AGU PUBLICATIONS

Water Resources Research

COMMENTARY

10.1002/2016WR019699

Key Points:

- Stochastic hydrogeology has overrelied on hydrologic measurements of K, neglecting low-K media that are inherent to geologic systems
- Subsurface characterization must merge hydrologic and geologic data to produce realistic, reliable characterizations
- Existing nonlocal theories of transport are not yet capable of modeling processes that are most relevant to groundwater quality management

Correspondence to:
G. E. Fogg,
gefogg@ucdavis.edu

Citation:
Fogg, G. E., and Y. Zhang (2016), Debates—Stochastic subsurface hydrology from theory to practice: A geologic perspective, *Water Resour. Res.*, *52*, doi:10.1002/2016WR019699.

Received 25 AUG 2016
Accepted 9 NOV 2016
Accepted article online 18 NOV 2016

Debates—Stochastic subsurface hydrology from theory to practice: A geologic perspective

Graham E. Fogg[1] **and Yong Zhang**[2]

[1]Department of Land, Air, and Water Resources, University of California, Davis, California, USA, [2]Department of Geological Sciences, University of Alabama, Tuscaloosa, Alabama, USA

Abstract A geologic perspective on stochastic subsurface hydrology offers insights on representativeness of prominent field experiments and their general relevance to other hydrogeologic settings. Although the gains in understanding afforded by some 30 years of research in stochastic hydrogeology have been important and even essential, adoption of the technologies and insights by practitioners has been limited, due in part to a lack of geologic context in both the field and theoretical studies. In general, unintentional, biased sampling of hydraulic conductivity (K) using mainly hydrologic, well-based methods has resulted in the tacit assumption by many in the community that the subsurface is much less heterogeneous than in reality. Origins of the bias range from perspectives that are limited by scale and the separation of disciplines (geology, soils, aquifer hydrology, groundwater hydraulics, etc.). Consequences include a misfit between stochastic hydrogeology research results and the needs of, for example, practitioners who are dealing with local plume site cleanup that is often severely hampered by very low velocities in the very aquitard facies that are commonly overlooked or missing from low-variance stochastic models or theories. We suggest that answers to many of the problems exposed by stochastic hydrogeology research can be found through greater geologic integration into the analyses, including the recognition of not only the nearly ubiquitously high variances of K but also the strong tendency for the good connectivity of the high-K facies when spatially persistent geologic unconformities are absent. We further suggest that although such integration may appear to make the contaminant transport problem more complex, expensive and intractable, it may in fact lead to greater simplification and more reliable, less expensive site characterizations and models.

1. Introduction

More than three decades of research in stochastic hydrogeology have contributed impressively to our understanding of the role of heterogeneity in flow and transport and how to better estimate or represent that heterogeneity through statistical and geological models. Most of the work on stochastic hydrogeology rests on a foundation of hydrogeologic data, especially hydraulic conductivity (K) measurements, and field experiments. Perhaps the most prominent among the field experiments are the Borden, Cape Cod, and MADE sites [*MacKay et al.*, 1986; *LeBlanc et al.*, 1991; *Boggs et al.*, 1992], all in clastic sedimentary geologic environments. Because the heterogeneity that gives rise to stochastic hydrogeology research is a direct consequence of geologic processes, it is important to ask the questions: “Do K measurements, especially those using various kinds of well tests, adequately represent the heterogeneity in K, and do the Borden, Cape Cod, and MADE sites represent a sufficiently broad cross section of geologic environments to allow much extrapolation from the experiments at these sites to groundwater phenomena at other sites?”

This commentary will address the two questions in the context of fundamental geologic concepts and characteristics of clastic sedimentary depositional systems in a hydrologic context. Because most of the stochastic hydrogeology literature has focused on clastic sedimentary aquifer settings, the scope of our comments will not include fractured rock or karst groundwater systems. We will close with suggestions on future research directions.

2. Sediment Texture and K

With the exception of eolian environments, clastic depositional systems deposit grain sizes that include clay, silt, sand, and gravel within the same formation. The spatial arrangement and geometry of these textural units depend on the particular fluvial, deltaic, shoreline, or marine depositional systems in which the sedimentary particles are eroded, transported, and deposited. Each depositional environment leads to predictable spatial patterns in textures or facies that can be used to great advantage in characterization of heterogeneity [*Galloway and Hobday*, 1983; *Fogg et al.*, 1998]. The range of textures stems directly from the fact that the water velocities in the depositional environment range from very low, such as on a floodplain ($10^{\circ}-10^{1}$ cm/s), to much higher, such as in a river channel ($10^{1}-10^{3}$ cm/s). For example, in a fluvial (riverine) depositional system, most of the deposition occurs during relatively rare, flood events in which high-velocity river flows move and deposit the coarse sand and/or gravel particles, while the overbank and floodplain portions of the system predominantly move and deposit the fines such as silt and clay. Exceptions can occur in glacio-fluvial outwash deposits, where glacial meltwater produces less flashy streamflow and hence less episodic deposition.

Given the range of water velocities during sediment deposition in geologic environments, most clastic sedimentary aquifer systems should be expected to contain significant fractions of fine sediments. Indeed, although proportions of coarse versus fine textures vary considerably, quite commonly, the aquifer systems contain more silt and clay beds than sand and gravel beds [*Fogg et al.*, 2000]. For example, it is typical for the geologic facies dominated by fines to make up 50–80% of the aquifer system [*Galloway and Hobday*, 1983; *Fogg et al.*, 1998; *Kaiser et al.*, 1978]. In other words, the fundamentals of earth surface processes, including the physics of sediment particle erosion, transport, and deposition, dictate that most clastic sedimentary deposits contain substantial fractions of silt and clay, and the subsurface geologic record unequivocally shows this.

