1 ***In-situ* falling-head test for hydraulic conductivity: evaluation in layered sediments of an**
2 **analysis derived for homogenous sediments**

3 Matthew C. Burnette^{a,b}, David P. Genereux^a, François Birgand^c

4 ^aDepartment of Marine, Earth, and Atmospheric Sciences, Campus Box 8208, North Carolina
5 State University, Raleigh, NC, USA, 27695, matthew.burnette@tetrattech.com;
6 genereux@ncsu.edu

7 ^bnow at Tetra Tech, Inc., 2110 Powers Ferry Rd. SE, Suite 202, Atlanta, GA, USA 30339

8 ^cDepartment of Biological and Agricultural Engineering, Campus Box 7625, North Carolina
9 State University, Raleigh, NC, USA, 27695, birgand@ncsu.edu

Abstract

The hydraulic conductivity (K) of streambeds is a critical variable controlling interaction of groundwater and surface water. The Hvorslev analysis for estimating K from falling-head test data has been widely used since the 1950s, but its performance in layered sandy sediments common in streams and lakes has not previously been examined. Our numerical simulations and laboratory experiments show that the Hvorslev analysis yields accurate K values in both homogenous sediment (for which the analysis was originally derived) and layered deposits with low-K sand over high-K sand. K from the Hvorslev analysis deviated significantly from true K only when two conditions were present together: (1) high-K sand was present over low-K sand, and (2) the bottom of the permeameter in which K was measured was at or very near the interface between high-K and low-K. When this combination of conditions exists, simulation and laboratory sand tank results show that *in-situ* Hvorslev K underestimates the true K of the sediment within a permeameter, because the falling-head test is affected by low-K sediment outside of (below the bottom of) the permeameter. In simulation results, the maximum underestimation (occurring when the bottom of the permeameter was at the interface of high K over low K) was by a factor of 0.91, 0.59, and 0.12 when the high-K to low-K ratio was 2, 10, and 100, respectively. In laboratory sand tank experiments, the underestimation was by a factor of about 0.83 when the high-K to low-K ratio was 2.3. Also, this underestimation of K by the Hvorslev analysis was about the same whether the underlying low-K layer was 2 cm or 174 cm thick (1% or 87% of the domain thickness). Numerical model simulations were useful in the interpretation of *in-situ* field K profiles at streambed sites with layering; specifically, scaling the model results to the maximum measured K at the top of the field K profiles helped constrain the likely ratio of high K to low K at field locations with layered heterogeneity. Vertical K values

are important in field studies of groundwater – surface water interaction, and the Hvorslev analysis can be a useful tool, even in layered media, when applied carefully.

Keywords

Hydraulic conductivity; permeameter; sediments; streambed; numerical modeling; groundwater – surface water interaction

1. Introduction

The hydraulic conductivity (K) of streambeds is a critical variable controlling interaction of groundwater and surface water. The magnitude and spatial patterns of water and chemical fluxes through streambeds are related to the magnitude and spatial pattern of streambed K [Gilmore et al. 2016a; Kennedy et al. 2009; Min et al. 2012; Sebok et al. 2014]. In addition to spatial variation, streambed K may also vary in time in association with time-varying controls such as erosion, deposition, or temporal variation in biogenic gases in streambed sediments [e.g., Genereux et al. 2008; Cuthbert et al. 2010; Kurtz et al. 2012]. A variety of *in-situ* approaches have been used to estimate streambed K, including falling head tests in field permeameters [Chen 2000; Landon et al. 2001; Genereux et al. 2008], constant head injection tests, or “CHIT” [Cardenas and Zlotnik 2003], slug tests [Cey et al. 1998; Springer et al. 1999; Sebok 2014], and regressions based on grain size [Alayamani and Sen 1993]. Landon et al. [2001] offer a comparison of several methods. Lu et al. [2012] adapted a falling head permeameter approach to estimate anisotropy in streambed K. Ong and Zlotnik [2011] incorporated an air-water manometer for accurate measurement of head differences into their CHIT device for estimation

of lakebed K . Other novel approaches include the small-scale pump tests and “piezo-seep meter” of Kelly and Murdoch [2003].

Often measurement of streambed K is motivated by the goal of determining water flux (v) through the streambed using K and associated measurements of vertical hydraulic head gradient in the streambed, J (i.e., $v = KJ$), and/or determining dissolved solute flux through the streambed as the product of v and solute concentration in the groundwater (for a gaining stream) or stream water (for a losing stream). The ideal K measurement for this flux application would be a value of vertical K measured over the same streambed depth interval as J , to give a consistent and meaningful pairing of K and J for estimation of v from Darcy’s equation. In this regard, the field permeameter approach has an advantage over methods that yield measures of horizontal K , given that sediments are commonly layered and thus anisotropic to some degree (e.g., Lu et al. [2012]).

Field permeameter tests involve inserting a thin-walled pipe (the permeameter) into a streambed, filling the pipe with water, and observing the rate at which the water level in the pipe falls. An analysis originally proposed by Hvorslev [1951; solution E, page 44] and widely used since [e.g., Cedergren 1989; McCarthy 1998; Chen 2000, 2004; Landon et al. 2001; Song et al. 2007; Lu et al. 2012] is most useful for data analysis when stated in the following form [Genereux et al. 2008]:

$$\ln H = -\frac{K}{\left(\frac{\pi D}{11M} + L\right)}t + \ln H_0 \quad (1)$$

where D is the permeameter diameter, L is the depth of penetration of the permeameter into the streambed (i.e., the length of the sediment column inside the bottom of the permeameter), M is

the square root of the ratio of horizontal to vertical K , H is the water level inside the permeameter relative to the ambient water level (the subscript 0 refers to the start of the test), and t is time. In his original work, Hvorslev assumed the porous medium is homogeneous, the water and sediment are not compressible, friction losses due to the permeameter itself are negligible, and that Darcy's equation holds true. Horizontal K does appear in Eqn. 1 (within M) but the analysis is very insensitive to horizontal K at high ratios of L to D [Chen, 2004]. For example, assuming $M = \infty$ (extreme anisotropy) when M is actually 1 (isotropic sediment), or vice versa, results in less than 5% error in K when L/D is about 5 [Genereux et al. 2008]. Thus, with adequate L/D (5 or more), the falling head test and Hvorslev analysis give the desired estimate of vertical K . In sandy streambeds the test is easy to perform and often many can be done in a single day.

A potential complication is that the Hvorslev analysis was developed for homogeneous porous media, but sediments are generally layered. In their analysis of the uncertainty in streambed K values derived from equation 1, Genereux et al. [2008] referred to the uncertainty arising from layering (i.e., the “model error” associated with using equation 1 in layered sediments) as the most difficult source of uncertainty to quantify in the Hvorslev analysis of falling-head field permeameter tests. To date, this source of uncertainty and the performance of equation 1 in layered sediments have not been directly assessed, despite widespread application of equation 1 since the 1950s.

