

1 **Solute transport in bounded porous media characterized by**

2 **Generalized Sub-Gaussian log-conductivity distributions**

3 **Guillem Sole-Mari^{1,2,3}, Monica Riva⁴, Daniel Fernàndez-Garcia^{2,3}, Xavier Sanchez-**
4 **Vila^{2,3}, Alberto Guadagnini⁴**

5 ¹Earth and Environmental Sciences Area, Geochemistry Department, Lawrence Berkeley
6 National Laboratory

7 ²Department of Civil and Environmental Engineering (DECA), Universitat Politècnica
8 de Catalunya

9 ³Associated Unit: Hydrogeology Group (UPC-CSIC)

10 ⁴ Department of Civil and Environmental Engineering (DICA), Politecnico di Milano

11
12
13 Corresponding author: Guillem Sole-Mari (gsolemari@lbl.gov)

14
15
16 **Highlights:**

- 17 • Log-conductivity (Y) is modeled as a Generalized Sub-Gaussian (GSG) field
- 18 • Average longitudinal spreading is smaller for GSG than for Gaussian Y fields
- 19 • Plume stretching in GSG fields is enhanced with respect to Gaussian fields
- 20 • Local dispersion masks identification of impacts of non-Gaussianity on transport
- 21 • The GSG nature of Y can be difficult to identify relying only on transport metrics

23 **Abstract**

24 There are increasing evidences that probability distributions and associated
25 statistical moments of a variety of hydrogeological and soil science variables and their
26 spatial increments display distinctive scale-dependent features that are not captured by a
27 typical Gaussian model. A Generalized Sub-Gaussian (GSG) model is able to capture key
28 aspects of this pattern. We present the results of a suite of computational analyses set in
29 a Monte Carlo framework and aimed at assessing the impact of a GSG structure of log
30 hydraulic conductivity (Y) on transport of a conservative solute through a three-
31 dimensional bounded porous medium under steady-state saturated Darcy flow. Our
32 results indicate that the longitudinal spreading of a plume is on average significantly
33 smaller for Sub-Gaussian than for Gaussian Y fields. Otherwise, the velocity field arising
34 from a Sub-Gaussian Y field induces enhanced plume stretching with respect to what can
35 be observed in a Gaussian Y setting, this aspect potentially influencing the strength of
36 solute mixing within these two types of conductivity domains. We also find that, in some
37 cases, it may be difficult to identify the nature of the underlying conductivity field relying
38 solely on observations of solute concentrations migrating within the system. In this
39 regard, we show that the action of local dispersion tends to mask the influence of Sub-
40 Gaussianity on major transport metrics.

41
42 **Keywords:** Porous media, flow and transport, heterogeneity, stochastic modeling, Sub-
43 Gaussian models

45

1. Introduction

46 Flow and transport in porous and fractured media are strongly affected by spatial
47 variability of hydraulic properties of the system. While it is virtually impossible to
48 characterize all details of such heterogeneity, a large body of literature has framed the
49 assessment of flow and transport processes taking place in heterogeneous porous media
50 on a (geo)statistical description of continuum-scale system attributes, major emphasis
51 being given on hydraulic conductivity (e.g., Dagan, 1984; Gelhar, 1993; Kitanidis, 1997;
52 Rubin, 2003). Stochastic analyses of the effects of heterogeneity on flow and transport in
53 porous media have been the subject of several investigations for over four decades.
54 Among the variety of approaches developed in this context, studies embedded in a
55 numerical Monte Carlo (MC) framework have enabled to explore subtle relationships
56 between the heterogeneous features of the subsurface and salient elements characterizing
57 flow and transport (e.g., Bellin et al., 1992; Naff et al., 1998a,b; Riva et al., 2008; de
58 Dreuzy et al., 2012; Pedretti et al., 2017; Siena and Riva, 2018 and references therein). A
59 typical Monte Carlo study comprises the analysis of a collection of outputs of interest that
60 are generated from one or multiple models, given uncertainty in their inputs. In the
61 context of flow and transport in heterogeneous porous media, this often involves the
62 generation of a collection of realizations of random hydraulic conductivity fields.

63 Several studies consider the logarithm of hydraulic conductivity ($Y(\mathbf{x}) = \ln K(\mathbf{x})$,
64 vector \mathbf{x} denoting space location) to be described through a Gaussian distribution,
65 characterized by a given variogram/covariance function describing the degree of spatial
66 dependence based only upon the distance between points; i.e., assuming second-order
67 stationarity. Rather, there is increasing evidence documenting the occurrence of distinct
68 non-Gaussian features characterizing the distributions of a variety of hydrogeological and
69 soil science variables. Key documented manifestations of such a behavior include the

70 observation that the distribution of spatial increments of such variables taken between
71 two points separated by a given spatial distance (lag) tend to be symmetric and to develop
72 heavier tails and sharper peaks as lag decreases. Such a behavior has been displayed
73 (among others) by log-hydraulic conductivity and permeability (Painter, 1996, 2001; Liu
74 and Moltz, 1997; Meerschaert et al., 2004; Siena et al., 2012, 2019; Riva et al., 2013a,
75 2013b; Guadagnini et al., 2018), electrical resistivity (Painter, 2001), vadose zone
76 hydraulic properties (Guadagnini et al., 2012, 2013, 2014), neutron porosity (Riva et al.,
77 2015a), sediment transport (e.g., Ganti et al., 2009), fully developed turbulence (Boffetta
78 et al., 2008), and micro-scale geochemical data (Siena et al., 2020).

79 The Generalized Sub-Gaussian (GSG) model introduced in the literature in the last
80 few years (Riva et al., 2015a, 2015b; Panzeri et al., 2016) has been shown to be capable
81 of capturing all of these aspects. According to the GSG model, the departure of the
82 distribution of a variable and its two-point increments from the Gaussian one is given by
83 the action of a (spatially uncorrelated) subordinator on an otherwise spatially correlated
84 Gaussian random field. This modeling strategy allows representing jointly within a
85 unique framework the above-documented behavior (as described by probability
86 distributions and/or moments) of a quantity and its incremental values and has been
87 successfully applied to the interpretation of main features displayed by various subsurface
88 attributes (Riva et al 2015a; Guadagnini et al., 2018; Siena et al., 2020, and references
89 therein).

90 These concepts have already been employed in preliminary analytical and
91 numerical studies of flow and transport in porous media whose log-conductivity is
92 characterized through a GSG model. Riva et al. (2017) present lead-order analytical flow
93 and transport solutions in unbounded GSG log-conductivity fields under mean-uniform
94 flow. Libera et al. (2017) rely on a numerical Monte Carlo framework to analyze the joint

95 effects of a GSG heterogeneous log-conductivity field and a temporally variable pumping
96 rate on solute breakthrough curves (BTCs) detected at a pumping well operating in a two-
97 dimensional domain. Besides spatial dimensionality, three major limitations can be
98 identified for the analytical study by Riva et al. (2017): (i) macrodispersion is the only
99 transport metric analyzed, (ii) the joint impact of local dispersivity combined with the
100 heterogeneous advection driven by Y is not evaluated, and (iii) output uncertainty due to
101 finite size of the medium is not considered.

102 Starting from the above studies, here we focus on the influence of randomly
103 heterogeneous GSG log-conductivity fields on transport processes taking place therein
104 and address the following two questions: (1) At what extent does Sub-Gaussianity (i.e.,
105 the degree of departure from a Gaussian behavior) of Y impact main features and
106 observables of solute transport driven by advection and local dispersion in three-
107 dimensional bounded heterogeneous porous media?; and (2) Would the Sub-Gaussian
108 nature of Y be clearly identifiable upon relying solely on the analysis of solute
109 concentration fields in these settings or could it be disguised into estimates of apparent
110 structural parameters of a Gaussian model? To address these issues, here we rest on a
111 suite of computational analyses framed in a Monte Carlo context and explore the effects
112 of a Sub-Gaussian distribution of Y on observable transport metrics associated with a
113 solute migrating within a bounded three-dimensional porous medium under the action of
114 advective and dispersive mechanisms and subject to a uniform (in the mean) flow.

115 The work is structured as follows. Section 2 is devoted to a synthesis of the
116 theoretical framework underlying the Sub-Gaussian model and of the computational
117 analyses performed. Presentation and discussion of the results are included in Section 3.
118 Finally, in Section 4 we enumerate our main conclusions.

119

2. Theoretical framework, methodology and set-up of scenarios

2.1 Generalized Sub-Gaussian model and stochastic generation of three-dimensional log hydraulic conductivity fields

We consider a three-dimensional porous medium where log-conductivity, $Y(\mathbf{x}) = \ln K(\mathbf{x})$, is described as a stationary random function of space with constant ensemble mean, $\langle Y \rangle$, and zero-mean random fluctuation, $Y'(\mathbf{x}) = Y(\mathbf{x}) - \langle Y \rangle$. The latter forms a Generalized Sub-Gaussian process, GSG, defined as (Riva et al. 2015a, b)

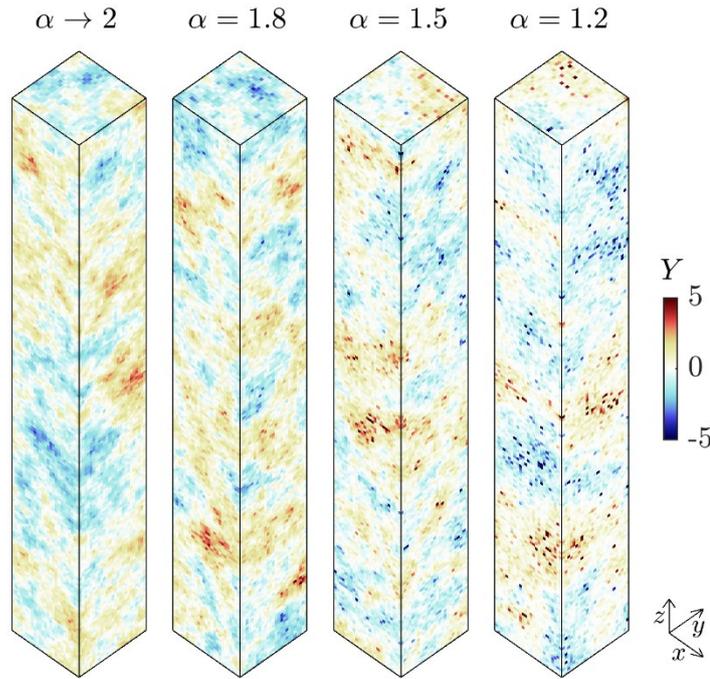
$$Y'(\mathbf{x}) = U(\mathbf{x})G(\mathbf{x}). \quad (1)$$

Here, $\mathbf{x} = [x, y, z]^T$, $G(\mathbf{x})$ is a zero-mean Gaussian spatially correlated second-order stationary field, and $U(\mathbf{x})$ is a random non-negative spatially uncorrelated stationary field that is independent of $G(\mathbf{x})$. The statistical properties of $Y'(\mathbf{x})$ vary with the subordinator $U(\mathbf{x})$ and with the Gaussian field $G(\mathbf{x})$. Details about the mathematical formulation and theoretical framework of the GSG model can be found in Riva et al. (2015a) and Siena et al. (2020), who present analytical expressions for probability density functions (*pdf*) and statistical moments of Y' and of the corresponding increments (ΔY) evaluated at various separation lags, including formulations for covariance and variogram functions as well as integral scale of Y' . It is worth noting that, regardless the distributional form of U , the variogram of Y' is characterized by an integral scale which is always smaller than the one associated with the Gaussian field G and a nugget effect, rendered by the product of the variance of G and the variance of U .

Here, we follow Riva et al. (2015a, 2017) and consider Monte Carlo realizations of $Y'(\mathbf{x})$ associated with a log-normal subordinator, i.e., $\ln(U) \sim \mathcal{N}[0, (2 - \alpha)^2]$, $\alpha < 2$. As such, $Y'(\mathbf{x})$ is characterized by two constant parameters (α, σ_G^2) and a correlation function ρ_G ; σ_G^2 and ρ_G being variance and correlation function of $G(\mathbf{x})$, respectively. Note that

144 $Y'(\mathbf{x})$ becomes Gaussian and coincides with $G(\mathbf{x})$ when $\alpha \rightarrow 2$. For completeness,
 145 Appendix A includes a brief summary of the key analytical formulations and properties
 146 of the GSG model we consider in this study.

147 We generate three-dimensional realizations of $Y(\mathbf{x})$ of size $L_x \times L_y \times L_z$ by setting
 148 $\langle Y \rangle = 0$, $\sigma_Y^2 = 1$ (where $\sigma_Y^2 = e^{2(2-\alpha)^2} \sigma_G^2$) and adopting an isotropic exponential
 149 correlation function for $G(\mathbf{x})$, i.e. $\rho_G = e^{-s/l_G}$; s and l_G being the separation distance
 150 (lag) between two locations and the integral scale of $G(\mathbf{x})$, respectively. We set $L_x =$
 151 $L_y = 5l_Y$ and $L_z = 40l_Y$ in all realizations (z being the dominant direction of flow, see
 152 Section 2.2), and $l_Y = e^{-(2-\alpha)^2} l_G$ denoting the integral scale of Y (see also Appendix A).
 153 The spatial grid employed for the generation involves 125,000 cubic blocks of size $l_Y/5$.
 154 Figure 1 depicts examples of $Y(\mathbf{x})$ realizations obtained for differing values of α . To
 155 appreciate the nature of the random log-conductivity fields considered, we illustrate in
 156 Appendix A the main features of the probability density functions associated with Y' and
 157 its spatial increments ΔY evaluated at various separation lags for the fields depicted in
 158 Figure 1. Figure A1 depicts the probability density function of Y' , $f_{Y'}$, for the three values
 159 of α considered in Figure 1 together with the Gaussian ($\alpha \rightarrow 2$) distribution.
 160 Corresponding *pdfs* of spatial increments, $f_{\Delta Y}$, evaluated at short, intermediate and large
 161 (with respect to l_Y) separation lags are depicted in Figure A2. The pattern associated with
 162 the behavior of peaks and tails of these *pdfs* can be described quantitatively by analyzing
 163 deviations from Gaussianity, clearly revealed by the excess standardized kurtosis,
 164 reported in Figure A3. It is clear that the peak of the *pdf* of Y' or ΔY grows sharper and
 165 the associated tails become heavier as α decreases, thus evidencing the deviations of the
 166 fields we consider from a Gaussian behavior.



167

168 *Figure 1: Example of $Y(x)$ individual realizations for various degrees of Sub-Gaussianity*
 169 *($\alpha \rightarrow 2$ being equivalent to a Gaussian distribution).*

170

171 As a remark, we note that the variance of Y in geologic porous media has sometimes
 172 been reported to attain markedly high values, e.g., $\sigma_Y^2 \geq 4$ (Fogg & Zhang, 2016). It is
 173 well recognized that this stems from a homogenization of conductivity values within a
 174 unique population, while they actually pertain to diverse geological facies. One may
 175 alternatively represent the system as a composite medium, whose internal architecture is
 176 characterized by the presence of distinct facies, within which hydraulic conductivity can
 177 be spatially heterogeneous. As such, conductivity variance within a given geomaterial
 178 can be mild (Winter & Tartakovsky, 2000, 2002). Hence, the selected mild value of the
 179 log-conductivity variance in this work ($\sigma_Y^2 = 1$) is deemed as representative of the natural
 180 variability contained within a geological unit, which can potentially be depicted through
 181 statistically stationary heterogeneity models (Winter et al., 2003) like the one we consider.
 182 The relative size of the cross-section (i.e., $L_x = L_y = 5I_Y$) is set to be large enough to
 183 capture the main features of the log-conductivity distributions, while being small enough

184 to generate variability in the results between realizations with identical configuration,
185 similar to what one would expect from a collection of finite-sized samples belonging to
186 the same geological material or unit.

187 We further remark that here we focus on the analysis of the joint effects on transport
188 of the strength of (i) the departure of Y from a Gaussian model (as rendered by the
189 parameter α related to the subordinator U), and (ii) the relative importance of advective
190 and dispersive processes in Sub-Gaussian Y fields. Future studies can tackle also the
191 potential influence of the degree of heterogeneity, as expressed in terms of σ_Y^2 , or consider
192 representing a log-conductivity field involving multiple subdomains, with diverse
193 geomaterials through a multimodal distribution, such as a mixture of Sub-Gaussian *pdfs*.

194 **2.2 Flow and transport model**

195 Steady-state single phase flow is described through mass balance and Darcy's
196 equations within the heterogeneous $K(\mathbf{x})$ fields generated as described in Section 2.1.

$$197 \quad \nabla \cdot \mathbf{q}(\mathbf{x}) = 0, \quad \mathbf{q}(\mathbf{x}) = -K(\mathbf{x})\nabla H(\mathbf{x}). \quad (2)$$

198 Here, $H(\mathbf{x})$ is hydraulic head and $\mathbf{q}(\mathbf{x})$ is Darcy flux. We set impermeable boundary
199 conditions on the lateral sides of the domain ($x = 0, L_x; y = 0, L_y$). Flow is forced along
200 direction z by setting a constant mean flow, \bar{q} , at the bottom of the sample, and by setting
201 a prescribed head at the top boundary. A homogeneous buffer zone of length $2l_Y$ and with
202 $Y = \langle Y \rangle$ is placed at both ends of the system (Figure 2) to stabilize the inflow condition.
203 These boundary conditions correspond to what one would typically impose in an
204 experimental set-up, head at the inlet being practically uniform in the transverse direction
205 due to the presence of the homogeneous buffer region. Imposing a fixed mean
206 longitudinal flow \bar{q} for all realizations enables us to study transport phenomena which are
207 driven by relative spatial fluctuations of $q(\mathbf{x})$ around the same mean value.

