PAPER

Revealing vertical aquifer heterogeneity and hydraulic anisotropy by pumping partially penetrating wells

Ruth Maier [1] · Carsten Leven [1] · Emilio Sánchez-León [1] · Daniel Strasser [2] · Maximilian Stoll [1] · Olaf A. Cirpka [1] ●

Abstract
The stratification of sedimentary aquifers introduces spatial variability in hydraulic conductivity, primarily between individual horizontal layers. On larger scales, the vertical heterogeneity enhances hydraulic anisotropy, with the horizontal conductivity typically exceeding the vertical one. In this study, the hydraulic anisotropy of a stratified aquifer is estimated from data of hydraulic tests in which water is sequentially extracted from well sections screened at different depths, and the hydraulic response is measured at various multilevel observation wells. The applicability of the method is demonstrated by field tests in a fluvial gravel aquifer in the Upper Rhine Valley, Germany. A homogeneous anisotropic model, and models with three and five anisotropic layers, are fitted to the measured drawdowns in the steady-shape regime, in which differences in hydraulic head between observation locations do not change over time even though the head values themselves change. The position of the five horizontal layers is based on the lithology of the drilling profile at the pumping-well location. The three-layer model is achieved by merging insensitive or similar layers with sensitive layers. The fits result in estimates of the radial and vertical hydraulic conductivities for all layers of the respective models, which are upscaled to effective parameters over the entire depth in the case of the multilayer models. The homogeneous model shows significantly higher errors than those of the heterogeneous models. The heterogeneous locally anisotropic models not only reveal vertical variability of hydraulic conductivity, but also lead to a three-times larger anisotropy ratio upon upscaling.

Keywords Anisotropy · Heterogeneity · Partially penetrating wells · Aquifer testing · Germany

Introduction

Resolving the main properties of the spatially variable hydraulic conductivity tensor $\mathbf{K}$ is a key issue in groundwater modeling. When setting up a groundwater flow model, a model needs to be selected that considers appropriate initial and boundary conditions but also constitutes a suitable representation of the main geological features of the investigated aquifer. In fluvial gravel aquifers, this implies considering bedded subsurface features defined by the deposition of sediments of different size, geometry and sorting (Borghi et al. 2015, Bennett et al. 2019). Since the stratification of sediments affects the spatial distribution of hydraulic properties (Koltermann and Gorelick 1996; Heinz and Aigner 2003; Heinz et al. 2003), a major concern when conceptualizing the conductivity distribution in a groundwater model of a fluvial aquifer is to identify and characterize an appropriate number of layers with different properties. The differences between individual strata also cause the formation-

✉ Olaf A. Cirpka
   olaf.cirpka@uni-tuebingen.de

   Ruth Maier
   ruth.maier@uni-tuebingen.de

   Carsten Leven
   carsten.leven-pfister@uni-tuebingen.de

   Emilio Sánchez-León
   emilio.sanchez@uni-tuebingen.de

   Daniel Strasser
   daniel.strasser@baw.de

   Maximilian Stoll
   maximilian.stoll@student.uni-tuebingen.de

1 Center for Applied Geoscience, University of Tübingen, Schnarrenbergstraße 94-96, 72076 Tübingen, Germany

2 Federal Waterways Engineering and Research Institute, Kußmaulstraße 17, 76187 Karlsruhe, Germany

✑ Springer

averaged hydraulic conductivity tensor **K** to be anisotropic because groundwater preferably flows in the direction of the layers rather than perpendicular to it (Bear 1972; Borghi et al. 2015). This implies that the principal directions of the effective hydraulic conductivity tensor $\mathbf{K}^{\text{eff}}$ are typically the horizontal and vertical directions which are aligned with the strata. In this study, the ratio of horizontal to vertical conductivity, when averaging over the horizontal layers, is denoted the anisotropy ratio ϑ. The hydraulic anisotropy is of importance whenever the vertical flow component is significant, for instance in flow close to partially penetrating wells, horizontal collector wells, or around objects that partially penetrate aquifers, or when considering river–groundwater exchange. While regional flow is predominantly horizontal, these specific boundary conditions induce a vertical-flow component that can be crucial in the overall design of groundwater management measures.

Many hydrogeological applications such as the design of remediation systems, depend not only on precise information on subsurface heterogeneity (e.g. Cardiff and Barrash 2011; Zschornack et al. 2013), but also require information on hydraulic anisotropy (e.g. Bair and Lahm 1996; Zlotnik and Ledder 1996). A specific example in which hydraulic anisotropy is relevant is the delineation of capture zones of partially penetrating wells (Bair and Lahm 1996). Several experimental methods have been developed for resolving the spatial variability of hydraulic conductivity at different scales and degrees of resolution. Hydraulic tomography, for example, is a common method that helps to obtain larger-scale (>10 m) three-dimensional (3D) information on hydraulic-conductivity variations (Gottlieb and Dietrich 1995; Yeh and Liu 2000; Bohling 2009; Hochstetler et al. 2016; Sanchez-Leon et al. 2016). Direct-push methods such as direct-push injection logging (Bohling et al. 2002; Butler et al. 2007; Dietrich et al. 2008; Lessoff et al. 2010) or the direct-push permeameter (Butler et al. 2007; Chen et al. 2008, 2010; Klammler et al. 2011; Zschornack et al. 2013) resolve apparent horizontal hydraulic conductivity with depth and are thus well suited to investigate local hydraulic-conductivity variations in the vertical direction at high resolution.

To resolve the ratio of horizontal to vertical hydraulic conductivity, different field methods have been investigated. Klammler et al. (2017) proposed a shape factor to estimate the bulk hydraulic anisotropy from measurements obtained with the direct-push permeameter but without considering the vertical variability of hydraulic conductivity. The tomographic slug test proposed and tested by Paradis et al. (2015, 2016) in a littoral aquifer seems to be more appropriate for resolving hydraulic anisotropy induced by heterogeneities at smaller scales, also in the horizontal direction, e.g., from cross-bedding. A specific limitation of the tomographic slug test, however, is the very small range of investigation in the horizontal direction (typically <10 m), especially in highly permeable aquifers (Paradis et al. 2015). Even though different studies have dealt with the investigation of anisotropic conductivity, a suitable field method for estimating the ratio of horizontal to vertical hydraulic conductivity on larger scales in fluvial gravel aquifers has not yet been tested. The aim of this study is to examine the viability and benefits of a method for resolving hydraulic anisotropy on larger scales induced by the vertical heterogeneity on smaller scales.

The present study builds upon a method introduced by Maier et al. (2020) to estimate hydraulic anisotropy by inverting steady-shape aquifer tests using a partially penetrating pumping well. The approach follows the basic principles of hydraulic tomography. That is, a series of pumping tests is performed, in which groundwater from different intervals of a single pumping well are sequentially extracted, and the hydraulic response is observed in surrounding observation wells, placed at different distances and depths. In contrast to the steady-state pumping regime, the absolute drawdowns are still changing in the steady-shape pumping regime, but the hydraulic-head differences between observation locations remain constant (Bohling et al. 2002, Bohling et al. 2007).

In this work, the method described by Maier et al. (2020) is modified and applied to a fluvial gravel aquifer located in the Upper Rhine Valley at the Germany-France border. Specifically, the question of how a homogeneous anisotropic groundwater flow model performs in comparison to models with several anisotropic layers is addressed, as well as how these layers should be defined.

While Maier et al. (2020) described the application of steady-shape aquifer tests to jointly optimize modeling and measurement strategies with a synthetic scenario, the present study considers a field application, including the design of the experiments.

This paper starts with a brief repetition of the underlying theory, followed by a description of the field application. Then the numerical models are described, and the principles used in model calibration are outlined. After presenting the site-specific results, the paper finishes with discussing the main findings and giving general recommendations.

Field application

Field site

Hydrogeological setting

The field site is located in the Upper Rhine Valley, north of the municipality Kappel-Grafenhausen, Baden-Württemberg, southwest Germany (Fig. 1). The study site is surrounded by an artificial flow channel in the west and by River Elz in the east. The general groundwater flow direction is from the

