Groundwater

Methods Note/

How to Find Aquifer Statistics Utilizing Pumping Tests? Two Field Studies Using `welltestpy`

by Sebastian Müller[1,2,3], Carsten Leven[4], Peter Dietrich[2,4], Sabine Attinger[2,3], and Alraune Zech[2,5]

Abstract

We present a workflow to estimate geostatistical aquifer parameters from pumping test data using the Python package `welltestpy`. The procedure of pumping test analysis is exemplified for two data sets from the Horkheimer Insel site and from the Lauswiesen site, Germany. The analysis is based on a semi-analytical drawdown solution from the upscaling approach Radial Coarse Graining, which enables to infer log-transmissivity variance and horizontal correlation length, beside mean transmissivity, and storativity, from pumping test data. We estimate these parameters of aquifer heterogeneity from type-curve analysis and determine their sensitivity. This procedure, implemented in `welltestpy`, is a template for analyzing any pumping test. It goes beyond the possibilities of standard methods, for example, based on Theis' equation, which are limited to mean transmissivity and storativity. A sensitivity study showed the impact of observation well positions on the parameter estimation quality. The insights of this study help to optimize future test setups for geostatistical aquifer analysis and provides guidance for investigating pumping tests with regard to aquifer statistics using the open-source software package `welltestpy`.

Introduction

The knowledge of aquifer heterogeneity is essential for predicting subsurface flow and transport but is hardly considered in practice (Fiori et al. 2016; Rajaram 2016), resulting in erroneous and uncertain predictions of contaminant distributions. The variation of hydraulic conductivity over many orders of magnitude is a well-known phenomenon (Bear 1972), but its detection and modeling still remains a challenge. Spatial resolution is usually inhibited by data scarcity due to limited accessibility and high investigation costs.

Geostatistical descriptions of aquifer properties provide information on spatial variability in hydraulic quantities. A well-established approach is transmissivity, being the vertically integrated hydraulic conductivity, which has been approximated as a log-normally distributed random variable with a spatial correlation (Dagan 1989; Gelhar 1993). Following this approach, heterogeneity can be described with a few characterizing parameters: mean μ, log-variance σ^2, and horizontal correlation length ℓ. The first two are essential to determine average flow velocity (Sánchez-Vila et al. 2006), which is the basis for reliable contaminant transport predictions for, for example, risk assessment and groundwater remediation. The strength of heterogeneity determines the spreading of solutes, thus providing σ^2 and ℓ can be used to approximate transport parameters such as dispersion coefficients (Dagan 1989).

[1]Corresponding author: Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany; sebastian.mueller@ufz.de
[2]Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany
[3]Institute of Environmental Science and Geography, University of Potsdam, 14476 Potsdam, Germany
[4]Department of Geoscience, University of Tübingen, 72074 Tübingen, Germany
[5]Department of Earth Science, Utrecht University, 3584 CS, Utrecht, The Netherlands

Article impact statement: We present a workflow to infer parameters of subsurface heterogeneity from pumping test data exemplified at two sites using `welltestpy`.

[The Legal statement line for this article was changed on 19 July 2021 after original online publication.]

Received August 2020, accepted June 2021.

doi: 10.1111/gwat.13121

A shortcoming of stochastic methods is the need for spatially distributed transmissivity data for parameter estimation. Classical geostatistical one-point and two-point analyses rely on multiple point measurements, for example, from flowmeter, permeameter or direct push, causing high cost due to the large amount of necessary data. A promising opportunity to overcome this bottleneck is using pumping test data to achieve a first guess on heterogeneity parameters. In this line, we present the missing link in form of tools to infer aquifer statistics such as log-variance σ^2 and horizontal correlation length ℓ by interpreting pumping tests in heterogeneous media.

Classical interpretation methods for pumping tests assume homogeneity, such as Thiem's or Theis' solutions. These methods provide estimates of average transmissivity when the setting allows flow to sample sufficient volume of the aquifer. However, they cannot cover the impact of heterogeneity close to the pumping well (Indelman and Abramovich 1994; Leven and Dietrich 2006). There is no constant transmissivity being representative for a heterogeneous media in the nonuniform flow setting of a pumping test. Distinct representative values emerge as function of the radial distance to the pumping well: the geometric mean $T_G = e^{\mu}$ for the far field and the harmonic mean $T_H = T_G e^{-\sigma^2/2}$ at the pumping well when considering a constant head boundary conditions there. Other boundary conditions lead to different representative T values (Sánchez-Vila et al. 2006).

