Recent Advances in Scalable Non-Gaussian Geostatistics: The Generalized Sub-Gaussian Model

Alberto Guadagnini[1,2], Monica Riva[1,2] and Shlomo P. Neuman[2]

[1]Dipartimento di Igegneria Civile Ambientale, Politecnico di Milano

Piazza L. Da Vinci 32, 20133 Milano, Italy

[2]Department of Hydrology and Atmospheric Sciences, University of Arizona

Tucson, AZ 85721, USA

ABSTRACT

Geostatistical analysis has been introduced over half a century ago to allow quantifying seemingly random spatial variations in earth quantities such as rock mineral content or permeability. The traditional approach has been to view such quantities as multivariate Gaussian random functions characterized by one or a few well-defined spatial correlation scales. There is, however, mounting evidence that many spatially varying quantities exhibit non-Gaussian behavior over a multiplicity of scales. The purpose of this minireview is not to paint a broad picture of the subject and its treatment in the literature. Instead, we focus on very recent advances in the recognition and analysis of this ubiquitous phenomenon, which transcends hydrology and the Earth sciences, brought about largely by our own work. In particular, we use porosity data from a deep borehole to illustrate typical aspects of such scalable non-Gaussian behavior, describe a very recent theoretical model that (for the first time) captures all these behavioral aspects in a comprehensive manner, show how this allows generating random realizations of the quantity conditional on sampled values, point toward ways of incorporating scalable non-Gaussian behavior in hydrologic analysis, highlight the significance of doing so, and list open questions requiring further research.

SCALABLE NON-GAUSSIAN DISTRIBUTIONS

Flow and transport in geologic media are known to be affected strongly by medium heterogeneity. As heterogeneity is difficult to map out deterministically, it has become common to characterize it geostatistically (e.g., Journel and Huijbregts, 1978; Kitanidis, 1997) and solve problems of flow and transport in spatially varying hydraulic conductivity environments stochastically (e.g., Dagan, 1989; Zhang, 2002; Rubin, 2003). The traditional approach has been

to treat log hydraulic conductivity, $\ln K(\mathbf{x})$, where $\mathbf{x}$ is position vector and K is hydraulic conductivity, as a multivariate (varying at multiple points $\mathbf{x}$) Gaussian random function characterized by one or a few well-defined spatial correlation scales (such that values at points separated by distances much larger than these scale are uncorrelated and, in the Gaussian case, statistically independent). Yet many hydrological (as well as diverse earth, environmental, ecological, biological, physical, social, financial and other) quantities vary in space (and/or time) in manners more complex than is allowed by such a simple Gaussian model. Many such variables exhibit behaviors similar to those of neutron porosity data from a deep vertical borehole (e.g. Guadagnini et al., 2015, and Riva et al., 2015) that we adopt for illustration purposes below.

Prior to describing this set of data, we note that our purpose here is not to provide a comprehensive review of statistical scaling and its treatment in the literature but rather to describe very recent advances brought about largely by our own work. We are motivated by the ubiquitous nature of the statistical scaling phenomena described in this minireview and by the fact that they transcend hydrology and Earth/environmental sciences. Our exposé assumes that the reader is comfortable with basic concepts and techniques of traditional statistical and geostatistical analyses as described, for example, in Journel and Huijbregts (1978) and Kitanidis (1997).

The data discussed below derive from a borehole drilled in the Maroon field of southwestern Iran within which oil and natural gas are produced. A total of N = 3,567 neutron porosity (P) data have been collected in the well at vertical intervals of $\Delta z = 0.1524$ m, ranging from 0 to 46.04% with mean 14.01% and standard deviation 6.40%. Figure 1 (top row, left) depicts the variation of deviations, $P' = P - P_a$, from the data average, P_a, with depth. The

deviations are seen to oscillate in an irregular manner with spurious peaks and valleys. This is reflected in their frequency distribution, plotted in Figure 1 (top row, right) on semi-logarithmic scale. The empirical frequency distribution exhibits a relatively sharp peak, asymmetry and slight bimodality which cannot be captured fully by a best (maximum likelihood, ML) fit of a Gaussian probability density function (pdf) to its empirical counterpart, depicted in the figure by a broken curve. Yet lack of Gaussianity becomes much more pronounced upon considering spatial increments $\Delta P(s) = P(z+s) - P(z)$ of P where z is elevation and s the vertical distance (separation scale, lag) between any pair of P values. For convenience we consider integer-valued lags, $s_n = s/\Delta z$, normalized by the vertical distance Δz between contiguous data. Figure 1 (rows 2 - 5, left) shows how increments $\Delta P(s_n)$ at four normalized lags (s_n = 1, 8, 48, 256) vary with sequential (integer) vertical position in the borehole. Frequency distributions of $\Delta P(s_n)$ at the same four lags are plotted on semi-logarithmic scale in rows 2 - 5 (Figure 1, right) respectively. The empirical frequency distributions exhibit pronounced symmetry with sharp peaks and heavy tails that decay gradually toward Gaussian shapes as lags increase. In other words, increment frequency distributions are non-Gaussian at short lags and scale with separation distance.

The reader may rightfully wonder whether the above phenomenon would be observable with a much smaller set of data. Our experience to date has shown it to be so for as few as 184 log permeability values obtained from 1 m-scale packer tests in multiple boreholes at an unsaturated fractured rock site in Arizona (Riva et al., 2013c).

Three types of mathematical models have been shown capable of mimicking the widely-observed scale-dependence of sharply-peaked, heavy-tailed non-Gaussian increment frequency distributions (and associated moments, as described below): multifractals (e.g., Mandelbrot,

1974; Monin and Yaglom, 1975; Lovejoy and Schertzer, 1995; Schertzer et al., 1997; Veneziano et al., 2006; Boffetta et al., 2008; Frisch, 2016), fractional Laplace (e.g. Meerschaert et al., 2004; Kozubowski et al., 2006, 2013) and sub-Gaussian representations (e.g. Samorodnitsky and Taqqu, 1994; Neuman, 2010; Guadagnini et al., 2012a; Riva et al., 2013). More about the first and the third later. Unfortunately, none of these three model types describe jointly, in a unified and comprehensive manner, statistical behaviors (in terms of probability distributions and/or moments) of a variable (in our case porosity) and its increments: instead, whereas multifractal and fractional Laplace models focus exclusively on increment statistics, the sub-Gaussian model can represent the statistics of either a function or its increments, but not of both. As such, these three models can neither characterize nor generate multiple realizations of a random function and its increments in a manner that is self-consistent and compatible with data. Remedying this fundamental deficiency of existing statistical frameworks has been one major motivation behind our very recent development (Riva et al., 2015a, 2015b) of the generalized sub-Gaussian (GSG) model on which we focus below.

