PAPER

Ensemble-based stochastic permeability and flow simulation of a sparsely sampled hard-rock aquifer supported by high performance computing

Johanna Bruckmann [1,2,3] · Christoph Clauser [1,2,3]

Abstract

Calibrating the heterogeneous permeability distribution of hard-rock aquifers based on sparse data is challenging but crucial for obtaining meaningful groundwater flow models. This study demonstrates the applicability of stochastic sampling of the prior permeability distribution and Metropolis sampling of the posterior permeability distribution using typical production data and measurements available in the context of groundwater abstraction. The case study is the Hastenrather Graben groundwater abstraction site near Aachen, Germany. A three-dimensional numerical flow model for the Carboniferous hard-rock aquifer is presented. Monte Carlo simulations are performed, for generating 1,000 realizations of the heterogeneous hard-rock permeability field, applying Sequential Gaussian Simulation based on nine log-permeability values for the geostatistical simulation. Forward simulation of flow during a production test for each realization results in the prior ensemble of model states verified by observation data in four wells. The computationally expensive ensemble simulations were performed in parallel with the simulation code SHEMAT-Suite on the high-performance computer JURECA. Applying a Metropolis sampler based on the misfit between drawdown simulations and observations results in a posterior ensemble comprising 251 realizations. The posterior mean log-permeability is −11.67 with an uncertainty of 0.83. The corresponding average posterior uncertainty of the drawdown simulation is 1.1 m. Even though some sources of uncertainty (e.g. scenario uncertainty) remain unquantified, this study is an important step towards an entire uncertainty quantification for a sparsely sampled hard-rock aquifer. Further, it provides a real-case application of stochastic hydrogeological approaches demonstrating how to accomplish uncertainty quantification of subsurface flow models in practice.

Keywords Numerical modeling · Hydraulic properties · Heterogeneity · Fractured rocks · Germany

Introduction

Aquifers often exhibit heterogeneous hydraulic properties (e.g. porosity and hydraulic conductivity or permeability) which are crucial for characterizing subsurface flow and transport. Especially in hard-rock aquifers, hydraulic properties tend to be strongly spatially variable. At the same time, only a limited amount of direct measurements or indirect data is available from boreholes or surface geophysics for characterizing subsurface flow. Yet, assigning adequate hydraulic properties to hydrological units is crucial for a proper groundwater flow and transport assessment by numerical flow models. With increased computational power and advanced numerical methods, numerical modeling has become an established tool in aquifer research (Prickett 1975; Carrera et al. 2005). It enables the understanding of the aquifer in three dimensions, beyond one- or two-dimensional measurements, as well as accounting for spatial heterogeneity and uncertainty (e.g. Marsily et al. 2005).

✉ Johanna Bruckmann
 jbruckmann@eonerc.rwth-aachen.de

1 Institute for Applied Geophysics and Geothermal Energy, E.ON Energy Research Center, RWTH Aachen University, Mathieustr. 10, 52074 Aachen, Germany

2 Centre for High Performance Scientific Computing in Terrestrial Systems (HPSC TerrSys), c/o Forschungszentrum Jülich GmbH in the Helmholtz Association, 52425 Jülich, Germany

3 Jülich-Aachen Research Alliance (JARA-HPC), RWTH Aachen University, 52056 Aachen, Germany

This study addresses the challenge of calibrating a hard-rock aquifer flow model in a case study of an aquifer used for drinking water production, based on available production data and auxiliary data recorded in this context. It demonstrates the need for incorporating the spatial variability of hydraulic parameters in order to generate realistic models, which reproduce the data trend and the hydraulic behavior. To this end, this paper provides a case study, which shows and discusses the application of the Monte Carlo approach to a sparsely sampled hard-rock aquifer in order to investigate the heterogeneity and prior uncertainty of the aquifer permeability. Here, the aim is to demonstrate how the inverse parameterization of flow models, in this case the permeability field, can be improved based on a limited amount of data, which is often the case in practice. In contrast, many studies and research projects benefit from extensive measurement campaigns or work with synthetic data sets that are tailored to their needs in terms of input data availability (e.g. Kurtz et al. 2014; Li et al. 2012).

Stochastic approaches based on the Monte Carlo method for estimating subsurface model parameters and their uncertainties are well established in theory (e.g. Tarantola 2005; Zhang 2002; Rubin 2003; Oliver et al. 2008; Zhou et al. 2014). However, they are not applied commonly to subsurface flow modeling in the fields of, e.g. hydrogeology or geothermics (e.g. Dagan 2002; Renard 2007; Rubin et al. 2018). Only recently, Rubin et al. (2018) pointed out and discussed this lack of practical applications of stochastic hydrogeology once again, which shows that the problem still exists. They identify lack of data and software issues as two of the main hurdles.

Stochastic Monte Carlo simulations are one option for addressing subsurface parameter heterogeneity and uncertainty in numerical models. They are based on a statistical analysis of a large number of randomly created, equally likely forward simulations (e.g. Zhang 2002; Gelhar 1993). The spatially variable hydraulic properties such as permeability, are treated as random variables that can be generated by a stochastic spatial process. If the hydraulic property (i.e., permeability) is measured on a scale, which exceeds the scale of average fracture spacing it can be defined reasonably over a continuum (Neuman and Depner 1988). The stochastic modeling of the permeability, based on geostatistical parameters which honor the spatial trend in borehole data, allows considering the heterogeneity between fractures and matrix and the spatial variability of permeability due to main fracture orientations in a single continuum. Neuman and Depner (1988) and Tsang et al. (1996) were the first to demonstrate the applicability of this stochastic-continuum concept to fractured hard-rocks. The continuum model approach avoids the large data requirements of discrete or equivalent fracture model approaches which model fracture networks explicitly or by using upscaling approaches (e.g. Chen et al. 2018; Neuman 2005; Baca et al. 1984; Warren and Root 1963).

Main stochastic inversion approaches for model calibration and uncertainty assessment are sampling and optimization approaches (e.g. Zhou et al. 2014; Linde et al. 2015). The sampling approach aims at sampling the posterior distribution based on the Markov Chain Monte Carlo (McMC) technique (e.g. Tarantola 2005). It means that the evaluation of each sample depends only on the state of the previous sample in the chain. It is relatively easy to implement in a post-processing workflow, but sampling methods require the evaluation of a thousand or more forward realizations, which is computationally expensive and results is a large amount of data. The simplest sampler is the rejection algorithm that compares the likelihood of one sample to the likelihood of the previous sample and only accepts it as member of the posterior if the likelihood is not deteriorated. This sampling procedure is accurate but requires a large number of prior samples in order to retain a statistically sufficient number of accepted samples in the posterior. In contrast, the Metropolis algorithm uses a probabilistic transition operator for deciding whether to accept a sample as member of the posterior or to reject it. Depending on the design of the decision operator, it samples the posterior more efficiently up to a certain error level (e.g. Tarantola 2005; Mosegaard and Tarantola 1995; Mariethoz et al. 2010).

The optimization approach is based on minimizing an objective function, which is the mismatch between simulation results and observation data. Stochastic optimization methods such as the gradual deformation method (Hu 2000) or the probability perturbation method (Caers 2003), preserve the prior model and match observation data. Another optimization approach is the Ensemble Kalman filter (Evensen 2003), which continuously updates the ensemble of model realizations as transient data become available for optimizing the history match. The prior model structure is not preserved by this data assimilation technique.

Generally, both approaches differ in their main purpose. While optimization methods aim at history-matching and finding the optimal model parameterization, the main focus of sampling approaches lies in the realistic uncertainty quantification (Park et al. 2013). The interested reader is referred to Zhou et al. (2014) and Linde et al. (2015) and references therein for detailed reviews of stochastic inverse methods in hydrogeology.

Recent studies demonstrated the successful application of Bayesian Evidential Learning (BEL) to hydrogeological applications for uncertainty quantification and prediction (Hermans et al. 2018, 2019). This recently developed prediction-based approach derives a direct relationship between data and prediction based on the prior ensemble. First applications showed promising results for estimating the posterior distribution of model states, but it has not been applied for estimating parameter distributions yet.

This study assesses the heterogeneous hard-rock aquifer permeability at the Hastenrather Graben study site, near Aachen, Germany, using a massive Monte Carlo approach (e.g. Zhang 2002) which is composed of three main steps. First, the prior permeability distribution is defined by geostatistical variogram analysis of sparse permeability data from wells, and sampled using the SGSim algorithm for generating an ensemble of 1,000 unconditional realizations of the aquifer permeability field. Second, the forward problem is solved for each realization, which is the simulation of hydraulic head drawdown during a long-term pumping test. Third, the resulting ensemble is analyzed and statistical post-processing is performed. The third step includes verification of the prior with state data and analysis of the prior uncertainty. Subsequently, a Metropolis sampling technique is applied for converging to the posterior probability distribution as proposed by Mosegaard and Tarantola (1995). The statistical post-processing yields information on the uncertainty of the aquifer permeability and associated uncertainty of simulated hydraulic heads for the prior ensemble as well as the posterior ensemble.

This study focuses on the stochastic investigation of the hard-rock aquifer permeability and its heterogeneity, because it is assumed the dominant parameter controlling the aquifer response to production and the parameter with the largest prior range of uncertainty. The aquifer response is expected to be less sensitive to other parameters such as porosity, specific storage or boundary conditions. This justifies limiting the stochastic analysis to permeability for reducing the complexity of the stochastic model. The assumption is justified by geological expertise. Alternatively, global sensitivity analysis would provide a sophisticated analysis of the model's sensitivity to different parameters (e.g. Scheidt et al. 2018; Hermans et al. 2019).

In the following, this paper describes the application of these steps to the Hastenrather Graben hard-rock aquifer model in detail and discusses the approach in light of data scarcity. The focus is on demonstrating the applicability of the approach as a practical way of assessing the heterogeneous permeability distribution and associated uncertainties in a hard-rock aquifer despite sparse database from a practitioner's point of view. The resulting prior distribution may be used for more sophisticated stochastic inversion approaches, subsequently.

