand forms an approximately 1.5 m thick horizontal layer. In the right part of the cross-section, this layer decomposes into disconnected lenses, which are only a few decimeters thick. The sand facies is underlain or surrounded by a sandy gravel or gravelly sand facies (GPR and CPT cluster 3) for which CPT and GPR data give an average porosity of $25 \pm 3\%$ and $26 \pm 2\%$, respectively.

It is interesting to note that in contrast to full-waveform inversion results at other sites, where a positive correlation between porosity and hydraulic conductivity was observed [Klotzsche et al., 2013], here, our results suggests a negative correlation between porosity and hydraulic conductivity. The reason for this is probably that the high-porosity sand facies has a relatively narrow pore structure, whereas the low-porosity gravel facies has larger individual pores which compensate for the lower total porosity and lead

to a higher hydraulic conductivity of the gravel facies compared to the sand facies. This shows that the relationship between porosity and hydraulic conductivity is site dependent, and it shows how important it is to combine complementary investigation tools for aquifer characterization. Based on the GPR inversion results alone, one would probably have assumed (following the findings of Klotzsche et al. [2013]) that the high permittivity zone represents the highest hydraulic conductivity in the aquifer. Instead, flowmeter and grain size data revealed that it is actually the low permittivity zone which can be expected to have the highest hydraulic conductivity at Krauthausen.

Another point we want to note is that the clustered GPR transect shows the spatial connectivity of the individual facies. This makes it possible, for instance, to identify the well connected gravel layer between 3 and 4 m depth or the disconnected sand lenses in the right part of the transect (Figure 11). As mentioned earlier, identifying connected structures in the subsurface and mapping their spatial distribution is highly valuable for aquifer characterization because connected structures can have a substantial impact on flow and transport [Bianchi et al., 2011; Zheng et al., 2011]. Full-waveform inversion of crosshole GPR data followed by cluster analysis seems to provide an applicable approach to characterize the lateral connectivity of subsurface facies in aquifers, which is difficult to achieve with traditional aquifer exploration techniques.

2.4 CONCLUSIONS

Crosshole ground penetrating radar (GPR) data, acquired along a transect of 20 m length and 10 m depth, were used to characterize the heterogeneous architecture of an alluvial aquifer. Full-waveform inversion of the GPR data yielded the spatial distribution of permittivity and electrical conductivity at the scale of decimeters. Comparison of GPR porosity estimates with co-located CPT porosity logs showed good agreement and gave a linear correlation coefficient of 0.8 and a regression line close to the 1:1 line. Comparison of electrical conductivity data derived from GPR and CPT gave a linear correlation coefficient of 0.67 and a regression line with a slope close to 1, but with a distinct nonzero intercept, caused by an offset in the absolute electrical conductivity values derived from GPR and CPT. The offset could be caused by the inherent differences of the two methods or by biased electrical conductivity starting models used in the GPR full-waveform inversion. Irrespective of the discrepancies in absolute values, we found that GPR and CPT gave relatively consistent results regarding the spatial variations in electrical conductivity.

Cluster analysis of the bivariate GPR inversion results defined three different facies in the aquifer cross-section. The number of facies was confirmed by the results of a cluster analysis conducted (independently) with multivariate CPT data. Moreover, the spatial distribution of facies in CPT profiles showed a good match with the spatial distribution of facies in the GPR transect, which suggests that GPR and CPT identified the same facies in the subsurface. Comparison of the facies distribution with co-located profiles from grain size analyses and flowmeter measurements showed that the derived facies boundaries correlate with changes in grain size and porosity, and to a lesser extent with changes in hydraulic conductivity.

In conclusion, our data suggest that full-waveform inversion of crosshole GPR data followed by cluster analysis is an applicable approach to identify hydrogeological facies and to map their detailed spatial distribution. A particular benefit of using GPR tomographic data is that it yields information about the subsurface on full two-dimensional cross-sections, which makes it possible to assess the spatial connectivity of subsurface structures. Such information is highly valuable for constraining the heterogeneity of stochastic models and can be used to construct geologically realistic aquifer models for numerical flow and transport predictions. Finally, our work shows the benefit of using complementary investigation tools for aquifer characterization. While the GPR data yielded detailed maps of the facies distribution in the aquifer, CPT, grain size and flowmeter data provided the basis for a hydrogeological interpretation of the obtained facies maps.

3 THE BENEFIT OF GPR FULL-WAVEFORM INVERSION FOR HYDROGEOLOGICAL SITE CHARACTERIZATION‡

ABSTRACT

Limited knowledge about the spatial distribution of aquifer properties typically constrains our ability to predict subsurface flow and transport. Here, we investigate the value of using high resolution full-waveform inversion of cross-borehole ground penetrating radar (GPR) data for aquifer characterization. By stitching together GPR tomograms from multiple adjacent crosshole planes, we are able to image, with a decimeter scale resolution, the dielectric permittivity and electrical conductivity of an alluvial aquifer along cross-sections of 50 m length and 10 m depth. A logistic regression model is employed to predict the spatial distribution of lithological facies on the basis of the GPR results. Vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data suggest that the GPR predicted facies classification is meaningful with regard to porosity and hydraulic conductivity, even though the distributions of individual facies show some overlap and the absolute hydraulic conductivities from the different methods (direct-push, flowmeter, grain size) differ up to approximately one order of magnitude. Comparison of the GPR predicted facies architecture with tracer test data suggests that the plume splitting observed in a tracer experiment was caused by a hydraulically low-conductive sand layer with a thickness of only a few decimeters. Because this sand layer is identified by GPR full-waveform inversion but not by conventional GPR ray-based inversion we conclude that the improvement in spatial resolution due to full-waveform inversion is crucial to detect small-scale aquifer structures that are highly relevant for solute transport.

3.1 INTRODUCTION

Predicting subsurface flow and transport is challenging due to the complex heterogeneity found in most geologic media. While it has been shown that flow processes can be reasonably well described using spatially averaged (or upscaled) properties, which are easily available from classical hydrogeological methods (e.g. pumping test), this approach fails for the description of transport processes, which inherently depend on the distribution of local properties [Dagan, 1989]. Because the detailed spatial characteriza-

‡ this chapter has been published as: Gueting N., Vienken T., Klotzsche A., van der Kruk J., Vanderborght J., Caers J., Vereecken H., Englert A., (2017), High resolution aquifer characterization using crosshole GPR full-waveform tomography: Comparison with direct-push and tracer test data, *Water Resources Research*, **53**, http://dx.doi.org/10.1002/2016WR019498.

tion of aquifers is unfeasible using classical hydrogeological methods, novel techniques are needed that are able to characterize aquifers with higher spatial resolution.

Recent advances in geophysical measurement and inversion techniques allow imaging the subsurface with unprecedented resolution and coverage [Binley et al., 2015]. A particularly promising method with regard to the detailed characterization of porous aquifers is full-waveform inversion of cross-borehole GPR data [Ernst et al., 2007b; Klotzsche et al., 2014; Gueting et al., 2015]. This method uses a transmitting and a receiving antenna which are placed in two adjacent boreholes, several meters apart from each other. The transmitting antenna emits an electromagnetic wave which propagates through the subsurface and is detected by the receiving antenna. If the wave encounters contrasting electrical properties in the subsurface, it will be reflected, refracted or scattered [Davis and Annan, 1989]. The waveform that arrives at the receiving antenna thus contains information about the medium the wave has passed. By repeating the measurement for multiple transmitter-receiver configurations, and inverting the measured data, two-dimensional tomograms of the subsurface electrical properties (dielectric permittivity and electrical conductivity) can be obtained.

Because the attenuation of GPR waves depends on electrical conductivity [Davis and Annan, 1989], GPR performance is optimal for settings where the subsurface electrical conductivity is relatively low, such as coarse grained porous aquifers with low groundwater salinity. If significantly higher electrical conductivities are encountered (due to clay-rich sediments or high groundwater salinity), the use of GPR for subsurface characterization may be limited because of strong signal attenuation.

Traditional inversion of crosshole GPR data focuses on the reconstruction of first arrival times, and ray-based forward simulators are used to approximate the propagation of radar waves in the ground [e.g. Holliger et al., 2001]. Ray-based models, however, strongly simplify the propagation of electromagnetic waves, which can lead to biased inversion results [Linde and Vrugt, 2013]. In addition, first arrival times constitute only a small portion of the information contained in the measured waveforms. Recently, full-waveform inversion approaches, which are commonly used in seismic inversion, have been introduced to crosshole GPR inversion [Ernst et al., 2007b; Meles et al., 2010]. Full-waveform inversion aims at reconstructing not only first arrival times but entire waveforms, and employs a numerical solution of Maxwell's equations to more accurately simulate the electromagnetic wave propagation. As a consequence, this approach exploits more accurately the information contained in the GPR data and allows characterizing the subsurface with a considerably improved spatial resolution [Belina et al., 2009; Yang et al., 2013; Klotzsche et al., 2013, 2014; Gueting et al., 2015; van der Kruk et al., 2015].

Although the potential benefit of utilizing high resolution geophysical data for hydrogeological site characterization has been demonstrated [e.g. Scheibe and Chien, 2003; Linde et al., 2006; Dafflon et al., 2011], its routine application is hindered by the lack of direct and universal relationships between geophysical and hydrogeological properties. Even though strong and linear relationships seem to exist at some sites, which has motivated direct estimation of hydraulic conductivity from e.g. seismic slowness [Hyndman et al., 2000] or from radar velocity [Hubbard et al., 2001; Linde et al., 2008], relationships at other sites are essentially unknown and may or may not exist [Binley et al., 2015].

Rather than relying on explicit petrophysical relationships, other studies have instead exploited the geometrical information obtained from geophysical data [Hyndman and Gorelick, 1996; Tronicke et al., 2004; Schmelzbach et al., 2011]. The underlying assumption is that geophysical variations delineate lithological zonation, which is in turn related to hydrogeological properties. As no explicit petrophysical relationship is needed, zonation approaches alleviate the difficulty of defining one specific quantitative relationship required for the direct estimation of hydraulic parameters from geophysical data. Hyndman and Gorelick [1996] jointly inverted seismic and tracer data to estimate the geometry and hydraulic conductivity of lithological zones. With the approximation that each zone can be represented by a constant hydraulic conductivity, they obtained a reasonably good match between observed and numerically modelled tracer breakthrough curves. Wainwright et al. [2014] developed a Bayesian hierarchical approach to jointly invert multiscale geophysical data sets and used this approach to estimate reactive facies at a contaminated site by integrating high resolution crosshole and low resolution surface seismic data. In a synthetic study, Linde et al. [2006] used crosshole GPR data to delineate GPR facies and inverted tracer data to infer pixel-wise hydraulic conductivities from radar velocity, allowing a different petrophysical relationship between radar velocity and hydraulic conductivity for each GPR facies. Dogan et al. [2011] and Schmelzbach et al. [2011] employed surface GPR measurements to define subsurface facies and derived individual hydraulic conductivity distributions for each facies from co-located direct-push hydraulic conductivity profiles.

In the present study, we follow a conceptually similar approach and employ crosshole GPR to characterize the heterogeneous architecture of an alluvial aquifer. But in contrast to the aforementioned studies, we analyze the GPR data by full-waveform inversion, which enables us to image the subsurface with increased spatial resolution. Full-waveform inversion of crosshole GPR data has been applied previously to characterize alluvial aquifers [Oberröhrmann et al., 2013; Klotzsche et al., 2013, 2014; Gueting et al., 2015], but these studies considered only relatively small spatial scales (i.e. one or a few crosshole planes). Here, we investigate a considerably larger domain by analyzing a total number of 15 crosshole planes that are connected across an aquifer volume of 50 m $\times$ 25 m $\times$ 10 m.

The main contribution of this study, however, lies in the comparison of the GPR results with hydrogeological data, which provides an excellent basis to assess the actual benefit of GPR full-waveform inversion for hydrogeological site characterization. Thereto, the GPR results are compared with the tracer breakthrough observed in a field experiment and with vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data.

The remaining paper is arranged as follows: In Section 3.2, we describe measurement and inversion of the GPR data, and we describe direct-push injection logs and direct-push slug tests which were conducted to characterize the vertical variations in hydraulic conductivity at selected profiles along the GPR transects. Because the newly conducted GPR and direct-push measurements are compared with data from previous studies at this site, we briefly revisit some of the data retrieved in previous studies to facilitate the later comparisons. In Section 3.3, we first explore the direct correlation between GPR results, porosity and hydraulic conductivity using crossplots and linear regression; then we derive a lithological zonation from the GPR tomograms using logistic regression. The

resulting GPR facies are characterized in terms of porosity and hydraulic conductivity using co-located direct-push, flowmeter and grain size data, and the facies distribution along the GPR transects is compared with the breakthrough of a tracer observed in a field experiment. Section 3.4 summarizes the main findings.

3.2 MATERIAL AND METHODS

3.2.1 Study site

The Krauthausen test site is situated in the Lower Rhine Embayment, approximately halfway between the cities of Cologne and Aachen. It was set up by the research center Jülich in 1993 with the goal to study the in-situ distribution of subsurface properties and their link to groundwater flow and solute transport. A detailed description of the test site and its instrumentation is given by Döring [1997] and Vereecken et al. [2000]. The test site has an extent of 200×50 m with its longitudinal axis approximately aligned to the regional groundwater flow direction (Figure 15). The site is equipped with 76 observation wells reaching to a depth of 10-13 m which corresponds to the base of the uppermost aquifer. The aquifer material is composed of alluvial terrace sediments, deposited by the Rur river, a local braided river system, on top of older Rhine and Maas sediments [Englert, 2003]. The aquifer is characterized by sandy to gravely grain size with only small amounts of clay. The mean aquifer porosity, based on laboratory analysis of core samples is 26% [Vereecken et al., 2000]. The averaged hydraulic conductivity derived from a large scale pumping test is 3×10^{-3} m/s [Englert, 2003].

Over the last decades, the Krauthausen site has been intensively investigated using a broad spectrum of methods, including tracer experiments [Vereecken et al., 2000; Vanderborght and Vereecken, 2001; Kemna et al., 2002; Müller et al., 2010], flowmeter and pumping tests [Li et al., 2007, 2008], geophysical imaging methods [Kemna et al., 2002; Müller et al., 2010; Gueting et al., 2015], cone penetration tests [Tillmann et al., 2008] and laboratory characterization of core samples [Döring, 1997]. These studies have brought together a rich set of information, which provides excellent conditions to test and validate novel methods.

3.2.2 Crosshole ground penetrating radar (GPR) tomography

Data acquisition

GPR data were acquired along 15 crosshole planes with individual borehole distances from 3 to 12 m (Figure 15). Crosshole planes with borehole distances smaller than 7 m were measured with 200 MHz antennas, crosshole planes with borehole distances larger than 7 m were measured with 100 MHz antennas because the signal detected with 200 MHz antennas was too weak. To obtain sufficient ray coverage at affordable acquisition times, measurements were carried out in a semi-reciprocal acquisition setup with transmitter and receiver spacing of 0.5 m and 0.1 m, respectively [Oberröhrmann et al., 2013]. Measurements were conducted from approximately 3 m depth (about 1 m below the groundwater level during the measurements) down to approximately

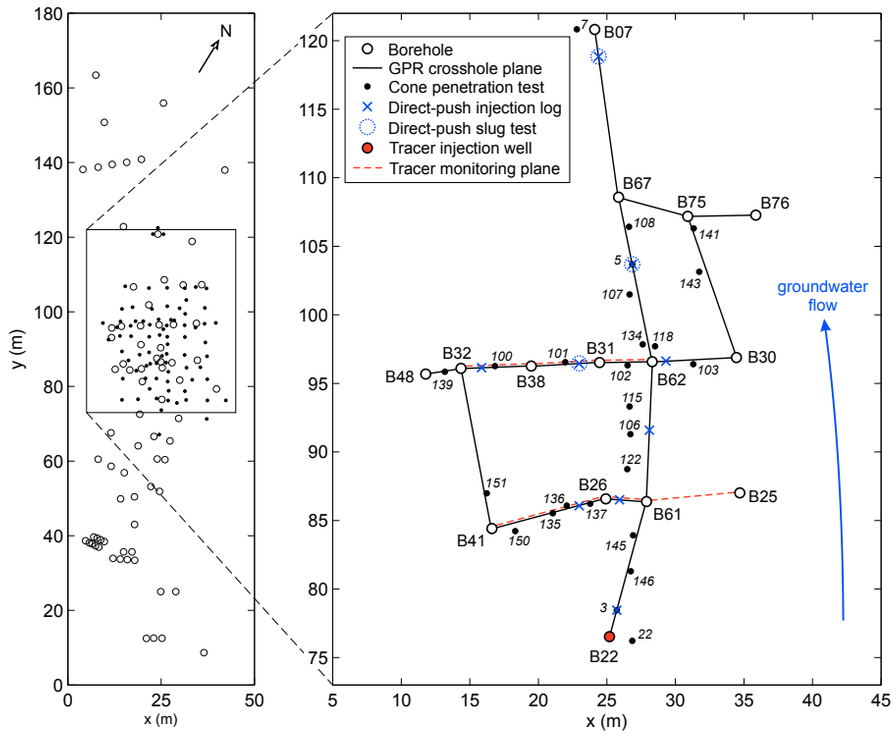


Figure 15: Map of the Krauthausen test site. For clarity, the close-up (right) shows only those boreholes and cone penetration tests that are located along the GPR transects.

10 m depth, depending on the depth of the boreholes. To avoid inversion artifacts due to imprecisely known transmitter and receiver positions [Linde et al., 2006], borehole deviations from the vertical were accurately determined by deviation log, prior to the GPR measurements.

Full-waveform inversion (FWI)

A full-waveform inversion approach was applied to infer the relative dielectric permittivity, ϵ, and the electrical conductivity, σ, from the measured GPR data. A brief description of the inversion scheme is given below, for a more detailed description of the inversion procedure and its implementation the reader is referred to Meles et al. [2010] and Klotzsche et al. [2010].

We used the inversion scheme introduced by Ernst et al. [2007b] and modified by Meles et al. [2010]. This approach employs a conjugate gradient method to iteratively minimize the misfit between observed and model-predicted radar traces, the latter being generated by a forward model based on a 2D finite difference time domain (FDTD) solution of Maxwell's equations. After each iteration, the parametrization of the forward model (i.e. the spatial distribution of ϵ and σ) is updated using local gradients de-

rived from a cross-correlation of the back-propagated residual wavefield and the model-predicted wavefield [Meles et al., 2010]. To avoid overfitting, the inversion is stopped when the root-mean squared (RMS) error between observed and model-predicted radar traces changes less than 1% from one iteration to the next.

An important requirement for the full-waveform inversion are adequate starting models for ϵ and σ. In particular, the ϵ starting model is of critical importance because proper convergence of the full-waveform inversion requires that for each transmitter-receiver combination the initial offset between observed and modeled wavefield is not greater than half the dominant wavelength, otherwise the inversion will converge to a local minimum [Klotzsche et al., 2014]. To derive adequate ϵ starting models, we conducted a ray-based traveltime inversion [Holliger et al., 2001] for each crosshole plane, and used the result of the ray-based inversion as starting model for the full-waveform inversion. For σ, we used one homogenous starting model ($\sigma = 13$ mS/m) for all crosshole planes. In principle, individual σ starting models for each crosshole plane could be obtained from ray-based amplitude inversion [Maurer and Musil, 2004]. However, we found the ray-based σ models obtained for a series of adjacent crosshole planes to show large discrepancies in absolute values, which is unlikely for adjacent planes measured on consecutive days on the same site. In fact, it has been shown that the estimation of σ using ray-based methods is not always straight-forward and, in particular, absolute σ-values can be biased [Holliger et al., 2001; Maurer and Musil, 2004]. We therefore took a different approach and tested, for a few crosshole planes, the convergence of the full-waveform inversion for a broad range of homogeneous σ starting models. The model which showed the best convergence was selected as starting model and used for all crosshole planes. Similar approaches have been applied by Klotzsche et al. [2013] and Keskinen et al. [2017].

To minimize inversion artifacts in the vicinity of the boreholes, we adopted the approach by Kurzmann et al. [2013] and implemented a gradient preconditioner that reduces the high sensitivity of the inversion close to transmitter and receiver positions. Details on the implementation are given in van der Kruk et al. [2015].

3.2.3 Complex refractive index model (CRIM)

Petrophysical models can be applied to translate the electrical properties obtained from GPR into hydrologically more relevant properties such as soil moisture or porosity. Among the various empirical, semi-empirical and theoretical models that have been proposed in the literature [e.g. Topp et al., 1980; Dobson et al., 1985; Steelman and Endres, 2011], we opted for the CRIM relationship [Birchak et al., 1974], which is based on the intuitive idea that the bulk dielectric permittivity, ϵ_b, of an n-phase medium can be expressed as the sum of the dielectric permittivities of the individual phases, ϵ_i, weighted by their volume fraction, V_i,

$$\epsilon_b^\alpha = \sum_{i=1}^{n} V_i \epsilon_i^\alpha. \tag{9}$$

The exponent α describes the spatial arrangement of the individual phases and takes a value of 0.5 for an isotropic phase arrangement [Birchak et al., 1974]. Although the CRIM does not account for the full complexity of natural porous media [Dobson et al., 1985; Brovelli and Cassiani, 2008], the CRIM has been found to yield robust porosity estimates for various types of sediments [Roth et al., 1990]. For a two-phase fully saturated porous medium with isotropic phase arrangement, Eq. (9) can be written as

$$\phi = \frac{\sqrt{\epsilon_b} - \sqrt{\epsilon_s}}{\sqrt{\epsilon_f} - \sqrt{\epsilon_s}}, \tag{10}$$

where ϕ is the porosity, ϵ_b is the bulk dielectric permittivity (obtained e.g. from GPR), and ϵ_s and ϵ_f are the dielectric permittivities of the solid grain matrix and of the pore fluid, respectively.

3.2.4 Logistic regression

Just like linear regression, logistic regression investigates the dependency between a response variable and one or more explanatory variables. But while linear regression applies to continuous response variables, logistic regression applies to binary or multinomial response variables. Given this difference, logistic regression may be able to identify relations between variables when linear regression does not. For example, Tesoriero et al. [1998] showed that, while it was not feasible to use linear regression to predict quantitative values for nitrate concentrations in groundwater from well depth data, it was still possible to use logistic regression to predict the probability of the concentration exceeding a critical threshold.

A brief description of the principle of logistic regression is given below, detailed discussions can be found in Hosmer and Lemeshow [2000] or in Helsel and Hirsch [2002]. In the field of geosciences, logistic regression has been used to predict landslide susceptibility on the basis of geomorphological parameters [Ayalew and Yamagishi, 2005; Van Den Eeckhaut et al., 2006], and groundwater vulnerability on the basis of land use data [Eckhardt and Stackelberg, 1995; Nolan, 2001]. The basic assumption of logistic regression is that the logit of the probability of an observation being in a response category can be modeled using a linear combination of explanatory variables

$$\text{logit}(\tau) = \ln\left(\frac{\tau}{1-\tau}\right) = \beta_0 + \beta X, \tag{11}$$

where τ is the probability, X is the vector of explanatory variables, and β_0 and β are the intercept and the vector of regression coefficients, respectively. For a multinomial response function with k categories and n explanatory variables, Eq. (11) becomes a set of k-1 equations

$$\ln\left(\frac{\tau_j}{\tau_k}\right) = \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + ... + \beta_{jn}X_n, \quad j = 1, ..., k-1, \tag{12}$$

where τ_j/τ_k represents the relative probability of belonging to category j versus belonging to the reference category k (here by definition the last category). The regres-

sion coefficients β can be determined using maximum likelihood estimation [Helsel and Hirsch, 2002]. Once β is determined, the probabilities of an observation belonging to each category can be calculated

$$\tau_j = \frac{e^{\beta_{j0}+\beta_{j1}X_1+\beta_{j2}X_2+...+\beta_{jn}X_n}}{1+\sum_{j=1}^{k-1} e^{\beta_{j0}+\beta_{j1}X_1+\beta_{j2}X_2+...+\beta_{jn}X_n}}, \quad j = 1, ..., k-1, \tag{13}$$

$$\tau_k = 1 - \sum_{j=1}^{k-1} \tau_j. \tag{14}$$

The individual model components (explanatory variables) can be tested for statistical significance using t-statistics, which are obtained by dividing each component of β by its standard error. To identify the 'best' model among a set of models with different (combinations of) explanatory variables, Akaike's Information Criterion (AIC) can be applied [Helsel and Hirsch, 2002]. The AIC calculates as $AIC = 2p - 2L$, with p the number of estimated model parameters and L the log-likelihood of the fitted model [Akaike, 1974]. Among a set of different models, the model that minimizes the AIC is the preferred model. The AIC, thus, rewards a high likelihood and penalizes a high number of model parameters.

