Investigating aquitard heterogeneity by inverse groundwater modelling at a drinking water production site

van Leer, Martijn D.; Zaadnoordijk, Willem J.; Zech, Alraune; Griffioen, Jasper; Bierkens, Marc F.P.

DOI
10.1016/j.envsoft.2025.106554

Publication date
2025

Document Version
Final published version

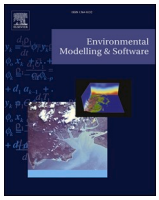
Published in
Environmental Modelling and Software

Environmental Modelling and Software 192 (2025) 106554

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Investigating aquitard heterogeneity by inverse groundwater modelling at a drinking water production site

Martijn D. van Leer [a,*], Willem J. Zaadnoordijk [b,c], Alraune Zech [d], Jasper Griffioen [b,e], Marc F.P. Bierkens [a,f]

[a] Department of Physical Geography, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, Netherlands
[b] TNO—Geological Survey of the Netherlands, P.O. Box 80015, 3508, TA, Utrecht, Netherlands
[c] Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands
[d] Department of Earth Sciences, Utrecht University, Princetonlaan 8a, 3584CB Utrecht, Netherlands
[e] Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a 3584 CB Utrecht, Netherlands
[f] Unit Subsurface and Groundwater Systems, Deltares, Daltonlaan 600, BK Utrecht, 3584, Netherlands

ARTICLE INFO

Keywords:
Aquitards
Heterogeneity
Upscaling
Calibration
Drinking water
Well field

ABSTRACT

This study investigates whether geostatistical properties of aquitards can be determined from data collected at a drinking water well field. A workflow adaptable to any drinking water extraction site with pumping and groundwater head data is developed. Using data from Budel, the Netherlands, a layered groundwater flow model is constructed and calibrated on hydraulic heads. A large number of realizations are generated for the aquitard, considering various heterogeneity parameters. These simulations are upscaled to the grid of the flow model, and their fit to observed heads is evaluated. Results show that optimal geostatistical parameters can be identified, even though many combinations reproduce observed heads. These parameters can be used to parameterize regional groundwater flow models. Particle tracking simulations show heterogeneity decreases contaminant breakthrough times while also decreasing the total flow through the aquitard. These findings emphasize the need to consider aquitard heterogeneity in risk assessments at drinking water production sites.

1. Introduction

Aquitards are low permeable layers that either separate or confine aquifers in multilayered aquifer-aquitard systems. Understanding their hydraulic properties is important for water resource management (Gurwin and Lubczynski, 2005), subsidence (Díaz-Nigenda et al., 2023; Soonthornrangsan et al., 2025) and contaminant transport (Filippini et al., 2020; Fjordbøge et al., 2017). Characterizing aquitards is challenging due to large variability in hydraulic conductivity (K) of the sediments that they consist of (clay, silt, peat) together with their low permeability. Lab measurements often differ by orders of magnitude for similar grain sizes (van Leer et al., 2023a) and tend to indicate lower K compared to in situ measurements, such as those from pumping tests (Hart et al., 2006; Zhuang et al., 2024). In addition, classical interpretations of pumping tests using analytical methods assume homogeneous layers with axisymmetric flow (e.g. Hantush and Jacob, 1955; Hemker and Maas, 1987) resulting into lumped parameters of which the physical meaning is unclear (Copty et al., 2008; Wu et al., 2005). Observation wells are typically placed only within the pumped aquifer, which makes aquitard parameters uncertain and also ambiguous because these might result from leakage through both the overlying and underlying aquitards. Many analytical solutions based on leaky aquifers do not account for drawdowns in overlying and underlying aquifers (Li and Neuman, 2007). With observation wells on both sides of the aquitard, tests still need to be run long enough for the drawdown to propagate through the aquitard (Neuzil, 1986) and to enable determination of both the storage coefficient and hydraulic conductivity of the aquitard (Fogg and Zhang, 2016; Van Der Kamp, 2001). Even if hydraulic properties of aquitards have been identified with conventional pumping tests, they do not provide information about spatial variability. It further remains unclear whether these values are appropriate for other flow patterns than radial flow towards the well (van Leer et al., 2023b).

To adequately characterize hydraulic properties of aquitards throughout spatial scales, a connection must be made between core scale measurements and local/regional scale data. Applying stochastic methods is a way to accomplish this by upscaling core scale

* Corresponding author
E-mail address: m.d.vanleer@uu.nl (M.D. van Leer).

https://doi.org/10.1016/j.envsoft.2025.106554
Received 14 February 2025; Received in revised form 9 May 2025; Accepted 3 June 2025
Available online 10 June 2025

measurements and borehole descriptions to model block scale. Geostatistics and upscaling methods have been extensively reviewed by De Marsily et al. (2005) and Sanchez-Vila et al. (2006), respectively. However, for aquitards specifically, it is difficult to determine the required parameters for this upscaling procedure, such as core scale hydraulic conductivity data, lithology distribution and variograms. Due to the variable nature of the hydraulic conductivity of core samples for low permeable materials, many samples are required to obtain an accurate probability density function per lithology or geological formation. The hydraulic conductivity of clays and silts are also sensitive to secondary processes such as compaction and soft soil deformation (van Leer et al., 2023a), which could cause deviations between e.g. the hydraulic conductivity distribution of the geological formation and the local distribution. Identifying the lithology distribution and estimating a variogram require a sufficient number of boreholes in the area of interest at different distances from each other, which is often not available. These variograms do however affect the percolation threshold and the presence of preferential flow paths, which have a strong effect on the effective hydraulic conductivity (Colecchio et al., 2021).

Many studies that used a form of stochastic analysis to obtain geostatistical parameters have focused on aquifers only (Demir et al., 2017; Firmani et al., 2006; Neuman et al., 2004; Zech et al., 2015). van Leer et al., 2023b were the first who identified horizontal and vertical correlation lengths for an aquitard with a pumping tests specifically designed for this purpose. However, such pumping tests are expensive to perform. At the same time, drinking water production sites are widespread with data routinely collected. As these sites are often made up of multiple wells where pumping rates between wells are often varied over time, they could possibly serve as an alternative to assess the heterogeneity of aquitard properties. Whether such data are indeed suitable for this purpose has, to our knowledge, not been investigated.

The main objective of this study therefore is to test whether it is possible to infer geostatistical parameters describing aquitard heterogeneity, specifically correlation lengths, from existing drawdown data around a drinking water production sites.

These sites are widespread, data is routinely collected, and aquitard heterogeneity is especially relevant there with regards to water sources, travel times and the risk of groundwater contamination. In addition, well fields that have been operational for extended time periods have had enough time for pumping induced drawdown to propagate through the surrounding aquitards. Moreover, the pumping in the individual wells usually varies in time, which makes the drawdown spatially variable that can be used as additional information. We address the following research questions.

- How can geostatistical parameters of aquitards be identified from a drinking water production site?
- How does heterogeneity of aquitards affect the groundwater flow pattern and travel times of water pumped at a drinking water production site?

2. Methods

2.1. Workflow

We developed a flexible modelling workflow adaptable to any drinking water extraction site in four steps. Step 1; a reference groundwater flow model, simulating hydraulic head responses to pumping based on initial parameter assumptions with relatively homogeneous aquitards and aquifers, is constructed and calibrated using field data. Step 2; heterogeneous realizations have been generated for the reference model setting using a range of geostatistical parameters. They replace the relatively homogeneous hydraulic conductivity in the aquitard in the calibrated reference model, the aquitard of interest. Step 3; the best performing realizations are selected based on their ability to reproduce observed head data. Step 4; particle tracking is run on the selected realizations to assess the hydrological implications of the heterogeneity of aquitards.

The workflow is implemented using Snakemake (Mölder et al., 2021). Snakemake was chosen for its ability to automate workflows, facilitate reproducibility, and manage computational resources, particularly for parallelizing the Monte Carlo simulations.

2.2. Data and site description

The workflow requires site-specific information, specifically the discharge per well, hydraulic heads at surrounding observation wells, borehole descriptions, and a multi-aquifer schematization. High resolution pumping and head data allow for the differentiation of pumping effects from individual pumping wells on observed hydraulic heads.

For the Budel well field, daily discharge per well and head data were available for a period of eight years. Observation wells have screens in the pumped aquifer, as well as in the overlying and underlying aquifers. Fig. 1 shows a hydrogeological profile with the arrangement of extraction and observation wells. Lithological descriptions are available for both the pumping and observation wells.

