Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site

Geoffrey C. Bohling,[1] Gaisheng Liu,[1] Steven J. Knobbe,[1] Edward C. Reboulet,[1] David W. Hyndman,[2] Peter Dietrich,[3] and James J. Butler Jr[1]

Received 14 April 2011; revised 1 December 2011; accepted 9 January 2012; published 21 February 2012.

[1] Spatial variations in hydraulic conductivity (K) provide critical controls on solute transport in the subsurface. Recently, new direct-push tools were developed for high-resolution characterization of K variations in unconsolidated settings. These tools were applied to obtain 58 profiles (vertical resolution of 1.5 cm) from the heavily studied macrodispersion experiment (MADE) site. We compare the data from these 58 profiles with those from the 67 flowmeter profiles that have served as the primary basis for characterizing the heterogeneous aquifer at the site. Overall, the patterns of variation displayed by the two data sets are quite similar, in terms of both large-scale structure and autocorrelation characteristics. The direct-push K values are, on average, roughly a factor of 5 lower than the flowmeter values. This discrepancy appears to be attributable, at least in part, to opposite biases between the two methods, with the current versions of the direct-push tools underestimating K in the highly permeable upper portions of the aquifer and the flowmeter overestimating K in the less permeable lower portions. The vertically averaged K values from a series of direct-push profiles in the vicinity of two pumping tests at the site are consistent with the K estimates from those tests, providing evidence that the direct-push estimates are of a reasonable magnitude. The results of this field demonstration show that direct-push profiling has the potential to characterize highly heterogeneous aquifers with a speed and resolution that has not previously been possible.

Citation: Bohling, G. C., G. Liu, S. J. Knobbe, E. C. Reboulet, D. W. Hyndman, P. Dietrich, and J. J. Butler Jr. (2012), Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site, *Water Resour. Res.*, *48*, W02525, doi:10.1029/2011WR010791.

1. Introduction

[2] A large body of theoretical and experimental research has identified the spatial distribution of hydraulic conductivity (K) as the most significant factor controlling the fate and transport of solutes in subsurface flow systems [*Gelhar and Axness*, 1983; *Dagan*, 1989; *Boggs et al.*, 1992; *Koltermann and Gorelick*, 1996; *Dagan and Neuman*, 1997; *Fogg et al.*, 2000]. The detailed characterization of heterogeneous aquifers is clearly needed to develop predictive models and improve our understanding of transport behavior [*Molz et al.*, 1986; *Sudicky*, 1986; *Killey and Moltyaner*, 1988]. Intensive studies of the Borden [*Freyberg*, 1986; *Mackay et al.*, 1986; *Sudicky*, 1986] and Cape Cod [*Garabedian et al.*, 1991; *LeBlanc et al.*, 1991; *Hess et al.*, 1992] sites have demonstrated that the classic advection-dispersion models can reasonably describe field-scale solute transport in mildly heterogeneous aquifers (variance of $\ln K < 1$; Borden = 0.29, Cape Cod = 0.26) using conventional methods to characterize the K field (e.g., geostatistical analysis of core permeameter data [*Sudicky*, 1986]).

[3] In contrast to studies of solute transport in mildly heterogeneous aquifers, efforts to simulate transport in moderately to highly heterogeneous aquifers (variance of $\ln K > 2$) using classic advection-dispersion models have not met with success [e.g., *Eggleston and Rojstaczer*, 1998; *Whittaker and Teutsch*, 1999]. The most frequently cited studies of transport in highly heterogeneous aquifers were carried out at the macrodispersion experiment (MADE) site on Columbus Air Force Base in Mississippi [*Boggs et al.*, 1992; *Zheng*, 2006; *Zheng et al.*, 2011]. Over the past 25 years, three large-scale tracer tests were conducted in the MADE aquifer (variance of $\ln K$ from flowmeter profiles at 49 wells ≈ 4.5 [*Rehfeldt et al.*, 1992]). Simulations of these tracer tests using advection-dispersion models have failed to reproduce the solute transport observed at MADE, which has stimulated widespread discussions on limitations of site characterization methods and solute-transport models for highly heterogeneous aquifers [*Barlebo et al.*, 2004; *Molz et al.*, 2006; *Hill et al.*, 2006].

[4] The results of the analyses of the MADE tracer tests indicate that the successful prediction of contaminant

[1]Kansas Geological Survey, University of Kansas, Lawrence, Kansas, USA.
[2]Department of Geological Sciences, Michigan State University, East Lansing, Michigan, USA.
[3]Department of Monitoring and Exploration Technologies, UFZ, Helmholtz Centre for Environmental Research, Leipzig, Germany.

0043-1397/12/2011WR010791

transport in highly heterogeneous aquifers requires detailed, site-specific characterization of the *K* field [*Zheng et al.*, 2011]. The key issue is how this can be accomplished. Practicing hydrologists cannot justify the costly, multiyear efforts that were required to assemble the data sets that have underlain the theoretical analyses of the Borden, Cape Cod, and MADE tracer tests [e.g., *Sudicky*, 1986; *Hess et al.*, 1992; *Rehfeldt et al.*, 1992]. Moreover, in highly heterogeneous aquifers, conventional field methods performed in existing wells cannot provide sufficiently detailed *K* information because of (1) the small number of wells at most sites, (2) the large averaging volumes of some of the methods, and (3) the sensitivity of many of the methods to the near-well disturbed zone and/or in-well hydraulics [*Butler*, 2005]. New field methods capable of rapid collection of high-resolution *K* data without the need for existing wells are clearly needed to improve transport prediction in heterogeneous aquifers [*Liu et al.*, 2009; *Lessoff et al.*, 2010]. In this paper we describe the results of a field demonstration of a new class of methods that has great potential for this purpose.

[5] Direct-push technology has been widely utilized to characterize shallow (depths <30 m) unconsolidated formations [*Dietrich and Leven*, 2006; *McCall et al.*, 2005]. In the last decade, a new series of direct-push tools has been developed for characterization of *K* variations in unconsolidated aquifers [*Liu et al.*, 2012]. The direct-push injection logger (DPIL) was developed for the rapid characterization of relative variations in *K* at a vertical resolution as fine as 0.015 m; a 15 m profile can be completed in less than 2 h [*Dietrich et al.*, 2008; *Liu et al.*, 2009; *McCall et al.*, 2009]. The direct-push permeameter (DPP) was developed to obtain reliable *K* estimates over a 0.4 m vertical interval with a series of short injection tests; testing of a moderate to high *K* interval can be completed in 10–20 min [*Butler et al.*, 2007]. Recently, the DPIL and DPP have been combined in the high-resolution *K* (HRK) tool [*Liu et al.*, 2009, 2012]. The HRK tool produces collocated DPIL and DPP profiling data, allowing the high-resolution DPIL data to be transformed into *K* estimates. The calibration procedure of *Liu et al.* [2009] uses a numerical model to account for the DPIL measurements at their acquired resolution, thereby circumventing the need to compare measurements at different support scales.

[6] These new approaches were recently applied at the MADE site. A series of direct-push profiles, using both DPIL-only and HRK tools, were performed across the area impacted by earlier tracer tests and where a large amount of *K* data had previously been acquired. These data were obtained from borehole flowmeter profiling in 67 wells at the MADE site [*Boggs et al.*, 1993]. *Rehfeldt et al.* [1992] present a geostatistical analysis of the 2187 *K* measurements, most representing vertical intervals of 0.15 m, from 49 of the 67 flowmeter profiles. The vast majority of the flowmeter wells were installed using a drive and wash method and were then developed for about 2 h [*Rehfeldt et al.*, 1989]. The time required to complete a flowmeter profile in each well, after well installation and development, was 2 to 8 h. In contrast, it was possible to complete a DPIL profile through the entire thickness of the MADE aquifer in 1 to 2 h, without the need to install a well. The 58 direct-push profiles described in this paper were obtained over four weeks at the site in 2008 and 2009. Under good conditions we were able to obtain six DPIL profiles per day.

[7] The primary objective of this article is to assess the efficacy of the new direct-push approaches through a descriptive assessment of *K* variations determined from 58 direct-push profiles (vertical resolution of 1.5 cm) at the MADE site, along with a comparison to the well-known flowmeter data from that site. We are comparing the direct-push *K* data to the flowmeter *K* data because the flowmeter data have served as the primary basis for characterizing the MADE site for two decades, not because we believe the flowmeter data to be flawless. Indeed, *Rehfeldt et al.* [1989] document a number of factors that contribute to uncertainty and bias in the flowmeter *K* measurements, including sensitivity to near-well conditions (as documented by large changes in *K* estimates with well development), sensitivity to parameters describing well losses, and errors in the measured discharge profiles, especially in lower *K* zones. Thus, discrepancies between flowmeter and direct-push *K* could result from errors in either or both techniques.

[8] Previous studies have demonstrated favorable agreement between direct-push *K* estimates and *K* estimates obtained using conventional techniques (e.g., multilevel slug tests and grain size analyses) at alluvial aquifer sites in Kansas and Germany [*Butler*, 2005; *Butler et al.*, 2007; *Dietrich et al.*, 2008; *Liu et al.*, 2012], but these new direct-push techniques have not yet been assessed at a site as extensively characterized as MADE.

[9] The paper begins with an overview of the different types of *K* data considered in the analysis, after which the assessment and comparison of the *K* variations are described. The comparison shows that the higher-resolution direct-push profiles, which were obtained in a much shorter time period with significantly less effort, depict the same general *K* field structure as the flowmeter profiles. Thus, direct-push profiling appears to provide the detailed *K* information required for risk assessments and remediation designs in a practically feasible manner.

