Estimating hydraulic conductivity correlation lengths of an aquitard by inverse geostatistical modelling of a pumping test

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

DOI
10.1007/s10040-023-02660-3

Publication date
2023

Document Version
Final published version

Published in
Hydrogeology Journal

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

Hydrogeology Journal
https://doi.org/10.1007/s10040-023-02660-3

PAPER

Estimating hydraulic conductivity correlation lengths of an aquitard by inverse geostatistical modelling of a pumping test

Martijn D. van Leer[1] · Willem Jan Zaadnoordijk[2,3] · Alraune Zech[4] · Jasper Griffioen[2,5] · Marc F. P. Bierkens[1,6]

Received: 5 December 2022 / Accepted: 12 June 2023

Abstract
Aquitards are common hydrogeological features in the subsurface. Typically, pumping tests are used to parameterize the hydraulic conductivity of heterogeneous aquitards. However, they do not take spatial variability and uncertainty into account. Alternatively, core-scale measurements of hydraulic conductivity are used in geostatistical upscaling methods, for which their correlation lengths are needed, but this information is extremely difficult to obtain. This study investigates whether a pumping test can be used to obtain the correlation lengths needed for geostatistical upscaling and account for the uncertainty about heterogeneous aquitard conductivity. Random realizations are generated from core-scale data with varying correlation lengths and inserted into a groundwater flow model which simulates the outcome of an actual pumping test. The realizations yielded a better fit to the pumping test data than the traditional pumping test result, assuming homogeneous layers are selected. Ranges of horizontal and vertical correlation lengths that fit the pumping-test well are found. However, considerable uncertainty regarding the correlation lengths remains, which should be considered when parameterizing a regional groundwater flow model.

Keywords Aquitards · Geostatistics · Heterogeneity · Hydraulic properties

Introduction

Aquitards are important hydrogeological features, as they play a key role in groundwater resource assessment (Gurwin and Lubczynski 2005), contamination transport (Ponzini et al. 1989), land subsidence (Zhuang et al. 2017), salinization (e.g. Van et al. 2022) subsurface energy storage (Sommer et al. 2015) and radioactive waste disposal (Hendry et al. 2015). Although the importance of aquitards is widely recognized (Hart et al. 2006; Keller et al. 1989), many stochastic evaluations of hydrogeological field studies focus on the characterization of aquifers and neglect aquitards (Fogg and Zhang 2016). This study aims to improve the stochastic parameterization of aquitards specifically.

To parameterize the hydraulic conductivity of hydrogeological units, typically in-situ-field-scale measurements, such as slug and pumping tests, are performed (Hantush and Jacob 1955; Keller et al. 1989; Neuman and Witherspoon 1972). However, the drawdown is generally only measured in the pumped aquifer and experiments are typically not run long enough to detect drawdowns outside the pumped aquifer, which results in poor estimates of aquitard parameters compared to aquifer parameters (Fogg and Zhang 2016; Neuzil 1986, 1994). Also, the aquifers and aquitards are typically assumed to be homogeneous and the flow to be axisymmetrical, yielding biased and ambiguous results (Berg and Illman 2015; Huang et al. 2011; Kuhlman et al. 2008; Wen et al. 2010; Wu et al. 2005). Thus, if identified at all, mean values of the aquitard's hydraulic resistance

Published in the special issue "Geostatistics and hydrogeology".

✉ Martijn D. van Leer
m.d.vanleer@uu.nl

1 Department of Physical Geography, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, The Netherlands

2 TNO - Geological Survey of the Netherlands, P.O. Box 80015, 3508 TA Utrecht, Netherlands

3 Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands

4 Department of Earth Sciences, Utrecht University, Princetonlaan 8a, 3584CB Utrecht, The Netherlands

5 Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, The Netherlands

6 Unit Subsurface and Groundwater Systems, Deltares, Daltonlaan 600, 3584 BK Utrecht, The Netherlands

Published online: 08 July 2023

do not provide information about spatial variability nor a measure of uncertainty of pumping test results. There are a few studies that performed stochastic analysis of pumping tests to identify correlation lengths (Demir et al. 2017; Firmani et al. 2006; Neuman et al. 2004; Zech et al. 2015); however, they focused on aquifers rather than aquitards. For cases where the description of the hydraulic conductivity of an aquitard needs to include spatial variability, methods have been developed to connect local measurements to a subregional variability or to regional-scale representative values within a geostatistical framework (De Marsily et al. 2005 and references therein). These local measurements typically consist of permeameter tests on core-scale samples (Alexander et al. 2011; Bierkens 1996; Keller et al. 1989). These tests may yield values that are not in line with field or model block-scale conductivities (Alexander et al. 2011; Batlle-Aguilar et al. 2016; Gerber et al. 2001; Hart et al. 2006; Zhao and Illman 2018), which can be caused by stress perturbations during collection, transportation and laboratory installation (Clark 1998) as well as geological heterogeneities which are not represented in the samples, such as fractures, dikes, sand streaks and interbeds (Hart et al. 2006; Keller et al. 1989; Van Der Kamp 2001; Bierkens and Weerts 1994; Meriano and Eyles 2009). Because of the latter, geostatistical upscaling is needed to relate the core-scale measurements to larger scales (Sanchez-Vila et al. 2006 and references therein), which is typically done by generating hydraulic conductivity realizations from core-scale measurement distributions with a certain spatial correlation, and running a groundwater flow model to derive the block-scale hydraulic conductivity (e.g. Pickup et al. 1994; Sarris and Paleologos 2004; Fleckenstein and Fogg 2008). In this way spatial variability in lithology and corresponding hydraulic conductivity can be accounted for; however, information about the vertical and lateral correlation lengths of the lithology or its corresponding hydraulic conductivity multi-point probability distributions needs to be known to adequately upscale the hydraulic conductivity from core to model block-scale within a geostatistical framework.