Zech et al. (2012, 2016) were the first to provide hydraulic head solutions, which reproduce the pumping test behavior in aquifers of log-normal random heterogeneity. They made use of the upscaling method *Radial Coarse Graining* to derive an effective transmissivity $T_{\text{eff}}(r)$ as function of the aquifer statistical parameters μ, σ^2, and ℓ and depending on the radial distance r following the lines of Neuman et al. (2004). Zech et al. (2012) and Zech et al. (2016) further derived semi-analytical solutions for the pumping test drawdowns $h_{\text{eff}}(r, t)$ by solving the steady state and transient groundwater flow equation under the use of the effective transmissivity $T_{\text{eff}}(r)$. Both solutions $h_{\text{eff}}(r, t)$ quantify the effect of heterogeneity on the drawdown through the aquifer statistics μ, σ^2, and ℓ. They can be interpreted as extensions of Thiem's and Theis' solutions to heterogeneous aquifers. However, tools are missing to apply the semi-analytical solutions to field data.

Closing the gap to applications, we provide implementations of these drawdown solutions $h_{\text{eff}}(r, t)$ for pumping tests. They are the first freely available forward models reproducing the effect of intermediate scale heterogeneity on drawdowns, particularly close to the pumping well where well flow is not ergodic. The key application of the code is the interpretation of pumping test data to infer aquifer statistics μ, σ^2 and ℓ representing aquifer heterogeneity at the scale of a few meter up to tens of meters related to the resolution of typical field pumping tests in fully penetrating wells. Although flow toward wells is inherently three-dimensional (3D) in heterogeneous media, drawdown signals are depth averaged thus providing depth-averaged aquifer characteristics (Sánchez-Vila et al. 2006; Zech et al. 2012).

We present a standardized workflow to estimate geostatistical parameters from pumping test data and demonstrate it for two field studies. We introduce the freely available open-source code `welltestpy` (Müller 2021) as a tool to manage and analyze pumping test data for quantifying heterogeneity aquifer. For each parameter, we specify its sensitivity toward test setup. Thus, reliability of parameter estimates can be inferred. Results provide valuable information on the design of future field campaigns in order to optimize parameter estimates.

Methods and Data

`welltestpy`

The Python package `welltestpy` (Müller 2021) provides a workflow to infer parameters of subsurface heterogeneity from pumping test data. It contains routines to handle, visualize, process and interpret field exploration campaigns based on well observations. Figure 1 shows the package organization. The structure and functionalities of sub-routines is outlines in the Supporting Information where also technical details on code administration are discussed. `welltestpy` is compatible with Python 3.5 and higher, and supports Linux, MacOS and Windows. It can be installed with `pip` or `conda`:

```
pip install welltestpy
conda install -c conda-
      forge welltestpy
```

`welltestpy` is part of the *GeoStat-Framework* (https://geostat-framework.org), which contains Python packages for geostatistical applications and subsurface flow and transport simulations. Currently, `welltestpy` is focused on pumping test campaigns but will be extended to other well based exploration techniques such as slug tests or dipole tests.

Any test campaign, once stored in the `welltestpy` format ".cpm," can be handled with a few lines of Python-code, including the run of estimation procedures.

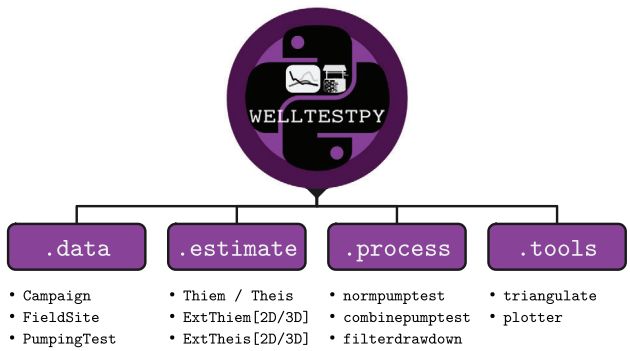

Figure 1. Organigram of `welltestpy` with its four sub-packages and their main functionalities.

For instance, the estimation results of this paper can be reproduced via:

```
import welltestpy as wtp
camp = wtp.load_campaign
      ("horkheim.cmp")
estimation = wtp.estimate.
      ExtTheis2D(
name="estimate", campaign=
      camp, generate=True)
estimation.run()
```

Details and tutorials on the use of welltestpy can be found in the online documentation under https://welltestpy.readthedocs.io.

Field Sites

We interpret pumping test data from two hydrogeological field sites in southern Germany: The Horkheimer Insel, close to the River Neckar in Heilbronn, is heterogeneous while the Lauswiesen site near Tübingen is more homogeneous. Key feature of both test campaigns for application here is the short distance of spatially distributed observations to the pumping well. Aquifer heterogeneity shows highest impact close to the pumping well. Thus, detection of heterogeneity characteristics, such as correlation length, requires drawdown observations close to the pumping well. In this regard, statistical estimates represent aquifer heterogeneity at the scale of the pumping test, being in the intermediate range in the order of a few tens of meters. Although both aquifers are unconfined, drawdowns show no signs of associated effects due to the short duration of pumping. Proof is given in diagnostic plots in the Supporting Information. Furthermore, both sites offer complementary estimates of aquifer statistics based on other characterization methods, such as flowmeter. We use these values for qualitative comparison given their dependence on method characteristics such as the particular support volume and spatial resolution.