In the present study, we applied a logistic regression of GPR results (explanatory variables) and lithological facies derived from grain size data (response categories). GPR and grain size data were available at co-located locations along 3 boreholes. The basic idea is to use the co-located data to derive an empirical relationship (i.e. the fitted logistic regression model), and utilize this relationship to predict the distribution of grain size facies at locations where no grain size data exist, solely on the basis of GPR. The following workflow was applied: prior to the logistic regression, the grain size data were classified into lithological facies using simple k-means clustering [MacQueen, 1967]. The idea behind this is to transform the multivariate grain size data into a categorical variable, which can be used in logistic regression. The logistic regression was conducted using the clustered grain size data as categorical response variable and the bivariate GPR results (dielectric permittivity and electrical conductivity) as explanatory variables. Maximum likelihood estimates for β were obtained using an iteratively weighted least squares algorithm. With β determined, Eqs. (13) and (14) were applied to calculate, for each GPR pixel, the probability to belong to each grain size facies.

To evaluate the benefit of GPR full-waveform inversion instead of classical GPR ray-based inversion, we compared the outcome of the logistic regression conducted with the full-waveform inversion results, and with the ray-based inversion results as explanatory variables.

3.2.5 Direct-push hydraulic measurements

Direct-push injection log (DPIL)

The DPIL method [Dietrich et al., 2008] allows to assess local variations in hydraulic conductivity. The method consists of advancing a small diameter direct-push rod, equipped with a filter screen, into the subsurface. At selected depths, the pushing is stopped, wa-

ter is injected through the screen, and the injection flow rate and the injection pressure are recorded. The ratio of injection rate and pressure, corrected for head loss in the tube, gives a proxy, K_{DPIL}, for the aquifer's local hydraulic conductivity [Dietrich et al., 2008]. Through calibration with absolute hydraulic conductivity data from direct-push slug tests (or any other method), K_{DPIL} can subsequently be converted into absolute hydraulic conductivities [Dietrich et al., 2008; Lessoff et al., 2010; Dietze and Dietrich, 2012]. At Krauthausen, DPIL profiles were measured at 9 locations (Figure 15). Since the focus is on a comparison with the GPR results, the DPIL measurements were conducted at locations where crosshole GPR results were available. DPIL measurements were conducted from 3 to 9 m depth with a vertical sampling interval of 20 cm. At each depth, 3 repeat measurements with different flow rates were conducted for quality control (e.g. to identify potential effects caused by opening of preferential flowpaths due to the water injection or water flow along the probing rods). During the advancement of the DPIL probe from one sampling depth to the next, water was continuously injected in order to prevent clogging of the screen.

Direct-push slug test (DPST)

In order to obtain absolute values of hydraulic conductivity, multi-level DPST [Butler et al., 2002] were performed at 3 profile locations (Figure 15). Because of the multitude of repeat measurements at each measurement interval that were performed for quality control (see Butler [1997]), the DPST measurements were relatively time consuming. Therefore, only 2 or 3 depth intervals were tested in each profile, resulting in a total number of 7 individual DPST. To enable calibration of the DPIL data, the DPST were carried out in close proximity (30 cm lateral distance) to DPIL locations; and depth intervals with fairly constant DPIL ratios were selected. Slug tests were performed in temporarily installed direct-push monitoring wells with a screen length of 1 m. During well installation, the screen was protected inside the rods to prevent clogging of the screen. At the selected depths, the screen was exposed and water was pumped until the pumped water was free of fines. The slug tests were initiated pneumatically [Butler et al., 2002], by pressurizing the air column above the water to create the initial head displacement and rapidly depressurizing the air column to initiate the tests. The recorded water table responses were analyzed using the approach of Butler et al. [2003], which is based on matching aquifer response type-curves to the measured head recovery. The obtained absolute hydraulic conductivities, K_{DPST}, were subsequently used as calibration data to translate the relative DPIL ratios, K_{DPIL}, into absolute hydraulic conductivities.

3.2.6 Previous studies at the Krauthausen site

To facilitate the later comparisons, we briefly revisit some of the data from previous studies that will be compared with the newly conducted crosshole GPR and direct-push hydraulic measurements.

Cone penetration tests (CPT)

More than 70 cone penetration tests have been conducted at Krauthausen and are described in detail by Tillmann et al. [2008]. At each CPT location, measurements of cone resistance and electrical resistivity were recorded while the CPT probe was continuously pushed into the ground. Subsequently, additional measurements of natural gamma, gamma-gamma and neutron activity were conducted. By calibration, gamma-gamma and neutron log were related to bulk density and water content, respectively [Tillmann et al., 2008]. Recently, Gueting et al. [2015] reviewed this calibration by comparing the water contents derived from CPT neutron log with the saturated porosity of core samples determined in the laboratory. They found that the CPT water contents overestimated the laboratory measurements by a constant value and applied a simple constant-shift correction to compensate for the overestimation. In the present study, we used the corrected CPT water contents, the electrical resistivity and the mechanic cone resistance for comparison with the GPR results. We focus only on those CPT that are located close to the GPR transects; the locations and ID numbers of the CPT considered in the present study are shown in Figure 15.

Hydraulic conductivity from flowmeter and grain size

Previous studies at Krauthausen determined local variations in hydraulic conductivity from flowmeter measurements and grain size [Vereecken et al., 2000; Li et al., 2008; Tillmann et al., 2008]. The flowmeter method consists of measuring, as function of depth, the vertical flow in a screened well subject to a constant pumping rate. From the increase in cumulative flow rate, ΔQ_i, along some depth interval, z_i, the hydraulic conductivity, K_i, of that interval can be estimated [Molz and Morin, 1989]

$$K_i = \frac{\Delta Q_i}{Q} \frac{b}{z_i} \overline{K}, \tag{15}$$

where Q is the total pumping rate, b is the screened thickness of the aquifer and $\overline{K}$ is the depth-averaged hydraulic conductivity obtained from a pumping test carried out in the same well.

Grain size based estimation of hydraulic conductivity is an indirect method (i.e. no flow process is measured) that relies on empirical or semi-empirical relationships between hydraulic conductivity and grain size distribution. Various relationships have been proposed in the literature (see e.g. Vienken and Dietrich [2011]). In the present study, we follow Tillmann et al. [2008], who employed the empirical formula of Bialas and Kleczkowski [1970] to derive hydraulic conductivity estimates for the Krauthausen sediments and found the obtained estimates to be of the same order of magnitude compared with the hydraulic conductivity derived from a pumping test. The Bialas and Kleczkowski [1970] formula is given by

$$K = 0.0036\, d_{20}^{2.3}, \tag{16}$$

with K the hydraulic conductivity in (m/s) and d_{20} the 20th percentile of the sediment particle size in (mm). Due to the empirical nature of this relationship the units are not balanced in Eq. (16).

Tracer experiments

Two consecutive tracer tests were carried out in the years 2002 and 2003 at the Krauthausen site [Müller et al., 2010]. In both experiments, tracer was injected into borehole B22 and tracer breakthrough was monitored along two downgradient detection planes using time-lapse electrical resistivity tomography (ERT) and multilevel groundwater samplers. The overall setup is shown in Figure 15. ERT electrodes were installed in boreholes B41, B26, B61 and B25 (first ERT transect), and in boreholes B48, B32, B31 and B62 (second ERT transect). Additional electrodes were placed at the surface [Müller et al., 2010]. In the first experiment, a 'positive' tracer with a higher electrical conductivity than the ambient groundwater was applied. In the second experiment, a 'negative' tracer with a lower electrical conductivity than the ambient groundwater was applied. The idea behind this was that if the two experiments yield consistent results, density effects related to the tracer application can be neglected. A consistent observation in both experiments was a splitting of the tracer plume in the first ERT transect (Figure 16, top). This splitting was also observed in the second ERT transect for the positive tracer experiment, but not for the negative tracer experiment (Figure 16, bottom). The reason why the negative tracer experiment did not show a plume splitting in the second ERT transect could be a smaller signal-to-noise ratio in the negative tracer experiment

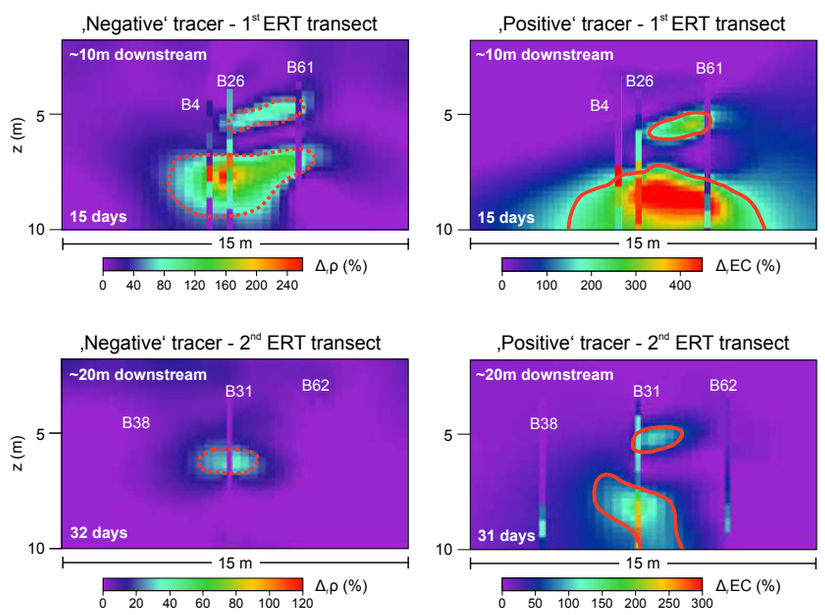


Figure 16: ERT observed tracer plume splitting (modified from Müller et al. [2010]). Colorbars refer to changes in bulk resistivity (left) and bulk electrical conductivity (right), related to the breakthrough of a resistive (negative) and a conductive (positive) tracer. Tracer plumes are outlined by the red dashed and solid lines. Vertical bars show results from multilevel groundwater samplers.

which possibly inhibited seeing detailed breakthrough patterns in the ERT images. For a detailed discussion of the results, the reader is referred to Müller et al. [2010].

In the present study, we compared the ERT observed tracer breakthrough with crosshole GPR data acquired along the same transects, in order to evaluate if the GPR tomograms give evidence about subsurface structures that caused the splitting of the tracer plume. Moreover, we derived estimates of hydraulic conductivity from the tracer arrival times in order to compare them with the hydraulic conductivity derived from direct-push, flowmeter and grain size data. Thereto, hydraulic conductivity estimates were derived from tracer velocities

$$K = \frac{v\,n}{i}, \tag{17}$$

with v the pixel-wise equivalent (streamtube) tracer velocities, that were determined by Müller et al. [2010] by fitting the solution of the 1D convection-dispersion equation to locally measured breakthrough curves (i.e. to the tracer breakthrough detected within a pixel of the ERT cross-sections). n and i are the aquifer porosity and the hydraulic gradient, respectively.

3.3 RESULTS AND DISCUSSION

3.3.1 GPR full-waveform inversion results

A full-waveform inversion was applied to infer the subsurface dielectric permittivity and electrical conductivity for a total number of 15 individual GPR crosshole planes. Depending on the depths of the boreholes, a single crosshole data set comprised approximately 1500 to 2000 individual radar traces. Each crosshole data set was inverted separately, i.e. independent from adjacent crosshole planes. A grid cell size of 3 cm $\times$ 3 cm was used for the forward modeling to ensure numerical accuracy [Holliger and Bergmann, 2002]. For the inversion, 9 forward modeling cells (3×3) were combined into one inversion cell [Ernst et al., 2007b]. Convergence of the full-waveform inversion was typically reached after 15 to 20 iterations. For each crosshole plane the full-waveform inversion lead to a reduction of the initial RMS misfit (obtained with the ray-based permittivity starting model and the uniform conductivity starting model) by about 50 to 70%.

Figure 17 shows a 3D plot of the full-waveform inversion results for all crosshole planes. Close-ups of ray-based and full-waveform dielectric permittivity along the main 2D transects are shown in Figure 18. Compared with the ray-based results the full-waveform tomograms contain spatial structures at greater detail. This is most notable in the longitudinal transect (Figure 18a, b), where the full-waveform inversion identifies several thin layers with high permittivity, that are not resolved by ray-based inversion.

An important finding is that the tomograms in Figure 17 and Figure 18 show relatively consistent spatial structures across boreholes, which indicates robust inversion results. For some boreholes, however, the tomograms left and right of the boreholes depart and/or the spatial structures in the tomograms can be observed to "fade out" close to the boreholes. This effect can most clearly be seen in Figure 18 along the boreholes B62, B61 and B38, both for ray-based and full-waveform inversion results. In the follow-

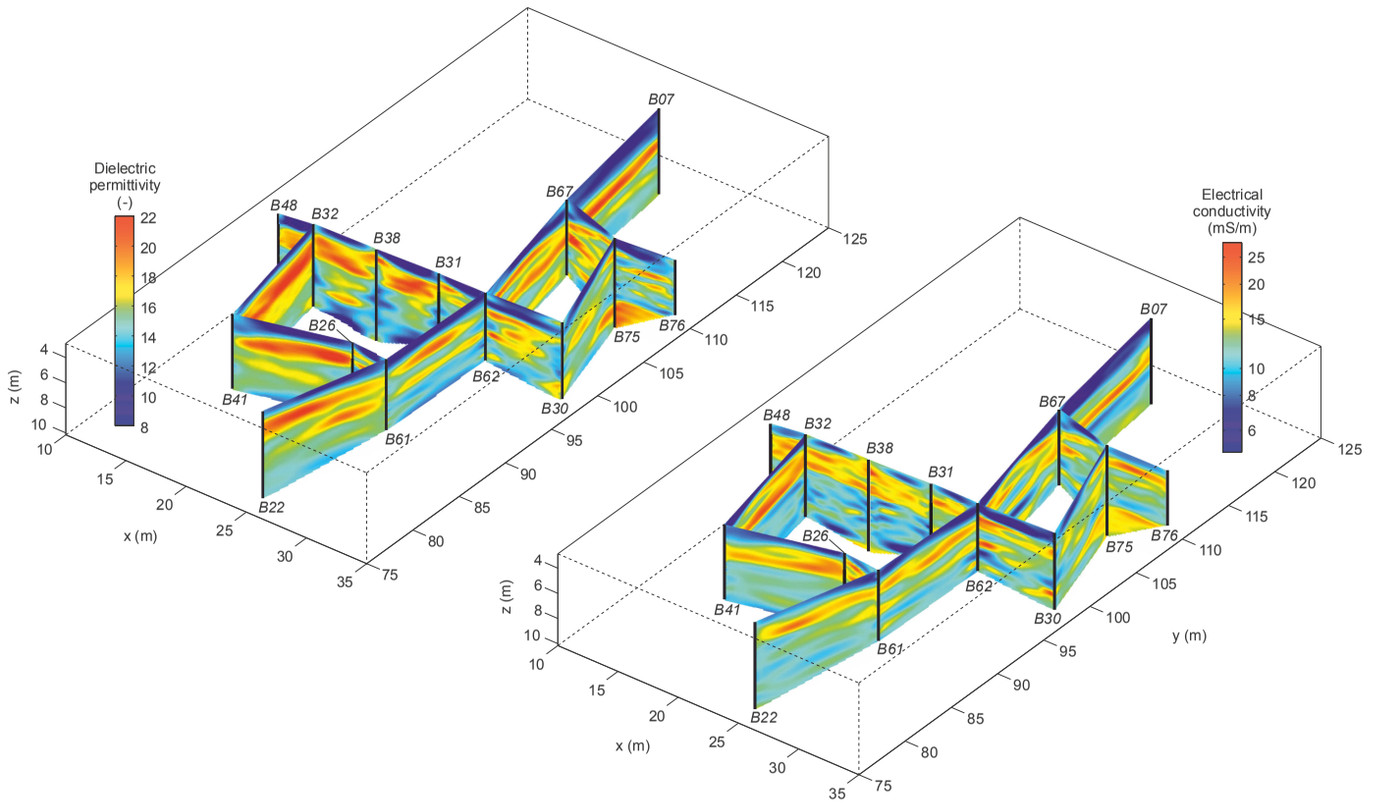


Figure 17: 3D plot of dielectric permittivity (left) and electrical conductivity (right) derived from full-waveform inversion of crosshole GPR data. Black vertical lines represent boreholes.

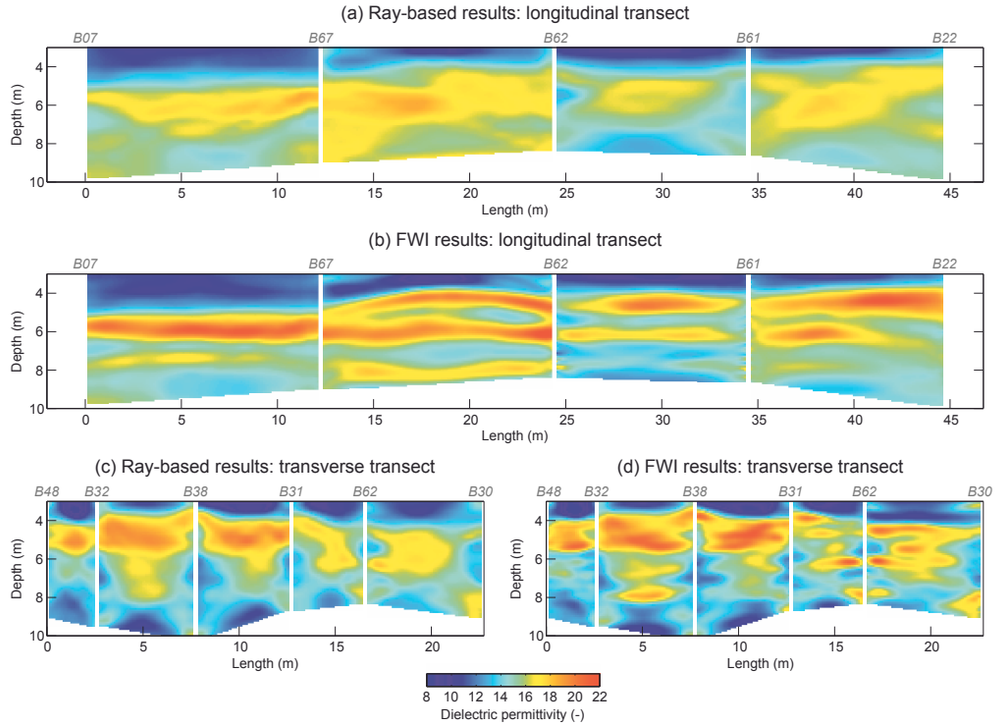


Figure 18: Comparison of ray-based (a, c) and FWI derived (b, d) dielectric permittivity along the two main transects (from borehole B07 to B22 and from B48 to B30; see Figure 15 for reference).

ing, we will briefly discuss potential causes for these inconsistencies as well as possible strategies against it. In fact, inversion artifacts close to boreholes are an issue commonly encountered in crosshole GPR (and seismic) tomography [e.g. Maurer and Green, 1997; Becht et al., 2004; Irving et al., 2007]. A well-known source of such artifacts are imprecisely known transmitter and receiver positions within the boreholes [Linde et al., 2006]. Therefore, it is common practice to perform borehole deviation log measurements prior to the crosshole measurements [e.g. Doetsch et al., 2010; Dafflon et al., 2011]. This is also what was done in the present study. However, Maurer and Green [1997] noted that measured borehole deviations can be flawed, and demonstrated that relatively minor coordinate errors can produce considerable inversion artifacts. As a strategy to account for potential errors in the measured coordinates they proposed to treat transmitter and receiver coordinates as uncertain and include them as additional inversion parameters. Applied to seismic data from the Grimsel site their approach was able to strongly reduce the inconsistencies between two adjacent tomograms.

Dafflon et al. [2011] analyzed radar traveltimes from a total number of 31 crosshole planes at the Boise Hydrogeophysical Research Site and showed that the consistency of intersecting tomograms could be greatly improved by simultaneously inverting all

crosshole data sets together instead of separately. Unfortunately, they also observed that by inverting all data sets jointly, a significantly smoother subsurface model was obtained (i.e. a model containing less details) than when each data set was inverted separately. As our primary goal is to image subsurface structures with the highest possible resolution we decided to invert each crosshole data set separately. An unwanted side effect of that might be that the obtained tomograms are not completely consistent with each other.

It should also be mentioned that the drilling of the boreholes may have locally disturbed the soil structure. In that case, discontinuities in the GPR tomograms would reflect actual zones of disturbance rather than inversion artifacts.

3.3.2 Comparison of GPR full-waveform inversion results with CPT neutron and electrical resistivity logs

As several CPT profiles are located in close proximity to the GPR transects (Figure 15), they provide an opportunity to validate the GPR results. It should be noted, however, that also the CPT profiles are not expected to be flawless, and thus, discrepancies between GPR and CPT could result from errors in either or both techniques. In the following, we compare the porosity from CPT neutron logs with porosity estimates derived from GPR; and we compare the electrical conductivity from CPT resistivity logs with the electrical conductivity derived from GPR. Altogether, 17 CPT porosity profiles and 9 CPT electrical conductivity profiles were available at locations co-located with the GPR transects. To obtain porosity estimates from GPR, the CRIM relation (Eq. (10)) was applied with ϵ_f=84, which represents the dielectric permittivity of water at 10 °C [Eisenberg and Kauzmann, 1969], and with ϵ_s=4.5, which represents the dielectric per-

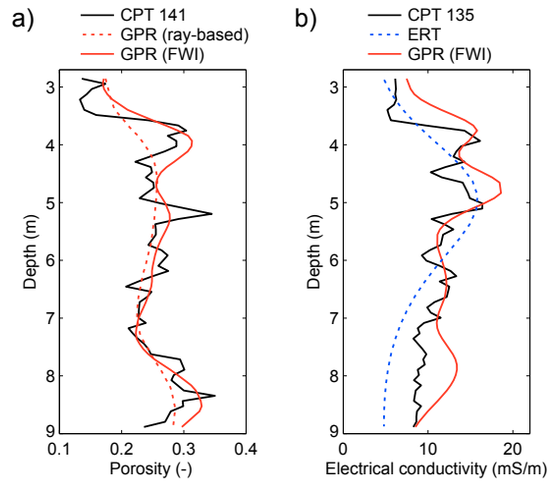


Figure 19: a) CPT porosity profile and comparison with co-located porosity estimated from GPR ray-based inversion and from GPR full-waveform inversion. b) CPT electrical conductivity profile and comparison with co-located electrical conductivity estimated from ERT and from GPR full-waveform inversion.