As the computation of several hundreds to thousands of realizations of reservoir scale transient flow simulations is computationally expensive, the Monte Carlo approach requires the use of a parallelized simulation code and of high-performance computing (HPC) resources or of cloud computing services. HPC resources are usually accessible for researchers from academia and industry via national or international centers such as PRACE (Partnership for advanced computing in Europe). Cloud computing services might be the right choice for practitioners from industry for performing computationally demanding simulations (e.g. Hayley 2017). The presented MC simulations are performed with the parallel simulation code SHEMAT-Suite on the supercomputer JURECA. The section 'Parallelization and parallel performance' provides information on the code's parallelization strategy and its parallel performance.

Hard-rock aquifer case study

The studied catchment area, the Hastenrather Graben, is located 15 km east of Aachen, Germany (Fig. 1). Geologically, it lies at the transition from the Rhenish Massif (to the south) to the Lower Rhine Embayment (to the north). The NNW–SSE trending graben structure holds a folded Carboniferous hard-rock aquifer limited it by the Sandgewand Fault to the SW and the Omerbach Fault to the NE. Within the graben center, thrusted Palaeozoic nappes as well as Cenozoic sediments cover the hard-rock aquifer (Fig. 1).

Since the 1950s, the Hastenrather Graben catchment area has been used for producing drinking water from this hard-rock aquifer. Common hydrogeological exploration and production data are available such as hydraulic conductivities inferred from pumping tests, drawdown curves from production tests and continuous pressure transducer records from observation wells (see Burs et al. 2016). This makes this study area well suited for demonstrating the approach of an ensemble-based stochastic investigation using data available from a commercial hydrogeological aquifer operation. Besides, the Carboniferous hard-rocks are of interest as potential geothermal energy reservoir rocks in Northwest Europe.

The numerical model is based on a comprehensive three-dimensional (3D) structural and conceptual model of the Hastenrather Graben presented by Burs et al. (2016) (Fig. 1c). Readers are referred to this reference (Burs et al. 2016) and to the references therein for additional and more detailed information on the geological and hydrogeological background, and the structural and conceptual model of the study area. The following paragraph will only point out some aspects of the hydrogeological model, which are crucial for understanding the numerical model set-up.

The Kohlenkalk formation is the rock unit from which water is produced; however, the numerical model jointly investigates the Walhorn and Stolberg Layers and the Kohlenkalk as one hard-rock aquifer unit. The hydrogeologic conceptual model of the Hastenrather Graben catchment area reveals that both units are connected hydraulically, caused by zones of rock disintegration and fractures. Pumping test reactions monitored in the Walhorn and Stolberg Layers, and hydrochemical samples from the Walhorn and Stolberg Layers with an increased amount of bicarbonate are proof of the hydraulic connection to the underlying Kohlenkalk. Additionally, hydraulic conductivities in both units are in the

Fig. 1　**a** Location of the study area in western Germany, near Aachen. **b** Location of the Hastenrather Graben catchment area in a regional geo-tectonic context (modified after Chatziliadou 2009). **c** 3D geological model of the Hastenrather Graben catchment area (adapted from Burs et al. 2016) with locations of used wells and indication of respective data types (circle, diamonds and crosses). The Carboniferous hard rocks constitute the main aquifer
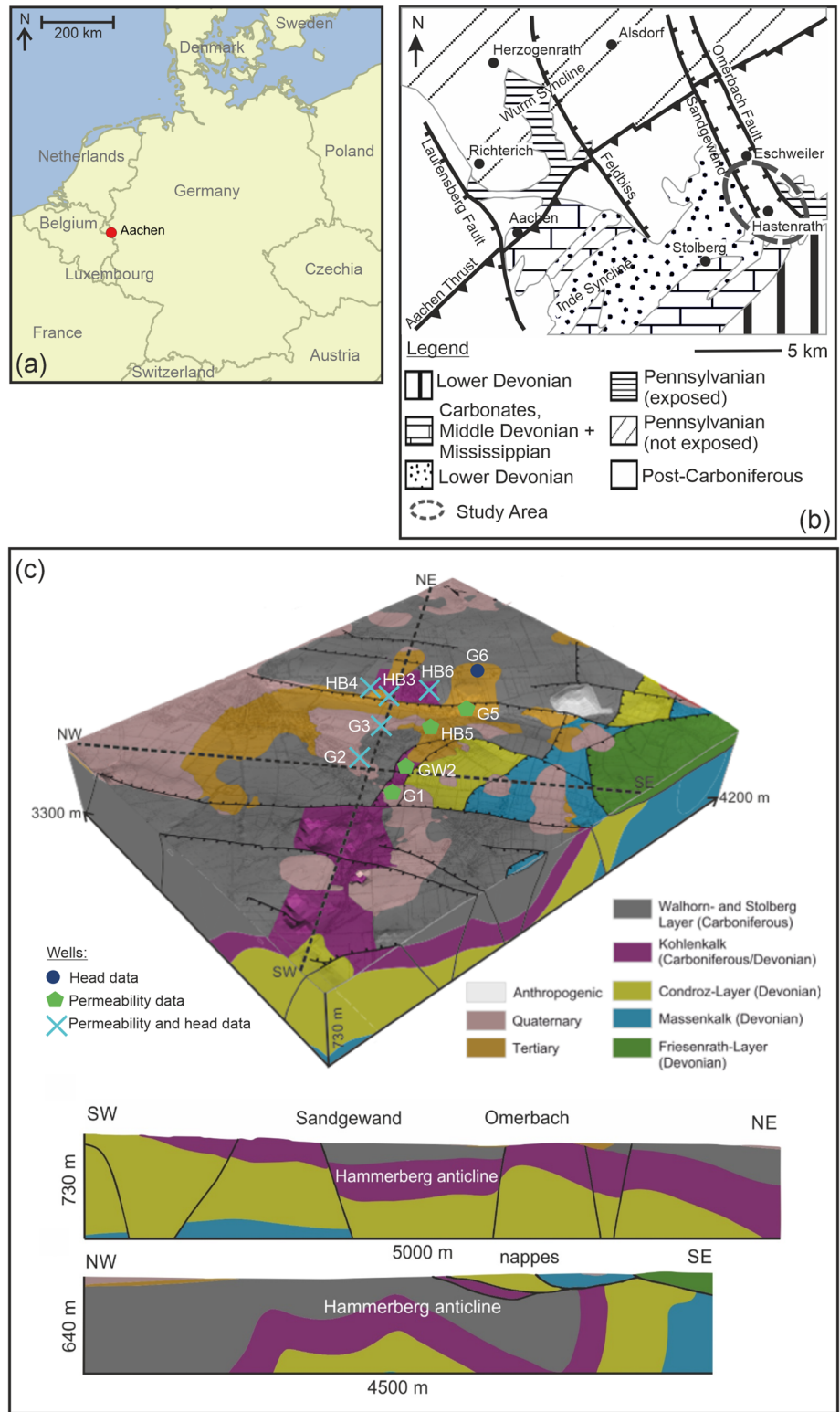
same order of magnitude; thus, it is reasonable to model them jointly as one aquifer unit, which has also practical reasons, because it enlarges the database for the aquifer. If the Kohlenkalk was considered individually, only four permeability data points and no monitoring wells would be available, which would prohibit the presented investigations.

Furthermore, the conceptual model assumes a hydraulic connection between the hard-rock aquifer and the overlying

sedimentary aquifer in some parts of the catchment area (Burs et al. 2016). Separating clay layers between the two aquifers were found occasionally in driving core samples, which lead to variably confined and unconfined conditions in the hard-rock aquifer. Since no comprehensive information on the distribution of the clay layers is available, they are not included in the geological and hence in the numerical model.

Besides some unconfined parts of the aquifer, the conceptual model identified the area around wells HB5 and G5 (Fig. 1c) as a separate hydraulic unit with a strong direct connection between the wells and partly artesian conditions. The small-scale geological situation around those wells, which may cause this singular hydraulic condition, is unclear.

Here, the aim is to model the permeability and hydraulic heads of the main hard-rock aquifer system, which is defined as the confined part of the joint Kohlenkalk and Walhorn and Stolberg Layers. Therefore, piezometric data from wells, which drill into areas with different hydraulic behavior, are not considered for model calibration. All available permeability data are considered from each well drilling into either the Walhorn and Stolberg Layers or the Kohlenkalk Formation. Figure 1c illustrates the location of the wells and respective usable data types.

The simulation code

Originally, the Simulator for HEat and MAss Transport (SHEMAT; Clauser 2003) was written as a forward simulation code for simulating reactive fluid flow and heat and species transport in geothermal and hydrogeological applications. During the last decade, the code has been extended and developed into a software package for solving forward as well as inverse problems for parameter estimation (Rath et al. 2006). Moreover, it has been transformed from a serial into a parallel application (Wolf 2011). This extended code package is called SHEMAT-Suite. The following sections describe in more detail those features of the code, which are used in the presented study.

The forward model

SHEMAT-Suite simulates groundwater flow by solving the partial differential equation for flow through porous media, which is a combination of the equation of continuity and Darcy's law:

$$\nabla \cdot \left[\frac{\rho_f g}{\mu_f} k \nabla h \right] + Q = S_s \frac{\partial h}{\partial t} \tag{1}$$

with hydraulic head h (m), permeability k (m^2), source term Q (m^3 s^{-1}), gravity g (m s^{-2}), pore fluid density ρ_f (kg m^{-3}), dynamic pore fluid viscosity μ_f (Pa s), time t (s), and specific storage coefficient S_s (m^{-1}). Equation (1) states that the

divergence of the specific discharge (i.e., the balance of total inflow and outflow) equals the sum of storage and sources and sinks. Darcy's law defines specific discharge v (m s^{-1}) as proportional to the product of the head gradient, fluid properties, gravity and rock permeability:

$$\nu = -\frac{\rho_f g}{\mu_f} k \nabla h \tag{2}$$

The specific storage coefficient describes how storage varies with hydraulic head. It is defined as the variation of water volume V_W per unit volume V_{tot} with that of hydraulic head:

$$S_s = \frac{1}{V_{tot}} \frac{\partial V_w}{\partial h} \tag{3}$$

This can be described by measurable physical properties as follows:

$$S_s = \gamma_w \left(\beta_m + \phi \, \beta_w \right) \tag{4}$$

where $\gamma_w = \rho_f \, g$ is the specific weight of water (N m^{-3}), ϕ is the porosity (−), β_m is the rock matrix compressibility (m^2 N^{-1}) and β_w is the compressibility of water (m^2 N^{-1}).