Horkheimer Insel

The fluvial unconfined aquifer consists of unconsolidated sediments with poorly sorted gravel and sand and has an average saturated thickness of $L = 3\,\mathrm{m}$ (Schad and Teutsch 1994). The prevailing hydraulic gradient is 0.1% toward the River Neckar. We make use of drawdown data of 4×2 transient pumping tests (Schad 1997). Each of the four pumping wells is used twice with different observation wells per test. Drawdown values are normalized to their pumping rate to combine pumping tests performed at the same well. Figure 2(left) shows the spatial distribution of wells and associated observation wells per test. The distances between wells range between 6.5 and 27 m.

Schad (1997) presented an extended hydrogeological analysis, including hydraulic conductivity estimates from flowmeter, permeameter and grain size analysis. An average transmissivity of $T = 3.1e-2\,\mathrm{m}^2/\mathrm{s}$ is reported for the pumping tests analyzed with Theis' method. One and two-point statistical results depend on the type of measurement method: log-conductivity variance ranges from $\sigma^2 = 1.57$ for flowmeter, $\sigma^2 = 2.32$ for grain size analysis to $\sigma^2 = 3.17$ for permeameter. Detailed variogram analyses suggested vertical and horizontal correlation lengths of $\ell_v = 0.15 - 0.25\,\mathrm{m}$ and $\ell_h = 6 - 10\,\mathrm{m}$, respectively.

Lauswiesen

The aquifer consists of unconsolidated, alluvial sediments with poorly sorted gravel and sand with a small amount of fines (<10%). The saturated thickness of the unconfined aquifer is $5-6\,\mathrm{m}$, while the average hydraulic gradient is 0.2% to 0.3% parallel to the Neckar valley. The data set stems from a series of five transient pumping tests (Leven 2020). The five fully screened wells are 5 to 25 m apart (Figure 2, right). Each well was pumped at a constant rate with automated drawdown recording at the pumping and all four observation wells at a sampling interval of 1 s.

The synopsis of all aquifer tests conducted at the site over the last 24 years led to transmissivities in the range of 0.002 to $0.22\,\mathrm{m}^2/\mathrm{s}$ with a mean of $0.017\,\mathrm{m}^2/\mathrm{s}$, resulting in an average hydraulic conductivity around $3e-3\,\mathrm{m/s}$. Depth resolved aquifer tests such as direct-push injection logs (DPIL) (Lessoff et al. 2010) revealed a subdivision into an upper more conductive zone of ≈ 2.5 m and a lower zone of ≈ 3 m. Lessoff et al. (2010) reported log-conductivity variances of $\sigma_u^2 = 0.1$, and $\sigma_l^2 = 0.93$ for the upper and lower unit, based on DPIL and multi-level slug tests. Estimated vertical and horizontal correlation lengths for the upper and lower units are $\ell_v = 0.23$ m, $\ell_h = 4.6$ m and $\ell_v = 0.32$ m, $\ell_h = 1.3$ m, respectively. Considering the aquifer as single 3D unit, Riva et al. 2006 reported a log-conductivity variance of $\sigma_Y^2 = 2.9$ and correlation lengths of $\ell_v = 0.9$ m and $\ell_h = 12$ m based on grain-size analysis.

Extended Theis' Solution

We analyze pumping tests with regard to their information content on heterogeneity making use of the extended Theis' solution of Zech et al. (2016). We will briefly outline the concept, for a detailed method description we refer to Zech and Attinger (2015) or Zech et al. 2016. Starting point is a spatially distributed transmissivity $T(x, y)$ being described by a log-normal distribution with characterizing parameters μ and σ^2 being the mean and variance of the underlying normal distribution of $\ln(T)$. We consider a Gaussian spatial correlation function $\rho(r) = \exp\left(-\frac{1}{2}(r/\ell)^2\right)$ with the correlation length ℓ.

The extended Theis solution $h_{\mathrm{eff}}(r, t)$ (Figure 3) is the result of the Radial Coarse Graining upscaling procedure. It follows as solution of the groundwater flow equation under pumping test initial and boundary conditions using an radial depending transmissivity $T_{\mathrm{eff}}(r) = T_{\mathrm{G}} \exp\left(-\frac{1}{2}\frac{\sigma^2}{1+1.6r^2/\ell^2}\right)$ which effectively represents heterogeneity. The semi-analytical expression $h_{\mathrm{eff}}(r, t)$ is a function of the radius r and of the statistical properties of transmissivity; behaving similar to Theis's solution

