A heterogeneity model comparison of highly resolved statistically anisotropic aquifers

Erica R. Siirila-Woodburn^{1,2*} and Reed M. Maxwell¹

¹Hydrologic Science & Engineering Program

Integrated Groundwater Modeling Center (IGWMC)

Department of Geology and Geological Engineering

Colorado School of Mines, Golden, CO 80401, United States

now at:

²Department of Geotechnical Engineering and Geosciences

Universitat Politècnica de Catalunya, UPC-BarcelonaTech

08034 Barcelona, Spain

*corresponding author: erica.siirila@upc.edu

1
2
3
4 **17 Abstract**

5
6 18 Aquifer heterogeneity is known to affect solute characteristics such as spatial
7
8
9 19 spreading, mixing, and residence time, and is often modeled geostatistically to address
10
11 20 aquifer uncertainties. While parameter uncertainty is often considered, the model
12
13
14 21 uncertainty of the heterogeneity structure is often ignored. In this high-resolution
15
16 22 heterogeneity model comparison, we perform a stochastic analysis utilizing spatial
17
18
19 23 moment and breakthrough curve (BTC) metrics on Gaussian (G), truncated Gaussian
20
21 24 (TG), and non-Gaussian, or “facies” (F) heterogeneous domains. Three-dimensional
22
23
24 25 plume behavior is rigorously assessed with meter (horizontal) and cm (vertical) scale
25
26 26 discretization over a ten-kilometer aquifer. Model differences are quantified as a
27
28
29 27 function of statistical anisotropy, ε , by varying the x-direction integral scale of hydraulic
30
31
32 28 conductivity, K , from 15-960 (m). We demonstrate that the model is important only for
33
34 29 certain metrics within a range of ε . For example, spreading is insensitive to the model
35
36 30 selection at low ε , but not at high ε . In contrast, center of mass is sensitive to the model
37
38
39 31 selection at low ε , and not at high ε . A conceptual model to explain these trends is
40
41 32 proposed and validated with BTC metrics. Simulations show that G model effective K ,
42
43 33 and 1st and 2nd spatial moments are much greater than that of TG and F models. A
44
45
46 34 comparison of G and TG models (which only differ in K -distribution tails) reveal
47
48
49 35 drastically different behavior, exemplifying how accurate characterization of the K -
50
51 36 distribution may be important in modeling efforts, especially in aquifers where extreme K
52
53
54 37 values are often not measured, or inadvertently overlooked.
55
56
57
58
59
60
61
62
63
64
65

40 1 Introduction

The number of heterogeneity model comparisons in the literature are surprisingly sparse (e.g. Gomez-Hernandez & Wen, 1998; Western et al., 2001; Zhang et al., 2006; Zinn & Harvey, 2003), especially in three dimensions (Bianchi et al., 2011;

dell'Arciprete et al., 2012; Lee et al., 2007, Berg and Illman, 2014) where differences in important behavior such as late-time tailing are observed. To accurately characterize plume transport and concentrations in natural settings, modeling of aquifer connectivity and potentially tortuous flow pathways with a three-dimensional, spatially correlated heterogeneity model is imperative, especially when analyzing macrodispersion (Gelhar & Axness, 1983). As demonstrated in previous work (Navarre-Sitchler et al., 2013; Siirila & Maxwell, 2012b; Siirila et al., 2012) there is a clearly linked relationship between aquifer connectivity, statistical anisotropy of K , solute residence time (and thus allowed time for a reaction to occur), and water quality assessment. The question of how to accurately model the three-dimensional structure of heterogeneous flow fields, and furthermore the sensitivity in the heterogeneity model, is an important one which deserves more analysis.

While there are a range of emerging approaches in spatial statistics, in this work we focus on the comparison of two of the most widely used representations of spatial persistence of K in hydrogeology: 1) Gaussian random fields and 2) non-Gaussian, or facies indicator approach (Koltermann & Gorelick, 1996). We contend that inherent in the choice of a heterogeneity model may be a choice in aquifer connectivity (e.g. Wen & Gómez-Hernández, 1998). For example, in a Gaussian random field, high and low K materials are spatially isolated due to the continuous nature of the way the heterogeneity is produced in this model. In contrast, the facies approach allows for the possibility of sharp interfaces between high and low K material, potentially causing differences in how connected the simulated aquifer is. LaBolle and co-authors found the importance of neighboring strata and hydrofacies (i.e. low and high K zones) to be controlling in

correctly dating post-1950, prebomb peak ^3H and ^3He water due to molecular diffusion and diffusive fractionation (LaBolle et al., 2006). The idea of low- K connections in an isotropic, Gaussian random field has also been explored in the context of mobile-immobile mass transfer (Zinn & Harvey, 2003), where flow fields with nearly identical lognormal univariate conductivity distributions but different patterns of high and low conductivity regions are connected (connected high K structures, connected intermediate K structures, and connected low K regions) result in different flow and transport. In a comprehensive analysis of facies fields, Zhang and co-authors (2013) found that the main factor affecting non-Fickian transport is the vertical the mean length of diffusion-limited layers, suggesting the importance of the structure and arrangement of high and low K material in aquifers is crucial.

To date, very few studies have analyzed the three dimensional differences in flow given Gaussian and non-Gaussian models. Most recently, Berg and Illman (2014) compared the performance of hydraulic tomography with traditional approaches (kriging, effective parameters, transition probability Markov chain models, geologic models, stochastic inverse modeling), and assessed performance using observed well-drawdown data. The analysis of the traditional methods were inconsistent; that is, for some drawdown tests the error was small, and for others it was substantial. In a Lawrence Livermore National Laboratory pump test simulation utilizing T-PROGS and sequential Gaussian random fields, the T-PROGS model yielded an aquifer with greater lateral connectivity and could better reproduce simulated drawdown response behavior (Lee et al., 2007). The simulated aquifer, however, was characterized (and conditioned) by a large number of facies samples in a highly heterogeneous aquifer (with a variance of

1
2
3
4 109 $\ln(K) \approx 25$), potentially leading to a naturally better fit with a geologically realistic, facies
5
6
7 110 model. A different study comparing a small, 6 (m) x 6 (m) x 6.2 (m), area of the
8
9 111 macrodispersion experiment site (MADE) showed 43% of the fastest particle paths' are
10
11 112 located within a high K zone in a Gaussian random field and as high as 69% in a T-
12
13 113 PROGS field (Bianchi et al., 2011). In this comparison preferential flow dictated
14
15 114 transport, where fast pathways in the variogram-based methods were not necessarily
16
17 115 through the highest K material, suggesting particle "jumps" or "leaks". To accurately
18
19 116 characterize these interactions over long transport distances, fine cell discretization and
20
21 117 large ensembles are needed to provide robust statistical conclusions. None of the
22
23 118 aforementioned three-dimensional studies investigates far-field interactions, where the
24
25 119 three analyses were restricted to sub-fifteen meter spatial extents,.

30
31 120 To avoid these limitations, comparisons of Gaussian and non-Gaussian random
32
33 121 fields in this analysis are 1) finely discretized (meter in the horizontal direction, cm in the
34
35 122 vertical direction), 2) large extent (ten kilometers), and 3) composed of an ensemble of
36
37 123 multiple realizations to minimize statistical error. To address aquifer connectivity, spatial
38
39 124 auto-correlation lengths of K (and therefore statistical anisotropy) are also varied in this
40
41 125 analysis, where integral scales vary from the tens of meters to the kilometer scale. With
42
43 126 this highly resolved, large-scale numerical setup, our primary objective is to quantify and
44
45 127 compare the effect of heterogeneity model on plume metrics such as peak and mean
46
47 128 arrival times, aquifer connectivity, and plume center of mass and spreading. Sensitivity
48
49 129 to model selection is assessed as a function of statistical anisotropy for a given travel
50
51 130 distance, and through the use of numerical solutions. Section 2.1 describes the
52
53 131 heterogeneity models used to generate the statistically anisotropic K fields and the range
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 132 of parameters assessed, followed by Section 2.2 which describes the numerical setup of
5
6 133 the flow and transport models used for all heterogeneity models. Lastly, Section 2.3
7
8
9 134 defines the metrics used to characterize plume behavior and to perform the heterogeneity
10
11
12 135 model comparison.
13
14 136