2.3. Groundwater flow model

A layered MODFLOW 6 model is created with FloPy (Hughes et al., 2024) and nlmod (Caljé et al., 2022). MODFLOW is a widely used modular groundwater modelling software that uses a control-volume finite-difference method to solve the groundwater flow equation. Nlmod is a Python package that uses open subsurface data of TNO – Geological Survey of the Netherlands to create MODFLOW 6 models. The hydrogeological characterization is based on the model 'H3O De Kempen' (Vernes et al., 2018), which is a transboundary hydrogeological model of the southern Netherlands and Flanders (Belgium) with the same format as the national hydrogeological model REGIS II (Hummelman et al., 2019), for which nlmod has been created. H3O De Kempen was selected instead of REGIS II, as the well field of Budel is located close to the Belgian border.

The groundwater flow model grid is 6.4 km by 6.4 km and unstructured with incrementally refined cells towards the pumping wells, ranging from 800 m at the boundaries to 25 m at the wells. An unstructured grid with finer cells near the wells allows for more detailed simulation of hydraulic gradients and flow patterns around the pumping wells while maintaining computational efficiency. A period of 4 months is selected from the available 8 years time series of observed heads and pumping discharges, to make sure constant General Head Boundaries (GHB) can be used at the horizontal boundaries of the domain and uncertain long-term dynamics are negligible. The initial heads of the GHB boundaries are based on long term average heads of the *LHM national groundwater flow model* (Lenssinck, 2023), to ensure compatibility with broader hydrogeological conditions in the region. The conductance is computed from the *H3O De Kempen* model as well, with a range of 5000 m outside the model domain. The transient model uses daily time steps, matching the temporal resolution of the observed head and discharge data.

2.4. Reference model calibration

The reference model is calibrated in two stages. First, the hydraulic conductivities (assuming a fixed anisotropy ratio of 10) and the heads of the GHB at the side boundary cells are calibrated in steady state. Second, the specific storage and hydraulic conductivity are calibrated in a transient simulation with fixed GHB heads from the steady-state calibration, focusing the optimization on specific storage linked to the dynamic head responses.

Hydraulic conductivity and specific storage are calibrated as multipliers on the initial values, so these will follow the regional trends of the original hydrogeological model *H3O De Kempen*, but do not include

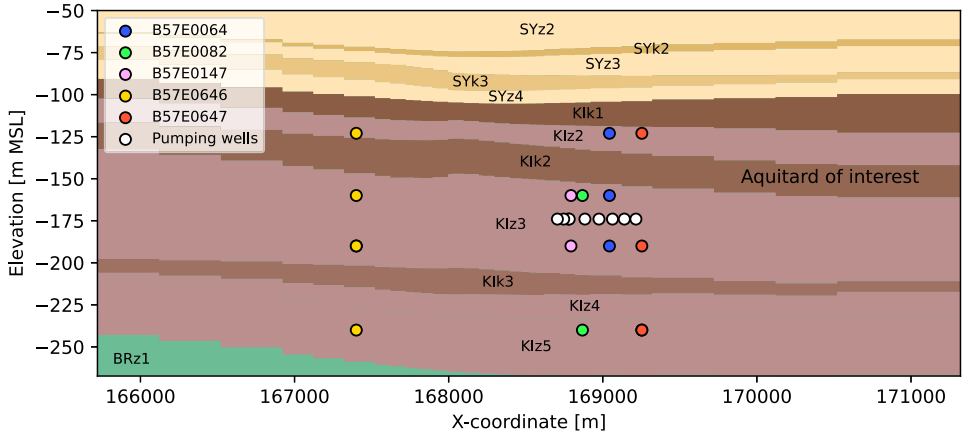


Fig. 1. West – East hydrogeological profile of the Budel well field (REGIS II, [20]). Shown are parts of the Early Pleistocene Stramproy (SY), Mio/Pliocene Kieseloolite (KI) and Miocene Breda (BR) Formations with aquitards (k, darker colours) and aquifer (z, lighter colours). Observation and extraction wells are represented as dots, even though the well screens of extraction wells cover most of the aquifer, while the observation wells are located either at the top or bottom of the aquifer. The x-coordinates are presented in Amersfoort/RD New format (EPSG: 28,992). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

variability at smaller scales. The parameters are optimized for the pumped aquifer as well as the aquitards and aquifers directly above and below the pumped aquifer. Parameters in other layers in the model are kept constant at the initial values. The Nelder-Mead downhill simplex algorithm (Gao and Han, 2012) is used to minimize the root mean squared error (RMSE) on observed heads in the calibrated aquifers.

2.5. Conditional simulations and tested parameters

We inversely infer the correlation lengths, core-scale conductivity, and lithology distribution of the aquitard using Monte Carlo simulations and subsequent realization selection. Using a brute force forward approach has two reasons: (i) correlation lengths are difficult to optimize using gradient-based methods, as altering the correlation length of a realization becomes a completely new realization, unrelated to the initial realization, and (ii) a correlation length which scores on average good might still result in low scoring realizations and vice versa, even if other parameters are kept constant.

Conditional geostatistical simulations are performed for the aquitard of interest, directly overlying the pumped aquifer. First, the section that lies within the aquitard is selected from the borehole description. Due to trends in the depth of the layer and its thickness, the depth of the borehole sections is flattened around the middle of the layer. This way values at the middle of the aquitard layer are spatially correlated with each other, rather than with overlying or underlying aquifers.

A density distribution of K values is constructed from core scale hydraulic conductivity values in the selected sections of the boreholes (Fig. 2). This distribution is bimodal with high K values related to sand and gravel, and low K values related to clay, silt and peat. The lithologies that correspond to the two modes of the distribution can be used as high and low conductivity indicators, which serve as the basis for the geostatistical simulations (cf. Bierkens, 1996).

Heterogeneous K-distributions are generated by three-dimensional sequential indicator simulations (Journel and Alabert, 1990) for a regular grid with a resolution of 25 m × 25 m x 1 m cells, covering the entire model domain. The simulations are conditioned on the borehole data, and established for a range of horizontal and vertical correlation lengths. A vertically anisotropic spherical indicator variogram model is used to capture the layered nature of the aquitard, and the correlation lengths we refer to are the ranges of the spherical horizontal (L_x) and spherical vertical variogram (L_z), assuming a geometric anisotropy. The parameter L_z was inferred from the borehole data (Appendix 1). We varied it with values of 7.5, 10, and 12.5 m to account for uncertainty in the

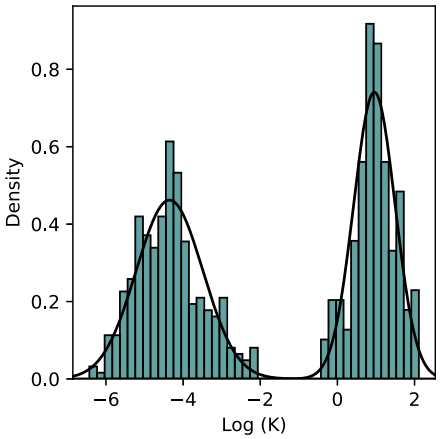


Fig. 2. Hydraulic conductivity distribution for the aquitard KIk2 at the Budel well field. For each lithological class K values are drawn from the K-distribution corresponding to the respective lithological class [20], weighted based on how frequently the lithologies occur in the borehole descriptions. Two distinct distributions emerge, one for low K values (clay, silt, peat) and one for high K values (sand, gravel).

variogram. We tested seven values for L_x ranging from 400 to 1200 m. In addition, the simulations are generated with varying fractions of the high and low conductivity indicators, to account for uncertainty of the fraction computed from the borehole data, we use the computed value, and adjustments of 0.05 in both positive and negative direction.

K values were randomly drawn from the high and low conductivity distributions and assigned to their respective indicators without correlation within the indicators. To account for uncertainty of the low K values in the core scale K distribution (K_{core}), the low K values $\log_{10}$(K) is adjusted with log values of 0.5, 1.5, 2 and 2.5. The high K distribution was kept constant as the uncertainty of K values for sands of the Kieseloolite Formation deposits is smaller, and the hydraulic resistance of the aquitard is mainly determined by that of the low conductivity sediments.

We generate 100 realizations for each combination of parameters, resulting in a total of 37,800 realizations. This ensures a wide range of geological and hydraulic uncertainties are tested while also accommodating for the variability between realizations within single parameter sets.

The realizations were generated for a structured three-dimensional

grid and upscaled to the unstructured two-dimensional aquitard layer grid in the MODFLOW model. We did this with a local flow model. Each local model spans the structured grid cells that fall within a cell of the unstructured flow model grid. We run MODFLOW again for the local upscaling with a simple flow model having a vertical constant head gradient to calculate the flux. These fluxes are then used with Darcy's Law to determine the hydraulic resistance c (ratio of head difference and flux) and equivalent K (equal to ratio of thickness and resistance). The latter form the upscaled hydraulic conductivity realization to be used in the calibrated MODFLOW model with the unstructured grid.

