UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year: 2024

Stochastic characterization of reactive processes in porous media

Recalcati Chiara

Recalcati Chiara, 2024, Stochastic characterization of reactive processes in porous media

Originally published at : Thesis, University of Lausanne
Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_872106D8975C9

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

POLITECNICO
MILANO 1863

Faculty of Geosciences
and Environment
Institute of Earth Sciences

PhD in Earth Sciences

Department of Civil and
Environmental Engineering

PhD in Environmental and
Infrastructure Engineering

Doctoral dissertation

Stochastic characterization of reactive processes in porous media

Chiara Recalcati

a.y. 2023 - 2024

Supervisors

Prof. Alberto **GUADAGNINI**, Politecnico di Milano
Prof. Pietro **DE ANNA**, Université de Lausanne

Co-supervisor

Prof. Martina **SIENA**, Politecnico di Milano

Tutor

Prof. Monica **RIVA**, Politecnico di Milano

UNIL | Université de Lausanne
Faculté des géosciences et de l'environnement
Géopolis
CH-1015 Lausanne

IMPRIMATUR

La Faculté des géosciences et de l'environnement de l'Université de Lausanne, vu le rapport du jury d'examen, autorise l'impression de la thèse de doctorat rédigée par

Chiara RECALCATI

intitulée

Stochastic characterization of reactive processes in porous media.

sans se prononcer sur les opinions exprimées dans cette thèse.

Directeur	Monsieur Pietro De Anna
Co-directeur	Monsieur Alberto Guadagnini

Lausanne, le 12.06.2024

Niklas Linde
Professeur Niklas Linde, Doyen

To the beloved memory
of my grandmother Jole

Acknowledgements

I would like the express my deepest gratitude to my advisors, Profs. Alberto Guadagnini and Monica Riva, for their unwavering support, immense patience, and wise guidance throughout these years. Their passion, dedication, and excitement for fundamental research inspired me every day of this long journey. This endeavor would not have been possible without them. I extend my heartfelt thanks to my co-advisor, Prof. Martina Siena, who supported me from the very beginning of the PhD path. Special appreciation goes to my advisor, Prof. Pietro de Anna, for introducing me to the exciting world of microfluidics and hosting me in Lausanne. I am also grateful to Prof. Gianlorenzo Bussetti for welcoming me to the wonderful SoLINano-Σ Lab.

I would like to thank my colleagues at Politecnico di Milano and at Université de Lausanne - Leonardo, Luca, Daniel, Panagiotis, Davide, Marta, Wenqiao, Andrea, and Laura - who made these years remarkably funny and filled with cherished memories. Thanks to my best friends, Francesco and Irene, for being there during good and bad times. I am grateful to my parents, Massimo and Patrizia, for always encouraging me to follow my dreams no matter what.

Abstract

The research is focused on exploring critical elements of the nature of flow and transport processes across porous geomaterials and their interactions/feedbacks with the host solid matrix. Natural rocks form the porous medium across which flow of water (and other fluids) and transport of chemicals dissolved therein take place across the Earth subsurface. In this broad context, dissolution is a key process driving mineral transformations taking place in the upper crust of the Earth. Such process contributes to drive formation of preferential pathways across the subsurface. Carbonate rocks are of particular interest. This is due to their abundance in the subsurface and to their high reactivity. They drive weathering of the surface of the Earth, interact with spreading of pollutants in groundwater aquifers, and favor carbon capture through mineralization. Key research questions tackled in this PhD dissertation are related to (a) enhancing our ability to directly observe precipitation/dissolution reaction rates through original nano-/microscale imaging experiments and (b) providing an interpretation of the observed rates through rigorous stochastic approaches capable of quantifying uncertainty. These are tackled with a unique blend of experimental and theoretical/modeling advancements. As such, methods are based on fundamental theoretical developments (with analytical and numerical approaches) and experiments. Theoretical approaches are of a stochastic nature, given the ubiquitous uncertainty about the mechanisms driving the dynamics of the alteration of the surface of minerals in contact with fluids. Original experiments are performed directly at the micro- and nanoscale to observe through imaging fundamental physics of the interaction between fluids and mineral surface. The dissertation is structured along the five Chapters briefly described in the following.

Chapter 1 provides an introduction to the work, including an overall description of the hydrogeochemical settings analyzed. **Chapter 2** provides an overview of the experimental setting considered and designed during the PhD path. The experimental protocols designed to acquire Atomic Force Microscopy (AFM) images mimicking natural settings typical of calcite dissolution processes taking place in various engineering/environmental scenarios are illustrated. A variety of experimental setups and protocols are designed to resemble condi-

tions associated with (*i*) diffusion-dominated and (*ii*) surface-controlled reactions that are typical of stagnant and flowing regions in porous geomaterials. Such protocols are documented to enable one to obtain experimental data that can be promptly employed for the evaluation of space-time distributions of surface topography of the mineral in contact with the fluid. These are in turn employed to evaluate the ensuing space-time distributions of reaction rates. The latter are then subject to an original stochastic characterization. In this sense, **Chapter 3** includes all of the details of the theoretical stochastic framework employed for the analysis and interpretation of the experimental evidences. The approach and ensuing formulations embed the joint assessment of the probability distribution of a target variable and its associated spatial increments, taken between locations separated by any given distance (or lag). The random field associated with reaction rates is interpreted through a generally non-Gaussian bimodal mixture model. The modes of the latter correspond to an indicator random field which is in turn related to the occurrence of different processes within the domain of observation. The model is seen to embed within a unique theoretical framework the main traits arising in the stochastic analysis of the system. **Chapter 4** is devoted to the presentation of the detailed results of the experimental and theoretical investigations, while conclusions and future perspectives are offered in **Chapter 5**.

Sommario

Questa tesi di dottorato è incentrata sullo studio di processi reattivi che avvengono all'interfaccia solido-liquido in mezzi porosi. Tra i diversi tipi di processi, particolare rilievo assumono quelli di precipitazione/dissoluzione. Questi sono infatti alla base dei fenomeni di erosione e alterazione chimica dei mezzi porosi naturali e determinano la formazione di percorsi preferenziali nel sottosuolo. Nel contesto dei geomateriali naturali, le roccie carbonatiche sono considerate di particolare interesse, in quanto costituenti principali della crosta superficiale terrestre. Esse sono caratterizzate da un'elevata reattività che ne favorisce l'interazione con numerosi inquinanti presenti nelle falde acquifere e sono considerate uno dei geomateriali ottimali per il sequestro di CO_2.

Le domande di ricerca affrontate in questa tesi di dottorato riguardano (a) l'avanzamento della capacità di osservare direttamente i tassi di reazione di precipitazione/dissoluzione mediante esperimenti originali alla nano-/microscala e (b) lo sviluppo di rigorose formulazioni matematiche finalizzate ad interpretare i tassi misurati in un contesto stocastico in grado di quantificarne l'incertezza associata. Questi due aspetti sono affrontati combinando metodi sperimentali innovativi e avanzamenti teorici/modellistici. Gli esperimenti sono effettuati mediante Microscopia a Forza Atomica (AFM). Questa tecnologia ad elevata risoluzione consente di osservare direttamente i processi meccanicistici che governano la cinetica di reazione. La modellazione dei risultati è basata su un approccio stocastico ispirato a formulazioni tipiche dell'ambito geostatistico. La tesi è articolata in cinque capitoli brevemente descritti nel seguito.

Il **Capitolo 1** costituisce l'introduzione al lavoro di tesi. Si descrivono le condizioni idro-geochimiche considerate e lo stato dell'arte. Il **Capitolo 2** offre una panoramica dei protocolli sperimentali sviluppati durante il percorso di dottorato. Questi ultimi consentono di riprodurre condizioni tipiche dei processi di dissoluzione in mezzi porosi ingegnerizzati e/o naturali. In particolare, si illustrano i principali aspetti dei setup sviluppati per simulare condizioni tipiche di regioni (i) stagnanti e (ii) caratterizzate da velocità non trascurabili. All'interno delle prime, i processi reattivi sono dominati dalla diffusione, delle seconde dalla

reattività della superficie. I setup sperimentali sviluppati permettono di acquisire mappe spaziali della topografia superficiale di un minerale sottoposto a reazione a diretto contatto con un fluido a diversi istanti temporali. Tali dati sono a loro volta utilizzati per ottenere campi spaziali di tassi di reazione, caratterizzati secondo un approccio stocastico. Il **Capitolo 3** illustra i dettagli delle formulazioni matematiche impiegate per l'interpretazione delle osservazioni sperimentali. L'approccio seguito considera in un'unica formulazione teorica la distribuzione di probabilità di una variabile di interesse e degli incrementi spaziali ad essa associati, valutati tra punti separati da una data distanza spaziale (o *lag*). Le mappe di tassi di reazione sperimentali sono considerate come funzioni random e sono interpretate mediante una miscela bimodale di campi non Gaussiani. Le diverse regioni spaziali che determinano la presenza di diversi modi nella densità di probabilità campionaria sono identificate mediante una variabile indicatore random. Quest'ultima è legata all'occorrenza di diversi processi meccanicistici all'interno della stessa finestra di osservazione. In tal senso, i modelli matematici sviluppati incorporano in un'unica formulazione teorica la presenza dei diversi fenomeni che contribuiscono alla reazione di dissoluzione. Nel **Capitolo 4** si discutono i risultati sperimentali e la loro interpretazione mediante un approccio stocastico. Il **Capitolo 5** presenta le principali conclusioni del lavoro di tesi e le prospettive future.

Résumé

La recherche présentée dans cette thèse se concentre sur l'exploration de la nature des processus de flux et de transport à travers les géomatériaux poreux et de leurs interactions/rétroactions avec la matrice solide. Les roches naturelles sont le milieu poreux à travers lequel les flux d'eau (et d'autres fluides) et le transport de solutés s'effectuent à travers le sous-sol terrestre. Dans ce contexte, la dissolution est un processus clé qui entraîne les transformations minérales qui ont lieu dans la croûte supérieure de la Terre. Ce processus contribue à la formation de voies préférentielles à travers le sous-sol. Les roches carbonatées sont d'un intérêt particulier. Cela est dû à leur abondance dans le sous-sol et à leur haute réactivité. Elles entraînent l'altération de la surface de la Terre, interagissent avec la propagation des polluants dans les aquifères souterrains, et favorisent la capture du carbone par processus comme la minéralisation. Les principales questions de recherche abordées dans cette thèse de doctorat sont liées à (*a*) l'amélioration de notre capacité à observer directement les taux de réaction de précipitation/dissolution grâce à des nouvelles expériences d'imagerie à l'échelle nano/micro et (*b*) la fourniture d'une interprétation des taux observés grâce à des approches stochastiques rigoureuses capables de quantifier l'incertitude. Ces questions sont abordées avec un mélange unique de théoriques/modélisation et d'expériences de laboratoire. En tant que tel, les méthodes sont basées sur des développements théoriques fondamentaux (avec des approches soit analytiques que numériques) et des manipulations. Les approches théoriques sont de nature stochastique, étant donné l'incertitude omniprésente sur les mécanismes qui entraînent la dynamique de l'altération de la surface des minéraux en contact avec les fluides. Des expériences originales sont effectuées directement à l'échelle micro et nano pour observer par imagerie la physique fondamentale de l'interaction entre les fluides et la surface minérale. La thèse est structurée le long des cinq chapitres brièvement décrits dans ce qui suit. Le **Chapitre 1** fournit une introduction au travail, y compris une description générale des paramètres hydro-géo-chimiques analysés. Le **Chapitre 2** donne un aperçu du cadre expérimental considéré et conçu pendant le parcours de doctorat. Ils sont illustrés les protocoles expérimentaux, conçus pour

acquérir des images de Microscopie à Force Atomique (AFM), imitant les paramètres naturels typiques des processus de dissolution de la calcite qui se produisant dans divers scénarios d'ingénierie/environnementaux. Une variété de configurations et de protocoles expérimentaux sont aussi conçus pour ressembler aux conditions associées à (*i*) des réactions dominées par la diffusion et (*ii*) des réactions contrôlées par la surface qui sont typiques des régions stagnantes ou d'écoulement dans les géo-matériaux poreux. Ces protocoles sont documentés pour permettre d'obtenir des données expérimentales qui peuvent être rapidement utilisées pour l'évaluation des distributions spatio-temporelles de la topographie de surface du minéral en contact avec le fluide en mouvement. Celles-ci sont à leur tour utilisées pour évaluer les distributions spatio-temporelles des taux de réaction qui s'ensuivent. Ces derniers sont alors soumis à des caractérisations stochastiques originales. En ce sens, le **Chapitre 3** comprend tous les détails du cadre théorique stochastique utilisé pour l'analyse et l'interprétation des manipulations expérimentales. L'approche, et les formulations qui en dérivent, intègrent l'évaluation conjointe de la distribution de probabilité d'une variable *target* et de ses incréments spatiaux associés, pris entre des emplacements séparés par une distance (*lag*) donnée. Le champ aléatoire associé aux taux de réaction est interprété à travers un modèle de mélange bimodal généralement non-Gaussien. Les modes de ce dernier correspondent à un champ aléatoire indicateur qui est à son tour lié à l'occurrence de différents processus à l'intérieur du domaine d'observation. Le modèle intègre dans un cadre théorique unique les principaux traits qui se dégagent de l'analyse stochastique du système. Le **Chapitre 4** est consacré à la présentation des résultats détaillés des investigations expérimentales et théoriques, tandis que les conclusions et les perspectives futures sont offertes dans le **Chapitre 5**.

Keywords

Dissolution Kinetics | Atomic Force Microscopy | Calcite | Rate Spectra | Spatial Heterogeneity | Stochastic Modeling

Contents

List of Figures

List of Tables

Acronyms

AFM Atomic Force Microscopy

CDF Cumulative Distribution Function

DHM Digital Holographic Microscopy

EM Expectation-Maximization

GMIX Gaussian Mixture

GEV Generalized Extreme Value

GSG Generalized sub-Gaussian

GSG-MIX Generalized sub-Gaussian Mixture

KIC Kashyap Information Criterion

ML Maximum Likelihood

mP monolayer etch pit

MP multilayer etch pit

NLL negative Log Likelihood

PDF Probability Density Function

RTM Reactive Transport Models

tfBm truncated fractional Brownian motion

TPV Truncated Power Variogram

VSI Vertical Scanning Interferometry

1 Introduction

This PhD dissertation is geared towards characterization and modeling of reactive phenomena taking place across a porous medium. Accurate modeling of such reactive processes is generally constrained by various sources of uncertainties. These arise from our lack of knowledge of mechanistic components driving reaction kinetics that, in turn, induce spatial (and temporal) variations of physical and/or chemical properties of the system. In this context, dissolution and precipitation phenomena taking place at the interface between minerals and fluids are key drivers of subsurface chemical weathering. They significantly contribute to modifying surface roughness of rocks, thereby influencing surface reactivity and/or leading to localized changes in flow paths (Fischer and Lüttge, 2017). These dynamic feedbacks contribute to shape the internal architecture of natural (and engineered) porous materials. Thus, quantitative assessment of environmental and industrial scenarios of pressing concern such as, e.g., long term stability of geogenic carbon sequestration (Fitts and Peters, 2013; Noiriel and Daval, 2017; Daval, 2018) or nuclear waste disposal systems (Arcos et al., 2008; Ewing, 2015), alteration of reservoir storage capacity (Mangane et al., 2013; Lamy-Chappuis et al., 2014), mobilization and ensuing migration of contaminants in aquifer systems (Dong et al., 2005; Harrison et al., 2017), geothermal energy exploitation (Erol et al., 2019), enhanced (conventional and unconventional) management of subsurface energy resources (Khather et al., 2020), morphology changes of fractures in deep reservoirs (Noiriel et al., 2007) with possible consequences on the overlying geological formations, as well as deterioration of concrete and/or marble structures (Kanellopoulou and Koutsoukos, 2003) eventually associated with cultural heritage settings, ultimately hinges on our ability to accurately assess dissolution/precipitation rates.

Bulk powder experiments are typically considered as a pillar for the estimation of dissolution rates (Lüttge et al., 2013a). These are inferred by monitoring changes in dissolved solute concentrations. By relying upon this technique, significant efforts have been devoted to evaluate reaction rates of a variety of minerals subject to diverse chemical conditions (see, e.g., Hellmann and Tisserand, 2006, Nagy and Lasaga, 1992, White and Brantley, 2003, and Arvidson et al., 2003). Rates estimated across such extensive experimental datasets exhibit a wide range of values, even as they are associated with a single mineral species and experiments are conducted upon relying on highly reproducible protocols (Arvidson et al., 2003; Bollermann and Fischer, 2020). Such variability can be only partly tied to the long-standing debate about the most appropriate type of surface measurement that should be employed to normalize bulk rates to obtain proper units, i.e., mass per unit area per unit time (Fischer et al., 2012; Arvidson et al., 2003). Drawbacks associated with this issue have been recently circumvented through the use of advanced high-resolution imaging techniques such as Atomic Force Microscopy (AFM), Vertical Scanning Interferometry (VSI), or Digital Holographic Microscopy (DHM). These enable one to measure directly the local growth/retreat of the mineral surface with respect to a reference plane. The latter corresponds to a portion of the surface where the reaction is inhibited by means of a non-reactive mask (see, e.g., Bouissonnié et al., 2018, Fischer and Lüttge, 2017, and Emmanuel, 2014). Deposition of such inert layer on a mineral sample is typically performed after mechanical polishing of the surface. This preliminary step is documented to impact dissolution rates by inducing the formation of additional cracks and dislocations in the mineral structure (Lucca et al., 2006). Measurements of average surface retreat with respect to the inert layer can be employed to estimate a mean dissolution rate, $\langle R \rangle$ $[\mathrm{mol} \cdot \mathrm{m}^{-2}\mathrm{s}^{-1}]$, as

$$\langle R \rangle = \frac{1}{V_m} \frac{\langle \Delta h \rangle}{\Delta t}, \tag{1.1}$$

where Δh [m] is the surface retreat evaluated with respect to the reference ($\langle \cdot \rangle$ denoting spatial averaging); Δt [s] is the duration of the experiment; and V_m $[\mathrm{m}^3\ \mathrm{mol}^{-1}]$ is calcite molar volume.

However, relying on average quantities of the kind associated with Eq. 1.1 leads to a complete loss of information about the often marked degree of heterogeneity of local surface topography (and hence rate) values that is revealed by high-resolution imaging techniques. The latter stems from local inhomogeneities and defects in the crystal lattice. These are usually regarded as sources of intrinsic variability and lead to uneven distributions of surface energy. Their impact on the dissolution process has been identified as the ultimate cause of the wide range of average rate estimates under given experimental conditions (Lüttge et al., 2013a).

The ensuing markedly heterogeneous distributions of rates that are then observed across a mineral surface cast some doubts about the robustness of a conceptual picture relying on a description of reaction kinetics based solely on average values. Since the density and the distribution of defects in the mineral lattice display traits which are typical of random processes, a shift of paradigm towards the reliance on a stochastic approach has been encouraged by several authors to capture the spatial variability of dissolution rates (Fischer et al., 2014, 2012; Lüttge et al., 2013a; Lüttge et al., 2019). Fischer et al. (2012) introduced the concept of *rate spectra*, i.e., sample probability density functions (PDFs) associated with spatially heterogeneous fields of reaction rate, $R(\mathbf{x})$, as a tool to preserve the richness of surface features contributing to the overall reactive process. In this context, Emmanuel (2014) relies on AFM to evaluate rate spectra of a polished dolostone sample subject to reaction at $p\text{H} = 3.5 - 4.5$ at temperature $T = 29.5 \pm 1°\text{C}$ under continuous flow. All of them exhibit similar statistical traits regardless of the pH and document long right tailing of sample PDFs of R. Similar right-skewed behaviors of sample frequency distributions are observed by Brand et al. (2017), who rely on DHM to measure the topography of a polished calcite sample subject to dissolution with deionized water under continuous flow in a flow through cell at $T = 23 \pm 1°\text{C}$. Additionally, multimodal traits of sample rate spectra are documented by Fischer and Lüttge (2017) and Bibi et al. (2018), relying upon VSI. Fischer and Lüttge (2017) induce dissolution of a polished polycrystalline calcite marble at $p\text{H} = 9.2$ in a flow-through cell. Bibi et al. (2018) image a freshly cleaved calcite surface subject to reaction at far-from-equilibrium conditions under quasi-continuum flow at $p\mathrm{H} = 8.82$ and temperature $T = 22°\mathrm{C}$. Other typical traits observed for rate spectra include (*i*) sharp peaks and heavy tails of the sample frequency distribution; (*ii*) a scaling tendency of sample distributions of associated spatial increments, $\Delta R(s) = R(\mathbf{x}) - R(\mathbf{x} + s)$, their shape depending on the separation distance (or *lag*), s, at which these are evaluated; (*iii*) multimodal traits of sample PDFs of ΔR when multiple mechanistic processes contribute to drive the reaction. Such distinctive features are not consistent with a description of the statistical behavior of rates resting on a standard Gaussian model. Preliminary modeling efforts aimed at capturing documented traits of sample rate spectra are grounded on a Generalized Extreme Value (GEV) framework (Brand et al., 2017; Emmanuel, 2014). While relying on the latter allows successfully capturing the tailing trend, it otherwise fails at representing the typically observed multimodal behaviors.

Non-Gaussian traits of the above mentioned kinds are also documented for a wide variety of other hydrogeological variables. These include, e.g., permeability (Painter, 1996; Riva et al., 2013), porosity (Painter, 1996; Guadagnini et al., 2014, 2015), hydraulic conductivity (Liu and Molz, 1997; Guadagnini et al., 2013; Meerschaert et al., 2004), electrical resistivity (Yang et al., 2009), soil and sediment texture (Guadagnini et al., 2014), rainfall (Kumar and Foufoula-Georgiou, 1993), and sediment transport rate (Ganti et al., 2009). In this context, Riva et al. (2015a) and Guadagnini et al. (2018) suggest that the way the PDF of spatial increments of a quantity of investigation scale with lag can be captured through a Generalized sub-Gaussian (GSG) model. In contrast to other modeling frameworks capable of reproducing the scale dependence of increments such as multifractals (e.g., Boffetta et al., 2008, Frisch, 2016, Lovejoy and Schertzer, 1995, and Veneziano et al., 2006), fractional Laplace (e.g., Kozubowski et al., 2006, Kozubowski et al., 2013, and Meerschaert et al., 2004) and sub-Gaussian (e.g., Samorodnitsky and Taqqu, 1994 and Guadagnini et al., 2012) formulations, the GSG model enables one to represent under a unique conceptual picture the documented behavior (as described by probability distributions and/or moments) of a random function and its increments. In this sense, joint analysis of the PDF of data and their increments within a unique theoretical framework that ensures consistency between these two types of information yields improved characterization of the quantity under investigation. It is noted that the GSG framework includes the Gaussian model as a particular case. Recent analyses by Siena et al. (2020) document the capability of the GSG model to interpret the statistics of VSI-based measurements of calcite surface topography subject to dissolution in close-to-equilibrium conditions. Under such chemical settings, dissolution of calcite takes place solely by step retreat (see, e.g., Teng, 2004), i.e., the reaction is driven by a single kinetic mechanism. Hence, a conceptual description relying upon a unimodal distribution such as the one encompassed in the GSG formulation is fully capable to capture the statistical traits of surface roughness induced by step retreat.

Otherwise, multimodal traits typically observed for rate spectra have been associated with the occurrence of diverse mechanisms contributing to the retreat of the crystal surface (Fischer et al., 2015; Fischer and Lüttge, 2017; Brand et al., 2017). In the hydrogeological context, multimodal behaviors of variables such as, e.g., hydraulic conductivity have been recognized to arise from a homogenization within a unique population of values that are otherwise linked to regions characterized by differing geological attributes (Journel, 1983; Desbarats, 1987; Rubin and Journel, 1991; Rubin, 1995). In this framework, a stochastic approach based on a model encompassing a unique scale of heterogeneity might not be adequate to represent such composite media within which various processes and/or geomaterials coexist across a given spatial window of observation. A description of a spatial random field as a statistically stationary system characterized through a multimodal model entails considering (*i*) the random geometry of the various regions (or clusters) identified across the system and (*ii*) the spatial distribution of the quantity of interest within each of these regions (Winter et al., 2003). In this setting, Rubin and Journel (1991) view the random function of interest as a sum of $m = 1, ..., M$ Gaussian components, each weighted by a (statistically homogeneous) random indicator function. These authors associate the latter with the spatial distribution of the M zones/clusters across the domain. Each $m-th$ random field is described through its spatial structure and is typically assumed to be independent from the others and from the indicator function. Rubin (1995) considers a porous system composed by $M = 2$ distinct modes and provides analytical formulations for the first two statistical moments (i.e., mean and variance) and for the covariance of the variable of investigation. Lu and Zhang (2002) further extend the above mentioned studies to include in the theoretical formulation a relationship between the covariance structure of the indicator and a characteristic length describing the spatial arrangement of the zones associated with the various geomaterials/processes. Such a length scale is characterized using a Markov chain model, as expressed by Carle and Fogg (1997). Recent applications of these concepts are illustrated by, e.g., Dai et al. (2020), to assess the impact of the internal architecture of a sedimentary porous medium on solute plume dispersion; Gournelos et al. (2020), for the interpretation of the statistical behavior of monthly water discharge and suspended sediment load; and Jia et al. (2022) for the simulation of synthetic long-term time series of streamflow data.

Modeling statistical traits of the kinds mentioned above requires experimental protocols that provide high quality and abundant data enabling an accurate assessment of extreme values driving tailing behavior of sample rate spectra. Having at our disposal a rich dataset is also critical when considering the process of deconvolution of the PDF for an accurate characterization of the statistical traits of each of its components. Experimental protocols should also be designed to replicate conditions that mimic those that are typical of relevant natural scenarios. Achieving this can be exceedingly challenging due to limitations of high-resolution imaging techniques. Atomic Force Microscopy (AFM) is characterized by horizontal and vertical resolutions of the order of $\sim 0.1\,\mathrm{nm}$ and $\sim 1\,\mathrm{nm}$, respectively. It offers exceptional insights into the phenomena underlying a reaction and its dynamics. Nevertheless, the vertical scanning range of the AFM scanner and the extent of the field of view are generally restricted to a few microns (a typical range is, e.g., $\pm 7\,\mu\mathrm{m}$ and $100\,\mu\mathrm{m}^2$ respectively, see, e.g., Eaton and West, 2010). Scanning artifacts can arise from abrupt changes in surface properties and roughness (Marinello et al., 2010). This still constitutes a challenge to obtaining high quality imaging of areas encompassing both inert and reactive layers. The presence of an inert mask is otherwise required to obtain absolute values of spatial topographies of crystal surfaces and, ultimately, of local reaction rates. An additional barrier still hampering the routine use of AFM for some environmentally relevant scenario is that the technique is well-suited for *in situ* measurements in static liquid settings. Otherwise, environmental scenarios sometimes entail reactive processes taking place at the interface of a solid surface and a flowing fluid. When it comes to continuous flow conditions, the quality of AFM measurements can deteriorate due to the interaction between the flow field and the movement of the cantilever. Thus, imaging is often conducted under intermittent flowing conditions by refreshing the solution in contact with the sample between each scanning session (see, e.g., Guren et al., 2020; Renard et al., 2018, 2019). On the other hand, VSI enables exploitation of wider fields and rougher surfaces with lower resolution than AFM. VSI and AFM are thus generally considered to address different aspects associated with reactive processes (Kurganskaya et al., 2012). The quantitative evaluation of spatial distributions of dissolution rates is typically accomplished through VSI (Bibi et al., 2018; Lüttge et al., 2019; Bouissonnié et al., 2018; Fischer and Lüttge, 2017), while AFM observations are usually employed to assess surface morphology and monitor its changes, as induced by variations in chemical conditions (see e.g., Ruiz-Agudo et al., 2009, Ruiz-Agudo et al., 2011, and Dong et al., 2020). Nanoscale quantitative rate estimates are generally limited to horizontal velocities of etch pit spreading/step retreat (see, e.g., Harstad and Stipp, 2007, and Ruiz-Agudo and Putnis, 2012). The sole example documenting the use of AFM to assess absolute reaction rates can be found in Emmanuel (2014).

This PhD dissertation integrates innovative experimental workflows and theoretical modeling developments keyed to (*i*) overcoming the above mentioned experimental barriers and (*ii*) developing a general interpretive stochastic framework. Our experimental and modeling workflows are tailored to evaluate and statistically characterize either dissolution or precipitation rates driven by various competing and/or coexisting processes acting at the nanoscale level upon fully embedding the possibility of recognizing such processes. As an illustrative example, we focus on spatial distributions of rates across the surface of a single calcite crystal subject to dissolution. Calcite is selected as an exemplary mineral because of its significance in environmental contexts as part of the carbon cycle (Bouissonnié et al., 2018) and as a key potential sink for toxic compounds such as Arsenic (Renard et al., 2019), Selenium (Heberling et al., 2014), Cadmium (Julia et al., 2023; Chada et al., 2005), Lead (Chada et al., 2005), Antimony (Renard et al., 2018), and Chromium (Guren et al., 2020). We induce the dissolution reaction at far-from-equilibrium conditions, i.e., the solution is highly undersaturated with respect to calcium. Under such setting, the process is governed by the nucleation of etch pits, which form in the presence of screw dislocations in the crystal lattice or randomly on flat terraces (Teng, 2004). Lateral spreading of these features occurs by radiation of steps that then travel across the surface (Lasaga and Lüttge, 2001, 2003).

Our investigations are grounded on the use of AFM. We design experimental approaches conducive to the establishment of experimental conditions that are representative of relevant environmental/engineering scenarios. These include (*i*) diffusion-dominated and (*ii*) surface-controlled reactions that are typical of stagnant and flowing regions in porous materials, respectively. Assessment of the reliability of the designed experimental setups rests on qualitative and quantitative analyses of an extensive collection of AFM images. We establish an experimental protocol to (*i*) fabricate and (*ii*) image through AFM mineral surfaces that are partially covered by an inert mask. As opposed to typical procedures relying upon deposition of an inert layer after mechanical surface polishing, the protocol developed in this PhD thesis enables direct fabrication of the mask on pristine crystal surfaces.

We propose comprehensive stochastic modeling frameworks capable to fully capture heterogeneous behaviors underpinning rate spectra and associated spatial increments. We start from the well established formulation of the GSG model by Riva et al. (2015a). We rely on the latter to characterize the statistical traits of a collection of dissolution rate fields obtained from topographic images of a sample subject to dissolution under diffusion-limited conditions. The GSG model fully succeeds to capture non-Gaussian heavy tailedness and peak sharpness of sample rate spectra, together with the scaling tendency of associated spatial increments. In contrast, it fails to reproduce observed multimodal behaviors arising from the homogenized description of the otherwise composite nature of dissolution rate fields as driven by multiple kinetic processes.

With the aim of capturing multimodal traits/features, we extend the theoretical framework underpinning a stochastic description of a composite random field through a stationary multimodal distribution. We start from the conceptualization of a bimodal Gaussian Mixture (GMIX) proposed by Lu and Zhang (2002). We derive rigorous formulations associated with the PDF of spatial increments to embed the observed scaling tendencies of such distributions within a unique analytical modeling approach. We then provide a general procedure for the estimation of all parameters embedded in the GMIX model. The GMIX model is employed to characterize the statistical behavior of rate spectra evaluated from AFM topographies subject to dissolution under surface-controlled settings. We then go beyond the limits imposed by relying on a Gaussian modeling framework and provide a theoretically sound generalization of the GMIX model to mixtures of Generalized sub-Gaussian fields (GSG-MIX). The latter modeling approach includes the GSG and the GMIX modeling frameworks as particular cases. As such, it enables one to encapsulate non Gaussian traits driving the level of heterogeneity of each component of the mixture (corresponding to each process contributing to the reaction) within the homogenized description of the composite random field.

This PhD dissertation is structured as follows. **Chapter 2** provides an overview of the designed experimental settings. **Chapter 3** includes all of the details of the theoretical stochastic framework employed for the analysis and interpretation of the experimental evidences. **Chapter 4** is devoted to the presentation of the detailed results of the experimental and theoretical investigations, while conclusions and future perspectives are offered in **Chapter 5**.

2 Experimental methods

This chapter illustrates key elements of the experimental platforms developed in the context of this PhD thesis. The design of the setups is keyed to (*a*) provide high quality and abundant data and (*b*) mimic conditions that are representative of relevant environmental/engineering scenarios.

The former requirement is critical to obtain data that can be readily employed for evaluation and statistical characterization of associated reaction rate maps. Our experimental investigation rests on AFM high resolution imaging of (non-polished) calcite samples cleaved along the $\{104\}$ cleavage plane (see Section 2.1). In collaboration with Dr. Monica Bollani[1], we design a protocol enabling direct fabrication of an inert mask on calcite surface. We recall that this element is necessary to evaluate absolute material fluxes taking place across the crystal surface. The reliability of point-by-point dissolution rate fields from AFM data also depends on the alignment in the x-y plane of topographic maps acquired at subsequent times. This could be hampered by lateral drifts of the AFM piezoelectric scanner. Even as experimental strategies are designed to reduce the instrument drift, a residual misalignment might affect experimental data. We rest upon an image registration procedure relying on reference *fiducial points* to assess and possibly correct such image misalignment.

Fluid regulation systems are designed to reproduce (*a*) diffusion- and (*b*) surface-limited dissolution processes that are typical of natural and/or engineered scenarios. In the former case, imaging is performed under static conditions. In the latter case, fluid flow is regulated

---

[1] Istituto di Fotonica e Nanotecnologie - Consiglio Nazionale delle Ricerche, Laboratory for Nanostructure Epitaxy and Spintronics on Silicon (LNESS), Via Anzani 42, 22100 Como, Italy

through two different systems enabling one to subject the crystal sample to (i) intermittent and (ii) continuous flow conditions. The reliability of the designed settings is assessed upon analysis of an extensive collection of AFM topographic data. Quality requirements of AFM images are evaluated on the basis of rigorous analyses of the distortions of a reference calibration sample imaged under identical conditions.

The Chapter is structured as follows. Section 2.1 describes the main mechanisms driving the evolution of the surface pattern of a calcite crystal subject to dissolution reaction. Preliminary data processing and formulations leading to the evaluation of spatial maps of dissolution rate are described in Section 2.2. Key elements of the experimental settings and of the fluid regulation systems are illustrated in Section 2.3. The assessment of the reliability of the latter is described in Section 2.4. Section 2.5 illustrates the inert mask fabrication procedure. The workflow keyed to image registration in the horizontal plane of AFM data acquired at subsequent times is described in Section 2.6.

2.1 | Calcite dissolution mechanisms

Calcite crystal surfaces may be composed of different crystallographic planes (Paquette and Reeder, 1995). These are generally indicated through the so-called *Miller indices*, $\{hkl\}$ (see, e.g., De Graef and McHenry, 2012). Among calcite cleavage planes, the $\{104\}$ surface (sketched in Fig. 2.1.D) is usually considered as a reference due to its stability (Massaro et al., 2008) and abundance in natural environments. Hence, the assessment of the overall properties of calcite is typically grounded on studies targeting the $\{104\}$ surface (Heberling et al., 2014). The natural flatness of the $\{104\}$ plane together with an optimal cleavage along this crystallographic direction contribute to its particular suitability for AFM imaging. A schematic depiction of calcite $\{104\}$ crystal structure and of typical lattice defects is offered in Fig. 2.1.A. On this face, stable step edges are oriented along $[\bar{4}41]_+$, $[48\bar{1}]_+$, $[\bar{4}41]_-$, and $[48\bar{1}]_-$ crystallographic directions. These are not orthogonal to the cleavage plane. Instead, they are at either acute (78°) or obtuse (102°) angles (see Fig. 2.1.E) with it. Depending on their orientation, these steps exhibit a distinct reactivity in terms of growth/dissolution and incorporation of trace elements (Ruiz-Agudo et al., 2011), the highest reactivity being typically documented for obtuse steps.

Dissolution of calcite $\{104\}$ involves three main mechanisms, i.e., (i) retreat of existing steps, (ii) defect-assisted dissolution, and (iii) spontaneous nucleation of etch pits (Teng, 2004; Ruiz-Agudo et al., 2011). The prevalence of one mechanism over the others is governed by the distance from chemical equilibrium, which is in turn quantified through the saturation state of the solution, Ω. The latter is defined as the ratio between the product of component activities in solution, $IAP = a_{Ca^{2+}} \cdot a_{CO_3^{2-}}$, and at equilibrium, $K_{sp} = a_{Ca^{2+},eq} \cdot a_{CO_3^{2-},eq}$, i.e., $\Omega = IAP/K_{sp}$. By definition, $\Omega \in (0,1)$ for dissolution reactions. Saturation then approaches unity at equilibrium. The dependence of the dissolution pattern on Ω has been extensively studied (e.g., Teng, 2004, Smith et al., 2013, and Bouissonnié et al., 2018). Nucleation of etch pits is highly unfavored at close to equilibrium conditions (i.e., $\Omega > 0.8$, see Fig. 2.1.C) and step retreat is the main reaction driver. Otherwise, the formation of etch pits is the dominant mechanism at far-from-equilibrium conditions (i.e., $\Omega < 0.1-0.2$,

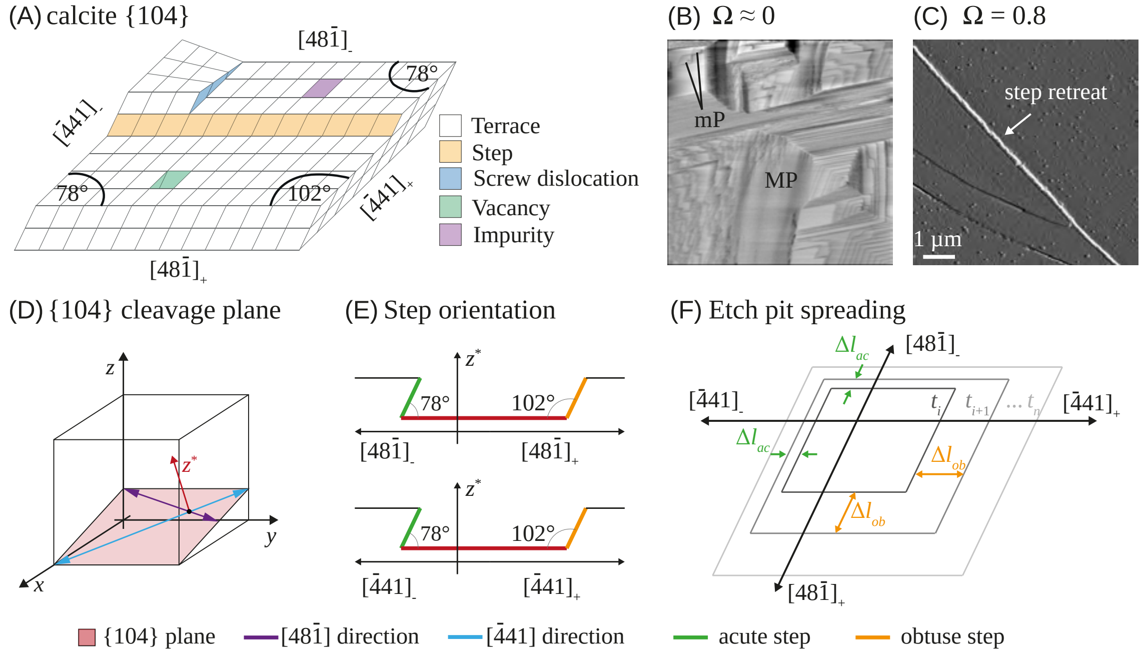

Figure 2.1: **(A)** Schematic representation of the {104} calcite rhombohedron with typical lattice defects. Exemplary AFM (friction signal) images of the {104} surface subject to dissolution at **(B)** far- and **(C)** close-to-equilibrium conditions, depicting mono- (mP) and multilayer (MP) etch pits and step retreat dominated surface patterns, respectively. **(E)** Vertical profile taken along $[48\bar{1}]$ and $[\bar{4}41]$ directions of (stable) step edges on the {104} crystallographic plane illustrated in **(D)**. **(F)** Sketch of stepwave radiation from the center of a MP.

see Fig. 2.1.B). These surface structures are characterized by a perfectly rhombohedral shape if $\Omega \sim 0$, whereas obtuse-obtuse edges become progressively round as saturation deviates from zero (Harstad and Stipp, 2007). The formation of a multilayer etch pit (MP) yields the removal of multiple layers of the crystal lattice, whereas the nucleation of a monolayer etch pit (mP) only dissolves one layer. Nucleation of the former features has been linked to screw dislocation and/or point defects in the crystal structure (e.g., Lüttge et al., 2019 and references therein). Defects of this kind induce a local excess of strain energy in the mineral lattice. The latter, in turn, leads to the development of a macroscopic feature, i.e., the MP. According to the stepwave model (Lasaga and Lüttge, 2001, 2003), the opening of an etch pit yields a local increase of the surface reactivity, as it leads to the formation of chains of steps or *stepwaves* that then travel across the surface of the crystal. These are illustrated in Fig. 2.1.F. Recent experimental observations by Fischer and Lüttge (2018) document that radiation of stepwaves follows a pulsating trend. This behavior has been firstly attributed by these authors to cyclic fluctuations in saturation levels at atomic steps. Recent investigations based on reactive transport and kinetic Monte Carlo modeling by Schabernack and Fischer (2024) suggest that the pulsating nature of dissolution reactions is driven by the inherent crystal reactivity at the atomistic level. If the reaction takes place under extremely far-from-equilibrium conditions (i.e., $\Omega < 0.007$), spontaneous nucleation of shallow etch pits at random locations on crystal terraces is also documented (see Fig. 2.1.B). As this mechanism is not defect-assisted, it may become dominant on dislocation-free surfaces.

2.2 | Evaluation of dissolution rates

Spatial maps of surface topography, $z(\mathbf{x}, t)$, acquired via AFM imaging at a given time t can be expressed as the sum of an average value, $\langle z(t) \rangle$, and a local zero-mean fluctuation, $z'(\mathbf{x}, t)$, i.e.,

$$z(\mathbf{x}, t) = \langle z(t) \rangle + z'(\mathbf{x}, t). \tag{2.1}$$

Hence, the spatial distribution of dissolution rates, $R(\mathbf{x}, t)$ [mol s^{-1} m^{-2}], can be obtained as

$$\begin{aligned} R(\mathbf{x}, t) &= \frac{1}{V_m} \frac{z(\mathbf{x}, t) - z(\mathbf{x}, t + \Delta t)}{\Delta t} \\ &= \frac{1}{V_m \Delta t} \left\{ \left[\langle z(t) \rangle - \langle z(t + \Delta t) \rangle \right] + \left[z'(\mathbf{x}, t) - z'(\mathbf{x}, t + \Delta t) \right] \right\} \\ &= \langle R(t) \rangle + R'(\mathbf{x}, t), \end{aligned} \tag{2.2}$$

where $V_m = 36.93$ cm^3 mol^{-1} is calcite molar volume and Δt [s] is the time interval between the acquisition of two topography images. The term $\langle R(t) \rangle$ represents an average dissolution rate across the whole surface, whereas the fluctuation term, $R'(\mathbf{x}, t)$, is informative about the spatial variability of the rates. AFM data often require a preliminary pre-processing step to remove spurious contributions. The latter generally include (*i*) tilting and (*ii*) an artificial curvature of the surface (see Fig. 2.2.B). The former is ascribed to the glue layer through which the crystal fragment is attached to the glass slide (see Section 2.3). The bow distortion observed across the $x - y$ plane arises from the swinging motion (schematically illustrated in Fig. 2.2.A) of the AFM piezoelectric tube during scanning (Ricci and Braga, 2004).

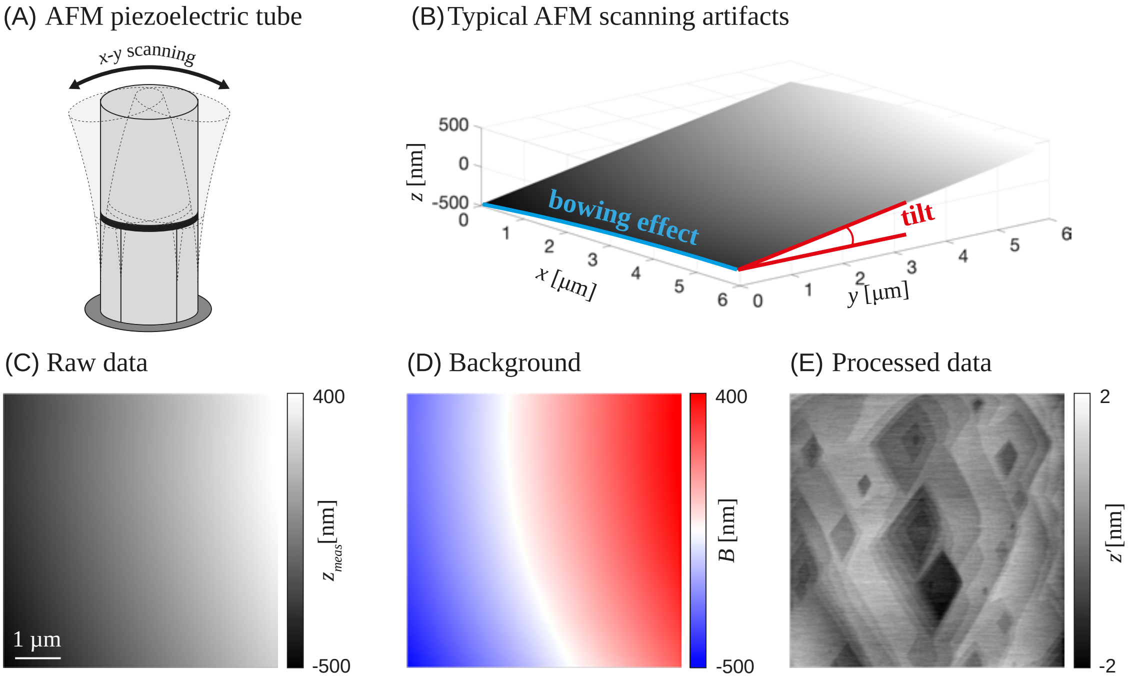
(A) AFM piezoelectric tube
(B) Typical AFM scanning artifacts
(C) Raw data
(D) Background
(E) Processed data

Figure 2.2: **(A)** Sketch of the swinging motion associated with the AFM piezoelectric tube during scanning across the $x - y$ plane. **(B)** Raw AFM data showing typical imaging artifacts, i.e., bowing effect and tilt. **(C)** Exemplary raw AFM topographic data, $z_{meas}(\mathbf{x}, t)$, acquired across a $6 \times 6\,\mu\text{m}^2$ observation window. **(E)** Spatial map of fluctuation of topography about the mean, $z'(\mathbf{x}, t)$, obtained after removal of the polynomial background ($B(\mathbf{x})$ depicted in **(D)**) from z_{meas}.

The actual surface topography/elevation, z, can be derived from data, according to $z(\mathbf{x}, t) = z_{meas}(\mathbf{x}, t) - S(\mathbf{x}, t)$, z_{meas} (Fig. 2.2.C) and S being the measured value and the overall distortion, respectively. The overall distortion can be inferred by fitting the measured data with a given polynomial function. If the entire surface of the sample is in contact with the solution, the reaction takes place at all locations on the crystal. Hence, the ensuing trend, $B(\mathbf{x}, t)$ (Fig. 2.2.D), necessarily includes the average elevation, i.e., $\langle z(t)\rangle$. This, in turn, implies that the detrended data, i.e.,

$$z_{meas}(\mathbf{x}, t) - B(\mathbf{x}, t) = z_{meas}(\mathbf{x}, t) - S(\mathbf{x}, t) - \langle z(t)\rangle, \tag{2.3}$$

correspond to $z'(\mathbf{x}, t)$ (Fig. 2.2.E) and the information about $\langle z(t)\rangle$ is lost. Therefore, the spatial map of rates obtained from the difference of the topography measurements at two diverse observation times corresponds to the dissolution rate fluctuations, i.e.,

$$R'(\mathbf{x}, t) = \frac{1}{V_m} \frac{z'(\mathbf{x}, t) - z'(\mathbf{x}, t + \Delta t)}{\Delta t}. \tag{2.4}$$

These types of information provide insights on the spatial variability of rates and are then subject to our statistical analyses. Note that the above-mentioned preliminary image processing is always required to quantify vertical retreats of the mineral surface.

Absolute dissolution rates, $R(\mathbf{x}, t)$, can then be obtained performing the above mentioned detrending procedure on an inert (i.e., non-reactive) region of the surface, whose elevation serves as a reference at all times. This enables one to remove spurious contributions associated with scanning artifacts and/or vertical drifts of the piezoelectric scanner while keeping the actual retreat. Such a region can be demarcated upon designing and applying an inert mask. Generally, mechanical polishing of the crystal surface is required prior to mask deposition (see, e.g., Bouissonnié et al., 2018, Fischer and Lüttge, 2017, and Emmanuel, 2014). However, the polishing procedure may lead to the formation of cracks at the level of the mineral lattice (see, e.g., Lucca et al., 2006). These, in turn, alter the dynamics of the mechanisms driving the dissolution reaction described in Section 2.1, that are otherwise naturally taking place on pristine calcite surfaces. In this context, we develop a novel experimental workflow enabling one to fabricate a metal mask on the sample surface without preliminary mechanical polishing (see Section 2.5).

It is further noted that resorting to a mask is not strictly required for the assessment of quantities that are defined along the horizontal plane such as, e.g., horizontal spreading rates of etch pits.

2.3 | Experimental setup

We design AFM scanning (Keysight 5500 apparatus) of a portion of the crystal surface, A_{obs}, across a uniform horizontal grid, forming a lattice of $n \times n$ elements. Scanning is performed either in contact-mode, similar to Renard et al. (2019), or in tapping mode. In the former case, we employ silicon tips (Bruker, RESPA-40) and an Al-coated cantilever characterized by an elastic constant $k = 5\,\mathrm{N/m}$. In the latter case, we rely on stiffer silicon tips (Al-coated; Nanosensors, PPP-NCHR, $k = 42$ Nm^{-1}). Acquisition parameters (i.e., size of the observation window, A_{obs}, number of grid elements, n, and scanning frequency, f_a) are selected as a trade-off between the need for (i) high image quality and abundance of data, i.e., high spatial resolution, and (ii) a temporal resolution, ΔT, enabling one to follow the temporal trends of the surface features. The former requirement is critical for accurate assessment of extreme values of sample PDFs and for characterizing the statistical behaviors of each identified dissolution mode. The latter enables one to capture the evolution of the surface morphology. Selected values for A_{obs}, n, and f_a are specified in Chapter 4 for each of the cases considered in this thesis work.

Preparation of a sample prior to starting the dissolution experiment begins by cleaving a crystal of Iceland spar (Mexico) along the {104} plane upon pressing with a razor blade to obtain a millimeter-size fragment with a height $< 1\,\mathrm{mm}$. Then, the sample is either directly used as such or subject to the mask fabrication procedure illustrated in Section 2.5, depending on the type of analysis to which experimental data are then subject. After preparation, the sample is placed on a glass slide and fluxed with nitrogen to remove any residues from cleavage. A fluid cell (of volume $V_{cell} \sim 2$ mL) open to air is mounted on the support plate and sealed with a Viton O-ring (Fig. 2.3.A). A schematic planar and lateral view of the cell is depicted in Fig. 2.3.B. The cell can be connected through two pipes (Tygon tubing, OD= $1/16''$, ID= $1/50''$) to each of the three different systems depicted in Figs. 2.3.C-E that can be employed to regulate fluid flow. All experiments are conducted at room temperature (i.e., $T = 22 \pm 1\,^{\circ}\mathrm{C}$) and at atmospheric pressure. In all settings, the dissolution process is started by injecting deionized MilliQ water (18.2 $\mathrm{M}\Omega \cdot \mathrm{cm}$) in the cell. Adopting *Setting 1* corresponds to consider a static fluid in the cell, promoting the formation of a diffusive boundary layer. Otherwise, *Settings 2* and *3* are employed to keep the sample at constant chemical conditions, favoring a surface-controlled reaction. This is achieved by minimizing the contact time between the solution and the sample. The key elements associated with each of these settings are described in the following.

