# Generalized Sub-Gaussian processes: theory and application to hydrogeological and geochemical data

**Martina Siena[1], Alberto Guadagnini[1], Arnaud Bouissonnié[2], Philippe Ackerer[2], Damien Daval[2], Monica Riva[1]**

[1]Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy

[2]Université de Strasbourg-CNRS ENGEES/EOST, Laboratoire d'Hydrologie et de Géochimie de Strasbourg, Strasbourg, France

Corresponding author: Martina Siena (martina.siena@polimi.it)

**Key Points:**

- We develop the theoretical formulation of the Generalized Sub-Gaussian (GSG) model for a general distributional form of the subordinator.
- The GSG formulations tested on laboratory- and field-scale data effectively capture the observed scale dependence of increments' statistics.
- Our formulations can improve flexibility and accuracy of the GSG model, supporting its applicability to a wide range of data.

## Abstract

We start from the well-documented scale dependence displayed by the probability distribution and associated statistical moments of a variety of hydrogeological and soil science variables and their spatial or temporal increments. These features can be captured by a Generalized Sub-Gaussian (GSG) model, according to which a given variable, *Y*, is subordinated to a (typically spatially correlated) Gaussian random field, *G*, through a subordinator, *U*. This study extends the theoretical framework originally proposed by Riva et al. (2015a) to include the possibility of selecting a general form of the subordinator, thus enhancing the flexibility of the GSG framework for data interpretation and modeling. Analytical expressions for the GSG process associated with lognormal, Pareto, and Gamma subordinator distributions are then derived. We demonstrate the ability of the GSG modeling framework to capture the way key features of the statistics associated with two datasets transition across scales. The latter correspond to variables which are typical of a geochemical and a hydrogeological setting, i.e., (*i*) data characterizing the micrometer-scale surface roughness of a crystal of calcite, collected within a laboratory-scale setting, resulting from induced mineral dissolution; and (*ii*) a vertical distribution of decimeter-scale porosity data, collected along a deep km-scale borehole within a sandstone formation and typically used in hydrogeological and geophysical characterization of aquifer systems.The theoretical developments and the successful applications of the approach we propose provide a unique framework within which one can interpret a broad range of scaling behaviors displayed by a variety of Earth and environmental variables in various scenarios.

## Plain Language Summary

Characterization of hydrogeological and geochemical systems aims at assessing the heterogeneity and scale dependency exhibited by their attributes and the associated key statistics. It has been shown that complex scaling features documented for the statistics of a wide range of Earth, environmental (and several other) variables and their spatial/temporal increments can be captured through a Generalized Sub-Gaussian (GSG) model. The latter relies on the subordination of a Gaussian random field through a subordinator. This study extends the theoretical framework originally proposed for the GSG model to include multiple choices of the subordinator distribution. We provide the theoretical formulation and discuss the main features of the GSG model resulting from (*i*) a general form of the subordinator and (*ii*) three selected distributional forms. We show

the effectiveness of the GSG modeling framework for the interpretation of real data encompassing a considerably wide range of scales by analyzing (*i*) a set of surface topography (roughness) data collected on a calcite sample in a laboratory-scale geochemical setting; and (*ii*) a field-scale distribution of porosity data, collected along a deep borehole within a sandstone formation.

## 1 Introduction

Geostatistical models adopted for the interpretation of key features of spatial heterogeneity of quantities related to subsurface flow and transport processes consider available observations of a variable of interest as samples from a random field with a given distribution. Analyses of a wide collection of datasets of hydrogeological attributes, including, *e.g.*, (log) hydraulic conductivity and permeability (Liu & Molz, 1997; Meerschaert et al., 2004; Painter, 2001; Painter, 1996, Riva et al., 2013a, 2013b; Siena et al., 2012, 2019), electrical resistivity (Painter, 2001), and neutron porosity (Guadagnini et al., 2015; Riva et al., 2015a) observations, clearly document the occurrence of distinct non-Gaussian features characterizing their distributions. Notably, it has been shown that spatial increments, $\Delta Y(\mathbf{s}) = Y(\mathbf{x}+\mathbf{s}) - Y(\mathbf{x})$, evaluated over separation distance (or lag) **s** (**x** being a position vector) of a given quantity $Y$ are characterized by distributions displaying peaks that become sharper and tails that tend to become heavier with decreasing lag. A similar behavior, corresponding to distributions transitioning from heavy tailed at small lags to seemingly-Gaussian at increased lags, is documented by analyses of a variety of spatial and/or temporal increments of environmental data, including sediment transport processes (*e.g.*, Ganti et al., 2009) and fully developed turbulence (Boffetta et al., 2008) as well as datasets of Earth, environmental and several other variables (see Neuman et al., 2013 and references therein). Such a scale dependence is directly imprinted to the associated (statistical) moments of increment distributions.

All of these evidences suggest that modeling the (spatially correlated) variability of $Y$ through a Gaussian model is not generally warranted. With specific reference to the spatial variability of hydrogeologic quantities, a number of studies evidence that the heterogeneity of natural aquifers is generally more complex than what can be captured through a Gaussian model (*e.g.*, among others, with reference to hydraulic conductivity, Gómez-Hernández & Wen, 1998; Haslauer et al., 2012; Mariethoz et al., 2010; Xu & Gómez-Hernández, 2015 and references therein).

In this context, it is also noted that attributes/properties of porous media that at some scale can be considered as composed by distinct facies/regions, each corresponding to a given material characterized by an internal degree of heterogeneity, could be represented through multi-modal distributions (see, *e.g.*, Desbarats, 1990; Lu & Zhang, 2002; Rubin, 1995; Russo, 2002, 2010; Winter et al., 2003 and references therein). The latter are representative of a conceptual (and mathematical) model that views the otherwise composite nature of the system as a unique continuum at the given scale of observation, natural variability within each region being characterized by a statistical behavior of the kind described above.

Riva et al. (2015a, b) show that the above illustrated scale-dependent behavior of the probability density function (pdf) of $\Delta Y$ can be captured through a Generalized Sub-Gaussian (GSG) model. This theoretical framework relies on the idea that the spatial random field $Y(\mathbf{x}) = \langle Y \rangle + Y'(\mathbf{x})$, $\langle Y \rangle$ and $Y'(\mathbf{x})$ being respectively the ensemble mean and a local zero-mean fluctuation, can be interpreted through the following model

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

where $G(\mathbf{x})$ is a zero-mean, Gaussian random field and $U(\mathbf{x})$ is a so-called *subordinator*, independent of *G*, consisting of statistically independent identically distributed (iid) non-negative random variables. The underlying Gaussian random field generally (but not necessarily) displays a multi-scale nature which can be captured, for example, through a geostatistical description based on a Truncated Power Variogram model (*e.g.*, Di Federico & Neuman, 1997; Neuman & Di Federico, 2003).

As opposed to mathematical models based on multifractals (*e.g.*, Boffetta et al., 2008; Frisch, 2016; Lovejoy & Schertzer, 1995; Mandelbrot, 1974; Monin & Yaglom, 1975; Veneziano et al., 2006) or fractional Laplace approaches (*e.g.*, Kozubowski et al., 2006, 2013; Meerschaert et al., 2004), which have been employed to mimic the above-mentioned pattern of increment frequency distributions, the GSG model enables one to represent jointly within a unique theoretical framework the documented behavior (as described by probability distributions and/or moments) of a quantity and its incremental values.

Riva et al. (2015a) provide the first analytical formulation of the GSG model, illustrating that the characteristic scaling behavior of the increments results from the decay of the correlation function of the underlying Gaussian random field with increasing lag. Riva et al. (2015b) illustrate an approach for the generation of unconditional random realizations of statistically isotropic or anisotropic GSG fields in multiple dimensions. Panzeri et al. (2016) develop an algorithm for the generation of GSG fields conditional to a given set of data. Siena et al. (2019) rely on the GSG model for the interpretation of the spatial variability of a set of air permeability data collected along a core of limestone. Guadagnini et al. (2018) present a 9-step procedure for the detection of GSG signatures in a given dataset. Notably, theoretical developments and applications to date rest solely on a lognormal distribution of $U$ characterized by a single parameter, thus limiting the range of possible applications of the GSG model.

The present study focuses on a generalization of the GSG framework by extending the formulation of Riva et al. (2015a) to include a generic subordinator. This enables us to enhance the flexibility of the model for data interpretation and modeling by taking into account specific features exhibited by the way statistics of a given dataset transition across scales. We then demonstrate the applicability of the general theoretical framework by considering a (*i*) two-parameter lognormal; (*ii*) Pareto; and (*iii*) Gamma distributional form of $U$ and developing the ensuing analytical expressions for the GSG process. We analyze in this context two datasets associated with differing processes and observation scales. The first application includes direct observations of μm-scale surface topography (or roughness) of mm-scale calcite crystals resulting from induced mineral dissolution. Calcite is a common mineral in the Earth's crust and is characterized by significant dynamics of its surface, depending on environmental conditions (*e.g.* Fischer et al., 2012; Jordan & Rammensee, 1998; Noiriel et al., 2009, 2020). Acquisition of the type of data we consider is subject to increased interest to characterize micro-scale geochemical processes deriving from interactions taking place at fluid-rock interfaces (*e.g.*, Bouissonnié et al., 2018; Pollet-Villard et al., 2016a, b and references therein). While the possibility of acquiring these direct observations is continuously enhanced through the use of modern atomic force microscopy and vertical scanning interferometry, statistical analyses of available datasets are still limited to standard variography (Pollet-Vilard et al., 2016a). As an additional test-bed, we analyze a vertical profile of neutron porosity data, collected along a deep borehole in a sandstone formation and encompassing a vertical depth of about 1 km at a 15-cm resolution (Dashtian et al., 2011). As these types of data are routinely available in (hydro)geological and geophysical subsurface exploration, they constitute a remarkable dataset to assess the applicability of our statistical scaling framework at such scales.

The work is structured as follows. Section 2 illustrates the key features of the GSG model and describes a moment-based method of inference of model parameters. The detailed original analytical formulation of the GSG model associated with a generic subordinator and the ensuing derivations for the three subordinators here considered are provided in Appendix A and B, respectively. In Section 3 we compare the performance of these three alternative GSG models for the interpretation of the two datasets illustrated above. Concluding remarks are provided in Section 4.

## 2 Generalized Sub-Gaussian model

### 2.1 Theoretical framework

Zero-mean fluctuations, $Y'$, at two spatial locations, $\mathbf{x}_1$ and $\mathbf{x}_2$, can be expressed as

$$Y'(\mathbf{x}_i) = U(\mathbf{x}_i) G(\mathbf{x}_i) = Y_i' = U_i G_i, \qquad \text{with } i = 1, 2. \tag{2}$$

The bivariate pdf of $Y_1'$ and $Y_2'$ is (Riva et al., 2015a)

$$f_{Y_1',Y_2'}(y_1', y_2') = \int_0^\infty \int_0^\infty f_{U_1}(u_1)\, f_{U_2}(u_2)\, f_{G_1G_2}\left(\frac{y_1'}{u_1}, \frac{y_2'}{u_2}\right) \frac{du_2}{u_2} \frac{du_1}{u_1}. \tag{3}$$

Here $f_{U_i}(u_i)$ is the pdf of $U_i$ and $f_{G_1G_2}$ is the bivariate pdf of $G_1$ and $G_2$, given by

$$f_{G_1G_2}\left(\frac{y_1'}{u_1}, \frac{y_2'}{u_2}\right) = \frac{e^{-\frac{1}{2\sigma_G^2(1-\rho_G^2)}\left(\frac{y_1'^2}{u_1^2} + \frac{y_2'^2}{u_2^2} - 2\rho_G \frac{y_1'}{u_1}\frac{y_2'}{u_2}\right)}}{2\pi\sigma_G^2\sqrt{1-\rho_G^2}}, \tag{4}$$

where, $\sigma_G^2$ is the variance of *G* and $\rho_G$ is the correlation coefficient between $G_1$ and $G_2$, which typically decreases as the separation distance (or lag) $s = |\mathbf{x}_1 - \mathbf{x}_2|$ increases. Starting from Riva et al. (2015a), who developed the analytical framework for the specific case of a single-parameter lognormal subordinator, we provide in Appendix A an original theoretical formulation of the GSG model considering a generic distributional form of *U*. It is worth noting that, regardless the distributional form of *U*, the variogram of $Y'$ is always characterized by a nugget effect (see Eq. A14), rendered by the product of the variance of *G* and the variance of *U*. This result implies that nugget effects, which are typically considered to appear due to variability of $Y'$ at scales smaller than the sampling interval and/or to measurement errors, may in fact be (at least in part) considered as a symptom of non-Gaussianity of the type embedded in the GSG theoretical framework.

The general framework introduced in Appendix A encompasses multiple possible formulations of the GSG model. In this context, we evaluate three possible alternative models for $U$, corresponding to a lognormal, Pareto, or Gamma distribution. Each of these models is characterized by $N_P = 2$ parameters, respectively controlling the shape (*shape* parameter) and the spreading (*scale* parameter) of the pdfs of the ensuing GSG formulation for $Y'$. Hereinafter, we denote the latter as LN-GSG, P-GSG, and Γ-GSG for the lognormal, Pareto, and Gamma subordinator, respectively. The theoretical formulation of each of these GSG models is provided in Appendix B.