A large amount of data is needed to derive semivariograms (Weerts and Bierkens 1993) and accurate probability density functions (Khan and Deutsch 2016). If this is not available, estimates have to be used based on expert geological knowledge. However, inaccurate semivariograms, and in particular semivariogram ranges or correlation lengths, result in inaccurate block-scale hydraulic conductivity values. This is especially an issue with aquitards, as the core-scale hydraulic conductivity values of aquitard material, such as clay and peat, can vary several orders of magnitude (Neuzil 1994).

The objective of this study is to combine pumping tests and geostatistical upscaling to investigate whether pumping test observations can be used to obtain information about the spatial correlation of the lithology within aquitards. The procedure can be used to inversely estimate aquitard geostatistics. The paper is organized as follows. In the method section, the pumping test setup and the geological architecture of aquitards and aquifers at the site are described. Next, the core-scale conductivity data deemed representative of the aquitard under study, the groundwater flow model used and the stochastic method to estimate unknown correlation lengths of the aquitard are introduced. Following, the results are presented, which are then discussed in detail. The paper closes with a summary and conclusions.

Methods

Pumping test site and setup

A pumping test was performed at a site in Schouwen-Duiveland, province of Zeeland, in the southwestern Netherlands (Figs. 1 and 2). Three aquifers and three aquitards are present at this location. A schematic profile of the geology and lithology at the test location is shown in Fig. 3. For the pumping test, one extraction well and four infiltration wells were installed in the second aquifer that lies directly below the aquitard under study (aquitard 2; Fig. 2). The wells were fully penetrating this aquifer. A total of 20 piezometers were installed in the first, second and third aquifers. Five pumping tests were performed with varying configurations concerning discharges and distribution of extracted water over infiltration wells. The discharges during the pumping periods are listed in Table 1.

The second aquitard, the one used for this study, is a Holocene tidal deposit stratigraphically classified as Wormer Member, which is part of the Naaldwijk Formation (TNO-GDN 2022). The well logs provide information about the depth of the aquitard's top. A gradual transition from sandy to more clayey sediments is observed, while the base of the aquitard is much more pronounced, with an assumed thickness of 3 m (Fig. 3). The probability distribution of the measured core-scale hydraulic conductivities for this unit is bimodal, consisting of an approximately log-normal distribution for sand samples, and 4–5 orders of magnitude smaller log-normal distribution of clay, sandy clay and clayey sand samples (Fig. 4). The core-scale samples have not been collected at the test site, but rather in other locations where the Wormer Member is present. Correlation lengths cannot be directly determined from the samples as they have not been collected at the pumping test location. Instead, they originate from distributed locations several kilometers away from each other, thereby also disabling variogram analysis. The borelog descriptions from the drilling for the pumping

Fig. 1 Location of the pumping test site

test contained insufficient information for determining correlation lengths. The distributions are deemed representative, because of the similar composition and depositional environments between the pumping test site and the Wormer Member at other locations.

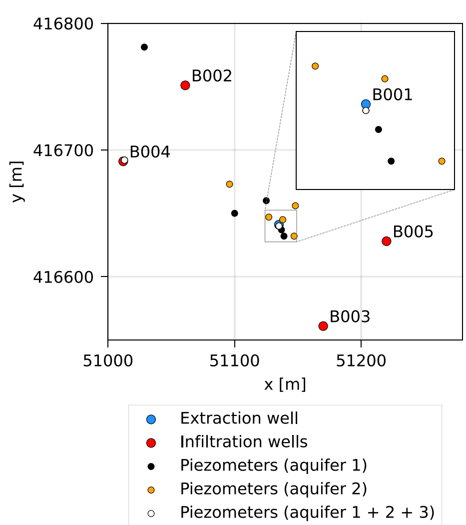


Fig. 2 Configuration of pumping test wells and piezometers. The area as shown in the inset is 25 m × 25 m

Groundwater flow models

Two groundwater models were created for evaluating the pumping test using MODFLOW 6 (Langevin et al. 2017) with FloPy (Bakker et al. 2016). The first, a reference model, consists of five layers with homogeneous hydraulic conductivity values; the second model is set up with a heterogeneous second aquitard, consisting of multiple layers.

Reference model

The model domain is 4,000 m × 4,000 m and consists of five layers. The computational grid has cells of 160 m × 160 m at the boundaries. Cell size gradually decreases to 5 m × 5 m for the inner domain of 500 m × 500 m around the pumping and extraction wells. Within a radius of 20 m around the wells, the cell size further decreases gradually down to 0.07 m at the wells. All cells that fall within the borehole diameter of 324 mm are given a constant hydraulic conductivity of 500 m/day in the pumped aquifer.