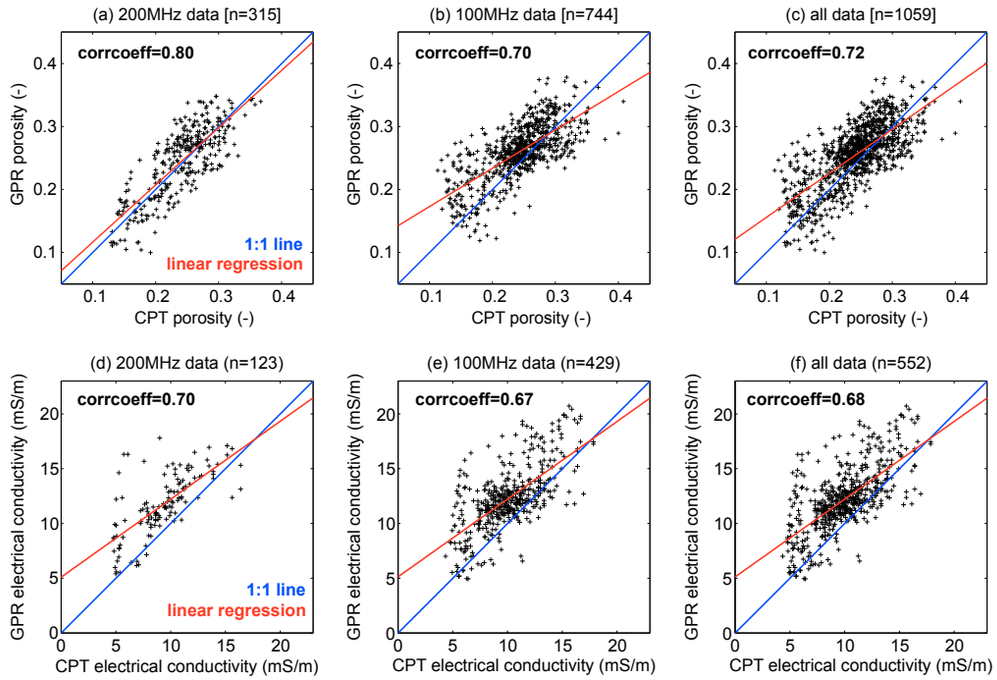


Figure 20: a-c) Crossplots of porosity from CPT and GPR FWI; d-f) Crossplots of electrical conductivity from CPT and GPR FWI. Data from crosshole planes measured with different GPR antenna frequencies are plotted separately and combined (a,d: 200 MHz; b,e: 100 MHz, c,f: together). n is the number of data points.

mittivity of quartz [Carmichael, 1982], i.e. no fitting was performed. Figure 19a shows a comparison of CPT and GPR derived porosity along CPT profile 141. Overall, the porosity from CPT and GPR shows a strong correlation. Compared with the ray-based inversion results the full-waveform inversion improves the reconstruction of small scale structures, such as the high porosity layer at about 4 m depth. The ray-based results give a good estimate of the basic trends in porosity, though.

Crossplots of CPT porosity versus GPR FWI porosity along all CPT profiles are shown in Figure 20a-c. Note that the data from crosshole planes measured with different GPR antenna frequencies are plotted both separately (Figure 20a,b) and combined (Figure 20c). As expected, the agreement between GPR and CPT is better (higher correlation coefficient, regression line closer to the 1:1 line) for the GPR data acquired with the 200 MHz antennas (Fig 20a), than for the GPR data acquired with the 100 MHz antennas (Fig 20b). This can be attributed to the higher spatial resolution that can be obtained from shorter wavelengths.

Figure 20d-f show crossplots of electrical conductivity from CPT and GPR FWI. Although the correlation is not as strong as it is for the porosity, GPR and CPT results still correlate fairly well. For the electrical conductivity data, a better correlation of the 200 MHz data compared to the 100 MHz data cannot be recognized, but this could also

be due to the limited amount of data (due to technical problems during the CPT survey, the electrical conductivity was measured only at approximately every second CPT profile). Generally, it can be observed that the majority of data points plot above the 1:1 line in Figure 20d-f, which indicates that the electrical conductivity obtained from GPR is slightly higher than the electrical conductivity obtained from CPT. Figure 19b shows a comparison of CPT and GPR derived electrical conductivity along CPT profile 135. Along this profile the CPT and GPR results can additionally be compared with the ERT results from Müller et al. [2010]. Overall, the three different methods identify fairly similar trends. However, some notable differences exist, particularly in the lower part of the profile. It is very interesting to see how the small scale variability is most pronounced in the CPT profile, less pronounced in the GPR profile, and virtually absent in the ERT profile, which highlights the differences in spatial resolution of the individual methods.

3.3.3 Comparison of GPR full-waveform inversion results with direct-push hydraulic conductivity

Prior to comparing the direct-push derived hydraulic conductivity with the GPR full-waveform inversion results, we converted the relative K_{DPIL} values into absolute hydraulic conductivity using the few K_{DPST} values as calibration data. Following Dietrich et al. [2008], we conducted a linear regression of co-located log K_{DPIL} and log K_{DPST} values and used the obtained regression equation to translate K_{DPIL} values into hydraulic conductivity (Figure 21).

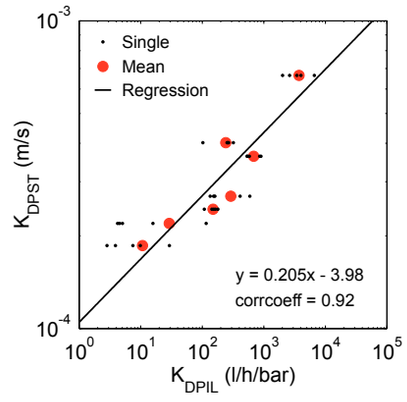


Figure 21: Linear regression of log K_{DPST} and log K_{DPIL}. The equation of the regression line and the correlation coefficient are given at the bottom of the crossplot.

Because the DPST were conducted with a filter screen length of 1 m, and the DPIL were measured with a vertical sampling interval of 20 cm, there are five DPIL measurements co-located with each DPST measurement. Assuming horizontal layering, we calculated an averaged (or upscaled) K_{DPIL} value for each DPST interval, by taking the arithmetic mean of the five associated K_{DPIL} values. The regression was conducted with

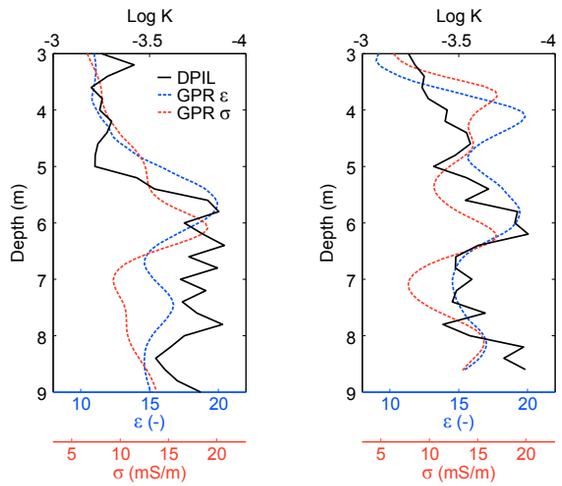


Figure 22: DPIL derived hydraulic conductivity (note the reversed axis for log K) and co-located dielectric permittivity and electrical conductivity from GPR full-waveform inversion.

the arithmetic means of the K_{DPIL} values and the corresponding K_{DPST} values. The high correlation coefficient of 0.92 suggests that the regression equation allows robust estimation of absolute hydraulic conductivity from K_{DPIL}.

The absolute hydraulic conductivity derived from DPIL and DPST were used to investigate if a correlation exists between hydraulic conductivity and GPR attributes. Figure 22 shows vertical profiles of hydraulic conductivity and GPR FWI results along DPIL profiles 2 and 5. Scatterplots of hydraulic conductivity (entire DPIL and DPST data set) versus GPR attributes are presented in Figure 23a, b. The scatterplots were constructed by averaging the values from GPR cells located along the DPST and the DPIL depth intervals, respectively. To reduce noise, a three-point moving average was applied to the

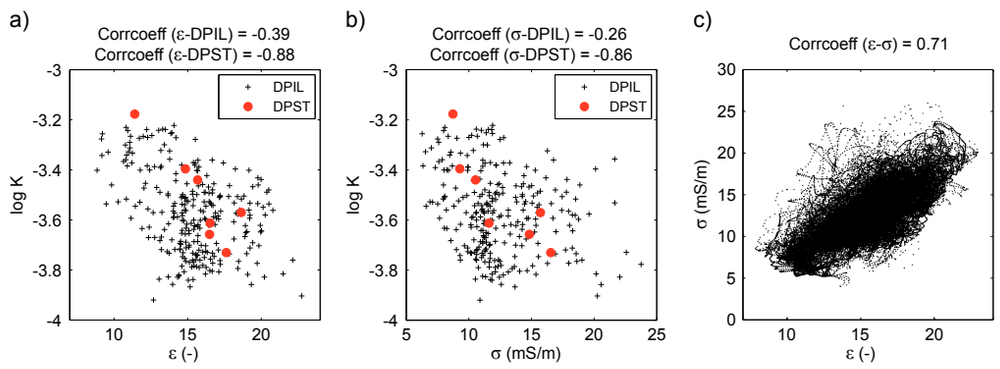


Figure 23: Scatterplots of direct-push hydraulic conductivity and GPR FWI attributes.

DPIL data. Overall, both GPR attributes show a negative correlation with hydraulic conductivity (note that in order to make it easier to see the negative correlation in Figure 22 the log K axis was reversed). The negative correlation is stronger for the (few) DPST hydraulic conductivity estimates than for the DPIL hydraulic conductivity estimates (Figure 23a, b).

Interestingly, the scatterplots for the two GPR attributes versus hydraulic conductivity look quite similar. The reason for this similarity is probably that the clay content at the Krauthausen site is so small that both the dielectric permittivity and the electrical conductivity primarily depend on porosity. According to grain size analyses, the average clay content of the part of the aquifer covered by GPR measurements is only 2%. If changes in clay content are minor, the bulk electrical conductivity of saturated media is controlled by the electrical conductivity of the pore fluid. Under such conditions, both the dielectric permittivity and the electrical conductivity obtained from GPR can be expected to depend mainly on saturated porosity, which is also reflected in the strong correlation of the two GPR attributes (Figure 23c). It is interesting that our data suggests a negative correlation between porosity and hydraulic conductivity, although intuitively one would expect a positive relationship. However, several studies showed that the relation between porosity and hydraulic conductivity is strongly site-dependent and a correlation between the two parameters can be positive, negative or absent [Nelson, 1994; Hubbard et al., 2001; Morin, 2006]. The reason for this is that the hydraulic conductivity is not a function of the total pore volume alone but also of the structure, size, tortuosity and connectivity of the individual pores within the pore network [Koltermann and Gorelick, 1996]. At the Krauthausen site, it appears that the structural features of the pore network that lead to a higher hydraulic conductivity are negatively correlated to the total pore volume. According to Tillmann et al. [2008] and Gueting et al. [2015], the Krauthausen aquifer comprises different lithological facies that are characterized by different average grain sizes (ranging from sand to gravel). In addition, these facies are characterized by different sorting, namely the sand-dominated facies is significantly better sorted than the gravel-dominated facies. This leads to a smaller total porosity of the gravel facies compared to the sand facies. Despite its smaller total porosity, however, we suspect that the gravel is characterized by (fewer but) larger individual pores which compensate for its smaller total porosity and lead to a higher hydraulic conductivity of the gravel facies compared to the sand facies. We believe that this is the reason why we observe a negative correlation between hydraulic conductivity and GPR attributes.

Yet despite this overall negative correlation, the relationship between DPIL derived hydraulic conductivity and GPR results is characterized by considerable scatter (Figure 23a, b). There are several possible explanations for that: 1) The small-scale variations in DPIL and/or GPR data are measurement scatter. 2) The GPR and DPIL estimates are not perfectly co-located so that a combination of a small spatial shift with large small-scale variations leads to considerable noise on the correlation between GPR and DPIL estimates. 3) The porosity does not explain the variations in hydraulic conductivity completely, which appears reasonable considering that the total porosity does not seem to exert primary control over hydraulic conductivity at the Krauthausen site, as was discussed above.

In conclusion, the analysis of co-located GPR full-waveform inversion results and direct-push hydraulic conductivity data suggests that an overall negative correlation

exists between porosity and hydraulic conductivity at the Krauthausen site. As such, this relationship provides the potential to estimate hydraulic conductivity on the basis of GPR attributes which are sensitive to porosity. Indeed the hydraulic conductivity derived from DPST correlates strongly with the GPR attributes. The weak correlation between the DPIL derived hydraulic conductivity and the GPR attributes, however, suggests that the use of GPR to estimate fine scale variations in hydraulic conductivity is limited at the Krauthausen site. Note that also a multi-linear regression approach, using both GPR attributes (or their principal components) simultaneously, did not lead to a significantly improved correlation between GPR and DPIL data (not shown here). In the following, we present an alternative facies-based approach in which the GPR attributes are not used to directly estimate hydraulic conductivity but instead to distinguish different aquifer units (facies), whose hydraulic conductivity is derived independently, based on direct-push, flowmeter and grain size data.

3.3.4 Facies classification

Lithological facies were defined on the basis of grain size data, available along 3 boreholes (B7, B22 and B32). During the drilling of these boreholes, soil samples were collected with a vertical sampling interval of 10 or 20 cm and analyzed for grain size distribution [Vereecken et al., 2000]. Co-located with the GPR data, a total number of 140 grain size distribution curves are available (Figure 24a).

To extract lithological facies from the ensemble of grain size distribution curves, we applied a simple k-means cluster analysis [MacQueen, 1967]. In this way, grain size distribution curves that are relatively similar to each other (i.e. belong to the same facies) are grouped together in clusters. Figure 24b-d show the results from clustering the multivariate grain size data (12 sieve fractions for each sediment sample) into three clusters. The division into three clusters was chosen based on the findings of Döring [1997] and Tillmann et al. [2008], who analyzed borehole logs and CPT profiles and found that the Krauthausen aquifer can be broadly divided into three lithological units. A similar finding was reported by Gueting et al. [2015], who analyzed multivariate GPR and CPT data and showed that both data sets were best represented by a division into three facies. It should be noted, however, that these facies are not expected to form clearly separated spatial structures in the Krauthausen aquifer, because sedimentation conditions in alluvial systems often change gradually [Miall, 1996], which leads to gradual instead of abrupt changes in subsurface properties (e.g. fining upward sequences). Indeed, it can be observed that the obtained clusters (Figure 24b-d) represent different subsets of a continuous ensemble rather than clearly separated units. Nevertheless, the facies based approach may be used to represent part of the spatial variability in subsurface properties since the variability of the grain size distribution within a certain cluster is reduced compared to the variability of the ensemble of all grain size distribution curves (Figure 24). Table 2 gives characteristic properties for each grain size facies. The main differences are reflected in the sand-to-gravel ratio and sorting.

As described in Section 3.2.4, we conducted a logistic regression of the clustered grain size data (response variable) and the co-located GPR data (explanatory variables). The results are given in Table 3. Note that we constructed four different logistic regression

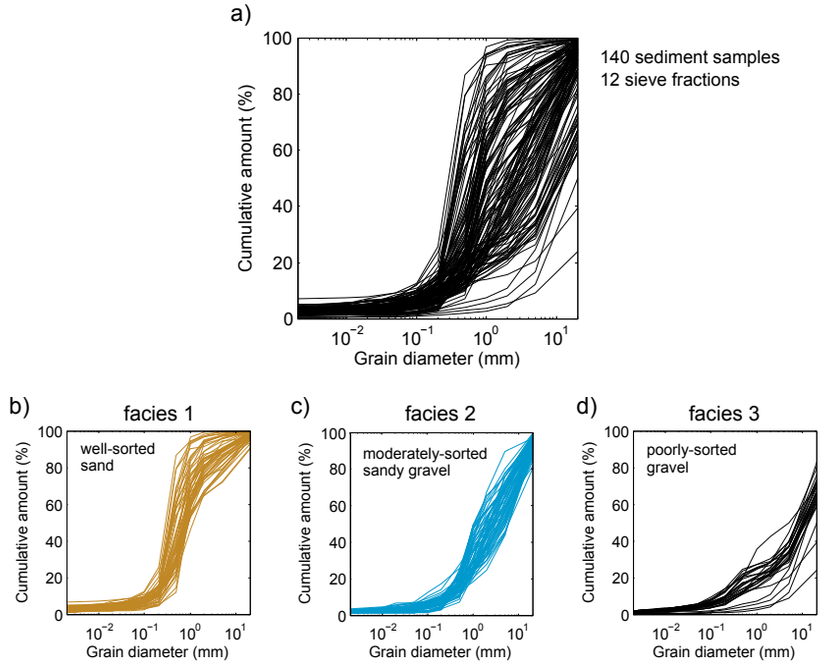


Figure 24: a) Ensemble of 140 grain size distribution curves. b-d) Results of k-means clustering.

models with different sets of explanatory variables. The variables contained in each model are listed in the second column of Table 3. The first model ('Intercept-Only') contains only a constant term, i.e. no GPR data at all. As a consequence, the probabilities derived from this model are constant, simply corresponding to the relative amounts of the individual facies in the grain size data. The reason why we included this most basic model is that any improvement due to incorporating GPR data can be seen by comparison with this basic model. The second model ('Ray-based') contains, in addition to a constant term, the GPR dielectric permittivity obtained from ray-based inversion. In the third model ('FWI (ϵ only)'), the ray-based dielectric permittivity is replaced by the full-waveform dielectric permittivity. The fourth model ('FWI') contains additionally the electrical conductivity obtained from full-waveform inversion.

Table 2: Characteristic properties of grain size facies with standard deviations (in brackets).

Facies	Clay (%)	Silt (%)	Sand (%)	Gravel (%)	U[a] (-)
1	3.1 (1.1)	2.8 (1.3)	76 (11)	18 (10)	5 (2)
2	2.0 (0.5)	2.4 (1.5)	41 (11)	55 (10)	19 (11)
3	1.4 (0.5)	2.9 (1.2)	18 (7)	78 (8)	60 (35)

[a] $U=d_{60}/d_{10}$, the coefficient of uniformity

Table 3: Results of the logistic regression of GPR and grain size data.

Model	Variables	β [a]		t-statistic		p-value		GSD[b]	AIC[c]
Intercept-Only	Constant	0.68 (0.15)	0.93 (0.14)	4.56	6.52	<0.05	<0.05	46%	780
Ray-based	Constant	-23.17 (2.29)	-9.16 (1.42)	-10.13	-6.43	<0.05	<0.05	69%	562
	Ray-based ϵ	1.60 (0.15)	0.73 (0.10)	10.43	7.02	<0.05	<0.05		
FWI (ϵ only)	Constant	-20.42 (1.98)	-8.81 (1.43)	-10.32	-6.14	<0.05	<0.05	73%	558
	FWI ϵ	1.41 (0.13)	0.70 (0.10)	10.48	6.73	<0.05	<0.05		
FWI	Constant	-19.43 (1.99)	-8.75 (1.47)	-9.73	-5.97	<0.05	<0.05	76%	545
	FWI ϵ	1.16 (0.15)	0.77 (0.12)	7.60	6.30	<0.05	<0.05		
	FWI σ	0.24 (0.12)	-0.10 (0.10)	2.08	-0.96	<0.05	0.34		

[a]Maximum likelihood estimates of regression coefficients and standard errors (in brackets)

[b]Proportion of cells with co-located GPR and grain size data, where the modelled facies classification (i.e. highest probability predicted by the logistic regression model) matches the observed facies classification (clustered grain size data along boreholes)

[c]Akaike´s Information Criterion (AIC) = $2p - 2L$, with p the number of model parameters and L the log-likelihood of the fitted model

The maximum likelihood coefficient estimates, β, obtained for each model (Table 3) represent matrices with as many rows as there are explanatory variables and as many columns as there are categories (not including the reference category, i.e. facies 3, for which the coefficients are by definition zero). The first column of β thus corresponds to the distinction between facies 1 and facies 3, while the second column corresponds to the distinction between facies 2 and facies 3 (compare Eq. (12)). The p-values, computed from the t-statistics of the individual components of β, suggest that all coefficients are statistically significant (i.e. $p<0.05$). The only exception is a p-value of 0.34 for FWI σ, which suggest that FWI σ does not provide additional significant information to distinguish between facies 2 and 3. But note that FWI σ is significant to distinguish between facies 1 and 3 (indicated by $p<0.05$, one position further left in Table 3).

The last two columns in Table 3 (GSD and AIC) give global measures that describe how well the different models explain the clustered grain size data. Based on these two metrics, the models listed in Table 3 get better from top to bottom, the FWI model best explains the grain size data. As expected, all models that contain GPR data as explanatory variable perform much better than the basic Intercept-Only model. The improvement using the GPR full-waveform instead of the ray-based inversion results seems only minor. As we will see later, however, this small improvement is quite important.

3.3.5 Comparison of GPR predicted facies distribution with CPT cone resistance logs

Figure 25 shows for one GPR transect the facies distribution predicted by the Ray-based, the FWI (ϵ only) and the FWI logistic regression model. Information about the reliability of the facies classification is included by scaling the color saturation proportional to the probability of facies assignment predicted by the logistic regression model (Eqs. (13) and (14)). Full color saturation refers to a probability of 1, zero color saturation refers to a probability of 0.33 (the lowest possible value for three facies).

Overall, the facies distribution obtained from the three different logistic regression models is quite similar for facies 3, but considerable differences exist regarding the distribution of facies 1 and 2. The GPR predicted facies distribution can be compared with CPT cone resistance logs (black lines in Figure 25) which were measured in close proximity to the GPR transect (Figure 15). While the CPT profiles show only very little correlation with the ray-based predicted facies distribution, they correlate quite well with the FWI predicted facies distribution: low cone resistance values coincide with the depth intervals occupied by facies 1; intermediate and high cone resistance values coincide with the depth intervals occupied by facies 2 and 3, respectively. Note that this order is also what we would expect based on the gravel contents of the individual grain size facies (Table 2).

It is noteworthy that the greatest part of the improvement of the FWI model relative to the ray-based model is already obtained when using the FWI (ϵ only) model. This shows that the improvement is not only due to the additional information from electrical conductivity but also due to the higher resolution of the dielectric permittivity obtained from full-waveform inversion. But including the electrical conductivity improves the consistency of facies across boreholes B62 and B61 (Figure 25c). This shows that the electrical conductivity contributes valuable information for facies estimation.