2. Site Overview

[10] *Zheng et al.* [2011] provide an overview of the 25 years of work aimed at characterizing the alluvial aquifer at MADE. The primary source of information for characterizing the *K* distribution at the site has been flowmeter profiles from a network of 67 wells [*Boggs et al.*, 1993], 49 of which were analyzed by *Rehfeldt et al.* [1992]; these wells have since been removed from the site. The direct-push profiles described in this study include 21 HRK profiles and 37 DPIL-only profiles (Figure 1), obtained during two field campaigns (two weeks each) in November 2008 and October 2009. The three smaller outlined areas in Figure 1 (ICA (intensively cored area) Cube, MLS (multilevel sampler) Cube, and Source Area) are used for more detailed investigations and are labeled here for cross referencing with other studies [e.g., *Liu et al.*, 2009; *Dogan et al.*, 2011]. The depths reported in this study are meters below the reference point for the site surveys conducted in 2008 and 2009. The elevation of that reference point, approximately 65.1 m above sea level (m asl), was estimated from the elevations by *Boggs et al.* [1993]. The flowmeter data

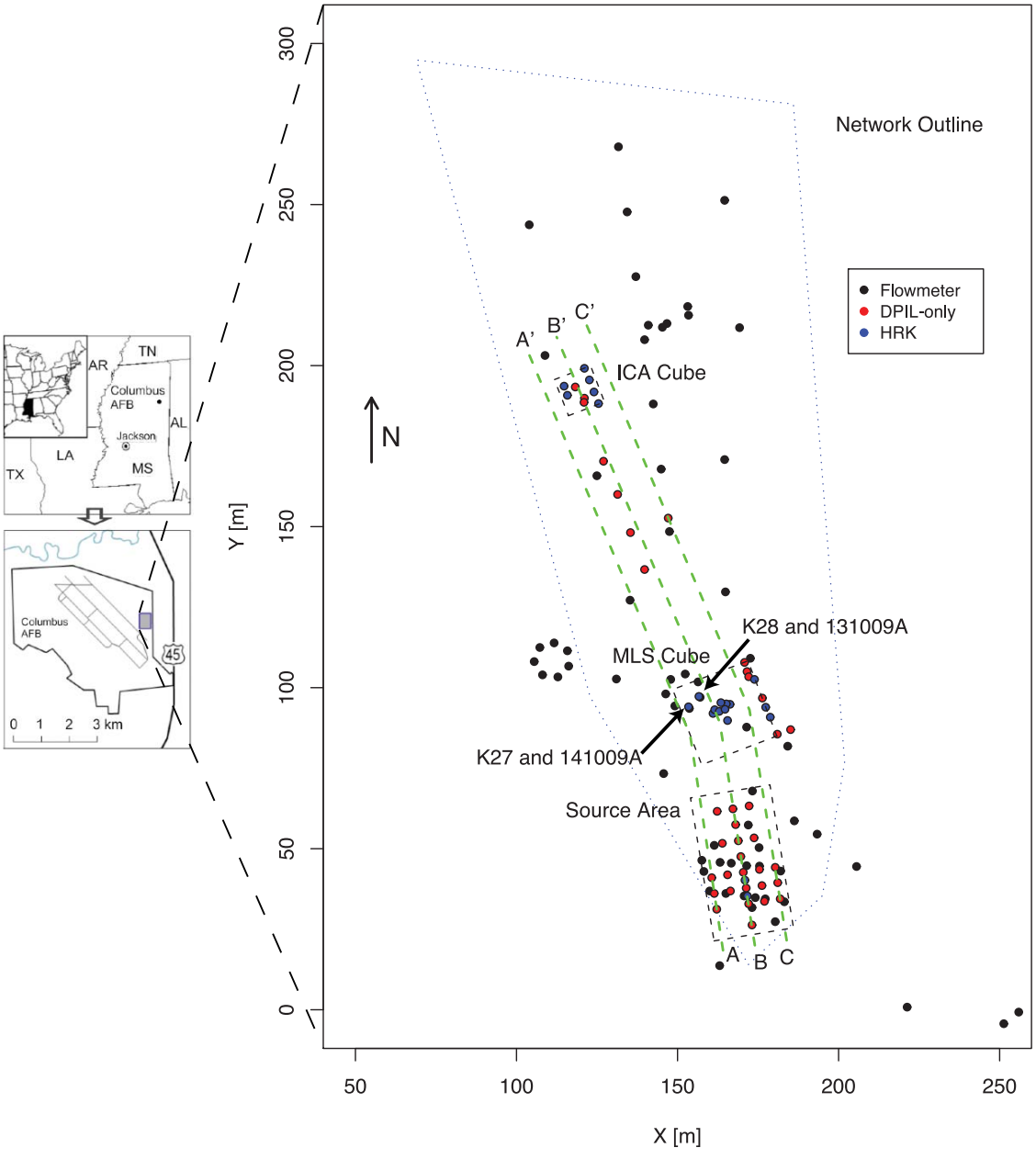

Figure 1. Locations of flowmeter wells (since removed from site) and direct-push profiles identified by profile type (DPIL-only and HRK [combined DPIL and DPP]), along with outlines of previously existing sampler network (blue), intensively sampled subareas (black), and locations of three cross sections displayed in Figure 7 (green). Locations of pairs of flowmeter and HRK profiles shown in Figure 3 are also noted. Coordinate system *Y* axis points due north, rather than 12 deg west of north like MADE-2 coordinate system [*Boggs et al.*, 1993]. The origin of the MADE-2 coordinate system (location of the center injection well) corresponds to $(X,Y) = (173\text{ m}, 28\text{ m})$ in this coordinate system.

cover a depth range of 2.5 to 12.9 m below the survey reference point. For this comparative study, we only employ DPIL data from the same depth range as the flowmeter data, thus excluding DPIL data from depths shallower than 2.5 m (with the exception of the direct comparison of adjacent profiles in Figure 3).

3. Direct-Push Data Acquisition

[11] The direct-push profiles discussed in this paper were obtained using two new direct-push tools. One tool is a direct-push injection logger (DPIL) probe coupled with an electrical conductivity probe. The DPIL component of the tool consists of a single screened port on the drive rod [*Dietrich et al.*, 2008; *Liu et al.*, 2009, 2012; *McCall et al.*, 2009]. Water is injected through the screen and pressure is monitored continuously as the rod is advanced. The injection pressure can be calculated from the monitored pressure using an estimate of the hydrostatic pressure at the injection port. The ratio of injection rate to injection pressure provides a high-resolution (0.015 m) indicator of vertical

variations in K. The second tool, referred to as the high-resolution K (HRK) tool, is a direct-push permeameter (DPP) probe modified to allow operation in DPIL mode during advancement [*Liu et al.*, 2009, 2012]. The HRK tool consists of a cylindrical screened port at the tip plus two transducer ports set into the side of the drive rod [*Liu et al.*, 2009]. The tool allows continuous injection and pressure monitoring through the transducer ports, so that DPIL-style data are collected at two points during advancement. In this study we employ the DPIL data obtained at the top pressure port to minimize the impact of drive-induced mechanical pressures. At certain depths, tool advancement is halted, the background hydraulic head is determined, and DPP tests are performed. The DPP test involves a short period of injection through the cylindrical tip screen and monitoring of pressure responses at both transducers. The spherical form of Darcy's law can be used to obtain an analytical estimate of K based on the DPP pressure responses [*Butler et al.*, 2007]; this estimate represents an effective horizontal K over the 0.4 m vertical interval between the injection screen and the upper pressure transducer [*Liu et al.*, 2008].

[12] In Figure 1 and hereafter (when it is appropriate to distinguish between the two tools) we refer to the first of these two tools using the term "DPIL-only". The DPIL data analyzed in this study are from both the DPIL-only and HRK profiles. The DPP data are from the HRK profiles.

[13] The primary logistical problem encountered during direct-push profiling at the MADE site was the need to replace transducer port screens after approximately every fifth profile. Due to their location on the side of the drive rod, these screens were subject to a considerable amount of mechanical wear during advancement through the angular gravels at the MADE site. Nevertheless, the screens can be replaced rapidly in the field. Only one transducer failed during the four weeks of profiling at MADE. Although the rod string is hammered during advancement, a drive cushion reduces the energy transmitted to the tool. In addition, the pressure transducers are connected to the rod through flexible tubing. Like the screens, the transducers can be replaced in the field, as the transducers and their port screens are readily accessible when the tool is pulled above the ground.

[14] Three factors that could impact the accuracy of direct-push measurements in the field are screen clogging, compaction of the sediments in the vicinity of the drive rod, and drive-induced mechanical pressure. With both of the tools used in this study, the continuous injection of water during advancement reduces the potential for screen clogging, which does not appear to have been a problem at MADE. A simulation study by *Liu et al.* [2008] demonstrated that DPP test responses are relatively insensitive to the effects of a narrow compacted zone around the drive rod, a result that is in accordance with the findings of *Lowry et al.* [1999]. The effect of sediment compaction on the DPIL measurements is more difficult to assess. In the following data analysis, we assume this effect remains relatively constant during driving, in which case the calibration process should compensate for it automatically. This assumption appears to be reasonable given the overall agreement between DPIL and DPP K values (see Figure 4 in *Liu et al.* [2009] and Figure 3 in this paper). The effect of drive-induced mechanical pressure on the DPIL measurements is assumed to be negligible at the position of the DPIL sensor port, which is 0.4 m above the end of drive rod on both the DPIL-only and HRK tools [*Liu et al.*, 2012].

4. Direct-Push Data Analysis

[15] Our procedure for transforming the high-resolution DPIL data into K estimates is based on the calibration of the DPIL data from the HRK profiles described by *Liu et al.* [2009]. The calibration involves finding optimal parameters a and b of a power transform for converting the ratio of DPIL injection rate (Q; kept relatively constant during profiling) to injection-induced pressure (P) into K:

$$K = e^{b}(Q/P)^{a} \quad \text{or} \quad \ln(K) = a\ln(Q/P) + b. \tag{1}$$

[16] The resulting K profile, represented directly at the 0.015 m resolution of the DPIL data, is used in a radial-vertical finite difference flow model [*Bohling and Butler*, 2001] to simulate the pressure responses to the DPP injection tests. The a and b parameters are then adjusted to minimize the sum of squared residuals between observed and simulated logarithmic pressures over all the DPP tests from the HRK profiles for a particular year (with different calibrations for each year due to tool modifications). This calibration approach accounts for the DPIL data at their acquired resolution, circumventing the need to upscale the DPIL data (0.015 m) for comparison with the analytical DPP K values (0.4 m). The DPIL and DPP K estimates are related through their mutual dependence on the DPP test data, but the DPIL data are not regressed against the DPP K estimates. The calibration equation developed from the HRK profiles from each year was also applied to the DPIL-only profiles from that year, as the profiling procedure was roughly the same and the dimensions of the DPIL components of the tools are identical for a given year.

