Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR038487

Characterization of Spatially Heterogeneous Environmental Variables Through Multi-Modal Generalized Sub-Gaussian Distributions

Chiara Recalcati[1], Alberto Guadagnini[1], and Monica Riva[1]

[1]Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy

Key Points:

- A theoretical framework enabling characterization of multi-modal random fields associated with non-Gaussian components is derived
- Our modeling framework accurately captures the main traits of probability distributions of calcite dissolution rates
- We document a strong link between parameters of our model and mechanisms driving dissolution reactions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Riva,
monica.riva@polimi.it

Citation:

Recalcati, C., Guadagnini, A., & Riva, M. (2025). Characterization of spatially heterogeneous environmental variables through multi-modal Generalized sub-Gaussian distributions. *Water Resources Research*, *61*, e2024WR038487. https://doi.org/10.1029/2024WR038487

Received 22 JUL 2024
Accepted 15 JAN 2025

Abstract We provide a sound theoretical framework for the characterization of randomly heterogeneous spatial fields exhibiting multi-modal, long-tailed probability densities. Multi-modal distributions are at the core of conceptual models employed to represent heterogeneity of hydrogeological or geochemical systems across which one can otherwise distinguish diverse regions whose location is uncertain. Within each region, the quantity of interest shows a distinct heterogeneous pattern that can be described through a generally non-Gaussian distribution. Our analytical model embeds the joint formulation of the probability density of the target variable and its spatial increments. The distributions of the latter scale with separation distance between locations at which increments are evaluated. This feature is in line with documented experimental observations of a variety of Earth system quantities. Our stochastic modeling framework integrates approaches based on unimodal non-Gaussian fields described through a Generalized Sub-Gaussian model and (multi-modal) distributions resulting from mixtures of Gaussian fields. These are recovered as specific instances within our comprehensive formulation. We apply this framework to an experimental data set consisting of a collection of dissolution rate fields obtained from high-resolution nanoscale measurements acquired through Atomic Force Microscopy and documenting the dissolution behavior of a calcite sample under continuous flow conditions. Our findings demonstrate the capability of our stochastic approach to elucidate key statistical traits and scaling features inherent in the heterogeneous distributions of these types of environmental variables.

Plain Language Summary Spatially heterogeneous fields of hydrogeological or geochemical quantities are typically distributed across distinct regions whose spatial arrangement within a given observation area can be either deterministically known or affected by uncertainty. These scenarios are sometimes represented through a class of approaches that rely on multi-modal distributions. These distributions emerge when values of the specific physical property, which are linked to various sub-domains, are clustered within a single population. When examining each of these regions separately, a variety of environmental variables therein are often documented to exhibit marked non-Gaussian behaviors. These manifest through a pronounced peak and heavy tails of the probability density of the target quantity and in the tendency of the frequency distributions of associated spatial increments to change with the separation distance at which these are evaluated. Our study introduces a detailed modeling framework that captures within a multi-modal distribution these non-Gaussian traits. We then apply our theoretical formulations to experimental data. These consist of spatial maps of dissolution rates evaluated from nanoscale measurements collected via Atomic Force Microscopy imaging of a calcite sample undergoing dissolution under continuous flow.

1. Introduction

This study aims to develop a robust modeling framework for the characterization of randomly heterogeneous non-Gaussian spatial fields that exhibit multi-modal and long-tail features in their probability density functions (PDFs). The study is motivated by the observation that (*a*) non-Gaussian patterns are ubiquitously documented for a wide variety of variables in the hydrological and hydrogeological context and (*b*) spatial fields encompassing regions associated with diverse and/or coexisting processes or system attributes are sometimes modeled upon lumping available data as a unique sample population (see, e.g., Winter et al., 2003, and references therein). The latter modeling approach is frequently employed, for example, to describe spatially heterogeneous distributions of permeabilities of porous media whose internal geological architecture comprises diverse geomaterials, each occupying a given region of the domain (e.g., Dai et al., 2004; Lu & Zhang, 2002; Massabó et al., 2008; Zhang et al., 2013). Grouping together data associated with such diverse geomaterials, regardless of material

category, yields a statistically homogenized conceptualization of the heterogeneous spatial distribution of the property considered. The otherwise composite nature of the spatial distribution of geomaterials is then reflected through a starkly multi-modal sample PDF of the spatially heterogeneous properties (Desbarats, 1987; Journel, 1983; Rubin, 1995; Rubin & Journel, 1991). A typical approach to modeling a given random function of interest, $Y(\mathbf{x})$ ($\mathbf{x}$ denoting a vector of spatial coordinates), exhibiting such multi-modal features is grounded on the use of Gaussian mixtures. In this context, $Y(\mathbf{x})$ is modeled as a sum of $m = 1, \dots, M$ independent Gaussian modes, $Y_m(\mathbf{x})$, that are identified through a statistically homogeneous indicator random field (Rubin & Journel, 1991). Building upon this framework, Rubin (1995) derives theoretical formulations for the first two statistical moments (i.e., mean and variance) and for the covariance structure of a bimodal (i.e., $M = 2$) Gaussian mixture. Lu and Zhang (2002) extend these formulations to embed a relationship linking the covariance of $Y(\mathbf{x})$ to a characteristic length associated with the observed spatial arrangement of the regions within which each of the random fields constituting the modes of the mixture are distributed within the spatial domain. This modeling framework has been further expanded by Siena et al. (2023). These authors provide analytical expressions for the PDF of spatial increments (ΔY) of a bimodal Gaussian mixture (hereafter termed GMIX). Increments are defined as differences of the variable of interest evaluated at spatial locations separated by a given (spatial) distance (or *lag*), $\mathbf{s}$, that is, $\Delta Y = Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{s})$. In this sense, the theoretical framework introduced by Siena et al. (2023) goes beyond the developments of Rubin (1995) and Lu and Zhang (2002) due to its inclusion under a unified theoretical framework of the behavior of Y, ΔY, and $q-th$ order statistical moments associated with these two quantities. As such, their formulation yields improved stochastic characterization of the main traits of the target spatial random field through the joint analysis of Y and ΔY.

Such a conceptual picture relying on the assumption that each mode can be modeled as a Gaussian distribution fails at accurately representing the stochastic behavior of Y in cases where any of the components of the mixture exhibits non-Gaussian traits. Typical evidences of non-Gaussian behaviors of Earth-system quantities include (a) the presence of sharp peaks and heavy tails characterizing their sample PDF and (b) a scaling tendency of the PDF of their spatial increments. The latter element manifests through a dependence of the PDF of incremental values on the lag at which increments are evaluated. These patterns are documented for numerous hydrological and hydrogeological variables, including, for example, permeability (Painter, 1996; Riva et al., 2013), porosity (Guadagnini et al., 2014, 2015; Painter, 1996), hydraulic conductivity (Guadagnini et al., 2013; Liu & Molz, 1997; Meerschaert et al., 2004), electrical resistivity (Yang et al., 2009), soil and sediment texture (Guadagnini et al., 2014), rainfall (Kumar & Foufoula-Georgiou, 1993), or sediment transport rate (Ganti et al., 2009). The above-mentioned features can be captured by the Generalized sub-Gaussian (GSG) model introduced by Riva et al. (2015). Such a theoretical model provides a joint description of the unimodal PDF of a variable of interest that is conceptualized as a randomly heterogeneous non-Gaussian field within a given region of a domain of interest as well as of its spatial increments evaluated therein. It embeds the Gaussian model as a particular case. Recent studies document the remarkable ability of the GSG model to interpret statistical behaviors observed across various spatial scales. Examples of application to a variety of spatial random fields include, for example, gas permeability of limestone cores (Siena et al., 2019), neutron porosity of deep vertical boreholes (Guadagnini et al., 2018), and (microscale) mineral surface roughness (Neuman et al., 2024; Stigliano et al., 2021) and dissolution rates (Siena et al., 2021).

The distinctive objective of our study is to extend the GMIX model of Siena et al. (2023) to include non-Gaussian modes described through the GSG model. The resulting Generalized sub-Gaussian Mixture (hereafter termed GSG-MIX) model is highly flexible. It enables one to treat as a unique population a field composed by diverse regions within which one can distinguish given system attributes and/or processes, each of these being characterized by a non-Gaussian nature. Thus, the GSG-MIX model enables one to capture (i) multiple sharp peaks and heavy tails of the sample PDF of the quantity of interest, together with (ii) scaling behaviors of the PDF of associated spatial increments. Our formulation is general and includes the above mentioned GSG and the GMIX models as special instances. We also design an operational workflow for the estimation of parameters embedded in the GSG-MIX model. The reliability of the latter is assessed upon relying on synthetically generated random fields.

We then employ the GSG-MIX model to characterize the statistical behavior of a novel experimental data set. The latter comprises a collection of spatially heterogeneous nanoscale dissolution rate maps evaluated from Atomic Force Microscopy (AFM) imaging of a calcite sample in contact with a continuously flowing fluid and subject to dissolution at far-from-equilibrium conditions. Our analysis yields an accurate characterization of spatially

heterogeneous mineral dissolution rates. Acquiring this ability is key to model processes associated with chemical weathering. The latter plays a major role in a variety of environmental and engineering scenarios as it drives, for example, contaminant release/trapping in natural rocks (Heberling et al., 2014; Hövelmann et al., 2018; Julia et al., 2023; Renard et al., 2018), changes in fracture aperture in deep reservoirs (Noiriel et al., 2007), or effectiveness of geogenic carbon (Daval, 2018; Noiriel & Daval, 2017) and nuclear waste (Ewing, 2015) strategies. Advanced high resolution imaging techniques reveal the coexistence of several local mechanisms driving the reaction across the mineral-fluid interface. The documented spatially heterogeneous strength of the actions of these mechanisms prevents the use of modeling frameworks grounded on average (bulk) rate values to provide accurate representations of material fluxes taking place at mineral-fluid interfaces (Lüttge et al., 2013). In this context, reliance upon a stochastic characterization of *rate spectra*, that is, sample PDFs of reaction rates, is suggested as a way to capture the variety of processes underpinning dissolution kinetics and giving rise to the intrinsic variability of observed rates (see, e.g., Bollermann & Fischer, 2020; Fischer et al., 2012, 2014). Typical patterns documented for rate spectra include (*i*) long right tailing of the sample PDF (Bibi et al., 2018; Brand et al., 2017); (*ii*) sharp peaks and heavy tails of sample frequency distributions (Siena et al., 2021); (*iii*) a scaling tendency of sample probability densities of associated spatial increments, their shape depending on the separation distance at which increments are evaluated (Siena et al., 2021). Multi-modal traits of sample PDFs of reaction rates, R, (Bollermann & Fischer, 2020; Fischer & Lüttge, 2017; Recalcati, Siena, et al., 2024; Siena et al., 2023) and their increments, ΔR, (Siena et al., 2023) are also documented when multiple processes contribute to drive the reaction within the observation domain. Preliminary efforts to model rate spectra rely on Generalized Extreme Value (GEV) (Brand et al., 2017; Emmanuel, 2014), GSG (Siena et al., 2021), and GMIX (Recalcati, Siena, et al., 2024; Siena et al., 2023) models. Whereas each of these models can grasp some of the aspects characterizing rate spectra, they do not incorporate all of them. The GEV formulation effectively captures tailing behaviors, while failing at interpreting multi-modal traits. Moreover, it only provides a statistical description of R without considering its spatial increments. Otherwise, the GSG model accurately describes the scaling behavior of ΔR and non-Gaussian features (i.e., peak and tails) of the PDF of R. Otherwise, it does not account for multi-modal attributes. Studies by Stigliano et al. (2021) and Neuman et al. (2024) document the effectiveness of the GSG model to characterize surface roughness of a calcite sample subject to dissolution at near-to-equilibrium conditions in the presence of a single mechanism driving mineral dissolution reaction. As opposed to the GEV and GSG formulations, the GMIX model has the ability to characterize multi-modal traits of the PDF of R and scaling tendencies of the PDFs of ΔR. However, Siena et al. (2023) observe some discrepancies between higher order moments of R and ΔR based on observed data and their theoretical GMIX-based counterparts. They attribute these differences to a possible non-Gaussian nature of the components of the mixture.

The GSG-MIX formulation we propose in this study offers a comprehensive conceptual and theoretical picture enabling one to capture all of the above-mentioned traits associated with the statistical behavior of a given variable and its spatial increments. Our study is structured as follows. Sections 2.1 and 2.2 illustrate our original theoretical formulation of the GSG-MIX model for a variable and its associated spatial increments, respectively. Section 2.3 presents the operational workflow tailored to the estimation of model parameters. Section 3 describes our experimental settings. Modeling results are discussed in Section 4. Section 5 outlines our conclusions.

2. Statistical Modeling Framework

2.1. Generalized Sub-Gaussian Mixture Model

We consider $Y(\mathbf{x})$ as a multi-modal spatial random field described according to (e.g., Lu & Zhang, 2002; Rubin, 1995)

$$Y(\mathbf{x}) = \sum_{m=1}^{M} I_m(\mathbf{x}) Y_m(\mathbf{x}), \tag{1}$$

where M is the number of independent and mutually exclusive modes, $Y_m(\mathbf{x})$, and $I_m(\mathbf{x})$ is an indicator random field independent of $Y_m(\mathbf{x})$ and following a Bernoulli distribution with parameter p_m. Focusing on a bimodal field (i.e., $M = 2$), Equation 1 reduces to

$$Y(\mathbf{x}) = I(\mathbf{x}) Y_A(\mathbf{x}) + [1 - I(\mathbf{x})] Y_B(\mathbf{x}), \tag{2}$$

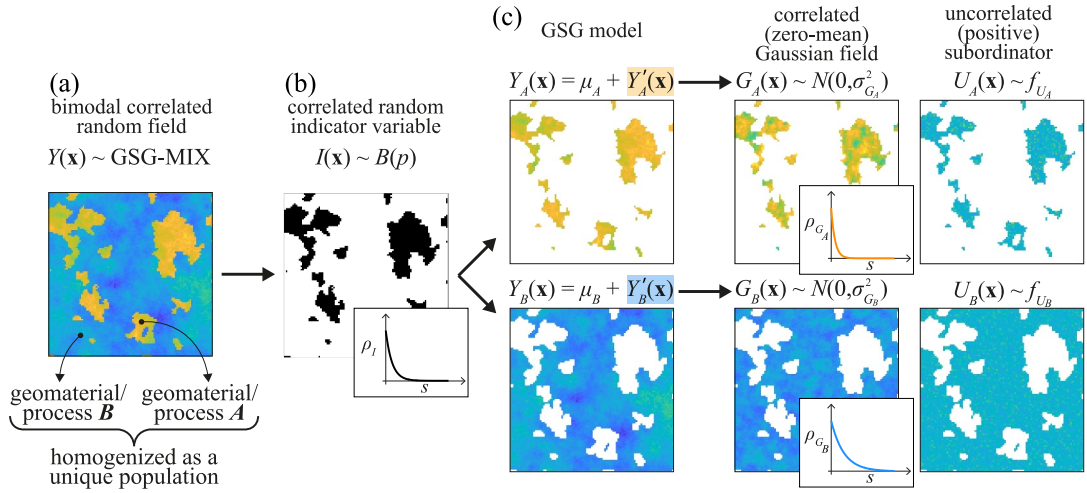

Figure 1. Graphical depiction of the conceptual framework underlying the mathematical developments encapsulated in Equations 1–7; (a) bimodal (spatially correlated) random field, $Y(\mathbf{x})$, distributed according to a GSG-MIX theoretical model; (b) spatially correlated random indicator variable $I(\mathbf{x})$ (distributed according to a Bernoulli distribution with parameter p, $B(p)$) describing the architecture of the two regions within the domain; (c) randomly heterogeneous spatial distribution of each component ($m = A, B$) of the GSG mixture across the domain.