Equation (A7) indicates that the pdfs $f_{\Delta Y}$ of incremental values ($\Delta Y$) corresponding to differing lags depend on (*i*) $\sigma_G^2$ and the $N_P$ parameters of $U$; and (*ii*) $\rho_G$. While the former parameters are constant for all lags, the correlation function of $G$ is lag-dependent, thus imprinting a scaling behavior, i.e., an intrinsic variability with lag, to the shape of the pdf of incremental values of $Y'$, independent of the GSG model considered. This feature is clearly illustrated in Figures 1a-c, where we depict $f_{\Delta Y}$ for selected values of the three GSG model parameters (analytical expressions being collected in Eq. (B8)) and three values of $\rho_G$ corresponding to short, intermediate and large lags. The pattern associated with the behavior of peaks and tails of the pdfs of $Y'$ and $\Delta Y$ can be described quantitatively by analyzing their standardized kurtosis, $\kappa_{Y'}$ (see Eq. (A6)) and $\kappa_{\Delta Y}$ (see Eq. (A11)), respectively, deviations from Gaussianity being clearly revealed by the excess kurtosis, $\kappa_{Y'} - 3$ and $\kappa_{\Delta Y} - 3$. As these quantities increase, the peak of the pdf of $Y'$ or $\Delta Y$ grows sharper and the associated tails become heavier. Figures 1d-f depict the excess kurtosis of $Y'$, as well as of $\Delta Y$, as a function of $\rho_G$ for selected values of the shape parameter $\alpha$ (for the LN-GSG model, Fig. 1d), $a$ (for the P-GSG model, Fig. 1e), and $k$ (for the Γ-GSG model, Fig. 1f). Inspection of these figures, together with Eqs. (B7) and (B11), suggests that for all GSG models (*i*) $\kappa_{Y'} - 3$ and $\kappa_{\Delta Y} - 3$ do not depend on the scale parameter of the subordinator and on the variance of $G$; (*ii*) for a given value of $\rho_G$, $\kappa_{Y'} - 3$ and $\kappa_{\Delta Y} - 3$ increase as the shape parameter of $U$ decreases; (*iii*) for a given value of the shape parameter, $\kappa_{\Delta Y} - 3$ increases as $\rho_G$ increases (or, equivalently, as lag decreases), i.e., the pdfs of $\Delta Y$ transition with lag. One can note that, in all cases, $\kappa_{\Delta Y} - 3$ exceeds zero by a significant margin at small lags (i.e., as $\rho_G \to 1$), even for the largest values of the shape parameter of $U$ considered. With reference to the LN-GSG model, Figure 1d and Eqs. (B7) and (B11) highlight that there is a threshold value of the shape parameter, corresponding to $\alpha_T = 2 - \sqrt{\ln 3} \approx 0.95$, such that (*i*) for $\alpha < \alpha_T$, the pdfs of $\Delta Y$ are characterized by lower peaks and lighter tails than those of $Y'$ at all lags; while (*ii*) for $\alpha > \alpha_T$, $\kappa_{\Delta Y} - 3$ is higher/lower than $\kappa_{Y'} - 3$ at small/large lags (see also Riva et al., 2015a). An analogous behavior is exhibited by the results associated with the Γ-GSG model (Fig. 1f), the threshold value, $k_T$, of the shape parameter being equal to 1.0. Otherwise, one can demonstrate analytically (see also Fig. 1e) that $\kappa_{\Delta Y} - 3$ is always larger than $\kappa_{Y'} - 3$ at small lags for the P-GSG model, regardless the value of the shape parameter *a*. Besides, the range of values of $\rho_G$ for which ($\kappa_{\Delta Y} - 3$) > ($\kappa_{Y'} - 3$) (i.e., the range of lags where the pdfs of the increments display sharper peaks and heavier tails than the pdf of $Y'$) tends to increase as *a* decreases.

### 2.2 Parameter estimation methods

The Method of Moment (MOM) is a straightforward way to infer model parameters from a dataset. Here, we illustrate two approaches to estimate model parameters through MOM. These are respectively based on (*i*) sample statistics of the parent variable (Method A) and (*ii*) sample statistics of both the parent variable and the incremental data at multiple lags (Method B). Sections 2.2.1 and 2.2.2 examine merits and drawbacks of these methods.

#### 2.2.1 Parameter estimation Method A

Method A (henceforth denoted as MOM_A) relies on the marginal frequency distribution and associated moments of $Y'$. Estimates of GSG model parameters are obtained by replacing $\langle Y'^2 \rangle$ and $\langle Y'^4 \rangle$ in Eqs. (A3) and (A6) with their sample counterparts, $M_2^{Y'}$ and $M_4^{Y'}$, inferred from data. The shape parameter of $U$ for the LN-GSG, P-GSG, and Γ-GSG models can be estimated by making use of Eq. (B7) as

$$\frac{M_4^{Y'}}{3\left(M_2^{Y'}\right)^2}=\begin{cases} e^{4(2-\alpha)^2}, & \text{with } \alpha<2 \quad \text{for LN-GSG} \\ \dfrac{(a-2)^2}{a(a-4)}, & \text{with } a>4 \quad \text{for P-GSG} \\ 1+\dfrac{4k+6}{k\left(1+k\right)}, & \text{with } k>0 \quad \text{for } \Gamma\text{-GSG} \end{cases} \tag{5}$$

Then, by making use of Eq. (B5), one can estimate the product between $\sigma_G$ and the scale parameter of *U* (*e.g.*, $e^{\mu}$, *b*, and *θ*, for LN-GSG, P- GSG and Γ-GSG model, respectively) as

$$\sigma_G^2 e^{2\mu}=\frac{M_2^{Y'}}{e^{2(2-\alpha)^2}} \qquad \text{for LN-GSG,} \tag{6}$$

$$\sigma_G^2 b^2 = M_2^{Y'}\,\frac{a-2}{a} \qquad \text{for P-GSG,} \tag{7}$$

$$\sigma_G^2 \theta^2 = \frac{M_2^{Y'}}{k\left(1+k\right)} \qquad \text{for } \Gamma\text{-GSG.} \tag{8}$$

It is noted that the analytical expressions of the marginal pdf (as well as its statistical moments) of $Y'$ for all GSG models are characterized by the scale parameter of the subordinator being always coupled with the scale parameter, $\sigma_G$, of the underlying Gaussian process (see Eqs. (B4) - (B6)). It then follows that the set of Eqs.(5)-(8) fully determines $f_{Y'}\left(y'\right)$, i.e., it is not necessary to estimate $\sigma_G$ and the scale parameter of *U* independently to determine $f_{Y'}\left(y'\right)$. As an additional remark, it is noted that one cannot estimate $\rho_G$ with the methodology here implemented. As such, its application, while straightforward, does not allow ascertaining the degree of spatial correlation of the random field $Y'$.

### 2.2.2 Parameter estimation Method B

Method B (henceforth denoted as MOM_B) yields estimates of GSG model parameters by relying jointly on sample statistics of $Y'$ and $\Delta Y$. For any given lag, one replaces $\left\langle Y'^2\right\rangle$, $\left\langle \Delta Y^2\right\rangle$ and $\left\langle \Delta Y^4\right\rangle$ in Eqs. (A3), (A8), and (A9) by their sample counterparts $M_2^{Y'}$, $M_2^{\Delta Y}$, and $M_4^{\Delta Y}$, respectively. Making use of Eqs. (B5), (B9) and (B11), the resulting systems of equations are

$$\begin{cases} M_2^{Y'} = \sigma_G^2 e^{2\mu} e^{2(2-\alpha)^2} \\ \dfrac{M_2^{\Delta Y}}{2M_2^{Y'}} = 1 - \dfrac{\rho_G}{e^{(2-\alpha)^2}} \\ \dfrac{M_4^{\Delta Y}}{\left(M_2^{\Delta Y}\right)^2} = 3e^{2(2-\alpha)^2} \left\{ 1 + \dfrac{1}{2} \left( \dfrac{e^{2(2-\alpha)^2} - 1}{e^{(2-\alpha)^2} - \rho_G} \right)^2 \right\} \end{cases} \quad \text{for LN-GSG,} \qquad (9)$$

$$\begin{cases} M_2^{Y'} = \sigma_G^2 b^2 \dfrac{a}{a-2} \\ \dfrac{M_2^{\Delta Y}}{2M_2^{Y'}} = 1 - \dfrac{a(a-2)\rho_G}{(a-1)^2} \\ \dfrac{M_4^{\Delta Y}}{\left(M_2^{\Delta Y}\right)^2} = \dfrac{3}{2} \left[ \dfrac{1}{a(a-4)} - \dfrac{4\rho_G}{(a-1)(a-3)} + \dfrac{1+2\rho_G^2}{(a-2)^2} \right] \left[ \dfrac{1}{(a-2)} - \dfrac{a\rho_G}{(a-1)^2} \right]^{-2} \end{cases} \quad \text{for P-GSG,} \qquad (10)$$

$$\begin{cases} M_2^{Y'} = \sigma_G^2 \theta^2 k \left(1+k\right) \\ \dfrac{M_2^{\Delta Y}}{2M_2^{Y'}} = 1 - \dfrac{k}{1+k} \rho_G \\ \dfrac{M_4^{\Delta Y}}{\left(M_2^{\Delta Y}\right)^2} = 3 \left[ 1 + \dfrac{1}{k} + \dfrac{(k+1)\left(2 + k + \rho_G^2 k - 2k\rho_G\right)}{k\left(k+1-\rho_G k\right)^2} \right] \end{cases} \quad \text{for } \Gamma\text{-GSG.} \qquad (11)$$

Equations (9)-(11) allow estimating all parameters characterizing the joint pdf of $\Delta Y$, i.e., (*i*) the product of the scale parameters of *G* and *U*, (*ii*) the shape parameter of *U*, and (*iii*) the correlation coefficient $\rho_G$, which enables us to diagnose the dependence on lag of increment statistics. We further note that relying on the joint use of $Y'$ and $\Delta Y$ data is recommended because it leads to an (often considerably) augmented set of data upon which sample moments are evaluated, thus improving the accuracy of the estimates. This approach yields a set of three parameter estimates for each investigated lag. Riva et al. (2015a) document that MOM_B provides results of similar quality to those one could obtain upon relying on parameter estimation through analyzing incremental data at various lags via Maximum Likelihood (ML). This element, together with the high computational demand associated with ML, leads us to rely on MOM_B for the purpose of our analyses.

According to our theoretical framework (see Section 2.1), we expect that values of the shape parameter and of the product between the scale parameters of $U$ and $G$ remain (approximately) constant with lag. It then follows that an additional benefit of relying on MOM_B, as opposed to MOM_A, is that it enables one to assess the consistency of the parameter estimation results with these theoretical requirements. As already noted for the pdf of $Y'$, the pdf of $\Delta Y$ (as well as its statistical moments) for all GSG models is also characterized by the scale parameter of $U$ being always coupled with $\sigma_G$ (see Eqs. (B8) - (B10)). Therefore, the inability to provide unique estimates of the scale parameters of $U$ and $G$ (while only their product is estimated) does not hamper the use of the results of the analysis for further applications, typically involving generations of a collection of realizations of a given random field to be employed in the context of studies on flow and transport processes in a Monte Carlo framework.

## 3 Application to laboratory- and field- scale datasets

The three alternative GSG models illustrated in Section 2 and in Appendix B are here considered for the characterization of the spatial variability of two datasets. These are selected to represent two differing observation scales, i.e., a laboratory- and a field- scale setting. Both systems are characterized by the availability of a considerable amount of observations, which is achievable with modern measurement techniques, and are therefore well suited for the analysis.

### 3.1 Micrometer-scale topography of a millimeter-scale calcite sample resulting from mineral dissolution

The first dataset we consider (hereinafter denoted as Dataset 1) comprises direct observations of surface topography collected on a (104) calcite cleavage plan. While calcite is the main rock-forming mineral of limestones and has a key role in a variety of geological and biological systems, its surface is characterized by remarkable dynamics when put in contact with aqueous fluid, which are still not completely characterized, the (104) surface plane being very common in natural settings. The sample consisted of a ~5mm-sized single crystal of calcite polished through a multi-step abrasive sequence. The initial arithmetic roughness of the surface was on the order of 50 nm. The sample was introduced in a mixed-flow reactor set-up. The crystal was subject to reaction for 8 days at room temperature and at a saturation index with respect to calcite of 0.8, corresponding to conditions where dissolution occurs while the nucleation of etch pits is thermodynamically impossible. Measurements of surface topography, $(x, y)$ being spatial coordinates in the horizontal plane, are collected by means of a vertical scanning interferometer (Zygo NewView 7300) with a vertical resolution of 3 nm, on a two-dimensional grid of $N_1 = 250\times250 = 62500$ cells, with lateral resolution $dl = 2.2$ µm. Additional details of the experimental set-up and procedure are offered in Bouissonnié et al. (2018). The surface is characterized by a slight curvature, resulting from the preliminary polishing of the sample. Mean-removed topography data, $Y'$, have been obtained by subtracting the best-fitting quadratic surface from the measurements. Figure 2a depicts the spatial distribution of $Y'$, the sample standard deviation being equal to $\sqrt{M_2^{Y'}} = 0.21$ µm.

### 3.2 Field-scale neutron porosity data

Dataset 2 is a collection of neutron porosity data sampled from a (km-scale) deep vertical borehole in southwestern Iran. The data are part of a wider dataset comprising multiple wells, some of which have been recently analyzed by Dashtian et al. (2011), Riva et al. (2015a), and Guadagnini et al. (2015). The borehole considered here is drilled in the Ahwaz field (see Dashtian et al., 2011), where oil and natural gas are produced from a sandstone formation. A large number ($N_2 = 6949$) of neutron porosity data collected at a uniform distance of $dz = 15$ cm is available. The one-dimensional profile of mean-removed porosity data is depicted in Fig. 2b, the associated sample standard deviation being equal to $\sqrt{M_2^{Y'}} = 8.35\%$.

### 3.3 Results and discussion

Figures 3a and 3b depict sample pdfs of $Y'$ for Dataset 1 and 2, respectively. Depictions are provided in linear and semi-logarithmic scales for ease of analysis. A slight bimodality and asymmetry are exhibited by the pdf of porosity observations in Dataset 2, the pdf of surface topography (Dataset 1) being left-skewed. A qualitative comparison (based on visual inspection) between each of these sample pdfs and a normal distribution with the corresponding variance, $M_2^{Y'}$, (also included in the figures) suggests deviation from Gaussianity for both variables. This qualitative result is also confirmed quantitatively by the outcomes of formal (Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling) tests performed on randomly-sampled subsets of data, which reject the Gaussian model at a significance level of 0.05 for both datasets.

We compute sample statistics of incremental data, $\Delta Y$, evaluated (*i*) along all directions in the *x-y* plane for Dataset 1 and (*ii*) along the *z* axis for Dataset 2. The pdfs of $\Delta Y$ at three diverse lags ($s$ = 1, 5, and 50 *dl* for Dataset 1; and $s$ = 5, 50, and 250 *dz* for Dataset 2) are depicted in Figs. 4a and 4b, respectively. As a term of comparison, corresponding normal distributions with the same variance are juxtaposed to the increment pdfs. These results illustrate that sample pdfs of increments (*i*) exhibit the characteristic scale dependence mentioned in Section 1; and (*ii*) progressively tend to distributions with lower peaks and lighter tails, resembling the Gaussian distribution as lag increases, this feature being particularly evident for Dataset 2.