SCALABLE STATISTICAL MOMENTS

The GSG model is comprehensive and unique in that it captures simultaneously observed scaling behaviors of (*a*) the probability density distribution and (*b*) (statistical) moments of a random function and its increments. It does so by incorporating the concept of a truncated fractal introduced into the literature by Di Federico and Neuman (1997) and Di Federico et al. (1999). A truncated fractal is formed by a Gaussian random function with a truncated power variogram. The variogram $\gamma(s)$ of a random function (where s is separation distance/scale or lag) is half the variance (semi-variance) of its increments. When the variogram is proportional to a power of the separation scale, $\gamma(s) \propto s^{2H}$ where $0 < H < 1$ is a so-called Hurst exponent, the function forms a

random fractal with fractal dimension defined in part by *H*. One of a fractal's most notable characteristics is the lack of a distinct (auto, spatial or temporal) correlation scale, there being instead a continuum of such scales ranging from zero to infinity. The concept was originally used by Neuman (1990, 1991, 1995) to help explain the widely observed dispersivity scale effect (for a recent discussion of this effect and its explanation see Neuman, 2017). Di Federico and Neuman (1997) and Di Federico et al. (1999) developed mathematical expressions for truncated power variograms representing a random fractal measured or resolved on a scale s_L and sampled across a larger domain (sampling window) of scale s_U. Their variogram closely approximates the power law $\gamma(s) \propto s^{2H}$ at lags intermediate between these cutoff scales, $s_L << s << s_U$, but deviates from the power law near the lower and upper cutoff scales, s_L and s_U. As variograms of most data show power law scaling over a limited range of lags (at most 2 - 3 log cycles of *s*), they are represented much more closely by truncated than by traditional fractal models. The truncation idea allows one to bridge statistical information about a fluctuating quantity (function) across cutoff (data measurement/resolution and sampling domain/window) scales: once the variogram (most notably its Hurst exponent *H*) of such a quantity has been established by measuring/resolving and sampling the quantity on given lower and upper cutoff scales, s_L and s_U, it can immediately be down and/or upscaled to any other choices of these scales. The concept and its early hydrogeologic applications have been reviewed by Neuman and Di Federico (2003). Further developments and applications based on the same concept are described by Neuman (2003, 2008, 2009) and Neuman et al. (2008).

Statistical moments of order *q* of absolute increments, designated here by S_N^q (*N* being the number of available increment data), are called structure functions; accordingly, a variogram

is a semi-structure function of order 2. Random fractals have power structure functions with exponents qH . Hence (*a*) plotting corresponding structure functions versus lag on log-log scale yields a straight line of slope qH and (*b*) plotting structure function exponents versus q yields a straight line of slope H. Yet experience has shown (see review by Neuman, 2010a) that (*a*) power-law behavior of structure functions generally breaks down at small and large lags, and (*b*) log-log plots of structure function exponents versus q often delineate curves rather than straight lines. Prior to illustrating some of this with our porosity data P we note that, as documented by Neuman, the literature commonly attributes (*a*) power-law breakdown at small lags to sampling noise, similar breakdown at large lags to undersampling, and (*b*) nonlinear variation of structure function exponents with q (nonlinear power-law scaling) to multifractality of the fluctuating quantity being analyzed. A multifractal is a non-Gaussian fractal with Hurst exponent H that varies continuously with q. Nonlinear power-law scaling is also exhibited by fractional Laplace functions (Meerschaert et al., 2004; Kozubowski et al., 2006, 2013) such as those applied to sediment transport data by Ganti et al. (2009). Yet Neuman (2010) and Siena et al. (2012) were able to demonstrate theoretically and numerically that both breakdown in power-law scaling at small/large lags (seen most clearly in Figure A1 of Siena et al.) and nonlinear power-law scaling (most clearly evident in Figures 8 and 12 of Siena et al.) are exhibited by truncated Gaussian fractals regardless of noise or undersampling. The most important implication of this finding is that the widespread tendency of researchers in many fields (including hydrology) to interpret nonlinear power-law scaling as an unambiguous indication of multifractality is unjustified.

Even though power-law scaling of structure functions is typically limited to intermediate ranges of lags, power-law scaling of structure functions ratios has been observed to hold in a wide variety of cases (for corroborating literature see Guadagnini et al., 2011, 2012b; Siena et

al., 2012) including turbulence, diffusion-limited aggregates, natural images, kinetic surface roughening, fluvial turbulence, sand wave dynamics, river morphometry, gravel-bed mobility, barometric pressure, low-energy cosmic rays, Martian topography, cosmic microwave background radiation, metal-insulator transition, irregularities in human heartbeat time series, turbulence in edge magnetized plasma of fusion devices and turbulent boundary layers of the Earth's magnetosphere, air permeability in fractured and sedimentary rocks. In particular, plotting $\log S^q$ versus $\log S^r$ for any positive orders q and r yields a straight line at all lags. The phenomenon, referred to in the literature as extended self-similarity (ESS), remained unexplained except for a relatively narrow class of functions satisfying a particular partial differential "Burger's" equation (as shown by Chakraborty et al. (2010) in the context of turbulence at high Reynolds numbers) until 2012 when Siena et al. proved it to be theoretically consistent with Gaussian truncated fractals. Their theory shows that, in the case of such fractals, the slope of the straight line resulting from plotting $\log S^q$ versus $\log S^r$ (when q and s are positive integers) is equal to q/r at all lags. We have seen that power-law scaling of structure functions, breakdown in such power-law scaling at small and large lags, and extended self-similarity are theoretically consistent with Gaussian truncated fractals. To account additionally for observed non-Gaussian heavy-tailed behavior, we (Neuman, 2010a, 2010b, 2011; Guadagnini et al., 2011, 2012a, 2012b, 2013, 2014, 2015; Riva et al. 2013, 2013b, 2013c) have initially (prior to introducing GSG) adopted and applied to a variety of hydrogeological and pedological data a sub-Gaussian model. This initial model represented a random function such as porosity deviations from the mean, P', by $P' = UG'$ where G' is a zero-mean Gaussian truncated fractal and U is a random variable (subordinator) totally skewed to the right of zero (hence non-negative) and statistically independent of G'. The model subordinates P' to G' through the

action of the subordinator *U*, rendering it a scale mixture of Gaussian truncated fractals with random variances proportional to U^2. The scale mixture is non-Gaussian with distribution that depends on the choice of *U* (we have in the past considered *U* to be alpha-stable or log normal). As such, it preserves the scaling properties of these fractals (as demonstrated theoretically for truncated power-law behavior by Neuman, 2010, and for extended self similarity by Guadagnini et al., 2012b and Neuman et al., 2013). The same is true about our more recent generalized sub-Gaussian model.