16 137 **2 Methodology**

18 138 *2.1. Model of heterogeneity*

20
21 139 In the facies approach, an integer code is assigned to lithologic or hydrologic units,
22
23
24 140 creating an indicator database to each facies type (Carle, 1996; Deutsch & Journel, 1998;
25
26 141 Johnson, 1995; Journel, 1983; Koltermann & Gorelick, 1996; Ritzi et al., 2000). Other
27
28
29 142 additional information can be included such as the volumetric proportion, mean lengths,
30
31 143 and juxtapositional tendencies of each indicator, yielding the statistical basis for the
32
33
34 144 facies model. Examples of facies models are SISIM within the GSLIB package (Deutsch
35
36 145 & Journel, 1998) and T-PROGS (Carle, 1999) which simulates integer-coded categorical
37
38
39 146 variables or continuous variables with an indicator database. The advantage of a facies
40
41 147 approach is a more geologically realistic model of spatial persistence, able to represent
42
43
44 148 features such as asymmetric patterns of heterogeneity, sharp interfaces, and an upward-
45
46 149 fining feature typical of many alluvial deposits (e.g. Fogg et al., 1998). Statistically, a
47
48 150 Gaussian random field can be generated in an equivalent way to a facies field, and could
49
50
51 151 thus be compared. Common models used to produce such geostatistical representations
52
53 152 include GSLIB (Deutsch & Journel, 1998) and the turning-bands approach (Journel &
54
55 153 Huijbregts, 1978; Mantoglou, 1987; Mantoglou & Wilson, 1982; Matheron, 1973;
56
57
58 154 Thompson et al., 1989). The turning bands algorithm generates spatially correlated
59
60
61
62
63
64
65

random fields from a normal distribution with zero mean and specified covariance structure. The method involves the generation of a series of one-dimensional random processes along lines radiating from a coordinate origin that are projected from these lines onto a three-dimensional space. The advantage of this approach is the computational efficiency of utilizing a one-dimensional solution in three-dimensional space, but requires a large number of lines in the solution to avoid spatial distortion. In this analysis, three-dimensional Gaussian and non-Gaussian models of spatially correlated random K fields are stochastically simulated and compared. A Gaussian random field assumes a log-normal distribution of the hydrologic parameter (i.e. K), and is completely characterized by the mean, variance, and semivariogram (e.g. Rubin, 2003).

Here, Gaussian fields are generated using the turning bands algorithm (Tompson, et al., 1989). The turning-bands algorithm enforces a semivariogram function through rotation of one-dimensional lines (or bands) through space, where each value in the random field is a weighted average of values contained within each band (Tompson, et al., 1989). An exponential model is used to define spatial correlation of K via a separation distance, ζ (m), and a directional integral scale, I (m):

$$R(\chi) = S^2 \exp^{-\chi/I} \quad (1)$$

The field produced is a continuous distribution of the hydraulic property (Koltermann & Gorelick, 1996). Non-Gaussian fields are generated using transition probability indicator simulations (T-PROGS) (Carle & Fogg, 1996; Carle & Fogg, 1997; Carle, 1996; Carle, 1999; Carle et al., 1998), also referred to here as the “facies” model. T-PROGS utilizes a Markov chain model with cokriging and a final quenching step in the interpolation scheme. In this study, non-Gaussian simulations are composed of three discrete facies:

High (H), Fill (F), and Low (L). Each facies type varies by an order of magnitude in the magnitude of K (K_H , K_F , K_L), and are equivalent in volumetric proportion (see Table 1). Given the geometric mean and variance of K in non-Gaussian fields, we impose identical global statistics on the Gaussian simulations (see Table 1). In addition, a two-sided truncated Gaussian model is compared. The truncated Gaussian model is identical to the Gaussian model except that tailing values of K greater than the discrete high facies (H) value or lower than the discrete low facies (L) value are “truncated” or set equivalent to K_H or K_L , respectively (see Figure 1). Again, while the Gaussian and facies models are both characterized by the same global statistics of K , the facies model is characterized by a discrete distribution of K whereas the Gaussian and truncated Gaussian models are characterized by continuous distribution of K (see Figure 1).

The effect of statistical anisotropy is analyzed by varying the directional integral scales, commonly referred to as the correlation length in a Gaussian model and as the mean length in a facies model. For the sake of generality, we adopt the integral scale terminology, defined as the directional length of K autocorrelation in the x, y, and z-directions: I_x (m), I_y (m), I_z (m). Because directional correlation lengths are typically estimated from borelog analysis where much more information concerning the K field correlation structure is known in the vertical rather than the horizontal directions, and because there is evidence in many geologic settings supporting directional correlation (Maxwell et al., 2008), we focus our sensitivity analysis in the horizontal direction parallel to flow (I_x). For both Gaussian and facies models, the magnitude of I_x is tested for sensitivity, ranging from 15.0 (m) to 960.0 (m) across seven scenarios. For all scenarios, the magnitude of I_y and I_z are held constant (see Table 1). Statistical

201 anisotropy of K is defined via the ratio: $\varepsilon = I_z / I_x (-)$, where a $\varepsilon = 1 (-)$ describes an
 202 isotropic aquifer and where $\varepsilon < 1 (-)$ describes a statistically anisotropic aquifer. For
 203 each of the aforementioned seven cases, corresponding ε values are defined (labeled ε_{1-7} ,
 204 see Table 1), and are used to differentiate each scenario of varying I_x . Because the
 205 number of integral scales traveled is a function of ε (for example, ε_1 plumes travel more
 206 correlation lengths than ε_7 plumes), our comparisons are focused on differences between
 207 model results for a fixed ε . An alternative model setup is to instead have a fixed number
 208 of integral scales traveled for all ε scenarios; inherent in this choice is the augmentation
 209 of the flow field with lower ε . We argue the distance to a receptor, or a point of interest,
 210 is typically known a priori in a contamination assessment, and therefore has little to zero
 211 associated uncertainty to warrant a sensitivity analysis. A second alternative to assess the
 212 impact of anisotropy is to vary I_z while keeping I_x constant. Again, because the
 213 uncertainty associated with I_x is almost always much greater than the uncertainty of I_z ,
 214 this second alternative numerical setup is also of less interest. For these reasons, and
 215 because our main objective is to assess the uncertainty resulting from the choice of
 216 heterogeneity models with the same global statistics, we utilize a fixed extent domain
 217 where the number of correlation lengths traveled is dependent on ε .

218 Lastly, we also investigate the effect of asymmetry in high and low facies I_x . That
 219 is, in a symmetric facies model the I_x of both facies H and facies L (i.e. $I_{x,H}$ and $I_{x,L}$) are
 220 varied as described above, and is most equivalent to a Gaussian model which also
 221 assumes symmetries in high and low K material. We denote this symmetric facies
 222 scenario “Facies A: $I_H = I_L$ ”. High K zones are observed to be better connected in many
 223 geologic environments (e.g. Sánchez-Vila, Carrera, & Girardi, 1996). Thus, two

asymmetric high and low I_x scenarios are examined: “Facies B: $I_H > I_{L=15}$ ” and “Facies C: $I_{H=15} < I_L$ ”. In the facies scenario B, $I_{x,L}$ is constant at the lowest integral scale, $I_L = 15.0$ (m), while $I_{x,H}$ is varied, depending on which ε is being simulated. Similarly, in facies scenario C $I_{x,H}$ is constant at the lowest integral scale, $I_H = 15.0$ (m), while $I_{x,L}$ is varied, depending on which ε is simulated (see Table 1). It should be noted that the ε_I -Facies A scenario is equivalent to Facies B and Facies C scenarios (i.e. when $I_{x,H} = I_{x,L} = 15.0$), thus comparisons made for statistical anisotropy at ε_I are only between the FA and TG models, whereas ε_{2-7} scenarios compare FA, FB, FC, and TG models.

The following nomenclature will be used for brevity: Gaussian (*G*), Truncated Gaussian (*TG*), Facies A (*FA*), Facies B (*FB*), and Facies C (*FC*). For each model and for each of the seven ε scenarios, fifty realizations of equally likely, yet spatially distinct random fields of K are simulated. Convergence of the K -fields, defined as percent error between K_G and the exponential of the sum of $\ln(K)/(n_x \cdot n_y \cdot n_z)$, is below 1.8% for all ensembles, where some ensembles (the low anisotropy ratios, and for all Facies ensembles) show excellent convergence of K , below 0.15%. Using the metrics defined in Section 2.3, each ensemble is compared and discussed in Section 3.

2.2 Flow and transport

Saturated aquifer flow is simulated using the parallel, three-dimensional groundwater model ParFlow, (Ashby & Falgout, 1996; Jones & Woodward, 2001; Kollet & Maxwell, 2006) utilizing a multi-grid preconditioned conjugate gradient solver. As described in Table 2, flow and transport is solved for large-extent (10 km), finely discretized aquifers to ensure an accurate representation of heterogeneity over multiple

247 integral scales, even with an I_x as large as nearly 1 (km). Per realization, the resulting
 248 domain is approximately 184 million cells, necessitating the use of parallel high
 249 performance computing. Two constant head boundaries at $x = 0.0$ (m) and $x = 10,002$
 250 (m) drive a groundwater gradient, where all other boundaries are no-flow (see Table 2).
 251 Using the flow fields generated with ParFlow, solute transport is simulated with the
 252 Lagrangian particle tracking model SLIM-FAST (Maxwell, 2010; Maxwell &
 253 Kastenbergh, 1999; Maxwell et al., 2007) that uses a quasi-analytical solution and linear
 254 interpolation for advection terms and a bi-linear velocity interpolation (LaBolle et al.,
 255 1996) for diffusion and dispersion terms. A 10.0 (m) cubic source of resident
 256 concentration particles is instantaneously released, then tracked as a function of time at a
 257 down gradient plane near the domain outlet (see Table 2). Although the choice of
 258 injection mode has shown to affect breakthrough curve tails and spatial moment results
 259 (Demmy et al., 1999; Janković & Fiori, 2010), our focus is on differences between
 260 transport behavior between different heterogeneity models for a fixed injection source,
 261 therefore not biasing our results. Solute transport is governed by the advection dispersion
 262 equation (Fetter, 1999):

$$\frac{\partial C}{\partial t} + v_j \frac{\partial C}{\partial x_j} = \frac{\partial}{\partial x_j} \left(D_{jl} \frac{\partial C}{\partial x_l} \right) \quad j, l = 1, 2, 3 \quad (2)$$

264 where C (mg L⁻¹) is the concentration in the $x(x_1, x_2, x_3)$ coordinate vector, θ (-) is the
 265 porosity, $v = -K \nabla h / \theta$ (m d⁻¹) is the local groundwater velocity, h is the change in head (m),
 266 and D (m² d⁻¹) is the small-scale (local) hydrodynamic dispersion coefficient. To be
 267 comprehensive in our model comparison, we first quantify the effect of dispersion from
 268 velocity variations in the flow field due to large-scale heterogeneities (i.e. $D_L = D_T = 0$).