2.6. Selection of realizations

The aquitard above the pumped aquifer is replaced by the upscaled realizations, and they are assessed based on RMSE and a modified version for groundwater heads of the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009). RMSE is sensitive to fitting well to the mean of the modelled and observed time series, while the KGE also incorporates the dynamics of the time series. We adjusted the KGE as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (1)$$

in which r is the Pearson correlation coefficient, $\alpha = \sigma_m/\sigma_o$ and $\beta = \frac{C-\mu_m-\mu_o}{C}$. The original β term represents the ratio of modelled and observed means, which is not meaningful for groundwater heads. Instead, we use the absolute difference normalized by a constant C, chosen as 0.5 m to make β dimensionless and obtain similar magnitudes compared to r and α. Realizations with RMSE lower than the reference model or KGE higher than the reference model are deemed an improvement over the relatively homogeneous reference model and are selected for further analysis. These criteria ensure that realizations are selected not only when they match observed mean heads but also whether they replicate temporal dynamics adequately.

2.7. Particle tracking

We investigate the value of using heterogeneous aquitards around drinking water well fields using MODPATH 7 (Pollock, 2016) for particle tracking, both forward and backward in a steady state model. Forward tracking is used to calculate travel times of groundwater from the overlying aquifer to the pumping wells, highlighting potential vulnerabilities to contamination. The particles are evenly distributed within the overlying aquifer to cover all potential flow paths through the aquitard. Only the particles that end up in the pumping wells are selected to compute travel times. With backward tracking, we quantify the proportion of extracted water that has flowed through the aquitard. This provides insight into the origin of the extracted water. Particles are released uniformly at the well screen to represent all water entering the well. By comparing the heterogeneous models to the reference model, we determined whether modelling with heterogeneity leads to faster travel times or alters the sources of extracted water.

3. Results

3.1. Calibration of the reference model

The calibration results are shown in Table 1. The aquitard of interest Kik2 has the largest deviation from the initial values with a factor 6. The specific storage values between aquifers and aquitards remain very similar, which can be explained by the depth of the well field, where the specific storage will be low and the effect on the drawdown will be limited. The GHB heads are all lower than the initial values especially in the pumped aquifer, suggesting the pumping induced drawdown is still present at the boundaries. The steady state and transient calibration steps result in a RMSE of 0.12 m. Fig. 3 presents the time series of observed and modelled heads across the different aquifers. In the pumped aquifer, the modelled heads closely match the observed data, with minimal error. Thus, we consider further calibration of the aquifer layers unnecessary. The pumping-induced fluctuations are well represented across all layers, indicating that the transient dynamics caused by pumping are captured adequately. However, background dynamics are relatively large in the overlying and underlying aquitards, compared to the pumping induced drawdown and are not fully captured in the model.

Table 1
Calibration results of the reference model; see Fig. 1 for the aquifer and aquitard layer codes.

Parameter	Layer type	Layer	Type	Value
Hydraulic conductivity K	Aquifer	KIz2	Multiplier	3.68
Hydraulic conductivity K	Aquitard	KIk2	Multiplier	6.07
Hydraulic conductivity K	Aquifer	KIz3	Multiplier	0.40
Hydraulic conductivity K	Aquitard	Kik3	Multiplier	1.28
Hydraulic conductivity K	Aquifer	KIz4/KIz5	Multiplier	0.47
Specific storage	Aquifer	All	Multiplier	1.92
Specific storage	Aquitard	All	Multiplier	2.58
GHB head	Aquifer	KIz2	Addition	−0.68 m
GHB head	Aquifer	KIz3	Addition	−2.21 m
GHB head	Aquifer	KIz4	Addition	−0.36 m

3.2. Performance of Monte Carlo simulations and realization selection

2253 out of the 37,800 flow simulations with heterogeneous hydraulic conductivity realizations for the aquitard have either an improved (i.e., lower) RMSE or improved (i.e., higher) adjusted KGE (Fig. 4). In the remainder of this paper these will be referred to as 'selected realizations'. The majority of these realizations only show an improvement in the adjusted KGE. Fig. 5 shows the value distributions of the four tested parameters for the improved realizations. There are improved scores for most of the individual parameter values and there is no clear optimal value when each parameter is viewed independently. However, optimal combinations appear when the parameters are viewed together (Fig. 6) because the parameters are not independent. L_x shows correlation with all other properties. Low values correspond to low L_z values (Fig. 6c) a pattern which has been found in previous studies (van Leer et al., 2023b), while high L_x values correspond to a negative clay fraction adjustment (Fig. 6c). When L_x is plotted against K_{core} adjustment, it converges to a single optimum, from which the number of realizations decreases in every direction (Fig. 6a). This indicates that while there may be an optimum set of values for some parameter combinations, it is not straightforward to derive a singular optimal value for all parameters.

We use an iterative pruning process as alternative method to identify the optimal parameter sets within the four-dimensional parameter space. In each iteration the parameter value with the fewest realizations relative to the parameter value with the most realizations is removed from the ensemble. The results of the iterative pruning are shown in Fig. 7.

The optimal parameter values ($K_{core} = 1.5$, $L_x = 1000$ m, $L_z = 7.5$ m, clay fraction $= -0.05$) identified through the pruning refinement differ from those that appeared optimal when considered individually (Fig. 5) for L_x, L_z and clay fraction adjustment. However, they do align with the parameter combinations in Figs. 6a, 8e and 8f.

Among the parameters, K_{core} adjustment was found to be the most sensitive, converging to a single value in fewer iterations than the other parameters. This was expected as K_{core} governs the K values that are assigned to the model cells, while the other parameters control the spatial distribution mainly. L_x appears the least sensitive parameter as it requires the most iterations to converge to a single optimal value in the pruning process, meaning there is limited difference in performance between parameter values.

To account for the uncertainty in the resulting optimal parameters,

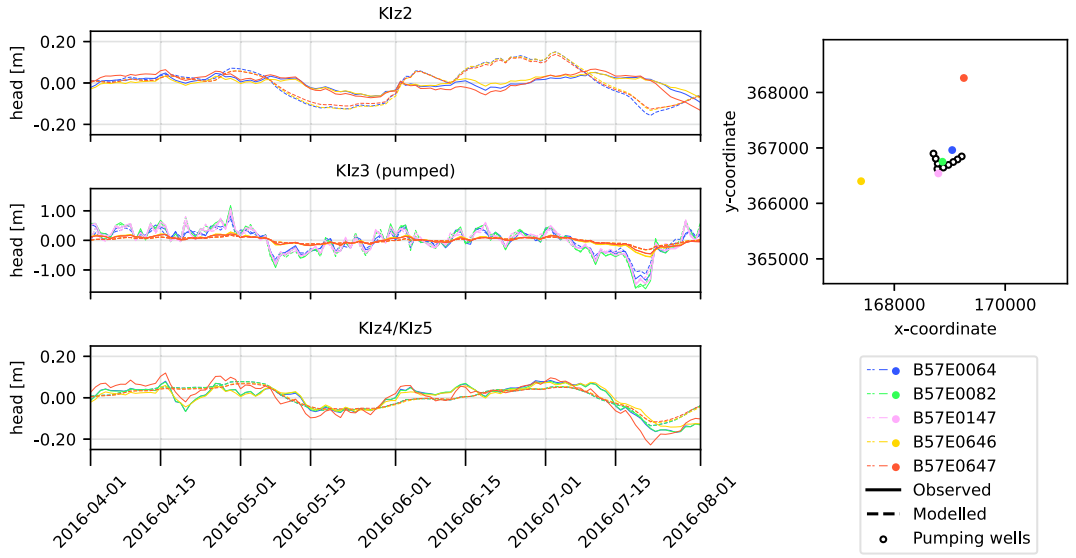


Fig. 3. Time series of observed and modelled heads in observation wells. The timeseries are grouped per aquifer (left). Top right shows the location of the observation wells. Coordinates are presented in Amersfoort/RD New format (EPSG: 28,992).

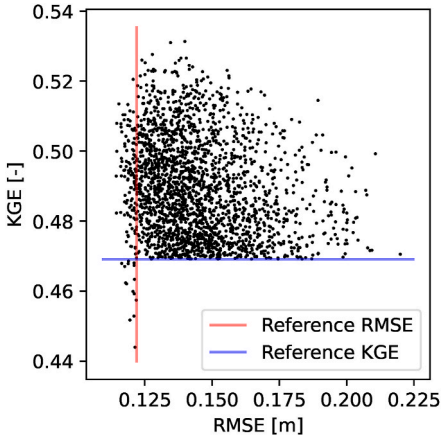


Fig. 4. Root Mean Squared Error (RMSE) and adjusted Kling-Gupta Efficiency (KGE, Equation (1)) of selected realizations. Each point represents a realization. The lines refer to the RMSE and KGE values of the reference model. Only realizations that show an improvement in these scores compared to the reference model are included.