To illustrate some of the above scaling behaviors, we plot in Figure 2 sample porosity structure functions $S_N^q(s_n)$ of orders q = 0.5, 1.0 and 2.0 versus s_n on log-log scale. Log-log regression lines fitted to the data separately at vertical distance scales $s_n < 10$ and $s_n > 12$ suggest, at relatively high levels of confidence (coefficients of determination, R^2, ranging from 0.98 to 0.99 at $s_n < 10$ and from 0.89 to 0.99 at $s_n > 12$), that $S_N^q(s_n)$ varies as a power of s_n in each of these two scale ranges. Power-law exponents are larger at small ($s_n < 10$) than at large ($s_n > 12$) lags. We thus have a crossover between two diverse power-law regimes at distance scales 1.5 - 1.8 m delineated in Figure 2 by a dashed red line. Such dual scaling behavior is not characteristic of sub-Gaussian functions; it is to be viewed as an exception rather than the rule. We interpret the power-law scaling of $S_N^q(s_n)$ at $s_n < 10$ as representative of variability within, and that at $s_n > 12$ as a reflection of variability between, sedimentary layers at the site. Similar dual-scaling behavior has been reported by Siena et al. (2014) vis-á-vis porosities and specific surface areas imaged using X-ray computer microtomography throughout a millimeter-scale block of Estaillades limestone, at a spatial resolution of 3.3 μm, as well as Lagrangian velocities computed by solving the Stokes equation in the rock sample pore space. Following an approach

outlined by Guadagnini et al. (2013, 2014) one can use the method of moments to obtain estimates, $\hat{H}_w$ and $\hat{H}_b$, of Hurst scaling exponents, H_w and H_b, characterizing the within- and between-layers scaling behaviors of neutron porosity increments in the well. Consistent with our view of P' as a sub-Gaussian random function, $\hat{H}_w$ and $\hat{H}_b$ are set equal to the slopes, 0.86 and 0.10, of regression lines fitted to $S_N^1(s_n)$ on log-log scale at $s_n < 10$ and $s_n > 12$, respectively. The relatively large value of $\hat{H}_w$ suggests that intra-layer variability is persistent in that large values tend to follow large values and small values tend to follow small values; the relatively small value of H_b implies that inter-layer variability is strongly antipersistent, small and large values tending to alternate rapidly. The latter is likely a manifestation of strong variations in environments responsible for the deposition of alternating sedimentary layers.

As no theory other than ours (Siena et al., 2012; Guadagnini et al., 2012b; Neuman et al., 2013) is known to explain extended self-similarity of variables that do not necessarily satisfy Burger's equation (Chakraborty et al., 2010), demonstrating that the above porosity data satisfy ESS is akin to verifying that these data conform to our theoretical scaling framework. That this is indeed the case becomes evident upon examining the high-confidence (R^2 = 0.91-0.99) straight-line relationships between log $S_N^{q+1}(s_n)$ and log $S_N^q(s_n)$, and corresponding power-law relationships between $S_N^{q+1}(s_n)$ and $S_N^q(s_n)$, at $s_n < 10$ and $s_n > 12$ in Figure 3 for q = 1, 2 and 3. According to Siena et al. (2012) the slopes $\log S_N^{q+1}(s_n) / \log S_N^q(s_n)$ of the straight lines are equal to (1 + 1/q) in the purely Gaussian case. In the two non-Gaussian cases depicted in Figure 3, the slopes correspondingly diminish from 1.97 through 1.47 to 1.31 when $s_n < 10$, and from 1.33 through 0.99 to 0.90 when $s_n > 12$, as q increases from 1 through 2 to 3).

At this stage one can compute functional relationships between power exponents $\xi_w(q)$ and $\xi_b(q)$, and the order *q*, of P' structure functions that scale as powers of lag. In the method of moments these powers are the slopes of regression lines fitted to log-log plots of $S_N^q(s_n)$ versus s_n, such as those depicted in Figure 2. In the case of ESS we use $\xi_w(q=1)$ and $\xi_b(q=1)$, determined by the method of moments, as reference values for the sequential computation of $\xi_w(q)$ and $\xi_b(q)$ at $q > 1$ based on known power-law relationships between $S_N^{q+1}(s_n)$ and $S_N^q(s_n)$, such as those given in Figure 3. Corresponding plots of $\xi_w(q)$ and $\xi_b(q)$ as functions of *q*, evaluated by the method of moments and ESS at $s_n < 10$ and $s_n > 12$, are presented in Figure 4. Results obtained by the two methods are, for the most part, very similar. In both cases $\xi_w(q)$ and $\xi_b(q)$ delineate convex functions that fall below straight lines having slopes $\hat{H}_w$ and $\hat{H}_b$, respectively, which pass through the origin. As already noted, tradition has it that whereas such straight lines are characteristic of monofractal random fields, nonlinear variations of power exponents such as those exhibited by $\xi_w(q)$ and $\xi_b(q)$ in Figure 4 are symptomatic of multifractals (or, as some might suggest, fractional Laplace functions). Yet we have seen that similar behavior is characteristic of random fields subordinated to truncated monofractals. For reasons described below, we believe that this is also the case with P'.

GENERALIZED SUB-GAUSSIAN MODEL

We noted earlier that no statistical framework known to us prior to 2015 (most notably multifractals, fractional Laplace functions, or the sub-Gaussian model) was able to capture jointly, in a unified and self-consistent manner, statistical scaling behaviors of a variable and its increments. The standard sub-Gaussian model fails to do so consistently because (*a*) multiplying

any one realization of G by a random number U renders all realizations of P' and ΔP Gaussian, rendering it impossible to generate more realistic non-Gaussian realizations, and (*b*) setting $\Delta P = U\Delta G$ where U is a random constant fails to ensure that both ensemble and sample distributions of ΔP scale with lag in the aforementioned manner. We likewise noted that without the ability to reconcile the statistical behaviors of a random function and its increments, one can neither characterize them consistently by sampling the function, its increments or both, nor generate random realizations of either the function or its increments consistent with each other and with available data samples.

The difficulty has been overcome by us (Riva et al., 2015a, 2015b) very recently through the development of a new generalized sub-Gaussian (GSG) model, which captures the behaviors of both a function and its increments jointly in a comprehensive and consistent manner. Our new model is written as

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

where $G'(\mathbf{x})$ is a zero-mean Gaussian random function characterized by standard deviation σ_G and correlation function ρ_G. This GSG model differs from its sub-Gaussian precursor in that U is now a non-negative random function rather than a random variable. Values of $U(\mathbf{x})$ at any two points in space (or time) are statistically independent of each other. This notwithstanding, the fact that U now varies spatially makes the consistent generation of non-Gaussian realizations possible. Starting with a log-normal subordinator $U(\mathbf{x})$, we were able to (*a*) derive analytical expressions for bivariate and marginal distributions of a random function and marginal distributions of its increments, (*b*) propose new approaches to infer parameters characterizing the GSG model based on samples of one or (more importantly and reliably) both of these variables,

(*c*) develop a multidimensional algorithm to generate statistically isotropic and anisotropic realizations of the function (which in turn allows generating corresponding realizations of its increments), and (*d*) demonstrate the accuracy of this generation algorithm by comparing ensemble statistics of the function and its increments (such as, mean, variance, variogram and probability density function) with those of Monte Carlo generated realizations. Figure 5 is a flowchart illustrating the 9 steps one would follow to detect GSG signatures in a given set of data, to estimate the parameters of one or more corresponding GSG model(s), generate random equally likely (conditional or unconditional) GSG realizations, and use these as inputs into stochastic process (e.g. flow, transport) models. The steps are described below. Though their illustration focuses on one-dimensional porosity data, similar steps apply (with slight modifications) to multidimensional (and statistically anisotropic) data sets (e.g., Riva et al., 2015b).