No anisotropy was assumed, as during a pumping test the flow in aquifers is predominantly horizontal, and the flow in aquitards is mainly vertical. The model is calibrated using five periods of pumping. It is assumed the

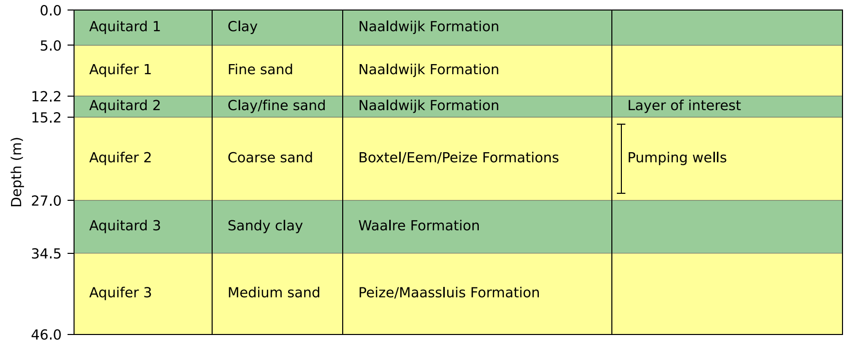


Fig. 3 Hydrogeological schematization of the subsurface at the pumping test site

test reaches steady state within each period's duration. The impact of this assumption is discussed in section 'Discussion', and results in five drawdown data points per piezometer. The hydraulic conductivity values for each model layer are obtained by calibration using the Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) to obtain hydraulic conductivity values of the model layers. As an objective function for the calibration, the root mean squared error (RMSE) is minimized:

$$\mathrm{RMSE} = \sqrt{\frac{1}{5n}\sum_{i=1}^{n}\sum_{j}^{5}\left(m_{ij} - o_{ij0}\right)^2} \tag{1}$$

where n is the number of piezometers, m_{ij} is the modelled drawdown and o_{ij} is the observed drawdown at piezometer i for configuration j. The piezometers in the active wells are not taken into account to prevent effects regarding non-Darcian flow and discretization issues due to the circular shape of the well, while the model grid consists of squared cells.

Heterogeneous model

The heterogeneous model is a refined version of the reference model. Heterogeneous distributions of hydraulic conductivity values were created for the aquitard of interest (aquitard 2) within the inner domain of 500 m × 500 m around the wells. The aquitard was also subdivided into 30 layers of 0.1 m to increase vertical resolution. All other layers were assigned the hydraulic conductivity values from the calibrated reference model.

The conductivity realizations for the aquitard section were generated by unconditional geostatistical simulation. According to borehole data in the area, the aquitard is composed of 56% clay (this class includes sandy clay and clayey sand) and 44% sand. These two contrasting lithologies result in two well-differentiated distributions of measured core-scale conductivities (Fig. 4). Due to the 4–5 orders of magnitude conductivity contrasts between the two classes, the lithology is assumed to be the main source of the conductivity heterogeneity of the aquitard. Thus, instead of generating hydraulic conductivities, the occurrence of sand and clay were simulated. A spherical indicator semivariogram model of sand/clay occurrence with geometric axisymmetric anisotropy was used, with a larger horizontal than vertical range. The semivariogram range is what is termed "correlation length" in the following. A sequential indicator simulation was applied (Journel and Alabert 1990) using gstat in R (Pebesma 2004) through the Python package rpy2. Horizontal and vertical correlation lengths ranging from 20 to 100 m and 0.5–2 m have been used, respectively. Examples of the generated fields are shown in Fig. 5. A total of 50 realizations were generated for each combination of correlation lengths adding up to 5,850 realizations in total. Because lithology is the dominant source of heterogeneity and because the variation of core-scale hydraulic conductivity within the lithoclasses can be assumed to be much smaller than the cell size (5 m × 5 m × 0.1 m; cf. Bierkens 1996), a constant representative value is assigned to each lithoclass. Here, an analytical formula was used to up scale the core-scale conductivity fluctuations that were proposed by Matheron (1967):

Table 1 Well discharges during pumping tests. Negative values denote extraction, positive values infiltration

Configuration	Duration (days)	Discharge of wells (m^3/h)				
		B001	B002	B003	B004	B005
1	70	–28.36	9.21	5.08	9.53	4.54
2	5	–22.51	0	11.64	0	10.87
3	3	–22.56	0	11.72	10.84	0
4	4	–19.56	9.59	0	0	9.97
5	10	–18.69	6.18	3.34	5.94	3.23

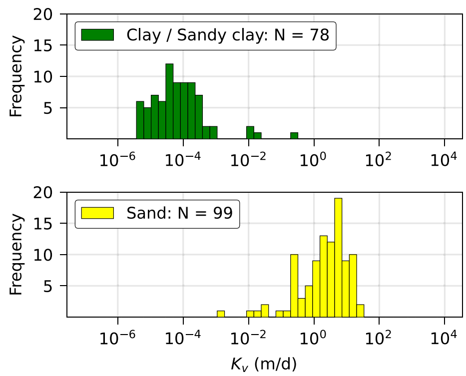


Fig. 4 Core-scale hydraulic conductivity measurements for clay-rich and sand samples for the Wormer member (aquitard 2)

$$K_{\text{cell}} = K_{\text{g}} \times \exp(\sigma_Y^2/6) \quad (2)$$

where K_{cell} is the resulting cell hydraulic conductivity, K_{g} is the geometric mean, and ${\sigma_Y}^2$ is the variance of the logarithm of the core-scale hydraulic conductivity distribution. This formula assumes the hydraulic conductivity variogram within the cell to be isotropic. Any anisotropy will follow from different horizontal and vertical correlation lengths at the scale of multiple cells and not from within the cells.

Assessment Framework

The performance of the simulated head distribution for each run of the heterogeneous aquifer model is evaluated and ranked according to the RMSE (Eq. 1) between the observed heads of the pumping test and each heterogeneous model realization.

Only the realizations with an RMSE lower than the RMSE of the reference model are evaluated. These realizations account for spatial variability and its uncertainty while outperforming the homogeneous model.