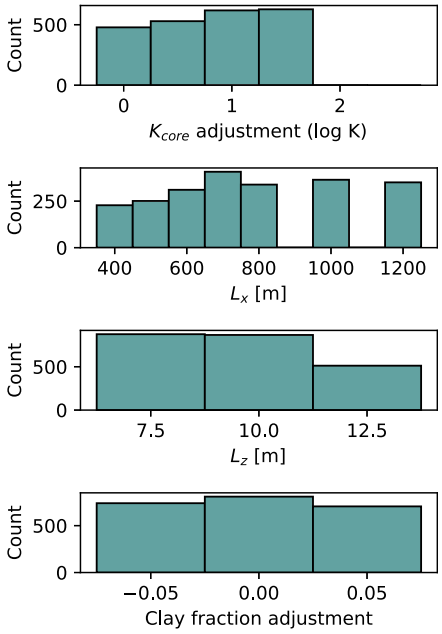


Fig. 5. Histograms of the parameter values (core scale hydraulic conductivity *Kcore*, horizontal correlation length *Lx*, vertical correlation length *Lz* and adjustment of clay fraction relative to the computed value) for the selected realizations that are an improvement compared to the reference model.

instead of selecting only the single optimal values, we selected an iteration where the relative difference between the minimum and maximum values was smallest and therefore the difference in performance between parameter values the smallest. In iteration 11, this criterion was met. The set of realizations remaining after iteration 10 was chosen as the 'optimal realizations' for further analysis with a range of values for L_x (700 m, 800 m, 1000 m and 1200 m) and clay fraction (−0.05 and 0), and single values for K_{core} (adjustment of 1.5) and L_z (7.5 m). This optimal set consists of 200 realizations.

3.3. Hydraulic resistance fields

The representative hydraulic conductivity K used in the MODFLOW simulations has been converted to a hydraulic resistance c of the aquitard using the aquitard thickness B via $c = B/K$. Fig. 8 shows the mean of the logarithm of the hydraulic resistance of the aquitard for the 2253 selected and the 200 optimal realizations.

Clear spatial patterns with high and low resistance areas can be identified, especially near wells. At locations without wells the values converge to the mean as none of the observation contain information regarding the hydraulic resistance at these locations.

The smallest grid cells around the wells have higher resistance values than surrounding, larger cells in both the selected realizations and the optimal realizations. The difference is approximately one order of magnitude in resistance values between the smallest (25 m × 25 m) and the largest scale (800 m × 800 m). We attribute this to the scale effect, where, in the upscaling procedure, larger cells show less resistance due to preferential flow paths at the smaller scale in the stochastic realization. This pattern is not visible in the original realizations before upscaling (SI 1). In such larger cells, flow tends to bypass the lowest K cells and goes through higher K cells during the upscaling process, leading to higher effective K values in the upscaled MODFLOW cells.

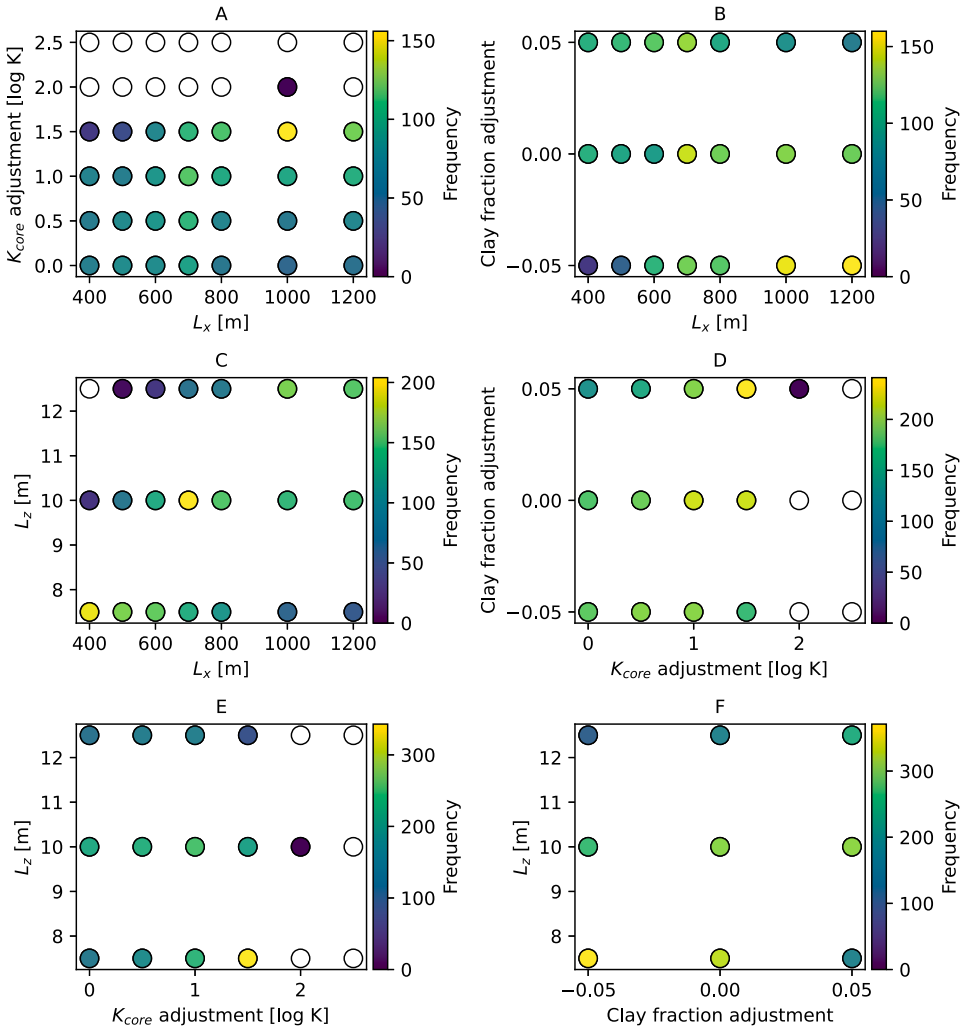


Fig. 6. Number of improved realizations in 2D histograms of the four aquitard properties that have been varied (core scale hydraulic conductivity *Kcore*, clay fraction, horizontal correlation length *Lx*, and vertical correlation length *Lz*). Yellow shows the largest number of improved realizations (note this value differs per histogram), blue the smallest. White dots are modelled properties that do not occur in the selected realizations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

This effect does not occur in the smallest MODFLOW cells, as these have only been upscaled in the vertical dimension and remained single cells in the horizontal dimensions.

The *c* values of the optimal realization set are lower than those of all selected realizations, and closer to the reference model. This shows that refining the selected realization set to the optimal realization set improves the usefulness of the realizations with regards to parameterization and upscaling.

The hydraulic resistance values of the three optimal realization (with the lowest RMSE) are shown in Fig. 9. Realizations before upscaling are shown in supplementary information (SI 2). The individual realizations do not show the same spatial pattern as the mean of the realizations. The scale effect is present but is less defined. While the spatial distribution of *c* values differ among the realizations, their distributions (Fig. 9c) are similar. This implies that further upscaling of these realizations will result in similar mean *c* values.

3.4. Particle tracking simulations

Figs. 10 and 11 show results of the particle tracking simulations with regard to the amount of water flowing through the aquitard and the travel times. When heterogeneous realizations of hydraulic conductivity in the aquitard are used, a smaller fraction of the abstracted groundwater originates from the aquitard above compared with the reference model (Fig. 10). However, water that does come from the aquitard above reaches the pumping wells more quickly than in the reference model (Fig. 11). This means that contaminants from the aquifers overlying or underlying the extraction reach the extraction wells faster than would be expected assuming a relatively homogeneous aquitard. Unfortunately, no data was available at the Budel site to validate these findings. However, as we argue in the discussion, the large differences between the homogeneous and heterogeneous results suggest that information on age distributions of the pumped water may help to further identify the correct heterogeneity.