where subscripts A and B denote the two components of the mixture, $I(\mathbf{x}) = I_A(\mathbf{x})$, and $I_B(\mathbf{x}) = 1 - I(\mathbf{x})$. The ensuing cumulative distribution function (CDF) and PDF of Y respectively read as

$$F_Y(y) = \Pr\{Y \le y\} = pF_A(y) + (1-p)F_B(y), \tag{3}$$

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial y} = pf_A(y) + (1-p)f_B(y). \tag{4}$$

Here, $p = p_A$ is the relative proportion of A across the domain and corresponds to the ensemble mean of the random indicator, that is, $p = \langle I(\mathbf{x}) \rangle$, $\langle \cdot \rangle$ denoting the expected value operator. We consider each mode $Y_m(\mathbf{x})$ of the mixture to be distributed according to a GSG model (Riva et al., 2015), that is,

$$Y_m(\mathbf{x}) = \mu_m + Y'_m(\mathbf{x}) \quad \text{with} \quad Y'_m(\mathbf{x}) = U_m(\mathbf{x})G_m(\mathbf{x}), \tag{5}$$

where μ_m is the mean of $Y_m(\mathbf{x})$, $G_m(\mathbf{x})$ is a zero-mean Gaussian random field, and $U_m(\mathbf{x})$, denoted as subordinator, is a positive spatially uncorrelated random function that is independent of $G_m(\mathbf{x})$. Thus, Y_m is distributed according to the following PDF (Riva et al., 2015)

$$f_{Y_m}(y) = f_{Y'_m}(y - \mu_m) = \frac{1}{\sqrt{2\pi}\sigma_{G_m}} \int_0^\infty f_{U_m}(u) e^{-\frac{(y-\mu_m)^2}{2\sigma^2_{G_m} u^2}} \frac{du}{u}, \tag{6}$$

where $\sigma^2_{G_m}$ is the variance of $G_m(\mathbf{x})$ and $f_{U_m}(u)$ corresponds to the distribution of $U_m(\mathbf{x})$. Substituting Equation 6 into Equation 4 yields

$$f_Y(y) = \frac{p}{\sqrt{2\pi}\sigma_{G_A}} \int_0^\infty f_{U_A}(u) e^{-\frac{(y-\mu_A)^2}{2\sigma^2_{G_A} u^2}} \frac{du}{u} + \frac{(1-p)}{\sqrt{2\pi}\sigma_{G_B}} \int_0^\infty f_{U_B}(u) e^{-\frac{(y-\mu_B)^2}{2\sigma^2_{G_B} u^2}} \frac{du}{u}. \tag{7}$$

A graphical depiction of the conceptual framework underlying the mathematical developments encapsulated in Equations 1–7 is offered in Figure 1.

The GSG-MIX model rendered by Equation 7 can be considered as a generalization of the GMIX model illustrated in Siena et al. (2023), as each of its modes corresponds to a Gaussian distribution when U_m is a deterministic quantity (i.e., f_{U_m} tends to a Dirac delta function).

Raw moments of order q of Y can be evaluated by making use of Equation 7 as

$$\langle Y^q \rangle = p\langle Y_A^q \rangle + (1-p)\langle Y_B^q \rangle. \tag{8}$$

Setting $q = 1$ in Equation 8 yields the mean of Y, that is,

$$\mu_Y = p\mu_A + (1-p)\mu_B. \tag{9}$$

The latter varies linearly with p. Central moments of Y can be evaluated as

$$\langle Y'^q \rangle = \langle (Y-\mu_Y)^q \rangle = \sum_{j=0}^{q} \binom{q}{j} (-1)^j \mu_Y^j \langle Y^{q-j} \rangle. \tag{10}$$

Variance, σ_Y^2, skewness, Sk_Y, and kurtosis, κ_Y, of Y respectively read

$$\sigma_Y^2 = p\sigma_{G_A}^2\langle U_A^2 \rangle + (1-p)\sigma_{G_B}^2\langle U_B^2 \rangle + p(1-p)(\mu_A-\mu_B)^2, \tag{11}$$

$$Sk_Y = \frac{p(1-p)}{\sigma_Y^3}(\mu_A-\mu_B)\left[(1-2p)(\mu_A-\mu_B)^2 + 3\left(\sigma_{G_A}^2\langle U_A^2 \rangle - \sigma_{G_B}^2\langle U_B^2 \rangle\right)\right], \tag{12}$$

$$\begin{aligned}\kappa_Y = \frac{1}{\sigma_Y^4}\Big\{ & 3\left[p\left(\sigma_{G_A}^4\langle U_A^4 \rangle - \sigma_{G_B}^4\langle U_B^4 \rangle\right) + \sigma_{G_B}^4\langle U_B^4 \rangle\right] + p(1-p)(\mu_A-\mu_B)^2 \\ & \left[(1-3p(1-p))(\mu_A-\mu_B)^2 + 6\left(\sigma_{G_A}^2\langle U_A^2 \rangle - p\left(\sigma_{G_A}^2\langle U_A^2 \rangle - \sigma_{G_B}^2\langle U_B^2 \rangle\right)\right)\right]\Big\},\end{aligned} \tag{13}$$

where $\langle U_m^q \rangle$ is the $q-th$ order raw moment of the subordinator associated with each $m-th$ mode of the mixture. The variance displays a quadratic dependence on p and attains a maximum at $p = (1+\beta)/2$, with $\beta = \left(\sigma_{G_A}^2\langle U_A^2 \rangle - \sigma_{G_B}^2\langle U_B^2 \rangle\right)/(\mu_A-\mu_B)^2$. Equations 12 and 13 indicate that the PDF of the mixture (i) is non-symmetric ($Sk_Y \neq 0$), even as each GSG mode is symmetric (i.e., $Sk_{Y_m} = 0$) and (ii) can be leptokurtic ($\kappa_Y > 3$) or platikurtic ($\kappa_Y < 3$), depending on model parameters. The shape of a GSG-MIX distribution depends on p and on the statistical parameters characterizing each mode Y_m, such as the difference between the component means, $\mu_A - \mu_B$, and the variance $\sigma_{G_m}^2$ ($m = A, B$), and on the PDF of the subordinator associated with each mode of the mixture.

Figure 2 provides an appraisal of the impact of the values of model parameters on the main traits of $f_Y(y)$. For illustration purposes, here we set $\mu_A = 3$, $\mu_B = 0$, and consider $U_m(\mathbf{x})$ as lognormally distributed, that is, $U_m \sim \log N\left(0,(2-\alpha_m)^2\right)$, $\alpha_m \in (0,2)$ (see Appendix A). A GSG-MIX formulation embedding such a distributional form of the subordinator tends to a GMIX model if $\alpha_m \to 2$, that is, each mode of the mixture is described by a Gaussian distribution. Otherwise, non-Gaussian traits of each component are heightened as α_m departs from 2. Previous applications of a unimodal GSG model relying on a lognormal subordinator yield accurate interpretations of the statistical behavior of environmentally relevant variables such as, for example, porosity (Guadagnini et al., 2018; Riva et al., 2015), electrical resistivity (Li et al., 2022), gas permeability (Siena et al., 2019), mineral surface roughness (Siena et al., 2020) and dissolution rate (Siena et al., 2021). In the latter cases, documented values for α_m are in the range 1.3−1.9. Figure 2 depicts the PDF of Y evaluated for different values of the proportion coefficient, p, and α_m while keeping $\sigma_{Y_A}^2$ and $\sigma_{Y_B}^2$ (with $\sigma_{Y_m}^2 = \sigma_{G_m}^2\langle U_m^2 \rangle$) constant. GMIX distributions (dashed curves) having the same mean and variance as the GSG-MIX model are also shown. Similar to their GMIX counterparts, GSG-MIX PDFs are non-symmetric and exhibit two peaks (located around μ_A and μ_B) and a local minimum comprised therein. Increasing the non-Gaussianity of the Y_m modes (i.e., decreasing α_m) leads to an increase of the heaviness of the tails of the resulting PDF of Y as well as of the magnitude of the two peaks (with a corresponding decrease of the local minimum). These results support the flexibility of the GSG-

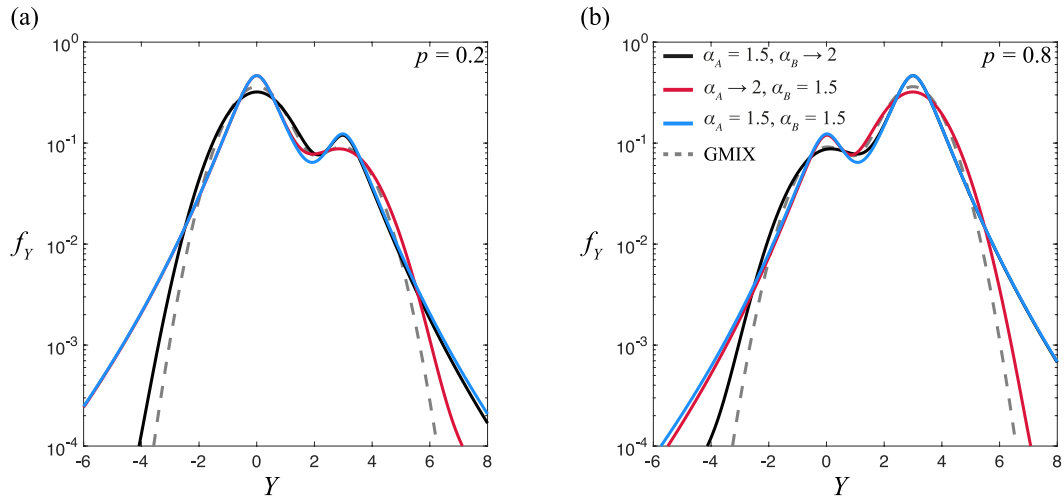

Figure 2. Probability density function, $f_Y(y)$, resulting from the GSG-MIX model evaluated considering a lognormal subordinator (Equation A2) upon setting $\mu_A = 3$, $\mu_B = 0$, $\sigma^2_{Y_A} = \sigma^2_{Y_B} = 1$, various combinations of α_m, and (a) $p = 0.2$, (b) $p = 0.8$ (continuous curves). Analytical expressions evaluated according to the GMIX model having the same mean and variance of their GSG-MIX counterparts are also depicted (dashed curves).

MIX model to describe the observed statistical behaviors of numerous environmental variables characterized by the presence of multiple sharp peaks and asymmetric long tails.

2.2. Spatial Increments of a Generalized Sub-Gaussian Mixture

Let Y_1 and Y_2 correspond to the GSG-MIX field at two spatial (vector) locations $\mathbf{x}_1$ and $\mathbf{x}_2$, that is,

$$Y_i = Y(\mathbf{x}_i) = I(\mathbf{x}_i)\, Y_A(\mathbf{x}_i) + [1 - I(\mathbf{x}_i)]\, Y_B(\mathbf{x}_i), \tag{14}$$

with $i = 1, 2$. The bivariate PDF of Y_1 and Y_2 reads (Siena et al., 2023)

$$\begin{aligned} f_{Y_1 Y_2}(y_1, y_2) = {} & \left[p^2 + C_I(\mathbf{x}_1, \mathbf{x}_2)\right] f_{Y_{A_1} Y_{A_2}}(y_1, y_2) \\ & + \left[(1-p)^2 + C_I(\mathbf{x}_1, \mathbf{x}_2)\right] f_{Y_{B_1} Y_{B_2}}(y_1, y_2) \\ & + \left[p(1-p) - C_I(\mathbf{x}_1, \mathbf{x}_2)\right] \left[f_{Y_{A_1}}(y_1) f_{Y_{B_2}}(y_2) + f_{Y_{B_1}}(y_1) f_{Y_{A_2}}(y_2)\right], \end{aligned} \tag{15}$$

where $C_I(\mathbf{x}_1, \mathbf{x}_2)$ is the spatial covariance of the indicator and $f_{Y_{m_1}, Y_{m_2}}(y_1, y_2)$ is the GSG bivariate PDF of the $m-th$ component of the mixture. The latter is defined as (Riva et al., 2015)

$$f_{Y_{m_1}, Y_{m_2}}(y_1, y_2) = \int_0^{+\infty} \int_0^{+\infty} f_{U_{m_1}}(u_1) f_{U_{m_2}}(u_2) f_{G_{m_1}, G_{m_2}}\left(\frac{y_1}{u_1}, \frac{y_2}{u_2}\right) \frac{du_1}{u_1} \frac{du_2}{u_2}. \tag{16}$$

Here, $f_{G_{m_1}, G_{m_2}}$ is the Gaussian bivariate PDF of the underlying G_m field, that is,

$$f_{G_{m_1}, G_{m_2}}\left(\frac{y_1}{u_1}, \frac{y_2}{u_2}\right) = \frac{1}{2\pi\, \sigma^2_{G_m} \sqrt{1 - \rho^2_{G_m}}} e^{-\frac{\left(\frac{y_1 - \mu_m}{u_1}\right)^2 + \left(\frac{y_2 - \mu_m}{u_2}\right)^2 - 2\rho_{G_m} \frac{(y_1 - \mu_m)(y_2 - \mu_m)}{u_1 u_2}}{2\sigma^2_{G_m}(1 - \rho^2_{G_m})}}, \tag{17}$$

where ρ_{Gm} is the correlation function of G_m. We assume second-order stationarity and isotropy of all fields, that is, $C_I(\mathbf{x_1}, \mathbf{x_2}) = C_I(s)$ and $\rho_{G_m}(\mathbf{x_1}, \mathbf{x_2}) = \rho_{G_m}(s)$, $s = |\mathbf{x}_1 - \mathbf{x}_2|$ being the lag at which spatial increments are evaluated. The PDF of increments, $\Delta Y(s) = Y_1 - Y_2$, can be obtained from the corresponding CDF, $F_{\Delta Y}(\Delta y)$. The latter reads

$$F_{\Delta Y}(\Delta y) = \Pr\{\Delta Y \le \Delta y\} = \int_{y_2=-\infty}^{+\infty} \int_{y_1=-\infty}^{\Delta y + y_2} f_{Y_1,Y_2}(y_1, y_2)\, dy_1 dy_2. \tag{18}$$

Recalling that $f_{\Delta Y}(\Delta y) = dF_{\Delta Y}(\Delta y)/d(\Delta y)$ and making use of Equation 15 yields

$$\begin{aligned} f_{\Delta Y}(\Delta y) = {} & \frac{p^2 + C_I(s)}{\sqrt{2\pi}\sigma_{G_A}} \int_0^{+\infty} \int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{A_2}}(u_2)}{r} e^{-\frac{\Delta y^2}{2r^2\sigma_{G_A}^2}} du_2 du_1 \\ & + \frac{(1-p)^2 + C_I(s)}{\sqrt{2\pi}\sigma_{G_B}} \int_0^{+\infty} \int_0^{+\infty} \frac{f_{U_{B_1}}(u_1) f_{U_{B_2}}(u_2)}{r} e^{-\frac{\Delta y^2}{2r^2\sigma_{G_B}^2}} du_2 du_1 \\ & + \frac{p(1-p) - C_I(s)}{\sqrt{2\pi}} \left[\int_0^{+\infty} \int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{B_2}}(u_2)}{h} e^{-\frac{(\Delta y - \mu_A + \mu_B)^2}{2h^2}} du_2 du_1 \right. \\ & \left. + \int_0^{+\infty} \int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{B_2}}(u_2)}{w} e^{-\frac{(\Delta y + \mu_A - \mu_B)^2}{2w^2}} du_2 du_1 \right], \end{aligned} \tag{19}$$

with $r = \sqrt{u_1^2 + u_2^2 - 2\rho_{G_m} u_1 u_2}$, $h = \sqrt{\sigma_{G_A}^2 u_1^2 + \sigma_{G_B}^2 u_2^2}$, and $w = \sqrt{\sigma_{G_B}^2 u_1^2 + \sigma_{G_A}^2 u_2^2}$.