1
2
3
4 269 The effect of local dispersion in the longitudinal and transverse directions (i.e. non-zero
5
6 270 D_L and D_T) given the same conditions to test for feedbacks between macro-scale (as
7
8
9 271 discussed here) and local-scale dispersion is reserved for another analysis focusing on the
10
11 272 effects on breakthrough curve (BTC) shape (Siirila-Woodburn et al., 2014, in review).
12
13
14 273 Here, non-reactive solutes are simulated, although the investigation of reactive solutes in
15
16 274 the context of this analysis is also an area of future study. Table 2 lists flow and transport
17
18
19 275 parameters used in this analysis.
20
21 276

23 277 *2.3 Characterization of plume behavior*

26 278 A combination of spatial moment analysis (Section 2.3.1) and BTC metrics
27
28 279 (Section 2.3.2) are used to evaluate the sensitivity in transport behavior given different
29
30
31 280 heterogeneity models and statistical anisotropy ratios. A combination of dimensional and
32
33 281 non-dimensional metrics is used. BTC metrics provide information such as plume arrival
34
35
36 282 times, important in real-world applications such as time-dependent risk assessment
37
38 283 (Siirila & Maxwell, 2012a), whereas spatial moments and dimensionless metrics provide
39
40
41 284 a larger context of this analysis in the field of stochastic contaminant transport.
42

43 285

45 286 *2.3.1 Spatial moment and effective hydraulic conductivity metrics*

48 287 The use of spatial moments, unlike point estimates of concentration, are global
49
50 288 measures of the plume's behavior, and are therefore not as error prone (Rubin, 2003).
51
52
53 289 The 0th, 1st, and 2nd spatial moments of the plume are calculated using an approximated
54
55 290 particle solution (e.g. Maxwell et al., 1999; Tompson & Gelhar, 1990). The 0th order
56
57
58 291 moment, or total mass within the domain Ω is defined as:

$$M = \int_{\omega} qC(x,t)dx = mN_p \quad (3)$$

where N_p is the total number of particles and m is the mass of each individual particle.

The 1st order moment, or the average trajectory of the plume center of gravity is defined as:

$$X_j(t) = \frac{1}{M} \int_{\omega} qx C(x,t) dx \quad j = 1, 2, 3 \quad (4a)$$

Or, evaluated with respect to an approximate particle solution (denoted by the p subscript):

$$X_j(t) \approx X_{j,p}(t) = \frac{1}{N_p} \sum_{p \in N_p} x_p^t \quad j = 1, 2, 3 \quad (4b)$$

where x_p^t is the particle location of particle p at time t . The average “moment of inertia”, or 2nd order spatial moment of the solute body with respect to its mean center of gravity, is defined relative to the total mass as:

$$S_{jl}(t) = \frac{1}{M(t)} \int_{\omega} \theta C(x,t) [x_j - X_j(t)][x_l - X_l(t)] dx \quad j, l = 1, 2, 3 \quad (5a)$$

Or, evaluated with respect to an approximate particle solution:

$$S_{jl}(t) \approx S_{jl,p}(t) = \frac{1}{N_p} \sum_{p \in N_p} x_p^t(t) x_p^t(t)^T - X_{j,p}(t) X_{l,p}(t)^T \quad j, l = 1, 2, 3 \quad (5b)$$

Spatial moment results are shown as ensemble averages, and assume each particle is equal in mass. To maintain statistically significant results, ensemble averages are only reported at times where 100% of realizations contain at least 95% of the initial mass.

Lastly, for each realization, r , effective hydraulic conductivity, K_{eff}^r (m d⁻¹), is determined from the numerical results of the first spatial moment, $X_I(t)$. As X_I is a

function of time, the longest sampling of the K -field is ensured by utilizing the latest time at which 100% of the particles are within the flow field, t_f . At this time, and for each realization, the K_{eff}^r representative of the longest travel distance is calculated with respect to the source location, S_x :

$$K_{eff,t=t_f}^r = \frac{(X_{11,t=t_f} - S_x)q}{t_f J} \quad (6)$$

Ensemble averaged, effective hydraulic conductivity $K_{eff,avg}^r$ (m d^{-1}), is calculated using a running average given r of N_{real} realizations:

$$K_{eff,avg}^r = \bar{a} \frac{K_{eff,t=t_f}^i}{r} \bigg|_{i=1}^{i=r} \bigg|_{r=1}^{r=N_{real}} \quad (7)$$

While $K_{eff,avg}^{r=50}$ is the overall representative effective hydraulic conductivity for a given ensemble, the evolution of $K_{eff,avg}^r$ is also discussed in terms of convergence.

2.3.2 BTC metrics

In addition to spatial moment analysis, dimensional and non-dimensional BTC metrics based on mass arrival at the down-gradient plane (see Table 2) are used. The first is the peak mass discharge at the plane, $mass_{d,pk}$ (mg y^{-1}), calculated as the greatest amount of mass discharge which crosses the plane at a given time step, as well as the associated time at which the peak mass discharge breaks through the plane, t_{pk} (y). From a practical standpoint, the peak mass and associated time are important in water quality issues. To assess mean plume transport and the shape of the BTC, the times which fifty and five percent of the plume discharge at the plane, t_{50} (y) and t_5 (y) respectively, are

used to define the second metric, the connectivity indicator, $CI = t_{50}/t_5$ (-) (e.g. Bianchi, et al., 2011; Knudby & Carrera, 2005). CI assesses the shape of the BTC, is used to quantify channeling through preferential flow pathways, and can relay information on higher order moments of the plume. A higher CI value signifies a BTC skewed towards earlier arrival times and significant tailing, or a greater presence of preferential flow pathways (Knudby & Carrera, 2005).

3 Results and discussion

3.1 Spatial moment and effective hydraulic conductivity analysis

Figure 2 shows the numerical results for the ensemble averaged, dimensionless 1st order spatial moment, or center of mass, in the $j = 1$ direction for each model of heterogeneity: TG , G , FA , FB , FC (see key for color and symbols). Differences in ε_{1-7} are shown in subsequent figures a-g, where the X_l/I_x axes decrease in scale by a factor of two. The dimensionless form of the 1st spatial moment can be thought of as the number of integral scales traveled by the plume's center of mass, shown here as a function of time. Although the G scenario only contains a sparse number of data points before a sufficient portion of the mass escapes the domain (recall that results are only shown for time-steps in which fifty of fifty realizations contain 95% of the initial mass), it is clear that the ensemble average X_l/I_x for the G scenario is much greater than the other models. Comparing the TG - X_l/I_x and G - X_l/I_x results (which are identical in heterogeneity model and only differ by the lack of tails in the K -distribution, see Figure 1), we can conclusively state that the rapidly moving plume in the G simulations is attributed to the highest and lowest values of the K -distribution. This X_l/I_x sensitivity to the tails of the K -

1
2
3
4 354 distribution demonstrates the necessity to accurately characterize the K -distribution and
5
6 355 variance, where truncation of the K -distribution tails leads to drastically differing results
7
8
9 356 in flow and transport behavior. For example, the G model center of mass travels almost
10
11 357 twice as many integral scales than the TG model center of mass within the same time
12
13
14 358 frame. This is especially important in real-world systems where the proportions of
15
16 359 extreme high or low K values are small, and are therefore either inadvertently or
17
18
19 360 advertently overlooked, or simply not measured. Further discussion regarding the fast
20
21 361 moving G model simulations will be detailed below.