Fig. 1 Overview map of the field site
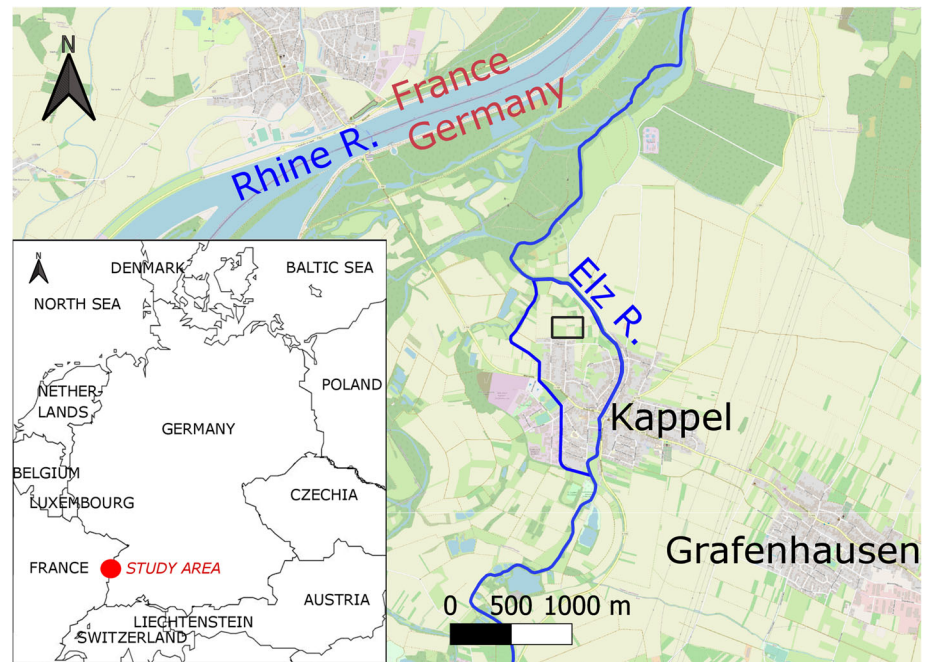
southeast to the northwest with a hydraulic gradient of 0.14%. The fluvial unconsolidated aquifer consists of Quaternary sediments of the Neuenburg-Formation (qN), and is characterized by fluviatile gravel deposits with varying amounts of sand and small amounts of silt (LGRB 2004; Wirsing and Luz 2007). The sequence of strata varies locally. The saturated aquifer of ~41 m thickness is overlain by a ~2-m-thick layer of alluvial fines and bounded by considerably less conductive Quaternary sediments from the Breisgau-Formation (qBS) at the aquifer base (LGRB 2004; Wirsing and Luz 2007). Results from a grain size analysis of an exploration drilling at the field site indicate hydraulic-conductivity values of the unconfined aquifer in the range of 6.7×10^{-5} to 2.6×10^{-1} m/s.

Monitoring network

Figure 2 shows a schematic representation of the field installation at the test site. The large-diameter pumping well (denoted R01, well-screen radius $r_w = 0.4$ m) reaches to a depth of ~21 m within the aquifer. The well was designed with three separate screen sections (I, II and III) of 2-m length each and with a spacing of 4.5 m in between, centered at elevations of 37.32 m (upper screen), 30.82 m (middle screen) and 24.32 m (lower screen) from the aquifer base. Note that all z-coordinates are given in reference to the aquifer bottom. The well was installed with a prepacked filter along the filter-screen sections and completed with coarse gravel to a 0.4-m-thick filter pack, extending by 0.5 m above and below the individual screen sections. Each filter-pack section is connected above

and below to a clay fill through a 0.5-m-thick secondary filter layer.

Three bundles of nested observation wells are placed each in the north, east, south, and west direction of the pumping well R01, at radial distances of 3.5, 6.5 and 10.5 m, respectively (Fig. 2). Two additional bundles of nested observation wells are placed at a radial distance of 21 m, one to the east and one to the west of well R01.

Each bundle of nested observation wells positioned north, south and west of R01 comprises three 1-in (2.54-cm) piezometers, each having a 0.3-m-long screen at its bottom. Two of the three piezometers are placed at depths between the elevation of the pumping-well screens I and II, and the third is at a depth between the elevation of the pumping-well screens II and III (Fig. 2). All observation wells to the east of R01 are continuous multichannel tubing wells (CMT wells) with seven individual channels with a diameter of ~10 mm. Their depths are above, at, and below the elevation of the pumping-well screens I, II and III. Starting from the top to the bottom, the screen openings are enumerated from 1 to 7 (Fig. 2). While screen openings 1 to 6 are 0.4 m long, the lowermost screen opening 7 has a length of 0.15 m. In total, the monitoring network consists of 58 observation points. Table S1 in the electronic supplementary material (ESM) summarizes details of the monitoring network.

Hydraulic tests

In three series of short-term pumping tests, water was successively extracted from the screen sections I, II, and III of pumping well R01 with a frequency-controlled submersible
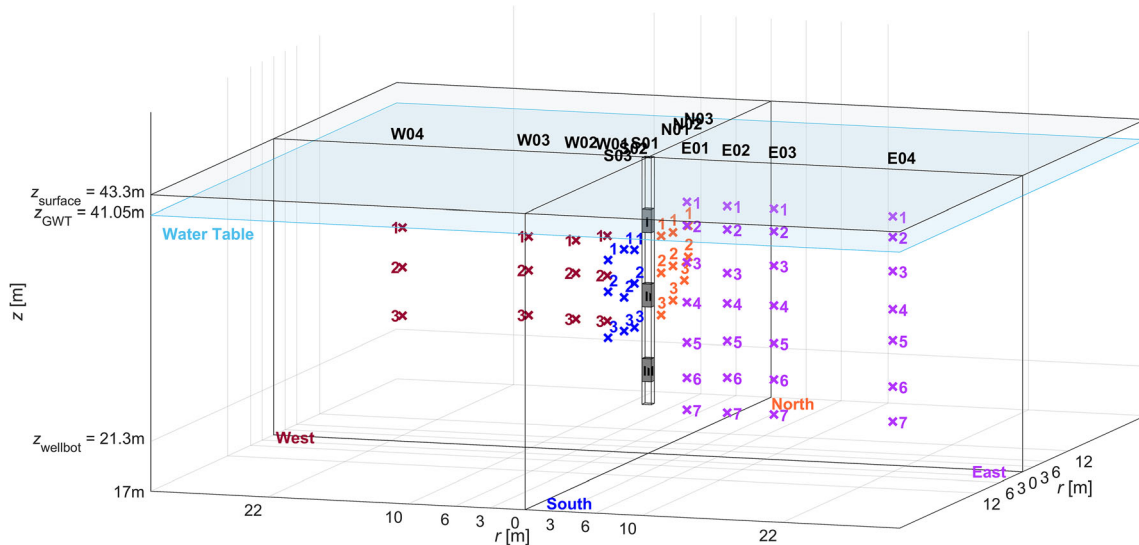
Fig. 2 Overview of the measurement network installed at the field site. The vertical coordinate is measured from the base of the aquifer, and the horizontal coordinates from the pumping well R01 in the middle, illustrated by its well screens. On account of the high permeability of the investigated unconfined aquifer and the short test durations, the numerical model considers a confined system

pump. To prevent water inflow from the adjacent screen sections, a customized straddle-packer system was introduced into the well and placed above and below the active screen section.

For each pumping depth, multiple pumping tests p_t with different pumping rates $Q(p_t)$ were performed. All experiments belonging to the same extraction interval constitute a specific hydraulic test y. Details on the number of pumping tests belonging to the individual hydraulic tests and the applied pumping rates are listed in Table 1.

The transient drawdown response of each pumping test was automatically measured in 44 observation points using fiber-optic pressure transducers and different types of piezoresistive data loggers of similar resolution, whereas the drawdown at the remaining 14 observation points were manually measured. Due to technical problems with one data logger the total number of active observation points in each hydraulic test was reduced to 57.

In order to monitor the stability of the pumping rate, flowmeter readings were regularly taken and the drawdown within the pumping well was observed using a pressure transducer. The latter in-well drawdown measurements are not considered

in the model calibration because these measurements are affected by pump-induced pressure variations and well-skin effects.

Data processing

The pumping tests resulted in a total of 1,334 drawdown curves, from which defective datasets caused by strong instabilities of pressure transducers or the pumping rate were eliminated. Section S2 of the ESM shows an exemplary dataset to illustrate the criterion for eliminating datasets and the state on the assessment of pumping rate stability. In addition, each remaining curve was corrected to consider barometric pressure variations. Considering that drawdown is linearly proportional to the pumping rate, the drawdown curves from pumping tests with different extraction rates were scaled to those of a harmonized rate Q_h of 0.01 m³/s. The scaling is based on the mean discharge, observed during the pumping phase up to timepoint t_{trim} (Table 1). These scaled curves were used to compare the multiple hydraulic responses observed at each observation point and for each hydraulic test, and to check the data reproducibility. Pumping tests in which the

Parameter	I Upper screen	II Middle screen	III Bottom screen
Hydraulic test, y	1	2	3
Number of pumping tests, p_t	5	10	7
Reduced number of pumping tests, p_r	4	7	4
Time to reach steady-shape behavior, t_{trim} [s]	1,800	1,680	1,700
Range of $Q(p_t)$ [L/s]	10.0–10.5	5.5–10.9	17.8–19.0

Table 1 Key data of the pumping test series

drawdown curves predominantly failed the reproducibility test were discarded, reducing the total number of pumping tests from p_t to p_r of each hydraulic test (Table 1).