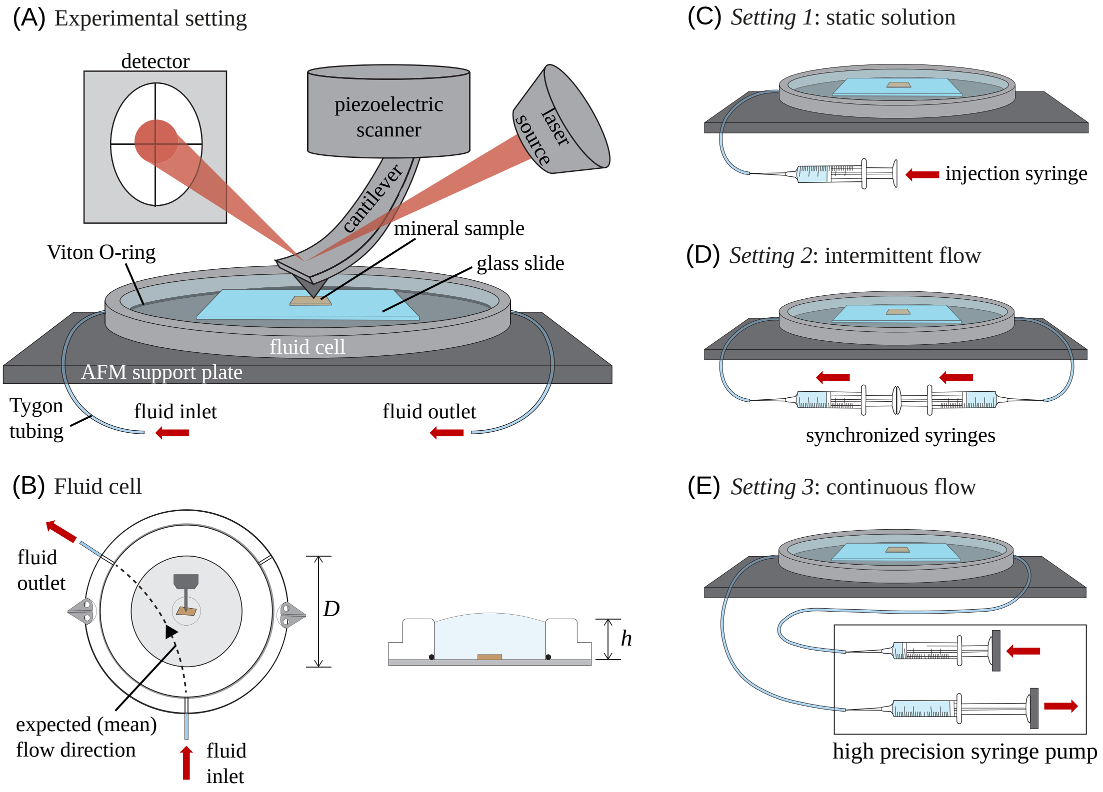

Figure 2.3: Schematic view of the experimental setting. Panel **(A)** depicts the main components of the AFM imaging system equipped with a fluid cell (of diameter $D = 22\,\mathrm{mm}$ and height $h = 5\,\mathrm{mm}$), whose planar and lateral views are depicted in **(B)**. Panels **(C)**, **(D)**, and **(E)** illustrate key elements of the systems employed to impose static, intermittent, and continuous flow conditions, respectively.

2.3.1 | *Setting 1*: STATIC

Fig. 2.3.C depicts experimental *Setting 1*. Prior to the beginning of the experiment, the sample is exposed to the solution for 1.5 h. The solution is then refreshed through a syringe

connected to the fluid inlet pipe and resides in the cell under static conditions for a time T_{c_1}. The dissolution of the mineral favors the development of a diffusive boundary layer around the sample where fluid saturation with respect to calcium increases in time. The surface pattern evolves in time as the main mechanism governing dissolution changes, driven by the increase of saturation. This setting mimics natural conditions that can be found in, e.g., dead-end pores in porous media, where the occurrence of chemical reactions is diffusion-dominated. Here, an appropriate modeling of reaction kinetics is critical as the dynamic system evolution is subject to variations of concentration due to diffusion. *Setting 1* is considered here as a reference for *Settings 2* and *3* to discriminate whether a given flow rate ensures consistency of the surface pattern in time (i.e., a flow rate is large enough to avoid development of the aforementioned diffusive boundary layer).

2.3.2 | *Setting 2*: INTERMITTENT FLOW

Experimental *Setting 2* is depicted in Fig. 2.3.D. The system comprises a mechanical device holding a pair of plastic syringes (of volume equal to 60 mL) whose pistons are mechanically coupled to withdraw and inject the same amount of solution. A given volume of fluid, V_s, is replaced in the cell between each AFM scanning. The (intermittent) flow rate, Q_2, is evaluated as

$$Q_2 = \frac{V_s}{T_{c_2}}, \tag{2.5}$$

where T_{c_2} is the residence time of the solution in the cell and coincides with the image acquisition time, ΔT. We set $V_s = 3\,\mathrm{mL} \sim 1.5 \cdot V_{cell}$ to ensure a complete replacement of the fluid in the cell. The fluid residence time $T_{c_2} = n/f_a$ is strongly linked to the AFM acquisition parameter set, n and f_a. Minimization of T_{c_2} can be achieved either by (*i*) setting a high scanning rate, f_a, or (*ii*) reducing the resolution, i.e., decreasing n (hence increasing the pixel size). Note that increasing the scanning rate yields a higher drift during AFM scanning. Otherwise, an increase in the pixel size results in loss of information due to coarsening of the spatial resolution. Hence, values of f_a and n are selected as a trade-off between between the need to have a short contact time and to obtain high quality

AFM images. We set the horizontal resolution to 11.7 nm and the scanning frequency to $f_a = 1.41\,\mathrm{Hz}$. This enables us to obtain a residence time of the fluid in the cell of $T_{c_2} = 6\,\mathrm{min}$. The resulting (intermittent) flow rate is $Q_2 = 8.33\,\mu\mathrm{L\ s}^{-1}$.

2.3.3 | *Setting 3*: CONTINUOUS FLOW

Figure 2.3.E depicts the key elements of experimental *Setting 3*. Here, the flow is controlled through a high precision syringe pump (CETONI, Nemesys S), equipped with two modules. These allow for independent injection and withdrawal of the solution in/from the system. A three way valve (CETONI, Countiflow Valve, 1/4″-28 UNF thread) is mounted on each module and a glass syringe (of volume equal to 1 mL) is directly screwed in one port of each valve. The other two ports are connected through flangeless fittings and ferrules (Elveflow, Flangeless PFA Fittings and ETFE Ferrules 1/4″-28 to 1/16″ OD) to the cell and to an external reservoir at atmospheric conditions, respectively. The delivering/withdrawing (pulse free) flow rate ranges between 0.003 and 100 $\mu\mathrm{L\ s}^{-1}$. The pump operating system is controlled through a custom code in the CETONI Elements software. We start by setting a continuous flow in the cell, upon delivering and withdrawing the solution with the same flow rate, Q_3, until the injection syringe is empty and the aspiration syringe is full. We then switch the valve to the external circuit, dispose the withdrawn solution and refill the new fluid from the corresponding reservoirs. Note that this second step is completely independent of the AFM acquisition. Therefore, this procedure can be performed at the maximum allowed flow rate, i.e., $Q = 100\,\mu\mathrm{L\ s}^{-1}$. Thus, the time required to restore the initial syringe conditions is equal to 10 s. This enables one to overcome limitations associated with syringe volumes.

The selection of Q_3 is a key element of this experimental setup, as it determines the contact time between the fluid and the mineral sample. In this context, we acquire temporal series of calcite topography associated with various flow rates and analyze qualitative and quantitative aspects associated with the observed dissolution pattern. We refer to the dimensionless Reynolds cell number defined as

$$Re = \frac{\rho L V}{\mu}, \tag{2.6}$$

where ρ [kg m^{-3}] and μ [Pa s] are fluid density and viscosity, respectively, V [m s^{-1}] is the average fluid velocity evaluated on the section where the sample is located, and L [m] is the characteristic length of the cell. We follow Schmidt and Alkire (1994) and consider L as the hydraulic diameter of the cell, i.e., $L = D_h$. The latter is defined as $D_h = (D \cdot h)/[2 \cdot (D + h)]$ where D [mm] and h [mm] are the diameter and the height of the cell, respectively (see Fig. 2.3.B). Our cell has $D = 22\,\text{mm}$ and $h = 5\,\text{mm}$. The average velocity, V, is then evaluated as $V = Q_3/(D \cdot h)$. Relying upon the evaluation of the Reynolds number enables one to replicate conditions in other experimental settings (i.e., using fluid cells associated with different geometries). Each temporal series is associated with an observation time of the crystal surface of 30 min. As documented by the results of *Setting 1* (see Section 2.3.1), such temporal window enables us to detect possible variations in the type of mechanisms driving the reaction. For each time series, we first analyze the qualitative features characterizing the dissolution pattern. At all times, these must be consistent with a process of dissolution at far-from-equilibrium conditions, i.e., traits associated with the development of a diffusive boundary layer (*Setting 1*) should not be observed. If saturation is stable, we expect that the reactivity of the surface remains (approximately) constant in time. This is assessed through analysis of the step retreat velocity, ν [nm s^{-1}]. Such quantity is defined as (Agudo and Putnis, 2012)

$$\nu = \frac{\Delta l_{mean}}{\Delta t}, \tag{2.7}$$

where $\Delta l_{mean} = (\Delta l_{ac} + \Delta l_{ob})/2$ [nm], Δl_{ac} and Δl_{ob} being the separation distance between acute (ac) or obtuse (ob) step edges at subsequent times (separated by a time step Δt [s]), respectively. As in Agudo and Putnis (2012), Δl_i is evaluated on the basis of the etch pit spreading rate rather than step retreat (see Fig. 2.1.F). At constant Ω, we expect slight variations of the spreading rate measured at different times, ν_t, with respect to its mean value, $\bar{\nu}$. Such variability is quantified through the coefficient of variation, $CV_\nu = \sigma_\nu/\bar{\nu}$, σ_ν

being the standard deviation of ν_t. We then analyze the way $\bar{\nu}$ and CV_ν vary with the Re, similar to Liang and Baer (1997).

The quality loss associated with the AFM signal as a result of the presence of fluid flow is assessed upon analyzing the distortion of reference sample images obtained with the same acquisition parameters and the same flow conditions as those associated with the analysis of the mineral. Such a sample is a calibration Silicon chip ($5 \times 5 \times 0.5$ mm^3, NanoTips, TGZ1) characterized by a regular two-dimensional (2D) pattern given by a parallel grating with a step height of 20 nm and period of 3 μm. We quantify image distortion as the variation in the horizontal length between two fixed points, $\hat{L}$, with respect to its reference value measured under static conditions, $\hat{L}_0$. We do so upon relying on the normalized root mean square error, NRMSE $= \sqrt{(\hat{L} - \hat{L}_0)^2}/\hat{L}_0$.

2.4 | Reliability of the designed settings

The reliability of the experimental settings introduced in Sections 2.3.1-2.3.3 is tested through acquisition and analysis of various temporal series of AFM data. These consist of collections of topography data acquired over a temporal observation window of 30 min. All experiments described in Sections 2.4.1-2.4.3 are performed employing unmasked freshly cleaved calcite samples. Experimental parameters are listed in Table 2.1 and include:

- the size of the observation window, A_{obs} [μm^2];
- the scanning frequency, f_a [Hz];
- the temporal resolution, ΔT [min];
- the flow rate, Q [μL s^{-1}];
- the Reynolds number, Re [$-$] (the latter is evaluated only for experiments performed resting on *Setting 3*);
- the contact time, T_c [min]. For *Setting 1* and *2*, the latter is defined as the time interval in which the fluid is left as static in the cell. When considering *Setting 3*, it is

defined as the time that a fluid particle takes to travel across the observation window, i.e., $T_{c_3} = L_{obs}/Q \cdot (D \cdot h)$, where L_{obs} is the lateral size of the observation window, D and h are the diameter and the height of the fluid cell, respectively;

- repetitions for each experiment, N_{exp}.

Setup	A_{obs} [μm^2]	f_a [Hz]	ΔT [min]	Q [μL s^{-1}]	Re [–]	T_c [min]	N_{exp} [–]
Setting 1	6	1.41	6	0	–	30	3
Setting 2	6	1.41	6	8.33 (*)	–	6	6
Setting 3	6	1.41	6	5	0.37	0.132	1
	6	1.41	6	8	0.59	0.083	1
	6	1.41	6	9	0.66	0.073	1
	6	1.41	6	10	0.74	0.066	3
	6	1.41	6	11	0.81	0.060	1
	6	1.41	6	12	0.89	0.055	3
	6	1.41	6	15	1.11	0.044	2

Table 2.1: Experimental parameter (i.e., size of the observation window, A_{obs}, scanning frequency, f_a, temporal resolution, ΔT, flow rate, Q, Reynolds number, Re, contact time, T_c, and repetitions for each experiment, N_{exp}) set to test the reliability of the designed experimental settings. The symbol (*) indicates that the flow rate is evaluated considering intermittent refreshment of the solution (see Section 2.4.2).

2.4.1 | *Setting 1*

We acquire temporal series of spatial topography. These comprise a collection of 5 images scanned along a 512×512 grid (pixel size 11.7 nm) upon setting the acquisition frequency to $f_a = 1.41$ Hz. This result in a temporal window of observation $T_{c_1} = 30$ min. Fig. 2.4.A-E depicts the pattern observed for a dissolution experiment performed under static solution conditions, i.e., upon relying on *Setting 1*. Each image is identified through the time elapsed from the injection of the solution. The area investigated comprises an acute step visible in the top-left corner (that is related to a deep etch pit of size larger than the width of the observation window) and a sequence of flat terraces separated by steps. Numerous monolayer etch pits nucleate at random locations on the terraces right after contact with deionized water (Fig. 2.4.A). This behavior is consistent with highly undersaturated conditions. As time elapses (Fig. 2.4.B), we observe spreading and coalescence of existing shallow etch pits rather than nucleation of new ones. Progressive rounding of obtuse-obtuse edges is also documented. Shallow etch pits are seen to merge with preexisting steps after $t_3 = 18\,\text{min}$ (Figs. 2.4.C-D). At time $t_5 = 30\,\text{min}$ (Fig. 2.4.E) almost all etch pits disappear and step retreat becomes the dominant mechanism. A similar temporal evolution of the surface pattern has been observed for other series of topography imaged on different calcite samples (see Appendix A). We also notice a temporal decrease in the reactivity of the surface. This is quantified through Fig. 2.4.G that depicts the spreading rate of the acute step edge in the top left corner, ν_{ac}, as a function of the time elapsed from solution injection. These results are obtained by making use of Eq. 2.7 upon considering only the separation distance between acute steps (depicted in Fig. 2.4.F for different times), i.e. Δl_{ac}. The dynamics of the surface pattern are strongly related to the local saturation at the fluid/solid interface. As such, the observed evolution of the surface is consistent with a local increase in the concentration of calcium within the developing boundary layer. No precipitates are observed at the last observation time, i.e., at $t_5 = 30\,\text{min}$. Hence, the local concentration of calcium is such that supersaturation has not been attained in the boundary layer.

2.4.2 | *Setting 2*

Fig. 2.5 depicts the temporal evolution of calcite surface imaged during experimental conditions representative of *Setting 2*. Similar features are observed at all times. These are consistent with the behavior expected at far-from-equilibrium conditions, i.e., nucleation, spreading and coalescence of mono- and multilayer etch pits, together with step retreat.

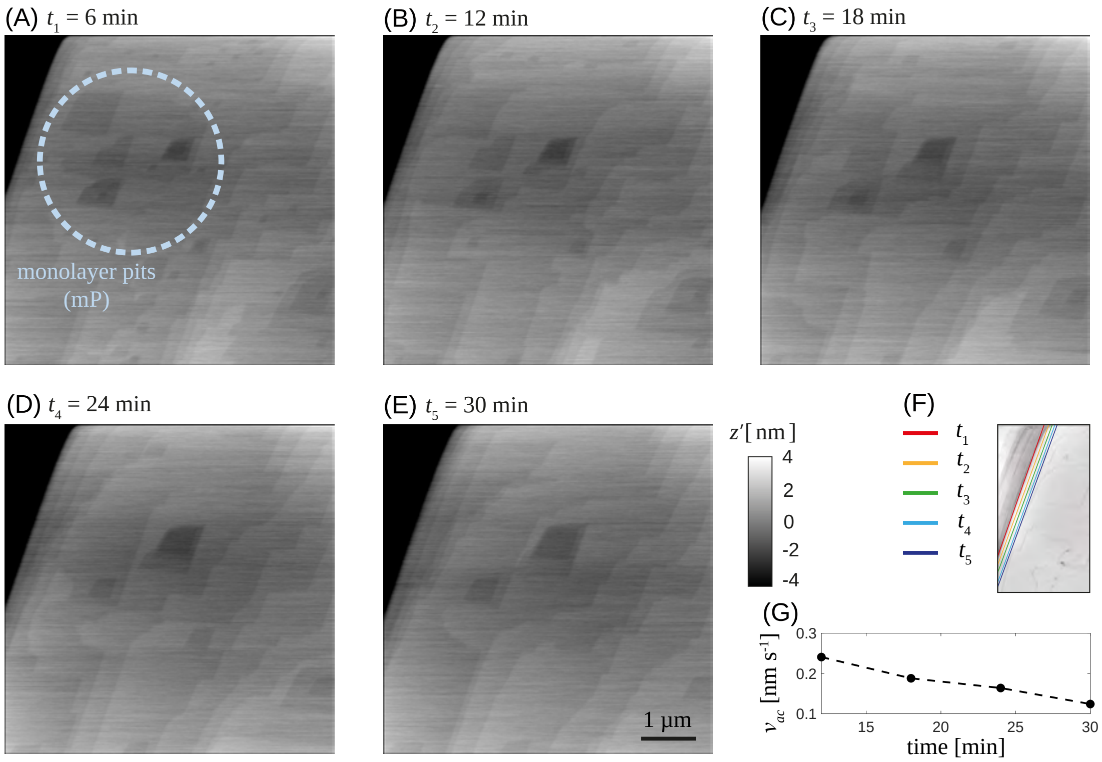

Figure 2.4: Exemplary topography of calcite surface imaged upon relying on *Setting 1* at observation times **(A)** t_1 = 6 min, **(B)** t_2 = 12 min, **(C)** t_3 = 18 min, **(D)** t_4 = 24 min, and **(E)** t_5 = 30 min (elapsed from first contact with the solution). Panel **(G)** depicts the spreading rate, ν_{ac}, associated with the acute step edges shown in **(F)** at various times.

Multilayer etch pits nucleate at the three preferential locations (A, B and C) depicted in Fig. 2.5.F. From these, trains of steps form, consistent with an interpretation based on the stepwave dissolution model (Lasaga and Lüttge, 2001, 2003). Diverse temporal evolutions are observed for each of the three regions. A deep etch pit nucleates in A at time t_1 (Fig. 2.5.A), whereas at locations B and C several small and shallow pits nucleate at the flat bottom of wide multilayer pits that have formed prior to the beginning of our acquisition. Horizontal spreading of the pit in A is observed at times t_2 and t_3 (Figs. 2.5.B-C). Simultaneously, shallow pits in B and C are observed to spread and coalesce, together with the formation of new small structures. The situation is reversed at time t_4, i.e., two deep etch pits nucleate at locations B and C, whereas the expansion of the pit in A leads to a flat bottom on which shallow pits spontaneously form. Numerous monolayer features are observed to nucleate at random locations on the terraces (Figs. 2.5.A-E) during the dynamic evolution of multilayer etch pits. These are ephemeral structures that either spread or merge with existing steps and/or other shallow pits as time progresses. These results suggest that the temporal interval and the resolution selected for image acquisition empower us with the ability to follow closely the evolution of the surface pattern.

Fig. 2.5.G depicts the temporal dependence of the number of shallow etch pits, n_{pit}, observed on the surface. This quantity remains approximately constant, consistent with the observation that chemical conditions in the cell are such that spontaneous nucleation of pits is promoted and that the degree of undersaturation is stable in time. Note that the density of etch pits on terraces is case specific, as it ultimately depends on the molecular structure of the crystal in the observed region. The etch pit density, ρ_{pit}, can be readily evaluated upon normalizing the number of pits by the area of the observed surface, i.e., $\rho_{pit} = n_{pit}/A_{obs}$ [cm^{-2}]. On average, we obtain $\rho_{pit} = 2.5 \cdot 10^8$ cm^{-2}, consistent with literature studies (Teng, 2004; Agudo and Putnis, 2012; Ruiz-Agudo et al., 2009). In this context, a joint analysis grounded on direct observations of the kind we provide and modeling tools such as, e.g., kinetic Monte Carlo simulations of dissolution reactions (see, e.g., Meakin and Rosso, 2008; Kurganskaya and Luttge, 2016; Rohlfs et al., 2018), might provide additional insights on the quantification of the relationship between the number of etch pit nucleating on the crystal surface and the amount of pre-existing defects in the crystal lattice. Rounding of obtuse-obtuse corners is also observed for multi- and monolayer etch pits. This behavior may be ascribed to (slightly) non-zero local saturations. Similar results have been obtained for other temporal series of topography imaged on different calcite samples (see Appendix B).

2.4.3 | *Setting 3*

Figs. 2.6.A-D depict four exemplary topography images selected from different temporal series associated with increasing values of Re. From a qualitative standpoint, we note that all images evidence nucleation of mono- and multilayer etch pits. Rounding of the

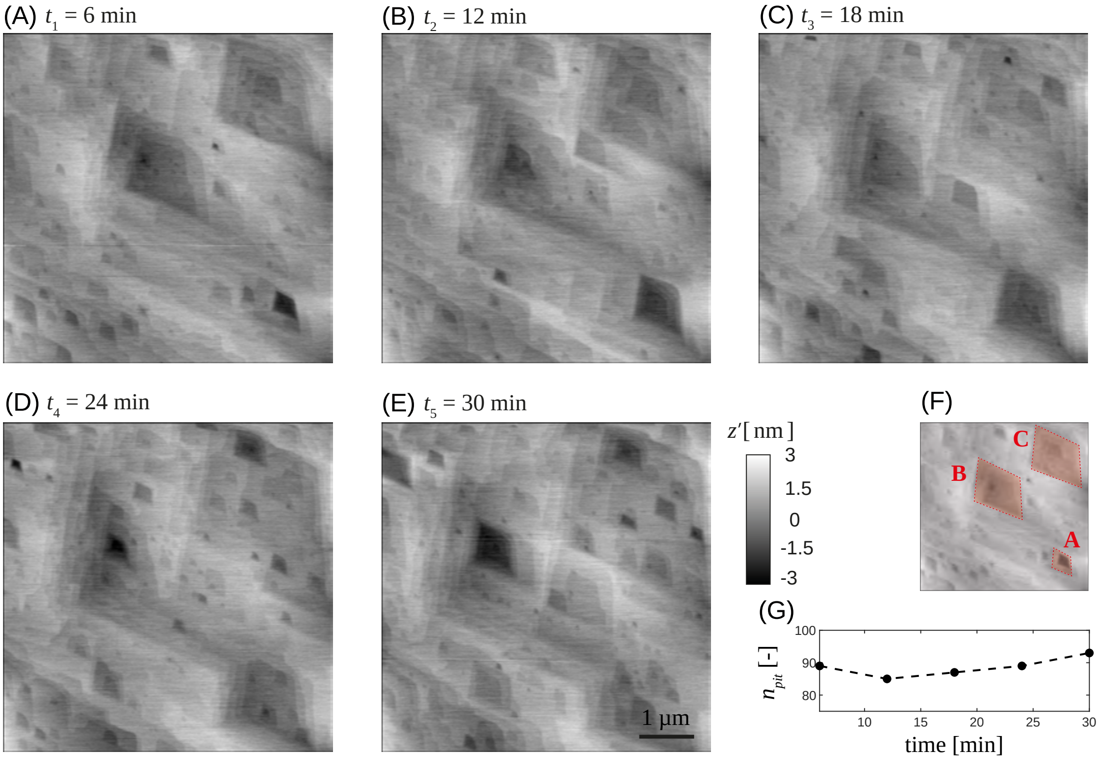

Figure 2.5: Exemplary topography of calcite surface imaged with *Setting 2* at observation times **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** highlights preferential nucleation regions of multilayer etch pits. The number of monolayer etch-pits, n_{pit}, versus time is depicted in **(G)**.

obtuse-obtuse edge is observed only for low values of the Reynolds number, i.e., $Re < 0.35$ (Fig. 2.6.A). Otherwise, etch pits display a perfectly rhombohedral shape for higher Re (Figs. 2.6.B-D). As expected, a higher amount of linear errors and a higher degree of stripe noise tend to emerge as Re increases, documenting a progressive degradation of the quality of AFM images with increasing flow rate. Note that, from a qualitative standpoint, measurement disturbances detected here upon increasing the flow rate are of the same kind as those evidenced by, e.g., Schmidt and Alkire (1994).

Figs. 2.6.E-H depict the reference sample imaged considering the same values of Re associated with calcite topographies shown above (Figs. 2.6.A-D). Following the approach suggested by Schmidt and Alkire (1994), we consider measurements (taken along the principal scanning direction) of horizontal distances between two points as references. We measure the size of the two steps, L^A and L^B (see Fig. 2.6.E). Such lengths are evaluated as the average among three measurements (L_k^j, $j = A, B$, $k = 1, 2, 3$) taken at different locations and are compared against their corresponding reference counterparts associated with static conditions, $L_{k,0}^j$. Fig. 2.6.L depicts the normalized root mean square error, NRMSE (see Section 2.3.3), as a function of Re. It is noted that NRMSE increases with Re for both L^A and L^B. Otherwise, it is also noted that the largest error at all Re is always less than 6%, documenting that AFM measurements taken along the principal scanning direction are not significantly altered by fluid flow. However, we observe that step edges in Figs. 2.6.E-H visually appear to be somehow distorted as Re increases. As edges are almost orthogonal to the principal scanning direction, such a behavior is ascribed to an increase in the AFM drift with Re. This might be attributed to the formation of local fluid re-circulation regions between the steps whose size and/or local velocity tends to increase with Re.

Fig. 2.6.I depicts the average spreading rate associated with each topography series, $\bar{\nu}$, and the ensuing coefficient of variation, CV_ν, as a function of Re. Otherwise, a decreasing trend followed by an increasing pattern seem to underlie the oscillating behavior of CV_ν with Re. This result is possibly a consequence of the increase in the drift of AFM images with the flow rate. Since no qualitative nor quantitative differences are documented in the surface pattern for $Re > 0.60$, whereas the amount of linear errors and the AFM drift (slightly) increase, we conclude that operating at $Re = 0.60 - 0.70$ can imbue us with confidence that dissolution occurs under stable chemical conditions while preserving good quality of topography images, i.e., reducing AFM drift and measurement noise. We complete the analysis by considering Figs. 2.7.A-E, which depicts an exemplary temporal evolution of calcite surface pattern imaged at $Re = 0.60$. Dissolution is initially ($t = 6\,\text{min}$) dominated by etch pits (locations A-F on Fig. 2.7.A). Only spreading of etch pit C is observed at the following observation time ($t = 12\,\text{min}$), while pits A-B and D-F merge with pre-existing steps and disappear. Subsequently ($t = 18 - 30$ min), the reaction is dominated by spreading of etch pit C and by the dynamics of small and shallow pits forming on its flat bottom. Exemplary

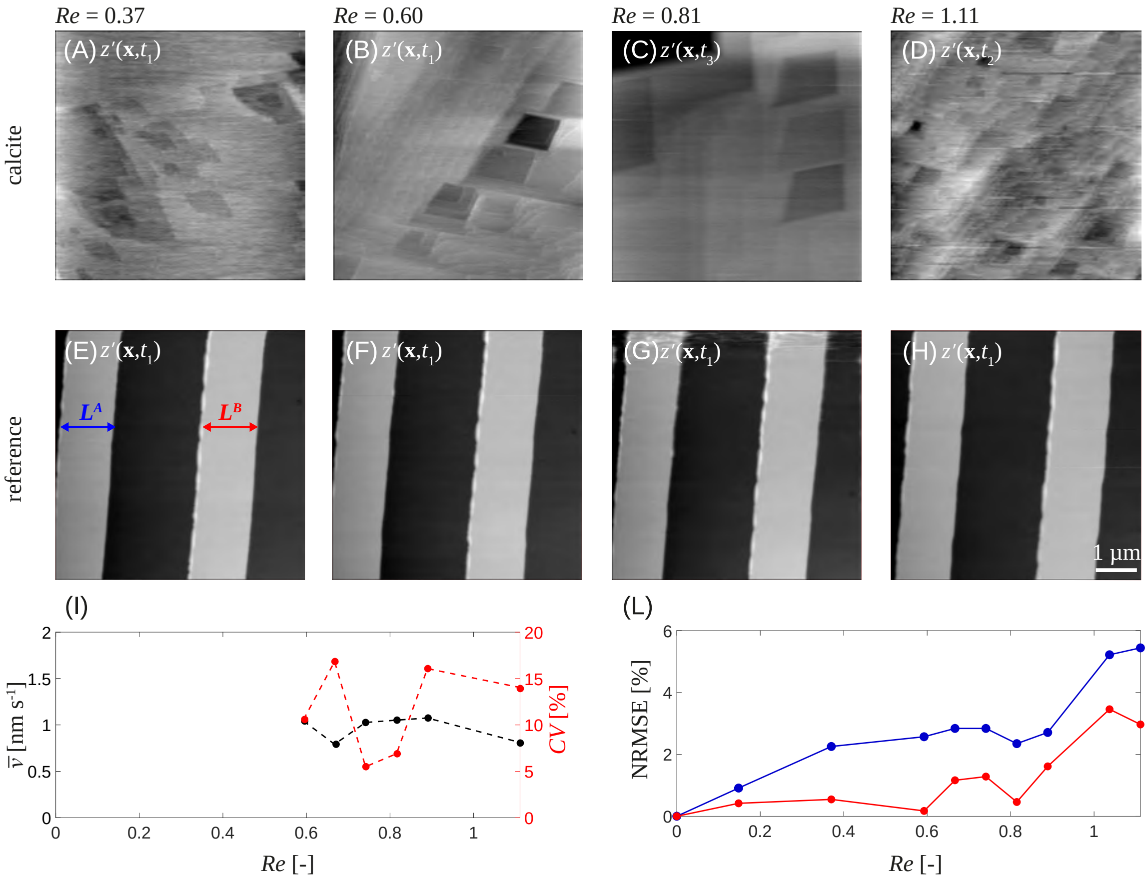

Figure 2.6: Exemplary topography images acquired for four different values of Reynolds number, Re. Top panels depict AFM topography signals of calcite surface imaged for **(A)** $Re = 0.37$, **(B)** $Re = 0.60$, **(C)** $Re = 0.81$, and **(D)** $Re = 1.11$. Panels **(E-H)** illustrate the topography of the reference Si grating sample imaged under flow conditions identical to those associated with **(A-D)**. Images of the calcite surface and of the sample grating are acquired at times t_i (with $i = 1, ..., 6$) from the beginning of the experimental observation window. Average spreading rate associated with various topography series, $\bar{\nu}$, and ensuing coefficient of variation, CV_ν, are depicted as a function of Re in panel **(I)**. Panel **(L)** shows the normalized root mean square error, NRMSE, associated with variations in the horizontal length between two fixed points of the reference sample (with respect to its reference value measured under static condition) versus Re.

temporal series of topographic images acquired at $Re = 0.37, 0.59, 0.74, 0.81, 0.89$, and 1.11 are included in Appendix C.

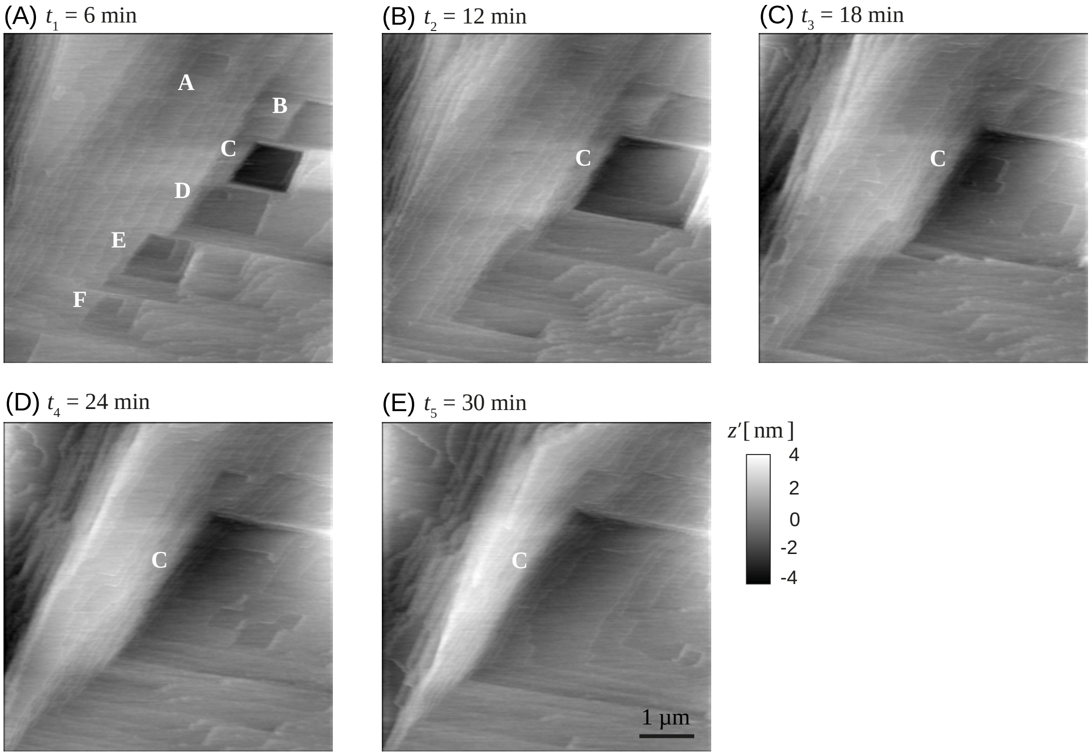

Figure 2.7: Exemplary topography of calcite surface subject to continuous fluid flow with $Re = 0.66$ acquired through *Setting 3* at observation times **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment.

2.5 | Inert mask fabrication

After clevage, the sample is subject to a cleaning treatment in acetone and isopropanol. Then, the metallic mask is fabricated through a combination of optical lithography and electron beam evaporation. The geometrical pattern of the mask is depicted in Fig. 2.8.E-F. This encompasses circles of diameter $\sim 80 - 300$ μm heterogeneously distributed on the surface. As such, this pattern enables one to investigate various regions of the crystal surface. A schematic depiction of the fabrication workflow is offered in Figs. 2.8.A-E. The procedure comprises a first stage where the adhesion promoter Ti Prime is spin-coated at 3000 rpm for 5 s and heated at 110°C for 1 min to evaporate the solvents. The positive resist AZ 5214E is then spin-coated in two different steps (at 750 rpm for 5 s and at 4000 rpm for

40 s) and finally annealed at 120°C for 2 min. These initial steps ensure that the entire surface is uniformly covered by the primer and the positive resist (Fig. 2.8.A). The sample is then subject to optical lithography and to a corresponding developing step in tetramethylammonium hydroxide solution for 50 s. These steps of the workflow enable one to transfer the desired geometrical pattern onto calcite surface. During the optical lithography stage, a glass mask comprising opaque and transparent regions is interposed between the sample and a UV light source (Fig. 2.8.B). Hence, only portions of the surface that correspond to the transparent glass mask are exposed to UV. The latter portions are removed from the surface by the subsequent developing phase (Fig. 2.8.C). The sample is then loaded in the electron beam evaporation chamber for Titanium deposition (Barri et al., 2020) (Fig. 2.8.D). The metal is uniformly deposited on the entire surface. It is finally subject to a lift-off phase in acetone (Fig. 2.8.E). This last step removes from the surface the photoresist and Ti layer attached to it and provides the desired geometry. Fig. 2.8.F depicts an exemplary image taken at the optical microscope of the sample surface after mask fabrication. Natural steps of calcite crystal are visible under the metal layer. The final Ti thickness is about 40 ± 1 nm (value also confirmed by AFM and VSI analysis, see Fig. 2.8.G).

Note that some precipitates form on calcite surface during the optical lithography phase of the mask fabrication workflow. Hence, the sample is subject to continuous flow in MilliQ water for a time ≥ 6 h before image acquisition. The first layers of the crystal are dissolved during this preliminary step together with precipitates attached to the calcite surface (and not covered by the inert layer) dissolve. Otherwise, precipitates that have been covered by the Ti layer during the metal deposition phase are preserved from dissolution. Hence, these points can serve as *fiducial points* (see Fig. 4.16) to correct any horizontal misalignment of images acquired at subsequent times (see Section 2.6).

2.6 | Alignment of topographic maps

Reliable estimates of point-by-point absolute dissolution rates can be obtained only upon exact horizontal alignment of topographic data acquired at subsequent times. These are

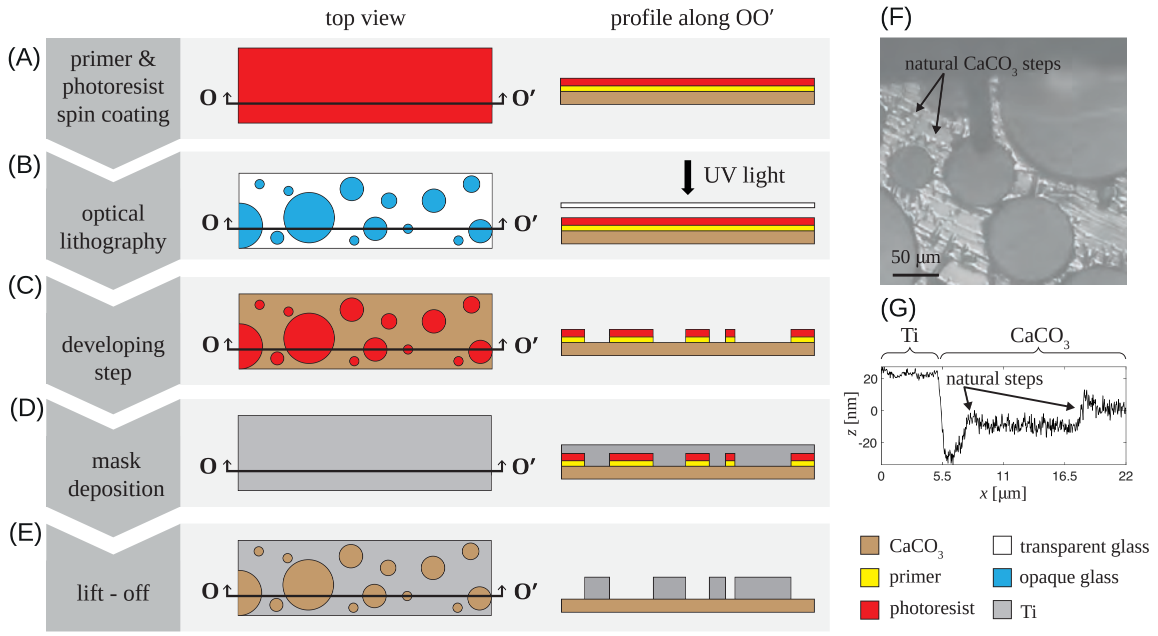

Figure 2.8: Top and profile (along cross section $O - O'$) sketch views of the steps included in the inert mask fabrication worlflow. These entail **(A)** primer and photoresist spin-coating; **(B)** optical lithography and corresponding **(C)** developing step; **(D)** Ti layer deposition; and **(E)** lift-off phase. **(F)** Optical microscope image depicting the geometrical pattern of the Ti mask on an exemplary calcite sample. **(G)** Vertical profile of a VSI-based image comprising the Ti mask and the pristine calcite surface acquired in air prior to exposure of the sample to the solution.

often affected by lateral drift due to thermal expansions of the piezoelectric scanner (Rahe et al., 2010, see, e.g.,). In the AFM system adopted in this PhD thesis, scanning in the x-y directions is associated with displacement of the tip, while the sample under investigation is fixed. This element, together with a warm-up phase of the instrument of about 1.5 h (Marinello et al., 2010), is documented to reduce the horizontal drift. The possible residual misalignment of topographic signals at various time can be corrected upon registering all images with respect to fixed points. Having at our disposal a portion of the surface that is covered by an inert layer and, in turn, does not react with the solution, has the additional advantage to provide fixed reference points that can be used as *fiducial points* to correct any misalignment in the horizontal plane. These coincide with the centroids of some precipitates that have formed on calcite surface during sample preparation and that have been covered by the Titanium layer during deposition (see Section 2.5). For freshly cleaved (non-masked) samples, fiducial points are identified as small precipitates spontaneously forming on the surface during the reaction. Note that the latter are visible only if AFM scanning is performed in tapping mode. Otherwise, the contact between the tip and the sample surface removes weakly attached particles forming on the surface when imaging is performed in contact mode (Guren et al., 2020; Renard et al., 2019).

3 Modeling Frameworks

Hydrogeological properties of natural porous systems are seen to be characterized by spatially heterogeneous behaviors. These hinder a description of such quantities though a deterministic approach. Therefore, characterization of hydrogeological attributes through geostatistical analyses has become a standard approach (Journel and Huijbregts, 1976; Kitanidis, 1997). In this context, the quantity of interest, $Y(\mathbf{x})$ ($\mathbf{x}$ denoting a spatial vector location), can be treated as a stationary random field and is defined as

$$Y(\mathbf{x}) = \langle Y \rangle + Y'(\mathbf{x}), \tag{3.1}$$

where $\langle Y \rangle$ is the (constant) ensemble average of Y and $Y'(\mathbf{x})$ is a zero-mean random fluctuation of $Y(\mathbf{x})$ about $\langle Y \rangle$ and drives the heterogeneity of $Y(\mathbf{x})$. In this work, we consider different modeling concepts to interpret the spatial heterogeneity of $Y(\mathbf{x})$ through characterization of $Y'(\mathbf{x})$. We start from the formulation of the GSG model by Riva et al. (2015a). Such formulation provides a joint description of the statistical behavior of Y and of it spatial increments, $\Delta Y(s)$, and has been seen to be well-suited to capture non-Gaussian behaviors of unimodal fields in a variety of contexts (see, e.g., Riva et al., 2015a, Guadagnini et al., 2018, Siena et al., 2019, 2020, and Li et al., 2022). Then, we focus of the characterization of fields exhibiting multimodal traits. We start from the formulation proposed by Lu and Zhang (2002) upon considering bimodal Gaussian mixtures. We extend such modeling approach by deriving rigorous formulations associated with the PDF of spatial increments to embed the observed scaling tendencies of such distributions within a unique analytical theoretical framework. Lastly, we generalize the formulation of the GMIX model to mixtures of

Generalized sub-Gaussian fields (herafter termed GSG-MIX). The latter modeling approach includes the GSG and the GMIX modeling frameworks as particular cases.

Chapter 3 is structured as follows. Mathematical notations and general formulations that are common to all modeling approaches considered are introduced in Section 3.1. Section 3.2 briefly illustrates the key analytical expressions of the GSG theoretical framework, its complete formulation being available in Riva et al. (2015a), Siena et al. (2020), and Neuman et al. (2024). Section 3.3 illustrates the theoretical formulation of the GMIX model and our original developments associated with the probability distributions of incremental values of a bimodal Gaussian random field. Section 3.4 extends the GMIX framework to describe a bimodal mixture composed by GSG modes. Parameter estimation schemes for (*i*) GSG and (*ii*) bimodal mixture models are respectively described in Sections 3.5 and 3.6.1-3.6.3. The reliability of parameter estimation procedures designed for bimodal mixtures is tested against synthetically generated data as detailed in Sections 3.6.2-3.6.4.

3.1 | General analytical formulation

We introduce the following notation to define Y' evaluated at two spatial locations $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$Y_1' = Y'(\mathbf{x}_1), \tag{3.2a}$$

$$Y_2' = Y'(\mathbf{x}_2). \tag{3.2b}$$

The bivariate Cumulative Distribution Function (CDF) of Y_1' and Y_2' reads

$$F_{Y_1'Y_2'}(y_1', y_2') = \Pr(Y_1' \leq y_1', Y_2' \leq y_2') = \int_{-\infty}^{y_2'} \int_{-\infty}^{y_1'} f_{Y_1'Y_2'}(v, w) dv dw, \tag{3.3}$$

$f_{Y_1'Y_2'}$ being the joint probability density of Y_1' and Y_2'. The latter can be obtained by differentiation of Eq. 3.3, leading to

$$f_{Y_1'Y_2'}(y_1', y_2') = \frac{\partial^2}{\partial y_1' \partial y_2'} F_{Y_1'Y_2'}(y_1', y_2'). \tag{3.4}$$

Following Eq. 3.4, the marginal PDF of Y' reads

$$f_{Y'}(y') = \int_{-\infty}^{+\infty} f_{Y_1'Y_2'}(y_1', y_2' = y') dy_1'. \tag{3.5}$$

We define spatial increments at lag $s = |\mathbf{x}_1 - \mathbf{x}_2|$ as $\Delta Y(s) = Y_1' - Y_2'$. The probability distribution of ΔY is (Riva et al., 2015a)

$$F_{\Delta Y}(\Delta y) = \Pr(\Delta Y \leq \Delta y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{y_2' + \Delta y} f_{Y_1'Y_2'}(y_1', y_2') dy_1' dy_2'. \tag{3.6}$$

The probability density of Δy is obtained from Eq. 3.6 as

$$f_{\Delta Y}(\Delta y) = \frac{d}{d(\Delta y)} F_{\Delta Y}(\Delta y) = \int_{-\infty}^{+\infty} f_{Y_1'Y_2'}(y_2' + \Delta y, y_2') dy_2'. \tag{3.7}$$

3.2 | Generalized sub-Gaussian model

In the following, we illustrate the key analytical expressions of the GSG theoretical framework, the complete set of details about their derivation being available in Riva et al. (2015a), Siena et al. (2020), and Neuman et al. (2024). According to the GSG formulation, the variable of interest, Y', is expressed as

$$Y'(\mathbf{x}) = U(\mathbf{x})G(\mathbf{x}). \tag{3.8}$$

where $G(\mathbf{x})$ and $U(\mathbf{x})$ are a zero-mean (typically correlated) Gaussian random field and a non-negative subordinator independent of G. The latter consists of statistically independent identically distributed non-negative random variables. The Gaussian field is characterized by a scale parameter, σ_G, i.e., the standard deviation of G, and a correlation function, ρ_G, that typically decreases as the separation distance between two points increases. It is noted that the GSG framework embeds the Gaussian model as a particular case when U tends to a deterministic constant. Considering the definition of Y' given by Eq. 3.8, the joint PDF of Y_1 and Y_2 (i.e., Eq. 3.4) is

$$f_{Y'_1,Y'_2}(y'_1, y'_2) = \int_0^{+\infty}\int_0^{+\infty} f_{U_1}(u_1) f_{U_2}(u_2) f_{G_1,G_2}\left(\frac{y'_1}{u_1}, \frac{y'_2}{u_2}\right)\frac{du_1}{u_1}\frac{du_2}{u_2}, \tag{3.9}$$

where $f_{U_i}(u_i)$ is the distributional form of the subordinator and f_{G_1,G_2} is the bivariate Gaussian PDF, i.e.,

$$f_{G_1,G_2}\left(\frac{y'_1}{u_1}, \frac{y'_2}{u_2}\right) = \frac{1}{2\pi\sigma_G^2\sqrt{1-\rho_G^2}} e^{-\frac{\frac{{y'_1}^2}{{u_1}^2}+\frac{{y'_2}^2}{{u_2}^2}-2\rho_G\frac{y'_1}{u_1}\frac{y'_2}{u_2}}{2\sigma_G^2(1-\rho_G^2)}}. \tag{3.10}$$

The marginal probability density function of Y' is obtained by replacing Eq. 3.9 into Eq. 3.5 as

$$f_{Y'}(y') = \frac{1}{\sqrt{2\pi}\sigma_G}\int_0^{\infty} f_U(u) e^{-\frac{y'^2}{2\sigma_G^2 u^2}}\frac{du}{u}, \tag{3.11}$$

Whereas odd order statistical moments of Y' are identically equal to zero, even order $q-th$ moments, $\langle Y'^q\rangle$, are given by

$$\langle Y'^q\rangle = \langle G^q\rangle\langle U^q\rangle \quad \text{with } q = 2, 4, 6..., \tag{3.12}$$

$\langle G^q \rangle$ and $\langle U^q \rangle$ being the $q-th$ order moment of the underlying Gaussian field and of the subordinator, respectively. Hence, the variance, σ_Y^2, and fourth order moment, $\langle Y'^4 \rangle$, can be respectively obtained as

$$\sigma_Y^2 = \langle Y'^2 \rangle = \sigma_G^2 \langle U^2 \rangle, \tag{3.13}$$

$$\langle Y'^4 \rangle = 3\sigma_G^4 \langle U^4 \rangle. \tag{3.14}$$

The kurtosis of Y, κ_Y, provides a measure of the degree of sharpness of the distribution peak and is evaluated from Eqs. 3.13 and Eq. 3.14 as

$$\kappa_Y = \frac{\langle Y'^4 \rangle}{\langle Y'^2 \rangle^2} = 3\frac{\langle U^4 \rangle}{\langle U^2 \rangle^2}. \tag{3.15}$$

Eq. 3.15 dictates that (i) κ_Y depends only on the subordinator U (and not on G) and (ii) the GSG PDF always displays a leptokurtic character ($\kappa_Y \to 3$ when U tends to a deterministic constant).