All statistical moments of ΔY can be evaluated from Equation 19. In particular, mean and odd-order moments of ΔY vanish. Second- and fourth-order moment respectively read

$$\begin{aligned} \langle \Delta Y^2 \rangle = {} & 2\{p^2 \sigma_{G_A}^2 \left(\langle U_A^2 \rangle - \langle U_A \rangle^2 \rho_{G_A}\right) + (1-p)^2 \sigma_{G_B}^2 \left(\langle U_B^2 \rangle - \langle U_B \rangle^2 \rho_{G_B}\right) \\ & + p(1-p)\left[(1-\rho_I)(\mu_A - \mu_B)^2 + \sigma_{G_A}^2 \left(\langle U_A^2 \rangle - \langle U_A \rangle^2 \rho_I \rho_{G_A}\right)\right. \\ & \left. + \sigma_{G_B}^2 \left(\langle U_B^2 \rangle - \langle U_B \rangle^2 \rho_I \rho_{G_B}\right)\right]\}, \end{aligned} \tag{20}$$

$$\begin{aligned} \langle \Delta Y^4 \rangle = {} & 6\Big\{p^2 \sigma_{G_A}^4 \left[\langle U_A^4 \rangle - 4\langle U_A \rangle \langle U_A^3 \rangle \rho_{G_A} + \langle U_A^2 \rangle^2 \left(1 + 2\rho_{G_A}^2\right)\right] \\ & + (1-p)^2 \sigma_{G_B}^4 \left[\langle U_B^4 \rangle - 4\langle U_B \rangle \langle U_B^3 \rangle \rho_{G_B} + \langle U_B^2 \rangle^2 \left(1 + 2\rho_{G_B}^2\right)\right] \\ & + p(1-p)\rho_I \Big[\sigma_{G_A}^4 \left(\langle U_A^4 \rangle - 4\langle U_A \rangle \langle U_A^3 \rangle \rho_{G_A} + \langle U_A^2 \rangle^2 \left(1 + 2\rho_{G_A}^2\right)\right) \\ & + \sigma_{G_B}^4 \left(\langle U_B^4 \rangle - 4\langle U_B \rangle \langle U_B^3 \rangle \rho_{G_B} + \langle U_B^2 \rangle^2 \left(1 + 2\rho_{G_B}^2\right)\right)\Big]\Big\} \\ & + 2p(1-p)(1-\rho_I)\Big[(\mu_A - \mu_B)^4 + 6(\mu_A - \mu_B)^2 \left(\sigma_{G_A}^2 \langle U_A^2 \rangle + \sigma_{G_B}^2 \langle U_B^2 \rangle\right) \\ & + 3\left(\sigma_{G_A}^4 \langle U_A^4 \rangle + \sigma_{G_B}^4 \langle U_B^4 \rangle + 2\sigma_{G_A}^2 \sigma_{G_B}^2 \langle U_A^2 \rangle \langle U_B^2 \rangle\right)\Big], \end{aligned} \tag{21}$$

where $\rho_I = C_I/\sigma_I^2$ ($\sigma_I^2 = p(1-p)$ being the variance of I) is the (spatial) correlation function of $I(\mathbf{x})$. The analytical expression for the kurtosis of incremental values, $\kappa_{\Delta Y} = \langle \Delta Y^4 \rangle / \langle \Delta Y^2 \rangle^2$, can be readily obtained from Equations 20 and 21.

Recalling that $C_Y = \sigma_Y^2 - \gamma_Y$ ($\gamma_Y = \langle \Delta Y^2 \rangle / 2$ being the variogram of Y), the covariance of Y is obtained from Equation 20 as

$$C_Y(s) = \begin{cases} \sigma_Y^2 & \text{if } s = 0 \\ p^2 C_{Y_A} + (1-p)^2 C_{Y_B} + C_I \left[(\mu_A - \mu_B)^2 + C_{Y_A} + C_{Y_B}\right] & \text{if } s > 0 \end{cases} \tag{22}$$

Here, $C_{Y_m} = \sigma_{G_m}^2 \langle U_m \rangle^2 \rho_{G_m}$ ($m = A, B$) is the covariance of the $m-th$ component of the mixture. The covariance C_Y is discontinuous at the origin, that is, it exhibits a nugget effect

$$\nu = p\sigma_{G_A}^2\sigma_{U_A}^2 + (1-p)\sigma_{G_B}^2\sigma_{U_B}^2. \tag{23}$$

Here, $\sigma_{U_m}^2 = \langle U_m^2\rangle - \langle U_m\rangle^2$ $(m = A,B)$ is the variance of $U_m(\mathbf{x})$. Hence, the variogram of Y, γ_Y, can be evaluated as

$$\begin{aligned}\gamma_Y &= \nu + p^2\langle U_A\rangle^2\gamma_{G_A} + (1-p)^2\langle U_B\rangle^2\gamma_{G_B} + \left(\mu_A-\mu_B\right)^2\gamma_I \\ &\quad + p(1-p)\left[\langle U_A\rangle^2\sigma_{G_A}^2\left(1-\rho_I\rho_{G_A}\right) + \langle U_B\rangle^2\sigma_{G_B}^2\left(1-\rho_I\rho_{G_B}\right)\right],\end{aligned} \tag{24}$$

where $\gamma_{G_m} = \sigma_{G_m}^2\left(1-\rho_{G_m}\right)$ is the variogram associated with the Gaussian field underlying each component of the mixture; and $\gamma_I = p(1-p)\left(1-\rho_I\right)$ is the variogram of the indicator field. The integral scale of the mixture, I_Y, is evaluated upon integrating the spatial correlation $\rho_Y = C_Y/\sigma_Y^2$. This yields

$$\begin{aligned}I_Y = \frac{1}{\sigma_Y^2}\Big\{&p^2\sigma_{G_A}^2\langle U_A\rangle^2\lambda_{G_A} + (1-p)^2\sigma_{G_B}^2\langle U_B\rangle^2\lambda_{G_B} \\ &+ p(1-p)\left[\left(\mu_A-\mu_B\right)^2\lambda_I + \sigma_{G_A}^2\langle U_A\rangle^2\lambda_{I-A}^{eq} + \sigma_{G_B}^2\langle U_B\rangle^2\lambda_{I-B}^{eq}\right]\Big\}.\end{aligned} \tag{25}$$

Here, λ_{G_m} $(m = A,B)$ and λ_I are the correlation lengths of G_m and of I, respectively; $\lambda_{I-m}^{eq} = \int_0^\infty \rho_I\rho_{G_m}ds$ is a representative length scale. The latter can be viewed as a measure of the way the strength of the correlation of G_m is modulated by the action of the correlation of the random indicator field (or vice versa) in a representation of Y as a bimodal random field.

Figures 3a–3c illustrate the effect of model parameters on the main traits of the PDF of ΔY. As in Section 2.1, we assume a lognormal distribution for the subordinator and set $\mu_A = 3, \mu_B = 0$. For the purpose of our illustration, we set $p = 0.2$, $\sigma_{Y_A}^2 = \sigma_{Y_B}^2 = 1$, and consider an exponential model for ρ_{G_m} and ρ_I (i.e., $\rho_{G_m} = e^{-s/\lambda_{G_m}}$ and $\rho_I = e^{-s/\lambda_I}$). We set $\lambda_{G_A} = 10$, $\lambda_{G_B} = 20$, and $\lambda_I = 7$. GMIX counterparts characterized by the same mean, variance, and correlation structure are also depicted (dashed curves). The height of the central peak of $f_{\Delta Y}$ is controlled by the first two terms in Equation A9. The third and fourth term in Equation A9 otherwise govern the height of the two secondary peaks. These terms are in turn weighted through the covariance of the indicator, C_I, which drives the relative proportion between the central and the lateral peaks. As s increases, the relative importance of the secondary peaks increases at the expenses of the dominant peak. At all lags, the central peak associated with the GSG-MIX model is less pronounced than its GMIX counterpart, the lateral peaks being otherwise heightened. Figure 3d depicts analytical variograms of Y evaluated through Equation A14 for various combinations of α_A and α_B. Consistent with Equations A15 and A16, increasing the degree of non-Gaussianity of each component of the mixture (i.e., decreasing α_m) yields a decrease of the integral scale of the mixture and an increase of the nugget effect. Otherwise, the latter vanishes if the two components are Gaussian (dashed curve). We further note that at small lags $\kappa_{\Delta Y}$ (Figure 3e) tends to increase as the two components of the mixture tend to a Gaussian (i.e., $\alpha_m \to 2$). In contrast, the asymptotic value attained by $\kappa_{\Delta Y}$ at large values of s increases as the two modes depart from the Gaussian case.

2.3. Estimation of Model Parameters

Estimation of model parameters relies on a custom algorithm grounded on a Bayesian classification approach (Duda et al., 2000; James et al., 2013). For illustration purposes, we describe in the following the estimation procedure considering the set of GSG-MIX formulations derived in Appendix A and corresponding to a lognormal subordinator. Prior to application to experimental observations, the performance of the designed operational scheme is assessed on synthetic data sets described in Appendix B. These are generated to resemble typical geometrical patterns observed for experimental dissolution rate data at the mineral-fluid interface (see Section 4). A schematic representation of the application of the parameter estimation workflow to a synthetic GSG-MIX field is offered in Figure 4.

The methodology requires an initial parameter set. This can be obtained upon relying on a preliminary classification, $I^{(0)}(\mathbf{x})$, of the field based on our prior knowledge of the spatial distribution of the regions within the

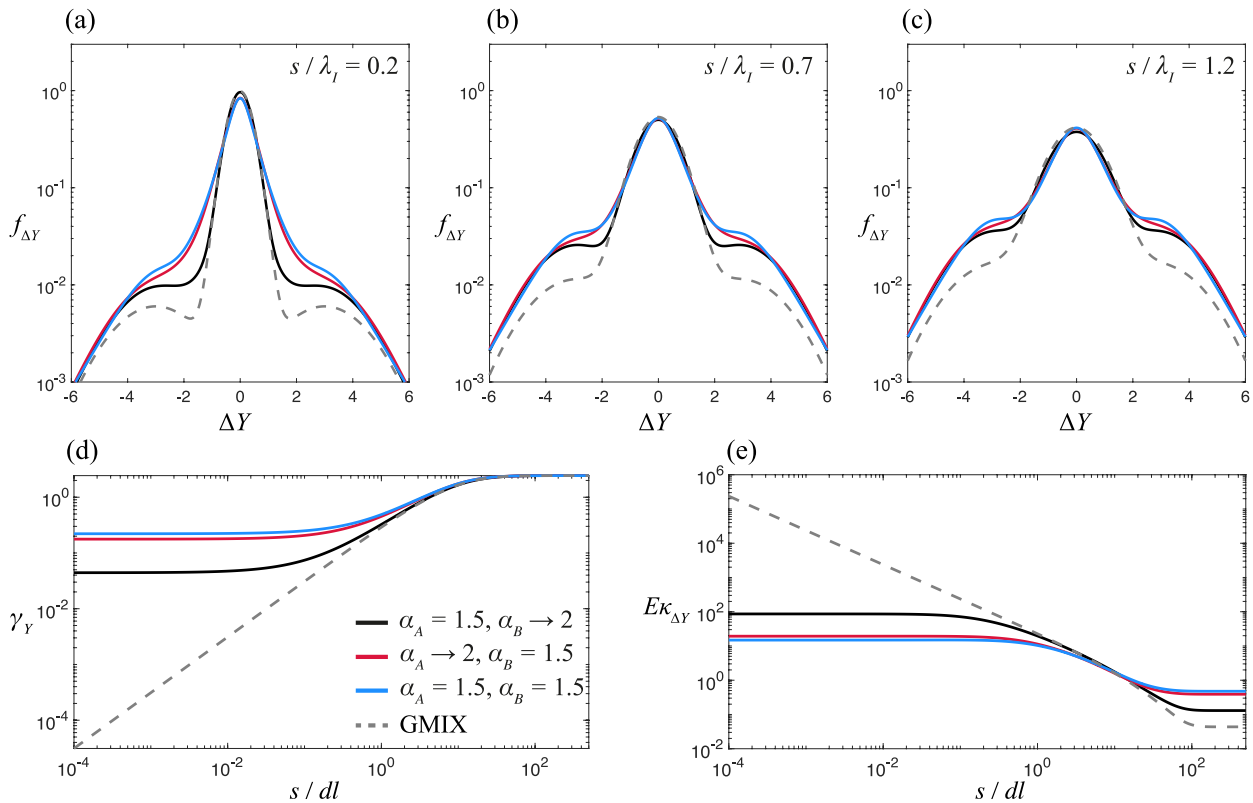

Figure 3. Probability density function, $f_{\Delta Y}(\Delta y)$, resulting from the GSG-MIX model evaluated considering a lognormal subordinator (Equation A9) upon setting $p = 0.2$, $\mu_A = 3$, $\mu_B = 0$, $\sigma^2_{Y_A} = \sigma^2_{Y_B} = 1$, various combinations of α_m, $\lambda_{G_A} = 10$, $\lambda_{G_B} = 20$, and $\lambda_I = 7$ for (a) $s/\lambda_I = 0.2$, (b) $s/\lambda_I = 0.7$, and (c) $s/\lambda_I = 1.2$. Associated (d) variogram, γ_Y, and (e) excess kurtosis, $E\kappa_{\Delta Y} = \kappa_{\Delta Y} - 3$, are depicted as a function of lag. Analytical expressions evaluated according to the GMIX model having the same mean and variance of their GSG-MIX counterparts are also depicted (dashed curves).