In these equations, the physical properties of rock matrix and fluid are kept constant as the study narrows down to simulating groundwater flow. SHEMAT-Suite uses the finite difference approach for solving the coupled equations. More details on the numerical approaches such as used linear solvers or time stepping schemes can be found in Clauser (2003).

Monte Carlo approach and random field generation

The classical Monte Carlo approach consists of three main steps (e.g. Zhang 2002) - (1) probabilistic generation of a prior ensemble of parameter field realizations based on geostatistical parameters, (2) numerical simulation of the deterministic forward model for each realization, and (3) statistical analysis of the ensemble's parameters and state variables. This approach is incorporated directly in the numerical simulation software package SHEMAT-Suite.

The Sequential Gaussian Simulation algorithm (SGSim; Deutsch and Journel 1998) is implemented for random field generation (1). It generates randomly equally likely log$_{10}$ permeability (from here on 'log-permeability') fields with respect to the spatial distribution of the log-permeability which is characterized by a probability distribution and direction-dependent correlation lengths and variogram parameters. The forward model (2) for simulating the hydraulic head is described in section 'The forward model'.

The resulting prior ensemble of log-permeability fields and corresponding forward flow simulations for each permeability field realization can be characterized by statistical moments

such as mean and standard deviation. The forward realizations are analyzed in terms of hydraulic head values. Their standard deviation is a measure of prior uncertainty.

Subsequently, in a post-processing step a sampling algorithm derived from the Metropolis sampler (Mosegaard and Tarantola 1995) is applied for sampling the posterior log-permeability distribution of the stochastic model and receiving a posterior ensemble of hydraulic head simulations.

Proceeding from the geostatistically generated prior ensemble, the sampling algorithm works as follows:

1. Randomly draw a model realization m_i from the prior ensemble and evaluate its misfit $S(m_i)$.
2. Iterate over i, until prior ensemble size is reached:

   a. Draw another realization $m*$ from the prior and evaluate its misfit $S(m*)$
   b. Accept the new realization $m*$ with probability $\alpha(m_i, m*)$
   c. If the new realization $m*$ is accepted, then m_{i+1} becomes $m*$, otherwise $m_{i+1} = m_i$

The acceptance probability is

$$\alpha\left(m_i, m*\right) = \begin{cases} \exp\left(-\dfrac{S(m*)-S(m_i)}{s^2}\right), & \text{if } S\left(m*\right) > S(m_i) \\ 1, & \text{if } S\left(m*\right) \leq S(m_i) \end{cases} \tag{5}$$

The error term s is a combination of measurement error and model errors. Model errors stem from uncertainty in boundary conditions, conceptual error, the uncertainty in the variogram model, discretization errors and further error sources. The misfit S is evaluated in terms of the equally weighted root mean square error (RMS) for the complete time series (i.e., drawdown curves), the difference between simulated and observed heads at each time step j, $h_{i,j}^{\text{sim}}$ and h_j^{obs}, respectively:

$$\text{RMS}_i = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(h_{i,j}^{\text{sim}}-h_j^{\text{obs}}\right)^2} \tag{6}$$

where n is the total number of time-steps.

This means that a realization is accepted for the posterior ensemble if it improves the quality of fit and accepted with a certain probability (Eq. 5) if it deteriorates the quality of fit. The resulting ensemble of all accepted realizations converges to the posterior distribution and yields the posterior uncertainty of the stochastic model.

Parallelization and parallel performance

Computationally expensive parts of the forward flow and transport model in SHEMAT-Suite use the OpenMP

programing paradigm for shared-memory parallel programming (Wolf 2011). This way, simulations can run in parallel using a moderate number of concurrent threads (e.g. Chapman et al. 2008); however, an optimal scalability on modern high-performance computers requires a parallelization for distributed-memory systems using a distributed-memory programming paradigm such as the Message Passing Interface (MPI; Snir et al. 1998; Gropp et al. 1998). Therefore, an MPI-parallel or hybridly parallelized code is essential for stochastic simulations of computationally expensive and memory intensive 3D models. Consequently, SHEMAT-Suite uses MPI parallelization for the sequential stochastic field generation. Each realization of an ensemble in the Monte Carlo simulation can be computed in parallel, distributed over a number of processes - for example, 504 processes (21 nodes) on the JURECA (Jülich Research on Exascale Cluster Architectures) supercomputer at Jülich Supercomputing Center (JSC) computed a Monte Carlo run with 504 realizations.

The usage of supercomputers requires the applied software to be scalable, i.e. to use the supercomputer resources efficiently. For memory-bound problems such as Monte Carlo simulations, it means that when the workload is increased in direct proportion to the number of processes, the application's run time should ideally stay constant. This is called weak scaling. A weak scaling test on the JURECA architecture demonstrates the code's weak scaling efficiency. Here, each node comprises two 12 core Intel Xeon E5–2680 v3 Haswell CPUs. In the weak scaling test, the ensemble size increases simultaneously with the number of cores from 24 to 576. The weak scaling efficiency E_{weak} scaled to 24 cores for p cores is given by

$$E_{\text{weak}_p} = \frac{t_{24}}{t_p} \cdot 100(\%) \tag{7}$$

where t_p is CPU time on p cores.

The aquifer model described in section 'Numerical model setup' comprises around 7.5 million grid cells and requires 10 GB memory for one realization. The Monte Carlo simulation with SHEMAT-Suite scales well for up to 600 processes (Table 1; Fig. 2): weak scaling efficiency is above 90% for 600 parallel processes, where memory demand limits the parallel simulation. The memory demand of the simulation increases with increasing number of realizations, until it exceeds the memory available on the JURECA nodes at an ensemble size of 600. Simulation time for one single realization or 600 parallel realizations is around 75 min meaning that 600 serial realizations would consume 1 month of computing time. An additional scaling test with a smaller synthetic model (around 10^5 grid cells, 182 MB memory per realization) shows a satisfying scaling behavior of over 70% for up to 2,000 processes

Table 1 Weak scaling efficiency E_{weak} of SHEMAT-Suite Monte Carlo simulations on Intel Xeon E5–2680 v3 Haswell CPUs of the JURECA system at JSC. This test was performed with 24 model realizations per node. Test model: 10^5 cells, 182 MB memory per realization. Aquifer model: $7.5 \cdot 10^6$ cells, 10 GB memory per realization

Model type	No. of nodes	No. of processes	CPU time (s)	E_{weak} (%)
Test model	1	24	354	100
	28	672	389	91
	42	1,008	399	89
	84	2,016	473	75
Aquifer model	1	24	4,436	100
	2	48	4,231	105
	4	96	4,557	97
	8	192	4,496	99
	16	384	4,828	92
	20	480	4,888	91
	24	576	4,837	92
	25	600	4,827	92
	26	624	4,901	91

(Fig. 2). Deviation from the ideal scaling can be caused by overhead associated with synchronization or hardware issues such as cache effects or memory bandwidth.

A prospective MPI parallelization of the forward simulation involving domain decomposition may prevent the observed limitation of parallel performance caused by memory demand of the model. For large model domains with big parameter fields, the memory needed for one forward flow model can exceed the memory available on one compute core. An MPI parallel forward simulation using domain decomposition would spread the model domain over several compute cores.
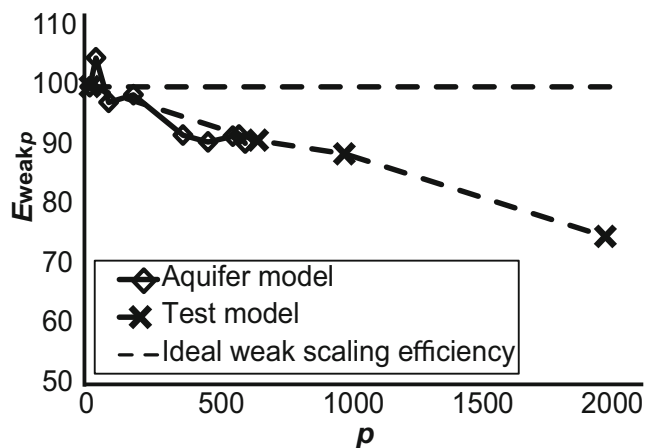
Fig. 2 Weak scaling efficiency E_{weak} of Monte Carlo simulations on Intel Xeon E5–2680 v3 Haswell CPUs of the JURECA system at JSC versus number of cores p. In addition, see Table 1 for test results and information on the models

Numerical model setup

The 3D geological model of the study area (Burs et al. 2016) is gridded in finite volume blocks for the numerical simulations. A subset of 3,623 m to the east, 2,870 m to the north and 720 m depth is divided into a structured, hexahedral grid with a uniform cell size of 10 m × 10 m × 10 m. The resulting grid has 362 cells in x-direction, 287 cells in y-direction, and 72 cells in z-direction, resulting in a total number of approximately 7.5 million grid cells. The model's reference depth z_0 = 0 m corresponds to a depth of –500 m above mean sea level (m asl). The model consists of six geological units (Table 2). The hard-rock aquifer system comprises the Kohlenkalk Formation and the Walhorn and Stolberg Layers (Fig. 1c, section 'Hard-rock aquifer case study'). Six monitoring points are placed within the aquifer, according to locations and filter depths of monitoring and production wells at the study site (Fig. 1c).

Initial steady-state model

A calibrated deterministic steady-state model with homogeneous zonation and confined conditions serves as the initial model for subsequent transient and stochastic simulations (Fig. 3). Each rock zone is characterized by average hydraulic parameters representative for the study area and inferred from measurements or the literature (see Burs et al. 2016). Permeability was calculated from hydraulic conductivity data for conditions at 10 °C. Hydraulic conductivities of the aquifer units result from several pumping tests in the area, which were commissioned by the local water company or conducted by researchers (see Burs et al. 2016 for details). No anisotropy is assumed between vertical and horizontal permeability.