[17] *Liu et al.* [2009] present the analysis of the 11 HRK profiles collected at MADE in 2008. For 54 DPP tests performed during these profiles, the correlation between observed and simulated logarithmic pressure head differences (Figure 2) is 0.70. This analysis demonstrated that the DPIL Q/P values were impacted by an upper threshold value corresponding to a K of roughly 1×10^{-4} m s^{-1}, because the small tubing diameter limited injection rates and, therefore, the ability to obtain measurable formation pressure responses in higher-K zones. This limitation is apparent in the systematic overestimation of small DPP head differences using the transformed DPIL K values (note the lack of simulated head differences less than 0.02 m in Figure 2), corresponding to an underestimation of K in the more permeable intervals (Figure 4 in *Liu et al.* [2009]). Improvements to both the tool and profiling procedure (primarily increasing tubing diameter and DPIL injection rate) allowed us to raise that threshold to about 7×10^{-4} m s^{-1} for the 2009 campaign, albeit at the cost of nonlinear flow losses in the tubing and the transducer port [*Liu et al.*, 2012]. For the 42 DPP tests performed in conjunction with the 10 HRK profiles obtained in 2009, the correlation between observed and simulated logarithmic pressure differences is 0.79 and the match to small pressure

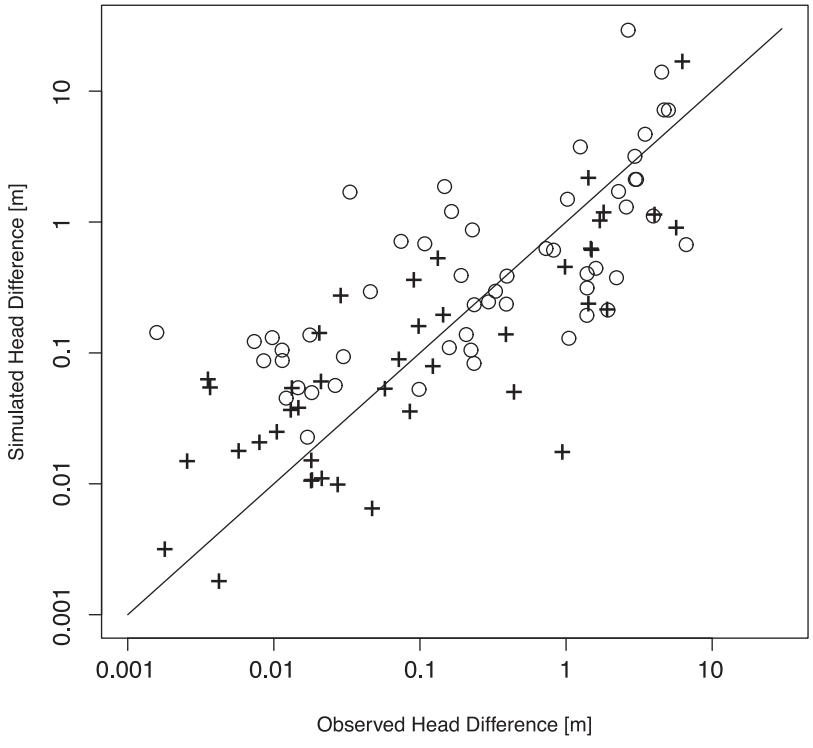

Figure 2. Crossplot of flow-induced pressure head differences observed during DPP tests performed in 2008 (circles) and 2009 (pluses) and those simulated using calibrated DPIL *K* profiles.

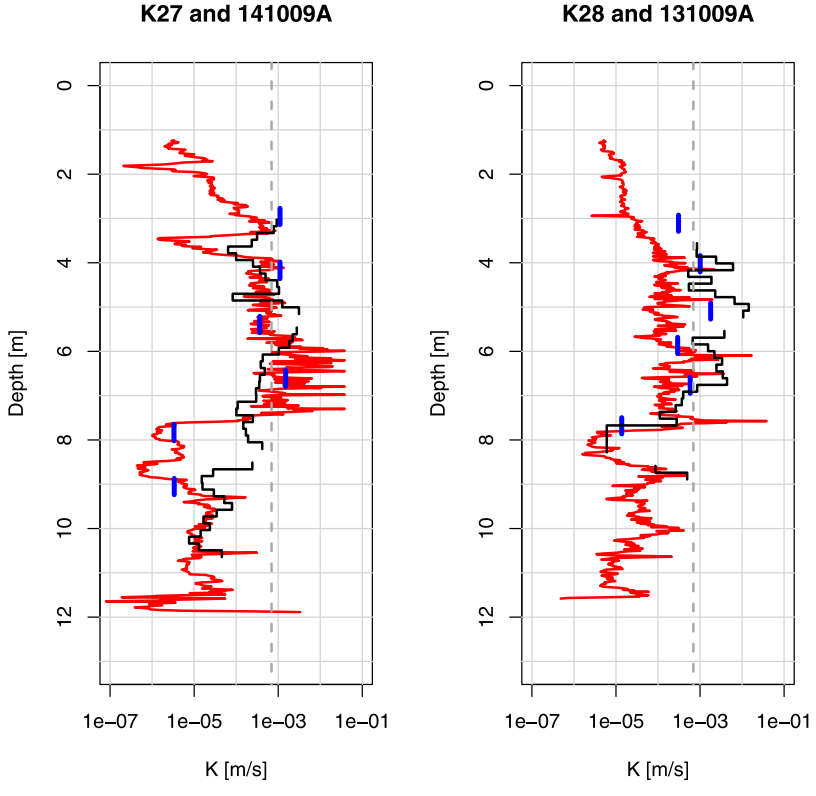

Figure 3. Comparison of adjacent flowmeter and 2009 HRK profiles (locations noted on Figure 1). Flowmeter *K* estimates for 0.15 m intervals are shown in black, analytical DPP *K* estimates are shown by blue line segments spanning 0.4 m interval associated with each test, and DPIL-*K* estimates at 0.015 m spacing are shown in red. The gray dashed line at 7×10^{-4} m s^{-1} represents a rough upper limit for reliable DPIL-*K* estimates in 2009.

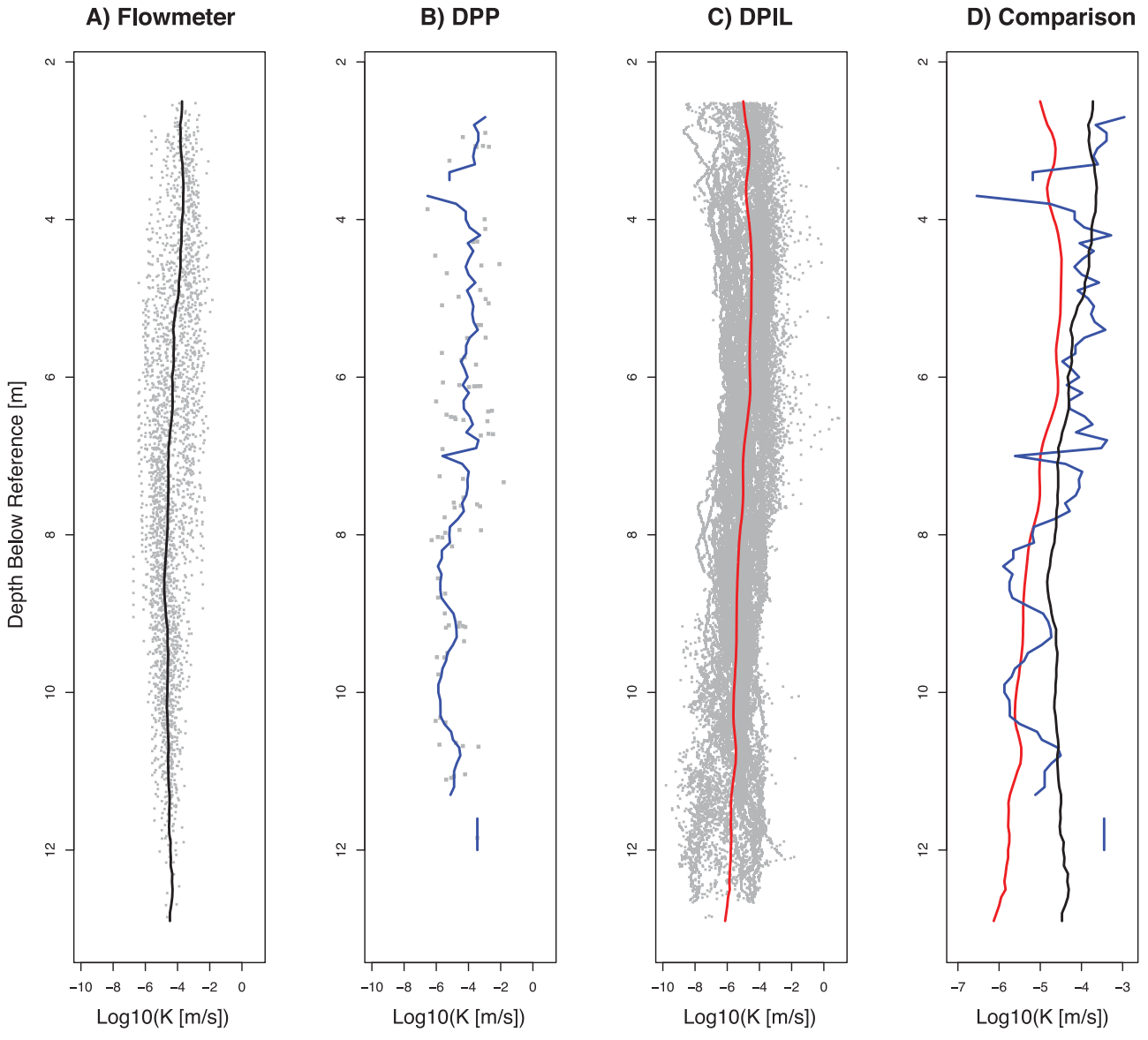

Figure 4. All (a) flowmeter, (b) DPP, and (c) DPIL Log10(K) values between 2.5 and 12.6 m depth (the depth range of the flowmeter data) with lines representing Gaussian kernel smooths with 0.5 m bandwidth. (d) Comparison of three smoothed (composite) profiles (note expanded horizontal axis).

differences is significantly improved (Figure 2). The 2009 DPIL profiles (both HRK and DPIL-only) exhibit two contrasting behaviors in high-K zones: (1) intervals in which DPIL Q/P values are constrained by the upper limit, leading to an underestimation of the K and a suppression of variance, and (2) anomalously noisy intervals leading to spuriously high K values and an inflation of variance.

[18] In addition to the upper limit on reliable DPIL measurements, another problem affecting the calibration is a lack of DPP tests in the lower-K zones, due to the extended time required for pressures to stabilize in these zones. Consequently, the lower DPIL-based ln$K$ values represent an extrapolation of the linear transformation of the ln(Q/P) values beyond the range for which we have constraining DPP pressure response data. Note that both of these problems would have impacted the calibration even if we had used the conceptually simpler approach of regressing DPIL data against the analytical DPP K estimates.

[19] Finally, a range of calibration parameters (a, b), and thus DPIL-based K values, yield an acceptable match between observed and simulated DPP test responses, leading to a fairly large calibration uncertainty. However, there would be little point in presenting a statistical assessment of this uncertainty without first addressing the systematic problems in the calibration (i.e., the upper limit on DPIL measurements and lack of DPP tests in lower-K zones). Furthermore, it is quite possible that the ideal relationship between ln(Q/P) and K exhibits significant nonlinearity and we are currently investigating a calibration procedure that would accommodate such nonlinearity.