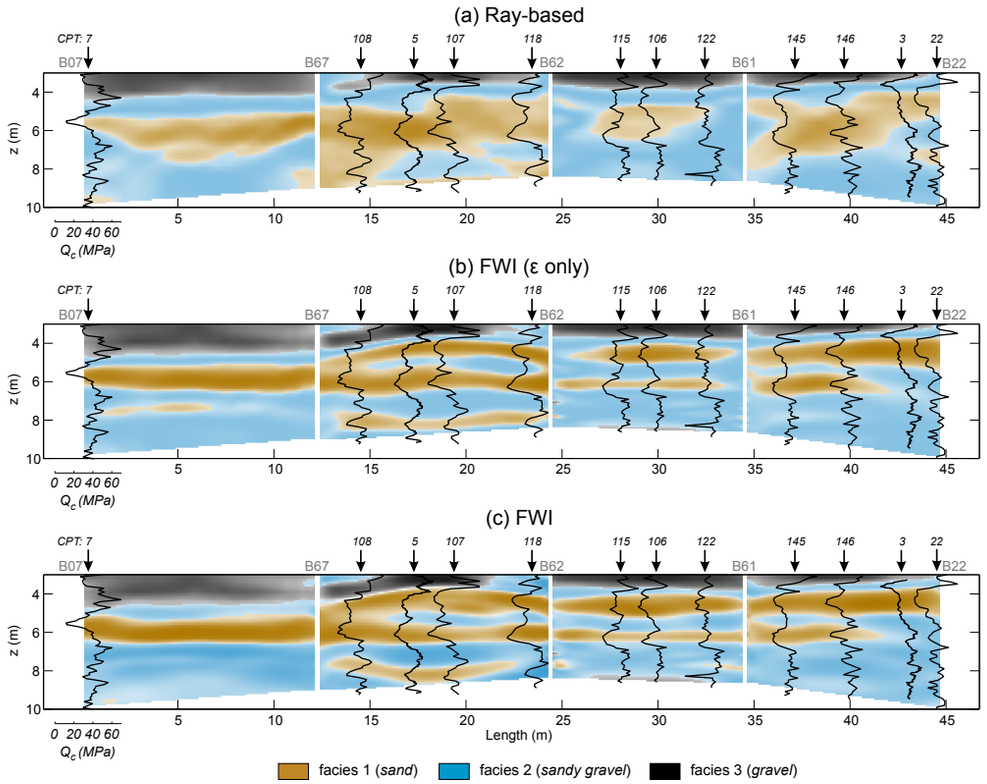


Figure 25: GPR predicted facies distribution using a) ray-based dielectric permittivity, b) full-waveform dielectric permittivity, c) full-waveform dielectric permittivity and electrical conductivity (see Table 3). Color saturation is proportional to the probability of facies assignment predicted by the logistic regression model. Vertical gaps represent boreholes. Black lines represent CPT cone resistance logs.

3.3.6 Comparison of GPR predicted facies distribution with tracer breakthrough

Figure 26a and b show, respectively, the ray-based and the FWI predicted facies distribution in the part of the aquifer where the two tracer experiments were conducted. The transects parallel to the x-axis are the same transects that were monitored with ERT during the tracer tests (Section 3.2.6). The red dashed and solid lines show the outlines of the ERT observed tracer breakthrough for the negative and positive tracer experiment, respectively (compare with Figure 16).

The main difference in Figure 26 between ray-based and FWI results is the distribution of facies 1. Note that facies 1 is the facies with the finest grain size distribution which can therefore be expected to be less hydraulically conductive than facies 2 and 3. The ray-based approach shows a thick facies 1 layer between approximately 5 and 8 m depth which is interrupted at some places by facies 2. The FWI approach, in contrast, suggests

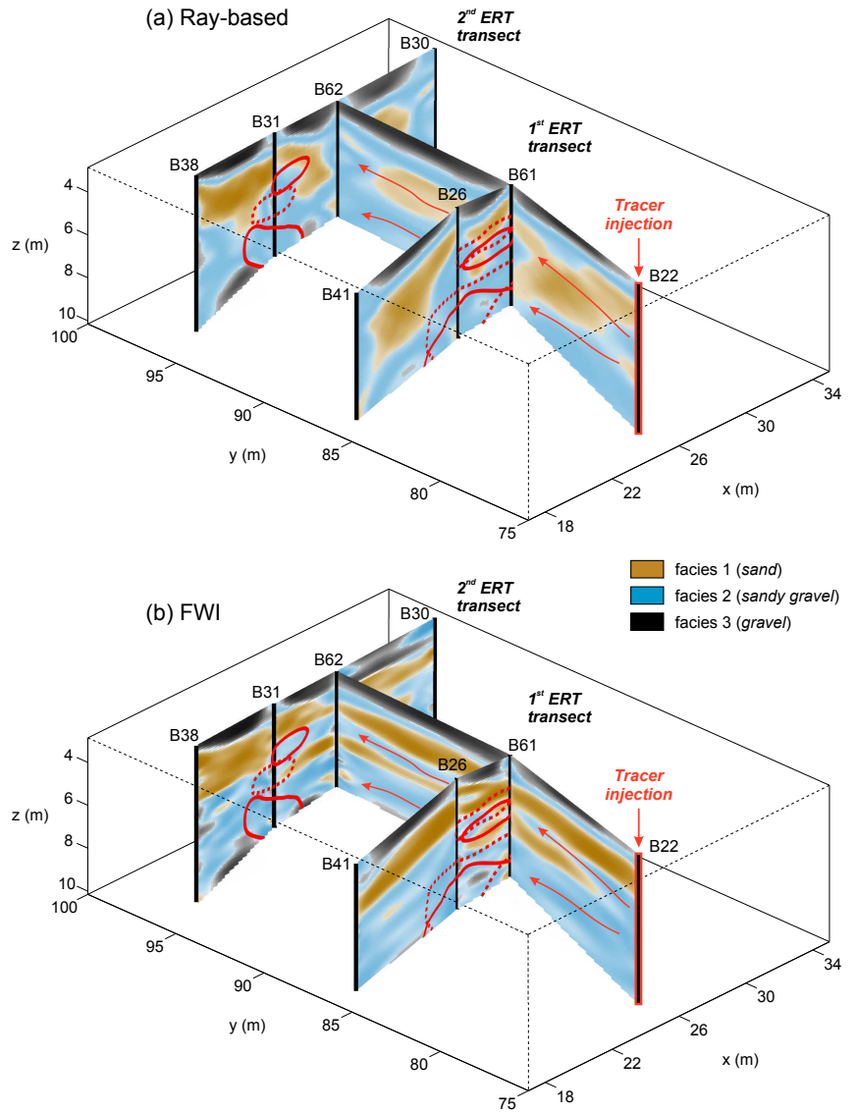


Figure 26: Comparison of ERT observed tracer breakthrough and GPR predicted facies distribution using a) ray-based dielectric permittivity, b) full-waveform dielectric permittivity and electrical conductivity. Color saturation is proportional to the probability of facies assignment predicted by the logistic regression model. Red dashed and solid lines show the outlines of the ERT observed tracer plumes projected on the GPR tomograms (compare with Figure 16).

that this layer is split up between boreholes B62 and B22 into two thinner layers of facies 1, and that between these two layers, there is a thin facies 2 layer at approximately 5-6 m depth that is laterally more or less continuous. At the bottom of this facies 2 layer, there is again a thin facies 1 layer found. The consequence of this facies structure derived from FWI for transport of a tracer that is injected into borehole B22 is that part of the tracer plume will be transported through the upper continuous facies 2 layer (between 5 and 6 m depth) whereas the lower facies 1 layer will lead to a split of the tracer plume (indicated by red arrows in Figure 26). In contrast, the facies structure derived from the ray-based approach would suggest tracer transport mainly through facies 2 below 7 m depth. This finding suggests that the improvement in resolution due to full-waveform inversion is crucial to detect small-scale aquifer structures which are highly important for solute transport. The correlation between the tracer breakthrough and the FWI predicted facies distribution suggests that the derived facies are meaningful with regard to flow and transport. In the following, we will test this hypothesis by investigating the porosity and the hydraulic conductivity of the GPR FWI facies.

3.3.7 Porosity of GPR facies

The porosity of the FWI predicted GPR facies was estimated from CPT neutron logs and from GPR dielectric permittivity (see Section 3.3.2). The porosity of each facies, based on GPR and CPT data, is shown in Figure 27. Note that the portion of the aquifer sampled by GPR and CPT is different, because CPT porosity is available along 17 vertical profiles, whereas GPR porosity is available for each and every GPR pixel. Nevertheless, the porosity estimates from GPR and CPT are relatively consistent. Both methods suggest that the porosity decreases from facies 1, to facies 2, to facies 3, and that the porosity distribution of each facies is distinct (i.e. different from the porosity distribution of the other two facies).

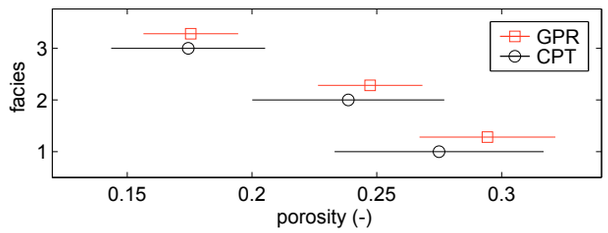


Figure 27: Porosity of the FWI predicted GPR facies according to CPT neutron logs and GPR dielectric permittivity. Horizontal lines represent two times the standard deviation.

3.3.8 Hydraulic conductivity of GPR facies

The hydraulic conductivity of the FWI predicted GPR facies was estimated from the DPST and DPIL profiles; and from flowmeter (Eq. (15)) and grain size data (Eq. (16)). Co-located with the GPR transects, flowmeter data were available along boreholes B22,

B26, B31 and B32; and grain size data were available along boreholes B7, B22 and B32. Unfortunately, there were no flowmeter data available at locations co-located with facies 3, because the flowmeter measurements were conducted only below approximately 4 m depth, and facies 3 occurs mainly above 4 m depth (Figure 26).

Figure 28 shows the hydraulic conductivity of each facies according to DPST, DPIL, flowmeter and grain size. Overall, it can be noted that the different methods show a fairly similar trend, i.e. an increase in hydraulic conductivity from facies 1, to facies 2, to facies 3. This suggests that the GPR derived facies are meaningful with regard to hydraulic conductivity, although the overlap of the hydraulic conductivity distributions for individual facies indicates that the GPR facies are not as well constrained by hydraulic conductivity as they are by porosity (Section 3.3.7).

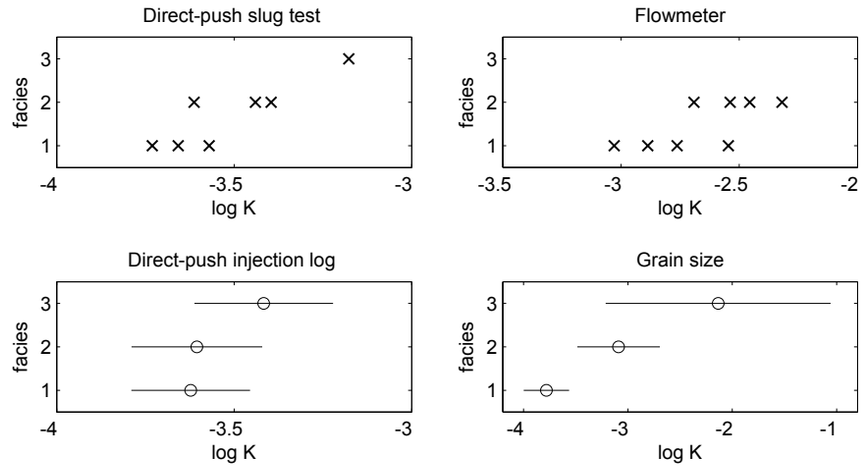


Figure 28: Hydraulic conductivity of the FWI predicted GPR facies according to direct-push slug test, direct-push injection log, flowmeter and grain size data (note the different x-axes!). Horizontal lines represent two times the standard deviation.

Interestingly, the highest hydraulic conductivity is obtained for facies 3, although no tracer breakthrough was observed above 4 m depth, where facies 3 primarily occurs (Figure 26b). This is most likely a consequence of the tracer injection into borehole B22, which is equipped with a filter screen ranging from 2.9 to 10.9 m depth. As facies 3 seems to occur only at the very top of the screened interval (the uppermost GPR pixels in Figure 26 correspond to a depth of 2.86 m depth), it can be expected that the amount of tracer injected into this uppermost layer was only small, which explains that no tracer mass was detected in this layer farther downstream.

The lowest hydraulic conductivity is obtained for facies 1, which is in agreement with the hypothesis that the facies 1 layer caused the splitting of the tracer plume (Section 3.3.6). It is intriguing, however, that the distinction of facies 1 and 2 (the two transport relevant facies) is not very clear when looking at their hydraulic conductivity distributions which overlap (in particular the DPIL derived distributions). Nevertheless,

the difference in the average properties of the two facies was apparently large enough to have an important impact on transport in the aquifer because the tracer tests do show a clear and consistent effect of the facies structure on the observed tracer transport. This indicates that the structure and especially the connectivity of the facies is important even though the measured variability of local hydraulic properties within a facies is large.

Although the hydraulic conductivity derived from direct-push, flowmeter and grain size data agree fairly well regarding the trend, they show quite some disagreement regarding the absolute values (note the different x-axes in Figure 28). In order to evaluate which method best explains the observed tracer arrival times, we derived hydraulic conductivity estimates from the tracer test data, by applying Eq. (17) with $i = 0.002$, the mean hydraulic gradient during the tracer experiments [Müller et al., 2010], and $n = 0.25$, the mean porosity of GPR FWI facies 2 (because the tracer breakthrough occurred primarily through facies 2 (Figure 26b)).

Figure 29 shows a comparison of the tracer test derived hydraulic conductivity and the hydraulic conductivity of facies 2 according to direct-push, flowmeter and grain size data. It can be seen that in order to explain the observed tracer arrival times, log hydraulic conductivities of around -3 are required. This is approximately half an order of magnitude higher than the direct-push derived hydraulic conductivity, and approximately half an order of magnitude lower than the flowmeter derived hydraulic conductivity. The grain size based estimates lie between the direct-push and flowmeter results and show the best agreement with the tracer test derived hydraulic conductivity.

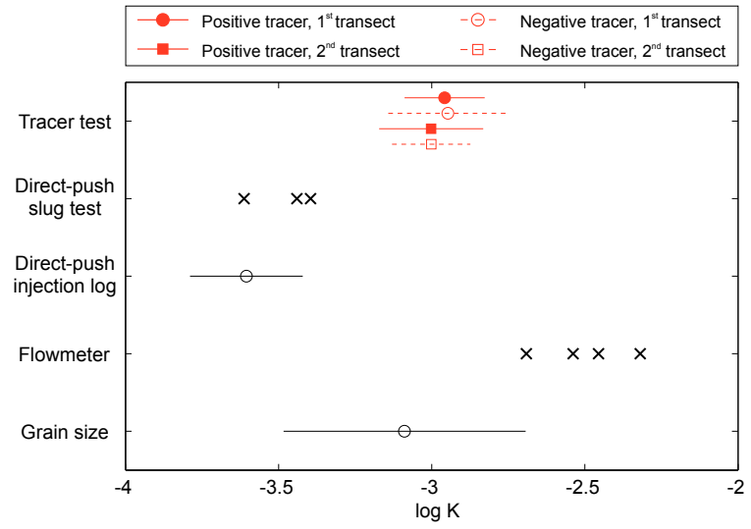


Figure 29: Tracer test derived hydraulic conductivity and comparison with the hydraulic conductivity of GPR FWI facies 2 according to direct-push slug test, direct-push injection log, flowmeter and grain size data. Horizontal lines represent two times the standard deviation.

The differences in absolute hydraulic conductivities obtained from the different methods may be attributed to different sensitivities, support volumes and biases of the different measurements techniques. Particularly at strongly heterogeneous sites, different methods have been reported to give deviating results. For example, Rehfeld et al. [1992] statistically analyzed hydraulic conductivity at the MADE site on the basis of flowmeter, grain size, slug tests and laboratory permeameter tests and found that the different methods yielded hydraulic conductivity distributions with comparable covariance parameters but largely varying mean values. Bohling et al. [2012] compared flowmeter and direct-push derived hydraulic conductivity and found that the two methods identified similar structures in the subsurface, but the direct-push estimates were on average about a factor 5 to 10 smaller compared with the flowmeter results.

One explanation for lower hydraulic conductivity measured with direct-push is that direct-push tools may locally compress the subsurface structure and thereby decrease hydraulic conductivity. Another practical difficulty consists in keeping the injection screen free from clogging because this can be misinterpreted as a decrease in hydraulic conductivity [Dietrich et al., 2008]. Although it was found that the continuous injection of water during DPIL tool advancement [Dietrich et al., 2008] and the use of shielded screens for DPST measurements [Butler et al., 2002] can mitigate or prevent screen clogging, under unfavorable conditions screen clogging may still occur and lead to an underestimation of hydraulic conductivity. Butler et al. [2002] discussed the implication of well development on DPST results and noted that insufficient development of the direct-push installed well can lead to significant underestimation of hydraulic conductivity by DPST.

Flowmeter derived hydraulic conductivities, on the other hand, are well known to be sensitive to skin effects, filter pack, ambient flow in the well and head losses in well and screen [Rehfeldt et al., 1989; Molz et al., 1994]. Ruud et al. [1999] demonstrated numerically that the combined effect of flowmeter-induced hydraulic head losses and cross-flow through the filter pack can easily lead to an overestimation of the true hydraulic conductivity by a factor 10.

Finally, we want to note that the method-based discrepancy in absolute hydraulic conductivity observed in the present study (approximately one order of magnitude) appears less dramatic if one considers the enormous variability (several orders of magnitude) of hydraulic conductivity in the subsurface.

3.4 CONCLUSIONS

High resolution tomographic images of the electrical properties of an alluvial aquifer were obtained from full-waveform inversion of cross-borehole GPR data. By analyzing multiple crosshole planes and stitching together the tomograms from adjacent planes, we were able to image, with a decimeter scale resolution, the dielectric permittivity and electrical conductivity along several cross-sections up to 50 m in length and 10 m in depth.

Porosity estimates derived from the GPR results using the Complex Refractive Index Model (CRIM) showed a strong correlation with porosity estimates derived from CPT

neutron logs (correlation coefficient of 0.8). A less pronounced correlation was observed between the GPR results and the hydraulic conductivity from direct-push injection logs (correlation coefficient of -0.4).

By employing a logistic regression model that relates the electrical properties obtained from GPR to lithological facies obtained from grain size data, we utilized the GPR full-waveform inversion results to predict the spatial distribution of lithological facies along the GPR transects. Vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data suggest that the GPR predicted facies are meaningful with regard to porosity and hydraulic conductivity, even though the distributions of individual facies show some overlap. The hydraulic conductivity estimates from the different methods (direct-push, flowmeter, grain size) were found to identify similar trends but deviating absolute values. Taking tracer test derived hydraulic conductivity as a reference, the direct-push derived hydraulic conductivities were approximately half an order of magnitude lower, and the flowmeter derived hydraulic conductivities were approximately half an order of magnitude higher. Somewhat surprisingly, the indirect (i.e. purely empirical) estimation of hydraulic conductivity from grain size data was found to best explain the observed tracer arrival times.

Comparison of the GPR predicted facies distribution with tracer test data suggests that the plume splitting observed in a tracer experiment was caused by a hydraulically low-conductive sand layer with a thickness of only a few decimeters. Because this sand layer was detected by GPR full-waveform inversion but not by conventional GPR ray-based inversion we conclude that the improvement in spatial resolution due to full-waveform inversion is crucial to detect small-scale aquifer structures that are highly important for solute transport.

We hope that the results presented here will stimulate further applications of GPR full-waveform inversion. Admittedly, the results from intensively studied test sites, such as the Krauthausen site, must be regarded as best case scenarios of what information a method can provide. Yet, full-waveform inversion of crosshole GPR data has been tested under different conditions at several sites now [Klotzsche et al., 2013, 2014; Gueting et al., 2015], and we believe that hydrological site characterization could often benefit from its application (provided that the fundamental requirement for any GPR application is fulfilled, i.e. signal attenuation due to electrical conductivity is not too high).

A particularly interesting application in our view would be to use tomograms from GPR full-waveform inversion as "training images" for multiple-point-statistical (MPS) simulations [e.g. Mariethoz and Caers, 2014]. In this framework, GPR full-waveform inversion would be used to collect site-specific high resolution information about the geological structures that exist at a site; and MPS simulations would be used to stochastically derive 3D subsurface models based on the structural information provided by the GPR tomograms. The resulting models could be parametrized for flow and transport simulations.

4 | 3-D AQUIFER MODEL GENERATION USING MULTIPLE-POINT STATISTICS§

ABSTRACT

Suitable training images (TIs) for multiple-point statistics (MPS) are difficult to identify in real-case 3-D applications, posing challenges for modelers trying to develop realistic subsurface models. This study demonstrates that high resolution 2-D geophysical images, when employed as training and conditioning data, can provide sufficient information for 3-D MPS simulations. The advantage of such data-driven approach is that it does not rely on any external (possibly inappropriate) TI. The disadvantage is that 3-D MPS simulations must be carried out based on 2-D information. We have tested three different approaches (two existing, one new) to overcome this problem. The two existing approaches rely on the reconstruction of incomplete images and on sequential 2-D simulations, respectively. The third approach is a new combination of the two former approaches. The three approaches are applied to model the 3-D facies structure of an alluvial aquifer based on high resolution ground penetrating radar cross-sections. We evaluate the quality of a simulation outcome based on the similarity of its multiple-point histogram (MPH) with reference MPHs, derived from the geophysical images. Our analysis reveals that the first approach (image reconstruction) performs well close to conditioning data, but farther away from the data the simulation results deteriorate. Quite conversely, the second approach (sequential 2-D) performs well when only few conditioning data exist, but with increasing simulation sequence the quality decreases. The newly proposed third approach integrates the benefits from both approaches and is found to reproduce the reference MPHs significantly better than each of the two approaches stand-alone.

4.1 INTRODUCTION

Flow and transport processes in the subsurface are controlled by the complex heterogeneity found in most geologic media. Geostatistical simulation techniques aim at mimicking geological heterogeneity to derive subsurface models that can be used to simulate and predict, for example, groundwater flow and contaminant transport [e.g. Deutsch and Journel, 1998]. Considerable attention has been devoted to the development of multiple-point statistical (MPS) simulation approaches [Guardiano and Srivastava, 1993; Strebelle, 2002; Straubhaar et al., 2011; Mariethoz et al., 2010; Mariethoz and Caers, 2014] because such approaches are able to model complex geometries and connectivity pat-

§ this chapter has been published as: Gueting N., Caers J., Comunian A., Vanderborght J., Englert A., (2017), Reconstruction of three-dimensional aquifer heterogeneity from two-dimensional geophysical data, *Mathematical Geosciences*, http://dx.doi.org/10.1007/s11004-017-9694-x

terns, that cannot be modeled using traditional variogram-based simulation techniques [Gómez-Hernández and Wen, 1998; Western et al., 2001; Renard and Allard, 2013]. The basic idea of MPS simulations is to generate models by reproducing the multiple-point spatial statistics derived from a user-defined training image (TI). Because the spatial structure of the TI is reproduced in the model, the choice of the TI is of most critical importance. As a consequence, applying MPS to real field cases can be challenging, because finding an appropriate TI is not always straight-forward, especially for 3-D cases [Huysmans and Dassargues, 2009; Le Coz et al., 2011].

One possible approach is to use some external, conceptual TI that is anticipated to reflect the type of spatial structures that (presumably) exist at the site being modeled. Conceptual models for different geological settings may be found in data bases [Comunian and Renard, 2008; Mariethoz and Caers, 2014]. Alternatively, the TI itself may be modeled using object-based methods [Maharaja, 2008] or process-imitating models [Comunian et al., 2013]. If the geological setting is unclear, Hermans et al. [2015] proposed to allow multiple alternative TIs, and narrow down the range of initially plausible TIs by validation/falsification with geophysical data.

In this study, we take a different approach and utilize data, instead of an external conceptual TI, to directly inform MPS simulations. The motivation for this data-driven approach is that data, collected directly at the site of interest, contains information about the actual field; and therefore it may be advantageous to rely on the data rather than on some external TI that is possibly not representative for the actual site being modeled [Ortiz and Deutsch, 2004; Mariethoz and Renard, 2010]. Until now, data-based TIs were obtained mainly from outcrops [Bayer et al., 2011; Huysmans and Dassargues, 2012; Pickel et al., 2015]. However, outcrops are often not available and, at best, located at the outer boundaries of the domain to be modeled. In contrast, state-of-the-art geophysical imaging techniques allow imaging the subsurface structure with high resolution and coverage directly at the location of interest. For example, Gueting et al. [2017] used cross-borehole ground penetrating radar (GPR) to characterize the spatial distribution of lithological facies in an alluvial aquifer. By applying a high resolution full-waveform inversion to analyze the tomograms from multiple crosshole planes, they were able to map the distribution of facies with a decimeter scale resolution, along cross-sections up to 50 m long and 10 m deep.