The work reported here addresses two principal questions: First, how applicable is the Hvorslev analysis (equation 1) when layered heterogeneity exists? Second, is there any practical guidance

to help with interpretation of field K values from equation 1 in the presence of layered heterogeneity? We began with numerical simulations of falling-head tests to evaluate how well equation 1 would estimate the K of layered sediments. Then, results from falling-head tests in a laboratory sand tank where the location of a K boundary was designed and known were compared to the numerical simulation results. Finally, field measurements were done in two streams in eastern North Carolina to determine the degree of vertical variability in K due to layering and to assess whether results from numerical simulations could be used to help interpret field estimates of K from equation 1 in the presence of layered heterogeneity.

2. Methods

2.1. Numerical Modeling

We used MODFLOW 2005 with its associated LAKE package [Merritt and Konikow 2000] to model falling-head field permeameter tests in streambeds of known K (both homogeneous and layered) under different conditions, to predict the water level (i.e., the head) in the permeameter as a function of time. Head results from each simulated falling head test were then analyzed with Eqn. 1, and the K value obtained was compared to the true K that we had assigned to the sediments in the model. Close agreement would indicate good performance of equation 1 in estimating K from falling head results. Equation 1 predicts a linear relationship between $\ln H$ and t , and K was calculated from the best-fit slope of this relationship, for model, laboratory sand tank, and field results:

$$K = -(slope) \left(\frac{\pi D}{11M} + L \right) \quad (2)$$

Together, MODFLOW and the LAKE package represent a fully-coupled groundwater – surface water model capable of conserving water mass during transient simulations. In our work, the

permeameter was modeled as a “lake” in a rectangular block of sandy sediment (Figure 1). The initial permeameter head was assigned to be 60 cm, 30 cm higher than the steady head in the sediment. Thus, the permeameter head fell from 60 toward 30 cm during each simulation. The 30 cm initial head difference is similar to that used in field permeameter work in streambeds; simulations with initial head differences of 1 cm and 60 cm produced results essentially identical to those with an initial head difference of 30 cm [Burnette 2013]. The USGS graphical user interface ModelMuse [Winston 2009] was used to create model input files. GW-Chart was used as a post-processor to analyze model output [Winston 2000].

The rectangular groundwater domain was a cube of volume 8 m^3 . Overall water budget discrepancies for simulations were about 0.2% for the 8 m^3 domain and 0.1% for a domain 4x larger (2 m thick, 4 m x 4 m horizontally), both negligible and well below the 1% limit recommended by others [e.g., Elkrail and Ibrahim 2008]. With 260 rows, 260 columns, and 101 layers, the 8 m^3 domain was quite large computationally (almost 7 million cells) and large enough to give confidence that the effects modeled were not sensitive to resolution (size of individual cells) or overall size of the model domain; identical results were obtained in comparison simulations between this grid and other grids that were either finer (smaller cells), or larger in overall dimensions, including one of almost 28 million cells [Burnette 2013].

A variable grid spacing was used to provide high resolution in areas of relatively high gradient near the simulated permeameter while also affording some computational efficiency in areas of lower gradient toward the outer edges of the domain (Figure 1). Grid cells were 2 cm thick and had lateral dimensions of 0.25 to 1.0 cm. Boundary conditions were an impermeable base and

constant head (30 cm) on the 4 vertical sides and the top of the domain, representing a block of streambed overlain by surface water that is 30 cm deep.

From this point forward in the paper, we refer to the horizontal high-K and low-K zones in the model grid as “layers”, and each of these layers was made up of multiple two-cm-thick “model grid layers” in the numerical grid. Two-layer simulations were done with high-K (20 m/d) over low-K (0.2, 2, or 10 m/d), or low-K over high-K, giving a K ratio of 2, 10, or 100 between the high-K and low-K layers. These K values span a range found in streambeds that are mainly sandy but have finer zones of lower K [e.g., Genereux et al. 2008].

The upper layer in two-layer simulations, whether of high or low K, was 26 cm thick (composed of 13 two-cm model grid layers) and the lower layer was 174 cm thick (87 two-cm model grid layers). Additionally, three-layer simulations were run with a single two-cm layer of low-K material (depth interval of 26-28 cm) within an otherwise high-K domain. In all layered simulations, the “true K” (for comparison with the Hvorslev K from equation 2) was defined as the harmonic mean K for the model grid layers penetrated by the simulated permeameter.

The walls of the permeameter were modeled by creating a ring, 7 cm in diameter, of impermeable cells ($K = 0$), running from the top layer down to the desired depth of penetration of the permeameter (20-50 cm) in the center of the domain [Burnette 2013].

The Geometric MultiGrid (GMG) solver was used and provided reliable convergence. All model simulations were run for 100 time steps, with each time step one second in length. This allowed

head in the permeameter to drop at least 1 cm and provided a significant number of points to calculate the slope of the $\ln H$ vs. t regression (equation 1).

In most simulations there was no flow through the sand except that caused by the water inflow from the falling head permeameter, though a few simulations were run with ambient upward groundwater flow from the base of the model domain. A few simulations were also run with a rising instead of falling permeameter head (initial head in the permeameter was set to 20 cm instead of 60 cm). These simulations showed that ambient upward groundwater flow in the domain and the direction of the permeameter response (falling or rising) did not affect the magnitude of the K value estimated by equation 2 [Burnette 2013].

In all of our work with equation 2 (model, lab, and field results), each best estimate of K was based on averaging the upper limit of K (from assuming $M = 1$) and lower limit of K (from assuming $M = \infty$) for a given falling-head simulation or test. This was true even when working with model-simulated falling head results, when M was prescribed, in order to represent what the outcomes would be in a realistic field case, in the absence of any information on M .

2.2. Laboratory Sand Tank

Laboratory investigations in a sand tank allowed falling-head tests to be performed in a controlled environment. The laboratory tank (Figure 2) had dimensions of 79.5 cm x 70.5 cm in the horizontal and 61 cm in the vertical, and was constructed of clear 3/8 inch (9.5 mm) Plexiglas sheets. The tank was constructed so that the sand rested on a perforated plate at the bottom that overlaid an open water-filled space with no sand; this bottom reservoir connected the main sand-

filled reservoir of the tank to a smaller side reservoir with no sand. The perforated plate contained 1/4 inch (6.4 mm) diameter holes spaced about 3 cm apart, and was covered by a piece of wire mesh to prevent sand from falling through the holes. To reduce the chances of preferential flow along the walls of the tank, the inside of each of the four walls of the main reservoir was lined with 13 horizontal Plexiglas strips (1 cm x 1 cm) spaced about 2 cm apart.