4. Discussion

4.1. Reference model calibration

The reference model is calibrated using multipliers on the hydrogeological model parameters, and therefore relatively homogeneous. This means the choice of reference model impacts the selection of well-fitting realizations. As the main forcing stems from the discharge in the pumping wells, and external forcing is limited (e.g. recharge), the reference model calibration dependance on initial values for the calibration is limited, similar to how parameters can be derived from

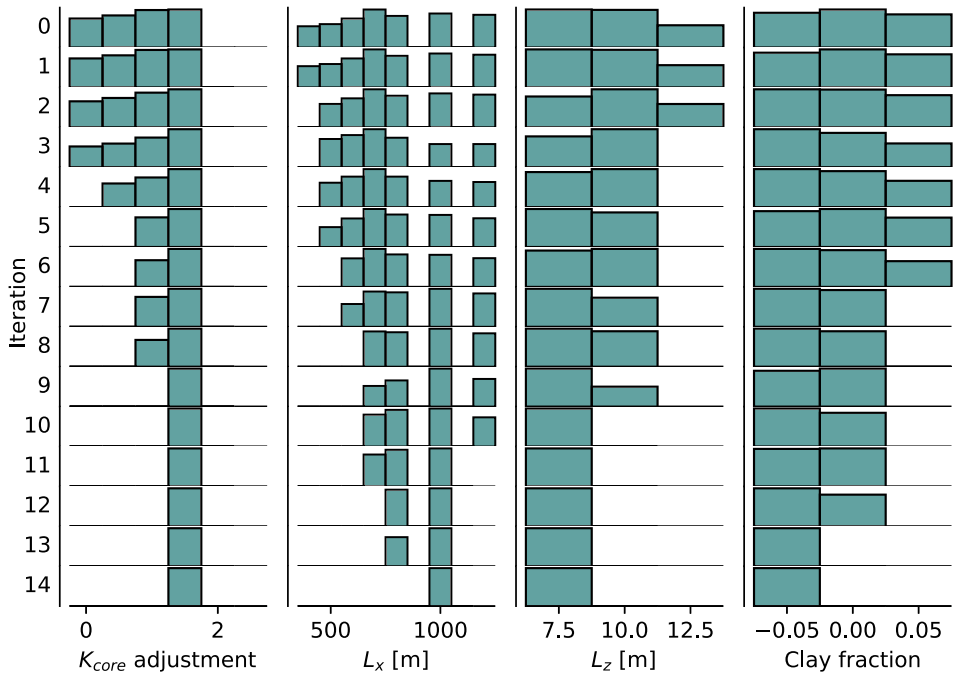


Fig. 7. Iterations of the successive refinement procedure. Iteration 0 shows histograms of the completed set of realizations that perform better than the reference model. The lowest scoring parameter value is removed in each iteration to obtain a defined optimum with only well performing parameters.

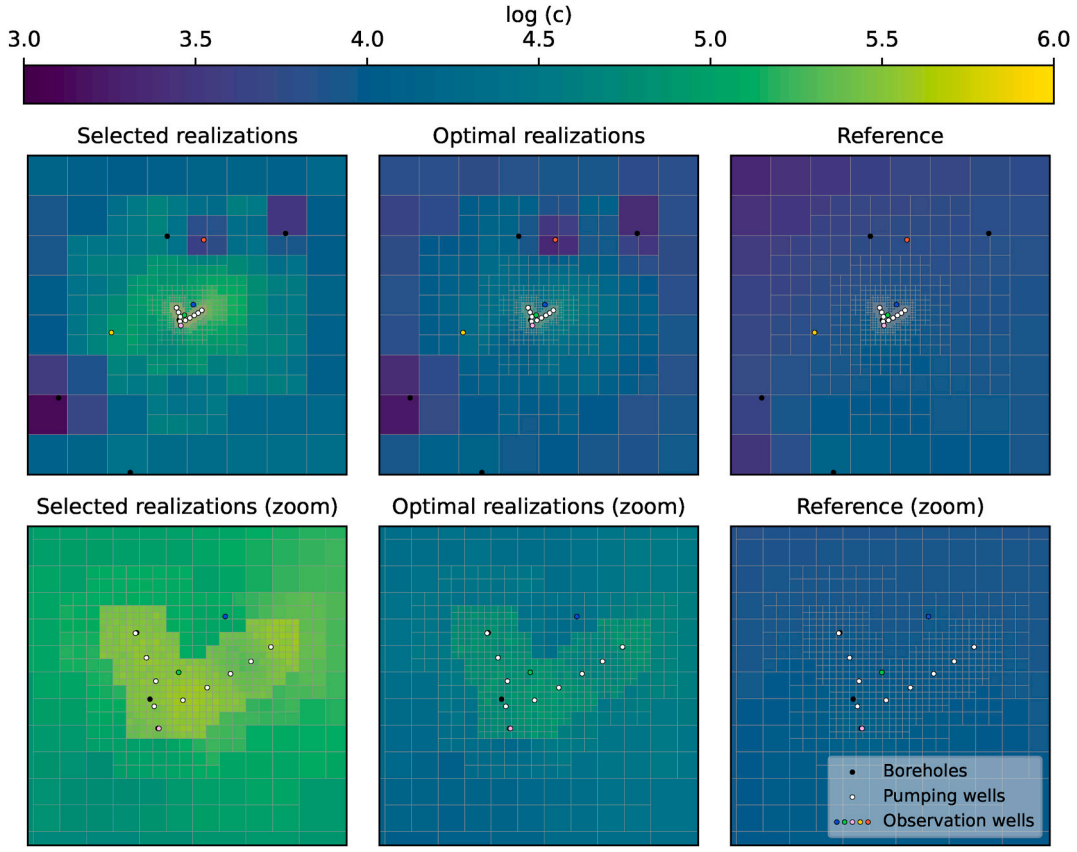


Fig. 8. Hydraulic log-resistance log(c) in the aquitard of interest. Left: mean over all selected realizations; middle: mean over optimal realizations; right: calibrated reference model. Top: full realization domain, bottom: zoom to the well field.

pumping tests unambiguously with analytical equations. The measured drawdown in observation wells however is mostly determined by the effective upscaled hydraulic conductivity of the aquitard. If the effective hydraulic conductivity of a realization is vastly different from the real value, it will heavily impact the observed drawdowns and therefore not result in a good performance score. This means that realizations that

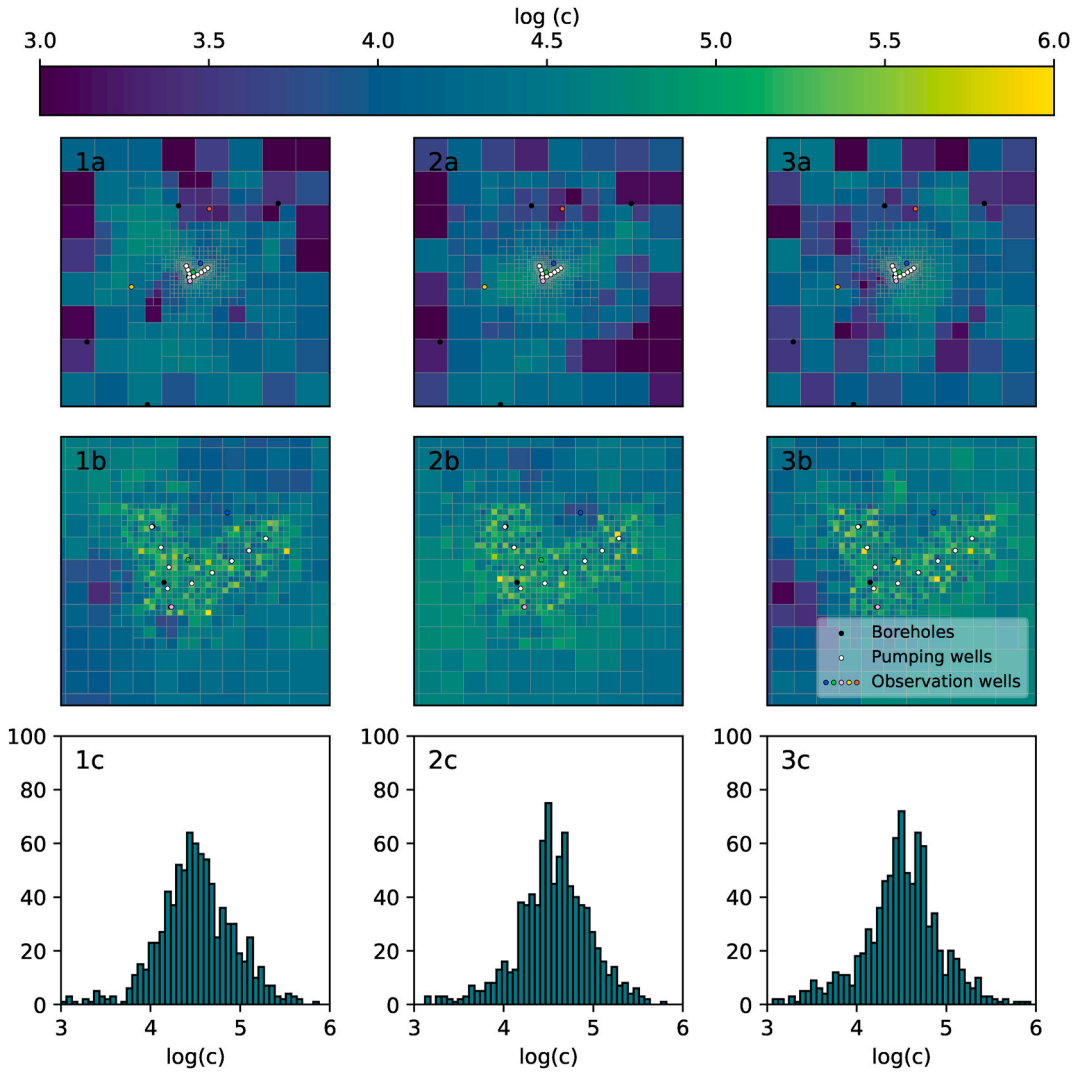


Fig. 9. Three realizations of the optimal realization set with the lowest RMSE. (a) Map of hydraulic resistance c for entire MODFLOW model, (b) zoom of the map around wells, (c) histogram of log (*c*) values.