22
23
24 362 In comparing the other model X_I/I_x results, the FA model yields the farthest X_I/I_x ,
25
26 363 followed by the TG and asymmetric models (FB and FC). Surprisingly, when compared
27
28
29 364 to the TG model, the FA model more closely resembles the G model X_I/I_x . Recall that the
30
31 365 range of the K -distribution is the same for the FA and TG models, but differ as the FA
32
33 366 model distribution is discrete whereas the TG model distribution is continuous. To
34
35
36 367 determine what mechanism is causing the FA model to travel faster than the TG model,
37
38 368 the peak mass and peak time BTC metrics for these ensembles are compared in Section
39
40
41 369 3.2.1. In terms of anisotropy, for greater ε differences between the TG and facies models
42
43
44 370 are small; as ε decreases, differences between model X_I/I_x ensembles increase. In other
45
46 371 words, for more isotropic aquifers the center of mass is insensitive to heterogeneity
47
48
49 372 models, but for more anisotropic aquifers, there is sensitivity in the choice of
50
51 373 heterogeneity model. A possible explanation is that when ε is low and the heterogeneity
52
53 374 structures are large, the effect on the bulk solute position (quantified here by X_I) is more
54
55
56 375 sensitive to the structure (and thus model) of the heterogeneity. This conceptual model is
57
58 376 also further explored in Section 3.2.1 with the peak mass and peak time BTC metrics.
59
60
61
62
63
64
65

Using these numerical results for the 1st spatial moment, effective hydraulic conductivity is derived (see equations 6 and 7). Table 3 shows the representative, ensemble effective hydraulic conductivity for $r=50$, as reported by the $K_{eff,avg}^{r=50}$. Consistent with first spatial moment results (Figure 2), the G model $K_{eff,avg}^r$ is much greater than the other models, followed by the FA , TG , and asymmetric facies models (FB and FC). G model, $K_{eff,avg}^{r=50}$ are as high as five times greater than the TG and facies model $K_{eff,avg}^{r=50}$ values, suggesting that sampled K -field, and the connected structures within the field, are much more conductive. To demonstrate convergence, $K_{eff,avg}^r$ is also shown as a function of the number of realizations used. That is, while all results in this analysis are reported in terms of an ensemble of fifty realizations, $K_{eff,avg}^r$ is shown as a running average to determine after how many realizations averaged over does the effective K reach a steady-state value. Figure 3 shows $K_{eff,avg}^r$ for the different models of heterogeneity (TG , G , FA , FB , and FC , see key for color and symbols) and for differences in ε_{l-7} (shown in subsequent figures a-g). Here, we show that the effective hydraulic conductivity reaches convergence after fifty realizations for almost all ensembles. The exceptions are the low ε - G models, which reach a quasi-steady state effective hydraulic conductivity. The TG and facies models reach convergence in effective hydraulic conductivity after a relatively small number of realizations (i.e. less than twenty-five).

To assess plume spreading, the 2nd spatial moment is compared for each model, normalized by the corresponding integral scale, S_{11}/I_x^2 . Figure 4 shows the numerical results for the ensemble averaged, dimensionless 2nd spatial moment in the $j_l = 11$ direction for each model of heterogeneity: TG , G , FA , FB , FC (see key for color and

symbols). Differences in ε_{I-7} are shown in subsequent figures a-g, and where S_{II}/I_x^2 axes decrease in scale by a factor of two. For all ε , numerical simulation S_{II}/I_x^2 results are greatest for the G model, followed by the TG , FA and asymmetric facies models (FB and FC). Interestingly, this trend with regards to heterogeneity model differs from X_I and $K_{eff,avg}^{r=50}$ results (Figure 2 and Table 3, respectively), which show the FA model to most closely resemble the G model results, followed by the TG model results. Here, we show that the Gaussian models (G and TG) result in the greatest amount of plume spreading compared to the facies models, and that faster moving plumes (shown here through X_I and $K_{eff,avg}^{r=50}$) don't necessarily correlate with greater plume spreading. Also contrary to X_I and $K_{eff,avg}^{r=50}$ results are S_{II} trends with regards to anisotropy. Figure 4 shows that the model of heterogeneity becomes less important in determining plume spreading as ε decreases. For example, in Figure 4 g (ε_7) the difference between the symmetric models, TG and FA , are nearly negligible. Recall that for the 1st spatial moment, the model of heterogeneity is *more* important as ε decreases. We show that when ε is low and the heterogeneity structures are large, the effect on the bulk location of the solute (quantified with X_I) is more sensitive to the structure (and thus model) of the heterogeneity. In contrast, plume spreading (quantified with S_{II}) is greatly affected by strong preferential flow pathways. Within these preferential pathways, differences in plume spreading between heterogeneity models are small. Here, we propose it is more important that the pathways exist, and not the structure (or the model) which created the pathway. Similar results are found for plume spreading in the vertical direction (quantified with S_{33} , graphs

not shown here). Vertical spreading is also greatest for the G model, highlighting sensitivity of heterogeneity model selection in three-dimensions.

422

423 3.2 BTC analysis

424 As discussed in Section 2.4.1, a combination of dimensional and dimensionless
425 BTC metrics are used to quantify mass breakthrough at a plane down-gradient from the
426 source. The G model results are not shown for BTC metrics, as the total mass passing the
427 plane is low for many of the ensembles. Interestingly, more anisotropic G models
428 resulted in a higher percentage of mass breakthrough when compared to more isotropic G
429 model ensembles. This suggests that 1) in three dimensions, more isotropic plumes are
430 more spatially expansive, resulting in particle mass loss vertically and transversely, and
431 2) even without long, connected structures, the G model may result in very different
432 transport behavior at this variance of K , and 3) again the tails of the K -distribution may
433 play a pivotal role in determining flow and transport behavior. Peak metrics are
434 discussed in Section 3.2.1 and BTC shape metrics are discussed in Section 3.2.2.

435

436 3.2.1 Peak mass and time analysis

437 Numerical results for the peak mass and corresponding breakthrough time are
438 shown in Figure 5 for each model of heterogeneity: TG , FA , FB , FC (see key for color
439 and symbols), where each symbol represents one realization, and each ensemble of points
440 is composed of fifty realizations. Differences in ε_{l-7} are shown in subsequent figures a-g.
441 Scatter plots of t_{50} versus $mass_{d,50}$ show trends very similar to t_{pk} and $mass_{d,pk}$ metrics, and

are therefore not shown here. Discussion in terms of the peak metric therefore reflects the trends observed in terms of the 50th percentile of mass breakthrough.

First, it should be noted that for a fixed ε , each heterogeneity model results in different quantities of peak mass and the peak occurs at different times. As mentioned previously, peak metrics are important in terms of risk assessment to determine exposure times, and as shown here, would vary considerably given the choice of heterogeneity model used. In general, all facies models (*FA*, *FB*, *FC*) result in greater $mass_{d,pk}$ breakthrough than the *TG* model. This is especially true for higher anisotropy ratios (ε_{1-4}) where ensemble scatter is low in the *TG* model. In these cases, $mass_{d,pk}$ is always higher in the facies models than in the *TG* model. Our results show that while we can generalize between facies and *TG* models with regards to peak mass, corresponding peak times need to be discussed with regards to symmetric (*FA*) and non-symmetric facies models (*FB*, *FC*).

In first comparing the *TG* and *FA* models, generally the *TG* models result in later t_{pk} than the *FA* model. Again, this is especially true at higher anisotropy ratios (ε_{1-4} Figure 5, a-d) where ensemble scatter is low in the *TG* model. These two observations regarding the *TG* model (lower peak mass and at later times) suggests that the solute plume is less spatially condensed when compared to the *FA* plumes. That is, *TG* model plumes may experience more tortuous pathways in the aquifer, resulting in dilution of the bulk of the solute, decreasing the plume $mass_{d,pk}$, and delaying the corresponding t_{pk} . An example illustration of two representative flow fields and resulting plumes is shown in Figure 6. Figure 6a shows a vertical cross section of one realization of a *FA* field, compared to Figure 6b, which shows a vertical cross section of one realization of a *TG*

field. The impact of the heterogeneity model on plume shape and behavior is illustrated in Figures 6c and 6d for the same FA and TG fields, respectively. This is consistent with the X_I and $K_{eff,avg}^{r=50}$ results (Figure 2 and Table 3, respectively), which show a faster center of mass and higher effective hydraulic conductivity for the F model when compared to the TG model. This conceptual model also explains why the F model more closely resembles the G model X_I results, and why the TG model more closely resembles the G model S_{II} results.

It is important to note that while this general trend is true for all anisotropy ratios, at the lowest anisotropy ratios (ε_{4-7} , Figure 5 e-g) some realizations in the TG model actually result in *earlier* t_{pk} with *greater* $mass_{d,pk}$. Thus, caution should be used in assuming the TG model will always result in later and lower mass breakthrough. As shown here, statistical anisotropy of the aquifer needs to be explicitly considered when comparing heterogeneity models, where given low ε , preferential flow in the aquifer may be very strong, resulting in peak mass breakthrough earlier and at a higher quantity. Recall that the effect of heterogeneity models on X_I results is greater when ε is small (Figure 2 e-g). Similarly, the effect of heterogeneity models on t_{pk} is greater when ε is small (Figure 5 e-g). In other words, as the structure of the heterogeneity is large (i.e. when ε is low), the effect on the bulk of the solute (quantified here by X_I , t_{pk} , and $mass_{d,pk}$) is more sensitive to the model that created that structure.