208 At the initial time ($t = 0$) a flux-averaged pulse injection of solute takes place
 209 across the bottom of the heterogeneous region, i.e., along the horizontal plane $z = 0$.
 210 Solute transport is described by the Advection Dispersion Equation (ADE)

$$211 \quad \frac{\partial c}{\partial t}(\mathbf{x}, t) = \nabla \cdot (\mathbf{D}(\mathbf{x})\nabla c(\mathbf{x}, t)) - \nabla \cdot (\mathbf{v}(\mathbf{x})c(\mathbf{x}, t)), \quad (3)$$

212 where $c(\mathbf{x}, t)$ is resident solute concentration (i.e., solute per unit volume of the porous
 213 medium), and $\mathbf{v}(\mathbf{x}) = \mathbf{q}(\mathbf{x})/\phi$, ϕ being porosity which we consider constant in this study.
 214 In the context of modeling transport in porous media, local dispersion is an upscaled
 215 apparent process that accounts for the combined effect of unresolved pore-scale velocity
 216 variations and molecular diffusion (e.g., Bear, 1972). At high pore-scale Péclet numbers,
 217 it is often considered to be proportional to the fluid velocity. To simplify the problem, we
 218 assumed an isotropic local dispersion tensor $\mathbf{D}(\mathbf{x})$,

$$219 \quad \mathbf{D}(\mathbf{x}) = D(\mathbf{x})\mathbf{1}_3, \quad D(\mathbf{x}) = a|\mathbf{v}(\mathbf{x})|, \quad (4)$$

220 where $\mathbf{1}_3$ is the 3×3 identity matrix, a is a constant dispersivity, and $|\cdot|$ denotes the norm
 221 of a vector. The advection and dispersion timescales for a characteristic distance l_Y can
 222 be compared through the Péclet number, $Pe = l_Y/a$, higher values of Pe corresponding
 223 to a lower relative significance of the local dispersion process. In this study we explore
 224 settings associated with values of Pe ranging from $Pe = 10$ to $Pe \rightarrow \infty$.

225 The steady-state flow problem (2) is solved by the widely tested finite difference
 226 code MODFLOW 2005 (Harabaugh, 2005) using the Preconditioned Conjugate Gradient
 227 method (PCG). The output cell-interface fluxes are then used to solve the ADE by means
 228 of the Random Walk Particle Tracking method suggested by LaBolle et al. (2000). The
 229 latter solves the so-called Generalized Stochastic Differential Equations (GSDE) and is
 230 based on the following particle movement

231
$$\mathbf{X}_p(t + dt) = \mathbf{X}_p(t) + \mathbf{v}(\mathbf{X}_p(t)) dt + \sqrt{2D(\mathbf{X}_p(t) + \mathbf{\Lambda}_p(t))} dt \boldsymbol{\xi}_p(t), \quad (5)$$

232
$$\mathbf{\Lambda}_p(t) = \sqrt{2D(\mathbf{X}_p(t))} dt \boldsymbol{\xi}_p(t), \quad (6)$$

233 where $\mathbf{X}_p(t)$ is the position of particle p at time t , and $\boldsymbol{\xi}_p(t)$ is a vector of 3 (the number
 234 of spatial dimensions) independent uncorrelated standard normal random numbers.
 235 Velocities at any point inside a numerical cell Ω are evaluated by linear interpolation

236
$$\mathbf{v}(\mathbf{x} \in \Omega) = (1 - \hat{\mathbf{x}}) \odot \mathbf{v}_{\Omega}^- + \hat{\mathbf{x}} \odot \mathbf{v}_{\Omega}^+, \quad (7)$$

237 where $\hat{\mathbf{x}} = (\mathbf{x} - \mathbf{x}_{\Omega}^-) \oslash (\mathbf{x}_{\Omega}^+ - \mathbf{x}_{\Omega}^-)$; operators “ \odot ” and “ \oslash ” correspond to the
 238 elementwise Hadamard product and division, respectively; \mathbf{x}_{Ω}^- and \mathbf{x}_{Ω}^+ are vectors
 239 containing the lower and upper cell interface location (for each direction); and \mathbf{v}_{Ω}^- and \mathbf{v}_{Ω}^+
 240 are vectors whose entries are the corresponding cross-interface flow velocities.

241 The GSDE method (5) is particularly accurate and efficient when representing
 242 isotropic local dispersion characterized by strong local variations. Therefore, this method
 243 is suitable for reproducing transport features even within porous media characterized by
 244 local highly non-uniform flow conditions. We refer to Salamon et al. (2006) for a
 245 comprehensive review of random-walk methods for solving the advection-dispersion
 246 equation in spatially heterogeneous settings.

247 Solution of the flow problem (2) may require using a spatial grid finer than that
 248 introduced in Section 2.1 and according to which log-conductivity is generated. The grid
 249 block size, Δ , as well as the number of particles, n , and the time step, Δt , considered for
 250 the implementation of the GSDE (5) have been selected as detailed in the following. The
 251 spatial grid resolution (i.e., $1/\Delta$) and the number of particles n are increased until
 252 convergence is attained, i.e., until no significant changes are observed in the numerical
 253 solution (as quantified through the metrics/indicators detailed in Section 3). The time step

254 Δt is adapted dynamically, to fulfill the condition which is most restrictive between $\Delta t \leq$
255 $\mathcal{C}\Delta/|\mathbf{v}|$ and $\Delta t \leq \mathcal{P}^2\Delta^2/2D$, \mathcal{C} and \mathcal{P} being numerical discretization parameters which
256 are adjusted in the same fashion as Δ and n . The numerical model has been found to give
257 an accurate solution of the flow and transport problem by setting $n = 2^{18}$ (= 262,144),
258 $\mathcal{C} = \mathcal{P} = 0.125$, and subdividing each conductivity block into 64 elements, i.e., $\Delta =$
259 $I_Y/20$, giving rise to a computational grid formed by 8 million cells.

260 All particles are assigned an identical mass M/n and are injected according to a
261 flux-weighting scheme on the horizontal plane at $z = 0$. Flow and transport are simulated
262 across $N = 5,000$ Monte Carlo realizations of $Y(\mathbf{x})$ (see Section 2.1) for $\alpha = 1.2, 1.5, 1.8,$
263 and $\alpha \rightarrow 2.0$ (i.e., Gaussian field), and for $Pe = 10, 100, 1000,$ and $Pe \rightarrow \infty$ (i.e., pure
264 advection), for a total of 80,000 simulations. The latter have been performed using 180
265 cores from the Marconi A1 partition of the high-performance computer center CINECA
266 (<https://www.cineca.it/en>). The total computational time required for our study is about
267 100,000 core hours.

268 In Section 3, we analyze the spatial and temporal moments, as well as additional
269 key metrics (see also Appendix B for details), of

$$270 \quad C_r(z, t) = \frac{\phi}{M} \int_0^{L_x} \int_0^{L_y} c_r(\mathbf{x}, t) \, dy \, dx, \quad (8)$$

$$271 \quad J_r(z, t) = \frac{\phi}{M} \int_0^{L_x} \int_0^{L_y} \mathbf{n}^T \cdot [\mathbf{v}_r(\mathbf{x})c_r(\mathbf{x}, t) - D_r(\mathbf{x})\nabla c_r(\mathbf{x})] \, dy \, dx, \quad (9)$$

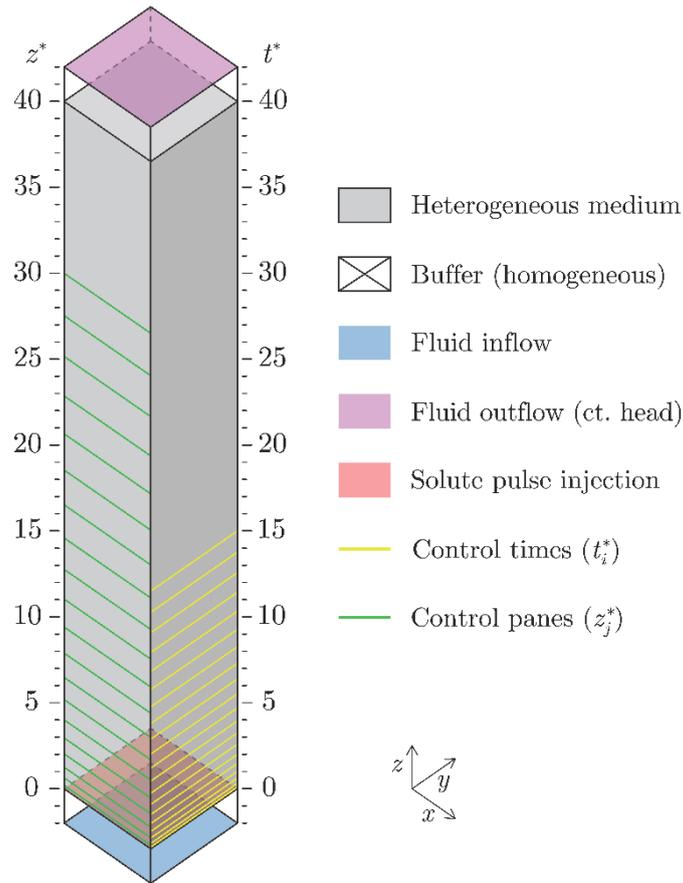
272 where $\mathbf{n}^T = [0,0,1]$, and subscript r refers to the realization number, i.e., $r = 1, \dots, N$.
273 The quantity $C_r(z, t)$ is the longitudinal density of solute mass normalized by the total
274 mass injected (hence it integrates to 1 along z). The quantity $J_r(z, t)$ is the net temporal
275 flux of mass through a horizontal plane, also normalized by the total mass (hence it

276 integrates to 1 with t). Hereafter, we refer to $C_r(z, t)$ as the section-integral
 277 concentrations, and to $J_r(z, t)$ as the breakthrough curve (BTC).

278 Results shown in Section 3 are illustrated in terms of dimensionless times, t^* , and
 279 longitudinal positions, z^* , defined as

$$280 \quad t^* = \frac{\bar{v}}{I_Y} t, \quad z^* = \frac{z}{I_Y}, \quad (10)$$

281 where $\bar{v} = \bar{q}/\phi$. Note that t^* corresponds to the number of integral scales of Y traversed
 282 by the center of mass of the plume. The evolution of the solute plume is monitored by
 283 recording particle positions at 20 increasing dimensionless times t_i^* , $i \in \{1, 2, \dots, 20\}$,
 284 with $t_1^* = 0.1$, $t_{20}^* = 15$, and $t_i^* = t_1^* i^{\log_2(t_{20}^*/t_1^*)}$. It has been verified that all particles
 285 are still located inside the heterogeneous domain in all MC realizations at t_{20}^* . We record
 286 particle passage times through selected planes perpendicular to the mean flow direction
 287 at distance z_j^* ($j \in \{1, 2, \dots, 20\}$), with $z_1^* = 0.2$, $z_{20}^* = 30$, and $z_j^* = z_1^* j^{\log_2(z_{20}^*/z_1^*)}$
 288 from the origin. The design of the setup is depicted in Figure 2.



289

290 *Figure 2: Design of the three-dimensional numerical simulations and scheme for*
 291 *recording the results. The control times are shown graphically in terms of the*
 292 *corresponding mean displacement of the fluid. Empty non-shaded regions at both ends*
 293 *correspond to the homogeneous buffers.*

294

295 Our simulations yield a collection (obtained by sampling across MC realizations)
 296 of (i) solute concentration profiles at a given observation time, $C_r(z, t_i)$, or (ii) solute
 297 breakthrough curves (BTCs) at a given transverse section, $J_r(z_j, t)$. We follow Liu and
 298 Müller (2004) and Lu and Stauffer (2012) and note that, in this context, a point-wise
 299 average is not a proper estimate of the mean concentration profile (or curve). For instance,
 300 such an average tends to overestimate the spread of the tails and may also underestimate
 301 the peak. This issue is addressed by employing a percentile average over the cumulative
 302 forms of all curves (see Lu and Stauffer (2012) for details). The shape of the ensuing
 303 average curves is generally more similar than the point-wise average to the one of curves

304 corresponding to individual realizations. Here, we evaluate percentile averages $\bar{C}(z, t_i)$
305 and $\bar{J}(z_j, t)$ upon organizing simulation results (given in terms of particle positions and
306 arrival times, respectively) in increasing order, and then averaging across realizations, to
307 finally reconstruct the density of data by Kernel Density Estimation (KDE). Details on
308 this procedure are offered in Appendix B.1 (see also Appendix D for details on KDE).
309 For brevity, hereafter we refer to $\bar{C}(z, t_i)$ and $\bar{J}(z_j, t)$ as the average longitudinal
310 concentration profile and the average BTC, respectively. Note that these percentile-
311 averaged ensemble quantities are denoted with an overline, as opposed to angular brackets
312 (i.e., $\langle \rangle$) indicating arithmetic ensemble average, which we employ for the scalar metrics
313 analyzed (see eq. (11) in Section 3).

314

3. Results and discussion

315 We start by analyzing our results in terms of average spatial and temporal dynamics
316 of the solute on the basis of the numerical Monte Carlo framework described in Section
317 2. We do so by considering the setting corresponding to pure advective mechanisms, i.e.,
318 $Pe \rightarrow \infty$ (Section 3.1.1), to then include the assessment of the impact of considering a
319 finite value for Pe (Section 3.1.2). Section 3.2 illustrates the analysis related to the
320 complete set of statistical distributions of the modeling goals of interest.

321

3.1. Average space-time dynamics of the solute

322 Hereafter, we resort to the notation $\langle b \rangle$ to indicate the arithmetic average of a metric
323 b of interest, as obtained across the collection of N Monte Carlo realizations, i.e.,

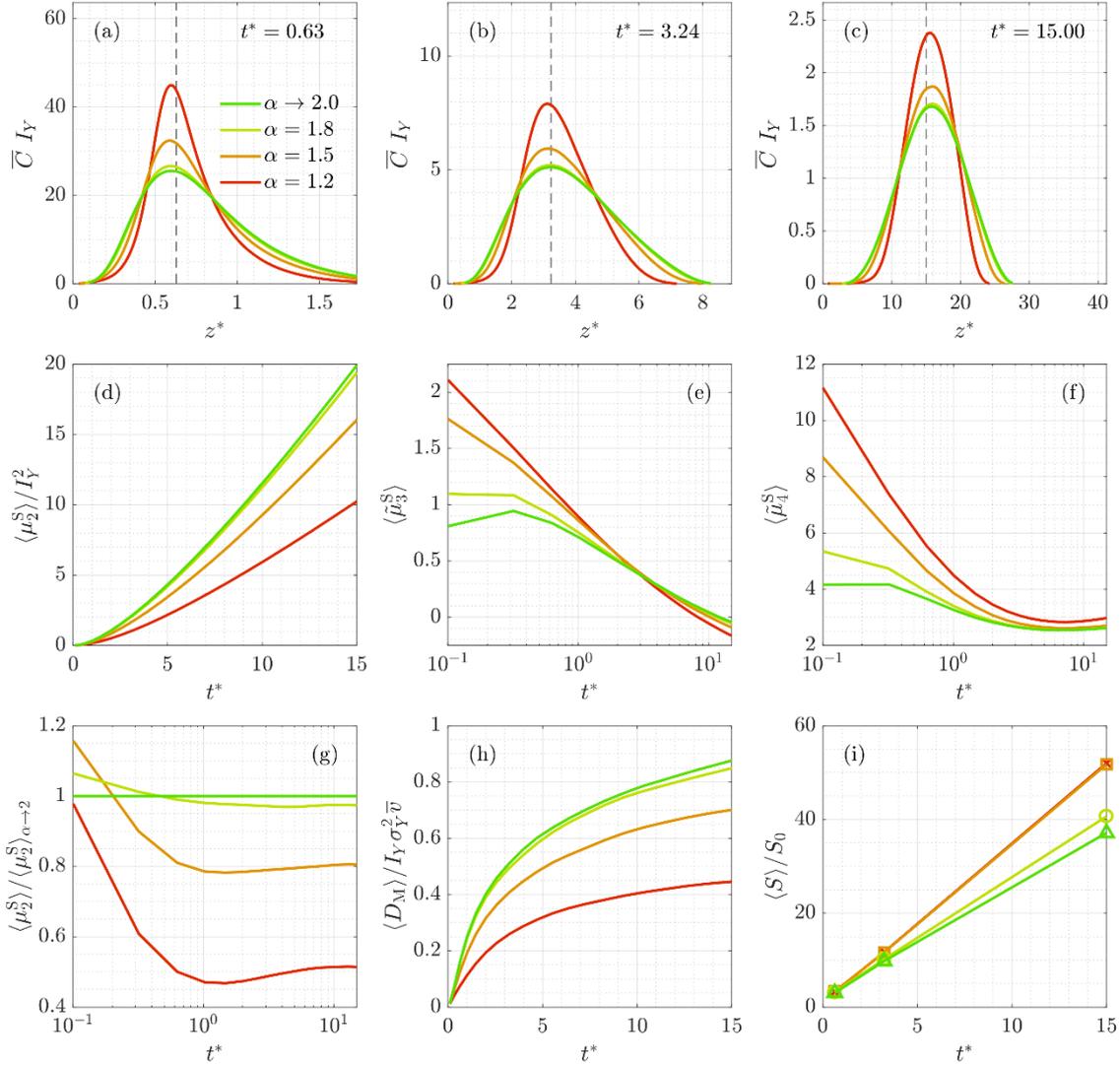
$$324 \quad \langle b \rangle = \frac{1}{N} \sum_{r=1}^N b_r, \quad (11)$$

325 b_r being the value of metric b in the r^{th} realization.