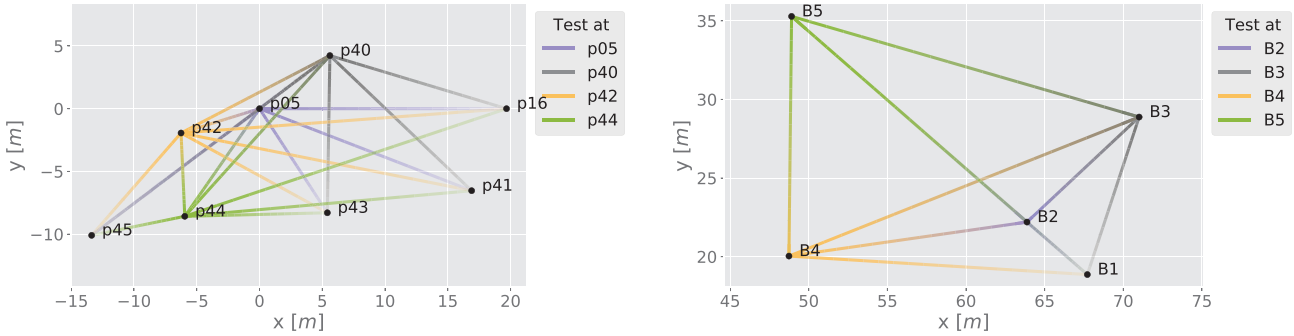

Figure 2. Overview of the well setup at Horkheimer Insel (left) and Lauswiesen (right) site. Pumping wells are indicated in legend, fading colored lines show connections to observations wells.

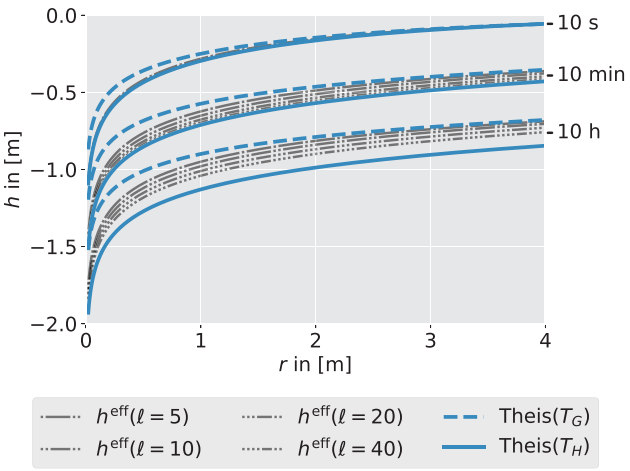

Figure 3. Transient pumping tests: extended Theis' solution $h_{\text{eff}}(r)$ for various correlation length ℓ (differently dashed lines) and Theis' solution for geometric (T_{G}) and harmonic (T_{H}) mean transmissivity at three times: 10 s, 10 min, 10 h.

far from the well for a uniform transmissivity being the geometric mean T_{G} and close to the well for Theis's solution with the harmonic mean $T_{\text{H}} = T_{\text{G}} \exp\left(-\frac{1}{2}\sigma^2\right)$. In between, $h_{\text{eff}}(r, t)$ interpolates as function of correlation length ℓ (Figure 3) similar to the effective transmissivity $T_{\text{eff}}(r)$. A comparison to numerical pumping test ensembles (Zech et al. 2012 for steady state and Zech et al. 2016 for transient well flow) showed that $h_{\text{eff}}(r, t)$ captures the entire range of the drawdown based on the transmissivity statistics. Implementations of the semi-analytical $h_{\text{eff}}(r, t)$ are provided in `AnaFlow` (Müller 2020) and subsequently used in `welltestpy`.

Inverse Estimation and Parameter Sensitivity

Inverse estimation of heterogeneity parameters is based on type curve fitting. We minimize the root mean square error (RMSE) between the pumping test observations and the extended Theis' solution. For optimization, we use the shuffled complex evolution (SCE) algorithm (Duan et al. 1992), implemented in `spotpy` (Houska et al. 2015). The optimal guess delivers the parameter set of T_G, σ^2, ℓ, and S simultaneously. Table 1 shows the parameter ranges used during estimation. Mean

Table 1
Parameter Ranges During Estimation

Parameter	Range	Initial Guess
$\ln S$	$(-13, -1)$	-7
$\ln T_{\text{G}}$	$(-16, -2)$	-9
σ^2	$(0, 10)$	5
ℓ	$(1, 50)$	25.5

Storativity (S) and transmissivity (T) are fitted in log-scale. Initial guesses relate to mean values of ranges.

transmissivity T_{G} and storativity S are fitted in log scale since they vary over orders of magnitude (Bear 1972; Gelhar 1993). The extended Theis' solution enables testing a broad range of log-transmissivity variance values. The maximal length scales ℓ exceeds the diameters of the well setups at each site by a factor of two.

The workflow comprises standardization of pumping test results by rescaling all drawdowns to a normalized pumping-rate of $Q = 1\frac{m^3}{s}$ and resampling of drawdown time series with quadratically increasing step size to account for the logarithmic time evolution. Details are provided in the Supporting Information.

We apply the Fourier amplitude sensitivity test (FAST) (Saltelli et al. 1999), implemented in `spotpy` (Houska et al. 2015), to infer the sensitivity of heterogeneity parameters during estimation. FAST is a variance based global sensitivity test, analyzing the Fourier series expansion of the target function. It computes sensitivities between 0 and 1 (low to high). Parameter interactions are taken into account by relating the ratios between values.