To avoid the uncertainty associated with the transient behavior of the hydraulic responses, and the intensive computational requirements of simulating transient groundwater flow, the drawdown curves were analyzed in the steady-shape regime, in which the absolute value of drawdown still changes but the hydraulic-head differences between measurement locations remains constant (Bohling et al. 2002, 2007). In the reduced set of pumping tests p_r, the steady-shape regime was defined by identifying the time (t_{trim}) at which changes in drawdown differences between observation locations could be neglected (see Table 1 and section S3 in the ESM). Then the drawdown s_{obs} at timepoint t_{trim} were averaged among all pumping tests p_r available in hydraulic test y. By that, a single averaged drawdown measurement s_{meas} for each observation point k in each hydraulic test y was obtained:

$$s_{\mathrm{meas}}(y,k) = \frac{1}{p_r} \sum_{i=1}^{p_r} s_{\mathrm{obs},i}(y,k) \qquad (1)$$

together with an associated standard deviation σ_{repr} as a metric of the reproducibility of each measurement:

$$\sigma_{\mathrm{repr}}(y,k) = \sqrt{\frac{1}{p_r-1} \sum_{i=1}^{p_r} \left(s_{\mathrm{obs},i}(y,k) - s_{\mathrm{meas}}(y,k)\right)^2} \qquad (2)$$

σ_{repr} denotes the reproducibility error, which contributes to the overall error of the measurements but does not include systematic errors in the data (e.g., due to the misplacement of observation points) or in the conceptual model (e.g., due to disregarding horizontal heterogeneity). To quantify the variability among the measurement points, the observation points that are arranged in different directions to the pumping well, but coincide by ±0.6 and ±0.5 m in their r- and z-coordinates, respectively, were clustered and the associated drawdown measurements s_{meas} compared. These cluster ranges are realistic, since intended measurement locations of observation points may be misplaced in the installation of observation wells (Maier et al. 2020). The mean value μ_{c_p} and standard deviation σ_{c_p} of all measurements n_{cp} available in each cluster c_p are computed by:

$$\mu_{c_p}(y) = \frac{1}{n_{cp}(c_p)} \sum_{i=1}^{n_{cp}(c_p)} s_{\mathrm{meas},i}(y,c_p) \qquad (3)$$

$$\sigma_{c_p}(y) = \sqrt{\frac{1}{n_{cp}(c_p)-1} \sum_{i=1}^{n_{cp}(c_p)} \left(s_{\mathrm{meas},i}(y,c_p) - \mu_{c_p}(y)\right)^2} \qquad (4)$$

Note that the data clustering was performed only to illustrate the data in a comprehensive way and to estimate the

variability of drawdown measurements among different but close-by observation points. The model fit utilizes the nonclustered data. The error model applied in the fitting procedure is discussed in the following.

Model setups

Governing equations

The starting point is the radial-symmetric groundwater-flow equation, describing the drawdown induced by water extraction with a partially penetrating well:

$$\frac{1}{r} \frac{\partial}{\partial r} \left(K_r r \frac{\partial s}{\partial r}\right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial s}{\partial z}\right) = S_0 \frac{\partial s}{\partial t} \qquad (5)$$

in which r [L] and z [L] are the radial and vertical coordinates, K_r [LT^{-1}] and K_z [LT^{-1}] are the radial and vertical hydraulic conductivities, s [L] is the drawdown, i.e., the change in hydraulic head induced upon pumping, t [T] denotes time, and S_0 [L^{-1}] is the specific storage. In this study, $K_r(z)$ and $K_z(z)$ are assumed to vary only in the vertical direction. At the outer radius R [L] of the model domain (much further away from the pumping well than any observation well) zero drawdown is assumed, and there are no-flow boundaries at the top z_{top} [L] and bottom z_{bot}[L]:

$$s(R,z) = 0 \forall z \qquad (6)$$

$$\left.\frac{\partial s}{\partial z}\right|_{z=z_{\mathrm{bot}}} = \left.\frac{\partial s}{\partial z}\right|_{z=z_{\mathrm{top}}} = 0 \forall r \qquad (7)$$

The considered extraction well has three separated well screens to extract groundwater from three different and isolated depth intervals. The pumping well has the radius r_w [L], and each well screen i_{scr} has the same length b [L], but is centered at a different vertical position z_c [L]. Along the well screen considered for pumping, the hydraulic head is constant, and the total extraction flux Q [L^{3}T^{-1}] for the well screen is constant during the hydraulic test. The model assumes no-flow boundary conditions at all depths below and above the respective active screen section, so that at $r = r_w$ the following boundary conditions hold:

$$\left.\begin{array}{l} s = \text{constant, and} \\ -\int_{z_c(i_{\mathrm{scr}})-\frac{b}{2}}^{z_c(i_{\mathrm{scr}})+\frac{b}{2}} 2\pi r K_r \frac{\partial s}{\partial r} dz = Q \end{array}\right\} \quad \text{if } z_c(i_{\mathrm{scr}}) - \frac{b}{2} \leq z \leq z_c(i_{\mathrm{scr}}) + \frac{b}{2}$$

$$\frac{\partial s}{\partial r} = 0 \qquad \qquad \text{otherwise}$$

$$(8)$$

A transient model would require initial conditions and a value of the specific storage, but the constant-shape regime of the pumping tests, in which the spatial profile of drawdown in the domain of interest changes only by a time-dependent constant in space, does not require this information. In this regime, drawdown differences between observation points can be simulated by the steady-state drawdown equation, that is, Eq. 5 with a right-hand side of zero.

The hydraulic tests were simulated with a finite-element model implemented in MATLAB (codes are available through the data portal at University of Tübingen 2022) which solves the axisymmetric steady-state groundwater flow equation on rectangular elements. The radial grid spacing increases logarithmically, whereas the vertical resolution is uniform with a grid spacing of 0.17 m. To obtain simulation results at observation points that did not fall onto nodes of the grid, bilinear interpolation consistent with the finite element formulation was applied.

Conceptual models

In total, three different groundwater-flow models were set up. The first model considers a single, homogeneous layer (1-layer model), whereas the second one contains five horizontal layers (5-layer model). In the third model, the number of layers was set to three (3-layer model). All models account for hydraulic anisotropy in each layer, that is, each layer has an individual radial and vertical hydraulic conductivity K_r and K_z, respectively. In all models the aquifer is treated as confined, which is a qualified assumption due to the short test durations and the small drawdowns in comparison to the total thickness of the aquifer.

Figure 3 shows the basic model setup. All models contain two separate isotropic units that represent the gravel pack (red units in Fig. 3, hydraulic conductivity $K_{gp} = 10^{-2}$ m/s) and clay fill (gray units in Fig. 3, hydraulic conductivity, $K_{cl} = 10^{-7}$ m/s) of the filter pack installed around the pumping well. Fitting the conductivity values of these units was tested, but the data were not sensitive to K_{gp} and K_{cl}, so reasonable fixed values were chosen.

The black-dashed horizontal lines in Fig. 3 indicate the layer boundaries considered in the five-layer model. The actual location of the layer boundaries in the 5-layer model is based on the existing field description of the drill core taken at the location of pumping well R01 (Figs. 3 and 4a). Sand layers were delineated when the sand fraction clearly exceeded that of the gravel over more than half a meter in thickness (Fig. 4b). The layers are numbered from the aquifer top to the bottom. In the 3-layer model, layer 2 was merged to layer 3, while layer 4 was merged to layer 5 of the 5-layer model. The reasoning for these choices is discussed in the following section.

Model calibration

All three models were independently calibrated by the Trust-Region Reflective Least-Squares method of the function lsqnonlin in the optimization toolbox of MATLAB (Coleman and Li 1996). To reduce the large data volume in model calibration, the averaged drawdown measurements s_{meas} from all three hydraulic tests were jointly considered in the calibration, leading to $n_{meas} = 3 \times 57 = 171$ drawdown observations.

As mentioned before, a steady-shape pumping regime was considered in the simulations, in which drawdown differences between observation locations remain constant. Typically, this requires the specification of pairs of observation points by either setting one observation location as the superordinate reference point (Maier et al. 2020) or by considering all feasible pairs of observation points (Bohling et al. 2002). Each field measurement, however, is subject to measurement errors of different types, including measurement noise or the misplacement of observation wells (Maier et al. 2020). In trials not reported here, the effect of considering different observation points as the reference point had been tested, yielding different model-calibration results due to measurement error. To avoid the propagation of uncertainties in the generation of pairs of observation points, the model calibration includes a virtual reference point. That is, for each hydraulic test, the simulated drawdown difference $s_{sim} = |s_t - s_{ref}|$ contained the simulated steady-state drawdown s_t and the drawdown at a virtual reference point s_{ref}, which is identical among all measurement points but needs to be estimated together with the hydraulic-conductivity values.

Then the differences between the simulated and measured drawdowns s_{sim} and s_{meas} were computed and normalized by the error σ_i of each measurement i, which is defined by an error model discussed below.

In the calibration, the objective function φ to be minimized is defined as the sum of squared normalized residuals:

$$\varphi = \sum_{i=1}^{n_{meas}} \left(\frac{s_{sim,i}(\mathbf{p}) - s_{meas,i}}{\sigma_i} \right)^2 \tag{9}$$

in which $\mathbf{p}$ is the parameter vector including the logarithms of K_r and K_z of all horizontal layers considered and the reference drawdown s_{ref} for each of the three hydraulic tests. Thus, in total, the 1-, 3- and 5-layer models include $n_{par} = 3$, $n_{par} = 7$, and $n_{par} = 11$ calibration parameters, respectively.