Typical K ranges for clay and silt are $10^{-11}-10^{-5}$ cm/s and for sand and gravel are $10^{-7}-10^{-2}$ [*Domenico and Schwartz*, 1990]. Given that most clastic sedimentary depositional systems contain both fine and coarse textured sediments, it is not surprising that K within the same deposit can vary many orders of magnitude. For example, the presence of both clay and sand or gravel within the same deposit easily leads to K values that would range over 4–7 orders of magnitude. Although K in such deposits may be multimodal, let us assume that the spectrum of K values within a deposit are part of a lognormal distribution and calculate ln-K variances ($\sigma^2_{\ln K}$). Based on the simple approximation of variance ($((\text{max-min})/4)^2$) [*Triola*, 2010], where in this case max and min are ln-K values, it can be shown that

$$\sigma^2_{\ln K} \approx \frac{R^2}{3}, \tag{1}$$

where R is the range between the max and min values of K in orders of magnitude (base 10). For example, if the K varies over 3 orders of magnitude, $\sigma^2_{\ln K} \approx 3.0$, and if K varies over 5 or 7 orders of magnitude, $\sigma^2_{\ln K} \approx 8.3$ or 16.3, respectively. If one estimates variances assuming multimodal K distributions based on different modes for different facies, then the $\sigma^2_{\ln K}$ will likely be higher than estimates based on an assumption of normality because of greater weighting of the tails.

Accordingly, for most clastic sedimentary deposits, we should anticipate values of $\sigma^2_{\ln K}$ extending well above 3, with values exceeding 10 being highly likely because of the near ubiquity of fine-grained deposits in the geologic deposits that compose most aquifer systems. For example, *Fogg* [1986] measured K values of the Wilcox Group fluvial system using pumping tests and laboratory core analysis and found K values ranging from 10^{-6} to 20 m/d (1.2×10^{-11} to 2.3×10^{-4} m/s), and although he did not publish a $\sigma^2_{\ln K}$ value, his data show it to be approximately 17 based on equation (1). Fitting of two Gaussian distributions to *Fogg*'s [1986] bimodal K distribution leads to a composite $\sigma^2_{\ln K}$ of 18.6, so the $\sigma^2_{\ln K}$ of this very typical, fluvial depositional system is around 17–19. It is noteworthy that Fogg's measurements of K targeted all of the facies, not just the coarse facies. Had he used only hydrologic K data produced by well testing, which, for reasons discussed below, are generally not capable of representing the full range of K values, he would have estimated a $\sigma^2_{\ln K}$ value of about 0.9.

Similarly, *Fogg et al.* [1998] used pumping and slug test data from the alluvial aquifer system underlying Lawrence Livermore National Laboratory to estimate K values of fluvial channel, overbank (e.g., levee deposits) and debris flow ranging from about 0.2 to 56 m/d (2×10^{-6} to 6.5×10^{-4} m/s). Together, those data approximately represent a lognormal distribution [*Fogg et al.*, 1998, Figure 4] with $\sigma^2_{\ln K} \approx 2$; however, the abundant core data show that fully 56% of the system is composed of tight, floodplain silt and clay facies that were initially estimated to have a K value of 1.73×10^{-4} m/d (2×10^{-9} m/s) based on the literature and site data. Calibration of the 3-D heterogeneous model using pumping test data with seven observation wells, however, resulted in a lower estimated K of these silt and clay facies of 4.32×10^{-5} (m/d) (5×10^{-10} m/s), which would be consistent with a system $\sigma^2_{\ln K} \approx 12.4$. In the four-facies, multimodal indicator model of the system, however, the calculated $\sigma^2_{\ln K}$ is nearly 24 [*Lee et al.*, 2007; *Fogg et al.*, 2000] because the multimodality is represented in the model; i.e., Lee's $\sigma^2_{\ln K}$ is not representative of a unimodal Gaussian. The point is that whether one represents such a multifacies system as multimodal or unimodal, the $\sigma^2_{\ln K}$ is obviously much larger than 1 or 2.

3. K Variance and Sampling Bias in Stochastic Hydrogeology

Most of the stochastic hydrogeology literature is based on the foundational assumption that $\sigma^2_{\ln K} < 1$ or 2 [*Gelhar*, 1993; *Rubin*, 2003], which would appear to contradict the much higher variances that we expect based on knowledge of geologic sediments and their documented ranges in K values. We suggest that this apparent contradiction stems from five factors: (1) observation scale, (2) presence of high-energy depositional systems, (3) data from soils and not geologic sediments, (4) K measurement methods, and (5) investigation bias toward aquifer hydrology rather than aquifer-aquitard hydrology as originally suggested by *Neuman and Witherpoon* [1972].

3.1. Observation Scale

Horizontal dimensions of the textural units, or hydrofacies, laid down by clastic sedimentary depositional systems are similar to the dimensions of the geomorphic systems such as rivers, river meander belts, and delta distributaries. For example, the rivers that deposited the Wilcox Group of Texas were on the scale of the Mississippi River and hence the channel sands that make up the aquifer portions of the aquifer system are 10^3 m in width [*Kaiser et al.*, 1978]. In contrast, the rivers that deposited the alluvial aquifer systems on the west side of the Central Valley of California were much smaller, and hence widths of the channel sands are 10^2 m or less. Width of the channel deposits in Livermore Valley, an alluvial basin within the Coast Ranges, appears to be on the order of 10^1 m [*Fogg et al.*, 2000]. Many point source groundwater contamination investigations and field experiments are on the scale of 10^1-10^3 m. Accordingly, the observation scale may not be expansive enough to encompass the full heterogeneity of the geologic system, and in the view of the local site investigator, the "aquifer" can appear to be rather homogeneous. Indeed, the scale at which the assumption of homogeneity and low $\sigma^2_{\ln K}$ is most likely applicable is the very local scale, which is fine until the plume ventures further out into the geologic system. Nevertheless, most fluvial channel deposits are not individual channels, but rather amalgamations of channel point bars and near-channel overbank deposits and therefore commonly contain ample amounts of fines.