Thus, one could argue that not only should ε be considered when comparing heterogeneity model effects on plume behavior, but also that heterogeneity model must be considered when assessing the impact of ε on plume behavior. At a similar variance of $\ln(K)$ as simulated here, and neglecting the impact of heterogeneity model, Zarlenga

488 and co-authors found a weak dependence of ε on BTC results (2013). Some key
 489 differences in numerical setups should also be noted. The ε in our analysis range over a
 490 much greater parameter space, where our most anisotropic aquifer is approximately sixty-
 491 five times more anisotropic of that of Zarlenga and co-authors; in fact, the ε of Zarlenga
 492 et al. is equivalent to the highest statistical anisotropy ratio in this study ($\varepsilon_I = 0.1$).
 493 Zarlenga et al. also demonstrate that the effect of ε varies as a function of the number of
 494 integral scales traveled; sensitivity to the number of integral scales traveled for our
 495 numerical experiment for infinite (as presented here) and additionally for a range of finite
 496 Peclet numbers is detailed in a second analysis (Siirila-Woodburn, et al., 2014, in review),
 497 showing the number of integral scales traveled impacts the influence of ε on the shape of
 498 the breakthrough curve. Thus, while Zarlenga et al. recommend the use of isotropic
 499 formation approximations when assessing longitudinal spreading and the impact of the
 500 breakthrough curve, stating “the anisotropy ratio of the hydraulic conductivity has
 501 generally a weak influence on the BTC for any degree of the spatial heterogeneity of the
 502 porous formation” we must caution the use of such approximations. Our results show
 503 that the impact of heterogeneity model in conjunction with statistical anisotropy, in
 504 addition to considerations with regards to the relative degree of anisotropy (and those
 505 covering a larger parameter space such as shown here) will likely impact breakthrough
 506 curve results. Additionally, total travel distance, and local dispersion may play an
 507 important role in solute behavior.

508 A central question with regards to models of heterogeneity BTC results still
 509 remains: what mechanisms causes the t_{pk} , X_I , and S_{II} metrics to vary between the *TG* and
 510 *FA* models, given that they are based on the same global statistics? One distinction is

511 based on the differences in K distribution (Figure 1). The intermediate values of K in the
 512 TG case (i.e. $K_L < K < K_F$ and $K_F < K < K_H$) are likely affecting the associated uncertainty
 513 about these metric results, especially given the sensitivity of the G model. The two
 514 dimensional analysis by Zhang and co-authors (2006) showed similar results in terms of
 515 peak times, where the Gaussian simulations resulted in peak breakthrough 1000 years
 516 earlier than the facies simulations, but also resulted in extensive tails, whereas the facies
 517 simulations did not. There is also the question of adjacency of high and low materials.
 518 Recall that the Gaussian models are limited to smoothly varying spatial changes in K
 519 where areas of high and low K are spatially isolated, whereas in the facies models high
 520 and low K material may be adjacent to one-another, depending on the statistics of the
 521 transition probabilities. For example, in the absence of local dispersion a parcel of solute
 522 moving through the facies aquifer in a low- K region has a much greater probability of
 523 encountering a high- K region, and thus a greater probability of faster travel time. In the
 524 Gaussian aquifer, a parcel of solute traveling in a low- K region has a low probability of
 525 encountering a high- K region nearby (and vice-versa if traveling in a high- K region there
 526 is a low probability of encountering a low- K region). Thus, in the absence of local
 527 dispersion it is more likely that the parcel of solute will not jump streamlines, and will
 528 continue traveling at the initial rate (e.g. Siirila and Maxwell, 2012). Therefore, flow in
 529 the facies models are dictated primarily by the K_H , whereas flow in the Gaussian models
 530 are dictated by a wider range of K values, and the probability of encountering the highest
 531 value (such as in the facies models) is much smaller. Obviously, the highest value of the
 532 K -distribution in the TG and FA models is the same (K_H); the distinction is that it may be
 533 more difficult for a parcel of solute to encounter a streamlines dictated by K_H in the TG

model due to the differences in flow field structure. We hypothesize the similarity in trends between the *FA* and *TG* models is due to their similarities in *K*-distribution (and highest *K* equivalent to K_H), but stress the importance of structure with regards to different plume metrics as a function of anisotropy as discussed above.

Lastly, in comparing the asymmetric facies cases (*FB* and *FC*) with the *TG* model results, asymmetric facies models always result in a greater $mass_{d,pk}$ than the *TG* model. This is true for all anisotropy ratios. With respect to peak time, the *FB* and *FC* ensemble t_{pk} generally fall within the ensemble t_{pk} times of the *TG* model. When compared to its symmetric counterpart (*FA*), *FB* and *FC* $mass_{d,pk}$ is approximately equal to that of the *FA* case for high anisotropy ratios (ε_{I-4} , Figure 5 a-d), but result in higher $mass_{d,pk}$ for low anisotropy ratios (ε_{4-7} , Figure 5 e-g). With respect to peak time *FB* and *FC* t_{pk} are always later than *FA* t_{pk} . Unlike the other model results (*FA* and *TG*), the *FB* and *FC* cases do not vary drastically as ε decreases, and are characterized by a very small ensemble scatter. This suggests that when the asymmetry in high or low material is present in the facies model, there is a lack of sensitivity to ε and a stable plume regardless of if the *H* or *L* facies types were adjusted to reflect differences in anisotropy. From these and the spatial moment observations, we discern that for facies model plumes to behave similar to *TG* model plumes, symmetry in the high and low material must be present in the flow field structure. No other trends are notable, for the *FB* and *FC* model results. We attribute this static behavior to one of the restrictions imposed in the facies model, where volumetric proportions of each of the three facies types (*H*, *F*, *L*) are equivalent (see Table 1). Thus, the spatial extent of I_H in scenario *FB* and I_L in scenario *FC* may not be as pronounced as in the symmetric facies case, *FA*. This result highlights the insensitivity

of ε given asymmetric integral scale facies models when the proportions of the facies types are equivalent, even when varying I_x from 15.0 (m) to 960 (m). This insensitivity can also be thought of as increased certainty in plume behavior, or more predictable conditions of flow and transport given an aquifer characterized similar to scenarios *FB* or *FC*.

3.2.2 BTC shape analysis

To assess BTC shape, the mass and time which fifty percent of the plume breaks through the plane, $mass_{d,50}$ (mg) and t_{50} (y), respectively, are also compared for different heterogeneity models. As noted before, scatter plots of t_{50} versus $mass_{d,50}$ show very similar trends as peak mass and time metrics (i.e. Figure 5 for the peak breakthrough), and are therefore not shown here. Instead, we focus on the ratio of the t_{50} and t_5 metrics (i.e. the connectivity indicator, *CI*) in the form of cumulative distribution functions (CDFs). Figure 7 shows the numerical *CI* results for each model of heterogeneity: *TG*, *FA*, *FB*, and *FC* (see key for color and symbols), where each ensemble is composed of fifty realizations. Differences in ε_{1-7} are shown in subsequent figures a-g.

CI is generally greatest for the *TG* model and lowest for the asymmetric facies models (*FA* and *FB*). The symmetric facies model (*FA*) is of intermediate connectivity, although yields *CI* values closer to that of the *TG* model than the asymmetric facies models. This is consistent with the 2nd spatial moment results, S_{11} , which show the same trends with regards to heterogeneity model, and confirms the proposed conceptual model that trends in preferential flow should coincide with trends in S_{11} . These trends are especially evident at lower anisotropy ratios (ε_{4-7} , Figure 7 e-g) where *TG CI* distributions

are greatest and *FB* and *FC* *CI* distributions are invariant to the decrease in anisotropy and yield a very low *CI*, regardless of ε . From this observation we can again interpret that when facies volumetric proportions are equivalent, symmetry in the heterogeneity correlation structure (i.e. symmetry in H and L facies) is required to ensure plume behavior similar to those of the Gaussian models. Preferential flow in *FB* and *FC* aquifers is considerably lower compared to the not only the *TG* model but also the symmetric facies counterpart (*FA*).

On close inspection of *CI* CDFs for low anisotropy ratios, a portion of the *TG* realizations have lower *CI* than the *FA* model. The number of realizations varies with ε between 8% (ε_4 , Figure 7 e) and 38% (ε_7 , Figure 7 g). In other words, while in general there is more preferential flow in the *TG* model when compared to the facies models, a percentage of plumes in the *TG* model are actually characterized by relatively low *CI* values, and less preferential flow than the *FA* model plumes. This indicates that some realizations are characterized by low preferential flow (where either the t_{50} is low and the bulk of the solute is moving slowly, or t_5 is high and there may be early arriving solute streamlines at the plane). To determine which metric (t_{50} or t_5) is causing a lower *TG* *CI* for these realizations, t_{50} and t_5 are further assessed to determine patterns in arrival times.