The PDF of increments can be evaluated by substituting Eq. 3.9 in Eq. 3.7 as

$$f_{\Delta Y}(\Delta y) = \frac{1}{\sqrt{2\pi}\sigma_G} \int_0^\infty \int_0^\infty f_{U_1}(u_1) f_{U_2}(u_2) \frac{e^{-\frac{\Delta y^2}{2\sigma_G^2 r^2}}}{r} \frac{du_2}{u_2} \frac{du_1}{u_1}, \tag{3.16}$$

with $r = \sqrt{u_1^2 + u_2^2 - 2\rho_G u_1 u_2}$. Whereas odd order moments of ΔY are identically zero, even order moments of order q are given by

$$\langle \Delta Y^q \rangle = \sum_{k=0}^{q} (-1)^{q-k} \binom{q}{k} \langle U^k \rangle \langle U^{q-k} \rangle \langle G(\mathbf{x})^k G(\mathbf{x}+\mathbf{s})^{q-k} \rangle \quad \text{with } q = 2, 4, 6... \tag{3.17}$$

From Eq. 3.17, second and fourth order moments of the increments are respectively obtained as

$$\langle \Delta Y^2 \rangle = 2\sigma_G^2 \left[\langle U^2 \rangle - \langle U \rangle^2 \rho_G \right], \tag{3.18}$$

$$\langle \Delta Y^4 \rangle = 6\sigma_G^4 \left[\langle U^4 \rangle - 4\langle U^3 \rangle \langle U \rangle \rho_G + \langle U^2 \rangle^2 (1 + 2\rho_G^2) \right]. \tag{3.19}$$

Making use of Eqs. 3.18 and 3.19, the kurtosis of ΔY reads

$$\kappa_{\Delta Y} = \frac{\langle \Delta Y^4 \rangle}{\langle \Delta Y^2 \rangle^2} = \frac{3}{2} \frac{\langle U^4 \rangle - 4\langle U^3 \rangle \langle U \rangle \rho_G + \langle U^2 \rangle^2 (1 + 2\rho_G^2)}{\left[\langle U^2 \rangle - \langle U \rangle^2 \rho_G \right]^2}. \tag{3.20}$$

As dictated by Eq. 3.20, $\kappa_{\Delta Y}$ only depends on U and on the correlation coefficient ρ_G. The covariance of Y, C_Y, can be obtained from Eq. 3.18 as $C_Y = \sigma_Y^2 - \langle \Delta Y^2 \rangle / 2$. This yields

$$C_Y = \begin{cases} \sigma_Y^2 & \text{if } s = 0, \\ \langle U \rangle^2 \sigma_G^2 \rho_G & \text{if } s > 0. \end{cases} \tag{3.21}$$

Hence, the covariance of the GSG field Y is discontinuous at the origin, $s = 0$, exhibiting a so-called nugget effect. The variogram of Y is expressed as

$$\gamma_Y = \frac{\langle \Delta Y^2 \rangle}{2} = \sigma_G^2 \left[\langle U^2 \rangle - \langle U \rangle^2 \rho_G \right] = \sigma_G^2 \sigma_U^2 + \langle U \rangle^2 \gamma_G. \tag{3.22}$$

Here, σ_U^2 is the variance of the subordinator and $\gamma_G = \sigma_G^2 (1 - \rho_G)$ is the variogram of the underlying Gaussian field. As dictated by Eq. 3.22, γ_Y is characterized by a nugget effect, $\nu = \sigma_G^2 \sigma_U^2$. The integral scale of Y can be obtained from Eq. 3.21 as

$$I_Y = \frac{\langle U \rangle^2}{\sigma_U^2 + \langle U \rangle^2} I_G. \tag{3.23}$$

Therefore, $0 < I_Y < I_G$ regardless of the type of subordinator one considers. An increase in σ_U^2 results in a decrease in the integral scale of Y'. The GSG model admits a variety of subordinator choices. In this thesis work, we follow Siena et al. (2019), Siena et al. (2020), Riva et al. (2015a), Guadagnini et al. (2018), and Li et al. (2022) and consider U as lognormally distributed according to

$$U \sim \ln N(0, (2 - \alpha)^2), \tag{3.24}$$

$\alpha \in (0, 2)$ being the shape parameter of U. The set of equations obtained upon this choice of the subordinator are included in Appendix D.

3.3 | Gaussian Mixture model

We consider $Y(\mathbf{x})$ to be a spatial random field exhibiting multimodal behavior across a given domain of interest, described as (e.g., Rubin, 1995, Lu and Zhang, 2002, and Dai et al., 2020)

$$Y(\mathbf{x}) = \sum_{m=1}^{M} I_m(\mathbf{x}) Y_m(\mathbf{x}), \tag{3.25}$$

where M is the number of independent and mutually-exclusive modes (or components) of $Y(\mathbf{x})$, $Y_m(\mathbf{x})$ is the $m-th$ component evaluated at (vector) location $\mathbf{x}$, and $I_m(\mathbf{x})$ is an indicator random field independent of Y_m and defined as

$$I_m(\mathbf{x}) = \begin{cases} 1 & \text{if component } m \text{ occurs at } \mathbf{x} \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

Note that I_m follows a Bernoulli distribution with mean $p_m(\mathbf{x}) = \langle I_m(\mathbf{x}) \rangle$ (which corresponds to the relative proportion of I_m across the domain, under ergodic conditions, $\langle \cdot \rangle$ denoting ensemble expectation), and variance, $\langle (I_m(\mathbf{x}) - \langle I_m(\mathbf{x}) \rangle)^2 \rangle = p_m(\mathbf{x})\left[1 - p_m(\mathbf{x})\right]$. Note also that the following constraint is satisfied

$$\sum_{m=1}^{M} I_m(\mathbf{x}) = 1, \tag{3.27}$$

at any location $\mathbf{x}$ in the system.

Focusing on a bimodal field (i.e., $M = 2$), Eq. 3.25 reduces to

$$Y(\mathbf{x}) = I(\mathbf{x}) Y_A(\mathbf{x}) + (1 - I(\mathbf{x})) Y_B(\mathbf{x}), \tag{3.28}$$

where subscripts A and B denote the two modes associated with the random field $Y(\mathbf{x})$. Setting $\langle I(\mathbf{x}) \rangle = p$, the cumulative distribution function (CDF) and the probability density function (PDF) of $Y(\mathbf{x})$ are respectively defined as

$$F_Y(y) = \Pr\{Y \leq y\} = p F_{Y_A}(y) + (1-p) F_{Y_B}(y), \tag{3.29}$$

and

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial y} = p f_{Y_A}(y) + (1-p) f_{Y_B}(y). \tag{3.30}$$

Here, $F_{Y_m}(y)$ and $f_{Y_m}(y)$ (with $m = A, B$) are the CDF and PDF of component m of the mixture, respectively.

If each component m of Y is characterized by a Gaussian distribution with mean μ_m and variance σ_m^2, i.e., $Y_m \sim N(\mu_m, \sigma_m^2)$, the field $Y(\mathbf{x})$ is a bimodal Gaussian mixture (GMIX), and Eq. 3.30 reads

$$f_Y(y) = \frac{p}{\sqrt{2\pi}\sigma_A} e^{-\frac{(y-\mu_A)^2}{2\sigma_A^2}} + \frac{(1-p)}{\sqrt{2\pi}\sigma_B} e^{-\frac{(y-\mu_B)^2}{2\sigma_B^2}}. \tag{3.31}$$

Making use of Eq. 3.30, the raw moment of Y of order q, $\langle Y^q \rangle$, can be computed as

$$\langle\, Y^q \,\rangle = p\langle\, Y_A^q \,\rangle + (1-p)\langle\, Y_B^q \,\rangle. \tag{3.32}$$

Therefore, the mean of Y can be derived by setting $q = 1$ in Eq. 3.32, as

$$\mu_Y = p\mu_A + (1-p)\mu_B, \tag{3.33}$$

and central moments of order q of Y can be evaluated as

$$\langle Y'^q \rangle = \langle (Y - \mu_Y)^q \rangle = \sum_{j=0}^{q} \binom{q}{j} (-1)^j \mu_Y^j \langle Y^{q-j} \rangle. \tag{3.34}$$

In particular, variance, σ_Y^2, skewness, Sk_Y, and kurtosis, κ_Y, associated with a bimodal GMIX field are evaluated by setting in Eq. 3.34 $q = 2, 3, 4$, respectively, as

$$\sigma_Y^2 = \langle\, Y'^2 \,\rangle = p\sigma_A^2 + (1-p)\sigma_B^2 + p(1-p)(\mu_A - \mu_B)^2, \tag{3.35}$$

$$Sk_Y = \frac{\langle Y'^3 \rangle}{\sigma_Y^3} = \frac{p}{\sigma_Y^3}(1-p)(\mu_A - \mu_B)\Big[(1-2p)(\mu_A - \mu_B)^2 + 3(\sigma_A^2 - \sigma_B^2)\Big], \tag{3.36}$$

$$\begin{aligned} \kappa_Y =& \frac{\langle Y'^4 \rangle}{\sigma_Y^4} = \frac{1}{\sigma_Y^4}\Big\{ 3p(\sigma_A^4 - \sigma_B^4) + 3\sigma_B^4 \\ &+ p(1-p)(\mu_A - \mu_B)^2\Big[\big(1 - 3p(1-p)\big)(\mu_A - \mu_B)^2 + 6\big(\sigma_A^2 - p(\sigma_A^2 - \sigma_B^2)\big)\Big]\Big\}. \end{aligned} \tag{3.37}$$

Eqs. 3.36 and 3.37 clearly show that the PDF of a GMIX field can be (i) non-symmetric (i.e., $Sk_Y \neq 0$) even though each component Y_m of Y is symmetric and/or (ii) leptokurtic ($\kappa_Y > 3$, corresponding to a heavy tailed distribution) or platikurtic ($\kappa_Y < 3$), even as components Y_m are mesokurtic (i.e., characterized by $\kappa_Y = 3$). In the hydrogeological context, examples of bimodal features documented for quantities of interest observed across heterogeneous systems include porosity, conductivity, permeability, vadose zone hydraulic properties, and electrical resistivity (e.g., Zhang et al., 2005, Zhang, 2009, Riva et al., 2013, Guadagnini et al., 2013, Guadagnini et al., 2015, Russo, 2002, and Li et al., 2022). The extent of these deviations from a Gaussian behavior is controlled by the difference between the component means (i.e., $\mu_A - \mu_B$), the component variances (i.e., σ_A^2 and σ_B^2), and p. In order to illustrate the main traits of the field considered, Fig. 3.1 shows the impact of p on the PDF (and related statistical moments) of a GMIX field characterized by $\mu_A - \mu_B = 2$, $\sigma_A^2 = 0.15$, and $\sigma_B^2 = 0.05$. The PDF of Y (see Fig. 3.1.A) exhibits two peaks and a local minimum located within the interval $y \in [\mu_B, \mu_A]$. As dictated by Eq. 3.33, the mean of Y varies linearly with p (see Fig. 3.1.B, note that μ_Y increases with p in our example, since $\mu_A > \mu_B$). The variance of Y exhibits a parabolic behavior with p (Fig. 3.1.C), as prescribed by Eq. 3.35. It attains a maximum value at $p = p_{max} = (1 + \eta)/2$, where $\eta = (\sigma_A^2 - \sigma_B^2)/(\mu_A - \mu_B)^2$ (with $\eta = 0.05$ in our example). This also implies that σ_Y^2 monotonically increases with p only when $\eta > 1$. The skewness of Y (see Eq. 3.36 and Fig. 3.1.B) vanishes for the two trivial cases $p = 0$ (where $Y = Y_B$) and $p = 1$ (where $Y = Y_A$) and when $p = p_3^{Sk=0} = (1 + 3\eta)/2$. Note that the PDF of Y is right-skewed ($Sk_Y > 0$) for $p \in (0, p_3^{Sk=0})$ and left-skewed ($Sk_Y < 0$) for $p \in (p_3^{Sk=0}, 1)$. When $|\eta| \geq 1/3$, then $p_3^{Sk=0} \notin (0, 1)$ and the PDF is right- (for $\eta > 1/3$) or left- (for $\eta < 1/3$) skewed regardless of the component proportions. Fig. 3.1.C also depicts the trend of the excess kurtosis ($E\kappa_Y = \kappa_Y - 3$) versus p. It can be shown from Eq. 3.37 that, in addition to the two trivial cases $p^{E\kappa=0} = 0, 1$, $E\kappa_Y$ vanishes for $p_{3,4}^{E\kappa=0} = 1/2 + \eta \pm \sqrt{(1 + 6\eta^2)/12}$. Hence, the PDF is platikurtic for $p \in (p_3^{E\kappa=0}, p_4^{E\kappa=0})$ and leptokurtic outside this range. If $|\eta| \geq (3+\sqrt{6})/3$, then $p_{3,4}^{E\kappa=0} \notin (0, 1)$ and the PDF is leptokurtic regardless of the component proportions.

3.3.1 | SPATIAL INCREMENTS OF A BIMODAL GAUSSIAN MIXTURE

For a bimodal GMIX, the notation introduced in Eq. 3.2 yields

$$Y_1 = Y(\mathbf{x}_1) = I(\mathbf{x}_1)Y_A(\mathbf{x}_1) + (1 - I(\mathbf{x}_1))Y_B(\mathbf{x}_1), \tag{3.38a}$$

$$Y_2 = Y(\mathbf{x}_2) = I(\mathbf{x}_2)Y_A(\mathbf{x}_2) + (1 - I(\mathbf{x}_2))Y_B(\mathbf{x}_2). \tag{3.38b}$$

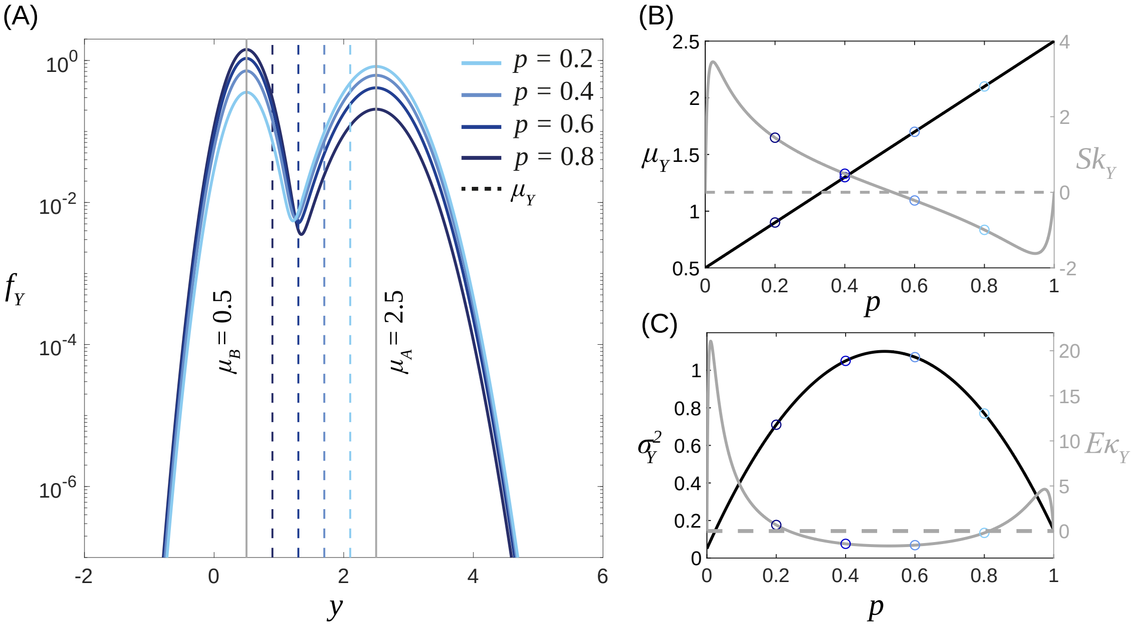

Figure 3.1: **(A)** Probability density functions (PDFs), $f_Y(y)$, of the GMIX model evaluated according to Eq. 3.31 for $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma_A^2 = 0.15$, $\sigma_B^2 = 0.05$, and four values of p. The associated **(B)** mean, μ_Y, and skewness, Sk_Y; **(C)** variance, σ_Y^2, and excess kurtosis, $E\kappa_Y$, are also depicted as a function of p. Empty circles in **(B)**-**(C)** correspond to statistical moments associated with the PDFs depicted in **(A)**.

We extend the approach illustrated by Rubin (1995) and obtain the joint PDF of Y_1 and Y_2 as

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) =& \Pr\{I(\mathbf{x}_1)=1, I(\mathbf{x}_2)=1\}\, f_{Y_{A,1},Y_{A,2}}(y_1,y_2) \\ &+\Pr\{I(\mathbf{x}_1)=0, I(\mathbf{x}_2)=0\}\, f_{Y_{B,1},Y_{B,2}}(y_1,y_2) \\ &+\Pr\{I(\mathbf{x}_1)=1, I(\mathbf{x}_2)=0\}\, f_{Y_{A,1},Y_{B,2}}(y_1,y_2) \\ &+\Pr\{I(\mathbf{x}_1)=0, I(\mathbf{x}_2)=1\}\, f_{Y_{B,1},Y_{A,2}}(y_1,y_2), \end{aligned} \tag{3.39}$$

where

$$f_{Y_{m,1},Y_{m,2}}(y_1,y_2) = \frac{e^{-r}}{2\pi\sigma_m^2\sqrt{1-\rho_m^2}}, \tag{3.40}$$

with

$$r = \frac{(y_1-\mu_m)^2 + (y_2-\mu_m)^2 - 2\rho_m(y_1-\mu_m)(y_2-\mu_m)}{2\sigma_m^2(1-\rho_m^2)} \text{ and } m = (A, B),$$

is the bivariate PDF of the Gaussian components of the mixture at the two locations. The joint PDF introduced in Eq. 3.40 is seen to depend on the spatial correlation $\rho_m = \rho_m(\mathbf{x}_1, \mathbf{x}_2)$ of each mode. We recall that, as mentioned above, the two components Y_A and Y_B are assumed to be uncorrelated. Hence, $f_{Y_{A,1},Y_{B,2}} = f_{Y_A}(y_1) f_{Y_B}(y_2)$ and $f_{Y_{B,1},Y_{A,2}} = f_{Y_B}(y_1) f_{Y_A}(y_2)$. Therefore, Eq. 3.39 leads to

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) =& \langle I(\mathbf{x}_1) I(\mathbf{x}_2)\rangle f_{Y_{A,1},Y_{A,2}}(y_1,y_2) \\ &+ \langle [1-I(\mathbf{x}_1)][1-I(\mathbf{x}_2)]\rangle f_{Y_{B,1},Y_{B,2}}(y_1,y_2) \\ &+ \langle I(\mathbf{x}_1)[1-I(\mathbf{x}_2)]\rangle f_{Y_A}(y_1) f_{Y_B}(y_2) \\ &+ \langle [1-I(\mathbf{x}_1)] I(\mathbf{x}_2)\rangle f_{Y_B}(y_1) f_{Y_A}(y_2). \end{aligned} \tag{3.41}$$

Considering that

$$\langle I(\mathbf{x}_1) I(\mathbf{x}_2)\rangle = \langle I(\mathbf{x})\rangle^2 + C_I(\mathbf{x}_1,\mathbf{x}_2) = p^2 + C_I(\mathbf{x}_1,\mathbf{x}_2), \tag{3.42}$$

where $C_I(\mathbf{x}_1,\mathbf{x}_2)$ is the covariance of the indicator field $I(\mathbf{x})$, Eq. 3.41 can be rewritten as

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) = & \left[p^2 + C_I(\mathbf{x}_1,\mathbf{x}_2)\right] f_{Y_{A,1},Y_{A,2}}(y_1,y_2) \\ & + \left[(1-p)^2 + C_I(\mathbf{x}_1,\mathbf{x}_2)\right] f_{Y_{B,1},Y_{B,2}}(y_1,y_2) \\ & + \left[p(1-p) - C_I(\mathbf{x}_1,\mathbf{x}_2)\right] \left\{ f_{Y_A}(y_1) f_{Y_B}(y_2) + f_{Y_B}(y_1) f_{Y_A}(y_2) \right\}. \end{aligned} \tag{3.43}$$

In the following we derive the analytical formulation for the PDF of the omnidirectional spatial increments $\Delta Y(s) = Y_1 - Y_2$ ($s =\mid \mathbf{x}_1 - \mathbf{x}_2 \mid$). Second-order stationarity is assumed for all random fields, i.e., $C_I(\mathbf{x}_1,\mathbf{x}_2) = C_I(s)$ and $\rho_m(\mathbf{x}_1,\mathbf{x}_2) = \rho_m(s)$. Replacing Eq. 3.43 in Eq. 3.4 leads to

$$\begin{aligned} f_{\Delta Y}(\Delta y) = & \frac{p^2 + C_I(s)}{\sqrt{4\pi\sigma_A^2\,(1-\rho_A)}} e^{-\frac{\Delta y^2}{4\sigma_A^2(1-\rho_A)}} + \frac{(1-p)^2 + C_I(s)}{\sqrt{4\pi\sigma_B^2\,(1-\rho_B)}} e^{-\frac{\Delta y^2}{4\sigma_B^2(1-\rho_B)}} \\ & + \frac{p(1-p) - C_I(s)}{\sqrt{2\pi(\sigma_A^2+\sigma_B^2)}} \left(e^{-\frac{(\Delta y - \mu_A + \mu_B)^2}{2(\sigma_A^2+\sigma_B^2)}} + e^{-\frac{(\Delta y + \mu_A - \mu_B)^2}{2(\sigma_A^2+\sigma_B^2)}} \right). \end{aligned} \tag{3.44}$$

The analytical expression of $f_{\Delta Y}(\Delta y)$ depends on (i) variances (σ_A^2 and σ_B^2) and correlation functions (ρ_A and ρ_B) associated with each of the mixture components, (ii) the difference between the component means ($\mu_A - \mu_B$) and (*iii*) mean (p) and covariance (C_I) of the indicator field. Fig. 3.2.A shows a graphical depiction of Eq. 3.44 for various lags, obtained upon relying on the exemplary set of parameters used for Fig. 3.1 and considering $p = 0.2$. For illustration purposes, we consider an isotropic exponential model to describe the above mentioned indicator covariance function, i.e., $C_I(s) = \sigma_I^2 \rho_I(s) = \sigma_I^2 e^{-s/\lambda_I}$, ($\lambda_I$ and $\sigma_I^2 = p(1-p)$ being the correlation length and variance of I, respectively) and for ρ_m, i.e., $\rho_m = e^{-s/\lambda_m}(m = A, B)$, λ_m being the correlation length of component Y_m. Here, for illustration purposes, we set $\lambda_A = \lambda_B = 6$ and $\lambda_I = 6.4$. It can be noted that the PDF of ΔY is always (*i*) symmetrical with respect to zero; and (*ii*) characterized by a dominant central peak (located at $\Delta y = 0$ and controlled by the first two terms in Eq. 3.44) and two lateral peaks (controlled by the last term in Eq. 3.44 and located at $\Delta y \approx \pm(\mu_A - \mu_B)$). Fig. 3.3.A also reveals that the relative importance of the lateral peaks increases (at the expense of the central peak) as lag increases. This behavior is driven by $C_I(s)$ and $-C_I(s)$ that are seen to multiply terms related to the central and lateral peaks in Eq. 3.44, respectively. As lag increases, $C_I(s)$ decreases and the difference between the height of the central and lateral peaks tends to be reduced. One can also see that the correlation functions ρ_A and ρ_B appear only within the first 2 terms of Eq. 3.44. Thus, their dependence on lag can only affect the central peak of the PDF of ΔY.

Statistical moments of the increments can then be readily evaluated from Eq. 3.44. Mean and all odd-order moments of ΔY are identically zero. The second moment of ΔY reads

$$\begin{aligned} \langle \Delta Y^2 \rangle = 2\big\{ & p^2\sigma_A^2(1-\rho_A) + (1-p)^2\sigma_B^2(1-\rho_B) \\ & + p(1-p)\left[(1-\rho_I)(\mu_A-\mu_B)^2 + \sigma_A^2(1-\rho_A\rho_I) + \sigma_B^2(1-\rho_B\rho_I)\right]\big\}. \end{aligned} \tag{3.45}$$

Since $C_Y = \sigma_Y^2 - \gamma_Y$, where $\gamma_Y = \langle \Delta Y^2 \rangle / 2$ is the variogram of Y, Eqs. 3.35 and 3.45 allow evaluating the covariance of Y as

$$C_Y = p^2\sigma_A^2\rho_A + (1-p)^2\sigma_B^2\rho_B + p(1-p)\rho_I\left[(\mu_A-\mu_B)^2 + \sigma_A^2\rho_A + \sigma_B^2\rho_B\right] \tag{3.46}$$

The integral scale of Y, I_Y, can be computed by integrating $\rho_Y = C_Y/\sigma_Y^2$. As also discussed by Lu and Zhang (2002), I_Y can be larger or smaller than the integral scale of the two modes and of the indicator, depending on the value of p, σ_A^2, σ_B^2 as well as on the correlation of the indicator field. Fig. 3.2.B depicts the variogram of Y as a function of lag for various values of p. Each curve attains the corresponding sill, σ_Y^2 (dashed horizontal lines), for large lags. Consistent with Fig. 3.1.B, $\sigma_Y^2(p = 0.6) > \sigma_Y^2(p = 0.4) > \sigma_Y^2(p = 0.8) > \sigma_Y^2(p = 0.2)$. It can be noted that the same ordering holds also for values attained by the variogram at any given lag.

The fourth-order moment of ΔY is

$$\begin{aligned}\langle \Delta Y^4 \rangle = 2\big\{&6p^2\sigma_A^4(1-\rho_A)^2 + 6(1-p)^2\sigma_B^4(1-\rho_B)^2 \\ &+ 6p(1-p)\rho_I\left[\sigma_A^4(1-\rho_A)^2 + \sigma_B^4(1-\rho_B)^2\right] \\ &+ p(1-p)(1-\rho_I)\left[(\mu_A-\mu_B)^4 + 3(\sigma_A^2+\sigma_B^2)(2(\mu_A-\mu_B)^2 + (\sigma_A^2+\sigma_B^2))\right]\big\}.\end{aligned} \tag{3.47}$$

The analytical expression for the kurtosis, $\kappa_{\Delta Y} = \langle \Delta Y^4 \rangle / \langle \Delta Y^2 \rangle^2$, associated with the increments of a Gaussian mixture can then be derived from Eqs. 3.45 and 3.47. We recall that this statistical moment quantifies the tailedness of the distribution and the sharpness of its peaks. Its dependence on lag is a distinctive element of the scaling behavior exhibited by the PDFs of increments of a GMIX field. Fig. 3.2.C depicts excess kurtosis, $E\kappa_{\Delta Y} = \kappa_{\Delta Y} - 3$, as a function of lag for various values of p. All curves exhibit a monotonic trend. The value of $E\kappa_{\Delta Y}$ is seen to increase (indicating tails that become heavier and peaks that become sharper) as s decreases. This pattern is starkly consistent with the behavior observed for several Earth and environmental variables (Riva et al., 2015a and references therein). These results clarify that the increments of a GMIX field exhibit clear non-Gaussian traits, despite each component of the mixture being Gaussian. As noted above, they also show that the PDFs of increments tend to change with lag due the action of the degree of spatial correlation of the two Gaussian components of the mixture and of the indicator field. Values of $E\kappa_Y$ are also depicted in Fig. 3.2.C (dashed horizontal lines) and are such that $E\kappa_Y(p = 0.2) > E\kappa_Y(p = 0.8) > E\kappa_Y(p = 0.4) > E\kappa_Y(p = 0.6)$. The same relative order is maintained also by the values of $E\kappa_{\Delta Y}$. Note that values of $E\kappa_{\Delta Y}$ are negative (i.e., indicating platikurtic distributions of increments) at large values of s for $p > 0.2$.

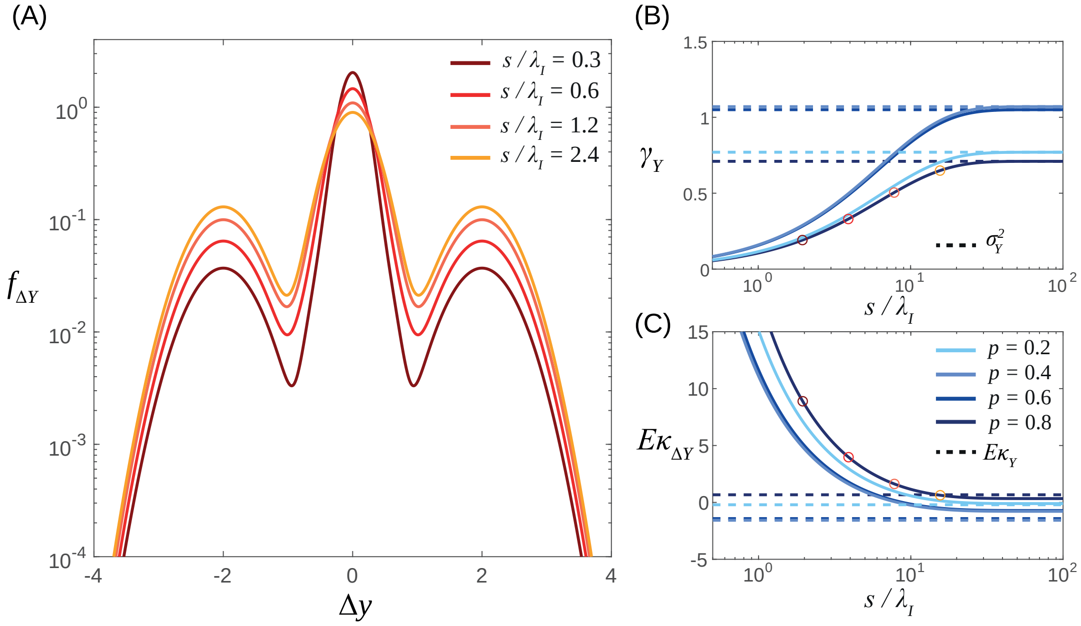

Figure 3.2: (A) Probability density functions (PDFs) of increments, $f_{\Delta Y}(\Delta y)$, evaluated according to the GMIX model (Eq. 3.44) for $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma_A^2 = 0.15$, $\sigma_B^2 = 0.05$, $p = 0.2$, $\lambda_A = \lambda_B = 6$ and $\lambda_I = 6.4$, at four values of the dimensionless lag, s/λ_I. The **(B)** variogram, γ_Y and **(C)** excess kurtosis, $E\kappa_{\Delta Y}$, are also depicted versus s/λ_I for four values of p. Empty circles in **(B-C)** correspond to the statistical moments associated with the PDFs depicted in **(A)**.

3.4 | Generalized sub-Gaussian Mixture model

We extend the formulation of the GMIX model by considering a bimodal mixture entailing Generalized sub-Gaussian components (GSG-MIX). We start from Eq. 3.30 and define each mode of the mixture, $Y_m(\mathbf{x})$, as

$$Y_m(\mathbf{x}) = \mu_m + Y'_m(\mathbf{x}) \tag{3.48}$$

where μ_m is the mean of each component of the mixture and $Y'_m(\mathbf{x})$ is the fluctuation about μ_m and is distributed according to Eq. 3.8, i.e.,

$$Y'_m(\mathbf{x}) \sim G_m(\mathbf{x}) U_m(\mathbf{x}). \tag{3.49}$$

Here, $G_m(\mathbf{x})$ is a zero-mean Gaussian field underlying each mode of the mixture (characterized by a scale parameter σ_{G_m} and a spatial correlation ρ_{G_m}); and $U_m(\mathbf{x})$ is a subordinator associated with each $m-th$ component of the mixture. Thus, Y_m is distributed according to

$$f_{Y_m}(y) = \frac{1}{\sqrt{2\pi}\sigma_{G_m}} \int_0^\infty f_{U_m}(u) e^{-\frac{(y-\mu_m)^2}{2\sigma_{G_m}^2 u^2}} \frac{du}{u}, \tag{3.50}$$

$f_{U_m}(u)$ corresponding to the distribution of the subordinator associated with each mode of the mixture. Replacing Eq. 3.50 into Eq. 3.30 yields

$$f_Y(y) = \frac{p}{\sqrt{2\pi}\sigma_{G_A}} \int_0^\infty f_{U_A}(u) e^{-\frac{(y-\mu_A)^2}{2\sigma_{G_A}^2 u^2}} \frac{du}{u} + \frac{(1-p)}{\sqrt{2\pi}\sigma_{G_B}} \int_0^\infty f_{U_B}(u) e^{-\frac{(y-\mu_B)^2}{2\sigma_{G_B}^2 u^2}} \frac{du}{u} \tag{3.51}$$

It can be noted that the GSG-MIX model rendered by Eq. 3.51 can be considered as a generalization of the GMIX model (Eq. 3.31) illustrated in Section 3.3, as each of its modes corresponds to a Gaussian distribution when U_m is deterministic and f_{U_m} tends to a Dirac delta function.

Similar to the GMIX framework, raw and central moments of order q of Y can be evaluated by making use of Eqs. 3.32 and 3.34, respectively. Setting $q = 1$ in Eq. 3.34, yields the mean of Y as defined by Eq. 3.33. Variance, σ_Y^2, skewness, Sk_Y, and kurtosis, κ_Y, of Y respectively read

$$\sigma_Y^2 = p\sigma_{Y_A}^2 + (1-p)\sigma_{Y_B}^2 + p(1-p)\left(\mu_A - \mu_B\right)^2, \tag{3.52}$$

$$Sk_Y = \frac{p}{\sigma_Y^3}(1-p)(\mu_A - \mu_B)\left[(1-2p)(\mu_A-\mu_B)^2 + 3(\sigma_{Y_A}^2 - \sigma_{Y_B}^2)\right], \tag{3.53}$$

$$\begin{aligned}\kappa_Y = \frac{1}{\sigma_Y^4}\Big\{ & p\left(\langle Y'^4{}_A\rangle - \langle Y'^4{}_B\rangle\right) + \langle Y'^4{}_B\rangle \\ & + p(1-p)(\mu_A-\mu_B)^2\Big[(1-3p(1-p))(\mu_A-\mu_B)^2 \\ & + 6\left(\sigma_{Y_A}^2 - p(\sigma_{Y_A}^2 - \sigma_{Y_B}^2)\right)\Big]\Big\},\end{aligned} \tag{3.54}$$

where $\sigma_{Y_m}^2$ and $\langle Y_m'^4\rangle$ are the variance and the fourth order central moment of the $m-th$ mode of the mixture and are defined through Eqs. 3.13 and 3.14. Eqs. 3.53 and 3.54 indicate that the PDF of the mixture (i) is non-symmetric ($Sk_Y \neq 0$) and (ii) can be leptikurtic ($\kappa_Y > 3$) or platikurtic ($\kappa_Y < 3$), depending on model parameters, even as each GSG mode is symmetric ($Sk_{Y_m} = 0$) and meso- or leptokurtic ($\kappa_{Y_m} \geq 3$, with $\kappa_{Y_m} \to 3$ when U_m is a deterministic constant, see Section 3.2). The extent of the deviation of a GSG-MIX distribution from its unimodal GSG counterpart depends on the difference between the component means, $\mu_A - \mu_B$, and variances, $\sigma^2_{Y_A} - \sigma^2_{Y_B}$.

Fig. 3.3 provides an appraisal of the impact of the model parameters on the main traits of $f_Y(y)$ and on the associated statistical moments. Here, for illustration purposes, we set $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma^2_{G_A} = 0.15$, and $\sigma^2_{G_B} = 0.05$. We consider $U_m(\mathbf{x})$ as lognormally distributed, i.e., $U_m \sim \log N(0, (2 - \alpha_m)^2)$, $\alpha_m \in (0, 2)$ being the shape parameter of the $m - th$ component. A GSG-MIX formulation embedding such a distributional form of the subordinator tends to a GMIX model if $\alpha_m \to 2$ ($m = A, B$), i.e., 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. The complete set of equations obtained upon this choice of U_m are reported in Appendix E. 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, e.g, porosity (Riva et al., 2015a; Guadagnini et al., 2018), electrical resistivity (Li et al., 2022), gas permeability (Siena et al., 2019), and mineral surface roughness (Siena et al., 2020). In the latter cases, documented values for α_m are in the range $1.3 - 1.9$. Fig. 3.3.A depicts the PDF of Y evaluated for different values of the proportion coefficient, p, upon setting $\alpha_A = 1.5$ and $\alpha_B = 1.7$. Analytical results corresponding to the GMIX model constituted by the G_m fields underlying the GSG-MIX model are also depicted as dashed curves. Similar to their GMIX counterparts, the GSG-MIX PDFs exhibit two peaks (centered at μ_A and μ_B, respectively) and a local minimum comprised therein. Figs. 3.3.B-C illustrate the dependence of odd and even order statistical moments of Y on the proportion coefficient, respectively. The mean of Y increases linearly with p. The variance displays a quadratic dependence on p and attains a maximum at $p = (1+\beta)/2$, with $\beta = (\sigma^2_{Y_A} - \sigma^2_{Y_B}) / (\mu_A - \mu_B)^2$. The skewness of Y vanishes for $p_1^{Sk=0} = 0$ (i.e., $Y = Y_B$), $p_2^{Sk=0} = 1$ (i.e., $Y = Y_A$), and $p_3^{Sk=0} = (1+3\beta)/2$. The PDF of Y is right-skewed for $p \in (0, p_3^{Sk=0})$ and left-skewed for $p \in (p_3^{Sk=0}, 1)$. If $|\beta| > 1/3$ the PDF is right- ($\beta > 1/3$) or left-skewed ($\beta < 1/3$) independent of p. The PDF is normokurtic (i.e., $E\kappa_Y = \kappa_Y - 3 = 0$) for $p_1^{E\kappa_Y=0} = 0$, $p_2^{E\kappa_Y=0} = 1$, and $p_{3,4}^{E\kappa_Y=0} = 1/2 + \beta \pm \sqrt{(1+6\beta^2)/12}$, leptokurtic for $p \in (0, p_3^{E\kappa_Y=0})$, $p \in (p_4^{E\kappa_Y=0}, 1)$, and platikurtic for $p \in (p_3^{E\kappa_Y=0}, p_4^{E\kappa_Y=0})$. It is otherwise leptokurtic regardless of p if $|\beta| > (3+\sqrt{6})/3$.

Figs. 3.3.D-F illustrate how deviations of each mode from a Gaussian behavior impact on variance, skewness, and excess of kurtosis of the mixture for fixed value of the proportion coefficient. The entity of deviations from Gaussianity are embedded in Eqs. 3.52, 3.53 and 3.54 through the statistical moment of the subordinator, $\langle U_m^q \rangle$ ($m = A, B$; $q = 2, 4$). For illustration purposes, we consider a lognormal subordinator and we set $p = 0.2$. Following this choice of the subordinator, if $\alpha_m \to 2$ (i.e., Y_m tends to a Gaussian), $\langle U_m^q \rangle \to 1$. The variance of Y attains a minimum for $\langle U_A^2 \rangle, \langle U_B^2 \rangle \to 1$. This is consistent with the observation that the GSG-MIX variance is always greater than its GMIX counterpart (depicted as a dashed black line in Fig. 3.3.C), regardless of the value of p. The skewness of the PDF of Y is minimized if $\langle U_A^2 \rangle \to 1$ and $\langle U_B^2 \rangle \to \left[(1-p)(\mu_A - \mu_B)^2 + \sigma_{G_A}^2(3-p)\right] / \left[(1-p)\sigma_{G_B}^2\right]$ and maximized if $\langle U_A^2 \rangle \to \left[p(\mu_A - \mu_B)^2 + \sigma_{G_B}^2(2+p)\right] / (p\sigma_{G_A}^2)$ and $\langle U_B^2 \rangle \to 1$ (yellow and green circles in Fig. 3.3.E). The kurtosis of Y depends on both $\langle U_m^2 \rangle$ and $\langle U_m^4 \rangle$. It attains a minimum when at least one of the components tends to a Gaussian (i.e., $\langle U_m^q \rangle \to 1$; yellow circle in Fig. 3.3.F). In particular, κ_Y is minimized along $\langle U_h^q \rangle \to 1$ ($q = 2, 4$), with $h = A$ if $p < 0.5$, $h = B$ otherwise.

3.4.1 | Spatial increments of a Generalized sub-Gaussian mixture

We start from the formulation of the PDF of spatial increments obtained in Section 3.3.1 and rendered by Eq. 3.43. Here, we replace $f_{Y_{m_1}, Y_{m_2}}(y_1, y_2)$ with the bivariate GSG PDF defined through Eq. 3.9. Similar to the GMIX case detailed in Section 3.3, we assume second-order stationarity of all fields (i.e., $C_I(\mathbf{x_1}, \mathbf{x_2}) = C_I(s)$; $\rho_G(\mathbf{x_1}, \mathbf{x_2}) = \rho_G(s)$). Under these

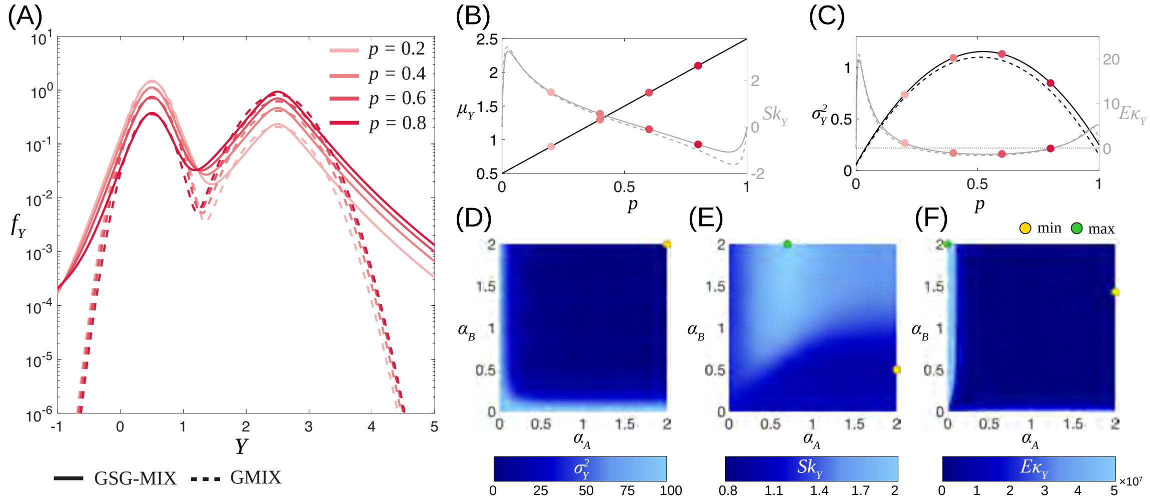

Figure 3.3: **(A)** Probability density functions (PDFs), $f_Y(y)$, of the GSG-MIX model evaluated considering a lognormal subordinator (Eq. E.1) upon setting $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma^2_{G_A} = 0.15$, $\sigma^2_{G_B} = 0.05$, $\alpha_A = 1.5$, $\alpha_B = 1.7$ and four values of p. Analytical expression of the GMIX PDFs constituted by the G_m fields underlying the GSG-MIX model are also reported as dashed lines. GSG-MIX statistical moments of **(B)** odd order, μ_Y (Eq. 3.33) and Sk_Y (Eq. E.3), and **(C)** even order, σ^2_Y (Eq. E.2) and $E\kappa_Y = \kappa_Y - 3$ (κ_Y defined by Eq. E.4), are plotted versus p. Circles depicted in **(B)** and **(C)** correspond to statistical moments of the PDFs represented in **(A)**. Contour plots of **(D)** variance, **(E)** skewness, and **(F)** excess of kurtosis as a function of α_A and α_B for fixed value of the proportion coefficient, $p = 0.2$. Minimum and maximum values attained by σ^2_Y, Sk_Y, and $E\kappa_Y$ are depicted as yellow and green circles, respectively.

assumptions, Eq. 3.7 yields the PDF of $\Delta Y(s)$ as

$$
\begin{aligned}
f_{\Delta Y}(\Delta Y) = {} & \frac{p^2 + C_I(s)}{\sqrt{2\pi}\sigma_{G_A}} \int_0^{+\infty}\int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{A_2}}(u_2)}{u_1 u_2 r} e^{-\frac{\Delta y^2}{2r^2\sigma^2_{G_A}}} du_2 du_1 \\
& + \frac{(1-p)^2 + C_I(s)}{\sqrt{2\pi}\sigma_{G_B}} \int_0^{+\infty}\int_0^{+\infty} \frac{f_{U_{B_1}}(u_1) f_{U_{B_2}}(u_2)}{u_1 u_2} e^{-\frac{\Delta y^2}{2r^2\sigma^2_{G_B}}} du_2 du_1 \\
& + \frac{p(1-p) - C_I(s)}{\sqrt{2\pi}} \int_0^{+\infty}\int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{B_2}}(u_2)}{u_1 u_2 v} e^{-\frac{(\Delta y - \mu_A + \mu_B)^2}{2v^2}} du_2 du_1 \\
& + \frac{p(1-p) - C_I(s)}{\sqrt{2\pi}} \int_0^{+\infty}\int_0^{+\infty} \frac{f_{U_{A_1}}(u_1) f_{U_{B_2}}(u_2)}{u_1 u_2 w} e^{-\frac{(\Delta y + \mu_A - \mu_B)^2}{2w^2}} du_2 du_1
\end{aligned}
\tag{3.55}
$$

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

The statistical moments of ΔY can be evaluated from Eq. 3.55. The second and fourth order moment of incremental values respectively read

$$\begin{aligned}\langle \Delta Y^2\rangle = 2\Big\{ & p^2\sigma^2_{G_A}\left(\langle U_A^2\rangle - \langle U_A\rangle^2\rho_{G_A}\right) + (1-p)^2\sigma^2_{G_B}\left(\langle U_B^2\rangle - \langle U_B\rangle^2\rho_{G_B}\right) \\ & + p(1-p)\Big[(1-\rho_I)(\mu_A-\mu_B)^2 + \sigma^2_{G_A}\left(\langle U_A^2\rangle - \langle U_A\rangle^2\rho_I\rho_{G_A}\right) \\ & + \sigma^2_{G_B}\left(\langle U_B^2\rangle - \langle U_B\rangle^2\rho_I\rho_{G_B}\right)\Big]\Big\},\end{aligned} \tag{3.56}$$

$$\begin{aligned}\langle \Delta Y^4\rangle = 6\Big\{ & p^2\sigma^4_{G_A}\left[\langle U_A^4\rangle - 4\langle U_A\rangle\langle U_A^3\rangle\rho_{G_A} + \langle U_A^2\rangle^2(1+2\rho^2_{G_A})\right] \\ & + (1-p)^2\sigma^4_{G_B}\left[\langle U_B^4\rangle - 4\langle U_B\rangle\langle U_B^3\rangle\rho_{G_B} + \langle U_B^2\rangle^2(1+2\rho^2_{G_B})\right] \\ & + p(1-p)\rho_I\Big[\sigma^4_{G_A}\left(\langle U_A^4\rangle - 4\langle U_A\rangle\langle U_A^3\rangle\rho_{G_A} + \langle U_A^2\rangle^2(1+2\rho^2_{G_A})\right) \\ & + \sigma^4_{G_B}\left(\langle U_B^4\rangle - 4\langle U_B\rangle\langle U_B^3\rangle\rho_{G_B} + \langle U_B^2\rangle^2(1+2\rho^2_{G_B})\right)\Big]\Big\} \\ & + 2\Big\{p(1-p)(1-\rho_I)\Big[(\mu_A-\mu_B)^4 + 6(\mu_A-\mu_B)^2\left(\sigma^2_{G_A}\langle U_A^2\rangle + \sigma^2_{G_B}\langle U_B^2\rangle\right) \\ & + 3\left(\sigma^4_{G_A}\langle U_A^4\rangle + \sigma^4_{G_B}\langle U_B^4\rangle + 2\sigma^2_{G_A}\sigma^2_{G_B}\langle U_A^2\rangle\langle U_B^2\rangle\right)\Big]\Big\}\end{aligned} \tag{3.57}$$

where $\rho_I = C_I/\sigma_I^2$ ($\sigma_I^2 = p(1-p)$ being the variance of I) is the (spatial) correlation of $I(\mathbf{x})$. The analytical expression of the kurtosis of incremental values, $\kappa_{\Delta Y} = \langle \Delta Y^4\rangle / \langle \Delta Y^2\rangle^2$, can be readily obtained from Eq. 3.56 and Eq. 3.57.

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 Eq. 3.56 as

$$C_Y = \begin{cases} \sigma_Y^2 & \text{if } s = 0, \\ p^2 C_{Y_A} + (1-p)^2 C_{Y_B} + C_I\left[(\mu_A-\mu_B)^2 + C_{Y_A} + C_{Y_B}\right] & \text{if } s > 0, \end{cases} \tag{3.58}$$

Here, C_{Y_m} is the covariance of the $m-th$ component of the mixture given by Eq. 3.21. The covariance C_Y is discontinuous at the origin, i.e., it exhibits a nugget effect

$$\nu = p\sigma^2_{G_A}\sigma^2_{U_A} + (1-p)\sigma^2_{G_B}\sigma^2_{U_B}, \tag{3.59}$$

at $s = 0$. Here, $\sigma^2_{U_m} = \langle U_m^2\rangle - \langle U_m\rangle^2$ ($m = A, B$) is the variance of $U_m(\mathbf{x})$. Hence, the variogram of Y, γ_Y, can be evaluated as

$$\begin{aligned}\gamma_Y = \nu & + p^2\langle U_A\rangle^2\gamma_{G_A} + (1-p)^2\langle U_B\rangle^2\gamma_{G_B} + (\mu_A-\mu_B)^2\gamma_I \\ & + p(1-p)\left[\langle U_A\rangle^2(1-\rho_I\rho_{G_A}) + \langle U_B\rangle^2(1-\rho_I\rho_{G_B})\right],\end{aligned} \tag{3.60}$$

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

$$I_Y = \frac{1}{\sigma^2_Y}\Big\{p^2\sigma^2_{G_A}\langle U_A\rangle^2\lambda_{G_A} + (1-p)^2\sigma^2_{G_B}\langle U_B\rangle^2\lambda_{G_B} \\ + p(1-p)\Big[(\mu_A - \mu_B)^2\lambda_I + \sigma^2_{G_A}\langle U_A\rangle^2\lambda^{eq}_{I-A} + \sigma^2_{G_B}\langle U_B\rangle^2\lambda^{eq}_{I-B}\Big]\Big\}. \tag{3.61}$$

Here, λ_{G_m} $(m = A, B)$ and λ_I are the correlation lengths of G_m and of I, respectively, and $\lambda^{eq}_{I-m} = \int_0^\infty \rho_I \rho_{G_m} ds$ 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.

Fig. 3.4.A illustrates the effect of model parameters on the main traits of the PDF of incremental values and on the variogram of Y. As in Section 3.4, we assume a lognormal distribution for the subordinator (see Appendix E) and set $p = 0.2$, $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma^2_{G_A} = 0.15$, and $\sigma^2_{G_B} = 0.05$. For the purpose of our illustration, we consider an exponential model for ρ_{G_m} and ρ_I (i.e., $\rho_{G_m} = e^{-s/\lambda_{G_m}}$ and $\rho_I = e^{-s/\lambda_I}$). The PDFs of ΔY are evaluated through Eq. E.5 at various lags upon setting $\alpha_A = 1.5$, $\alpha_B = 1.7$, $\lambda_{G_A} = 10$, $\lambda_{G_B} = 20$, and $\lambda_I = 7$. Corresponding GMIX PDFs (Eq. 3.44) are also depicted (dashed curves) for completeness. Similar to the GMIX case, the GSG-MIX PDFs are (*i*) symmetric and (*ii*) characterized by a central dominant peak (located at $\Delta y = 0$) and two secondary peaks at $\Delta y \approx \pm(\mu_A - \mu_B)$. The height of the central peak is controlled by the first two terms in Eq. 3.55. The third and fourth term in Eq. 3.55 otherwise govern the height of the two secondary peaks. These terms are in turn weighted on 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.