FAST is also used to analyze the sensitivity of each parameter toward the radial distance to the pumping well. Target metric here is the deviation of the extended Theis type curve from the Theis solution. This procedure is independent of the pumping test observations.

Results and Discussion

Estimated Parameters

Results of the type-curve fitting procedure for both sites are summarized in Figure 4. We estimated the

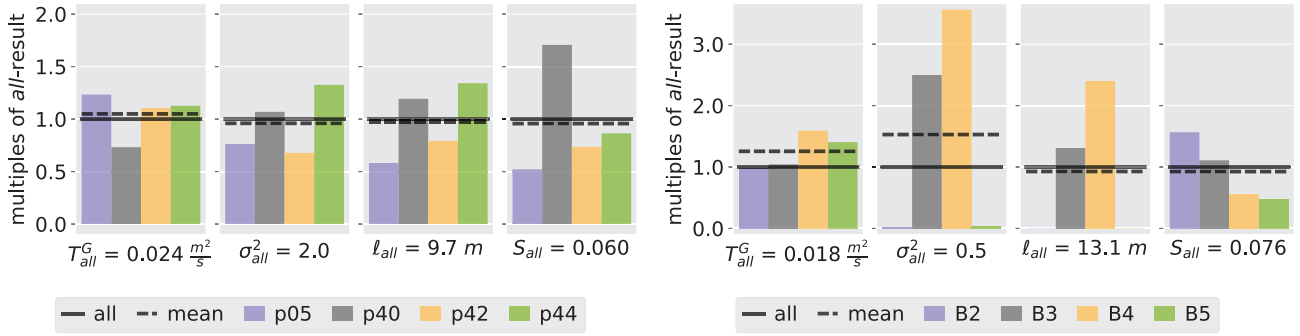

Figure 4. Parameter estimates for geometric mean T_G, log-transmissivity variance σ^2, correlation length ℓ and storativity S for the Horkheimer Insel site (left) and the Lauswiesen site (right): Colored bars represent estimates for individual pumping tests at specified wells. The dashed lines are means over individual estimates. Black line denotes optimal estimates for a combined fitting of *all* tests, to which values are scaled in the *y*-axis.

geostatistical parameters (T_G, σ^2, ℓ) and storativity S for each individual pumping test as well as in one fitting run combining all tests at a site in standardized form. The extended Theis' solution with the optimal parameter set, that is, from the combination of tests, is displayed along the drawdown observations in Figure 5. Time series observations are well covered by the extended Theis' solution. A comparison to the results of the classical Theis solution can be found in the Supporting Information.

Estimates of statistics (Figure 4) vary between tests showing the natural local heterogeneity. Exceptions are estimates of σ^2 and ℓ for the tests at $B2$ and $B5$ of the Lauswiesen site, which correspond to an apparent homogeneous drawdown. The lack of contrast of drawdown behavior at the well and in the far field inhibits the estimation of σ^2 and ℓ. The mean over all individual test estimates (for each of the four parameters) is quasi the same as the estimate for the combined tests, proving the robustness of the estimation procedure.

The optimal parameter estimates at both sites are reasonably in line with results of previous geostatistical investigations based on other field methods ("Field Sites" section). Given the larger support volume of depth averaged drawdown signals, variances are expected to be smaller than those resulting from aquifer characterization methods at smaller scale. $\sigma^2_Y = 0.5$ for Lauswiesen and $\sigma^2_Y = 2$ for Horkheim relate well to those estimates from multilevel slug and flowmeter tests. Depth averaging hardly impacts horizontal correlation length. Thus, the estimates of $\ell = 13.1\,\mathrm{m}$ and $\ell = 9.7\,m$ from Lauswiesen and Horkheim pumping test data are in the same range as those of previous investigations. The slight increase of values can be related again to the larger support volume of pumping test, being appropriate to detect sedimentary connections at larger scales and distances. At the same time, estimates refer to the comparably small scale of the test settings with distances of a few tens of meters between wells. Thus, the characterized heterogeneity parameters refer to intermediate scale aquifer heterogeneity and should not be extrapolated to greater scale.

Sensitivity

Figure 6 shows the general sensitivity of parameter estimates toward the distance to the pumping well. The log-transmissivity mean and the log-storativity show a high sensitivity value over the whole domain as expected given their impact also on homogeneous aquifer drawdowns. Storativity controls the time evolution of the drawdown whereas mean transmissivity controls the spatial behavior. The log-transmissivity variance has the highest sensitivity close to the pumping well since strong deviations of transmissivity have the highest effect on drawdown where the head gradients are largest. This behavior is represented by the effective transmissivity at the pumping wellbeing the harmonic mean $T_H = T_G \exp(\sigma^2)$.