Experiments were performed with three different types of sand packs in the tank: (1) homogenous high-K sand, (2) low-K sand over high-K sand, and (3) high-K sand over low-K sand. The high-K sand was “multipurpose sand” purchased from The Home Depot in Raleigh, North Carolina (NC), and its grain size distribution closely resembled that of sediment found in West Bear Creek [Leahy 2007], one of the two streams in which streambed cores were collected. The low-K sand was donated by the Unimin silica mine located in Marston, NC. This sand, used in the making of glass products, was finer grained than the high-K sand and contained no particles larger than 250 μm , with 90% of the distribution ranging from 80 to 100 μm . The high-K sand contained particles larger than 300 μm (6% by weight) with 60% of the distribution ranging from 100 to 300 μm .

The tank was filled with sand by gradually adding the sand to water to simulate settling in a field environment and to prevent sand “bridging” (the formation of small gaps with no sand). In all experiments the tank was filled with 55 cm of sand, with 3 cm of open water above the sand surface to simulate stream water above a streambed.

The first experiments were done with homogenous high-K sand. First, a constant-head test to estimate K was performed on the whole tank [Burnette 2013], using a set-up in which a constant-head difference of about 10 cm was maintained between the main sand-filled reservoir and the smaller side reservoir of the sand tank (Figure 2). K for the whole tank was then estimated as:

$$K = \frac{QL}{Adh} \quad (3)$$

where Q is the volumetric flow rate through the sand, L is the depth of the sand (55 cm), A is the horizontal cross-sectional area of the sand (5641 cm²), and dh is the head difference between the main and side reservoirs. Each time the tank was packed with sand, the constant-head whole tank K test was repeated 10 times to yield an average value.

Following the whole tank constant head tests, falling-head tests were performed inside four permeameters built of thin-walled polycarbonate pipe 7 cm in inner diameter [Leahy 2007; Genereux et al. 2008] that had been inserted to depths of 20, 24, 26, and 30 cm in the sand in the tank (these tests conducted within the sand tank are referred to here as “*in-situ*” falling-head tests). Equation 1 (the Hvorslev analysis) was used to estimate K from the *in-situ* falling-head test data. Each permeameter was inserted vertically into the sand by direct push.

All four permeameters were inserted in the tank, and then the first test was run on the shallowest (20 cm). The head inside this permeameter was then allowed to equilibrate with the standing water outside the permeameter, and the next *in-situ* falling-head test was run in the next deepest permeameter (24 cm). This process was continued until *in-situ* falling-head tests had been completed in all four permeameters.

Then the permeameters, with intact sand cores, were dug out of the sand tank and set up as laboratory permeameters for estimation of K by falling-head tests [Burnette 2013]. To keep the sand and water intact and undisturbed inside each permeameter during removal from the sand tank and set up in the lab, three-inch diameter plumbing test plugs were inserted into the top of each permeameter and tightened, and a piece of sheet metal with rubber glued to one side (for a watertight seal to the permeameter) was slid underneath the bottom of each permeameter.

The outlet of each permeameter was covered with a piece of fine mesh cloth and supported with a wire screen. The traditional analysis for a falling head lab permeameter [e.g., Domenico and Schwartz 1997, p.44; Fetter 2001, p.92] was used to estimate K . Solving for K in terms of the slope of the regression of $\ln H$ vs. t yields: $K = -(\text{slope})L$.

The falling-head lab permeameter tests on each core from the sand tank provided estimates of K that were independent of the Hvorslev analysis, equation 1. Because the head at the outlet of the permeameter is known and the test is not affected by sand outside the permeameter, the lab permeameter falling-head test probably offers the closest estimate of “true K ”.

For sand cores with a penetration depth greater than 20 cm, one or more sections at the bottom of the core were extruded and discarded so that K measurements could be made on multiple cores covering identical depth intervals, to assess lateral homogeneity in the sand tank. For example, after the falling-head test on the 24 cm core, the bottom 4 cm were extruded and discarded, and a falling-head test was run on the remaining 20 cm. Similarly, for the 26 cm core, tests were run on the full 26 cm, the upper 24 cm, and the upper 20 cm; for the 30 cm core, tests were run for

the full 30 cm, and then upper 26, 24, and 20 cm. Thus, with 20, 24, 26, and 30 cm cores collected from the sand tank, we ultimately ran falling-head lab permeameter tests on one 30 cm core, two 26 cm cores, three 24 cm cores, and four 20 cm cores.

After the measurements described above were performed on the homogenous tank (whole tank constant-head K test, *in-situ* falling-head K tests, and falling-head lab permeameter K tests), the tank was refilled with a second homogenous sand pack, and the entire set of measurements was repeated (the two sets of measurements on the two homogeneous sand packs are referred to as “run 1” and “run 2” in the Results section).

Work with layered sand packs was then begun. The upper 26 cm of high-K sand was removed from the tank and replaced with 26 cm of low-K sand (the bottom 29 cm of high-K sand remained in place). The same suite of measurements carried out in the homogeneous tank was then carried out on this layered sand tank. As with the homogeneous tank, the full suite of measurements was carried out on two different low-K over high-K sand packs (runs 1 and 2). Finally, the same steps were carried out on a layered sand pack with high-K sand in the top 26 cm of the tank over low-K sand in the bottom 29 cm, again with two complete runs 1 and 2.

For each of the three scenarios (homogeneous, low K over high K, and high K over low K), we compared the K values from the whole-tank measurement, the Hvorslev analysis (equation 1) of *in-situ* falling-head tests, and the falling-head lab permeameter tests, and also compared sand tank and numerical simulation results.

2.3. Field Measurements

2.3.1. Study Site

The field portion of this study was conducted in Bear Creek and West Bear Creek in the North Carolina coastal plain (Figure 3) in July and August of 2013. West Bear Creek has been described in previous publications concerning groundwater interaction with surface water [Leahy 2007; Genereux et al. 2008; Kennedy et al. 2009a,b, 2010; Gilmore et al. 2016a,b]. The study reaches were both located in Wayne County, NC, approximately 6 km southeast of Goldsboro. The study reach in West Bear Creek was just upstream of the State Road 1719 bridge that crosses West Bear Creek; the Bear Creek study area was located just downstream of the confluence with a small tributary (Seneca Branch).