Figs. 3.4.B-E document the effect of model parameters on γ_Y and I_Y. Fig. 3.4.B depicts analytical variograms evaluated through Eq. E.9 for different combinations of α_A and α_B

(here, we consider $\alpha_m = 1.0, 1.5, 1.9$ with $m = A, B$). These values correspond to highly-, mildly- and quasi-Gaussian modes, respectively. Consistent with Eq. 3.52 and Eq. 3.59, variograms associated with lower values of α_m attain greater sills and are characterized by a higher nugget, ν. The latter increases linearly with p (see Fig. 3.4.C) and vanishes if the two components are Gaussian (i.e., if $\langle U_A^2 \rangle, \langle U_B^2 \rangle \to 1$; yellow point in Fig. 3.4.D). The largest value for the nugget is observed when the two components deviate the most from the Gaussian case (green point in Fig. 3.4.D). The behavior of I_Y with p is depicted in Fig. 3.4.C for fixed values of the shape parameter of the subordinator associated with each mode. Here, we set $\alpha_A = 1.5$ and $\alpha_B = 1.7$ for illustration purposes. Analytical results associated with the GMIX scenario are also depicted. We note that resting on an exponential model for the correlation structures of G_m and I yields a value of λ_{I-m}^{eq} (Eq. 3.61) that is proportional to the harmonic mean of λ_{G_m} and λ_I (i.e., $\lambda_{I-m}^{eq} = (\lambda_{G_m}\lambda_I)/(\lambda_{G_m} + \lambda_I)$). On the basis of Eq. 3.61, one can note that I_Y is always lower for the GSG-MIX than for the GMIX model, regardless of the value of p. Therefore, I_Y is maximized if both components tend to a Gaussian distributions (green point in Fig. 3.4.E). The value of I_Y is equal to λ_{Y_B} and λ_{Y_A} for $p = 0$ and $p = 1$, respectively, λ_{Y_m} being the correlation scale of each mode (Eq. 3.23). Within the range $p \in (0, 1)$, I_Y is either comprised between λ_{Y_A} and λ_{Y_B} or is lower/greater than these. For fixed values of μ_m and $\sigma_{Y_m}^2$, the correlation length of the mixture attains a maximum (if $\lambda_{G_A} < \lambda_I$ and $\lambda_{G_B} < \lambda_I$) or a minimum (if $\lambda_I < \lambda_{Y_A}$ and $\lambda_I < \lambda_{G_B}$). Otherwise, if $\lambda_I \leq \lambda_{G_A}$ and $\lambda_I \geq \lambda_{G_B}$ or $\lambda_I \geq \lambda_{G_A}$ and $\lambda_I \leq \lambda_{G_B}$, I_Y exhibits maximum and minimum values at the two boundaries (i.e., for $p = 0$ and $p = 1$).

3.5 | GSG parameter estimation

As the GSG model captures the behaviors of both the random function $Y'(\mathbf{x})$ and its increments $\Delta Y(s)$, it allows estimating mode parameters either from sample statistics of $Y'(\mathbf{x})$ alone or sample statistics of both $Y'(\mathbf{x})$ and $\Delta Y(s)$ at multiple lags. In the following, we briefly illustrate how to estimate GSG parameters on the basis of $Y'(\mathbf{x})$ (method MOM_A), or $Y'(\mathbf{x})$ and $\Delta Y(s)$ (method MOM_B) relying upon the method of moments. A detailed

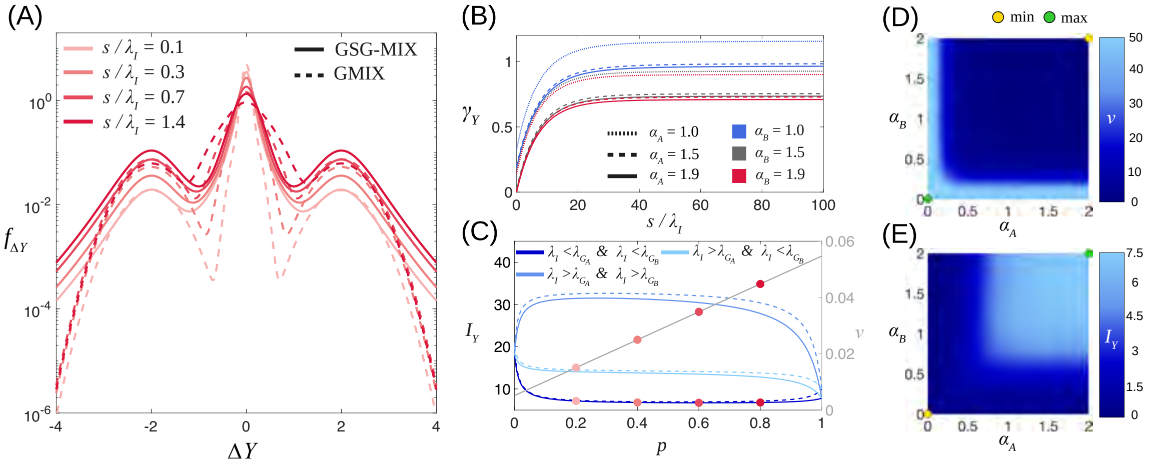

Figure 3.4: **(A)** Probability density functions (PDFs), $f_{\Delta Y}(\Delta y)$, of the GSG-MIX model evaluated considering a lognormal subordinator according to Eq. E.5 setting $p = 0.2$, $\mu_A = 2.5$, $\mu_B = 0.5$, $\sigma^2_{G_A} = 0.15$, $\sigma^2_{G_B} = 0.05$, $\alpha_A = 1.5$, $\alpha_B = 1.7$, $\lambda_A = 10$, $\lambda_B = 20$, and $\lambda_I = 7$ for four different lags. Analytical expression evaluated according to the GMIX model (Eq. 3.44) are also reported as a dashed lines. **(B)** variogram of Y, γ_Y, evaluated for different combinations of α_A and α_B. **(C)** integral scale, I_Y (Eq. E.11), and nugget effect, ν (Eq. E.10), of the mixture against p. Contour plots of **(D)** ν and **(E)** I_Y as a function of α_A and α_B. Minimum and maximum values attained by ν and I_Y are depicted as yellow and green circles, respectively.

description of GSG parameter estimation schemes is provided in Riva et al. (2015a).

3.5.1 | PARAMETER ESTIMATION BASED ON Y' SAMPLE DATA (MOM_A)

To estimate GSG parameters based on sample data of a random function $Y'(\mathbf{x})$, $\langle Y'^2\rangle$ and $\langle Y'^4\rangle$ are replaced by corresponding sample moments, $M_2^{Y'}$ and $M_4^{Y'}$, inferred from the data. Here, we refer to the formulation of the GSG with a lognormal subordinator, as detailed in Appendix D. Considering a lognormal distributional form for the subordinator, MOM_A yields the following system of equation:

$$\begin{cases} \hat{\alpha} = 2 - \frac{1}{2}\sqrt{\ln \frac{M_4^{Y'}}{3\left(M_2^{Y'}\right)^2}}, \\ \hat{\sigma}_G^2 = M_2^{Y'} e^{-\frac{1}{2}\ln \frac{M_4^{Y'}}{3\left(M_2^{Y'}\right)^2}}. \end{cases} \tag{3.62}$$

The estimated shape and scale parameters, $\hat{\alpha}$ and $\hat{\sigma}_G$, enable one to fully determine the marginal PDF expression of $Y'(\mathbf{x})$ in Eq. D.2. On the other hand, $Y'(\mathbf{x})$ sample data alone are not sufficient to estimate ρ_G, without which the covariance structure of $Y(\mathbf{x})$ remains undefined.

3.5.2 | Parameter estimation based on Y' and ΔY sample data (MOM_B)

All GSG parameters embedded in Eq. D.2 and D.6 can be estimated based on sample data of the random function $Y'(\mathbf{x})$ and of its increments, $\Delta Y(s)$. Again, we refer here to the lognormal subordinator case. For any given lag s, one approximates $\langle Y'^2 \rangle$, $\langle \Delta Y^2 \rangle$ and $\langle \Delta Y^4 \rangle$ by their corresponding sample moments $M_2^{Y'}$, $M_2^{\Delta Y}$, and $M_4^{\Delta Y}$, derived from the available data. Substituting these into Eqs. D.4, D.7 and D.8 yields

$$\begin{cases} M_2^{Y'} = e^{2(2-\hat{\alpha})^2} \hat{\sigma}_G^2, \\ M_2^{\Delta Y} = 2\hat{\sigma}_G^2 e^{(2-\hat{\alpha})^2} \left[e^{(2-\hat{\alpha})^2} - \hat{\rho}_G \right], \\ M_4^{\Delta Y} = 6\hat{\sigma}_G^4 e^{4(2-\hat{\alpha})^2} \left[e^{4(2-\hat{\alpha})^2} + 1 - 4e^{4(2-\hat{\alpha})^2} \hat{\rho}_G + 2\hat{\rho}_G^2 \right]. \end{cases} \tag{3.63}$$

Note that values of the scale and shape parameters are expected to be (approximately) constant with s, while the estimated Gaussian field correlation, $\hat{\rho}_G$, typically decreases as s increases.

3.6 | Mixture parameter estimation

Estimation of parameters associated with mixture models can be performed through soft- or hard-clustering of data. The former approach consists in assigning each observation to a components of the mixture with a given probability rather than to a unique mode. Otherwise, hard clustering of the data is required to assess incremental values associated (i) with each $m-th$ region and (ii) with the indicator. These data are necessary to estimate the parameters characterizing the correlation structure of Y_m $(m = A, B)$ and of I embedded in the formulation of $f_{\Delta Y}(\Delta y)$. We propose two different schemes for the estimation of mixture parameters. We consider a standard Expectation-Maximization (EM) algorithm to estimate parameters included in the GMIX model formulation. We then consider an estimation workflow grounded on a Bayesian classification approach that is initialized starting from our prior knowledge of the mechanisms driving the dissolution reaction. This approach is designed to embed the estimation of GMIX and GSG-MIX model parameters. A transparent assessment of the parameter estimation strategies is then provided though application of the designed workflows to synthetically generated GMIX and GSG-MIX fields.

3.6.1 | Parameter estimation based on Expectaction Maximization

We infer the 5 parameters of Y (i.e., μ_A, μ_B, σ_A^2, σ_B^2, and p in Eq. 3.31) by relying on a well-established Maximum Likelihood (ML) approach, implemented through an iterative EM procedure (see, e.g., McLachlan and Krishnan, 2008; Gournelos et al., 2020). According to the latter, each iteration consists of (*i*) the Expectation step (*E-step*), aimed at evaluating the (posterior) probability that each observation belongs to the mixture components, on the basis of an initial GMIX parameter set (or the parameter set obtained at the previous iteration); and (*ii*) the Maximization step (*M-step*), which uses the information from the E-step to estimate the GMIX parameter set maximizing the likelihood function. The algorithm stops when the increase of the likelihood function between two subsequent iterations is smaller than a prescribed threshold. Note that EM suffers from the typical issues associated with ML approaches, i.e., uniqueness, identifiability, and stability (Carrera and Neuman, 1986). To address the issue of the sensitivity of results to parameter initialization, application of the EM algorithm is repeated n times, each with a new set of initial parameters. Estimates of model parameters are then considered to correspond to the parameter set providing the highest likelihood among the n runs (McLachlan and Krishnan, 2008). The number of runs, n, is case specific and must be set through a stability analysis of the algorithm output. For the scenario considered in this PhD thesis, the selection of $n = 40$ allows obtaining stable estimates of μ_m, σ_m^2, and p. Note that this procedure can be employed to obtain a fuzzy (or soft) clustering of the data: each observation is assigned to each mixture component with a given probability rather than to a unique component, as it would result from a hard-clustering approach. Hence, we refer here to a hard clustering approach by assigning each observation to the mixture component with which the largest posterior PDF is associated. Then, we compute spatial increments ΔY_m $(m = A, B)$ within the regions associated with each Gaussian component and the corresponding sample correlation function $\hat{\rho}_m(s)$. An estimate of λ_m can obtained by fitting $\hat{\rho}_m(s)$ with a suitable theoretical model. An estimate of λ_I can be obtained according to the following two approaches:

- *method 1* - evaluate spatial increments associated with the indicator random field, I; fit $\hat{\rho}_I(s)$ with a suitable theoretical model (e.g., an exponential model or other);
- *method 2* - evaluate the mean length of the indicator field, l_A, and estimate λ_I as $\hat{\lambda}_I = (1 - p)l_A$ (Lu and Zhang, 2002).

3.6.2 | Reliability of the EM algorithm

Multiple realizations of synthetic GMIX fields are generated to provide a transparent assessment of the reliability of the parameter estimation strategy described in Section 3.6.1. Details about the generation procedure are offered in Appendix F. Our analysis relies upon $N = 100$ synthetic fields. Each realization is treated as a dataset to which the parameter estimation procedures can be applied. An exemplary realization of the GMIX is depicted in Fig. 3.5.A. Comparison between estimated and input model parameters enables us to assess the reliability of the proposed inference methodology. Fig. 3.5.B depicts the binary categorical (i.e., indicator) field that is inferred from the EM and clustering procedure applied to the synthetic dataset in Fig. 3.5.A. Given the indicator field and the ensuing sample correlation function $\hat{\rho}_I(s)$, an estimate of λ_I can then be obtained according to *method 1* and/or *2* introduced in Section 3.6.1. Note that one of the categories needs to be characterized in terms of its mean length, l_m, when considering *method 2*. The white portion of the domain (category B) in Fig. 3.5.B corresponds to the so-called *background category*. In a categorical random field, this term is commonly adopted to identify the category that fills in the space within which other categories are distributed. As an example, in geostatistical applications associated with hydrogeological scenarios, categories are represented by the various lithofacies of a depositional environment and the background geomaterial is typically associated with the category characterized by the lowest deposition energy (Carle and Fogg, 1997). In our datasets, we evaluate the mean length l_A of the category that is not in the background (black regions in Fig. 3.5.B, associated with category A) by averaging over all of the connected sets of A (i) the length of the sides of the bounding box (see green lines in Fig. 3.5.B) and (ii) the diameter of the inscribed maximal balls (yellow circle in Fig. 3.5.B).

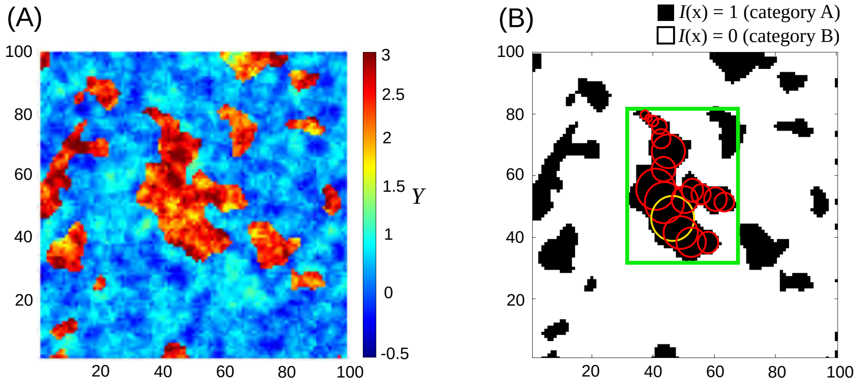

Figure 3.5: **(A)** Synthetic realization of a GMIX field with the set of parameters used in Fig. 3.1; **(B)** Indicator field associated with the GMIX realization depicted in **(A)**. Green lines and red circles represent the sides of the bounding box and the inscribed maximal balls for a connected cluster, respectively.

Fig. 3.6 collects values of GMIX parameters estimated for all the $N = 100$ synthetic realizations. The average of the estimates is always satisfactorily close to the corresponding input value. The mean squared relative deviation (MSRD) between input and estimated parameter values is evaluated over the whole collection of realizations. On the basis of this metric one can note that estimates are overall more accurate for (i) the parameters of the Gaussian distribution associated with category B (except for the mean value), that occupies a larger portion of the domain as compared to category A and (ii) the correlation scale $\hat{\lambda}_I$ obtained via *method 2* as compared against its counterpart based on *method 1*. In light of these results, *method 2* is considered for the estimation of λ_I in the context of the experimental datasets analyzed in Section 4.2.

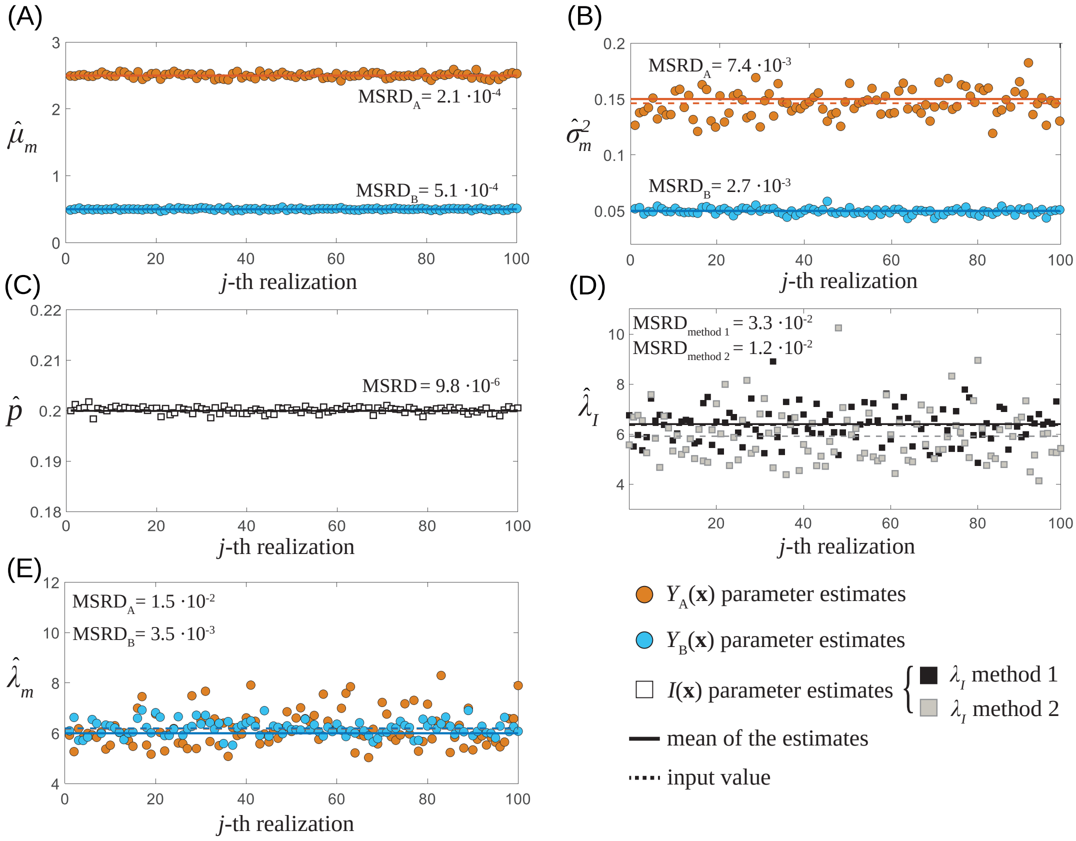

Figure 3.6: Results of the GMIX parameter estimation approach applied on the collection of synthetic datasets generated with the set of parameters used in Fig. 3.2. Mean of the estimates (dashed lines), input values used in the generation (solid lines) and mean squared relative deviation (MSRD) between input and estimated parameter values are also reported.

3.6.3 | PARAMETER ESTIMATION BASED ON BAYESIAN CLASSIFICATION

We design a custom algorithm grounded on a Bayesian classification approach to estimate parameters embedded in the GMIX and GSG-MIX modeling frameworks. As the GMIX is a specific case of the more general GSG-MIX model, the estimation procedure is here reported in details considering the latter. Prior to application to real data, the performance of the designed computational scheme is assessed on synthetic datasets. These are generated to resemble typical geometries observed for experimental dissolution rate data (see Appendix F). A schematic representation of the application of the parameter estimation workflow to a synthetic GSG-MIX field is offered in Figs. 3.8. We start by setting a random path visiting all $k = 1, ..., N_c$ cells, N_c being the total number of data (Fig. 3.8.B). At each k step, the algorithm assigns a given observation $y = Y(\mathbf{x}_k)$ to the category ω_m ($m = A, B$) for which a discriminant function, $g_m(y)$, is maximized. Relying on a Bayesian classifier with minimum-error-rate, the latter is considered as proportional to the posterior probability $P(\omega_m|y)$ (Duda et al., 2000; James et al., 2013), i.e.,

$$g_m(y) \propto p(y|\omega_m)P(\omega_m). \tag{3.64}$$

Here, $p(y|\omega_m)$ is the likelihood of class ω_m given y and $P(\omega_m)$ is the prior probability of ω_m. Considering each mode as distributed according to a GSG and making use of Eq. 3.50, Eq. 3.64 yields

$$g_m(y) \propto \frac{P(\omega_m)}{\sqrt{2\pi}\sigma_{G_m}} \int_0^\infty f_{U_m}(u) e^{-\frac{(y-\mu_m)^2}{2\sigma_{G_m}^2 u^2}} \frac{du}{u}, \tag{3.65}$$

where $P(\omega_A) = p$ and $P(\omega_B) = 1 - p$ (see Fig. 3.8.C). If the subordinator associated with each region tends to a deterministic constant (i.e., the GSG-MIX tends to a GMIX), Eq. 3.65 can be simplified to

$$g_m(r) = -\frac{1}{2}(r - \mu_m)^2 \sigma_m^{-2} - \frac{1}{2} \ln \sigma_m^2 + \log P(\omega_m). \tag{3.66}$$

The parameter set is then updated for the $(k + 1) - th$ step. For the general GSG-MIX case, the parameter estimation scheme depends on the choice of the distributional form of the subordinator. In the following, we describe the estimation procedure considering the set of GSG-MIX equations derived in Appendix E and corresponding to a lognormal subordinator U_m. At each k step, we rely upon method MOM_A (see Section 3.5.1) to estimate parameters associated with each $m - th$ component of the mixture, Y_m, and embedded in Eq. E.1 (i.e., $p, \mu_m, \sigma_{G_m}, \alpha_m$ with $m = A, B$). Setting the analytical first order raw moments of (i) I and of (ii) Y_m, and the (iii) second and (iv) fourth order central moments of Y_m equal to their sample counterparts yields the system of equations

$$\begin{cases} \hat{p}^{(k,n)} = \langle I^{(k,n)} \rangle, \\ \hat{\mu}_m^{(k,n)} = \langle Y_m^{(k,n)} \rangle, \\ \hat{\alpha}_m^{(k,n)} = 2 - \frac{1}{2}\sqrt{\ln \frac{M_{m,4}^{(k,n)}}{3\left(M_{m,2}^{(k,n)}\right)^2}}, \\ \hat{\sigma}_{G_m}^{(k,n),2} = M_{m,2}^{(k,n)} e^{-\frac{1}{2}\ln \frac{M_{m,4}^{(k,n)}}{3\left(M_{m,2}^{(k,n)}\right)^2}}, \end{cases} \tag{3.67}$$

where $n = 1, ..., N_r$ is the number of iteration of the algorithm. For the GMIX case, the system of equations 3.67 simplifies to

$$\begin{cases} \hat{p}^{(k,n)} = \langle I^{(k,n)} \rangle, \\ \hat{\mu}_m^{(k,n)} = \langle Y_m^{(k,n)} \rangle, \\ \hat{\sigma}_m^{(k,n),2} = M_{m,2}^{(k,n)}. \end{cases} \tag{3.68}$$

At each repetition, we start from the parameter set associated with the $n-1$ replica. We stop the iterations when convergence of the proportion coefficient, p, is attained (see Fig. 3.8.E). This choice for the exit flag of the algorithm is grounded on the results of the application of the scheme to synthetic cases, as detailed in Section 3.6.4. The Bayesian classifier also provides as output the posterior probability that y belongs to class ω_m. This is defined as

$$P(\omega_m|y) = \frac{p(y|\omega_m)P(\omega_m)}{p(y)}, \tag{3.69}$$

where $p(y)$ is evaluated via Eqs. E.1 and 3.31 for the GSG-MIX and GMIX case, respectively (Fig. 3.8.E). After the last repetition, we evaluate spatial increments associated with each of the two regions, i.e., $Y_A^{(N_r)}(\mathbf{x}) = I^{(N_r)}(\mathbf{x}) \cdot Y(\mathbf{x})$ and $Y_B^{(N_r)}(\mathbf{x}) = \left(1 - I^{(N_r)}(\mathbf{x})\right) \cdot Y(\mathbf{x})$. We then estimate the correlation of the underlying Gaussian field of each mode, $\hat{\rho}_{G_m}$ ($m = A, B$), by setting $M_2^{\Delta Y_m^{(N_r)}} = \langle \Delta Y_m^{(N_r),2} \rangle$. This yields

$$\hat{\rho}_{G_m} = \frac{1}{\langle U_m \rangle^2} \left[\langle U_m^2 \rangle - \frac{M_2^{\Delta Y_m^{(N_r)}}}{2\sigma_{G_m}^2} \right], \tag{3.70}$$

where $M_2^{\Delta Y_m^{(N_r)}}$ is the sample second order moment of incremental values of the $m-th$ component of the mixture. Note that Eq. 3.70 simplifies to $\hat{\rho}_m = 1 - M_2^{\Delta Y_m^{(N_r)}}/2\sigma_m^2$ for the GMIX particular case. Finally, we evaluate the increments of the indicator field associated with the last repetition, $I^{(N_r)}$, and estimate its correlation as $\hat{\rho}_I = 1 - \Delta M_2^{I^{(N_r)}} / \left[2\hat{p}(1-\hat{p})\right]$ (Fig. 3.8.F).

First guess values of the model parameter set are required to initialize the algorithm. These are obtained upon relying on a preliminary classification of the field. The more accurate the latter, the fewer the number of repetitions required to achieve convergence of the algorithm output. We perform the preliminary classification starting from our prior knowledge on the physical mechanisms driving the dissolution reaction. We associate regions characterized by high rates with spreading of younger dissolution stepwaves across the crystal surface. Otherwise, areas characterized by low rates correspond to spatial locations where the reaction is governed by the propagation and/or merging of older dissolution stepwaves. Identification of pit boundaries and/or steps in topography images $z(\mathbf{x}, t)$ and $z(\mathbf{x}, t+\Delta t)$ is therefore instrumental to perform a preliminary classification of R. Detecting such features relying on phase contrast images, $\phi(\mathbf{x}, t)$, is generally more straightforward than employing the ensuing topography maps. This is related to the nature of phase contrast signals, which represent the shift in the oscillation phase between the excitation force and the cantilever response during scanning in tapping mode. This effect is enhanced where the crystal surface experiences abrupt changes in the topographic profile (see, e.g., Garcıa and Perez, 2002 and references therein), i.e., where a younger dissolution pulse occurs. Otherwise, if imaging relies upon contact AFM, one could consider deflection or friction signals to perform such preliminary classification. Therefore, we extract profiles of steps associated with MP spreading at different times from phase signals and consider the area between these as the first estimate of class ω_A (high rates; see Fig. 3.7.A-B). Otherwise, we isolate portions of the surface where dissolution is controlled by low-rate mechanisms (i.e., merging of older stepwaves, shallow etch pits dynamics or vertical retreat) and identify (as first guess) ω_B as the area within which the dissolution pattern does not change between two consecutive observation times (see Fig. 3.7.A-B). This procedure enables one to obtain an initial indicator field (Fig. 3.7.C) and, in turn, initial regions A and B (Fig. 3.7.D).

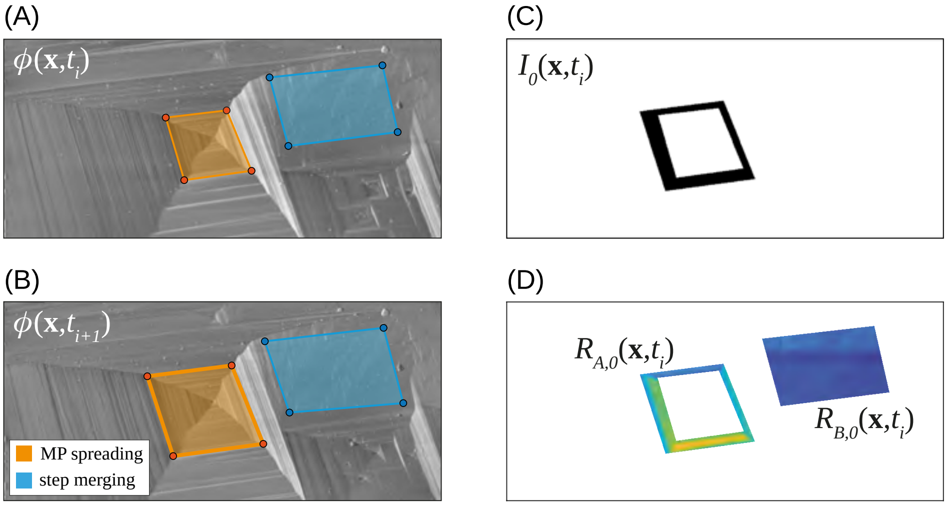

Figure 3.7: (A-B) Identification of pit margins and terraces/step merging regions in AFM phase shift signals, $\phi(\mathbf{x}, t)$, acquired at subsequent times by relying upon AFM tapping mode. Associated initial **(C)** indicator field, I_0, and **(D)** rates of region m, $R_{m,0}$ ($m = A, B$).

3.6.4 | RELIABILITY OF THE BAYESIAN CLASSIFICATION ALGORITHM

The methodology for parameter estimation is tested on synthetic fields generated on a 450 × 450 grid. Such a grid size is selected to be similar to the one of experimental data. The mixture is obtained by generating (i) two spatial random fields distributed according to a unimodal GSG and (ii) an indicator field resembling geometries that are typically observed for experimental dissolution rate fields. The latter is constructed to resemble the typical rhombohedral shape of stepwaves emanated from an etch pit (Fischer and Lüttge, 2018). Additional information on the generation scheme are offered in Appendix F. A schematic representation of the synthetic field obtained is depicted in Fig. 3.8.A. The fields used for the algorithm initialization are shown in Fig. 3.8.B. We mimic a classification error that could be performed when dealing with laboratory data. For region A, we artificially introduce a slight misclassification by selecting (i) a region having a lateral extent corresponding to

$\sim 60\%$ (black area in Fig. 3.8.A) of the real one (orange pattern in Fig. 3.8.A) and (*ii*) a small portion ($\sim 1\%$) of region B (blue pattern in Fig. 3.8.A) located next to the edges of the actual region A. On the other hand, 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 and low rates. Identification of region B is generally straightforward because regions that are only experiencing normal surface retreat, i.e., terrace regions, are associated with the lowest dissolution rate (see, e.g. Brand et al., 2017, Bibi et al., 2018, and Bollermann and Fischer, 2020) 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 only a few layers and/or older dissolution stepwaves are characterized by a high material flux. Therefore, we introduce the above-mentioned misclassification to resemble this additional difficulty in the preliminary assessment of region A. The results of the algorithm are compared with generation parameters in Fig. 3.8.D. Here, we plot the trend of $\hat{p}, \hat{\mu}_m, \hat{\alpha}_m$, and $\hat{\sigma}_{G_m}$ versus the number of repetitions, n, and represent the reference generation parameters as a dashed line. Errors between estimated and real values are evaluated considering the results obtained at the last repetition of the algorithm and are also included in Fig. 3.8.D. We notice that parameters associated with the highest error are μ_B, σ_{G_A}, and α_A. 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 to region B some cells that actually belong to region A induces an increases of $\hat{\mu}_B$, even as the error associated with the estimation of p is low (e.g., $err_p = 0.53\%$ in this case). Otherwise, the estimation error related to these two parameters can be attributed to the a low sample size of region A. Estimation of α_A and σ_{G_A} relies on higher order statistical moments of region A, i.e., $M_{A,2}$ and $M_{A,4}$ in Eq. 3.67. Due to the small lateral extent of region A, even a small error in $\hat{p}$ results in including outliers in region A. This, in turns, reduces the reliability of (sample) higher order moments associated with region A and results in higher error in parameter estimation. Still, the overall estimation error of these parameters 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 $\hat{p}$ (Fig. 3.8.D). 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.

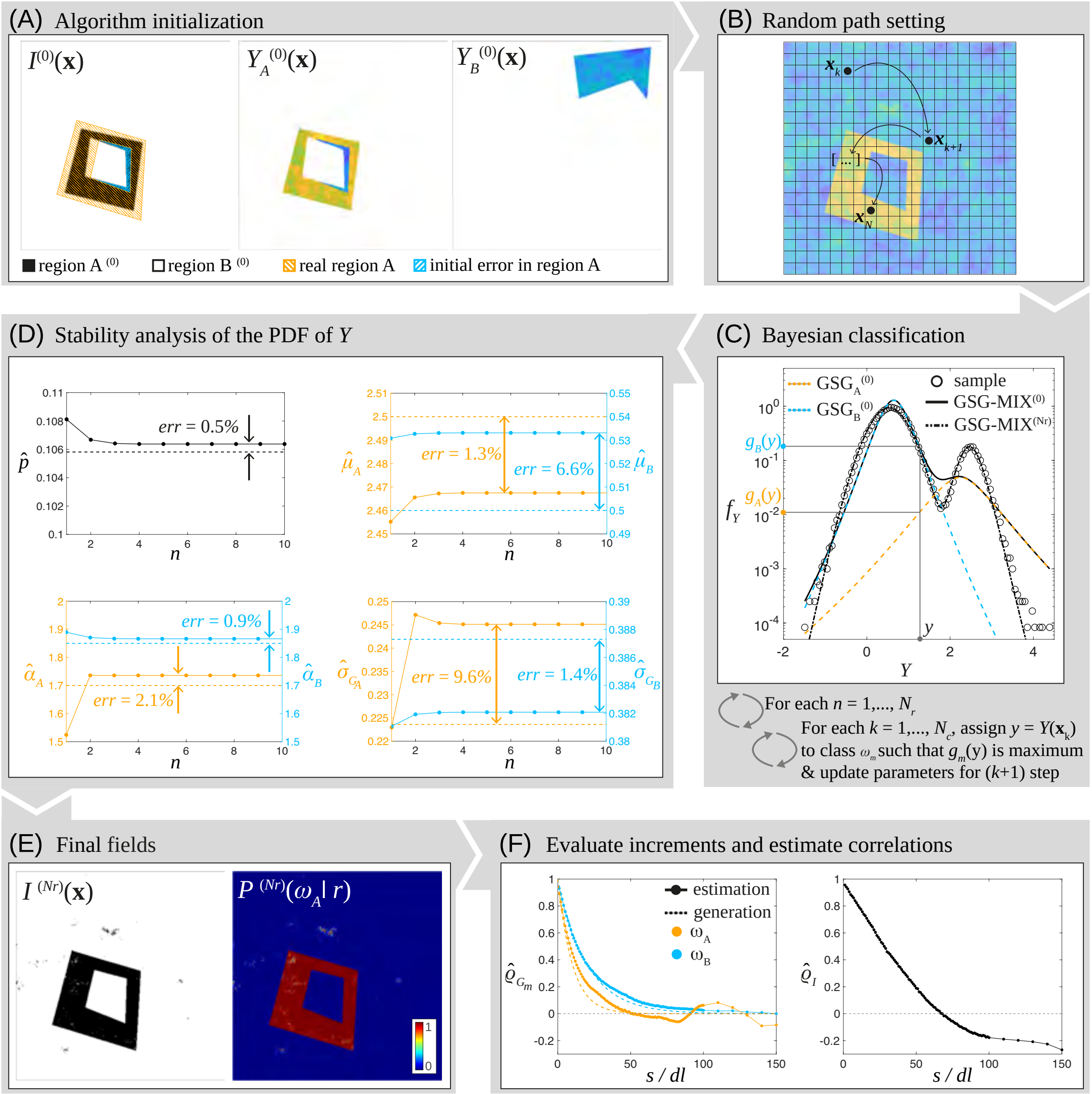

Figure 3.8: Schematic representation of the parameter estimation algorithm resting upon a Bayesian Classification approach. The estimation workflow includes **(A)** a preliminary classification of the field for algorithm initialization; **(B)** setting a random path visiting all N cells; **(C)** assigning each cell to a region and update of model parameters for the next step; **(D)** replication of the classification algorithm until parameter stability with the number of repetition, n; **(E)** evaluation of the posterior probability field associated with the last repetition, N_R; and **(F)** evaluation of correlation parameters of each $m-th$ region, $\hat{\rho}_{G_m}$, and of the indicator, $\hat{\rho}_I$.

4 Results

This Chapter illustrates the main experimental and modeling results of this PhD dissertation. The Chapter is structured as follows. Section 4.1 describes the application of the GSG modeling framework introduced in Section 3.2 to characterize the spatial heterogeneity of rate fields obtained from a calcite sample subject to dissolution under diffusion-controlled conditions (i.e., *Setting 1*, see Section 2.3.1). Section 4.2 illustrates the interpretation through the GMIX model of a collection of rate fields obtained from AFM images acquired through *Setting 2* (see Section 2.3.2), i.e., under intermittent flowing conditions. Section 4.3 includes the analysis of absolute rates evaluated from measurements of a calcite sample whose surface is partially covered by an inert layer and acquired under continuous flowing conditions (*Setting 3*, see Section 2.3.3). The heterogeneous behavior of associated dissolution rate fields is interpreted upon relying on the GMIX model. A GSG-MIX based interpretation of rate spectra associated with topographies of a freshly cleaved calcite sample acquired under continuous flowing conditions is illustrated in Section 4.4.

4.1 | Results of experimental *Setting 1*

In the following, we discuss results obtained upon relying on experimental *Setting 1* introduced in Section 2.3.1. A freshly cleaved (non-masked) calcite sample is subject to dissolution under static conditions leading to the development of a diffusive boundary layer in the close proximity of the sample surface. Imaging is performed in contact mode over a scanning window of $6 \times 6\ \mu\text{m}^2$ (comprising an area of 512×512 pixels; pixel size $dl = 11.7\,\text{nm}$). The acquisition frequency is here set to $f_a = 1.7\,\text{Hz}$, yielding a time interval $\Delta t = 5\,\text{min}$ between subsequent frames. The contact time between the fluid and the sample is $T_{c_1} = 30\,\text{min}$, similar to the dataset illustrated in Section 2.4.1 and Appendix A. Analysis of the spatial heterogeneity of associated dissolution rate fields is grounded on the Generalized sub-Gaussian (GSG) modeling framework. The results reported in the following are illustrated in Siena et al. (2021).

4.1.1 | EVOLUTION OF CALCITE DISSOLUTION PATTERNS

Fig. 4.1 depicts the AFM (friction) image acquired after a preliminary exposure of the calcite surface to MilliQ water for 1 hour and 30 minutes. The observation time associated with Fig. 4.1 is hereafter denoted as $t = t_0$ and corresponds to the time immediately before the renewal of the solution in contact with the mineral surface. The most evident feature is the presence of precipitates superimposed to the crystallographic steps that can be seen in the background. The formation of precipitates is compatible with the development of a supersaturated zone within a boundary layer at the solid/fluid interface (Renard et al., 2019). Fig. 4.2 collects the images acquired at a uniform time step ($\Delta t = 5$ min). Each plot is associated with the time t_i ($i = 1, ..., 6$) elapsed between the fluid replacement operation and the end of the acquisition of the $i - th$ image. The key features associated with these results are illustrated in the following.

- At time $t_1 = 5$ min (Fig. 4.2.A), the precipitates observed at time t_0 have been dissolved and the underlying steps of the crystal become clearly visible. Several rhombic

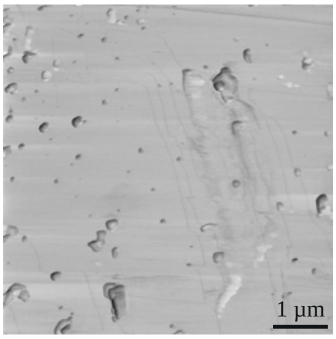

Figure 4.1: *In situ* AFM friction image acquired on a $6 \times 6 \mu m^2$ portion of the calcite sample surface at $t = t_0$, after 1 hour and 30 minutes of contact with stagnant MilliQ water.

monolayer pits (mP), whose appearance is typically associated with high undersaturation, begin to develop (with particular reference to the wider terraces). The edges of a large multilayer pit (hereafter denoted as MP1) are visible on the lower-right corner of the image.

- Time $t_2 = 10$ min (Fig. 4.2.B) is characterized by an increased density of mPs that have also widened with respect to t_1, new mPs nucleating at seemingly random locations and in the center of the existing ones. A new multilayer pit (denoted as MP2) starts to develop at the top left corner. Terrace steps on the surface appear to be stable.
- No newly formed mPs are detectable at time $t_3 = 15$ min (Fig. 4.2.C). The existing ones widen and coalesce, their shape mutating from rhombic to triangular. The acute step of MP1 is expanding slowly, whereas MP2 has spread considerably along the obtuse step direction and the margin of a third multilayer pit (MP3) is seen to invade the domain of investigation right below MP2. The step profiles close to these two pits begin to evolve.
- Most of the mPs are no longer visible at time $t_4 = 20$ min (Fig. 4.2.D), as they are essentially merged with other pits or with the edges of retreating steps. Compared to the previous image, the shape of the step profiles has changed significantly and in an irregular fashion, due to the progressive inclusion of mPs of differing sizes. The

margin of a high step (visible along the top of the investigated region) seems to prevent further spreading of the obtuse step $[48\bar{1}]_+$ of MP2. The expansion of the $[48\bar{1}]_+$ edge of MP3 results in the merging of this MP with MP2.

- No isolated mPs can be seen at time $t_5 = 25$ min (Fig. 4.2.E). The spreading of MPs slows down considerably. Dissolution is basically occurring solely by step retreat.
- At time $t_6 = 30$ min (Fig. 4.2.F) no relevant differences with respect to the pattern observed at t_5 can be seen, with the only exception of the evolution of the steps, the distance between neighboring ones being generally reduced.

The dissolution pattern on the crystal surface is strongly related to the driving dissolution mechanisms detailed in Section 2.1, these being in turn affected by the calcite concentration at the fluid/solid interface. The evolution described above is consistent with the expected temporal increase of concentration within the boundary layer. As opposed to what can be observed at $t = t_0$ (Fig. 4.1), no precipitates are detected at $t = t_6$. This could suggest that local supersaturation of the boundary layer has not been attained yet. Otherwise, this could also be ascribed to the effect of the scanning probe on the investigated area, as the AFM tip may displace precipitates that are weakly connected to the surface (see also Guren et al., 2020 and Renard et al., 2019). Spreading of MP1 takes place only along the acute step and is characterized by a lower extent than what can be noted for MP2 and MP3. Based on the behavior observed for MP2 at t_3 and t_4, we can assume that its growth is limited by the presence of larger steps or pits, even as the overall width of our observation window does not enable us to completely verify this hypothesis.

4.1.2 | EVALUATION OF ETCH PIT SPREADING RATE

A measure of surface reactivity is given by the spreading rate (ν [nm s^{-1}]) of the etch pits. The latter can be estimated through Eq. 2.7 by considering the variation of the separation distance between opposite sides of the pit within a given time interval. We rely on pairs of consecutive images and consider all isolated (i.e., not aggregated) monolayer pits to obtain

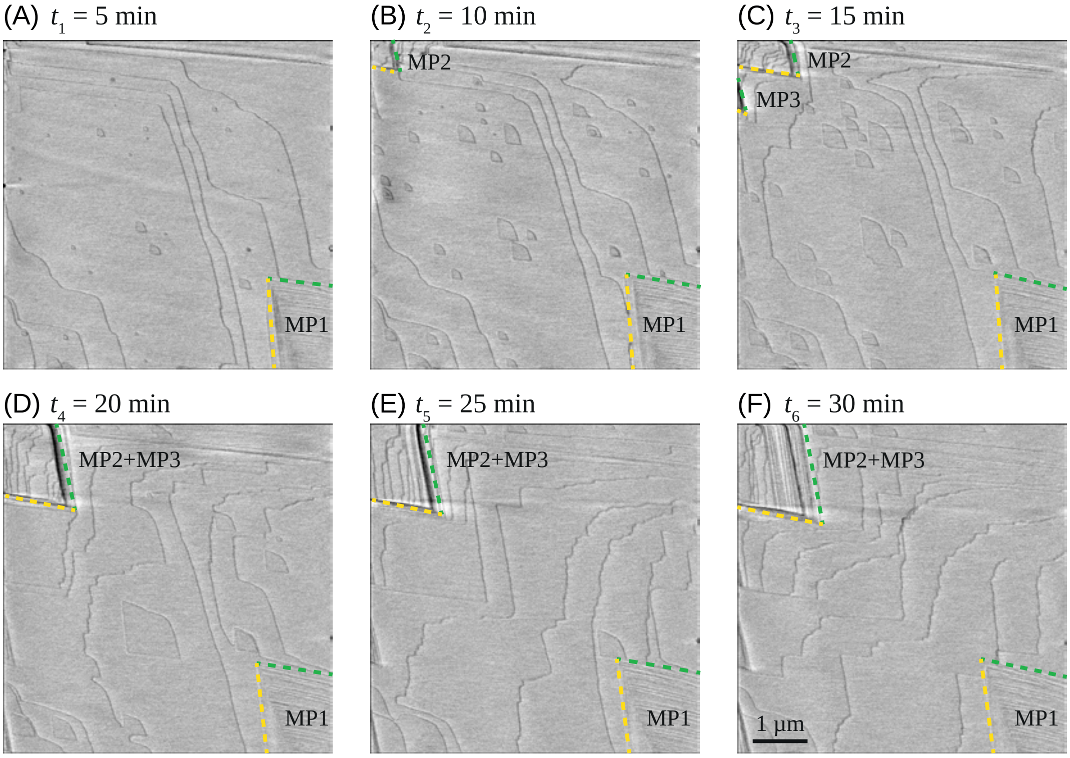

Figure 4.2: *In situ* AFM friction images acquired on a $6 \times 6\,\mu\mathrm{m}^2$ portion of the calcite sample surface at regular time intervals of 5 min after $t = t_0$. The edges of the multilayer etch pits (MP1, MP2, and MP3) are highlighted, green and yellow lines representing the obtuse and acute steps respectively.

an average value, $\bar{\nu}$, and the associated standard deviation, σ_ν. Table 4.1 lists the results of the evaluation of the spreading rate. As highlighted by Fig. 4.2, most of the mPs are short-lived. As such, the number of elements upon which ν can be evaluated decreases sharply over time, none of these structures being detected for $t_3^* > 20\,\mathrm{min}$ (with $t_j^* = t_i + \Delta t$, $i = j = 1, ..., 5$). The results listed in Table 4.1 show that $\bar{\nu}$ and σ_ν monotonically decrease in time, the observed values being consistent with those documented in previous experiments concerning calcite $\{104\}$ dissolution in MilliQ water (Guren et al., 2020; Harstad and Stipp, 2007). The rate of expansion of multilayer pits can be used as an additional element according to which one can estimate the spreading rate on the investigated surface. Even as

Time	t_1^*	t_2^*	t_3^*
Number of mPs	17	15	4
$\bar{\nu}$ [nm/s]	0.44	0.41	0.27
σ_ν [nm/s]	0.11	0.08	0.05

Table 4.1: Mean, $\bar{\nu}$, and standard deviation, σ_ν, of the spreading rate evaluated through Eq. 2.7 over all isolated monolayer pits (mPs) at various times. No isolated mPs can be found in our experiments for $t_4 > 20\,\text{min}$.

they do not fall entirely within our observation domain (Fig. 4.2), these structures are more persistent than their monolayer counterparts, thus enabling one to obtain at least a rough estimate of step retreat velocity throughout the whole temporal window of the experiments. The retreat velocity evaluated by considering the spreading rates of acute and obtuse edges (denoted with yellow and green lines in Fig. 4.2, respectively) and then averaging these over the three MPs, is depicted as a function of time in Fig. 4.3.A. Fig. 4.3.B illustrates the location of the MP margins at various observation times. As opposed to what observed for the shallow pits, the trend observed for these rates is not monotonically decreasing. It increases sharply between t_2^* and t_3^*, while decreasing abruptly for longer times. Consistent with the findings of Harstad and Stipp (2007), MP spreading rates are considerably larger than those listed in Table 4.1 for mP spreading rates.

4.1.3 | Analysis and statistical modeling of dissolution rates

Fig. 4.4 depicts spatial maps of rates, $R'(\mathbf{x}, t^*)$, evaluated through Eq. 2.4 considering consecutive pairs of topography images. Analysis of these results clearly highlights that (*i*) values of R' associated with MP spreading are much larger than their counterparts related to mPs expansion and step retreat across the (overall) flat terrace in between the MPs (see in particular Fig. 4.4.C); and (*ii*) the largest support spanned by sample PDFs of rates is observed at an intermediate time, i.e., for $R'(t_4^*)$. Sample probability densities of R' for each of the sub-plots of Fig. 4.4 are depicted in Fig. 4.5. These PDFs exhibit common features,

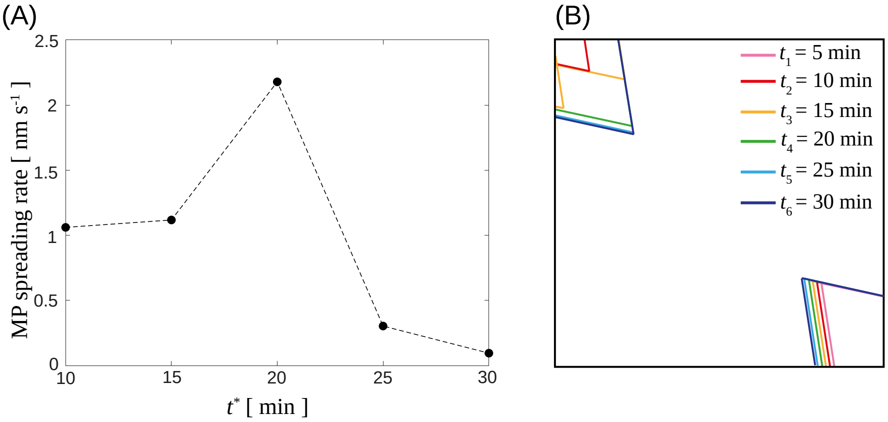

Figure 4.3: (A) Spreading rate evaluated on the basis of the displacements (inferred from the locations of multilayer etch pit (MP) margins at acquisition times t_i with $i = 1, ..., 6$ in **(B)**) of acute and obtuse steps of multilayer pits (MPs) versus time, t^*.

which have been detected also in previous studies (Bibi et al., 2018; Trindade Pedrosa et al., 2019). These include (*i*) the presence of a dominant peak, eventually accompanied by multiple local peaks; and (*ii*) a positive skewness. Consistent with our discussion about Fig. 4.4, values of standard deviation tend to increase from t_1^* to t_3^* and then decrease for longer times. We consider the statistical characterization of the spatial increments, ΔR, evaluated at various separation distances between pairs of locations. We consider omnidirectional increments, i.e., we set $s = |\mathbf{s}|$, deferring the analysis of possible anisotropic behaviors to future studies. Fig. 4.7 depicts sample PDFs of incremental data for three selected lags ($s = 1, 10$, and $100\ dl$). These distributions are generally more symmetric than their counterparts associated with R' (Fig. 4.5). It is clear that the shape of these PDFs varies with separation distance, their peak becoming sharper with decreasing lag, a feature which is common to corresponding results associated with a variety of other Earth science variables (see, e.g., Guadagnini et al., 2018, Siena et al., 2020 and references therein). All distributions are persistently leptokurtic, with $\kappa_{R'} \gg 3$ even at large lags ($s = 100\ dl$). We estimate GSG model parameters relying upon method MOM_B detailed in Section 3.5.2. Fig. 4.6 depicts estimated shape, $\hat{\alpha}$, and scale, $\hat{\sigma}_G$, parameters versus lag, s. As expected, these parameters do not change significantly with s. We consider average values of $\hat{\alpha}$ and

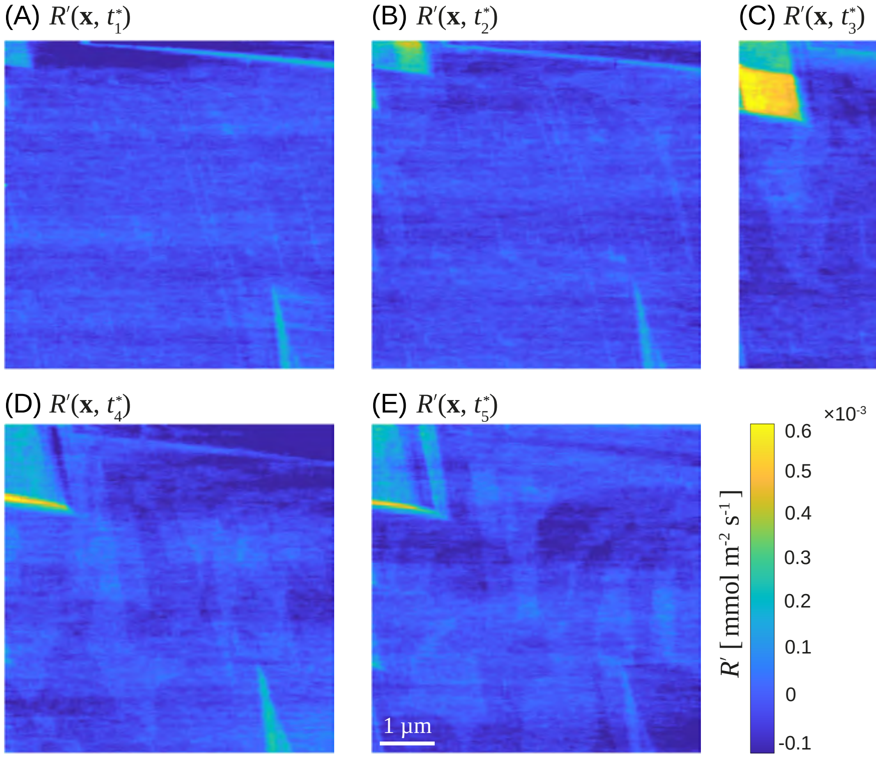

Figure 4.4: Spatial fields of dissolution rates, $R'(\mathbf{x}, t^*)$, evaluated for each consecutive pair of topography images (taken at uniform temporal intervals $\Delta t = 5$ min).