The impact of the correlation length ℓ rises within the first meters and then stays constant at the same low level as the variance. ℓ characterizes the transition behavior of the effective drawdown from near well to far field (Figure 3). Thus, its estimation depends on the σ^2 (near well) and T_G (far field). This is also confirmed by the evolution of parameter estimates during optimization which is discussed in the Supporting Information. As a side effect, the length scale is getting more influential for higher values of variance. On the other side, if the drawdown is apparently homogeneous, that is, it can be represented by Theis' solution with T_G, makes the estimation of a length scale impossible given the implication of zero variance. However, this behavior does not imply that the site is homogeneous. Instead, it can be a random effect that the local harmonic mean of transmissivity at the pumping well is identical to the geometric mean of the field.

Figure 7 shows the individual sensitivities of each parameter estimate at the two field sites. Parameter sensitivities are similar among tests and at both sites. The high sensitivity of T_G is due to the general sensitivity of the parameter and related to the amount of observations in the far-field. Storativity and σ^2 show medium sensitivity. The correlation length ℓ shows a very low sensitivity for these particular well settings. ℓ relies on drawdown observations in the transition zone to the far field,

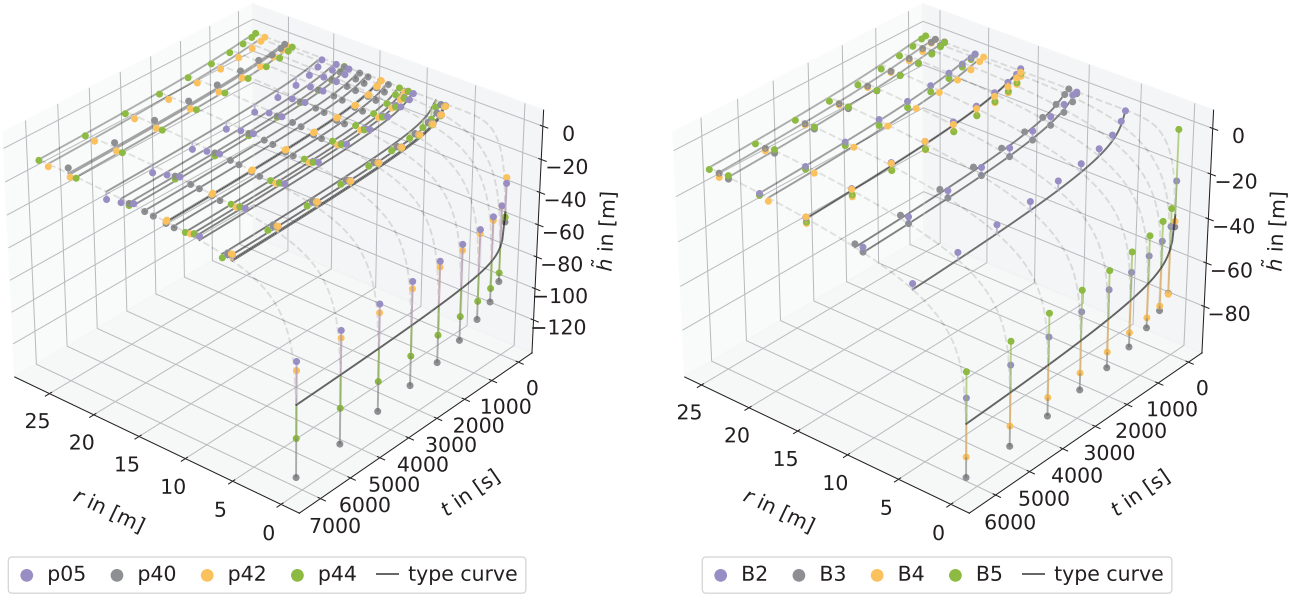

Figure 5. Standardized drawdown data and extended Theis solution with optimal estimates for the Horkheimer Insel site (left) and the Lauswiesen test site (right). Points indicate observations at each well (radial position) for all tests (distinction by colors). The lines represent type curves for the optimal estimates of *all* tests combinated.

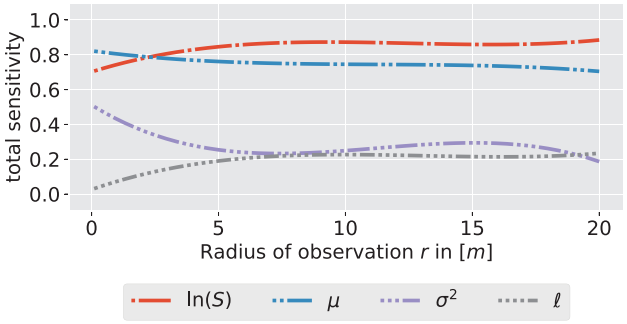

Figure 6. Sensitivity of all four parameters as function of radial distance to the pumping well based on the FAST method (with 1 being a high and 0 low sensitivity).

particularly the distance of a few meters from the pumping well. Figure 2 reveals that both campaigns lack of observation wells located within less than 5 m distance.