Table 2 Optimization results of homogeneous pumping test model

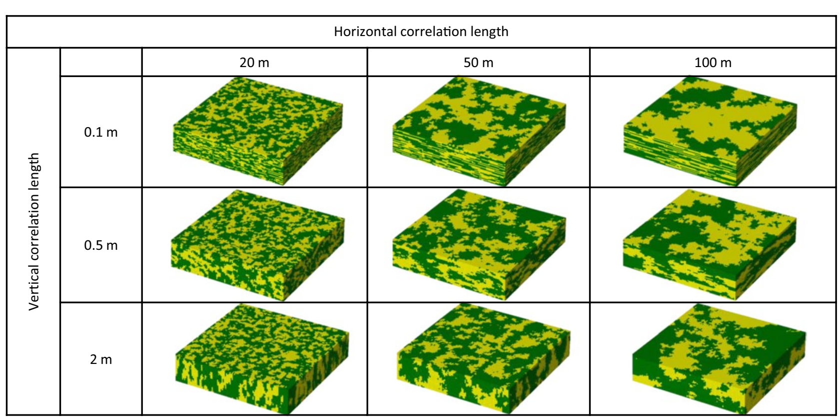
Layer	Hydraulic conductivity K (m/day)	Relative standard error of ln(K)
Aquifer 1	3.89	8.91%
Aquitard 1	6.09 e-3	4.54%
Aquifer 2	23.7	0.53%
Aquitard 2	6.44 e-2	14.38%
Aquifer 3	6.77	9.96%

The representative hydraulic conductivities in the aquitard that belong to the best fitting realizations are determined in two ways. First, the drawdown results of each of the realizations are used as a synthetic reality, which are used to calibrate the reference model again to obtain a representative hydraulic conductivity for the entire aquitard. In addition, the representative vertical block hydraulic conductivity of the realizations was computed with a traditional numerical upscaling method. For this, a MODFLOW model with constant head boundaries on the top and bottom of the aquitard at hand is used to model flow through the aquitard under uniform flow conditions. The hydraulic conductivity is calculated from the average flux through the aquitard via Darcy's law.

Results

Table 2 shows the details of the calibration of the homogenous model. The RMSE of the calibrated homogeneous model is 0.054 m, and there are 209 (out of 5,850) realizations of the heterogeneous model with a lower RMSE than this. Their correlation lengths (L) fall mostly within a range of L_{H} = 20–60 m and L_{V} = 0.875–2.0 m for horizontal and vertical directions, respectively. Details are displayed in Fig. 6. Note that the mean RMSE for each combination of correlation lengths is within a small range from 0.052 to 0.054 m. The maximum number of

Fig. 5 Examples realizations of the aquitard's sand and clay distribution for three vertical and three horizontal correlation lengths used in the heterogeneous aquitard model. Green represents clay/sandy clay/clayey sand, and yellow represents sand



realizations with identical horizontal and vertical correlation lengths performing better than the homogeneous model is 10 with $L_H = 50$ m and $L_V = 1.625$ m. Low average RMSE values are also mainly located close to these values.

The representative aquitard conductivities K of the 209 best fitting realizations are calculated through two upscaling schemes. First, synthetic pumping tests are conducted on the heterogeneous fields and the corresponding representative conductivity for each realization identified. The second way is to run a flow simulation on each realization with a spatially uniform head distribution across the aquitard and identify the corresponding mean conductivity. The resulting frequency distributions of representative K values for the 209 realizations are shown in Fig. 7.

The results show that the hydraulic conductivity values from the synthetic pumping tests are close to the value found in the actual field pumping test which was identified by calibrating the homogeneous model. However, the upscaling scheme, assuming uniform flow conditions on average (typically the standard procedure), results in representative values spread out over several orders of magnitude, with higher values than found in the pumping tests.

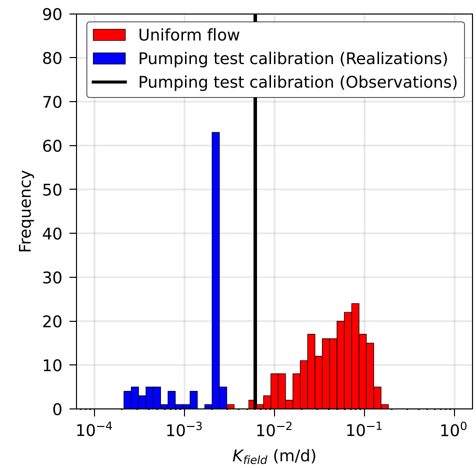


Fig. 7 Histogram of hydraulic conductivity for the best fitting realizations using a spatially uniform head difference across the aquitard, and the hydraulic conductivity obtained with the interpretation of a synthetic pumping test. The black line shows the hydraulic conductivity value obtained by calibrating the homogeneous model for the observed drawdowns

Discussion

Best fits

The suite of best fitting realizations (Fig. 6) belongs to a range of correlation length values rather than a single optimum. There were also realizations with the same combination of correlation lengths that fit the pumping test worse than the homogeneous model. This is partly due to the probabilistic nature of the approach and partly related to the fact that a finite number of realizations have been used in the Monte Carlo approach. A larger number of realizations might have made the location of the optimal values more obvious, but the general pattern is clear. In addition, drawdowns were only measured above and below the aquitard so, in the model, the aquitard may be considered as a single layer. Averaging the internal heterogeneity causes large uncertainties in the correlation lengths. This was also observed in other studies that attempted to determine correlation lengths from field data such as Xiao et al. (2021), Abellan and Noetinger (2010); Gautier and Noetinger (2004); Hoeksema and Kitanidis (1984). The small variation in mean RMSE values in Fig. 6 is partly caused by the use of all piezometers, including those in the pumped aquifer (aquifer 2) and those in the aquifers above (aquifer 1) and below (aquifer 3; see Figs. 2 and 3). Particularly, drawdowns in aquifer 3 are insensitive to changes in the composition of the aquitard. In addition, the average RMSE values in Fig. 6 contain all fits better than the homogeneous model, including realizations that are just slightly better. The results show that for a horizontal correlation length, a limited range of vertical correlation lengths results in a good fit. Lower horizontal correlation lengths are associated with lower vertical correlation lengths; however, the broad range of results indicate that a single semivariogram model does not give the full uncertainty range of the model-block-scale hydraulic conductivity, as the uncertainty of the correlation lengths itself are usually not taken into account in upscaling procedures.