The steady-state model represents the aquifer under natural conditions without any pumping activities. It provides the initial conditions for further transient simulations (Fig. 3).

Model boundaries coincide with no-flow boundaries in the east and west — in the east it is the catchment boundary and in the southwest it is the hydraulically active Sandgewand Fault and the catchment boundary. At the southern and northern boundaries, Dirichlet boundary conditions are used for adjusting the corresponding inflow and outflow. Their constant head values depend on the date, which the steady-state model represents, because they were inferred from hydraulic head data at a certain time. The situation at 16 July 2006 represents natural conditions before the beginning of water withdrawal during the long-term production test.

Head contour lines are constructed from piezometer data from six available wells which have been drilled into the main hard-rock aquifer system and are located in the model center (see Fig. 1c). These contour lines are extrapolated linearly to the northern and southern model boundaries, according to their mean northward flow gradient of 0.004. Extrapolation

Table 2 Average porosity and permeability values for the six model units as inferred from measurements or the literature (Burs et al. 2016). Arithmetic means were calculated where measurement data were available. Hard-rock aquifer units are in italic type

Model unit	Matrix porosity (−)	Permeability (m²)	Hydraulic conductivity (m s⁻¹)
Quaternary sediments	0.10	$1.32 \cdot 10^{-10}$	$1.00 \cdot 10^{-3}$
Tertiary sediments (Köln Layer)	0.30	$6.61 \cdot 10^{-12}$	$5.00 \cdot 10^{-5}$
Walhorn and Stolberg Layers	*0.01*	*$1.51 \cdot 10^{-12}$*	*$1.14 \cdot 10^{-5}$*
Kohlenkalk	*0.08*	*$1.10 \cdot 10^{-11}$*	*$8.31 \cdot 10^{-5}$*
Condroz Layer	0.02	$1.01 \cdot 10^{-12}$	$7.60 \cdot 10^{-6}$
Massenkalk	0.08	$1.06 \cdot 10^{-11}$	$8.00 \cdot 10^{-5}$
Friesenrath Layer	0.01	$1.32 \cdot 10^{-14}$	$1.00 \cdot 10^{-7}$

Table 3 Steady-state model calibration results for constant head boundaries of 154 m asl in the north and 165 m asl in the south at six monitoring wells in the hard-rock aquifer system. The root mean square error is 0.34

Well	Observed head (m asl)	Simulated head (m asl)	Head difference (m)
G2	158.76	158.50	−0.26
G3	158.30	158.33	0.03
G6	159.14	158.42	−0.72
HB3	157.55	157.55	0.00
HB4	156.90	157.20	0.30
HB6	158.20	158.29	0.09

This underlines the need for incorporating a heterogeneous permeability into the model.

results in Dirichlet boundaries of 154 m asl in the north and 165 m asl in the south. There is no other flow boundary such as local recharge considered in the model since there is no evidence for it — see also Burs et al. (2016) for more details on the hydrogeological conceptual model.

The steady-state simulation results in an RMS error of 0.34 for confined conditions (Table 3; Fig. 4) for the calibration at the six monitoring wells. The northward flow gradient of 0.004 yields a volume flow of $2.6 \cdot 10^{-1}$ m³ s⁻¹ or $8.2 \cdot 10^{6}$ m³ year⁻¹ through the model domain in the confined case. Figure 5 shows contour lines of the simulated head and model boundaries in a slice at 615 m model reference depth for confined conditions.

The head at well G2 is not matched well by the model because permeability measured at this location is one order of magnitude lower than the average permeability which is used in this deterministic steady-state model. The even higher head deviation at well G6 is probably also caused by a permeability deviation, but no permeability data is available here.

Transient production test model

The inflow rates at the southern and outflow rates at the northern boundary resulting from steady-state model calibration are used as Neumann boundary conditions in subsequent transient simulations. The specific storage coefficient is constant for each rock zone and computed from porosities (Table 2) according to Eq. (4), assuming default rock and fluid compressibility of 10^{-10} Pa⁻¹ and $5 \cdot 10^{-8}$ Pa⁻¹ respectively. This results in specific storage coefficients of $4.02 \cdot 10^{-5}$ for the Walhorn and Stolberg Layers and $5.89 \cdot 10^{-6}$ for the Kohlenkalk.

The transient model simulates a multiple step production test, performed by the water supplier in the Hastenrather Graben between July 2006 and November 2006. Table 4 lists the varying production rates within the four production wells HB3, HB4, HB5, and HB6 during the different stages. The simulation of the production test starts 3 days before the beginning of stage 0. This way, the model reaches a quasi-steady-state initial condition before the production starts.

Fig. 3 Numerical model showing initial steady-state conditions and lateral boundary conditions (see Fig. 1c for color code of model units and location of wells). The y-axis points towards the north
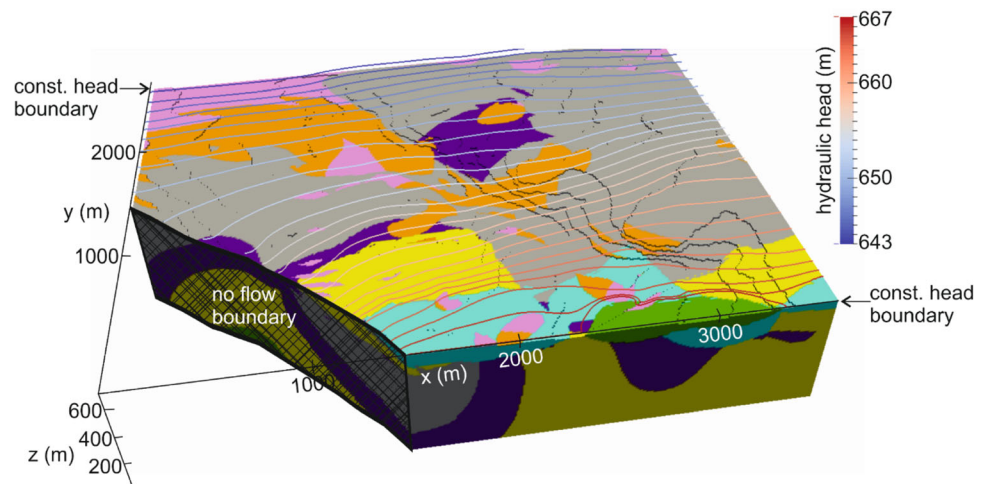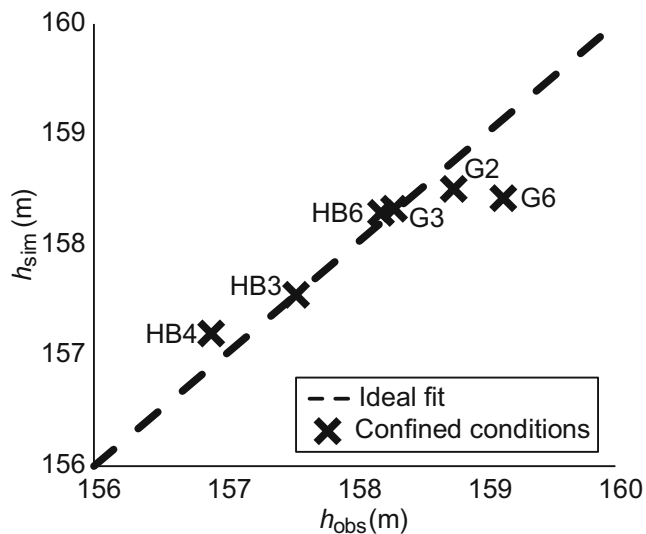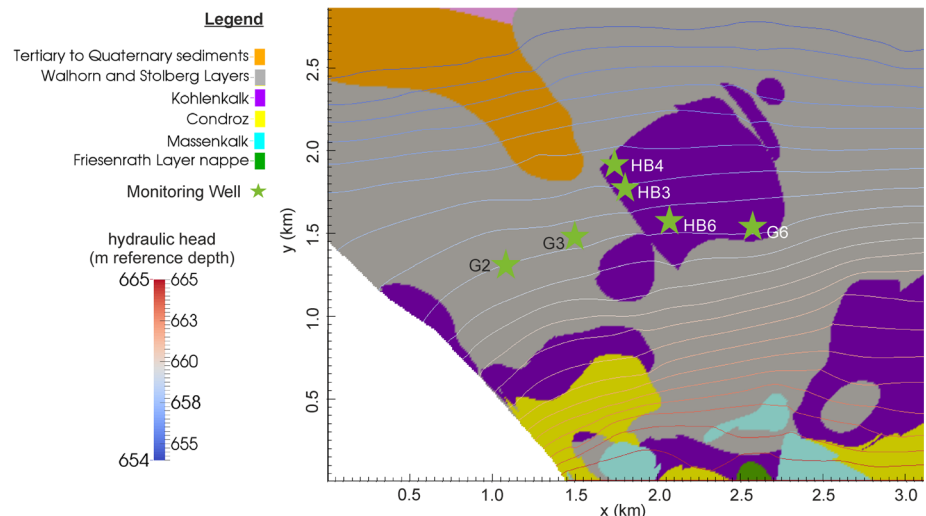
Fig. 4 Observed hydraulic heads h_{obs} versus simulated hydraulic heads h_{sim} at six monitoring wells within the hard-rock aquifer system of the steady-state model with mean permeabilities. Constant heads at the northern and southern model boundaries are 654 m above reference depth and 665 m above reference depth, respectively

Table 4 Different production stages and pumped volume flowrate of the different wells during the production test in 2006

Stage	Duration (days)	Volume flowrate (m^3 s^{-1})			
		HB3	HB4	HB5	HB6
0	45	0.025	0.000	0.000	0.000
I	14	0.025	0.014	0.000	0.000
II	14	0.025	0.014	0.000	0.008
III	7	0.025	0.014	0.000	0.013
IV	7	0.025	0.019	0.000	0.013
V	14	0.025	0.019	0.007	0.013
Recovery	15	0.025	0.000	0.000	0.000

This is necessary especially for models with a heterogeneous permeability field where the initial situation differs from that of the homogeneous permeability scenario. Production rates scaled to the cell size (i.e., multiplied by 10^{-1}) are applied as time-dependent Neumann boundary conditions in cells where the filter sections are located. The pumping test period of 116 days is divided into 46 time steps. When a new pumping stage starts, the discretization is half a day for 1 day, and 1 day for the next day. During the distinct pumping levels, the discretization is around 4 days.

