Multiscale Hydraulic Conductivity Characterization in a Fractured Granitic Aquifer: The Evaluation of Scale Effect

Shuangpo Ren1,2,3, Ye Zhang3, Tian-Chyi Jim Yeh4, Yuli Wang4, and Bradley J. Carr4

1Key Laboratory of Theory and Technology of Petroleum Exploration and Development in Hubei Province, Wuhan, China, 2Key Laboratory of Tectonics and Petroleum Resources, China University of Geosciences, Ministry of Education, Wuhan, China, 3Department of Geology and Geophysics, University of Wyoming, Laramie, WY, USA, 4Department of Hydrology and Atmospheric Science, University of Arizona, Tucson, AZ, USA

Abstract We characterized horizontal hydraulic conductivity (K) of a fractured granitic aquifer using single- and cross-hole hydraulic tests to evaluate “scale effect.” For selected boreholes, K estimates were obtained using single-hole FLUTe liner and slug tests. Several cross-hole pumping tests were carried out at various durations. Drawdown responses were first interpreted using analytical well-test solutions to obtain an effective horizontal conductivity (K_eff) assuming a homogeneous and infinite aquifer. The same drawdowns were then numerically inverted using transient hydraulic tomography (THT) to delineate spatial distributions of K and storativity in the area encompassing the boreholes. Papadopulos (1965) and a nonlinear least squares minimization method produced a similar principal K_eff direction that is consistent with the dominant fracture strike observed from outcrop and borehole televiewer data. However, principal direction and magnitude of this K_eff depend on the pumping test duration and the number of monitoring boreholes used in the interpretation. As a group, K obtained from cross-hole tests is larger than that obtained from single-hole tests. However, because cross-hole tests stressed the aquifer at both interwell and larger scales, K_eff obtained from interpreting cross-hole data is observed to decrease with pumping time, likely due to the dominance of less permeable fractures at larger scale. This lateral reduction of mean K is also revealed by THT as low K zones surrounding the test boreholes. Overall, K is found to increase from single-hole to the interwell scale and then decrease at larger scale, exhibiting a non-monotonic scale effect.

Plain Language Summary Knowledge of hydrological properties, particularly hydraulic conductivity (K), which controls groundwater flow in aquifers, is limited by the lack of subsurface data. Scientists have long observed that K estimated using field-testing methods that stimulate different volumes of aquifers can vary by several orders of magnitude, which creates a conundrum for how groundwater flow should be analyzed. To determine if scale effect is apparent in a fractured granitic aquifer in the Laramie Range, Wyoming, testing methods that stressed the aquifer with different water volumes were employed. Our results indicate that K increases from single-hole to the interwell scale but decreases at larger scale, with an overall nonmonotonic, albeit mild, scale effect. Compared to similar studies performed at other fractured rock sites, scale effect appears to be site-specific and is not generally transferable across different aquifers.

1. Introduction

In fractured crystalline aquifers, rock matrix generally has negligible porosity and fractures constitute the main conduits to groundwater flow and storage. In such aquifers, fracture hydraulic conductivity controls fluid flow, mass transfer, and energy transfer; therefore, its characterization is needed for quantitative analyses to assist evaluation, management, and remediation of groundwater resources. Hydraulic conductivity of crystalline fractured aquifers is controlled by individual fracture properties (e.g., aperture, length, and orientation) as well as properties of the fracture network (e.g., fracture density and connectivity). At both the individual and fracture network scales, hydraulic conductivity can exhibit heterogeneity and anisotropy, which impact the pathway and rate of flow (e.g., Berkowitz, 2002; Neuman, 2005). To estimate this parameter, both laboratory and field hydraulic testing methods have been developed. However, hydraulic
conductivity estimated using testing methods that stimulate different volumes of the aquifer can vary by several orders of magnitude (e.g., Brace, 1980, 1984; Clauser, 1992; Hyun et al., 2002; Jazayeri Noushahadi et al., 2011; Le Borgne et al., 2006; Martínez-Landa & Carrera, 2005; Winkler & Reichl, 2014; Yang et al., 2017; Zhan et al., 2016). This “scale effect” poses a challenge for the analysis of fluid flow and transport in fractured aquifers, as conductivity estimated at different aquifer volumes leads to scale-dependent estimates of transport velocities.

The “scale effect” of hydraulic conductivity is observed in both porous and fractured aquifers with an ongoing debate about the underlying cause (e.g., Butler & Healey, 1998a), Neuman (1994) and Schulze-Makuch et al. (1999) proposed generalized relationships between conductivities measured at different support scales from a theoretical point of view. Asymptotic behavior has also been suggested. For example, Clauser (1992) and Rovey and Cherkauer (1995) found that conductivity increased with a measurement scale until a critical distance, beyond which a constant regional value prevailed. Arguments against scale dependency include that the observed behavior could be caused by artifacts of the testing methods (e.g., Butler & Healey, 1998a, 1998b) or undersampling of open fractures in laboratory measurements (Hsieh, 1998; Renshaw, 1998; Zlotnik et al., 2000).

In fractured crystalline aquifers, $K$ estimates are often found to increase with the measurement scale (e.g., Le Borgne et al., 2006). At a fractured granite aquifer in Switzerland, which is located in a similar geological setting as our study site, it is shown that large-scale $K$ was controlled by a few interconnected flowing fractures while the majority of small-scale tests were performed in the matrix intervals (Martínez-Landa & Carrera, 2005). Accordingly, large-scale $K$ was over two orders of magnitude greater than the small-scale $K$. However, Maréchal et al. (2004) reported the absence of a scale effect in the weathered-fractured layer of Maheshwaram watershed, India. For a crystalline aquifer in central Arizona, USA, Illman (2006) reported that cross-hole $K$ was likely controlled by the connectivity of flowing fractures along specific directions, which increased with the volume of the aquifer tested. However, at the Mirror Lake fractured rock site, Hsieh (1998) reported a decrease of $K$ with scale due to the dominance of less permeable fractures at the large scale. Recently, a decrease of hydraulic properties from the tested well to some hundreds of meters is revealed based on a multiscale pumping experiment (only a single pumping well) conducted in a fractured formation (Guihéneuf et al., 2020).

This research uses single- and cross-hole hydraulic testing and interpretation methods to characterize hydraulic conductivity of a fractured granitic aquifer in the field to evaluate the “scale effect.” The field-testing methods, which interrogated increasing aquifer volumes, include: (a) single-hole FLUTe liner profiling of $K$ with a test duration of several hours; (b) single-hole slug test, including several slug-in and slug-out tests, with a test duration of several hours; and (c) cross-hole pumping test with a test duration of hours to days. In the following subsection, each testing method and associated interpretation techniques are discussed.

### 1.1. FLUTe Liner Profiling

FLUTe profiling is a high-resolution borehole method for rapidly estimating $K$ profiles along open boreholes (see Keller et al. (2014) for details). In our field-testing program, three open holes were profiled with FLUTe, whereas aquifer responses during profiling were analyzed using the analytical Thiem solution to infer a near-wellbore isotropic $K$ for every 0.3 m of the borehole interval (Ren et al., 2019). Although FLUTe results may be less accurate compared to short-interval straddle packer tests due to unresolved small groundwater velocity changes when the descent velocity of the liner is relatively large (e.g., Quinn et al., 2015). On the other hand, packer tests are more time-consuming and expensive than the FLUTe profiling method.

### 1.2. Single-Hole Slug Test

Compared to FLUTe profiling, conventional slug tests are widely carried out for aquifer characterization due to its speed and ease of execution. While cross-hole slug tests are reported for high-conductivity formations, single-hole response is more common in low-conductivity aquifers or in formations instrumented with small-diameter piezometers. A single-hole slug test imposes a small volume displacement and can provide an isotropic $K$ estimate in the vicinity of the borehole over an entire open hole intersecting many fractures (Butler, 1998). The interpretation of slug tests yielded a $K$ that reflected an average response over
the open hole. Depending on $K$ of the tested interval and well completion techniques, such interpretation can be affected by non-Darcian flow and skin effect, which may alter $K$ around the borehole due to incomplete well development (Butler et al., 1996).