326

3.1.1. Average spatial and temporal behavior under pure advection

We start by considering a fully advective transport scenario ($Pe \rightarrow \infty$).



328

329 *Figure 3: Average temporal evolution of the solute distribution in the longitudinal*
 330 *direction: (a-c) Longitudinal concentration profile at various times (dashed gray line*
 331 *corresponds to $z = \bar{v}t$), (d) spatial variance, (e) spatial skewness, (f) spatial kurtosis, (g)*
 332 *spatial variance normalized by the Gaussian case, (h) macrodispersion coefficient, and*
 333 *(i) advective stretching.*

334

335 Figures 3(a-c) depict the average longitudinal concentration profile, \bar{C} , evaluated as
 336 detailed in Section 2.2 and Appendix B.1, for various values of α at increasing values of
 337 t^* (the vertical dashed gray lines correspond to the average fluid displacement, $z = \bar{v}t$).
 338 In the following, we discuss the effect of α on the temporal evolution of the plume by

339 analyzing these concentration profiles in combination with various metrics of interest
340 depicted within Figures 3(d-i).

341 The longitudinal extent of the region where one can find (on average) significant
342 concentration values tends to decrease with departure of Y from Gaussianity (i.e.,
343 decreasing α). A quantification of this qualitative behavior can be offered through the
344 average centered second moment of the longitudinal profile of section-integral
345 concentrations, $\langle \mu_2^S \rangle$, which is an effective measure of longitudinal dispersion (the latter
346 being different from the local dispersion, as quantified by \mathbf{D} in (3)), and whose temporal
347 evolution is depicted in Figure 3(d). Details on the definition and computation of spatial
348 and temporal moments in each MC realization are provided in Appendix B.2. Stronger
349 degrees of Sub-Gaussianity (i.e., lower values of α) are associated with decreased values
350 of $\langle \mu_2^S \rangle$. Figure 3(g) also reveals that the ratio between any of the $\langle \mu_2^S \rangle$ curves and the one
351 corresponding to the Gaussian case $\langle \mu_2^S \rangle_{\alpha \rightarrow 2}$ tends to become constant after a few
352 characteristic times (i.e., for $t^* > 1$). Quantities $1 - \langle \mu_2^S \rangle / \langle \mu_2^S \rangle_{\alpha \rightarrow 2}$ and $(2 - \alpha)^2$ display
353 a nearly perfect linear correlation at late times (with a slope of ~ 0.78 , details not shown).
354 We note that dispersion in Gaussian Y fields is, under some assumptions, proportional to
355 σ_Y^2 (Dagan, 1984). As such, our results imply that, at long times, relying solely on an
356 analysis of the macrodispersive behavior of a solute transported by pure advection within
357 a saturated porous formation may not allow to distinguish whether the underlying log
358 hydraulic conductivity distribution is (i) Sub-Gaussian ($\alpha < 2$) with variance σ_Y^2 , or (ii)
359 Gaussian with some variance σ_Y^{*2} being lower than σ_Y^2 . Yet, as shown in the following,
360 these two models lead to quite different estimates in a number of observables.

361 Figure 3(h) depicts the evolution of the average longitudinal (macro)dispersion
362 coefficient, $\langle D_M \rangle$, computed for each realization as

363
$$D_{M,r}(t_i) = \frac{1}{2} \left. \frac{\partial \mu_{2,r}^S}{\partial t} \right|_{t_i} \cong \frac{\mu_{2,r}^S(t_i) - \mu_{2,r}^S(t_{i-1})}{2(t_i - t_{i-1})}. \quad (12)$$

364 This quantity is normalized by $\sigma_Y^2 I_Y \bar{v}$, which is the long-time asymptotic value of D_M in
 365 an infinite medium with Gaussian log-conductivities characterized by low spatial
 366 heterogeneity. Note that an asymptotic dispersion regime is not fully reached by $t^* = 15$.
 367 Riva et al. (2017) show that $D_M/\sigma_Y^2 I_Y \bar{v} \rightarrow 1$ (at a rate which depends on α) in two-
 368 dimensional Sub-Gaussian Y fields and in the long-time limit, this result corresponding
 369 to analytical findings by Dagan (1982, 1984, 1989) and Rubin (1990a,b) for Gaussian Y
 370 fields under the same assumptions on flow. Assuming that a corresponding long-time
 371 limit holds also in three-dimensional unbounded systems, the results depicted in Figure
 372 3(h) suggest that the impact of the presence of boundaries on the dynamics of D_M (i.e.,
 373 on the departure of $D_M/\sigma_Y^2 I_Y \bar{v}$ from 1) tends to increase with decreasing α .

374 Further inspection of Figures 3(a-c) reveals that low values of α yield a more
 375 pronounced peak for longitudinal concentration profiles. Additionally, the standardized
 376 distribution of vertical locations of the solute exhibits longer tails as α decreases (not
 377 shown). These features are markedly more evident at early times ($t^* = 0.63$). This result
 378 may be related to the observation that distributions of Y increments (which are related to
 379 spatial variations of velocities, that ultimately control longitudinal solute spreading) in
 380 Sub-Gaussian fields display heavy tails and sharp peaks at short distances (Riva et al.,
 381 2015a). Differences between curves associated with differing values of α become less
 382 apparent at late times ($t^* = 15$). One can also detect an increasing degree of asymmetry
 383 of the plume as α decreases, both at early and late times.

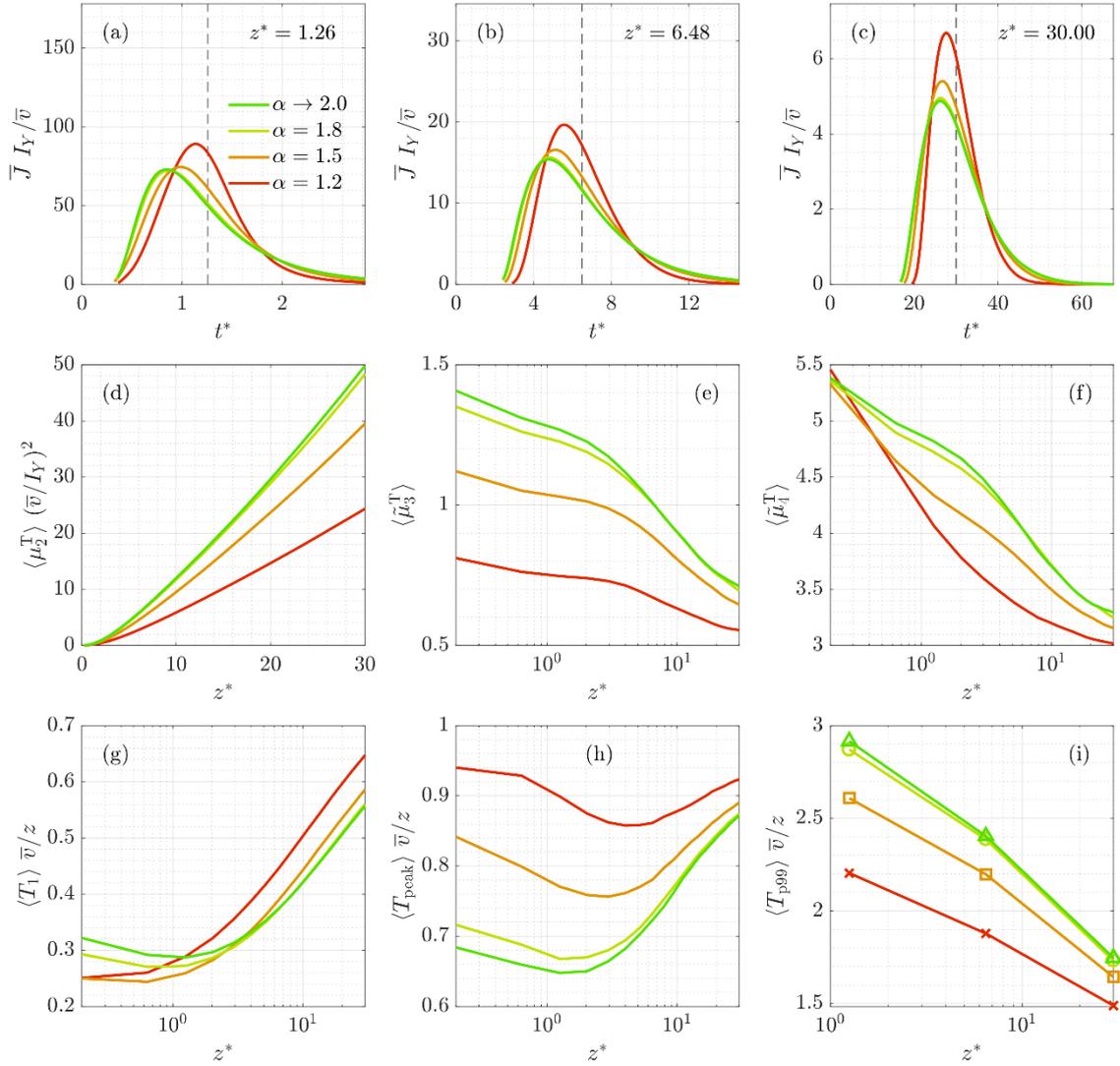
384 All of these qualitative observations are quantified by the analysis of the temporal
 385 evolutions of the normalized spatial moments of the plume, namely the spatial skewness
 386 $\tilde{\mu}_3^S$ (Figure 3(e)) and kurtosis $\tilde{\mu}_4^S$ (Figure 3(f)) (see Appendix B.2 for details about their

387 evaluation). We note that low α values tend to promote more skewed (i.e., asymmetric)
388 plumes, being positively skewed at early times (a behavior mainly driven by high
389 velocities) and (slightly) negatively skewed at late times (mainly driven by low
390 velocities). Figure 3(f) suggests that lower values of α are associated (on average) with
391 more kurtotic plumes. Differences between (average) spatial skewness and kurtosis of
392 plumes evolving in Gaussian and Sub-Gaussian cases are substantial at early times and
393 tend to become mild as time increases, consistent with the observation that the distribution
394 of Y spatial increments in a Sub-Gaussian field tends to approach the Gaussian one as lag
395 increases.

396 Figure 3(i) depicts the temporal evolution of the average advective stretching, i.e.,
397 $\langle S \rangle / S_0$. This metric quantifies hydrodynamic deformation through the evaluation of the
398 aggregated temporal growth of a fluid surface, S , represented by a collection of particles
399 displaced by advection and originally arranged on the horizontal plane $z = 0$ (i.e.,
400 corresponding to $S_0 = L_x L_y$). All details are provided in Appendix B.4. On average, the
401 velocity field associated with a Sub-Gaussian Y field induces more stretched plumes than
402 those evolving through a Gaussian setting. It is worth noting that this takes place despite
403 the longitudinal spreading (represented by μ_2^S) being (sometimes significantly) smaller
404 for Sub-Gaussian than for Gaussian fields. One may compare these results with those
405 reported by Le Borgne et al. (2013, 2015) in two-dimensional Gaussian Y fields.
406 Numerical simulations performed by these authors document that the mean elongation of
407 a solute line displays a power-law temporal behavior with an exponent ranging between
408 0.65 and 1.25 for σ_Y^2 increasing from 0.25 to 4.0. Interpreting the temporal evolution of
409 $\langle S \rangle / S_0$ through a power-law behavior (as in the above-mentioned Gaussian Y settings),
410 its (best-fit) exponent increases for decreasing values of α (while $\sigma_Y^2 = 1$ is kept
411 constant), from 0.85 in the Gaussian case to 0.95 for $\alpha \leq 1.5$. These results are probably

412 related to the observation that low values of α promote marked changes in longitudinal
413 velocity values at very short transverse distances (details not shown here for brevity), thus
414 inducing enhanced degrees of short-range fluid stretching. This effect seems to be capped
415 at strong degrees of Sub-Gaussianity (i.e., $\alpha = 1.2$), for which the average surface
416 increase is almost identical to the case of moderate-to-high Sub-Gaussian behavior (i.e.,
417 $\alpha = 1.5$).

418 The stark differences observed between the effects of a Sub-Gaussian Y field on the
419 longitudinal dispersion (Figure 3(d)) and on advective stretching (Figure 3(i)) could
420 indicate that the increased longitudinal dispersion associated with Gaussian Y fields with
421 respect to a Sub-Gaussian setting may not necessarily translate into enhanced mixing, the
422 latter being strongly linked to stretching (see, e.g., Dentz et al., 2011; Le Borgne et al.,
423 2013, 2015).



424

425 *Figure 4: Average spatial evolution of the solute BTCs across planes orthogonal to the*
 426 *mean flow direction: (a-c) BTCs at various distances from the inlet (dashed gray line*
 427 *corresponds to $t = z/\bar{v}$), (d) temporal variance, (e) temporal skewness, (f) temporal*
 428 *kurtosis, (g) first arrival, (h) arrival of the peak, and (i) arrival of the 99% of the total*
 429 *mass.*

430

431 Figures 4(a-c) depict the average temporal evolution of the solute mass flux,
 432 representing the solute breakthrough curve (BTC), for various values of α and at
 433 increasing distances from the inlet. The average BTCs are visibly narrower for lower
 434 values of α . This is related to the observation that the spatial evolution of the centered
 435 second temporal moment, depicted in Figure 4(d), has been found to be qualitatively very

436 similar to its temporal evolution counterpart (Figure 3(d)) in all cases analyzed, a feature
437 which was also noted by Fernández-García et al. (2005) for transport in three-
438 dimensional, Gaussian Y fields. The average BTC is also visibly more symmetric for
439 lower values of α , the arrival of the peak being less advanced in time with respect to the
440 mean arrival time z/\bar{v} (marked on Figures 4(a-c) by a gray vertical dashed line) than its
441 counterpart based on Gaussian Y fields. This latter feature is particularly evident at short
442 distances from the inlet ($z^* = 1.26$).

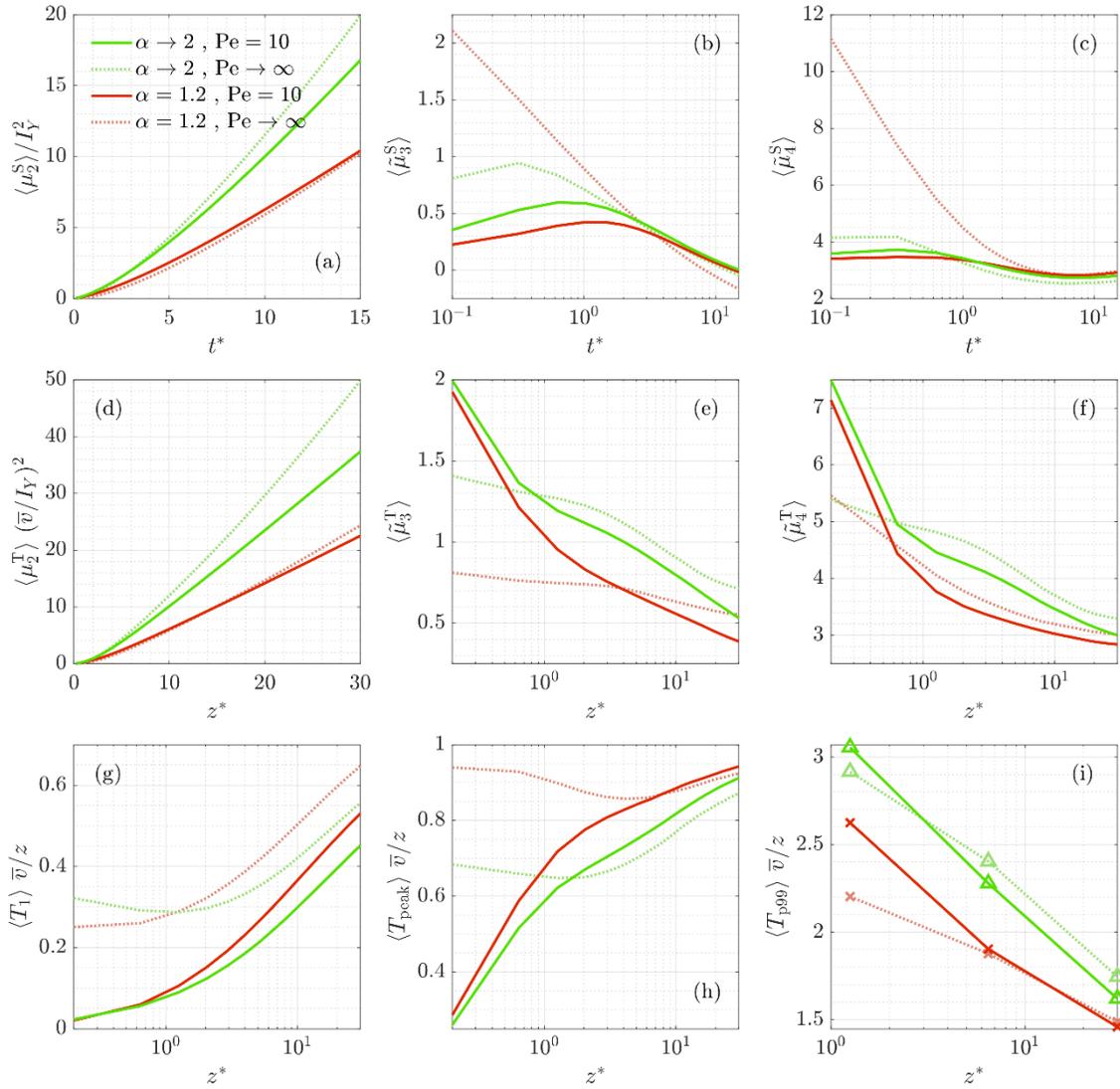
443 For all values of α , the BTCs become more symmetric as z^* increases. This
444 behavior can be quantified through the (average) skewness of the BTCs, i.e., $\langle \hat{\mu}_3^T \rangle$, whose
445 spatial evolution along the column (Figure 4(e)) denotes a decreasing trend of the
446 temporal skewness (*i*) with increasing distance from the inlet and (*ii*) with departure from
447 Gaussian behavior of the underlying Y field. A similar feature is observed in Figure 4(f)
448 for the average temporal kurtosis, $\langle \hat{\mu}_4^T \rangle$.