Steps 1 - 3. One starts by analyzing the frequency distributions of available data (P, Step1) and space (or time) increments, ΔP (Step 2). If values of ΔP are available at a sufficiently wide range of small to large separation scales, one examines (Step 3) whether or not increment frequency distributions vary with lag, as they do in the right column of Figure 1. If they do, analysis by means of a GSG model is indicated and one proceeds to Step 4. If they do not, the data may be amenable to analysis by classical methods of geostatistics.

Steps 4 - 6. Step 4 consists of selecting a GSG subordinator $U(\mathbf{x})$. To date, we have worked solely with a log-normal subordinator (Riva et al., 2015b), but other choices are possible. Start by assuming that U is lognormally distributed according to $\ln N\left(0,(2-\alpha)^2\right)$ where $0 < \alpha \leq 2$ is a parameter. The GSG model

is then characterized by two constant parameters, α and σ_G, and one lag-dependent parameter, ρ_G. One way to estimate these parameters is to first compute the sample variance M_2^P of *P'* and the sample variance and kurtosis $M_2^{\Delta P}$ and $M_4^{\Delta P}$, respectively, of ΔP (Step 5), then evaluate the parameters (Step 6) according to

$$\xi^4 - 2\xi^2\left(1-\left(\frac{M_2^{\Delta P}}{2M_2^P}\right)^2\right)+1-\frac{M_4^{\Delta P}}{6\left(M_2^P\right)^2}=0; \qquad \sigma_G^2=\frac{M_2^P}{\xi^2}; \qquad \rho_G=\xi\left(1-\frac{M_2^{\Delta P}}{2M_2^P}\right) \tag{2}$$

where $\xi = \mathrm{e}^{(2-\alpha)^2}$. Another way is to (Steps 5 - 6) estimate α and ρ_G upon fitting the theoretical probability density function of ΔP

$$f_{\Delta P}(\Delta p)=\frac{1}{2\pi^2\left(2-\alpha\right)^2}\sqrt{\frac{\pi}{2}}\int_0^\infty\int_0^\infty \frac{e^{-\frac{1}{2}\left[\frac{1}{(2-\alpha)^2}\left(\ln^2\frac{u_1}{\sigma_G}+\ln^2\frac{u_2}{\sigma_G}\right)+\frac{(\Delta p)^2}{u_1^2+u_2^2-2u_1u_2\rho_G}\right]}}{u_2u_1\sqrt{u_1^2+u_2^2-2u_1u_2\rho_G}}\,du_2du_1 \tag{3}$$

to frequency distributions of ΔP at all available lags by Maximum Likelihood (ML) while setting the variance of *P'* equal to its sample value counterpart, M_2^P. Riva et al. (2015) have shown that both options yield similar parameter estimates though the second is computationally much more demanding than the first.

Step 7. Once GSG parameters have been estimated, one is in a position to verify (Step 7) that (*a*) corresponding GSG pdfs of both *P'* and ΔP (at all considered lags) provide adequate representations of corresponding empirical frequency distributions (as is seen to happen on the right hand side of Figure 1) and (*b*) estimates of α and σ_G show no systematic variations with lag (as is the case, in the context of our porosity data, in Figs. 12 and 13 of Riva et al., 2015a).

Otherwise one may in principle (as we have not yet done so) repeat Steps 4 - 6 with a different subordinator $U(\mathbf{x})$ (note however that Eqs. (2) - (3) would change accordingly) as indicated by the dashed box in Figure 5.

Step 8. Once a suitable GSG model has thus been established, one can generate scalable, non-Gaussian, equally likely unconditional (Riva et al., 2015b) or conditional (Panzeri et al., 2016) Monte Carlo realizations of corresponding GSG functions. Unconditional realizations are obtained by (*i*) generating a zero-mean stationary Gaussian random function, $G'(\mathbf{x})$, on a grid and (*ii*) multiplying each discrete value of $G'(\mathbf{x})$ by a (non-negative) random draw of $U(\mathbf{x})$. Generating realizations conditioned on noisy measurements remains a partially resolved challenge (Panzeri et al., 2016). The reason is that we do not presently have a closed-form expression for the conditional pdf of GSG, rendering its evaluation for more than just a few conditioning points computationally infeasible. Though Panzeri et al. were able to come up with two approximate, computationally less demanding ways to partially overcome this constraint, the efficient generation of conditional GSG realizations remains an open challenge.

Step 9. Random GSG functions generated in the above manner can be used in Step 9 to conduct Monte Carlo simulations of associated processes such as flow and transport in randomly heterogeneous, scale-variable non-Gaussian environments. Potential applications, such as those concerning environmental risk assessment and management, are virtually endless. An early step in this direction was taken by Riva et al. (2017) who had explored analytically lead-order effects that non-Gaussian log hydraulic conductivity described by a GSG model could have on the

stochastic description of subsurface flow and transport. Libera et al. (2017) presented a numerical Monte Carlo study assessing the influence of a GSG structure of log-transmissivities on uncertainty associated with concentrations of nonreactive solutes detected at a pumping well operating under transient flow conditions in a simplified two-dimensional system.

Just like its sub-Gaussian precursor, our GSG model retains all scaling properties of statistical moments discussed in the previous section. Remarkably, GSG structure functions exhibit a nugget effect that renders them nonzero at zero lag, $s = 0$. The variogram of P is given by

$$\gamma_P = \frac{\langle \Delta P^2 \rangle}{2} = \sigma_G^2 \left[\sigma_U^2 + \langle U \rangle^2 \left(1 - \rho_G \right) \right] \tag{4}$$

where σ_U^2 is the variance of U. The nugget effect, $\sigma_G^2 \sigma_U^2$, disappears when U is a constant (as in the Gaussian case). This suggests to us that the widely-observed nugget effect, commonly attributed in the geostatistical literature to unresolved fluctuations and/or noise in measured values of a function P', may in fact be wholly or in part due to (and thus indicative of) the generalized non-Gaussian nature of P'.