### 1.3. Cross-Hole Pumping Test and Hydraulic Tomography

By inducing a larger rate of water withdrawal over a typically longer time, the pumping test induces cross-hole responses that are used to determine $K$ of the aquifer at the interwell scale. Pumping tests are commonly analyzed using classic analytical well-test solutions, which yield an effective horizontal hydraulic conductivity ($K_{\text{eff}}$) between a pair of source and monitoring wells while aquifer is assumed homogeneous and laterally infinite. However, if the aquifer exhibits conductivity heterogeneity, the estimated $K_{\text{eff}}$ varies significantly among the source-monitoring pairs (Wen et al., 2010; Wu et al., 2005). While common analytical well-test solutions yield an isotropic $K_{\text{eff}}$ when drawdowns are recorded at more than 3 monitoring boreholes, a tensor $K_{\text{eff}}$ with principal direction and anisotropy can be estimated using the “three-observation borehole method” (Papadopulos, 1965) or the distance-drawdown minimization approach (Wen et al., 2010; Wu et al., 2005).

If the cross-hole pumping tests are carried out in a sequential manner, that is, pumping a sequence of wells while monitoring responses in all other wells, hydraulic conductivity of the aquifer can be interpreted using a numerical inversion technique such as hydraulic tomography (HT) (e.g., Yeh & Liu, 2000; Zhu & Yeh, 2005). HT is a relatively new technique for obtaining spatially distributed hydraulic conductivity and storativity as well as their uncertainty estimates. Unlike the analytical solutions discussed above, recorded drawdowns from all pumping tests are jointly analyzed using inversion to obtain the spatial distribution of aquifer parameters (e.g., Cardiff & Barrash, 2011; Fischer et al., 2020; Wang et al., 2017; Yeh & Liu, 2000; Zhu & Yeh, 2005). An attractive feature of this technique is that conductivity is resolved at high resolution and even in 3D if drawdowns are collected at high spatial and temporal frequencies. Alternatively, if the tested boreholes are completed with single screens that span the thickness of the aquifer, HT can be used to estimate a spatially distributed and vertically averaged $K$ map. For fractured rocks, HT has been successfully tested using high-resolution drawdown and borehole flux measurements sampled from a numerical model (Zha et al., 2014). Field applications of this technique to fractured aquifers have also shown results indicating that fracture connectivity may be resolved (e.g., Castagna et al., 2011; Illman et al., 2009; Zha et al., 2016).

Our previous field-testing program (i.e., FLUTE and slug tests at three boreholes) in a granitic aquifer revealed a heterogeneous distribution of vertically averaged $K$ at individual borehole locations and among the boreholes (Ren et al., 2018, 2019). However, due to the short duration and small hydraulic forcing used, interwell connectivity, which controls fracture flow at the field scale, could not be delineated. In this work, a series of cross-hole pumping tests were carried out at selected bedrock wells at the same site. Cross-hole drawdown responses were first analyzed using analytical well-test solutions to infer an isotropic or anisotropic $K_{\text{eff}}$ at the interwell and larger scales. Using 2D transient hydraulic tomography (THT), cross-hole responses from multiple pumping tests were jointly analyzed using numerical inversion that yielded (locally isotropic) $K$ and storativity maps with uncertainty estimates. A sensitivity analysis of THT was conducted to identify if results were sensitive to assumptions of boundary conditions used for inversion. From both analytical and THT inversions, we examine if the dominant fracture orientations can be identified from the cross-hole drawdown data. The change in scale of the measurement is examined by comparing $K$ estimated using single- versus cross-hole tests. This study therefore contributes site-specific data at a fractured crystalline aquifer for the ongoing discussion about “scale effects.”

### 2. Study Site

The Blair Wallis fractured rock hydrology research well field is located at the Crow Creek watershed of the Laramie Range in southeastern Wyoming (Figure 1). It lies in a mountain watershed in a semiarid region with snowmelt-driven hydrology. The well field, approximately 500 m by 300 m, lies on an exposed hillslope in a US Forest Service land ~21 km southeast of Laramie, Wyoming. Geological studies suggest that the well field is part of the Rocky Mountain surface (Bradley, 1987; Chapin & Kelley, 1997; Eggler
et al., 1969; Evanoff, 1990) that originated from the Eocene time (Gregory & Chase, 1994; Mears, 1993; Moore, 1960; Scott & Taylor, 1986). The bedrock of the Laramie Range is highly fractured due to a series of mountain-building events, to the extent that the northern portion of the Range may have rotated southward (Blackstone, 1996). Mapping studies have also identified numerous surface lineaments in the region (Ochsner, 2014). From airborne surveys (unpublished data), several NNE striking lineaments can be seen south of the well field while they become indistinct at the well field. Approximately 10–18 m of weathered granitic soil (i.e., saprolite) overlies granite bedrock at the well field. The granite consists of microcline, quartz, plagioclase, biotite, hornblende, and ilmenite, although biotite and porphyritic granite also exist (Edwards & Frost, 2000; Frost et al., 1999). Due to the composition variation, granite at the site is divided into two facies, namely “Sherman” and “Lincoln” granite. The bedrock, which lies in the saturated zone, is the focus of this investigation.

From granite outcrops surrounding the well field, fracture length is estimated to vary from 3 to 30 m (Ren et al., 2018). Core samplings, geophysical logging, and hydraulic tests confirmed that many bedrock fractures are open in the subsurface, allowing water storage and transmission at the kilometer scale (Flinchum et al., 2018; Ren et al., 2018). From borehole image logs, fracture density varies from less than 5 to over 20 per foot interval.

From 2014 to 2017, nine boreholes, named BW1-BW9 (Figure 1), have been drilled into the fractured granite to investigate fractures and their impact on subsurface flow. All boreholes are cased ∼5 feet below the bottom of the saprolite but remain open in the granite bedrock. A schematic well completion diagram in relation to the position of saprolite/bedrock at the site is provided in Ren et al. (2018). Note that BW4 lies in a sulfide-rich facies (Lincoln granite) while the rest of the wells lie in the Sherman granite (Hayes, 2016). A transition between these facies thus occurs between BW 4 and BW1.

At BW1-5, after the casing was set, cores were recovered using a wireline coring system (cores were not recovered from BW6-9). The recovered cores were photographed using a Geotek core logger with a Sigma 105-mm microlens, which identified fractures in both subvertical and sub-horizontal to horizontal directions. For some fractures, mineral fillings, which reduced their porosity, have been observed. For all nine boreholes, a Mount Sopris Optical and Acoustic Televiewers (Mt. Sopris Instruments, Denver, CO) were deployed to provide 360° unwrapped borehole images. From the image logs, preferential fracture orientation and dominant fracture strike among the subvertical fractures were identified to be NNE (Novitsky et al., 2018; also see Figure S1), which is consistent with the orientation of surface lineaments observed

**Figure 1.** Location and light detecting and ranging (LIDAR) topography of the Blair Wallis fractured rock research well field in southeastern Wyoming, USA (star in the inset). The red dotted box indicates the area in which cross-hole pumping tests were conducted for this study.
south of the well field from Google Earth. Given the core data, borehole image logs, and well test responses, the same subvertical lineaments are believed to exist beneath the well field, explaining observed poor fracture connectivity among some of the wells. Moreover, borehole image and flowmeter logging revealed that only a subset of the fractures is flowing (Ren et al., 2018).

At each bedrock borehole, water level and temperature monitoring data have been collected every 15 min since 2015. Depth-to-water varies between 5 and 15 m bgs (below ground surface) from borehole water level measurement, and the inferred water table follows the topographic slope in the local watershed (Ren et al., 2018). The background water level fluctuates in response to annual snowmelt recharge and lasts till dry season, although this fluctuation is generally less than 1.5 m. Specifically, snowmelt in early spring dominates bedrock recharge, while recharge from rainfall infiltration later in spring is negligible, likely due to high evapotranspiration and low soil moisture content above the saturated zone during the growing season, limiting bedrock recharge. Overall, the water table lies within the granite bedrock, which remains primarily saturated over the annual recharge cycle. Though the granite bedrock is unconfined, individual fractures in the bedrock below the water table are confined. For these reasons, our investigation targets the dry seasons.