Summary and Conclusion

We present a workflow to infer parameters of subsurface heterogeneity from pumping test data exemplified at two sites using the open source Python package `welltestpy`. Geostatistical analysis of transmissivity is based on the extended Theis' solution of Zech et al. 2016. Beside mean transmissivity and storativity, `welltestpy` allows gaining robust estimates of log-transmissivity variance σ^2 and horizontal correlation length ℓ. These parameters characterizing spatial variability are essential to improve predictions of groundwater flow and contaminant transport for, for example, risk assessment and remediation under uncertainty induced by the inherent heterogeneity of aquifer structure and properties.

The application of the workflow shows that statistical information of transmissivity can be successfully estimated from standard pumping test measurements at heterogeneous field sites. Estimates of mean transmissivity and storativity are as reliable as with Theis' solution. Estimates of log-transmissivity variance σ^2 and correlation-length ℓ require specific positioning of observation wells.

The sensitivity analysis indicates optimal locations of observation well to infer aquifer statistical parameters, providing valuable information for pumping test designs. Storativity and its sensitivity depends on the confinement of the aquifer. We expect a smaller radius of impact and higher sensitivity of storativity for unconfined conditions while the cone of depression will spread farther and more rapidly in confined systems. When the field study's focus is on reliably inferring statistical parameters, key points are: (1) tests at multiple pumping wells are necessary to infer the log-transmissivity variance σ^2 of the overall field, (2) reliable values of correlation length ℓ require observations close to the pumping well (i.e., a few meter distance). Observations at and close to the pumping well provide most information on heterogeneity since pressure gradients are the steepest here and drawdowns are deviate from mean behavior.

With pumping test being a standard field investigation method, the `welltestpy`-workflow provides a quick, simple and cost-efficient option to infer first estimates of aquifer statistics, particularly at highly heterogeneous sites. The workflow can be easily transferred and applied to other field campaigns considering the following key aspects: (1) the project objectives is on detecting aquifer heterogeneity statistics leading to a test design which allows observing drawdowns at and close to the pumping well ("transition zone") where observations are not reproduced by "homogeneous" solutions (e.g., that of Theis');

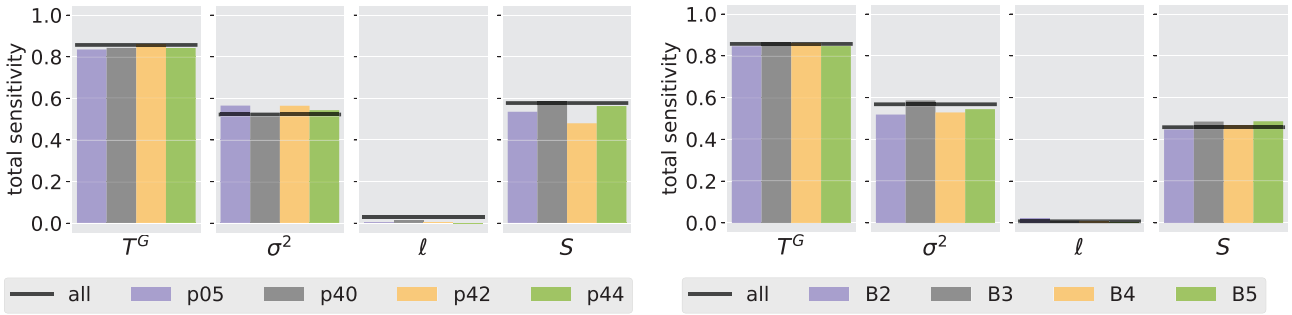

Figure 7. Sensitivities of the parameters during estimation at the Horkheimer Insel site (left) and the Lauswiesen site (right) based on the FAST method (with 1 being a high and 0 low sensitivity): colored bars represent results for individual pumping tests, the solid line is the sensitivity for the combined estimation.

(2) the interest is in intermediate scale heterogeneity characteristics, for example, correlation length and variance of facies structures at the scale of a few up to several tens of meters, being in line with the scale and support volume of the observation method; (3) the test setting is adapted to confinement conditions with smaller transition zones for unconfined aquifers.

Acknowledgment

Sebastian Müller was funded by the German Federal Environmental Foundation (DBU). Open Access funding enabled and organized by Projekt DEAL.

Code and Data Availability

`welltestpy`: As part of the Geostat Framework, the code of `welltestpy` is developed on https://github.com/GeoStat-Framework/welltestpy. The documentation includes a quickstart guide, in-depth tutorials, and a complete overview over the API. It can be accessed via https://welltestpy.readthedocs.io/.

Data: The underlying data and scripts to reproduce the results of this paper can be found at https://github.com/GeoStat-Examples/welltestpy-field-site-analysis and are archived in Müller and Zech 2021. The pumping test data from the Lauswiesen site are available at the research data portal of the University of Tübingen at http://www.hdl.handle.net/10900.1/bfb0b0f7-7065-4a24-8a91-92ad8aa8fc40.

Authors' Note

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Appendix S1 Supporting Information.

References

Bear, J. 1972. *Dynamics of Fluids in Porous Media*. New York: Elsevier.