[20] In the present study we simply acknowledge the provisional nature of the calibration and focus primarily on

a comparison of patterns of variation indicated by the DPIL and flowmeter data. Because the ln$K$ values are a linear transformation of the ln(Q/P) values, the patterns of variation of the two are the same. In particular, the semivariogram of the DPIL-based ln$K$ values is the same as that of the ln(Q/P) values, apart from a scaling of the sill (variance) that is dictated by the value of a. Thus, comparisons of the shapes (e.g., exponential) and correlation lengths of the DPIL and flowmeter semivariograms are not influenced by the values of a and b. Despite these provisos regarding the calibration, the DPIL K values compare favorably with the bulk K estimates obtained from pumping tests in two areas of the site, providing evidence that the a and b estimates are reasonable.

5. Results

5.1. Direct Comparison of Direct-Push and Flowmeter Data

[21] In order to obtain a direct comparison of flowmeter and direct-push estimates, we placed a direct-push profile within approximately a meter from the estimated location of a former flowmeter well at two locations (see Figure 1). K27 and K28 are flowmeter profiles and 131009A and 141009A are HRK profiles obtained in 2009. While the paired flowmeter and direct-push profiles (Figure 3) show some agreement, they also show some discrepancies that we attribute primarily to opposite biases of the two techniques. The flowmeter tends to overestimate K values in low-K zones, such as indicated in the K27–141009A comparison from 7.5 to 9 m depth, and the DPIL tends to underestimate K in high-K zones, such as indicated in the K28–131009A comparison from 4 to 5.5 m depth. *Rehfeldt et al.* [1992] report that the lower threshold of flow measurement for the impeller flowmeter used at MADE corresponds to a K of approximately 1×10^{-6} m s^{-1}. Even though the 141009A DPIL K estimates are not substantially below this limit between 7.5 and 10 m depth, it seems plausible that low flow rates at these depths could lead to an overestimate of K by the flowmeter. The DPIL estimates, on the other hand, are limited on the high end due to the problems mentioned earlier. These profiles demonstrate the contrasting effects of this upper threshold mentioned above. In particular, the DPIL K estimates between 4 and 6 m depth in 131009A seem to be suppressed, while the interval between 6 and 7.5 m in 141009A is an example of an anomalously noisy section including spuriously high values.

[22] The two DPIL K profiles are notably more consistent with each other than are the two flowmeter profiles. The distance between the two profiles is only 5.7 m, smaller than the horizontal correlation length (12.2 m) estimated by *Rehfeldt et al.* [1992]. Given the relatively short distance, the lack of consistency between the flowmeter profiles indicates that the flowmeter K estimates may be subject to a higher degree of noise than the DPIL estimates.

[23] To get a better sense of how the flowmeter, DPP, and DPIL K profiles compare on average, we computed a composite profile for each type (Figure 4). The composite profiles represent Gaussian kernel smoothing of the data from all profiles of each type using a vertical bandwidth (twice the interquartile range) of 0.5 m in each case. A comparison of the three composite profiles (Figure 4d) clearly shows that (1) in the upper portions of the aquifer, where the DPIL K estimates have an upper limit as discussed earlier, the flowmeter and DPP K estimates agree well on average; and (2) in the lower portions of the aquifer, where the flowmeter measurements are probably impacted by a lower limit on accurate measurement of incremental flowrate, the DPP and DPIL K estimates agree well on average. Note that the DPP K estimates are also somewhat biased due to the lack of DPP tests in the lowest K intervals. This bias represents a shortage of DPP tests in the low-K zones lower in the aquifer and a shortage of HRK profiles in the Source Area, which has a significantly lower K than regions further north. There are only two HRK profiles in the Source Area, compared to 23 DPIL-only profiles, whereas the two profile types are more equally distributed in other portions of the site (Figure 1). Thus, the composite DPIL profile is more heavily weighted toward the lower-K Source Area than is the composite DPP profile.

[24] The discrepancy between DPIL and flowmeter K values (Figure 4d) is attributable, at least in part, to the DPIL's underestimation of high-K values in the upper portion of the aquifer and the flowmeter's overestimation of low-K values in the lower portion of the aquifer. Given the apparent biases in the DPIL and flowmeter methods, each method alone probably underestimates the K contrast between the upper and lower portions of the aquifer. The composite DPP profile does show a stronger contrast between upper and lower portions of the aquifer, but this result may be due in part to the limited lateral coverage of the DPP tests. Of the 21 HRK profiles, 13 are in the MLS Cube, 6 are in the ICA Cube, and 2 are in the source area (Figure 1). Consequently, the composite DPP profile is less "mixed" with regard to lateral variations than either the DPIL or flowmeter composite profiles.

5.2. Comparisons of Univariate Distributions

[25] Graphical comparisons of the univariate distributions of the log10(K) values for each profile type are shown in Figure 5, both as nonparametric kernel density estimates and as normal quantile-quantile (QQ) plots (see Appendix). Representations of the DPIL data shifted to the same logarithmic mean as the flowmeter data have been included on the plots to facilitate comparison of the shapes of the two distributions. Apart from the difference in means, the shapes of the central portions of the flowmeter and DPIL K distributions are similar, although the DPIL data clearly exhibit significantly heavier tails. A lower limit to the flowmeter K estimates is readily apparent in the QQ plot, but the upper limit on reliably measureable DPIL K values is not readily apparent, due in large part to the noisy intervals contributing spuriously high-K values. The DPP K distribution is notably more bimodal than either the flowmeter or DPIL distribution, a result that is consistent with the comparison of composite profiles (Figure 4).

[26] Table 1 compares univariate statistics for the flowmeter, DPP, and DPIL K data. These statistics are presented with 95% confidence intervals that were computed using the same approach as *Rehfeldt et al.* [1992]. This approach, described in the Appendix of this paper, provides estimates for the sample variances of the ln$K$ mean and variance estimates after accounting for the effect of spatial autocorrelation, which reduces the amount of

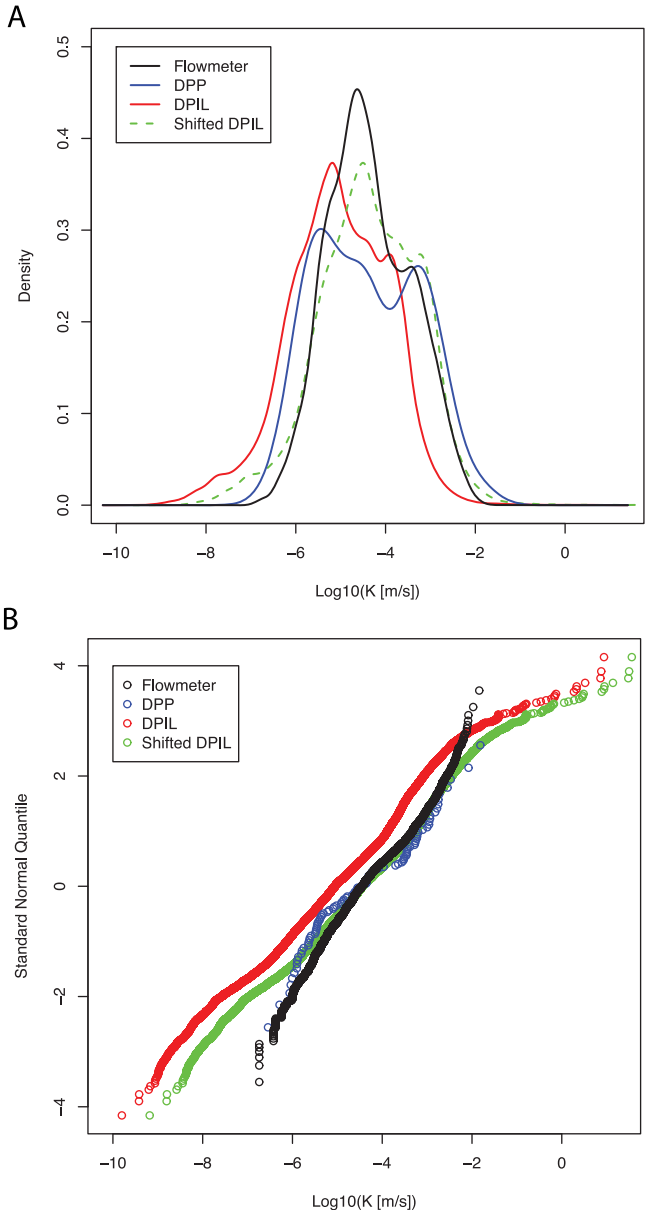

Figure 5. Univariate distributions of log10(K [m s^{-1}]) for each profile type displayed as (a) nonparametric density curves, and (b) normal quantile-quantile plots for measurements between 2.5 and 12.6 m depth (depth range of flowmeter data). Green dashed line and green circles represent DPIL data shifted to same logarithmic mean as flowmeter data, to ease comparison of distribution shapes.

independent information in the data. The third column in Table 1 (N^*) represents the approximate number of independent data for each data type. The more numerous DPIL data are deemed to contain less independent information than the flowmeter data as a consequence of the more limited areal coverage of the DPIL profiles; as already mentioned, the areal coverage of the DPP data (from the HRK profiles) is even more limited. The confidence intervals in Table 1 only account for the effects of sample variance, ignoring the larger and probably more important uncertainties associated with systematic problems in either the flowmeter or direct-push data. Therefore, we caution the reader against taking away a false sense of accuracy from this rather limited uncertainty assessment.

[27] The DPP K geometric mean is well within the 95% confidence interval for the flowmeter geometric mean. The DPP values have a higher ln$K$ variance (6.9 compared to 4.4) than the flowmeter values, although the confidence intervals for these estimates overlap. Again, it is important to consider the differences in lateral spatial coverage associated with these data. However, as explained in the discussion of Figure 4, it is also possible that both the flowmeter and DPIL measurements underestimate the K contrast between the upper and lower portions of the aquifer, a primary contributor to the ln$K$ variance. The DPP data, although more sparse, may provide the most accurate representation of this contrast.