### 3. Methods

#### 3.1. Field Testing Methods

##### 3.1.1. Single-Hole Tests

Slug test and FLUTe liner profiling test (see Keller et al. (2014) for details) were carried out to estimate near-wellbore $K$ at three boreholes by Ren et al. (2018, 2019). In this study, additional slug tests were performed at BW1, BW8, and BW9 (Table 1). We also tested well skin and non-Darcian flow effects. We used two sizes of solid slugs to generate two different initial water level displacements (i.e., Large $H_0$ and Small $H_0$) at each well. We refer to our earlier work for the test design (Ren et al., 2018). Details of the slug test configuration are summarized in Tables S1 and S2.

<table>
<thead>
<tr>
<th>Boreholes</th>
<th>Slug test</th>
<th>FLUTe profiling</th>
<th>Pumping test</th>
<th>Monitoring boreholes</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW1</td>
<td>05/31/2017 (6 hours and 37 minutes)</td>
<td>/</td>
<td>10/25/2017 (35 mins, ~7 gpm)</td>
<td>BW4, BW6, BW7, BW8, and BW9</td>
</tr>
<tr>
<td>BW4</td>
<td>/</td>
<td>/</td>
<td>11/18-19/2015 (28 hours, variable pumping rate)</td>
<td>BW1</td>
</tr>
<tr>
<td>BW5</td>
<td>06/07/2017 (4 hours and 21 minutes)</td>
<td>2016/10/11 (~4 hours)</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>BW6</td>
<td>05/17/2017 (10 hours and 20 minutes)</td>
<td>2016/10/11 (~6 hours)</td>
<td>11/16/2017 (246 mins, ~5 gpm)</td>
<td>BW1, BW4, BW7, BW8, and BW9</td>
</tr>
<tr>
<td>BW7</td>
<td>05/24/2017 (4 hours and 46 minutes)</td>
<td>2016/10/12 (~6 hours)</td>
<td>08/22-24/2017 (44 hours, ~4.0 gpm)</td>
<td>BW1, BW4, BW6, BW8, and BW9</td>
</tr>
<tr>
<td>BW8</td>
<td>06/02/2017 (7 hours and 11 minutes)</td>
<td>/</td>
<td>10/04/2017 (95 mins, ~3.0 gpm)</td>
<td>BW1, BW4, BW6, BW7, and BW9</td>
</tr>
<tr>
<td>BW9</td>
<td>06/01/2017 (4 hours and 12 minutes)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

*Note. Please refer to Tables S1 and S2 for more details about the borehole configuration. Slug tests at BW1, BW8, and BW9 and all the pumping tests are new to the current study while the FLUTe profiling and slug tests at BW5, BW6, and BW7 have been reported previously in Ren et al. (2018). Cross-hole pumping tests in the grey shaded region were used for HT experiment.*
3.1.2. Cross-Hole Pumping Tests

The well field and nearby drainage areas were covered by snow in winter, which did not melt until early spring (April to June). Therefore, the pumping tests were conducted during the summer-to-early-winter timeframe when recharge to the aquifer was negligible. This condition is considered ideal for well-test interpretation techniques that do not consider a recharge flux. The recharge boundary condition (BC) is generally unknown and is difficult to estimate from borehole data (Scanlon et al., 2006). During the well test seasons in 2015–2017, five pumping tests were conducted to evaluate the magnitude of (or lack thereof) cross-hole responses (Table 1).

We pumped a selected borehole at a nearly constant rate for each pumping test and measured changes in water levels at the pumping and monitoring boreholes using downhole pressure transducers (In-Situ Level-Troll 400 and 500) and a water level meter. At the pumping borehole, the transducer was placed ~12.7 cm above the pump, which is set ~1 m above the borehole bottom. In the monitoring boreholes, the transducers were placed sufficiently deep below the initial water level to ensure that all drawdowns can be measured. All transducers recorded either absolute pressure (i.e., gauge pressure plus barometric pressure) or water pressure every minute. A separate HOBO barologger, installed in BW2, recorded air pressure every 30 min. The recorded pressure data at BW8 and BW9 (in which the LevelTroll loggers recorded the absolute pressure) were also examined for barometric effects, which were found insignificant (no major storms passed). Water level data collected before and after the tests do not suggest significant external recharge during the five pumping tests.

During the well test, manual water level measurements using the water level meter were taken according to the frequency recommendation of Weight (2008). The manual water level measurements were used to (a) monitor the water level in the pumping well during the test to prevent the water level from dropping below the pump intake; (b) verify and calibrate all transducer data recorded during the tests. Depth-to-water (DTW) data computed from the estimated water pressure at transducers were shifted against manual DTWs until they fit in a least square sense. Following the advice of Zhao and Illman (2018), a 15-point centrally weighted moving average method was then applied to these data to remove the noise before drawdowns were used for parameter estimation.

3.2. Methods of Analysis

The following analytical interpretation methods for single- and cross-hole tests assume a two-dimensional, depth-averaged, homogeneous, and unbounded aquifer with either isotropic or anisotropic hydraulic conductivity. THT analysis of the cross-hole tests conceptualized the aquifer as a highly parameterized heterogeneous, locally isotropic, bounded, two-dimensional plane porous medium.

3.2.1. Slug Test Interpretation

We used the Kansas Geological Survey (KGS) (Hyder et al., 1994) model, which assumes a fully or partially penetrating borehole in unconfined or nonleaky confined aquifer, to estimate $K$, which reflects an average value over open hole at each test borehole. In this interpretation, well-skin effect and non-Darcian flow were found negligible (Ren et al., 2018). After $K$ was estimated at each borehole, Kriging (Matheron, 1967) was used to generate an interpolated $K$ map.

3.2.2. Cross-Hole Drawdowns Interpreted With Analytical Methods

From the observed drawdowns in a monitoring borehole, Papadopulos and Cooper (1967)’s method was first used to estimate an isotropic $K_{\text{eff}}$ between the pumping and the monitoring boreholes. This model assumes no leakage during a pumping test in a homogeneous and isotropic confined aquifer of infinite areal extent. Furthermore, the generalized radial flow (GRF) model for fractured rocks, which generalizes the flow dimension to nonintegral values, while retaining the assumptions of radial flow and homogeneity (Barker, 1988) is also used to estimate $K_{\text{eff}}$ and $S_{w}$. The GRF model introduces a fractional dimension $n$, which characterizes the change of the flow section geometry according to the distance from the pumping borehole (Doe, 1991).
Based on drawdowns collected at three or more monitoring boreholes, the Papadopulos (1965)’s semi-log straight-line method, which assumes a homogeneous and anisotropic infinite aquifer, was used to determine an anisotropic $K_{\text{eff}}$ tensor. Generally, the monitoring boreholes are required to be at different angles from the pumping well in order to determine $K_{\text{eff}}$. Using BW6’s pumping test as an example, the straight-line procedure is summarized as follows:

1. Establish a Cartesian coordinate with the origin at BW6 and record the coordinates of each monitoring borehole (i.e., BW1, BW7, BW8, and BW9);
2. Choose three of the 4 monitoring boreholes and plot drawdowns ($s$; linear) versus the elapsed time ($t$; logarithmic) on a semi-log paper for each borehole;
3. Calculate the best-fit slope of the late-time straight-line portion of the semi-log plot for each monitoring borehole, then average the three slopes;
4. Use the average slope to re-fit the straight-line portions of the 3 semi-log plots to obtain the t-intercept ($t_0$) for each monitoring borehole;
5. The average slope is related to transmissivity components through the following equation:

$$\Delta s \over \text{logcycle} = \frac{2.303 Q}{4 \pi \sqrt{T_{xx} T_{yy} - T_{xy}^2}}$$

(1)

where $\Delta s$ is the average slope of the 3 semi-log time-drawdown plots, $T_{xx}$, $T_{yy}$, and $T_{xy}$ are components of the transmissivity tensor [L/T], and $Q$ is the pumping rate [L/T].