The error model accounts for the combined effects of the reproducibility error, a potential measurement bias (e.g., due to misplacement of the observation points), and most importantly the model-conceptual error (e.g., due to suboptimal definition of layers or lacking 3-D heterogeneity). In essence, none of the defined models are claimed to be perfect representations of reality so that misfits that are bigger than the error of
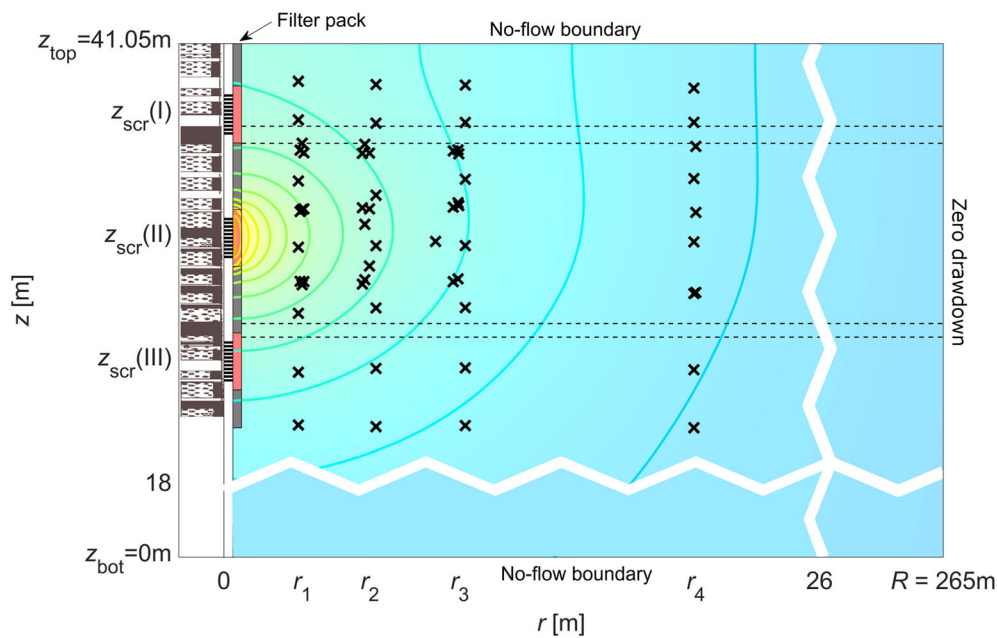
Fig. 3 Conceptual representation of the 1-layer and 5-layer models, distinguished by dashed horizontal lines. The drilling profiles are displayed left of the pumping well showing lithologic information at the pumping well location. Also, the gravel pack (red) and clay seal (gray) are shown right of the pumping well. The black crosses correspond to individual observation points, placed at approximate radial distances r_1 = 3.5, r_2 = 6.5, r_3 = 10.5 and r_4 = 21 m. The model domain reaches from the saturated aquifer thickness at z_{top} = 41.05 m to the aquifer bottom at z_{bot} = 0 m. Note that axes are not to scale and that the contour lines of drawdown correspond to an exemplary response in a homogeneous aquifer system, i.e. considering a 1-layer model

the measurements themselves are accepted for the sake of keeping the hydrogeological models comparably simple and the fitted parameters meaningful. In this framework, a heteroscedastic error model is needed that has a set of parameters that become part of the fitting procedure. As different models have different deficiencies, they have different model errors, and judging the quality of the different models is based on the fitted coefficients of the error model. After testing different error models, which for the sake of brevity are not presented here, the following parameterization appeared to represent the behavior of the residuals reasonably well:

$$\sigma = a + \frac{b \cdot s_{meas}^2}{c + s_{meas}} \quad (10)$$

with a, b and c being the error-model parameters. This specific error model starts off with a constant error with $\lim_{s_{meas} \to 0} \sigma = a$ corresponding to the absolute error, then shows a quadratic increase with the measurement, and converges to a linear dependence on s_{meas} for large values with $\lim_{s_{meas} \to \infty} \frac{\sigma}{s_{meas}} = b$ corresponding to the relative error. The parameter c quantifies how quickly the error model converges from the measurement-independent to the linear regime.

The error model parameters are determined by calibrating the 1-, 3-, and 5-layer models according to the expectation-maximization method (Dempster et al. 1977). The scheme involves iteratively minimizing the objective function with

the Trust-Region Reflective Least-Squares method of the function lsqnonlin in the optimization toolbox of MATLAB (Coleman and Li 1996) with given coefficients of the error model and updating the error-model parameters by performing a least-squares fit of the error model to the absolute residuals $|s_{sim}(\mathbf{p}) - s_{meas}|$ of the model fit to the measured drawdown s_{meas}. With this, the error-model parameters a, b and c, as well as all model parameters, are included in the optimization process. The iterative calibration procedure is completed when the change in all model and error parameters is less than 1%. The comparison of the different models is now not based on meeting the observations within the measurement error but on the magnitude of the model error needed to accept the different models. In the following, the goodness of fit is assessed by comparing the resulting absolute and relative errors between the 1-, 3-, and 5-layer models.

After fitting the models, the associated standard deviation $\hat{\sigma}_{p_i}$ of estimation of the model parameter i are first computed by linearized error propagation:

$$\hat{\sigma}_{p_i} = \sqrt{\mathbf{C_{pp}}(i,i)} \quad (11)$$

with the parameter covariance matrix $\mathbf{C_{pp}}$ computed by:

$$\mathbf{C_{pp}} = \frac{\varphi}{n_{meas} - n_{par}} \left(\mathbf{J}^T \mathbf{\Sigma}^{-1} \mathbf{J} \right)^{-1} \quad (12)$$

in which the Jacobian $\mathbf{J}$ contains the partial derivatives of all simulated measurements with respect to all parameters,

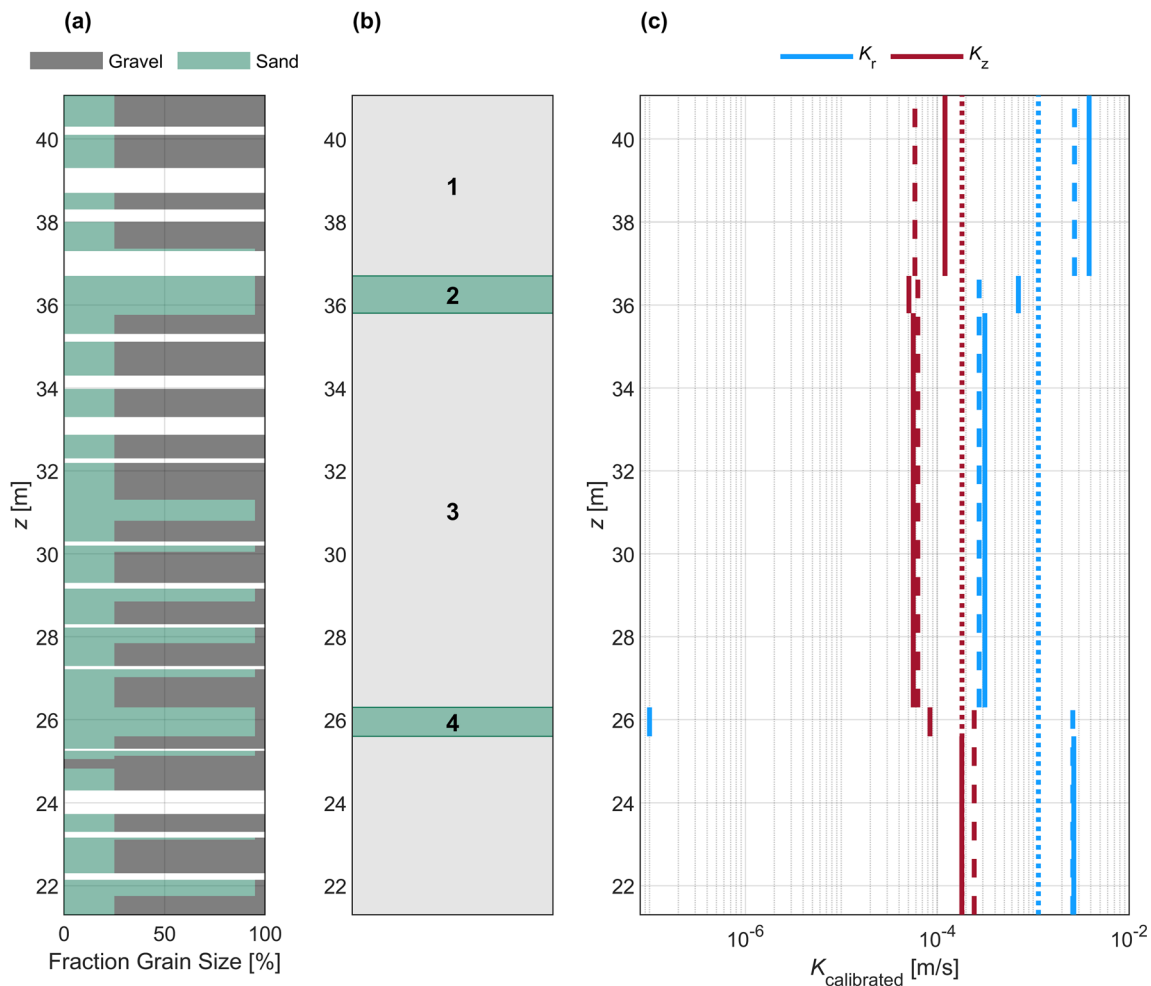**(a)**   **(b)**   **(c)**

Fig. 4 **a** Relative gravel and sand fractions from the field description of the drilling core of the large-diameter well R01 at the field site Kappel-Grafenhausen (white gaps = core loss). **b** Definition of five horizontal layers considering depths of more than half a meter in thickness in which the sand fraction clearly exceeds the gravel fraction. **c** Calibrated radial (blue) and vertical hydraulic conductivities (red) of the 1-layer (dotted lines), 3-layer (dashed lines), and 5-layer model (solid lines). In the 3-layer model, layer 2 was merged to layer 3, while layer 4 was merged to layer 5 of the 5-layer model

and Σ is the diagonal matrix of the squared errors according to the error model. Because the parameters a and b of the error model are bigger if the model shows larger misfits, the resulting parameter standard deviations of estimation are also bigger.