[28] The DPIL K data have a significantly lower geometric mean K (8.9×10^{-6} m s^{-1}) than the flowmeter data (4.3×10^{-5} m s^{-1}), consistent with the offset between the two methods seen in the composite profiles (Figure 4d). The ln$K$ variance for the DPIL data (6.6) is notably higher than that for the flowmeter data (4.4), although the 95% confidence intervals for these two estimates do overlap. One might imagine that the DPIL data would exhibit a higher variance due to the smaller vertical sample spacing (smaller support volume) relative to the flowmeter data. We investigated this issue by upscaling the DPIL data to the same sample interval as the flowmeter data, using a simple arithmetic averaging of the DPIL K values (10 in most cases) within 0.15 m vertical intervals in each profile, on the premise that the flowmeter K values represent estimates of effective horizontal K. These upscaled values actually exhibited a slightly higher ln$K$ variance (7.0) than the original data. Although not conclusive, this result provides some evidence that the difference in variance is not due to support volume. This difference is more likely due to the stronger K contrast between upper and lower portions of the aquifer in the DPIL data than in the flowmeter data.

Table 1. K Statistics in Bold With 95% Confidence Intervals by Profile Type for Data Between 2.5 and 12.9 m Depth Below Reference (Depth Range of Flowmeter Data)

Type	N[a]	N^*[b]	Geometric Mean K (m s^{-1})[c]	ln$K$ Variance
Flowmeter	2611	163	3.1×10^{-5}; **4.3×10^{-5}**; 6.0×10^{-5}	3.4; **4.4**; 5.4
DPP	95	25	1.4×10^{-5}; **3.9×10^{-5}**; 1.1×10^{-4}	3.0; **6.9**; 10.7
DPIL	31,123	120	5.6×10^{-6}; **8.9×10^{-6}**; 1.4×10^{-5}	4.9; **6.6**; 8.3

[a]Number of data.
[b]Approximate number of independent data after accounting for spatial autocorrelation.
[c]95% confidence intervals for geometric mean represent backtransform of 95% confidence intervals on ln$K$ mean.

Dividing the aquifer at a depth of 8 m (based on the composite profiles in Figure 4d), the flowmeter data indicate a factor of three contrast in geometric mean K between upper and lower portions of the aquifer while the DPIL data indicate a factor of 6 contrast.

[29] There have been two pumping tests performed at the MADE site. One test was conducted using a well at the center of the ICA Cube [*Molz et al.*, 2006]. The pumping well is 0.1 m in diameter and 9.75 m deep, with a screened interval 3.66–9.75 m below land surface and a filter pack consisting of commercial well-sorted medium sands. The depth-averaged K value estimated from this pumping test was 3.9×10^{-5} m s^{-1}. The arithmetically averaged K value for ICA Cube DPIL profiles was very similar at 4.0×10^{-5} m s^{-1} over the 4–10 m depth interval and 3.6×10^{-5} m s^{-1} over the 2.5–12.9 m depth interval. Arithmetic averaging yields the appropriate effective K for horizontal flow through a stratified K distribution, which is probably a reasonable approximation of the flow configuration in the vicinity of the pumping well. Clearly these estimates compare favorably, providing evidence that the DPIL K estimates are of a reasonable magnitude.

[30] There are no flowmeter profiles in the immediate vicinity of the ICA Cube. However, the arithmetically averaged K value from the three flowmeter profiles that form a triangle around the ICA Cube is 1.3×10^{-4} m s^{-1}. From the center of the ICA Cube, these three profiles are approximately 15 m to the northwest, 23 m to the east, and 27 m to the south and the respective measurement depth ranges are 2.9 to 12.2 m, 3.0 to 10.9 m, and 2.6 to 11.0 m. The individual average K values for these three profiles are 3.0×10^{-5}, 2.2×10^{-4}, and 1.6×10^{-4} m s^{-1}, respectively, so the nearest flowmeter profile shows a slightly lower average K than the pumping test, although the average from all three profiles is higher than the pumping test K by a factor of about 3.

[31] *Boggs et al.* [1992] report the results of a second pumping test (referred to as AT2 in their paper) that was conducted in the vicinity of the ring of flowmeter wells at the northwest corner of the MLS Cube. The estimated depth-averaged horizontal K from that test, which apparently involved pumping over a 10 m screened interval, was 2.0×10^{-4} m s^{-1}. In comparison, the arithmetic average of the MLS Cube DPIL K values is 2.2×10^{-4} m s^{-1} over the entire depth interval with measurements (about 1.2 to 13 m) and 2.4×10^{-4} m s^{-1} over the 2.5–12.9 m interval. In contrast, the average of the K values for the seven flowmeter profiles in this ring, with measurement depths ranging from 2.6 to 12.0 m, is 1.0×10^{-3} m s^{-1}. The average for the 10 flowmeter profiles in the immediate vicinity of the MLS Cube, including the seven in the ring plus three profiles to the east and south of that ring, is 8.9×10^{-4} m s^{-1}. The DPIL K values show much better agreement with this pumping test K estimate than do the flowmeter K values.

5.3. Comparison of Semivariograms

[32] Despite the differences in mean and variance between the flowmeter and DPIL K data, their semivariograms (Figure 6) are strikingly similar, indicating that both methods are conveying the same information regarding the K autocorrelation structure. The vertical semivariograms were computed using a lag spacing equal to the typical sample spacing for each profile type, 0.15 m for the flowmeter data and 0.015 m for the DPIL data. Following *Rehfeldt et al.* [1992], we computed horizontal semivariograms using a lag spacing of 5 m with a vertical tolerance angle of $\pm 5^\circ$ and vertical bandwidth of ± 0.16 m. In Figure 6 the semivariograms are displayed in scaled form, each divided by the global ln$K$ variance for the respective data set. For a stationary property, dividing by the global variance would scale the semivariogram to a unit sill. The scaled flowmeter and DPIL semivariograms do indeed appear to reach a unit sill before diverging at larger lags.

[33] A semivariogram is not influenced by the global mean. As a result, the DPIL ln$K$ semivariogram is independent of the calibration parameter b. Furthermore, scaling by the variance removes the dependence on a. Thus, the scaled semivariogram is independent of the calibration parameters. That is, it is the same as the scaled semivariogram of the "raw" DPIL ln(Q/P) values. This scaled semivariogram is therefore not impacted by errors or uncertainties in the calibration process.

[34] For both profile types, there are "nonideal" aspects of data distributions (e.g., lower limit of flowmeter values or upper limit of DPIL values). We explored the impact of these aspects by computing semivariograms of the normal score transforms of both the DPIL and the flowmeter data. The normal score transform (see Appendix) replaces a data set with one whose univariate distribution is perfectly normal but which has the same spatial configuration as the original data set with regard to ranks (sorting order) of the data values. For both profile types, the semivariograms of the normal score transformed data are essentially identical to the scaled semivariograms shown in Figure 6, indicating that the nonideal aspects of the data distributions have minimal impact on the computed semivariograms.

[35] The vertical semivariograms of the flowmeter and DPIL data, which are both well defined due to the regular sample spacing, are nearly identical for lags up to about 4.5 m, the distance at which both semivariograms reach the global variance value (and the practical range of both model semivariograms). The semivariograms diverge notably beyond 4.5 m, with the DPIL semivariogram increasing and the flowmeter semivariogram decreasing. *Rehfeldt et al.* [1992] interpreted this decrease as a possible indication of cyclicity. Indeed, the composite flowmeter profile (Figure 4) could be interpreted as cyclical in the sense that it shows a minimum at about 8.5 m in depth, and increases in value both above and below that depth. The composite DPIL profile, on the other hand, indicates a stronger K contrast between the upper and lower portions of the aquifer, with a steadily decreasing trend below 6.5 or 7 m depth. The increasing semivariance beyond a lag of 4.5 m reflects this trend. The horizontal semivariograms are less well defined due to the relative sparseness and irregularity of the sampling in the horizontal direction. Nevertheless, they are also similar, overall.

[36] Following *Rehfeldt et al.* [1992], we fit exponential models to the empirical semivariograms for the flowmeter and DPIL data, with confidence intervals on the estimated correlation lengths, by setting the sill values to the estimated ln$K$ variance or its lower or upper confidence limit, and then adjusting the vertical and horizontal correlation lengths in each case to produce the best match with the empirical semivariogram with the model sill fixed at the specified variance.

Table 2 shows the estimated correlation lengths, with their 95% confidence intervals. Both profile types yield the same estimate, 1.5 m, for the vertical correlation length. The horizontal correlation length estimates, 12.3 m for the flowmeter data and 10.2 m for the DPIL data, are similar relative to the uncertainties in these estimates.

5.4. Comparison of Interpolated Values

[37] Kriging interpolations were performed for both the flowmeter and DPIL data along the three cross sections shown in Figure 1 to compare K values at common

Table 2. Vertical and Horizontal Correlation Lengths of Exponential Models Fit to Flowmeter and DPIL Semivariograms, in Bold, With 95% Confidence Intervals

Type	Vertical Correlation Length (m)	Horizontal Correlation Length (m)
Flowmeter	1.1; **1.5**; 2.0	7.8; **12.3**; 17.5
DPIL	1.0; **1.5**; 2.0	6.2; **10.2**; 14.6