6. Apply Equation 2 to the fitted data of each monitoring borehole:

$$t_0 = \frac{S}{2.25} \left( \frac{T_{xx} x^2 + T_{yy} y^2 - 2T_{xy} xy}{T_{xx} T_{yy} - T_{xy}^2} \right)$$

(2)

where $S$ is storage coefficient [-], $x$ and $y$ are coordinates of a monitoring borehole, and $t_0$ is its t-intercept. Combined with Equation 1, four equations are developed to solve for 4 unknowns: $S$, $T_{xx}$, $T_{yy}$, and $T_{xy}$.

7. Substitute $T_{xx}$, $T_{yy}$, and $T_{xy}$ into Equations 3–5 to determine the principal transmissivities $T_{zz}$ and $T_{qq}$ and orientations of the principal axes.

$$T_{zz} = \frac{1}{2} \left( T_{xx} + T_{yy} \right) + \left( T_{xx} - T_{yy} \right)^2 + 4T_{xy}^2 \right)^{1/2}$$

(3)

$$T_{qq} = \frac{1}{2} \left( T_{xx} + T_{yy} \right) - \left( T_{xx} - T_{yy} \right)^2 + 4T_{xy}^2 \right)^{1/2}$$

(4)

$$\theta = \arctan \left( \frac{T_{zz} - T_{xx}}{T_{yy}} \right)$$

(5)

where $\theta$ is a counterclockwise angle from $x$ to the $z$ axis.

8. Divide transmissivity components and $S$ by saturated thickness at the pumping borehole to obtain the $K_{\text{eff}}$ tensor and specific storage ($S$). Average the principal components of $K_{\text{eff}}$, using geometric mean ($K_\text{g}$) to compute an isotropic mean $K_{\text{eff}}$.

9. Choose another combination of 3 monitoring boreholes and repeat the above.

For an ideal homogeneous anisotropic aquifer, when the semi-log time-drawdown data of all monitoring boreholes are compared, slopes of the late-time drawdowns should be the same. However, this would not be true if aquifer exhibits significant heterogeneity or drawdowns contain noise. In this case, the “three-observation-borehole” method can yield a randomly variable $K_{\text{eff}}$. Wen et al. (2010) suggested that drawdowns from all monitoring boreholes could be used to obtain a representative $K_{\text{eff}}$. In this study, a nonlinear least squares minimization method was used to obtain an optimized $K_{\text{eff}}$ at different elapsed times since the start of a pumping test. For a given elapsed time $t$, a best-fit $K_{\text{eff}}$ was found by minimizing the difference between observed and calculated drawdowns using the following equations:

$$s(\xi, t) = \frac{Q}{4 \pi \sqrt{T_{xx} T_{yy} - T_{xy}^2}} W(u_\xi)$$

(6)
To initialize SimSLE, a set of prior mean assumed fully known during inversion. Flow rate—constant or temporally variable—from the field. Thus, source/sink for each pumping test was sides of the model. Recharge flux was set to zero. Each pumping test was simulated using the measured conditions yielded similar results). Constant head boundary conditions (2,465 m) were assigned to all four water was static at the start of each pumping test (inversion assuming a steady state flow field as initial developed by Yeh et al. (1993). Drawdown-time selection of the THT analysis followed suggestions from previous THT studies. According to Sun et al. (2013), measured drawdown at the interpolated time at which theoretical drawdown is zero in the Cooper–Jacob log-linear approach is highly correlated with $S_i$. In contrast, late-time drawdowns are strongly correlated to $K$. Furthermore, for estimating $S_i$ using SimSLE, Yeh et al. (2011) and Mao et al. (2013) suggested that head change over at least two time steps was necessary. Besides, drawdowns are highly correlated over time, Zhu and Yeh (2005) suggested that 4–5 points sampled from each observation well hydrograph were sufficient to condition the inversion, reducing the computational burden. In this study, we investigated all the sampling strategies proposed above to condition inversion.

The conditional $L_2$ norm of the heads quantifies the estimation errors at the end of each iteration during inversion:

$$u_c = \frac{s}{4t} \left( T_{xx} y^2 - 2T_{xy} xy + T_{yy} x^2 \right)$$

$$\text{minimize} \sum_i \left[ s(r_i,t) - s^*(r_i,t) \right]^2$$

where $s(r, t)$ is the theoretical drawdown calculated by Equation 6 for monitoring well $i$, $s^*(r, t)$ is observed drawdown at this monitoring well at distance $r$, from the pumping well at elapsed time $t$, and $W(u_c)$ is the well function.

### 3.2.3. Interpretation of Cross-Hole Drawdowns With THT

For imaging hydraulic conductivity distribution in synthetic (e.g., Hao et al., 2008) and natural porous and fractured aquifers (e.g., Illman et al., 2009; Zha et al., 2016), numerical inversion techniques based on the HT concept and the successive linear estimator (SLE) have been developed (Yeh et al., 1996; Yeh & Liu, 2000; Zhu & Yeh, 2005). The inversion program is available at www.tian.arizona.edu/ download, while details on the theory and its implementation are described in the above references.

In this study, four cross-hole pumping tests were performed among six bedrock wells, the measured drawdowns thus reflect vertically averaged values from fractures intersected at the borehole locations. A 2D THT analysis was performed in plane view to estimate the spatial distribution of vertically averaged horizontal $K$ and $S_i$. To obtain these images, transient drawdowns from the four pumping tests were inverted simultaneously using the SimSLE algorithm (Xiang et al., 2009).

For the THT analysis, a 160 × 80 m² model domain was created to include all pumping wells and was discretized into 80 × 40 uniform square (2 × 2 m²) cells. During inversion, a forward model simulated each transient pumping test using Variably Saturated Flow and Transport in 2D (VSAFT2), a finite element code developed by Yeh et al. (1993). Initial conditions for the forward model were set by assuming that groundwater was static at the start of each pumping test (inversion assuming a steady state flow field as initial conditions yielded similar results). Constant head boundary conditions (2,465 m) were assigned to all four sides of the model. Recharge flux was set to zero. Each pumping test was simulated using the measured flow rate—constant or temporally variable—from the field. Thus, source/sink for each pumping test was assumed fully known during inversion.

To initialize SimSLE, a set of prior mean $K$ and $S_i$ fields and their spatial correlation scales are required. In this work, exponential correlation functions were assumed to describe the unconditional (pre-inversion) spatial structures of ln$K$ and ln$S_i$: the correlation scales for both parameters were set to 15 m initially along both the simulation $x$ and $y$ axes. The unconditional, prior variances of ln$K$ and ln$S_i$ were set to 5.0 and 2.0, respectively, representing the initial uncertainties of these parameters. The inversion started with a cokriging step using all available measurements of hydraulic properties and hydraulic heads to produce the conditional parameter fields. The conditional means of these parameters were then populated into the forward model to predict heads at selected observation locations and times. The linearly estimated parameter fields of ln$K$ and ln$S_i$ and their residual variances were then iteratively updated by SimSLE by minimizing the differences between observed and simulated heads. This iteration aims to overcome the limitation of the linear estimator for the nonlinear relationship between $K$ and the head $h$ (see Yeh et al., 1996; Yeh & Liu, 2000).

Drawdown-time selection of the THT analysis followed suggestions from previous THT studies. According to Sun et al. (2013), measured drawdown at the interpolated time at which theoretical drawdown is zero in the Cooper–Jacob log-linear approach is highly correlated with $S_i$. In contrast, late-time drawdowns are strongly correlated to $K$. Furthermore, for estimating $S_i$ using SimSLE, Yeh et al. (2011) and Mao et al. (2013) suggested that head change over at least two time steps was necessary. Besides, drawdowns are highly correlated over time, Zhu and Yeh (2005) suggested that 4–5 points sampled from each observation well hydrograph were sufficient to condition the inversion, reducing the computational burden. In this study, we investigated all the sampling strategies proposed above to condition inversion.
\[ L_2^{(i)} = \frac{1}{m} \sum_{i=1}^{m} \left( h_i^{*} - \hat{h}_i^{(i)} \right)^2 \] (9)

where \( h_i^{*} \) and \( \hat{h}_i^{(i)} \) are observed and simulated heads at \( r \)th iteration, respectively. The subscript \( i \) is the number index for the observations; \( m \) is the total number of head observations used in inversion.