Figures 5a and 5b depict the dependence of sample values of $\kappa_{\Delta Y}-3$ on lag for Dataset 1 and Dataset 2, respectively, dashed horizontal lines denoting values of excess kurtosis of the parent variable $Y'$. For both sets, incremental data excess kurtosis is significantly larger than zero at small lags. Excess kurtosis (EK) of (omnidirectional) incremental data associated with Dataset 1 decreases rapidly as lag increases and tends to attain a quite stable value of ≈ 3.5 at large lags. Otherwise, values of EK for Dataset 2 tend to consistently decrease across the whole range of lags considered, attaining values smaller than 1 (i.e., approaching a Gaussian distribution, consistent with the qualitative result depicted in Fig. 4b) from $s$ = 400 *dz*.

To provide an appraisal of the accuracy associated with the sample estimates of EK, we apply a standard bootstrapping technique (Efron, 1992) to each set of incremental data. This procedure relies on sampling (with replacement) from a collection of $\Delta Y$ data related to a given lag a total of $m$ (here we set $m$ = 10,000) sets, each characterized by the same number of elements of the original collection of $\Delta Y$. The same procedure is then repeated for all lags considered. Figures 5a and 5b depict the 95%-confidence intervals, CI, associated with the estimates of EK at four representative lags. Uncertainties associated with EK estimates are (in general) negligible. Threfore, we consider the observed overall decrease of EK with the lag to be significant for both datasets. We note that $\kappa_{\Delta Y}-3 > \kappa_{Y'}-3$ at small lags for Dataset 1 (Fig. 5a), implying that frequency distributions of $\Delta Y$ exhibit sharper peaks and heavier tails than does that of $Y'$, whereas the opposite behavior is documented at large lags. Otherwise, $\kappa_{\Delta Y}-3 > \kappa_{Y'}-3$ over the whole range of lags considered for Dataset 2 (Fig. 5b). Considering the type of analyses documented in Figs. 1d-f, the behavior observed for both datasets is consistent with our theoretical models for (*i*) $0.95 < \alpha < 2$ in the case of LN-GSG; (*ii*) $a > 4$ for P-GSG, and (*iii*) $k > 1$ for Γ-GSG.

Estimates of (*i*) the shape parameter and (*ii*) the product of the scale parameters of $U$ and $G$ (henceforth denoted only as *global scale parameter* for conciseness) obtained via MOM_A and MOM_B for each GSG model formulation are depicted in Fig. 6 (Dataset 1) and Fig. 7 (Dataset 2) as a function of normalized lag. These results are complemented by Table 1 where we list parameter estimates obtained via MOM_A, together with mean and coefficient of variation (cv) evaluated over all lags of MOM_B estimates, obtained for all GSG model formulations and both datasets.

Considering Dataset 1, results obtained via MOM_B for LN-GSG (i.e., $\alpha$ in Fig. 6a and $e^{\mu}\sigma_G$ in Fig. 6d) and P-GSG (i.e., $a$ in Fig. 6b and $b\sigma_G$ in Fig. 6e) do not vary appreciably with lag (cv ≈ 2-3%), consistent with our theoretical framework. Otherwise, MOM_B estimates of $k$ and $\theta\sigma_G$ (Figs. 6c and 6f, respectively) associated with Γ-GSG are characterized by stronger oscillations around an average value, as indicated by larger values of the corresponding coefficient of variation, as compared to the other models. Nevertheless, values of cv range between 18% (for the shape parameter) and 22% (for the global scale parameter), which (also in view of ubiquitously present experimental uncertainties) can still be considered as a good approximation of the constraints associated with theoretical requirements. Figure 6 and Table 1 also document that MOM_A estimates are consistent with their counterparts obtained via MOM_B for all models.

Results for Dataset 2 (Fig. 7) obtained through MOM_B generally reveal more pronounced oscillations around a constant value and larger values of cv than those observed for Dataset 1, in particular considering the Γ-GSG model. We remark that the two considered datasets are associated with differing dimensionalities (Dataset 1 and Dataset 2 being two- and one-dimensional, respectively) and considering that $N_1 / N_2 \approx 9$, statistics of incremental data for Dataset 2 are evaluated on a much smaller sample of data as compared to Dataset 1. We regard this as the main reason related to the (slightly) increased deviations from the expected theoretical pattern.

We rely on the bootstrapping procedure mentioned above to evaluate the uncertainty associated with the GSG parameter estimates obtained via MOM_B. Figures 6-7 include depictions of the 95% CIs related to the GSG parameter estimates evaluated at four representative lags. The width of these intervals is in general very limited. The results obtained via MOM_A (see Table 1 and dashed lines in Fig. 7) tend to overestimate all parameters, as compared to their

MOM_B-based counterparts (except for $\theta\sigma_G$ in Fig. 7f), a notable discrepancy between the two estimation methods being observed for the shape parameters of P-GSG (Fig. 7b) and Γ-GSG (Fig. 7c).

Results collected in Table 1 also evidence that estimates of the shape parameter stemming from the application of each GSG model to Dataset 1 are smaller than their counterparts related to Dataset 2. This finding is indicative of a stronger non-Gaussian signature in the former data set, a behavior that can also be inferred from the increased values of excess kurtosis exhibited by Dataset 1 (see Figs. 5a and 5b).

Figures 8a and 8b depict estimates of $\rho_G$ as a function of lag obtained for Dataset 1 and 2, respectively. These results show that the correlation function of the underlying Gaussian process is quite insensitive to the choice of subordinator adopted in the GSG model, in particular considering Dataset 1. Figure 8b suggests that the width of the 95% CIs for Dataset 2 is particularly wide in the range of lags where the results associated with the three models do not overlap. This observation suggests that differences observed between $\rho_G$ estimates obtained with the three GSG models may not be particularly significant in this dataset and can be related to effects of the limited size of this sample. This result (*i*) is in agreement with the theoretical framework according to which the subordinator should be statistically independent of $G$ and (*ii*) suggests that the correlation structure provided by the underlying Gaussian process can be considered as a distinctive signature of the system.

Figure 9 depicts sample pdfs of the parent variables (Figs. 9a, 9c) and their increments (Figs. 9b, 9d) corresponding to two separation lags included in Fig. 4 and presented here for the sake of comparison against theoretical pdfs corresponding to the various GSG models considered. In these plots, $f_{Y'}$ and $f_{\Delta Y}$ associated with GSG models are evaluated respectively on the basis of (*i*) parameters estimated via MOM_A and (*ii*) the mean values of shape and global scale parameters obtained via MOM_B, $f_{\Delta Y}$ also including the lag dependent parameter, $\rho_G$, computed with MOM_B and depicted in Fig. 8. From a qualitative comparison between Figs. 9a-d and Figs. 3 and 4, it can be appreciated that all GSG models are generally in better agreement with the target sample pdfs than the Gaussian model. The degree of similarity between sample and analytical pdfs is quantified through the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951), $D_{KL}$. The latter is a measure of the information lost when a given distribution is used to approximate a target one. As such, smaller values of $D_{KL}$ are associated with reduced loss of information. Considering the pdf of $Y'$, (*i*) for Dataset 1 we obtain $D_{KL}$ = 0.048 (for LN-GSG), 0.013 (for P-GSG), and 0.071 (for Γ-GSG), thus suggesting P-GSG as the best among the models considered; (*ii*) $D_{KL} \approx 0.068$ for Dataset 2, regardless the subordinator employed. This latter outcome is consistent with Fig. 9c, where all GSG pdf are virtually overlapping. Therefore, when considering Dataset 2 the sole analysis of the parent data population does not allow discriminating between alternative GSG models. We finally evaluate $D_{KL}$ between sample and theoretical pdfs of incremental data for diverse lags. Figure 10a depicts $D_{KL}$ versus lag for Dataset 1, Fig. 10b showing a corresponding depiction for Dataset 2. These results highlight that, considering Dataset 1, the P-GSG model provides the highest degree of similarity between sample and theoretical pdfs of increments at almost all lags ($s$ > 25 $dl$), and is consistent with the results obtained for the parent variable as well as with those collected in Table 1 and Fig. 6. Considering Dataset 2, Fig. 10b suggests that the three models provide results of similar quality for lags $s$ > 200 $dz$, a feature that can also be noted from the almost overlapping analytical results depicted in Fig. 9d for $s$ = 250 $dz$. Otherwise, LN-GSG and Γ-GSG outperform P-SGS in the range 0 < $s$ < 100 $dz$. This observation, in conjunction with the analysis performed in Fig. 7 and Table 1, leads to favoring LN-GSG for Dataset 2.

Overall, our results support the ability of the GSG model to provide a theoretical interpretation of characteristic features associated with the statistics of both investigated datasets. We note that having at our disposal these tools forms the basis to achieve the overarching goal to quantify the way one can transfer the key statistics of a variable (and its increments) across scales, with direct implications on uncertainty quantification. With reference to the spatial distribution of surface roughness, these results constitute an important step to bridge across characterizations of reactive phenomena at microscopic and laboratory scales. In this context, there is documented and growing interest in the application of statistical methods (Fischer et al., 2012; Lüttge et al., 2013; Pollet-Villard et al., 2016; Trindade Pedrosa et al., 2019) to firmly ground the multiscale nature of such processes on rigorous theoretical bases. The quality of our results is encouraging to promote further studies targeting statistically-based descriptions of the temporal evolution of the surface topography of calcite minerals subject to precipitation/dissolution processes acting at diverse scales. We envision addressing this objective in the future by coupling our theoretical approach with direct in situ observations through, *e.g.*, time-lapse nanoscale imaging. In this context, characterizing porosity of natural porous media has the clear potential to link geochemical processes acting at small scales with descriptions of flow and transport at scales compatible with a continuum description of the system. Hydraulic conductivity is intimately related to porosity. As mentioned in the Introduction, statistics of its spatial increments have also been documented to display a behavior consistent with what we have observed here for porosity. These concepts have already been employed in the context of preliminary analytical and numerical studies of flow and transport in porous media associated with such a statistical description by Riva et al. (2017) and Libera et al. (2017).

## 4 Concluding remarks

We extend the Generalized Sub-Gaussian (GSG) stochastic model proposed by Riva et al. (2015a) by providing theoretical formulations of the GSG for a generic subordinator *U*. Properties of such an extended and more general model are analyzed and alternative formulations of the GSG model, derived for three selected subordinator forms, are considered to interpret observations associated with two datasets: (*i*) a set of observations characterizing the surface-roughness resulting from the dissolution of a crystal of calcite, collected in a geochemical laboratory-scale setting under given environmental conditions (Dataset 1); and (*ii*) a field-scale spatial distribution of porosity data, collected along a deep borehole within a sandstone formation (Dataset 2). Our study leads to the following key conclusions.

1. For any subordinator type associated with the GSG, the analytical formulation of standardized kurtosis, $\kappa_{Y'}$ and $\kappa_{\Delta Y}$, governing the behavior of peaks and tails of the pdf of $Y'$ and $\Delta Y$, respectively, does not depend on scale parameters of *U* and *G*. Values of $\kappa_{Y'}$ and $\kappa_{\Delta Y}$ increase as the shape parameter of *U* decreases, $\kappa_{\Delta Y}$ decreasing as the separation distance (or lag) at which increments are evaluated increases. Thus, GSG models are suitable to capturing the extensively documented peculiar features of Earth and environmental variable whose distributions transition from heavy tailed at small lags to seemingly-Gaussian at increased lags.
2. The proposed theoretical framework successfully captures the main features of the distributions of the variables analyzed as well as their spatial increments. Results of statistical analyses performed on both datasets are consistent with theoretical expectations: (*i*) estimates of shape and (global) scale parameters of the GSG models are nearly constant

with lag; (*ii*) the correlation coefficient ( $\rho_G$ ) of the underlying Gaussian process decreases as lag increases, according to a trend that is almost insensitive to the type of subordinator considered. The latter results suggest that the correlation structure provided by the Gaussian process underlying the GSG field can be considered as a distinctive signature of the system behavior.

3. The Kullback-Leibler (KL) divergence is adopted to evaluate degree of similarity between theoretical (i.e., based on the various GSG model formulations) and sample $Y'$ and $\Delta Y$ pdfs in each dataset. Our results indicate that the implementation of multiple subordinators within the GSG framework can enhance the flexibility of the model and improve the accuracy of the interpretation of statistical behavior of a given dataset.

The approach and theoretical developments we propose provide a unique framework within which one can interpret a broad range of scaling behaviors displayed by a variety of Earth and environmental variables in various settings. The successful demonstration we present imbues us with confidence about research applications targeting hydrogeological and geochemical scenarios upon leveraging on modern experimental investigation techniques leading to characterize natural systems across a diverse range of scales. These include, for example, further experiments and theoretical analyses devoted to the assessment of micro-scale reaction rates taking place at rock-liquid interfaces.

| | | **Dataset 1** | | | **Dataset 2** | | |
|---|---|---|---|---|---|---|---|
| | | LN-GSG | P-GSG | Γ-GSG | LN-GSG | P-GSG | Γ-GSG |
| **Shape parameter** | MOM_A | 1.34 | 4.14 | 2.10 | 1.86 | 9.05 | 46.12 |
| | MOM_B (mean) | 1.43 | 4.36 | 1.76 | 1.56 | 4.96 | 3.96 |
| | MOM_B (cv) | 0.02 | 0.02 | 0.18 | 0.04 | 0.06 | 0.48 |
| **Global scale parameter** | MOM_A | 0.16 | 0.16 | 0.08 | 8.18 | 7.37 | 0.18 |
| | MOM_B (mean) | 0.15 | 0.16 | 0.11 | 6.88 | 6.44 | 2.12 |
| | MOM_B (cv) | 0.02 | 0.03 | 0.22 | 0.05 | 0.02 | 0.30 |

**Table 1.** Parameter estimates obtained via MOM_A; mean and coefficient of variation (cv) evaluated over all lags of MOM_B estimates obtained for all tested GSG model formulations and both datasets.

## Appendix A: Analytical formulation of the GSG model for a general distributional form of the subordinator

The theoretical framework of the GSG model is here presented considering a general distributional form of the subordinator. We do so by deriving analytical expressions for (*i*) pdf, statistical moments and standardized kurtosis of the parent variable $Y'$ and of increments, $\Delta Y$, as a function of separation lag; and (*ii*) covariance and variogram functions as well as integral scale of $Y'$.