The West Bear Creek study reach had an average width of 7 m and the water depth ranged from about 24 to 44 cm during measurements and core collection. The Bear Creek study reach had an average width of 3 m and water depth of about 27 to 42 cm. For agricultural drainage purposes, both channels were dredged and straightened sometime in the 1950s or 1960s [Leahy 2007]. As a result, the channel banks were steep and extended 2-3 m above the streambed. The West Bear Creek study reach was bordered on both sides by a wooded riparian buffer that ranged from 5 to 10 m wide on the north bank and 20 to 25 m wide on the south bank. On the north side of the stream, the wooded riparian area included a grassy maintenance road (built up with the dredge spoils) that was approximately 3 m wide. The streambed was comprised mostly of sand [Genereux et al. 2008]. Finer silty and clayey material was found along the channel banks, probably as a result of bank failure. The mean K of the upper 36 cm of streambed in the study reach was 15.5 m/d and ranged from 0.01 to 66.2 m/d [Genereux et al. 2008]. The Bear Creek

study reach was bordered on both sides by a wooded riparian buffer that ranged from 2 to 5 m wide on both banks. On the east bank, the riparian area included a grassy maintenance road (also probably built from dredge spoils). The Bear Creek streambed was comprised mostly of sand, but was finer than that found in West Bear Creek.

2.3.2. Field Methods

In-situ falling-head tests were performed and cores collected at 8 streambed locations in West Bear Creek and 4 in Bear Creek. Falling-head tests were performed at 5 depths in the streambed at each location (20, 24, 28, 32, and 36 cm). In West Bear Creek, the name of each measurement location was the distance upstream of the State Road 1719 bridge followed by the letter L, C, or R to indicate left side, center, or right side of the channel (left was the north side). Similarly, in Bear Creek, the location name was the distance downstream of the confluence of Bear Creek and Seneca Branch followed by L, C, or R (left was the east side).

At each measurement location, we first measured vertical head gradient (J) in the streambed over each depth interval in which K was to be measured, to be used later in estimating the ambient pre-test head inside the permeameter as the sum of head outside the permeameter (in the stream) plus the product of J times penetration depth; the ambient head inside the permeameter is needed to compute H values for equation 1. J was measured with a PVC piezomanometer as described by Kennedy et al. [2007].

After the J profile measurements had been completed at each location, falling-head K tests were performed at the same depths. The permeameter was inserted to a depth of 20 cm to the right of

the piezomanometer. Stream water head outside the permeameter was read in a vertical section of 1/4 inch ID clear tubing affixed to the outside of the permeameter (the tubing acted as a stilling well). Next, compaction was measured inside the permeameter as was done in the laboratory sand tank tests. A falling-head test was then performed inside the permeameter. Following the test at 20 cm, the permeameter was then advanced to depths of 24, 28, 32, and 36 cm, with a falling-head test run at each depth.

Following completion of the K measurement at a depth of 36 cm, the permeameter (with intact streambed sediment core inside) was dug out of the streambed for transport back to the lab. Before extraction from the streambed, the permeameter was filled with water and a 3 inch plumbing test plug was inserted in the top of the permeameter to provide an air lock to prevent water drainage from the core. A metal plate was placed beneath the bottom of the permeameter, and the bottom was wrapped with a rubber sheet. The cores were held vertical and remained fully saturated, with standing water above the top of each core, during transport to the lab [Burnette 2013].

Once back in the lab, falling-head lab permeameter tests were run on the sediment cores in the same way described above for cores from the laboratory sand tank. The first measurement in the lab permeameter was for the full 36 cm long core collected in the field, followed by sequential extrusion of the bottom of each core and additional falling-head tests (the target was to run tests for the top 32, 28, 24, and 20 cm of each 36 cm field core).

3. Results

3.1. Numerical Modeling

3.1.1. Homogenous Simulations

The first set of MODFLOW simulations were run with a homogenous porous medium with $K = 20$ m/d (Figure 4), a value near the mean K of the streambed in West Bear Creek [Genereux et al. 2008]. Three simulations were also run with a homogenous porous medium having $K = 2$ m/d. Simulations were run with the permeameter penetrating 20-50 cm, a range of depths typical for falling-head permeameter tests in the field. On all plots of K vs. depth, the plotted points represent the depth of the bottom of the permeameter, though the K value applies to the whole column of sediment above the plotted point.

For both series of homogenous simulations ($K = 20$ m/d and $K = 2$ m/d), K estimated with the Hvorslev analysis (equation 1) was about 2.6% lower than the true model-assigned K at a penetration depth $L = 20$ cm and 1.2% lower at $L = 50$ cm. Thus, K estimates from equation 1 deviated only slightly from true K (Figure 4), suggesting that equation 1 gives a good representation of the physics of the *in-situ* falling-head test (provided those physics are well represented by the numerical model, as expected).

3.1.2. Heterogeneous (Layered) Simulations

Falling-head test simulations in layered domains were run with the same permeameter penetration values as in the homogenous simulations (20-50 cm), with the layer interface at a depth of 26 cm.

Simulations with low-K sand over high-K sand showed excellent agreement between true K and the Hvorslev K estimated from equation 1 (Figure 5). For simulated permeameter penetrations <26 cm (i.e., above the layer interface), the Hvorslev analysis underestimated true K by less than 2%. For simulated permeameter penetrations at and below the layer interface, the Hvorslev analysis overestimated true K by less than 5%. The maximum overestimation occurred at a simulated permeameter penetration of 26 cm, when the bottom of the permeameter was located exactly at the layer interface. At this depth, the true K is simply equal to that of the low-K sand (the permeameter penetrates only low-K sand), but the escape of water from the permeameter during the falling-head test is influenced in part by the high-K material at the bottom of the permeameter (i.e., the falling-head test has a small but non-zero sensitivity to the presence of the high-K sand immediately below the permeameter). Model results suggest this sensitivity would introduce at most a minor error (<5%) in the field estimation of K with equation 1, even when the lower sand has K 100x greater than that of the upper sand, and only if the bottom of the permeameter is at the layer interface (Figure 5).

The potential error from equation 1 is more significant in media with high-K sand over low-K sand, and depends on the K contrast between the layers. For these simulations, the Hvorslev K estimated from the model output is lower than the true K, because of the sensitivity of the falling-head test to the K of material outside of and below the permeameter (K of this material is lower than the true K in the depth interval penetrated by the permeameter, if the permeameter extends to or below the layer interface in a medium of high K over low K). The degree of underestimation was small when the bottom of the permeameter was a few cm or more above or below the layer interface, became larger as the bottom of the permeameter came closer to the

layer interface, and was largest when the bottom of the permeameter was exactly at the layer interface (Figure 6). The amount of underestimation increased as the ratio of high-K to low-K increased; with the bottom of the permeameter at the layer interface, Hvorslev K underestimated true K by 9.3% when the K ratio was 2, 41.9% when the ratio was 10, and 88.4% when the ratio was 100 (Figure 6).