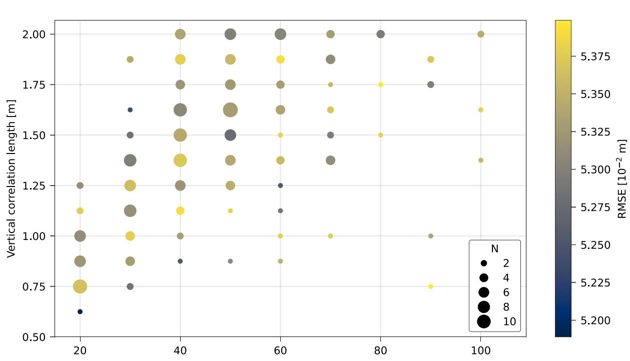


Fig. 6 Number of realizations that perform better than the homogeneous model (in terms of RMSE) for each combination of horizontal and vertical correlation length. The size of the dots indicates the number, while the colour indicates the mean RMSE of these realizations

The modelled correlation lengths range from below to above the percolation threshold, which is the point where connected pathways of sand exist from the top to the bottom of the aquitard, forming a connected object (Colecchio et al. 2020, 2021). The results show that the aquitard is close to the percolation threshold as part of the ensemble realizations have percolation, but not in all of them, thus explaining the wide range of correlation lengths that result in good fits. Percolation at larger distances from the well can partially explain the difference between the two representative conductivities in Fig. 7, as it will result in high conductivities for the complete field with uniform flow, but is not observed in the pumping test. In aquitards where the hydraulic conductivity of the lithoclasses is not as well differentiated, percolation will be less of an issue and might lead to smaller ranges of correlation lengths that result in good fits. The representative hydraulic conductivity of the aquitard also depends on the flow pattern, as the interpreted hydraulic conductivity from a pumping test is not identical to that derived from uniform flow (Fig. 7). The representative conductivity from a pumping test is dominated by the heterogeneity close to the pumping wells due to the steep gradients during pumping. Thus, the selection of the best fitted realizations is insensitive to the existence of preferential flow paths in the aquitard at large distances (Copty et al. 2008). However, such high conductivity will increase the representative hydraulic conductivity under uniform flow conditions, which underlines previous findings (Sanchez-Vila et al. 2006; Bierkens and Van der Gaast 1998) that representative hydraulic conductivities are very sensitive to the flow geometry applied and essentially a non-local problem. As shown, this particularly applies to aquitards were differences between representative conductivity for uniform and radial flow can be of orders of magnitude. It is also akin to previous work on aquifer conductivities that distinguish between apparent values close to the pumping test being different than representative conductivities for larger scales (Dagan 2001). Thus, care should be taken when assessing the representative hydraulic conductivity of aquitards from a pumping test with a limited sphere of influence. It is likely that a larger pumping test would be needed to find representative conductivities for strictly vertical flow through the aquitard. For instance, with an aquitard thickness of ~3 m and an aquifer thickness of 11.8 m, the conductivities in Table 2 result in a leakage factor $\sqrt{\mathrm{KD} \times c} = 317$ m (in which KD is the transmissivity and c is the hydraulic resistance), while the dimension of the test site being monitored is 250 m × 250 m. It should be noted that the infiltration wells decrease the area in which vertical flow occurs, diminishing part of this discrepancy. However, the discrepancy suggests that a larger area should be monitored to find representative subregional-scale aquitard conductivities.

Model assumptions

Lithology and geometry of the aquitard

Lithology realizations with a fixed ratio between clay and sand are generated, with a single fixed ratio. The ratio was determined from well logs over the depth interval of the aquitard layer. Assuming this ratio as fixed is the source of uncertainty, as the composition of the subsurface, similar to the locations of the wells, is not randomly selected. Figure 7 shows a discrepancy between the hydraulic conductivity interpreted as a pumping test and with uniform flow conditions. This discrepancy could partially be caused by an overrepresentation of clay near the wells, which is compensated at a larger distance from the wells by an overrepresentation of sand.

A similar reasoning holds for the thickness of the aquitard, which is assumed constant. Assuming the aquitard to be thicker, i.e. counting the entire transition zone to the overlying aquifer, would result in a larger assumed sand fraction. In contrast, assuming a smaller thickness would result in a larger clay fraction, which would change the resulting correlation lengths.

In addition, the hydraulic conductivity is assumed to be constant within the lithology classes in each cell, based on the conductivity distributions of the Wormer Member. In reality the conductivity will differ throughout lithologies, which means additional information about the correlation within lithoclasses needs to be known. However, the variation between the conductivity distributions between the classes clay and sand is much larger than the variation within these classes, entailing that the correlation within the lithoclasses is of less importance for the hydraulic conductivity at field scale.