449 Figures 4(g), 4(h) and 4(i) complement the description by depicting the average
450 longitudinal distribution of key arrival times, corresponding to the time of arrival of the
451 first particle, T_1 , the BTC peak, T_{peak} , and the 99th percentile, T_{p99} , respectively. These
452 quantities are normalized by the mean arrival time (i.e., z/\bar{v}) for ease of interpretation. It
453 is generally recognized that first arrival times are strongly linked to connectivity metrics,
454 such as, e.g., the least resistance path, which is the path connecting two points or surfaces
455 with the largest associated effective hydraulic conductivity (e.g., Gómez-Hernández and
456 Wen, 1998; Rizzo and De Barros, 2017). In this context, Figure 4(g) suggests that Sub-
457 Gaussian Y fields are more connected than their Gaussian counterparts at short distances
458 ($z^* \leq 1$), the trend being reversed at longer distances ($z^* \gg 1$). This finding could be
459 related to the aforementioned nugget effect displayed by the longitudinal velocity
460 covariances associated with low α values, which would increase the likelihood of

461 velocities changing abruptly at very short distances, and possibly decrease the likelihood
462 of occurrence of persistent high velocity paths.

463 As α decreases, the peak of the BTC (Figure 4(h)) is on average closer to the mean
464 arrival time, a finding in agreement with the aforementioned reduced temporal skewness.
465 This implies that in Sub-Gaussian fields the largest concentrations tend to be observed
466 (on average) appreciably later than in Gaussian fields, especially for short distances (i.e.,
467 $z^* \approx 1$). The average longitudinal distribution of the 99th percentile of arrival times, T_{p99}
468 (Figure 4(i)), shows that the BTC tends to be more compacted for lower values of α . All
469 of these findings suggest that, as compared to a Gaussian distribution of log-
470 conductivities, in the presence of a Sub-Gaussian Y field one can (on average) expect (i)
471 a generally delayed arrival of the mass of a flux-injected solute to crossing planes located
472 a few integral scales away from the injection plane (i.e., for $z^* > 5$), as expressed in terms
473 of both the first arrival and the arrival of the peak; (ii) more pronounced peaks; and (iii)
474 an earlier breakthrough of the majority (99%) of the solute mass.

475 **3.1.2. Analysis of the impact of local dispersion**



476

477 *Figure 5: Effect of local dispersion ($Pe = 10$) compared to pure advection ($Pe \rightarrow \infty$) on*
 478 *the results with Gaussian and Sub-Gaussian ($\alpha = 1.2$) Y fields, for (a) spatial variance*
 479 *(macrodispersion), (b) spatial skewness, (c) spatial kurtosis, (d) temporal variance, (e)*
 480 *temporal skewness, (f) temporal kurtosis, (g) first arrival, (h) arrival of the peak, and (i)*
 481 *arrival of the 99% of the solute mass.*

482

483 Here, we concentrate on the scenario corresponding to the lowest value of α
 484 considered (i.e., $\alpha = 1.2$; strongly non-Gaussian Y), compared to the Gaussian Y scenario
 485 (i.e., $\alpha \rightarrow 2$), to clearly illustrate the effects of local dispersion (i.e., finite values of Pe)
 486 on transport. In Figure 5 we compare the spatial and temporal variance, skewness and
 487 kurtosis as well as the longitudinal distribution of key arrival times (defined in Section

488 3.1.1) obtained for $Pe \rightarrow \infty$ (dotted curves) with those corresponding to $Pe = 10$ (solid
489 curves).

490 As one could expect, local dispersion typically reduces the difference between the
491 results for Gaussian and Sub-Gaussian settings. In the Gaussian case (green curves), one
492 can see that $\langle \mu_2^S \rangle$ (Figure 5a) decreases due to the effect of local dispersion. This effect is
493 increasingly evident for decreasing values of Pe (not shown). Such a behavior has been
494 documented (analytically, numerically, and experimentally) in previous studies (e.g.,
495 Dentz et al., 2000, 2002, 2003; Dartois et al., 2018; Gist et al., 1990) and is related to the
496 observation that local dispersion favors particles to sample a wider range of velocities,
497 thus reducing the likelihood of sampling only extreme (low or high) velocity paths.
498 Conversely, this effect is not apparent in the strongly Sub-Gaussian fields we examine.
499 However, the aforementioned similarity between the temporal evolution of $\langle \mu_2^S \rangle$ and the
500 spatial evolution of $\langle \mu_2^T \rangle$ and the results of Figure 5(d) suggest that the reduction in
501 longitudinal dispersion observed for $Pe = 10$ with respect to $Pe \rightarrow \infty$ will eventually take
502 place in time also for $\alpha = 1.2$.

503 While it was noted that the average longitudinal skewness of the plume, $\langle \tilde{\mu}_3^S \rangle$, tends
504 to deviate more strongly from zero both at early and late times as α decreases when $Pe \rightarrow$
505 ∞ , Figure 5(b) clearly documents that this trend is reversed at early times in the presence
506 of local dispersion. A similar behavior is documented in Figure 5(c) for the average spatial
507 kurtosis, $\langle \tilde{\mu}_4^S \rangle$, the results corresponding to Gaussian and Sub-Gaussian settings tending
508 to become very similar under the action of local dispersion.

509 The average higher-order moments of the BTC, i.e., $\langle \tilde{\mu}_3^T \rangle$ (Figure 5(e)) and $\langle \tilde{\mu}_4^T \rangle$
510 (Figure 5(f)), are highly influenced by local dispersion at early times. At later times, local
511 dispersion appears to induce a (downward) shift in the results, with a pattern that is similar

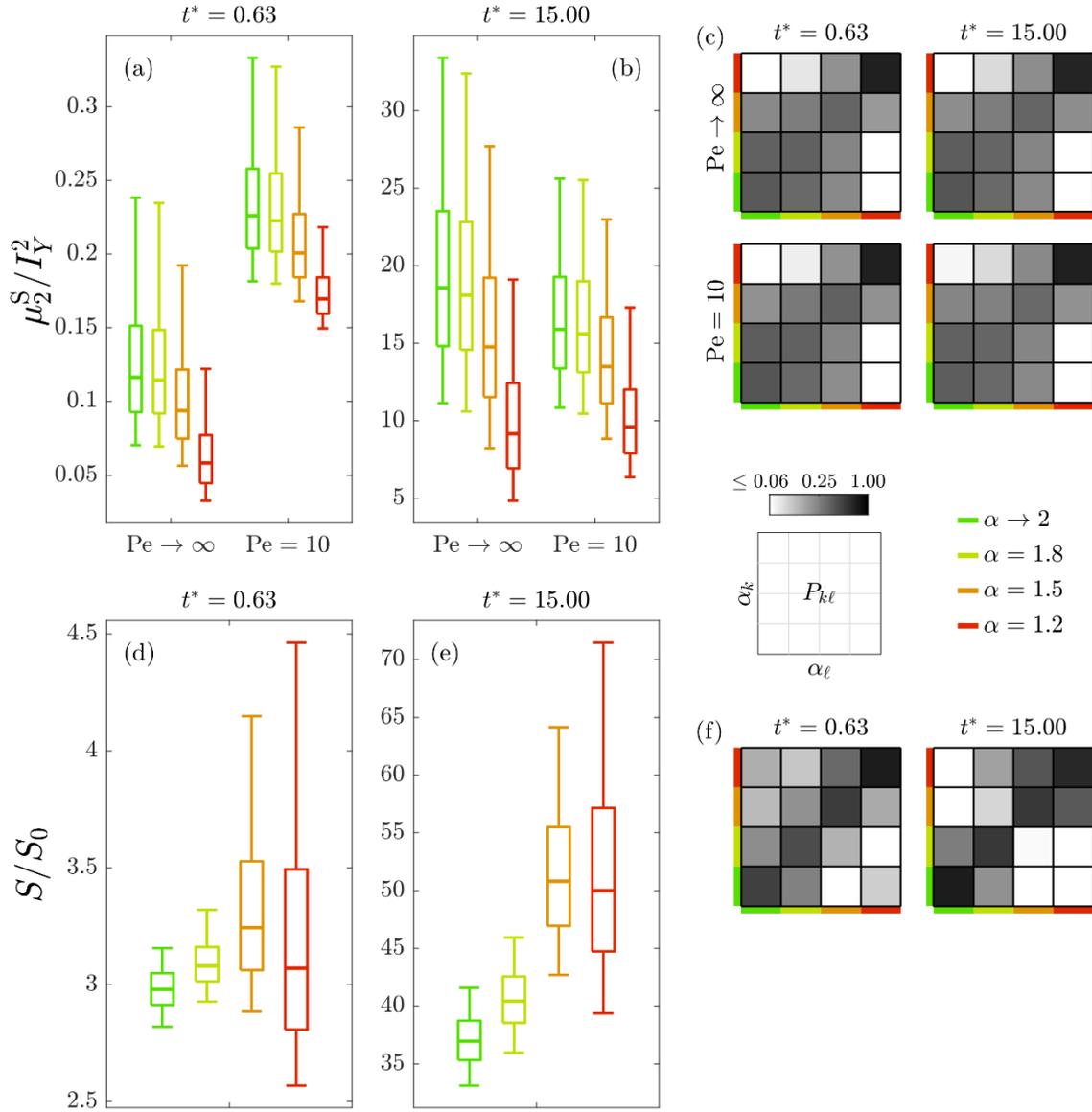
512 for $\alpha \rightarrow 2$ and $\alpha = 1.2$. This feature is also observed for the first arrival time, $\langle T_1 \rangle$.
513 Otherwise, the mass flux peak (Figure 5(h)) is delayed in time (on average) by the action
514 of a finite Pe, which also reduces the difference between the cases corresponding to $\alpha \rightarrow$
515 2 and $\alpha = 1.2$. The effect of dispersion on the (average) 99th percentile arrival time,
516 $\langle T_{p99} \rangle$, is analogous to that on $\langle \mu_2^T \rangle$. As such, one can see that these two metrics and $\langle \mu_2^S \rangle$
517 appear to contain very similar information about the effects of α and Pe on the processes
518 analyzed.

519 **3.2. Monte Carlo-based statistical distributions of metrics characterizing transport**

520 Here, we discuss the results obtained in terms of the full distributions of quantities
521 of interest across the collection of Monte Carlo realizations analyzed. We focus on the
522 system behavior at early and late stages of the plume deformation and confine the
523 discussion to the scenarios corresponding to $Pe \rightarrow \infty$ and $Pe = 10$. As a first result, we
524 observe that the empirical MC-based distributions associated with most of the analyzed
525 metrics are not well differentiated for the various values of α considered (i.e., there is
526 some overlap between distributions). As such, we investigate the significance of relying
527 on the metrics considered to discriminate the extent at which the log-conductivity field
528 deviates from the Gaussian behavior (as reflected by the value of α) when only results
529 from a single realization of Y are available. In other words, we analyze how significantly
530 each metric of solute transport is impacted by the value of α characterizing a given Y
531 realization. We do so upon relying on the Bayes classifier (e.g., James et al., 2013)
532 combined with Kernel Density Estimation.

533 Following the developments detailed in Appendix C, we quantify the degree of
534 overlap between sample distributions of a given metric by means of the index $P_{k\ell} :=$
535 $P(\alpha = \alpha_k | \hat{\alpha} = \alpha_\ell)$, which represents the likelihood of the actual value of α being equal

536 to α_k , given that the classification method infers an estimate $\hat{\alpha} = \alpha_\ell$ (from comparing the
537 result of a realization to the corresponding Monte Carlo collection of results). We
538 consider the discrete set of α values corresponding to $\alpha_k, \alpha_\ell \in \{1.2, 1.5, 1.8, \dots, 2\}$. The
539 values of index $P_{k\ell}$ are then seen as the entries of a (4×4) matrix, which we refer to as
540 overlap matrix. Note that one can obtain an overlap matrix for each of the transport
541 metrics analyzed. Values on the diagonal (i.e., $P_{k\ell}$ with $k = \ell$) of this matrix indicate the
542 extent to which we can rely on the distribution of results for a given metric to identify the
543 actual value of α associated with the underlying Y field. High values of the index
544 concentrated on the diagonal of the overlap matrix indicate mild overlap (i.e., strong
545 differentiation) of population densities, values evenly distributed across columns
546 suggesting strong overlap (i.e., poor differentiation). The strongest overlaps are observed
547 for high-order moments, whereas advective stretching is the best-differentiable metric.
548 For most metrics, overlap is strengthened at later times and for lower values of Pe . A
549 detailed analysis follows below.



550

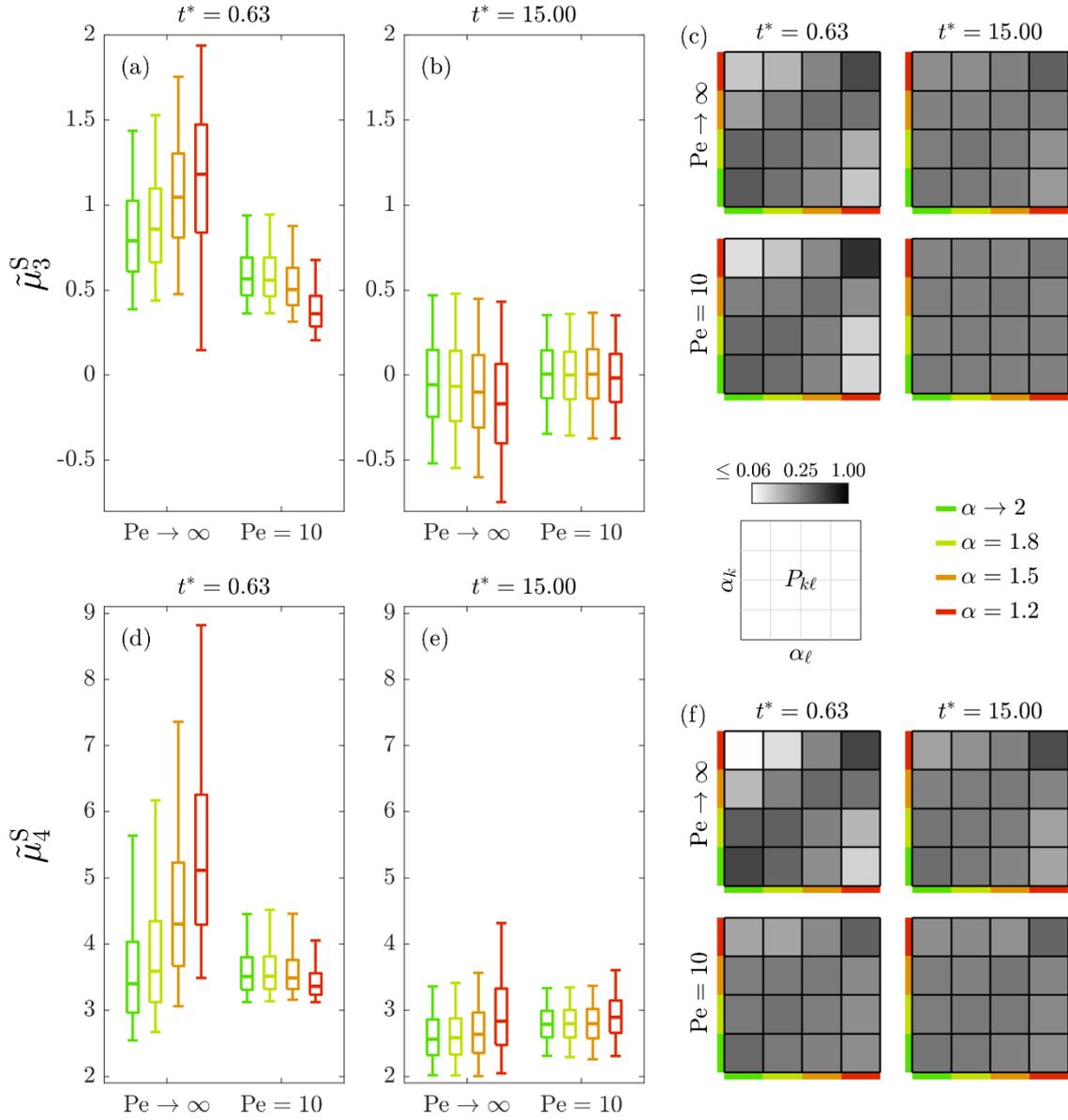
551 *Figure 6: Box plot of (a)-(b) μ_2^S for $Pe \rightarrow \infty$ and $Pe = 10$; and (d)-(e) S/S_0 for $Pe \rightarrow \infty$ at*
 552 *an early and at a late time. The corresponding overlap quantification, $P_{k\ell}$, is also*
 553 *included (c, f).*

554

555 Figure 6 depicts boxplots related to the populations of centered second moment of
 556 the section-integral concentrations μ_2^S (Figure 6(a-b)) and advective stretching S/S_0
 557 (Figure 6(d-e)) for the various settings examined. These results are complemented by the
 558 graphical depiction of results stemming from the overlap analysis (see implementation
 559 details in Appendix C) for the two combinations of Pe (except for S/S_0 , where $Pe \rightarrow \infty$

560 only) and various stages of plume development. Corresponding depictions related to other
561 selected transport metrics are presented in Figures 7 and 8.