Three-layer simulations ($K = 0.2$ m/d at a depth of 26-28 cm, $K = 20$ m/d above and below that) had good agreement between Hvorslev K and true K except when the permeameter bottom was at the first layer interface (26 cm depth; Figure 7). Here, Hvorslev K was 84.7% lower than true K, very similar to the 88.4% mentioned above. There was not a significant difference between Hvorslev K and true K when the permeameter bottom was at the second layer interface (28 cm depth) where low K overlies high K, in accord with low K over high K results in Fig. 5. In other words, as the bottom of the permeameter approached an interface (either high K over low K, or low K over high K), the accuracy of the Hvorslev falling-head analysis was the same whether the layer below the interface was thin (2 cm) or thick (effectively infinite, in extending to the bottom of the groundwater domain). Finally, with regard to the thin low-K layer, the Hvorslev estimated K value slightly underestimated true K above the top of the layer ($\leq 6.3\%$) and slightly overestimated true K below the top of the layer ($\leq 5.7\%$; Figure 7).

These results show that Hvorslev K (equation 1) accurately predicted true K for a homogenous porous medium and for a two-layer porous medium with low-K over high-K (maximum error $< 5\%$). Hvorslev K was significantly less than true K when high-K material overlaid low-K material, whether the underlying low-K layer was 2 cm or 174 cm thick (1% or 87% of the

domain thickness), but only when the bottom of the permeameter was at or very near (within 2 cm of) the layer interface.

3.2. Laboratory Sand Tank

3.2.1. Homogenous Experiments

The first work in the laboratory sand tank was performed on homogenous sand packs. Prior to inserting the permeameters for the falling-head permeameter tests, whole tank constant-head tests were performed. The mean K and standard deviation from the whole tank constant-head tests (10 replicates) were 95.5 m/d and 2.7 m/d, respectively, for run 1, and 87.9 m/d and 2.7 m/d, respectively, for run 2.

The *in-situ* and corresponding lab permeameter K results for each full core (Table 1) differed from each other by about 11% on average, and differed from the whole tank K estimates by, on average, about 10% (*in-situ* K) and 13% (lab permeameter K). These differences and those between runs 1 and 2 (two different packs of the tank using the same sand) were most likely due to random error and small variations in sand packing. These findings give a frame of reference for the inherent variability in K values from the three different K measurement methods, in homogeneous laboratory sand packs. It is also possible that the differences between the two permeameter methods and the whole tank estimate could be due in part to “core shortening” within the permeameters during insertion (minor compaction, or exclusion of material, as the permeameter was advanced into the sand pack in the tank, described in detail by Burnette [2013]).

3.2.2. Low-K over High-K Experiments

The first set of layered tests in the laboratory sand tank was performed on a low-K over high-K sand pack. The mean K and standard deviation from the whole tank constant-head tests (10 replicates) were 45.7 m/d and 1.3 m/d, respectively, for run 1, and 42.5 m/d and 1.0 m/d, respectively, for run 2.

The *in-situ* and corresponding lab permeameter K results for each full core (Table 2) differed from each other by about 9% on average, similar to the average difference between these methods in the homogenous sand pack (11%). We calculated the expected harmonic mean K for the top 30 cm of this sand tank, assuming that the top 26 cm of low-K sand had $K = 36.1$ m/d (average of the lab permeameter K values for all the 20, 24, and 26 cm cores containing only low-K sand), and that the 4 cm of high-K sand from 26 to 30 cm deep had $K = 83.3$ m/d (average of all the lab permeameter K values for cores from the homogenous tank with high-K sand). The calculated harmonic mean K of the top 30 cm was 39.1 m/d. The *in-situ* Hvorslev K for the 30 cm core was 2.1% lower than the harmonic mean K for run 1 and 4.3% lower for run 2. Thus, the Hvorslev analysis yielded a reasonable, reliable estimate of true K in a layered system with low-K over high-K, whether or not the permeameter penetrated both sand layers.

3.2.3. High-K over Low-K Experiments

The second set of layered tests in the sand tank was performed on a sand pack with high-K over low-K. The mean K and standard deviation from the whole tank constant-head tests (10 replicates) were 44.6 m/d and 1.4 m/d, respectively, for run 1, and 45.2 m/d and 1.5 m/d, respectively, for run 2.

466

467 Here, the *in-situ* and lab permeameter K estimates agreed well with one another at depths of 20,
468 24, and 30 cm (-3.7 to 11%; Table 3). However, the *in-situ* K estimate at 26 cm (with the bottom
469 of the permeameter at the layer interface) was 16.5 to 23.5% lower than the corresponding lab
470 permeameter K estimate. At 26 cm, the *in-situ* K was affected by the presence of low-K sand
471 just below the bottom of the permeameter, but the lab permeameter K was not (there was no sand
472 outside the bottom of the lab permeameter).

473

474 The *in-situ* falling-head tests for the laboratory sand tank with high-K over low-K were modeled
475 with MODFLOW using the average K values mentioned above for the high-K and low-K sands.
476 In all three cases (run 1, run 2, and the model), Hvorslev-estimated *in-situ* K was farthest from
477 the true or best estimate of K when the bottom of the permeameter was at the interface between
478 high and low K (Figure 8).

479

480 Thus, the numerical model and sand tank experimental results are in agreement that when high-K
481 material overlies low-K material, the *in-situ* K estimate from equation 1 will be too low if the
482 bottom of the permeameter is at the interface between the two materials. This suggests that if a
483 permeameter were inserted in the field such that the bottom was at or near an interface of high K
484 over low K, the *in-situ* estimate of K with equation 1 could be significantly lower than the K of
485 the sediment actually contained within the permeameter. It seems possible that this effect may at
486 least partly explain the findings by Lu et al. [2012, their Figure 3] of numerous instances of *in-*
487 *situ* K lower than lab permeameter K (they referred to the latter as “on-site K”) for the same
488 cores.

3.3. Field Results

Of the twelve measurement locations in West Bear Creek and Bear Creek, eight locations showed no evidence of layering (in the form of significant variation in K with depth) and three locations did show strong evidence of layering (Figure 9). One site, WBC479R, was ignored because of complications caused by a piece of woody debris in the sediment that became stuck in the permeameter during coring [Burnette 2013].

At the three streambed measurement locations that did show strong evidence of layering (WBC356.25R, WBC400C, and WBC504L), K decreased by a factor of 4.1, 15.8, and 1.9, respectively, between 20 cm and 24 cm (Figure 9). For these three locations, we considered how the results from the MODFLOW simulations could be used to help interpret the *in-situ* streambed K profiles determined from equation 1.

We found (Figure 10) that the *in-situ* profile at WBC356.25R was very similar to the two-layer simulation results in the middle graph of Figure 6, when the simulation results were (1) multiplied by a factor of $21.6/20$ so the maximum simulated K matched the maximum field K (21.6 m/d), and (2) shifted upward 4 cm so that interface between high and low K in the simulation results shifted to 22 cm, the approximate depth of the high-low K interface in the field (Figure 9 shows this interface must have been between 20 and 24 cm in the field). Thus, comparing the *in-situ* K profile to the numerical simulations suggests that the field data at WBC356.25R are reasonably consistent with a two-layer streambed having K about 10x higher in the upper layer, above 22 cm in depth. The same approach showed the *in-situ* K profile at

WBC400C was consistent with a two-layer streambed with K about 100x greater in the upper layer (Figure 10), based on comparison to the simulation results (scaled by a factor of 3.6/20) in which K in the upper layer was 100x greater than K in the lower layer (lower graph in Figure 6).