observation window (see Figure 4a). Note that such preliminary classification is (generally) affected by error (see also Appendix B). We then set a random path visiting all $k = 1, \dots, N_c$ cells, N_c being the total number of data points (Figure 4b). Relying on a Bayesian classifier, at each step k we assign a given observation $y = Y(\mathbf{x}_k)$ to category m ($m = A, B$) by maximizing the discriminant function, $g_m(y)$ (see also Figure 4c)

$$g_m(y) = p(y|m)P(m), \tag{26}$$

where $p(y|m)$ is the likelihood of y given m and $P(m)$ is the prior probability of m. In our setting we assign category m to location $\mathbf{x}_k$ according to

$$I(\mathbf{x}_k) = \begin{cases} I_A = 1 \text{ if } g_A \geq g_B \\ I_B = 1 \text{ if } g_A < g_B, \end{cases} \tag{27}$$

where, making use of Equation 6,

$$g_m(y) = \frac{P(m)}{\sqrt{2\pi}\sigma_{G_m}} \int_0^\infty f_{U_m}(u) e^{-\frac{(y-\mu_m)^2}{2\sigma^2_{G_m} u^2}} \frac{du}{u}, \tag{28}$$

with $P(A) = p$ and $P(B) = 1-p$. The estimate, $\hat{\xi}$, of each parameter embedded in Equation A2 (i.e., $\xi = p, \mu_m, \sigma_{G_m}, \alpha_m$, with $m = A, B$) at step k is evaluated upon relying on the method of moments as detailed below. As already mentioned, the methodology begins by defining an initial parameter set. To ensure a solution

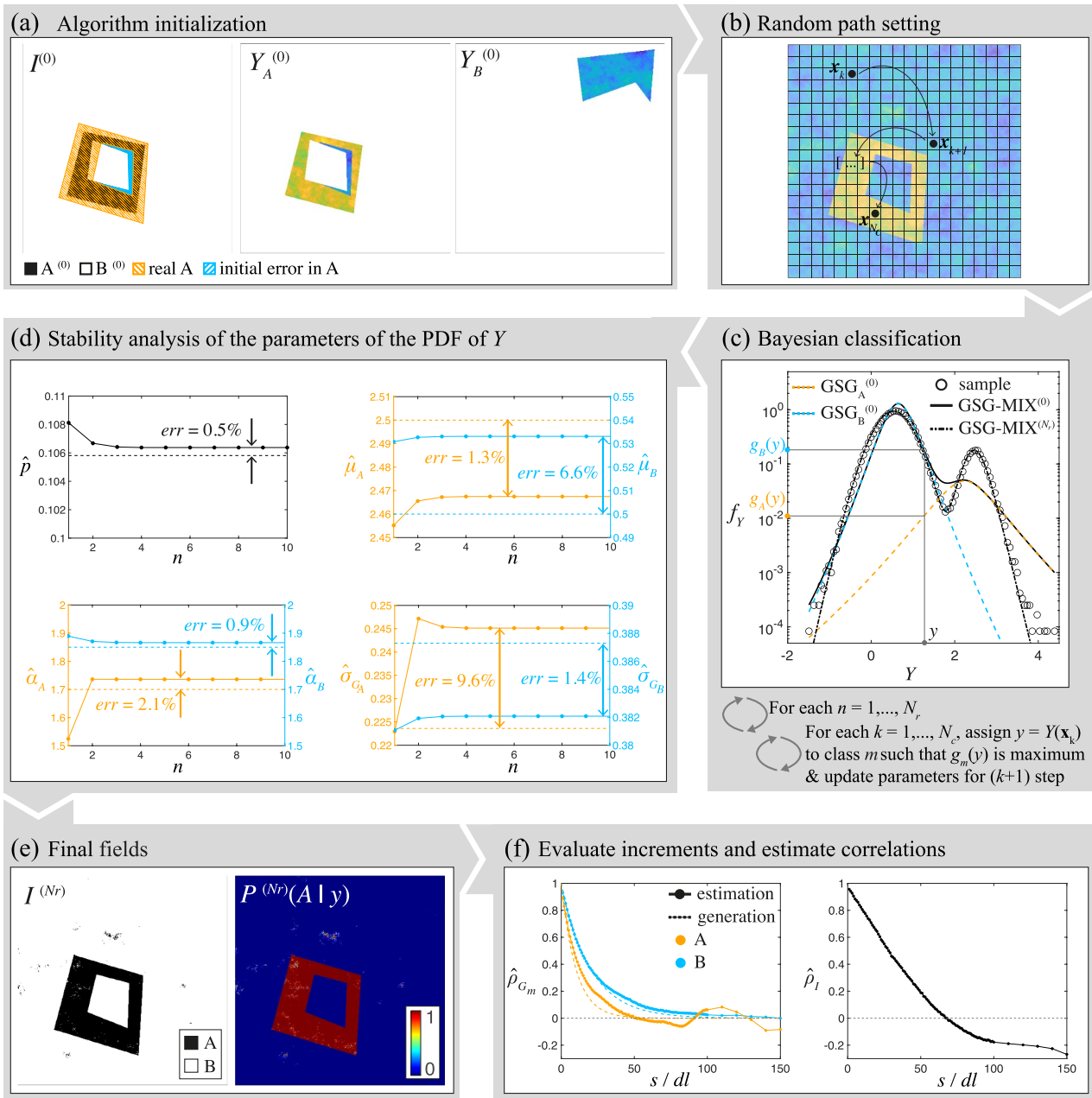

Figure 4. Schematic representation of the parameter estimation algorithm resting upon a Bayesian classification approach. The estimation workflow includes (a) a preliminary classification of the field for algorithm initialization; (b) setting a random path visiting all N_c cells; (c) assigning each cell to a region and updating of model parameters for the next step; (d) replication of the classification algorithm until stabilization of parameter values is achieved; (e) evaluation of the posterior probability of the indicator associated with the last repetition, $P^{(N_r)}(A|\,y)$; and (f) evaluation of correlation functions $\hat{\rho}_{G_m}$ and $\hat{\rho}_I$.

that is independent of the initial parameters, the procedure is repeated n times, with $n = 1, \ldots, N_r$. Starting parameter values at repetition $n - th$ are those estimated with the $(n-1)-th$ repetition. A new random path is selected at each repetition. Estimates of parameters at step k during repetition n, $\hat{\xi}^{(k,n)}$, are obtained by setting (i) the ensemble mean of I, p, equal to its sample counterpart, $M_{I,1}^{(k,n)}$, (ii) the ensemble mean of Y_m, μ_m, equal its sample counterpart, $M_{m,1}^{(k,n)}$, and (iii) replacing the second- and fourth-order central moments of Y_m in Equations A7 and A8 by their sample counterparts, $M_{m,2}^{(k,n)}$ and $M_{m,4}^{(k,n)}$, respectively. This procedure leads to the following system of equations

$$\begin{cases} \hat{p}^{(k,n)} = M_{I,1}^{(k,n)}, \\ \hat{\mu}_m^{(k,n)} = M_{m,1}^{(k,n)}, \\ \hat{\alpha}_m^{(k,n)} = 2 - \frac{1}{2}\sqrt{\ln \frac{M_{m,4}^{(k,n)}}{3(M_{m,2}^{(k,n)})^2}}, \\ \hat{\sigma}_{G_m}^{(k,n),2} = \left(M_{m,2}^{(k,n)}\right)^2 \sqrt{\frac{3}{M^{(k,n)}{}_{m,4}}}, \end{cases} \tag{29}$$

The value of N_r is obtained when $\hat{p}$ stabilizes (Figure 4d, see also Appendix B). The convergence rate of $\hat{p}$ is case-specific and is mainly influenced by the shape of the PDF of the mixture as well as by the available sample size. Enhanced performance is observed when the peaks of the mixture are clearly identifiable and well separated. Then, we evaluate the spatial increments associated with each of the two regions identified at the last repetition, that is, $\Delta Y_m^{(N_r)}$ $(m = A, B)$. The correlation function of the underlying Gaussian field of each mode, $\hat{\rho}_{G_m}$, is estimated by replacing $\langle \Delta Y_m^2 \rangle$ in Equation A12 by its sample counterpart, $M_2^{\Delta Y_m^{(N_r)}}$, that is,

$$\hat{\rho}_{G_m} = \frac{1}{e^{(2-\hat{\alpha}_m)^2}} \left[e^{2(2-\hat{\alpha}_m)^2} - \frac{M_2^{\Delta Y_m^{(N_r)}}}{2\hat{\sigma}_{G_m}^2} \right], \tag{30}$$

Finally, we evaluate the sample variance of increments of the indicator field associated with the last repetition, $M_2^{\Delta I^{(N_r)}}$, and estimate the correlation function of I as (Figure 4f)

$$\hat{\rho}_I = 1 - \frac{M_2^{\Delta I^{(N_r)}}}{2\hat{p}(1-\hat{p})}. \tag{31}$$

The Bayesian classifier also provides as output the (posterior) probability of finding class m at any given position in the domain as (see Figure 4e)

$$P(m|\, y) = \frac{p(y|m)P(m)}{p(y)} = \frac{g_m(y)}{p(y)}, \tag{32}$$

where $p(y)$ is evaluated via Equation A2 (Figure 4e).

3. Experimental Setup and Evaluation of Dissolution Rate Maps

Spatial maps of calcite dissolution rates are obtained from direct investigations of the topography of a crystal sample subject to reaction under continuous flowing conditions. Measurements are performed at the nanoscale upon relying on AFM imaging. The sample is freshly cleaved from a calcite crystal of Iceland Spar (Mexico) along the {104} plane and then glued onto a glass slide. The latter is mounted on a support plate that is magnetically attached to the AFM, enabling in situ and real-time imaging of the mineral-fluid interface. A fluid cell (volume $V \sim 2$ mL) is screwed to the support plate and sealed with a Viton O-ring. The cell is connected to a high precision syringe pump (Cetoni, Nemesys S) employed to regulate the fluid flow in the system. The flow rate is set to $Q = 10$ μL s^{-1}. This value ensures that chemical conditions are constant for the entire duration of the experiment, while limiting imaging distortions associated with fluid flow (Recalcati et al., 2023). A 6×6 μm^2 portion of the crystal surface is imaged across a 512×512 horizontal grid (pixel size $dl = 11.7$ nm). Imaging is performed in tapping mode by employing silicon tips with an aluminum coated cantilever (Nanosensors, PPP-NCHR, $k = 42$ Nm^{-1}). The scan rate is set to 1.28 Hz, leading to an acquisition time of $\Delta t = 400$s ($\sim$6.6 min).

Dissolution of the sample is induced at ambient conditions (i.e., $p = p_{atm}$ and $T = 22$°C) by using Milli-Q water. Under such chemical conditions, calcite dissolves by forming rhombohedral mono- and multilayer etch pits (see, e.g., Bouissonnié et al., 2018; Teng, 2004). The former are shallow ephemeral surface features that can nucleate randomly on crystal terraces or at point defects and have a depth of a single crystal layer (Harstad & Stipp, 2007).

The latter arise in the presence of screw dislocations and spread laterally and vertically, removing several layers of the lattice. At such locations, the material flux is enhanced by an excess of strain of the crystal lattice that results in the radiation of chains of steps or *stepwaves* (Lasaga & Lüttge, 2001, 2003; Lüttge et al., 2019). These then travel across the crystal surface inducing a layer-by-layer removal of the material. Recent work by Fischer and Lüttge (2018) provides experimental evidence that stepwave radiation at screw dislocations follows a pulsating mechanism. The complex interaction and/or superposition of these trains of steps yields markedly heterogeneous spatial distributions of reaction rates, $R(\mathbf{x},t^*)$ [mmol m^{-2}s^{-1}], across the mineral surface. These are obtained from differences of topographies, $z(\mathbf{x},t)$ [m], measured at subsequent times as (e.g., Bouissonnié et al., 2018; Emmanuel, 2014)

$$R\left(\mathbf{x},t_j^*\right)=\frac{z(\mathbf{x},t_i)-z(\mathbf{x},t_i+\Delta t)}{V_m\Delta t}, \tag{33}$$

with $t_j^*=t_i+\Delta t$, $V_m=36.93\cdot10^{-9}$ m^3 mmol^{-1} being calcite molar volume.

Similar to Siena et al. (2021), we express the reaction rate as the sum of a mean spatial value, $\bar{R}\left(t_j^*\right)=[\bar{z}(t_i)-\bar{z}(t_i+\Delta t)]/(V_m\Delta t)$, and a local zero-mean fluctuation, $R'\left(\mathbf{x},t_j^*\right)=[z'(\mathbf{x},t_i)-z'(\mathbf{x},t_i+\Delta t)]/(V_m\Delta t)$. The former represents the average material flux across the mineral surface, whereas the latter is informative about the spatial variability of reaction rates. Pre-processing is generally required to recover the actual topography, $z(\mathbf{x},t)$, from raw AFM data, $z_{meas}(\mathbf{x},t)$. These are often affected by lateral drift due to thermal expansions of the piezoelectric scanner (see, e.g., Rahe et al., 2010) and/or by a bowing effect associated with the scanning mechanism (Ricci & Braga, 2004). The former may cause a misalignement of topographic signals at different times. This effect is corrected upon registering all images with respect to fixed points. We identify the latter as small precipitates spontaneously forming on the surface during the reaction. Such image registration procedure restricts our observation window to 5.3 × 5.3 μm^2 (corresponding to 450 × 450 cells). The bowing effect is removed upon subtracting the best fitting second-order polynomial to $z_{meas}(\mathbf{x},t)$. As detailed in Siena et al. (2021), this pre-processing phase also removes the average value of $z_{meas}(\mathbf{x},t)$, $\bar{z}(t)$. Hence, we assess spatial maps of fluctuation of the topography about $\bar{z}(t)$, $z'(\mathbf{x},t)$, and, in turn, of the reaction rate, $R'(\mathbf{x},t^*)$. This enables us to provide insights on the spatial variability of rates.