$\hat{\sigma}_G$ (dashed lines in Fig. 4.6) to evaluate analytical GSG expressions.

The results related to the theoretical formulations of the GSG PDFs of R' and ΔR with parameters estimated through the method of moments show a remarkable agreement with their sample counterparts (see Figs. 4.5 and 4.7, respectively). As an additional element to support the analysis, we note that, while the method of moments is simple and straightforward in its application, parameter estimates obtained through ML applied on R' and ΔR data provide results of similar quality (not shown). Results obtained upon relying on a GEV model, i.e.,

$$f_{R'}^{GEV}(r) = \frac{1}{\sigma}\left[1 + k\left(\frac{r-\mu}{\sigma}\right)\right]^{-1-\frac{1}{k}} e^{-\left[1+k\left(\frac{r-\mu}{\sigma}\right)\right]^{-\frac{1}{k}}} \tag{4.1}$$

that has been adopted in previous works for the interpretation of rate spectra (Brand et al., 2017; Emmanuel, 2014) are also depicted in Fig. 4.5 as a further term of comparison. Here,

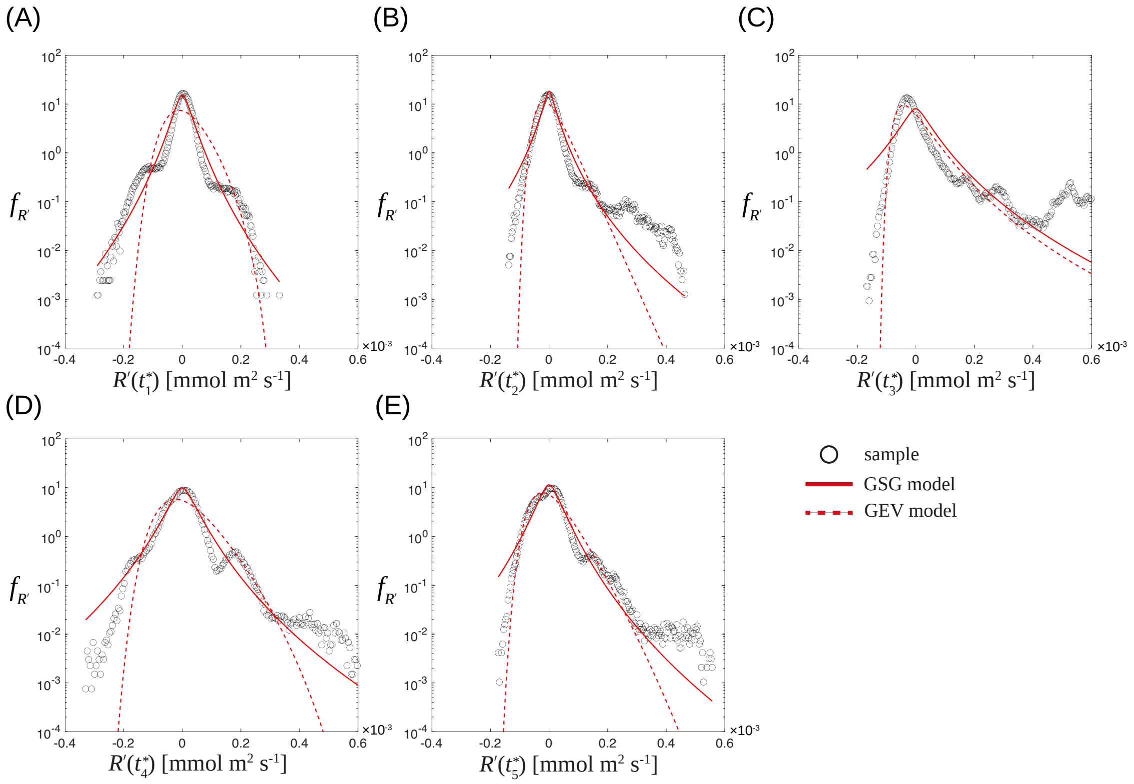

Figure 4.5: Sample probability densities (PDFs) (symbols) associated with the five reaction rate datasets depicted in Fig. 4.4. Interpretive models based on GSG (Eq. D.2) and GEV (Eq. 4.1) distributions are also depicted.

k, σ, and μ correspond to the shape, scale, and location parameter, respectively, and are estimated through a classical ML procedure. Visual inspection of Fig. 4.5 suggests that relying on a GEV model can provide results of similar quality to those that can be obtained through the GSG model with reference to sample PDFs of R'. Otherwise, we note that the GEV model does not include information about the statistical behavior of incremental values, a feature which is naturally embedded in the GSG framework. The Kullback-Leibler Divergence (Kullback and Leibler, 1951), D_{KL}, is then employed to compare quantitatively the performance of the two interpretive models. This metric provides a measure of the amount of information lost by representing the empirical distribution associated with the available data with a given theoretical model. Hence, low values of D_{KL} correspond to

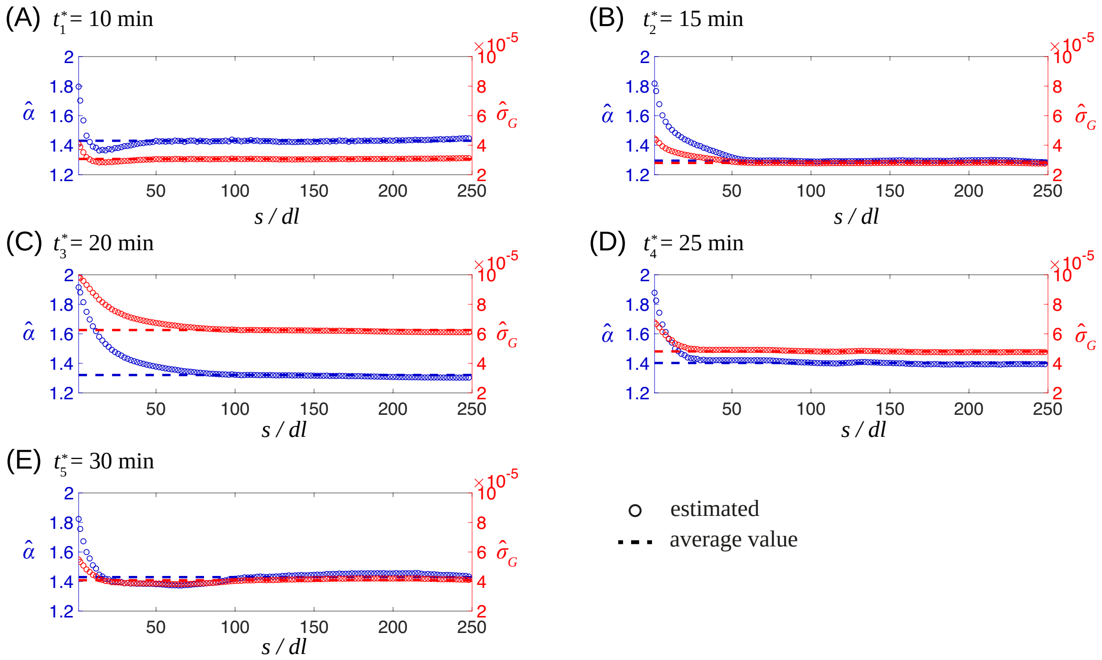

Figure 4.6: GSG shape, $\hat{\alpha}$, and scale, $\hat{\sigma}_G$, parameters estimated by relying upon method MOM_B at **(A)** $t_1^* = 10\,\text{min}$, **(B)** $t_2^* = 15\,\text{min}$, **(C)** $t_3^* = 20\,\text{min}$, **(D)** $t_4^* = 25\,\text{min}$, and **(E)** $t_5^* = 30\,\text{min}$ versus lag. Average values are depicted as dashed lines.

high degrees of similarity between sample and modeled distributions. Table 4.2 lists the results of this analysis, suggesting that the GSG model outperforms the GEV model for all observation times, with the exception of $t_3^* = 20\,\text{min}$, where one can see that both analytical models provide a good representation of the upper tail while the GEV captures the peak of the sample PDF more closely that its GSG-based counterpart.

Fig. 4.8 depicts the temporal behavior of estimated shape (Fig. 4.8.A) and scale (Fig. 4.8.B) parameters obtained for the GSG and GEV models. Shape parameters, $\hat{\alpha}$ and $\hat{k}$, linked to the tailedness of the distributions, respectively display a minimum and a maximum for $t^* \approx 15 - 20$ min. This pattern is consistent (for $\hat{k}$) with the one exhibited by the spreading rate of the MPs (Fig. 4.3.A), the corresponding results documented for $\hat{\alpha}$ being characterized by a trough, reflecting symmetrically the temporal behavior of $\hat{k}$. Therefore, these results suggest that the time evolution of these structures has a major effect on the statistical

Time	t_1^*	t_2^*	t_3^*	t_4^*	t_5^*
D_{KL} sample PDF and GSG	**0.048**	**0.049**	0.331	**0.033**	**0.051**
D_{KL} sample PDF and GEV	0.373	0.145	**0.133**	0.173	0.070

Table 4.2: Kullback-Leibler divergence (D_{KL}) evaluated from the comparison of sample PDFs of R' and GSG (Eq.D.2) or GEV (Eq. 4.1) analytical models.

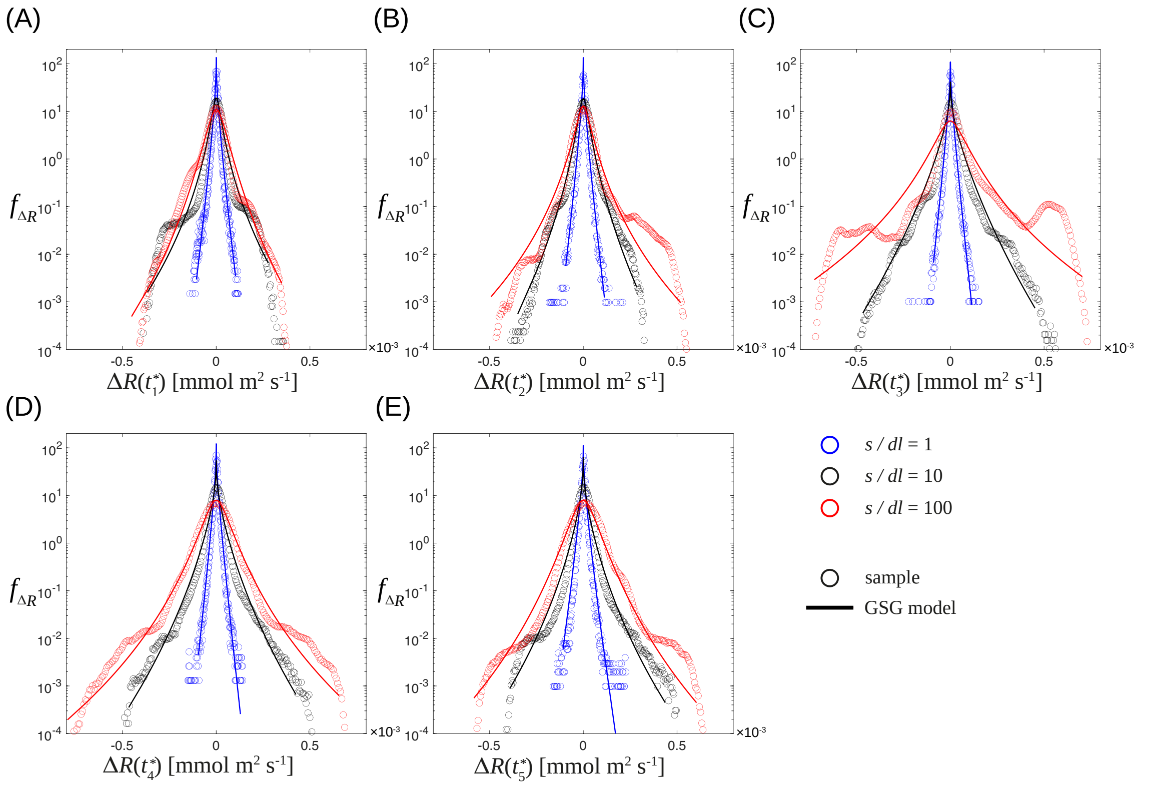

Figure 4.7: Sample probability densities (PDFs) of rate increments, $\Delta R(t_j^*)$ ($j = 2, ..., 6$), evaluated at three selected lags, s. Interpretive models based on GSG distributions for incremental data (Eq. D.6) are also depicted.

distribution of R'. The temporal trend exhibited by $\hat{\sigma}_G$ indicates an increased spreading of the distribution of rates at t_3^*, whereas the peak of the GEV scale parameter, $\hat{\sigma}$, is attained at a later time (i.e., t_4^*). The evolution of the spatial correlation of rates can be inferred from the analysis of the estimated $\hat{\rho}_G$ versus lag at various observation times (Fig. 4.9).

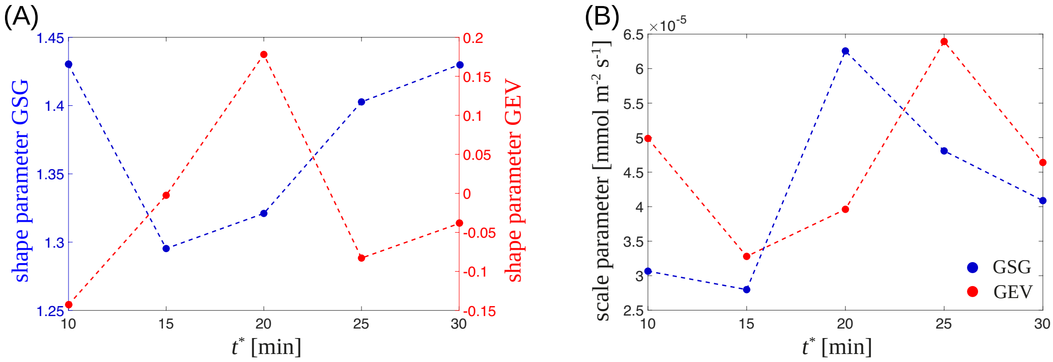

Figure 4.8: **(A)** Shape and **(B)** scale parameters embedded in GSG and GEV models versus time.

The skill of a single-parameter exponential model, i.e.,

$$\rho_G(s) = e^{-s/a} \tag{4.2}$$

to represent the observed correlation is compared against the results obtained through two selected formulations of nested structures, whose components are associated with standard stationary variograms. We consider two nested structures entailing (i) a spherical and a Gaussian model (Eq. 4.3) and a spherical and a hole-effect model (Eq. 4.4), i.e.,

$$\rho_G(s) = \begin{cases} 1 - \left\{ \frac{c}{\sigma_G^2} \left[\frac{3}{2}\frac{s}{a_1} - \frac{1}{2}\left(\frac{s}{a_1}\right)^3 \right] + \frac{\sigma_G^2 - c}{\sigma_G^2} \left[1 - e^{-\left(\frac{s}{a_2}\right)^2} \right] \right\} & \text{for } s < a_1 \\ 1 - \frac{c}{\sigma_G^2} - \frac{\sigma_G^2 - c}{\sigma_G^2} \left[1 - e^{-\left(\frac{s}{a_2}\right)^2} \right] & \text{for } s \geq a_1. \end{cases} \tag{4.3}$$

$$\rho_G(s) = \begin{cases} 1 - \left\{ \frac{c}{\sigma_G^2} \left[\frac{3}{2}\frac{s}{a_1} - \frac{1}{2}\left(\frac{s}{a_1}\right)^3 \right] + \frac{\sigma_G^2 - c}{\sigma_G^2} \left[1 - \cos\frac{\pi s}{a_2} \right] \right\} & \text{for } s < a_1 \\ 1 - \frac{c}{\sigma_G^2} - \frac{\sigma_G^2 - c}{\sigma_G^2} \left[1 - \cos\frac{\pi s}{a_2} \right] & \text{for } s \geq a_1. \end{cases} \tag{4.4}$$

Each of these models is defined in terms of three parameters: (i) the range of each component, a_1 and a_2, and (ii) the coefficient c, which determines the relative contribution (or weight) of each component. We note that these formulations appear to be more complex than the exponential model (Eq. 4.2). Nevertheless, we prefer to consider also these formulations due to the usefulness of nested variogram structures to interpret spatially heterogeneous settings where multiple processes, each characterized by their own degree of spatial persistence, can jointly contribute to the resulting heterogeneous pattern of the overall system.

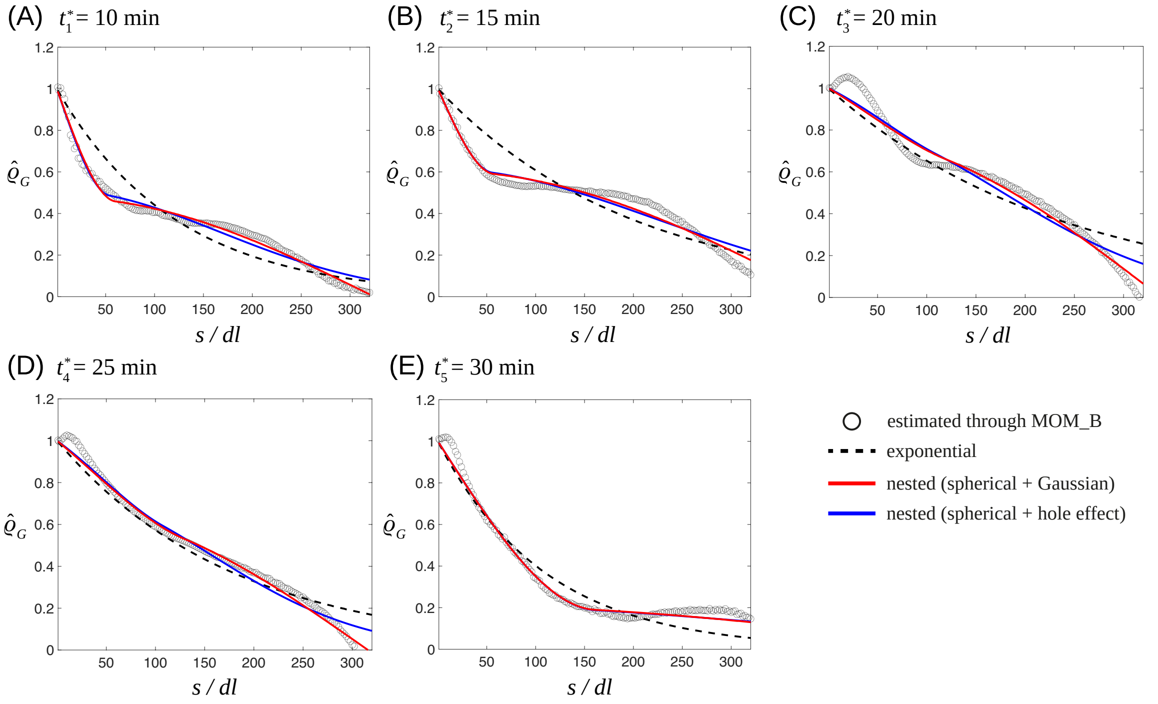

Figure 4.9: Estimated values of the correlation of G, $\hat{\rho}_G$, versus lag obtained from MOM_B at all observation times. Results from calibration of exponential (Eq. 4.2) and nested (Eqs. 4.3 and 4.4) analytical models are also depicted.

The effectiveness of the models corresponding to Eqs. 4.2, 4.3, and 4.4 is evaluated within a ML framework. While estimates of model parameters are obtained through minimization of the Negative Log-Likelihood (NLL) (see e.g., Carrera and Neuman, 1986) criterion, the performance of a given model is ranked according to the well-established Kashyap Information Criteria (KIC) for model discrimination (Kashyap, 1982):

$$\mathrm{KIC} = \mathrm{NLL} - N_P \ln 2\pi - \ln |\mathbf{Q}|. \tag{4.5}$$

Here, N_P is the number of parameters associated with a given model and $|\mathbf{Q}|$ is the determinant of the covariance matrix of the ML parameter estimation errors. Relying on KIC enables one to consider the quality of model fit to observations (via NLL) while jointly penalizing models with large N_P and fully considering the quality of parameter estimates. The latter aspect is embedded in $|\mathbf{Q}|$, which acts as a term penalizing models with small variance (i.e., large expected information content per observation) of parameter estimates (Ye et al., 2008). In this context, model ranking is performed according to increasing values of KIC, lower values of the latter corresponding to more skillful models. Figs. 4.9.A-E depict the values of $\hat{\rho}_G$ estimated at various times along the experiment together with the results obtained through the calibrated theoretical models. A qualitative inspection of these plots suggests that the nested formulations outperform the simple exponential model. This is quantitatively supported by the ML results collected in Table 4.3, where one can note that a nested structure composed by a spherical and a hole effect model (Eq. 4.4) is always ranked as best among the three models assessed. This result imbues us with confidence about the possibility to include in our interpretive model the richness of mechanisms underlying the system evolution through the superimposition of diverse correlation structures. We note that the contribution of a given component of a nested structure to acting processes can be assessed in simple large scale sedimentary settings, where these can be related to temporal sequences of depositional processes (e.g., Salamon et al., 2007 and references therein). As this is the first time, to the best of our knowledge, that these types of analyses are performed for scenarios of the kind we focus upon, disentangling the way components of nested structures can be associated with the action of kinetic mechanisms described in Section 4.1.1 poses significant challenges. We present preliminary interpretations in the following, noting that future works will be keyed to further support these results with additional data. Following the results illustrated above, we focus on the parameters of the most highly ranked model (Eq. 4.4) and analyze their evolution in time. Figs. 4.10.A-B depict the behavior of $\left(c_1 = \frac{c}{a_1}, a_1\right)$ and $\left(c_2 = 1 - \frac{c}{\sigma_G}, a_2\right)$, i.e., relative contribution and range of component 1 and 2 in Eq. 4.4, respectively. It can be noted that $a_2 \gg a_1$ at all times, components 1 and 2 being respectively related to short- and long-range correlation. These results show that a_1 tends to increase with time, ranging from 60 to 160 dl, indicating that short-range correlations increase as the dissolution pattern on the terraces evolves from being monolayer-pit

Time	t_1^*		t_2^*		t_3^*		t_4^*		t_5^*	
	NLL	KIC	NLL	KIC	NLL	KIC	NLL	KIC	NLL	KIC
Eq. 4.2	-333.8	-337.8	-298.6	.303.6	-311.3	-316.6	-377.0	-381.3	-408.6	-411.9
Eq. 4.3	-613.2	-612.1	-613.2	-613.8	-429.3	-434.5	-507.3	-506.4	-615.3	-618.7
Eq. 4.4	**-756.2**	**-753.1**	**-646.6**	**-646.5**	**-515.0**	**-518.8**	**-672.0**	**-672.3**	**-619.4**	**-623.8**

Table 4.3: Values for negative Log-Likelihood (NLL) and Kashyap (KIC) model identification criteria obtained from Maximum Likelihood parameter estimation for the selected analytical models of $\hat{\rho}_G$.

to step-retreat dominated. Otherwise, the long-range correlation parameter, a_2, displays an oscillatory behavior within the range $640 - 800$ dl, showing a steep increase only at t_5^*. Length scales in this range are comparable with the distance between the multi-layer pits developing on the opposite corners of the observation window. Additional evidence about a possible relationship between the dynamics of multilayer pits and the long-range component of $\hat{\rho}_G$ can be inferred from the analysis of parameters c_1 and c_2. Our results document that (i) both components have similar weight ($c_1 \approx c_2 \approx 0.5$) at t_1^*; and (ii) the long-range component becomes dominant at $t_3^* = 20$ min, its relevance decreasing for $t > t_3^*$ (attaining the value $c_2 = 0.22$ at $t_5^* = 30\,\text{min}$). Comparison of these results with those depicted in Fig. 4.3.A evidences that the relative importance of the long-range component mirrors the trend displayed by the MPs spreading rate. Fig. 4.10 provides a clear indication supporting a conceptual picture according to which the dynamics of multilayer pits markedly and quantifiably affect the frequency distribution as well as the spatial correlation of rates. Our findings further support the benefit of relying on a modeling framework capable of jointly embedding the statistical behavior of rates and of the associated increments. The importance of characterizing the spatial correlation of key variables driving mineral dissolution processes was highlighted by Pollet-Villard et al. (2016). These authors develop a numerical model to describe dissolution and ground model calibration on the comparison between sample variograms evaluated on experimental data of surface topography and its numerically based counterpart. Our results reveal the impact that the diverse dissolution patterns can have on the correlation structure of reaction rates. This information can potentially lead to the development of future flexible numerical models, which can have the capability of taking into account multiple length scales resulting from the occurrence of diverse reaction mechanisms.

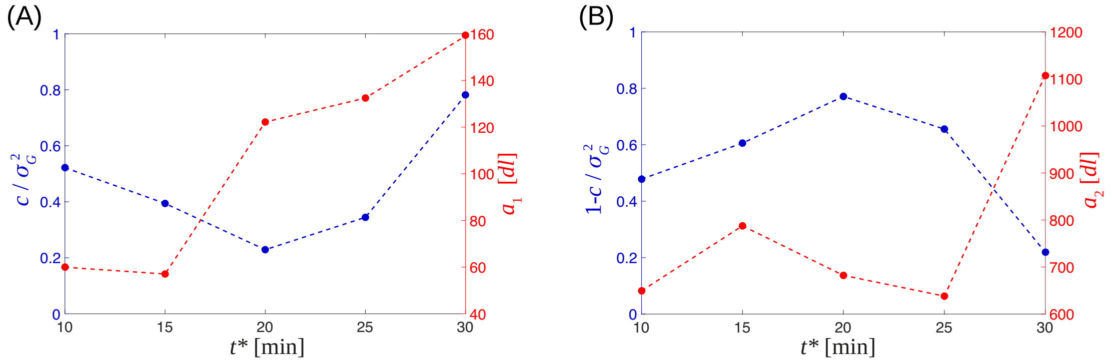

Figure 4.10: Time evolution of the parameters of the most highly ranked model interpreting ρ_G (Table 4.3): relative contribution and range of **(A)** component 1 and **(B)** component 2 of the nested structure (Eq. 4.4).

4.2 | Results of experimental *Setting 2*

We discuss the application of the Gaussian Mixture (GMIX) model to interpret the statistical behavior of spatially heterogeneous fields of dissolution rate. These are obtained from topographic measurement of a (non-masked) sample acquired under intermittent flow conditions, i.e., relying upon experimental *Setting 2* (see Section 2.3.2). Imaging is performed in contact mode over a scanning window of $6 \times 6\ \mu\text{m}^2$ across a 512×512 pixels (pixel size $dl = 11.7\,\text{nm}$). The acquisition frequency is here set to $f_a = 1.41\,\text{Hz}$. This set of acquisition parameters yields a scanning time (coinciding with the fluid residence time in the cell) $T_{c_2} = 6\,\text{min}$, similar to the datasets discussed in Section 2.4.2 and Appendix B. Refreshing of the fluid in contact with the sample surface is performed between each scanning for a time $\approx 30\,\text{s}$. Hence, the time interval between subsequent frames is $\Delta t \approx 6.5\,\text{min}$. Dissolution rate fields are evaluated through Eq. 2.4 from topographic maps taken at two times separated by a temporal interval equal to $2\Delta t = 13\,\text{min}$. The results reported in the following are included in Siena et al. (2023).

4.2.1 | Evolution of calcite dissolution patterns

Fig. 4.11 collects maps of $z'(\mathbf{x}, t)$ acquired during the dissolution experiment. These provide a qualitative appraisal of the temporal evolution of the crystal surface during the reaction. We observe two main topography patterns. These are respectively related to (*i*) the spreading of a multilayer (deep) etch-pit in the bottom left corner; and (*ii*) the nucleation, spreading and coalescence of several monolayer (shallow) etch-pits taking place on the terrace. As a consequence, topography maps can be subdivided into two regions, hereafter termed *Multilayer Region* and *Terrace Region*, respectively. These patterns are consistent with published literature studies regarding dissolution in far-from-equilibrium conditions described in Section 2.1 (e.g., Teng, 2004 and Bouissonnié et al., 2018).

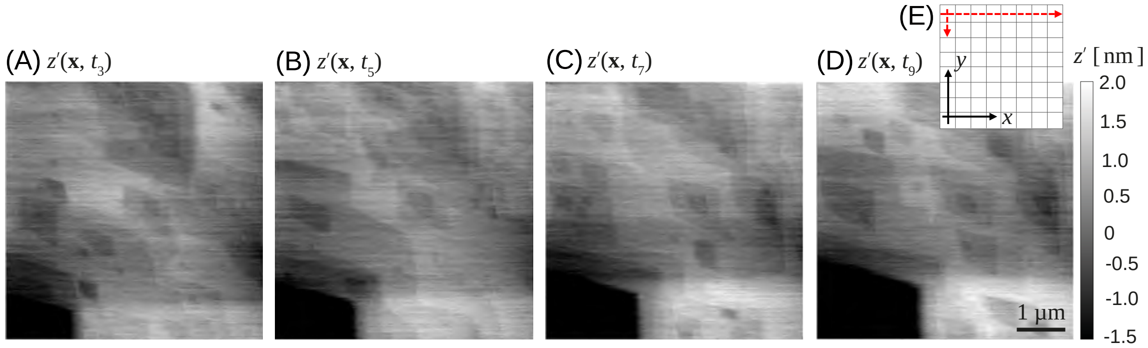

Figure 4.11: Images of fluctuation of calcite topography about its mean, $z'(\mathbf{x}, t_i)$, acquired via AFM at **(A)** $t_3 = 26\,\text{min}$, **(B)** $t_5 = 39\,\text{min}$, **(C)** $t_7 = 52\,\text{min}$ and **(D)** $t_9 = 65\,\text{min}$ from the beginning of the experiment. Scanning directions are depicted in **(E)**.

4.2.2 | Analysis and statistical modeling of dissolution rates

Figs. 4.12.A-D depict spatial distributions of $R'(\mathbf{x}, t_j^*)$, with $t_j^* = t_i + 2\Delta t$ ($j = i = 1, ..., 8$). The corresponding sample PDFs of R' are depicted in Figs. 4.12.E-H. All PDFs exhibit a prominent peak at $R' \approx 0$ and a secondary peak for $R' > 0$. From a qualitative standpoint, these results suggest that all points that belong to the same topography region (either Terrace or Multilayer) at times t_i and $t_i + 2\Delta t$ contribute to the highest peak. Otherwise, the secondary peak is driven by values of rate that are associated with locations that transition from one topography region to the other during the time interval $2\Delta t$, following the spreading of the multilayer etch pit. The observed bimodal trait of the sample PDFs of R' is consistent with an interpretation based on the GMIX stochastic framework introduced in Section 3.3. We denote hereafter as components A and B those associated with the peak at $R' > 0$ and at $R' \approx 0$, respectively.

We compute spatial increments of dissolution rate, ΔR, at various separation distances. Sample statistics are evaluated considering omnidirectional increments, with the only exception of the direction parallel to the AFM acquisition (denoted as x in Fig. 4.11.E), to avoid spurious correlation originated from measurement artifacts. This is consistent with the study of Marinello et al. (2010) who show that AFM measurements are often affected by

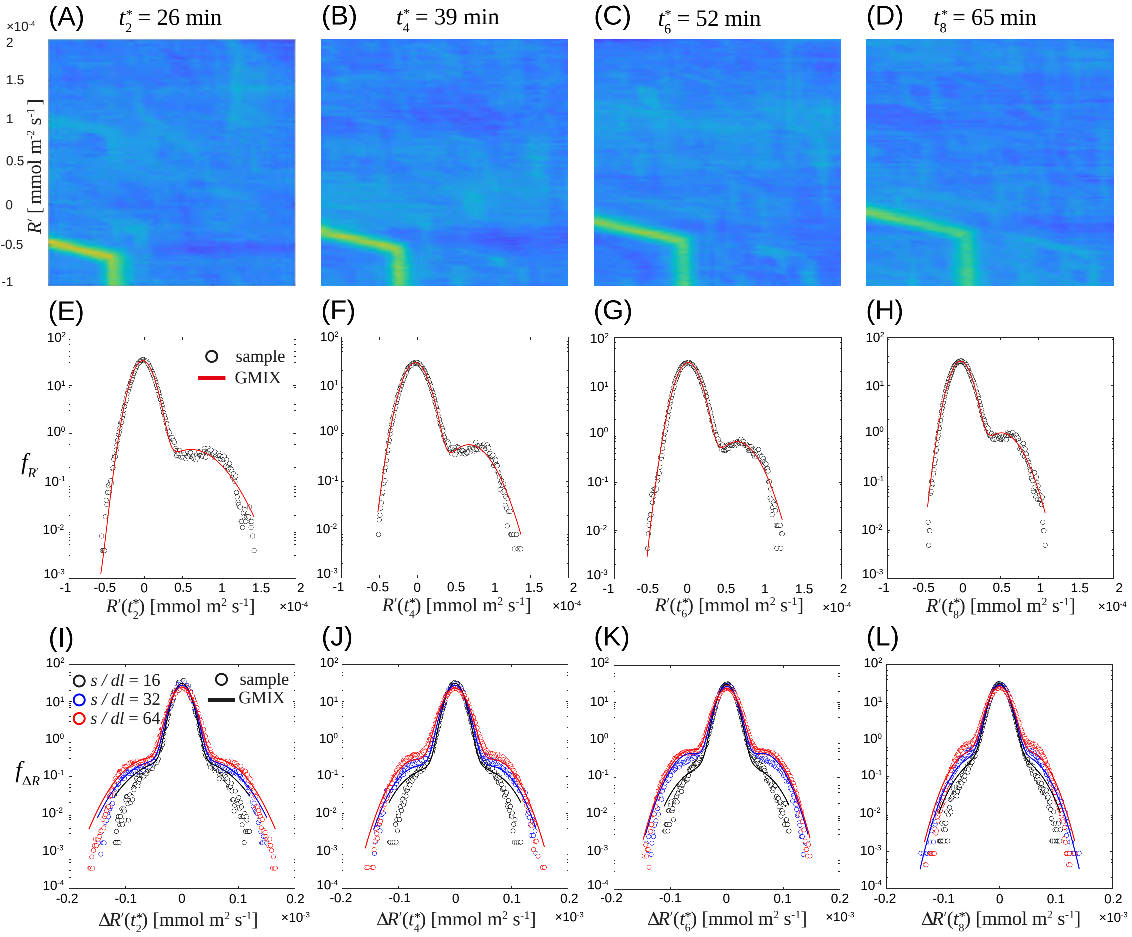

Figure 4.12: Dataset and statistical results associated with observed dissolution rate maps, R'. Spatial maps of $R'(\mathbf{x}, t_j^*)$ evaluated with Eq. 2.4 from AFM topography measurements depicted in Fig. 4.11 are shown for the considered times **(A-D)**. Sample PDFs of **(E-H)** R' and **(I-L)** ΔR are also depicted. Analytical results for the PDFs of reaction rate (Eq. 3.31) and its spatial increments (Eq. 3.44) are juxtaposed to the experimental data.

stripe noise, i.e., a distortion of the signal occurring along the principal scanning direction. Figs. 4.12.I-L depict sample PDFs of increments ΔR for lags s = 16, 32, 64 dl, encompassing short and large distances relative to the size of the domain. All of these PDFs share some common features with their counterparts described in Section 3.3 in the context of the GMIX framework, i.e., they display (i) an overall symmetric behavior, (ii) the presence of a dominant peak coupled with lateral peaks, and (*iii*) a tendency to change their main traits with lag, denoting a scaling behavior.

The GMIX parameters as well as those associated with the distribution of $f_{\Delta R}$ are assessed according to the procedure detailed in Section 3.6.1, λ_I being estimated via *method 2*. The analytical formulations of $f_{R'}$ and $f_{\Delta R}$ (Eqs. 3.31 and 3.44) obtained upon considering the GMIX parameters estimated at each time are juxtaposed to their sample counterparts in Figs. 4.12.E-H and 4.12.I-L, revealing a remarkably satisfactory agreement. The analysis of the GMIX parameters at different times provides insights on the temporal evolution of the mechanisms driving the dissolution reaction. Temporal variations of parameter values are mainly linked to component A of the mixture. This is related to the observation that the area associated with category A is subject to higher relative variations than the corresponding one related to category B, with an average variation of $\sim 36\%$ and $\sim 4\%$, respectively, across the total temporal window analyzed. The overall temporal increase of $\hat{p}$ and $\hat{\lambda}_I$ (Fig. 4.13.A and Fig. 4.13.B) reflects a progressive growth of the area associated with category A. Such increasing trend is consistent with the (approximately) constant horizontal spreading rate, ν, of the MP (see Fig. 4.13.C) evaluated through Eq. 2.7. Here, we can only evaluate the spreading rate of MP acute steps ν_{ac} because no obtuse step fall inside the observation window. Hence, ν_{ac} is estimated from the separation distance between etch pit edges at subsequent times (reported in Fig. 4.13.D). It can be noted that the order of magnitude of the results depicted in Fig. 4.13.C is consistent with existing results recently documented in the literature (e.g., Guren et al., 2020; Dong et al., 2020). The mean of both components A and B remains almost constant with time (Fig. 4.13.E). Otherwise, a decreasing temporal trend is observed for the variance of component A, whereas $\hat{\sigma}_B^2$ remains almost constant (Fig. 4.13.F). The documented pattern suggests that values of the second moment associated with component A progressively becomes more similar to its counterpart related to component B. This is also consistent with the observed temporal dampening of the multimodal behavior displayed by the PDF of R'. This trend is also revealed by an observed temporal decrease for Sk_R and $E\kappa_R$ (not shown).

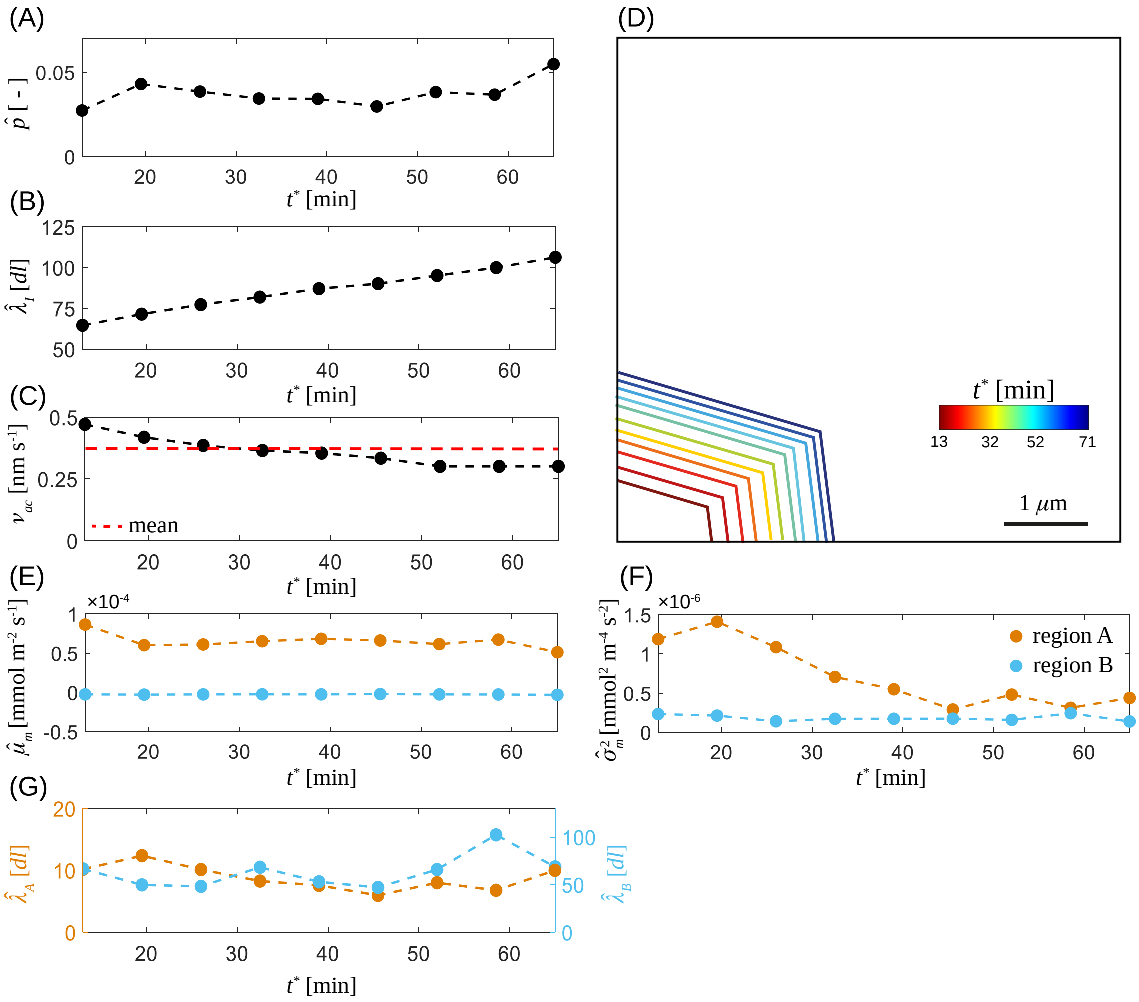

Figure 4.13: Temporal trend of parameters of the GMIX model. The evolution of estimated parameters associated with the indicator random field ($\hat{p}$ and $\hat{\lambda}_I$) is depicted in **(A)** and **(B)**, respectively. The behavior of mean, $\hat{\mu}_m$ **(E)**, variance, $\hat{\sigma}^2_m$ **(F)**, and correlation length, $\hat{\lambda}_m$ **(G)** for component $m = (A, B)$ is illustrated. Panel **(C)** depicts the horizontal spreading rate of acute steps, ν_{ac}, associated with the etch-pit edges shown in **(D)** at various times.

The temporal evolution of the spatial correlation structure of each component of the mixture is inferred from the analysis of $\hat{\rho}_m$ $(m = A, B)$. Fig. 4.14 depicts the sample spatial correlation associated with regions A and B. The following common traits can be noted at all times: (i) an oscillating behavior at large separation distances for $\hat{\rho}_A$ and (ii) the presence of a nugget effect for both $\hat{\rho}_A$ and $\hat{\rho}_B$. We relate the oscillations in $\hat{\rho}_A$ to the small number of points separated by large lags for region A. Otherwise, the second trait could be attributed to the persisting stripe noise, which might especially influence short lags. We consider the exponential with nugget as interpretive model for $\hat{\rho}_m$. Fig. 4.14 juxtaposes theoretical ρ_m values and their experimentally-based counterparts. The analytical formulation enables one to grasp the main features associated with the experimental setting. Fig. 4.13.G depicts $\hat{\lambda}_A$ and $\hat{\lambda}_B$ versus time. An oscillatory behavior can be noticed, in particular for component B. We relate this trend to the dynamics of the monolayer etch-pits nucleating and spreading on the crystal terrace. Fig. 4.15 juxtaposes the analytical curves associated

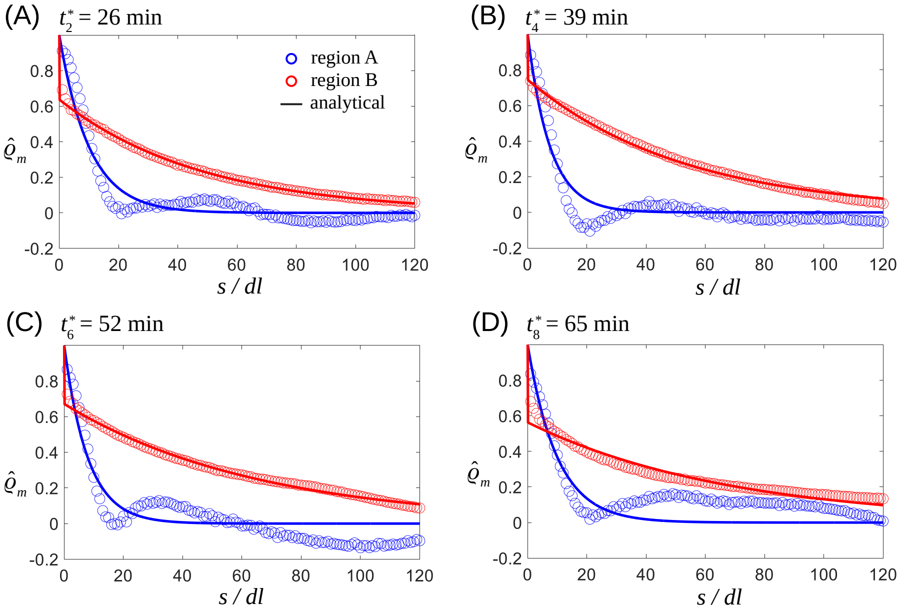

Figure 4.14: Spatial correlation of components A and B associated with the rate maps depicted in Fig. 4.12.A-D as a function of separation distance, s. The analytical interpretive model, i.e., exponential model with nugget, is also depicted.

with the GMIX correlation function (i.e., $\rho_R = 1 - \langle \Delta R^2 \rangle / 2\sigma_R^2$, $\langle \Delta R^2 \rangle$ being evaluated through Eq. 3.45) and excess kurtosis, $E\kappa_{\Delta R}$ (evaluated making use of Eqs. 3.45 and 3.47), of ΔR to their experimental sample counterparts. Theoretical spatial correlation structures shown in Figs. 4.15.A-D exhibit a satisfactory agreement with their sample counterparts, discrepancies being mainly visible at time $t_6^* = 52$ min, for intermediate lags. Here, we notice that sample PDFs of ΔR evaluated for component A appear to deviate, albeit slightly, from a Gaussian behavior (not shown). Such a deviation from Gaussianity could also be at the core of the imperfect agreement observed between sample and theoretical values of $E\kappa_{\Delta R}$.

The results obtained here through the GMIX modeling framework are remarkably promising for the interpretation of high resolution geochemical data at the microscale. They show that model parameters are strictly linked to the temporal evolution of the surface features driving the dissolution reaction.

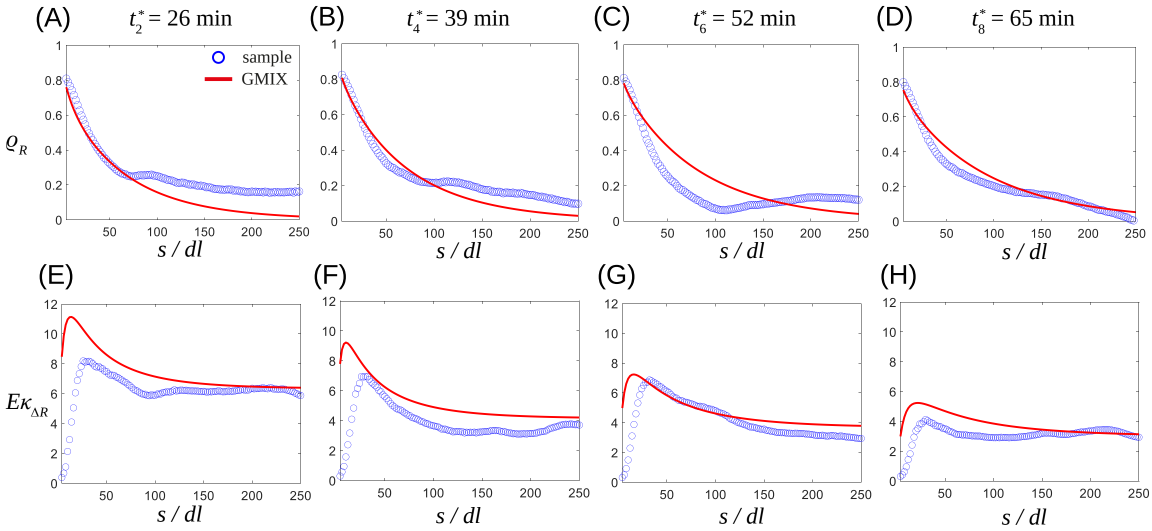

Figure 4.15: Statistical moments associated with the spatial increments of R'. Analytical expressions resulting from the GMIX formulation are juxtaposed to **(A-D)** sample correlation function, ρ_R, and **(E-H)** excess of kurtosis, $E\kappa_{\Delta R}$ associated with the rate maps shown in Fig. 4.12.A-D.

4.3 | Results of experimental *Setting 3* - Masked samples

We discuss discuss experimental and modeling results obtained by imaging a masked calcite sample subject to continuous flow conditions (i.e., experimental *Setting 3*, see Section 2.3.3). Fabrication of the inert mask relies upon the workflow described in Section 2.5. We scan a $20 \times 20\,\mu\text{m}^2$ area of the sample surface, comprising a portion covered by the non-reactive mask $\approx 45\%$ of the observation window (see Fig. 4.17.A). Imaging is performed in tapping mode and the acquisition frequency is here set to $f_a \sim 1.1\,\text{Hz}$, yielding a temporal resolution $\Delta t \sim 17\,\text{min}$. The pixel size is set to $dl = 19.7$ nm. The degree of spatial heterogeneity of associated absolute dissolution rate fields is modeled upon resting on the Gaussian Mixture (GMIX) framework (see Section 3.3). The alignment of topographic data measured at subsequent times is assessed upon relying on *fiducial points* coinciding with centroids of precipitates underneath the inert mask (see Section 2.6). Fig. 4.16.A depicts the phase shift signal associated with the topography measured at t_1. Here, clusters of precipitates are indicated by white arrows. The comparison of the positions of such points at subsequent temporal instants is illustrated in Fig. 4.16.B. These results confirm that the horizontal drift between topography maps is negligible (i.e., below our lateral resolution) as centroids are overlapping at all times. The results illustrated in the following are included in Recalcati et al. (2024).