The *in-situ* field K profile at WBC504L was bounded by the two-layer simulation results in which K of the upper layer was between 2x and 10x greater than K of the lower layer (the upper two graphs of Figure 9, scaled by a factor of 45.7/20), though the uniform field K profile below 24 cm (Figure 10) contrasts with the gradual decrease in K with increasing depth seen in the simulation results. The actual vertical distribution of K at location WBC504L may be more complex than the two-layered simulation models. For example, streambed K may decrease sharply between 20 and 24 cm but then increase again such that the harmonic mean K of the sediments in the permeameter remained roughly constant as the permeameter advanced from 24 to 36 cm. A similar complexity in layering might have been present at WBC356.25R and WBC400C, to a smaller degree than at WBC504L. Additional simulations with other variations in layering may better constrain the exact vertical distributions of K in the streambed sediment. But exact knowledge of vertical K profiles is not needed for accurate calculation of streambed water and solute fluxes from streambed K values [e.g., Gilmore et al. 2016a,b; Kennedy et al. 2009a,b], only accurate mean K values for the full depth interval of interest are needed. Figure 10 suggests that simulation results can be useful in at least approximately constraining the possible vertical distributions of streambed K that might produce observed profiles of *in-situ* K from equation 1. In this way, model simulations can help with interpretation of the *in-situ* field results.

4. Conclusions

Based on numerical simulation results and laboratory sand tank experimental results, the Hvorslev analysis for estimating hydraulic conductivity (K) from falling-head test data yielded accurate K values in both homogenous sediment (for which the analysis had been originally derived) and layered sediment with low K above high K. K from the Hvorslev analysis deviated significantly from true K only when two conditions were present together: (1) high-K material was present over low-K material, and (2) the outlet of the permeameter was at or very near (within 2 cm of) the interface between high-K and low-K. When this combination of conditions exists, results from numerical simulations and laboratory sand tank experiments show that in-situ Hvorslev K (from equation 1) underestimates the true K of the sediment within a permeameter, because the falling-head test is affected by low-K sediment outside of (below the bottom of) the permeameter. In simulation results, the maximum underestimation (occurring when the bottom of the permeameter was at the layer interface) was by a factor of 0.91, 0.59, and 0.12 when the high-K to low-K ratio was 2, 10, and 100, respectively. In laboratory sand tank experiments, the underestimation was by a factor of about 1.2 when the high-K to low-K ratio was 2.3. Simulation results with the permeameter bottom at an interface of high K over low K showed that the underestimation of K by the Hvorslev analysis was about the same whether the underlying low-K layer was 2 cm or 174 cm thick (1% or 87% of the domain thickness).

Only 3 of 11 *in-situ* streambed K profiles showed evidence of strong layering, each with high-K over low-K. Numerical model simulations were useful in the interpretation of *in-situ* field K profiles at the streambed sites with layering; specifically, scaling the model results to the maximum measured K at the top of the field K profiles, and shifting model results vertically so

that the K interface in the model domain was at the same depth as the interface in the field, helped constrain the likely ratio of high K to low K at the field measurement locations with layered heterogeneity.

The work reported here suggests recommendations for future field campaigns that require point estimates of streambed K. After a K estimate has been performed at the desired measurement depth, a second K estimate should be performed at the same location but at a slightly deeper depth (2-3 cm). If the K estimate at the deeper depth is approximately the same or higher, then the shallower K estimate was likely not affected by layered heterogeneity in the sediment (our results show that, for K estimates based on Hvorslev analysis of *in-situ* falling-head tests, the only possible significant effect of layered heterogeneity is to cause a K estimate to be lower, not higher, than one made just above it). If however the K value at the deeper depth is more than about 15% lower than the shallower K estimate, the shallower K estimate could have been affected by low K sediments beneath the bottom of the permeameter (our results show a variation up to 11% in Hvorslev estimates of K from *in-situ* falling-head tests in homogeneous sand tanks, meaning that differences this small or smaller are not likely to be reliable indicators of an influence of layered heterogeneity on field estimates of K).

In the latter case, one should continue insertion of the permeameter to one or more deeper depths and run a falling-head test at each depth. If K is found to level out with depth (e.g., lower part of the WBC356.25R field profile, Fig. 10), the plateau K value could be used along with a vertical head gradient measured over the same depth interval in the streambed to calculate vertical water flux through the streambed with Darcy's equation. The plateau K value should give reasonable

accuracy for this calculation if the bottom of the field permeameter is more than several cm from an interface between high and low K. For example, even with sediment of high K on top of sediment of lower K (the most problematic combination for Hvorslev analysis), and a factor of 10 between the high and low K values, Hvorslev K was within about 12% of true K when the bottom of the permeameter was 10 cm below the interface between high and low K, and the agreement improved as the permeameter went deeper (Fig. 6, middle graph) [Burnette 2013].

Also, if information about the K structure of the sediment layering is desired, numerical simulation results could be used as in Figure 10 to see if field results are generally consistent with a model of two or more layers, and if so, to constrain the K values of the layers. For example, K in streambed measurement location WBC356.25R seemed to level out at about 5 m/day in the depth range 24-36 cm, but comparison to scaled simulation results (Fig. 10) suggested that the measured field profile is consistent with sediment of $K \approx 22$ m/day over deeper layer of $K \approx 2$ m/day (the “plateau value” of about 5 m/day at depth was actually a slowly decreasing K of 5.3 to 4.6 m/day from 24 to 36 cm).

Finally, a core could be collected as we did by digging the permeameter out of the streambed, and a falling-head test could be done on the core in the lab (where there will be no possibility of influence from sediment outside the permeameter).

The work reported here includes specific new insights and practical findings for studies in groundwater - surface water interaction. Vertical K values are important in this field of study, and the Hvorslev analysis (equation 1) can be a useful and relatively simple field tool, even in

604 layered media, when applied carefully. Future research efforts could focus on performing tests
605 (laboratory and numerical) in porous media with layering scenarios beyond those investigated
606 here.

607

608 **Acknowledgements**

609 The research was funded by the U.S. National Science Foundation under award EAR-1045162 to
610 North Carolina State University. The authors gratefully acknowledge the assistance of Troy
611 Gilmore with the field work.