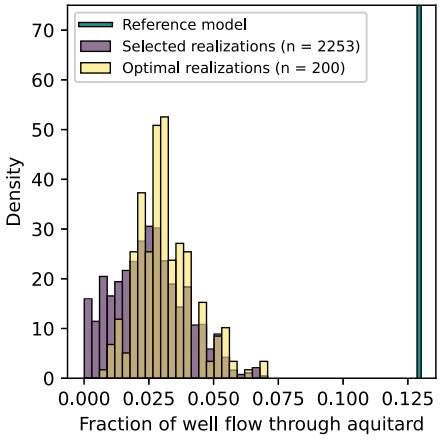


Fig. 10. Relative frequency of fraction of water flowing through the aquitard. These values result from the backwards particle tracking from the well. The reference model is shown in blue, the selected realizations (n = 2253) in yellow and the optimal realizations (n = 157) in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

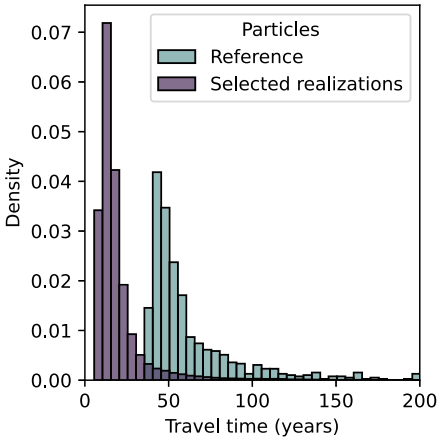


Fig. 11. Results of the forwards tracking simulation. The histogram shows the travel time for particles flowing through the aquitard of interest for both particles in the selected well-performing heterogeneous realizations (purple) and particles in the relatively homogeneous reference model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

perform well with a certain reference model are also likely to score well in reference models that are calibrated in a different way, or with different priors. In addition, due to the large number of selected realizations, it is unlikely that a change in the exact cutoff value changes the statistics of the parameters of the selected realizations.

4.2. Geostatistical model

The geostatistical model is developed from in situ data at the extraction site, such as lithologies from the borehole descriptions, and hydraulic conductivity distributions per lithology. In the case of Budel, drawing log-conductivity values based on the borehole descriptions results in two distinct distributions (Fig. 2). Both of these distributions are approximately Gaussian, which leads to the use of indicator simulations. Bi-modal distributions are not uncommon in aquitards (van Leer et al., 2023b), but other sites might exhibit a single-peak Gaussian distribution of log-hydraulic conductivity, where a sequential Gaussian simulation can be used.

In the case study of Budel, we use the correlation length that refers to one of the indicators. In the case of a log-Gaussian distribution this can also be the correlation length of the hydraulic conductivity values directly. With multi-modal distributions more correlation lengths exist, which makes calibration more challenging due to the increase in the number of parameters.

Other assumptions in the geostatistical model are stationarity and horizontal isotropy. However, effects of non-stationarity and anisotropy are partly covered by the fact that the realizations are conditioned on boreholes. If borehole data shows a trend, this will be taken into account in the realizations due to the conditioning. In practice it is difficult to know whether anisotropy is present, although this could potentially be calibrated, too.

4.3. Impact of heterogeneity in aquitards on model performance

The implementation of heterogeneous hydraulic conductivity in an aquitard within a groundwater flow model only slightly improves the model fit to data when calibrating on heads. However, it impacts the distribution of groundwater travel times and the origin of the groundwater in the pumping wells. The small difference in calibration measures, such as RMSE and KGE to the reference model is attributed to the insensitivity of heads to heterogeneity in general. Many realizations result in the same head distribution. This well-known phenomenon is confirmed in this study, as realizations with a wide range of parameters fit the observed heads well. However, further conditioning on flow velocities or travel times potentially reduces the range of plausible realizations, although such data is rarely available at drinking water extraction sites (Visser et al., 2013).

Gradient-based methods (e.g., Levenberg-Marquardt, Nelder-Mead (Gao and Han, 2012; Marquardt, 1963)) allow for finer optimization of parameters than brute force Monte Carlo. Monte Carlo methods, are better suited for exploring the parameter space for heterogeneity parameters but result in less specific optimal realizations compared to gradient-based calibration since their computational demands allow for a limited number of parameter values only. More efficient sampling schemes, such as Markov-Chain Monte Carlo (MCMC) were not deemed necessary for this study, as the Monte Carlo method resulted in good results with manageable computation times. Combining the strengths of both Bayesian and gradient-based approaches, such as generating realizations first and refining key parameters (K_{core}, high/low K fractions) with gradient-based calibration, potentially yields improved results but requires significant computational resources.

We applied adjusted KGE and RMSE as assessment metrics for the realizations. A larger number of realizations showed improvements in the adjusted KGE compared to RMSE. We attribute this to the fact that the reference model is calibrated with RMSE as objective function, leaving room for improvement in KGE. Using both metrics provides a more comprehensive evaluation, identifying realizations that better match observed dynamics (via KGE) and mean states (via RMSE), ultimately reducing noise in optimal parameter value distributions.

4.4. Scale effects

Within this study, variance in hydraulic conductivity between scales stems from spatial correlation and variance at smaller scales. We deal with three spatial scales: the core scale of lab measurements (~0.1 m), the model block scale of the stochastic simulation (25 m × 25 m x 1 m) and the model block scale of the groundwater flow model (25–800 m × 27 m). We assumed no difference between the core scale and the model block scale of the stochastic simulation, even though it is expected that K in these cells differ from the original 0.1 m scale conductivities due to scale effects. We computed effective conductivity values within the model cells as representative K values at the 25 m scale. The results show K increases by 2 orders of magnitude compared to the original core scale conductivities, partly due to inaccurate prior conductivity values, but potentially also due to the scale difference. While variance typically decreases with increasing scale due to averaging, this effect is not occurring in our method. However, this is not a real limitation, as variance within lithologies is generally several orders of magnitude smaller than variance between lithologies in aquitards.

We use an unstructured model grid which also showcases how K varies with scale. The difference between K values of the smallest (25 m) and the largest cells (800 m) is on average approximately 1 order of magnitude, suggesting that the scale effect is strong in aquitards. This can be attributed to the large variance in core scale hydraulic conductivities specifically in aquitards, especially between lithologies. With strong scale effects, horizontal variability in hydraulic resistance is an important factor and should be considered when upscaling from core scale to local/regional scale.

To summarize, the results of this study depend on the model block size and might change with other model grid resolutions. The dependence of hydraulic conductivity with scale is large for aquitards, which should be taken into account in the parameterization of regional groundwater flow models.

4.5. Application at other sites

The site chosen for testing the workflow features a relatively thick aquitard with large mean hydraulic resistance. As a result, pumping-induced drawdown in the overlying and underlying aquifers was small and highly sensitive to external dynamics (i.e., boundary conditions). Although the method successfully identified a range of optimal parameters for this site, sites with more permeable aquitards may be even better suited for this approach. At such sites, the calibration will be more sensitive to pumping-induced drawdown in the overlying aquifer, thereby reducing uncertainty regarding aquitard parameters.

For successful application of the workflow at other locations, sufficient piezometers must be present in both the pumped and overlying aquifers. This ensures the ability to identify leakage from the overlying and underlying aquifers and provides additional data points to characterize heterogeneity in hydraulic conductivity within the aquitard. Lithological descriptions are commonly available at the locations of piezometers, since these are installed in boreholes, although quality of the lithology description could vary depending on drilling method, personnel and availability of borehole logging. Other information may also be available like geophysical borehole logs providing information on (vertical) heterogeneity. In this study we could not differentiate whether high conductivity zones were identified based solely by conditioning on borehole data, calibrating on head observations, or a combination of both. However, geostatistical or statistical parameters such as correlation lengths of lithological classes and K_{core} distributions were derived primarily on the heads, as borehole data alone could not reliably provide this information at this site (Appendix 1).