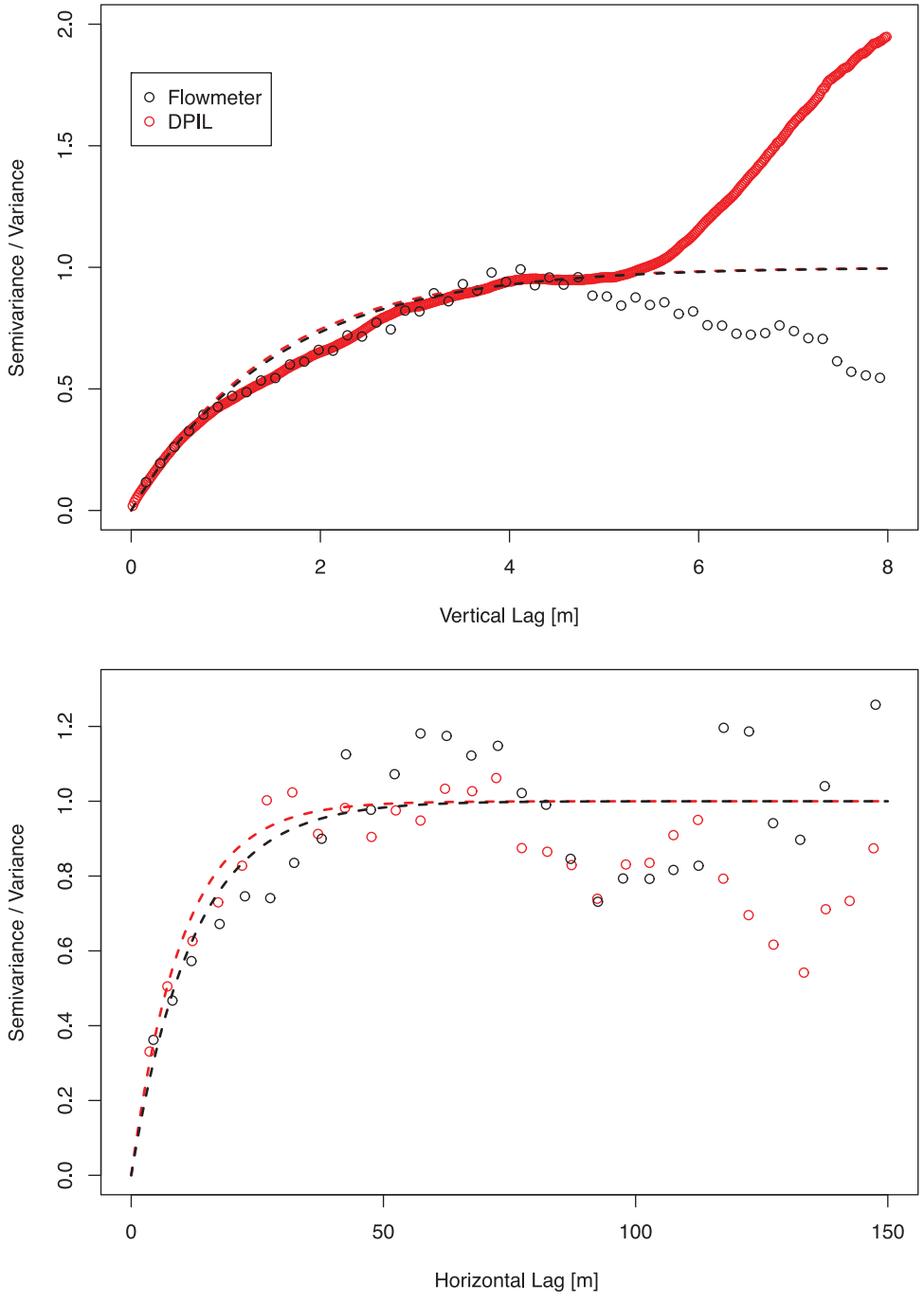

Figure 6. Flowmeter (black) and DPIL (red) ln$K$ empirical semivariograms (points), each scaled to respective global variance, with fitted models (dashed lines).

locations and reduce the impacts of differences in spatial sampling distributions. The interpolated values in each case are based on kriged ln$K$ values, evaluated over a grid of cells on each cross section, which are then recast in terms of log10(K) for ease of interpretation (Figure 7). Ordinary kriging was performed with the GSLIB program kt3d [*Deutsch and Journel*, 1998], using the semivariogram model parameters in Tables 1 (variances = semivariogram sills) and 2 (correlation lengths). The kriging standard deviations (Figures 7c and 7d) reflect the proximity to data (flowmeter or DPIL profile locations) along each cross section, providing a sense of the relative accuracy of the interpolation in different regions.

[38] There is roughly one order of magnitude contrast between the interpolated flowmeter and DPIL data; the geometric mean K is 7.2×10^{-5} m s^{-1} for the interpolated flowmeter values and 6.0×10^{-6} m s^{-1} for the interpolated DPIL values. The ln$K$ variances of the interpolated values are also markedly different, 1.7 for the flowmeter data and 5.4 for the DPIL data. Although one might expect some reduction in variance relative to the original data, due to the smoothing inherent to the interpolation process, the contrast in variance between the interpolated and original flowmeter data (4.4) is striking. This is probably attributable in part to differences in areal coverage: the cross sections have been designed to represent the regions occupied

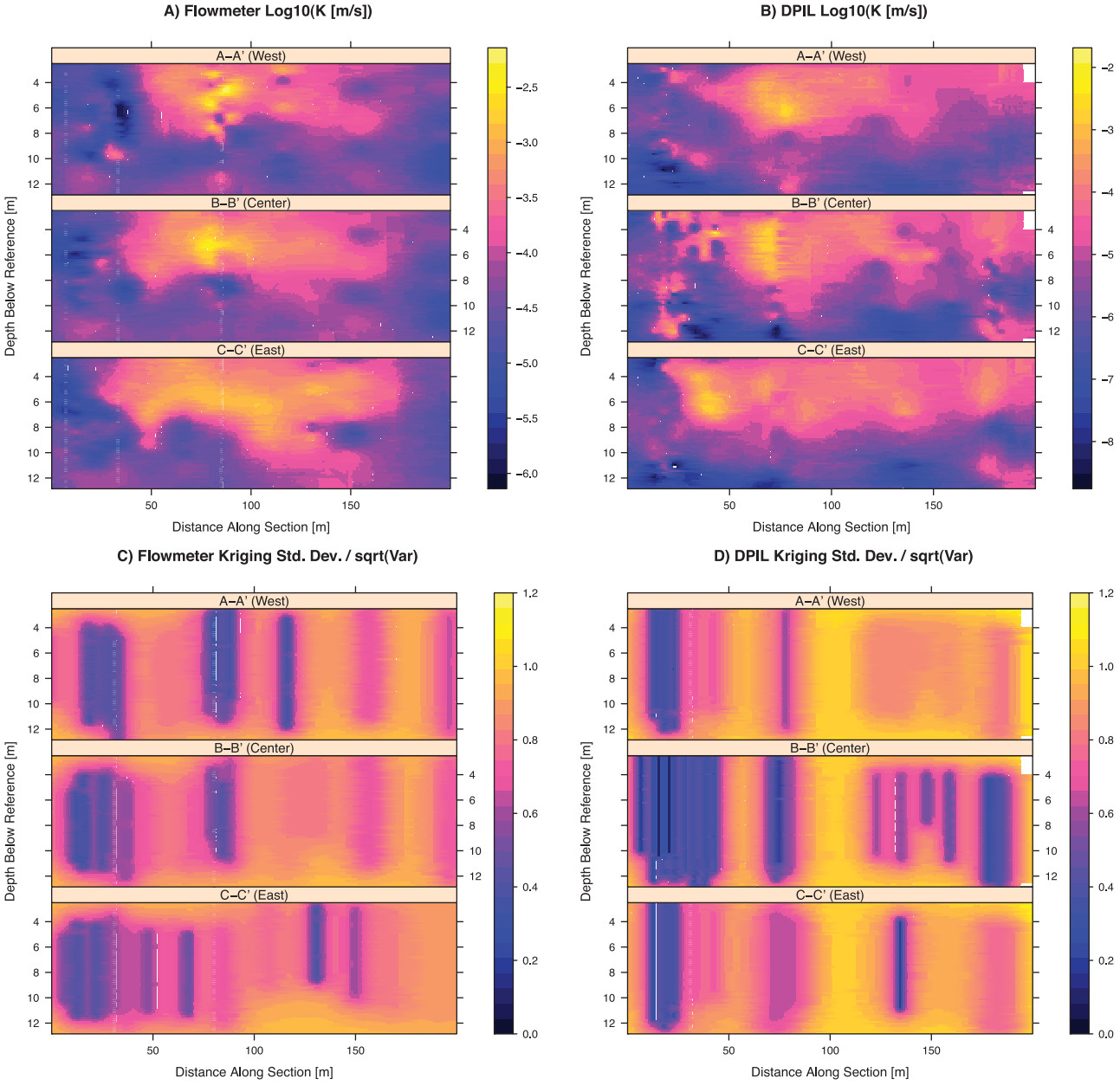

Figure 7. Log10(K [m s^{-1}]) values kriged to three cross sections shown on the map in Figure 1 for (a) flowmeter and (b) DPIL data, along with kriging standard deviation scaled by global standard deviation for (c) flowmeter and (d) DPIL data. In (a) and (b), color scale is centered at mean log10(K) value of respective interpolated data set and spans 3.5 standard deviations to either side of mean.

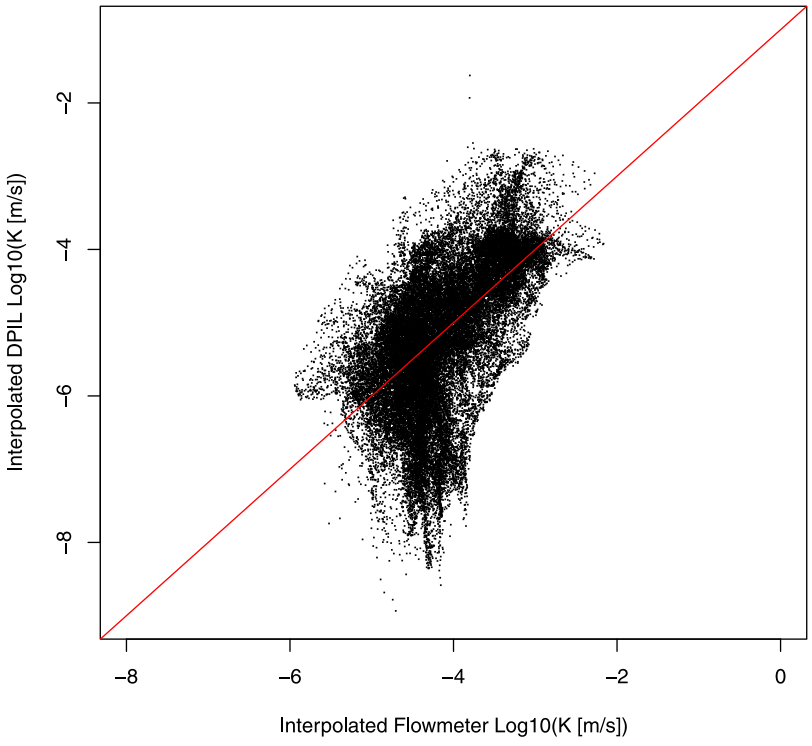

Figure 8. Crossplot of interpolated DPIL and flowmeter logarithmic K values shown in Figure 7. The red line represents DPIL log10(K) = flowmeter log10(K) − 1.

by DPIL profiles and consequently are not representative of the entire area covered by flowmeter profiles. Clearly, despite the differences in mean and variance, the large-scale patterns of K variation revealed by the interpolated DPIL and flowmeter data are similar. A crossplot of the two interpolated fields (Figure 8) helps clarify the relationship between them. The correlation between the two interpolated logarithmic K fields is 0.58 and the bulk of the data follow a trend represented by DPIL log10(K) = flowmeter log10(K) – 1, which is essentially identical to the best-fit linear regression of the interpolated DPIL versus interpolated flowmeter values (a regression that is highly significant, even though it only "explains" 33% of the variation in the DPIL data). Apart from the overall shift in the mean, the most notable discrepancy between the two fields is that the DPIL data indicate a stronger contrast between the upper and lower portions of the aquifer, with a more continuous low-K region in the lower portion of the aquifer and with the high-K zone in the upper portion extending further north than indicated by the flowmeter data. Both profile types indicate that the source area, at the south end of the profiles, has a significantly lower K than regions further north and both show the highest K values in the vicinity of the MLS Cube.