To address nonlinearity, the uncertainty estimate of the model parameters is refined by applying a Markov-Chain Monte Carlo (MCMC) method for the hydraulic-conductivity values with Metropolis-Hastings sampling, starting with the best estimate of the preceding optimization and keeping the coefficients of the error model as well as the reference drawdown values s_{ref} fixed. This leads to a sample of 1,000 parameter realizations for each model, drawn from the posterior distribution. The results of the MCMC sampling are given in section S6 of the ESM.

Finally, the radial and vertical conductivities K_r and K_z are upscaled to the full aquifer thickness, resulting in the effective radial and vertical conductivities K_r^{eff} and K_z^{eff}, defined as the arithmetic and harmonic means of layer-specific values, respectively:

$$K_r^{\mathrm{eff}} = \frac{1}{z_{\mathrm{top}} - z_{\mathrm{bot}}} \int_{z_{\mathrm{bot}}}^{z_{\mathrm{top}}} K_r(\zeta) d\zeta \tag{13}$$

$$K_z^{\mathrm{eff}} = (z_{\mathrm{top}} - z_{\mathrm{bot}}) \left(\int_{z_{\mathrm{bot}}}^{z_{\mathrm{top}}} \frac{1}{K_z(\zeta)} d\zeta \right)^{-1} \tag{14}$$

From this, the anisotropy ratio ϑ is calculated by:

$$\vartheta = \frac{K_r^{\mathrm{eff}}}{K_z^{\mathrm{eff}}} \tag{15}$$

The calculation of K_r^{eff}, K_z^{eff}, and ϑ is also performed for each realization of the MCMC ensemble, resulting in distributions of these quantities.

Results and discussion

Field measurements

Figure 5 shows the drawdown measurements s_{meas} rescaled to reflect a harmonized pumping rate of 0.01 m³/s belonging to the hydraulic tests performed at the top (Fig. 5a), the middle (Fig. 5b) and the bottom (Fig. 5c) screen of the pumping well. The measurements are displayed as a function of the radial distance to the pumping well (different colors in Fig. 5a–c), while for each radial distance the observations are aligned with elevation (bar placement along the z-axis in Fig. 5). As expected, Fig. 5 shows that in each hydraulic test the observed drawdown decreases with increasing radial distance to the pumping well and are higher at elevations close to the pumped screen interval.

The drawdown observations s_{meas} range between millimeters and meters. The signal strength differs between the three tests even after correcting for different pumping rates. When extracting water from the lower screen (Fig. 5c), the drawdown does not reach the high values observed in the other tests, whereas the strongest responses result from the hydraulic test with water extraction from the middle screen (Fig. 5b). These differences may be caused by vertical variations of hydraulic conductivity. In particular, a lower-conductivity layer at a depth close to that of the middle screen could explain higher drawdown values when water is extracted from this screen.

Measurement reproducibility and horizontal heterogeneity

Figure 6a shows the rescaled drawdown measurements s_{meas} together with the errors obtained during the reproducibility test σ_{repr}. Again, Fig. 6a shows that the measurement signal varies between the three hydraulic tests (different colors in Fig. 6a).

A comparison of the errors between the four directions (north, east, south, and west of the pumping well, see different marker symbols in Fig. 6a) reveals no significant spatial pattern with respect to reproducibility.

Figure 6b shows the errors associated with measurements obtained at similar radial distance and depth but in different directions to the pumping well. While the reproducibility errors σ_{repr} do not exceed values of 3 cm, the errors associated with horizontal heterogeneity are higher and reach values of up to 10 cm.

Goodness of model calibration

Figure 7a–c shows the absolute differences between simulated and measured drawdowns $|s_{sim} - s_{meas}|$ versus the measured drawdowns s_{meas} of the 1-, 3-, and 5-layer model calibrations considering the final error-model update. Comparing parts a, b and c of Fig. 7 shows that the differences between simulated and measured drawdowns are significantly higher for the 1-layer than for the two multilayer models. In all cases, the error-
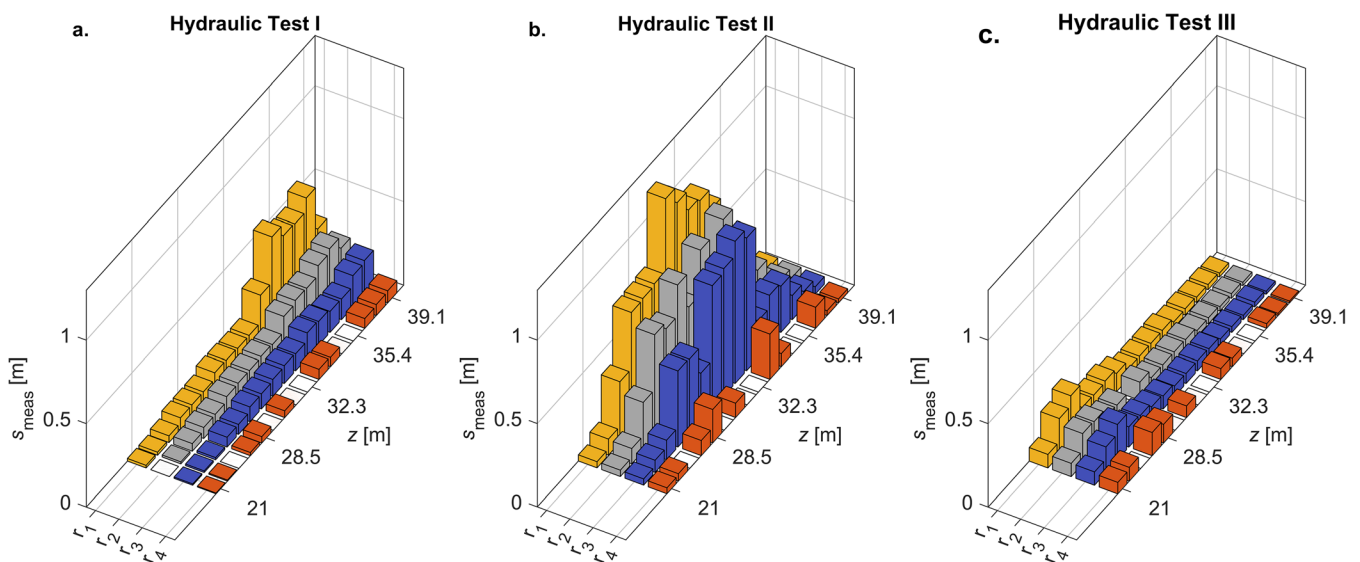

Fig. 5 Drawdown measurements of the hydraulic tests performed at **a** the top, **b** the middle and **c** the bottom screen of the pumping well. The yellow, grey, blue and orange colors correspond to the radial distances $r_1 \in [3.3, 3.6]$, $r_2 \in [6.2, 6.8]$, $r_3 \in [10.2, 10.7]$ and $r_4 \in [20.9, 21]$, respectively