Substituting Eq. (4) into Eq. (3) yields

$$f_{Y_1',Y_2'}\left(y_1',y_2'\right)=\frac{1}{2\pi\sigma_G^2\sqrt{1-\rho_G^2}}\int_0^\infty\int_0^\infty f_{U_1}\left(u_1\right)f_{U_2}\left(u_2\right)e^{-\frac{1}{2\sigma_G^2\left(1-\rho_G^2\right)}\left(\frac{y_1'^2}{u_1^2}+\frac{y_2'^2}{u_2^2}-2\rho_G\frac{y_1'}{u_1}\frac{y_2'}{u_2}\right)}\frac{du_2}{u_2}\frac{du_1}{u_1}. \tag{A1}$$

The marginal pdf of $Y'$ can then been obtained from Eq. (A1) as

$$f_{Y'}\left(y'\right)=\int_{-\infty}^{\infty}f_{Y_1',Y_2'}\left(y_1',y_2'=y'\right)dy_1'=\frac{1}{\sqrt{2\pi}\sigma_G}\int_0^\infty f_U\left(u\right)e^{-\frac{1}{2\sigma_G^2}\frac{y'^2}{u^2}}\frac{du}{u}. \tag{A2}$$

All odd-order statistical moments of $Y'$ identically vanish, whereas variance, kurtosis and (in general) *q*-th even order moments can be respectively expressed as

$$\left\langle Y'^2\right\rangle=\int_{-\infty}^{+\infty}y'^2 f_{Y'}(y')dy'=\sigma_G^2\int_0^\infty u^2 f_U\left(u\right)du=\sigma_G^2\left\langle U^2\right\rangle, \tag{A3}$$

$$\left\langle Y'^4\right\rangle=\int_{-\infty}^{+\infty}y'^4 f_{Y'}(y')dy'=3\sigma_G^4\int_0^\infty u^4 f_U\left(u\right)du=3\sigma_G^4\left\langle U^4\right\rangle, \tag{A4}$$

$$\left\langle Y'^q\right\rangle=\int_{-\infty}^{+\infty}y'^q f_{Y'}(y')dy'=\frac{2^{\frac{q}{2}}}{\sqrt{\pi}}\Gamma\left(\frac{1+q}{2}\right)\sigma_G^q\int_0^\infty u^q f_U\left(u\right)du=\left\langle G^q\right\rangle\left\langle U^q\right\rangle. \tag{A5}$$

The standardized kurtosis of $Y'$ is then given by

$$\kappa_{Y'}=\frac{\left\langle Y'^4\right\rangle}{\left\langle Y'^2\right\rangle^2}=\frac{3\left\langle U^4\right\rangle}{\left\langle U^2\right\rangle^2} \tag{A6}$$

and depends only on the subordinator (and not on *G*).

The pdf of incremental values, $\Delta Y(\mathbf{s})=Y'(\mathbf{x}+\mathbf{s})-Y'(\mathbf{x})$, can be evaluated as

$$f_{\Delta Y}(\Delta y) = \int_{-\infty}^{\infty} f_{Y_1' Y_2'}(\Delta y + y_2', y_2') dy_2' = \frac{1}{\sqrt{2\pi}\sigma_G} \int_0^{\infty}\int_0^{\infty} f_{U_1}(u_1) f_{U_2}(u_2) \frac{e^{-\frac{\Delta y^2}{2\sigma_G^2 r^2}}}{r} du_2 du_1, \tag{A7}$$

with $r = \sqrt{u_1^2 + u_2^2 - 2\rho_G u_1 u_2}$. Odd-order moments of $\Delta Y$ are identically zero, whereas variance, kurtosis, and moments of even order $q$ can be respectively expressed as

$$\langle \Delta Y^2 \rangle = \sigma_G^2 \int_0^{\infty}\int_0^{\infty} r^2 f_{U_1}(u_1) f_{U_2}(u_2)\; du_2 du_1 = 2\sigma_G^2 \left[ \langle U^2 \rangle - \langle U \rangle^2 \rho_G \right], \tag{A8}$$

$$\langle \Delta Y^4 \rangle = 3\sigma_G^4 \int_0^{\infty}\int_0^{\infty} r^4 f_{U_1}(u_1) f_{U_2}(u_2)\; du_2 du_1 = 6\sigma_G^4 \left[ \langle U^4 \rangle - 4\langle U^3 \rangle \langle U \rangle \rho_G + \langle U^2 \rangle^2 \left(1 + 2\rho_G^2\right) \right], \tag{A9}$$

$$\begin{aligned} \langle \Delta Y^q \rangle &= \frac{\sigma_G^q}{\sqrt{\pi}} 2^{\frac{q}{2}} \Gamma\left(\frac{q+1}{2}\right) \int_0^{\infty}\int_0^{\infty} r^q f_{U_1}(u_1) f_{U_2}(u_2) du_2 du_1 \\ &= \sum_{k=0}^{q} (-1)^{q-k} \binom{q}{k} \langle U^k \rangle \langle U^{q-k} \rangle \langle G(x)^k G(x+s)^{q-k} \rangle. \end{aligned} \tag{A10}$$

The standardized kurtosis of $\Delta Y$ is derived from Eqs. (A8) and (A9) as

$$\kappa_{\Delta Y} = \frac{\langle \Delta Y^4 \rangle}{\langle \Delta Y^2 \rangle^2} = \frac{3}{2} \frac{\langle U^4 \rangle - 4\langle U^3 \rangle \langle U \rangle \rho_G + \langle U^2 \rangle^2 \left(1 + 2\rho_G^2\right)}{\left[ \langle U^2 \rangle - \langle U \rangle^2 \rho_G \right]^2}, \tag{A11}$$

the latter depending on the subordinator and on the correlation coefficient $\rho_G$ (but not on $\sigma_G^2$).

The Covariance of $Y'$ between two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is

$$\begin{aligned} C_{Y'}(\mathbf{x_1}, \mathbf{x_2}) &= \langle Y'(\mathbf{x_1}) Y'(\mathbf{x_2}) \rangle \\ &= \langle U(\mathbf{x_1}) U(\mathbf{x_2}) \rangle \langle G(\mathbf{x_1}) G(\mathbf{x}_2) \rangle = \langle U(\mathbf{x_1}) U(\mathbf{x_2}) \rangle \sigma_G^2 \rho_G(\mathbf{x_1}, \mathbf{x_2}). \end{aligned} \tag{A12}$$

From Eq. (A12), one derives

$$C_{Y'}(0) = \sigma_{Y'}^2 = \langle U^2 \rangle \sigma_G^2, \qquad C_{Y'}(s > 0) = \langle U \rangle^2 \sigma_G^2 \rho_G. \tag{A13}$$

Note that according to Eq. (A13) the covariance $C_{Y'}$ of the Sub-Gaussian field is discontinuous at the origin, i.e., at $s = 0$, thus exhibiting a nugget effect. The variogram of $Y'$ can be evaluated from Eq. (A8) as

$$\gamma_{Y'} = \frac{\langle \Delta Y^2 \rangle}{2} = \sigma_G^2 \left[ \langle U^2 \rangle - \langle U \rangle^2 \rho_G \right] = \sigma_G^2 \left[ \sigma_U^2 + \langle U \rangle^2 - \langle U \rangle^2 \rho_G \right] = \sigma_G^2 \sigma_U^2 + \langle U \rangle^2 \gamma_G \tag{A14}$$

and is characterized by a nugget effect, quantified by $\sigma_G^2\sigma_U^2$, $\gamma_G = \sigma_G^2\left(1-\rho_G\right)$ and $\sigma_U^2$ being the variogram of $G$ and the variance of $U$, respectively.

The integral scale of $Y'$ can be obtained by making use of Eq.(A12) as

$$I_{Y'} = \frac{\langle U\rangle^2}{\langle U^2\rangle} I_G = \frac{\langle U\rangle^2}{\sigma_U^2 + \langle U\rangle^2} I_G, \tag{A15}$$

so that one can recognize that $0 < I_{Y'} < I_G$, independent of the type of subordinator considered. An increase of $\sigma_U^2$ results in a decrease of the (integral) correlation scale of $Y'$.

## Appendix B: GSG formulation for lognormal, Pareto, and Gamma distribution of *U*

Here, we consider $U_1$ and $U_2$ to be described by (*i*) a lognormal distribution, $U_i \sim \ln N\left(\mu,(2-\alpha)^2\right)$, (*ii*) a Pareto distribution, $U_i \sim PD\left(a,b\right)$, and (*iii*) a Gamma distribution, $U_i \sim \Gamma\left(k,\theta\right)$, i.e.,

$$f_{U_i}\left(u_i\right) = \frac{e^{-\frac{\left(\ln u_i - \mu\right)^2}{2\left(2-\alpha\right)^2}}}{\sqrt{2\pi}u_i\left(2-\alpha\right)} \qquad \text{with } \alpha < 2;\ u_i \in \left(0,+\infty\right) \tag{B1a}$$

$$f_{U_i}\left(u_i\right) = \frac{ab^a}{u^{a+1}} \qquad \text{with } a > 0;\ b > 0;\ u_i \in \left[b,+\infty\right) \tag{B1b}$$

$$f_{U_i}\left(u_i\right) = \frac{u_i^{k-1}e^{-\frac{u_i}{\theta}}}{\Gamma\left(k\right)\theta^k} \qquad \text{with } k > 0;\ \theta > 0;\ u_i \in \left(0,+\infty\right);\ \Gamma\left(k\right) = \int_0^\infty x^{k-1}e^{-x}\,dx \tag{B1c}$$

here, $i$ = 1, 2; $\alpha$, $a$, and $k$ are shape parameters, while $e^\mu$, $b$, and $\theta$ are scale parameters. Note that the exponential distribution can be obtained from Eq. (B1c) by setting $k$ = 1.

The $q$-th order raw moment of $U$ is

$$\langle U^q \rangle = \begin{cases} e^{q\mu + \frac{q^2}{2}(2-\alpha)^2} & if\ U_i \sim \ln N\left(\mu,(2-\alpha)^2\right), \\ \dfrac{ab^q}{a-q} & if\ U_i \sim PD(a,b), \\ \dfrac{\Gamma(k+q)\theta^q}{\Gamma(k)} & if\ U_i \sim \Gamma(k,\theta), \end{cases} \tag{B2}$$

the variance being equal to

$$\sigma_U^2 = \begin{cases} e^{2\mu+(2-\alpha)^2}\left(e^{(2-\alpha)^2}-1\right) & if\ U_i \sim \ln N\left(\mu,(2-\alpha)^2\right), \\ ab^2/\left[(a-1)^2(a-2)\right] & if\ U_i \sim PD(a,b), \\ k\theta^2 & if\ U_i \sim \Gamma(k,\theta). \end{cases} . \tag{B3}$$

As specified in Section 2.2, the application of Method of Moment (MOM) requires $U_i$ to have finite raw moments up to order $q = 2N_P$ (thus implying $a > 4$ in (B2)).

Substituting Eq. (B1) into Eq. (A2) yields the following marginal pdf of $Y'$

$$f_{Y'}(y) = \begin{cases} \dfrac{1}{2\pi\,\sigma_G e^{\mu}(2-\alpha)} \displaystyle\int_0^\infty e^{-\frac{1}{2}\left[\left(\frac{\ln u}{2-\alpha}\right)^2 + \left(\frac{y}{\sigma_G e^{\mu} u}\right)^2\right]} \dfrac{du}{u^2} & \text{for LN-GSG}, \\ \dfrac{a}{\sqrt{2\pi}\left(b\sigma_G\right)^{-a}} \displaystyle\int_{b\sigma_G}^\infty e^{-\frac{1}{2}\left(\frac{y}{u}\right)^2} \dfrac{du}{u^{a+2}} & \text{for P-GSG}, \\ \dfrac{1}{\sqrt{2\pi}\,\sigma_G\theta\,\Gamma(k)} \displaystyle\int_0^\infty u^{k-2} e^{-\left[u+\frac{y^2}{2\sigma_G^2\theta^2u^2}\right]} du & \text{for Γ-GSG}. \end{cases} . \tag{B4}$$

Note that LN-GSG coincides with a normal-lognormal distribution (NLN) when $\mu = 0$. The latter has been shown to well represent some financial (Clark, 1973) and environmental (Guadagnini et al., 2015) data. Making use of Eqs. (A3) - (A6) and (B2), variance, kurtosis and standardized kurtosis of $Y'$ are respectively given by

$$\langle Y'^2 \rangle = \sigma_G^2 \begin{cases} e^{2\mu} e^{2(2-\alpha)^2} & \text{for LN-GSG}, \\ b^2 \dfrac{a}{a-2} & \text{for P-GSG}, \\ \theta^2 k(1+k) & \text{for Γ-GSG}. \end{cases} \tag{B5}$$

$$\left\langle Y'^4 \right\rangle = 3\sigma_G^4 \begin{cases} e^{4\mu} e^{8(2-\alpha)^2} & \text{for LN-GSG,} \\ b^4 \dfrac{a}{a-4} & \text{for P-GSG,} \\ \theta^4 \prod_{i=0}^{3} (k+i) & \text{for } \Gamma\text{-GSG,} \end{cases} \tag{B6}$$

$$\kappa_{Y'} = 3 \begin{cases} e^{4(2-\alpha)^2} & \text{for LN-GSG,} \\ \dfrac{(a-2)^2}{a(a-4)} & \text{for P-GSG,} \\ 1 + \dfrac{4k+6}{k(k+1)} & \text{for } \Gamma\text{-GSG.} \end{cases} \tag{B7}$$

Substituting Eq. (B1) into Eq. (A7) yields the following expressions for the pdf of $\Delta Y$

$$f_{\Delta Y}(\Delta y) = \frac{1}{\sqrt{2}} \begin{cases} \dfrac{\sqrt{\pi}}{2\pi^2 (2-\alpha)^2 \sigma_G e^{\mu}} \displaystyle\int_0^{\infty}\int_0^{\infty} e^{-\frac{1}{2}\left[\frac{1}{(2-\alpha)^2}\left(\ln^2 u_1 + \ln^2 u_2\right) + \frac{\Delta y^2}{\sigma_G^2 e^{2\mu} r^2}\right]} \dfrac{du_2 du_1}{u_2 u_1 r} & \text{for LN-GSG,} \\ \dfrac{a^2}{\sqrt{\pi}\left(b\sigma_G\right)^{-2a}} \displaystyle\int_{\sigma_G b}^{\infty}\int_{\sigma_G b}^{\infty} e^{-\frac{1}{2}\frac{\Delta y^2}{r^2}} \dfrac{du_2 du_1}{u_1^{a+1} u_2^{a+1} r} & \text{for P-GSG,} \\ \dfrac{1}{\sqrt{\pi}\left(\sigma_G \theta\right)^{2k} \Gamma^2(k)} \displaystyle\int_0^{\infty}\int_0^{\infty} \left(u_1 u_2\right)^{k-1} e^{-\left[\frac{1}{\sigma_G \theta}(u_1+u_2) + \frac{1}{r^2}\frac{\Delta y^2}{2}\right]} \dfrac{du_2 du_1}{r} & \text{for } \Gamma\text{-GSG,} \end{cases} \tag{B8}$$

with $r = \sqrt{u_1^{\,2} + u_2^{\,2} - 2\rho_G u_1 u_2}$ .