Stochastic model

For addressing parameter heterogeneity within the hard-rock aquifer, the spatially variable permeability is treated as a

random variable that can be generated by a stochastic process in space. The stochastic modeling of permeability is based on geostatistical parameters which honor measurement data. This allows considering the heterogeneity between fractures and matrix and the spatial variability of permeability due to main fracture orientations in a single continuum (e.g. Neuman and Depner 1988; Tsang et al. 1996). Geostatistical parameters of the log-permeability values are required for generating a random log-permeability field by Sequential Gaussian Simulation (SGSim; Deutsch and Journel 1998). In total, nine log-permeabilities are available from pumping tests within the hard-rock aquifer (Burs et al. 2016) which can be used for a variogram analysis. They clearly exhibit an increasing spatial trend from southwest (SW) to northeast (NE), which agrees with one of the two dominant fracture sets identified in the study area (Becker et al. 2014). This trend was removed before variogram analysis. Therefore, the data positions were projected onto a vector in SW–NE direction and a linear regression was performed on the projected data (Fig. 6). The linear trend revealed from the data is curtailed in the western and eastern parts of the model area in order to avoid unrealistically low and high permeabilities.
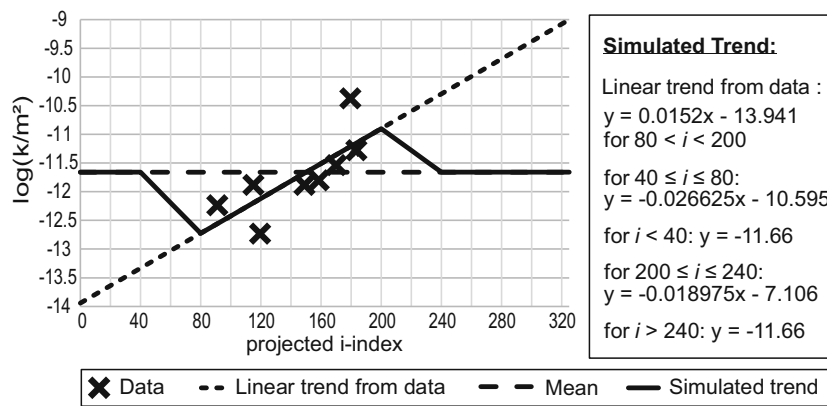
Fig. 5 Model slice at 615 m above reference depth showing locations of wells (green stars) and contour lines of the steady-state hydraulic head solution with confined conditions

Simulated Trend:

Linear trend from data :
y = 0.0152x - 13.941
for 80 < i < 200

for 40 ≤ i ≤ 80:
y = -0.026625x - 10.595

for i < 40: y = -11.66

for 200 ≤ i ≤ 240:
y = -0.018975x - 7.106

for i > 240: y = -11.66

✖ Data　- - Linear trend from data　- - Mean　── Simulated trend

Fig. 6 Log-permeability data of the hard-rock aquifer versus their position projected onto a vector in SW–NE direction (x-markers). The position is given as *i*-index of the numerical model grid. The broken line shows the linear trend. The unbroken line shows the curtailed trend which later is added again to the residuals field for calculating the random log-permeability field. The SGSim simulation is unconditioned, since according to Vogt (2013) unconditioned simulation is preferable to conditioned simulation in case of only few spatial data points

The variogram analysis of the log-permeability residuals from this linear trend relies on a sparse database: nine permeability data points, resulting in 36 data pairs for the variogram. This is not a satisfying number in terms of statistic reliability, but reflects a realistic situation in groundwater projects, especially for nonshallow aquifers. Consequently, a pragmatic approach for receiving an experimental variogram which is as meaningful as possible is to define not less than five lags with at least four data pairs per lag. This resulted in several alternative experimental variograms with five or six lag-bins and maximal lag distances between 1,000 and 1,300 m (i.e., lags between 183 and 260 m), hence leading to several alternative variogram models. For investigating their variability, four reasonable experimental variograms were fitted with spherical variogram models each. The models' range varies between 426 and 664 m, and their sill varies between 0.18 and 0.2. This comparison reveals a robust sill, whereas the range is sensitive to the choice of the experimental variogram. This bears an additional source of uncertainty for

the aquifer permeability that could be considered explicitly in the modeling process by performing several Monte Carlo simulations with different variogram parameters (i.e., geostatistical models) each, resulting in different priors (e.g. Hermans et al. 2015). However, the shortcoming of this approach is the additional computational effort of computing several massive Monte Carlo ensembles. Considering a practical approach, this study uses a deterministic geostatistical model which has the shortcoming of not quantifying this so-called scenario uncertainty in the prior model (e.g. Park et al. 2013).

For the subsequent Monte Carlo simulation, one possible experimental variogram with five lags of 260 m distance each is fitted by trial and error and visual inspection (Fig. 7). Applying more objective mathematical techniques for fitting the variogram model based on extensive statistical approaches (e.g. Marchant and Lark 2004) is not reasonable in case of data scarcity. The resulting nested variogram model consists of two variance regions: a Gaussian variogram with a variance contribution of 0.1278 and an exponential variogram with a variance contribution of 0.0547 (Eq. 8; Fig. 7), both with a horizontal range of 575 m. No horizontal anisotropy could be considered because of the sparse data density. The vertical range is assumed one third of the horizontal range. The error of this deterministic variogram model is expressed in terms of the standard deviation of semivariance at each lag bin (Fig. 7). The first two lag bins are very certain and fitted well by the model as the standard deviation of 0.002 reveals. For the other three lag bins, the model is at least at one end of the uncertainty range. The uncertainty of lag bins 3 and 5 is around 30 times higher than for the first two lag bins.

$$\gamma(d) = 0.1278 \cdot \left(1 - e^{\frac{-3d^2}{575^2}}\right) + 0.0547 \cdot \left(1 - e^{\frac{-3d}{575}}\right) \qquad (8)$$
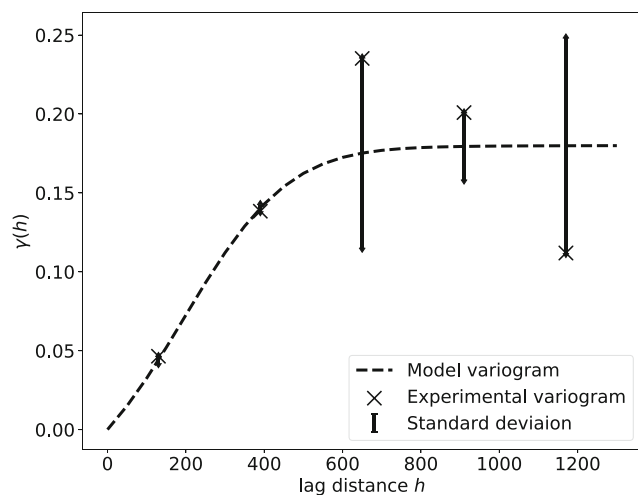

Fig. 7 Experimental variogram (x-markers) and variogram model (dashed line) for the log-permeability residuals. Error bars depict the standard deviation of semivariance at each lag bin

Because of the necessary detrending, the stochastic modeling has to be performed with the log-permeability residuals as

Fig. 8 Simulated drawdown h (broken lines) during the production test for a model with homogeneous permeability in each unit and confined flow conditions compared to monitored drawdown curves (unbroken lines) at six wells versus pumping test duration D
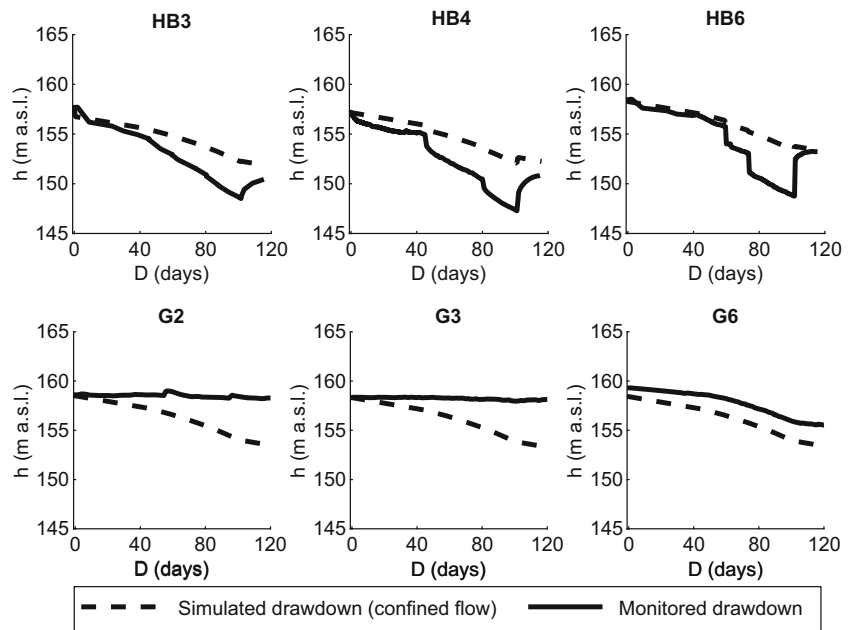
well. After random field generation, these residuals have to be transformed into corresponding permeabilities for solving the flow equation by adding the linear trend which is shown in Fig. 6. This processing step has been implemented directly into the SHEMAT-Suite code.

Additionally, the random field generation with SGSim requires a probability distribution of the log-permeability residuals. As data sparsity does not allow for inferring this distribution directly from measurement data, a random log-normal distribution around a mean residual of zero is assumed. The standard deviation of the log-permeability residuals from pumping test results is 0.4, but as the few data might underestimate standard deviation, the used standard deviation for the log-normal

Fig. 9 Prior log-permeability distribution (solid curve) and posterior log-permeability distribution (dashed curve) of the main hard-rock aquifer system. The latter is inferred from 251 posterior samples

distribution is 0.8. The trimming limits of the distribution are set to three times the data standard deviation, which is 1.2, for avoiding unrealistically small or large residuals.