The present study builds upon the work by Gueting et al. [2017] and addresses the question if the GPR cross-sections provide enough information to reconstruct the 3-D aquifer structure using MPS. The main problem we are facing with this approach is that 3-D MPS simulations have to be carried out on the basis of 2-D information. This kind of problem has already been investigated by Mariethoz and Renard [2010] and Comunian et al. [2012] who developed two approaches that enable 3-D MPS simulations but do not require a 3-D TI. Both approaches have proven their ability to generate 3-D models when only 2-D data are available [Mariethoz and Renard, 2010; Comunian et al., 2012; Bayer et al., 2015; Pickel et al., 2015], but only little is known about when these approaches work or fail, and under which conditions one approach should be preferred over the other. In the present study, we quantitatively compare the performance of the two approaches based on a proxy for simulation quality that is derived from the similarity of simulated and observed spatial patterns. Using this proxy, we discuss the benefits

and limitations of the two existing approaches and we propose a new approach that integrates the benefits from both approaches.

4.2 MATERIAL AND METHODS

4.2.1 Study site and GPR data set

The Krauthausen test site, set up by the research center Jülich in 1993 [Vereecken et al., 2000], is situated in the southern part of the Lower Rhine Embayment, close to the city of Düren. The 200 m × 50 m test site is equipped with a total number of 76 boreholes that reach down to the base of a shallow unconfined aquifer at approximately 10 m depth. The aquifer is composed of alluvial terrace sediments that were deposited by a local braided river system on top of older Rhine and Maas sediments [Englert, 2003]. The aquifer material comprises sand and gravel in varying proportions, with only small amounts of clay.

In 2013, an extensive cross-borehole GPR survey was conducted in the central part of the test site [Gueting et al., 2015, 2017]. The GPR data were inverted for dielectric permittivity and electrical conductivity using a recently developed full-waveform inversion approach that is preferable, with respect to spatial resolution, over conventional ray-based inversion approaches [Klotzsche et al., 2010, 2013]. To translate the electrical properties derived from GPR into hydrogeologically meaningful facies, Gueting et al. [2017] conducted a logistic regression of GPR results and co-located grain size data, and used the fitted logistic regression model to predict the distribution of lithological facies along the GPR transects. A 3-D plot of the obtained facies distribution along the various 2-D GPR transects is shown in Figure 30. Overall, the facies are arranged in layer-like

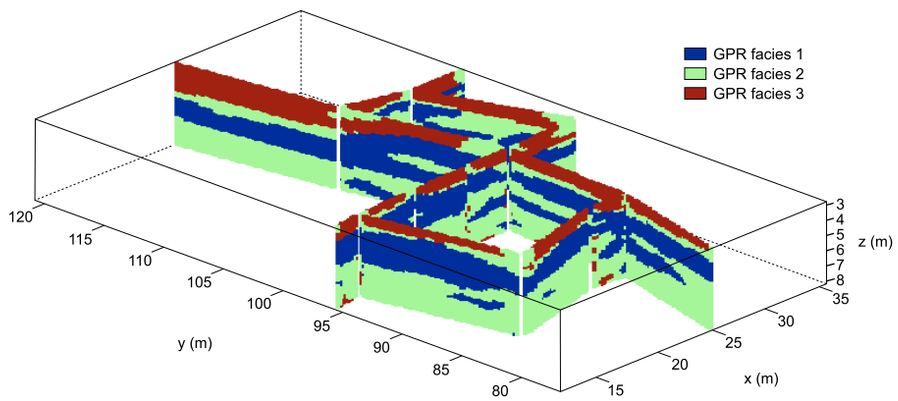


Figure 30: GPR derived facies distribution along 15 individual 2-D cross-borehole planes. Details on the inversion of the GPR data and the classification into facies are given by Gueting et al. [2017].

Table 4: Porosity n and log hydraulic conductivity Y of the GPR facies according to different methods (CPT = Cone Penetration Test; GPR = Ground Penetrating Radar; DPST = Direct-Push Slug Test; DPIL = Direct-Push Injection Log; FM = Flowmeter; GSD = Grain Size Distribution). Details on the methods and their application at the Krauthausen site can be found in Gueting et al. [2017].

Facies	n_{CPT}	n_{GPR}	Y_{DPST}	Y_{DPIL}	Y_{FM}	Y_{GSD}
1 (sand)	0.27	0.29	-3.65	-3.62	-2.81	-3.78
2 (sandy gravel)	0.24	0.25	-3.48	-3.60	-2.50	-3.09
3 (gravel)	0.17	0.18	-3.18	-3.42	n.a.	-2.14

structures that appear plausible for a braided river alluvial setting. Note that despite the generally layered architecture, significant lateral variations in number and thickness of layers can be observed, which indicates that conceptualizing the aquifer as a perfectly stratified medium (as one may be tempted to do if only 1D vertical borehole logs were available) would be an oversimplification of the true aquifer structure. This highlights the value of crosshole geophysical imaging techniques, which yield information in 2-D, along the vertical *and* the lateral direction.

Vertical profiles of porosity and hydraulic conductivity from direct-push tests, flowmeter measurements and grain size data, were used to hydrogeologically characterize the GPR facies. Table 4 gives the porosity and hydraulic conductivity estimates for each facies, according to the different methods. Although the absolute values from the different methods show some disagreement, there is a clear and consistent trend that the hydraulic conductivity increases from facies 1 to facies 2 to facies 3, whereas the porosity decreases. Gueting et al. [2017] compared the GPR derived facies distribution with the tracer breakthrough observed by Müller et al. [2010] and found a clear effect of the facies structure on the observed transport, which confirmed the hydrogeological relevance of the GPR derived facies. In the present study we take the 2-D facies structure along the GPR transects obtained by Gueting et al. [2017] as given, and address the problem of estimating the 3-D facies structure. In other words, the facies distribution along the 2-D GPR transects (Figure 30) is assumed to reflect the true aquifer structure; the goal is to reconstruct the 3-D volume between these transects.

4.2.2 Multiple-point-statistical simulations

Basic principle

This section briefly reviews the principle idea of multiple-point-statistical (MPS) simulation and its implementation in the *impala* algorithm [Straubhaar et al., 2011] and in the *direct sampling* algorithm [Mariethoz et al., 2010], which were used in the present study.

Consider an attribute S with K possible states. S may represent a categorical variable which identifies, for instance, the occurrence of different geological facies; it may also represent a continuous variable, such as porosity or permeability, whose range was divided into K classes using some kind of classification technique. The objective of MPS simulations is to assign a value $S(u)$ to each node u in a 2-D or 3-D simulation

grid, by taking into account the spatial position of u relative to already informed nodes (i.e. conditioning data or already simulated nodes). The data values and geometry of N informed nodes in the neighborhood of u can be expressed as the data event d_u

$$d_u = \{S(u + h_1), ..., S(u + h_N)\} \tag{18}$$

where $h_{1,...,N}$ are 2-D or 3-D vectors that describe the spatial position of the informed grid nodes with respect to the central node u. To simulate $S(u)$ for a given data event d_u we are interested in the conditional probability distribution function (cpdf)

$$\text{Prob}\{S(u) = k \mid d_u\} = f(u; k \mid d_u), \qquad k = 1, 2, ..., K \tag{19}$$

The basic idea of MPS is to infer this cpdf from a user-defined Training image (TI) that is supposed to be representative for the geological setting being modeled [Strebelle, 2002].

In the *impala* algorithm [Straubhaar et al., 2011], the cpdf is built by scanning the TI prior to simulation, and storing the number of occurrences of individual patterns in a catalog. Then the nodes in the simulation grid are sequentially visited and, for each node u (characterized by the data event d_u), the number of replicates of d_u in the TI is retrieved from the catalog, and a value for $S(u)$ is drawn randomly based on the histogram of the central node values for all replicates found in the TI.

A different strategy is used in the *direct sampling* algorithm [Mariethoz et al., 2010]. Instead of explicitly building the cpdf by scanning the TI prior to the simulation, this step is skipped by directly sampling the TI during the simulation. As soon as a TI pattern is found which either matches exactly d_u, or whose mismatch is below a specified threshold, the scanning is stopped and the central node of the found TI pattern is directly pasted into the simulation grid. Because the TI is scanned randomly, this strategy is completely equivalent to drawing a random value from the cpdf, yet it increases simulation speed.

3-D reconstruction approach

Recognizing that an important practical limitation of MPS modeling consists in finding an appropriate TI, Mariethoz and Renard [2010] developed a data-driven image reconstruction approach that uses multiple-point statistics but does not require a TI. Instead of extracting multiple-point statistics from a TI, their approach extracts multiple-point statistics directly from the available data (i.e. the informed grid nodes), which serve as training and conditioning data.

In the present study, we applied the approach by Mariethoz and Renard [2010], from here on called the *3-D reconstruction approach*, to derive 3-D facies models for the Krauthausen aquifer using the GPR derived facies distribution along the various 2-D crosshole planes as training and conditioning data (Figure 30). For the simulation, the approximately 25 m $\times$ 50 m $\times$ 6 m aquifer volume shown in Figure 30 is discretized on a grid of 131 $\times$ 244 $\times$ 30 nodes. This results in approximately 10^6 nodes, of which about 2% are informed by the GPR planes. The simulations were carried out with the *direct sampling* algorithm using a maximum neighborhood of 35 nodes, an acceptance threshold of 0 (this means that the TI is initially scanned for patterns that exactly match the data event; if no such pattern is found, the TI pattern with the lowest mismatch is

accepted) and a simulation path that ensures that only nodes with at least one informed neighbor are simulated. This parametrization was chosen based on the examples given by Mariethoz and Renard [2010] who used the same parameter values for their "Real-case 3-D application" (Example 3.4 in their paper).

Sequential 2-D approach

In field studies, providing a 3-D training image will often be more challenging than providing a 2-D training image because 2-D information can often be obtained with less effort e.g. from outcrops or geophysical surveys. This has motivated Comunian et al. [2012] to develop a method that allows conducting 3-D MPS simulations on the basis of 2-D training images. Two perpendicular 2-D training images need to be specified. These TIs are used to perform multiple sequential 2-D MPS simulations in a 3-D grid. Because the simulations are conducted along perpendicular directions, the simulated slices cross each other, and the nodes along the intersection lines with already simulated slices can be used as conditioning data for the following 2-D simulations. In this way, a 3-D volume is eventually generated "slice-by-slice" without performing an actual 3-D simulation at any time. As a consequence, their approach requires only 2-D TIs instead of 3-D TIs.

Figure 31: 2-D perpendicular training images used for the sequential 2-D approach.

In the present study, two perpendicular TIs were derived from the longest longitudinal GPR transect and the longest transverse GPR transect, respectively (Figure 31). The 2-D TIs were used to apply the approach by Comunian et al. [2012], from here on called the *sequential 2-D approach*, to simulate the facies distribution in the $131 \times 244 \times 30$ nodes grid introduced already in the previous section. Following Comunian et al. [2012], we used 3 multigrids and z as an auxiliary variable to conduct 2-D simulations with the *impala* algorithm. Figure 32 illustrates the simulation sequence for the first 10 slices. At the beginning, the four outer surfaces are simulated (black lines). In the following, the simulation proceeds by dividing the remaining "empty spaces" into half (i.e. the slices

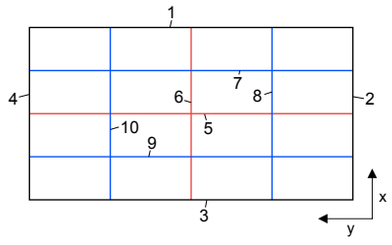


Figure 32: Simulation sequence for the first 10 slices using the sequential 2-D approach (top view). The numbers refer to the order in which individual slices are simulated. Each 2-D simulation considers the nodes along the intersection lines with already simulated slices as conditioning data.

along the central red lines are simulated next, followed by the slices along the blue lines). Note that while the TIs comprise only the two main GPR transects, all GPR transects are included as conditioning data (Figure 30).

Combined approach

As will be shown in Section 4.3.1 and 4.3.2, our results suggest that the 3-D reconstruction approach is preferable over the sequential 2-D approach when dense conditioning data are available; but when sparse conditioning data are available, the sequential 2-D approach is preferable over the 3-D reconstruction approach. With the goal to integrate the benefits from both approaches, we propose a combined application of the two approaches. The basic idea is to employ each approach under the conditions where its performance is optimal: the sequential 2-D approach in the beginning, when only few conditioning data exist; the 3-D reconstruction approach later, when already simulated nodes contribute more and more conditioning data. We propose the following 2-step workflow:

1. The first step consists of starting the sequential 2-D approach exactly in the same manner as already described. After a user-defined number of 2-D simulations, however, the sequential 2-D approach is stopped, and the partly informed simulation grid is saved. At this stage, the simulation grid contains (a) the initially available conditioning data and (b) the already simulated 2-D slices.

2. This is the starting point for the second step, which consists of running the 3-D reconstruction approach using the partly informed simulation grid from the first step, as training and conditioning data.

Note that this workflow is not at all restricted to the data set used in the present study, but can be applied, without any additional implementations, whenever the sequential 2-D approach can be applied.

4.2.3 Evaluating the quality of a simulation outcome

We evaluate the quality of a simulation outcome based on its multiple-point histogram (MPH). The MPH indicates the frequency of spatial patterns in a grid, for a user-defined template [Boisvert et al., 2007]. For example, for a variable with 3 possible states and a 3×3 nodes template, there would be a total number of 19683 (3^9) possible patterns. The MPH is obtained by scanning the entire grid for each pattern, and recording their number of occurrences [Boisvert et al., 2007].

MPS simulations aim at reproducing the spatial patterns found in a TI. Because grids with similar spatial patterns are characterized by similar MPHs, successful pattern reproduction will result in an MPH that is very similar to the MPH of the TI. However, as the simulation proceeds and the simulation grid gets increasingly filled by informed nodes, the simulation may produce data events that are incompatible with the TI. Such data events deteriorate the simulation and lead to a deviation of the simulated MPH from the MPH of the TI. Taking the MPH of the TI as a reference and quantifying the deviation of a simulated MPH from this reference MPH, thus provides information about

the quality of a simulation outcome. A quantitative metric for the dissimilarity of two MPHs, p and q, is the Jensen-Shannon (JS) divergence [Tan et al., 2014]

$$d(p,q) = \frac{1}{2}\sum_i p_i \log\left(\frac{p_i}{q_i}\right) + \frac{1}{2}\sum_i q_i \log\left(\frac{q_i}{p_i}\right) \tag{20}$$

If more than two MPHs shall be compared, a distance matrix D can be obtained by computing Eq. (20) for all possible pairs of MPHs. In the present study, we use the MPH-distance to the GPR derived facies distribution to evaluate the quality of an MPS simulation outcome. More precisely, we use the 15 available GPR planes to obtain an ensemble of 15 reference MPHs, and we assume that the quality of a simulation outcome is reflected in the distance of its MPH to these reference MPHs. A small distance indicates consistent spatial structures, and thus, high-quality simulation results. A large distance indicates inconsistent structures, and thus, poor simulation results.

Because the simulation outcome is 3-D, but the GPR planes are 2-D, we decompose the simulated 3-D block into 2-D vertical slices. As the simulation grid has dimensions of 131 $\times$ 244 $\times$ 30 nodes, this results in 131 slices along the longitudinal direction (yz) and in 244 slices along the transverse direction (xz). Because we want to investigate in detail if the simulation quality differs with the location in the simulation domain we additionally split each longitudinal slice into 7 segments, and each transverse slice into 4 segments. This results in a total number of 1893 segments with a length of approximately 6 m. By calculating the MPH (for a 3 $\times$ 3 template) for each simulated segment and for each GPR plane, and by computing Eq. (20) for all possible pairs, a distance matrix D is obtained which describes the dissimilarity among all simulated segments and GPR planes. The information in D is visualized using multi-dimensional scaling (MDS) [Caers, 2011]. MDS approximates the MPH-distance by a lower-dimensional Euclidean distance in Cartesian space, which facilitates the visualization of D.

In addition, we propose to use the minimal MPH-distance to a GPR plane as a quantitative proxy for the quality of a simulated segment. As there are 15 GPR planes, there are 15 individual distances $d_{1,\dots,15}$ available for each simulated segment (one distance to each GPR plane). By taking the smallest distance $d_{min} = \min(d_{1,\dots,15})$ as a quality measure, a simulated segment is assigned a high quality if it is similar to any one of the GPR planes. Alternatively, one could use the sum of distances to all GPR planes, but this would mean that if several GPR planes are similar to each other, a simulated segment which is similar to them is better than a simulated segment which is similar to a GPR plane that is different from other GPR planes.

4.3 RESULTS

4.3.1 3-D reconstruction approach (stand-alone)

The facies model obtained from the 3-D reconstruction approach is shown in Figure 33. The simulated facies distribution exhibits 3-D structures that are locally consistent with the conditioning data, i.e. the spatial structures in the GPR planes continue in the simulation grid (Figure 33b). Overall, however, the facies distribution appears patchy. Dis-

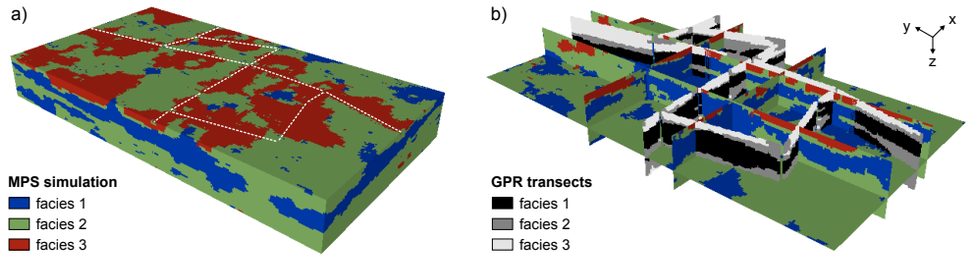


Figure 33: Result of the 3-D reconstruction approach. a) Block view of the simulated 3-D volume. White dashed lines show the locations of the GPR transects, which were used as training and conditioning data. b) Cross-sectional slices through the simulated 3-D volume and comparison with the GPR planes. For clarity, a different color-scale is prescribed to simulated nodes and to GPR informed nodes.

connected structures and isolated pixels can be observed. Particularly further away from the conditioning GPR planes, the simulated facies structures seem to deteriorate. This is most obvious for the distribution of facies 3, which seems to follow the GPR transects in the simulation grid (Figure 33a). In the rest of the simulation domain, facies 3 is less abundant. This is also reflected in the global facies proportions (Table 5). While facies 3 accounts for 14% of the conditioning data, it accounts for only 7% of the simulated 3-D volume. For all three facies, the differences between reference and simulated facies proportions sum up to 24%.

As described in Section 4.2.3, the consistency between the spatial structures in the simulation grid and the spatial structures in the GPR images is assessed by decomposing the simulated 3-D block into multiple 2-D segments and calculating the JS-divergence (Eq. (20)) between the MPHs of these segments and the MPHs of the GPR images. The resulting distance matrix is visualized using MDS. Figure 34 shows the MDS plot for the 3-D reconstruction result. The GPR images are plotted in red. The 2-D segments extracted from the simulated 3-D block are plotted in black. On the one hand, it can be observed that the distribution of black data points comprises the area where the GPR images plot. This shows that some part of the simulated segments are consistent with the GPR data (meaning that they are characterized by an MPH similar to the MPH of the GPR images). On the other hand, the black data points are scattered over a much larger

Table 5: Facies proportions for the conditioning data and for the simulated 3-D models shown in Figure 33a, 36a and 39a

Data / Simulation result	Facies 1	Facies 2	Facies 3	Deviation[a]
GPR conditioning data (reference)	36 %	50 %	14 %	-
3-D reconstruction approach	31 %	62 %	7 %	24 %
Sequential 2-D approach	38 %	51 %	11 %	6 %
Combined approach	38 %	49 %	13 %	4 %

[a] Total deviation from reference (i.e. the sum of deviation over all facies)

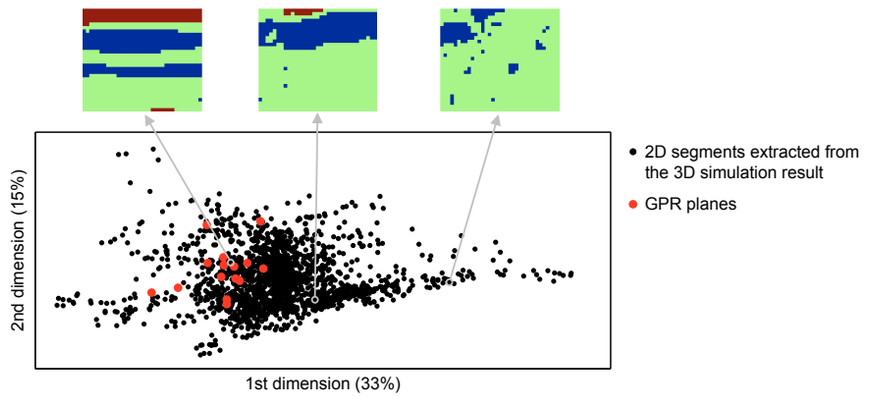


Figure 34: MDS plot for the 3-D reconstruction approach. The distance is the JS-divergence between the MPHs of simulated segments/GPR planes. Percentages in paranthesis quantify the amount of variance explained by a dimension.

area, including regions far away from where the GPR images plot. These simulated segments are not consistent with the GPR data (i.e. they contain spatial structures which are not contained in the GPR images).

Figure 34 includes images of three individual 2-D segments that plot in different regions of the MDS map. As expected, the spatial structures in the segments can be seen to deteriorate with increasing distance to the GPR planes in the MDS plot. This supports the idea that the quality of a simulated segment can be evaluated based on the segment's distance (in terms of the MPH) to GPR planes. As described in Section 4.2.3, we use the minimal MPH-distance to a GPR plane, d_{min}, as a proxy for the quality of a simulated segment. Note that we also tested using the arithmetic and harmonic mean

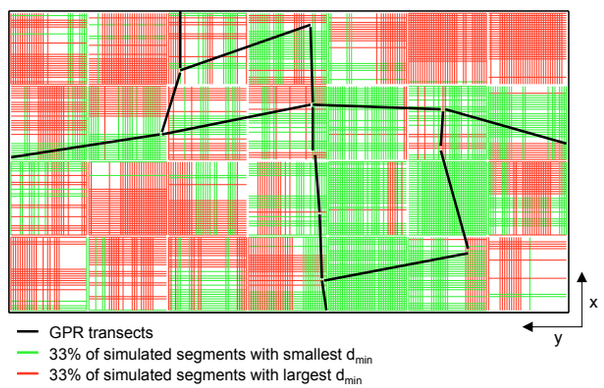


Figure 35: Spatial distribution of 2-D segments with "high quality" (green) and "low quality" (red) derived from the 3-D reconstruction approach. As a measure of quality, the minimal distance d_{min} between the MPH of a simulated segment and the MPH of a GPR plane is used.

of the MPH-distances to GPR planes as a quality proxy, but the results were very similar to the results obtained using the minimal distance d_{min}. Here we show only the results that were obtained with d_{min}. Figure 35 shows in green and red color, the 33% of the longitudinal and transverse segments that are characterized by the smallest and largest d_{min}-values, respectively. While the green segments are distributed closely around the GPR planes, the red slices tend to plot farther away from the GPR planes (e.g. in the corners of the domain). This suggests that the 3-D reconstruction approach yields better results close to the conditioning data than farther away from the conditioning data. This observation is in agreement with the findings of Mariethoz and Renard [2010], who discussed the benefits and limitations of the 3-D reconstruction approach and noted that only "at locations where sufficient information is present, complex shapes can be inferred from the data" [Mariethoz and Renard, 2010]. Our study provides quantitative evidence for this behavior using the JS-divergence between simulated and observed MPH as a proxy for simulation quality.