Making use of Eqs. (A8)-(A11) and (B2), variance, kurtosis and standardized kurtosis of $\Delta Y$ are respectively given by

$$\left\langle \Delta Y^2 \right\rangle = 2\sigma_G^2 \begin{cases} e^{2\mu} e^{(2-\alpha)^2} \left[ e^{(2-\alpha)^2} - \rho_G \right] & \text{for LN-GSG,} \\ b^2 a \left[ \dfrac{1}{a-2} - \dfrac{a\rho_G}{(a-1)^2} \right] & \text{for P-GSG,} \\ k\theta^2 \left[ 1 + k - \rho_G k \right] & \text{for } \Gamma\text{-GSG,} \end{cases} \tag{B9}$$

$$\left\langle \Delta Y^4 \right\rangle = 6\sigma_G^4 \begin{cases} e^{4\mu} e^{4(2-\alpha)^2}\left[1+e^{4(2-\alpha)^2}-4e^{(2-\alpha)^2}\rho_G+2\rho_G^2\right] & \text{for LN-GSG,} \\ b^4 a^2 \left[\dfrac{1}{a(a-4)}-\dfrac{4\rho_G}{(a-1)(a-3)}+\dfrac{\left(1+2\rho_G^2\right)}{(a-2)^2}\right] & \text{for P-GSG,} \\ 2\theta^4 k^2 (k+1)\left[\dfrac{3}{k}+k+3+\rho_G\left(\rho_G(k+1)-2(k+2)\right)\right] & \text{for } \Gamma\text{-GSG,} \end{cases} \tag{B10}$$

$$\kappa_{\Delta Y} = 3\begin{cases} e^{2(2-\alpha)^2}\left\{1+\dfrac{1}{2}\left(\dfrac{e^{2(2-\alpha)^2}-1}{e^{(2-\alpha)^2}-\rho_G}\right)^2\right\} & \text{for LN-GSG,} \\ \dfrac{1}{2}\left[\dfrac{1}{a(a-4)}-\dfrac{4\rho_G}{(a-1)(a-3)}+\dfrac{1+2\rho_G^2}{(a-2)^2}\right]\left[\dfrac{1}{(a-2)}-\dfrac{a\rho_G}{(a-1)^2}\right]^{-2} & \text{for P-GSG,} \\ 1+\dfrac{1}{k}+\dfrac{(k+1)\left(2+k+\rho_G^2 k-2k\rho_G\right)}{k\left(k+1-\rho_G k\right)^2} & \text{for } \Gamma\text{-GSG.} \end{cases} \tag{B11}$$

The variogram, $\gamma_{Y'}$, covariance, $C_{Y'}(s)$, for $s>0$ (note that $C_{Y'}(s=0)$ coincides with $\left\langle Y'^2 \right\rangle$ evaluated in Eq. (B5)) and integral scale, $I_{Y'}$, of $Y'$, can be derived from Eqs. (A13) – (A15) and (B2) as

$$\gamma_{Y'} = \begin{cases} \sigma_G^2 e^{2\mu+(2-\alpha)^2}\left(e^{(2-\alpha)^2}-1\right)+e^{2\mu+(2-\alpha)^2}\gamma_G & \text{for LN-GSG,} \\ \sigma_G^2 \dfrac{b^2 a}{(a-1)^2(a-2)}+\dfrac{a^2 b^2}{(a-1)^2}\gamma_G & \text{for P-GSG,} \\ \sigma_G^2 k\theta^2 + k^2\theta^2\gamma_G & \text{for } \Gamma\text{-GSG,} \end{cases} , \tag{B12}$$

$$C_{Y'}(s>0) = \sigma_G^2\rho_G \begin{cases} e^{2\mu+(2-\alpha)^2} & \text{for LN-GSG,} \\ \dfrac{a^2 b^2}{(a-1)^2} & \text{for P-GSG,} \\ k^2\theta^2 & \text{for } \Gamma\text{-GSG,} \end{cases} \tag{B13}$$

$$I_{Y'} = I_G \begin{cases} e^{-(2-\alpha)^2} & \text{for LN-GSG,} \\ a\dfrac{a-2}{(a-1)^2} & \text{for P-GSG,} \\ 1-\dfrac{1}{1+k} & \text{for } \Gamma\text{-GSG.} \end{cases} \quad . \tag{B14}$$

It is thus seen that when the pdf of $U$ tends to the Dirac delta function (i.e., when $\alpha \to 2$ for LN-GSG; $a \to \infty$ for P-GSG; or $k \to \infty$ for $\Gamma$-GSG), then $I_{Y'} \to I_G$. Otherwise, $I_{Y'}$ is smaller than $I_G$ (regardless the subordinator adopted), while never vanishing. The range of values which can be undertaken by $I_{Y'}$ depends on the type of subordinator employed and on the threshold values of the shape parameters (see Section 2.1). The broadest range of variability of $I_{Y'}$ is associated with the LN-GSG, where $0.33 < I_{Y'} / I_G < 1$. Otherwise, the smallest interval is obtained through P-GSG, where $0.89 < I_{Y'} / I_G < 1$, $\Gamma$-GSG being associated with $0.5 < I_{Y'} / I_G < 1$.

## Data

Datasets are available at: https://data.mendeley.com/datasets/trdgwfwsvn/draft?a=ee55e214-386a-48f4-88bf-461c5ddaf7ec

## References

Boffetta, G., Mazzino, A., & Vulpiani, A. (2008). Twenty-five years of multifractals in fully developed turbulence: a tribute to Giovanni Paladin. Journal of Physics A: Mathematical and Theoretical, 41(36), 363001. https://doi.org/10.1088/1751-8113/41/36/363001

Bouissonnié, A., Daval, D., Marinoni, M., & Ackerer, P. (2018). From mixed flow reactor to column experiments and modeling: Upscaling of calcite dissolution rate. Chemical Geology, 487, 63–75. https://doi.org/10.1016/j.chemgeo.2018.04.017

Clark, P. K. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica, 41(1), 135–155. https://doi.org/10.2307/1913889

Dashtian, H., Jafari, G. R., Sahimi, M., & Masihi, M. (2011). Scaling, multifractality, and long-range correlations in well log data of large-scale porous media. Physica A: Statistical

Mechanics and Its Applications, 390(11), 2096–2111. https://doi.org/10.1016/j.physa.2011.01.010

Desbarats, A. J. (1990). Macrodispersion in sand-shale sequences. Water Resources Research, 26(1), 153–163. https://doi.org/10.1029/WR026i001p00153

Di Federico, V., & Neuman, S. P. (1997). Scaling of random fields by means of truncated power variograms and associated spectra. Water Resources Research, 33(5), 1075–1085. https://doi.org/10.1029/97WR00299

Efron, B. (1992). Bootstrap Methods: Another Look at the Jackknife. Breakthroughs in Statistics, Springer, pp. 569-593.

Fischer, C., Arvidson, R. S., & Lüttge, A. (2012). How predictable are dissolution rates of crystalline material? Geochimica et Cosmochimica Acta, 98, 177–185. https://doi.org/10.1016/j.gca.2012.09.011

Frisch, U. (2016). The collective birth of multifractals. Journal of Physics A: Mathematical and Theoretical, 49(45), 451002. https://doi.org/10.1088/1751-8113/49/45/451002

Ganti, V., Singh, A., Passalacqua, P., & Foufoula-Georgiou, E. (2009). Subordinated Brownian motion model for sediment transport. Physical Review E, 80(1), 011111. https://doi.org/10.1103/PhysRevE.80.011111

Gómez-Hernández, J. J., & Wen, X.-H. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. Advances in Water Resources, 21(1), 47–61. https://doi.org/10.1016/S0309-1708(96)00031-0

Guadagnini, A., Neuman, S. P., Nan, T., Riva, M., & Winter, C. L. (2015). Scalable statistics of correlated random variables and extremes applied to deep borehole porosities. Hydrology and Earth System Sciences, 19(2), 729–745. https://doi.org/10.5194/hess-19-729-2015

Guadagnini, Alberto, Riva, M., & Neuman, S. P. (2018). Recent advances in scalable non-Gaussian geostatistics: The generalized sub-Gaussian model. Journal of Hydrology, 562, 685–691. https://doi.org/10.1016/j.jhydrol.2018.05.001

Haslauer, C. P., Guthke, P., Bárdossy, A., & Sudicky, E. A. (2012). Effects of non-Gaussian copula-based hydraulic conductivity fields on macrodispersion. Water Resources Research, 48(7). https://doi.org/10.1029/2011WR011425

Jordan, G., & Rammensee, W. (1998). Dissolution Rates of Calcite (104) Obtained by Scanning Force Microscopy: Microtopography-Based Dissolution Kinetics on Surfaces with

Anisotropic Step Velocities. Geochimica et Cosmochimica Acta, 62(6), 941–947. https://doi.org/10.1016/S0016-7037(98)00030-1

Kozubowski, T. J., Meerschaert, M. M., & Podgórski, K. (2006). Fractional Laplace motion. Advances in Applied Probability, 38(2), 451–464. https://doi.org/10.1239/aap/1151337079

Kozubowski, Tomasz J., Podgórski, K., & Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields. Journal of Multivariate Analysis, 113, 59–72. https://doi.org/10.1016/j.jmva.2012.02.010

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Libera, A., de Barros, F. P. J., Riva, M., & Guadagnini, A. (2017). Solute concentration at a well in non-Gaussian aquifers under constant and time-varying pumping schedule. Journal of Contaminant Hydrology, 205, 37–46. https://doi.org/10.1016/j.jconhyd.2017.08.006

Liu, H. H., & Molz, F. J. (1997). Comment on "Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations" by Scott Painter. Water Resources Research, 33(4), 907–908. https://doi.org/10.1029/96WR03788

Lovejoy, S., & Schertzer, D. (1995). Multifractals and rain. In New Uncertainty Concepts in Hydrology and Water Resources (pp. 61–103). Cambridge University Press. https://doi.org/10.1017/CBO9780511564482.009

Lu, Z., & Zhang, D. (2002). On stochastic modeling of flow in multimodal heterogeneous formations. Water Resources Research, 38(10), 8-1-8–15. https://doi.org/10.1029/2001WR001026

Lüttge, A., Arvidson, R. S., & Fischer, C. (2013). A stochastic treatment of crystal dissolution kinetics. Elements, 9, 183–188. https://doi.org/10.2113/gselements.9.3.183

Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. Journal of Fluid Mechanics, 62(2), 331–358. https://doi.org/10.1017/S0022112074000711

Mariethoz, G., Renard, P., & Straubhaar, J. (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46(11). https://doi.org/10.1029/2008WR007621

Meerschaert, M. M. (2004). Fractional Laplace model for hydraulic conductivity. Geophysical Research Letters, 31(8), L08501. https://doi.org/10.1029/2003GL019320

Monin, A. S., & Yaglom, A. M. (1975). Statistical Fluid Mechanics. Volume 2. Mechanics of Turbulence. Cambridge, Massachusetts: MIT Press.

Neuman, S. P., & Di Federico, V. (2003). Multifaceted nature of hydrogeologic scaling and its interpretation. Reviews of Geophysics, 41(3), 1014. https://doi.org/10.1029/2003RG000130

Neuman, S. P., Guadagnini, A., Riva, M., & Siena, M. (2013). Recent Advances in Statistical and Scaling Analysis of Earth and Environmental Variables. In Advances in Hydrogeology (pp. 1–25). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-6479-2_1

Noiriel, C., Luquot, L., Madé, B., Raimbault, L., Gouze, P., & van der Lee, J. (2009). Changes in reactive surface area during limestone dissolution: An experimental and modelling study. Chemical Geology, 265(1–2), 160–170. https://doi.org/10.1016/j.chemgeo.2009.01.032

Noiriel, C., Oursin, M., & Daval, D. (2020). Examination of crystal dissolution in 3D: A way to reconcile dissolution rates in the laboratory? Geochimica et Cosmochimica Acta, 273, 1–25. https://doi.org/10.1016/j.gca.2020.01.003

Painter, S. (1996). Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations. Water Resources Research. https://doi.org/10.1029/96WR00286

Painter, S. (2001). Flexible scaling model for use in random field simulation of hydraulic conductivity. Water Resources Research, 37(5), 1155–1163. https://doi.org/10.1029/2000WR900394

Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. P. (2016). Theory and generation of conditional, scalable sub-Gaussian random fields. Water Resources Research, 52(3), 1746–1761. https://doi.org/10.1002/2015WR018348

Pollet-Villard, M., Daval, D., Ackerer, P., Saldi, G. D., Wild, B., Knauss, K. G., & Fritz, B. (2016a). Does crystallographic anisotropy prevent the conventional treatment of aqueous mineral reactivity? A case study based on K-feldspar dissolution kinetics. Geochimica et Cosmochimica Acta, 190, 294–308. https://doi.org/10.1016/j.gca.2016.07.007

Pollet-Villard, M., Daval, D., Fritz, B., Knauss, K. G., Schäfer, G., & Ackerer, P. (2016b). Influence of etch pit development on the surface area and dissolution kinetics of the orthoclase (001) surface. Chemical Geology, 447, 79–92. https://doi.org/10.1016/j.chemgeo.2016.09.038

Riva, M., Neuman, S. P., Guadagnini, A., & Siena, M. (2013a). Anisotropic Scaling of Berea Sandstone Log Air Permeability Statistics. Vadose Zone Journal, 12(3), vzj2012.0153. https://doi.org/10.2136/vzj2012.0153

Riva, M., Neuman, S. P., & Guadagnini, A. (2013b). Sub-Gaussian model of processes with heavy-tailed distributions applied to air permeabilities of fractured tuff. Stochastic Environmental Research and Risk Assessment, 27(1), 195–207. https://doi.org/10.1007/s00477-012-0576-y