We noted earlier that G', a truncated Gaussian fractal, exhibits a breakdown in power-law scaling of structure functions caused by lower and upper cutoff scales, s_L and s_U. An upper cutoff causes structure functions of G', and correspondingly of GSG functions P', to flatten at $s >> s_U$. Such flattening is widely observed and nicely illustrated on cloud radiance and daily runoff accumulation data by Tessier et al. (1993). Neuman et al. (2008) demonstrated theoretically and illustrated on log hydraulic conductivity data that it may cause misinterpretation of G' and/or P' as stationary random fields. A lower cutoff causes structure functions of G',

and correspondingly of P', to steepen at $s << s_L$ as illustrated theoretically for G' in Figure A1 of Siena et al. (2012), and exemplified by river runoff data reported by Koscielny-Bunde et al. (2006). Yet many data (e.g., those of Tessier et al., 1993) show a flattening rather than a steepening of structure functions at $s << s_L$. The flattening is caused by a nugget effect, which we saw earlier characterizes GSG functions regardless of data resolution or noise. The question whether GSG structure functions steepen or flatten at $s << s_L$ is thus seen to depend on an interplay between the steepening effect of s_L and the flattening effect of a nugget.

SUMMARY

Following is a brief summary of our main points:

1. Many hydrologic and a wide variety of other randomly fluctuating quantities exhibit scalable non-Gaussian behaviors. Commonly observed features of this behavior include (*a*) non-Gaussian distribution of the quantity with mild or sharp peaks, light tails, some asymmetry and bimodality, (*b*) symmetric distributions of increments with sharp peaks and heavy tails that decay (scale) toward Gaussian shapes with increasing lags, (*c*) power law scaling of structure functions (statistical moments of absolute increments) in midranges of lags, (*d*) breakdown in power-law scaling manifested by steepening or flattening of structure functions at small lags and flattening at large lags, (*e*) extended power law scaling (linear relations between log structure functions of diverse orders) at all lags, and (*f*) nonlinear scaling of power-law exponents with order of structure function.
2. On one hand, the tendency of increment distributions toward Gaussian shapes at all but short lags, and the flattening of structure functions at large lags, explain why many scalable non-Gaussian quantities appear and are often interpreted to be monoscale Gaussian. On the other

hand, nonlinear scaling of power-law exponents is often interpreted to imply multifractality. These two interpretations are mutually exclusive.

3. No statistical framework known to us prior to 2015 (e.g., multifractals, fractional Laplace functions, sub-Gaussian representations) has been able to capture jointly, in a self-consistent and comprehensive manner, scalable non-Gaussian behaviors of a random function and its increments. To our knowledge, the generalized sub-Gaussian (GSG) model proposed by us in 2015 is the first to do so.
4. The most important novel feature of our GSG model is its unique ability to reconcile the statistical behaviors of a scalable non-Gaussian random function and its increments. This, for the first time, allows one to (*a*) characterize both the function and its increments jointly and consistently by sampling one or both, as we have illustrated here, and (*b*) generate random realizations of both in a manner consistent with data and each other. Doing so is a prerequisite for the proper embedding of scalable non-Gaussian quantities, such as porosity and log hydraulic conductivity, in stochastic models of phenomena such as subsurface flow and transport. Our motivation in writing this minireview is to alert the hydrologic and broader scientific communities to (*a*) the need for a new geostatistical framework to account for ubiquitous non-Gaussian scaling behaviors of random functions and their increments and (*b*) to describe and illustrate a very recent framework that does so in a unified and comprehensive manner.
5. Open issues awaiting future research and development include (*a*) the effect on GSG of subordinators other than log-normal, (*b*) GSG based statistical characterization of diverse hydrologic and other quantities exhibiting scalable non-Gaussian behaviors in multidimensional space and/or time, (*c*) computationally efficient generation of

multidimensional and/or temporal GSG realizations conditional on sampled data, and (*d*) analytical and/or computational solution of stochastic problems involving GSG functions.

Acknowledgement

Funding from MIUR (Italian ministry of Education, University and Research, Water JPI, WaterWorks 2014, Project: WE-NEED- Water NEEDs, availability, quality and sustainability) and from the European Union's Horizon 2020 Research and Innovation programme (Project "Furthering the knowledge Base for Reducing the Environmental Footprint of Shale Gas Development" FRACRISK - Grant Agreement No. 636811) are acknowledged. We are grateful to Professor Muhammad Sahimi, University of Southern California, for having shared with us borehole geophysical log data some of which we have discussed in this study.

REFERENCES

Boffetta, G., Mazzino, A., Vulpiani, A., 2008. Twenty- five years of multifractals in fully developed turbulence: a tribute to Giovanni Paladin. J. Phys. A: Math. Theor. 41, 363001.

Chakraborty, S., Frisch, U., Ray, S.S., 2010. Extended self-similarity works for the Burgers equation and why. J. Fluid Mech. 649, 275285. doi:10.1017/S0022112010000595.

Dagan, G., 1989. Flow and Transport in Porous Formations. Springer-Verlag, New York.

Di Federico, V., Neuman, S.P., 1997. Scaling of random fields by means of truncated power variograms and associated spectra. Water Resour. Res. 33(5), 1075-1085.

Di Federico, V., Neuman, S.P., Tartakovsky, D.M., 1999. Anisotropy, lacunarity, upscaled conductivity and its covariance in multiscale fields with truncated power variograms. Water Resour. Res. 35(10), 2891-2908.

Frisch, U., 2016. The collective birth of multifractals. J. Phys. A: Math. Theor. 49, 451002. doi:10.1088/1751-8113/49/45/451002.

Ganti, V., Singh, A., Passalacqua, P., Foufoula-Georgiu, E., 2009. Subordinated Brownian motion model for sediment transport. Phys. Rev. E 80, 011111. doi:1539-5663755/2009/80(1)/011111(8), 2009.

Guadagnini, A., Neuman, S.P., 2011. Extended power-law scaling of self-affine signals exhibiting apparent multifractality. Geophys. Res. Lett. 38, L13403. doi:10.1029/2011GL047727.

Guadagnini, A., Neuman, S.P., Riva, M., 2012a. Numerical investigation of apparent multifractality of samples from processes subordinated to truncated fBm. Hydrol. Processes 26(19), 2894-2908. doi:10.1002/hyp.8358.

Guadagnini, A., Riva, M., Neuman, S.P., 2012b. Extended power-law scaling of heavy-tailed random air-permeability fields in fractured and sedimentary rocks. Hydrol. Earth Syst. Sci. 16, 3249-3260. doi:10.5194/hess-16-3249-2012.

Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2013. Anisotropic statistical scaling of vadose zone hydraulic property estimates near Maricopa, Arizona. Water Resour. Res. 49, 8463-8479. doi:10.1002/2013WR014286.

Guadagnini, A., Neuman, S.P., Schaap, M.G., Riva, M., 2014. Anisotropic statistical scaling of soil and sediment texture in a stratified deep vadose zone near Maricopa, Arizona. Geoderma 214-215, 217-227. doi:10.1016/j.geoderma.2013.09.008.

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. Hydrol. Earth Syst. Sci. 19, 729-745. doi:10.5194/hess-19-729-2015.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic, San Diego, Calif.

Kitanidis, P., 1997. Introduction to Geostatistics, Applications in Hydrogeology. Cambridge Univ. Press, Cambridge, U. K.