4. Results

4.1. Estimation of \( K \) From Single-Hole Tests

The slug tests and their analysis at boreholes BW5, 6, and 7 were presented in Ren et al. (2018). They concluded that skin effect and non-Darcian flow were negligible during the slug tests in all boreholes. This is not unexpected at this well field because the non-Darcian flow can be ignored when the slug-test-estimated \( K \) varies from \( 10^{-7} \) m/s to \( 10^{-5} \) m/s. \( K \) estimated with the KGS model were summarized in Table S3. A Kriged map of the slug-test estimated \( K \) further exhibits a clear decreasing trend from NE to SW (Figure S2). Moreover, for boreholes BW5, 6, and 7, a FLUTE liner was used to profile \( K \) for a 21.02, 43.69, and 55.76 m interval below the well casing, respectively (Ren et al., 2019).

4.2. Visual Analysis of Drawdown-Time Behaviors From Cross-Hole Pumping Tests

BW1 has a shallow depth compared to the other wells (Table S1) and its pumping test lasted a short duration. We did not detect responses at the other wells, likely due to the limited development of the cone of depression. Consequently, BW1 pumping test was neither used in the cross-hole analytical nor THT analyses. On the other hand, during a 28-h pumping test at BW4 (a deep well), only BW1 was monitored—BW6, BW7, BW8, and BW9 were not installed yet—and BW1 showed no response to this test. Nonetheless, this data set was included in the THT analysis because it indicates disconnection between BW4 and BW1. On the other hand, this test was excluded from the cross-hole analytical analysis.

During pumping tests at BW1, BW6, BW7, and BW8, we did not observe drawdown response in BW4. We surmise that there must exist low fracture connectivity between BW4 and the rest of the well field, which is consistent with the known geology (BW4 lies in the sulfide-rich “Lincoln” granite) (Hayes, 2016) and lineaments mapped in the region. The inferred poor connectivity between BW4 and the rest of the well field is likely due to this facies transition.

Figure 2. Normalized drawdowns versus normalized elapsed time during the pumping test of BW7 (a), BW8 (b), and BW6 (c). \( Q \) is the pumping rate of a test, \( s \) is the drawdown of a monitoring borehole, and \( r \) is the distance between the pumping borehole and a given monitoring borehole.

Time-drawdown data at the monitoring boreholes during the pumping tests of BW7, 8, and 6 are displayed in Figure 2. For each test, drawdowns at a monitoring borehole were normalized by the pumping rate \( Q \), while the elapsed time \( t \) was normalized by the squared distance between the pumping and monitoring boreholes, \( r^2 \). These normalized hydrographs attenuate the effect of distance and pumping rate on the absolute drawdown response at each monitoring well, thus reveal any connectivity between the pumping borehole and the other boreholes.

Figure 2a shows the normalized hydrographs at BW1, BW6, BW8, and BW9 during the BW7 pumping test. Water level at borehole BW9 responded first, followed by BW8 and BW6. BW1 showed almost no response during this test (the total drawdown of BW1 is 3.40 cm). The rapid response at BW9 suggests that BW9 is most connected with BW7 while fracture connectivity is likely poor between BW1 and BW7. It is worth noting that the irregular drawdowns at BW6 and BW8 between 216 and 360 min (i.e., \( t/r^2 \) is between 0.52 and
0.87 min/m²) were likely due to a variable pumping rate when pumping BW7. Sediments were mobilized as the pressure cone of depression expanded from BW7 into connected fractures further away. This mobilization partially clogged the intake of the submersible pump (which was verified post-test), causing its rate to fluctuate. Though a constant rate is not required by THT interpretation, this fluctuation will introduce estimation errors into the interpretation with the analytical cross-hole methods, which assume a constant rate.

When pumping BW8, BW7 showed the fastest response followed by BW6 (Figure 2b). BW9 showed responses at a later time while BW1 showed no response. BW8 is connected with BW7, while low fracture connectivity likely exists between BW8 and BW1, BW6, and BW9. Note that during the first 30 min of this pumping test, drawdown data recorded at BW9 was negative while manual measurements using a water level meter at the same time did not indicate cross-hole responses in this well (Figure S3). Because determining DTW at this well required the calibration of the recorded absolute pressure by the air pressure series, different data collection frequency between In-situ LevelTroll transducer and HOBO barologger had to be resolved using temporal interpolation. This interpolation could have introduced an error. Hence, these negative drawdowns were excluded from the data interpreted by either analytical methods or THT.

During the pumping test of BW6, BW7 and BW1 showed responses at nearly the same normalized time, followed by BW8 and BW9 (Figure 2c), suggesting that compared to BW8 and BW9, BW1 and BW7 are more connected to BW6. It is worth noting that drawdown at BW1 occurred earliest, but at late time drawdown at BW1 is less than BW9. In a homogeneous infinite aquifer undergoing constant-rate pumping, the key factor controlling the drawdown arrival time is hydraulic diffusivity ($D = K/\phi S_s$, where $S_s$ is specific storage [1/L]), while the factor controlling drawdown magnitude is largely $K$. In a heterogeneous aquifer in the field, it may be difficult to state such a simple relation because of heterogeneity in both $K$ and $S$ and the possible influence of boundaries. On the one hand, we have evidence of heterogeneity: Slug test results indicate that BW9 has the lowest near-wellbore $K$ (Table S3) and once the pressure cone reached BW9 during a pumping test, water level at BW9 dropped faster than the other boreholes. Drawdown at BW8 also shows a similar trend at late time due to its lower near-wellbore $K$ (see Table S3). On the other hand, observed drawdowns at a well theoretically depend not only on the drawdown’s sensitivity to the anomaly’s location but also on its magnitude (Wen et al., 2019). These facts highlight the need for a THT analysis.

### 4.3. Estimation of $K_{\text{eff}}$ Between a Pair of Pumping and Observation Boreholes

Besides the above visual inspection and intuitive analysis, we used the Papadopulos and Cooper (1967) model’s solution to determine $K_{\text{eff}}$ from drawdown data collected from observation wells during pumping tests at BW6, 7, and 8 (Table 2). The BW1 and BW4 pumping tests were not analyzed because they did not produce noticeable responses at other wells.

Overall, the estimated $K_{\text{eff}}$ for different pumping and monitoring pairs varies by up to a factor of 7 (Table 2). In particular, the estimated $K_{\text{eff}}$ between pumping well BW7 and three monitoring boreholes varies within a factor of two. Using BW8 as the pumping well, we found the estimated $K_{\text{eff}}$ using drawdowns at BW7 five times greater than that using drawdowns at BW9. Using BW6 as the pumping well, the estimated $K_{\text{eff}}$ based on the other four wells’ responses vary within a factor of 3 (Table 2). Such variations in the estimates suggest the highly heterogeneous nature and complex distribution of fractures at the site. Moreover, the anisotropy between the same two boreholes in different directions seems to be negligible. The estimated $T_{\text{eff}}$ between BW7-BW8 and BW7-BW6 are similar for both BW7 and BW8 pumping tests, as well as for BW7 and BW6 pumping tests. Moreover, the duration of each test was different, and different volumes of the aquifer were stressed, which will result in different estimated $T_{\text{eff}}$. The variability observed among multiple pumping tests is likely a result of different fracture network connectivity between boreholes (Straface et al., 2007; Wen et al., 2010; Wu et al., 2005). The larger $T_{\text{eff}}$ indicates the fractures are more connected between borehole pairs.