Riva, M., Neuman, S. P., & Guadagnini, A. (2015a). New scaling model for variables and increments with heavy-tailed distributions. Water Resources Research, 51(6), 4623–4634. https://doi.org/10.1002/2015WR016998

Riva, M., Panzeri, M., Guadagnini, A., & Neuman, S. P. (2015b). Simulation and analysis of scalable non-Gaussian statistically anisotropic random functions. Journal of Hydrology, 531, 88–95. https://doi.org/10.1016/j.jhydrol.2015.06.066

Riva, M., Guadagnini, A., & Neuman, S. P. (2017). Theoretical analysis of non-Gaussian heterogeneity effects on subsurface flow and transport. Water Resources Research, 53(4), 2998–3012. https://doi.org/10.1002/2016WR019353

Rubin, Y. (1995). Flow and Transport in Bimodal Heterogeneous Formations. Water Resources Research, 31(10), 2461–2468. https://doi.org/10.1029/95WR01953

Russo, D. (2002). Stochastic analysis of macrodispersion in gravity-dominated flow through bimodal heterogeneous unsaturated formations. Water Resources Research, 38(7), 19-1-19–14. https://doi.org/10.1029/2001WR000850

Russo, D. (2010). First-order and numerical analyses of flow and transport in heterogeneous bimodal variably saturated formations. Water Resources Research, 46(6). https://doi.org/10.1029/2009WR008307

Siena, M., Guadagnini, A., Riva, M., & Neuman, S. P. (2012). Extended power-law scaling of air permeabilities measured on a block of tuff. Hydrology and Earth System Sciences, 16(1). https://doi.org/10.5194/hess-16-29-2012

Siena, M., Riva, M., Giamberini, M., Gouze, P., & Guadagnini, A. (2019). Statistical modeling of gas-permeability spatial variability along a limestone core. Spatial Statistics, 34, 100249. https://doi.org/10.1016/j.spasta.2017.07.007

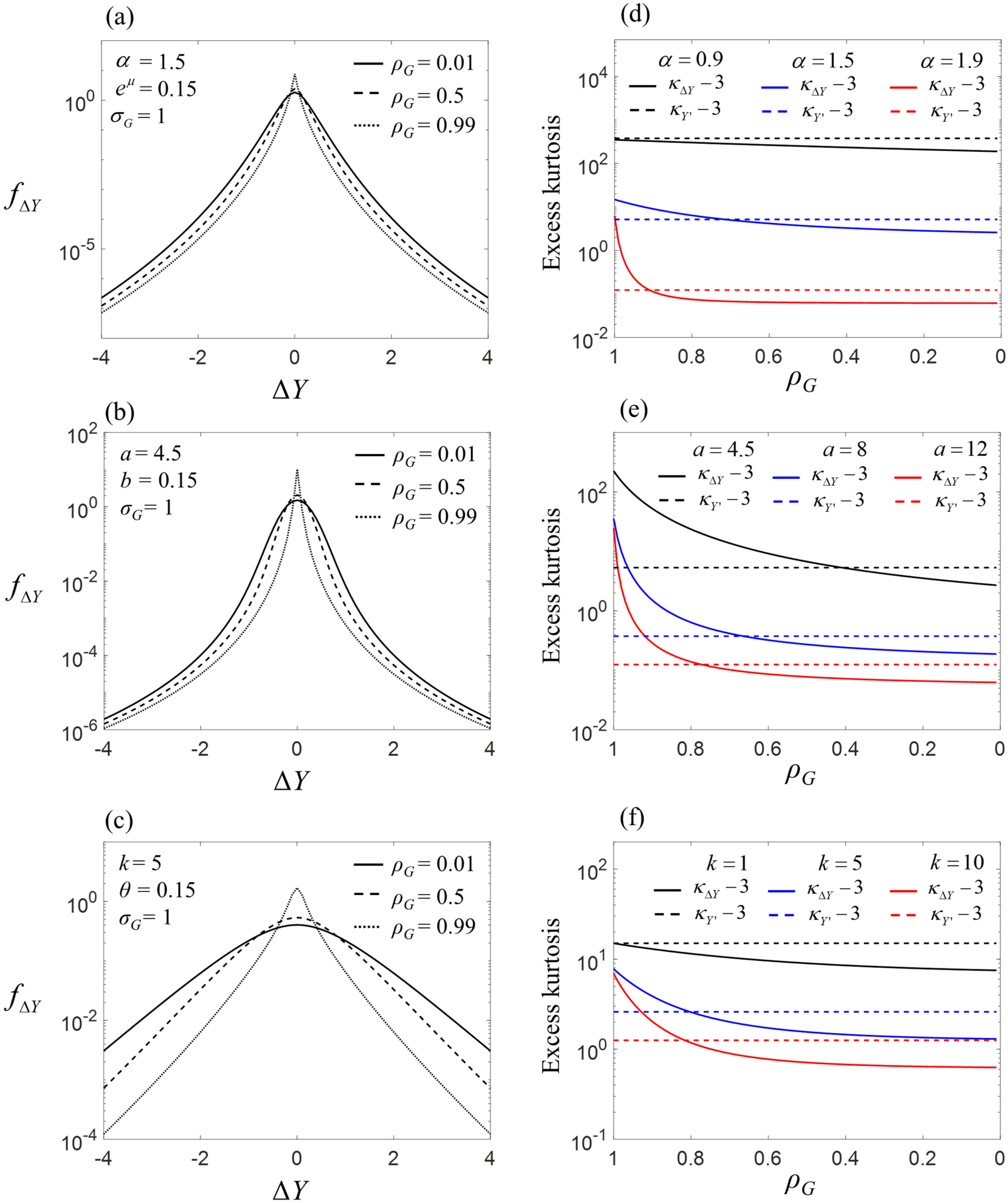
Trindade Pedrosa, E., Kurganskaya, I., Fischer, C., & Luttge, A. (2019) A Statistical Approach for Analysis of Dissolution Rates Including Surface Morphology. Minerals, 9, 458. https://doi.org/10.3390/min9080458

Veneziano, D., Langousis, A., & Furcolo, P. (2006). Multifractality and rainfall extremes: A review. Water Resources Research, 42(6). https://doi.org/10.1029/2005WR004716

Winter, C. L., Tartakovsky, D. M., & Guadagnini, A. (2003). Moment differential equations for flow in highly heterogeneous porous media. Surveys in Geophysics, 24, 81–106. https://doi.org/10.1023/A:1022277418570

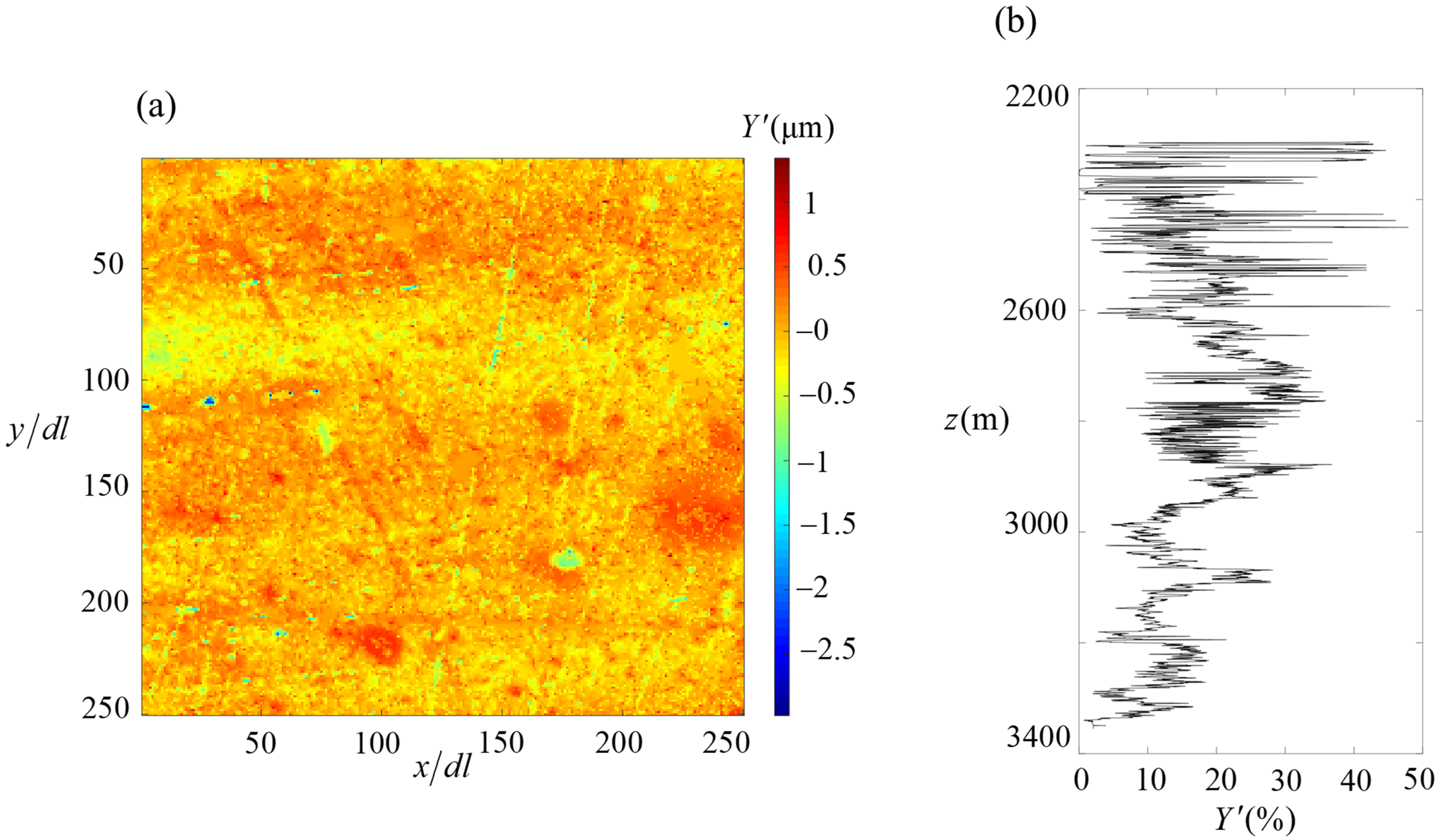
Xu, T., & Gómez-Hernández, J. J. (2015). Inverse sequential simulation: A new approach for the characterization of hydraulic conductivities demonstrated on a non- <scp>G</scp> aussian field. Water Resources Research, 51(4), 2227–2242. https://doi.org/10.1002/2014WR016320

**Figure 1.**



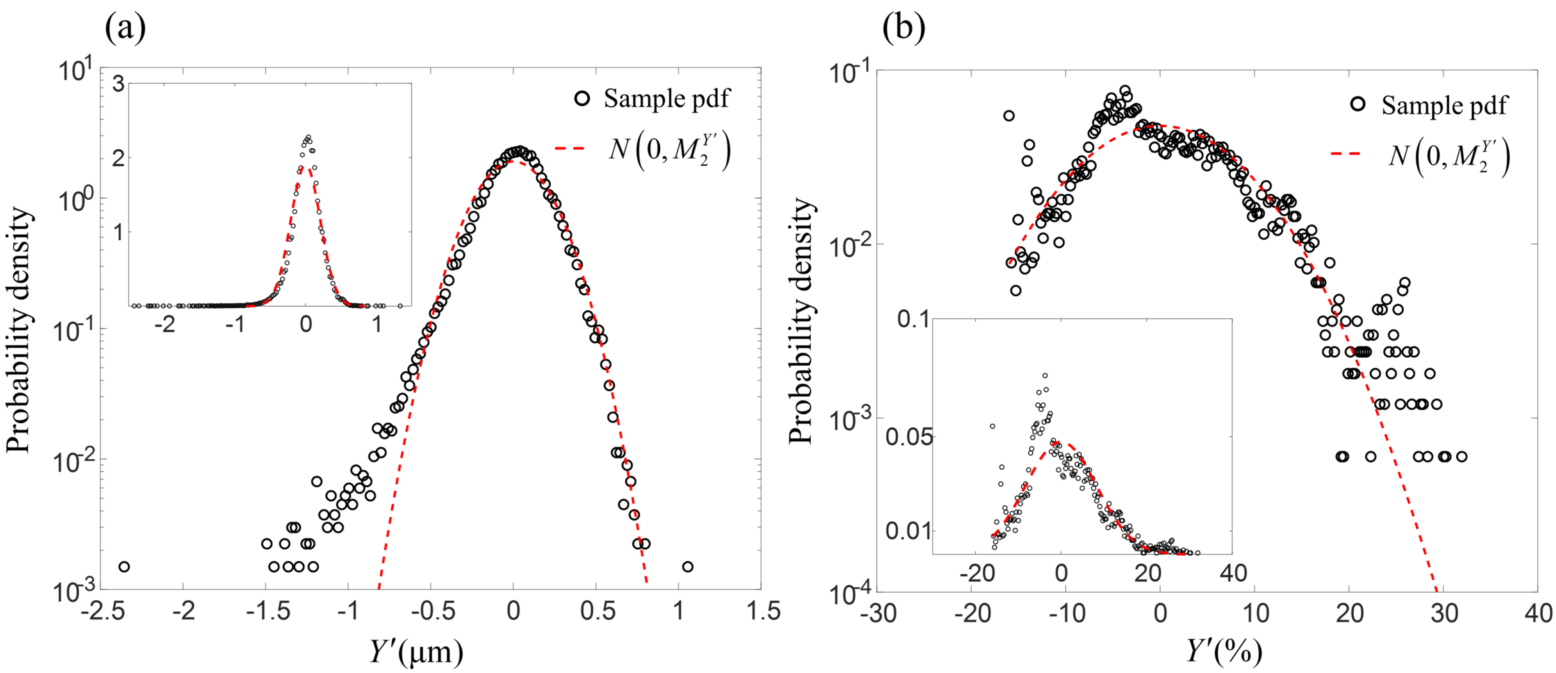
(a)
$\alpha = 1.5$
$e^{\mu} = 0.15$
$\sigma_G = 1$
$\rho_G = 0.01$
$\rho_G = 0.5$
$\rho_G = 0.99$
$f_{\Delta Y}$
$\Delta Y$
(b)
$a = 4.5$
$b = 0.15$
$\sigma_G = 1$
$\rho_G = 0.01$
$\rho_G = 0.5$
$\rho_G = 0.99$
$f_{\Delta Y}$
$\Delta Y$
(c)
$k = 5$
$\theta = 0.15$
$\sigma_G = 1$
$\rho_G = 0.01$
$\rho_G = 0.5$
$\rho_G = 0.99$
$f_{\Delta Y}$
$\Delta Y$
(d)
$\alpha = 0.9$
$\alpha = 1.5$
$\alpha = 1.9$
$\kappa_{\Delta Y} - 3$
$\kappa_{Y'} - 3$
Excess kurtosis
$\rho_G$
(e)
$a = 4.5$
$a = 8$
$a = 12$
$\kappa_{\Delta Y} - 3$
$\kappa_{Y'} - 3$
Excess kurtosis
$\rho_G$
(f)
$k = 1$
$k = 5$
$k = 10$
$\kappa_{\Delta Y} - 3$
$\kappa_{Y'} - 3$
Excess kurtosis
$\rho_G$