4. Results and Discussion

Figure 5 depicts spatial maps of fluctuations of topography, $z'(\mathbf{x},t_i)$, measured at various times, t_i ($i=1,\ldots,7$). At all instants sampled during our experiment, we observe an evolution of the surface pattern driven by the mechanisms described in Section 3, that is, step retreat and nucleation and spreading of shallow and/or deep etch pits. At t_1, the reaction is entirely dominated by the dynamics of shallow etch pits and by step retreat (Figure 5a). A multilayer etch pit (MP1) enters our observation window at time t_2 (Figure 5b) and grows laterally at t_3 (Figure 5c). Horizontal spreading of MP1 at time t_4 is coupled with nucleation of a new deep etch pit (MP2) at another dislocation that is positioned at the center of our observation window (Figure 5d). This pit considerably grows laterally at t_5 and t_6 (Figures 5e and 5f). At these times, the retreat of steps emanated from the center of MP2 favors exposure of crystal terraces to the fluid solution. In turn, numerous shallow etch pits form on such terrace. The edges of MP2 start to display an irregular (spatial) pattern at t_7. This feature is typical of the final phase of the action of the portion of the screw dislocation parallel to the z direction and associated with MP2 in driving the material flux. As younger dissolution pulses are not radiated anymore from the center of MP2, the reaction proceeds by retreat of the steps that have previously formed. A new etch pit (MP3) is seen to nucleate at t_7.

This documented evolution of the surface pattern leads to highly heterogeneous spatial distributions of reaction rates, $R'(\mathbf{x},t^*)$, evaluated through Equation 33 and depicted in Figures 6a–6f. The lateral expansion of MP1, MP2, and MP3 yields dissolution stepwaves that are schematically delineated in orange, blue and green in Figure 6g, respectively. The corresponding regions identify portions of the surface that experience enhanced material fluxes. From a qualitative standpoint, reaction rates within such regions are seen to decrease as time progresses and MPs spread horizontally. This observation is particularly striking for MP2, whose entire temporal evolution is captured within our observation window. The increase of the areal extent of the spatial region

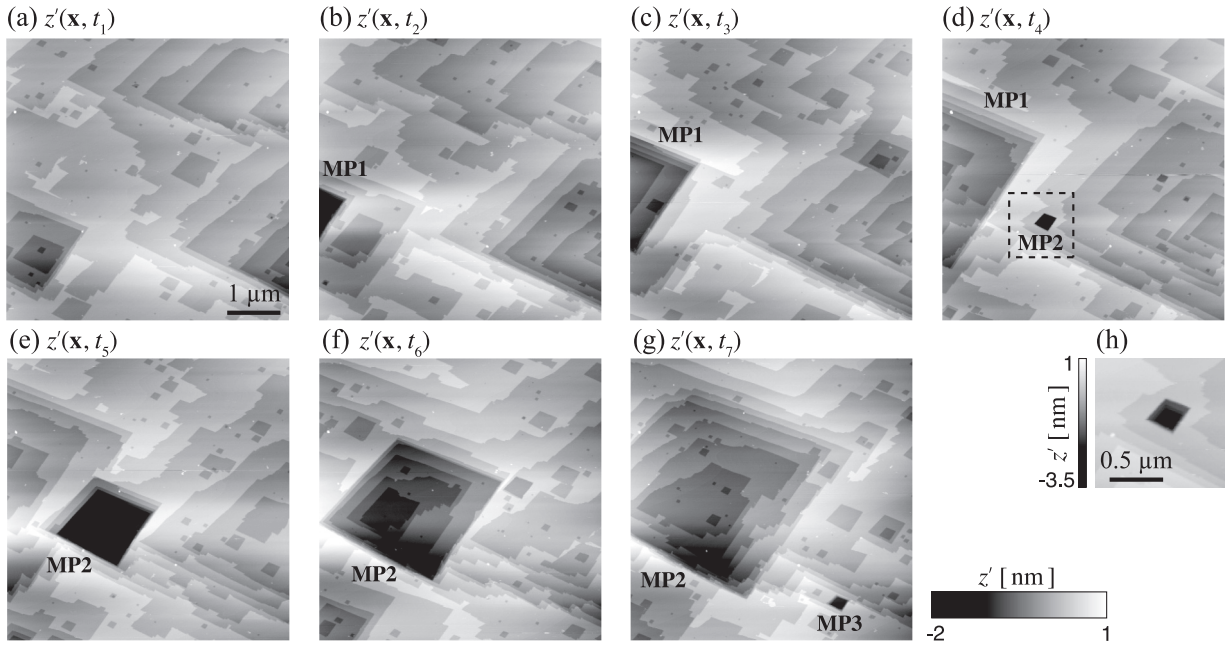

Figure 5. Spatial distributions of fluctuations (about the mean) of calcite topography, $z'(\mathbf{x}, t_i)$, acquired at times (a) $t_1 = 6.6$ min, (b) $t_2 = 13.2$ min, (c) $t_3 = 19.8$ min, (d) $t_4 = 26.4$ min, (e) $t_5 = 33.0$ min, (f) $t_6 = 39.6$ min, and (g) $t_7 = 46.2$ min from the beginning of the experiment; (h) enlarged view of the multilayer etch pit MP2 nucleating within time interval $t_3 - t_4$ and enclosed in the dashed box depicted in (d).

comprised between the edges of MP2 at subsequent times (Figures 6c–6f) is associated with a decrease of the strength of R' therein. Such a temporal evolution of R' is consistent with a description of the dissolution process as driven by stepwave emanation (Fischer & Lüttge, 2018). Otherwise, lower values of rate are attained at spatial locations where the reaction is governed by shallow etch pits and step dynamics.

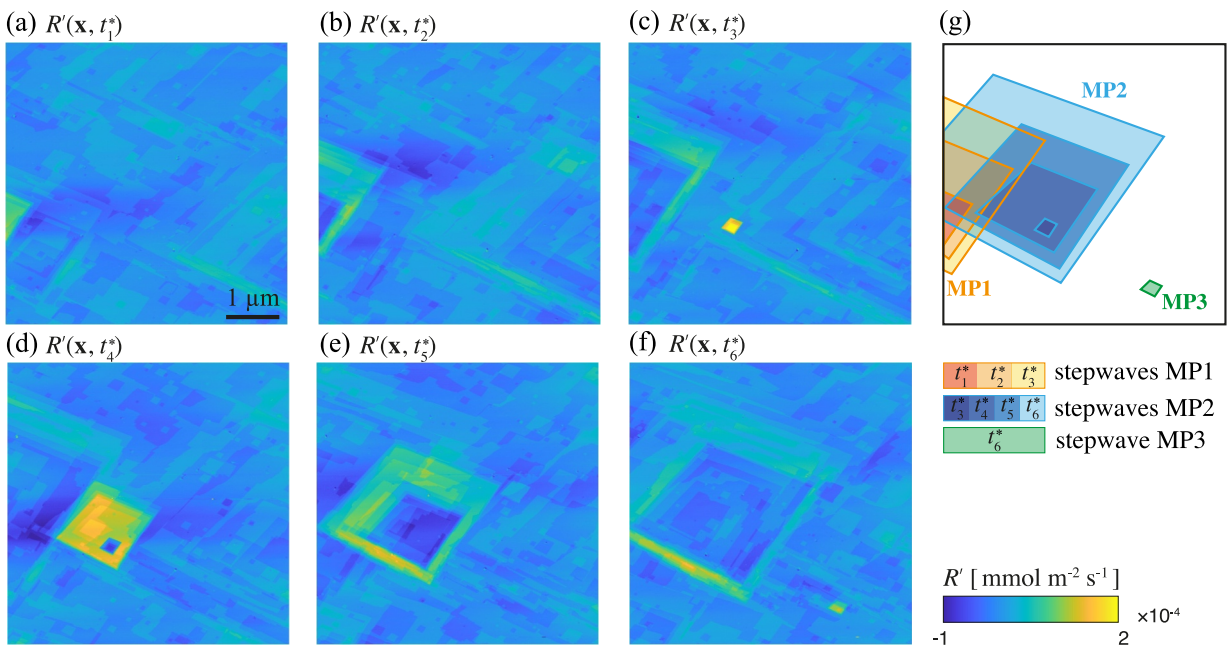

Figure 6. Spatial maps of $R'(\mathbf{x}, t_j^*)$ evaluated with Equation 33 from AFM topography measurements depicted in Figure 5 at times (a) $t_1^* = 13.2$ min, (b) $t_2^* = 19.8$ min, (c) $t_3^* = 26.4$ min, (d) $t_4^* = 33.0$ min, (e) $t_5^* = 39.6$ min, and (f) $t_6^* = 46.2$ min; (g) schematic depiction of dissolution stepwaves emanated from MP1, MP2, and MP3.

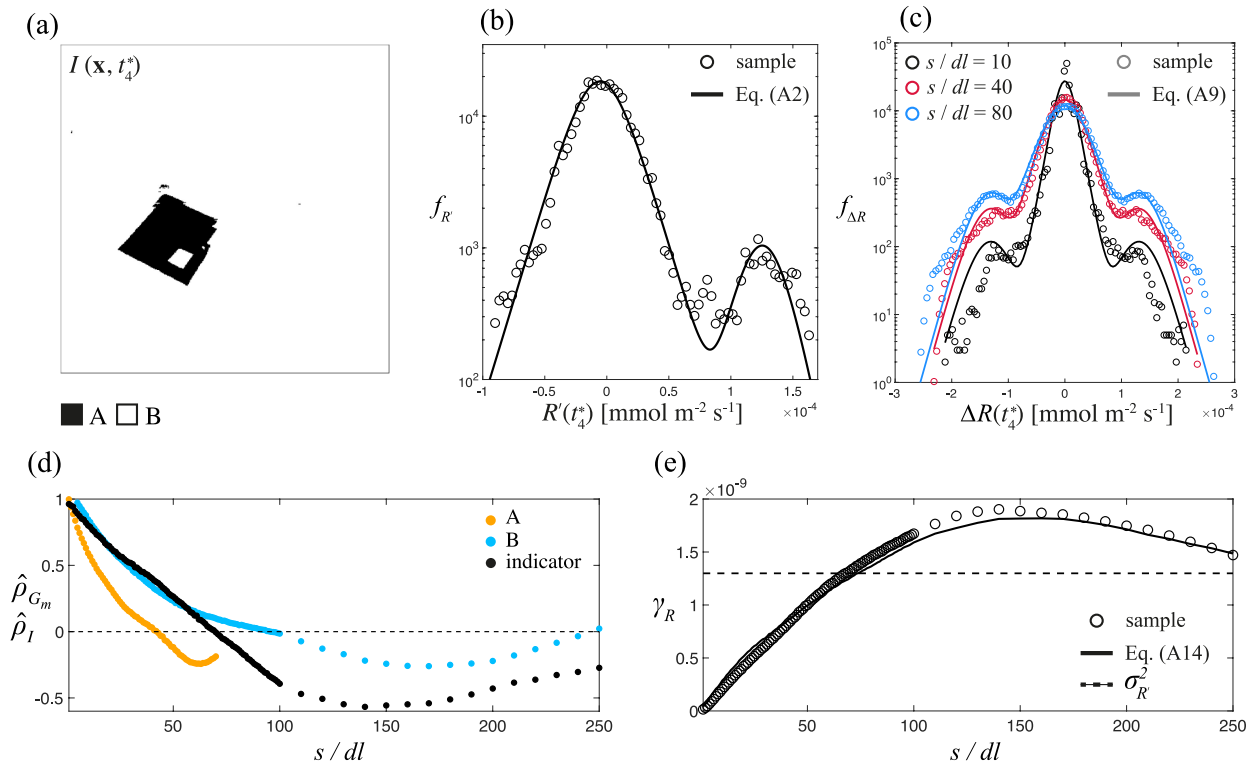

Figure 7. Application of the GSG-MIX modeling framework to the spatial field $R'(\mathbf{x}, t_4^*)$. (a) Indicator field, $I(\mathbf{x}, t_4^*)$, resulting from the classification resting on the algorithm illustrated in Section 2.3. Sample PDFs of (b) R' and (c) ΔR (evaluated at lags $s/dl = 10$, 40, and 80). (d) Estimated correlations of the underlying Gaussian field associated with each *m-th* mode of the mixture, $\hat{\rho}_{G_m}$ $(m = A, B)$, and of the indicator field, $\hat{\rho}_I$. (e) Variogram of R', γ_R. In (b, c, e), GSG-MIX model results (continuous curves) are juxtaposed to sample data (symbols).

As a consequence of the presence of diverse regions, each associated with a given dissolution mechanism, sample probability densities of $R'(\mathbf{x}, t^*)$ exhibit marked multi-modal traits. These emerge whenever distinct mechanistic processes (each characterized by its kinetics) drive dissolution and the (otherwise distinct) populations of the ensuing rates are then lumped within a unique population. Similar patterns are documented for dissolution rate PDFs associated with experimental settings and/or mineral-fluid systems other than those here considered (e.g., Emmanuel, 2014; Fischer & Lüttge, 2017; Fischer et al., 2015; Lange et al., 2021; Noiriel et al., 2018, 2020). As an illustrative example, Figure 7b shows the sample PDF associated with the spatial field $R'(\mathbf{x}, t_4^*)$ depicted in Figure 6d. This is obtained by grouping data into 100 (uniformly spaced) bins. The number of data associated with each of the sample PDF values depicted in Figure 7b ranges between a minimum of 150 and a maximum of 12 911. The PDF is characterized by a dominant peak centered at $R' \sim 0$ and a secondary peak located at $R' > 0$. We relate the former with kinetic processes taking place on crystal terraces, that is, the dynamic evolution of shallow etch pits and step retreat. Otherwise, the latter is associated with high material fluxes due to stepwaves radiating from MP2. A similar behavior is observed at all times (see Figures S1–S5b in Supporting Information S1). Modes associated with high and low values of reaction rate are hereafter denoted as *A* and *B*, respectively (see Figure 7a). Figure 7c depicts sample probability densities of incremental values evaluated at three lags, $s/dl = 10, 40, 80$. At each lag, sample PDFs display a central dominant peak at $\Delta R \approx 0$ and lateral secondary peaks that are symmetric with respect to the origin. These arise from differences of R' values evaluated between two points located within region *A* and *B*, respectively. Sample PDFs of ΔR are characterized by a pronounced scaling tendency with separation distance. The shape of the PDFs changes depending on the lag at which increments are evaluated. As lag decreases, the peaks of the PDFs become sharper and the tails become heavier. All of the above-mentioned traits of the PDFs of R' and ΔR are consistent with a description of the statistical behavior of the dissolution rate as described by the GSG-MIX modeling framework. In the following, we rely on a

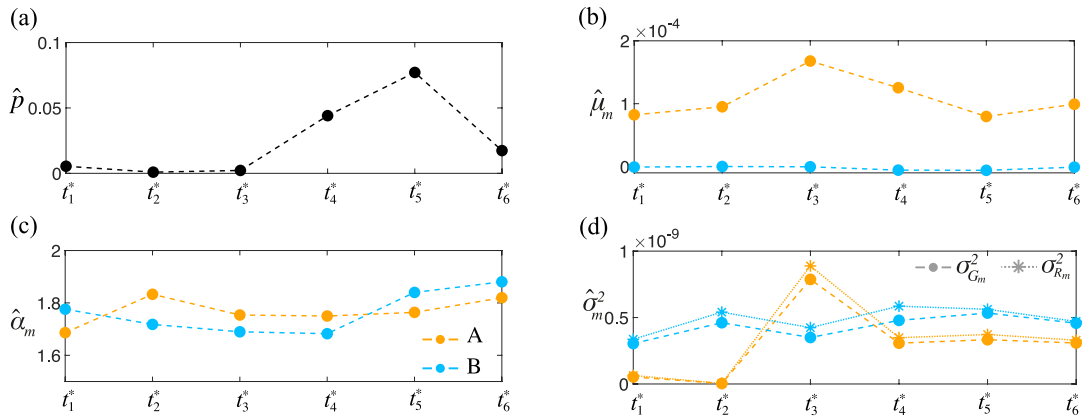

Figure 8. Temporal trends of estimated GSG-MIX parameters embedded in Equation A2, that is, (a) $\hat{p}$, (b) $\hat{\mu}_m$, (c) $\hat{\alpha}_m$, and (d) $\hat{\sigma}^2_{G_m}$ $(m = A, B)$.

lognormal distributional form of the subordinator associated with each component of the mixture (see Appendix A) for our analyses.