Koscielny-Bunde, E., Kantelhardt, J.W., Braun, P., Bunde, A., Havlin, S., 2006. Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. J. Hydrol. 322 120-137.

Kozubowski, T.J., Meerschaert, M.M., Podgorski, K., 2006. Fractional Laplace motion. Adv. Appl. Probab. 38, 451-464. doi:10.1239/aap/1151337079.

Kozubowski, T.J., Podgorski, K., Rychlik, I., 2013. Multivariate generalized Laplace distribution and related random fields. J. Multivar. Anal. 113, 59-72.

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. J. Contam. Hydrol., in press.

Lovejoy, S., Schertzer, D., 1995. Multifractals and rain. In New Uncertainty Concepts in Hydrology and Hydrological Modelling, Kundzewicz A.W. (Ed.). Cambridge University Press: New York 61-103.

Mandelbrot, B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. J. Fluid Mech. 62, 331-58.

Meerschaert, M.M., Kozubowski, T.J., Molz, F.J., Lu, S., 2004. Fractional Laplace model for hydraulic conductivity. Geophys. Res. Lett. 31, L08501.

Monin, A.S., Yaglom, A.M., 1975. Statistical Fluid Mechanics: Mechanics of turbulence. MIT Press: Cambridge 874.

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

Neuman, S.P., 1994. Generalized scaling of permeabilities: Validation and effect of support scale. Geophys. Res. Lett. 21(5) 349-352. doi:10.1029/94GL00308.

Neuman, S.P., 1995. On advective transport in fractal permeability and velocity fields, Water Resour. Res. 31(6). 1455-1460.

Neuman, S.P., 2003. Relationship between juxtaposed, overlapping and fractal representations of multimodal spatial variability. Water Resour. Res. 9(8), 1205. doi:10.1029/2002WR001755.

Neuman, S.P., 2008. Multiscale relationships between fracture length, aperture, density and permeability. Geophys. Res. Lett. 35, L22402. doi:10.1029/2008GL035622.

Neuman, S.P., 2009. Hortonian scaling and effect of cutoffs on statistics of self-affine river networks. Water Resour. Res. 45, W12408. doi:10.1029/2009WR008124.

Neuman, S.P., 2010a. Apparent/spurious multifractality of data sampled from fractional Brownian/Lévy motions. Hydrol. Processes 24, 2056-2067. doi:10.1002/hyp.7611.

Neuman, S.P., 2010b. Apparent/spurious multifractality of absolute increments sampled from truncated fractional Gaussian/ Lévy noise. Geophys. Res. Lett. 37, L09403. doi:10.1029/2010GL043314.

Neuman, S.P., 2011. Apparent multifractality and scale-dependent distribution of data sampled from self-affine process. Hydrol. Process 25(11), 1837-1840. doi:10.1002/hyp.7967.

Neuman, S. P., 2016. Comment on ‘‘Advective transport in heterogeneous aquifers: Are proxy models predictive?’’ by A. Fiori, A. Zarlenga, H. Gotovac, I. Jankovic, E. Volpi, V. Cvetkovic, and G. Dagan. Water Resour. Res. 52, 5701–5702, doi:10.1002/2016WR019093.

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

Neuman, S.P., Riva, M., Guadagnini, A., 2008. On the geostatistical characterization of hierarchical media. Water Resour. Res. 44, W02403. doi:10.1029/2007WR006228.

Neuman, S.P., Guadagnini, A., Riva, M., Siena, M., 2013. Recent advances in statistical and scaling analysis of earth and environmental variables. In: Mishra, P.K., Kuhlman, K.L. (Eds.), Advances in Hydrogeology. Springer, New York, pp. 1-15.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S.P., 2016. Theory and generation of conditional, scalable sub-Gaussian random fields. Water Resour. Res. 52, 1746-1761. doi:10.1002/2015WR018348.

Riva, M., Neuman, S.P., Guadagnini, A., 2013a. On the identification of Dragon Kings among extreme-valued outliers. Nonlin. Processes Geophys. 20, 549-561, doi:10.5194/npg-20-549-2013.

Riva, M., Neuman, S.P., Guadagnini, A., Siena, M., 2013b. Anisotropic scaling of berea sandstone log air permeability statistics. Vadose Zone J. 12, 1-15. doi:10.2136/Vzj2012.0153, 2013b.

Riva, M., Neuman, S.P., Guadagnini, A., 2013c. Sub-Gaussian model of processes with heavy-tailed distributions applied to air permeabilities of fractured tuff. Stoch. Env. Res. Risk A. 27, 195-207. doi:10.1007/s00477-012-0576-y, 2013c.

Riva, M., Neuman, S.P., Guadagnini, A., 2015a. New scaling model for variables and increments with heavy-tailed distributions. Water Resour. Res. 51, 4623-4634. doi:10.1002/2015WR016998.

Riva, M., Panzeri, M., Guadagnini, A., Neuman, S.P., 2015b. Simulation and analysis of scalable non-Gaussian statistically anisotropic random functions. J. Hydrol. 531, 88-95, doi: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 Resour. Res. 53, 2998-3012, doi:10.1002/2016WR019353.

Rubin, Y., 2003. Applied Stochastic Hydrogeology. Oxford University Press, Oxford.

Samorodnitsky, G., Taqqu, M.S., 1994. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman and Hall, New York.

Schertzer, D., Lovejoy, S., Schmitt, F., Chigirinskaya, Y., Marsan, D., 1997. Multifractal cascade dynamics and turbulent intermittency. Fractals 5 (3), 427-471.

Siena, M., Guadagnini, A., Riva, M., Neuman, S.P., 2012. Extended power-law scaling of air permeabilities measured on a block of tuff. Hydrol. Earth Syst. Sci. 16, 29-42. http://dx.doi.org/10.5194/hess-16-29-2012.

Siena, M., Guadagnini, A., Riva, M., Bijeljic, B., Pereira Nunes, J.P., Blunt, M.J., 2014. Statistical scaling of pore-scale Lagrangian velocities in natural porous media. Phys. Rev. E 90, 023013. doi:10.1103/PhysRevE.90.023013.

Tessier, Y., Lovejoy, S., Schertzer, D., 1993. Universal multifractals: theory and observations for rain and clouds. J. Applied Meteorology 32, 223-250.

Veneziano, D., Langousis, A., Furcolo, P., 2006. Multifractality and rainfall extremes: a review. Water Resour. Res. 42, W06D15. doi:10.1029/2005WR004716.

Zhang, D.Z., 2002. Stochastic Methods for Flow in Porous Media. Academic, San Diego, Calif.