These drawdown data are further analyzed using the GRF model (Table S4). The obtained flow dimensions ($n$) are around 2, suggesting a 2D cylindrical flow in an isotropic media. Overall, the estimated $K_{\text{eff}}$ using the GRF model are similar to those estimated using the Papadopulos and Cooper (1967) solution (except for the two estimates from BW8 and BW9 when pumping BW7), but the estimated $S_s$ using the GRF model are more variable comparing with those estimated using the Papadopulos and Cooper (1967) solution.
4.4. Estimation of Hydraulic Conductivity Anisotropy

For the BW6 pumping test, drawdowns at monitoring boreholes were jointly interpreted with the semi-log straight-line method of Papadopulos (1965) to determine interwell horizontal conductivity anisotropy. For this test, semi-log drawdowns versus time at all four monitoring boreholes are shown in Figure 3. Straight-line (late-time) portions of the drawdowns at BW1, BW7, and BW8 have similar slopes, while that of BW9 is greater (Figure 3), suggesting the heterogeneity in hydraulic diffusivity, and the different t-intercept \( t_0 \) indicate that compared to BW8 and BW9, BW1 and BW7 are more connected to BW6. Using data from three of the four monitoring boreholes, hydraulic parameters were estimated for four monitoring well combinations (Table 3). The maximum principal transmissivity lies in NNE (except for Case 2), consistent with the dominant strike of fractures observed from outcrop and borehole televiewer data (Novitsky et al., 2018; Figure S1). However, for cases 2 and 3, a computed \( (ST_{xx}) (ST_{yy}) (ST_{xy})^2 \) is negative, indicating that transmissivity cannot be determined from drawdowns at these monitoring boreholes (Maslia & Randolph, 1987). While

![Figure 3](image_url)

*Figure 3.* Semi-log drawdowns versus time at 4 monitoring boreholes (dots) during the BW6 pumping test. The fitted dashed lines to late-time drawdowns have slopes of 0.072 m for both BW7 and BW8, 0.098 m for BW1 and 0.201 m for BW9.
measurement noise can result in a discrepancy between theoretical and observed drawdowns, fracture heterogeneity (e.g., difference in fracture connectivity between boreholes) is likely the major cause. Because of the issue, a minimization approach to determine the anisotropy is presented next.

Using the least squares approach (Equations 6–8), we have obtained the estimates for Cases 2 and 3, intractable using the analytical method. The results are listed in Table 3 in gray-shaded columns. As shown in Table 3, even though the pumping location is fixed at BW6, estimates from either analytical or least squares methods change with different combinations of three observation boreholes from the four. Such a change in the estimate due to the change of one of the three observation boreholes signifies the explanation of THT’s logic by Wen et al. (2019). That is, adding a new observation borehole provides new information about heterogeneity to THT analysis.

We also used the nonlinear least squares minimization to estimate the parameters using drawdowns from the four monitoring boreholes induced by pumping at BW6 and those by pumping at BW7. For the BW6 pumping test, the optimized parameters are $T_{xx} = 5.00 \times 10^{-4}$ m$^2$/s, $T_{yy} = 4.00 \times 10^{-4}$ m$^2$/s, $T_{xy} = 2.00 \times 10^{-4}$ m$^2$/s, and $S = 0.37$. The corresponding $T_{xx}$, $T_{yy}$, and $	heta$ are 6.56 $\times 10^{-4}$ m$^2$/s, 2.44 $\times 10^{-4}$ m$^2$/s, and 38$^\circ$, respectively. The isotropic means are: $T_{eff}^{xy} = 4.12 \times 10^{-4}$ m$^2$/s, $K_{eff}^{xy} = 8.65 \times 10^{-6}$ m/s, which differ from those estimated with Papadopulos (1965) by a factor less than 2. We believe that this difference is small, largely due to the similarity in the equations solved by these methods. Furthermore, $T_{eff}^{xy}$ determined using data from the four monitoring boreholes is close to the geometric mean of $T_{eff}^{xy}$ (3.95 $\times 10^{-4}$ m$^2$/s) of the same four monitoring boreholes obtained with the Papadopulos and Cooper (1967)’s solution. However, $S$ determined with the four monitoring boreholes is two orders of magnitude greater than those estimated with Papadopulos (1965) and Papadopulos and Cooper (1967) method. Given the sparse wells we have, storativity estimate for the fractured granite could suffer some inaccuracies.

For the BW7 pumping test with BW1, BW6, BW8, and BW9 as monitoring wells, the best-fit parameters are $T_{xx} = 1.30 \times 10^{-3}$ m$^2$/s, $T_{yy} = 3.40 \times 10^{-3}$ m$^2$/s, $T_{xy} = 2.10 \times 10^{-3}$ m$^2$/s, and $S = 0.21$. The corresponding $	heta$, $T_{eff}^{xy}$, and $K_{eff}^{xy}$ are 58$^\circ$, 1.00 $\times 10^{-4}$ m$^2$/s, and 1.64 $\times 10^{-6}$ m/s, respectively. For the BW8 pumping test, drawdowns were only observed at BW7 and BW9, which were insufficient to solve for the tensor transmissivity and $S$.

<table>
<thead>
<tr>
<th>Monitoring boreholes used</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW1, BW7, BW8</td>
<td>0.081</td>
<td>/</td>
<td>0.124</td>
<td>/</td>
</tr>
<tr>
<td>BW1, BW7, BW9</td>
<td></td>
<td></td>
<td>0.124</td>
<td>/</td>
</tr>
<tr>
<td>BW1, BW8, BW9</td>
<td></td>
<td></td>
<td></td>
<td>0.115</td>
</tr>
</tbody>
</table>

Table 3

Hydraulic Parameters Estimated Using Papadopulos (1965) and Nonlinear Least-Squares Minimization Methods for the BW6 Pumping Test Using Three Monitoring Boreholes

Note. Pumping rate $Q = 3.15 \times 10^{-4}$ m$^3$/s. N/A: computed ($ST_{xy}$)($ST_{xy}$) is negative so transmissivity could not be determined from the monitored drawdown data using Papadopulos (1965) analytical solution. $T_{eff} = \sqrt{T_{xx}T_{yy}}$, while $K_{eff} = \sqrt{K_{xx}K_{yy}}$. Values in the gray shaded region are estimated using nonlinear least squares minimization methods.
4.5. Estimated $K$ and $S$ Distributions With THT

Finally, the observed differential cross-hole responses during the sequential pumping tests of BW4, BW7, BW8, and BW6 were interpreted jointly using THT to image the distribution of $K$ and $S$. The inversion was initialized with homogeneous mean fields of $K = 1.0 \times 10^{-5} \text{ m/s}$ and $S = 1.0 \times 10^{-4} \text{ m}^{-1}$, which were average values obtained from interpreting the same cross-hole tests with the analytical solutions. The inverse model domain extends to BW4, where no drawdown was detected during the pumping tests at BW1, 6, 7, and 8. The zero drawdown implies that BW4 is hydraulically disconnected from the other wells. This piece of information is useful. Thus, BW4 was included in the THT analysis and the model domain included BW4.

Following Mao et al. (2013), Sun et al. (2013), Yeh et al. (2011), and Zhu and Yeh. (2005), observed heads were sampled from the monitored drawdowns, which included both early- and late-time data to condition the inversion. To achieve convergence, the number of sampling points depends on the characteristics of each monitoring well hydrograph. Fewer points could be selected if the hydrograph exhibits low variability; conversely, more points were used if the hydrograph varies significantly. For the BW7 pumping test, 4, 12, 11, and 9 heads were sampled from the hydrograph of BW1, BW6, BW8, and BW9, respectively. For the pumping tests of BW8 and BW6, however, only four heads were sampled from the hydrograph of each monitoring well, which yielded satisfactory results without incurring an unduly high number of inversion iterations. Figure 4 shows the sampling points along with the observed and calibrated drawdowns.

According to Table 3, the long principal hydraulic conductivity direction is NNE, consistent with the findings reported by Novitsky et al. (2018). Thus, we have selected the correlation lengths to 7.5 and 15 m in $x$ and $y$ directions for both $\ln K$ and $\ln S$, as the prior information for the inversion.

The $K$ and $S$ tomograms estimated with SimSLE after assimilating the four pumping tests are shown in Figure 5 along with $D$ map, which is the ratio of $K$ to $S$. The $K$ map yielded two “high-$K$ zones,” which indicate a greater connectivity of the flowing fractures between BW7 and BW9 and between BW7 and BW8 (Figure 5a), while there is low connectivity of flowing fractures between BW7 and BW1 and between BW8 and BW1. The estimated $S$ tomogram is displayed in Figure 5b. In general, $S$ tends to be smaller at boreholes with rapid drawdown responses. Both the estimated $K$ and $S$ tomograms are considered plausible because they explain the largest drawdown responses observed in the field: (a) BW9 recorded large and rapid responses during the pumping of BW7; (b) BW7 recorded rapid responses during the pumping of BW8; and (c) BW7 recorded rapid responses during the pumping of BW6.