References

- Alayamani, M. S. and Sen, Z. (1993). Determination of hydraulic conductivity from complete grain-size distribution curves. *Groundwater* 31(4): 551-555.
- Burnette, M. C. (2013). Applicability of the Hvorslev solution in estimating streambed hydraulic conductivity from falling-head tests. M.S. Thesis, Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC, 125pp.
- Cardenas, M. B. & Zlotnik, V. A. (2003a). A Simple Constant-Head Injection Test for Streambed Hydraulic Conductivity Estimation. *Groundwater* 41(6): 867-871.
- Cedergren, H.R. (1989). Seepage, Drainage, and Flow Nets. 3rd Edition. John Wiley and Sons, New York, 465 pages.
- Cey, E. E., Rudolph, D. L., Parkin, G. W., and Aravena, R. (1998). Quantifying groundwater discharge to a small perennial stream in southern Ontario, Canada. *J. Hydrology* 210(1–4), 21–37. doi:10.1016/S0022-1694(98)00172-3
- Chen, X. H., (2000). Measurement of Streambed Hydraulic Conductivity and its Anisotropy. *Environmental Geology* 39(12): 1317-1324.
- Chen, X. H., (2004). Streambed Hydraulic Conductivity for Rivers in South-Central Nebraska. *Journal of the American Water Resources Association* 40(3): 561-573.
- Cuthbert, M. O., Mackay, R., Durand, V., Aller, M. F., Greswell, R. B., & Rivett, M. O. (2010). Impacts of river-bed gas on the hydraulic and thermal dynamics of the hyporheic zone. *Advances in Water Resources* 33: 1347-1358.
- Domenico, P. A. & Schwartz, F. W. (1997). Physical and Chemical Hydrogeology. John Wiley & Sons, New York. 506pp.
- Elkhalil, A. B. & Ibrahim, A. E. (2008). Regional Groundwater Flow Modeling of Gash River

635 Basin, Sudan. *Journal of Applied Sciences in Environmental Sanitation*. 3(3): 157-167.

636 Fetter, C. W. (2001). Applied Hydrogeology. 4th ed. *Prentice-Hall*. Upper Saddle River, NJ.

637 598pp.

638 Genereux, D. P., Leahy, S., Mitsova, H., Kennedy, C. D., & Corbett, D. R. (2008). Spatial and

639 temporal variability of streambed hydraulic conductivity in West Bear Creek, North

640 Carolina, USA. *J. Hydrology* 358(3–4), 332–353. doi:10.1016/j.jhydrol.2008.06.017

641 Gilmore, T. E., D. P. Genereux, D. K. Solomon, J. E. Solder, B. A. Kimball, H. Mitsova, & F.

642 Birgand. (2016a). Quantifying the fate of agricultural nitrogen in an unconfined aquifer:

643 stream-based observations at three measurement scales. *Water Resources Research*, in

644 press.

645 Gilmore, T. E., D. P. Genereux, D. K. Solomon, & J. E. Solder. (2016b). Groundwater transit

646 time distribution and mean from streambed sampling in an agricultural coastal plain

647 watershed, North Carolina, USA. *Water Resources Research*, in press.

648 Hvorslev, M. J. (1951). Time Lag and Soil Permeability in Groundwater Observations.

649 Waterways Experiment Station, Corps of Engineers, *US Army Bulletin 36*, Vicksburg

650 MS.

651 Kelly, S. E., & Murdoch, L. C. (2003). Measuring the hydraulic conductivity of shallow

652 submerged sediments. *Groundwater* 41(4):431-439.

653 Kennedy, C. D., Genereux, D. P., Corbett, D. R., & Mitsova, H. (2007). Design of a light-oil

654 piezomanometer for measurement of hydraulic head differences and collection of

655 groundwater samples. *Water Resources Research* 43, W09501,

656 doi:10.1029/2007WR005904

657 Kennedy, C. D., Genereux, D. P., Corbett, D. R., & Mitsova, H. (2009a). Relationships among

groundwater age, denitrification, and the coupled groundwater and nitrogen fluxes
through a streambed. *Water Resources Research* 45, W09402,
doi:10.1029/2008WR007400

Kennedy, C. D., Genereux, D. P., Corbett, D. R., & Mitasova, H. (2009b). Spatial and temporal
dynamics of coupled groundwater and nitrogen fluxes through a streambed in an
agricultural watershed. *Water Resources Research* 45, W09401,
doi:10.1029/2008WR007397

Kennedy, C. D., Murdoch, L. C., Genereux, D. P., Corbett, D. R., Stone, K., Pham, P., &
Mitasova, H. (2010). Comparison of Darcian flux calculations and seepage meter
measurements in a sandy streambed in North Carolina, United States. *Water Resources
Research* 46, W09501, doi:10.1029/2009WR008342

Kurtz, W., Hendricks Franssen, H. J., & Vereecken, H. (2012). Identification of time-variant
river bed properties with the ensemble Kalman filter. *Water Resources Research* 48,
W10534, doi:10.1029/2011WR011743

Landon, M. K., Rus, D. L., & Harvey, F. E. (2001). Comparison of instream methods for
measuring hydraulic conductivity in sandy streambeds. *Groundwater* 39(6): 870-885.

Leahy, S. (2007). Spatial and temporal variability in streambed hydraulic conductivity of West
Bear Creek, North Carolina. M.S. Thesis, Department of Marine, Earth, and Atmospheric
Sciences, North Carolina State University, Raleigh, NC, 117pp.

Lu, C., Chen, X., Cheng, C., Ou, G., & Shu, L. (2012). Horizontal hydraulic conductivity of
shallow streambed sediments and comparison with the grain-size analysis results.
Hydrological Processes, 26(3):454-466, doi: 10.1002/hyp.8143.

680 McCarthy, D.F. (1998). Essentials of Soil Mechanics and Foundations: Basic Geotechnics. 5th
681 Edition. Prentice Hall, Upper Saddle River, New Jersey, 730 pages.

682 Merritt, L. M. & Konikow, L. F. (2000). Documentation of a Computer Program to Simulate
683 Lake-Aquifer Interaction Using the MODFLOW Ground-Water Flow Model and the
684 MOC3D Solute-Transport Model: *U.S. Geological Survey Water-Resources*
685 *Investigations Report 00-4167*: 1-146.