In other words, even within channel deposits the expected range in local K values is several orders of magnitude or more. An excellent example of this can be found at the MADE site [*Boggs et al.*, 1992; *Zheng et al.*, 2011], which sits within a Quaternary fluvial deposit that contains the full range of textures, including tight clays. Importantly, however, the entire experimental site is in a fluvial sand and gravel deposit that, contrary to frequent claims, is not "highly heterogeneous" at all because, unequivocally, it is a coarse-grained unit [*Zheng et al.*, 2011]. The MADE site sand unit, however, contains enough of a range in sand and gravel grain sizes and silt that one gets a somewhat higher $\sigma^2_{\ln K}$ than at Borden or Cape Cod ($\sigma^2_{\ln K} \approx 4.5$–6.6) [*Rehfeldt et al.*, 1992; *Fiori et al.*, 2013].

The above discussion supports a geologic explanation for the observed growth in dispersivity with scale [*Neuman*, 1990; *LaBolle and Fogg*, 2001]. That is, growth of dispersivity is consistent with plumes starting out within an individual hydrofacies, such as a channel sand where $\sigma^2_{\ln K}$ is low, but increasingly encountering the other hydrofacies of the geosystem, including silt and clay dominated facies. In essence, the $\sigma^2_{\ln K}$ contacted by the plume grows with time and distance of transport, easily reaching values on the order of 5–15. A prime example of the presence of finer grained units just beyond the zone of measurement is the Borden site, where a mere 4 m below the very homogenous sand within which the Borden experiments were run [*Mackay et al.*, 1986, Figure 1] lies a unit described as "clay, silty, pebbly."

3.2. High-Energy Depositional Systems

Some clastic deposits are the product of large-scale, high-velocity (high-energy) fluid flow with an ample source of sediment supply. Geologic examples include eolian sands and glacio-fluvial outwash, wherein river flows are driven by relentless supply of water and sediment caused by melting of glaciers over thousands of years or more. Although not devoid of fines, the strong supply of water and sediment over long timescales can lead to substantial volumes of sand and gravel-rich deposits, and in some cases, relatively uniform sands. Hydrologic examples of glacio-fluvial outwash sands are the Borden and Cape Cod experimental sites [*Mackay et al.*, 1986; *LeBlanc et al.*, 1991] where $\sigma^2_{\ln K}$ estimates are low. Nevertheless, such high-energy depositional systems still commonly contain sediments that range from the very coarsest gravels to fine sand and silt, resulting in K values that range over 4–5 orders of magnitude. Regardless, any scientific conclusions drawn from experiments at relatively clean, sandy, homogeneous sites such as Borden and Cape Cod should be couched in terms of the glacio-fluvial outwash environment so that potential applicability to other aquifer environments, particularly those containing greater ranges in sediment sizes, can be ascertained.

3.3. Soils and Not Geologic Sediments

Soil is not the same as geologic sediment, even though the terms are often used interchangeably and geologic sediments and rocks are the parent materials for soils. Soils typically comprise the upper 1–2 m of the subsurface and have undergone hundreds to thousands of years of pedogenesis that includes physical, biological, and geochemical weathering. Without that pedogenesis, current levels of agro-ecosystem and natural ecosystem productivity would not have been possible.

The physical weathering includes infiltration of fine-grained particles during precipitation and overland flow, wherein a geologic deposit that starts out as a sand can evolve into a sandy loam that is lower in K because of the infiltrated fines [*Weissmann et al.*, 1999]. The biological weathering includes the growth of plant roots and action of burrowing by micro and macrofauna which can affect K of all the parent sediments, but most noteworthy, can create pathways through otherwise tight geologic silts and clays. The geochemical weathering includes dissolution and precipitation as well as kaolinization of feldspars, wherein feldspars, which are common in clastic sediments, start out as sand size particles, but through geochemical weathering are converted fully or partly to kaolinite clays that occlude the pore spaces of sand deposits. Accordingly, after the geologic sediment is deposited, pedogenesis can alter substantially the K of both the coarse and fine fractions, decreasing K of the former through infiltrated fines and weathering of sand-size particles to clay-size particles, and increasing K of the latter through biological activity. In other words, pedogenesis tends to reduce K of coarse-grained parent sediments and increase K of fine-grained parent sediments, in turn decreasing $\sigma^2_{\ln K}$ of the parent sediments in most geologic settings. This explains why studies of paleosols (paleo-soil profiles that are buried beneath more recent alluvial deposition) show that ancient soils derived from the same parent sediment as the overlying and underlying alluvial deposits act as aquitards with respect to the sandy portions of those alluvial aquifers [*Weissmann et al.*, 1999, 2002].

Many in the groundwater community apparently assume that because both soils and the deeper subsurface in clastic sedimentary systems consist of sediment particles, that they are the same thing. To the contrary, most sedimentary deposits that make up the vadose zone (below the soil zone) and aquifer systems were deposited and buried sufficiently rapidly in major flood events or glacial melt episodes, that they never underwent appreciable pedogenesis [*Galloway and Hobday*, 1983]. Therefore, we should anticipate that geologic sediments that are deposited as sands, gravels, silts, and clays have higher values of $\sigma^2_{\ln K}$ than soils because they have not undergone the pedogenesis processes that can reduce the high-end K's while increasing the low-end K's. In other words, $\sigma^2_{\ln K}$ values from soils are likely not comparable to those from geologic sediments.