6. Summary and Conclusions

[39] Direct-push (DP) technology has been widely utilized to characterize shallow (depths <30 m) unconsolidated formations [*Dietrich and Leven*, 2006; *McCall et al.*, 2005]. Recently, a new series of direct-push tools has been developed for the high-resolution characterization of hydraulic conductivity (K) variation in unconsolidated aquifers [*Liu et al.*, 2012]. In this work we focus on two tools: the direct-push injection logger (DPIL), which provides profiles of relative variations in K at a vertical resolution as fine as 0.015 m; and the HRK, which incorporates the DPIL and the direct-push permeameter (DPP) into a single tool. The DPP provides reliable K estimates over a 0.4 m vertical interval and its incorporation in the HRK tool allows the high-resolution DPIL data to be transformed into K estimates. The effectiveness of these tools was demonstrated through a field comparison at the extensively studied MADE site.

[40] In this study, the direct-push profiling data were statistically analyzed and compared with previously collected borehole flowmeter data. Our results indicate that the direct-push profiles and the flowmeter profiles provide similar information about the general structure of the K field. Despite an overall shift of roughly one half to one order of magnitude between the K distributions estimated from flowmeter and direct-push profiling, the two techniques yield similar depictions of the autocorrelation structure (Figure 6) and the large-scale distribution of K at the MADE site (Figure 7). After the overall difference in central tendency, the most notable difference between the two distributions is that the DPIL data indicate a stronger and more laterally persistent contrast between the higher-K, shallower portion of the aquifer and the lower-K, deeper portion than do the flowmeter data. This difference is reflected in the differing behaviors of the vertical semivariograms at large lags, with the flowmeter semivariogram decreasing and the DPIL semivariogram increasing for vertical lags beyond 4.5 m (Figure 6), and in the univariate distributions (Figure 5), with the DPIL data exhibiting, in particular, a heavier lower tail than the flowmeter distribution.

[41] Based on the comparison of the composite flowmeter, DPIL, and DPP profiles (Figure 4), it seems likely that the offset between flowmeter and direct-push K estimates is at least in part a result of opposite biases in the measurement techniques compounded by the vertical structure of the K distribution: the DPIL profiles underestimate the high K values in the upper portion of the aquifer and the flowmeter profiles appear to overestimate the low-K values in the lower portion of the aquifer. Nevertheless, arithmetically averaged DPIL K values are similar to K values estimated from two pumping tests at the site, providing evidence that the DPIL K estimates are of a reasonable magnitude. In contrast, the average K value estimated from flowmeter profiles in the vicinity of one of these tests is notably higher than the pumping test K. The comparison between pumping-test and flowmeter K values for a second test is less conclusive, due to a lack of flowmeter profiles in the immediate vicinity of that test, but also hints at the possibility that the flowmeter K values are higher than expected based on the pumping test results.

[42] The direct-push profiling methods used in this study can characterize highly heterogeneous aquifers at a speed and resolution that has not previously been possible. A profile through the 12 m thick MADE aquifer can be completed in 1 to 2 h, a small fraction of the time needed to obtain a flowmeter profile, when the drilling, installation, and development of the required wells are considered. Although the K range of the DPIL tool used in this work was limited in the high-K intervals, we are working on modifying tool designs and field protocols to extend the range of reliably measureable K values (e.g., *Liu et al.* [2012]). In addition, we are investigating a calibration procedure that would accommodate nonlinearity in the relationship between ln$K$ and the logarithm of the DPIL flux to pressure ratio in an adaptive fashion, using a nonparametric relationship that is developed during the course of the inversion, similar to the approach described by *Bohling* [2008].

[43] For over a quarter century, groundwater hydrologists have recognized that spatial variations in hydraulic conductivity provide critical controls on solute transport in subsurface flow systems. However, the field characterization of such variations at the resolution needed for effective transport models and remediation system designs required an investment of time, labor, and money that is not feasible for the vast majority of site investigations. Thus, the practical impact of a large body of work in the research community was limited. The new direct-push profiling methods used here enable high-resolution K data to be acquired in unconsolidated settings in a practically feasible manner. These methods, particularly when combined with surface geophysics (e.g., *Dogan et al.* [2011]; *Schmelzbach et al.* [2011]), provide a new means to increase the practical impact of the vast body of previous research on subsurface flow and transport (e.g., *Lessoff et al.* [2010]).

Appendix: Statistical Methods

[44] Apart from the ordinary kriging, which was performed using the GSLIB program kt3d [*Deutsch and Journel*, 1998], the statistical and geostatistical computations in this study were performed in the R statistical computing environment (http://www.r-project.org).

[45] The kernel density estimates shown in Figure 5a were computed by R's *density* function using Gaussian smoothing kernels. The default bandwidths (standard deviation of the kernel) computed by the *density* function using a rule based on the number of data [*Silverman*, 1986] are 0.13 log10(K) units for the DPIL data and 0.17 for the flowmeter data. For consistency in the comparison, we used the average of these two defaults, 0.15, for both the flowmeter and DPIL data. This was too small a bandwidth for the considerably less abundant DPP data (resulting in an overly noisy density curve), so in that case we used the default bandwidth of 0.4.

[46] A normal QQ plot (Figure 5b) shows sorted data values plotted versus the standard normal quantiles evaluated at the empirical probability level associated with each data point [$p_i = (i - 0.5)/n$, where i is the index of the data point in the sorted list and n is the number of data points]. A normally distributed data set will plot roughly as a straight line on a normal QQ plot, while two data sets following similarly shaped non-normal distributions will exhibit similar deviations from linearity. A concept closely related to the normal QQ plot is the normal score transform [*Deutsch and Journel*, 1998], in which each point in a data set is replaced by the standard normal deviate associated with the data point's empirical probability level, resulting in a data set whose univariate distribution is a perfect standard normal, but whose basic spatial configuration, in terms of locations of the ranked values, is the same as the original data set. Although not shown in this paper, we computed semivariograms of normal-score transformed versions of the flowmeter and DPIL datasets to assess the impact of nonidealities (limits, spurious values) on the semivariograms. The normal-score transformed semvariograms are essentially identical to the variance-scaled semivariograms of the original data (Figure 6), indicating that these nonidealities have little impact on the assessment of the spatial autocorrelation structure.

[47] The geostatistical computations presented in this study were performed using the gstat [*Pebesma*, 2004] addin for R. For the sake of consistency we followed the protocols described by *Rehfeldt et al.* [1992] for computing empirical semivariograms. The vertical semivariograms were computed using a lag spacing equal to the typical sample spacing for each profile type, 0.15 m for the flowmeter data and 0.015 m for the DPIL data. Following *Rehfeldt et al.* [1992], we computed horizontal semivariograms using a lag spacing of 5 m with a vertical tolerance angle of $\pm 5^\circ$ and vertical bandwidth of ± 0.16 m. The resulting flowmeter semivariograms are essentially the same as those presented by *Rehfeldt et al.* [1992], indicating that any differences due to computation procedures or to use of the additional 18 flowmeter wells [*Boggs et al.*, 1993] are minor.

[48] We followed the procedures described by *Rehfeldt et al.* [1992] to compute 95% confidence intervals on the estimated ln$K$ means and variances for each data type accounting for the impact of spatial autocorrelation. The variance of the mean estimate $\overline{x}$ is given by

$$\mathrm{Var}[\overline{x}] = \frac{s^2}{N^*}, \tag{A1}$$

where s^2 is the estimated sample variance, and the variance of s^2 is

$$\mathrm{Var}[s^2] = \frac{2s^4}{N^*}, \tag{A2}$$

where N^* represents the number of independent data values. As discussed by *Rehfeldt et al.* [1992], the number of independent data values is less than the total number of data N due to spatial autocorrelation. They estimated N^* from

$$N* = \frac{N}{M} = \frac{N}{M_v M_h}, \tag{A3}$$

where M is a correction factor to account for spatial correlation of the observations, given as the product of vertical and horizontal correction factors. *Rehfeldt et al.* [1992] estimated the vertical correction factor from the ratio of the vertical correlation length (λ_v) to vertical sample spacing (Δ_v):

$$M_v = \frac{\lambda_v}{\Delta_v}. \tag{A4}$$

[49] Both the flowmeter and DPIL data indicate a vertical correlation length of about 1.5 m, resulting in $M_v = 10$ (*Rehfeldt et al.* [1992] used $M_v = 9.8$) for the flowmeter data, sampled at 0.15 m intervals, and $M_v = 100$ for the DPIL data, sampled at 0.015 m intervals. The DPP tests were performed at varying vertical spacings, but have a median spacing of 1.2 m, corresponding to a vertical correction factor of $M_v = 1.25$.

[50] *Rehfeldt et al.* [1992] estimated the horizontal correction factor from

$$M_h = \frac{W_t}{W_u}, \tag{A5}$$

where W_t is the total number of wells and W_u is the number of well clusters. A well cluster is a set of wells separated by less than the horizontal correlation length λ_h; any well that is further than λ_h from any other well is a one-well cluster. Essentially, the horizontal correction factor states that each well cluster should be counted as a single well in terms of the amount of information it contributes regarding horizontal variation. The flowmeter and DPIL data indicate horizontal correlation lengths of 12.3 and 10.2 m. For the sake of identifying well clusters, we will use a threshold distance of 10 m. Doing so yields an estimate of approximately 42 clusters for the 67 flowmeter wells ($M_h = 1.6$; *Rehfeldt et al.* [1992] used $M_h = 1.7$), 22 clusters for the 58 DPIL (HRK and DPIL-only) profiles ($M_h = 2.6$), and 7 clusters for the 21 HRK profiles ($M_h = 3.0$). These numbers reflect the reduced lateral coverage of the direct-push profiles relative to the flowmeter profiles and it is this difference, primarily, that results in a lower number of independent direct-push data than flowmeter data.

[51] Following *Rehfeldt et al.* [1992], we calculated model semivariograms for the flowmeter and DPIL data, with confidence intervals on the estimated correlation lengths, by setting the sill values to the estimated $\ln K$ variance or its lower or upper confidence limit, and then adjusting the vertical and horizontal correlation lengths in each case to produce the best match between the empirical and model semivariograms with the sill fixed at the specified variance. Correlation lengths were estimated using the nonlinear regression procedure implemented in the *fit.variogram* function in gstat.