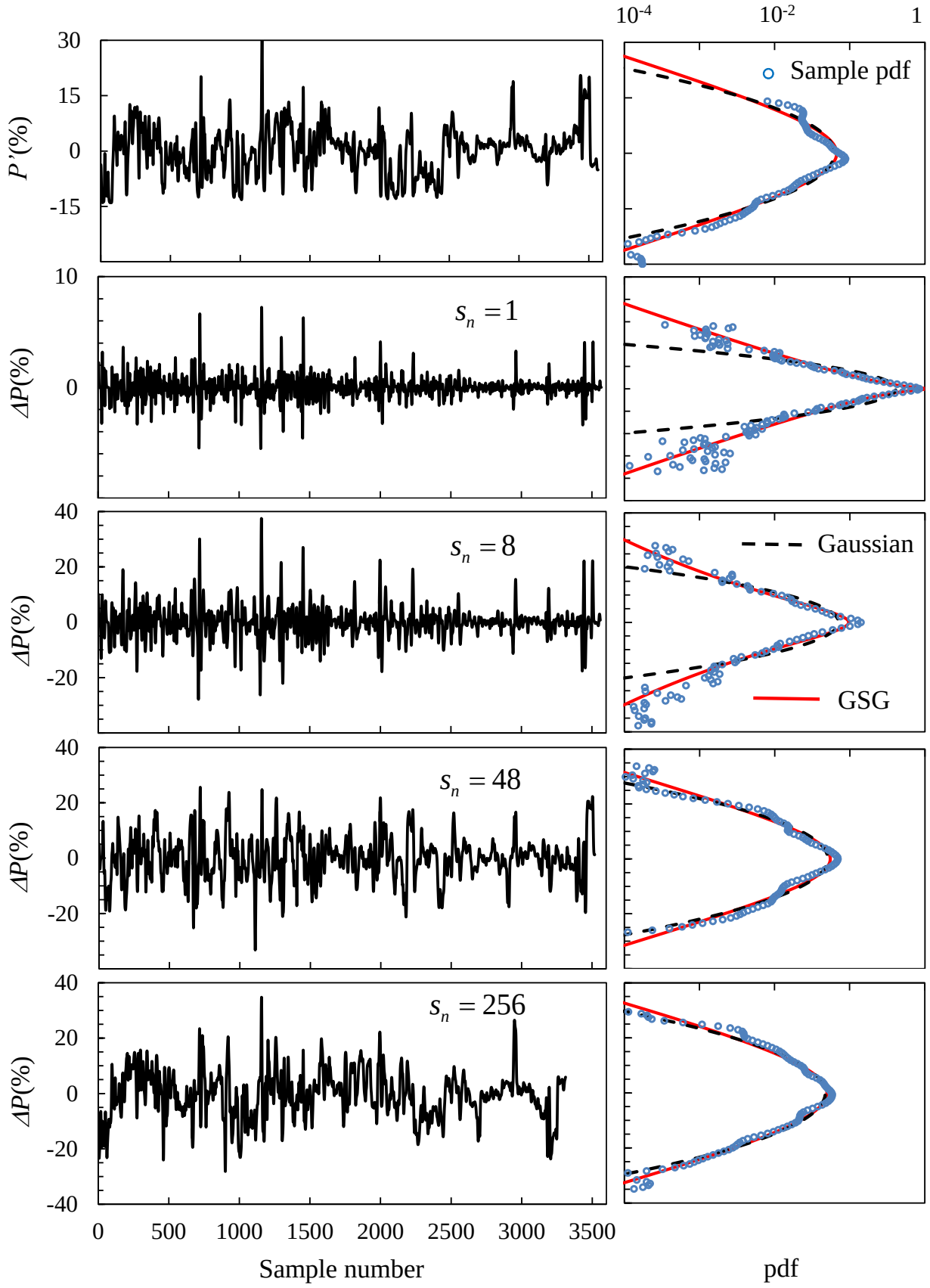

Figure 1 Left column: P' and $\Delta P(s_n)$ versus sequential (integer) vertical position. Right column: frequency distributions (sample pdfs) of P' and $\Delta P(s_n)$. Also shown are ML fits of Gaussian (dashed) and GSG (solid red) pdfs

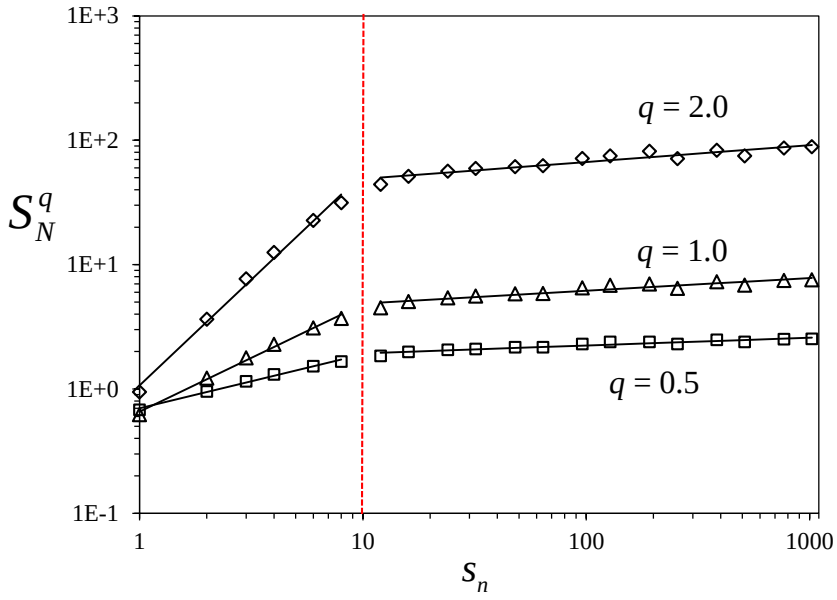

Figure 2 S_N^q versus s_n for q = 0.5, 1.0, 2.0. Red dashed line demarcates breaks in power-law scaling regimes.

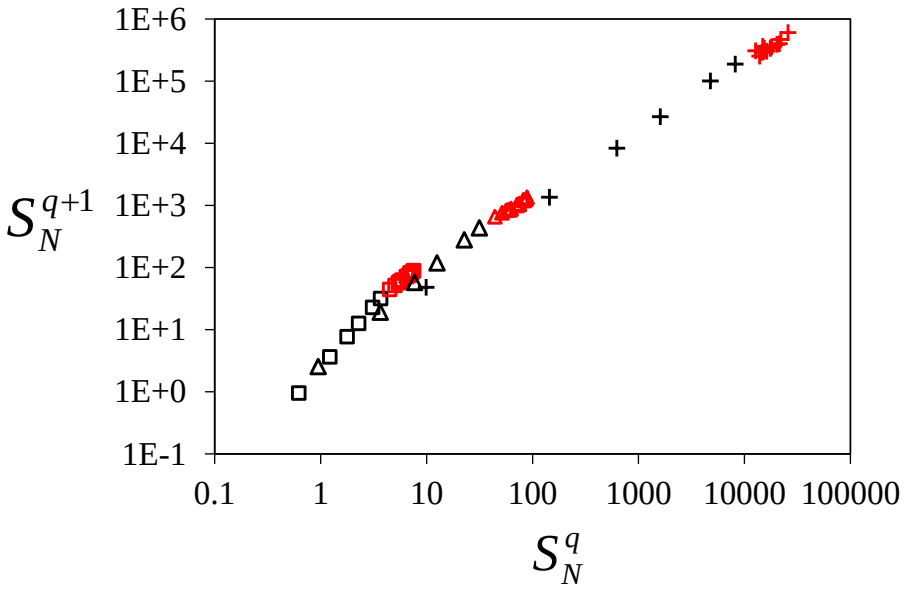

Figure 3 S_N^{q+1} versus S_N^q for q = 1.0 (□), 2.0 (△), 3.0 (+) for $s_n < 10$ (black symbols) and $s_n > 12$ (red symbols).