Results

Deterministic production test simulation

Figure 8 illustrates the simulation results of the transient, confined forward model with mean permeabilities in each model unit. The simulated drawdown is too high at all monitoring wells (G2, G3, G6). At wells G2 and G3 it is up to 5 m higher than the observed drawdown and also much steeper. In reality, both wells do not show any reaction to the pumping. In contrast, monitoring well G6 shows a slight reaction to the pumping with a total drawdown of nearly 5 m, which is nearly captured by the simulation; however, the absolute simulated head is around 1–2 m lower than the observed head.
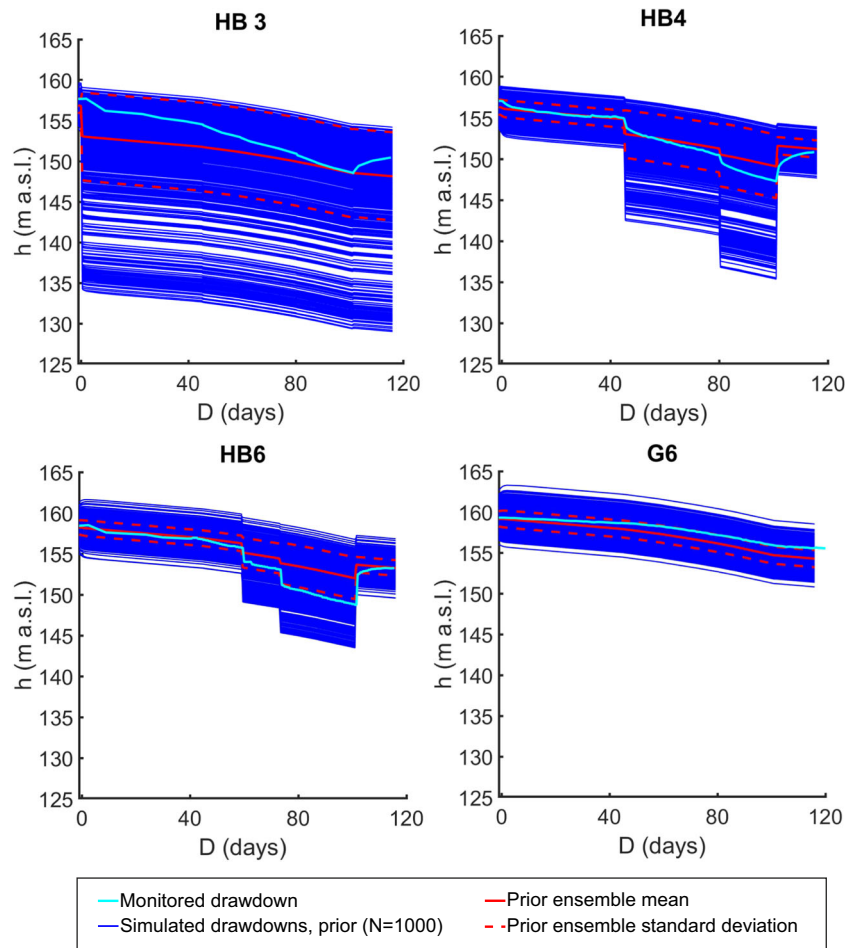
The pumping wells (HB3, HB4, HB6) show pronounced and direct reactions to the pumping test in the observed hydraulic heads. The simulation captures the direct reactions on changes in the pumping stages as well, but they are much less pronounced. Observed head drawdowns are in the range of several meters, whereas the simulated head differs only few centimeters as a reaction to changing pumping rates. These results emphasize the need for implementing heterogeneous permeability into the model, as the model with homogeneous unit permeability does not reproduce the real aquifer behavior.

Conceptually, the aquifer may be unconfined in some parts of the study area; therefore, the transient model was simulated with unconfined flow conditions, too. The unconfined simulation results in a very low drawdown at all monitoring points, which

represents the transient flow behavior observed in wells G2 and G3 satisfactorily. It can therefore be concluded that wells G2 and G3 drill into an unconfined part of the aquifer, which is why they are not used as monitoring points for further stochastic investigation of the confined main hard-rock aquifer system.

Ensemble-based stochastic production test simulations

An ensemble of 1,000 model realizations was generated honoring permeability data and their spatial variability as defined by the prior model (section 'Stochastic model'). Figure 9 shows the prior log-permeability distribution with mean log-permeability of

–11.66 and a standard deviation of 1.34; subsequently, the transient production test model was simulated for each realization in parallel. The simulations ran on JURECA. Resulting drawdown curves at the monitoring and observation wells within the main hard-rock aquifer system are plotted in Fig. 10 together with the prior ensemble mean and standard deviations and in comparison to the observed drawdown curves. The average RMSE of the prior drawdown realizations at the four wells is 1.49 m with 1.6 m standard deviation (Table 5). The average standard deviation of the prior drawdown simulation over the complete time series is 2.5 m (Table 5).

Observed drawdown is within the range covered by the prior ensemble simulations for the whole time-series. The

Table 5 Root mean square errors (*RMSE*) of the deterministic realization with homogeneous permeability, the prior ensemble mean and the posterior ensemble mean alongside with average standard deviation (*SD*) of the prior and posterior drawdown realizations respectively.
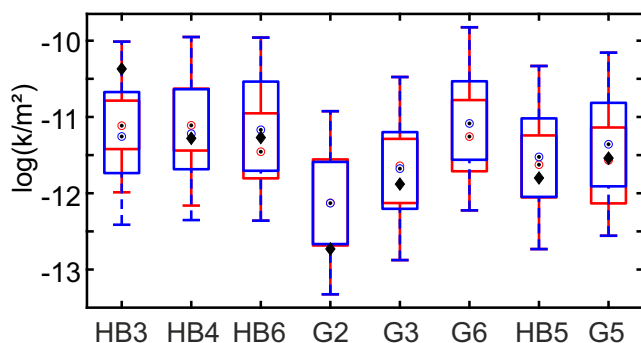
Well	RMSE (m) deterministic	RMSE (m) prior ensemble mean	Prior SD (m)	RMSE (m) posterior ensemble mean	Posterior SD (m)
HB3	3.84	2.13	5.0	1.74	1.2
HB4	4.29	1.06	2.3	1.81	1.0
HB6	3.07	1.86	1.5	2.1	1.2
G6	1.86	0.89	1.0	1.1	0.9
Mean	3.27	1.49 ± 1.6	2.5	1.69 ± 0.69	1.1

Fig. 11 Posterior ensemble of 251 drawdown curves (blue area/lines) at four monitoring and observation wells which were sampled from the prior. What appears as white lines are gaps between ensemble members. Solid red lines are the ensemble means and dashed red lines are the ensemble standard deviations. Cyan lines depict the observed drawdown curves for comparison to simulation results
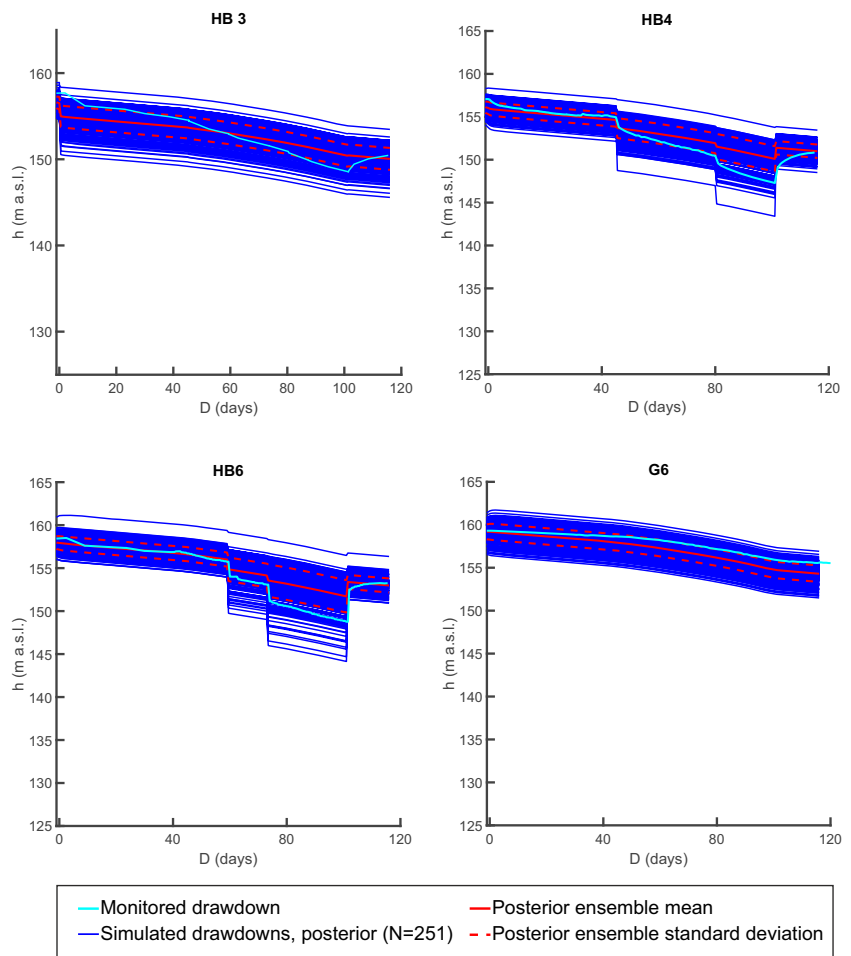
prior ensemble comprises the magnitude of the observed drawdown during the production test at all four wells and covers the temporal behavior (i.e., reaction to pumping) well at wells HB4, HB6 and G6. At well HB3, observations show a steeper decrease and more pronounced recovery of hydraulic head than the prior realizations. The prior realizations are

Fig. 12 Boxplots of the prior ensemble ($N = 1,000$) log-permeabilities at eight monitoring points (blue boxplots) compared to log-permeability data from pumping tests (black diamonds). No permeability data is available for well G6. The red boxplots represent log-permeabilities of the posterior ensemble ($N = 251$). Note that sampling of the posterior was based on head drawdown observations from the main hard-rock aquifer system only (i.e., wells HB3, HB4, HB6 and G6)

characterized by a continuous linear drawdown, whereas the observed drawdown is becoming steeper with time until the recovery phase. A continuous production rate is applied in well HB3; changes in slope of the drawdown are a direct reaction to the onset or increase of production in wells HB4 and HB6. Obviously, the prior model does not reproduce that particular aquifer response exactly. Additionally, the spread of prior drawdown realizations is largest at well HB3; however, the prior ensemble does as well encompass the average drawdown behavior over the complete time-series at well HB3.