The map of hydraulic diffusivity (Figure 5c) shows how BW8, BW9, BW6, BW7, and BW1 are connected. BW7/BW9, BW6/BW1, and BW9/BW6 are most likely connected by fracture zones (the red zone in Figure 5c). In contrast, BW8/BW9 and BW7/BW6 are not completely connected. This observation seems to agree well with Table 2. As explained in Section 4.2, a larger $D$ leads to an earlier arrival of the first arrival. The $D$ map supports the qualitative interpretation of the observed drawdowns (Figure 2).

Overall, the inversion result is satisfactory, and THT with an equivalent porous media model can capture the dominant fractures at test boreholes (Dong et al., 2019). Note that it would be challenging to use THT results to precisely decipher the estimates from the earlier analysis of the cross-hole tests using equivalent homogeneous models or vice versa. The difficulty stems from the fact that the estimated properties using equivalent homogeneous models are a weighted average of $K$ anomalies from many parts of the site. The weights can vary spatially and temporally depending on the magnitude of drawdown (see Figures 2 and 3 in Sun et al., 2013).

Additionally, we notice that two abnormal “high-$D$ zones” appear at the model’s upper left and right side boundaries, where the constant head was assigned. These anomalies may indicate the existence of inflow and outflow boundaries (Daranond et al., 2020; Liu et al., 2020; Sun et al., 2013). Flinchum et al. (2018) reported a regional flow map for this site, where flow is inferred from northwest to east, which corroborates this result. Such a result suggests that the observed drawdowns also bear the influence of significant anomalies far from the well field where the correlation or sensitivity is low.

Comparing the $K$ tomogram to the Kriging-interpolated $K$ field from the slug tests (Figure S2), the tomogram reveals more details, such as the preferential pathways between BW7 and BW8 and a “low-$K$ zone”
between BW1 and BW7. Again, THT considers the actual flow connectivity rather than the spatial statistical relationships between boreholes.

The calibrated heads from THT and the observed heads were compared (Figure 6a). A linear model was fitted to the scatter plot. As shown by the fit statistics, simulated heads and observed heads are in good agreement, suggesting that the estimated parameters yielded unbiased head fields. The scattering around the 1:1 line reflects the effects of noise and unresolved heterogeneity in the domain. Figure 6b shows the conditional $L_2$ norm versus the number of iterations during the joint inversion of the four pumping tests. $L_2$ norm decreases with increasing iterations and stabilizes after iteration #21. At this point, variances of the estimated $\ln K$ and $\ln S_s$ fields also stabilized. By jointly using the $L_2$ norm and variances of the estimated parameters as convergence criteria, overexploitation of noisy measurements is avoided.

Residual variances of $\ln K$ and $\ln S_s$ (the likely deviation of actual values from the THT estimates) were computed at each model grid cell. Uncertainty in the estimated $K$ and $S_s$ is lower in the region encompassing the pumped and observation well locations as expected (Figure 7). Compared to the priori variances, THT reduced variances by more than 60% at these boreholes. Estimation uncertainty increases away from the

![Figure 4. Hydrographs of the monitoring boreholes when BW7 (a) and BW6 (b) was pumped: THT sampling points (dots), observed (solid curve), and simulated (dash curve) drawdowns are shown.](image-url)
observation locations and becomes the highest near model boundaries. This result is anticipated, given that no observations were available along these boundaries. In general, fracture connectivity and hydraulic conductivity can only be reliably inferred in the interwell region, where estimation uncertainty is the lowest. Overall, the uncertainty of both parameters in the area enclosed by BW1, BW6, BW7, BW8, and BW9 is reasonable compared to those reported in previous HT studies of fractured rock aquifers (e.g., Dong et al., 2019; Tiedeman & Barrash, 2020).


For the slug tests, the lateral radius of influence was estimated to be $\sim$15 m for BW6 and $\sim$20 m for BW7 using the Bouwer and Rice (1976) method. For the FLUTE tests at the same wells, the lateral radius of influence was estimated at 10–60 m (Keller et al., 2014). Given the uncertainty in radius estimation and the fact that both methods led to similar vertically averaged $K$ values, we believe that these methods likely
interrogated similar lateral volume. However, $K$ profiles were estimated for every 0.3 m of the open-hole interval using FLUTe liner, thus these discrete values may better represent hydraulic conductivity of the fractures. Clearly, compared to slug test, FLUTe test investigated smaller vertical scale: the actual support volume (i.e., interrogated volume) of the individual discrete $K$ are thus smaller if the lateral scale is similar

Figure 6. (a) Simulated versus observed heads after jointly inverting the four sequential pumping tests. (b) Conditional $L_2$ norm of hydraulic head, $\text{Var}(\ln K)$, and $\text{Var}(\ln S_s)$ versus the number of iterations during the simultaneous inversion of the 4 pumping tests.

Figure 7. The residual variance of estimated $\ln K$ (a) and $\ln S_s$ (b) for the simultaneous inversion of four cross-hole pumping tests.
as discussed above. Slug test results indicate that BW1 has the largest $K$, followed by BW7 and BW6, while BW8 and BW9 have the lowest $K$ (Table S3). These values are consistent with THT tomograms.

Compared to the slug-test-derived $K$, pumping-test-derived $K_{\text{eff}}$ (i.e., the Papadopulos and Cooper (1967)'s solution) is about 1–2 orders of magnitude greater at BW8, approximately 5–10 times greater at BW6, and approximately 4–6 times greater at BW7 (Tables S3 and 2). For BW1, BW6, BW7, BW8, and BW9, $K_{\text{g}}$ of the slug-test-derived $K$ is $8.02 \times 10^{-7}$ m/s while $K_{\text{g}}$ of pumping-test-derived $K_{\text{eff}}$ is $8.38 \times 10^{-6}$ m/s (Figure 8), an order of magnitude greater. However, due to no drawdown being observed at BW1 during the BW7 pumping test, the average $K$ between BW7 and BW1 cannot be estimated using the Papadopulos and Cooper (1967) solution, but THT inverted a low-$K$ zone between the two boreholes. Therefore, $K_{\text{g}}$ of pumping-test-derived $K_{\text{eff}}$ for the area enclosed by BW1, BW6, BW7, BW8, and BW9 is smaller than $8.38 \times 10^{-6}$ m/s. Using a statistical two-sample t-test for comparing sample means (Snedecor & Cochran, 1989), the null hypothesis of equal mean was rejected at a significance level of 0.05, thus we conclude that a scale effect of $K$ exists at single- and cross-hole scales at Blair Wallis. Compared to those reported for similar crystalline aquifers (Le Borgne et al., 2006; Martinez-Landa & Carrera, 2005), however, this scale effect is mild which we attribute to the relatively well-developed fracture network at the interwell scale.

5. Discussion

5.1. Sensitivity Analysis for THT

Because BCs of the studied aquifer are unknown, a sensitivity analysis was carried out to evaluate if THT inversion is sensitive to the BC assigned to the inverse model. In Section 4.5, a constant head (2,465 m) was assigned to all model boundaries. Here, we conducted two inversions with two different BC assumptions in the THT analysis. One case considered that the left boundary was no-flux and the other three boundaries a constant head (2,465 m). The other case considered that the left and right boundaries were specified with a constant head (2,465 m) while the top and bottom no-flux. All other inversion parameters were held the
same. Results indicate that there is not much difference among the three inversion outcomes based on different BCs (see Figures 5 and S4–S5). All three inversions also identified a similar $K$ pattern in the vicinity of the boundaries. Thus, inversion in this work appears insensitive to the BC assumption.

Using numerical experiments, Daranond et al. (2020), Liu et al. (2020), and Sun et al. (2013), recommended the constant head boundaries for THT analysis. In particular, Daranond et al. (2020) demonstrated that THT can identify actual no-flux boundaries as zones of low conductivities at the boundaries if a constant head BC was assigned. Likewise, Liu et al. (2020) showed that THT yielded a high-conductivity zone near the boundary if a constant head boundary is used for the actual flux boundary. Interestingly, the THT analysis of this field data confirms their findings, supported by the regional flow map (Flinchum et al., 2018).