Dagan, G. 1989. *Flow and Transport in Porous Formations*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-75015-1

Duan, Q., S. Sorooshian, and V. Gupta. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28, no. 4: 1015–1031. https://doi.org/10.1029/91WR02985

Fiori, A., V. Cvetkovic, G. Dagan, S. Attinger, A. Bellin, P. Dietrich, A. Zech, and G. Teutsch. 2016. Debates – Stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for? *Water Resources Research* 52, no. 12: 9228–9234. https://doi.org/10.1002/2015WR017525

Gelhar, L.W. 1993. *Stochastic Subsurface Hydrology*. New York: Prentice Hall, Englewood Cliffs.

Houska, T., P. Kraft, A. Chamorro-Chavez, and L. Breuer. 2015. SPOTting model parameters using a ready-made python package. *PLoS ONE* 10, no. 12: e0145180. https://doi.org/10.1371/journal.pone.0145180

Indelman, P., and B. Abramovich. 1994. Nonlocal properties of nonuniform averaged flows in heterogeneous media. *Water Resources Research* 30, no. 12: 3385–3393. https://doi.org/10.1029/94WR01782

Lessoff, S.C., U. Schneidewind, C. Leven, P. Blum, P. Dietrich, and G. Dagan. 2010. Spatial characterization of the hydraulic conductivity using direct-push injection logging. *Water Resources Research* 46, no. 12: W12502. https://doi.org/10.1029/2009WR008949

Leven, C. 2020. *Pumping Tests in Wells B1-B5 at the Hydrogeological Research Site Lauswiesen*. Tübingen: Eberhard Karls Universität Tübingen http://hdl.handle.net/10900.1/bfb0b0f7-7065-4a24-8a91-92ad8aa8fc40.

Leven, C., and P. Dietrich. 2006. What information can we get from pumping tests? Comparing pumping test configurations using sensitivity coefficients. *Journal of Hydrology* 319: 199–215. https://doi.org/10.1016/j.jhydrol.2005.06.030

Müller, S. 2021. GeoStat-Framework/`welltestpy`: v1.0.3. Zenodo. https://doi.org/10.5281/zenodo.4549297 (accessed February 19, 2021).

Müller, S. 2020. GeoStat-Framework/`AnaFlow` v1.0.1. Zenodo. https://doi.org/10.5281/zenodo.3738268 (accessed May 31, 2020).

Müller, S., and A. Zech. 2021. GeoStat-Examples/welltestpy-field-site-analysis: v1.1.0. Zenodo. https://doi.org/10.5281/zenodo.4724118 (accessed April 27, 2021).

Neuman, S.P., A. Guadagnini, and M. Riva. 2004. Type-curve estimation of statistical heterogeneity. *Water Resources Research* 40: W04201. https://doi.org/10.1029/2003WR002405

Rajaram, H. 2016. Debates—Stochastic subsurface hydrology from theory to practice: Introduction. *Water Resources Research* 52, no. 12: 9215–9217. https://doi.org/10.1002/2016WR020066

Riva, M., L. Guadagnini, A. Guadagnini, T. Ptak, and E. Martac. 2006. Probabilistic study of well capture zones distribution at the Lauswiesen field site. *Journal of Contaminant Hydrology* 88, no. 1: 92–118. https://doi.org/10.1016/j.jconhyd.2006.06.005

Saltelli, A., S. Tarantola, and K.P.-S. Chan. 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41, no. 1: 39–56. https://doi.org/10.1080/00401706.1999.10485594

Sánchez-Vila, X., A. Guadagnini, and J. Carrera. 2006. Representative hydraulic conductivities in saturated groundwater flow. *Reviews of Geophysics* 44: RG3002. https://doi.org/10.1029/2005RG000169

Schad, H. 1997. Variability of hydraulic parameters in non-uniform porous media: Experiments and stochastic modeling at different scales. PhD thesis, University Tübingen.

Schad, H., and G. Teutsch. 1994. Effects of the investigation scale on pumping test results in heterogeneous porous aquifers. *Journal of Hydrology* 159: 61–77. https://doi.org/10.1016/0022-1694(94)90249-6

Zech, A., S. Müller, J. Mai, F. Heße, and S. Attinger. 2016. Extending Theis' solution: Using transient pumping tests to estimate parameters of aquifer heterogeneity. *Water Resources Research* 52, no. 8: 6156–6170. https://doi.org/10.1002/2015WR018509

Zech, A., and S. Attinger. 2015. Technical note: Analytical solution for the mean drawdown of steady state pumping tests in two-dimensional isotropic heterogeneous aquifers. *Hydrology and Earth System Sciences* 12: 6921–6944. https://doi.org/10.5194/hessd-12-6921-2015

Zech, A., C.L. Schneider, and S. Attinger. 2012. The extended Thiem's solution—Including the impact of heterogeneity. *Water Resources Research* 48, no. 10: W10535. https://doi.org/10.1029/2012WR011852