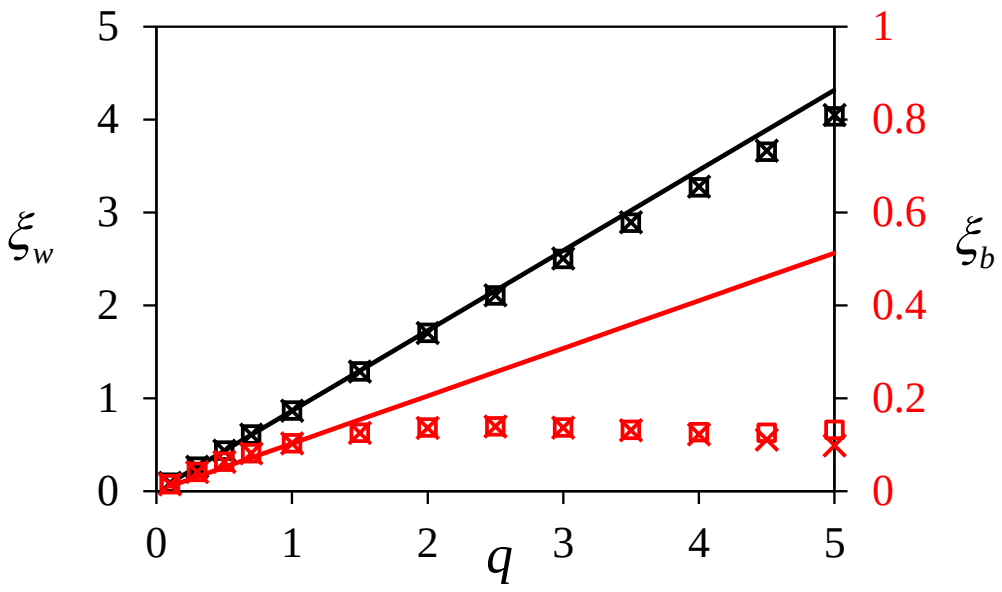

Figure 4 ξ_w (black) and ξ_b (red) versus q evaluated by the method of moments (×) and ESS (□).

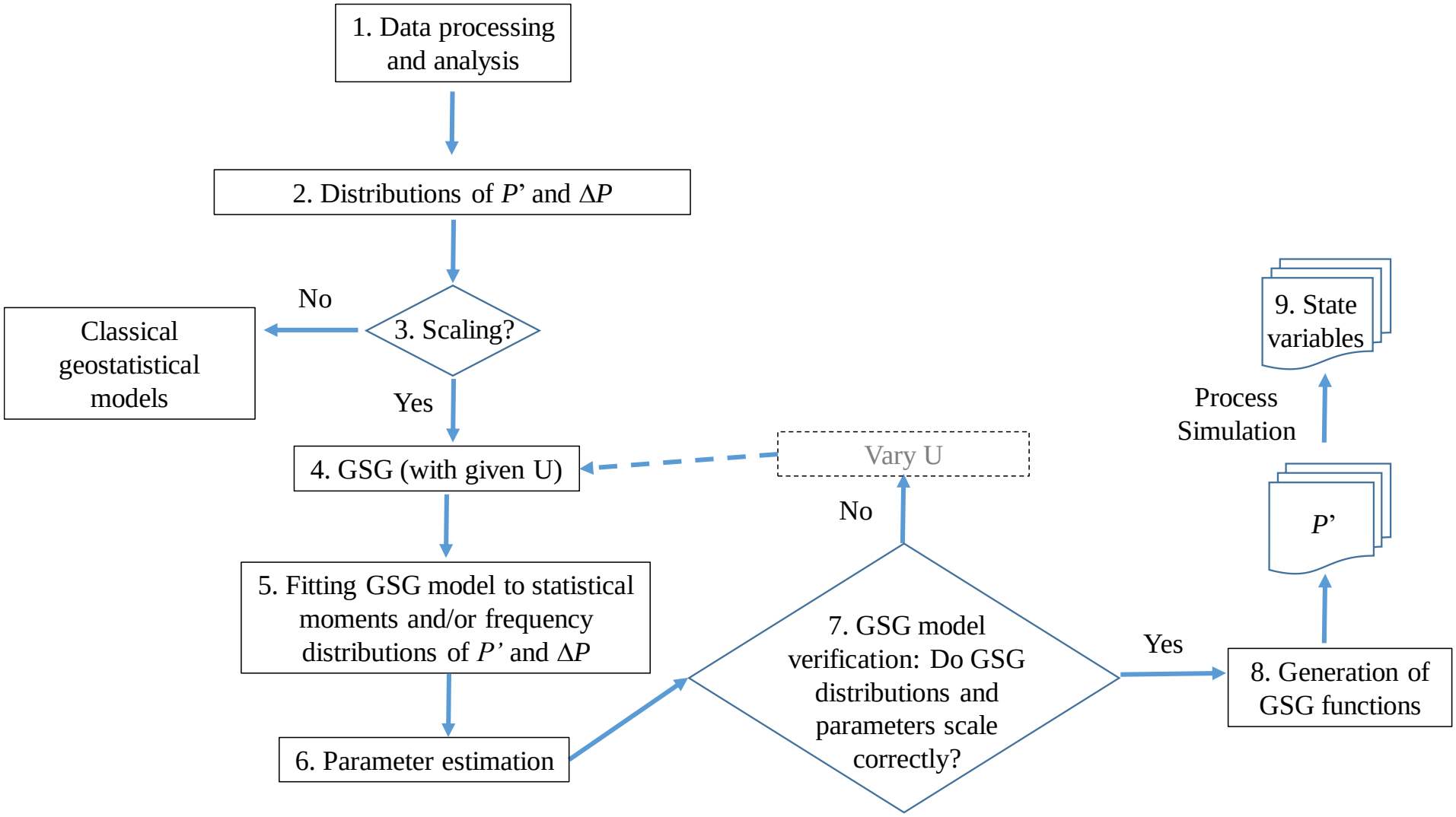

Figure 5 Flowchart illustrating the 9 steps followed to detect and model GSG signatures in a given data set, generate random (conditional or unconditional) GSG realizations, and use these as inputs into stochastic process (e.g. flow, transport) models.