**Figure 2.**



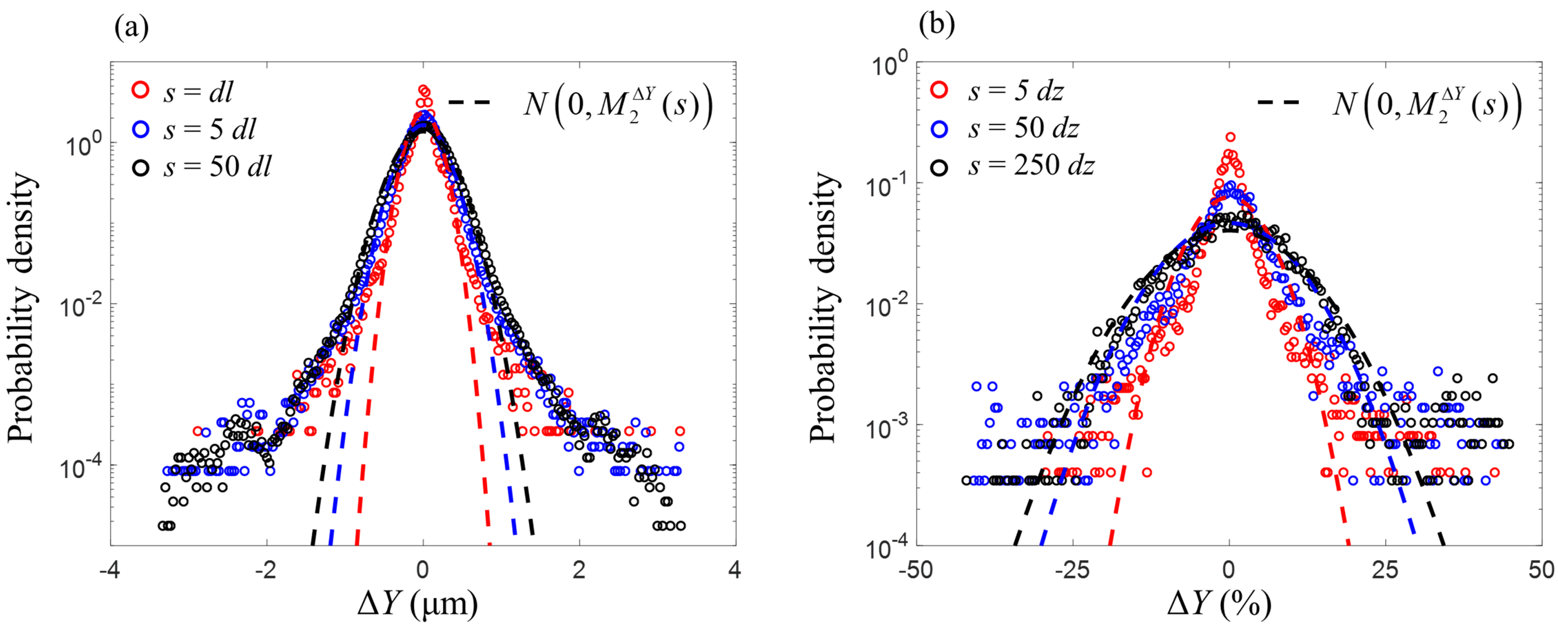
(a)
Y′(μm)
1
0.5
−0
−0.5
−1
−1.5
−2
−2.5
50
100
150
200
250
y/dl
50
100
150
200
250
x/dl
(b)
2200
2600
3000
3400
z(m)
0
10
20
30
40
50
Y′(%)

**Figure 3.**



(a)
Sample pdf
$N\left(0, M_2^{Y'}\right)$
Probability density
$Y'$(μm)
(b)
Sample pdf
$N\left(0, M_2^{Y'}\right)$
Probability density
$Y'$(%)

**Figure 4.**



(a)
$s = dl$
$s = 5\ dl$
$s = 50\ dl$
$N\left(0, M_2^{\Delta Y}(s)\right)$
Probability density
$10^{0}$
$10^{-2}$
$10^{-4}$
-4
-2
0
2
4
$\Delta Y$ (μm)
(b)
$s = 5\ dz$
$s = 50\ dz$
$s = 250\ dz$
$N\left(0, M_2^{\Delta Y}(s)\right)$
Probability density
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
-50
-25
0
25
50
$\Delta Y$ (%)

**Figure 5.**

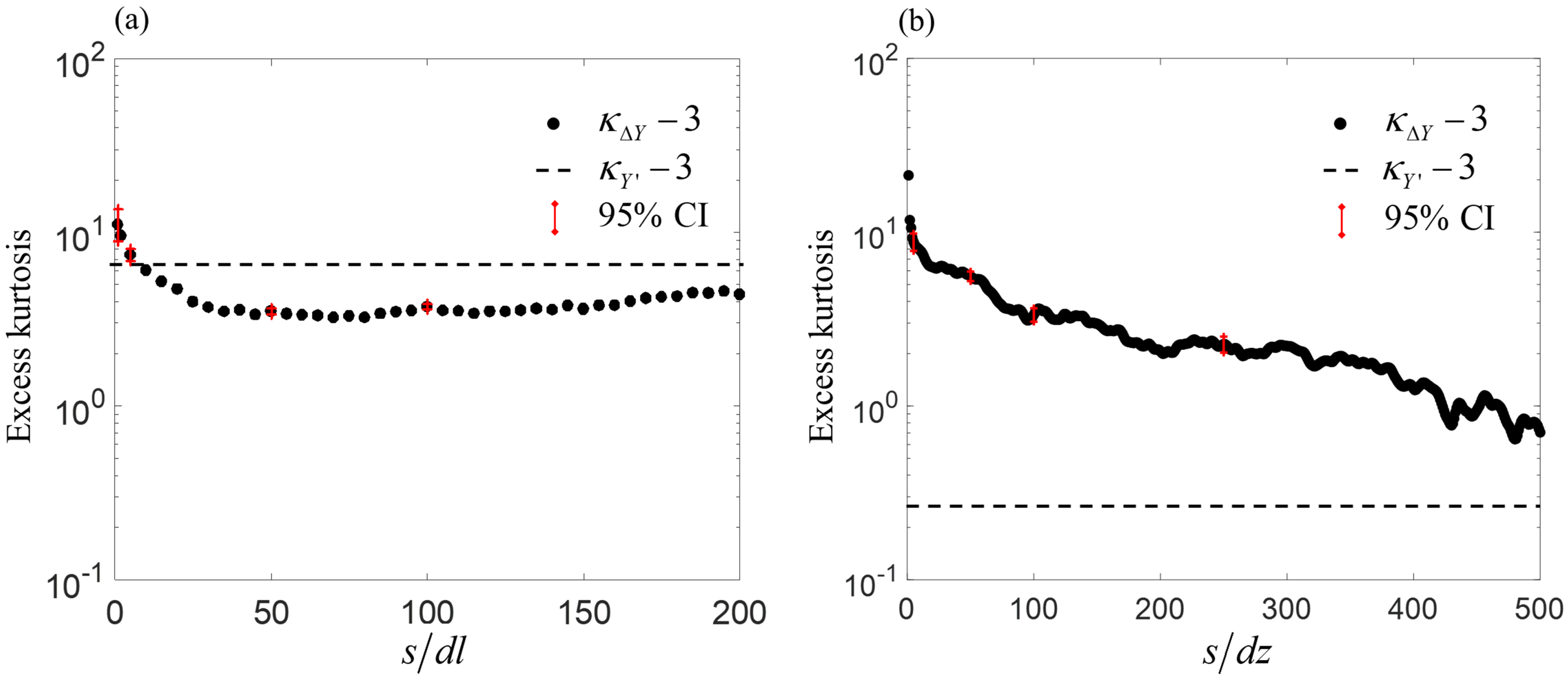
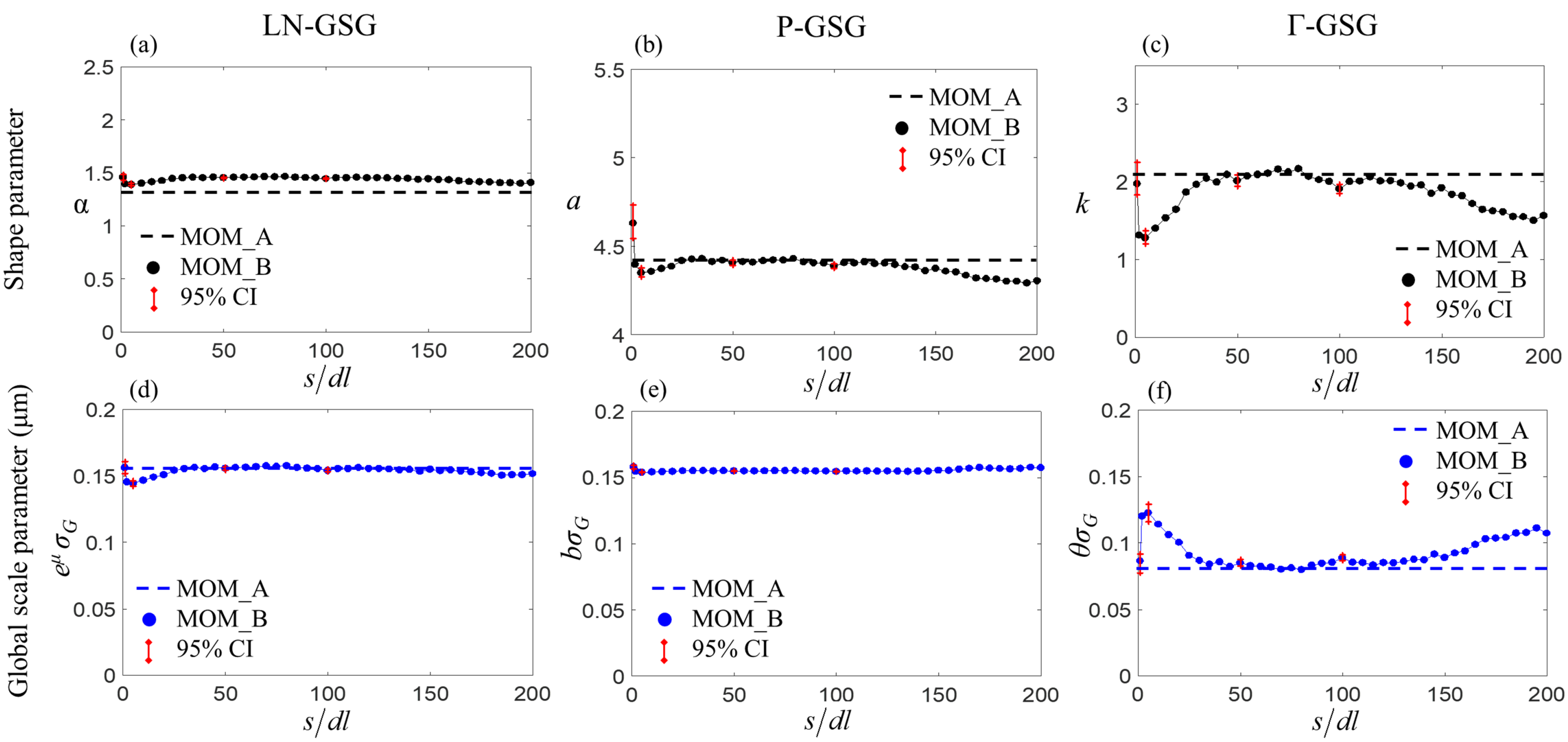
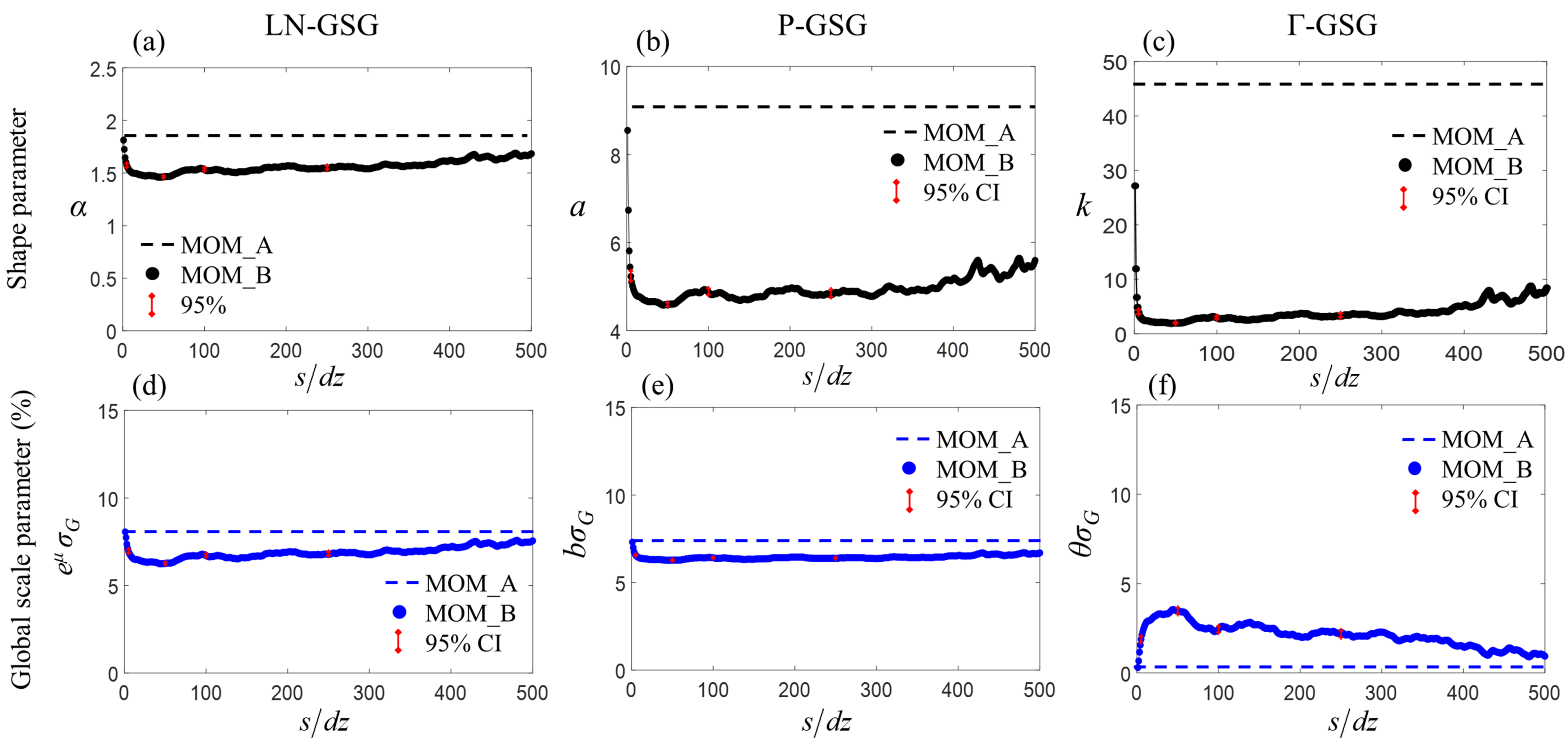