3.4. K Measurement Methods Bias

The main methods of K measurement in hydrology are pumping tests, slug tests, borehole flowmeter tests, and laboratory permeameter tests. The producing interval of any well that can be pumped contains aquifer materials (e.g., sands or gravels), hence even if aquitard materials are present within the screened interval, the effective K will be dominated by the high-K materials [*Fogg*, 1986], and in turn, the $\sigma^2_{\ln K}$ of pumping test K values from any aquifer system will be biased low. Similarly, although slug tests are capable of measuring lower K values than are pumping tests, the resulting measured K values will be biased toward the highest-K media that intersect the well screen, and any $\sigma^2_{\ln K}$ based on slug test K data will be biased low. Borehole flowmeter tests are an admirable improvement on pumping and slug tests when one seeks to better represent the variability in local K along a borehole, but when such methods are applied in the presence of interbedded coarse and fine media, the borehole fluid velocities adjacent to the latter are too small to be resolved sufficiently to allow reliable K estimation [e.g., *Riva et al*., 2012]. In general, it appears that borehole flowmeter methods tend to overestimate the low K values and underestimate the high K values [e.g., *Whittaker and Teutsch*, 1999; *Barahona-Palomo et al*., 2011]. Consequently, borehole flowmeter data are also biased against detecting the full range in K values and tend to result in $\sigma^2_{\ln K}$ estimates that are biased low. Lastly, laboratory permeameter tests on core or repacked sediment samples are capable of measuring both high and low K values, but are seldom used in groundwater investigations. In reality, some combination of borehole and core permeameter methods would be needed to measure the full range of naturally occurring K values, yet in groundwater studies, the borehole methods tend to dominate. Core permeameter methods are common in soil studies, but as explained above, soils are not a good proxy for representation of heterogeneity in sedimentary clastic aquifer systems.

3.5. Investigation Bias Toward Aquifer Hydrology

Just as biologists are inclined to study biology, surface water hydrologists are inclined to study surface water, and aquifer hydrologists are inclined to study aquifer materials. Compared to aquitards, the aquifers can be probed and tested relatively easily via wells. Indeed, after performing one of the few studies of aquitard hydrology, including development of new techniques for field testing of aquitard properties, *Neuman and Witherspoon* [1972] called for the recognition of the important yet neglected role of aquitards in subsurface hydrology. Nevertheless, the difficulty of measuring K of aquitards and the inclination of groundwater hydrologists to study the aquifer portions of the subsurface even if they occupy a minority of the subsurface volume in many systems, has led to a tendency for field data collection campaigns that target aquifer material more than aquitard materials.

4. Joint Use of Conventional Hydrologic Measurements and Site Geology Is Essential

As outlined at the beginning of this paper, the geology and physics of sediment erosion, transport, and deposition dictates that subsurface systems will typically have broad assortments of grain sizes that lead to many orders of magnitude of variability of K within the same aquifer system. When we carefully account for K of both the aquifer materials and the nonaquifer materials that are inherent to our aquifer systems, values of $\sigma^2_{\ln K}$ that are dramatically larger than 2 or 3 are evident. This contradicts the pervasively low values of $\sigma^2_{\ln K}$ published in two excellent books on stochastic hydrogeology [*Gelhar*, 1993, Table 6.1; *Rubin*, 2003, Table 2.2]. A careful reading of the literature used for the data sources in those two tables, however, reveals the following.

Of the total of 45 $\sigma^2_{\ln K}$ values reported by *Gelhar* [1993] and *Rubin* [2003], about one third (16) are from borehole tests, another 12 are from soils, and another 7 are from studies that were targeted to specific facies either because the scale of the investigation was too small to encompass the other facies in the depositional system, or there was an inherent bias toward investigating the aquifer materials instead of aquitards. Thus, 35 of the 45 studies produce $\sigma^2_{\ln K}$ "measurements" that should be expected to be biased toward the low end. The remaining 10 include 1 case in an eolian sand, 3 cases in glacio-fluvial outwash, 2 cases where the methods of K measurement could not be determined with sufficient confidence, and 4 cases where none of the above sources of bias appeared to apply. In those latter 4 cases, two of the $\sigma^2_{\ln K}$ values were for alluvial materials and registered 2.7 and 4.5 at the MADE site [*Rehfeldt et al*., 1992], and the other two $\sigma^2_{\ln K}$ values were for volcanic ash flow tuff and registered 0.6 and 3.61 [*Istok et al*., 1994].

The above leads to the following, inevitable deductions. If one characterizes the distribution or variance of K in an aquifer system *solely through the use of borehole measurements of K or measurements of only the aquifer materials*, and without additional information from a sound geologic characterization of the site, the degree of heterogeneity is likely to be underestimated because (1) the borehole K measurements will typically reflect only the upper or central portions of the K distribution and (2) lack of knowledge about the existence of geologic materials that would not be represented in the borehole K measurements eliminates the possibility of correcting the bias of those measurements through targeted sampling and analysis of the unrepresented facies (e.g., aquitard lenses or layers). In other words, characterizing heterogeneity of a site solely with hydrologic methods and without geologic methods is like a blind man characterizing an elephant from the feel of its tusk—you are, unknowingly or not, sensing only a part of the reality and are not able to compensate for the incomplete picture.

5. Consequences

The consequences of ignoring or overlooking the geology and of overreliance on conventional K measurements of the aquifer (not aquitard) materials range from broad-based to specific. Broadly speaking, the development and promotion of oversimplified subsurface transport theories and models means that the scientific significance and transferability is much more limited than authors may suggest. Similarly, transferability of research results obtained at research field sites is limited by the character of the geology at those sites. For example, results from the Borden and Cape Cod sites are transferable to local-scale transport in coarse facies within glacio-fluvial outwash sediments, and results from the MADE site are transferable to local-scale transport in sandy facies in a fluvial system. By not framing the conclusions about the significance of the results in such a geologic context, one invites gross misconceptions about the nature of transport and real-world heterogeneity.

A hint of such transferability problems arose when researchers found that the macrodispersion theory which adequately modeled plumes at the Borden and Cape Cod sites was unable to reproduce important characteristics of the MADE plume. Unlike plumes in relatively homogeneous systems, the MADE plume exhibited considerable mass holdback near the source zone as well as more rapid downfield transport than predicted [*Zheng et al.*, 2011]. Such behavior, with heavy tails on both the proximal and distal portions of the plume, is entirely consistent with the connected-network paradigm [*Fogg et al.*, 2000] in which large $\sigma^2_{\ln K}$ within aquifer-aquitard complexes, together with good connectivity of the high-K materials [*Harter*, 2005; *Bianchi and Zheng*, 2016] produce plumes with similar, heavy tailed characteristics [*LaBolle and Fogg*, 2001].