Fig. 6 a Measurement errors σ_{repr} resulting from the reproducibility analysis as function of the drawdown measurements s_{meas}. Colors indicate extraction from the different pumping intervals at different depths, and marker styles indicate horizontal orientation of the measurement points. **b** Measurement errors σ_{c_p} resulting from measurements at similar radial distances and depths but different directions to the pumping well
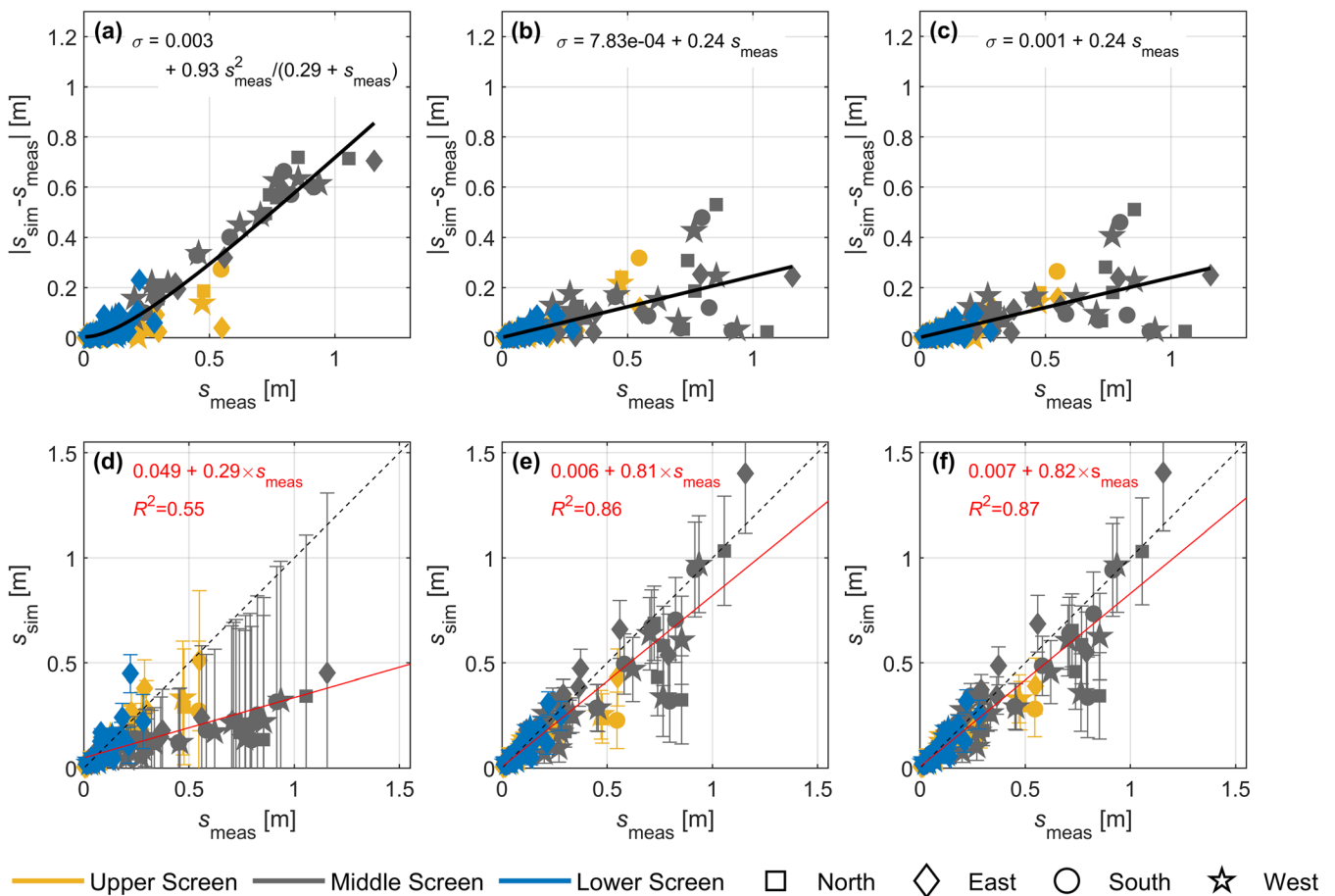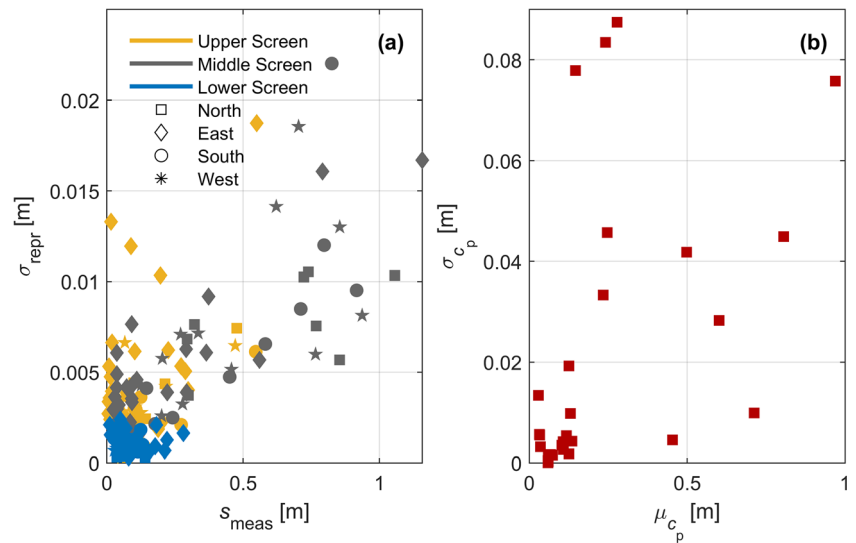

Fig. 7 Assessment of model results. **a–c** Absolute difference between simulated and measured drawdown versus the measured drawdowns of the 1-layer model (**a**), 3-layer model (**b**), and 5-layer model (**c**) and the thereto fitted error models. **d-f** Field measurements versus simulated results of the best fitting 1-layer model (**d**), 3-layer model (**e**), and 5-layer model (**f**) with error-bars according to the error model. The black-dashed diagonal lines represent the 1:1-identity lines. The red lines are linear regressions with corresponding equations and coefficients of determination R^2. The same plots are presented with log-log scales in section S4 of the ESM

model fit according to Eq. 10 captures the majority of measurements and its errors, while some outliers exist.

Figure 7d–f presents a comparison between measured and simulated drawdown values, s_{meas} and s_{sim}, for the best-fitting 1-, 3-, and 5-layer models associated with the error models determined in Fig. 7a–c, respectively. Figure 7d reveals that the 1-layer model systematically underestimates the drawdown in the second hydraulic test, in which water was extracted from the middle screen section, whereas the two multilayer models can decently fit all three hydraulic tests (Fig. 7e,f). As discussed earlier, the drawdown values were higher in the second test series than in the tests where water was extracted from the bottom and top screen, respectively. The multilayer models can reproduce this pattern by fitting a lower horizontal hydraulic-conductivity value to the middle depth of the aquifer (see the following), whereas the 1-layer model can either fit the high drawdown values of the second hydraulic test or the smaller drawdown measurements of the first and third tests. Linear regressions of the fitted versus the measured drawdown values confirm that the 1-layer model has systematic difficulties (slope of 0.29, coefficient of determination R^2 of 0.55), whereas the 3-layer model (slope of 0.81, $R^2 = 0.86$) and the 5-layer model (slope of 0.82, $R^2 = 0.87$) show a similar, overall satisfactory performance.

In general, the two multilayer models meet the majority of measured drawdowns (Fig. 7e,f), with only a few measurements falling far off the 1:1-identity line. As indicated by the different marker styles in Fig. 7e,f, there is no clear relationship between the directions of the measurement locations (north, east, west, and south of the pumping well) and the tendency towards over- or underestimating the measured drawdown values, obviating horizontal anisotropy.

Fitted parameter values

Tables 2 and 3 list the parameters estimated for all models. As mentioned before, the set of calibrated parameters include the radial and vertical hydraulic conductivities K_r and K_z of each horizontal layer considered in the model and for each extraction depth the drawdown s_{ref} at a virtual reference point to avoid computing the drawdown differences between true observation points in the steady-shape regime. While Table 3 contains the fitted values of s_{ref}, they do not have any real physical meaning.

Figure 4c includes the radial (blue lines) and vertical (red lines) hydraulic conductivities estimated for the 1-layer model (dotted lines), the 3-layer model (dashed lines), and the 5-layer model (solid lines). In both cases, the radial hydraulic conductivities are higher than the vertical counterparts, except for layer 4 of the 5-layer model (Table 2).

In the 5-layer model the radial and vertical hydraulic log-conductivities estimated for the sand layer 4 have considerably higher associated standard deviations $\hat{\sigma}_{\ln K}$ of estimation than all other conductivity estimates (Table 2), indicating that the measurements are insensitive to the conductivities estimated for that layer. Also, the fitted conductivity values of the 5-layer model reveal that the sand layer 2 is quite similar to layer 1. This hints that the sand layers, which were delineated by grain-size analysis, may not necessarily constitute distinct individual layers, but rather represent transition zones between the three main aquifer segments. This is the reason why the 3-layer model was set up. In this model, the investigated aquifer portion is subdivided into three main sections, with overall similar effective behavior as in the 5-layer model but without the need to fit insensitive parameters. The horizontal hydraulic conductivity K_r for the top and bottom of the investigated aquifer portion show significantly higher values than the middle section, whereas the fitted vertical conductivity K_z systematically increases with depth. The reduced horizontal conductivity of the middle section (layer 3 in the 5-layer model, layer 2 in the 3-layer model) can explain the larger drawdown values in the second test series, in which groundwater is extracted from the middle screen.