686 Min, L., Yu, J., Liu, C., Zhu, J., & Wang, P. (2012). The spatial variability of streambed vertical
687 hydraulic conductivity in an intermittent river, northwestern China. *Environ. Earth Sci.*
688 *69*:873-883. doi:10.1007/s12665-012-1973-8

689 Ong, J. B. & Zlotnik, V. A. (2011). Assessing lakebed hydraulic conductivity and seepage flux
690 by potentiomanometer. *Groundwater* 49(2):270-274. doi 10.1111/j.1745-
691 6584.2010.00717.x

692 Sebok, E., Duque, C., Engesgaard, P., & Bogh, E. (2014). Hydraulic conductivity of streambed
693 sediments of contrasting geomorphological environments. *Hydrological Processes*,
694 *29*:458-472. doi:10.1002/hyp.10170

695 Song, J. X., Chen, X., Cheng, C., Summerside, S. & Wen, F. (2007). Effects of Hyporheic
696 Processes on Streambed Vertical Hydraulic Conductivity in Three Rivers of Nebraska.
697 *Geophysical Research Letters* 34, L07409. doi: 10.1029/2007GL029254

698 Springer, A. E., Petroustson, K. W. F., & Semmens, B. A. (1999). Spatial and temporal variability
699 of hydraulic conductivity in active reattachment bars of the Colorado River, Grand
700 Canyon. *Groundwater* 37(3):338-344.

701 Winston, R. B. (2000). Graphical user interface for MODFLOW, Version 4. U.S. Department of

702 the Interior, U.S. Geological Survey; Branch of Information Services, *Open-File Report*
703 2000-315, 27pp.

704 Winston, R. B. (2009). ModelMuse – A graphical user interface for MODFLOW – 2005 and
705 PHAST. *U.S. Geological Survey Techniques and Methods 6-A29*, 52pp, available on-line
706 at <http://pubs.usgs.gov/tm/tm6A29/>.
707

Figure Captions

Figure 1. Map view of the top layer (top), and overall schematic (bottom) of the model domain used to numerically simulate falling-head tests. The map-view image shows zones of different horizontal grid spacing: blue areas had the largest cells (1 cm x 1 cm), the black area in the center had the smallest cells (0.25 cm x 0.25 cm), and the white, green, and red areas had rectangular cells with lateral dimensions varying between 0.25 cm and 1 cm. The overall schematic shows the spatial extent and boundary conditions of the domain (h = hydraulic head).

Figure 2. Schematic vertical cross section through the laboratory sand tank (third dimension normal to cross section was 70.5 cm) showing a single permeameter penetrating into low-K sand beneath high-K sand (see text for details of permeameter placement and tank construction). Permeameter dimensions were 121.75 cm long x 7 cm in diameter (the permeameter length is not shown to scale but all other dimensions are). Water in the side reservoir to the right was in direct contact with the open water below the sand pack but not the standing water above the sand pack or the sand itself. Additional tests were performed with a homogenous sand pack in the tank (a single 55 cm thick layer of high-K sand) and with low-K sand overlying high-K sand.

Figure 3. Aerial photo from 2010 (North Carolina OneMap, downloaded August 8th, 2013) of the West Bear Creek and Bear Creek watersheds. The study reaches are outlined in the red rectangles. State Road (SR) 1719 is shown with a black line.

Figure 4. Hvorslev K (calculated using the head vs. time results from the MODFLOW falling-head simulation) and true K (20 m/d) as a function of permeameter penetration for numerical simulations in homogenous sand.

Figure 5. Hvorslev K calculated from the MODFLOW model output, and true K, vs. permeameter penetration. The K values assigned to the model domain were 20 m/d for the lower 174 cm of the model domain, and 10 m/d, 2 m/d, or 0.2 m/d for the upper 26 cm of the model domain.

Figure 6. Hvorslev K calculated from MODFLOW output, and true K, vs. permeameter penetration. The K values assigned to the model domain were 20 m/d for the upper 26 cm of the model domain, and 10 m/d, 2 m/d, or 0.2 m/d for the lower 174 cm of the model domain.

Figure 7. Hvorslev K calculated from the MODFLOW model output, and true K, vs. permeameter penetration. The K values assigned to the model domain were 0.2 m/d in the depth range from 26 cm to 28 cm (the 2 cm layer indicated by horizontal lines on the graph) and 20 m/d for the rest of the model domain.

Figure 8. Results from the sand tank with high-K sand over low-K sand: measured K values from the *in-situ* Hvorslev tests and lab permeameter tests (runs 1 and 2 in the tank), and the K values (true K and Hvorslev K) from the MODFLOW simulations of this tank. In all three cases (run 1, run 2, and model simulations), Hvorslev-estimated K (filled circle) is farthest from the true or best estimate of K (open circle) when the bottom of the permeameter is at the layer interface (26 cm).

Figure 9. Profiles of field (*in-situ* Hvorslev) K and lab permeameter K for the three streambed locations that showed a significant drop in K between 20 and 24 cm. The uncertainty bars were

748 calculated from propagation of uncertainties in the three measurands (L, D, and slope in equation
749 2).

750 **Figure 10.** Profiles of field (*in-situ* Hvorslev) K and scaled Hvorslev K calculated from
751 MODFLOW simulation results (see text for explanation). 2x, 10x, and 100x refer to the ratios of
752 high to low K in the two-layer model domains used for simulations. The uncertainty bars were
753 calculated from propagation of uncertainties in the three measurands (L, D, and slope in equation
754 2).

755

Table 1. *In-situ* K and lab permeameter K values for each full core in or from the homogenous sand tank.

Measurement Target Depth (cm)	Run 1		Percent Difference (%)	Run 2		Percent Difference (%)
	<i>In-situ</i> K (m/d)	Lab K (m/d)		<i>In-situ</i> K (m/d)	Lab K (m/d)	
20	76.5	67.6	12.4	92.9	77.6	17.9
24	90.7	79.8	12.7	72.8	78.5	-7.5
26	87.4	76.9	12.8	80.1	84.7	-5.5
30	101.7	110.4	-8.2	64.4	58.9	8.9

Table 2. *In-situ* K and lab permeameter K values for each full core in or from the tank packed with low-K sand over high-K sand.

Measurement Target Depth (cm)	Run 1		Percent Difference (%)	Run 2		Percent Difference (%)
	<i>In-situ</i> K (m/d)	Lab K (m/d)		<i>In-situ</i> K (m/d)	Lab K (m/d)	
20	39.2	32.7	18.1	32.9	27.8	16.8
24	34.5	32.8	4.8	34.4	34.3	0.2
26	36.9	33.0	11.0	31.2	31.8	-1.9
30	38.3	35.5	7.3	37.5	32.1	15.5

Table 3. *In-situ* K and lab permeameter K values for each full core in or from the tank packed with high-K sand over low-K sand.

Measurement Target Depth (cm)	Run 1		Percent Difference (%)	Run 2		Percent Difference (%)
	<i>In-situ</i> K (m/d)	Lab K (m/d)		<i>In-situ</i> K (m/d)	Lab K (m/d)	
20	77.7	80.6	-3.7	97.6	87.4	11.0
24	85.0	80.9	4.9	105.3	100.6	4.6
26	62.0	73.1	-16.5	66	83.5	-23.5
30	41.5	43.2	-3.9	46.3	48.2	-3.9