Indeed, the estimates of $\sigma^2_{\ln K}$ at MADE are ~4.5–6.6 [*Rehfeldt et al.*, 1992; *Fiori et al.*, 2013], but interestingly, all the available texture data show that, unequivocally, the entire experimental zone at MADE is in a fluvial sand and gravel unit with only small percentages of fines [*Zheng et al.*, 2011; *Bianchi and Zheng*, 2016]. Moreover, laterally or vertically beyond this sand unit we should anticipate presence of the fine-grained fluvial facies, or aquitards, that most plumes will encounter as they migrate laterally or vertically beyond the 250 m long by ~10 m thick zone of this local experiment. In other words, although described as such in virtually every paper about MADE, the portion of the site where the experiments were conducted is unequivocally not "highly heterogeneous," as is very clear from knowledge of depositional systems in general, and the MADE site geology in specific. Further, the local alluvial deposit is not a major aquifer in the region, but is a very thin deposit that rests on much finer grained Cretaceous rocks [*U.S. Geological Survey*, 2010]. Just to the west of the MADE site, lie the massive fluvial-deltaic deposits of the Gulf of Mexico Basin, where sediments like those supporting the MADE experiments occur locally and form major aquifer systems, within which the silts and clays comprise the majority of the subsurface.

Why then do researchers refer to MADE as "highly heterogeneous?" This misconception appears to arise from their frame of reference, which is strongly influenced by the published $\sigma^2_{\ln K}$ values and experiences at the Borden and Cape Cod sites, which are actually at the "highly homogeneous" end of the spectrum. Clearly, as any plume at a site like MADE migrates beyond that local scale, the encountered heterogeneity would reflect much larger values of $\sigma^2_{\ln K}$, such as 10 or more. In this case to be consistent, the appropriate adjective would have to change from "highly heterogeneous" to "astronomically heterogeneous." We do not advocate such preposterous language, but we think it is important to realize that, in a geologic context, the MADE site is merely moderately heterogeneous. This in turn, is highly relevant to the meaning and transferability of all the exciting transport modeling theories that have been based on or inspired by MADE.

Is this misconception about MADE a problem? Yes, because there is now broad, tacit acceptance that MADE is highly heterogeneous, and considerable research has been devoted to development of transport theories capable of representing heavy-tailed plumes, for which MADE is seen as about as complex as it gets. If one assumes that MADE is highly heterogeneous, the next logical assumption is that any transport theories developed from it would represent the upper end of the complexity spectrum. Given, however, that migrating plumes would encounter still other hydrofacies and still greater variations in K, it is questionable whether the transport models capable of fitting or modeling the local-scale MADE plumes would be any more capable of making the jump to the next scale than the older macrodispersion theory was capable of representing observed MADE site conditions.

We find that many hydrogeologists, when faced with such challenges of complexity and scale as described above, jump to the conclusion that the characterization and transport prediction problems are intractable.

To the contrary, there is ample evidence from other hydrogeologic and petroleum investigations [*Galloway and Hobday*, 1983; *Fogg*, 1986; *Fogg et al*., 1998; *Bianchi and Zheng*, 2016] that as one moves beyond local scales, one's ability to recognize, detect, and predict the general distribution, arrangements, and connectivity of hydrofacies improves because the relevant textural units approach the scale at which geomorphic elements such as channel, overbank and floodplain deposits become recognizable components of readily observable, modern depositional systems.

6. Plume Prediction, Cleanup, Nonpoint Source, and Future Directions

Contaminant transport analysis in hydrogeology is important because we must predict plume migration, cleanup plumes, and predict effects of nonpoint-source contamination on basin-scale groundwater sustainability [*Fogg and LaBolle*, 2006]. Perspective on how stochastic hydrogeology could better address these three problems provides some direction for future work.

We suggest that measurement and modeling of plume movement on the 100 m scale is important and has resulted in tremendous advances in contaminant hydrogeology; however, predicting forward plume movement on this scale with sufficient accuracy to ascertain risks to human health and the environment is not a grand challenge, because plume prediction on the 100 m scale, especially within predominantly sandy facies, is not particularly difficult. Even with the MADE plumes, predictions of ultimate plume extents were sufficient for ascertaining general risks.

We further suggest that because the geologic variability sampled by plumes beyond the 10^2-10^3 m scale is dramatically more variable, that this scale of transport is both more important and more challenging. After all, that is the scale at which substantial growth in apparent dispersivity has been observed in measured and inferred data [*Neuman*, 1990], and it is evident to us that much of the above discussion provides fresh ground for reconciling such observations with the nature of the porous medium. Although the past 30 or so years of stochastic hydrogeology research have focused on the 10^2 m scale, we need to break out into the larger scale with field and modeling studies that encompass a fuller range of geologic variability.

Plume remediation and control is paramount, but as demonstrated in the most recent review of the state of the art [*National Research Council (NRC)*, 2012], despite billions in spending on site cleanup, precious little progress has been achieved even for conservative or nonreactive solutes. The mantra among field practitioners when decades of pump-and-treat shows little progress is something like: "The solutes diffuse into the low-K parts of the system, rendering pump-and-treat (PAT) much less effective." Unfortunately, this inefficiency of PAT has not been addressed through modeling, but instead, almost exclusively through trial-and-error field PAT activities. The reason for this is that almost no transport models of contamination sites have been constructed that are capable of simulating the typical, decades-long cleanup times for typical plumes on the 10^2-10^3 m scale. Why? We believe that this is because the models have not contained sufficiently broad ranges of K, including both advectively dominated aquifer facies and diffusion dominated aquitard facies, to reproduce the long PAT times. The only exception of which we are aware is *LaBolle and Fogg* [2001].