Table 2 Calibrated radial and vertical hydraulic conductivities and the associated standard deviations $\hat{\sigma}$ of estimation of each horizontal layer in the 1-, 3-, and 5-layer models. Best est.: value determined by gradient-based optimization (lsqnonlin); Lin. Prop.: linearized uncertainty propagation; MCMC: geometric mean and standard deviation determined by Markov-Chain Monte-Carlo method

Model	Layer	K_r [m/s]		$\hat{\sigma}_{\ln K_r}$		K_z [m/s]		$\hat{\sigma}_{\ln K_z}$	
		Best est.	MCMC	Lin. Prop.	MCMC	Best est.	MCMC	Lin. Prop.	MCMC
1-layer model	–	1.1×10^{-3}	1.1×10^{-3}	0.055	0.010	1.8×10^{-4}	1.8×10^{-4}	0.135	0.084
3-layer model	1	2.7×10^{-3}	2.7×10^{-3}	0.087	0.068	5.9×10^{-5}	6.8×10^{-5}	0.261	0.230
	2 (2∪3 of 5-layer model)	2.7×10^{-4}	2.7×10^{-4}	0.119	0.110	6.3×10^{-5}	6.3×10^{-5}	0.116	0.077
	3 (4∪5 of 5-layer model)	2.6×10^{-3}	2.6×10^{-3}	0.116	0.047	2.4×10^{-4}	2.3×10^{-4}	0.338	0.282
5-layer model	1	3.8×10^{-3}	3.7×10^{-3}	0.164	0.100	1.2×10^{-4}	1.3×10^{-4}	0.533	0.441
	2	7.0×10^{-4}	5.7×10^{-4}	1.213	0.829	5.1×10^{-5}	4.5×10^{-5}	0.357	0.298
	3	3.1×10^{-4}	3.2×10^{-4}	0.123	0.103	5.6×10^{-5}	5.5×10^{-5}	0.131	0.101
	4	1.0×10^{-7}	5.1×10^{-6}	571	2.425	8.4×10^{-5}	9.2×10^{-5}	1.038	0.508
	5	2.7×10^{-3}	2.7×10^{-3}	0.129	0.057	1.8×10^{-4}	1.7×10^{-4}	0.371	0.300

Table 3 Effective parameters and additional calibration results of the locally anisotropic 1-, 3-, and 5-layer models. The effective conductivities K_r^{eff} and K_z^{eff} as well as the anisotropy ratio ϑ are given as arithmetic means $\pm$ their standard deviations as obtained from the Markov-Chain Monte-Carlo simulations

Parameter	1-layer model	3-layer model	5-layer model
K_r^{eff} [m/s]	$1.1\times10^{-3}\pm1.1\times10^{-5}$	$2.0\times10^{-3}\pm6.6\times10^{-5}$	$2.2\times10^{-3}\pm8.3\times10^{-5}$
K_z^{eff} [m/s]	$1.8\times10^{-4}\pm1.6\times10^{-5}$	$1.1\times10^{-4}\pm1.5\times10^{-5}$	$1.0\times10^{-4}\pm1.5\times10^{-5}$
ϑ [−]	6.21 ± 0.48	18.0 ± 2.62	21.1 ± 3.25
$s_{\mathrm{ref,I}}/s_{\mathrm{ref,II}}/s_{\mathrm{ref,III}}$ [m]	0.071/0.064/0.062	0.032/0.028/0.022	0.028/0.023/0.019
a [m]	0.003	7.8×10^{-4}	0.001
b [−]	0.93	0.24	0.24
c [m]	0.29	0	0

Figure 8 shows the ensemble results of the effective horizontal and vertical hydraulic conductivities, K_r^{eff} and K_z^{eff}, respectively, and the anisotropy ratio $\vartheta = K_r^{\mathrm{eff}}/K_z^{\mathrm{eff}}$ for all three models. In the 1-layer model, the fitted horizontal and vertical conductivities K_r and K_z correspond to the effective values K_r^{eff} and K_z^{eff}, respectively. Upon upscaling of the 3- and 5-layer models, the estimated effective horizontal conductivities K_r^{eff} of the multilayer models is about twice as high as the estimated value of the 1-layer model (see Table 3), while the upscaled vertical conductivity K_z^{eff} of the multilayer model is about half the fitted value of the 1-layer model. As a consequence, the anisotropy ratio $\vartheta = K_r^{\mathrm{eff}}/K_z^{\mathrm{eff}}$ differs between the one- and multilayer models by a factor of more than 3 ($\vartheta \approx 6$ versus $\vartheta \approx 20$). Given the uncertainty of the estimates (see distributions in Fig. 8 and standard deviations in Table 3), the effective anisotropy of the 3- and 5-layer models is about the same. Note that both the small anisotropy ratio estimated by the 1-layer model and the large one by the multilayer models are within

reasonable ranges expected for fluvial deposits (Freeze and Cherry 1979; Kruseman and de Ridder 1994). Applying the upscaled anisotropic conductivity in a homogeneous model leads to similar misfits of the drawdown data as the fitted 1-layer model (data shown in section S5 of the ESM). That is, meeting the observations clearly requires a profile of hydraulic conductivity with vertical differences.

Table 3 contains the determined coefficients of the error models related to all three models. While the 1-layer model shows an absolute error of $a = 3$ mm and a relative error of $b = 93\%$, the multilayer models have similar absolute errors of ≤ 1 mm and a considerably lower relative error of 24%, proving to be the preferred model choice. Also, the specific error model of Eq. (10), with a smooth transition from an error that does not depend on the magnitude of the measurement to a linear dependence, is only needed for the 1-layer model. The errors of the two multilayer models can be expressed by standard expressions involving an absolute and a relative error only.
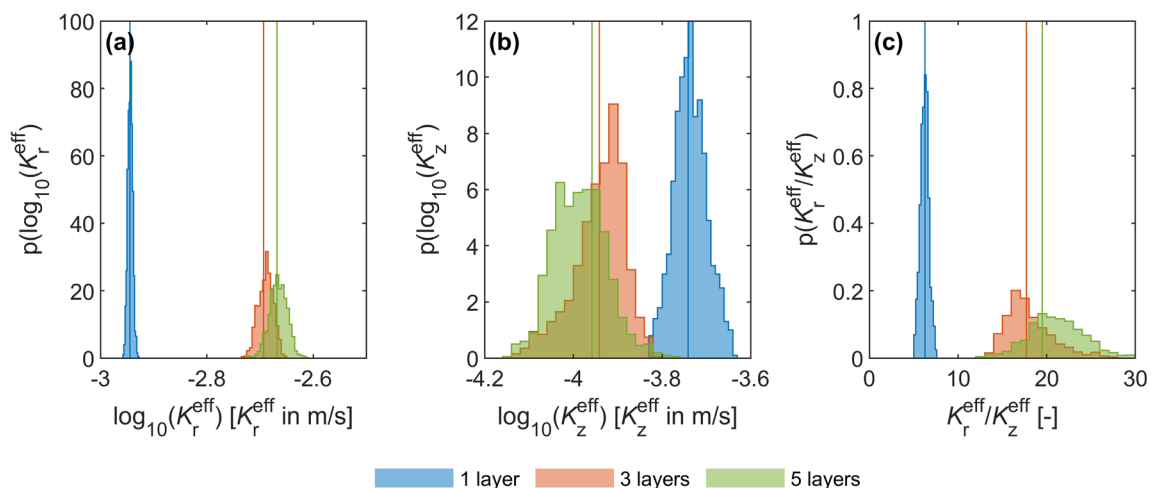

Fig. 8 Results of Markov-Chain Monte-Carlo sampling. Probability density functions of **a** effective horizontal hydraulic conductivity K_r^{eff}, **b** effective vertical hydraulic conductivity K_z^{eff}, and **c** anisotropy ratio $\vartheta = K_r^{\mathrm{eff}}/K_z^{\mathrm{eff}}$. Vertical lines: Best estimate of the gradient-based optimization

Conclusions

This work has tested an approach for estimating the hydraulic anisotropy induced by vertical heterogeneity in stratified aquifers. The approach is based on calibrating groundwater flow models using data of sequential hydraulic tests with partially penetrating wells, in which water is extracted from different aquifer depths and the hydraulic response is measured at different radial and vertical distances to an extraction screen (Maier et al. 2020). Pumping-test series with three extraction depths were performed in a fluvial gravel aquifer in South-West Germany, measuring more than 1,000 transient drawdown responses with a monitoring network of 58 observation points. These data were used to fit an anisotropic homogeneous model as well as locally anisotropic 3-layer and 5-layer models. The main target parameters were the radial and vertical hydraulic conductivities of each horizontal layer. The 3- and 5-layer models could reproduce the observed drawdown measurements considerably better than the 1-layer model, particularly because one of the three test series showed larger drawdown values, which could be attributed to pumping from a less permeable layer in the multilevel models, whereas the uniform model showed a systematic bias.

Based on the presented investigations, the following general recommendations for the design and analysis of pumping tests targeting hydraulic anisotropy are proposed:

1. The key element of the pumping tests is to extract water from a partially penetrating well, which induces a strong vertical flow component (at least in the vicinity of the pumping well), which is required to resolve the directional dependence of hydraulic conductivity in stratified aquifers.
2. The development of the pumping well considered in this study follows the development of an extraction well used for dewatering measures in a large construction pit. Performing the pumping tests is not restricted to such a large-diameter well or to the screen lengths of the well considered in the present study. The well diameter should be dimensioned based on the objective to induce a sufficiently large cone of depression which at the same time is within a measurable signal range.
3. Stressing the aquifer by extraction in different depths is mandatory. If water had been extracted only from a single depth (e.g., using the bottom well screen), the general vertical profile of hydraulic conductivity would most likely not have been detected.
4. Checking the reproducibility of the performed pumping tests by repeating the tests with different pumping rates and then rescaling the results to a common rate is highly recommended. Averaging over the repetitive tests has reduced the large data volume.