562 Figure 6(a) suggests that the early-time ($t^* = 0.63$) distributions of values of μ_2^S
563 obtained for the analyzed values of α strongly differ for the two values of Pe considered.
564 This difference decreases as time increases. Otherwise, the degree of overlap between
565 distributions for diverse values of α , as quantified by the results in Figure 6(c), is not
566 significantly modified by local dispersion. The degree of overlap between the results for
567 Gaussian ($\alpha \rightarrow 2$) and slightly Sub-Gaussian ($\alpha = 1.8$) Y fields is markedly significant,
568 implying that the results of single realizations with $\alpha \rightarrow 2$ or $\alpha = 1.8$ are essentially
569 indistinguishable in terms of longitudinal dispersion. This tendency is documented for
570 most of the metrics analyzed. Results very similar to those corresponding to μ_2^S are
571 obtained in the analysis of the populations and overlaps of μ_2^T and T_{p99} (not shown here
572 for brevity). Figures 6(d) and 6(e) show that the range of values undertaken by the surface
573 growth due to hydrodynamic deformation, S/S_0 , increases with the departure of Y from
574 Gaussianity, this metric being the most affected by slight differences in α . This is also
575 evidenced by the value of $P_{k\ell}$ in Figure 6(f). At late times ($t^* = 15$), there is only a slight
576 overlap between the moderately-to-strongly Sub-Gaussian cases ($\alpha \leq 1.5$) and the
577 Gaussian / slightly Sub-Gaussian cases ($\alpha \geq 1.8$).



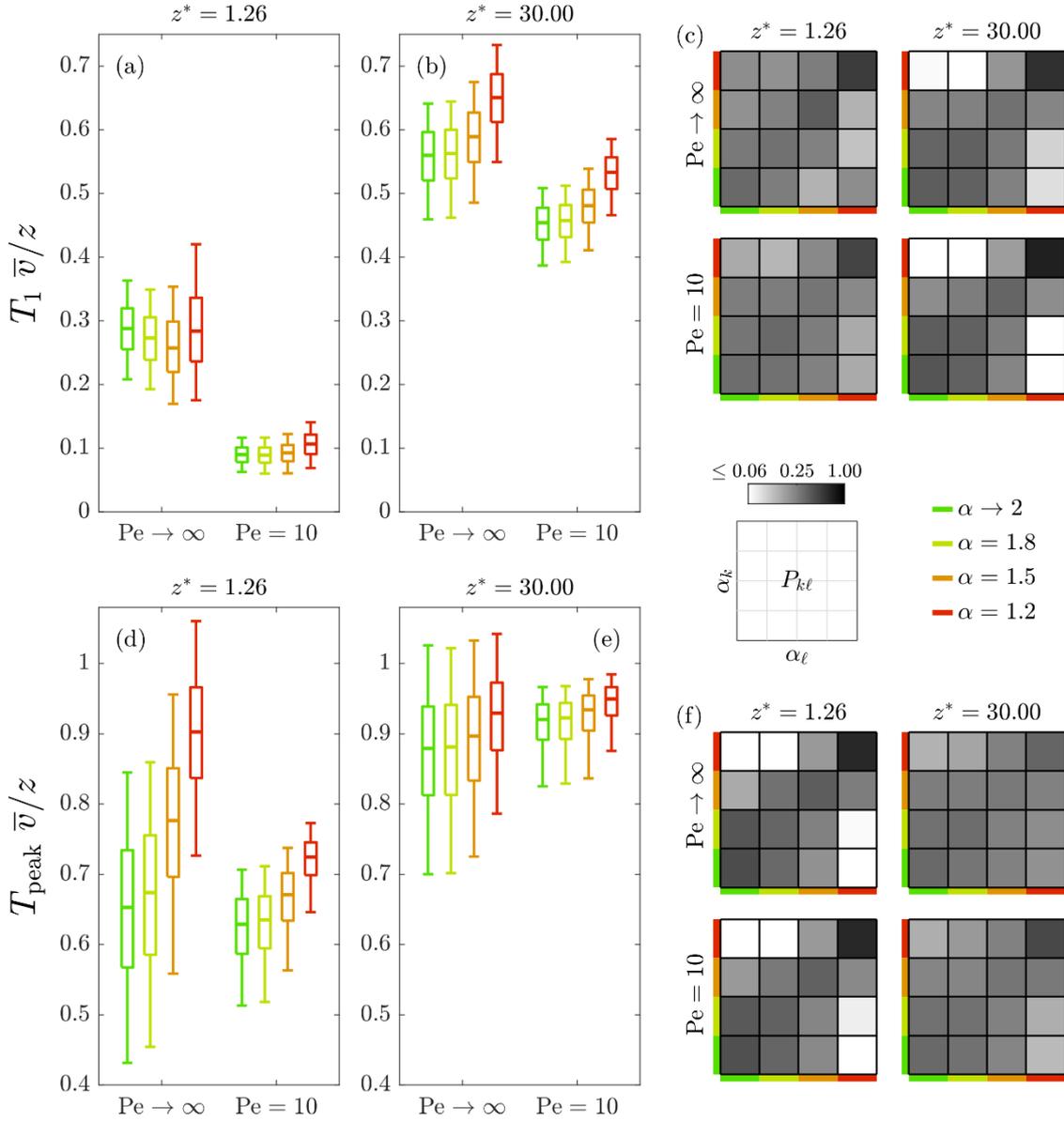
578

579 *Figure 7: Box plot of (a)-(b) $\tilde{\mu}_3^S$ and (d)-(e) $\tilde{\mu}_4^S$ for $Pe \rightarrow \infty$ and $Pe = 10$ at an early and at*
 580 *a late time. The corresponding overlap quantification, $P_{k\ell}$, is also included (c, f).*

581

582 Similar to what is observed for the centered second moment of the longitudinal
 583 concentration profile, a finite value of Pe narrows the range of values undertaken by the
 584 higher-order spatial moments across the collection of MC realizations (Figure 7). One
 585 can see that the distribution overlaps are significant (even for strongly Sub-Gaussian
 586 fields) for both skewness and kurtosis, especially at late times ($t^* = 15$) and for $Pe = 10$.
 587 These results lead us to conclude that the average behavior of the higher-order moments

588 associated with diverse values of α which we document in Section 3.1.1 would most
 589 probably be undetectable from the analysis of the evolution of solute concentrations
 590 across individual realizations of the Y field. Similar conclusions are drawn from the
 591 analysis of the populations of higher-order temporal moments (not shown).

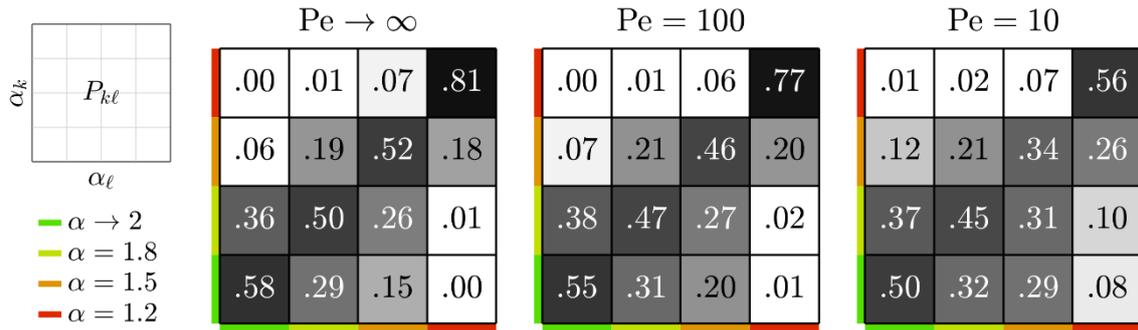


592

593 *Figure 8: Box plot of (a)-(b) T_1 and (d)-(e) T_{peak} for $Pe \rightarrow \infty$ and $Pe = 10$ at a close and*
 594 *at a far distance from the inlet. The corresponding overlap quantification, $P_{k\ell}$, is also*
 595 *included (c, f).*

596 Figure 8 depicts the boxplots and overlap quantification corresponding to the first
 597 arrival time (T_1) and the peak arrival time (T_{peak}). At close distance from the inlet ($z^* =$

598 1.26), the range of values of T_1 (i) narrows due to the effect of a finite Pe and (ii) is
 599 virtually insensitive to the value of α . These results suggest that the behavior of T_1 is
 600 chiefly dominated by local dispersion, this impact somewhat persisting for long travel
 601 distances (i.e., $z^* = 30$). These results support the conjecture that, even as the overlap
 602 with respect to the Gaussian case is not very high at long distances for moderate to strong
 603 Sub-Gaussian Y fields, it would be hard to discriminate the effect of α from that of Pe if
 604 both parameters are not known a priori. It is also worth noting that, while the time at
 605 which the BTC peak occurs is (on average) closer to the mean travel time as α decreases
 606 (Section 3.1), this tendency can hardly be detected at long distances from single
 607 realizations. This is clearly seen in Figure 8(f), which shows that the overlap is nearly
 608 total at $z^* = 30$.



609

610 *Figure 9: Overall overlap quantification, $P_{k\ell}$, between the distributions of the results*
 611 *corresponding to various values of α and for three selected Péclet numbers. The*
 612 *overlap is defined as $P_{k\ell} := P(\alpha = \alpha_k | \hat{\alpha} = \alpha_\ell)$, where $\hat{\alpha}$ in this case is the result of a*
 613 *Quadratic Discriminant Analysis (see Appendix C) combining the information embedded*
 614 *in the evolution of all spatial and temporal moments as well as T_1 , T_{peak} , and T_{p99} .*

615

616 Our analyses suggest that metrics that would typically be measurable in a real
 617 setting, such as the (spatial or temporal) moments describing the evolution of the plume,
 618 or the key arrival times, would not be able by themselves (i.e., without relying on
 619 additional observations about the Y field) to reflect with certainty the Sub-Gaussian (or

620 Gaussian) nature of the heterogeneous porous medium across which solute migrates. This
621 is evidenced even assuming that σ_Y^2 , I_Y , and Pe are known.

622 We further explore this aspect by analyzing the distributions of the results obtained
623 for all of these metrics jointly. We do so by relying on the classification technique known
624 as Quadratic Discriminant Analysis (QDA) to obtain an overall quantification of the
625 degree of differentiation of the sample distributions of the multiple metrics considered.
626 Details about the theoretical elements underpinning the approach are illustrated in
627 Appendix C. We ground QDA on the overall collections of results pertaining to spatial
628 and temporal moments ($\mu_2^S, \tilde{\mu}_3^S, \tilde{\mu}_4^S, \mu_2^T, \tilde{\mu}_3^T, \tilde{\mu}_4^T$), as well as the arrival times T_1, T_{peak} and
629 T_{p99} , and jointly considering results for $t^* = 0.63, 3.24, 15$ and $z^* = 1.26, 6.48, 30$, for
630 a total of 27 sample populations. We exclude S/S_0 from the analysis, as we focus on
631 solute transport metrics that could be detected directly in a real setting. This type of
632 analysis yields a global appraisal of our ability to discriminate results associated with a
633 unique realization through the assessment of an overlap matrix whose entries $P_{k\ell}$
634 represent the likelihood of the actual value of α being equal to α_k when the QDA
635 classification yields a value equal to α_ℓ . To broaden the range of scenarios, we perform
636 this analysis for Pe = 10, 100, and Pe $\rightarrow \infty$. The results are depicted in Figure 9. The
637 degree of differentiation between the overall behavior of individual realizations is visibly
638 higher than what one could infer by considering each of the metrics separately (Figures
639 6-8). Nevertheless, incorrect classifications are still abundant, especially for low values
640 of Pe. The results of our joint analysis of spatial moments, temporal moments and key
641 arrival times at multiple stages further support the observation that relying solely on
642 observations of concentrations and/or travel times for the considered sample size does not
643 allow to clearly and uniquely discriminate between values of α .

- 667 • As compared to a Gaussian distribution of log-conductivities, a Generalized Sub-
668 Gaussian (GSG) field of Y yields (on average) (i) a delayed first time of arrival of
669 the solute mass to crossing planes located at a distance of a few integral scales
670 from the solute injection plane; (ii) enhanced and delayed concentration peaks;
671 and (iii) a reduced breakthrough time associated with the majority of the solute
672 mass migrating in the system.
- 673 • On average, we document (i) an increasing degree of asymmetry and (ii) an
674 enhanced kurtosis (resulting in heavier tails) of the plume with log-conductivity
675 departing from a Gaussian behavior. Differences between the effects imprinted by
676 Gaussian and strongly Sub-Gaussian ($\alpha \leq 1.5$) Y domains on spatial skewness
677 and kurtosis of evolving plumes are marked at early times and tend to become
678 mild as time increases.
- 679 • The velocity field associated with a GSG Y field promotes plume stretching while
680 reducing longitudinal plume spreading compared to its Gaussian counterpart.
681 Therefore, the increased longitudinal dispersion associated with Gaussian Y fields
682 with respect to a GSG setting may not necessarily translate into enhanced solute
683 mixing. This finding can have important implications on reactive transport; for
684 example, in the context of mixing-limited reactive transport, it might imply that a
685 higher mixing rate and reactivity can take place in Sub-Gaussian fields as
686 compared to their Gaussian counterparts. Detailed analyses of this aspect will be
687 subject of future research.

688 Concerning one's ability to clearly identify a possible underlying Sub-Gaussian
689 nature of Y upon relying solely on the analysis of solute concentration fields in the
690 settings analyzed, our study leads to the following major conclusions.

- 691 • Even as one has information about the values of variance and integral scale of Y ,
692 the effects of a Sub-Gaussian nature of Y on hydrodynamic dispersion might
693 hardly be detectable in a (finite) heterogeneous setting, especially if characterized
694 by a low degree of Sub-Gaussianity (i.e., corresponding to $1.8 \leq \alpha < 2.0$ for the
695 type of subordinator considered). The action of local dispersion (i.e., finite values
696 of Pe) can then contribute to further mask the influence of Sub-Gaussianity on
697 major transport metrics, as also seen from a Quadratic Discriminant Analysis
698 performed on spatial and temporal moments and arrival times at various evolution
699 stages of the plume.
- 700 • While a GSG-based model can be consistent with scaling patterns of the
701 probability distribution of log-conductivity and its spatial increments at various
702 lags, relying solely on observations of solute concentrations migrating within the
703 system can in some cases hamper our ability to discern the nature of the
704 underlying conductivity field. To characterize the latter, one would need to
705 consider also a joint analysis of Y data and their increments.

706 We note that the present study focuses on mildly heterogeneous bounded domains
707 (representing a spatially heterogeneous single geological unit) and is mostly based on
708 spatially or temporally aggregated transport metrics. Additional studies are required to
709 fully assess the extent to which the emergence of strictly local features associated with
710 transport behavior (such as those related to strong channeling effects within highly
711 heterogeneous, virtually unbounded media) might contribute to distinguish GSG log-
712 conductivity domains from their Gaussian counterparts.