Also, the hydraulic conductivity at the test location might differ from the distribution of all samples within the Wormer Member. However, to develop a regional groundwater model, some assumptions have to be made; therefore, although the initial assumptions affect the correlation lengths, the proposed method results in correlation lengths that fit the assumed sand and clay fractions with the corresponding aquitard thickness and conductivity distributions.

Discretization

The cell size in the heterogeneous realizations is 5 m × 5 m × 0.1 m. Variation of hydraulic conductivity below that size are not resolved. Correlation lengths, particularly the best fitting ones $L_{\mathrm{H}} = 20$–60 m and $L_{\mathrm{V}} = 0.875$–2.0 m (Fig. 6), are well resolved over several cells to properly represent the variation in lithoclasses. A too small cell size would not only have resulted in much larger run times but would also have invalidated the assumption of a single representative hydraulic conductivity within a cell of a given lithoclass, which is based on the assumption that the correlation length of small-scale variation of conductivity within a lithoclass is smaller than the cell size.

Steady-state assumption

Steady state conditions are assumed, to forego the need for calibrating the specific storage. Reducing the number of parameters reduces the risk of nonuniqueness of the calibration result. The assumption is in line with the pumping test setup. The combination of extraction and re-infiltration results in a steady state after a shorter time period than a conventional pumping test with only extraction. Also, the first time period is long enough to reach steady state. The drawdown in the vicinity of the single extraction well does not change significantly between time periods; thus, the drawdown does not have to completely develop through the aquitard for every configuration.

Pumping test setup

The setting of the pumping test studied here, with multiple infiltration wells and several configurations, increases the model sensitivity to spatial variation in the aquitard beyond the pumping well, whereas it reduces the strong dependence on the composition of the aquitard close to the pumping well as described by Copty et al. 2008. A conventional pumping test, with a single extraction well, favours heterogeneous realizations where the harmonic mean of the hydraulic conductivity values directly at the well equal the conductance value, instead of inferring information about the larger-scale spatial variability. The use of multiple wells results in multiple sensitive regions, which narrows the range of possible heterogeneous realizations that fit the observed drawdowns.

The horizontal correlation lengths of the best fitting realizations are shorter than the distance between the wells, which leaves some uncertainty to the estimate of the correlation length. This uncertainty could be further reduced by increasing the number of infiltration and extraction wells at shorter distance as well as additional (independent) configurations with variations in infiltration and extraction wells—in a similar fashion to hydraulic tomography (Yeh and Liu 2000; Zhao and Illman 2018; Berg and Illman 2013; Poduri and Kambhammettu 2021). Also, the extent of the observation network should exceed the radius in which vertical flow caused by the wells occurs, as determined by the leakage factor, if the intent is to obtain a representative value of aquitard conductivity that represents subregional vertical flow.

The method can be improved for future application by taking additional measurements to assess heterogeneous realizations not only on drawdown observations, but also by including flow velocity and travel time data. This can be achieved by using, e.g., fiber optics (des Tombe 2021) and electromagnetic methods (e.g. Bonnett et al. 2019).

Implications for upscaling and regional parameterization

Correlation lengths of lithologies or hydraulic conductivities are useful information for the stochastic parameterization of heterogeneous aquitards. They also serve to assess the expected range of hydraulic conductivity values at locations where no field-scale data is available to include uncertainty.

The results of this study show that the representative hydraulic conductivity value obtained from a pumping test for a heterogeneous aquitard is not necessarily representative of other flow patterns. Particularly, it does not equal the representative value for uniform flow which is typically needed in regional groundwater flow models. The representative hydraulic conductivity values for uniform flow spread over several orders of magnitude for a limited number of realizations. Thus, its uncertainty is high and typically underestimated by conventional parameterization methods using a single correlation length.

The optimum correlation lengths identified may be transferred to unmonitored locations. A prerequisite is a similar distribution of lithoclasses and the same geological depositional environment. Uncertainty about correlation lengths should be taken into account.

Conclusions

The hydraulic conductivity of aquitards plays an important role in groundwater flow and transport models. The hydraulic parameterization of aquitards is typically done either by interpreting pumping tests, resulting in a single resistivity value, or by geostatistical upscaling, for which information about the scales of spatial variability is needed.

A method is presented to derive correlation lengths of lithoclasses with varying hydraulic conductivities in aquitards from a pumping and injection test. That way, the lithological variation and associated hydraulic conductivity of the aquitard can be parameterized, including spatial variability and uncertainty. Results show that it is not a single value for horizontal and vertical correlation lengths that fits best, but rather a range of equally eligible values. Thus, some uncertainty about the correlation lengths remains.

This information about spatial variability scale and uncertainty is transferable to aquitards of a similar depositional environment and can be used to improve the parameterization in regional groundwater flow models elsewhere.

This study also shows that the representative hydraulic conductivity for an aquitard obtained through a pumping test is not directly representative for other flow patterns. It should therefore be used with care in regional groundwater flow models assuming uniform flow. Thus, it is better to upscale the simulated hydraulic conductivity realizations to representative model block conductivities under the flow conditions expected to be simulated with the groundwater model.

Acknowledgements The authors want to thank the province of Zeeland for their financial support with regards to the pumping test. The constructive comments of the reviewers are gratefully acknowledged.

Declarations

Conflict of interest No conflict of interest is reported by the authors.