We assert that models capable of simulating site remediation for nonreactive as well as for reactive contaminants are needed, not only to ascertain how the process and costs can be optimized, but also to demonstrate to regulators those cases where cleanup times of decades or even centuries can be expected, and hence more cost-effective plume control measures should be implemented. In the context of the Borden, Cape Cod and MADE sites, the research community should be asking itself: "How do the experiments at these sites help advance the grand challenge of remediation?" Here again, the scale and geologic settings of those studies is limiting, because they all are in coarse sediments. That is, remediation on the 100 m scale in very to moderately uniform sand and gravel facies is not very difficult; hence, it does not appear to us that the conditions at those field sites would provide appreciable insight into solute transport when you reverse the boundary conditions from a case of forward plume movement to a case of PAT. In fact, we believe that none of the low to moderate $\sigma^2_{\ln K}$ models of these sites would have enough diffusion or slow advection-dominant media to generate long PAT times. This is yet another reason for stochastic hydrogeology to branch out into larger scales that encompass enough realistic geologic variability to match with real-world problems.

Lastly, let us point out that while point-source plumes are a definite threat to local groundwater sustainability, water quality in the entirety of many groundwater basins is in jeopardy because of nonpoint-source contamination [*Fogg and LaBolle*, 2006]. Addressing this problem requires basin-scale, groundwater quality management models. Unfortunately, no such reliable models yet exist because, as well demonstrated by much of the MADE site-related research, we are still struggling to model transport phenomena, including preferential flow in connected networks and mass exchange between aquitards and bona fide aquitards, at the 10^3-10^4 m scales. Indeed, regional-scale modeling of transport and groundwater quality management is likely the biggest stochastic contaminant hydrogeology challenge of all time!

7. Postscript

The diverse and rich perspectives provided by our colleagues in the companion articles of *Cirpka and Valocchi* [2016], *Fiori et al*. [2016], and *Sanchez-Vila and Fernandez-Garcia* [2016] have been illuminating and most enjoyable to read. From the collective opinions of this series, we believe the following themes emerge: (1) role of geology, (2) reasons for lack of adoption of stochastic hydrogeology research by practitioners, (3) reactive transport, and (4) connectivity.

7.1. Role of Geology

All three papers explicitly mention the critical role of geology [*Cirpka and Valocchi*, 2016; *Sanchez-Vila and Fernandez-Garcia*, 2016] or the importance of identifying the "structure" [*Fiori et al*., 2016] which is very much in agreement with our perspective. Moreover, *Cirpka and Valocchi* [2016] and *Sanchez-Vila and Fernandez-Garcia* [2016] indicate or state directly that K fields need to be conditioned on not only K measurements, but also on geologic information. This is consistent with our assertion that overreliance on K measurements and not enough on geologic information in the stochastic hydrogeology community has led to broad mischaracterization of K variability and structure.

Our colleagues suggest that new measurement techniques such as direct-push and hydrogeophysics are promising for better characterizing the relevant geologic characteristics. We agree with *Cirpka and Valocchi* [2016], however, that most of the past developments in hydrogeophysics characterization are at too local a scale to be relevant in most contaminant investigations that are on the scale of a basketball court or larger. Importantly, we think our colleagues may be underestimating the value of older geologic characterization methods based on drill cuttings, grain size analyses, core and geophysical logs, all of which have been used to great advantage by practitioners and researchers, many of whom are cited in our reference list.

In other words, although we applaud advancements such as those in direct-push technologies and geophysics for improving our shallow investigations, we think the problem of geologic characterization in stochastic hydrogeology is not strongly impeded by technological barriers. For example, *Bianchi and Zheng*'s [2016] recent paper on the MADE site demonstrates how a simple geostatistical facies model based on the sediment texture, without any of the voluminous data sets on geophysics, direct-push, and borehole flowmeter K values, reproduces the important characteristics of the plume in an ADE model. The key feature in this model is the recognition of a relatively coarse gravel hydrofacies and realistic simulation of it geostatistically in 3-D. This textural feature had been overlooked in past characterizations, but not because of a lack of straightforward technology to detect and model it.

7.2. Reasons for Lack of Adoption by Practitioners

We applaud our colleagues for going into considerable depth on possible reasons for much of the stochastic hydrogeology research penetrating so little into practice. We fully agree with *Cirpka and Valocchi* [2016] and *Sanchez-Vila and Fernandez-Garcia* [2016] that there are many processes besides spatially variable K that influence or dominate contaminant transport, often limiting applicability of the stochastic transport theories to actual problems. These processes include spatially variable and transient boundary conditions in both the flow and the transport systems, such as the fact that the contaminant source terms are almost never Dirac or instantaneous, are often unknown, and can last for decades.

Relative to the question of incorporating the geology, *Sanchez-Vila and Fernandez-Garcia* [2016] made the astute observation that the practitioners routinely incorporate geologic details such as geometry and location of aquitard confining beds and lenses into their deterministic models, "yet, we [stochastic hydrogeologists] routinely build our stochastic model on simplistic geometrical depictions and hope that the SRF framework will be smart enough to take over." We believe this tells much of the story behind the adoption problem—practitioners at consulting firms and agencies include geologists who map geologic units, including nonaquifer materials, that do not fit neatly into much of the stochastic theory, and hence they apparently do not see a role for that theory.

To produce greater relevance in stochastic hydrogeology research, *Cirpka and Valocchi* [2016] and *Fiori et al*. [2016] suggested a more goal-based approach in which there is more overlap with the goals of practitioners. We agree enthusiastically, and we also think that much of the past stochastic hydrogeology research is not aligned with those goals for the following reasons. In our experience, the goals of groundwater contaminant site practitioners are to (1) predict plume movement under various hydraulic and solute boundary conditions that are commonly not based on much knowledge of the specific location, strength, and temporal variability of the source term, (2) cleanup or control the plume through pumping extraction wells, reactive barriers or other means, and (3) predict the combined effects of physical transport and biochemical reactions on natural or engineered attenuation of the contaminants.