Figure 8 shows ensemble scatter plots of t_{50} versus t_5 for each model of heterogeneity: *TG*, *FA*, *FB*, and *FC* (see key for color and symbols), where each ensemble is composed of fifty realizations. Differences in ε_{1-7} are shown in subsequent figures a-g. *TG* realizations in which the *CI* of the *TG* model is lower than that of the *FA* model are marked with a pink cross (Figure 8 e-g). While realizations with these markers are among the highest t_5 of the ensemble, suggesting the cause of the lower *CI* value is

603 due to early arriving solute streamlines at the plane is the physical mechanisms affecting
 604 the calculation of high CI , some marked realizations also reflect a low t_{50} , suggesting a
 605 slow arrival of the bulk of the solute as the mechanisms affecting the CI metric. Further
 606 analysis of the t_{50} and t_5 metrics show that these breakthrough times are correlated,
 607 especially when ε is high (see Figure 8). Therefore, because a higher t_{50} (thus a higher
 608 CI) will likely correspond to a higher t_5 (thus a lower CI) and vice versa, there may be a
 609 bias in merely evaluating the ratio of the breakthrough times. We propose an alternative
 610 approach to the CI by simply taking the differential of these two breakthrough times (i.e.
 611 $t_{50} - t_5$) to remove the bias due to correlated breakthrough times as shown in Figure 8.
 612 The differential in time $t_{50} - t_5$ can conceptually be thought of as half the duration (or
 613 residence) of the plume travel time. This alternative definition is shown in Figure 9,
 614 where the numerical $t_{50} - t_5$ results for each model of heterogeneity: TG , FA , FB , and FC
 615 (see key for color and symbols) are shown in ensemble CDFs composed of fifty
 616 realizations. Differences in ε_{1-7} are shown in subsequent figures a-g.

617 While the trends in Figure 9 are the same as those shown by the CI metric (Figure
 618 7), the differential in time $t_{50} - t_5$ highlights the drastic travel time differences given the
 619 choice in heterogeneity model. For example, the asymmetric facies models (FB and FC)
 620 always have a $t_{50} - t_5$ time less than 150 (y), regardless of ε . In comparison, the lowest t_{50}
 621 - t_5 time for the TG model is always greater than this time, where the lowest TG $t_{50} - t_5$ is
 622 165 (y), (Figure 9a), and the highest TG $t_{50} - t_5$ is as high as 344,000 (y) (Figure 9g).
 623 Even in comparing FA and TG model results for the same ε , FA $t_{50} - t_5$ are typically half
 624 the duration of the TG $t_{50} - t_5$. These differences in duration are substantial, and highlight
 625 the sensitivity in model selection.

Similar to the *CI* metric (Figure 7) there is a $t_{50} - t_5$ CDF crossover point between the *TG* and *FA* models for low ε (ε_{5-7} , Figure 5 f-g). For $t_{50} - t_5$ results, the crossover occurs at a lower anisotropy ratio (ε_5 opposed to ε_4 for *CI* results), and affects a lower percentage of the ensemble (less than 6% of the realizations for ε_6 and 10% of the realizations for ε_7). *TG* realizations in which the $t_{50} - t_5$ is lower than that of the *FA* model are marked with a blue cross on the scatterplot (Figure 8 f-g). Similar to the *CI* analysis, realizations with these markers are among the highest t_5 of the ensemble. Using this second metric, we elucidate the mechanism that causes *TG* model realizations to result in a lower *CI* than the *F* model is a slow arrival of the bulk of the plume, and not early arrivals of streamlines at the plane. As discussed before, since t_{50} and t_5 are found to be correlated in these simulations (Figure 8), the degree to which this physical trapping mechanism is occurring given a low ε may be over-estimated by the *CI* metric, and may be lower, (although still present) as shown by the $t_{50} - t_5$ metric.

639

640 **4 Conclusions**

In this work, different models of heterogeneity are compared. A stochastic analysis utilizing a combination of spatial moment and BTC metrics on Gaussian (*G*), Truncated Gaussian (*TG*), and non-Gaussian (i.e. “facies” *FA*, *FB*, *FC*) heterogeneous flow fields was performed. This analysis expands on previous small, two-dimensional work with high-resolution simulations in three dimensions where heterogeneity is discretized with meter (horizontal) and cm (vertical) scale discretization over a ten-kilometer regional aquifer. Three-dimensional plume behavior is quantified with an array of metrics such as plume center of mass, plume spreading, arrival mass and time, and

connectivity. We also address the question of symmetric and asymmetric high and low K material by varying the statistical anisotropy of each modeled aquifer, which is adjusted by varying the x-direction integral scale (symmetrically and asymmetrically in high and low K material) between 15.0 (m) and 960.0 (m) to provide sensitivity over a vast parameter space of integral scales for each heterogeneity model. Sensitivity to the x-direction integral scale is of interest in our sensitivity analysis as characterization in the vertical direction is often much greater than in the horizontal, and thus has a higher associated uncertainty. Whereas others studies have focused on the effect of anisotropy as a function of variance of $\ln(K)$ (e.g. Zarlenga et al., 2013), our results focused on this sensitivity at a moderate, fixed variance of $\ln(K)$. Main results include:

1. G model transport is much faster than its TG and FA model counterparts. When compared to the TG results (which are identical in heterogeneity model and global statistics but differ in the presence of the K distribution tails) plume behavior is drastically different, where the effective hydraulic conductivity ($K_{eff,avg}^{r=50}$) of the G models is as high as five times faster. This result exemplifies how accurate characterization of the K distribution may be important in modeling efforts, especially important in real-world systems where the extreme high or low K values are often not measured, or advertently overlooked.
2. G and TG models result in the greatest amount of plume spreading compared to the facies models, and a faster moving plume (shown here through the first spatial moment, X_I and $K_{eff,avg}^{r=50}$) doesn't necessarily correlate with greater plume spreading. The latter is more impacted by preferential flow in the aquifer.

- 1
2
3
4 671 3. 1st and 2nd spatial moments (X_I and S_{II} , respectively) show the model of heterogeneity
5
6 672 is important, but is a function of the anisotropy ratio, ε . For the cases examined, X_I is
7
8 673 sensitive to the model of heterogeneity, especially when ε is low. In contrast, S_{II} is
9
10 674 sensitive to the model of heterogeneity, especially when ε is high. An explanation for
11
12 675 these differences is that when ε is low (and the heterogeneity structures are large), the
13
14 676 effect on the bulk solute position, quantified here through X_I , is more sensitive to the
15
16 677 structure (and thus model) of the heterogeneity. Also at low ε , preferential flow is
17
18 678 high, and plume spreading is confined by large structures; thus when quantifying S_{II} ,
19
20 679 the model which created preferential flow is less important, only that preferential
21
22 680 flow exists.
23
24 681 4. Trends with regards to the heterogeneity model differ between the 1st and 2nd spatial
25
26 682 moments. For X_I , the G model is much greater than the other models, followed by
27
28 683 the FA , TG , and asymmetric facies models (FB and FC). The same is observed with
29
30 684 regards to derived effective hydraulic conductivity values, and peak times. In
31
32 685 contrast, G model S_{II} is greater than the other models, but is followed by the TG , FA ,
33
34 686 and asymmetric facies models (FB and FC). When compared to the FA model, the
35
36 687 TG model results in lower peak mass and at later times, suggests that a TG plume is
37
38 688 less spatially condensed when compared to FA plumes. Conceptually, TG model
39
40 689 plumes may experience more tortuous pathways in the aquifer, resulting in dilution of
41
42 690 the bulk of the solute, and delaying the corresponding breakthrough time.
43
44 691 5. Breakthrough times t_{50} and t_5 were analyzed to assess connectivity, and for the
45
46 692 simulations in the analysis, show clear correlation in arrival times. Because a higher
47
48 693 t_{50} (thus a higher degree of connectivity, $CI = t_{50}/t_5$) will likely correspond to a higher
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

t_5 (thus a lower *CI*) and vice versa, there may be a bias in merely evaluating the ratio of the breakthrough times. Of interest are specific *TG* realizations where the *TG CI* is less than the *FA CI*. Although some realizations are also characterized by a relatively slow t_{50} , these realizations are also characterized by a fast t_5 . To remove the bias in correlation, the differential of these two breakthrough times (i.e. $t_{50} - t_5$) shows realizations where the *TG* differential is less than the *FA* differential are characterized by a slow t_{50} , and is therefore the main physical mechanism linked to a lapse in trend.

6. In scenarios where there is not a clear preference for one heterogeneity model over another and only global statistics are known, differences in transport may arise from the selection of Gaussian versus non-Gaussian models. Facies model plumes of equivalent proportion facies behave similarly to Gaussian model plumes only when symmetry in the high and low material is present in the flow field structure. Unlike the other model results, the asymmetric facies models (*FB* and *FC*) X_I and S_{II} results are the lowest of the models analyzed, BTC metrics results do not vary drastically as ε decreases, and very small ensemble scatter characterizes BTC metrics.