References

Abellan A, Noetinger B (2010) Optimizing subsurface field data acquisition using information theory. Math Geosci 42:603–630. https://doi.org/10.1007/s11004-010-9285-6

Alexander M, Berg SJ, Illman WA (2011) Field study of hydrogeologic characterization methods in a heterogeneous aquifer. Groundwater 49(3):365–382. https://doi.org/10.1111/J.1745-6584.2010.00729.X

Bakker M, Post V, Langevin CD, Hughes JD, White JT, Starn JJ, Fienen MN (2016) Scripting MODFLOW model development using python and FloPy. Groundwater 54(5):733–739. https://doi.org/10.1111/gwat.12413

Batlle-Aguilar J, Cook PG, Harrington GA (2016) Comparison of hydraulic and chemical methods for determining hydraulic conductivity and leakage rates in argillaceous aquitards. J Hydrol. https://doi.org/10.1016/j.jhydrol.2015.11.035

Berg SJ, Illman WA (2013) Field study of subsurface heterogeneity with steady-state hydraulic tomography. Ground Water 51(1):29–40. https://doi.org/10.1111/j.1745-6584.2012.00914.x

Berg SJ, Illman WA (2015) Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site. Groundwater 53(1):71–89. https://doi.org/10.1111/gwat.12159

Bierkens MFP (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

Bierkens MFP, van der Gaast JW (1998) Upscaling hydraulic conductivity: theory and examples from geohydrological studies. Nutr Cycling Agroecosyst 50(1):193–207. https://doi.org/10.1023/A:1009740328153

Bierkens MFP, Weerts HJ (1994) Block hydraulic conductivity of crossbedded fluvial sediments. Water Resour Res 30(10):2665–2678. https://doi.org/10.1029/94WR01049

Bonnett B, Mitchell B, Frampton M, Hayes M (2019) Low-noise instrumentation for electromagnetic groundwater flow measurement. I2MTC 2019–2019 IEEE International Instrumentation and Measurement Technology Conference, Proceedings, Auckland, New Zealand, May 2019. https://doi.org/10.1109/I2MTC.2019.8827136

Clark JI (1998) The settlement and bearing capacity of very large foundations on strong soils: 1996 R.M. Hardy keynote address. Can Geotech J 35(1):131–145. https://doi.org/10.1139/t97-070

Colecchio I, Boschan A, Otero AD, Noetinger B (2020) On the multiscale characterization of effective hydraulic conductivity in random heterogeneous media: a historical survey and some new perspectives. Adv Water Resour 140:103594. https://doi.org/10.1016/j.advwatres.2020.103594

Colecchio I, Otero AD, 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 NK, Trinchero P, Sanchez-Vila X, Sarioglu MS, Findikakis AN (2008) Influence of heterogeneity on the interpretation of pumping test data in leaky aquifers. Water Resour Res 44(11):11419. https://doi.org/10.1029/2008WR007120

Dagan G (2001) Effective, equivalent, and apparent properties of heterogeneous media. In: Mechanics for a new millennium. Springer, Dordrecht, The Netherlands, pp 473–486

De Marsily G, Delay F, Goncalves 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 MT, Copty NK, 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

des Tombe B (2021) Measuring horizontal groundwater flow with distributed temperature sensing along cables installed with direct-push equipment. https://doi.org/10.4233/UUID:92565BDB-CF5A-4110-ABF3-1B4298720466

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

Fleckenstein JH, Fogg GE (2008) Efficient upscaling of hydraulic conductivity in heterogeneous alluvial aquifers. Hydrogeol J 16(7):1239–1250. https://doi.org/10.1007/s10040-008-0312-3

Fogg GE, 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

Gautier Y, Nœtinger B (2004) Geostatistical parameters estimation using well test data. Oil Gas Sci Technol 59:167–183. https://doi.org/10.2516/ogst:2004013

Gerber R, Boyce J, Howard K (2001) Evaluation of heterogeneity and field-scale groundwater flow regime in a leaky till aquitard. Hydrogeol J 9(1):60–78. https://doi.org/10.1007/s100400000115

Gurwin J, Lubczynski M (2005) Modeling of complex multi-aquifer systems for groundwater resources evaluation: Swidnica study case (Poland). Hydrogeol J 13:627–639. https://doi.org/10.1007/S10040004-0382-9/FIGURES/9

Hantush MS, Jacob CE (1955) Non-steady Green's functions for an infinite strip of leaky aquifer. EOS Trans Am Geophys Union 36(1):101–112. https://doi.org/10.1029/TR036i001p00101

Hart DJ, Bradbury KR, Feinstein DT (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

Hendry MJ, Solomon DK, Person M, Wassenaar LI, Gardner WP, Clark ID, Mayer KU, Kunimaru T, Nakata K, Hasegawa T (2015) Can argillaceous formations isolate nuclear waste? Insights from