In addition to the BC, the correlation scale of $\ln K$ is also uncertain. Two inversion cases were carried out assuming increasing correlation scales ($x = 15\text{m}, y = 30\text{m}; x = 30\text{m}, y = 60\text{m}$) as prior information. In both cases, constant heads ($2465\text{m}$) were specified along model boundaries. Inversion outcomes are presented in Figures S9 and S10. Compared to Figure 5, both “high-$K$” and “low-$K$” zones became more continuous with the increased correlation scales, as expected. When correlation scales grew, a “low-$K$ zone” area in the southwest region of the model and a “high-$K$ zone” in the southeast region grew larger. Clearly, to determine which continuity pattern is the most likely, more constraint on the correlation scales is needed. Such constraints may come from additional drilling or correlation with geophysical measurements. More importantly, additional observation wells are needed, as explained in Yeh, Khaleel et al. (2015), Yeh, Mao, et al. (2015).

### 5.2. Effect of Sample Location and Fracture Connectivity

For single-hole tests, the derived $K$ is likely highly dependent on well location: if we sample at a low-$K$ zone (blue color in Figure 5a; southwest of BW8 and BW9), the single-hole test derived $K$ will be low; if we sample at a high-$K$ zone (e.g., red color in Figure 5a; southwest of BW7), a higher $K$ estimate will be obtained. For the cross-hole pumping test, both the pumping location and monitoring location affected the estimate $K$.

From borehole logging observations, flowing fractures are extensively developed at our well field (Ren et al., 2018, 2019), thus they were likely sampled more frequently by the single-hole tests. Moreover, the obtained flow dimensions using the GRF model are around 2, which implies good connectivity of the fracture network (Maréchal et al., 2004). Although Butler and Healey (1998a, 1998b) attributed scale effect between pumping and slug-test interpreted $K$ to skin effect of the slug-test due to incomplete well development, well-skin for the slug tests was found negligible at our site (Ren et al., 2018). Therefore, the observed scale effect could be explained by changing the connectivity of flowing fractures as the test volume is expanded.

### 5.3. Scale Dependence of Hydraulic Conductivity

Figure 8 summarizes $K$ estimates from FLUTE test, slug test, cross-hole test, and THT analysis. As a group, $K$ obtained from cross-hole tests is larger than that obtained from single-hole tests. The difference in support volume likely explains the variability in the estimated $K$. For example, the isotropic $K_{\text{est}}$ estimated from cross-hole tests varies within one order of magnitude. Slug-test-derived $K$ varies approximately two orders of magnitude. On the other hand, FLUTE-derived $K$, which involves a smaller vertical support (i.e., 0.3 m) than the slug test, varies over three orders of magnitude. The greater variability in the FLUTE-derived $K$ likely reflects various fractures and matrices encountered at the FLUTE measurement scale.

Previous studies also indicate that hydraulic conductivity estimation is affected by the duration of aquifer stimulation (e.g., Wen et al., 2010), related to the interrogated volume of the aquifer. As shown in Figure 8, the estimated $K_{\text{est}}$ using drawdowns of four monitoring boreholes induced by pumping at BW6 is more than five times greater than the similarly estimated $K_{\text{est}}$ for the BW7 pumping test. The testing time for BW7 was 10 times longer and a larger aquifer volume was stressed. When drawdowns of the same pumping duration (i.e., 240 min) were used, the estimated $K_{\text{est}}$ are $8.20 \times 10^{-6} \text{m/s}$ and $5.08 \times 10^{-6} \text{m/s}$ for the BW6 and BW7 tests, respectively, with the difference reduced to less than two. Moreover, due to the dominance of less permeable fractures at larger scale, which is revealed by THT inversion (i.e., one low-$K$ zone in the northeast of BW7 and another low-$K$ zone in the southwest of BW8 and 9; see Figure 5a), the estimated $K_{\text{est}}$ is observed to
decrease with pumping time for both tests. Overall, K is found to increase from single-hole to the interwell scale and then decrease at larger scale with a nonmonotonic, albeit mild, scale effect.

Pumping-test-derived $K_{\text{eff}}$ may or may not represent a field-wide average value depending on whether the stressed volume reaches the representative elementary volume (REV) for the fractured medium (Bear, 1972). For the BW6 and BW7 pumping tests, $K_{\text{eff}}$ tensors and their principal directions (estimated using drawdowns at four monitoring boreholes) were computed for increasing testing time to examine the existence of a field-scale REV (Figure 9). Figures 9a and 9c show the estimated $K_{\text{eff}}$ tensor components as a function of pumping time at BW6 and Figures 9b and 9d for BW7. Both $K_{xy}$ and $K_{yz}$ from the BW6 pumping test appear to converge to stable values with increasing test time, while the maximum principal horizontal axis (i.e., $\theta$) converges to approximately 40°. For the BW7 pumping test, $K_{xy}$ and $K_{yz}$ approach stable values much later (after 500 min) than those observed for the BW6 pumping test. The maximum principal axis converges to approximately 60°. The estimated $K_{\text{eff}}$ tensors from the two tests vary significantly, suggesting statistical heterogeneity, and therefore, it is questionable that a single field-scale REV can be defined. Neuman (1987) also pointed out that it is impossible to define an REV for a typical fractured rock site.

An interesting observation is that for both pumping tests, $K_{\text{eff}}$ principal components appear to converge for a time before switching to a different value. This temporary convergence may be due to the emergence of a local statistical equivalent homogeneity before the cone of depression encountered new (flowing or non-flowing) fractures with different conductivity characteristics.

6. Conclusions

The general findings are summarized below:

1. Borehole televiewer and outcrop data suggest that subsurface fractures are preferentially oriented, which is confirmed by cross-hole hydraulic tests. Assuming a homogeneous conceptual model, $K_{\text{eff}}$ estimated using cross-hole drawdowns at different orientations from pumping well can reveal this anisotropy. Both
the Papadopulos (1965) and nonlinear least squares minimization methods produced a similar maximum principal $K_{\text{eff}}$ direction consistent with the dominant outcrop and borehole fracture strike. To estimate $K_{\text{eff}}$, the nonlinear minimization approach is recommended, as the Papadopulos (1965) method can lead to physically incorrect values when drawdowns from 3 monitoring wells are used.

2. Several high hydraulic diffusivity zones were identified by THT, which may indicate preferential flows in connected fractures among the boreholes. Inversion result is insensitive to the assumed model boundary conditions, but is sensitive to the assumed lnK correlation scale. As revealed by inversion, the central well cluster (BW 1, 6, 7, 8, and 9) is bordered by several low-K zones. This may explain the observed decrease of the analytically estimated $K_{\text{eff}}$ with pumping time, as pressure in the central cluster propagates into this region with less permeable fractures.

3. As a group, $K$ obtained from cross-hole tests is greater than that obtained from single-hole tests. However, because cross-hole tests stressed the aquifer at both interwell and larger scales, $K_{\text{eff}}$ obtained from interpreting cross-hole data is observed to decrease with pumping time, likely due to the dominance of less permeable fractures at larger scale. This lateral reduction of mean $K$ is also revealed by THT as low-K zones surrounding the test boreholes.

At the fractured rock aquifer site in Laramie Range, Wyoming, our field testing program found that $K$ increases from single-hole to the interwell scale and then decreases at larger scale, exhibiting a nonmonotonic, albeit mild, scale effect.

**Data Availability Statement**

Data reported in this work can be downloaded from [http://doi.org/10.5281/zenodo.5031319](http://doi.org/10.5281/zenodo.5031319).

**Acknowledgments**

Field component of the research described in this article was supported by Wyoming Center for Environmental Hydrology and Geophysics under NSF EPSCoR grant (EPS 1208909). The first author acknowledges the support by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (No. G1323520265), and the Open Foundation of Key Laboratory of Tectonics and Petroleum Resources (China University of Geosciences, Wuhan), Ministry of Education (No. TPR-2019-17). The authors thank the associate editor Matthew Becker and three anonymous reviewers for their critical and constructive comments, which helped to improve the organization and clarity of this article.

**References**