5. To avoid the challenges related to analyzing transient data or of reaching steady-state drawdown in field applications, the steady-shape analysis (Bohling et al. 2002; Bohling et al. 2007) is advantageous. Implementing a virtual reference point to compute drawdown differences is a reasonable alternative to the computation of drawdown differences based on pairs of true observation points, for which inherent measurement errors are propagated.
6. If sufficient data are available, it is preferable to resolve the main vertical structure of hydraulic conductivity over fitting a uniform effective conductivity tensor. A better agreement between simulated and measured drawdowns, avoiding systematic bias, was achieved with the multilayer models than with the single-layer model. Upon upscaling, the anisotropy ratio resulting from the multilayer model was considerably larger. Also, identifying layers of preferential flow may be important both in solute-transport applications and in flow applications in which the vertical flow component occurs mainly in a specific depth, as in the dewatering scenario considered in the authors' preceding theoretical study (Maier et al. 2020).
7. Selecting the right number and vertical positions of multiple layers is a challenge and may be prone to confirmation bias. As Zhao and Illman (2018) have illustrated, the use of information from prior hydrogeological investigations benefits model calibration. In this study, available lithologic information from the drilling profile of the pumping well proved to be a plausible decision guide for narrowing down potential layers, but in hindsight one layer per extraction screen turned out to be sufficient. Most likely, performing several flowmeter or direct-push injection-logging tests to see whether consistent layers of higher or lower conductivities exist across several vertical profiles would have been better for delineating hydraulically relevant layers than the grain-size data used here.
8. The true hydraulic conductivity in an aquifer will always be a spatially variable full 3×3 tensor. On the scale of pumping tests, however, horizontal variability is often smaller than the differences among the vertical layers. To justify the assumption of radial symmetry (neglecting horizontal heterogeneity and/or anisotropy), it was important to install observation wells in several directions from the pumping well.

Overall, the study has demonstrated the applicability of the proposed approach targeting the vertical variability and anisotropy of potentially stratified aquifers. Of course, the experimental effort of installing a multisection partially penetrating well and multilevel observation wells is considerably higher than the effort associated with fully-screened wells. This extra effort may only be justified in applications in which

either significant vertical flow is to be expected such as in riverbank-filtration setups or in the design of horizontal collector wells, or when the identification of preferential-flow layers is crucial, like in solute-transport applications.

Declarations

Conflict of interest None.

References

Bair ES, Lahm TD (1996) Variations in capture-zone geometry of a partially penetrating pumping well in an unconfined aquifer. Ground Water 34(5):842–852. https://doi.org/10.1111/j.1745-6584.1996.tb02079.x

Bear J (1972) Dynamics of fluids in porous media. Dover, New York

Bennett JP, Haslauer CP, Ross M, Cirpka OA (2019) An open, object-based framework for generating anisotropy in sedimentary subsurface models. Groundwater 57(3):420–429. https://doi.org/10.1111/gwat.12803

Bohling GC (2009) Sensitivity and resolution of tomographic pumping tests in an alluvial aquifer. Water Resour Res 45:W02420. https://doi.org/10.1029/2008wr007249

Bohling GC, Butler JJ, Zhan YY, Knoll MD (2007) A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities. Water Resour Res 43 (5):W05430. https://doi.org/10.1029/2006wr004932

Bohling GC, Zhan XY, Butler JJ, Zheng L (2002) Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. Water Resour Res 38 (12):1324. https://doi.org/10.1029/2001wr001176

Borghi A, Renard P, Courrioux G (2015) Generation of 3D spatially variable anisotropy for groundwater flow simulations. Groundwater 53(6):955–958. https://doi.org/10.1111/gwat.12295

Butler JJ, Dietrich P, Wittig V, Christy T (2007) Characterizing hydraulic conductivity with the direct-push permeameter. Ground Water 45(4):409–419. https://doi.org/10.1111/j.1745-6584.2007.00300.x

Cardiff M, Barrash W (2011) 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. Water Resour Res 47:W12518. https://doi.org/10.1029/2010wr010367

Chen XH, Burbach M, Cheng C (2008) Electrical and hydraulic vertical variability in channel sediments and its effects on streamflow depletion due to groundwater extraction. J Hydrol 352(3–4):250–266. https://doi.org/10.1016/j.jhydrol.2008.01.004

Chen XH, Song JX, Wang WK (2010) Spatial variability of specific yield and vertical hydraulic conductivity in a highly permeable alluvial aquifer. J Hydrol 388(3–4):379–388. https://doi.org/10.1016/j.jhydrol.2010.05.017

Coleman TF, Li YY (1996) An interior trust region approach for nonlinear minimization subject to bounds. SIAM J Optim 6(2):418–445. https://doi.org/10.1137/0806023

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. J Royal Stat Soc Series B-Methodol 39(1):1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dietrich P, Butler JJ, Faiss K (2008) A rapid method for hydraulic profiling in unconsolidated formations. Ground Water 46(2):323–328. https://doi.org/10.1111/j.1745-6584.2007.00377.x

Freeze RA, Cherry JA (1979) Groundwater. Prentice Hall, Englewood Cliffs, NJ

Gottlieb J, Dietrich P (1995) Identification of the permeability distribution in soil by hydraulic tomography. Inverse Problems 11(2):353–360. https://doi.org/10.1088/0266-5611/11/2/005

Heinz J, Aigner T (2003) Hierarchical dynamic stratigraphy in various quaternary gravel deposits, Rhine glacier area (SW Germany): implications for hydrostratigraphy. Int J Earth Sci 92(6):923–938. https://doi.org/10.1007/s00531-003-0359-2

Heinz J, Kleineidam S, Teutsch G, Aigner T (2003) Heterogeneity patterns of quaternary glaciofluvial gravel bodies (SW-Germany): application to hydrogeology. Sediment Geol 158(1–2):1–23. https://doi.org/10.1016/S0037-0738(02)00239-7

Hochstetler DL, Barrash W, Leven C, Cardiff M, Chidichimo F, Kitanidis PK (2016) Hydraulic tomography: continuity and discontinuity of high-K and low-K zones. Ground Water 54:171–185. https://doi.org/10.1111/gwat.12344

Klammler H, Hatfield K, Nemer B, Mathias SA (2011) A trigonometric interpolation approach to mixed-type boundary problems associated with permeameter shape factors. Water Resour Res 47:W03510. https://doi.org/10.1029/2010wr009337

Klammler H, Layton L, Nemer B, Hatfield K, Mohseni A (2017) Theoretical aspects for estimating anisotropic saturated hydraulic conductivity from in-well or direct-push probe injection tests in uniform media. Adv Water Resour 104:242–254. https://doi.org/10.1016/j.advwatres.2017.04.010

Koltermann CE, Gorelick SM (1996) Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour Res 32(9):2617–2658. https://doi.org/10.1029/96wr00025

Kruseman GP, de Ridder NA (1994) Analysis and evaluation of pumping test data, vol 47. International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands

Lessoff SC, Schneidewind U, Leven C, Blum P, Dietrich P Dagan G (2010) Spatial characterization of the hydraulic conductivity using

direct-push injection logging. Water Resour Res 46:W12502. https://doi.org/10.1029/2009wr008949

LGRB (2004) Geological map of Baden-Württemberg, Germany. Landesamt für Geologie, Rohstoffe und Bergbau Baden-Württemberg, Germany

Maier R, Gonzalez-Nicolas A, Leven C Nowak W, Cirpka OA (2020) Joint optimization of measurement and modeling strategies with application to radial flow in stratified aquifers. Water Resour Res 56 (7):2019WR026872. https://doi.org/10.1029/2019WR026872

Paradis D, Gloaguen E, Lefebvre R, Giroux B (2015) Resolution analysis of tomographic slug test head data: two-dimensional radial case. Water Resour Res 51(4):2356–2376. https://doi.org/10.1002/2013wr014785

Paradis D, Gloaguen E, Lefebvre R, Giroux B (2016) A field proof-of-concept of tomographic slug tests in an anisotropic littoral aquifer. J Hydrol 536:61–73. https://doi.org/10.1016/j.jhydrol.2016.02.041

Sanchez-Leon E, Leven C, Haslauer CP, Cirpka OA (2016) Combining 3D hydraulic tomography with tracer tests for improved transport characterization. Groundwater 54(4):498–507. https://doi.org/10.1111/gwat.12381

University of Tübingen (2022) Forschungsdatenportal FDAT [Research data portal FDAT]. http://hdl.handle.net/10900.1/a67994b5-19be-4934-9bdb-12fdc0aaef23. Accessed January 1, 2022

Wirsing G, Luz A (2007) Hydrogeologischer Bau und Aquifereigenschaften der Lockergesteine im Oberrheingraben (Baden-Württemberg) [Hydrogeological structure and aquifer properties of the unconsolidated rocks in the upper Rhine rift (Baden-Württemberg)]. Freiburg im Breisgau, Germany

Yeh TCJ, Liu SY (2000) Hydraulic tomography: development of a new aquifer test method. Water Resour Res 36(8):2095–2105. https://doi.org/10.1029/2000wr900114

Zhao ZF, Illman WA (2018) Three-dimensional imaging of aquifer and aquitard heterogeneity via transient hydraulic tomography at a highly heterogeneous field site. J Hydrol 559:392–410. https://doi.org/10.1016/j.jhydrol.2018.02.024

Zlotnik V, Ledder G (1996) Theory of dipole flow in uniform anisotropic aquifers. Water Resour Res 32(4):1119–1128. https://doi.org/10.1029/95wr03813

Zschornack L, Bohling GC, Butler JJ, Dietrich P (2013) Hydraulic profiling with the direct-push permeameter: assessment of probe configuration and analysis methodology. J Hydrol 496:195–204. https://doi.org/10.1016/j.jhydrol.2013.05.036