4.3.2 Sequential 2-D approach (stand-alone)

The 3-D facies model obtained from the sequential 2-D approach is shown in Figure 36. Overall, the facies model is characterized by layer-like structures that appear to be connected across the entire simulation grid. The conditioning GPR planes are well integrated in the simulation grid (Figure 36b). However, particularly in the inner volume of the 3-D block (Figure 36b), noisy simulation results with unclean facies boundaries can be observed. Compared with the 3-D reconstruction result (Figure 33), the facies distribution in the sequential 2-D model is more continuous. Most notably, a more or less continuous top layer of facies 3 is obtained, which was not obtained from the 3-D reconstruction approach. As a consequence, the global facies proportions of the sequential 2-D model better match the facies proportions of the conditioning data. The deviations for all three facies sum up to only 6%, which is a distinct improvement relative to the 24% deviation obtained with the 3-D reconstruction approach (Table 5).

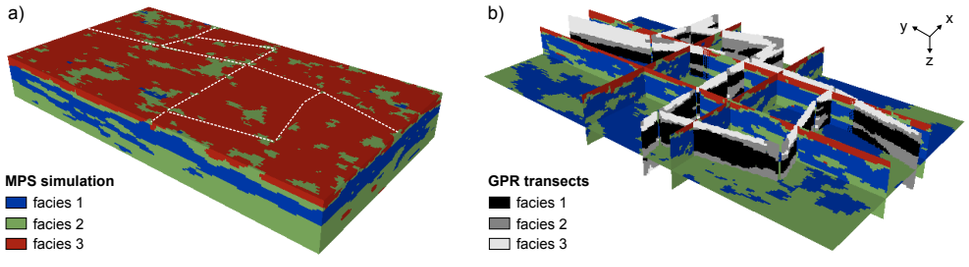


Figure 36: Result of the sequential 2-D approach using the training images shown in Figure 31. a) Block view of the simulated 3-D volume. White dashed lines show the locations of the GPR transects, which were used as conditioning data. b) Cross-sectional slices through the simulated 3-D volume and comparison with the GPR planes. For clarity, a different color-scale is prescribed to simulated nodes and to GPR informed nodes.

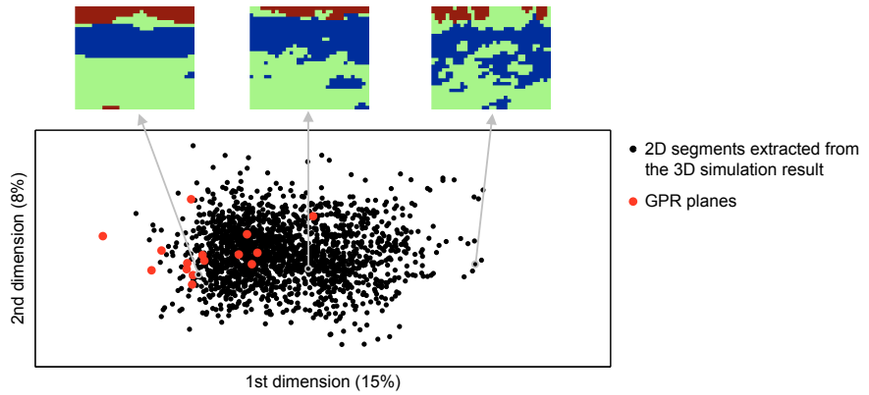


Figure 37: MDS plot for the sequential 2-D approach. The distance is the JS-divergence between the MPHs of simulated slices / GPR planes.

Figure 37 shows the MDS plot for the sequential 2-D simulation results. The distribution of the data points shows that the ensemble of 2-D segments extracted from the 3-D simulated block comprises both segments that are consistent (in terms of the MPH) with the GPR planes, but also segments that are not consistent with the GPR planes. This is similar to what was observed for the 3-D reconstruction approach. What is different for the sequential 2-D simulation result, is how the consistent/inconsistent segments are distributed in the model domain. This is shown in Figure 38a, where again the 33% of the longitudinal and transverse segments with the smallest and largest d_{min}-values are plotted in green and red color, respectively. In contrast to the 3-D reconstruction result, where the green segments were clearly aligned along the GPR planes (Figure 35), here, the green and red segments are quite evenly distributed over the entire model domain.

The quality of the sequential 2-D results, thus, does not seem to be controlled by the proximity to conditioning data. Instead, the quality seems to be a function of the simulation sequence (Figure 38b). This can be explained as follows. The sequential 2-D approach generates 3-D volumes by performing multiple 2-D simulations, considering already simulated slices as conditioning data for the following simulations. With increasing simulation sequence, the 2-D simulations become more and more constrained by the already simulated pixels along the intersection lines with previously simulated slices. As discussed by Comunian et al. [2012] this increasing amount of conditioning data eventually produces data events that are incompatible with the TIs. Such incompatible data events deteriorate the simulation results and are the reason why we see a correlation between simulation sequence and quality (Figure 38b).

In conclusion, the MPH based analysis of the facies models obtained from the 3-D reconstruction approach and from the sequential 2-D approach, highlighted some important differences between the two approaches. The 3-D reconstruction approach was found to yield good results close to the conditioning data, but further away from the conditioning data, the quality of the simulation results decreased. In contrast, the sequential 2-D approach was able to produce good results far away from the conditioning data, but with increasing simulation sequence (i.e. increasing amount of conditioning

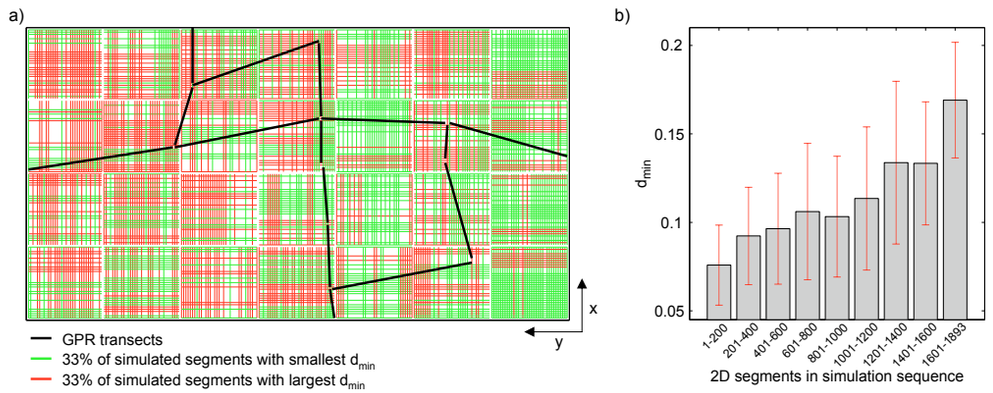


Figure 38: a) Spatial distribution of 2-D segments with "high quality" (green) and "low quality" (red) for the sequential 2-D approach. As a measure of quality, the minimal distance, d_{min} between the MPH of a segment and the MPH of a GPR plane is used. b) Quality of the segments as function of the simulation sequence. Red lines represent two times the standard deviation.

data), the quality of the simulation results decreased. So basically, while the presence of conditioning data improves the 3-D reconstruction results, it deteriorates the sequential 2-D simulation results. Given these complimentary requirements, it appears promising to combine the two approaches in order to benefit from the advantages of both approaches.

4.3.3 Combined approach

As described in Section 4.2.2, a combined application of the two approaches was proposed by starting with the sequential 2-D approach, and after a couple of 2-D simulations, switching to the 3-D reconstruction approach. Obviously, the question arises regarding the optimal point to switch from the one approach to the other. We will address this question later in Section 4.4. First, we will analyze the results obtained by choosing a switching point after 10 slices simulated by the sequential 2-D approach. Figure 39 shows the resulting 3-D facies model. Visual comparison with the models obtained from the two approaches stand-alone (Figure 33 and Figure 36) indicates that the combined approach might indeed be an improvement because the obtained model shows neither the disconnected patches nor the noisy facies boundaries, which were the main issues in the models obtained from the two approaches stand-alone. The combined approach also better reproduces the reference facies proportions (Table 5).

To compare the facies models from the three different approaches in more detail, MDS plots were constructed by calculating the MPH-distance between the 2-D segments extracted from all three models and plotting them together in the same MDS plot. The resulting MDS maps are shown in Figure 40a. Individual segments from the three models are compared in Figure 40b. Overall, it can be observed that the segments obtained from the combined approach tend to plot closer to the GPR planes in the

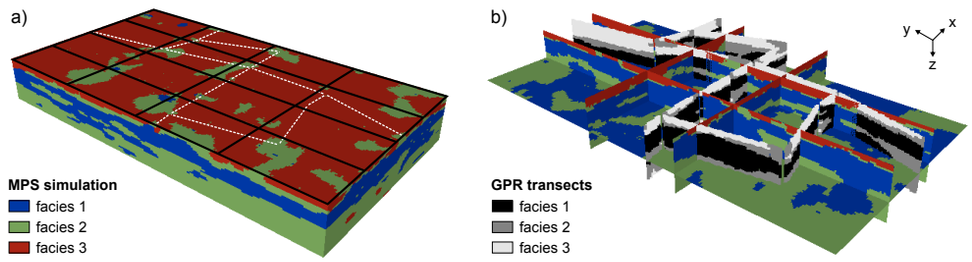


Figure 39: Result of the combined approach. a) Block view of the simulated 3-D volume. White dashed lines indicate the GPR transects. Black lines indicate the slices simulated with the sequential 2-D approach before switching to the 3-D reconstruction approach. b) 2-D slices through the simulated 3-D volume. For clarity, a different color-scale is prescribed to simulated nodes and to GPR informed nodes.

MDS maps than the segments obtained from the other two approaches. This suggests that, of the three approaches, the combined approach produces spatial structures that best resemble the GPR images. However, due to the strongly overlapping data points, it is difficult to decide on the basis of the MDS maps if one approach outperforms another. Moreover, the distances in the MDS maps are not the actual distances but

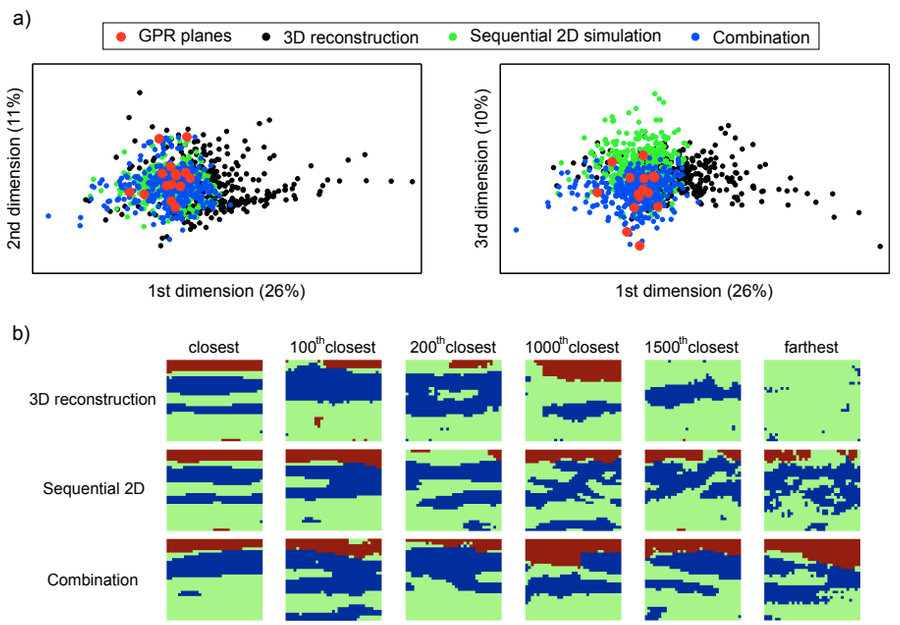


Figure 40: a) MDS map projections along the first three dimensions. For clarity, only every fifth segment extracted from each model is plotted. b) Comparison of individual segments (with different d_{min}) from the different models.

their Euclidean approximations. It appears more appropriate to use the actual distances for quality assessment. Recall that in Section 4.3.1 and 4.3.2 we have used the actual distance d_{min} to identify differences in the quality of segments belonging to the same 3-D model (Figure 35 and 38). Now we will use d_{min} to assess differences in the quality of segments belonging to different 3-D models. Figure 41 shows d_{min} for all segments extracted from (a) the 3-D reconstruction model, (b) the sequential 2-D model and (c) the combined model. By sorting the individual segments (x-axis) with respect to d_{min}, distribution curves for d_{min} are obtained that allow for a quantitative comparison of the quality of different models. Because low values for d_{min} express a strong statistical similarity between the simulated spatial structures and the structures in a GPR image, they indicate high quality simulation results. In contrast, high values for d_{min} indicate low quality simulation results. It is obvious in Figure 41, that the combined approach leads to the distribution curve that is characterized by the lowest d_{min} values. This shows that the combined approach produces spatial structures that are in better agreement with the GPR images than both the 3-D reconstruction and the sequential 2-D approach stand-alone. The distinct improvement is particularly intriguing because the d_{min}-curves from the two approaches stand-alone look quite similar (with slight advantages for the sequential 2-D approach, though). Due to the similarity of the two curves, it appears useless to combine the two approaches to get a better result. But the point is that due to the opposite effect that conditional data has on the two approaches (Figure 35 and 38), the two approaches produce "good" and "bad" segments in different spatial parts of the simulation domain. Therefore, the effect of combining the two approaches is that the "bad" segments of the one approach are replaced by the "good" segments of the other approach. This eventually leads to the distinct improvement that can be observed in Figure 41 for the combined approach.

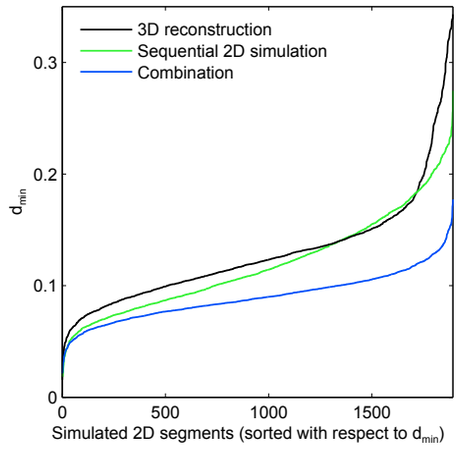


Figure 41: Characteristic distribution curves for d_{min} obtained from the different approaches.

4.4 DISCUSSION

4.4.1 General validity of findings

In Section 4.3 we have analyzed in detail one simulation outcome obtained from each approach. But MPS offers a stochastic modeling framework, i.e. one simulation outcome is only one out of multiple realizations that can be obtained from each approach. Therefore, it is important to test if our observations are valid for other realizations. Figure 42 shows the distribution of d_{min} for three realizations obtained from each approach. The distribution curves from the individual realizations are strikingly similar. Note that this similarity is not a consequence of nearly identical 3-D facies models. In fact, the spatial arrangement of facies in the simulation grid looks quite different from one realization to the other (not shown here). Nevertheless, in terms of d_{min}, which expresses the MPH-based similarity to the GPR images regardless of the exact location of specific structures, the realizations yield nearly identical results. This finding strongly suggests that the results presented in Section 4.3 are generally valid, even though a number of only three realizations is arguably not sufficient to ascertain general validity.

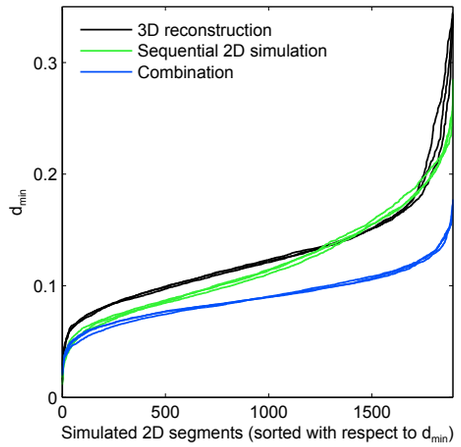


Figure 42: Distribution of d_{min} for three different realizations obtained from each approach.

4.4.2 Optimal switching point

As foreshadowed in Section 4.3.3, we will investigate if there is an optimal point to switch from the sequential 2-D approach to the 3-D reconstruction approach. Principally, an optimal switching point should be not too early, because otherwise there is not sufficient conditioning data available for the 3-D reconstruction approach, but also not too late, because otherwise flawed simulation results, obtained with increasing simulation sequence from the sequential 2-D approach, are irrevocably included into the simulation grid. In fact, avoiding the latter is critical because, once simulated, defective structures get reproduced by the 3-D reconstruction approach, which uses the already simulated

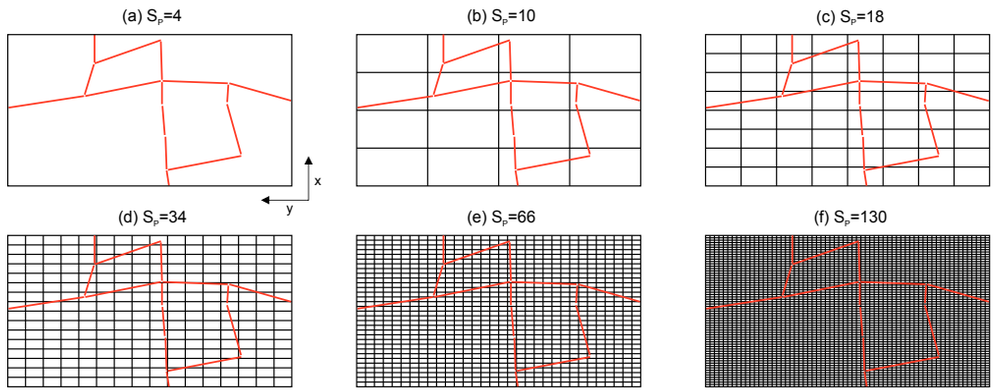


Figure 43: Different switching points. S_P refers to the number of slices simulated with the sequential 2-D approach before switching to the 3-D reconstruction approach. The slices simulated with the sequential 2-D approach are represented by black lines. The conditioning GPR planes are represented by red lines.

slices as training data. To identify an optimal switching point, the performance of the combined approach was tested for different switching points. This was achieved by running the sequential 2-D approach and saving the partly informed simulation grid after 4, 10, 18, 34, 66 and 130 simulated slices (Figure 43). The partly informed grids were then used as starting points for the 3-D reconstruction approach.

Figure 44 shows the distribution of d_{min} obtained for the different switching points S_P. While relatively high values for d_{min} are obtained for very early and late switching points, lower values for d_{min} are obtained for the intermediate switching points. This

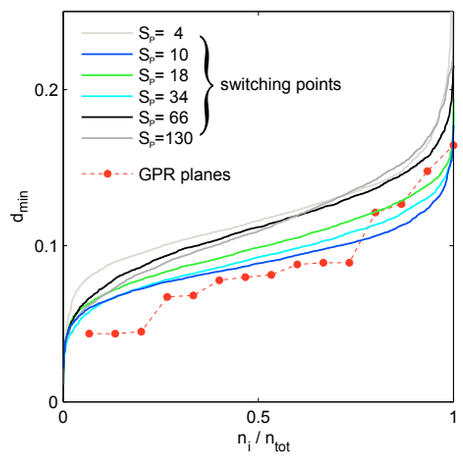


Figure 44: Effect of different switching points on simulation results and comparison with the (reference) distribution of d_{min} derived from the GPR images.

shows that the intermediate switching points yield results that are in better agreement (in terms of the MPH) with the GPR images than the results obtained with the early and late switching points.

What is additionally included in red color in Figure 44 is the distribution of d_{min} derived for the GPR images themselves by calculating for each GPR image the minimal JS-divergence to another GPR image. To enable the comparison of the distribution of d_{min} for simulation results and GPR images, the x-axis in Figure 44 is expressed as the rank number n_i of the sorted segments/images divided by the total number n_{tot} of segments/images (i.e. n_{tot}=1893 for the simulated segments and n_{tot}=15 for the GPR images). This comparison is based on the idea that lower values for d_{min} might not always indicate better simulation results. Consider the extreme case that d_{min} is zero for each 2-D segment extracted from a simulated 3-D model. In this case, each 2-D segment would be a mere copy of one of the GPR images. Clearly, this should not be considered a good simulation result. If we assume that the 15 GPR images are representative for the entire aquifer, then the distribution of d_{min} for the GPR images can be considered a reference distribution, and the distribution of d_{min} for the simulated segments should ideally follow this reference distribution. In Figure 44, the majority of d_{min} values obtained from the GPR images plot below the distribution curves from the simulations, which shows that the similarity among the GPR images is more pronounced than the similarity between simulation results and GPR images. Nevertheless, the GPR image based d_{min} distribution is not distinctly different from the simulation results, and particularly the simulation results obtained with the switching point S_P=10 show a fairly good match with the GPR image based reference distribution.

4.4.3 Possible future research directions

More research should be carried out to further test and optimize the combined approach presented here. One potential improvement would be to develop a strategy against incompatible "simulation fronts" originating from different conditioning slices. This problem was already recognized by Mariethoz and Renard [2010] who applied the 3-D reconstruction approach to reconstruct a synthetic 3-D test case and observed unrealistic sharp transitions in the simulated facies distribution at the equidistance of two parallel conditioning slices. They attributed these artifacts to the custom path (i.e. the sequence in which nodes are simulated) used for the 3-D reconstruction. This path randomly visits nodes that have at least one informed neighbor. This means that first the direct neighborhood of the conditioning data is simulated, then the direct neighborhood of the simulated nodes is simulated, etc. As a result, the simulation gradually moves away from the conditioning data. When simulated nodes originating from distant conditioning slices meet, incompatibilities can occur and lead to sharp transitions in the simulation grid. In the present study, the same custom path was used for the 3-D reconstruction, and occasionally, sharp transitions in the simulated structures between two conditioning slices could be observed. One example for such a sharp transition can be seen in Figure 40b, where the 1000^{th} closest segment for the combined approach shows an upper blue facies layer that abruptly ends as if vertically cut. Probably, this is due to incompatible simulation fronts originating from conditioning slices located left and

right of this segment. As noted by Mariethoz and Renard [2010], a possible strategy to avoid such incompatible simulation fronts would be to use larger neighborhoods for the simulations because this would increase the probability that simulated nodes coming from another slice belong to the neighborhood of the node to be simulated. Alternatively, one could use syn-processing [Mariethoz et al., 2010] to remove inconsistent patterns through recursive un- and re-simulation of nodes. Unfortunately, both strategies would greatly increase the computational demands. Another possibility would be to relax the condition for the path so that not only nodes in the direct neighborhood of informed nodes are simulated but also nodes which are some (small) distance away from informed nodes.

Another potential improvement lies in modifying the sequence of slices simulated with the sequential 2-D approach. In the present study, alternating simulations along xz and yz were performed to gradually fill up the simulation domain (compare Figure 43). As noted by Comunian et al. [2012], it can be beneficial when conditioning data are available, to define a customized sequence by preferentially simulating slices that include as many conditioning data as possible. In addition, when conditioning data are available along 2-D planes (such as from GPR) it appears particularly useful to avoid the simulation of slices located in close distance to conditioning planes with approximately parallel orientation, because in this case, it is very likely that the simulated slice and the conditioning plane contain inconsistent structures, which are difficult to "link" in the later simulations.