We utilized four months of daily data, even though eight years of hourly records were available. Daily data was sufficient to capture pumping induced drawdown and variations in drawdown patterns from changing well discharges. We tested longer time series, but that introduced more external dynamics that are difficult to capture in the model. While four months is significantly longer than the duration of typical pumping tests, which usually last several hours to some days, the extended time frame allowed for sufficient dynamics on a daily scale to differentiate pumping-induced drawdown from external influences and to differentiate drawdowns of different wells. At drinking water extractions, it is challenging to establish the base level of heads that would occur if there was no pumping, especially if they have been active for a long time. Therefore, it is essential to have a time frame and resolution of head and discharge data that clearly capture pumping-induced head dynamics. As long as such dynamics are observed, it is possible to calibrate a model effectively for the well field. The necessary time frame and frequency of pumping and head data might vary per site, depending on local hydrogeological conditions and the available drawdown and discharge data, as well as the operational management of the groundwater pumps.

Once a homogeneous model is calibrated adequately, the workflow with heterogeneous aquitard realizations can be applied. Even with a limited number of observation wells, it is possible to identify a range of geostatistical parameters that reproduce the effective conductivity value of the reference model, albeit with significant uncertainty. Increasing the number of observation wells improves the identification of specific high or low conductivity zones within the area, improving the overall characterization of aquitard heterogeneity. We demonstrated this by the low resistance zones in the northern part of the domain where more data is are available, compared to the southeast part were the data is limited.

4.6. Implications for groundwater resources management

Our study has local and general implications with regards to regional-scale hydrogeological parameterization. Locally, modelling flow with heterogeneous aquitard representations improves model fit to observed heads. Apart from that, it impacts the water balance, since less water flows through the aquitard compared to the reference model. It also provides critical information for risk assessment of the quality of the extracted water as the travel time of part of the extracted water, and thus that of contaminants, is reduced. This implies that if the extraction is at risk of contamination, the time when a contaminant front arrives at the well might be severely overestimated. Since the uncertainty of the travel times is large, further investigations are recommended, e.g. with increased monitoring of water quality in wells or tracer tests. Potential management strategies in case of contamination risk include decreasing extraction rates, moving the wells to deeper aquifers or improving water treatment as well as restrictions on land use to prevent contamination.

Regionally, the optimal geostatistical parameter ranges can be used to parameterize groundwater flow models at various scales, as they are used to upscale from core scale to larger scales. Having more accurate regional flow models with better established uncertainty regarding spatial variability can improve water management decision making and operational management of the drinking water extraction.

5. Conclusion

We present a workflow for determining structural hydrogeological parameters of aquitards, such as lithology correlation lengths, core scale conductivity distributions and lithology fractions with observation data from drinking water well field. These parameters are challenging to derive from lithological descriptions or laboratory data, but they are critical for conducting risk assessments related to water quality and ensuring consistent scaling from core to regional flow.

Through model calibration and Monte Carlo simulations, realizations were identified that had a better fit between modelled and observed hydraulic heads. These realizations provided optimal value ranges for horizontal and vertical correlation lengths, core-scale hydraulic conductivity, and lithology fractions.

Particle tracking using high-scoring realizations revealed that incorporating heterogeneous aquitards results in a smaller fraction of water originating from above the aquitard within the model domain. However, the water that locally flows through the aquitard reaches the pumping wells more quickly. This finding underscores an increased risk of contamination at drinking water extraction sites due to faster transport of contaminants from overlying or underlying aquifers, compared to a simpler model using relatively homogeneous aquitards. The proposed workflow is adaptable to other drinking water production sites, provided that enough borehole descriptions and piezometers data are available, and sufficient pumping-induced dynamics are observable for both the pumped aquifer and the aquifers above and below aquitards. In addition, a sufficiently accurate subsurface characterization is required.

This approach offers a tool for improving the representation of heterogeneity in aquitards, contributing to better risk assessments and water management practices at and around drinking water production sites.

Further improvements in the parameterization of aquitards using drinking water extraction sites could be made with additional data types to condition the calibration, such as observations of flow velocity using e.g. active temperature sensing (Bakx et al., 2023) or tracer tests (Hendry et al., 2000).

Software and data availability section

The code that has been developed for this modelling framework are based on Python (version 3.11), and mainly relies on the packages flopy and nlmod. The source code can be found at https://github.com/MartijnVanLeer/ExtractionCalibrator. This repository is created by Martijn van Leer (m.d.vanleer@uu.nl) in 2024 and published in 2025. The required input and output data can be found at https://doi.org/10.5281/zenodo.14869963 (500 MB).

CRediT authorship contribution statement

Martijn D. van Leer: Writing – original draft, Visualization, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Willem J. Zaadnoordijk:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Alraune Zech:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Jasper Griffioen:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Marc F.P. Bierkens:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

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

6. Acknowledgements

The authors thank Brabant Water for providing the drawdown and discharge data from the Budel drinking water well field and TNO – Geological Survey of The Netherlands for funding this research.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106554.

Appendix A. Variogram derivation

Variogram models for lithologies within hydrogeological formations are difficult to derive. Figure A1 shows variograms and fitted spherical variogram models for the Budel site.

Vertical correlation is can be fitted decently, as each lag distance has multiple data points within each borehole. For lag distances of 15–20 m a worse correlation occurs. This can be attribute to the fact that clay, which mostly occurs in the center of the layer, is not correlated with the sand that is present on either side of the aquitard. The correlation improves again for lag distances of 20–25 m. This can be attributed to correlation between the sand on the top and bottom of the layer being correlated. Some uncertainty remains around whether this pattern for lag distances >10 should be accounted for in the variogram model, which can, subjectively, be chosen to have shorter correlation lengths than the best fit of 13.7. Note that the R^2 is relatively high with 0.94.

The horizontal variogram shows a different pattern. Each lag distance bin contains a limited number of values, due to the limited number of boreholes. This causes an overrepresentation of correlation between single boreholes. Badly correlated boreholes skew the variogram as there is only a limited number of sets of boreholes that have a certain lag size. Increasing the bin size to have adequate average values of semivariance for each bin size would result in bin sizes that are large compared to expected correlation lengths. Fitting a variogram model to this pattern results in a correlation length close to the smallest lag distance and a relatively poor fit ($R^2 = 0.38$).

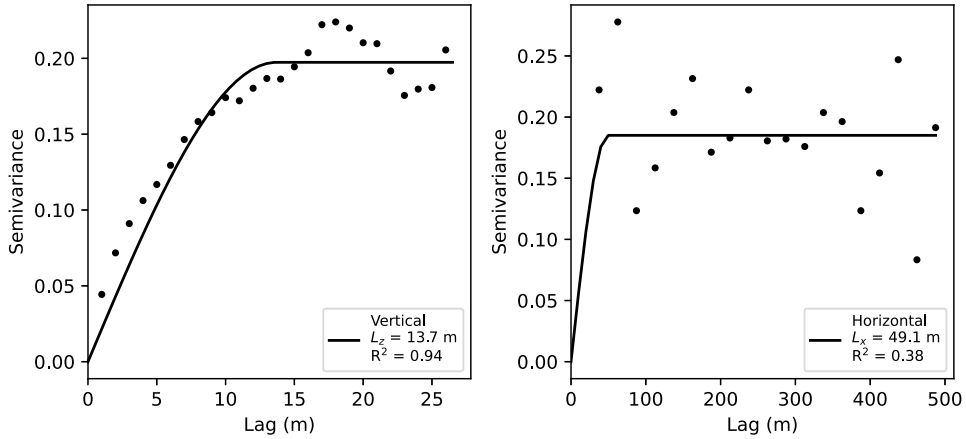


Figure 12. Horizontal (top) and vertical (bottom) semivariograms for high and low conductivity indicators. A spherical variogram without nugget is fitted through binned lags of borehole data.

Data availability

The data and source code are publicly available.

References

Bakx, W., Bense, V.F., Karaoulis, M., Oude Essink, G.H.P., Bierkens, M.F.P., 2023. Measuring groundwater flow velocities near drinking water extraction wells in unconsolidated sediments. Water (Lond. 1974) 15 (12). https://doi.org/10.3390/w15122167. Article 12.

Bierkens, M.F.P., 1996. Modeling hydraulic conductivity of a complex confining layer at various spatial scales. Water Resour. Res. 32 (8), 2369–2382. https://doi.org/10.1029/96WR01465.

Caljé, R., Schaars, F., Brakenhoff, D., Ebbens, O., 2022. Open source grondwatermodellering met MODFLOW 6. Stromingen 3. https://www.nhv.nu/stroming/nieuws/stromingen-20223-essay-calje/.

Colecchio, I., Otero, A.D., Noetinger, B., Boschan, A., 2021. Equivalent hydraulic conductivity, connectivity and percolation in 2D and 3D random binary media. Adv. Water Resour. 158, 104040. https://doi.org/10.1016/j.advwatres.2021.104040.