4.3.1 | Evolution of calcite dissolution patterns

Figs. 4.17.F-H depict snapshots of topographies (scaled with respect to the inert layer, see Section 2.2) at three observation times. The reaction is dominated by vertical deepening and horizontal spreading of two multilayer etch pits oriented according to Fig. 4.17.C and schematically depicted in Fig. 4.17.B. The two pits exhibit different temporal behaviors, yielding a complex surface pattern. The deep etch pit on the left of Figs. 4.17.F-H (denoted as MP-L) evolves by forming trains of steps. The latter behavior is consistent with a description based on a stepwave model (Lasaga and Lüttge, 2001, 2003). The latter conceptualizes material fluxes from the surface as governed by spreading of steps originating

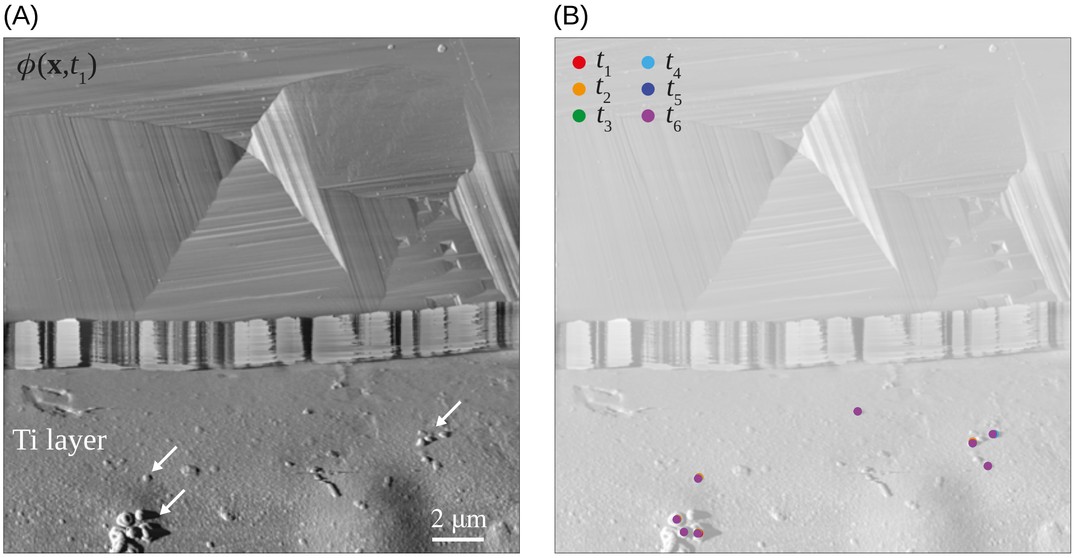

Figure 4.16: **(A)** AFM phase shift signal measured at t_1. Clusters of precipitates that have formed during the optical lithography phase of the mask fabrication workflow and have been covered by the Ti layer are indicated through white arrows. **(B)** Locations of centroids of the precipitates at times $t_1 - t_6$, documenting that results are virtually indistinguishable.

from etch pits, which then travel across the crystal surface. In contrast, the spreading of the etch pit on the right of Figs. 4.17.F-H (denoted as MP-R) remains limited. It is hypothesized that such a distinct behavior arises from the presence of different phases across which the etch pit deepening processes evolves. An etch pit displaying a MP-R type pattern results from the coalescence of two multilayer etch pits. Each of these is related to screw dislocation lines that were parallel to the z direction (and thus active in promoting stepwave emanation from pit center and, in turn, vertical deepening of the pit) within upper layers of the crystal lattice. These layers have been dissolved prior to the beginning of our acquisition. During our temporal observation window, dislocation lines associated with pits that have coalesced to form MP-R might not be oriented anymore along the z direction. Hence, emanation of young dissolution pulses could not be observed at MP-R. Details about the role of dislocations as drivers for etch pit formation and evolution are available in, e.g., Pollet-Villard et al. (2016) and MacInnis and Brantley (1992). Therefore, only the spreading of old stepwaves marking the edges of MP-R can be documented in the experiments. The screw dislocation associated with MP-L remains active throughout our entire observation window. The flat bottom of MP-R acts as a nucleation site for other small etch pits that spread and coalesce. Fig. 4.17.D provides a qualitative appraisal of such small-scale features by depicting an enlargement of the AFM phase shift signal associated with $z(\mathbf{x}, t_3)$ within box 1 of Fig. 4.17.G. Merging of stepwaves radiated in the acute direction from MP-L and MP-R yields a complex pattern of steep rhombohedral steps in the region comprised between these two main features (Fig. 4.17.E).

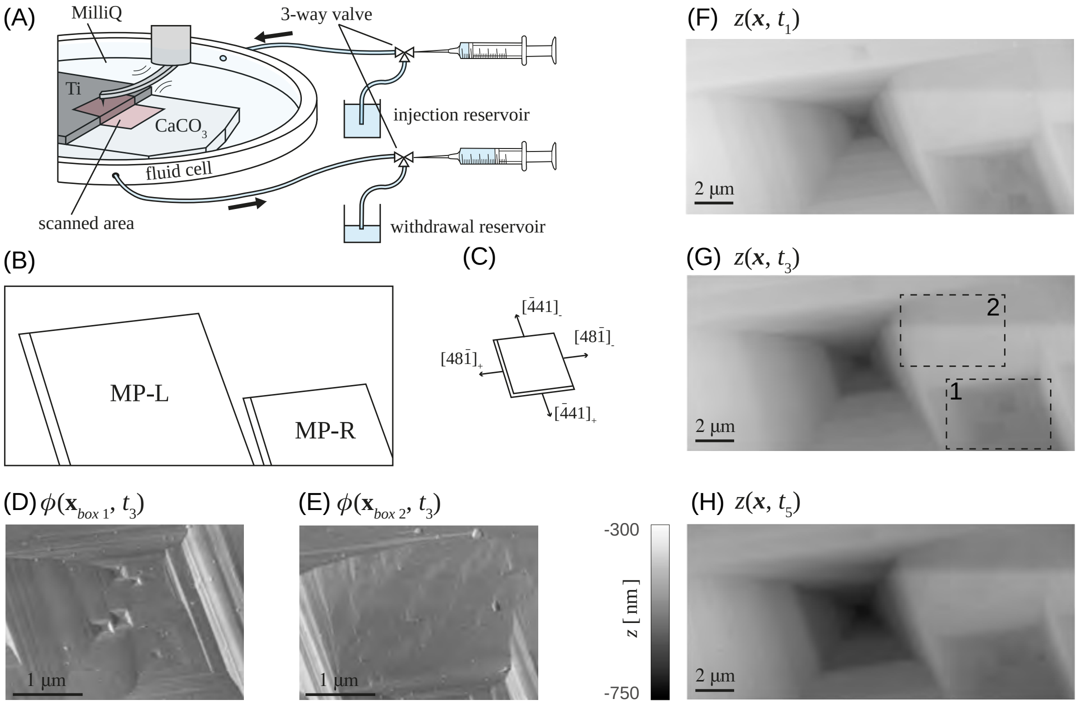

Figure 4.17: **(A)** Sketch of the experimental setup. Topography images at observation times **(F)** $t_1 = 17$ min, **(G)** $t_3 = 51$ min, and **(H)** $t_5 = 85$ min from the beginning of the experiment. **(B)** Representation of multilayer etch pits governing the surface pattern evolution and **(C)** their crystallographic orientation. **(D)** and **(E)** enlargements of the AFM phase shift signal, $\phi(\mathbf{x}, t_3)$, associated with $z(\mathbf{x}, t_3)$ within box 1 and 2 of **(G)**, respectively.

Fig. 4.18 depicts spatial distributions of $R(\mathbf{x}, t_j^*)$ ($t_j^* = t_i + \Delta t$, $j = i = 1, ..., 5$) evaluated from the acquired sequence of experimental topographic data though Eq. 2.2. Rate maps exhibit traits that are consistent with recent experimental findings by Fischer and Lüttge (2018) (see Section 2.1). Figs. 4.18.F-L and Figs. 4.18.M-Q depict values of $R(\mathbf{x}, t^*)$ evaluated along segments AA' and BB', respectively. These sections are oriented along the $[48\bar{1}]$ crystallographic direction (see Fig. 4.17.E) and their analysis supports the aforementioned different behavior of the two etch pits. A clear sequence of stepwaves emanating from the center of MP-L (identified as a red dot in Figs.4.18.F-L) is observed along section AA'. At each time, stepwaves located farther away from the dislocation center correspond to older stepwaves, exhibiting lower peak heights as compared to younger stepwaves closer to the center of the etch pit. Three prominent trains of steps can be clearly distinguished at time t_1^* (Fig.4.18.F) (labeled as c-d, b-e, and a-f, where c, b, and a represent obtuse-oriented stepwaves, while d, e, and f correspond to acute-oriented stepwaves). As the reaction progresses (Fig.4.18.G), these stepwaves propagate across the surface and move away from the dislocation center. Acute stepwaves e and f gradually disappear (Fig.4.18.H-I and 4.18.L) as they merge with stepwaves originating from MP-R along the $[\bar{4}41]_-$ direction. Sections BB' display a primary stepwave (labeled as g in Figs. 4.18.M-Q), that propagates in the $[48\bar{1}]_-$ direction, together with the flat bottom of MP-R. Fluctuations of R observed at times t_1^*, t_2^*, and t_5^* are related to smaller stepwaves emanating from etch pits formed at the bottom of MP-R. No obtuse stepwaves corresponding to g are observed for MP-R. It is hypothesized that these are annihilated due to the spreading of MP-L in the $[48\bar{1}]_-$ direction.

4.3.2 | ANALYSIS AND STATISTICAL MODELING OF DISSOLUTION RATES

Evaluation and analysis of sample probability densities is critical to obtain information about the probability of occurrence of given ranges of rate values across the mineral surface. Unlike standard approaches grounded on the assessment of an average rate (i.e., a constant rate value), which is then considered as a unique value characterizing the bulk

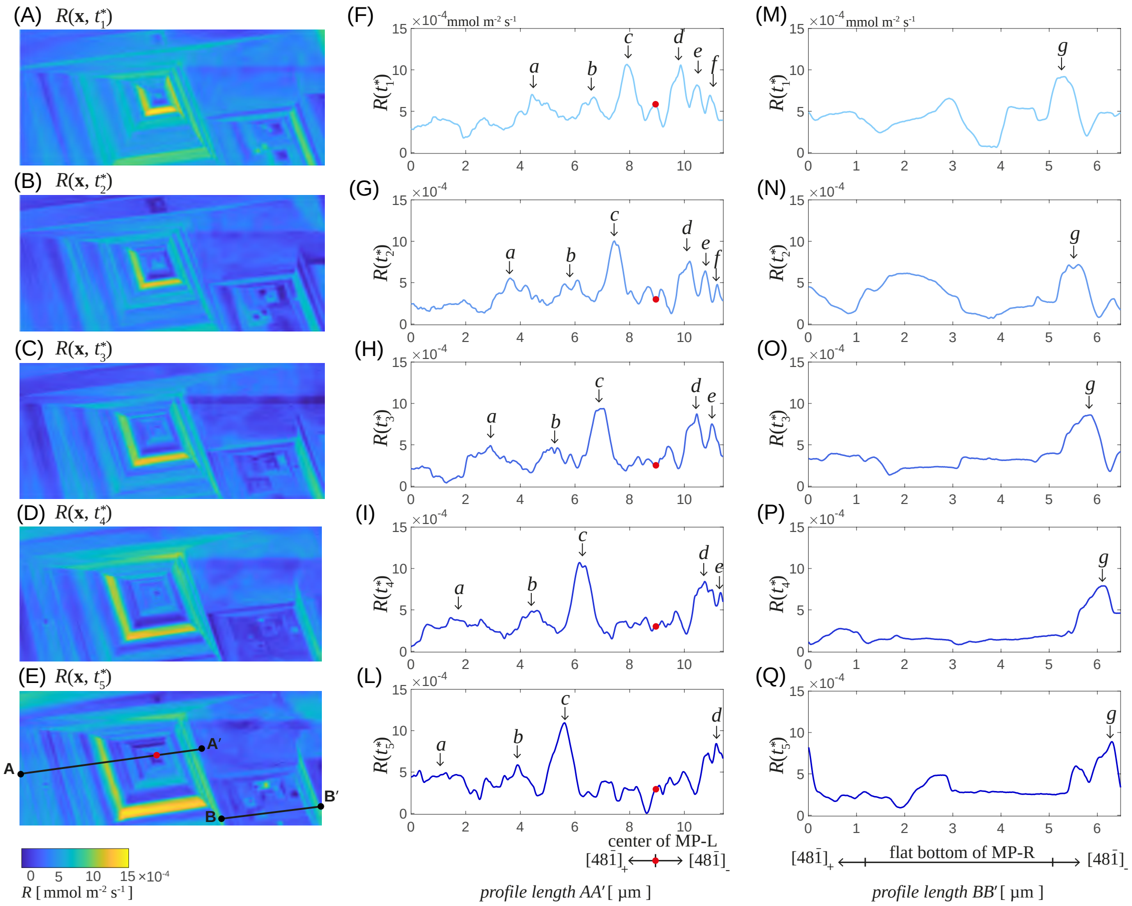

Figure 4.18: Spatial heterogeneity of absolute dissolution rate, $R(\mathbf{x}, t^*)$, at **(A)** $t_1^* = 34$ min, **(B)** $t_2^* = 51$ min, **(C)** $t_3^* = 68$ min, **(D)** $t_4^* = 85$ min, and **(E)** $t_5^* = 102$ min. Reaction rate evaluated along profiles AA' **(F-L)** and BB' **(M-Q)** as indicated in **(E)**.

surface retreat, relying on a stochastic approach enables one to capture the richness of information governing the complexity underlying the evolution of mineral surface reactivity (Lüttge et al., 2013b; Fischer et al., 2012). Fig. 4.19 depicts sample PDFs associated with spatial maps of R detected at five observation times. The overall width of the support of measured values of R (in terms of their order of magnitude) is consistent with data reported for similar chemical conditions. The red circle in Figs. 4.19.A-E denotes the average dissolution rate measured by Arvidson et al. (2003) through ex-situ VSI observations of a calcite crystal exposed to a solution of deionized water and Na_2CO_3 at pH $= 8.8$ at

$T = 25$°C in a stirred reactor. The yellow circle in Figs. 4.19.A-E corresponds the value of the rate evaluated by Bouissonnié et al. (2018) upon relying on ex-situ VSI measurements of the topography of a polished calcite crystal subject to dissolution in a mixed flow reactor ($Q = 0.24\,\mathrm{mL\,min^{-1}}$) in a solution of MilliQ water, $NaHCO_3$, and NaCL at $p\mathrm{H} = 8$. Sample distributions exhibit pronounced multimodal traits leading to long right tails. Similar features have been observed for calcite and other carbonate minerals through AFM analyses (Emmanuel, 2014; Siena et al., 2021, 2023) as well as at larger spatial scales typical of VSI (Bibi et al., 2018; Bollermann and Fischer, 2020), DHM (Brand et al., 2017), or X-ray microtomography (Noiriel et al., 2018). Modeling results based on Eq. 3.31 juxtaposed to sample PDFs are depicted in Figs.4.19.A-E, documenting the remarkable ability of our theoretical framework to capture the essential elements of the experimental observations. Vertical orange and blue lines depict estimated component means, $\hat{\mu}_m$, along with corresponding intervals of semi-width equal to one standard deviation, $\hat{\sigma}_m$ ($m = A, B$). These identify ranges for low and high rates that are fully consistent with those obtained by Brand et al. (2017), which are associated with identical chemical conditions to those considered here, even as acquired at a larger spatial scale via DHM observations of a polished calcite surface (green rectangles in Figs. 4.19.A-E).

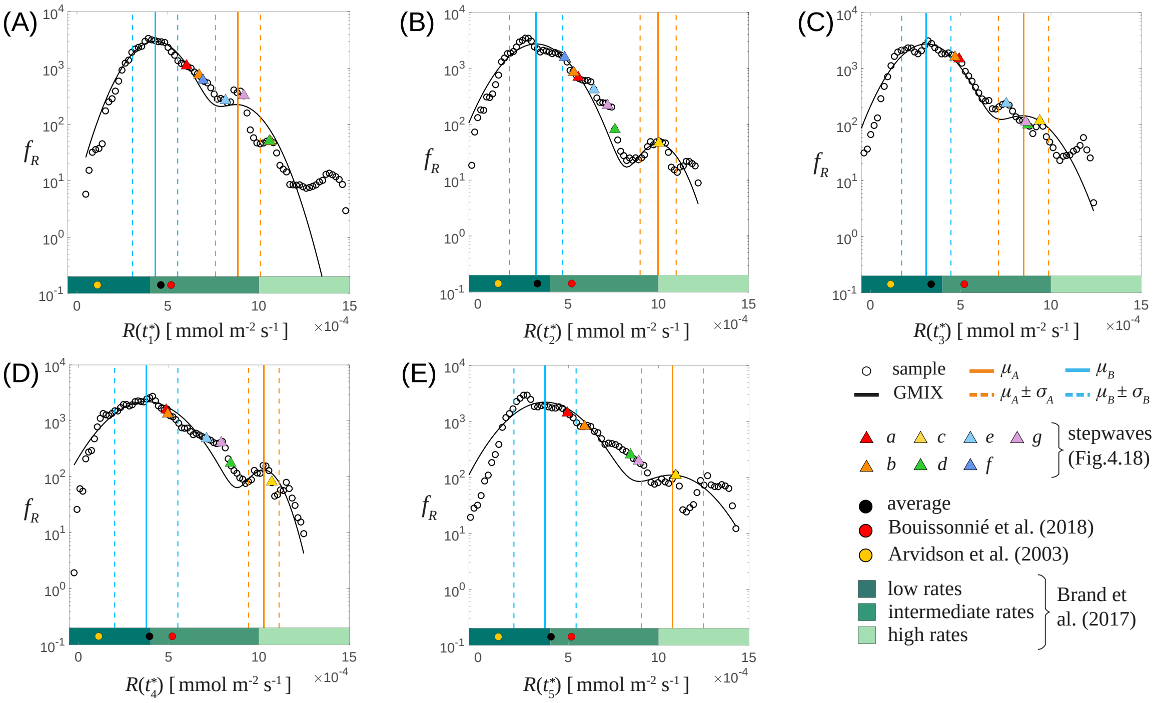

Figure 4.19: Sample PDFs and GMIX model (Eq. 3.31) of dissolution rates at five observation times. Vertical orange and blue lines depict estimated component means, $\hat{\mu}_m$, along with the intervals of semi-width corresponding to one standard deviation, $\hat{\sigma}_m$ ($m = A, B$). Average value (black circle) and literature data (red and yellow circles) are also depicted. Green rectangles delineate intervals of low, medium, and high rates as evaluated by Brand et al. (2017). Triangles depicted on **(A-E)** correspond to peak values attained by R evaluated along profiles AA' and BB' and illustrated in Figs. 4.18.F-Q.

4.4 | Results of experimental *Setting 3* - Unmasked samples

In the following, we discuss experimental and modeling outcomes obtained from the investigation of the evolution of the surface of a (non-masked) calcite sample. A portion of the surface of extent $6 \times 6\,\mu\text{m}^2$ is imaged across a 512×512 grid (pixel size $dl = 11.7\,\text{nm}$) in tapping mode. The acquisition frequency is set to $f_a = 1.28\,\text{Hz}$, yielding a temporal resolution $\Delta t = 6.6\,\text{min}$. Image alignment in the x-y plane at subsequent times is performed upon considering as *fiducial points* the centroids of precipitates spontaneously forming on the surface during the reaction (see black arrow depicted in Fig. 4.20.A). Such image registration procedure restricts our observation window to $5.3 \times 5.3\ \mu\text{m}^2$ (corresponding to 450×450 cells). Modeling of the statistical behavior of dissolution rate fields, R', and of associated spatial increments, ΔR, relies upon the Generalized sub-Gaussian mixture (GSG-MIX) model introduced in Sections 3.4 and 3.4.1.

4.4.1 | EVOLUTION OF CALCITE DISSOLUTION PATTERNS

Fig. 4.20 depicts spatial maps of topography, $z'(\mathbf{x}, t_i)$, measured at various times, t_i ($i = 1, ..., 7$). At all instants sampled during our experiment, we observe an evolution of the surface pattern driven by the mechanisms described in Section 2.1, i.e., 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 (Fig. 4.20.A). A multilayer etch pit (MP1) enters our observation window at time t_2 from the left boundary (Fig. 4.20.B) and grows laterally at t_3 (Fig. 4.20.C). 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 (Fig. 4.20.D). This pit considerably grows laterally at t_5 and t_6 (Figs. 4.20.E-F). 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. Our hypothesis is that the dislocation line originating MP2 is no more oriented along the z direction starting from t_6. 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 at the bottom right of our field of view. Even as we only have partial observations concerning the evolution of MP1 and MP3, our temporal observation window encompasses the entire evolution of MP2.

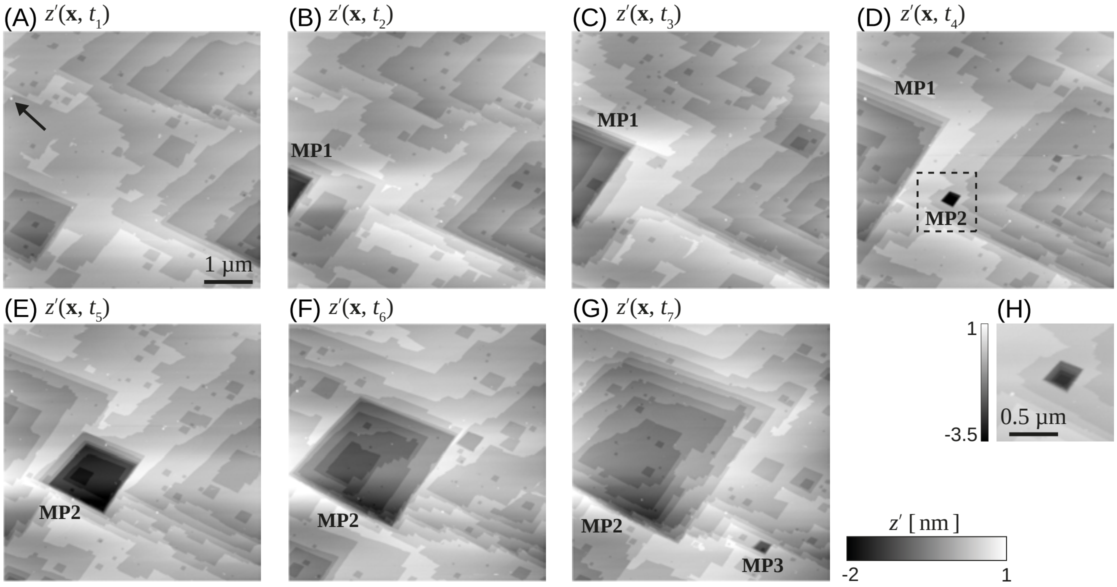

Figure 4.20: Spatial distributions of fluctuations (about their mean) of calcite topography, $z'(\mathbf{x}, t_i)$, acquired at times **(A)** $t_1 = 6.6\,\mathrm{min}$, **(B)** $t_2 = 13.2\,\mathrm{min}$, **(C)** $t_3 = 19.8\,\mathrm{min}$, **(D)** $t_4 = 26.4\,\mathrm{min}$, **(E)** $t_5 = 33.0\,\mathrm{min}$, **(F)** $t_6 = 39.6\,\mathrm{min}$, and **(G)** $t_7 = 46.2\,\mathrm{min}$ 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)**.

This documented evolution of the surface pattern leads to highly heterogeneous spatial distributions of reaction rates, $R'(\mathbf{x}, t_j^*)$. These are evaluated through Eq. 2.4 at times $t_j^* = t_i + \Delta t$ $(j = i = 1, ..., 6)$ and are depicted in Figs. 4.21.A-F. The lateral expansion of MP1, MP2, and MP3 yields dissolution stepwaves that are schematically delineated in orange, blue and green in Fig. 4.21.G, respectively. These regions correspond to portions of the surface that experience enhanced material fluxes. From a qualitative standpoint, reaction rates within such regions are seen to decrease as time proceeds 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 comprised between the edges of MP2 at subsequent times (Figs. 4.21.C-F) is associated with a decrease of the strength of R'. Such a temporal evolution of R' is consistent with a description of the dissolution process as driven by stepwave emanation (Lasaga and Lüttge, 2001, 2003; Fischer and Lüttge, 2018). Otherwise, lower values of rate are attained at spatial locations where the reaction is governed by shallow etch pits and step dynamics.

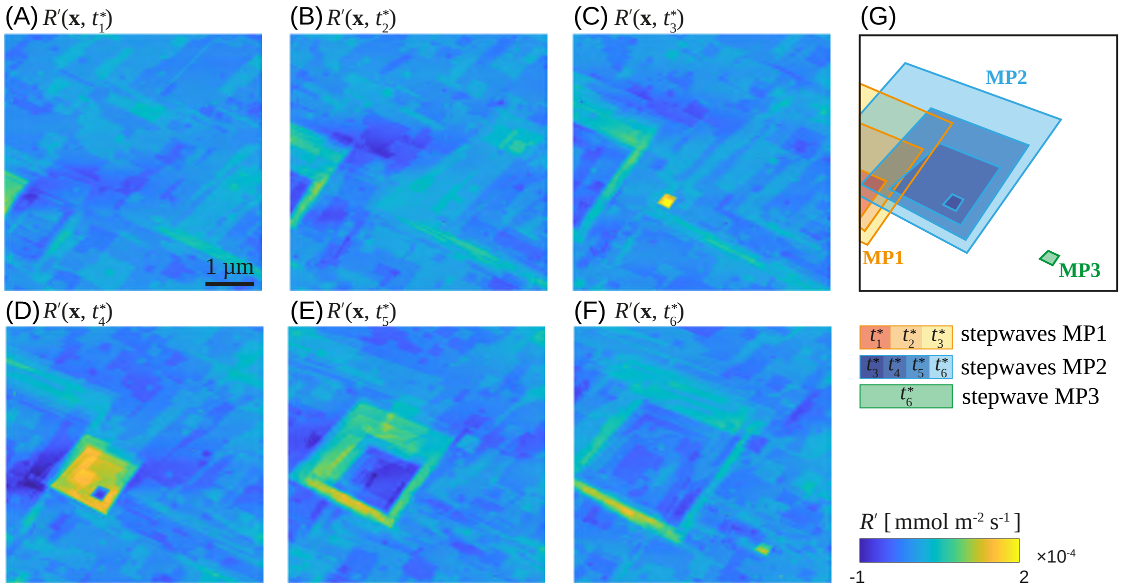

Figure 4.21: Spatial maps of $R'(\mathbf{x}, t_j^*)$ evaluated with Eq. 2.4 from AFM topography measurements depicted in Fig. 4.20 at times **(A)** $t_1^* = 13.2$ min, **(B)** $t_2^* = 19.8$ min, **(C)** $t_3^* = 26.4$ min, **(D)** $t_4^* = 33.0$ min, **(E)** $t_5^* = 39.6$ min, and **(F)** $t_6^* = 46.2$ min; **(G)** schematic depiction of dissolution stepwaves emanated from MP1, MP2, and MP3.

4.4.2 | ANALYSIS AND STATISTICAL MODELING OF DISSOLUTION RATES

As a consequence of the presence of diverse regions, each characterized by a given dissolution mechanism, sample probability densities of $R'(\mathbf{x}, t_j^*)$ exhibit marked multimodal traits.

As an illustrative example, Fig. 4.22.B shows the sample probability density (PDF) associated with the spatial field $R'(\mathbf{x}, t_4^*)$ depicted in Fig. 4.21.D. 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, i.e., the dynamic evolution of shallow etch pits and the retreat of steps. Otherwise, the latter is associated with high material fluxes due to stepwaves radiating from MP2. A similar behavior is observed at all times (see Figs. G.1-G.5.B in Appendix G). Similar to the other experimental settings here considered, modes associated with high and low values of rate are hereafter denoted as A and B, respectively. Fig. 4.22.C depicts sample probability densities of incremental values evaluated at three lags, $s = 10, 40, 80$. 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 locations, one of which is associated with region A, the other one being located in region B. Sample PDFs of ΔR are characterized by a pronounced scaling tendency with separation distance, in the sense that they change depending on the lag at which increments are evaluated. All of the above-mentioned traits of the PDFs of R' and ΔR are consistent with a description of the behavior of the spatial field of the dissolution rate grounded on the GSG-MIX modeling framework introduced in Section 3.4 and 3.4.1. We rely on a lognormal distributional form of the subordinator associated with each component of the mixture (see Appendix E) for our analyses.

We rest on the parameter estimation scheme described in Section 3.6.3 and obtain a classification of the observed R' fields in two regions together with the estimate of GSG-MIX model parameters. Initialization of the algorithm relies upon the identification of (i) terrace regions and (ii) dissolution stepwaves emanated by MP1, MP2, and MP3. Fig. 4.22.A shows the indicator field $I(\mathbf{x}, t_4^*)$ resulting from the classification of $R'(\mathbf{x}, t_4^*)$. Here, the portion of the surface involved in stepwave radiation is assigned to region A. Results of similar quality are obtained at all times $t_1^* - t_6^*$ (see Figs. G.1-G.5.A in Appendix G). Estimates of the 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 Fig. 4.22.D. 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 contexts of application (see, e.g., Mazzoli and Favoni, 2012, Lolo et al., 2023, and 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 Figs. G.1-G.5.D in Appendix G). This behavior arises from the geometrical pattern characterizing the indicator field and is strictly 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 Fig. 4.22.E for $\gamma_R(t_5^*)$ and Figs. G.1-G.5.E in Appendix G for all of the other observation times). Analysis and interpretation of these types of correlation behaviors could rely, e.g., upon nested models entailing a hole-effect component (see Section 4.1.3). 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 Eq. E.1 and Eq. E.5 are juxtaposed to sample PDFs of R' and ΔR in Figs. 4.22.B and C, respectively. Visual inspection of these results documents a remarkable agreement between modeled and sample statistics. For completeness, the analytical expression of the GSG-MIX variogram (Eq. E.9) is juxtaposed to sample data in Fig. 4.22.E. Results of similar quality are obtained at all times (see Figs. G.1-G.5.B,C, and E).

Fig. 4.23 depicts temporal trends of estimated model parameters embedded in Eq. E.1 (i.e., $\hat{p}$, $\hat{\mu}_m$, $\hat{\sigma}^2_{G_m}$, and $\hat{\alpha}_m$, with $m = A, B$), along with the temporal evolution of the statistical moments of the mixture (i.e., variance, σ^2_R, skewness Sk_R, and of kurtosis, κ_R). Being strictly linked to the dissolution mechanisms, model parameters can be viewed as

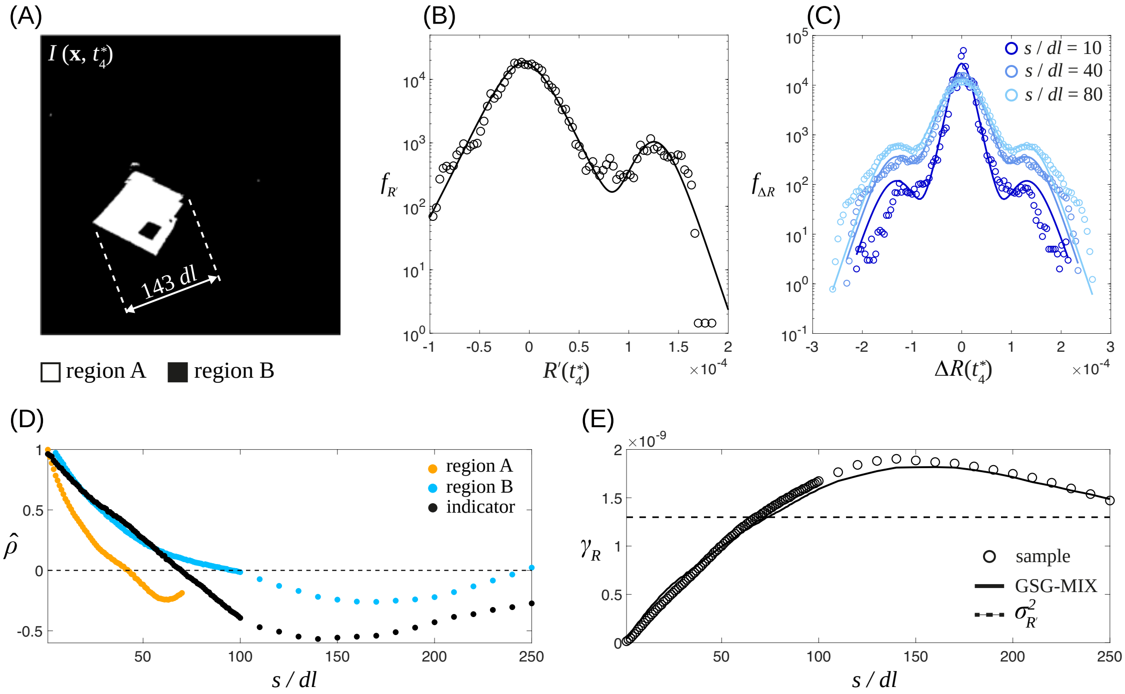

Figure 4.22: Application of the GSG-MIX modeling framework to the spatial field $R'(\mathbf{x}, t^*_4)$. **(A)** indicator field, $I(\mathbf{x}, t^*_4)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m-th$ mode of the mixture, $\hat{\rho}_{G_m}$ ($m = A, B$), and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9.

aggregate indicators 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 $\hat{p}$ (Fig. 4.23.A) displays a certain degree of symmetry (with respect to a horizontal axis) to the behavior of $\hat{\mu}_A$, a peak in the former corresponding to a trough in the latter (Fig. 4.23.B). At $t^*_1 - t^*_2$, $\hat{\mu}_A$ is almost constant. Similarly, the estimated proportion coefficient $\hat{p}$ is virtually constant within this temporal window and is nearly zero. This behavior reflects the ending of the action of the dislocation originating MP1 as parallel to the z direction. The maximum 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, i.e., 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 (Lasaga and Lüttge, 2001, 2003; Schabernack and Fischer, 2024). 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, and then drops at t_6^*. The latter time corresponds to the ending of stepwave radiation from MP2. Hence, at t_6^* only a part of the area comprised between pit edges at subsequent times is attributed to the high-rate component of the mixture (see Fig.G.5.A). 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 Fig. 4.23.C. These results suggest that the degree of non-Gaussianity of the statistics within each region is (i) mild ($\hat{\alpha}_m \sim 1.7 - 1.8$), ($ii$) similar for the two regions, and (iii) does not change significantly with time. We recall that the variance ($\sigma^2_{R_m}$) of each component of the mixture is jointly governed by parameters α_m and $\sigma^2_{G_m}$. Hence, temporal variations of $\sigma^2_{R_m}$ are essentially controlled by changes of the variance of the underlying Gaussian fields, $\sigma^2_{G_m}$. The estimated variances of each $m - th$ mode and of G_m are depicted against time as dotted and dashed lines, respectively, in Fig. 4.23.D. Whereas $\hat{\sigma}^2_{R_B}$ remains constant at all times, $\hat{\sigma}^2_{R_A}$ exhibits pronounced variations. As far as MP1 is the only etch pit driving the reaction (i.e., at t_1^* and t_2^*), significant changes of $\hat{\sigma}^2_{R_A}$ are not observed. Similarly, at t_4^*, t_5^*, and t_6^*, MP2 is governing the dissolution process and $\hat{\sigma}^2_{R_A}$ is almost constant. Such temporal trend reflects that the degree of spatial heterogeneity within region A at these times is ascribed to the nucleation of shallow etch pits on terraces that are freshly exposed to the solution by the retreat of steps originating from MP2. This interpretation is supported by the small difference observed at $t_4^* - t_6^*$ between $\hat{\sigma}^2_{R_A}$ and $\hat{\sigma}^2_{R_B}$. Indeed, the variance of region B is also governed by the formation of shallow etch pits and is almost constant at all times. Otherwise, $\hat{\sigma}^2_{R_A}$ is considerably higher at t_4^* than at all of the other times. At this particular time, the region that is identified as contributing to high rates is seen to comprise two sub-regions. The presence of these could be observed due to the temporal resolution of our experimental setting. Values of $R'(\mathbf{x}, t_3^*)$ are assessed starting from topographies $z'(\mathbf{x}, t_3)$ and $z'(\mathbf{x}, t_4)$. As no evidence of the presence of MP2 at t_3 is observed, 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 topographic 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 $R'(\mathbf{x}, t_3^*)$ 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 $(t_3 - t_4)$. Fig. 4.20.H provides an enlargement of the topography measured at t_4 (corresponding to the dashed box in Fig. 4.20.D) and illustrates these trains of steps.

Sample statistical moments of second, third and fourth order evaluated across the entire field are depicted against time in Figs. 4.23.E-G, respectively. Corresponding analytical expressions evaluated through Eq. E.2, E.3, and E.4 are also included as red asterisks and document a remarkable level of agreement with experimental results. The statistical moments are controlled by (*i*) the square difference of the means, $(\mu_A - \mu_B)^2$, (*ii*) the difference of the variances, $(\sigma^2_{R_A} - \sigma^2_{R_B})$, and of the fourth order moments, $\left(\langle {R'}^4{}_A \rangle - \langle {R'}^4{}_B \rangle\right)$ of the two components of the mixture, together with (*iv*) the proportion coefficient, p. As shallow etch pit dynamics drive heterogeneity within each region, statistical moments of second and fourth order of R'_m $(m = A, B)$ attain similar values at all times (see Appendix H). Therefore, the temporal evolution of the statistical moments characterizing the PDF of R'

is essentially controlled by the behavior of $(\mu_A - \mu_B)^2$ and p. Fig. 4.23.E illustrates the temporal trend of the variance of R', σ_R^2. Overall, one can observe that the temporal behavior of σ_R^2 mirrors the trend exhibited by $\hat{\mu}_A$ with a temporal shift corresponding to Δt. At times $t_1^* - t_3^*$, σ_R^2 approximately coincides with $\sigma_{R_B}^2$, as the relative proportion of region A across the overall domain is approximately zero (Fig. 4.23.A). A sudden increase of σ_R^2 is then observed at t_4^*, followed by a decreasing trend at $t_5^* - t_6^*$. The latter reflects the homogenization of regions A and B that takes place as a consequence of the end of the action of the dislocation originating MP2 along the vertical direction. The temporal behavior of σ_R^2 documents a delay between the actual formation of the MPs driving the reaction and the time at which this propagates the statistical behavior of the entire field. Indeed, even as the difference of the means is maximized at t_3^* (Fig. 4.23.B), the small lateral extent of the region involved in stepwave emanation at the initial phase of MP2 evolution somehow tempers its effect. For the same reason, the stepwave arising at MP3 at t_7 does not affect σ_R^2 at t_6^*. The skewness of R', Sk_R, is depicted against time in Fig. 4.23.F. A decreasing trend is documented at the beginning, corresponding to the end of the effect of the action of the dislocation emanating MP1. The PDF of R' is almost symmetric at time t_2^* (see also Fig. G.2.B in Appendix G), consistent with the nearly zero value of $\hat{p} \approx 0$ observed at this instant. The strength of the asymmetry of the distribution then increases from t_3^* (i.e., from the beginning of the effect of MP2). Following Eq. E.3 and similar to what we observe for σ_R^2, the third order central moment R' (i.e., the numerator of Eq. E.3) is jointly driven by $(\mu_A - \mu_B)^2$ and p when $(\sigma_{R_A}^2 - \sigma_{R_B}^2) \to 0$. Therefore, since the emission of stepwaves at MP2 affects both terms similarly, Sk_R exhibits oscillations around a constant value slightly greater than 1. Fig. 4.23.G depicts the kurtosis of R', κ_R, as a function of time. The behavior of this moment closely resembles the pattern observed in the mean of region A. In agreement with the behavior observed for $\hat{\alpha}_m$ and $\hat{\sigma}_{G_m}^2$ (see Figs. 4.23.C-D), $\left(\langle R'^4{}_A\rangle - \langle R'^4{}_B\rangle\right) \to 0$ in Eq. E.4, similarly to $(\sigma_{R_A}^2 - \sigma_{R_B}^2) \to 0$. Hence, κ_R is controlled by (i) the fourth order central moment of R_B and (ii) the joint action of p and $(\mu_A - \mu_B)^2$. In this case, $\langle {R'_B}^4\rangle$ is almost constant in time. Therefore, changes in κ_R are controlled by variations of $(\mu_A - \mu_B)^2$ and p. In this specific case, the difference between the means of the two components counteracts the effect of p and drives the temporal behavior of κ_R.

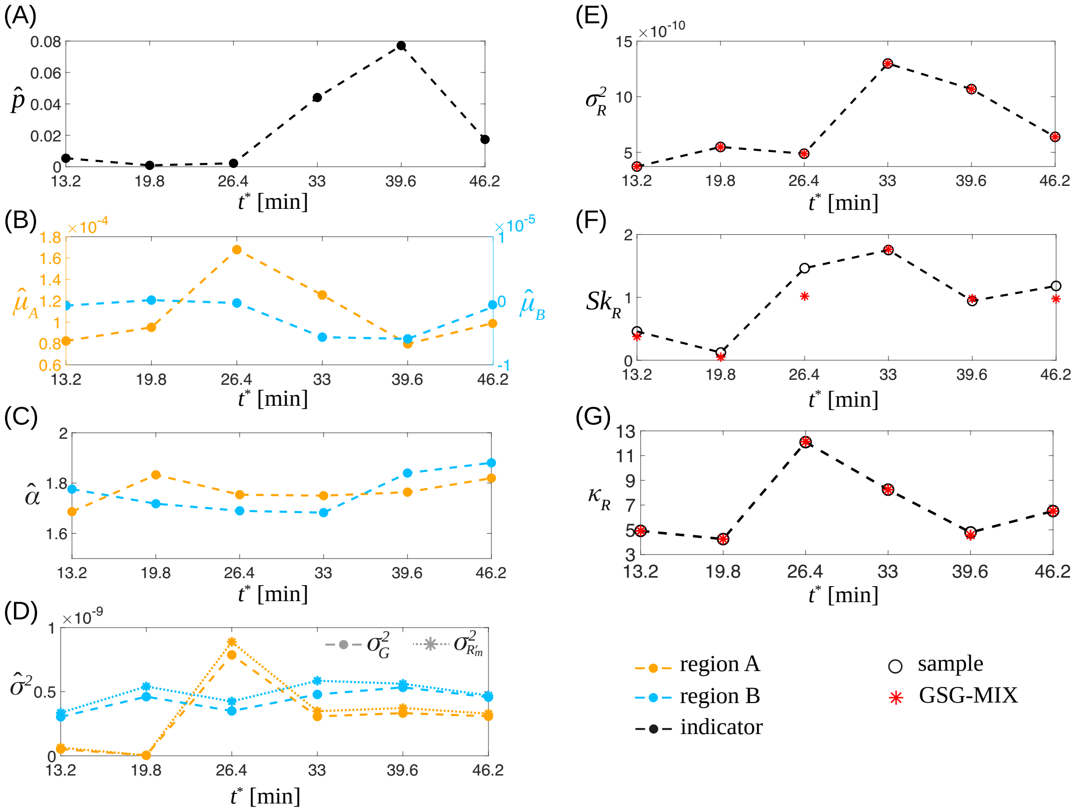

Figure 4.23: Temporal trends of estimated GSG-MIX parameters embedded in Eq. E.1, i.e., **(A)** $\hat{p}$, **(B)** $\hat{\mu}_m$, **(C)** $\hat{\alpha}_m$, and **(D)** $\hat{\sigma}_{G_m}$ $(m = A, B)$, and of **(E)** second, **(F)** third, and **(G)** fourth order statistical moments of the mixture. Analytical expressions for the latter evaluated according to Eqs. E.2, E.3 and E.4 are also included.

5 Conclusions

This PhD dissertation contributes to (*a*) strengthening our ability to directly observe reaction rates through original nano-/microscale imaging experiments and (*b*) providing an interpretation of the observed rates through rigorous stochastic approaches. The work is grounded on an original combination of experimental and theoretical/modeling advancements. Major conclusions stemming from the work are illustrated in the following. These are structured with emphasis on the experimental as well as on the theoretical/modeling components.

Experimental protocols are developed to enable direct investigation of mineral reactive kinetics associated with a high fidelity reproduction of diffusion-limited and surface-controlled conditions. We target dissolution reactions of non-polished calcite {104} under far-from-equilibrium conditions. We design original platforms for sample preparation enabling one to acquire absolute topographic measurements (referenced to an engineered non-reacted surface). These, in turn, allow documenting the spatial heterogeneity of the surface of a mineral in contact with a flowing fluid and subject to dissolution. Imaging is grounded on high resolution Atomic Force Microscopy (AFM). The ensuing high-quality big datasets can be readily employed to obtain spatial fields of corresponding absolute reaction rate with nanoscale resolution. Coupling the high horizontal resolution allowed by AFM with the tailored experimental procedure for the application of an inert metallic mask on the crystal surface offers an unprecedent opportunity for direct observation and stochastic analyses of absolute material fluxes and of the details of their spatial heterogeneity.

We design different setups for fluid flow regulation during *in-situ* experiments. *Setting*

1 (Section 2.3.1) is designed to mimic typical processes associated with (extremely) low velocity/stagnant regions in porous systems where reactions are dominated by diffusion. Otherwise, flow-through conditions are reproduced through *Settings 2* (Section 2.3.2) and *3* (Section 2.3.3), thus empowering us with the ability to examine surface-controlled processes. The designed experimental settings include the following key elements.

- *Setting 1* corresponds to promoting the formation of a diffusive boundary layer. The selected temporal observation window allows following transitions from a surface- to a diffusion-controlled reaction. The observed dissolution pattern documents (*i*) a change of the main mechanism driving the reaction from etch-pit nucleation to step retreat following local increase of saturation in the boundary layer and (*ii*) a temporal decrease in the surface reactivity.

- *Setting 2* involves the use of two synchronized syringes connected to the fluid cell. Flow-through conditions are established by intermittent refreshment of the solution between each AFM image, whereas the fluid stagnates in the cell during scanning. The contact time between the solution and the sample is linked to the AFM acquisition parameters, i.e., the number of grid elements, n, the scanning frequency, f_a, being selected as a trade-off between resolution and quality of images. Surface topography images acquired with this setting document a dynamic evolution of mono- and multilayer etch pits on the surface that is consistent with highly undersaturated conditions. Rounding of obtuse-obtuse pit edges is also observed, thus suggesting that (slight) variations in dissolved calcium saturation are still present.

- *Setting 3* is grounded on the use of a high precision syringe pump enabling scanning the system in the presence of a continuous flow in the cell. The fluid/solid contact time is independent of acquisition parameters as is determined solely by the flow rate. Otherwise, the force field associated with the flow may induce disturbances that result in a degradation of the quality of AFM images. The flow rate is selected upon resting on (*i*) analysis of qualitative aspects and target experimental quantities associated with calcite dissolution pattern in highly unsaturated conditions and (*ii*) assessment of the quality loss due to fluid flow. Temporal series of topography images of calcite surface are acquired for various values of the Reynolds number, Re. The latter is defined on the basis of the flow rate and of the geometry of the fluid cell. Relying upon such dimensionless quantity enables one to generalize results and replicate conditions in other experimental settings. Consistency of the dissolution pattern with far-from-equilibrium conditions and constant surface reactivity are observed for $Re > 0.60$. The quality loss resulting from the presence of a fluid flow is assessed upon evaluation of the distortion of the topography of a reference calibration grating imaged with identical acquisition parameters. While quality loss (slightly) increases with Re, we conclude that operating at $Re = 0.60 - 0.70$ ensures that dissolution takes place under stable chemical conditions, while preserving good quality of the images.

We model the heterogeneous field of dissolution rate as a random spatial field. We define the rate as the sum of an average value, $\langle R \rangle$, and a random fluctuation about the mean, R'. The former represents an average material flux across the whole surface, whereas the latter is informative about the spatial variability of the rate and is subject to our stochastic analysis. Modeling of the statistical behavior of dissolution rate field obtained through the designed experimental settings is grounded on robust original theoretical models. These provide a unified framework of analysis for the probability distributions of R' and its spatial increments (ΔR) evaluated at various separation distances (or lags). In this sense, our theoretical frameworks enable one to infer distributions of quantities of interest through a joint analysis of data about values of the target quantity and its (spatial) increments. We then ensure consistency between these two sets of observations.

We rely on the well-established Generalized sub-Gaussian (GSG) formulation by Riva et al. (2015a). This flexible and versatile model has been documented to accurately capture non-Gaussian behaviors of a unimodal variable and of its associated spatial increments at other spatial scales (Riva et al., 2015a; Guadagnini et al., 2018; Siena et al., 2019, 2020; Li et al., 2022). Typical non-Gaussian traits entail (*i*) sharp peaks and (*ii*) heavy tails of sample PDFs of the target variable, together with (*iii*) scaling tendencies of associated spatial increments.

Otherwise, multimodal behaviors are observed in sample probability density functions (PDFs) of R' (Fischer et al., 2015; Fischer and Lüttge, 2017; Brand et al., 2017) and ΔR. Key elements observed for such distributions include (*i*) a slight to moderate asymmetry in the distribution of R, resulting from the presence of multiple peaks, and (*ii*) the occurrence of a dominant peak together with multiple secondary peaks in the distribution of ΔR. The relative importance of these peaks tends to vary with the lag at which increments of R are taken across the system. This gives rise to a stark and observable scaling behavior of the PDF of ΔR. We start by focusing on the particular case of a bimodal Gaussian mixture, whose modes are identified through an indicator random field. These traits of sample PDFs arise from a unified description of the otherwise composite nature of a field that encompasses various regions, each being associated with a different degree of internal heterogeneity. The presence of such distinct regions in spatial fields of dissolution rates is related to the coexistence and/or competition of different kinetic mechanisms underpinning dissolution reactions. We provide theoretical frameworks enabling one to capture these statistical traits. When considering a bimodal Gaussian mixture model (GMIX), we extend mathematical formulations by Lu and Zhang (2002) to embed the stochastic nature and behavior of spatial increments. Then, we further extend our modeling framework to consider bimodal mixtures associated with Generalized sub-Gaussian modes (GSG-MIX). The latter is a general formulation that includes the GMIX and the GSG as particular cases. It enables one to capture bimodal traits encompassing non-Gaussian patterns exhibited by each component of the mixture. The implementation of the GSG and GSG-MIX models adopted in this thesis relies on a log-normal subordinator. The latter has been tested for the interpretation of the spatial statistics of a wide range of data (e.g., Riva et al., 2015a, Siena et al., 2019, Li et al., 2022). It is emphasized that the theoretical framework illustrated in this thesis includes the possibility of selecting a general form of the subordinator. The application of alternative formulations of the GSG and GSG-MIX models on dissolution rates will be the subject of future investigations.

We propose two general procedures to estimate model parameters, which include partitioning the domain into the two components of the mixture, A and B. A first methodology relies on a well-established Expectation-Maximization (EM) algorithm. The latter estimation scheme is otherwise grounded on a Bayesian classification approach. The robustness of the proposed methodologies is assessed through extensive tests on a collection of synthetically generated random fields. Application of the ensuing workflows enable one to characterize the statistical traits of each component of the PDF through mode deconvolution. Note that the mathematical formulations employed (GSG) or developed (GMIX and GSG-MIX) in the context of this PhD thesis can be readily applied to interpret other key environmental variables at different spatial scales and is not restricted to nano-/microscale spatial observation supports/windows. As an example, the GMIX yields accurate characterization of sample PDFs and statistical moments associated with air-permeability data obtained by Tidwell and Wilson (1999) on a centimeter scale block of tuff (Siena et al., 2023).

Experimental topographic data obtained upon relying on the designed settings can be promptly employed for a stochastic analysis of spatial maps of reaction rates through the modeling approaches described above.