Site characterization activities are costly and can be aided by understanding the differences in plume behavior for different models of heterogeneity. For example the difference in plume geometry, rates of spreading and breakthrough times shown the different models might aid in guiding the choice of hydrogeologic conceptual model at a given site. Future work includes the investigation of different models of heterogeneity at this high resolution and spatial extent, for example comparisons with multipoint geostatistics, (e.g. dell’Arciprete et al., 2012). The effect of *K* distribution variance, the number of facies types, and the corresponding facies transition probabilities are potential

1
2
3
4 717 next steps in assessing the effect of heterogeneity model on plume behavior. Further
5
6 718 analysis in asymmetry of high and low material includes simulations in which the
7
8
9 719 volumetric proportions of each of the three facies types (H , F , L) are not-equivalent.
10
11 720 While the results presented here are valid for finite Peclet numbers only (i.e. where $\alpha_L =$
12
13 721 0.0, $\alpha_T = 0.0$), the question of feedbacks in scales of heterogeneity is addressed in a
14
15 722 second analysis, where the effect of local dispersion is analyzed with regards to Gaussian
16
17 723 and non-Gaussian models, and also as a function of statistical anisotropy, with special
18
19 724 attention on the shape of the BTC (Siirila-Woodburn, et al., 2014, in review). We thus
20
21 725 caution generalizing the impact of anisotropy on plume behavior (e.g. Zarlenga et al.,
22
23 726 2013), where considerations with regards to the relative degree of anisotropy, total travel
24
25 727 distance, local dispersion, variance of $\ln(K)$, and model of heterogeneity will all likely
26
27 728 impact results.
28
29
30
31
32
33
34

35 730 **Acknowledgments**

36
37
38 731 Data for this paper is available on request from the corresponding author. Funding for
39
40 732 this work was provided by DOE NETL Grant No. DE-FE0002059 and EPA STAR Grant
41
42 733 No. RD-83438701-0. This research was supported in part by the Golden Energy
43
44 734 Computing Organization at the Colorado School of Mines using resources acquired with
45
46 735 financial assistance from the National Science Foundation and the National Renewable
47
48 736 Energy Laboratory. This research has been supported in part by a grant from the US
49
50 737 Environmental Protection Agency's Science to Achieve Results (STAR) program.
51
52 738 Although the research described in the article has been funded wholly or in part by the
53
54 739 US Environmental Protection Agency's STAR program through Grant RD-83438701-0,
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

740 it has not been subjected to any EPA review and therefore does not necessarily reflect the
741 views of the Agency, and no official endorsement should be inferred.

742

743

744

References

- 745 Andričević, R., Cvetković, V. (1996). Evaluation of Risk from Contaminants Migrating
746 by Groundwater. *Water Resour. Res.*, 32(3), 611-621.
- 747 Ashby, S., & Falgout, R. (1996). A parallel multigrid preconditioned conjugate gradient
748 algorithm for groundwater flow simulations. *Nuclear Science and Engineering*,
749 124(1), 145-159.
- 750 Berg, S.J. & Illman, W. A.(2014). Comparison of hydraulic tomography with traditional
751 methods at a highly heterogeneous site. *Groundwater*, doi: 10.1111/gwat.12159.
- 752 Bianchi, M., Zheng, C., Wilson, C., Tick, G. R., Liu, G., & Gorelick, S. M. (2011).
753 Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential
754 flow paths. *Water Resour. Res.*, 47(5), W05524.
- 755 Bolster, D., Barahona, M., Dentz, M., Fernandez-Garcia, D., Sanchez-Vila, X., Trinchero,
756 P. (2009). Probabilistic risk analysis of groundwater remediation strategies. *Water*
757 *resources research*, 45(6), n/a-n/a.
- 758 Carle, S., & Fogg, G. (1996). Transition probability-based indicator geostatistics.
759 *Mathematical Geology*, 28(4), 453-476.
- 760 Carle, S., & Fogg, G. (1997). Modeling Spatial Variability with One and
761 Multidimensional Continuous-Lag Markov Chains. *Mathematical Geology*, 29(7),
762 891-918.
- 763 Carle, S. F. (1996). *A transition probability-based approach to geostatistical*
764 *characterization of hydrostratigraphic architecture*. University of California,
765 Davis.
- 766 Carle, S. F. (1999). T-PROGS: Transition probability geostatistical software. *University*
767 *of California, Davis, CA*.
- 768 Carle, S. F., Labolle, E. M., Weissmann, G. S., Van Brocklin, D., & Fogg, G. E. (1998).
769 Conditional simulation of hydrofacies architecture: a transition
770 probability/Markov approach. *Hydrogeologic models of sedimentary aquifers,*
771 *concepts in hydrogeology and environmental geology*, 1, 147-170.
- 772 Dagan, G., & Neuman, S. P. (2005). *Subsurface flow and transport: a stochastic*
773 *approach*: Cambridge University Press.
- 774 dell'Arciprete, D., Bersezio, R., Felletti, F., Giudici, M., Comunian, A., & Renard, P.
775 (2012). Comparison of three geostatistical methods for hydrofacies simulation: a
776 test on alluvial sediments. *Hydrogeology Journal*, 20(2), 299-311.
- 777 Demmy, G., Berglund, S., & Graham, W. (1999). Injection mode implications for solute
778 transport in porous media: Analysis in a stochastic Lagrangian Framework. 35(7),
779 1965-1973.
- 780 Deutsch, C. V., & Journel, A. G. (1998). *GSLIB geostatistical software library and user's*
781 *guide*. 2nd ed. New York: Oxford University Press.
- 782 Fetter, C. (1999). *Contaminant hydrogeology*: Prentice hall Upper Saddle River, NJ.
- 783 Fogg, G. E., Noyes, C. D., & Carle, S. F. (1998). Geologically based model of
784 heterogeneous hydraulic conductivity in an alluvial setting. *Hydrogeology*
785 *Journal*, 6(1), 131-143.

- Gelhar, L. W., & Axness, C. L. (1983). Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resour. Res.*, 19(1), 161-180.
- Gomez-Hernandez, J. J., & Wen, X.-H. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1), 47-61.
- Janković, I., & Fiori, A. (2010). Analysis of the impact of injection mode in transport through strongly heterogeneous aquifers. *Special Issue on Novel Insights in Hydrological Modelling Rome-2009*, 33(10), 1199-1205.
- Johnson, N. M. (1995). Characterization of alluvial hydrostratigraphy with indicator semivariograms. *Water Resour. Res.*, 31(12), 3217-3227.
- Jones, J. E., & Woodward, C. S. (2001). Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. [doi: DOI: 10.1016/S0309-1708(00)00075-0]. *Advances in Water Resources*, 24(7), 763-774.
- Journel, A. (1983). Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15(3), 445-468.
- Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*: Academic Press San Diego.
- Knudby, C., & Carrera, J. s. (2005). On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, 28(4), 405-421.
- Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945-958.
- Koltermann, C. E., & Gorelick, S. M. (1996). Heterogeneity in Sedimentary Deposits: A Review of Structure-Imitating, Process-Imitating, and Descriptive Approaches. *Water Resour. Res.*, 32(9), 2617-2658.
- LaBolle, E. M., Fogg, G. E., & Eweis, J. B. (2006). Diffusive fractionation of 3H and 3He in groundwater and its impact on groundwater age estimates. *Water resources research*, 42(7), W07202.
- LaBolle, E. M., Fogg, G. E., & Tompson, A. F. B. (1996). Random-Walk Simulation of Transport in Heterogeneous Porous Media: Local Mass-Conservation Problem and Implementation Methods. *Water Resour. Res.*, 32(3), 583-593.
- Lee, S.-Y., Carle, S. F., & Fogg, G. E. (2007). Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation. *Advances in Water Resources*, 30(9), 1914-1932.
- Mantoglou, A. (1987). Digital simulation of multivariate two-and three-dimensional stochastic processes with a spectral turning bands method. *Mathematical Geology*, 19(2), 129-149.
- Mantoglou, A., & Wilson, J. L. (1982). The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, 18(5), 1379-1394.
- Matheron, G. (1973). The intrinsic random functions and their application. *Applied Probability Trust*, 5, 439-468.
- Maxwell, R., Carle, S., & Tompson, A. (2008). Contamination, risk, and heterogeneity: on the effectiveness of aquifer remediation. [10.1007/s00254-007-0955-8]. *Environmental Geology*, 54(8), 1771-1786.
- Maxwell, R. M. (2010). SLIM-FAST: A User's Manual V.4, GWMI.