In conclusion, the prior model is consistent with observation data and cannot be falsified; thus, subsequent conditioning of the prior to the hydraulic head drawdown observations in order to converge to the posterior ensemble is a valid step. A Markov Chain Monte Carlo method is used for estimating the posterior permeability distribution and the respective posterior ensemble of drawdown curves. Applying the algorithm described in section 'Monte Carlo approach and random field generation', the prior model is conditioned to the drawdown observations. Each prior realization is evaluated based on its misfit in terms of RMSE (Eq. 6) and accepted for the posterior ensemble according to the acceptance

probability given in Eq. (5). The error term s combining measurement error and model errors is assumed 0.5 m. This reflects the remaining uncertainties in, e.g. boundary conditions or the variogram and at the same time results in a reasonable acceptance rate of 251 realizations for the posterior ensemble.

Figure 11 shows the resulting 251 posterior realizations alongside the observed drawdown at the four monitoring and production wells. The spread of the posterior drawdown ensemble decreased compared to the prior; yet, the posterior ensemble embraces the observed drawdown at all four wells, thus providing a meaningful posterior uncertainty quantification. The average standard deviation of the posterior drawdown simulation over the complete time series is 1.1 m.

The constant modest drawdown at well G6 is reproduced well by the posterior ensemble. The same holds for the comparable behavior at wells HB4 and HB6 during the first 45 days of the production test. It is a reaction to the constant pumping rate in well HB3; however, the reactions to pumping and changes in pumping rates at wells HB4 and HB6 are more pronounced in reality than in the posterior mean, but there are realizations within the posterior that are able to reproduce the observed aquifer response. Uncertainty in terms of standard deviation is increased during those periods of direct reactions to changes in pumping rates.

The resulting posterior log-permeability distribution reveals with −11.66 almost the same mean as the prior, but has a lower standard deviation of 0.83 (Fig. 9). For investigating the resulting permeabilities more closely, Fig. 12 depicts the prior and posterior distribution at the well positions as boxplots. Average log-permeability of the prior follows the applied trend of decreasing log-permeability from NE to SW (see Fig. 6). The spread is reduced in the posterior compared to the prior at wells where conditioning data were available and at two other wells (G3 and HB5). Measured log-permeabilities at wells HB3 and G2 are not captured by the ensemble median but are still within the log-permeability range of the full ensemble. This somewhat bad fit can likely be eliminated by conditioning the geostatistical prior model to the available log-permeability data.

Overall, the posterior reduces the uncertainty both in permeability and in simulated drawdown compared to the prior: uncertainty of log-permeability is reduced by 0.5 and uncertainty of the drawdown simulation is reduced by an average of 1.4 m in terms of standard deviation. Besides enabling the uncertainty assessment of the aquifer permeability and of related drawdown simulations, a comparison of the RMSE (Table 5) illustrates the improvement of the average model fit by a stochastically simulated heterogeneous permeability ensemble compared to homogeneous average permeability. The average drawdown of the posterior ensemble matches the observed drawdown with a mean RMSE of 1.69 m, which shows how stochastic simulations

supported by high performance computing can be applied to hydrogeological problems for improving groundwater flow models.

Discussion

Ensemble-based stochastic calibration improves the hard-rock aquifer flow model of the Hastenrather Graben catchment area. Advances in computational sciences enable such massive Monte Carlo simulations. Conditioning the model to steady-state hydraulic head data does not capture the heterogeneity of the hard rock's permeability. Instead, transient drawdown data from a production test in several wells distributed over the model domain accounts for heterogeneity more sufficiently. Commonly, such production test data become available through aquifer testing; similarly, concentrations from tracer tests can be used for inversion and model calibration.

A stochastic model always depends on geostatistical parameters. For this study, some educated assumptions had to be made in order to define all geostatistical parameters necessary for the ensemble generation. Data scarcity made robust variogram modeling difficult and results in high uncertainty of the geostatistical model. More wells or a different spatial distribution of the wells across the model domain could have improved the determination of the spatial correlation. This is worth considering when placing monitoring wells in a catchment area where stochastic simulations are planned, as stressed by, e.g. Júnez-Ferreira et al. (2019). Bogaert and Russo (1999) propose, for example, an optimization approach for spatial sampling design in order to increase the quality of the variogram estimation. Alternatively, one could introduce this scenario uncertainty into the estimation process by considering several potential variogram models simultaneously instead of using one deterministic model (e.g. Rubin et al. 2018). So far, the presented study did not account for it explicitly; thus, the results do not fully capture the prior uncertainty. Possible extension of the approach is at the cost of higher computational burden because several prior ensembles would have to be computed (e.g. Park et al. 2013; Hermans et al. 2015).

The few available data points are spread over a relatively large area; therefore, the geostatistical model covers relatively large-scale correlations (lag distance in the order of 10^2), smaller scale correlations cannot be resolved. Thus, the resulting flow model is not able to simulate effects, which are caused by small scale heterogeneities, which might be one reason why the reaction to pumping in the surrounding wells is not reproduced by the ensemble realizations at well HB3 since small-scale permeability variations and, hence, possible flow paths between the wells are not modeled correctly. However, even the

introduced level of heterogeneity improves the model considerably compared to a deterministic model with homogeneous zonation.

In addition, the data show a spatial trend which had to be removed before variogram analysis. The applied detrending is another source of uncertainty and of possible errors for the modeling process — for example, the bad permeability match of the prior and posterior at wells HB3 and G2 are most likely explained by the fact that it is not well fitted by the trend; however, data scarcity did not allow for removing any data from the trend model and from variogram modeling and to consider them separately. Besides, conditioning the geostatistical prior simulation to the available permeability data would have improved the match at the wells, which should be considered in any potential further stochastic investigations of the Hastenrather Graben catchment area.

The study focuses on the aquifer permeability which is assumed to be the main hydraulic parameter to affect the drawdown. Nevertheless, specific storage partly controls the aquifer's reaction on pumping as well. Deviations from the average specific storage applied in the simulation are very likely in a heterogeneous hard-rock aquifer. Its spatial variability could be incorporated and investigated in a similar manner and would likely improve the simulation; however, here, this was neglected for the sake of simplicity.

If more spatial data points on permeability are available in the future, the model could be improved by analyzing the Kohlenkalk and the Walhorn and Stolberg Layers separately with distinct geostatistical models. Additionally, one has to be aware that the model yields valid results only for the central model domain, where the calibration data points are located and which is not influenced by boundary conditions — for example, the simulation yields high head values which are above elevation in the northern model domain. This may be caused by the uncertain northern boundary condition that was extrapolated from the central model area as no data are available for the north. The high heads can also be explained by confined conditions due to the existence of the separating clay layer in that area. Nevertheless, model results in the outer areas of the model domain could not be validated by any data.

An improvement of the quality of fit might even be achieved by using advanced data assimilation methods such as the Ensemble Kalman Filter (e.g. Gómez-Hernández et al. 1997; Vogt et al. 2012; Kurtz et al. 2014; Keller et al. 2018). For models of this size (i.e., memory, discretization, number of realizations), a distributed-memory parallelism of the EnKF is necessary such as that provided by the Parallel Data Assimilation Framework (PDAF; Nerger and Hiller 2013); however, this is not yet integrated into the SHEMAT-Suite code.

Conclusions

A model with six homogeneous permeability zones was sufficient for calibrating the natural steady-state behavior of the Hastenrather Graben hard-rock aquifer; however, for simulating the aquifer's behavior during production, the interaction between the wells has to be captured by introducing heterogeneous permeability. Quantitative statements on preferential flow paths within the hard-rock aquifer will only be possible by direct investigation of conduits (i.e., karst structures and the fault and fracture network) in the subsurface and their integration into a numerical model. Geostatistical analysis of hydraulic data enabled stochastic simulations of the hard-rock aquifer's heterogeneous permeability with a Monte Carlo approach. The computationally expensive simulations were performed on a high-performance computer using the parallelized simulation software SHEMAT-Suite.

This paper provides a possible geostatistical model of the hard-rock aquifer permeability at the Hastenrather Graben catchment area despite data scarcity. Monte Carlo simulations of 1,000 model realizations allowed for verifying the prior model and quantifying prior uncertainty of the aquifer permeability and of drawdown simulations. Due to data scarcity, the variogram model is not very robust and relatively uncertain. Nevertheless, drawdown observations verify the prior model and subsequent Markov Chain Monte Carlo sampling could be performed for converging to the posterior model. A posterior mean log-permeability of -11.66 with an uncertainty of 0.83 is obtained, whereas the corresponding average uncertainty of the drawdown simulation is 1.1 m, while some sources of uncertainty such as variogram uncertainty, remain unquantified. Still, this study is an important step towards an entire uncertainty quantification for a sparsely sampled hard-rock aquifer, which is crucial for the adequate management of groundwater resources. It can potentially be the basis for further stochastic investigations with advanced methods.

The stochastic simulation of the hard-rock aquifer permeability improved the flow model compared to a model with homogeneous permeability. Although permeability data in the study area are sparse and the hard-rock aquifer is highly heterogeneous, the average fit to the observed drawdown curves is 1.7 m in terms of RMSE. This could be achieved by the stochastic approach without adding any information on the fracture network of the hard-rock aquifer, which makes it a practical approach if fracture network data are sparse. The resulting posterior model of the Hastenrather Graben catchment area can be used for simulating other production scenarios providing drawdown predictions and according uncertainty.