713 **Acknowledgements**

714 The authors would like to thank the EU, MIUR, and MINECO for funding, in the frame
715 of the collaborative international Consortium (WE-NEED) financed under the ERA-NET

716 WaterWorks2014 Cofunded Call. This ERA-NET is an integral part of the 2015 Joint
717 Activities developed by the Water Challenges for a Changing World Joint Programme
718 Initiative (Water JPI). Part of the work was developed while Prof. A. Guadagnini was at
719 the University of Strasbourg with funding from Région Grand-Est and Strasbourg-
720 Eurométropole through the *Chaire Gutenberg*. The UPC authors acknowledge funding
721 from AGAUR Research Groups, 2017 SGR 1485.
722 Data available on Zenodo (<https://doi.org/10.5281/zenodo.3356708>).

723 **References**

- 724 Bear, J. (1972). *Dynamics of fluids in porous media*. Elsevier, New York.
- 725 Bellin, A., Salandin, P., & Rinaldo, A. (1992). Simulation of dispersion in
726 heterogeneous porous formations: Statistics, first-order theories, convergence of
727 computations. *Water Resources Research*, 28(9), 2211–2227.
728 <https://doi.org/10.1029/92WR00578>
- 729 Boffetta, G., Mazzino, A., & Vulpiani, A. (2008). Twenty-five years of multifractals in
730 fully developed turbulence: A tribute to Giovanni Paladin, *J. Phys. A Math. Theor.*,
731 41, 363001
- 732 Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via
733 diffusion. *Ann. Statist.*, 38(5), 2916–2957. <https://doi.org/10.1214/10-AOS799>
- 734 Dagan, G. (1982). Stochastic modeling of groundwater flow by unconditional and
735 conditional probabilities: 2. The solute transport. *Water Resources Research*,
736 18(4), 835–848. <https://doi.org/10.1029/WR018i004p00835>
- 737 Dagan, G. (1984). Solute transport in heterogeneous porous formations. *Journal of*
738 *Fluid Mechanics*, 145, 151–177. <https://doi.org/10.1017/S0022112084002858>
- 739 Dagan, G. (1989). *Solute transport in porous formations*. Springer, Berlin.
- 740 Dartois, A., Beaudoin, A., & Huberson, S. (2018). Impact of local diffusion on
741 macroscopic dispersion in three-dimensional porous media. *Comptes Rendus*
742 *Mécanique*, 346(2), 89–97. <https://doi.org/10.1016/j.crme.2017.12.012>
- 743 de Dreuzy, J.-R., Méheust, Y., & Pichot, G. (2012). Influence of fracture scale
744 heterogeneity on the flow properties of three-dimensional discrete fracture
745 networks (DFN). *Journal of Geophysical Research: Solid Earth*, 117(B11).
746 <https://doi.org/10.1029/2012JB009461>

- 747 Dentz, M., Kinzelbach, H., Attinger, S., & Kinzelbach, W. (2003). Numerical studies of
748 the transport behavior of a passive solute in a two-dimensional incompressible
749 random flow field. *Physical Review E - Statistical Physics, Plasmas, Fluids, and*
750 *Related Interdisciplinary Topics*. <https://doi.org/10.1103/PhysRevE.67.046306>
- 751 Dentz, M., Kinzelbach, H., Attinger, S., & Kinzelbach, W. (2000). Temporal behavior
752 of a solute cloud in a heterogeneous porous medium: 1. Point-like injection. *Water*
753 *Resources Research*, 36(12), 3591–3604. <https://doi.org/10.1029/2000WR900162>
- 754 Dentz, M., Kinzelbach, H., Attinger, S., & Kinzelbach, W. (2002). Temporal behavior
755 of a solute cloud in a heterogeneous porous medium 3. Numerical simulations.
756 *Water Resources Research*, 38(7), 13–23. <https://doi.org/10.1029/2001WR000436>
- 757 Dentz, M., Le Borgne, T., Englert, A., & Bijeljic, B. (2011). Mixing, spreading and
758 reaction in heterogeneous media: A brief review. *Journal of Contaminant*
759 *Hydrology*. <https://doi.org/10.1016/j.jconhyd.2010.05.002>
- 760 Fernández-García, D., & Sanchez-Vila, X. (2011). Optimal reconstruction of
761 concentrations, gradients and reaction rates from particle distributions. *Journal of*
762 *Contaminant Hydrology*, 120–121, 99–114.
763 <https://doi.org/10.1016/j.jconhyd.2010.05.001>
- 764 Fernández-García, D., Illangasekare, T. H., & Rajaram, H. (2005). Differences in the
765 scale-dependence of dispersivity estimated from temporal and spatial moments in
766 chemically and physically heterogeneous porous media. *Advances in Water*
767 *Resources*, 28(7), 745–759. <https://doi.org/10.1016/j.advwatres.2004.12.011>
- 768 Fogg, G. E., & Zhang, Y. (2016). Debates—Stochastic subsurface hydrology from
769 theory to practice: A geologic perspective. *Water Resources Research*, 52(12),
770 9235–9245. <https://doi.org/10.1002/2016WR019699>
- 771 Ganti, V., Singh, A., Passalacqua, P., & Fofoula-Georgiou, E. (2009). Subordinated
772 Brownian motion model for sediment transport, *Phys. Rev. E*, 80, 011111,
773 doi:1539-5663/2009/80(1)/011111(9)
- 774 Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Prentice-Hall.
- 775 Gist, G. A., Thompson, A. H., Katz, A. J., & Higgins, R. L. (1990). Hydrodynamic
776 dispersion and pore geometry in consolidated rock. *Physics of Fluids A: Fluid*
777 *Dynamics*, 2(9), 1533–1544. <https://doi.org/10.1063/1.857602>
- 778 Gómez-Hernández, J. J., & Wen, X.-H. (1998). To be or not to be multi-Gaussian? A
779 reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1), 47–61.
780 [https://doi.org/10.1016/S0309-1708\(96\)00031-0](https://doi.org/10.1016/S0309-1708(96)00031-0)
- 781 Guadagnini, A., Riva, M., & Neuman, S.P. (2012). Extended power-law scaling of
782 heavy-tailed random air-permeability fields in fractured and sedimentary rocks,
783 *Hydrol. Earth Syst. Sci.*, 16, 3249–3260, doi:10.5194/hess-16-3249-2012

- 784 Guadagnini, A., Riva, M., & Neuman, S. P. (2018). Recent advances in scalable non-
785 Gaussian geostatistics: The generalized sub-Gaussian model. *Journal of*
786 *Hydrology*, 562, 685–691. <https://doi.org/10.1016/j.jhydrol.2018.05.001>
- 787 Guadagnini, A., Neuman, S.P., Schaap, M.G., & Riva, M. (2013). Anisotropic statistical
788 scaling of vadose zone hydraulic property estimates near Maricopa, Arizona, *Water*
789 *Resour. Res.*, 49, doi:10.1002/2013WR014286
- 790 Guadagnini, A., S.P. Neuman, M.G. Schaap, and M. Riva (2014), Anisotropic
791 Statistical Scaling of Soil and Sediment Texture in a Stratified Deep Vadose Zone
792 near Maricopa, Arizona, *Geoderma*, 214-215, 217-227,
793 doi:10.1016/j.geoderma.2013.09.008.
- 794 Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey Modular
795 Ground-Water Model - the Ground-Water Flow Process. U.S. Geological Survey
796 Techniques and Methods 6-A16
- 797 James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical*
798 *Learning: With Applications in R*. Springer Publishing Company, Incorporated.
- 799 Jankovic, I., Maghrebi, M., Fiori, A., & Dagan, G. (2017). When good statistical models
800 of aquifer heterogeneity go right: The impact of aquifer permeability structures on
801 3D flow and transport. *Advances in Water Resources*, 100, 199–211.
802 <https://doi.org/10.1016/J.ADVWATRES.2016.10.024>
- 803 Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hydrogeology*.
804 Cambridge University Press. <https://doi.org/10.1017/CBO9780511626166>
- 805 LaBolle, E.M., Quastel, J., Fogg, G.E., & Gravner, J. (2000). Diffusion processes in
806 composite porous media and their numerical integration by random walks:
807 Generalized stochastic differential equations with discontinuous coefficients,
808 *Water Resour. Res.*, 36, 3, doi: 10.1029/1999WR900224
- 809 Le Borgne, T., Dentz, M., & E. Villemaux (2013). Stretching, coalescence, and mixing
810 in porous media. *Physical Review Letters*, 110, 20,
811 doi:10.1103/PhysRevLett.110.204501
- 812 Le Borgne, T., Dentz, M., & E. Villemaux (2015). The lamellar description of mixing in
813 porous media. *Journal of Fluid Mechanics*, 770, 458-498,
814 doi:10.1017/jfm.2015.117
- 815 Libera, A., de Barros, F. P. J., Riva, M., & Guadagnini, A. (2017). Solute concentration
816 at a well in non-Gaussian aquifers under constant and time-varying pumping
817 schedule. *Journal of Contaminant Hydrology*, 205, 37–46.
818 <https://doi.org/https://doi.org/10.1016/j.jconhyd.2017.08.006>
- 819 Liu, H. H., & Molz, F. J. (1997). Comment on “Evidence for non-Gaussian scaling
820 behavior in heterogeneous sedimentary formations” by Scott Painter. *Water*
821 *Resources Research*, 33(4), 907–908. <https://doi.org/10.1029/96WR03788>

- 822 Liu, X., & Müller, H.-G. (2004). Functional Convex Averaging and Synchronization for
823 Time-Warped Random Curves. *Journal of the American Statistical Association*,
824 99(467), 687–699. <https://doi.org/10.1198/016214504000000999>
- 825 Lu, Z., & Stauffer, P. H. (2012). On estimating functional average breakthrough curve
826 using time-warping technique and perturbation approach. *Water Resources*
827 *Research*, 48(5). <https://doi.org/10.1029/2011WR011506>
- 828 Meerschaert, M. M., T. J. Kozubowski, F. J. Molz, and S. Lu (2004), Fractional Laplace
829 model for hydraulic conductivity, *Geophys. Res. Lett.*, 31, L08501,
830 doi:10.1029/2003GL019320
- 831 Naff, R. L., Haley, D. F., & Sudicky, E. A. (1998a). High-resolution Monte Carlo
832 simulation of flow and conservative transport in heterogeneous porous media: 1.
833 Methodology and flow results. *Water Resources Research*, 34(4), 663–677.
834 <https://doi.org/10.1029/97WR02712>
- 835 Naff, R. L., Haley, D. F., & Sudicky, E. A. (1998b). High-resolution Monte Carlo
836 simulation of flow and conservative transport in heterogeneous porous media: 2.
837 Transport results. *Water Resources Research*, 34(4), 679–697.
838 <https://doi.org/10.1029/97WR02711>
- 839 Painter, S. (1996). Evidence for Non-Gaussian Scaling Behavior in Heterogeneous
840 Sedimentary Formations. *Water Resources Research*, 32(5), 1183–1195.
841 <https://doi.org/10.1029/96WR00286>
- 842 Painter, S. (2001), Flexible scaling model for use in random field simulation of
843 hydraulic conductivity, *Water Resources Research*, 37, 1155–1163.
844 <https://doi.org/10.1029/2000WR900394>
- 845 Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. P. (2016). Theory and generation
846 of conditional, scalable sub-Gaussian random fields. *Water Resources Research*,
847 52(3), 1746–1761. <https://doi.org/10.1002/2015WR018348>
- 848 Pedretti, D., Fernández-García, D., Bolster, D., & Sanchez-Vila, X. (2013). On the
849 formation of breakthrough curves tailing during convergent flow tracer tests in
850 three-dimensional heterogeneous aquifers. *Water Resources Research*, 49(7),
851 4157–4173. <https://doi.org/10.1002/wrcr.20330>
- 852 Pedretti, D., & Fernández-García, D. (2013). An automatic locally-adaptive method to
853 estimate heavily-tailed breakthrough curves from particle distributions. *Advances*
854 *in Water Resources*, 59, 52–65. <https://doi.org/10.1016/j.advwatres.2013.05.006>
- 855 Pedretti, D., Masetti, M., & Beretta, G. Pietro. (2017). Stochastic analysis of the
856 efficiency of coupled hydraulic-physical barriers to contain solute plumes in highly
857 heterogeneous aquifers. *Journal of Hydrology*, 553, 805–815.
858 <https://doi.org/10.1016/j.jhydrol.2017.08.051>
- 859 Riva, M., Guadagnini, A., Fernández-García, D., Sanchez-Vila, X., & Ptak, T. (2008).
860 Relative importance of geostatistical and transport models in describing heavily

- 861 tailed breakthrough curves at the Lauswiesen site. *Journal of Contaminant*
862 *Hydrology*, 101(1–4), 1–13. <https://doi.org/10.1016/j.jconhyd.2008.07.004>
- 863 Riva, M., Guadagnini, A., & Neuman, S. P. (2017). Theoretical analysis of non-
864 Gaussian heterogeneity effects on subsurface flow and transport. *Water Resources*
865 *Research*, 53(4), 2998–3012. <https://doi.org/10.1002/2016WR019353>
- 866 Riva, M., Neuman, S. P., & Guadagnini, A. (2015a). New scaling model for variables
867 and increments with heavy-tailed distributions. *Water Resources Research*, 51(6),
868 4623–4634. <https://doi.org/10.1002/2015WR016998>
- 869 Riva, M., Panzeri, M., Guadagnini, A., & Neuman, S. P. (2015b). Simulation and
870 analysis of scalable non-Gaussian statistically anisotropic random functions.
871 *Journal of Hydrology*, 531, 88–95.
872 <https://doi.org/10.1016/J.JHYDROL.2015.06.066>
- 873 Riva, M., Neuman, S. P., & Guadagnini, A. (2013a). Sub-Gaussian model of processes
874 with heavy-tailed distributions applied to air permeabilities of fractured tuff.
875 *Stochastic Environmental Research and Risk Assessment*, 27(1), 195–207.
876 <https://doi.org/10.1007/s00477-012-0576-y>
- 877 Riva, M., Neuman, S. P., Guadagnini, A., & Siena, M. (2013b). Anisotropic Scaling of
878 Berea Sandstone Log Air Permeability Statistics. *Vadose Zone Journal*, 12.
879 <https://doi.org/10.2136/vzj2012.0153>
- 880 Rizzo, C. B., & de Barros, F. P. J. (2017). Minimum Hydraulic Resistance and Least
881 Resistance Path in Heterogeneous Porous Media. *Water Resources Research*,
882 53(10), 8596–8613. <https://doi.org/10.1002/2017WR020418>
- 883 Rubin, Y. (1990a). Stochastic modeling of macrodispersion in heterogeneous porous
884 media. *Water Resources Research*, 26(1), 133–141.
885 <https://doi.org/10.1029/WR026i001p00133>
- 886 Rubin, Y. (1990b). Correction to “Stochastic modeling of macrodispersion in
887 heterogeneous porous media” by Yoram Rubin. *Water Resources Research*,
888 26(10), 2631. <https://doi.org/10.1029/WR026i010p02631>
- 889 Rubin, Y. (2003). *Applied stochastic hydrogeology*. New York : Oxford University
890 Press.
- 891 Salamon, P., Fernández-García, D., & Gómez-Hernández, J. J. (2006). A review and
892 numerical assessment of the random walk particle tracking method. *Journal of*
893 *Contaminant Hydrology*, 87(3–4), 277–305.
894 <https://doi.org/10.1016/j.jconhyd.2006.05.005>
- 895 Siena, M., Guadagnini, A., Riva, M., & Neuman, S. P. (2012). Extended power-law
896 scaling of air permeabilities measured on a block of tuff. *Hydrology and Earth*
897 *System Sciences*, 16(1), 29–42. <https://doi.org/10.5194/hess-16-29-2012>

- 898 Siena, M., & Riva, M. (2018). Groundwater withdrawal in randomly heterogeneous
 899 coastal aquifers. *Hydrology and Earth System Sciences*, 22(5), 2971–2985.
 900 <https://doi.org/10.5194/hess-22-2971-2018>
- 901 Siena, M., Riva, M., Giamberini, M., Gouze, P., & Guadagnini, A. (2019). Statistical
 902 modeling of gas-permeability spatial variability along a limestone core, *Spatial*
 903 *Statistics*, 39, 100249, doi: 10.1016/j.spasta.2017.07.007.
- 904 Siena, M., Guadagnini, A., Bouissonnié, A., Ackerer, P., Daval, D., & Riva, M. (2020).
 905 Generalized Sub-Gaussian processes: theory and application to hydrogeological
 906 and geochemical data, *Water Resources Research*, WRCR24714,
 907 doi:10.1029/2020WR027436
- 908 Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*.
 909 *ChapMan & HALL/CRC* (Vol. 37). <https://doi.org/10.2307/2347507>
- 910 Sole-Mari, G., & Fernández-Garcia, D. (2018). Lagrangian Modeling of Reactive
 911 Transport in Heterogeneous Porous Media With an Automatic Locally Adaptive
 912 Particle Support Volume. *Water Resources Research*, 54(10).
 913 <https://doi.org/10.1029/2018WR023033>
- 914 Sole-Mari, G., Fernández-Garcia, D., Rodríguez-Escales, P., & Sanchez-Vila, X.
 915 (2017). A KDE-Based Random Walk Method for Modeling Reactive Transport
 916 With Complex Kinetics in Porous Media. *Water Resources Research*, 53(11).
 917 <https://doi.org/10.1002/2017WR021064>
- 918 Sole-Mari, G., Bolster, D., Fernández-Garcia, D., & Sanchez-Vila, X. (2019). Particle
 919 Density Estimation with Grid-Projected and Boundary-Corrected Adaptive
 920 Kernels. *Advances in Water Resources*, 103382.
 921 <https://doi.org/10.1016/j.advwatres.2019.103382>
- 922 Winter, C. L., Tartakovsky, D. M., & Guadagnini, A. (2003). Moment Differential
 923 Equations for Flow in Highly Heterogeneous Porous Media. *Surveys in*
 924 *Geophysics*, 24(1), 81–106. <https://doi.org/10.1023/A:1022277418570>
- 925 Winter, C. L., & Tartakovsky, D. M. (2000). Mean Flow in composite porous media.
 926 *Geophysical Research Letters*, 27(12), 1759–1762.
 927 <https://doi.org/10.1029/1999GL011030>
- 928 Winter, C. L., & Tartakovsky, D. M. (2002). Groundwater flow in heterogeneous
 929 composite aquifers. *Water Resources Research*, 38(8), 11–23.
 930 <https://doi.org/10.1029/2001WR000450>

931 **Appendix A: Analytical formulation of the GSG model and key properties**

932 Riva et al. (2015a,b) introduce the Generalized Sub-Gaussian (GSG) model
 933 according to which the random function $Y'(\mathbf{x})$ defined by (1) is described considering the

934 subordinator $U(\mathbf{x})$ to be lognormally distributed according to $\ln(U) \sim \mathcal{N}[0, (2 - \alpha)^2]$,
 935 i.e.,

$$936 \quad f_U(u) = \frac{e^{-\frac{(\ln u)^2}{2(2-\alpha)^2}}}{\sqrt{2\pi u(2-\alpha)}}, \quad (\text{A1})$$

937 with $\alpha < 2$. Here, we include a brief summary of the key analytical formulations and
 938 scaling properties of the isotropic GSG model introduced by Riva et al (2015a).
 939 Corresponding formulations for anisotropic fields have been derived by Riva et al
 940 (2015b).