In the context of goal (1), the uncertainty of the source term as well as poorly known past or future hydraulic gradients, recharge and pumping can swamp out any uncertainty stemming from spatially varying K, leaving one with little alternative besides conservative, worst-case, deterministic calculations. When the plume is entirely within very sandy facies, such as those at Borden, Cape Cod, and MADE, predicting maximum plume extent and early breakthrough is quite feasible using standard well test measurements of the bulk, effective K (e.g., MADE site pumping test data from *Boggs et al*. [1990] and *Bianchi and Zheng* [2016]). This is consistent with our suggestion that modeling of forward plume movement should not necessarily be considered the primary goal of stochastic hydrogeology. Yet all of the model development at MADE has focused on modeling forward plume movement using nonlocal methods that are able to mimic or represent the important distal and proximal tails. The development of such models would in theory have special application for addressing goal (2) on site cleanup and plume management. Unfortunately, however, the nonlocal models, with the possible exception of *de Barros et al*. [2013], appear unable to represent the relevant transport phenomena when the boundary conditions switch from forward plume movement to plume extraction or management via pumping and injection wells or the like [e.g., *Neuman and Tartakovsky*, 2009]. Moreover, if the models are not capable of representing the diffusion or slow-advection dominant transport that makes plumes ubiquitously resistant to removal under transient hydraulic conditions influenced by pumping, such theories and models also do not help practitioners with the site cleanup and management problems. Consequently, the practitioners are left with insights and models that do not appreciably help them with two of their primary tasks.

Goal (3) on modeling or determining natural or engineered attenuation relates to both physical transport processes and reactions, which are covered in the next section.

7.3. Reactive Transport

All the authors point out reactive transport as a major challenge, and we strongly agree. *Cirpka and Valocchi* [2016] and *Fiori et al*. [2016] eloquently describe the problem in the context of mixing and spreading. *Sanchez-Vila and Fernandez-Garcia* [2016] describe the strong advantages of Lagrangian modeling methods for developing efficient, reliable models of the physical transport processes in heterogeneous systems, but point out the difficulty merging those particle methods with reactions in a way that is still efficient and representative of effects of mixing all the way to the molecular level. We believe that the solution will at least partly come from ongoing and future advances in our ability to model the mixing processes that control the reactions in a Lagrangian framework that will also remain true to the physical transport processes, including molecular diffusion. Then we will perhaps have tools that help practitioners address their goal number (3) on modeling natural and engineered attenuation of contaminants.

7.4. Connectivity

Fiori et al. [2016] and *Sanchez-Vila and Fernandez-Garcia* [2016] refer to connectivity of high-K facies as important to predictive transport and representation of the distal or early time tails. We agree. They also indicate or infer that methods are needed for detecting connectivity. We agree such methods would be useful, but we think that detection of *lack* of connectivity is both more feasible and more important.

Let us explain as follows. Research in stochastic hydrogeology and percolation theory [e.g., *Harter*, 2005; *Fogg et al*., 2000] demonstrates that the upper 12–20 percent of the K distribution will strongly tend to fully percolate in 3-D in correlated random fields, regardless of whether they are generated with Gaussian or Markov chain models of transition probability with geologically rigorous, asymmetric correlation structures. Perusal of 3-D geologic descriptive models in geologic textbooks [e.g., *Galloway and Hobday*, 1983] also provides intuitive, visual cues on how this could occur. Accordingly, we assert that in the absence of laterally or vertically persistent structures (i.e., geologic unconformities) that will interrupt the connectivity, laterally and vertically extensive pathways formed by percolating high-K facies should be considered more the rule than the exception, as long as the size of the domain is 5–100 times greater than the serial mean length of the facies [*Harter*, 2005].

Interestingly, this connected network paradigm of heterogeneity helps explain many scale-dependent phenomena, including low variance of pumping test K values, growth of apparent dispersivity with scale [*LaBolle and Fogg*, 2001], and the frequently observed phenomena in which solute exhibits apparent early arrivals or power law leading edges (e.g., MADE site) [*Bianchi and Zheng*, 2016; D. Benson, personal communication, 2016].

Let us return to the issue of predicting *lack* of connectivity caused by presence of geologic unconformities such as faults, erosional unconformities (e.g., paleosols) [*Weissmann et al*., 1999, 2002], and depositional systems transitions. It is important to recognize that such unconformities are relatively easy to identify and map in the geologic record. The predictable tendency for strong connectivity in hydrofacies complexes, together with the relatively identifiable and mappable unconformities that can interrupt connectivity, simplifies the characterization problem. This also provides a self-organizing framework for anticipating hydraulic and mass transport behavior in heterogeneous systems, without requiring exorbitant amounts of data and unsubstantiated inference.

Summarizing, we suggest that armed with the field and modeling experiences offered by the various experimental sites and modeling studies, and armed with a modern understanding of 3-D geologic systems and connectivity, our stochastic hydrogeology community may be closer to both theoretically profound and practical insights than previously thought.

Acknowledgments
We are grateful to the editor and reviewers for their insightful comments and edits that helped us improve the manuscript. We also benefited from discussions with Drs. Gary Weissmann, Marco Bianchi, Eric LaBolle, and William Galloway. This article integrates observations gained over many years of research, starting at The University of Texas Bureau of Economic Geology 1978–1989 and more recently at UC Davis, particularly in the Superfund Basic Research Program supported by award P42ES004699 from the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health. This paper does not analyze any new data.

References

Barahona-Palomo, M., M. Riva, X. Sanchez-Vila, E. Vazquez-Sune, and A. Guadagnini (2011), Quantitative comparison of impeller-flowmeter and particle-size-distribution techniques for the characterization of hydraulic conductivity variability, *Hydrogeol. J.*, *19*, 603–612, doi:10.1007/s10040-011-0706-5.

Bianchi, M., and C. Zheng (2016), A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer, *Water Resour. Res.*, *52*, 552–565, doi:10.1002/2015WR018186.