Copty, N.K., Trinchero, P., Sanchez-Vila, X., Sarioglu, M.S., Findikakis, A.N., 2008. Influence of heterogeneity on the interpretation of pumping test data in leaky aquifers. Water Resour. Res. 44 (11). https://doi.org/10.1029/2008WR007120.

De Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. Hydrogeol. J. 13 (1), 161–183. https://doi.org/10.1007/s10040-004-0432-3.

Demir, M.T., Copty, N.K., Trinchero, P., Sanchez-Vila, X., 2017. Bayesian estimation of the transmissivity spatial structure from pumping test data. Adv. Water Resour. 104, 174–182. https://doi.org/10.1016/j.advwatres.2017.03.021.

Díaz-Nigenda, J.J., Morales-Casique, E., Carrillo-García, M., Vázquez-Peña, M.A., Escolero-Fuentes, O., 2023. Importance of aquitard response time for groundwater management in multi-aquifer systems subject to land subsidence. Water Resour. Manag. 37 (13), 5367–5378. https://doi.org/10.1007/s11269-023-03611-z.

Filippini, M., Parker, B.L., Dinelli, E., Wanner, P., Chapman, S.W., Gargini, A., 2020. Assessing aquitard integrity in a complex aquifer – aquitard system contaminated by chlorinated hydrocarbons. Water Res. 171, 115388. https://doi.org/10.1016/j.watres.2019.115388.

Firmani, G., Fiori, A., Bellin, A., 2006. Three-dimensional numerical analysis of steady state pumping tests in heterogeneous confined aquifers. Water Resour. Res. 42 (3). https://doi.org/10.1029/2005WR004382.

Fjordbøge, A.S., Janniche, G.S., Jørgensen, T.H., Grosen, B., Wealthall, G., Christensen, A.G., Kerrn-Jespersen, H., Broholm, M.M., 2017. Integrity of clay till aquitards to DNAPL migration: assessment using current and emerging characterization tools. Groundwater Monitoring & Remediation 37 (3), 45–61. https://doi.org/10.1111/gwmr.12217.

Fogg, G.E., Zhang, Y., 2016. Debates—stochastic subsurface hydrology from theory to practice: a geologic perspective. Water Resour. Res. 52 (12), 9235–9245. https://doi.org/10.1002/2016WR019699.

Gao, F., Han, L., 2012. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Comput. Optim. Appl. 51 (1), 259–277. https://doi.org/10.1007/s10589-010-9329-3.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Gurwin, J., Lubczynski, M., 2005. Modeling of complex multi-aquifer systems for groundwater resources evaluation—swidnica study case (Poland). Hydrogeol. J. 13 (4), 627–639. https://doi.org/10.1007/S10040-004-0382-9/FIGURES/9.

Hantush, M.S., Jacob, C.E., 1955. Non-steady radial flow in an infinite leaky aquifer. Eos, Transactions American Geophysical Union 36 (1), 95–100. https://doi.org/10.1029/TR036I001P00095.

Hart, D.J., Bradbury, K.R., Feinstein, D.T., 2006. The vertical hydraulic conductivity of an aquitard at two spatial scales. Ground Water. https://doi.org/10.1111/j.1745-6584.2005.00125.x.

Hemker, C.J., Maas, C., 1987. Unsteady flow to wells in layered and fissured aquifer systems. J. Hydrol. 90 (3–4), 231–249. https://doi.org/10.1016/0022-1694(87)90069-2.

Hendry, M.J., Wassenaar, L.I., Kotzer, T., 2000. Chloride and chlorine isotopes (36Cl and δ37Cl) as tracers of solute migration in a thick, clay-rich aquitard system. Water Resour. Res. 36 (1), 285–296. https://doi.org/10.1029/1999WR900278.

Hughes, J.D., Langevin, C.D., Paulinski, S.R., Larsen, J.D., Brakenhoff, D., 2024. FloPy workflows for creating structured and unstructured MODFLOW models. Groundwater 62 (1), 124–139. https://doi.org/10.1111/gwat.13327.

Hummelman, J., Maljers, D., Menkovic, A., Reindersma, R., Vernes, R., Stafleu, J., 2019. Totstandkomingsrapport hydrogeologisch model (REGIS II). www.tno.nl.

Journel, A.G., Alabert, F.G., 1990. New method for reservoir mapping. J. Petrol. Technol. 42 (2), 212–218. https://doi.org/10.2118/18324-PA.

Lenssinck, L., 2023. Veranderingsrapportage LHM 4.3.

Li, Y., Neuman, S.P., 2007. Flow to a well in a five-layer system with application to the oxnard basin. Groundwater 45 (6), 672–682. https://doi.org/10.1111/j.1745-6584.2007.00357.x.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11 (2), 431–441. https://doi.org/10.1137/0111030.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., 2021. Sustainable data analysis with Snakemake. F1000Research 10 (33). https://doi.org/10.12688/f1000research.29032.1.

Neuman, S.P., Guadagnini, A., Riva, M., 2004. Type-curve estimation of statistical heterogeneity. Water Resour. Res. 40 (4). https://doi.org/10.1029/2003WR002405.

Neuzil, C.E., 1986. Groundwater flow in low-permeability environments. Water Resour. Res. 22 (8), 1163–1195. https://doi.org/10.1029/WR022I008P01163.

Pollock, D.W., 2016. User guide for MODPATH Version 7—a particle-tracking model for MODFLOW. In: *Open-File Report* (2016–1086). U.S. Geological Survey. https://doi.org/10.3133/ofr20161086.

Sanchez-Vila, X., Guadagnini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. (1985) 44 (3), 3002. https://doi.org/10.1029/2005RG000169.

Soonthornrangsan, J.T., Bakker, M., Vossepoel, F.C., 2025. Linked data-driven, physics-based modeling of pumping-induced subsidence with application to bangkok, Thailand. Groundwater 63 (2), 145–159. https://doi.org/10.1111/gwat.13443.

Van Der Kamp, G., 2001. Methods for determining the in situ hydraulic conductivity of shallow aquitards – an overview. Hydrogeol. J. 9 (1), 5–16. https://doi.org/10.1007/S100400000118.

van Leer, M.D., Zaadnoordijk, W.J., Zech, A., Buma, J., Harting, R., Bierkens, M.F.P., Griffioen, J., 2023a. Dominant factors determining the hydraulic conductivity of sedimentary aquitards: a random forest approach. J. Hydrol. 627, 130468. https://doi.org/10.1016/j.jhydrol.2023.130468.

van Leer, M.D., Zaadnoordijk, W.J., Zech, A., Griffioen, J., Bierkens, M.F.P., 2023b. Estimating hydraulic conductivity correlation lengths of an aquitard by inverse geostatistical modelling of a pumping test. Hydrogeol. J. 31 (6), 1617–1626. https://doi.org/10.1007/s10040-023-02660-3.

Vernes, R., Deckers, J., Bakker, M.A.J., Bogemans, F., De Ceukelaire, M., Doornenbal, J. C., den Dulk, M., Dusar, M., Van Haren, T.F.M., Heyvaert, V.M.A., Kiden, P., Kruisselbrink, A.F., Lanckacker, T., Menkovic, A., Meyvis, B., Munsterman, D.K., Reindersma, R., Rombaut, B., ten Veen, J.H., et al., 2018. Geologisch en hydrogeologisch 3D model van het Cenozoïcum van de Belgisch-Nederlandse grensstreek van Midden-Brabant/De Kempen.

Visser, A., Broers, H.P., Purtschert, R., Sültenfuß, J., De Jonge, M., 2013. Groundwater age distributions at a public drinking water supply well field derived from multiple age tracers (^{85}Kr,^{3}H/^{3}He, and^{39}Ar). Water Resour. Res. 49 (11), 7778–7796. https://doi.org/10.1002/2013WR014012.

Wu, C.-M., Yeh, T.-C.J., Zhu, J., Hau Lee, T., Hsu, N.-S., Chen, C.-H., Folch Sancho, A., Yeh, T.J., Zhu, J., Lee, T.H., Hsu, N., Chen, C., Sancho, A.F., 2005. Traditional analysis of aquifer tests: comparing apples to oranges? Water Resour. Res. 41 (9), 1–12. https://doi.org/10.1029/2004WR003717.

Zech, A., Arnold, S., Schneider, C., Attinger, S., 2015. Estimating parameters of aquifer heterogeneity using pumping tests—implications for field applications. Adv. Water Resour. 83, 137–147. https://doi.org/10.1016/j.advwatres.2015.05.021.

Zhuang, C., Yan, L., Kuang, X., Zhan, H., Illman, W.A., Dou, Z., Zhou, Z., Wang, J., 2024. Statistical characteristics of aquitard hydraulic conductivity, specific storage and porosity. J. Hydrol. 643, 132066. https://doi.org/10.1016/j.jhydrol.2024.132066.