941 The marginal *pdf* of $Y'(\mathbf{x})$, $f_{Y'}(y')$, is fully defined by α and the standard deviation
 942 of $G(\mathbf{x})$, σ_G , as

$$943 \quad f_{Y'}(y') = \frac{1}{2\pi\sigma_G(2-\alpha)} \int_0^\infty e^{-\frac{1}{2}\left[\left(\frac{\ln u}{2-\alpha}\right)^2 + \left(\frac{y'}{\sigma_G u}\right)^2\right]} \frac{du}{u^2}. \quad (\text{A2})$$

944

945 The latter corresponds to a normal-lognormal (NLN) distribution whose variance, $\sigma_{Y'}^2$, and
 946 standardized kurtosis, $\kappa_{Y'}$, of $Y'(\mathbf{x})$ are given by

$$947 \quad \sigma_{Y'}^2 = \sigma_G^2 e^{2(2-\alpha)^2}, \quad \kappa_{Y'} = 3e^{4(2-\alpha)^2}. \quad (\text{A3})$$

948 It is noted that the lognormal distribution (A1) tends to a delta function when $\alpha \rightarrow$
 949 2 and (A2) reduces to the Gaussian distribution. Figure A1 depicts $f_{Y'}$ for the three values
 950 of α examined in our study and by setting $\sigma_{Y'}^2 = 1$. The Gaussian *pdf* having the same
 951 mean and variance as $Y'(\mathbf{x})$ and corresponding to $\alpha \rightarrow 2$ is also depicted for
 952 completeness. While $f_{Y'}$ is close to the Gaussian *pdf* for $\alpha = 1.8$, one can observe that $f_{Y'}$
 953 exhibits peaks and tails which become sharp and long, respectively, as α decreases. These
 954 features, which represent a clear deviation from a Gaussian distribution, is also quantified
 955 by the excess kurtosis, $EK_{Y'} = \kappa_{Y'} - 3$, whose value increases (deviating from zero, which
 956 corresponds to the scenario for which $\alpha \rightarrow 2$) as α decreases.

957 The *pdf* of incremental values of $Y'(\mathbf{x})$, $\Delta Y = Y'(\mathbf{x}) - Y'(\mathbf{x} + \mathbf{s})$, is given by

958
$$f_{\Delta Y}(\Delta y) = \frac{1}{2\pi\sigma_G(2-\alpha)^2\sqrt{2\pi}} \int_0^\infty \int_0^\infty e^{-\frac{1}{2}\left[\frac{\ln^2 u_1 + \ln^2 u_2 + \frac{\Delta y^2}{\sigma_G^2 r^2}}{(2-\alpha)^2}\right]} \frac{du_2 du_1}{u_2 u_1 r}. \quad (\text{A4})$$

959 Here, $r = \sqrt{u_1^2 + u_2^2 - 2\rho_G u_1 u_2}$ and $\rho_G(s)$ is the correlation function of $G(\mathbf{x})$ evaluated at
 960 separation distance (or lag) $s = |s|$. The standardized kurtosis of ΔY is rendered as

961
$$\kappa_{\Delta Y} = 3e^{2(2-\alpha)^2} \left[1 + \frac{1}{2} \left(\frac{e^{2(2-\alpha)^2} - 1}{e^{(2-\alpha)^2} - \rho_G} \right)^2 \right]. \quad (\text{A5})$$

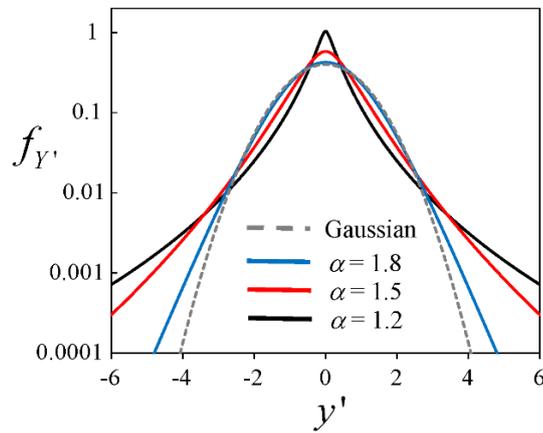
962 It is remarked that the dependence of $\rho_G(s)$ on lag induces a corresponding
 963 dependence of the shape of $f_{\Delta Y}$. For example, $\kappa_{\Delta Y} \rightarrow 3$ and the distribution of ΔY
 964 tends to the Gaussian one as $\alpha \rightarrow 2$. Otherwise, the shape of $f_{\Delta Y}$ scales with lag (or,
 965 equivalently, with ρ_G) with peak which tends to sharpen and tails to become heavier (i.e.,
 966 corresponding to increased $\kappa_{\Delta Y}$) as a function of ρ_G . Figure A2 depicts $f_{\Delta Y}$ for three
 967 exemplary lags and $\alpha = 1.8$ (Fig. A2a), 1.5 (Fig. A2b), and 1.2 (Fig. A2c). All GSG fields
 968 are characterized by the same variance ($\sigma_Y^2 = 1$) and integral scale ($I_Y = I_G e^{-(2-\alpha)^2}$, I_G
 969 being the integral scale of $G(\mathbf{x})$), and lags are normalized with respect to I_Y . Also shown
 970 for comparison are Gaussian distributions having the same variance as ΔY . In all cases,
 971 $f_{\Delta Y}$ exhibits sharp peaks and heavy tails at small lags. Increasing lag, $f_{\Delta Y}$ tends to
 972 become Gaussian when α is large (i.e., $\alpha = 1.8$), otherwise remaining heavy tailed for
 973 smaller values of α . This feature is also illustrated in Figure A3, depicting excess kurtosis
 974 of Y' and ΔY versus lag for the three values of α considered in our study. The excess
 975 kurtosis of ΔY , $EK_{\Delta Y} = \kappa_{\Delta Y} - 3$, decreases as lag increases, rendering the peak of $f_{\Delta Y}$
 976 less sharp and its tails lighter. When $\alpha = 1.8$, $EK_{\Delta Y}$ is seen to tend to an asymptotic value
 977 which is $\ll 1$, so that $f_{\Delta Y}$ becomes virtually Gaussian. Included in Figure A3 are
 978 horizontal lines depicting the excess of kurtosis of Y' , $EK_{Y'}$. In all investigated cases,
 979 $EK_{Y'} > EK_{\Delta Y}$ at small lags, i.e. $f_{\Delta Y}$ has sharper peaks and heavier tails than does $f_{Y'}$,
 980 the opposite being true at large lags.

981 The variogram, γ_Y , of Y' , is given by

982
$$\gamma_Y = \sigma_G^2 e^{(2-\alpha)^2} (e^{(2-\alpha)^2} - 1) + e^{(2-\alpha)^2} \gamma_G, \quad (\text{A6})$$

983 γ_G being the variogram of $G(\mathbf{x})$. Note that according to (A6) γ_Y is discontinuous at the
 984 origin, i.e., at $s = 0$, thus exhibiting a nugget effect. Figure A4 compares (on arithmetic
 985 and logarithm scales) γ_Y computed with three values of α ($= 1.2, 1.5, \text{ and } 1.8$) and setting
 986 $\sigma_Y^2 = 1$ and $I_Y = 1$. Also shown for comparison is the variogram obtained within a
 987 Gaussian field (i.e., corresponding to $\alpha \rightarrow 2$) characterized by the same variance and
 988 integral scale as Y' and by an exponential correlation function. The variogram associated
 989 with $\alpha = 1.8$ virtually coincides with its counterpart associated with a Gaussian field
 990 everywhere, with the exception of small lags where the GSG variogram is characterized
 991 by a nugget (clearly visible on logarithm scale).

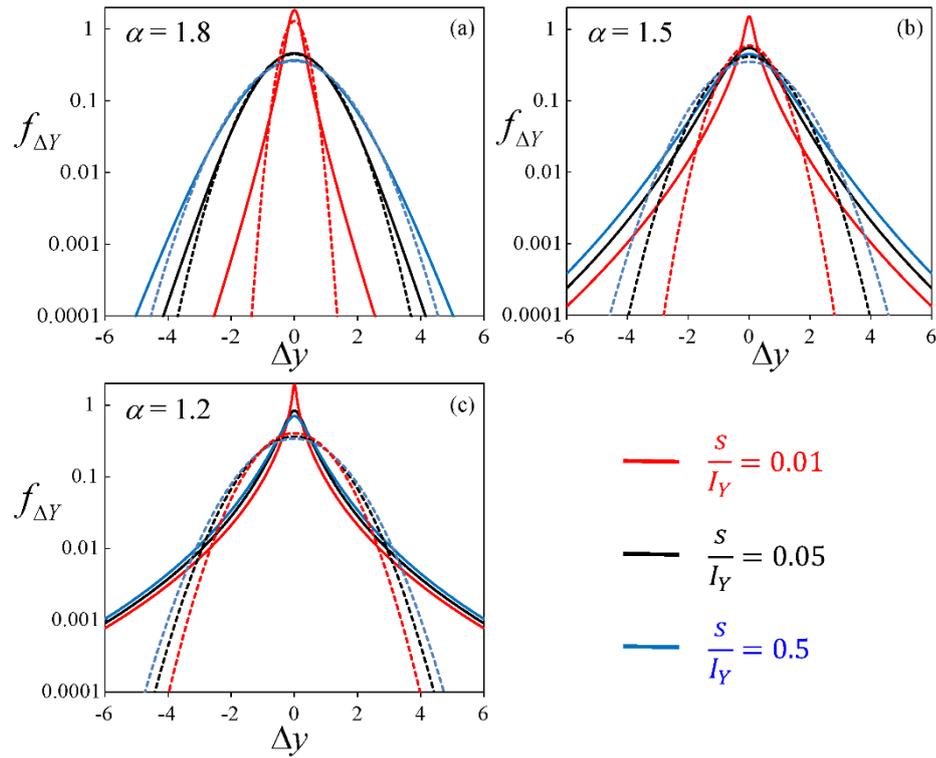
992



993

994 *Figure A1: Probability density function of Y' for $\sigma_Y^2 = 1$ and three values of α*
 995 *(solid curves). Also shown is the Gaussian pdf having the same mean and variance as Y'*
 996 *(dashed curve).*

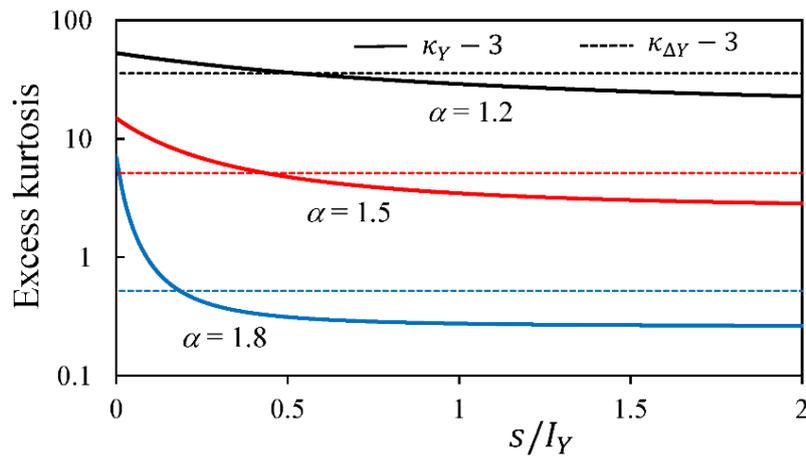
997



998

999 *Figure A2: Probability density function of ΔY for $\sigma_Y^2 = 1$, three lags and (a) $\alpha = 1.8$, (b)*
 1000 *1.5, and (c) 1.2 (solid curves). Also shown are Gaussian pdfs having the same variance*
 1001 *as ΔY (dashed curves). Lags are normalized with respect to I_Y .*

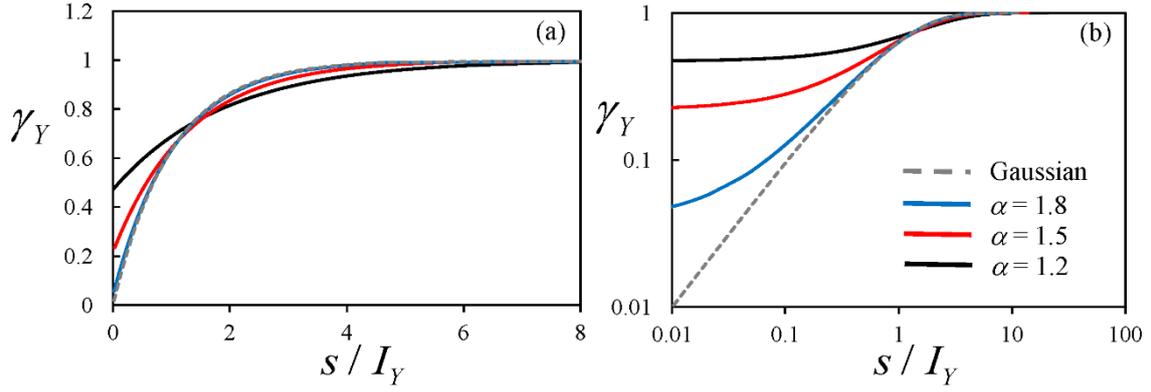
1002



1003

1004 *Figure A3: Excess kurtosis of Y' (dashed horizontal lines) and ΔY (solid curves) versus*
 1005 *normalized lag for $\alpha = 1.2, 1.5,$ and 1.8 .*

1006



1007

1008 *Figure A4: Variogram of Y' obtained for $\alpha = 1.2, 1.5,$ and 1.8 on (a) arithmetic and (b)*
 1009 *logarithmic scales (solid curves). Also shown for comparison the variogram associated*
 1010 *with a Gaussian field (and corresponding to $\alpha \rightarrow 2$) characterized by the same variance*
 1011 *and integral scale as Y' (dashed curve).*

1012

1013

Appendix B: Transport Metrics

1014

B.1. Percentile average of concentration curves

1015

1016

1017

1018

1019

1020

1021

We obtain percentile-averaged curves over all MC realizations of (i) section-integral longitudinal profiles of concentration at a given observation time and (ii) temporal distributions of mass flux through a given horizontal plane (or BTCs), here denoted as \bar{C} and \bar{J} , respectively. Following the terminology used by Lu and Stauffer (2012) for BTCs, the percentile average results from averaging over the times of the cumulative distribution of \bar{J} at fixed percentiles. By analogy, here we also extend this technique to longitudinal concentration profiles.