- Maxwell, R. M., & Kastenberg, W. E. (1999). Stochastic environmental risk analysis: an integrated methodology for predicting cancer risk from contaminated groundwater. *Stochastic Environmental Research and Risk Assessment*, 13(1-2), 27-47.
- Maxwell, R. M., Kastenberg, W. E., & Rubin, Y. (1999). A methodology to integrate site characterization information into groundwater-driven health risk assessment. *Water Resour. Res.*, 35(9), 2841-2855.
- Maxwell, R. M., Welty, C., & Harvey, R. W. (2007). Revisiting the Cape Cod Bacteria Injection Experiment Using a Stochastic Modeling Approach. *Environmental Science & Technology*, 41(15), 5548-5558.
- Navarre-Sitchler, A. K., Maxwell, R. M., Siirila, E. R., Hammond, G. E., & Lichtner, P. C. (2013). Elucidating geochemical response of shallow heterogeneous aquifers to CO₂ leakage using high-performance computing: Implications for monitoring of CO₂ sequestration. *Advances in Water Resources*, 53(0), 45-55.
- Ritzi, R. W., Dominic, D. F., Slesers, A. J., Greer, C. B., Reboulet, E. C., Telford, J. A., (2000). Comparing statistical models of physical heterogeneity in buried-valley aquifers. *Water resources research*, 36(11), 3179-3192.
- Rubin, Y. (2003). *Applied stochastic hydrogeology*: Oxford University Press, USA.
- Rubin, Y., Cushey, M. A., & Bellin, A. (1994). Modeling of transport in groundwater for environmental risk assessment. *Stochastic Hydrology and Hydraulics*, 8(1), 57-77.
- Siirila, E. R., & Maxwell, R. M. (2012a). A new perspective on human health risk assessment: Development of a time dependent methodology and the effect of varying exposure durations. *Science of The Total Environment*, 431(0), 221-232.
- Siirila, E. R., & Maxwell, R. M. (2012b). Evaluating effective reaction rates of kinetically driven solutes in large-scale, statistically anisotropic media: Human health risk implications. *Water Resour. Res.*, 48(4), W04527.
- Siirila, E. R., Navarre-Sitchler, A. K., Maxwell, R. M., & McCray, J. E. (2012). A quantitative methodology to assess the risks to human health from CO₂ leakage into groundwater. *Advances in Water Resources*, 36(0), 146-164.
- Siirila-Woodburn, E. R., Sanchez-Vila, X., & Fernández-García, D. (2014, in review). On the formation of multiple local peaks in breakthrough curves. *Water resources research*
- Sánchez-Vila, X., Carrera, J., & Girardi, J. P. (1996). Scale effects in transmissivity. *Journal of Hydrology*, 183, 1-22.
- Tartakovsky, D. M. (2007). Probabilistic risk analysis in subsurface hydrology. *Geophys. Res. Lett.*, 34(5), L05404.
- Tompson, A. F. B., Ababou, R., & Gelhar, L. W. (1989). Implementation of the three-dimensional turning bands random field generator. *Water Resour. Res.*, 25(10), 2227-2243.
- Tompson, A. F. B., & Gelhar, L. W. (1990). Numerical simulation of solute transport in three-dimensional, randomly heterogeneous porous media. *Water Resour. Res.*, 26(10), 2541-2562.
- Wen, X.H., Gómez-Hernández, J.J.,(1998). Numerical modeling of macrodispersion in heterogeneous media: a comparison of multi-Gaussian and non-multi-Gaussian models. *Journal of Contaminant Hydrology*, 30(1-2), 129-156.
- Western, A. W., Blöchl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resour. Res.*, 37(1), 83-97.

878 Zarlenga, A., Janković, I., & Fiori, A. (2013). Advective Transport in Heterogeneous
879 Formations: The Impact of Spatial Anisotropy on the Breakthrough Curve.
880 *Transport in Porous Media*, 96(2), 295-304.

881 Zhang, Y., Green, C.T., Fogg, G.E., (2013). The impact of medium architecture of
882 alluvial settings on non-Fickian transport. *Advances in Water Resources*, 54, 78-
883 99.

884 Zhang, H., Harter, T., & Sivakumar, B. (2006). Nonpoint source solute transport normal
885 to aquifer bedding in heterogeneous, Markov chain random fields. *Water Resour.*
886 *Res.*, 42(6), W06403.

887 Zinn, B., & Harvey, C. F. (2003). When good statistical models of aquifer heterogeneity
888 go bad: A comparison of flow, dispersion, and mass transfer in connected and
889 multivariate Gaussian hydraulic conductivity fields. *Water Resour. Res.*, 39(3),
890 1051.
891
892

Figure 1. Discrete versus continuous hydraulic conductivity (K) distributions for the models of heterogeneity used in this analysis: G , TG , FA , FB , FC .

Figure 2. Ensemble averaged, dimensionless 1st order spatial moment in the $jl = 11$ direction for each model of heterogeneity (see key for color and symbols). Model comparisons as a function of anisotropy (e_{1-7}) are shown in a-g, where X_l/I_x axes decrease in scale by a factor of two.

Figure 3. Numerical results for the effective hydraulic conductivity as a function of the number of realizations used. Model comparisons as a function of anisotropy (ε_{1-7}) are shown in a-g, and for model of heterogeneity (see key for color and symbols).

Figure 4. Ensemble averaged, dimensionless 2nd spatial moment in the $jl = 11$ direction as a function of time. Heterogeneity model comparisons (see key for color and symbols) are shown as a function of anisotropy (ε_{l-7}) in a-g, where S_{11}/I_x^2 axes decrease in scale by a factor of two.

Figure 5. Numerical results for the peak mass and associated time metrics for each model of heterogeneity (see key for color and symbols) where each ensemble represents the statistics of fifty realizations. Model comparisons as a function of anisotropy (ϵ_{1-7}) are shown in a-g.

Figure 6. Illustration of two representative flow fields and the resultant plumes. Vertical cross-sections of hydraulic conductivity, K (m/d) are shown for a) one realization of a FA field and b) one realization of TG field. Resulting concentration plumes for the same FA and TG fields are shown in c) and d), respectively.

Figure 7. Numerical results for the connectivity indicator metric for each model of heterogeneity (see key for color and symbols)

where each ensemble represents the statistics of fifty realizations. Model comparisons as a function of anisotropy (ϵ_{1-7}) are shown in a-g.

Figure 8. Correlations between t_{50} and t_5 for each model of heterogeneity (see key for color and symbols) where each ensemble represents the statistics of fifty realizations. Model comparisons as a function of anisotropy (ϵ_{1-7}) are shown in a-g.

Figure 9. CDFs of time differential $t_{50} - t_5$ for each model of heterogeneity (see key for color and symbols) where each ensemble represents the statistics of fifty realizations. Model comparisons as a function of anisotropy (e_{1-7}) are shown in a-g. Note that $t_{50} - t_5$ is in log-space.

Table 1. Heterogeneity parameters

Anisotropy Ratio	Value	Units
$\varepsilon_1 = 0.1$ (-)	$I_x = 15.0$	(m)
$\varepsilon_2 = 0.05$ (-)	$I_x = 30.0$	(m)
$\varepsilon_3 = 0.025$ (-)	$I_x = 60.0$	(m)
$\varepsilon_4 = 0.0125$ (-)	$I_x = 120.0$	(m)
$\varepsilon_5 = 0.00625$ (-)	$I_x = 240.0$	(m)
$\varepsilon_6 = 0.003125$ (-)	$I_x = 480.0$	(m)
$\varepsilon_7 = 0.0015625$ (-)	$I_x = 960.0$	(m)
*For all scenarios	$I_y = 15.0, I_z = 1.5$	(m), (m)

Note for the scenarios:

“Facies A” - $I_{x,H}$ and $I_{x,L}$ are both varied for each ε

“Facies B” - $I_{x,H}$ is only varied for each ε while $I_{x,L}=15$

“Facies C” - $I_{x,L}$ is only varied for each ε while $I_{x,H}=15$

Parameter	Value	Units
Geometric mean of K	$K_G = 1.0$	(m/d)
Standard deviation of $\ln(K)$	$\sigma = 2.3$	(-)
Facies K values and proportions	$K_H = 10.0, P_H = 0.33$	(m/d), (-)
	$K_F = 1.0, P_F = 0.34$	(m/d), (-)
	$K_L = 0.1, P_L = 0.33$	(m/d), (-)
Facies off-diagonal transition probabilities (all directions)	$r_{HL} = r_{LH} = 0.5$	(-)

Table 2. Flow and transport parameter values.

Parameter	Value	Units
Domain Size (x, y, z)	(10,002 x 501 x 99)	(m)
Cell discretization ($\Delta x, \Delta y, \Delta z$)	(3.0 x 3.0 x 0.3)	(m)
Number of cells (nx, ny, nz)	(3334 x 167 x 330)	(-)
Centered location of source (S_x, S_y, S_z)	(250.0, 250.0, 50.0)	(m)
Number of particles	400,000	(-)
Local dispersivity	$\alpha_L = 0.0, \alpha_T = 0.0$	(m)
Porosity	$\theta = 0.3$	(-)
Plane location (x, y, z)	(9,750, 0 – 501, 0 – 99)	(m)
Change in head and gradient	$\Delta h = 25.0$	(m)
	$J = 0.0025$	(-)

Table 3. Ensemble averaged effective hydraulic conductivity after 50 realizations, $K_{eff,avg}^{r=50}$ (m/d).

Model:	<i>G</i>	<i>TG</i>	<i>FA</i>	<i>FB</i>	<i>FC</i>
ϵ_1	3.95	2.14	2.24		
ϵ_2	5.56	2.48	2.83	2.55	2.38
ϵ_3	8.06	2.72	3.21	2.66	2.32
ϵ_4	9.67	2.83	3.48	2.65	2.21
ϵ_5	11.66	2.88	3.57	2.40	2.61
ϵ_6	11.96	2.86	3.60	2.65	2.49
ϵ_7	12.29	2.89	3.59	2.74	2.41