We rest on the parameter estimation scheme described in Section 2.3 and obtain a classification of the observed R' fields in two regions together with the estimation of GSG-MIX model parameters. Regions characterized by a higher material flux are associated with dissolution stepwaves emanated by MP1, MP2, and MP3 (see region A in Figure 7a). Otherwise, low reaction rates are identified within regions where the mineral-fluid interface is only experiencing a vertical retreat, that is, terrace regions (see region B in Figure 7a). Estimates of the spatial correlation of the underlying Gaussian field, $\hat{\rho}_{G_m}$ $(m = A, B)$, and of the indicator field, $\hat{\rho}_I$, are depicted versus lag in Figure 7d. We evaluate spatial increments within region A only up to separation distances corresponding to half of the characteristic length of this region. We estimate the latter by relying upon the maximum Feret diameter (Walton, 1948). This metric is typically employed to evaluate an equivalent representative size to characterize irregular shapes in various application contexts (see, e.g., Lolo et al., 2023; Mazzoli & Favoni, 2012; Weber et al., 2014). It is defined as the maximum difference between two parallel tangents to the convex hull enclosing an area (i.e., the biggest connected cluster of region A in our case). The estimated correlation of I exhibits a marked oscillating behavior about 0 at all times (see also Figures S1–S5d in Supporting Information S1). This behavior arises from the geometrical pattern characterizing the indicator field and is linked to stepwave emanation from multilayer etch pits. Similar traits are also observed for the correlation function $\hat{\rho}_{G_B}$, although this is associated with a less pronounced oscillating behavior. We relate such pattern to the dynamic evolution of monolayer etch pits taking place across crystal terraces. Oscillating behaviors at large lags also emerge for the variogram related to the mixture, γ_R (see Figure 7e for $\gamma_R(t_4^*)$ and Figures S1–S5e in Supporting Information S1 for all of the other observation times). Analysis and interpretation of these types of correlation behaviors could rely upon nested models entailing a hole-effect component (see, e.g., Siena et al., 2019, 2021). While we leave this modeling analysis to future developments, here we directly employ $\hat{\rho}_{G_m}$ $(m = A, B)$ and $\hat{\rho}_I$ to specialize the analytical GSG-MIX formulations. Modeling results obtained through Equations A2 and A9 are juxtaposed to sample PDFs of R' and ΔR in Figures 7b and 7c, respectively. Visual inspection of these results documents a remarkable agreement between modeling results and their sample counterparts. As an additional element of analysis, modeling results displayed in Figure 7b are compared against their GSG- and GMIX-based counterparts in Text S2 in Supporting Information S1. The analytical expression of the GSG-MIX variogram (Equation A14) is then juxtaposed to sample data in Figure 7e. Results of similar quality are obtained at all times (see Figures S1–S5b, S5c, and S5e in Supporting Information S1), demonstrating the ability of the novel GSG-MIX model to capture the experimental behavior of both R' and ΔR within a unified theoretical framework.

Figure 8 depicts temporal trends of model parameters embedded in Equation A2 (i.e., $\hat{p}$, $\hat{\mu}_m$, $\hat{\sigma}^2_{G_m}$, and $\hat{\alpha}_m$, with $m = A, B$). Each mode of the PDF of the reaction rates is inherently tied to a mechanism driving the evolution of the mineral-fluid interface. Therefore, analyzing the temporal behavior of each component of the mixture provides valuable insights into the way each mechanistic process contributes to drive the overall reaction dynamics.

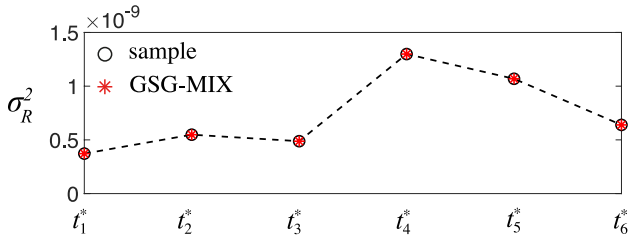

Figure 9. Temporal trend of the variance of R', σ_R^2. Analytical expressions evaluated through Equation A4 are depicted as red asterisks.

We do so by considering the temporal patterns of model parameters, which can, in turn, be viewed as aggregate descriptors of the evolution of the kinetic processes driving the reaction. Analysis of their temporal trend can therefore increase our knowledge on the competition among different mechanistic components and/or on the evolution of dissolution stepwaves. As such, it can assist one to unravel the effect of these mechanisms on the statistical traits exhibited by R'. The temporal trend of $\hat{p}$ (Figure 8a) displays a certain degree of correlation with the behavior of $\hat{\mu}_A$, a peak in the former corresponding to a trough in the latter (Figure 8b). Within the temporal window $t_1^* - t_2^*$, the estimated proportion coefficient $\hat{p}$ is virtually constant, as well as $\hat{\mu}_A$. This behavior reflects the ending phase of the action of the dislocation originating MP1. The largest value of $\hat{\mu}_A$ is attained at time t_3^*. The latter corresponds to the emanation of young dissolution pulses from MP2. As such, material fluxes are higher at this instant than at other times. However, high values of R' are localized in a limited portion of the mineral surface, that is, close to the center of the dislocation. Then, the value of $\hat{\mu}_A$ decreases at times $t_4^* - t_5^*$. This is consistent with the expected evolution of a dissolution stepwave (Fischer & Lüttge, 2018). Even as the horizontal spreading rate of MP2 remains approximately constant in time, the intensity of the material flux in the region comprised within pit edges at subsequent times (i.e., the region associated with the stepwave) decreases as the trains of steps move further away from the dislocation. Otherwise, $\hat{p}$ increases in this time interval, following the lateral expansion of the pit. Finally, time t_6^* corresponds to the ending of stepwave radiation from MP2 and only a part of the area comprised between pit edges at subsequent times is attributed to the high-rate component of the mixture (see Figure S5a in Supporting Information S1). The formation of a new stepwave at MP3 induces a slight increase in $\hat{\mu}_A$ at t_6^*. The mean of the low-rate component, $\hat{\mu}_B$, exhibits a slightly oscillating behavior in time that is ascribed to the dynamic evolution of shallow etch pits. The temporal trend of the estimated shape parameter, $\hat{\alpha}_m$ $(m = A, B)$, is depicted in Figure 8c. These findings suggest that the degree of non-Gaussianity of the data populations associated with each region is (i) significant $(\hat{\alpha}_m \sim 1.7 - 1.9)$, (ii) similar for the two regions, and (iii) remains virtually stable over time. Given that the variance $\left(\sigma_{R_m}^2\right)$ of each $m-th$ component of the mixture is influenced by both α_m and $\sigma_{G_m}^2$, its temporal variations are primarily driven by changes in the variance of the underlying Gaussian fields, $\sigma_{G_m}^2$, as illustrated in Figure 8d. Note that $\hat{\sigma}_{R_B}^2$ remains almost constant at all times. As far as MP1 is the only etch pit driving the reaction (i.e., at t_1^* and t_2^*), there are no significant changes observed in $\hat{\sigma}_{R_A}^2$. Similarly, at t_4^*, t_5^*, and t_6^*, MP2 governs the dissolution process and $\hat{\sigma}_{R_A}^2$ remains nearly constant. This temporal trend indicates that the degree of spatial heterogeneity within region A during these times is attributed to the nucleation of shallow etch pits on terraces freshly exposed to the solution by the retreat of steps originating from MP2. This interpretation is supported by the relatively small difference observed at $t_4^* - t_6^*$ between $\hat{\sigma}_{R_A}^2$ and $\hat{\sigma}_{R_B}^2$. Indeed, the degree of spatial heterogeneity (as expressed though the variance) within region B is also governed by the formation of shallow etch pits and remains almost constant at all times. Otherwise, $\hat{\sigma}_{R_A}^2$ is considerably higher at t_3^* than at all of the other times. At this particular time, the region that is identified as contributing to high reaction rates is seen to comprise two sub-regions. Values of $R'\left(\mathbf{x}, t_3^*\right)$ are assessed from topographies $z'(\mathbf{x}, t_3)$ and $z'(\mathbf{x}, t_4)$ (see Figures 5c, 5d, and 5h). As there is no evidence of the presence of MP2 at t_3, the dislocation originating MP2 is exposed to the solution in the time interval comprised between t_3 and t_4. Since the initial stage of stepwave formation at etch pits is extremely fast, our temporal resolution does not enable us to capture the first phases of this evolution. Therefore, the topography map $z'(\mathbf{x}, t_4)$ can be considered as a temporal average of the processes taking place on the surface in the time interval between t_3 and t_4. These encompass (i) the nucleation of MP2 and (ii) the emanation of initial dissolution pulses. Hence, the spatial map $R'\left(\mathbf{x}, t_3^*\right)$ includes (i) a sub-region associated with the emanation of a new stepwave and (ii) a sub-region arising from the spreading of the dissolution pulses that have been radiated during the time interval $t_3 - t_4$.

Figure 9 depicts the variance associated with the entire field, σ_R^2, versus time. Corresponding analytical expressions evaluated through Equation A4 are also included and document a remarkable level of agreement with experimental results. Values of σ_R^2 are virtually constant in the interval $t_1^* - t_3^*$. They attain a maximum at t_4^*, followed by an approximately linear decreasing trend. Such temporal dynamics somehow resemble the trend exhibited by μ_A (see Figure 8b), with a temporal shift corresponding to the acquisition time Δt. Given that GSG-

MIX parameters are linked to the evolution of the surface pattern, the analysis of their partial contribution to σ_R^2 is informative about the relative influence of the kinetic processes to the level of spatial heterogeneity, as expressed through the value of σ_R^2. On the basis of Equation A4, the variance of the mixture is controlled by (i) the square difference of the means, $(\mu_A - \mu_B)^2$, (ii) the difference of the variances, $\sigma_{R_A}^2 - \sigma_{R_B}^2$, and (iii) the variance of the background region, $\sigma_{R_B}^2$, together with (iv) the proportion coefficient, p. As shallow etch pit dynamics drive the heterogeneity within each region, variances of R_m $(m = A, B)$ attain similar values at all times (see Figure 8d). Therefore, the temporal evolution of σ_R^2 is essentially controlled by the behavior of $(\mu_A - \mu_B)^2$ and p. At times $t_1^* - t_3^*$, σ_R^2 approximately coincides with $\sigma_{R_B}^2$, as the value of p is negligible (Figure 8a). A sudden increase of σ_R^2 is then observed at t_4^*, due to the increase of p coupled with a non negligible value of $(\mu_A - \mu_B)^2$. Note that even as $(\mu_A - \mu_B)^2$ is maximized at t_3^* (Figure 8b), the small lateral extent of the region involved in stepwave emanation at the initial phase of MP2 evolution somehow tempers the effect of such difference. The decreasing trend observed at $t_5^* - t_6^*$ reflects the homogenization of regions A and B that takes place as a consequence of the final stages of the action of the dislocation originating MP2. Hence, the temporal dynamics of σ_R^2 document a delay between the actual formation of the MPs driving the reaction and the time at which this effect propagates onto the statistical behavior of the entire field.

5. Conclusions

We provide a theoretical framework, termed GSG-MIX, and ensuing formulations enabling one to characterize the spatial distribution of a random field, Y, encompassing different regions across an observation domain, each of these being characterized by a given degree of heterogeneity. Our modeling framework extends the Gaussian Mixture (GMIX) model to consider mixtures entailing Generalized sub-Gaussian (GSG) modes. We emphasize that the modeling approach illustrated in this work offers a remarkable flexibility as it admits a general form of the subordinator associated with each GSG mode and embeds the GSG and the GMIX models as particular cases.

Our analytical developments focus on a bimodal mixture (with components A and B) and provide rigorous formulations ensuring consistency between the statistics of Y and its spatial increments, ΔY. Each mode of the mixture is identified through an indicator random variable. The GSG-MIX modeling framework is general and could be applied to the analysis of a variety of hydrological/hydrogeological quantities exhibiting multiple sharp peaks and heavy tails and distributed across distinct (non-overlapping) regions and scales of investigation. The presence of such regions can be linked to the occurrence of different processes and/or geomaterials within the domain of observation. Examples include, for example, spatial distributions of cm-scale air-permeability data (Lu & Zhang, 2002; Siena et al., 2023), hydraulic conductivity of km-scale sedimentary aquifers associated with lithofacies architectures (Dai et al., 2020), or neutron porosity data (Guadagnini et al., 2015). Here, we exemplify its application to the statistical characterization of dissolution rate fields obtained from high quality nanoscale AFM measurements of the topography of a calcite sample subject to dissolution under continuous flowing conditions.

Our analysis leads to the following key conclusions.

1. The PDF of Y, as characterized through a GSG-MIX model, exhibits peaks that tend to become sharper as each component of the mixture departs from a Gaussian distribution. The probability density of ΔY is characterized by a dominant central peak and two secondary peaks. The importance of the central peak with respect to the lateral ones changes with the separation distance at which increments are evaluated. This pattern reflects a scaling behavior that is driven by the correlation structure associated with the indicator random field.
2. We propose and validate an operational workflow for the estimation of model parameters relying on a Bayesian classification approach. The Bayesian classification scheme offers the additional advantage to provide an estimate of the level of uncertainty associated with the clustering of data. Inspection of the posterior probability field associated with the classification imbues us with confidence on the reliability of the parameter estimation, sharp transition boundaries in the posterior probability fields indicating a high level of fidelity of the classification. We recall that an appropriate classification hinges on the availability of a sufficient amount of data. This requirement is met in laboratory-scale scenarios of the kind we illustrate here. Otherwise, applications related to intermediate or larger-scale settings, such as those encountered in various environmental and hydrological/hydrogeological scenarios, might suffer from these types of limitations. As a drawback, the

parameter estimation workflow is potentially associated with a high computational burden. We circumvent this issue by evaluating model parameters analytically upon resting on the method of moments. While computationally efficient, this methodology does not provide a rigorous and complete assessment of the uncertainty of all parameter estimates. Future investigations will focus on exploring alternative (computationally feasible) methodologies that can provide rigorous parameter estimation uncertainty associated with random fields that can be characterized through a GSG-MIX distribution.