1022

1023

The first step of the procedure is to rearrange in ascending order the collection of particle positions / arrival times corresponding to each realization, i.e.,

1024

$$Z_{p,r}(t_i) < Z_{p+1,r}(t_i), \quad \forall p \in \{1, 2, \dots, (n-1)\}, \quad (\text{B1})$$

1025

$$T_{p,r}(z_j) < T_{p+1,r}(z_j), \quad \forall p \in \{1, 2, \dots, (n-1)\}, \quad (\text{B2})$$

1026 By computing the arithmetic average of these quantities across MC realizations as
 1027 in (11), each particle index p is associated with a value of $\langle Z_p(t_i) \rangle$ or $\langle T_p(z_j) \rangle$. According
 1028 to the above-mentioned definition of percentile average, these values correspond to
 1029 regular increments in the cumulative (i.e., integral) forms of $\bar{C}(z, t_i)$ and $\bar{J}(z_j, t)$, i.e.,

$$1030 \quad \int_0^z \bar{C}(\zeta, t_i) d\zeta = \frac{1}{n} \sum_{p=1}^n \mathcal{H}(\langle Z_p(t_i) \rangle - z), \quad (\text{B3})$$

$$1031 \quad \int_0^t \bar{J}(z_j, \tau) d\tau = \frac{1}{n} \sum_{p=1}^n \mathcal{H}(\langle T_p(z_j) \rangle - t), \quad (\text{B4})$$

1032 where \mathcal{H} is the Heaviside function. Differentiation of (B3) and (B4) yields

$$1033 \quad \bar{C}(z, t_i) = \frac{1}{n} \sum_{p=1}^n \delta(\langle Z_p(t_i) \rangle - z), \quad (\text{B5})$$

$$1034 \quad \bar{J}(z_j, t) = \frac{1}{n} \sum_{p=1}^n \delta(\langle T_p(z_j) \rangle - t), \quad (\text{B6})$$

1035 where δ is the Dirac delta. Since n is a finite number, these expressions need to be
 1036 smoothed to produce adequate estimates. This is accomplished by replacing δ in (B5) -
 1037 (B6) with a smoothing kernel W characterized by a smoothing bandwidth h . This
 1038 constitutes the basis of Kernel Density Estimation (KDE), which is summarized in
 1039 Appendix D.

1040 B.2. Moments

1041 We study the evolution of the spatial and temporal moments of the section-integral
 1042 concentrations (C_r) and the breakthrough curves (J_r), respectively, in each realization r ,
 1043 as rendered through

$$1044 \quad \mu_{1,r}^S(t_i) = \int_0^{L_z} z C_r(z, t_i) dz, \quad \mu_{\kappa,r}^S(t) = \int_0^{L_z} (z - \mu_{1,r}^S(t_i))^\kappa C_r(z, t_i) dz, \quad \kappa \geq 2, \quad (\text{B7})$$

$$1045 \quad \mu_{1,r}^T(z_j) = \int_0^\infty t J_r(z_j, t) dt, \quad \mu_{\kappa,r}^T(z) = \int_0^\infty (t - \mu_{1,r}^T(z_j))^\kappa J_r(z_j, t) dt, \quad \kappa \geq 2, \quad (\text{B8})$$

1046 where κ is the moment order, and superscripts S and T stand for *spatial* and *temporal*,
 1047 respectively. In practice, integrals in (B7) and (B8) are estimated directly from the
 1048 discrete particle positions and arrival times, respectively, as

$$1049 \quad \mu_{1,r}^S(t_i) = \frac{1}{n} \sum_{p=1}^n Z_{p,r}(t_i), \quad \mu_{\kappa,r}^S(t_i) = \frac{1}{n} \sum_{p=1}^n [Z_{p,r}(t_i) - \mu_{1,r}^S(t_i)]^\kappa, \quad \kappa \geq 2, \quad (\text{B9})$$

$$1050 \quad \mu_{1,r}^T(z_j) = \frac{1}{n} \sum_{p=1}^n T_{p,r}(z_j), \quad \mu_{\kappa,r}^T(z_j) = \frac{1}{n} \sum_{p=1}^n [T_{p,r}(z_j) - \mu_{1,r}^T(z_j)]^\kappa, \quad \kappa \geq 2. \quad (\text{B10})$$

1051 We also define the standardized moments (i.e., skewness for $\kappa = 3$, kurtosis for
 1052 $\kappa = 4$)

$$1053 \quad \tilde{\mu}_{\kappa,r}^S(t_i) = \frac{\mu_{\kappa,r}^S(t_i)}{[\mu_{2,r}^S(t_i)]^{\frac{\kappa}{2}}}, \quad \tilde{\mu}_{\kappa,r}^T(z_j) = \frac{\mu_{\kappa,r}^T(z_j)}{[\mu_{2,r}^T(z_j)]^{\frac{\kappa}{2}}}, \quad \kappa \geq 3. \quad (\text{B11})$$

1054 We noted that temporal moments computed in our study from the full set of particle
 1055 arrival times (see (B10)) were mainly controlled by the tails of the associated
 1056 distributions. BTCs in heterogeneous porous media often display heavy tails that decay
 1057 slowly with a power-law behavior (e.g., Pedretti et al., 2013). Even truncated power laws
 1058 may have very high values of higher-order moments that would hardly be accurately
 1059 inferred from simulations with a reasonable number of particles. To overcome this issue,
 1060 temporal moments were computed by excluding the 1% highest arrival times.

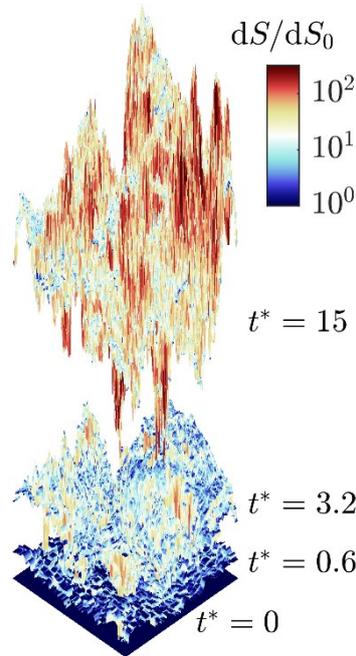
1061 All realizations were checked to fulfill the following expressions, corresponding to
 1062 a flux-weighted injection, for the spatial and temporal first moments at all evaluation
 1063 times t_i and all evaluation positions z_j , respectively:

$$1064 \quad \frac{\partial \mu_{1,r}^S}{\partial t}(t_i) \geq \bar{v}, \quad \mu_{1,r}^T(z_j) = \frac{z_j}{\bar{v}}. \quad (\text{B12})$$

1065 B.3. Key arrival times

1066 As a complement to temporal moments, we evaluate some additional BTC features
 1067 that may provide further insight on transport behavior. Namely, we compute: the first
 1068 arrival time ($T_{1,r}$), the arrival time corresponding to the peak of the corresponding
 1069 concentration curve ($T_{\text{peak},r}$) and the 99th percentile arrival time ($T_{\text{p99},r}$). The peak is
 1070 identified in each realization r from the KDE reconstruction of the BTC (see Appendix
 1071 D), whereas the two other measures are obtained directly from particle arrival times.

1072 B.4. Advective stretching



1073

1074 *Figure B1. Graphical depiction of the evolving shape of a material plane under*
 1075 *heterogeneous advection, and mapping of the local surface growth.*

1076

1077 We quantify hydrodynamic deformation (advective stretching) by tracking a mesh
 1078 of triangular surfaces defined by the initial relative position of particles in the scenario
 1079 corresponding to $Pe \rightarrow \infty$. Figure B1 illustrates the local growth of the individual
 1080 triangular surfaces dS at three observation times in a selected realization of Y . In Section
 1081 3 we focus on the ratio S/S_0 , where S is the sum of all triangular dS , and $S_0 = L_x L_y$.

1082 **Appendix C: Quantification of Distribution Overlaps**

1083 Here, we provide a synthesis of the basic concepts of Bayesian classification,
 1084 focusing on the methods we employ in our study.

1085 **C.1. The Bayes classifier**

1086 The Bayes classifier assigns observations of a (multi-dimensional random) vector
 1087 \mathbf{R} , with values \mathbf{r} , to a class K (which has a finite number N_k of possible values k)
 1088 according to data-driven estimates of the conditional probability density functions
 1089 $f_k(\mathbf{r}) = p(\mathbf{R} = \mathbf{r} | K = k)$. Ideally, it is the classification method with the least associated
 1090 error provided that the true *pdfs*, f_k , are known or reproduced correctly. Given an
 1091 observation with a value \mathbf{r} , this method assigns it to the category k for which $g_k(\mathbf{r}) =$
 1092 $P(K = k | \mathbf{R} = \mathbf{r})$ is highest among all possible k . According to Bayes theorem

$$1093 \quad g_k(\mathbf{r}) = \frac{p_k f_k(\mathbf{r})}{\sum_{\ell=1}^{N_k} p_\ell f_\ell(\mathbf{r})}, \quad (\text{C1})$$

1094 where $p_k = P(K = k)$ is the prior probability of any observation to belong to class k . If
 1095 no prior information is available, all $p_\ell = p_k$ ($\forall \ell \neq k$), and the classifier will assign the
 1096 observation \mathbf{r} to the class k for which f_k is highest. If we assume that all f_k are Gaussian,
 1097 with mean $\hat{\mathbf{m}}_k$ and covariance matrix $\hat{\Sigma}_k$ estimated from data, then the classification is
 1098 equivalent to assigning observation \mathbf{r} to the class k for which

$$1099 \quad \delta_k(\mathbf{r}) = -(\mathbf{r} - \hat{\mathbf{m}}_k)^T \hat{\Sigma}_k^{-1} (\mathbf{r} - \hat{\mathbf{m}}_k) - \log |\hat{\Sigma}_k| + 2 \log p_k \quad (\text{C2})$$

1100 is highest. This approach is known as Quadratic Discriminant Analysis (QDA).

1101 For one-dimensional variables, R , it may in some cases be feasible to reconstruct
1102 $f_k(r)$ from data, instead of assuming that it is Gaussian. We do so in our study upon
1103 relying on Kernel Density Estimation (see Appendix D). Each observation r is then
1104 assigned to the class k with the highest $g_k(r)$ (see (C1)). We term this approach as *KDE*
1105 *discriminant*.

1106 C.2. Confusion matrix and Overlap matrix

1107 In discriminant analysis, the Confusion Matrix compares the obtained classification
1108 (rows) with the actual classes (columns) of a number of random observations, so that the
1109 entries on the diagonal of the matrix correspond to correct classifications. Normalization
1110 of the columns by the total number of observations of each class yields a normalized
1111 confusion matrix. The latter informs us on the likelihood of correct (diagonal) and
1112 incorrect (outside diagonal) classification. In other words, if we define variable \widehat{K} as the
1113 estimated class (and K as the true class), the entries of the normalized confusion matrix
1114 are $P'_{\ell k} = P(\widehat{K} = \ell | K = k)$.

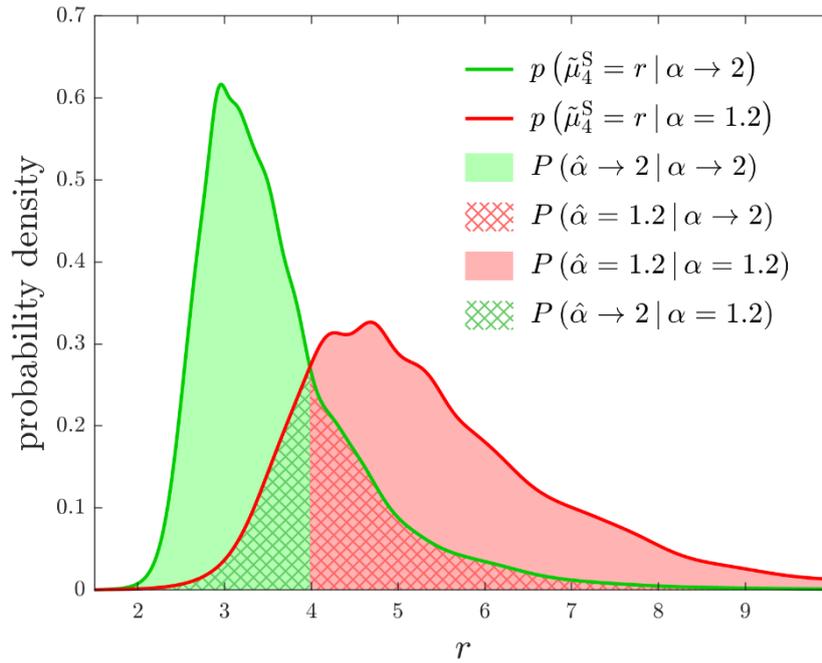
1115 We define the overlap of a *pdf*, f_k , (conditioned to $K = k$) with f_ℓ (conditioned to
1116 $K = \ell$) as the likelihood that, if the class given by the discriminant analysis, \widehat{K} , is ℓ , the
1117 actual class, K , is k ; i.e., $P_{k\ell} = P(K = k | \widehat{K} = \ell)$. The quantity $P_{k\ell}$ (and in particular, its
1118 values for $k = \ell$) indicates the level of reliability associated with relying on the *pdfs* to
1119 estimate the actual class of an individual realization. Mapping $P'_{\ell k}$ onto $P_{k\ell}$ is
1120 accomplished through Bayes theorem

$$1121 P_{k\ell} = \frac{p_k P'_{\ell k}}{\sum_{m=1}^{N_k} p_m P'_{\ell m}} \quad (C3)$$

1122

1123

C.3. Implementation



1124

1125 *Figure C1: Graphical depiction exemplifying the integration procedure for the*
 1126 *evaluation of the $P'_{\ell k}$ entries of the normalized confusion matrix using the KDE classifier*
 1127 *for the early-time spatial kurtosis.*

1128

1129 In our study, the classes are the four studied values of α , and the random
 1130 observations are the selected metrics evaluated from the solute transport results. We use
 1131 the KDE discriminant for the analysis of each of these metrics, whereas we apply QDA
 1132 for the joint analysis of several metrics. In both cases, the final result of interest is the
 1133 overlap matrix defined in Section C.2, obtained from the normalized confusion matrix by
 1134 (C3). In the case of the KDE discriminant, the entries of the normalized confusion matrix
 1135 can be obtained by numerical integration of the relevant parts of the *pdfs* (see Figure C1
 1136 as a graphical example), whereas the confusion matrix for the QDA is obtained by
 1137 explicitly classifying all observations.

1138 Note that the KDE approach and the QDA approach require a training dataset (for
 1139 the KDE reconstruction of the *pdf* in the former and for the estimation of the mean and

1140 covariance in the latter). Relying on the same dataset for training and for testing (i.e., for
1141 building the confusion matrix) could slightly bias the results towards correct classification
1142 (that is, towards a non-overlap result). However, we expect this effect to be negligible
1143 given the rather large size of the dataset (i.e., 5000 observations of each class).

1144 **Appendix D: Kernel Density Estimation**

1145 Here, we provide a brief introduction to univariate Kernel Density Estimation
1146 (KDE), which we use to reconstruct curves, $f(x)$, which are (or can be treated as)
1147 probability density functions, relying on a limited number of samples X_p , $p = 1, \dots, n$.
1148 The KDE is formulated as

$$1149 \quad f(x) \cong \hat{f}(x) = \frac{1}{n} \sum_{p=1}^n W(X_p - x; h), \quad (\text{D1})$$

1150 where W is a kernel or smoothing function, which we choose to be Gaussian, i.e.,

$$1151 \quad W(u; h) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{u^2}{2h^2}\right), \quad (\text{D2})$$

1152 h being the kernel bandwidth.

1153 When X_p are particle longitudinal positions or arrival times, $f(x)$ is the longitudinal
1154 density of solute mass (8) or the breakthrough curve (9), respectively (Fernández-García
1155 and Sánchez-Vila, 2011; Sole-Mari et al., 2017, 2019; Sole-Mari and Fernández-García,
1156 2018; Pedretti and Fernández-García, 2013). Alternatively, the collection of samples X_p
1157 can be formed by evaluations of a scalar metric that varies across MC realizations and in
1158 this case $f(x)$ is the probability density function of the metric x (see Appendix C).

1159 A key aspect of the accuracy of KDE is to select an optimal value for the bandwidth
1160 h . A widely popular technique employed for this aim relies on minimizing the Asymptotic

1161 Mean Integrated Squared Error (AMISE). It can be shown that, for a Gaussian W , in the
1162 limit when $nh \rightarrow \infty$ and $h \rightarrow 0$ (e.g. Silverman, 1986)

$$1163 \quad \text{AMISE} \approx (2\sqrt{\pi}nh)^{-1} + \frac{h^4}{4} \int_{\mathbb{R}} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 dx. \quad (\text{D3})$$

1164 By setting $\frac{\partial \text{AMISE}}{\partial h} = 0$, the absolute minimum AMISE can be shown to be obtained for the
1165 following choice of bandwidth

$$1166 \quad h = \left[2\sqrt{\pi} n \int_{\mathbb{R}} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 dx \right]^{-\frac{1}{5}}. \quad (\text{D4})$$

1167 Since $\int_{\mathbb{R}} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 dx$ is *a priori* unknown, several iterative methodologies have been
1168 developed to estimate it from the data and solve equation (D4). Here, we apply one such
1169 methodology, namely the Improved Sheater-Jones (ISJ) algorithm developed by Botev et
1170 al. (2010).