Setting 1 yields quantitative spatial distributions of fluctuations of dissolution rates, R', at various observation times elapsed from the renewal of the fluid (deionized water) in contact with the mineral surface. The evolution of surface patterns evidenced by AFM images is consistent with the temporal increase of the solution saturation in a boundary layer at the fluid/solid interface. The spreading rate evaluated on monolayer etch pits is monotonically decreasing in time. Otherwise, the multilayer etch pit expansion rate is highest at an intermediate time during the experiment and then displays a steep decrease. A similar trend is also documented by (i) the shape parameter and (ii) the scale parameter of the GSG model employed for the joint characterization of the statistical behavior of R' and ΔR. We provide qualitative and quantitative results about the relationship between the parameters of the GSG stochastic model and the dynamics of multilayer pits, documenting that the evolution of these structures significantly affects the statistical features of dissolution rates. Relying on the Kullback-Leibler Divergence metric, we find that the GSG model generally shows a higher fidelity to the interpretation of the sample probability distribution of R' than the Generalized Extreme Values (GEV) model, which has been previously used (e.g., Brand et al., 2017). We emphasize that the GSG formulation offers the additional advantage of fully embedding the features of the probability distributions of both R' and ΔR in a unified and consistent manner, an element which is not included in the above mentioned interpretations based on the GEV model. The stochastic characterization of incremental data yields critical information about the spatial correlation of the rate field through the correlation function (ρ_G) associated with the GSG model. Among the various theoretical models analyzed for the interpretation of ρ_G, we find that a nested structure with a short-range and a long-range correlation component (see Eq. 4.4) is consistently ranked as best according to rigorous model identification criteria. The temporal behavior of these characteristic length scales appears to be linked to the evolutionary dynamics of step retreat/monolayer pits and multilayer pit structures documented in the experiments. The importance of characterizing the spatial correlation of key variables driving mineral dissolution processes was highlighted by Pollet-Villard et al. (2016). These authors develop a numerical model to describe dissolution and ground model calibration on the comparison between sample variograms evaluated on experimental data of surface topography and its numerically based counterpart. Our results reveal the impact that the diverse dissolution patterns can have on the correlation structure of reaction rates. This information can potentially lead to the development of future flexible numerical models, which can imbue us with the ability of taking into account multiple length scales resulting from the occurrence of diverse reaction mechanisms.

Upon relying on *Setting 2*, we collect spatial maps of R' comprising two distinct regions associated with high and low values of dissolution rate, respectively. The presence of the former region arises from the horizontal spreading of a multilayer etch pit, whereas the latter is associated with the dynamic evolution of shallow etch pits on crystal terraces.

Our homogenized view of the otherwise composite nature of these fields leads to a clear bimodal character of sample PDFs. The latter is amenable to an interpretation based on a Gaussian mixture model composed by two modes, hereafter termed A or B. These are respectively associated with high and low material fluxes taking place across the mineral surface. Our theoretical framework also enables us to capture the behavior of sample statistical moments associated with increments (as rendered through correlation function, ρ_R, and excess kurtosis, $E\kappa_{\Delta R}$) at different lags, although some discrepancies between sample and modeled values are observed for a specific range of separation distances. Here, we notice that sample PDFs of (spatial) increments of reaction rates evaluated for component A appear to display slight deviations from a Gaussian behavior. These deviations could also be at the basis of the imperfect agreement observed between sample and theoretical values of $E\kappa_{\Delta R}$. The temporal behavior of the GMIX model parameters is seen to be closely related to the evolution of the observed dissolution patterns.

Setting 3 is employed to obtain spatial maps of (a) absolute dissolution rates and (b) fluctuations of R about the mean. The constant saturation conditions guaranteed by continuous flow imposed throughout the duration of the experiment enable us to observe pulsating behaviors associated with emanation of stepwaves from etch pits. Such a pulsating nature of the dissolution reaction has been recently observed by Fischer and Lüttge (2018) for the dissolution of calcite and Zinc Oxyde at a higher spatial scale typical of VSI. Recent findings by Schabernack and Fischer (2024) attribute the formation of dissolution pulses to the inherent variability of the surface reactivity arising from the nano-topography at the lattice level.

We rely on *Setting 3* to acquire spatial maps of the topography of a portion of the crystal surface encompassing the reactive mineral and the inert layer. Rescaling reactive data to the latter enables one to obtain spatial maps of absolute reaction rates at the high spatial resolution offered by AFM imaging. Associated sample frequency distributions are characterized by an overall width of the corresponding support that aligns with data from VSI and DHM observations related to similar (Bouissonnié et al., 2018; Arvidson et al.,

2003) or identical (Brand et al., 2017) chemical conditions. Sample PDFs display multimodal traits arising from the coexistence of various dissolution mechanisms and/or different phases across the temporal evolution of the same kinetic process. We view these mechanisms as contributing to two major statistical modes, A and B, upon relying on the GMIX framework. Modeling results show a remarkable agreement between modeled and sample statistics.

Considering *Setting 3*, our temporal observation window captures the entire evolution of a stepwave radiated from an etch pit. Sample probability densities of R' are characterized by a stark bimodal tendency that mirrors stepwave evolution. Distributions of ΔR show (i) a central dominant peak coupled to lateral secondary peaks and (ii) a pronounced scaling behavior with the separation distance at which increments are evaluated. Each of the two modes observed exhibits 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 the aforementioned traits are embedded in the GSG-MIX formulation. Modeling results denote a striking agreement between sample and theoretical values of the PDFs and of the statistical moments of R' and ΔR. The strong link between dissolution mechanisms and GSG-MIX model parameters enables one to view the latter as aggregate indicators of the evolution of the surface processes governing the dissolution reaction. Analysis of their temporal trend can therefore provide critical knowledge on the evolution of kinetic processes and on how these impact on the behavior of R, as viewed in a stochastic context.

While we exemplify the joint use of our experimental findings and statistical modeling frameworks to document (nanoscale) spatial distributions of rates observed across the surface of a calcite crystal subject to dissolution, our experimental and theoretical frameworks are flexible and readily transferable to a variety of minerals to quantify the rates and mechanisms governing chemical weathering thereof. Future investigations are envisioned to target coupled dissolution-precipitation reactions taking place at solid-fluid interfaces. The dissolution of a mineral substrate leads to the availability in the liquid phase of dissolved compounds. In turn, these might react with chemicals in the solution to form a new solid phase on the surface. Coupled chemical phenomena of this kind underpin critical processes such as, e.g., geological sequestration of CO_2 or incorporation of heavy metals in natural geomaterials (see, e.g., Renard et al., 2019; Putnis and Putnis, 2022). Understanding timescales and process parameters governing such coupled reactions at a fundamental level is key to engineer efficient design strategies of target environmental scenarios. Recent experimental efforts devoted to study these processes yield detailed information about nanoscale interactions of, e.g., a carbonate solid phase and Cadmium (Julia et al., 2023), Chromium (Guren et al., 2020) or Antimony (Renard et al., 2018). In these works, temporal changes in the surface reactivity ascribed to the competition between dissolution and precipitation processes are quantified through average values such as etch pit spreading rates. In contrast, the experimental setup and the mask fabrication procedure designed in the context of this PhD thesis would allow to obtain quantitative spatial maps of absolute reaction rates.

The modeling frameworks employed (GSG) or developed (GMIX and GSG-MIX) within the context of this PhD thesis respond to the critical need to include small scale processes in the interpretation of sample rate PDFs to provide a comprehensive description of the reaction kinetics. Having at our disposal modeling tools capable of encapsulating the dynamics of the physical mechanisms taking place at the solid-liquid interface within a robust theoretical framework providing a joint accurate description of the statistics of the variable and its increments can be beneficial to transfer information to other spatial scales. In this context, we envision to investigate the statistical scaling nature of R through analysis of sample structure functions. These are defined as sample moments of order q of spatial increments, i.e.,

$$S_N^q(s) = \frac{1}{N(s)} \sum_{n=1}^{N(s)} |\Delta R(s)|^q, \tag{5.1}$$

$N(s)$ being the number of incremental values associated with a given lag s. Structure functions exhibit a power law scaling if

$$S_N^q \propto s^{\xi(q)}, \tag{5.2}$$

$\xi(q)$ being a scaling exponent depending solely on q. The nature of scaling is reflected in the functional relationship between $\xi(q)$ and q. If $\xi(q)$ is linearly proportional to q such that $\xi(q) = Hq$, then the field is a (self-affine) monofractal random field/process (i.e., a fractional Brownian motion), H being the (constant) Hurst exponent. In contrast, if $\xi(q)$ is nonlinear in q such that $\xi(q) \leq Hq$, the field is considered to be multifractal (e.g., Veneziano and Yoon, 2013). Significantly diverse behaviors are attributed to monofractal and multifractal fields/processes. Whereas the former arise from additive components, the latter are associated with multiplicative phenomena characterized by continuous variations of the Hurst exponent with q. Despite attempts to associate such multiplicative phenomena with universal multiplicative energy cascades (Schertzer and Lovejoy, 1988) or fractional Laplace motion (Meerschaert et al., 2004), there is no rigorous and universally valid theoretical basis underpinning multifractal behaviors. Apparent multifractal traits might otherwise arise from truncation of a monofractal field (see, e.g., Guadagnini et al., 2012; Guadagnini and Neuman, 2011). This is in turn consistent with a theoretical description based on the truncated fractional Brownian motion (tfBm), as defined for statistically isotropic functions by Di Federico and Neuman (1997) and for statistically anisotropic functions by Di Federico et al. (1999). The correlation structure of a tfBm is characterized by a Truncated Power Variogram (TPV) associated with lower and upper cutoffs proportional to the measurement (or resolution) scale of the data and to the scale of the domain across which the data are sampled, respectively (Di Federico and Neuman, 1997). Leveraging on the GSG model and viewing the underlying Gaussian field as a tfBm, one could incorporate such scaling features of a field in the unique theoretical framework offered by the GSG model. We envision to consider such a conceptualization to characterize (*i*) nanoscale dissolution rate fields obtained through AFM upon relying on the setups designed in the context of this thesis and (*ii*) VSI-based observations taken at a larger spatial scale. We will investigate whether sample structure functions at these two scales exhibit traits that are consistent with apparent multifractal behaviors. We will then rely on the GSG-MIX framework developed in this thesis and view each of the component of the mixture as a tfBm subordinated to $U_m(\mathbf{x})$. Interpretation of the ensuing correlation length scale associated with each underlying Gaussian field will be grounded on the TPV model. We plan to investigate the temporal trend of lower-and upper-cutoffs of the TPV to link these to mechanistic processes driving the reaction, as well as to the data support scale. Such a modeling workflow would provide a first direct comparison of the statistical behavior of R at two different spatial scales.

A parallel line of research that might stem from this PhD thesis work will address the long-standing challenge of deriving accurate predictive Reactive Transport Models (RTM) at the continuum level embedding the inherent surface reactivity. High fidelity topographic measurements taken at various times and ensuing reaction rate fields can be readily employed to assess model formulations. Analyses of this type have been recently performed by Schabernack and Fischer (2022, 2024). These authors implement a RTM in COMSOL Multiphysics using VSI data of polished calcite from Bibi et al. (2018) and Fischer and Lüttge (2018) as targets. Parametrization of the surface reactivity is grounded on the local topographic gradient. We envision to further advance these formulations based on a so-called empirical *surface slope factor* to derive a theoretically robust relationship between local rates and spatial gradients of crystal surface topography. Such a mathematical formulation could include information and/or constraints associated with the crystallographic structure of the mineral through model parameters that could be inferred by taking advantage of the high resolution offered by AFM. The inherent reactivity of the crystal surface could also be encapsulated in a RTM by solving the latter in a typical Monte Carlo framework. In this context, a collection of random rate fields could be generated according to the GSG, GMIX, or GSG-MIX models.

Appendices

A | Reliability of *Setting 1*

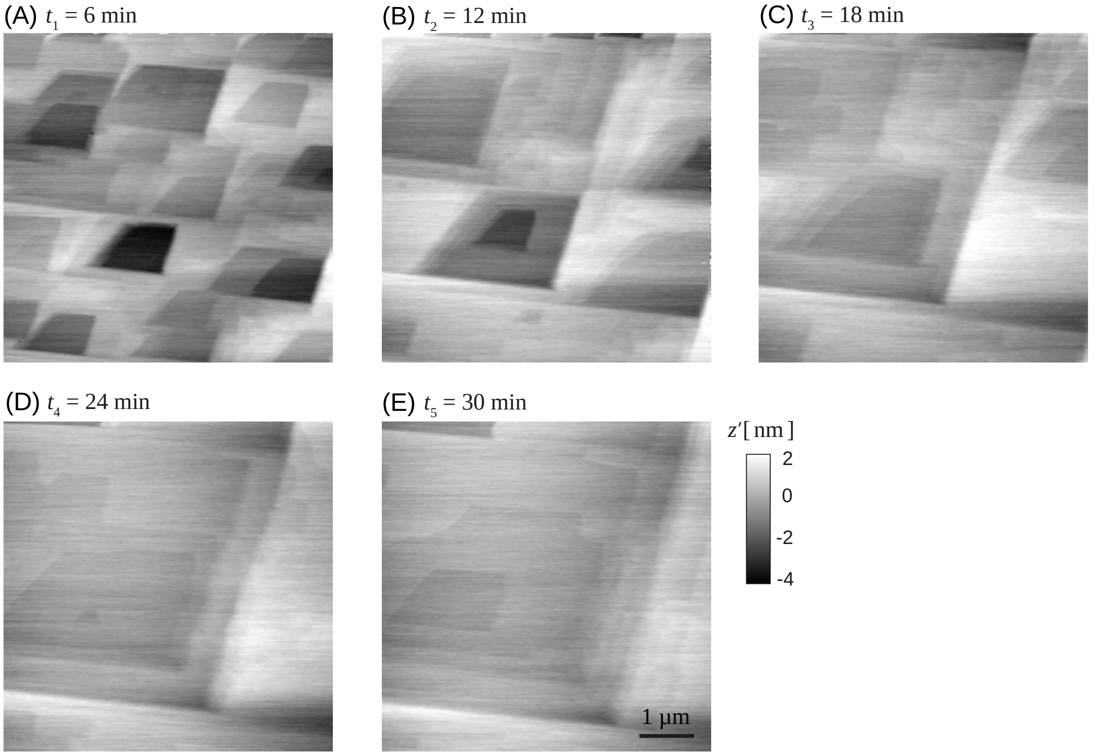

Figure A.1: Exemplary evolution (*Series 2*) of the topography of calcite surface imaged upon relying on experimental *Setting 1* at **(A)** t_1 = 6 min, **(B)** t_2 = 12 min, **(C)** t_3 = 18 min, **(D)** t_4 = 24 min, and **(E)** t_5 = 30 min from the beginning of the experiment.

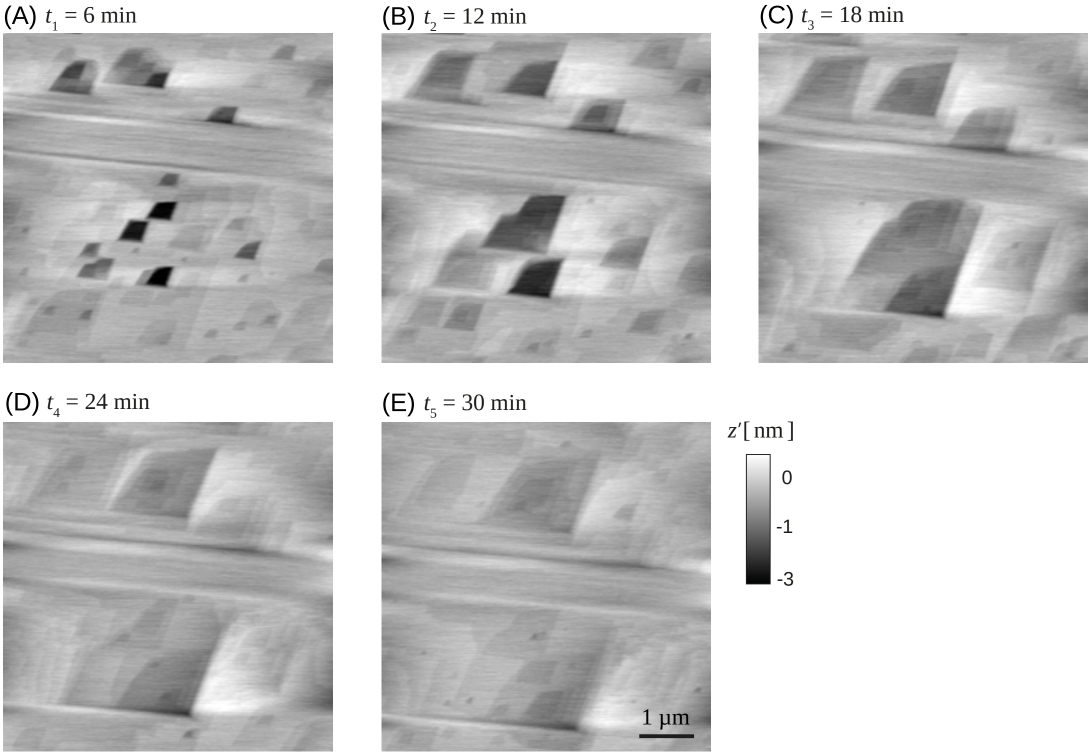

Figure A.2: Exemplary evolution (*Series 3*) of the topography of calcite surface imaged upon relying on experimental *Setting 1* at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment.

B | Reliability of *Setting 2*

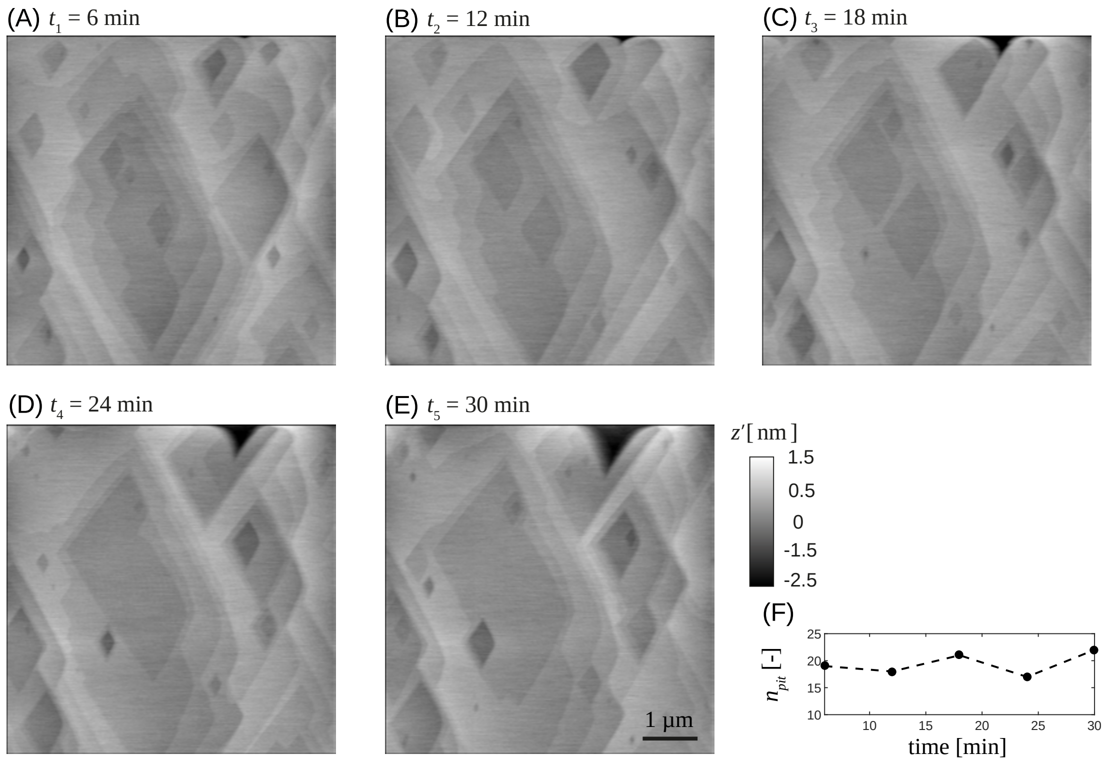

Figure B.1: Exemplary evolution (*Series 2*) of the topography of calcite surface imaged upon relying on Experimental Setting 2 at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the number of monolayer etch pits, n_{pit}, versus time.

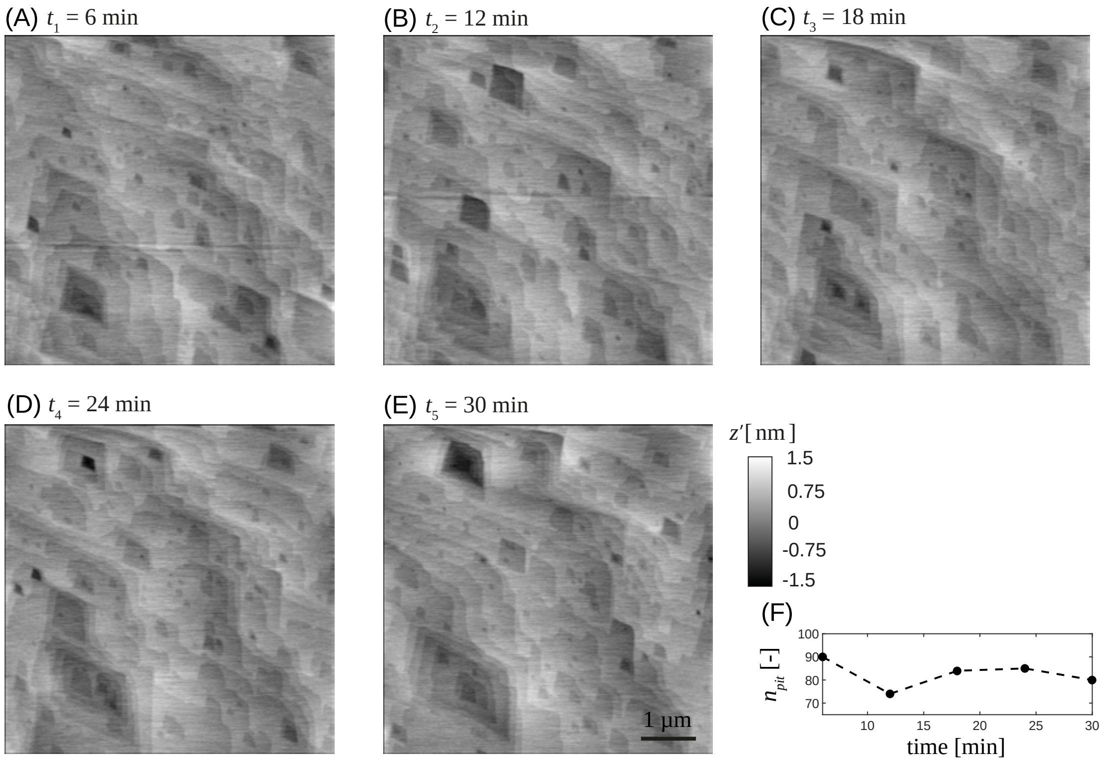

Figure B.2: Exemplary evolution (*Series 3*) of the topography of calcite surface imaged upon relying on Experimental Setting 2 at **(A)** t_1 = 6 min, **(B)** t_2 = 12 min, **(C)** t_3 = 18 min, **(D)** t_4 = 24 min, and **(E)** t_5 = 30 min from the beginning of the experiment. Panel **(F)** depicts the number of monolayer etch pits, n_{pit}, versus time.

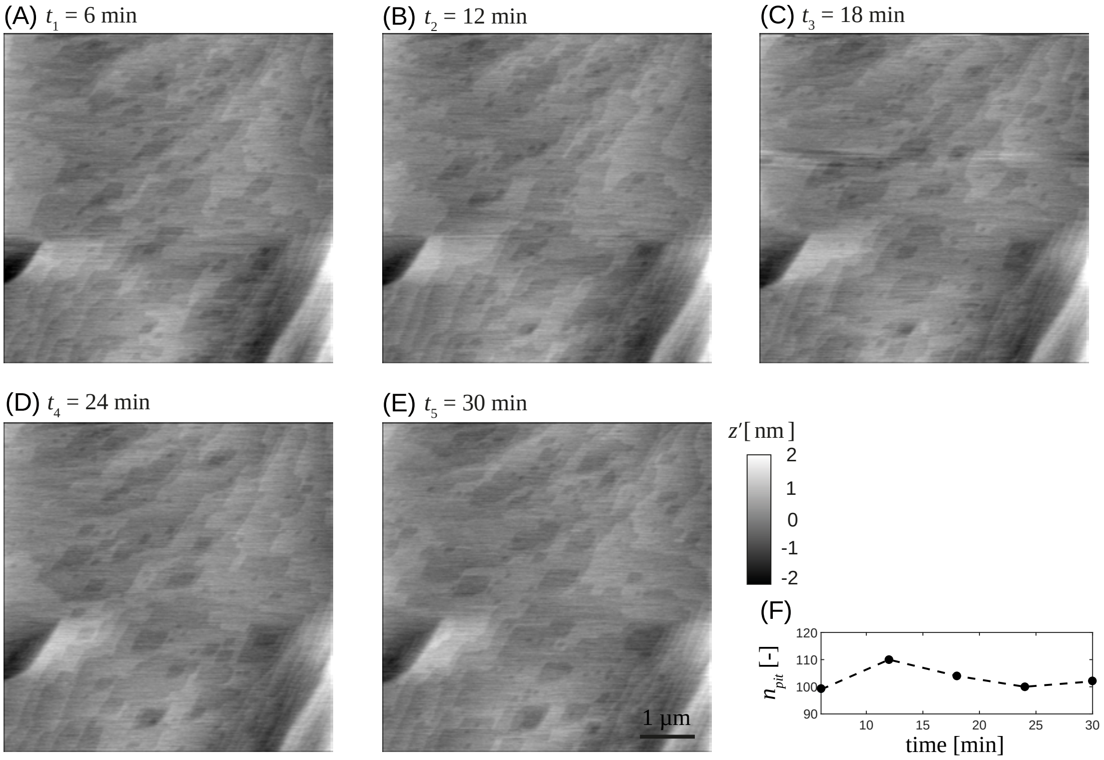

Figure B.3: Exemplary evolution (*Series 4*) of the topography of calcite surface imaged upon relying on Experimental Setting 2 at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the number of monolayer etch pits, n_{pit}, versus time.

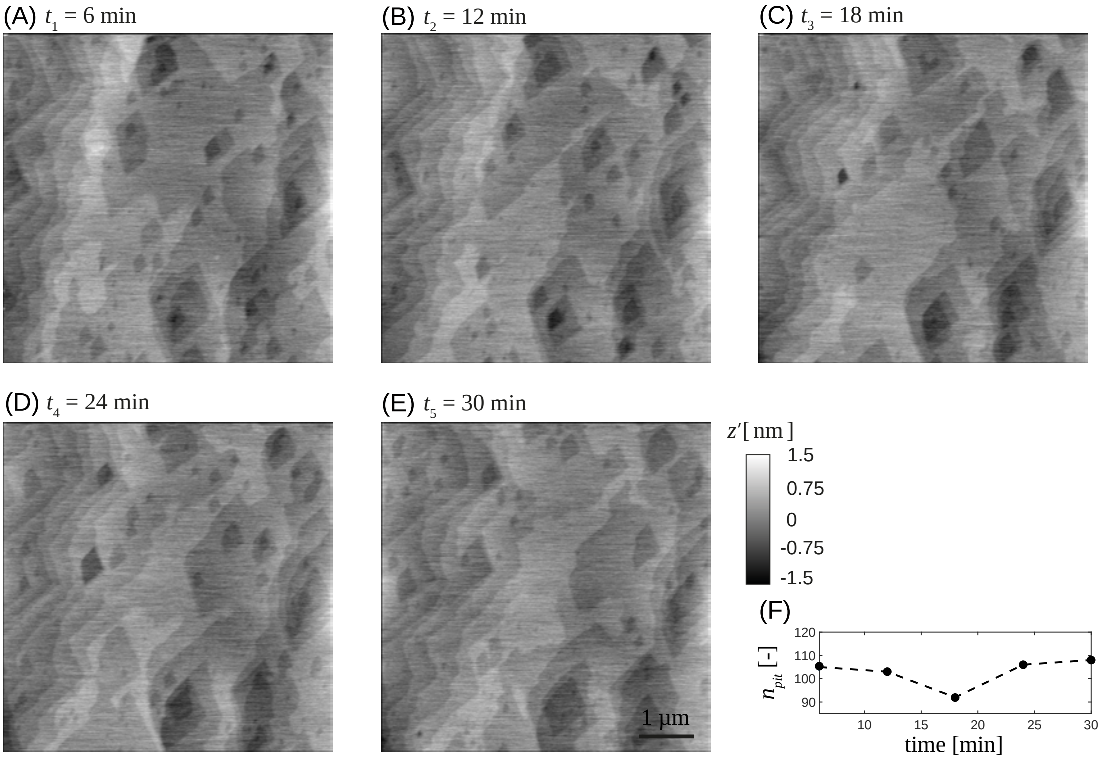

Figure B.4: Exemplary evolution (*Series 5*) of the topography of calcite surface imaged upon relying on Experimental Setting 2 at **(A)** t_1 = 6 min, **(B)** t_2 = 12 min, **(C)** t_3 = 18 min, **(D)** t_4 = 24 min, and **(E)** t_5 = 30 min from the beginning of the experiment. Panel **(F)** depicts the number of monolayer etch pits, n_{pit}, versus time.

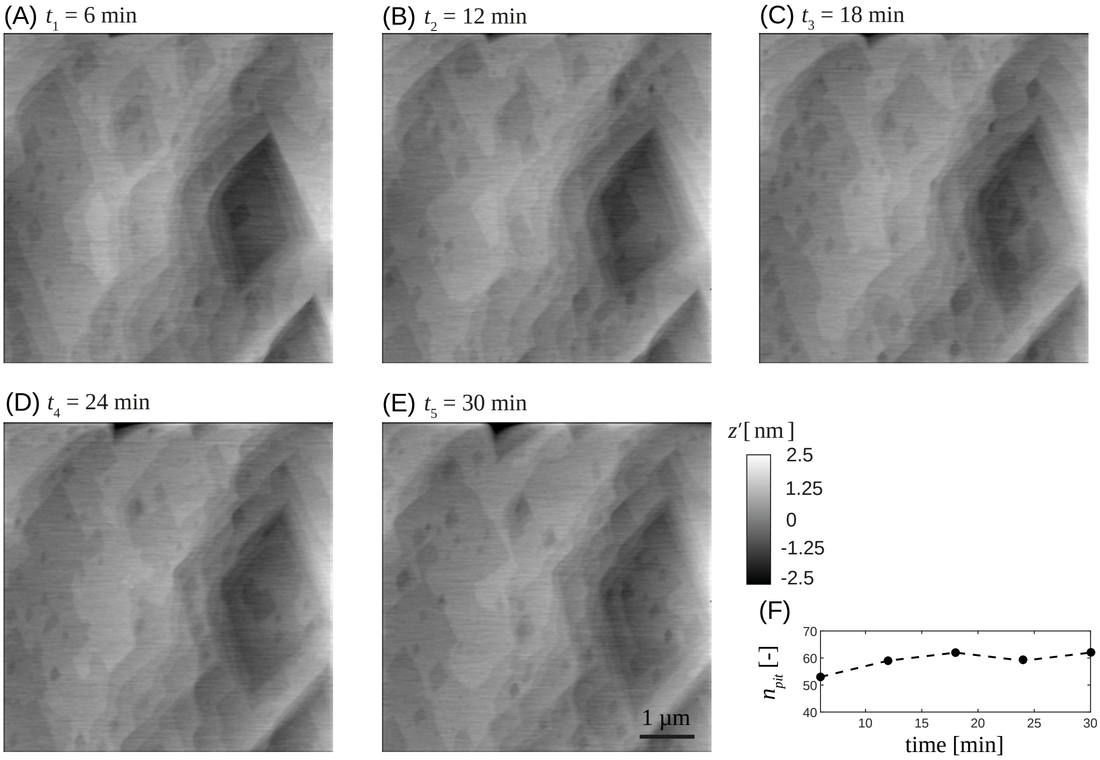

Figure B.5: Exemplary evolution (*Series 6*) of the topography of calcite surface imaged upon relying on Experimental Setting 2 at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the number of monolayer etch pits, n_{pit}, versus time.

C | Reliability of *Setting 3*

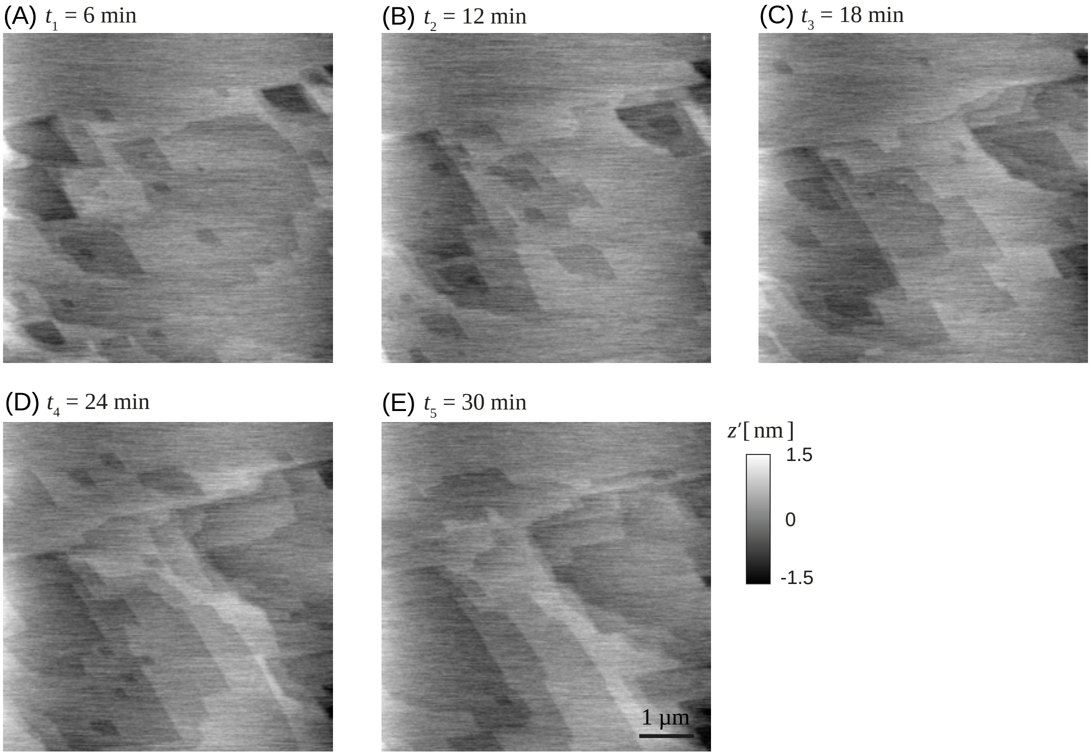

Figure C.1: Exemplary evolution of the topography of calcite surface imaged upon relying on *Setting 3* at $Q = 5\,\mu\text{Ls}^{-1}$ ($Re = 0.37$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment.

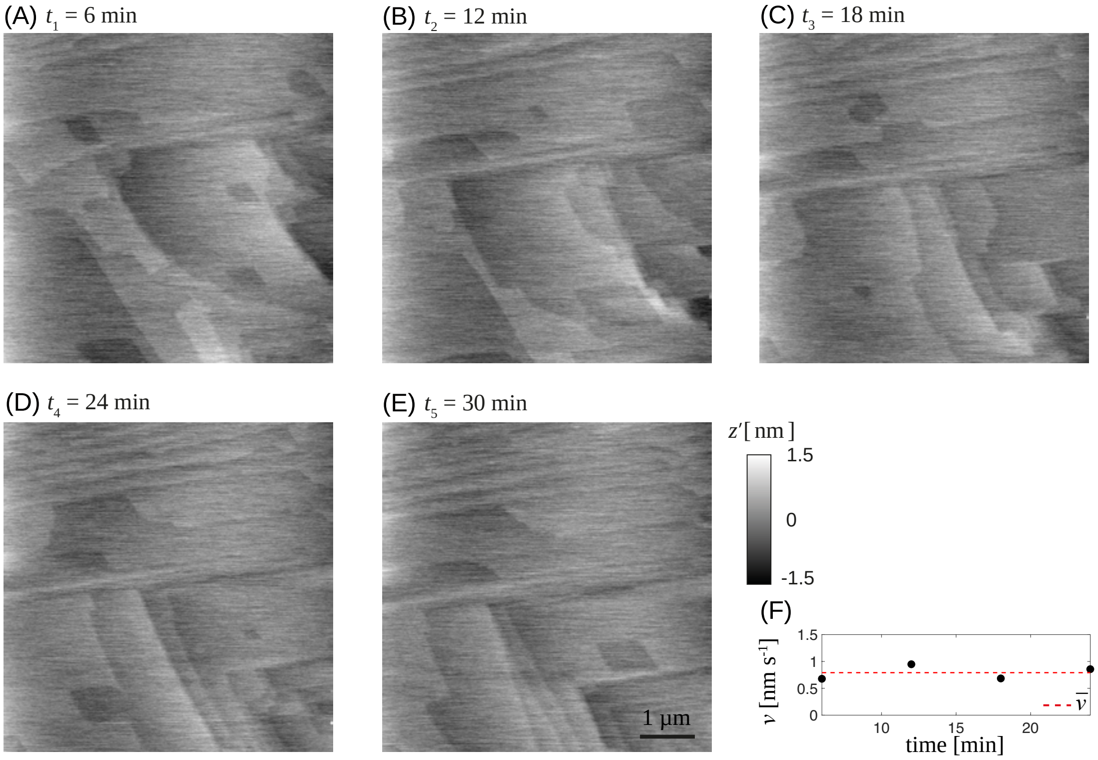

Figure C.2: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 8\,\mu\mathrm{Ls}^{-1}$ ($Re = 0.59$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

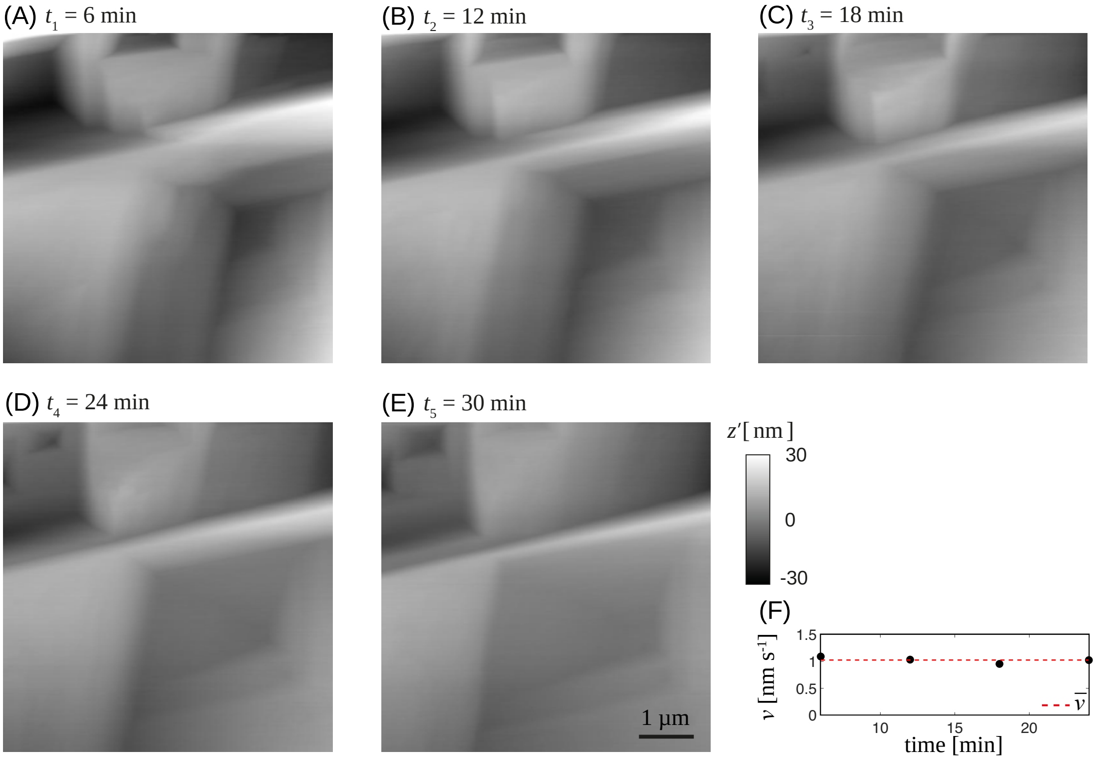

Figure C.3: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 9\,\mu\text{Ls}^{-1}$ ($Re = 0.66$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

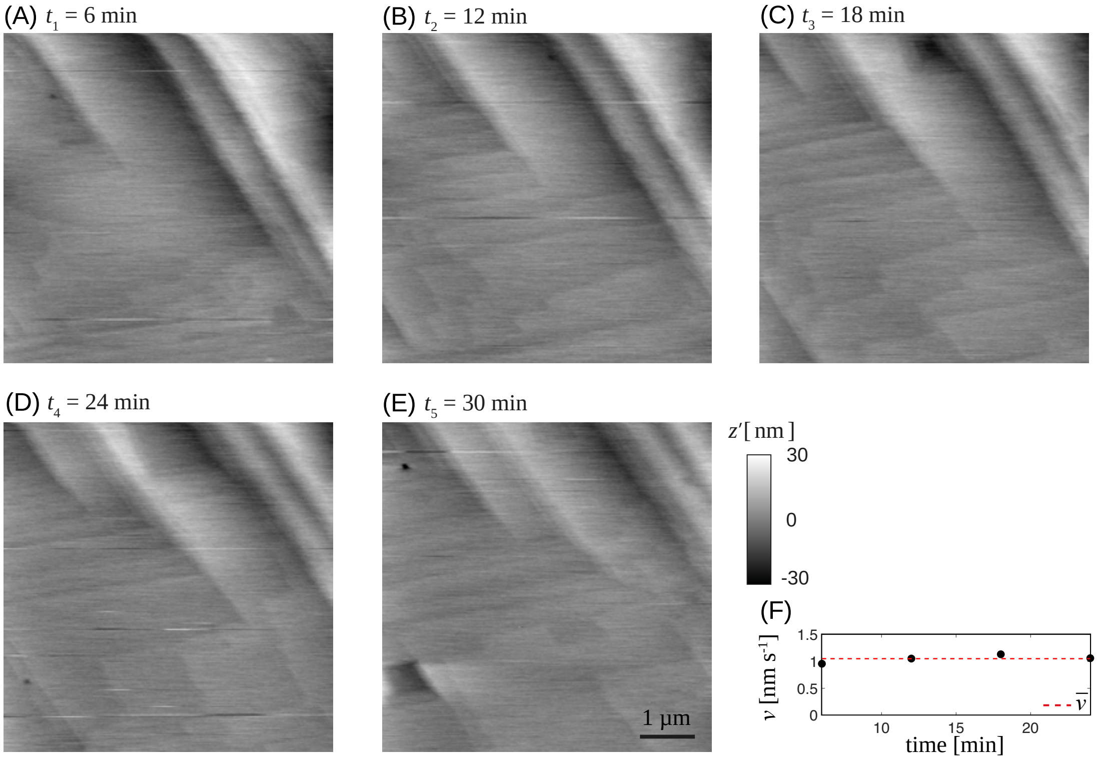

Figure C.4: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 10\,\mu\text{Ls}^{-1}$ ($Re = 0.74$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

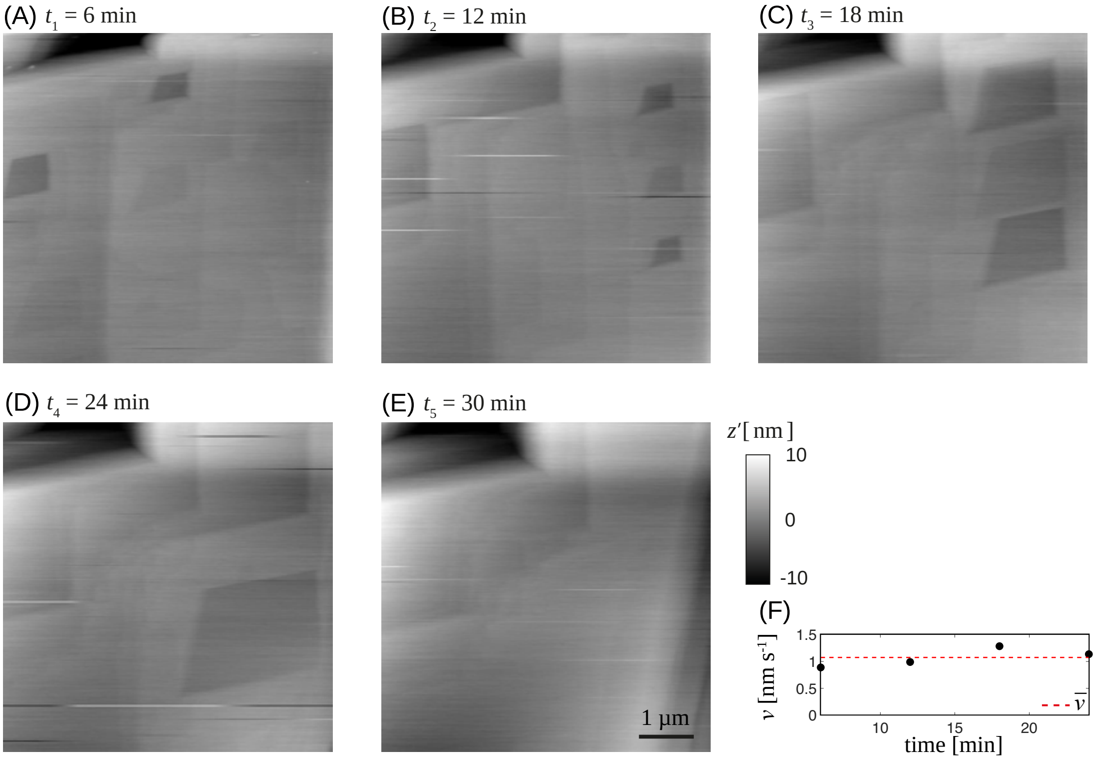

Figure C.5: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 11\,\mu\text{Ls}^{-1}$ ($Re = 0.81$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

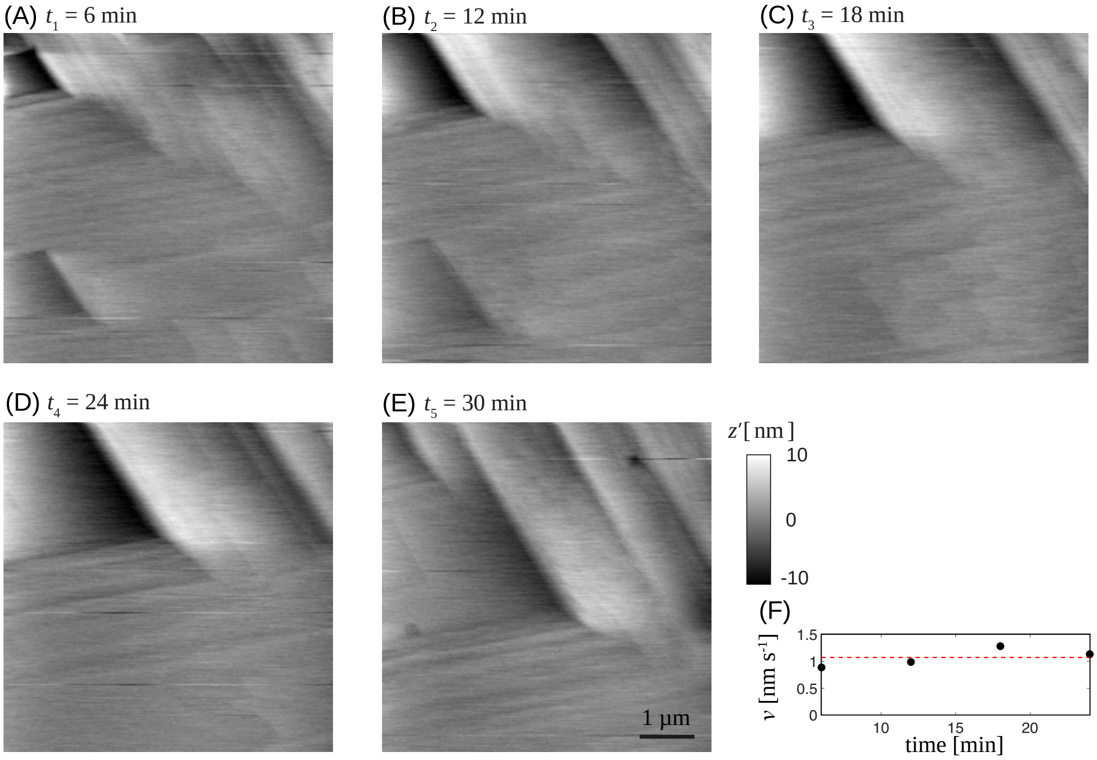

Figure C.6: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 12\,\mu\text{Ls}^{-1}$ ($Re = 0.89$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

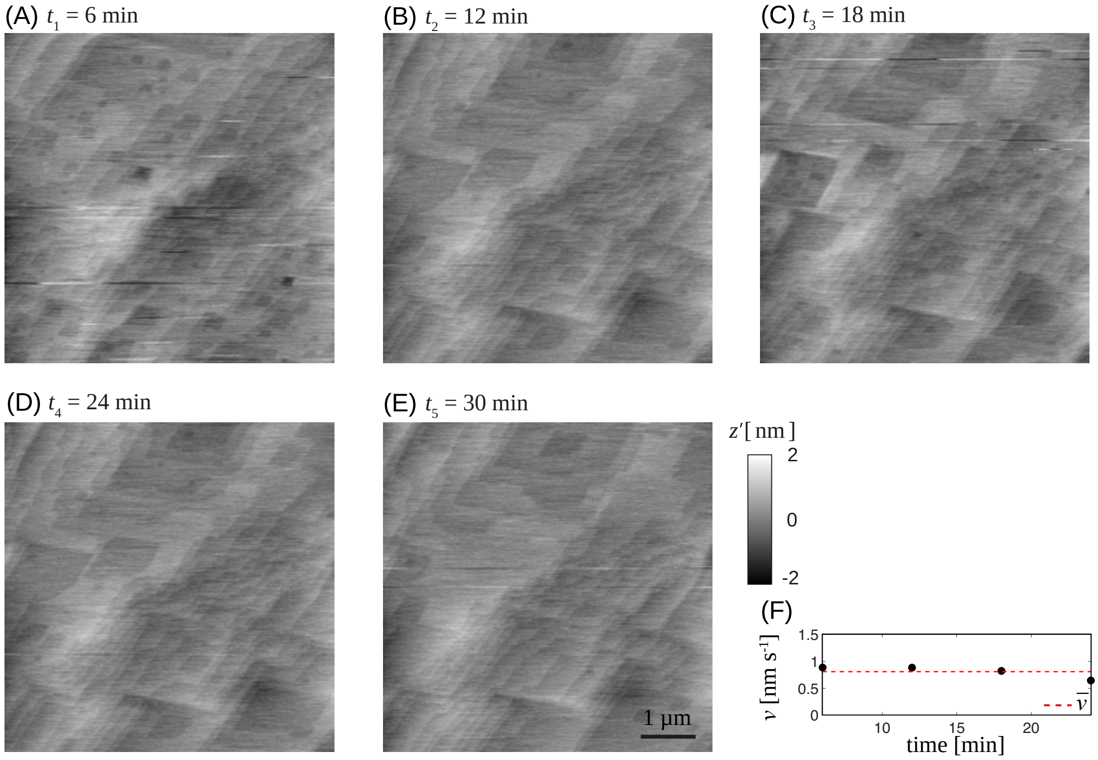

Figure C.7: Exemplary evolution of the topography of calcite surface imaged upon relying on experimental *Setting 3* at $Q = 15\,\mu\mathrm{Ls}^{-1}$ ($Re = 1.11$) at **(A)** $t_1 = 6$ min, **(B)** $t_2 = 12$ min, **(C)** $t_3 = 18$ min, **(D)** $t_4 = 24$ min, and **(E)** $t_5 = 30$ min from the beginning of the experiment. Panel **(F)** depicts the etch pit spreading rate, ν_t, as a function of time. The average value, $\bar{\nu}$, is reported as a dashed red line.