3. With reference to the nanoscale setting we considered, sample probability densities of dissolution rate fields, R', exhibit a stark bimodal tendency. Distributions of increments, ΔR, show (*i*) a central dominant peak coupled to lateral secondary peaks and (*ii*) a pronounced scaling behavior with the separation distance at which ΔR are evaluated. Each of the two modes of the mixture exhibits clear non-Gaussian traits such as heavy tails, sharp peak and scaling of the PDFs of increments evaluated within each of the two spatial regions identified. All of these traits are effectively capture by our GSG-MIX formulation. Modeling results exhibit a striking agreement between sample and theoretical PDFs as well as sample and theoretical moments of R' and ΔR. The strong link between dissolution mechanisms and GSG-MIX model parameters enables one to view these as aggregate descriptors of the evolution of the surface processes driving the dissolution reaction. As such, future research efforts will focus on establishing a functional relationship between model parameters and temporal evolution of mechanistic processes. Knowledge of such a relationship could then be used to yield predictions of dissolution rate statistics at unsampled times, thus enriching our fundamental knowledge of nano-scale dissolution processes. These insights are critical for developing robust (physically based) upscaled dissolution models to be employed in field scale scenarios. Statistical descriptors associated with the GSG-MIX modeling framework can provide information on aspects of the physics underpinning dissolution reactions, such as, for example, the degree of persistence of reaction rates related to pulsating dynamics of etch pits, as driven by the typically random spatial variability of crystal defects.

Appendix A: GSG-MIX Formulation for Lognormal Distribution of U

In the following, we consider U_m as described by a lognormal PDF, that is,

$$f_{U_m}(u) = \frac{e^{-\frac{\ln^2 u}{2(2-\alpha_m)^2}}}{\sqrt{2\pi}u(2-\alpha_m)} \quad \text{with} \quad u > 0 \quad \text{and} \quad \alpha_m \in (0,2). \tag{A1}$$

Substituting Equation A1 into Equation 7 yields

$$\begin{aligned} f_Y(y) = {} & \frac{p}{2\pi\sigma_{G_A}(2-\alpha_A)}\int_0^\infty e^{-\frac{1}{2}\left[\left(\frac{\ln u}{2-\alpha_A}\right)^2+\left(\frac{y-\mu_A}{\sigma_{G_A}u}\right)^2\right]}\frac{du}{u^2} \\ & + \frac{1-p}{2\pi\sigma_{G_B}(2-\alpha_B)}\int_0^\infty e^{-\frac{1}{2}\left[\left(\frac{\ln u}{2-\alpha_B}\right)^2+\left(\frac{y-\mu_B}{\sigma_{G_B}u}\right)^2\right]}\frac{du}{u^2}. \end{aligned} \tag{A2}$$

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

$$\langle U_m^q \rangle = e^{\frac{q^2}{2}(2-\alpha_m)^2}. \tag{A3}$$

Variance, skewness, and kurtosis of Y are respectively given by

$$\sigma_Y^2 = p\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} + (1-p)\sigma_{G_B}^2 e^{2(2-\alpha_B)^2} + p(1-p)(\mu_A-\mu_B)^2, \tag{A4}$$

$$\begin{aligned} Sk_Y = {} & \frac{p(1-p)}{\sigma_Y^3}(\mu_A-\mu_B)\Big\{(1-2p)(\mu_A-\mu_B)^2 \\ & + 3\left[\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} - \sigma_{G_B}^2 e^{2(2-\alpha_B)^2}\right]\Big\}, \end{aligned} \tag{A5}$$

$$\kappa_Y = \frac{1}{\sigma_Y^4}\Big\{3p\Big[\sigma_{G_A}^4 e^{8(2-\alpha_A)^2} - \sigma_{G_B}^4 e^{8(2-\alpha_B)^2}\Big] + 3\sigma_{G_B}^4 e^{8(2-\alpha_B)^2} + p(1-p)\big(\mu_A-\mu_B\big)^2\Big[(1-3p(1-p))\big(\mu_A-\mu_B\big)^2 + 6\Big(\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} - p\Big(\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} - \sigma_{G_B}^2 e^{2(2-\alpha_B)^2}\Big)\Big)\Big]\Big\}. \tag{A6}$$

Note that second- and fourth-order central moment of each component of the mixture respectively read

$$\langle Y_m^{'2}\rangle = \sigma_{G_m}^2 e^{2(2-\alpha_m)^2}, \tag{A7}$$

$$\langle Y_m^{'4}\rangle = 3\sigma_{G_m}^4 e^{8(2-\alpha_m)^2}. \tag{A8}$$

Substituting Equation A1 in Equation 19 leads to

$$\begin{aligned} f_{\Delta Y}(\Delta Y) = {} & \frac{p^2 + C_I(s)}{(2\pi)^{3/2}(2-\alpha_A)^2\sigma_{G_A}}\int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1 + \ln^2 u_2}{(2-\alpha_A)^2} + \frac{\Delta y^2}{\sigma_{G_A}^2 r^2}\right]}\frac{du_2 du_1}{u_1 u_2 r} \\ & + \frac{(1-p)^2 + C_I(s)}{(2\pi)^{3/2}(2-\alpha_B)^2\sigma_{G_B}}\int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1 + \ln^2 u_2}{(2-\alpha_B)^2} + \frac{\Delta y^2}{\sigma_{G_B}^2 r^2}\right]}\frac{du_2 du_1}{u_1 u_2 r} \\ & + \frac{p(1-p) - C_I(s)}{(2\pi)^{3/2}(2-\alpha_A)(2-\alpha_B)}\Bigg\{\int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\left(\frac{\ln u_1}{2-\alpha_A}\right)^2 + \left(\frac{\ln u_2}{2-\alpha_B}\right)^2 + \left(\frac{\Delta y - \mu_A + \mu_B}{h}\right)^2\right]}\frac{du_2 du_1}{u_1 u_2 h} \\ & + \int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\left(\frac{\ln u_1}{2-\alpha_B}\right)^2 + \left(\frac{\ln u_2}{2-\alpha_A}\right)^2 + \left(\frac{\Delta y + \mu_A - \mu_B}{w}\right)^2\right]}\frac{du_2 du_1}{u_1 u_2 w}\Bigg\}. \end{aligned} \tag{A9}$$

Making use of Equation A9, second- and fourth-order moment of the increments read

$$\begin{aligned} \langle \Delta Y^2\rangle = {} & 2\Big\{p^2\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\Big[e^{(2-\alpha_A)^2} - \rho_{G_A}\Big] + (1-p)^2\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\Big[e^{(2-\alpha_B)^2} - \rho_{G_B}\Big] \\ & + p(1-p)\Big[(1-\rho_I)\big(\mu_A-\mu_B\big)^2 + \sigma_{G_A}^2 e^{(2-\alpha_A)^2}\Big(e^{(2-\alpha_A)^2} - \rho_I\rho_{G_A}\Big) \\ & + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\Big(e^{(2-\alpha_B)^2} - \rho_I\rho_{G_B}\Big)\Big]\Big\}, \end{aligned} \tag{A10}$$

$$\begin{aligned} \langle \Delta Y^4\rangle = {} & 6\Big\{p^2\sigma_{G_A}^4 e^{4(2-\alpha_A)^2}\Big[1 + e^{4(2-\alpha_A)^2} - 4e^{(2-\alpha_A)^2}\rho_{G_A} + 2\rho_{G_A}^2\Big] \\ & + (1-p)^2\sigma_{G_B}^4 e^{4(2-\alpha_B)^2}\Big[1 + e^{4(2-\alpha_B)^2} - 4e^{(2-\alpha_B)^2}\rho_{G_B} + 2\rho_{G_B}^2\Big] \\ & + p(1-p)\rho_I\sigma_{G_A}^4 e^{4(2-\alpha_A)^2}\Big[1 + e^{4(2-\alpha_A)^2} - 4e^{(2-\alpha_A)^2}\rho_{G_A} + 2\rho_{G_A}^2\Big] \\ & + \sigma_{G_B}^4 e^{4(2-\alpha_B)^2}\Big[1 + e^{4(2-\alpha_B)^2} - 4e^{(2-\alpha_B)^2}\rho_{G_B} + 2\rho_{G_B}^2\Big]\Big\} \\ & + 2p(1-p)(1-\rho_I)\Big\{\big(\mu_A-\mu_B\big)^4 + 6\big(\mu_A-\mu_B\big)^2\Big[\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} + \sigma_{G_B}^2 e^{2(2-\alpha_B)^2}\Big] \\ & + 3\Big[\sigma_{G_A}^4 e^{8(2-\alpha_A)^2} + \sigma_{G_B}^4 e^{8(2-\alpha_B)^2} + 2\sigma_{G_A}^2\sigma_{G_B}^2 e^{2\left((2-\alpha_A)^2 + (2-\alpha_B)^2\right)}\Big]\Big\}. \end{aligned} \tag{A11}$$

Note that second-order moment of the increments associated with each component of the mixture reads

$$\langle \Delta Y_m^2\rangle = 2\sigma_{G_m}^2 e^{(2-\alpha_m)^2}\Big[e^{(2-\alpha_m)^2} - \rho_{G_m}\Big]. \tag{A12}$$

The covariance and the variogram of Y respectively read

$$C_Y(s) = \begin{cases} \sigma_Y^2 & \text{if } s = 0 \\ p^2\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\rho_{G_A} + (1-p)^2\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\rho_{G_B} + & \text{if } s > 0 \\ +p(1-p)\rho_I\left[(\mu_A-\mu_B)^2 + \sigma_{G_A}^2 e^{(2-\alpha_A)^2}\rho_{G_A} + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\rho_{G_B}\right] \end{cases} \tag{A13}$$

$$\begin{aligned} \gamma_Y &= \nu + p^2 e^{(2-\alpha_A)^2}\gamma_{G_A} + (1-p)^2 e^{(2-\alpha_B)^2}\gamma_{G_B} + (\mu_A-\mu_B)^2\gamma_I \\ &\quad + p(1-p)\left[\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\left(1-\rho_I\rho_{G_A}\right) + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\left(1-\rho_I\rho_{G_B}\right)\right], \end{aligned} \tag{A14}$$

where

$$\nu = p\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\left(e^{(2-\alpha_A)^2}-1\right) + (1-p)\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\left(e^{(2-\alpha_B)^2}-1\right) \tag{A15}$$

is the nugget. The integral scale of the mixture is obtained as

$$\begin{aligned} I_Y = \frac{1}{\sigma_Y^2}\Big\{ & p^2\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\lambda_{G_A} + (1-p)^2\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\lambda_{G_B} \\ & + p(1-p)\left[(\mu_A-\mu_B)^2\lambda_I + \sigma_{G_A}^2 e^{(2-\alpha_A)^2}\int_0^\infty \rho_I\rho_{G_A}ds + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\int_0^\infty \rho_I\rho_{G_B}ds\right]\Big\}. \end{aligned} \tag{A16}$$

Appendix B: Performance Assessment of the Designed Operational Scheme on Synthetic Datasets

The methodology for parameter estimation outlined in Section 2.3 is tested on synthetic fields generated on a 450×450 grid. Such a grid size is selected to match that of experimental data presented in Section 3. The mixture is obtained by generating (*i*) two spatial random fields distributed according to a unimodal GSG and (*ii*) an indicator field resembling geometries typically observed for experimental dissolution rate fields. Unimodal GSG fields associated with Y_A and Y_B are generated using an exponential correlation model for G_m, that is, $\rho_{G_m}(s) = e^{-s/\lambda_{G_m}}$ $(m = A, B)$, and setting: $\mu_A = 2.5, \sigma_{G_A}^2 = 0.05, \alpha_A = 1.7, \lambda_{G_A} = 10$ and $\mu_B = 0.5, \sigma_{G_B}^2 = 0.15, \alpha_B = 1.85, \lambda_{G_B} = 20$. The system is characterized by a proportion $p = 0.1$. The mixture is then evaluated making use of Equation 2.

A schematic representation of the application of the parameter estimation procedure is offered in Figure 4. The fields used for the algorithm initialization are shown in Figure 4a where we mimic the occurrence of a classification error that (often) occurs when dealing with experimental data. For region A, we introduce a misclassification by selecting (*a*) a region having an area corresponding to $\sim$60% (black area in Figure 4a) of the real one (orange pattern in Figure 4a) and (*b*) a small portion ($\sim$1%) of region B (blue pattern in Figure 4a) located next to the edges of the actual region A. Region B is initialized by selecting a random region with an areal extent of $\sim$7% of the real region B. Such a choice is grounded on our ability to preliminary identify mechanistic processes contributing to high (region A) and low (region B) rates. Identification of region B is generally straightforward because regions that are only experiencing vertical surface retreat, that is, terrace regions, are associated with the lowest dissolution rate (see, e.g., Bibi et al., 2018; Bollermann & Fischer, 2020; Brand et al., 2017) and are therefore always attributed to the lowest-mean component of the mixture. On the other hand, even as younger dissolution stepwaves could be reasonably associated with high rates, it is also possible that etch pits characterized by a depth comprising a few layers and/or older dissolution stepwaves are associated with a high material flux. The performance of the parameter estimation algorithm is assessed in Figure 4d, where the estimated parameters (i.e., $\hat{p}, \hat{\mu}_m, \hat{\alpha}_m$, and $\hat{\sigma}_{G_m}$) are compared with their actual reference values. Errors between estimated and reference parameter values are evaluated at the last repetition of the algorithm and are also included in Figure 4d. We notice that μ_B, σ_{G_A}, and α_A are associated with the highest errors. The error associated with μ_B is related to the

distance between the components means, $(\mu_A - \mu_B)$. If this distance is relevant, assigning some cells that actually belong to region A to region B induces an increases of $\hat{\mu}_B$, even as the error associated with the estimation of p is low (e.g., $= 0.5\%$ in the considered case). The estimation error related to σ_{G_A} and α_A can be attributed to the a low sample size of region A. Notably, the overall estimation error of these parameters still remains below 10%.

Analysis of the dependence of the results on the number of repetitions of the algorithm documents that the parameter set characterizing each component Y_m does not change anymore after convergence of $\hat{p}$ (Figure 4d). Therefore, we stop replicating the algorithm upon reaching variations of $\hat{p}$ that are smaller than 2%. The performance of the algorithm is also tested on synthetic fields generated starting from the same Y_m distributions and different geometrical arrangements of the indicator (not shown). These resemble other surface pattern evolution types, such as the nucleation of multiple smaller etch pits. The results obtained are of similar quality as those shown here.

Data Availability Statement

Data are freely available at Recalcati, Riva, and Guadagnini (2024). Matlab scripts specifically developed for the research are available at https://doi.org/10.5281/zenodo.14713743.