Finally, future studies should investigate the importance of the work presented here by conducting flow and transport simulations. In the present study we have evaluated the quality of simulated facies models based on spatial structures (i.e. based on the consistency, in terms of MPH, with available data). Often, the ultimate goal of a subsurface facies model is to enable flow and transport modeling (e.g. in the context of groundwater extraction, oil production, contaminant transport and remediation, etc.). Therefore, it would be interesting to evaluate the quality of a facies model on the basis of flow and transport simulations instead of its spatial structure only. In fact, the facies models derived in the present study provide an excellent opportunity to carry out such flow/transport based evaluations, because porosity and hydraulic conductivity values are available for each facies (Table 4) and it would be straight-forward to use them to parametrize the obtained facies models for flow and transport simulations. Simulation results could be compared with pumping and tracer test data available from previous studies at the Krauthausen site [Li et al., 2007, 2008; Müller et al., 2010]. This comparison would provide the opportunity to verify/falsify individual facies models and, on this basis, decide which approach yields the most realistic facies distribution.

4.5 CONCLUSIONS

The present study demonstrates that high resolution 2-D geophysical images can provide enough information to derive reasonable 3-D models of the subsurface using MPS simulations. The advantage of such data-driven approach is that no external TI (which possibly does not properly reflect the spatial structure of the actual field being mod-

eled) is required. The disadvantage is that 3-D MPS simulations have to be carried out on the basis of 2-D information. We have tested three different approaches (two existing, one new) to overcome this problem. To quantitatively compare the performance of the three approaches, we approximate the quality of a simulation outcome by the similarity of its multiple-point histogram (MPH) with reference MPHs, derived from the geophysical images. In this way, a simulation outcome is considered as "good" if the simulated spatial patterns are consistent with the type of spatial patterns observed along the geophysical transects. Using this MPH-based analysis we can show that the 3-D reconstruction approach proposed by Mariethoz and Renard [2010] produces good results close to the conditioning data, but farther away from the conditioning data the simulation results deteriorate. Quite conversely, the sequential 2-D approach proposed by Comunian et al. [2012] produces good results when only few conditioning data exist, but with increasing simulation sequence the quality decreases because the simulations become too strongly constrained by conditioning data.

To integrate the benefits from both approaches, we propose a combination of the two existing approaches. The combination is achieved by starting with the sequential 2-D approach, and after a couple of 2-D simulations, switching to the 3-D reconstruction approach. In this way, each approach is employed under the conditions where its performance is optimal: the sequential 2-D approach is used when only few conditioning data exist; the 3-D reconstruction approach is used when more and more conditioning data become available. Our results clearly show that the combined approach is able to reproduce the reference MPHs better than each of the two approaches stand-alone. As the combined approach does not require any additional implementations, it is easy to use and straight-forward to apply. The only requirement for its applicability is that suitable 2-D training images are available (i.e. similar requirements as for the sequential 2-D approach). More research should be carried out to further test and optimize the combined approach presented here. In particular, its benefits should be evaluated on the basis of flow and transport simulations.

5 CONCLUSIONS AND OUTLOOK

5.1 CONCLUSIONS

Overall, this thesis presents an approach that combines high resolution geophysical imaging and multiple-point statistical modeling to estimate the 3-D structure of aquifers. Chapters 2 and 3 investigate the value of crosshole GPR full-waveform inversion for hydrogeological site characterization. Chapter 4 explores different approaches to utilize the 2-D information from GPR to generate 3-D aquifer models using multiple-point statistics. In the following, we will summarize the principle findings from the individual Chapters and we will draw some general conclusions.

Chapter 2 describes the measurement and inversion of GPR data along five adjacent crosshole planes. Each crosshole plane is inverted separately, i.e. independent from the adjacent planes. Nevertheless, consistent structures in the permittivity and electrical conductivity distribution can be observed left and right of the boreholes, which indicates robust inversion results. To infer lithological facies from the GPR results we present a zonation approach that is based on cluster analysis: The k-means algorithm is used to partition the inverted GPR tomograms into clusters, and the variance ratio criterion is used to determine the optimal number of clusters. Our results suggest three clusters in the GPR data set. Mapped back into the aquifer cross-section, these clusters delineate spatial zones that integrate the information from both GPR attributes (dielectric permittivity and electrical conductivity). A second cluster analysis, conducted with multivariate CPT data, confirms the number of clusters and their spatial distribution in the aquifer. This clearly suggests that GPR and CPT independently identified the same facies. Grain size data available from one of the boreholes show that the obtained facies boundaries correlate with changes in grain size, sorting and porosity. These results demonstrate that full-waveform inversion of crosshole GPR data followed by simple k-means cluster analysis is an applicable approach to delineate lithological facies in alluvial aquifers. A limitation of this approach consists in the crisp classification scheme provided by k-means (each data point is uniquely assigned to one cluster). While this is a useful approach to detect sharp contrasts in subsurface properties, it is less well suited to characterize gradual changes. In our study this manifests in a better agreement of GPR and CPT clusters in the upper aquifer cross-section where an erosional-surface boundary clearly separates sand from gravel; and in a less well agreement in the lower aquifer cross-section where the aquifer properties change more gradually.

Chapter 3 focuses on the hydrogeological relevance of the information that can be obtained from GPR. We show that porosity estimates, derived from GPR dielectric permittivity, correlate strongly with the porosity independently determined from direct-push neutron logs. The dielectric permittivity from full-waveform inversion is found to more accurately reproduce sharp contrasts and small-scale variations in porosity than the dielectric permittivity from conventional ray-based inversion. A logistic regression

approach is used to predict the distribution of lithological facies along the GPR transects. In contrast to the cluster analysis based zonation approach presented in Chapter 2, the logistic regression approach uses secondary information from grain size distributions to define facies. Although this requires additional data, it has the advantage that it is clearly defined what kind of sediment each facies represents. Moreover, the logistic regression approach assigns each GPR pixel a probability to belong to each facies, which provides additional information about the reliability of the classification of GPR pixels. Cone resistance logs, located along the GPR transects, confirm the existence of the GPR mapped facies. Vertical profiles of porosity and hydraulic conductivity from direct-push, flowmeter and grain size data show that the GPR facies are meaningful with regard to porosity and hydraulic conductivity, even though the absolute hydraulic conductivities from the different methods (direct-push, flowmeter, grain size) differ up to approximately one order of magnitude. Comparison of the GPR predicted facies architecture with tracer test data suggests that the plume splitting observed in a tracer experiment was caused by a hydraulically low-conductive sand layer with a thickness of only a few decimeters. This sand layer was not detected using conventional ray-based GPR data inversion, demonstrating the strength of full-waveform inversion for hydrogeological site characterization.

The facies distribution along the various GPR transects provides valuable information about the architecture and connectivity of subsurface structures. In Chapter 4, we demonstrate that this information can be used to generate stochastic realizations of the 3-D aquifer structure using multiple-point statistics (MPS). The main problem encountered with this approach is that 3-D MPS simulations must be carried out on the basis of 2-D information. Three different approaches to this problem (two existing, one new) are tested and compared. We characterize a simulation outcome in terms of its multiple-point histogram (MPH) and develop a quantitative proxy for simulation quality by comparing the simulated MPHs with reference MPHs, derived from the GPR facies maps. In this way, the similarity between simulated and observed spatial patterns is assessed, regardless of their exact spatial location. Based on this proxy, we are able to evaluate the benefits and limitations of the two existing approaches, and we develop a new combined approach that integrates the benefits from both approaches. Using this combined approach, we obtain reasonable 3-D facies models that reproduce the reference MPHs better than each of the two approaches stand-alone.

In conclusion, this thesis demonstrates that crosshole GPR full-waveform inversion provides information that is highly valuable for hydrogeological site characterization. The particular strength of this method lies in its spatial resolution, which goes well beyond the resolution of classical geophysical imaging techniques such as ray-based GPR data inversion [e.g. Dafflon et al., 2011] or ERT imaging [e.g. Müller et al., 2010]. Equally important is that this method yields information on full cross-sections because this allows one to assess the specific connectivity structure of an aquifer. This information, in turn, is crucial to develop realistic groundwater models because knowledge about the connectivity structure of an aquifer is the key to predict flow and transport in that aquifer [Zheng et al., 2011; Renard and Allard, 2013]. We have presented a multiple-point statistical modeling approach that allows for generating 3-D aquifer models while explicitely accounting for the GPR inferred aquifer structure. In this way, we are able to model the site-specific heterogeneity of an aquifer, which provides an excellent ba-

sis to set up realistic groundwater models for numerical flow and transport prediction. Notably, the proposed workflow is stochastic and yields multiple realizations, which is highly useful for the assessment of model uncertainty.

5.2 OUTLOOK

The final outcome of the approach presented in this thesis are 3-D facies models that reflect the specific architecture and connectivity of the facies inferred from GPR. The next logical step would be to parametrize these facies models for numerical flow and transport simulations. The resulting groundwater models could be verified/falsified by comparing their output with pumping test data and tracer test data that are available from previous studies at the Krauthausen site [e.g. Li et al., 2007, 2008; Müller et al., 2010]. In the following, we will outline possible approaches how the derived facies models could be transformed into groundwater models, and we will indicate some research questions that could be adressed.

A simple approach would be to assign a uniform porosity and hydraulic conductivity to each facies. The required values could be retrieved directly from Table 4. By comparing the resulting flow and transport models, one could evaluate which method (direct-push, flowmeter, grain size) yielded the hydraulic conductivities that best reproduce the observed tracer transport. A particularly interesting question is if the contrast between the measured hydraulic conductivities of the sand and the sandy gravel facies is strong enough to reproduce the observed splitting of the tracer plume.

Instead of assuming internally homogeneous facies, one could also try to account for the heterogeneity within facies (i.e. each facies is assumed to be represented by a distribution of porosity and hydraulic conductivity rather than by uniform values). Wainwright et al. [2014] proposed to model intra-facies heterogeneity by (1) specifying specific parameter distributions for each facies, and (2) assigning pixel-wise parameter values by randomly drawing values from the corresponding distributions. A similar approach could be applied to parametrize the Krauthausen facies models: Specific porosity distributions for each facies are available from CPT neutron logs and from GPR dielectric permittivity (Figure 27); specific distributions of hydraulic conductivity are available from direct-push injection logs and from grain size (Figure 28). One could argue that randomly drawing values from a distribution (i.e. without considering spatial correlation/connectivity) contradicts our premise to properly account for structure and connectivity in aquifer models. But one could also argue that the spatial connectivity is already imposed by the facies structure, so that neglecting connectivity at the intra-facies scale is acceptable. This is anyway a very interesting direction for future research.

Finally, the hydraulic conductivity of facies could be estimated by inverse modeling. In fact, this would represent an interesting follow-up study of the work conducted by Li et al. [2008] who jointly inverted flowmeter and pumping test data to estimate the 3-D hydraulic conductivity structure at the Krauthausen site. Unfortunately, despite considering a huge amount of data in their inversion (175 measurements of steady-state drawdown and 808 flowmeter measurements) they found the inferred hydraulic conductivity field to show distinct 3-D structures only in the close vicinity of those

boreholes with flowmeter data. Some meters away from the boreholes, the inferred hydraulic conductivity field was practically homogeneous. Deriving 3-D facies models as described in this thesis, and using the structural information from these models to additionally constrain inverse modeling, might be the key to inferring more realistic 3-D hydraulic conductivity structures (at Krauthausen and elsewhere).

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19(6):716–723.

Anderson, M. (1989). Hydrogeologic facies models to delineate large-scale spatial trends in glacial and glaciofluvial sediments. *Geological Society of America Bulletin*, 101(4):501–511.

Anderson-Mayes, A. (2002). Strategies to improve information extraction from multivariate geophysical data suites. *Exploration Geophysics*, 33:57–64.

Ayalew, L. and H. Yamagishi (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, 65:15–31.

Barrash, W. and T. Clemo (2002). Hierarchical geostatistics and multifacies systems: Boise Hydrogeophysical Research Site, Boise, Idaho. *Water Resources Research*, 38(10):1196.

Bayer, P., A. Comunian, D. Höyng, and G. Mariethoz (2015). High resolution multi-facies realizations of sedimentary reservoir and aquifer analogs. *Scientific data*, 2:150033.

Bayer, P., P. Huggenberger, P. Renard, and A. Comunian (2011). Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. *Journal of Hydrology*, 405:1–9.

Becht, A., J. Tronicke, E. Appel, and P. Dietrich (2004). Inversion strategy in crosshole radar tomography using information of data subsets. *Geophysics*, 69(1):222–230.

Belina, F., J. Ernst, and K. Holliger (2009). Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches. *Journal of Applied Geophysics*, 68(1):85–94.

Belina, F. A., J. Irving, J. R. Ernst, A. Member, and K. Holliger (2012). Waveform Inversion of Crosshole Georadar Data : Influence of Source Wavelet Variability and the Suitability of a Single Wavelet Assumption. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4610–4625.

Bialas, Z. and A. Kleczkowski (1970). Über den praktischen Gebrauch von einigen empirischen Formeln zur Berechnung des Durchlässigkeitskoeffzienten. *K. Archivum Hydrotechniki, Warschau*.

Bianchi, M., C. Zheng, C. Wilson, G. R. Tick, G. Liu, and S. M. Gorelick (2011). Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths. *Water Resources Research*, 47:W05524.

Binley, A., S. Hubbard, J. Huisman, A. Revil, D. Robinson, K. Singha, and L. Slater (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, 51. doi:10.1002/2015WR017016.

Birchak, J., C. Gardner, J. Hipp, and J. Victor (1974). High dielectric-constant microwave probes for sensing soil-moisture. *Proceedings of the IEEE*, 62(1):93–98.

Bleistein, N. (1986). Two-and-one-half dimensional in-plane wave propagation. *Geophysical Prospecting*, 34(5):686–703.

Boggs, J., S. Young, and L. Beard (1992). Field study of dispersion in a heterogeneous aquifer, 1. overview and site description. *Water Resources Research*, 28(12):3281–3291.

Bohling, G. C., G. Liu, S. J. Knobbe, E. C. Reboulet, D. W. Hyndman, P. Dietrich, and J. J. Butler (2012). Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site. *Water Resources Research*, 48:W02525.

Boisvert, J. B., M. J. Pyrcz, and C. V. Deutsch (2007). Multiple-point statistics for training image selection. *Natural Resources Research*, 16(4):313–321.

Bowling, J. C., A. B. Rodriguez, D. L. Harry, and C. Zheng (2005). Delineating alluvial aquifer heterogeneity using resistivity and GPR data. *Ground Water*, 43(6):890–903.

Brovelli, A. and G. Cassiani (2008). Effective permittivity of porous media: A critical analysis of the complex refractive index model. *Geophysical Prospecting*, 56(5):715–727.

Brown, W. (1956). *Dielectrics, in Encyclopedia of Physics, Vol. 17*. Springer Verlag, New York.

Busch, S., J. van der Kruk, J. Bikowski, and H. Vereecken (2012). Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data. *Geophysics*, 77(6):H79–H91.

Butler, J., E. Garnett, and J. Healey (2003). Analysis of slug tests in formations of high hydraulic conductivity. *Ground Water*, 41(5):620–630.

Butler, J. J. (1997). *The design, performance, and analysis of slug tests*. Lewis Publishers, Boca Raton.

Butler, J. J., J. M. Healey, G. W. McCall, E. J. Garnett, and S. P. Loheide (2002). Hydraulic tests with direct-push equipment. *Ground Water*, 40(1):25–36.

Caers, J. (2011). *Modeling Uncertainty in the Earth Sciences*. Wiley.

Calinski, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3:1–27.

Carmichael, R. S. (1982). *Handbook of Physical Properties of Rocks*. Boca Raton, Florida: CRC Press.

Comunian, A., S. K. Jha, B. M. S. Giambastiani, G. Mariethoz, and B. F. J. Kelly (2013). Training Images from Process-Imitating Methods. *Mathematical Geosciences*, 46(2):241–260.

Comunian, A. and P. Renard (2008). Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. *Hydrogeology Journal*, 17(2):481–489.

Comunian, A., P. Renard, and J. Straubhaar (2012). 3D multiple-point statistics simulation using 2D training images. *Computers & Geosciences*, 40:49–65.

Comunian, A., P. Renard, J. Straubhaar, and P. Bayer (2011). Three-dimensional high resolution fluvio-glacial aquifer analog – Part 2: Geostatistical modeling. *Journal of Hydrology*, 405(1-2):10–23.

Dafflon, B., J. Irving, and W. Barrash (2011). Inversion of multiple intersecting high-resolution crosshole GPR profiles for hydrological characterization at the Boise Hydrogeophysical Research Site. *Journal of Applied Geophysics*, 73:305–314.

Dagan, G. (1989). *Flow and transport in porous formations*. Springer Verlag.

Davis, J. and A. Annan (1989). Ground-Penetrating Radar for High-Resolution Mapping of Soil and Rock Stratigraphy. *Geophysical prospecting*, 37:531–551.

Deutsch, C. and A. Journel (1998). *GSLIB: Geostatistical Software Library and User's Guide, 2nd Ed.* Oxford University Press.

Dietrich, P., J. Butler, and K. Faiß(2008). A Rapid Method for Hydraulic Profiling in Unconsolidated Formations. *Ground Water*, 46(2):323–328.

Dietrich, P., T. Fechner, J. Whittaker, and G. Teutsch (1998). An integrated hydrogeophysical approach to subsurface characterization. *In: Herbert, M.,Kovar, K. (Eds.), Groundwater Quality: Remediation and Protection. IAHS Publication no. 250, Wallingford, UK, pp.*, Pp. 513–519.

Dietrich, P. and J. Tronicke (2009). Integrated analysis and interpretation of cross-hole P- and S-wave tomograms: A case study. *Near Surface Geophysics*, 7(2):101–109.

Dietze, M. and P. Dietrich (2012). Evaluation of vertical variations in hydraulic conductivity in unconsolidated sediments. *Ground Water*, 50(3):450–6.

Dobson, M., F. Ulaby, M. Hallikainen, and M. Elrayes (1985). Microwave dielectric behavior of wet soil - Part II: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, 23(1):35–46.

Doetsch, J., N. Linde, and I. Coscia (2010). Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. *Geophysics*, 75(6):G53–G64.

Dogan, M., R. L. Van Dam, G. C. Bohling, J. J. Butler, and D. W. Hyndman (2011). Hydrostratigraphic analysis of the MADE site with full-resolution GPR and direct-push hydraulic profiling. *Geophysical Research Letters*, 38:L06405.

Döring (1997). *Transport der reaktiven Stoffe Eosin, Uranin und Lithium in einem heterogenen Grundwasserleiter*. PhD thesis, Christian Albrecht Universität Kiel.

Eckhardt, D. V. and P. E. Stackelberg (1995). Relation of Ground-Water Quality to Land Use on Long Island, New York. *Ground Water*, 33(6):1019–1033.

Eisenberg, D. S. and W. Kauzmann (1969). *The structure and properties of water*. Oxford University Press.

Englert, A. (2003). *Measurement, Estimation and Modelling of Groundwater Flow Velocity at Krauthausen Test Site*. PhD thesis, RWTH Aachen.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger (2007a). Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics*, 72(5):J53–J64.

Ernst, J. R., H. Maurer, A. G. Green, and K. Holliger (2007b). Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations. *IEEE Transactions on Geoscience and Remote Sensing*, 45(9):2807–2828.

Ester, M., H. Kriegel, J. Sander, and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. Portland, OR*, Pp. 10–15.

Everitt, B., S. Landau, and M. Leese (2001). *Cluster Analysis*, A Hodder Arnold Publication. Wiley.

Feyen, L. and J. Caers (2006). Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Advances in Water Resources*, 29(6):912–929.

Fogg, G. E. (1986). Groundwater flow and sand body interconnectedness in a thick, multiple-aquifer system. *Water Resources Research*, 22(5):679–694.

Fogg, G. E., S. Carle, and C. Green (2000). Connected-network paradigm for the alluvial aquifer system. *Geological Society of America*, Special Paper 348:25–42.

Gómez-Hernández, J. J. and X.-h. Wen (1998). Technical Note To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1):47–61.

Guardiano, F. B. and R. M. Srivastava (1993). *Multivariate Geostatistics: Beyond Bivariate Moments. In: Soares, A (ed) Geostatistics Tróia '92: Volume 1*, Pp. 133–144. Dordrecht: Springer Netherlands.

Gueting, N., A. Klotzsche, J. van der Kruk, J. Vanderborght, H. Vereecken, and A. Englert (2015). Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests. *Journal of Hydrology*, 524:680–695.

Gueting, N., T. Vienken, A. Klotzsche, J. van der Kruk, J. Vanderborght, J. Caers, H. Vereecken, and A. Englert (2017). High resolution aquifer characterization using crosshole GPR full-waveform tomography: Comparison with direct-push and tracer test data. *Water Resources Research*, 53. doi:10.1002/2016WR019498.

Haendel, F. and P. Dietrich (2012). Relevance of Deterministic Structures for Modeling of Transport: The Lauswiesen Case Study. *Ground Water*, 50(6):935–942.

Harbaugh, A. W. (2005). *MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process*. US Department of the Interior, US Geological Survey Reston, VA, USA.

Helsel, D. and R. Hirsch (2002). *Statistical Methods in Water Resources*. U.S. Geological Survey.

Hermans, T., F. Nguyen, and J. Caers (2015). Uncertainty in Training-Image Based Inversion of Hydraulic Head Data Constrained to ERT Data: Workflow and Case Study. *Water Resources Research*, 51:5332–5352.

Hermans, T. J. (2014). *Integration of near-surface geophysical, geological and hydrogeological data with multiple-point geostatistics in alluvial aquifers*. PhD thesis, University of Liege.

Holliger, K. and T. Bergmann (2002). Numerical modeling of borehole georadar data. *Geophysics*, 67:1249–1257.

Holliger, K., M. Musil, and H. Maurer (2001). Ray-based amplitude tomography for crosshole georadar data: A numerical assessment. *Journal of Applied Geophysics*, 47:285–298.

Hosmer, D. W. and S. Lemeshow (2000). *Applied logistic regression, 2nd Edition*. Wiley-Interscience Publication.

Hubbard, S., J. Chen, J. E. Peterson, E. L. Mayer, K. H. Williams, D. J. Swift, B. Mailloux, and Y. Rubin (2001). Hydrogeological characterization of South Oyster Bacterical Transport Site using geophysical data. *Water Resources Research*, 37(10):2431–2456.

Huysmans, M. and A. Dassargues (2009). Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). *Hydrogeology Journal*, 17(8):1901–1911.

Huysmans, M. and A. Dassargues (2012). Modeling the effect of clay drapes on pumping test response in a cross-bedded aquifer using multiple-point geostatistics. *Journal of Hydrology*, 450-451:159–167.

Huysmans, M., L. Peeters, G. Moermans, and A. Dassargues (2008). Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *Journal of Hydrology*, 361:41–51.

Hyndman, D. W. and S. M. Gorelick (1996). Estimating lithologic and transport properties in three dimensions using seismic and tracer data: The Kesterson aquifer. *Water Resources Research*, 32(9):2659–2670.

Hyndman, D. W., J. M. Harris, and S. M. Gorelick (2000). Inferring the relation between seismic slowness and hydraulic conductivity in heterogeneous aquifers. *Water Resources Research*, 36(8):2121–2132.

Irving, J., M. Knoll, and R. Knight (2007). Improving crosshole radar velocity tomograms: A new approach to incorporating high-angle traveltime data. *Geophysics*, 72(4):J31–J41.

Isaaks, E. H. and M. R. Srivastava (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, USA.

Javandel, I. and P. Witherspoon (1969). A method of analyzing transient fluid flow in multilayered aquifers. *Water Resources Research*, 5(4):856–869.

Journel, A. and T. Zhang (2006). The Necessity of a Multiple-Point Prior Model. *Mathematical Geology*, 38(5):591–610.

Kemna, A., J. Vanderborght, B. Kulessa, and H. Vereecken (2002). Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. *Journal of Hydrology*, 267:125–146.