Boggs, J. M., S. C. Young, D. J. Benton, and Y. C. Chung (1990), Hydrogeologic characterization of the MADE site, *Interim Rep. EN-6915*, Electr. Power Res. Inst., Palo Alto, Calif.

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.

Cirpka, O. A., and A. J. Valocchi (2016), Debates—Stochastic subsurface hydrology from theory to practice: Does stochastic subsurface hydrology help solving practical problems of contaminant hydrogeology?, *Water Resour. Res.*, doi:10.1002/2016WR019087, in press.

de Barros, F. P. J., D. Fernandez-Garcia, D. Bolster, and X. Sanchez-Vila (2013), A risk-based probabilistic framework to estimate the endpoint of remediation: Concentration rebound by rate-limited mass transfer, *Water Resour. Res.*, *49*, 1929–1942, doi:10.1002/wrcr.20171.

Domenico, P. A., and F. W. Schwartz (1990), *Physical and Chemical Hydrogeology*, John Wiley, New York.

Fiori, A., G. Dagan, I. Jankovic, and A. Zarlenga (2013), The plume spreading in the MADE transport experiment: Could it be predicted by stochastic models?. *Water Resour. Res.*, *49*, 2497–2507, doi:10.1002/wrcr.20128.

Fiori, A., V. Cvetkovic, G. Dagan, S. Attinger, P. Dietrich, A. Zech, and G. Teusch (2016), Debates—Stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for?, *Water Resour. Res.*, doi:10.1002/2015WR017525, in press.

Fogg, G. E. (1986), Groundwater flow and sand-body interconnectedness in a thick, multiple-aquifer system, *Water Resour. Res.*, *22*(5), 679–694.

Fogg, G. E., and E. M. LaBolle (2006), Motivation of synthesis, with an example on groundwater quality sustainability, *Water Resour. Res.*, *42*, W03S05, doi:10.1029/2005WR004372.

Fogg, G. E., C. D. Noyes, and S. F. Carle (1998), Geologically-based model of heterogeneous hydraulic conductivity in an alluvial setting, *Hydrogeol. J.*, *6*(1), 131–143.

Fogg, G. E., S. F. Carle, and C. Green (2000), A connected-network paradigm for the alluvial aquifer system, *Spec. Pap. Geol. Soc. Am.*, *348*, 25–42.

Galloway, W. E., and D. K. Hobday (1983), *Terrigenous Clastic Depositional Systems*, Springer, Berlin.

Gelhar, L. W. (1993), *Stochastic Subsurface Hydrology*, Prentice Hall, Upper Saddle River, N. J.

Harter, T. (2005), Finite-size scaling analysis of percolation in three-dimensional correlated binary Markov chain random fields, *Phys. Rev. E*, *72*(2), 26120, doi:10.1103/PhysRevE.72.026120.

Istok, J. D., C. A. Rautman, L. E. Flint, and A. L. Flint (1994), Spatial variability in hydrologic properties of a volcanic tuff, *Ground Water*, *32*(5), 751–760.

Kaiser, W. R., L. E. Johnston, and W. N. Bach (1978), Sand body geometry and the occurrence of lignite in the Eocene of Texas, *Geol. Circ. Univ. Tex. Austin Bur. Econ. Geol.*, *78-4*, 1–19.

LaBolle, E. M., and G. E. Fogg (2001), Role of molecular diffusion in contaminant migration and recovery in an alluvial aquifer system, *Transp. Porous Media*, *42*, 155–179.

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.

Lee, S.-Y., S. F. Carle, and G. E. Fogg (2007), Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation, *Adv. Water Resour.*, *30*(9), 1914–1932.

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.

National Research Council (NRC) (2012), *Alternatives for Managing the Nations Complex Contaminated Groundwater Sites*, 422 pp., Natl. Acad. Press, Washington, D. C.

Neuman, S. P. (1990), Universal scaling of hydraulic conductivities and dispersivities in geologic media, *Water Resour. Res.*, *26*(8), 1749–1758.

Neuman, S. P., and D. M. Tartakovsky (2009), Perspective on theories of non-Fickian transport in heterogeneous media, *Adv. Water Resour.*, *32*(5), supplement, 670–680.

Neuman, S. P., and P. A. Witherspoon (1972), Field determination of the hydraulic properties of leaky multiple aquifer systems, *Water Resour. Res.*, *8*(5), 1284–1298, doi:10.1029/WR008i005p01284.

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.

Riva, M., P. Ackerer, and A. Guadagnini (2012), Interpretation of flowmeter data in heterogeneous layered aquifers, *J. Hydrol.*, *452*, 76–82, doi:10.1016/j.jhydrol.2012.05.040.

Rubin, Y. (2003), *Applied Stochastic Hydrology*, 391 pp., Oxford Univ. Press, New York.

Sanchez-Vila, X., and D. Fernandez-Garcia (2016), Debates—Stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners?, *Water Resour. Res.*, doi:10.1002/2016WR019302, in press.

Triola, M. F. (2010), *Elementary Statistics*, 11th ed., Pearson Educ., Boston, Mass.

U.S. Geological Survey (2010), Groundwater atlas of the United States, Arkansas, Louisiana, Mississippi, Map HA 730-F, Reston, Va. [Available at http://pubs.usgs.gov/ha/ha730/ch_f/.]

Weissmann, G. S., S. F. Carle, and G. E. Fogg (1999), Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics, *Water Resour. Res.*, *35*(6), 1761–1770.

Weissmann, G. S., J. F. Mount, and G. E. Fogg (2002), Glacially-driven cycles in accumulation and sequence stratigraphy of a stream-dominated alluvial fan, San Joaquin Valley, California, *J. Sediment. Res. A*, *72*(2), 241–250.

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, doi:10.1016/S0309-1708(98)00056-6.

Zheng, C., M. Bianchi, and S. M. Gorelick (2011), Lessons learned from 25 years of research at the MADE site, *Groundwater*, *49*, 649–662, doi:10.1111/j.1745-6584.2010.00753.x.