[52] **Acknowledgments.** This work was supported by the Hydrologic Sciences Program of the National Science Foundation (grants EAR-0738955 and 0738938). Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the NSF. We thank Thomas Vienken, Uwe Schneidewind, Ludwig Zschornak, and Helko Kotas of UFZ-Leipzig, Geoffrey Tick of the Univ. of AL, and Shane Reed of Columbus Air Force Base for their assistance. We also thank Geoprobe Systems for supporting this work through technical assistance. Finally, we thank the anonymous reviewers for their thoughtful comments, which helped us to improve the manuscript considerably.

References

Barlebo, H. C., M. C. Hill, and D. Rosbjerg (2004), Investigating the macrodispersion experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model, *Water Resour. Res.*, *40*(4), W04211, doi:10.1029/2002WR001935.

Boggs, J. M., S. C. Young, and L. M. Beard (1992), Field study of dispersion in a heterogeneous aquifer, 1, Overview and site description, *Water Resour. Res.*, *28*(12), 3281–3291.

Boggs, J. M., L. M. Beard, S. E. Long, M. P. McGee, W. G. MacIntyre, C. P. Antworth, and T. B. Stauffer (1993), *Database for the Second Macrodispersion Experiment (MADE-2)*, *Tech. Rep. TR-102072*, Electric Power Res. Inst., Palo Alto, California.

Bohling, G. C. (2008), Information fusion in regularized inversion of tomographic pumping tests, in *Quantitative Information Fusion for Hydrol. Sci. (Studies in Computational Intelligence, Volume 79)*, edited by X. Cai and T.-C. J. Yeh, pp. 137–162, Springer, Berlin.

Bohling, G. C., and J. J. Butler Jr. (2001), Lr2dinv: A finite-difference model for inverse analysis of two-dimensional linear or radial groundwater flow, *Comput. Geosci.*, *27*, 1147–1156.

Butler, J. J., Jr. (2005), Hydrogeological methods for estimation of hydraulic conductivity, in *Hydrogeophysics*, edited by Y. Rubin and S. Hubbard, pp. 23–58, Springer, The Netherlands.

Butler, J. J., Jr., P. Dietrich, V. Wittig, and T. Christy (2007), Characterizing hydraulic conductivity with the direct-push permeameter, *Ground Water*, *45*(4), 409–419, doi:10.1111/j.1745-6584.2007.00300.x.

Dagan, G. (1989), *Flow and Transport in Porous Formations*, 465 p., Springer, New York.

Dagan, G., and S. P. Neuman (1997), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge Univ. Press, Cambridge, UK.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford University Press, New York.

Dietrich, P., and C. Leven (2006), Direct push technologies, in *Groundwater Geophysics*, edited by R. Kirsch, pp. 321–340, Springer, Berlin.

Dietrich, P., J. J. Butler Jr., and K. Faiß (2008), A rapid method for hydraulic profiling in unconsolidated formations, *Ground Water*, *46*(2), 323–328, doi:10.1111/j.1745-6584.2007.00377.x.

Dogan, M., R. L. Van Dam, G. C. Bohling, J. J. Butler Jr., and D. W. Hyndman (2011), Hydrostratigraphic analysis of the MADE site with full-resolution GPR and direct-push hydraulic profiling, *Geophys. Res. Lett.*, *38*, L06405, doi:10.1029/2010GL046439.

Eggleston, J., and S. Rojstaczer (1998), Identification of large-scale hydraulic conductivity trends and the influence of trends on contaminant transport, *Water Resour. Res.*, *34*(9), 2155–2168.

Fogg, G. E., S. F. Carle, and C. Green (2000), Connected network paradigm for the alluvial aquifer system, in *Theory, Modeling, and Field Investigation in Hydrogeology: A Special Volume in Honor of Shlomo P. Neuman's 60th Birthday*, edited by D. Zhang and C. L. Winter, *Geol. Soc. Am. Spec. Pap., 348*, 25–42.

Freyberg, D. L. (1986), A natural gradient experiment on solute transport in a sand aquifer: 2, Spatial moments and the advection and dispersion of nonreactive tracers, *Water Resour. Res.*, *22*(13), 2031–2046.

Garabedian, S. P., D. R. LeBlanc, L. W. Gelhar, and M. A. Celia (1991), Large-scale natural gradient tracer test in sand and gravel, Cape Cod,

Massachusetts: 2, Analysis of spatial moments for a nonreactive tracer, *Water Resour. Res.*, *27*(5), 911–924.

Gelhar, L. W. and C. L. Axness (1983), Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resour. Res.*, *19*(1), 161–180.

Hess, K. M., S. H. Wolf, and M. A. Celia (1992), Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 3, Hydraulic conductivity variability and calculated macrodispersivities, *Water Resour. Res.*, *28*(8), 2011–2027.

Hill, M. C., H. C. Barlebo, and D. Rosbjerg (2006), Reply to comment by F. Molz et al. on "Investigating the macrodispersion experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model," *Water Resour. Res.*, *42*, W06604, doi:10.1029/2005WR004624.

Killey, R. W. D., and G. L. Moltyaner (1988), Twin Lake tracer tests: Methods and permeabilities, *Water Resour. Res.*, *24*(10), 1585–1613.

Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, *32*(9), 2617–2658.

LeBlanc, D. R., S. P. Garabedian, K. M. Hess, L. W. Gelhar, R. D. Quadri, K. G. Stollenwerk, and W. W. Wood (1991), Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 1, Experimental design and observed tracer movement, *Water Resour. Res.*, *27*(5), 895–910.

Lessoff, S. C., U. Schneidewind, C. Leven, P. Blum, P. Dietrich, and G. Dagan (2010), Spatial characterization of the hydraulic conductivity using direct-push injection logging, *Water Resour. Res.*, *46*, W12502, doi:10.1029/2009WR008949.

Liu, G., G. C. Bohling, and J. J. Butler Jr. (2008), Simulation assessment of the direct-push permeameter for characterizing vertical variations in hydraulic conductivity, *Water Resour. Res.*, *44*, W02432, doi:10.1029/2007WR006078.

Liu, G., J. J. Butler Jr., G. C. Bohling, E. Reboulet, S. Knobbe, and D. W. Hyndman (2009), A new method for high-resolution characterization of hydraulic conductivity, *Water Resour. Res.*, *45*, W08202, doi:10.1029/2009WR008319.

Liu, G., J. J. Butler Jr., E. C. Reboulet, and S. Knobbe (2012), Hydraulic conductivity profiling with direct-push methods, *Grundwasser*, doi:10.1007/s00767-011-0182-9, in press.

Lowry, W., N. Mason, V. Chipman, K. Kisiel, and J. Stockton (1999), In-situ permeability measurements with direct push techniques: Phase II topical report, *Rep. SEASF-TR-98-207*, 102 pp., Fed. Energy Technol. Cent., Dep. of Educ., Morgantown, WV.

MacKay, D. M., D. L. Freyberg, P. V. Roberts, and J. A. Cherry (1986), A natural gradient experiment on solute transport in a sand aquifer: 1, Approach and overview of plume movement, *Water Resour. Res.*, *22*(13), 2017–2029.

McCall, W., D. M. Nielsen, S. Farrington, and T. M. Christy (2005), Use of direct-push technologies in environmental site characterization and ground-water monitoring, in *The Practical Handbook of Environmental Site Characterization and Ground-Water Monitoring*, 2nd ed., edited by D. M. Nielsen, pp. 345–472, CRC, Boca Raton, FL.

McCall, W., T. M. Christy, T. Christopherson, and H. Isaacs (2009), Application of direct push methods to investigate uranium distribution in an alluvial aquifer, *Ground Water Monitoring Remediation*, *29*(4), 65–76.

Molz, F. J., O. Guven, J. G. Melville, R. D. Crocker, and K. T. Matteson (1986), Performance, analysis, and simulation of a two-well tracer test at the Mobile site, an examination of scale-dependent dispersion coefficients, *Water Resour. Res.*, *22*(7), 1031–1037.

Molz, F. J., C. Zheng, S. M. Gorelick, and C. F. Harvey (2006), Comment on "Investigating the Macrodispersion Experiment (MADE) site in Columbus, Mississippi, using a three-dimensional inverse flow and transport model" by H. C. Barlebo, Mary C. Hill, and D. Rosbjerg, *Water Resour. Res.*, *42*, W06603, doi:10.1029/2005WR004265.

Pebesma, E. J. (2004), Multivariable geostatistics in S: The gstat package, *Comput. Geosci.*, *30*, 683–691.

Rehfeldt, K. R., P. Hufschmied, L. W. Gelhar, and M. E. Schaefer (1989), *Measuring Hydraulic Conductivity With the Borehole Flowmeter*, *Topical Report EPRI EN-6511*, Electric Power Research Institute, Palo Alto, CA.

Rehfeldt, K. R., J. M. Boggs, and L. W. Gelhar (1992), Field study of dispersion in a heterogeneous aquifer, 3, Geostatistical analysis of hydraulic conductivity, *Water Resour. Res.*, *28*(12), 3309–3324.

Schmelzbach, C., J. Tronicke, and P. Dietrich (2011), Three-dimensional hydrostratigraphic models from ground-penetrating radar and direct-push data, *J. Hydrol.*, *398*, 235–245, doi:10.1016/j.hydrol.2010.12.023.

Silverman, B. W. (1986), *Density Estimation*, Chapman and Hall, London.

Sudicky, E. A. (1986), A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process, *Water Resour. Res.*, *22*(13), 2069–2082.

Whittaker, J., and G. Teutsch (1999), Numerical simulation of subsurface characterization methods: Application to a natural aquifer analogue, *Adv. Water Resour.*, *22*(8), 819–829.

Zheng, C. (2006), Accounting for aquifer heterogeneity in solute transport modeling: A case study from the macrodispersion experiment (MADE) site in Columbus, Mississippi, in *Handbook of Groundwater Engineering*, 2nd ed., edited by J. W. Delleur, Chap. 26, pp. 1–18, CRC, Boca Raton, FL.

Zheng, C., M. Bianchi, and S. M. Gorelick (2011), Lessons learned from 25 years of research at the MADE site, *Ground Water*, *49*(5), 649–662.

G. C. Bohling, J. J. Butler Jr., S. J. Knobbe, G. Liu, and E. C. Reboulet, Kansas Geological Survey, University of Kansas, 1930 Constant Ave., Lawrence, KS 66047, USA. (geoff@kgs.ku.edu)

P. Dietrich, Department of Monitoring and Exploration Technologies, UFZ, Helmholtz Center for Environmental Research, Permoserstrasse 15, 04318 Leipzig, Germany.

D. W. Hyndman, Department of Geological Sciences, Michigan State University, 206 Natural Science Building, East Lansing, MI 48824, USA.