Keskinen, J., A. Klotzsche, M. C. Looms, J. Moreau, J. van der Kruk, K. Holliger, L. Stemmerik, and L. Nielsen (2017). Full-Waveform Inversion of Crosshole GPR Data: Implications for Porosity Estimation in Chalk. *Journal of Applied Geophysics*. in press.

Kinzelbach, W. (1992). *Numerische Methoden zur Modellierung des Transports von Schadstoffen im Grundwasser, 2. Edition*. R. Oldenbourg Verlag, München.

Klotzsche, A., J. V. D. Kruk, G. A. Meles, J. Doetsch, H. Maurer, and N. Linde (2010). Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland. *Near Surface Geophysics*, 8:635–649.

Klotzsche, A., J. van der Kruk, J. Bradford, and H. Vereecken (2014). Detection of spatially limited high-porosity layers using crosshole GPR signal analysis and full-waveform inversion. *Water Resources Research*, 50:6966–6985.

Klotzsche, A., J. van der Kruk, N. Linde, J. Doetsch, and H. Vereecken (2013). 3-D characterization of high-permeability zones in a gravel aquifer using 2-D crosshole GPR full-waveform inversion and waveguide detection. *Geophysical Journal International*, 195(2):932–944.

Klotzsche, A., J. van der Kruk, G. Meles, and H. Vereecken (2012). Crosshole GPR full-waveform inversion of waveguides acting as preferential flow paths within aquifer systems. *Geophysics*, 77(4):H57–H62.

Klotzsche, A., J. van der Kruk, A. Mozaffari, N. Gueting, and H. Vereecken (2015). Crosshole GPR full-waveform inversion and waveguide amplitude analysis: Recent developments and new challenges. In *2015 8th International Workshop on Advanced Ground Penetrating Radar (IWAGPR)*, Pp. 1–6.

Knudby, C. and J. Carrera (2006). On the use of apparent hydraulic diffusivity as an indicator of connectivity. *Journal of Hydrology*, 329:377–389.

Koltermann, C. E. and S. M. Gorelick (1996). Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, 32(9):2617–2658.

Kurzmann, A., A. Przebindowska, D. Kohn, and T. Bohlen (2013). Acoustic full waveform tomography in the presence of attenuation: a sensitivity analysis. *Geophysical Journal International*, 195(2):985–1000.

Le Coz, M., P. Genthon, and P. M. Adler (2011). Multiple-point statistics for modeling facies heterogeneities in a porous medium: The komadugu-yobe alluvium, lake chad basin. *Mathematical Geosciences*, 43(7):861.

Lessoff, S. C., U. Schneidewind, C. Leven, P. Blum, P. Dietrich, and G. Dagan (2010). Spatial characterization of the hydraulic conductivity using direct-push injection logging. *Water Resources Research*, 46:W12502.

Li, W., A. Englert, O. a. Cirpka, J. Vanderborght, and H. Vereecken (2007). Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests. *Water Resources Research*, 43:W04433.

Li, W., A. Englert, O. a. Cirpka, and H. Vereecken (2008). Three-dimensional geostatistical inversion of flowmeter and pumping test data. *Ground Water*, 46(2):193–201.

Linde, N., S. Finsterle, and S. Hubbard (2006). Inversion of tracer test data using tomographic constraints. *Water Resources Research*, 42:W04410.

Linde, N., A. Tryggvason, J. E. Peterson, and S. S. Hubbard (2008). Joint inversion of crosshole radar and seismic traveltimes acquired at the South Oyster Bacterial Transport Site. *Geophysics*, 73(4):G29–G37.

Linde, N. and J. A. Vrugt (2013). Distributed Soil Moisture from Crosshole Ground-Penetrating Radar Travel Times using Stochastic Inversion. *Vadose Zone Journal*, 12(1):1–16.

MacKay, D. (2002). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, USA.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, eds., volume 1, Pp. 281–297. University of California Press.

Maharaja, A. (2008). TiGenerator: Object-based training image generator. *Computers & Geosciences*, 34(12):1753–1761.

Mariethoz, G. and J. Caers (2014). *Multiple-point Geostatistics: Stochastic Modeling with Training Images*. Wiley-Blackwell.

Mariethoz, G. and P. Renard (2010). Reconstruction of incomplete data sets or images using direct sampling. *Mathematical Geosciences*, 42(3):245–268.

Mariethoz, G., P. Renard, and J. Straubhaar (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46:W11536.

Maurer, H. and A. Green (1997). Potential coordinate mislocations in crosshole tomography: Results from the grimsel test site, switzerland. *Geophysics*, 62(6):1696–17097.

Maurer, H. and M. Musil (2004). Effects and removal of systematic errors in crosshole georadar attenuation tomography. *Journal of Applied Geophysics*, 55(3-4):261–270.

Meier, P. M., J. Carrera, and S. Xavier (1998). An evaluation of Jacob ' s method for the interpretation of pumping tests in heterogeneous formations. *Water Resources Research*, 34(5):1011–1025.

Meles, G. A., S. Greenhalgh, J. van der Kruk, A. Green, and H. Maurer (2011). Taming the non-linearity problem in GPR full-waveform inversion for high contrast media. *Journal of Applied Geophysics*, 73(2):174–186.

Meles, G. A., J. van der Kruk, S. A. Greenhalgh, J. R. Ernst, H. Maurer, and A. G. Green (2010). A New Vector Waveform Inversion Algorithm for Simultaneous Updating of Conductivity and Permittivity Parameters From Combination Crosshole/Borehole-to-Surface GPR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12):3391–3407.

Miall, A. (1996). *The geology of fluvial deposits*. Springer Verlag.

Miall, A. D. (1985). Architectural-element analysis: A new method of facies analysis applied to fluvial deposits. *Earth-Science Reviews*, 22(4):261–308.

Molz, F., G. Boman, S. Young, and W. Waldrop (1994). Borehole flowmeters: Field application and data analysis. *Journal of Hydrology*, 163:347–371.

Molz, F. and R. Morin (1989). The impeller meter for measuring aquifer permeability variations: Evaluation and comparison with other tests. *Water Resources Research*, 25(7):1677–1683.

Morin, R. H. (2006). Negative correlation between porosity and hydraulic conductivity in sand-and-gravel aquifers at Cape Cod, Massachusetts, USA. *Journal of Hydrology*, 316:43–52.

Müller, K., J. Vanderborght, A. Englert, A. Kemna, J. A. Huisman, J. Rings, and H. Vereecken (2010). Imaging and characterization of solute transport during two tracer tests in a shallow aquifer using electrical resistivity tomography and multilevel groundwater samplers. *Water Resources Research*, 46:W03502.

Nelson, P. (1994). Permeability-Porosity Relationships in Sedimentary Rocks. *The Log Analyst*, 3:38–62.

Nolan, B. T. (2001). Relating nitrogen sources and aquifer susceptibility to nitrate in shallow ground waters of the United States. *Ground Water*, 39(2):290–299.

Oberröhrmann, M., A. Klotzsche, H. Vereecken, and J. van der Kruk (2013). Optimization of acquisition setup for cross-hole GPR full-waveform inversion using checkerboard analysis. *Near Surface Geophysics*, 11(2):197–209.

Ortiz, J. M. and C. V. Deutsch (2004). Indicator simulation accounting for multiple-point statistics. *Mathematical Geology*, 36(5):545–565.

Paasche, H., J. Tronicke, and P. Dietrich (2010). Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy c-means cluster analysis algorithm. *Geophysics*, 75(3):P11–P22.

Paasche, H., J. Tronicke, and P. Dietrich (2012). Zonal cooperative inversion of partially co-located data sets constrained by structural a priori information. *Near Surface Geophysics*, 10:103–116.

Pickel, A., J. D. Frechette, A. Comunian, and G. S. Weissmann (2015). Building a training image with Digital Outcrop Models. *Journal of Hydrology*, 531:53–61.

Rehfeld, K., J. Boggs, and L. Gelhar (1992). Field study of dispersion in a heterogeneous aquifer 3. Geostatistical analysis of hydraulic conductivity. *Water Resources Research*, 28(12):3309–3324.

Rehfeldt, K. R., P. Hufschmied, L. Gelhar, and M. Schaefer (1989). Measuring hydraulic conductivity with the borehole flowmeter. Technical report, Electric Power Research Inst., Palo Alto, CA (USA); Tennessee Valley Authority, Norris, TN (USA). Engineering Lab.; Massachusetts Inst. of Tech., Cambridge, MA (USA). Dept. of Civil Engineering.

Renard, P. and D. Allard (2013). Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51:168–196.

Riva, M., L. Guadagnini, A. Guadagnini, T. Ptak, and E. Martac (2006). Probabilistic study of well capture zones distribution at the Lauswiesen field site. *Journal of Contaminant Hydrology*, 88:92–118.

Robinson, D. A., S. B. Jones, J. M. Wraith, D. Or, and S. P. Friedman (2003). A Review of Advances in Dielectric and Electrical Conductivity Measurement in Soils Using Time Domain Reflectometry. *Physical Review*, 2:444–475.

Rogiers, B. (2013). *Multi-scale aquifer characterization*. PhD thesis, KU Leuven.

Ronayne, M. J., S. M. Gorelick, and J. Caers (2008). Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach. *Water Resources Research*, 44:W08426.

Roth, K., R. Schulin, H. Fluhler, and W. Attinger (1990). Calibration of time domain reflectometry for water-content measurement using a composite dielectric approach. *Water Resources Research*, 26(10):2267–2273.

Ruud, N. C., Z. J. Kabala, and F. J. Molz (1999). Evaluation of flowmeter-head loss effects in the flowmeter test. *Journal of Hydrology*, 224:55–63.

Scheibe, T. D. and Y. J. Chien (2003). An evaluation of conditioning data for solute transport prediction. *Ground Water*, 41(2):128–141.

Scheidt, C., C. Jeong, T. Mukerji, and J. Caers (2015). Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: Application to a turbidite reservoir case. *Geophysics*, 80(5):M89–M100.

Schmelzbach, C., J. Tronicke, and P. Dietrich (2011). Three-dimensional hydrostratigraphic models from ground-penetrating radar and direct-push data. *Journal of Hydrology*, 398:235–245.

Schneider, M. (1995). Charakterisierung der Variabilität der Permeabilität in einem quartären Grundwasserleiter. Master's thesis, RWTH Aachen.

Steelman, C. M. and A. L. Endres (2011). Comparison of Petrophysical Relationships for Soil Moisture Estimation using GPR Ground Waves. *Physical Review*, 10(1):270–285.

Straubhaar, J., P. Renard, G. Mariethoz, R. Froidevaux, and O. Besson (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–22.

Tan, X., P. Tahmasebi, and J. Caers (2014). Comparing training-image based algorithms using an analysis of distance. *Mathematical Geosciences*, 46(2):149–169.

Tesoriero, A. J., E. L. Inkpen, and F. D. Voss (1998). Assessing ground-water vulnerability using logistic regression. In *Proceedings for the Source Water Assessment and Protection 98 Conference, Dallas, TX*, Pp. 157–165.

Tillmann, A., A. Englert, Z. Nyari, I. Fejes, J. Vanderborght, and H. Vereecken (2008). Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the krauthausen test site using cone penetration test. *Journal of contaminant hydrology*, 95:57–75.

Topp, G., J. Davis, and A. Annan (1980). Electromagnetic determination of soil-water content - measurements in coaxial transmission-lines. *Water Resources Research*, 16(3):574–582.

Trefry, M. G. and C. Muffels (2007). Feflow: A finite-element ground water flow and transport modeling tool. *Ground Water*, 45(5):525–528.

Tronicke, J., K. Holliger, W. Barrash, and M. D. Knoll (2004). Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation. *Water Resources Research*, 40:W01519.

Van Den Eeckhaut, M., T. Vanwalleghem, J. Poesen, G. Govers, G. Verstraeten, and L. Vandekerckhove (2006). Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). *Geomorphology*, 76:392–410.

van der Kruk, J., N. Gueting, A. Klotzsche, G. He, S. Rudolph, C. von Hebel, X. Yang, L. Weihermüller, A. Mester, and H. Vereecken (2015). Quantitative multi-layer electromagnetic induction inversion and full-waveform inversion of crosshole ground penetrating radar data. *Journal of Earth Science*, 26(6):844–850.

Vanderborght, J. and H. Vereecken (2001). Analyses of locally measured bromide breakthrough curves from a natural gradient tracer experiment at Krauthausen. *Journal of Contaminant Hydrology*, 48:23–43.

Vereecken, H., U. Döring, H. Hardelauf, U. Jaekel, U. Hashagen, O. Neuendorf, H. Schwarze, and R. Seidemann (2000). Analysis of solute transport in a heterogeneous aquifer: the krauthausen field experiment. *Journal of Contaminant Hydrology*, 45:329–358.

Vienken, T. and P. Dietrich (2011). Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology*, 400:58–71.

Wada, Y., L. P. H. van Beek, C. M. van Kempen, J. W. T. M. Reckman, S. Vasak, and M. F. P. Bierkens (2010). Global depletion of groundwater resources. *Geophysical Research Letters*, 37(20). L20402.

Wainwright, H., J. Chen, D. Sassen, and S. Hubbard (2014). Bayesian hierarchical approach and geophysical data sets for estimation of reactive facies over plume scales. *Water Resources Research*, 50:4564–4584.

Webb, E. and M. Anderson (1996). Simulation of preferential flow in three-dimensional, heterogeneous conductivity fields with realistic internal architecture. *Water Resources Research*, 32(3):533–545.

Western, A. W., G. Böschl, and R. B. Grayson (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, 37(1):83–97.

WHO (2006). *World Health Organization - Protecting Groundwater for Health*. IWA Publishing, London.

Yang, X., A. Klotzsche, G. Meles, H. Vereecken, and J. van der Kruk (2013). Improvements in crosshole gpr full-waveform inversion and application on data measured at the boise hydrogeophysics research site. *Journal of Applied Geophysics*, 99:114–124.

Zheng, C., M. Bianchi, and S. M. Gorelick (2011). Lessons learned from 25 years of research at the MADE site. *Ground Water*, 49(5):649–62.

Zheng, C. and S. Gorelick (2003). Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. *Ground Water*, 41(2):142–155.

Zinn, B. and C. F. Harvey (2003). When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields. *Water Resources Research*, 39(3). doi:10.1029/2001WR001146.

ACKNOWLEDGEMENTS

Many thanks to all the great people at Jülich, Bochum, Stanford and elsewhere, whom I had the privilege to meet and work with during my PhD.

Foremost, I want to thank Andreas Englert who initiated this PhD project nearly four years ago and who has been supporting me and caring about every detail of this project ever since. Thanks for all the insights and ideas you shared with me.

I am equally grateful to Jan Vanderborght for supporting and advising me in Jülich. Thanks for always asking the right questions and for contributing the ideas that nobody else would have thought of.

Many thanks to Anja Klotzsche who invested so much time into me at the beginning of my PhD; and to Harry Vereecken and Jan van der Kruk for their valuable feedback and advice.

I would also like to thank Peter Dietrich and Thomas Vienken at UFZ for sharing their perspectives and for making the direct-push measurements at Krauthausen possible. This data certainly enriched my project.

I am deeply grateful to Jef Caers who warmly welcomed me at Stanford and provided me with all the support I could ever have wished for. Many thanks also to Philippe Renard, Julien Straubhaar and Gregoire Mariethoz for their ideas and for sharing the *impala* and *direct sampling* codes with me. Special thanks to Alessandro Comunian for providing feedback and guidance on the sequential 2-D MPS simulations.

This work certainly wouldn't have been possible without the support of Anja, Carly, Luka, Rainer, Manuel and Helko, whose assistance during the field measurements is highly appreciated.

It was a great pleasure working and socializing with the other PhD students and post-docs at Jülich and Stanford. Thanks to Asta, Khalid, Igor, Tao, Magda, Pasha, Luka, Manuela, Chris, Anna, Nina, Jan, Max, Sean, Kathi, Laura, Francois, Thomas, Kim, Emanuel, Celine, Thomas. Special thanks to Jannis for spending hundreds of lunches and coffee breaks over the last four years with me – has always been my greatest pleasure.

A special mention to my parents, my brother and my friends on whom I can always count. Finally, thanks to Nina for sharing everything with me and being the most important person in my life.

LIST OF PUBLICATIONS

i) The second Chapter of this thesis has been published as:

Gueting N., Klotzsche A., van der Kruk J., Vanderborght J., Vereecken H., Englert A. (2015), Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests, *Journal of Hydrology* 524, 680–695.

The GPR data were acquired in 2013 by Anja Klotzsche, Carly Hyatt and myself. The CPT data were available from a previous study [i.e. Tillmann et al., 2008]. The genuine idea for this paper was developed in several discussions among all authors. I did the analysis and I wrote the manuscript. Upon submission, the manuscript was peer-reviewed by two reviewers.

ii) The third Chapter of this thesis has been published as:

Gueting N., Vienken T., Klotzsche A., van der Kruk J., Vanderborght J., Caers J., Vereecken H., Englert A., (2017), High resolution aquifer characterization using crosshole GPR full-waveform tomography: Comparison with direct-push and tracer test data, *Water Resources Research*, 53, doi:10.1002/2016WR019498.

The GPR data were acquired in 2013 by Anja Klotzsche, Carly Hyatt and myself. The DPIL and DPST data were acquired in 2015 by Manuel Kreck and Helko Kotas from the UFZ Leipzig. The tracer test data were available from a previous study [i.e. Müller et al., 2010]. The genuine idea for this paper was developed in several discussions among all authors. Thomas Vienken did the quality-check of the DPIL and DPST raw data. I analyzed the data and I wrote the manuscript. Upon submission, the manuscript was peer-reviewed by three reviewers.

iii) The fourth Chapter of this thesis has been published as:

Gueting N., Caers J., Comunian A., Vanderborght J., Englert A., (2017), Reconstruction of three-dimensional aquifer heterogeneity from two-dimensional geophysical data, *Mathematical Geosciences*, doi:10.1007/s11004-017-9694-x

The genuine idea for this paper was developed in several discussions among all authors. I conducted the simulations and analyzed the results. I wrote the manuscript. Upon submission, the manuscript was peer-reviewed by two reviewers.

LIST OF FIGURES

LIST OF TABLES

Band / Volume 370
Performance of Plasma Facing Materials under Thermal and Plasma Exposure
I. Steudel (2017), XVI, 150 pp
ISBN: 978-3-95806-226-9

Band / Volume 371
The Impact of Transient Thermal Loads on Beryllium as Plasma Facing Material
B. Spilker (2017), XII, 134 pp
ISBN: 978-3-95806-227-6

Band / Volume 372
Analysis and Simulation of Macroscopic Defects in Cu(In,Ga)Se_2 Photovoltaic Thin Film Modules
B. Misic (2017), iv, 147 pp
ISBN: 978-3-95806-228-3

Band / Volume 373
Chemical and physical properties of sodium ionic conductors for solid-state batteries
M. Guin (2017), ix, 126 pp
ISBN: 978-3-95806-229-0

Band / Volume 374
Prediction of Oxidation Induced Life Time for FCC Materials at High Temperature Operation
R. Duan (2017), vi, 180 pp
ISBN: 978-3-95806-230-6

Band / Volume 375
Microstructure Evolution of Laves Phase Strengthened Ferritic Steels for High Temperature Applications
J. K. Lopez Barrilao (2017), XVI, 134 pp
ISBN: 978-3-95806-231-3

Band / Volume 376
Drying front formation in topmost soil layers as evaporative restraint
Non-invasive monitoring by magnetic resonance and numerical simulation
S. Merz (2017), xxii, 108 pp
ISBN: 978-3-95806-234-4

Band / Volume 377
Low Temperature Thin-Film Silicon Solar Cells on Flexible Plastic Substrates
K. Wilken (2017), 194 pp
ISBN: 978-3-95806-235-1

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment

Band / Volume 378
Dissolution Behaviour of Innovative Inert Matrix Fuels for Recycling of Minor Actinides
E. L. Mühr-Ebert (2017), xii, 164 pp
ISBN: 978-3-95806-238-2

Band / Volume 379
Charakterisierung und Modifizierung von Kupferoxid- und Kupfersulfid-Nanopartikeln für Dünnschichtsolarzellen
J. Flohre (2017), 141, iii pp
ISBN: 978-3-95806-241-2

Band / Volume 380
Einzelfaserkomposite aus Pulvermetallurgischem Wolfram-faserverstärktem Wolfram
B. Jasper (2017), v, 92, XVIII pp
ISBN: 978-3-95806-248-1

Band / Volume 381
Untersuchungen zur Deckschichtbildung auf $LiNi_{0,5}Mn_{1,5}O_4$-Hochvoltkathoden
Die Kathoden/Elektrolyt-Grenzfläche in Hochvolt-Lithium-Ionen-Batterien
K. Wedlich (2017), xvi, 157, xvii-xxvi pp
ISBN: 978-3-95806-249-8

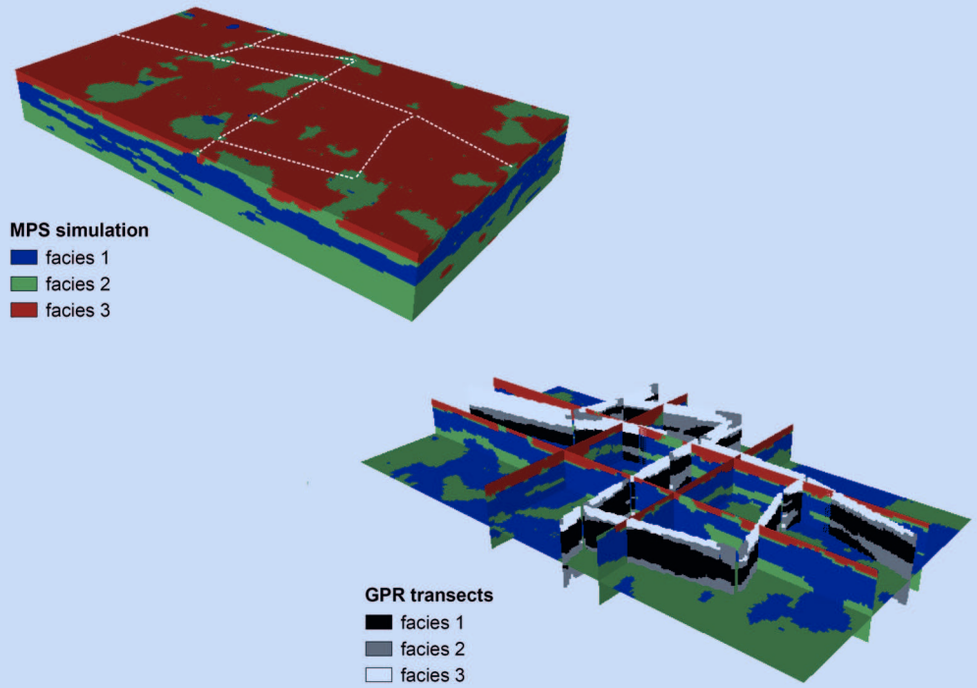
Band / Volume 382
Charakterisierung gradierter Eisen/Wolfram-Schichten für die erste Wand von Fusionsreaktoren
S. Heuer (2017), x, 235 pp
ISBN: 978-3-95806-252-8

Band / Volume 383
High resolution imaging and modeling of aquifer structure
N. Güting (2017), viii, 107 pp
ISBN: 978-3-95806-253-5

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

High resolution imaging and modeling of aquifer structure

Nils Güting



Energie & Umwelt / Energy & Environment
Band / Volume 383
ISBN 978-3-95806-253-5

JÜLICH FORSCHUNGSZENTRUM

Energie & Umwelt / Energy & Environment
Band / Volume 383
ISBN 978-3-95806-253-5

JÜLICH FORSCHUNGSZENTRUM