Acronyms

AFM	atomic force microscopy
CDF	cumulative distribution function
GMIX	Gaussian mixture
GSG	generalized sub-Gaussian
GSG-MIX	generalized sub-Gaussian mixture
PDF	probability density function

Acknowledgments
Funding from the European Union Next-Generation EU (National Recovery and Resilience Plan—NRRP, Mission 4, Component 2, Investment 1.3—D.D. 1243 2/8/2022, PE0000005) in the context of the RETURN Extended Partnership is acknowledged. The experiments have been performed at the Solid-Liquid Interface Nanomicroscopy and Spectroscopy laboratory (SoLINano-Σ LAB) at Politecnico di Milano. Open access publishing facilitated by Politecnico di Milano, as part of the Wiley - CRUI-CARE agreement.

References

Bibi, I., Arvidson, R. S., Fischer, C., & Lüttge, A. (2018). Temporal evolution of calcite surface dissolution kinetics. *Minerals*, *8*(6), 256. https://doi.org/10.3390/min8060256

Bollermann, T., & Fischer, C. (2020). Temporal evolution of dissolution kinetics of polycrystalline calcite. *American Journal of Science*, *320*(1), 53–71. https://doi.org/10.2475/01.2020.04

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

Brand, A. S., Feng, P., & Bullard, J. W. (2017). Calcite dissolution rate spectra measured by in situ digital holographic microscopy. *Geochimica et Cosmochimica Acta*, *213*, 317–329. https://doi.org/10.1016/j.gca.2017.07.001

Dai, Z., Ritzi, R. W., Jr., Huang, C., Rubin, Y. N., & Dominic, D. F. (2004). Transport in heterogeneous sediments with multimodal conductivity and hierarchical organization across scales. *Journal of Hydrology*, *294*(1–3), 68–86. https://doi.org/10.1016/j.jhydrol.2003.10.024

Dai, Z., Zhan, C., Dong, S., Yin, S., Zhang, X., & Soltanian, M. R. (2020). How does resolution of sedimentary architecture data affect plume dispersion in multiscale and hierarchical systems? *Journal of Hydrology*, *582*, 124516. https://doi.org/10.1016/j.jhydrol.2019.124516

Daval, D. (2018). Carbon dioxide sequestration through silicate degradation and carbon mineralisation: Promises and uncertainties. *npj Materials Degradation*, *2*(1), 11. https://doi.org/10.1038/s41529-018-0035-4

Desbarats, A. J. (1987). Numerical estimation of effective permeability in sand-shale formations. *Water Resources Research*, *23*(2), 273–286. https://doi.org/10.1029/wr023i002p00273

Duda, R. O., E, H. P., & G, S. D. (2000). *Pattern classification* (2nd ed.). John Wiley & Sons.

Emmanuel, S. (2014). Mechanisms influencing micron and nanometer-scale reaction rate patterns during dolostone dissolution. *Chemical Geology*, *363*, 262–269. https://doi.org/10.1016/j.chemgeo.2013.11.002

Ewing, R. C. (2015). Long-term storage of spent nuclear fuel. *Nature Materials*, *14*(3), 252–257. https://doi.org/10.1038/nmat4226

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

Fischer, C., Finkeldei, S., Brandt, F., Bosbach, D., & Lüttge, A. (2015). Direct measurement of surface dissolution rates in potential nuclear waste forms: The example of pyrochlore. *ACS Applied Materials & Interfaces*, *7*(32), 17857–17865. https://doi.org/10.1021/acsami.5b04281

Fischer, C., Kurganskaya, I., Schäfer, T., & Lüttge, A. (2014). Variability of crystal surface reactivity: What do we know? *Applied Geochemistry*, *43*, 132–157. https://doi.org/10.1016/j.apgeochem.2014.02.002

Fischer, C., & Lüttge, A. (2017). Beyond the conventional understanding of water–rock reactivity. *Earth and Planetary Science Letters*, *457*, 100–105. https://doi.org/10.1016/j.epsl.2016.10.019

Fischer, C., & Lüttge, A. (2018). Pulsating dissolution of crystalline matter. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(5), 897–902. https://doi.org/10.1073/pnas.1711254115

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

Guadagnini, A., Neuman, S., Schaap, M., & Riva, M. (2013). Anisotropic statistical scaling of vadose zone hydraulic property estimates near Maricopa, Arizona. *Water Resources Research*, *49*(12), 8463–8479. https://doi.org/10.1002/2013wr014286

Guadagnini, A., Neuman, S., Schaap, M., & Riva, M. (2014). Anisotropic statistical scaling of soil and sediment texture in a stratified deep vadose zone near Maricopa, Arizona. *Geoderma*, *214*, 217–227. https://doi.org/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. *Hydrology and Earth System Sciences*, *19*(2), 729–745. https://doi.org/10.5194/hess-19-729-2015

Guadagnini, A., 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

Harstad, A., & Stipp, S. (2007). Calcite dissolution: Effects of trace cations naturally present in Iceland spar calcites. *Geochimica et Cosmochimica Acta*, *71*(1), 56–70. https://doi.org/10.1016/j.gca.2006.07.037

Heberling, F., Bosbach, D., Eckhardt, J.-D., Fischer, U., Glowacky, J., Haist, M., et al. (2014). Reactivity of the calcite–water-interface, from molecular scale processes to geochemical engineering. *Applied Geochemistry*, *45*, 158–190. https://doi.org/10.1016/j.apgeochem.2014.03.006

Hövelmann, J., Putnis, C. V., & Benning, L. G. (2018). Metal sequestration through coupled dissolution–precipitation at the brucite–water interface. *Minerals*, *8*(8), 346. https://doi.org/10.3390/min8080346

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, *15*(3), 445–468. https://doi.org/10.1007/bf01031292

Julia, M., Putnis, C. V., King, H. E., & Renard, F. (2023). Coupled dissolution-precipitation and growth processes on calcite, aragonite, and Carrara marble exposed to cadmium-rich aqueous solutions. *Chemical Geology*, *621*, 121364. https://doi.org/10.1016/j.chemgeo.2023.121364

Kumar, P., & Foufoula-Georgiou, E. (1993). A multicomponent decomposition of spatial rainfall fields: 2. Self-similarity in fluctuations. *Water Resources Research*, *29*(8), 2533–2544. https://doi.org/10.1029/93wr00549

Lange, I., Toro, M., Arvidson, R. S., Kurganskaya, I., & Lüttge, A. (2021). The role of crystal heterogeneity in alkali feldspar dissolution kinetics. *Geochimica et Cosmochimica Acta*, *309*, 329–351. https://doi.org/10.1016/j.gca.2021.06.032

Lasaga, A. C., & Lüttge, A. (2001). Variation of crystal dissolution rate based on a dissolution stepwave model. *Science*, *291*(5512), 2400–2404. https://doi.org/10.1126/science.1058173

Lasaga, A. C., & Lüttge, A. (2003). A model for crystal dissolution. *European Journal of Mineralogy*, *15*(4), 603–615. https://doi.org/10.1127/0935-1221/2003/0015-0603

Li, K., Wu, J., Nan, T., Zeng, X., Yin, L., & Zhang, J. (2022). Analysis of heterogeneity in a sedimentary aquifer using generalized sub-Gaussian model based on logging resistivity. *Stochastic Environmental Research and Risk Assessment*, *36*(3), 767–783. https://doi.org/10.1007/s00477-021-02054-5

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

Lolo, F.-N., Walani, N., Seemann, E., Zalvidea, D., Pavón, D. M., Cojoc, G., et al. (2023). Caveolin-1 dolines form a distinct and rapid caveolae-independent mechanoadaptation system. *Nature Cell Biology*, *25*(1), 120–133. https://doi.org/10.1038/s41556-022-01034-3

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*(3), 183–188. https://doi.org/10.2113/gselements.9.3.183

Lüttge, A., Arvidson, R. S., Fischer, C., & Kurganskaya, I. (2019). Kinetic concepts for quantitative prediction of fluid-solid interactions. *Chemical Geology*, *504*, 216–235. https://doi.org/10.1016/j.chemgeo.2018.11.016

Massabó, M., Bellin, A., & Valocchi, A. (2008). Spatial moments analysis of kinetically sorbing solutes in aquifer with bimodal permeability distribution. *Water Resources Research*, *44*(9). https://doi.org/10.1029/2007wr006539

Mazzoli, A., & Favoni, O. (2012). Particle size, size distribution and morphological evaluation of airborne dust particles of diverse woods by scanning electron microscopy and image processing program. *Powder Technology*, *225*, 65–71. https://doi.org/10.1016/j.powtec.2012.03.033

Meerschaert, M. M., Kozubowski, T. J., Molz, F. J., & Lu, S. (2004). Fractional Laplace model for hydraulic conductivity. *Geophysical Research Letters*, *31*(8). https://doi.org/10.1029/2003gl019320

Neuman, S. P., Riva, M., Guadagnini, A., Siena, M., & Recalcati, C. (2024). Statistical scaling of randomly fluctuating hierarchical variables. In *Oxford Research Encyclopedia of Environmental Science*. Oxford University Press.

Noiriel, C., & Daval, D. (2017). Pore-scale geochemical reactivity associated with CO_2 storage: New frontiers at the fluid–solid interface. *Accounts of Chemical Research*, *50*(4), 759–768. https://doi.org/10.1021/acs.accounts.7b00019

Noiriel, C., Madé, B., & Gouze, P. (2007). Impact of coating development on the hydraulic and transport properties in argillaceous limestone fracture. *Water Resources Research*, *43*(9). https://doi.org/10.1029/2006wr005379

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

Noiriel, C., Oursin, M., Saldi, G., & Haberthür, D. (2018). Direct determination of dissolution rates at crystal surfaces using 3D x-ray micro-tomography. *ACS Earth and Space Chemistry*, *3*(1), 100–108. https://doi.org/10.1021/acsearthspacechem.8b00143

Painter, S. (1996). Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations. *Water Resources Research*, *32*(5), 1183–1195. https://doi.org/10.1029/96wr00286

Rahe, P., Bechstein, R., & Kühnle, A. (2010). Vertical and lateral drift corrections of scanning probe microscopy images. *Journal of Vacuum Science & Technology B*, *28*(3), C4E31–C4E38. https://doi.org/10.1116/1.3360909

Recalcati, C., Riva, M., & Guadagnini, A. (2024). Dataset of the paper "characterization of spatially heterogeneous environmental variables through multi-modal generalized sub-Gaussian distributions" [Dataset]. https://doi.org/10.17632/xgjxbprt3g.1

Recalcati, C., Siena, M., Riva, M., Bollani, M., & Guadagnini, A. (2024). Stochastic assessment of dissolution at fluid-mineral interfaces. *Geophysical Research Letters*, *51*(7), e2023GL108080. https://doi.org/10.1029/2023gl108080

Recalcati, C., Siena, M., Riva, M., & Guadagnini, A. (2023). Experimental assessment of calcite dissolution patterns through atomic force microscopy. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier.

Renard, F., Putnis, C. V., Montes-Hernandez, G., King, H. E., Breedveld, G. D., & Okkenhaug, G. (2018). Sequestration of antimony on calcite observed by time-resolved nanoscale imaging. *Environmental Science and Technology*, *52*(1), 107–113. https://doi.org/10.1021/acs.est.7b04727

Ricci, D., & Braga, P. C. (2004). Recognizing and avoiding artifacts in AFM imaging. *Atomic Force Microscopy: Biomedical Methods and Applications*, 25–37. https://doi.org/10.1385/1-59259-647-9:25

Riva, M., Neuman, S. P., & Guadagnini, A. (2015). 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., Neuman, S. P., Guadagnini, A., & Siena, M. (2013). Anisotropic scaling of Berea sandstone log air permeability statistics. *Vadose Zone Journal*, *12*(3), 1–15. https://doi.org/10.2136/vzj2012.0153

Rubin, Y. (1995). Flow and transport in bimodal heterogeneous formations. *Water Resources Research*, *31*(10), 2461–2468. https://doi.org/10.1029/95wr01953

Rubin, Y., & Journel, A. G. (1991). Simulation of non-Gaussian space random functions for modeling transport in groundwater. *Water Resources Research*, *27*(7), 1711–1721. https://doi.org/10.1029/91WR00838

Siena, M., Bussetti, G., Recalcati, C., Riva, M., Duò, L., & Guadagnini, A. (2021). Statistical characterization of heterogeneous dissolution rates of calcite from in situ and real-time AFM imaging. *Transport in Porous Media*, *140*(1), 291–312. https://doi.org/10.1007/s11242-021-01624-z

Siena, M., Guadagnini, A., Bouissonnié, A., Ackerer, P., Daval, D., & Riva, M. (2020). Generalized sub-Gaussian processes: Theory and application to hydrogeological and geochemical data. *Water Resources Research*, *56*(8), e2020WR027436. https://doi.org/10.1029/2020wr027436

Siena, M., Recalcati, C., Guadagnini, A., & Riva, M. (2023). A Gaussian-mixture based stochastic framework for the interpretation of spatial heterogeneity in multimodal fields. *Journal of Hydrology*, *617*, 128849. https://doi.org/10.1016/j.jhydrol.2022.128849

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

Stigliano, L., Siena, M., Ackerer, P., Guadagnini, A., & Daval, D. (2021). Statistical description of calcite surface roughness resulting from dissolution at close-to-equilibrium conditions. *ACS Earth and Space Chemistry*, *5*(11), 3115–3129. https://doi.org/10.1021/acsearthspacechem.1c00226

Teng, H. H. (2004). Controls by saturation state on etch pit formation during calcite dissolution. *Geochimica et Cosmochimica Acta*, *68*(2), 253–262. https://doi.org/10.1016/s0016-7037(03)00423-x

Walton, W. (1948). Feret's statistical diameter as a measure of particle size. *Nature*, *162*(4113), 329–330. https://doi.org/10.1038/162329b0

Weber, F., Kowarik, I., & Säumel, I. (2014). Herbaceous plants as filters: Immobilization of particulates along urban street corridors. *Environmental Pollution*, *186*, 234–240. https://doi.org/10.1016/j.envpol.2013.12.011

Winter, C., Tartakovsky, D., & Guadagnini, A. (2003). Moment differential equations for flow in highly heterogeneous porous media. *Surveys in Geophysics*, *24*(1), 81–106. https://doi.org/10.1023/A:1022277418570

Yang, C.-Y., Hsu, K.-C., & Chen, K.-C. (2009). The use of the levy-stable distribution for geophysical data analysis. *Hydrogeology Journal*, *17*(5), 1265–1273. https://doi.org/10.1007/s10040-008-0411-1

Zhang, Y., Green, C. T., & Fogg, G. E. (2013). The impact of medium architecture of alluvial settings on non-Fickian transport. *Advances in Water Resources*, *54*, 78–99. https://doi.org/10.1016/j.advwatres.2013.01.004