(a)
(b)
Excess kurtosis
$\kappa_{\Delta Y}-3$
$\kappa_{Y'}-3$
95% CI
$s/dl$
$s/dz$

**Figure 6.**



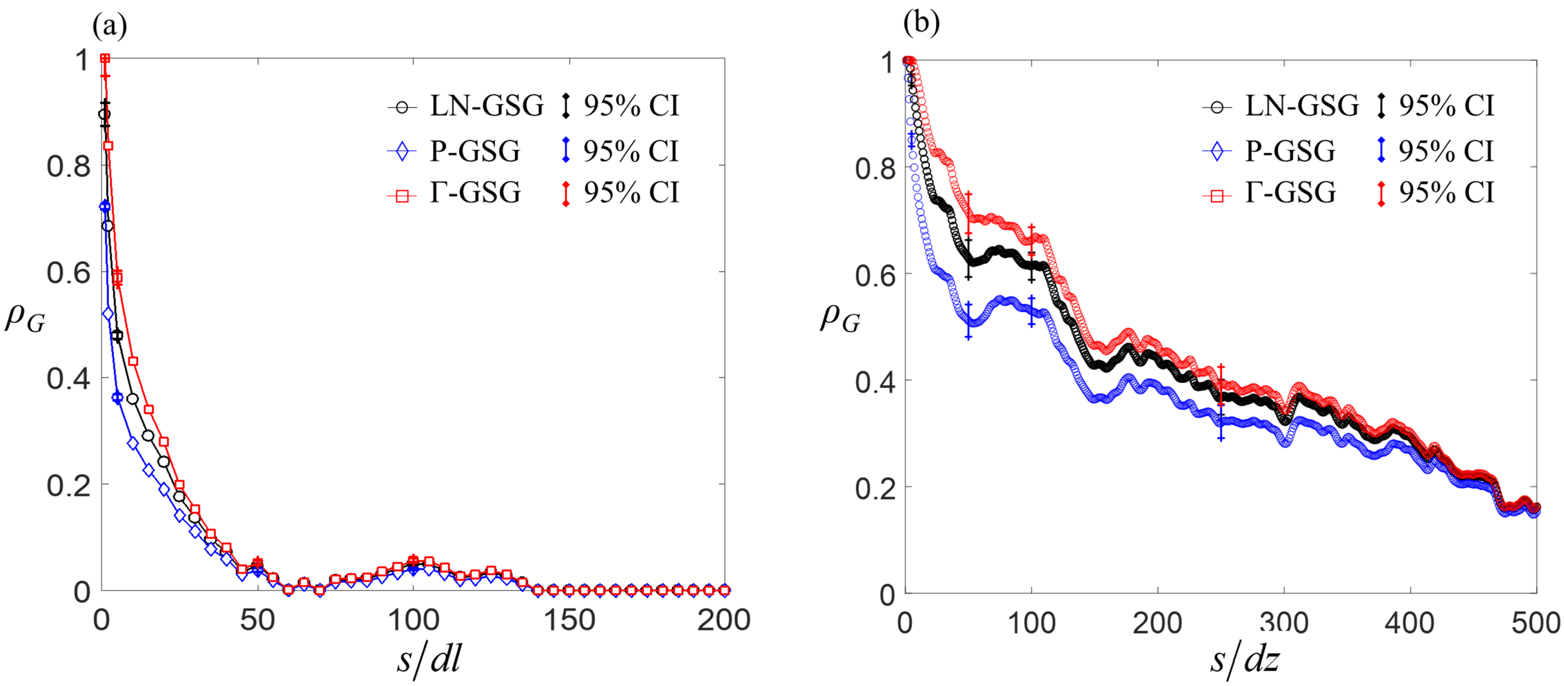
LN-GSG
P-GSG
Γ-GSG
(a)
(b)
(c)
(d)
(e)
(f)
Shape parameter
Global scale parameter (μm)
α
a
k
$e^{\mu}\sigma_G$
$b\sigma_G$
$\theta\sigma_G$
s/dl
MOM_A
MOM_B
95% CI

**Figure 7.**



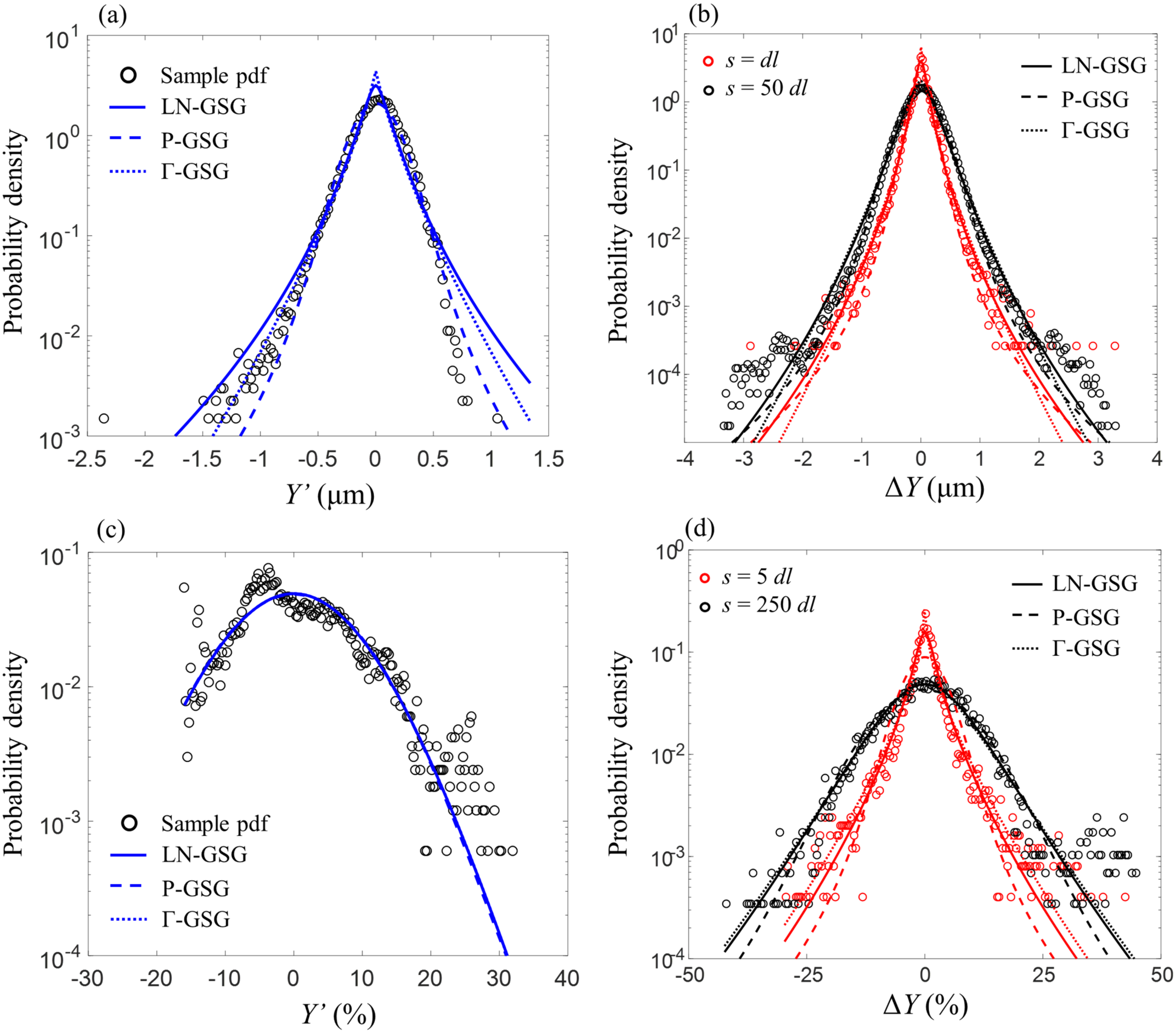
LN-GSG
P-GSG
Γ-GSG
(a)
(b)
(c)
(d)
(e)
(f)
Shape parameter
Global scale parameter (%)
α
a
k
$e^{\mu}\sigma_G$
$b\sigma_G$
$\theta\sigma_G$
s/dz
MOM_A
MOM_B
95%
95% CI

**Figure 8.**



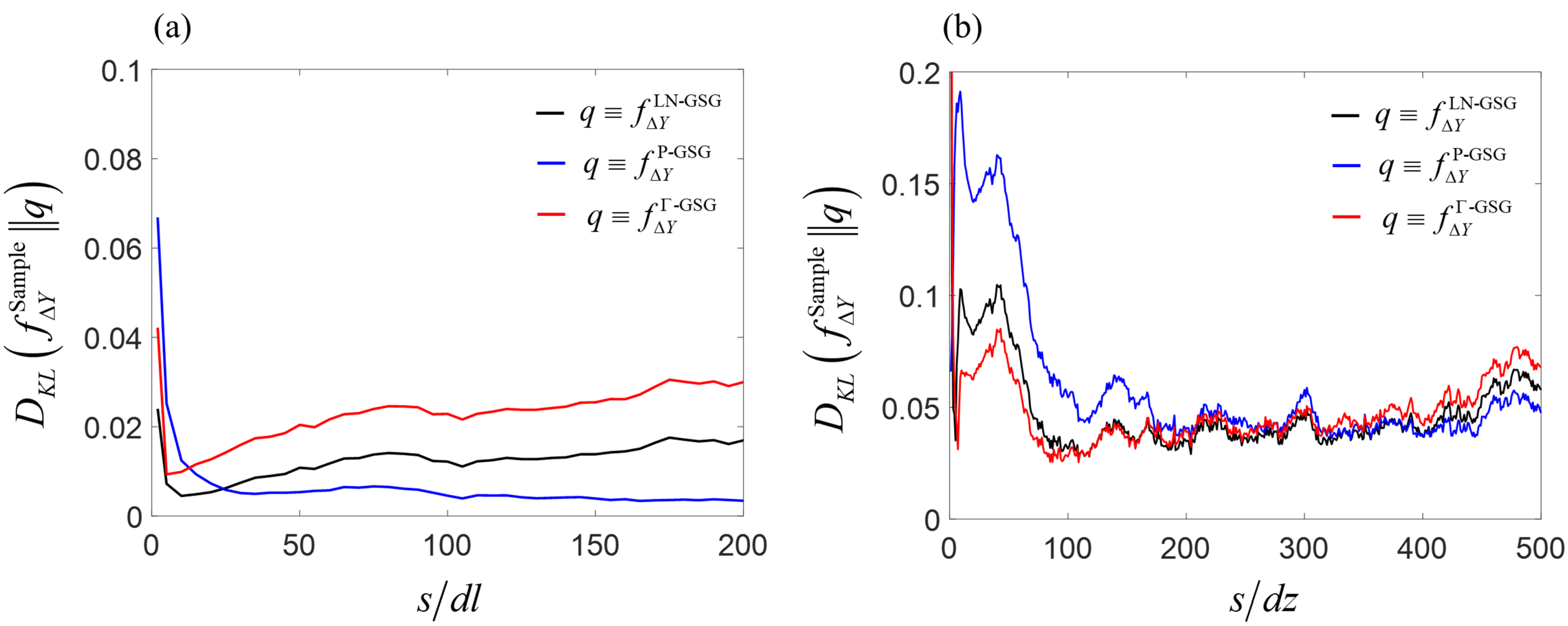
(a)
LN-GSG
95% CI
P-GSG
95% CI
Γ-GSG
95% CI
$\rho_G$
$s/dl$
0
50
100
150
200
0
0.2
0.4
0.6
0.8
1
(b)
LN-GSG
95% CI
P-GSG
95% CI
Γ-GSG
95% CI
$\rho_G$
$s/dz$
0
100
200
300
400
500

**Figure 9.**



(a)
Sample pdf
LN-GSG
P-GSG
Γ-GSG
Probability density
Y' (μm)
(b)
s = dl
s = 50 dl
LN-GSG
P-GSG
Γ-GSG
Probability density
ΔY (μm)
(c)
Sample pdf
LN-GSG
P-GSG
Γ-GSG
Probability density
Y' (%)
(d)
s = 5 dl
s = 250 dl
LN-GSG
P-GSG
Γ-GSG
Probability density
ΔY (%)

**Figure 10.**


(a)
(b)
$q \equiv f_{\Delta Y}^{\text{LN-GSG}}$
$q \equiv f_{\Delta Y}^{\text{P-GSG}}$
$q \equiv f_{\Delta Y}^{\Gamma\text{-GSG}}$
$D_{KL}\left(f_{\Delta Y}^{\text{Sample}} \| q\right)$
$s/dl$
$s/dz$
0.1
0.08
0.06
0.04
0.02
0
50
100
150
200
0.2
0.15
0.1
0.05
0
100
200
300
400
500