In summary, geostatistical simulations and Markov Chain Monte Carlo sampling were applied successfully to a sparsely sampled hard-rock aquifer for analyzing prior and posterior

uncertainty of the aquifer permeability and its response to water production. The demonstration of this real-case application of stochastic hydrogeological approaches to a complex hard-rock aquifer might motivate practitioners to apply stochastic methods for uncertainty quantification of subsurface flow models.

References

Baca RG, Arnett RC, Langford DW (1984) Modelling fluid flow in fractured-porous rock masses by finite-element techniques. Int J Numer Methods Fluids 4(4):337–348. https://doi.org/10.1002/fld.1650040404

Becker S, Nguyen HT, Nollet S, Fernandez-Steeger TM, Laux D, Hilgers C (2014) Methods to analyse fracture orientation patterns in a lower carboniferous carbonate reservoir analogue in the Voreifel, Germany. German J Geosci 165(3):319–330. https://doi.org/10.1127/1860-1804/2014/0078

Bogaert P, Russo D (1999) Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. Water Resour Res 35(4):1275–1289

Burs D, Bruckmann J, Rüde TR (2016) Developing a structural and conceptual model of a tectonically limited karst aquifer: a hydrogeological study of the Hastenrather graben near Aachen, Germany. Environ Earth Sci 75(18). https://doi.org/10.1007/s12665-016-6039-x

Caers J (2003) Efficient gradual deformation using a streamline-based proxy method. J Pet Sci Eng 39(1–2):57–83

Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ (2005) Inverse problem in hydrogeology. Hydrogeol J 12:206–222. https://doi.org/10.1007/s10040-004-0404-7

Chapman B, Jost G, Van der Pas R, Kuck DJ (2008) Using OpenMP: portable shared memory parallel programming. MIT Press, Cambridge

Chatziliadou M (2009) Rb–Sr Alter und Sr–Pb Isotopencharakteristik von Gangmineralisationen in paläozoischen Gesteinen am Nordrand des linksrheinischen Schiefergebirges (Raum Stolberg-Aachen-Kelmis) und Vergleich mit den rezenten Thermalwässern von Aachen-Burtscheid [Rb-Sr age and Sr-Pb isotope characteristics of vein mineralisations in Paleozoic rocks at the northern margin of the "linksrheinisches Schiefergebirge" (area Stolberg-Aachen-Kelmis) and comparison with the recent thermal waters from Aachen-Burtscheid]. PhD Thesis, RWTH Aachen University, Aachen. http://publications.rwth-aachen.de/record/51191/files/Chatziliadou_Maria.pdf. Accessed 1 July 2019

Chen T, Clauser C, Marquart G, Willbrand K, Hiller T (2018) Upscaling permeability for three-dimensional fractured porous rocks with the multiple boundary method. Hydrogeol J 26:1903–1916. https://doi.org/10.1007/s10040-018-1744-zDO

Clauser C (ed) (2003) Numerical simulation of reactive flow in hot aquifers: SHEMAT and Processing SHEMAT. Springer, Heidelberg, Germany

Dagan G (2002) An overview of stochastic modeling of groundwater flow and transport: from theory to applications. EOS Trans AGU 83(53):621

Deutsch CV, Journel AG (1998) GSLIB: geostatistical software library and user's guide. Oxford University Press, Oxford

Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367

Gómez-Hernández JJ, Sahuquillo A, Capilla JE (1997) Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. theory. J Hydrol 203(1–4):162–174

Gropp W, Huss-Lederman S, Lumsdaine A, Lusk EL, Nitzberg B, Saphir W, Snir M (1998) MPI: the complete reference, vol 2—the MPI-2 extensions. MIT Press, Cambridge

Hayley K (2017) The present state and future application of cloud computing for numerical groundwater modeling. Groundwater 55(5):678–682. https://doi.org/10.1111/gwat.12555

Hermans T, Nguyen F, Caers J (2015) Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: workflow and case study. Water Resour Res 51(7):5332–5352. https://doi.org/10.1002/2014WR016460

Hermans T, Nguyen F, Klepikova M, Dassargues A, Caers J (2018) Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian evidential learning. Water Resour Res 54:2931–2948

Hermans T, Lesparre N, De Schepper G, Robert T (2019) Bayesian evidential learning: a field validation using push-pull tests. Hydrogeol J 27(5):1661–1672. https://doi.org/10.1007/s10040-019-01962-9

Hu JY (2000) Gradual deformation and iterative calibration of Gaussian-related stochastic models. Math Geol 32(1):87–108

Júnez-Ferreira HE, Herrera GS, Saucedo E, Pacheco-Guerrero A (2019) Influence of available data on the geostatistical-based design of optimal spatiotemporal groundwater-level monitoring networks. Hydrogeol J 27:1207–1227. https://doi.org/10.1007/s10040-018-01921-w

Keller J, Hendricks-Franssen H-J, Marquart G (2018) Comparing seven variants of the ensemble Kalman filter: how many synthetic experiments are needed? Water Resour Res 54(9):6299–6318. https://doi.org/10.1029/2018WR023374

Kurtz W, Hendricks-Franssen H-J, Kaiser H-P, Vereecken H (2014) Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river–aquifer interactions. Water Resour Res 50:1665–1688. https://doi.org/10.1002/2013WR014823

Li L, Zhou H, Gómez-Hernández JJ, Hendricks Franssen H-J (2012) Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter. J Hydrol 428–429: 152–169. https://doi.org/10.1016/j.jhydrol.2012.01.037

Linde N, Renard P, Mukerji T, Caers J (2015) Geological realism in hydrogeological and geophysical inverse modeling: a review. Adv Water Resour 86:86–101. https://doi.org/10.1016/j.advwatres.2015.09.019

de Marsily G, Delay F, Gonçalvès J, Renard P, Teles V, Violette S (2005) Dealing with spatial heterogeneity. Hydrogeol J 13:161–183. https://doi.org/10.1007/s10040-004-0432-3

Marchant BP, Lark RM (2004) Estimating variogram uncertainty. Math Geol 36(8)

Mariethoz G, Renard P, Caers J (2010) Bayesian inverse problem and optimization with iterative spatial resampling. Water Resources Research 46(11)

Mosegaard K, Tarantola A (1995) Monte Carlo sampling of solutions to inverse problems. J Geophys Res 100(B7):12431–12447

Nerger L, Hiller W (2013) Software for ensemble-based data assimilation systems: implementation strategies and scalability. Comput Geosci 55:110–118. https://doi.org/10.1016/j.cageo.2012.03.026

Neuman SP, Depner JA (1988) Use of variable-scale pressure test data to estimate the log hydraulic conductivity covariance and dispersivity of fractured granites near Oracle, Arizona. J Hydrol 102:475–501

Neuman SP (2005) Trends, prospects and challenges in quantifying flow and transport through fractured rocks. Hydrogeol J 13(1):124–147. https://doi.org/10.1007/s10040-004-0397-2

Oliver DS, Reynolds AC, Liu N (2008) Petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge

Park H, Scheidt C, Fenwick D, Boucher A, Caers J (2013) History matching and uncertainty quantification of facies models with multiple geological interpretations. Comput Geosci 17(4):609–621. https://doi.org/10.1007/s10596-013-9343-5

Prickett TA (1975) Modeling techniques for groundwater evaluation. In: Ven Chow T (ed) Advances in hydrosciences, vol 10. Academic, New York, pp 1–143

Rath V, Wolf A, Bücker HM (2006) Joint three-dimensional inversion of coupled groundwater flow and heat transfer based on automatic differentiation: sensitivity calculation, verification, and synthetic examples. Geophys J Int 167:453–466

Renard P (2007) Stochastic hydrogeology: what professionals really need? Ground Water 45(5):531–541

Rubin Y (2003) Applied stochastic hydrogeology. Oxford University Press, Oxford

Rubin Y, Chang C-F, Chen J, Cucchi C, Harken B, Heße F, Savoy H (2018) Stochastic hydrogeology's biggest hurdles analyzed and its big blind spot. Hydrol Earth Syst Sci 22:5675–5695. https://doi.org/10.5194/hess-22-5675-2018

Scheidt C, Li L, Caers J (2018) Quantifying uncertainty in subsurface systems. Wiley and AGU, Hoboken, NJ and Washington, DC

Snir M, Otto SW, Huss-Lederman S, Walker DW, Dongarra J (1998) MPI: the complete reference, vol 1: the MPI core, 2nd edn. MIT Press, Cambridge, MA

Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. Society of Industrial and Applied Mathematics, Philadelphia, PA

Tsang YW, Tsang SF, Hale FV (1996) Tracer transport in a stochastic continuum model of fractured media. Water Resour Res 32:3077–3092

Vogt C (2013) Optimization of geothermal energy reservoir modeling using advanced numerical tools for stochastic parameter estimation and quantifying uncertainties. PhD Thesis, RWTH Aachen University, Aachen. https://publications.rwth-aachen.de/record/211527/files/4508.pdf. Accessed 24 January 2020

Vogt C, Marquart G, Kosack C, Wolf A, Clauser C (2012) Estimating the permeability distribution and is uncertainty at the EGS demonstration reservoir Soultz-sous-Forêts using the ensemble Kalman filter. Water Resour Res 48. https://doi.org/10.1029/2011WR011673

Warren JE, Root PJ (1963) The behavior of naturally fractured reservoirs. SPE J 3(03):245–255. https://doi.org/10.2118/426-PA

Wolf A (2011) Ein Softwarekonzept zur hierarchischen Parallelisierung von stochastischen und deterministischen Inversionsproblemen auf modernen ccNUMA-Plattformen unter Nutzung automatischer Programmtransformation [A software concept for the hierarchical parallelization of stochastic and deterministic inverse problems on modern ccNUMA architectures]. PhD Thesis, RWTH Aachen University, Aachen, Germany. http://publications.rwth-aachen.de/record/64281/files/3766.pdf. Accessed 1 July 2019

Zhang D (2002) Stochastic methods for flow in porous media. Academic, San Diego

Zhou H, Gómez-Hernández J-J, Li L (2014) Inverse methods in hydrogeology: evolution and recent trends. Adv Water Resour 63:22–37. https://doi.org/10.1016/j.advwatres.2013.10.014