isotopic, noble gas, and geochemical profiles. https://doi.org/10.1111/gfl.12132
Hoeksema RJ, Kitanidis PK (1984) An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. Water Resour Res 20:1003–1020. https://doi.org/10.1029/WR020i007p01003
Huang SY, Wen JC, Yeh TCJ, Lu W, Juan HL, Tseng CM, Lee JH, Chang KC (2011) Robustness of joint interpretation of sequential pumping tests: numerical and field experiments. Water Resour Res 47(10):10530. https://doi.org/10.1029/2011WR010698
Journel AG, Alabert FG (1990) New method for reservoir mapping. J Petrol Technol 42(02):212–218. https://doi.org/10.2118/18324-PA
Keller CK, Van Der Kamp G, Cherry JA (1989) A multiscale study of the permeability of a thick clayey till. Water Resour Res 25(11):2299–2317. https://doi.org/10.1029/WR025I011P02299
Khan KD, Deutsch CV (2016) Practical incorporation of multivariate parameter uncertainty in geostatistical resource modeling. Nat Resour Res 25(1):51–70. https://doi.org/10.1007/S11053015-9267-Y/FIGURES/17
Kuhlman KL, Hinnell AC, Mishra PK, Yeh TCJ (2008) Basinscale transmissivity and storativity estimation using hydraulic tomography. Ground Water 46(5):706–715. https://doi.org/10.1111/j.1745-6584.2008.00455.x
Langevin CD, Hughes JD, Banta ER, Niswonger RG, Panday S, Provost AM (2017) Documentation for the MODFLOW 6 groundwater flow model. US Geol Surv Techniques Methods. Book 6, Modeling Techniques, 197 pp. https://doi.org/10.3133/TM6A55
Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. Q Appl Math 2(2):164–168. https://doi.org/10.1090/qam/10666
Marquardt DW (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
Matheron G (1967) Elements pour une theorie del milieux poreux [Elements for a theory of porous media]. Masson et Cie, Paris
Meriano M, Eyles N (2009) Quantitative assessment of the hydraulic role of subglaciofluvial interbeds in promoting deposition of deformation till (Northern Till, Ontario). Quatern Sci Rev 28(7–8):608–620. https://doi.org/10.1016/j.quascirev.2008.08.034
Neuman SP, Witherspoon PA (1972) Field determination of the hydraulic properties of leaky multiple aquifer systems. Water Resour Res 8(5):1284–1298. https://doi.org/10.1029/WR008I005P01284
Neuman SP, Guadagnini A, Riva M (2004) Type-curve estimation of statistical heterogeneity. Water Resour Res 40. https://doi.org/10.1029/2003WR002405
Neuzil CE (1994) How permeable are clays and shales? Water Resour Res 30(2):145–150. https://doi.org/10.1029/93WR02930
Neuzil CE (1986) Groundwater flow in low-permeability environments. Water Resour Res 22(8):1163–1195. https://doi.org/10.1029/WR022I008P01163
Pebesma EJ (2004) Multivariable geostatistics in S: the GSTAT package. Comput Geosci 30(7):683–691. https://doi.org/10.1016/j.cageo.2004.03.012
Pickup GE, Ringrose PS, Jensen JL, Sorbie KS (1994) Permeability tensors for sedimentary structures. Math Geol 26(2):227–250. https://doi.org/10.1007/BF02082765
Poduri S, Kambhammettu BV (2021) On the performance of pilot-point based hydraulic tomography with a geophysical a priori model. Groundwater 59(2):214–225. https://doi.org/10.1111/gwat.13053
Ponzini G, Crosta G, Giudici M (1989) The hydrogeological role of an aquitard in preventing drinkable water well contamination: a case study. Environ Health Perspect 83:77–95. https://doi.org/10.1289/EHP.898377
Sanchez-Vila X, Guadagnini A, Carrera J (2006) Representative hydraulic conductivities in saturated groundwater flow. Rev Geophys 44(3):3002. https://doi.org/10.1029/2005RG000169
Sarris TS, Paleologos EK (2004) Numerical investigation of the anisotropic hydraulic conductivity behavior in heterogeneous porous media. Stoch Env Res Risk Assess 18(3):188–197. https://doi.org/10.1007/s00477-003-0171-3
Sommer W, Valstar J, Leusbrock I, Grotenhuis T, Rijnaarts H (2015) Optimization and spatial pattern of large-scale aquifer thermal energy storage. Appl Energy 137:322–337. https://doi.org/10.1016/j.apenergy.2014.10.019
TNO-GDN (2022) Naaldwijk Formation. Stratigraphic Nomenclature of the Netherlands. http://www.dinoloket.nl/en/stratigraphic-nomenclature/naaldwijk-formation. Accessed November 2022
Van HH, Larsen F, Quy NP, Vu LT, Thanh GNT (2022) Recharge mechanism and salinization processes in coastal aquifers in Nam Dinh province, Vietnam. J Earth Sci 44:213–238. https://doi.org/10.15625/2615-9783/16864
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
Weerts HJ, Bierkens MFP (1993) Geostatistical analysis of overbank deposits of anastomosing and meandering fluvial systems: Rhine-Meuse delta, The Netherlands. Sediment Geol 85(1–4):221–232. https://doi.org/10.1016/0037-0738(93)90085-J
Wen JC, Wu CM, Yeh TCJ, Tseng CM (2010) Estimation of effective aquifer hydraulic properties from an aquifer test with multiwell observations (Taiwan). Hydrogeol J 18(5):1143–1155. https://doi.org/10.1007/S10040-010-0577-1/FIGURES/11
Wu CM, Yeh TCJ, Zhu J, Hau Lee T, Hsu NS, Chen CH, Sancho AF (2005) Traditional analysis of aquifer tests: Comparing apples to oranges? Water Resour Res 41(9):1–12. https://doi.org/10.1029/2004WR003717
Xiao S, Xu T, Reuschen S, Nowak W, Hendricks Fransen HJ (2021) Bayesian inversion of multi-gaussian log-conductivity fields with uncertain hyperparameters: an extension of preconditioned Crank-Nicolson Markov chain Monte Carlo with parallel tempering. Water Resour Res 57:313. https://doi.org/10.1029/2021WR030313
Yeh TC, Liu S (2000) Hydraulic tomography: development of a new aquifer test method. Water Resour Res 36(8):2095–2105. https://doi.org/10.1029/2000WR900114
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
Zhao Z, 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
Zhuang C, Zhou ZF, Li ZF, Guo QN (2017) A method for determining hydraulic parameters of an overconsolidated aquitard. Rock Soil Mech. https://doi.org/10.16285/j.rsm.2017.01.008