D | Lognormal Generalized sub-Gaussian model

In the following we specialize the derivation of the GSG model considering the subordinator as lognormally distributed. Therefore, the distributional form of U reads

$$f_U(u) = \frac{1}{\sqrt{2\pi}u(2-\alpha)} e^{-\frac{\ln^2 u}{2(2-\alpha)^2}}. \tag{D.1}$$

Substituting Eq. D.1 into Eq. 3.11 yields

$$f_{Y'}(y') = \frac{1}{2\pi(2-\alpha)} \int_0^\infty \frac{1}{u^2} e^{-\frac{1}{2}\left(\frac{1}{(2-\alpha)^2}\ln^2\frac{u}{\sigma_G} + \frac{y'^2}{u^2}\right)} du. \tag{D.2}$$

The $q-th$ order raw moment of U can be evaluated as

$$\langle U^q \rangle = e^{\frac{q^2}{2}(2-\alpha)^2}. \tag{D.3}$$

Eq. D.3 dictates that $\langle U^q \rangle \to 1$ if $\alpha \to 2$. Variance and kurtosis respectively read

$$\sigma_Y^2 = e^{2(2-\alpha)^2}\sigma_G^2, \tag{D.4}$$

$$\kappa_Y = 3e^{4(2-\alpha)^2}. \tag{D.5}$$

Substituting Eq. D.1 in Eq. 3.16 leads to

$$f_{\Delta Y}(\Delta Y) = \frac{1}{2\pi^2(2-\alpha)^2}\sqrt{\frac{\pi}{2}} \int_0^\infty \int_0^\infty \frac{e^{-\frac{1}{2}\left[\frac{1}{(2-\alpha)^2}\left(\ln^2\frac{u_1}{\sigma_G} + \ln^2\frac{u_2}{\sigma_G}\right) + \frac{(\Delta y)^2}{r}\right]}}{r} \frac{du_1}{u_1}\frac{du_2}{u_2}. \tag{D.6}$$

with $r = \sqrt{u_1^2 + u_2^2 - 2\rho_{G_m} u_1 u_2}$. Making use of Eq. D.6, second and fourth order moment and kurtosis of the increments read

$$\langle \Delta Y^2 \rangle = 2\sigma_G^2 e^{(2-\alpha)^2}\left[e^{(2-\alpha)^2} - \rho_G\right], \tag{D.7}$$

$$\langle \Delta Y^4 \rangle = 6\sigma_G^4 e^{4(2-\alpha)^2}\left[e^{4(2-\alpha)^2} + 1 - 4e^{4(2-\alpha)^2}\rho_G + 2\rho_G^2\right], \tag{D.8}$$

$$\kappa_{\Delta Y} = 3e^{2(2-\alpha)^2}\left\{1 + \frac{1}{2}\left[\frac{e^{2(2-\alpha)^2} - 1}{e^{(2-\alpha)^2} - \rho_G}\right]^2\right\}. \tag{D.9}$$

The covariance and the variogram of Y respectively are expressed as

$$C_Y = e^{(2-\alpha)^2} C_G, \tag{D.10}$$

$$\gamma_Y = \nu + e^{(2-\alpha)^2}\gamma_G. \tag{D.11}$$

where $C_G = \sigma_G^2 \rho_G$ is the covariance of the G and $\nu = \sigma_G^2 e^{(2-\alpha)^2}\left[e^{(2-\alpha)^2} - 1\right]$ is the nugget effect. The integral scale of Y' is obtained as

$$I_Y = e^{-(2-\alpha)^2} I_G. \tag{D.12}$$

E | Lognormal Generalized sub-Gaussian Mixture model

In the following we specialize the derivation of the GSG-MIX model considering the subordinator as lognormally distributed according to Eq. D.1. Substituting Eq. D.1 into Eq. 3.31 yields

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

The $q-th$ order raw moment of U_m can be evaluated through Eq. D.3. Substituting the latter into Eqs. 3.52, 3.53, and 3.54 yields

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

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

$$\begin{aligned} \kappa_Y = \frac{1}{\sigma_Y^4}\Big\{ & 3p\left[\sigma_{G_A}^4 e^{8(2-\alpha_A)^2} - \sigma_{G_B}^4 e^{8(2-\alpha_B)^2}\right] + 3\sigma_{G_B}^4 e^{8(2-\alpha_B)^2} \\ & + p(1-p)(\mu_A-\mu_B)^2\Big[(1-3p(1-p))(\mu_A-\mu_B)^2 \\ & + 6\left(\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} - p\left(\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} - \sigma_{G_B}^2 e^{2(2-\alpha_B)^2}\right)\right)\Big]\Big\}. \end{aligned} \tag{E.4}$$

Eq. E.2 dictates that the variance of Y attains a minimum for $[\alpha_A^{\min(\sigma_Y^2)} \to 2, \alpha_B^{\min(\sigma_Y^2)} \to 2]$. The skewness of the PDF of Y is minimized for

$$\begin{cases} \alpha_A^{\min(Sk_Y)} \to 2 \\ \alpha_B^{\min(Sk_Y)} = 2 - \sqrt{\frac{1}{2}\ln\frac{(1-p)(\mu_A-\mu_B)^2+(3-p)\sigma_{G_A}^2}{(1-p)\sigma_{G_B}^2}} \end{cases},$$

and maximized for

$$\begin{cases} \alpha_A^{\max(Sk_Y)} = 2 - \sqrt{\frac{1}{2}\ln\frac{p(\mu_A-\mu_B)^2+(2+p)\sigma_{G_B}^2}{p\sigma_{G_A}^2}} \\ \alpha_B^{\max(Sk_Y)} \to 2 \end{cases}.$$

The kurtosis of Y is minimized along $\alpha_h \to 2$, with $h = A$ if $p < 0.5$, $h = B$ otherwise.

Substituting Eq. D.1 in Eq. 3.55 leads to

$$\begin{aligned}
f_{\Delta Y}(\Delta Y) = {} & \frac{p^2 + C_I(s)}{(2\pi)^{3/2}(2-\alpha_A)^2\sigma_{G_A}} \int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1 + \ln^2 u_2}{(2-\alpha_A)^2} + \frac{\Delta y^2}{2\sigma_{G_A}^2 r^2}\right]} \frac{du_2 du_1}{u_1 u_2 r} \\
& + \frac{(1-p)^2 + C_I(s)}{(2\pi)^{3/2}(2-\alpha_B)^2\sigma_{G_B}} \int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1 + \ln^2 u_2}{(2-\alpha_B)^2} + \frac{\Delta y^2}{2\sigma_{G_B}^2 r^2}\right]} \frac{du_2 du_1}{u_1 u_2 r} \\
& + \frac{p(1-p) - C_I(s)}{(2\pi)^{3/2}(2-\alpha_A)(2-\alpha_B)} \int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1}{(2-\alpha_A)^2} + \frac{\ln^2 u_2}{(2-\alpha_B)^2} + \frac{(\Delta y - \mu_A + \mu_B)^2}{v^2}\right]} \frac{du_2 du_1}{u_1 u_2 v} \\
& + \frac{p(1-p) - C_I(s)}{(2\pi)^{3/2}(2-\alpha_A)(2-\alpha_B)} \int_0^{+\infty}\int_0^{+\infty} e^{-\frac{1}{2}\left[\frac{\ln^2 u_1}{(2-\alpha_B)^2} + \frac{\ln^2 u_2}{(2-\alpha_A)^2} + \frac{(\Delta y + \mu_A - \mu_B)^2}{w^2}\right]} \frac{du_2 du_1}{u_1 u_2 w}
\end{aligned} \tag{E.5}$$

with $r = \sqrt{u_1^2 + u_2^2 - 2\rho_{G_m} u_1 u_2}$ $(m = A, B)$, $v = \sqrt{\sigma_{G_A}^2 u_1^2 + \sigma_{G_B}^2 u_2^2}$, and $w = \sqrt{\sigma_{G_B}^2 u_1^2 + \sigma_{G_A}^2 u_2^2}$. Making use of Eq. E.5, second and fourth order moment of the increments respectively read

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

$$\begin{aligned}
\langle \Delta Y^4 \rangle = {} & 6\Big\{p^2\sigma_{G_A}^4 e^{4(2-\alpha_A)^2}\left[1 + e^{4(2-\alpha_A)^2} - 4e^{(2-\alpha_A)^2}\rho_{G_A} + 2\rho_{G_A}^2\right] \\
& + (1-p)^2\sigma_{G_B}^4 e^{4(2-\alpha_B)^2}\left[1 + e^{4(2-\alpha_B)^2} - 4e^{(2-\alpha_B)^2}\rho_{G_B} + 2\rho_{G_B}^2\right] \\
& + p(1-p)\rho_I\sigma_{G_A}^4 e^{4(2-\alpha_A)^2}\left[1 + e^{4(2-\alpha_A)^2} - 4e^{(2-\alpha_A)^2}\rho_{G_A} + 2\rho_{G_A}^2\right] \\
& + \sigma_{G_B}^4 e^{4(2-\alpha_B)^2}\left[1 + e^{4(2-\alpha_B)^2} - 4e^{(2-\alpha_B)^2}\rho_{G_B} + 2\rho_{G_B}^2\right]\Big\} \\
& + 2\Big\{p(1-p)(1-\rho_I)\Big[(\mu_A - \mu_B)^4 + 6(\mu_A - \mu_B)^2\left(\sigma_{G_A}^2 e^{2(2-\alpha_A)^2} + \sigma_{G_B}^2 e^{2(2-\alpha_B)^2}\right) \\
& + 3\left(\sigma_{G_A}^4 e^{8(2-\alpha_A)^2} + \sigma_{G_B}^4 e^{8(2-\alpha_B)^2} + 2\sigma_{G_A}^2\sigma_{G_B}^2 e^{2\left((2-\alpha_A)^2 + (2-\alpha_B)^2\right)}\right)\Big]\Big\}.
\end{aligned} \tag{E.7}$$

The covariance and the variogram of Y respectively yield

$$C_Y = \begin{cases} \sigma_Y^2 & \text{if } s = 0 \\ p^2\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\rho_{G_A} + (1-p)^2\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\rho_{G_B} + & \text{if } s > 0 \\ +p(1-p)\rho_I\left[(\mu_A-\mu_B)^2 + \sigma_{G_A}^2 e^{(2-\alpha_A)^2}\rho_{G_A} + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\rho_{G_B}\right] & \end{cases} \tag{E.8}$$

$$\begin{aligned} \gamma_Y = \nu + p^2 e^{(2-\alpha_A)^2}\gamma_{G_A} + (1-p)^2 e^{(2-\alpha_B)^2}\gamma_{G_B} + (\mu_A-\mu_B)^2\gamma_I \\ + p(1-p)\left[e^{(2-\alpha_A)^2}(1-\rho_I\rho_{G_A}) + e^{(2-\alpha_B)^2}(1-\rho_I\rho_{G_B})\right], \end{aligned} \tag{E.9}$$

where the nugget effect, ν, reads

$$\nu = p\sigma_{G_A}^2\left[e^{(2-\alpha_A)^2}\left(e^{(2-\alpha_A)^2}-1\right)\right] + (1-p)\sigma_{G_B}^2\left[e^{(2-\alpha_B)^2}\left(e^{(2-\alpha_B)^2}-1\right)\right]. \tag{E.10}$$

The integral scale of the mixture is obtained as

$$\begin{aligned} I_Y = \frac{1}{\sigma_Y^2}\Big\{ & p^2\sigma_{G_A}^2 e^{(2-\alpha_A)^2}\lambda_{G_A} + (1-p)^2\sigma_{G_B}^2 e^{(2-\alpha_B)^2}\lambda_{G_B} \\ & + p(1-p)\Big[(\mu_A-\mu_B)^2\lambda_I + \sigma_{G_A}^2 e^{(2-\alpha_A)^2}\int_0^\infty \rho_I\rho_{G_A}ds + \sigma_{G_B}^2 e^{(2-\alpha_B)^2}\int_0^\infty \rho_I\rho_{G_B}ds\Big]\Big\}. \end{aligned} \tag{E.11}$$

F | Generation of synthetic bimodal fields

Multiple realizations of synthetic GMIX fields are generated to provide a transparent assessment of the reliability of the parameter estimation strategies described in Sections 3.6.1 and 3.6.3.

The generation of synthetic bimodal mixtures of Gaussian and Generalized sub-Gaussian fields is structured according to the four steps illustrated in the following and is perfomed on a two-dimensional regular grid formed by $n \times n$ nodes.

- *Step 1.* We generate an unconditional realization of the random indicator field, I.
- *Step 2.* We generate an unconditional realization of Y_A.
- *Step 3.* We generate an unconditional realization of Y_B.
- *Step 4.* We generate a bimodal field by setting (*i*) $Y = Y_A$ (see Step 2) at locations where $I = 1$ (see Step 1) and (*ii*) $Y = Y_B$ (see Step 3) at locations where $I = 0$ (see Step 1).

Steps 1-4 are repeated N times.

F.1 | GENERATION OF SYNTHETIC BIMODAL GAUSSIAN FIELDS

To assess the reliability of the EM algorithm as a classification methodology for GMIX fields, we generate $N = 100$ fields across a 100×100 grid. A schematic representation of the generation of GMIX fields according to steps 1-4 is depicted in Fig. F.1. An unconditional realization of the indicator field (step 1; Fig. F.1.A) is generated employing a Transition Probability simulation approach (which takes advantage of the widely tested code T-PROGS; e.g., Carle and Fogg, 1996, 1997). Here, we set $p = 0.2$ and $l_A = 8$. The Gaussian fields $Y_A(\mathbf{x})$ and $Y_B(\mathbf{x})$ are generated through a sequential Gaussian simulation framework (based on the broadly used and tested code SGSIM; e.g., Deutsch and Journel, 1998). Y_A and Y_B are obtained by two sets of N unconditional realizations of Gaussian random fields characterized by and exponential covariance function with λ_A= $\lambda_B = 6$, $\mu_A = 2.5$, $\mu_B = 0.5$,

$\sigma_A^2 = 0.15$ and $\sigma_B^2 = 0.05$ (step 2-3; Fig. F.1.B-C). The value of Y in each node of the grid is then computed via Eq. 3.28 (step 4; Fig. F.1.D).

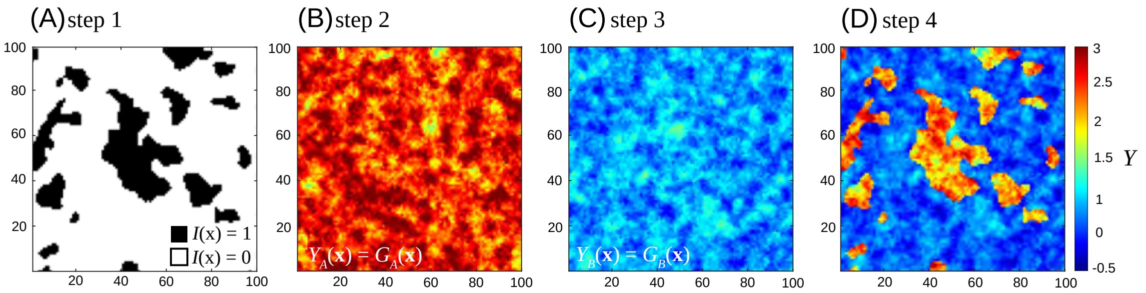

Figure F.1: Generation procedure of an exemplary GMIX random field; **(A)** synthetic realization of the indicator field, I; **(B-C)** synthetic realizations of the Gaussian fields, Y_A and Y_B; **(D)** final realization of the GMIX field obtained through Eq. 3.28.

F.2 | GENERATION OF SYNTHETIC BIMODAL GENERALIZED SUB-GAUSSIAN FIELDS

Analysis of the reliability of the algorithm resting on a Bayes classifier is grounded on the generation of fields resembling patterns that are typically observed for experimental dissolution rate fields. The generation procedure is schematically depicted in Fig. F.2 and is here detailed for the more general GSG-MIX case. We recall that the latter model includes the GMIX model as a particular case. The indicator field is constructed to resemble the typical rhombohedral shape of stepwaves emanated from an etch pit (Fischer and Lüttge, 2018) (step 1; Fig. F.2.A). The system is characterized by a proportion $p = 0.1$. Unconditional GSG fields associated with Y_A and Y_B are generated through a modified version of the SGSIM code (Riva et al., 2015b) (Steps 2-3; Fig. F.2.B-C). We rely upon an exponential model for the spatial correlation of G_m, i.e., $\rho_{G_m}(s) = e^{-s/\lambda_{G_m}}$ $(m = A, B)$, and known parameter sets ($\mu_A = 2.5, \sigma_{G_A}^2 = 0.05, \alpha_A = 1.7, \lambda_{G_A} = 10$ and $\mu_B = 0.5, \sigma_{G_B}^2 = 0.15, \alpha_B = 1.85, \lambda_{G_B} = 20$). Similar to the GMIX case, the bimodal Generalized sub Gaussian field is then obtained through Eq. 3.28 (step 4; Fig. F.2.D).

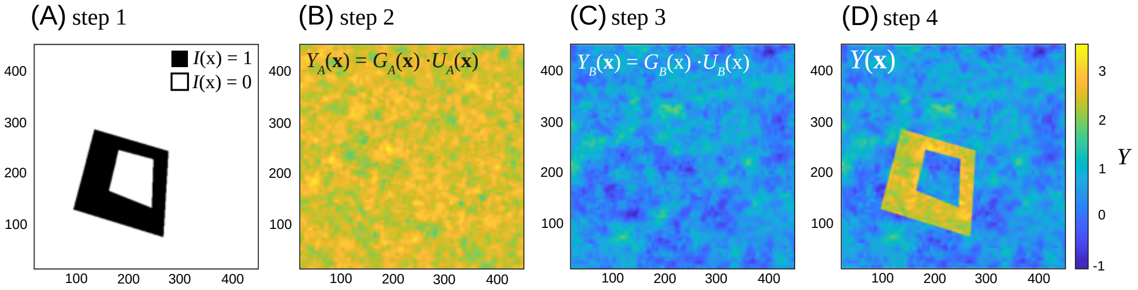

Figure F.2: Generation procedure of a GSG-MIX random field; **(A)** generation of an indicator random field, I, resembling geometrical patterns observed for dissolution rate maps; **(B-C)** synthetic realizations of the Generalized sub-Gaussian fields, Y_A and Y_B; **(D)** final realization of the GSG-MIX field obtained through Eq. 3.28.

G | GSG-MIX modeling of R'

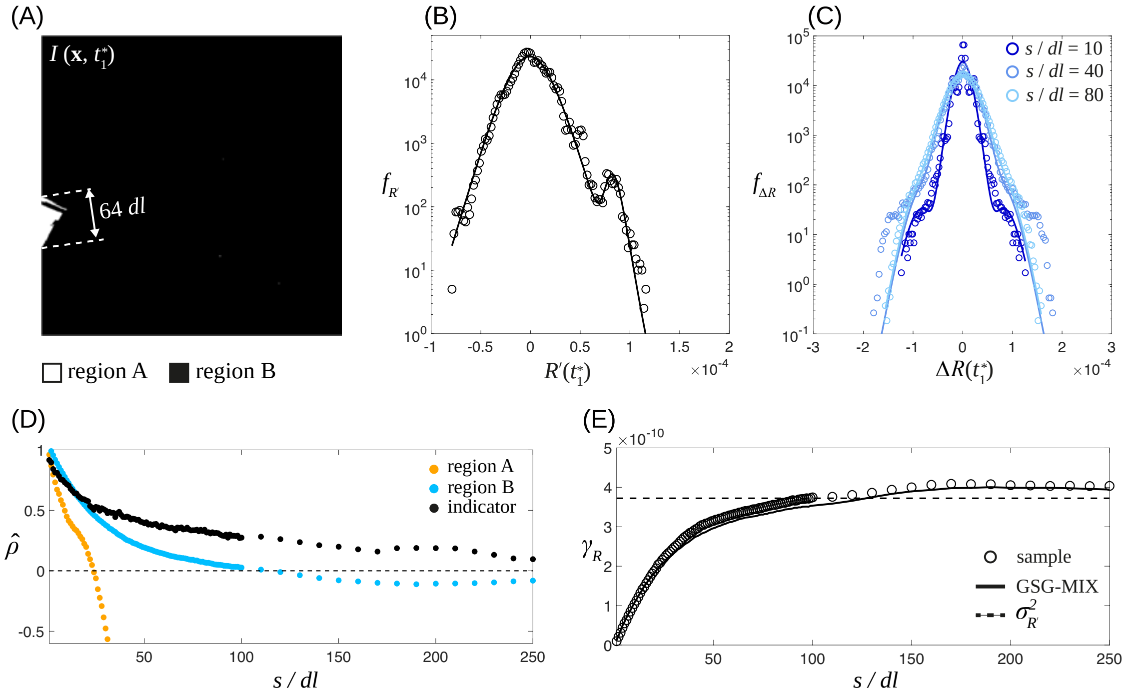

Figure G.1: Results of the GSG-MIX modeling framework applied to $R'(\mathbf{x}, t_1^*)$. **(A)** indicator field, $I(\mathbf{x}, t_1^*)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m-th$ mode of the mixture, $\hat{\rho}_{G_m}$ ($m = A, B$), and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9

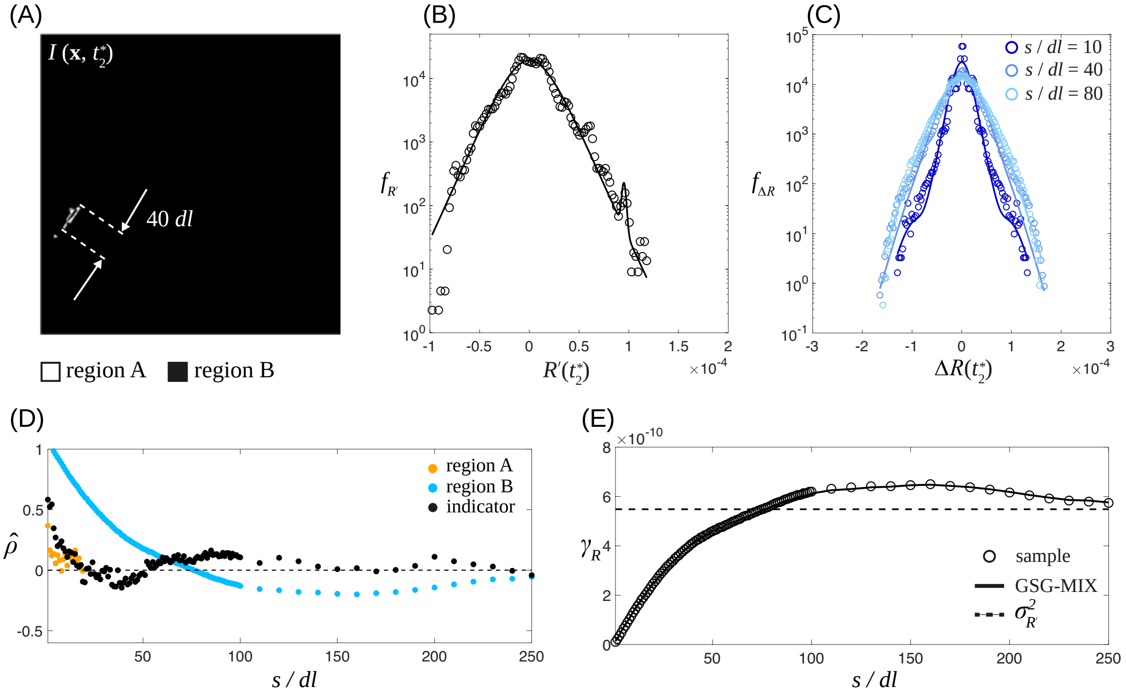

Figure G.2: Results of the GSG-MIX modeling framework applied to $R'(\mathbf{x}, t_2^*)$. **(A)** indicator field, $I(\mathbf{x}, t_2^*)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m-th$ mode of the mixture, $\hat{\rho}_{G_m}$ $(m = A, B)$, and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9.

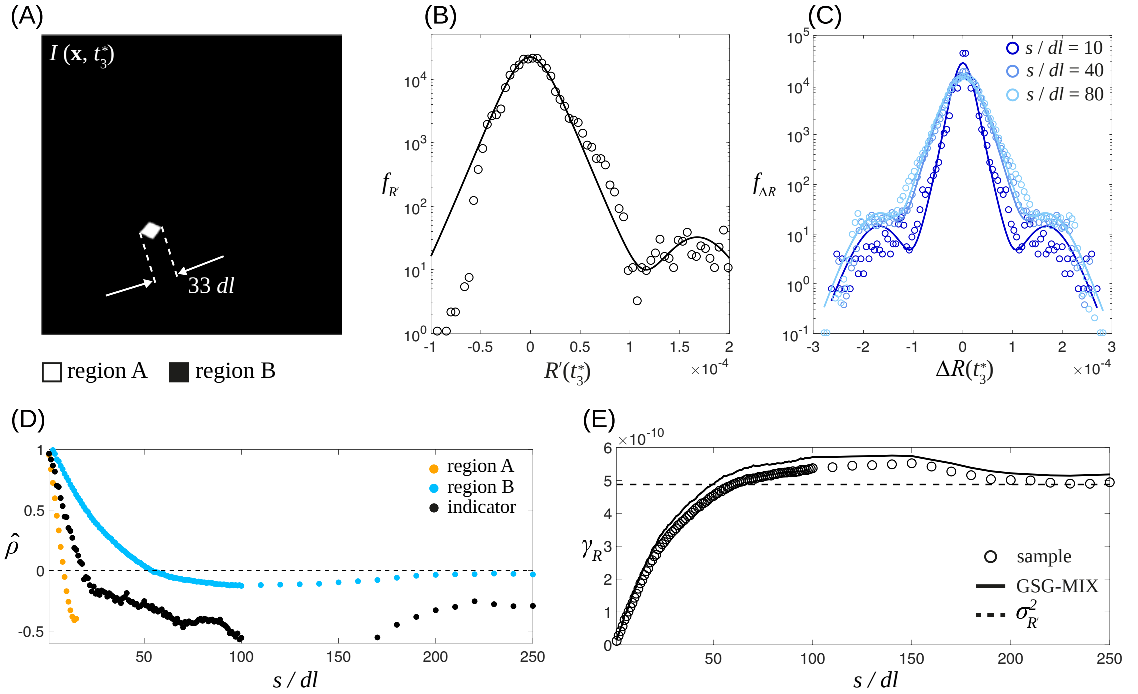

Figure G.3: Results of the GSG-MIX modeling framework applied to $R'(\mathbf{x}, t_3^*)$. **(A)** indicator field, $I(\mathbf{x}, t_3^*)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m - th$ mode of the mixture, $\hat{\rho}_{G_m}$ ($m = A, B$), and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9.

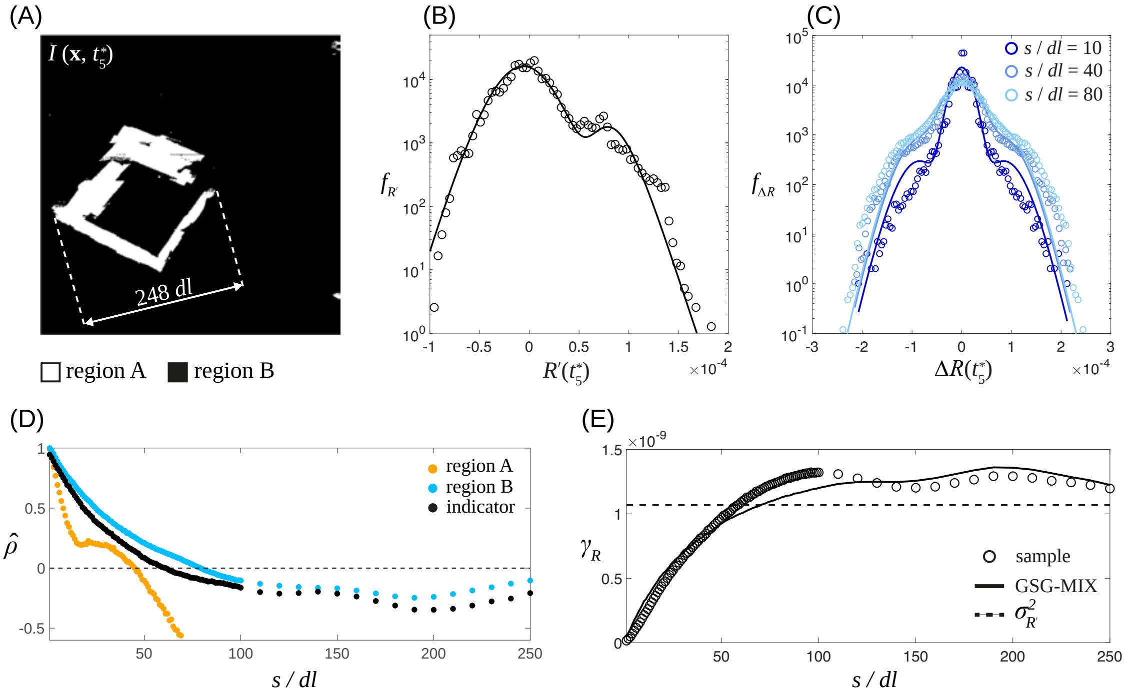

Figure G.4: Results of the GSG-MIX modeling framework applied to $R'(\mathbf{x}, t_5^*)$. **(A)** indicator field, $I(\mathbf{x}, t_5^*)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m-th$ mode of the mixture, $\hat{\rho}_{G_m}$ ($m = A, B$), and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9.

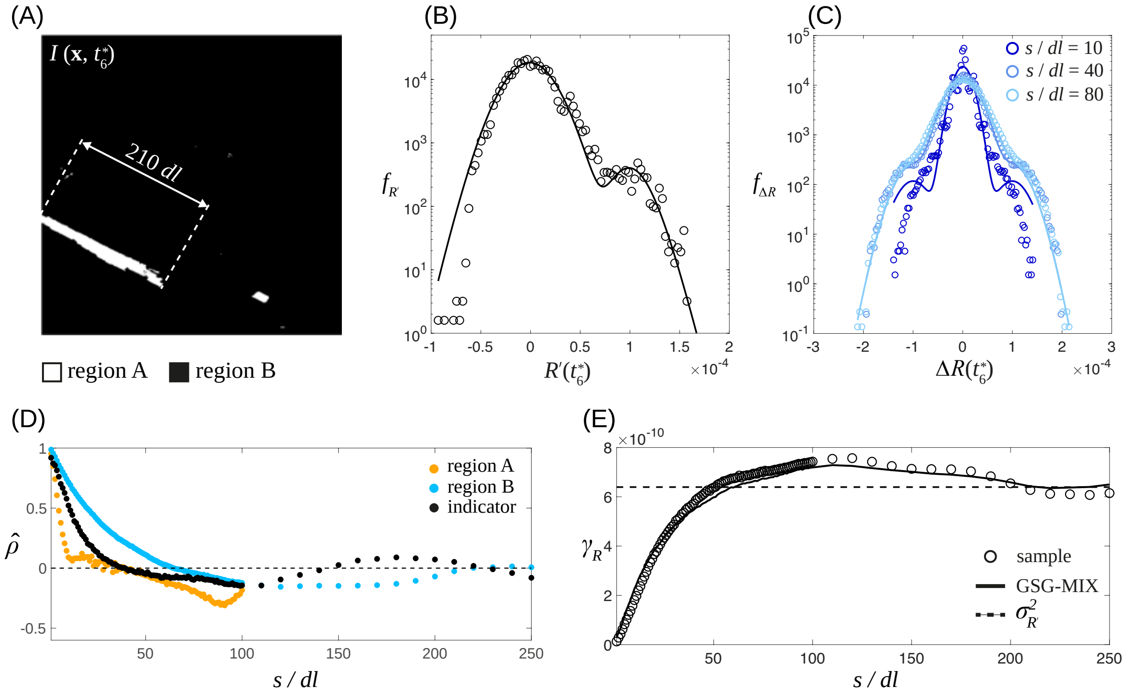

Figure G.5: Results of the GSG-MIX modeling framework applied to $R'(\mathbf{x}, t_6^*)$. **(A)** indicator field, $I(\mathbf{x}, t_6^*)$, resulting from the classification resting on the algorithm illustrated in Section 3.6.3. Sample and GSG-MIX PDFs of **(B)** R' and **(C)** ΔR (evaluated at lags $s = 10, 40$ and 80). **(D)** Estimated correlations of each $m-th$ mode of the mixture, $\hat{\rho}_{G_m}$ $(m = A, B)$, and of the indicator field, $\hat{\rho}_I$. **(E)** Sample and analytical GSG-MIX variogram, γ_R, evaluated through Eq. E.9.

H | Temporal trend of moments of R'

The statistical moments of the GSG-MIX model are controlled by (i) the square difference of the means, $\varepsilon^2 = (\mu_A - \mu_B)^2$, the difference of (ii) the variances, $\zeta = \sigma^2_{R_A} - \sigma^2_{R_B}$, and of the fourth order central moments, $\theta = 3\left(\langle {R'}^4_A\rangle - \langle {R'}^4_B\rangle\right)$ of the two components of the mixture, together with (iv) the variance of the indicator variable, $\sigma^2_I = p(1-p)$. Eqs. 3.52, 3.53 and 3.54 can be rewritten upon highlighting the latter terms as

$$\sigma^2_R = \sigma^2_{R_B} + p\zeta + \sigma^2_I\varepsilon^2, \tag{H.1}$$

$$Sk_R = \frac{1}{\sigma^3_R}\left\{\sigma^2_I\varepsilon\left[(1-2p)\varepsilon^2 + 3\zeta\right]\right\}, \tag{H.2}$$

$$\kappa_R = \frac{1}{\sigma^4_R}\left\{p\theta + \langle {Y'}^4_B\rangle + \sigma^2_I\varepsilon^2\left[\left(1-3\sigma^2_I\right)\varepsilon^2 + 6\left(\sigma^2_{Y_A} - p\zeta\right)\right]\right\}. \tag{H.3}$$

Analysis of the partial contribution to σ^2_Y, Sk_Y and κ_Y of each of the term embedded in Eqs. H.1, H.2, and H.3 can assist one to discriminate which GSG-MIX parameters have the highest impact on the statistical moments of the mixture. As GSG-MIX parameters are strictly linked to the evolution of the surface pattern, this is informative about the effect on the statistics of R' of the kinetic processes driving the reaction. Second, third, and fourth order moments of the mixture, together with their partial contributions, are depicted versus time in Fig. H.1. As the mechanism that is driving the level of heterogeneity of R_A and R_B is the same (i.e., dynamic evolution of shallow etch pits), $\zeta, \theta \to 0$ at all times (orange and dark red lines in Fig. H.1). Analysis of model parameters characterizing region B (Figs. 4.23.B-D) documents that statistical moments associated with this mode of the mixture are constant in time (blue lines in Fig. H.1). Furthermore, the proportion of region A with respect to the overall field is < 0.1 at all times (Figs. 4.23.A). Hence, all terms in Eqs. H.1, H.2, and H.3 that are multiplied by p are negligible with respect to the others (pink and green lines in Fig. H.1). As such, the temporal behavior of σ^2_R, Sk_R, and κ_R is controlled by σ^2_I and ε^2 (light blue lines in Fig. H.1).

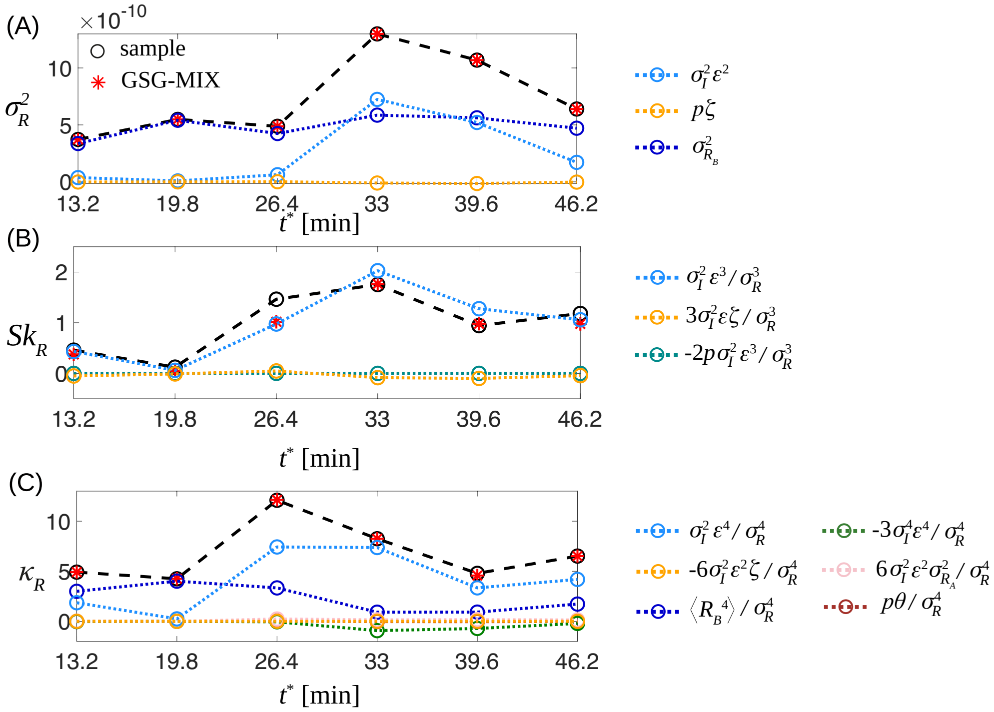

Figure H.1: Statistical moments of **(A)** second, **(B)** third, and **(C)** fourth order of R'. Each of the term embedded in Eqs. H.1, H.2 and H.3 and contributing to σ_R^2, Sk_R, and κ_R is depicted as a dotted line.

Bibliography

Agudo, E. R. and Putnis, C. V. (2012). Direct observations of mineral fluid reactions using atomic force microscopy: the specific example of calcite. *Mineralogical Magazine*, 76(1):227–253.

Arcos, D., Grandia, F., Domènech, C., Fernández, A. M., Villar, M. V., Muurinen, A., Carlsson, T., Sellin, P., and Hernán, P. (2008). Long-term geochemical evolution of the near field repository: Insights from reactive transport modelling and experimental evidences. *Journal of Contaminant Hydrology*, 102(3-4):196–209.

Arvidson, R. S., Ertan, I. E., Amonette, J. E., and Luttge, A. (2003). Variation in calcite dissolution rates: A fundamental problem? *Geochimica et cosmochimica acta*, 67(9):1623–1634.

Barri, C., Mafakheri, E., Fagiani, L., Tavani, G., Barzaghi, A., Chrastina, D., Fedorov, A., Frigerio, J., Lodari, M., Scotognella, F., et al. (2020). Engineering of the spin on dopant process on silicon on insulator substrate. *Nanotechnology*, 32(2):025303.

Bibi, I., Arvidson, R. S., Fischer, C., and Lüttge, A. (2018). Temporal evolution of calcite surface dissolution kinetics. *Minerals*, 8(6):256.

Boffetta, G., Mazzino, A., and Vulpiani, A. (2008). Twenty-five years of multifractals in fully developed turbulence: a tribute to giovanni paladin. *Journal of Physics A: Mathematical and Theoretical*, 41(36):363001.

Bollermann, T. and Fischer, C. (2020). Temporal evolution of dissolution kinetics of polycrystalline calcite. *American Journal of Science*, 320(1):53–71.

Bouissonnié, A., Daval, D., Marinoni, M., and Ackerer, P. (2018). From mixed flow reactor to column experiments and modeling: Upscaling of t calcite dissolution rate. *Chemical Geology*, 487:63–75.

Brand, A. S., Feng, P., and Bullard, J. W. (2017). Calcite dissolution rate spectra measured by in situ digital holographic microscopy. *Geochimica et cosmochimica acta*, 213:317–329.

Carle, S. and Fogg, G. (1996). Transition probability-based indicator geostatistics. *Math Geol*, 28(4):453–477.

Carle, S. and Fogg, G. (1997). Modelling spatial variability with one and multidimensional continuous-lag markov chains. *Math Geol*, 29(7):891–918.

Carrera, J. and Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. *Water Resources Research*, 22(2):211–227.

Chada, V. G. R., Hausner, D. B., Strongin, D. R., Rouff, A. A., and Reeder, R. J. (2005). Divalent cd and pb uptake on calcite {101$^-$ 4} cleavage faces: An xps and afm study. *Journal of Colloid and Interface Science*, 288(2):350–360.

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

Daval, D. (2018). Carbon dioxide sequestration through silicate degradation and carbon mineralisation: promises and uncertainties. *npj Materials Degradation*, 2(1):1–4.

De Graef, M. and McHenry, M. E. (2012). *Structure of materials: an introduction to crystallography, diffraction and symmetry*. Cambridge University Press.

Desbarats, A. J. (1987). Numerical estimation of effective permeability in sand- shale formations. *Water Resources Research*, 23:273–286.

Deutsch, C. V. and Journel, A. G. (1998). Gslib, geostatistical software library and user's guide. *Oxford University Press, New York*.

Di Federico, V. and Neuman, S. P. (1997). Scaling of random fields by means of truncated power variograms and associated spectra. *Water Resources Research*, 33(5):1075–1085.

Di Federico, V., Neuman, S. P., and Tartakovsky, D. M. (1999). Anisotropy, lacunarity, and upscaled conductivity and its autocovariance in multiscale random fields with truncated power variograms. *Water Resources Research*, 35(10):2891–2908.

Dong, S., Berelson, W. M., Adkins, J. F., Rollins, N. E., Naviaux, J. D., Pirbadian, S., El-Naggar, M. Y., and Teng, H. H. (2020). An atomic force microscopy study of calcite dissolution in seawater. *Geochimica et Cosmochimica Acta*, 283:40–53.

Dong, W., Ball, W. P., Liu, C., Wang, Z., Stone, A. T., Bai, J., and Zachara, J. M. (2005). Influence of calcite and dissolved calcium on uranium (vi) sorption to a hanford subsurface sediment. *Environmental science & technology*, 39(20):7949–7955.

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

Eaton, P. and West, P. (2010). *Atomic force microscopy*. Oxford university press.

Emmanuel, S. (2014). Mechanisms influencing micron and nanometer-scale reaction rate patterns during dolostone dissolution. *Chemical Geology*, 363:262–269.

Erol, S., Fowler, S. J., Nehler, M., De Boever, E., Harcouët-Menou, V., and Laenen, B. (2019). An analytical algorithm of porosity–permeability for porous and fractured media: extension to reactive transport conditions and fitting via flow-through experiments within limestone and dolomite. *Transport in Porous Media*, 129:343–383.

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

Fischer, C., Arvidson, R. S., and Lüttge, A. (2012). How predictable are dissolution rates of crystalline material? *Geochimica et Cosmochimica Acta*, 98:177–185.

Fischer, C., Finkeldei, S., Brandt, F., Bosbach, D., and 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.

Fischer, C., Kurganskaya, I., Schäfer, T., and Lüttge, A. (2014). Variability of crystal surface reactivity: what do we know? *Applied Geochemistry*, 43:132–157.

Fischer, C. and Lüttge, A. (2017). Beyond the conventional understanding of water–rock reactivity. *Earth and Planetary Science Letters*, 457:100–105.

Fischer, C. and Lüttge, A. (2018). Pulsating dissolution of crystalline matter. *Proceedings of the National Academy of Sciences*, 115(5):897–902.

Fitts, J. P. and Peters, C. A. (2013). Caprock fracture dissolution and co2 leakage. *Reviews in Mineralogy and Geochemistry*, 77(1):459–479.

Frisch, U. (2016). The collective birth of multifractals. *Journal of Physics A: Mathematical and Theoretical*, 49(45):451002.

Ganti, V., Singh, A., Passalacqua, P., and Foufoula-Georgiou, E. (2009). Subordinated brownian motion model for sediment transport. *Physical Review E*, 80(1):011111.

Garcıa, R. and Perez, R. (2002). Dynamic atomic force microscopy methods. *Surface science reports*, 47(6-8):197–301.

Gournelos, T., Kotinas, V., and Poulos, S. (2020). Fitting a gaussian mixture model to bivariate distributions of monthly river flows and suspended sediments. *Journal of Hydrology*, 590:125166.

Guadagnini, A., Neuman, S., Schaap, M., and Riva, M. (2013). Anisotropic statistical scaling of vadose zone hydraulic property estimates near maricopa, arizona. *Water Resources Research*, 49(12):8463–8479.

Guadagnini, A., Neuman, S., Schaap, M., and Riva, M. (2014). Anisotropic statistical scaling of soil and sediment texture in a stratified deep vadose zone near maricopa, arizona. *Geoderma*, 214:217–227.

Guadagnini, A. and Neuman, S. P. (2011). Extended power-law scaling of self-affine signals exhibiting apparent multifractality. *Geophysical research letters*, 38(13).

Guadagnini, A., Neuman, S. P., Nan, T., Riva, M., and 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.

Guadagnini, A., Neuman, S. P., and Riva, M. (2012). Numerical investigation of apparent multifractality of samples from processes subordinated to truncated fbm. *Hydrological Processes*, 26(19):2894–2908.

Guadagnini, A., Riva, M., and Neuman, S. P. (2018). Recent advances in scalable non-gaussian geostatistics: The generalized sub-gaussian model. *Journal of hydrology*, 562:685–691.

Guren, M. G., Putnis, C. V., Montes-Hernandez, G., King, H. E., and Renard, F. (2020). Direct imaging of coupled dissolution-precipitation and growth processes on calcite exposed to chromium-rich fluids. *Chemical Geology*, 552:119770.

Harrison, A. L., Jew, A. D., Dustin, M. K., Thomas, D. L., Joe-Wong, C. M., Bargar, J. R., Johnson, N., Brown Jr, G. E., and Maher, K. (2017). Element release and reaction-induced porosity alteration during shale-hydraulic fracturing fluid interactions. *Applied Geochemistry*, 82:47–62.

Harstad, A. and Stipp, S. (2007). Calcite dissolution: Effects of trace cations naturally present in iceland spar calcites. *Geochimica et Cosmochimica Acta*, 71(1):56–70.

Heberling, F., Bosbach, D., Eckhardt, J.-D., Fischer, U., Glowacky, J., Haist, M., Kramar, U., Loos, S., Müller, H. S., Neumann, T., Pust, C., Schäfer, T., Stelling, J., Ukrainczyk, M., Vinograd, V., Vučak, M., and Winkler, B. (2014). Reactivity of the calcite–water-interface, from molecular scale processes to geochemical engineering. *Applied Geochemistry*, 45:158–190.

Hellmann, R. and Tisserand, D. (2006). Dissolution kinetics as a function of the gibbs free energy of reaction: An experimental study based on albite feldspar. *Geochimica et Cosmochimica Acta*, 70(2):364–383.

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

Jia, B., Zhou, J., Tang, Z., Xu, Z., Chen, X., and Fang, W. (2022). Effective stochastic streamflow simulation method based on Gaussian mixture model. *Journal of Hydrology*, 605:127366.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 15:445–468.

Journel, A. G. and Huijbregts, C. J. (1976). *Mining geostatistics.* Academic press, Harcourt Brace Jovanovich.

Julia, M., Putnis, C. V., King, H. E., and 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.

Kanellopoulou, D. G. and Koutsoukos, P. G. (2003). The calcitic marble/water interface: Kinetics of dissolution and inhibition with potential implications in stone conservation. *Langmuir*, 19(14):5691–5699.

Kashyap, R. L. (1982). Optimal choice of ar and ma parts in autoregressive moving average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(2):99–104.

Khather, M., Saeedi, A., Myers, M. B., and Giwelli, A. (2020). Effects of co2-saturated brine on the injectivity and integrity of chalk reservoirs. *Transport in Porous Media*, 135(3):735–751.

Kitanidis, P. K. (1997). *Introduction to geostatistics: applications in hydrogeology*. Cambridge university press.

Kozubowski, T. J., Meerschaert, M. M., and Podgorski, K. (2006). Fractional laplace motion. *Advances in applied probability*, 38(2):451–464.

Kozubowski, T. J., Podgórski, K., and Rychlik, I. (2013). Multivariate generalized laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Kumar, P. and Foufoula-Georgiou, E. (1993). A multicomponent decomposition of spatial rainfall fields: 2. self-similarity in fluctuations. *Water Resources Research*, 29(8):2533–2544.

Kurganskaya, I., Arvidson, R. S., Fischer, C., and Luttge, A. (2012). Does the stepwave model predict mica dissolution kinetics? *Geochimica et cosmochimica acta*, 97:120–130.

Kurganskaya, I. and Luttge, A. (2016). Kinetic monte carlo approach to study carbonate dissolution. *The Journal of Physical Chemistry C*, 120(12):6482–6492.

Lamy-Chappuis, B., Angus, D., Fisher, Q., Grattoni, C., and Yardley, B. W. (2014). Rapid porosity and permeability changes of calcareous sandstone due to co2-enriched brine injection. *Geophysical Research Letters*, 41(2):399–406.

Lasaga, A. and Lüttge, A. (2001). Variation of crystal dissolution rate based on a dissolution stepwave model. *Science*, 291(5512):2400–2404.

Lasaga, A. and Lüttge, A. (2003). A model for crystal dissolution. *Eur. J. Mineral.*, 15(4):603–615.

Li, K., Wu, J., Nan, T., Zeng, X., Yin, L., and 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.

Liang, Y. and Baer, D. (1997). Anisotropic dissolution at the $caco_3(10\bar{1}4)$-water interface. *Surface Science*, 373(2-3):275–287.

Liu, H. H. and 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.

Lolo, F.-N., Walani, N., Seemann, E., Zalvidea, D., Pavón, D. M., Cojoc, G., Zamai, M., Viaris de Lesegno, C., Martínez de Benito, F., Sánchez-Álvarez, M., et al. (2023). Caveolin-1 dolines form a distinct and rapid caveolae-independent mechanoadaptation system. *Nature cell biology*, 25(1):120–133.

Lovejoy, S. and Schertzer, D. (1995). Multifractals and rain, in new uncertainty concepts in hydrology and water resources. *AW Kundzewicz*, pages 61–103.

Lu, Z. and Zhang, D. (2002). On stochastic modeling of flow in multimodal heterogeneous formations. *Water Resources Research*, 38(10):8–1–8–15.

Lucca, D., Shao, L., Wetteland, C., Misra, A., Klopfstein, M., and Nastasi, M. (2006). Subsurface damage in (1 0 0) znse introduced by mechanical polishing. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 249(1-2):907–910.

Lüttge, A., Arvidson, R. S., and Fischer, C. (2013a). A stochastic treatment of crystal dissolution kinetics. *Elements*, 9(3):183–188.

Lüttge, A., Arvidson, R. S., and Fischer, C. (2013b). A stochastic treatment of crystal dissolution kinetics. *Elements*, 9(3):183–188.

Lüttge, A., Arvidson, R. S., Fischer, C., and Kurganskaya, I. (2019). Kinetic concepts for quantitative prediction of fluid-solid interactions. *Chemical Geology*, 504:216–235.

MacInnis, I. N. and Brantley, S. L. (1992). The role of dislocations and surface morphology in calcite dissolution. *Geochimica et Cosmochimica Acta*, 56(3):1113–1126.

Mangane, P. O., Gouze, P., and Luquot, L. (2013). Permeability impairment of a limestone reservoir triggered by heterogeneous dissolution and particles migration during co2-rich injection. *Geophysical Research Letters*, 40(17):4614–4619.

Marinello, F., Carmignato, S., Voltan, A., Savio, E., and Chiffre, L. D. (2010). Error sources in atomic force microscopy for dimensional measurements: Taxonomy and modeling. *Journal of Manufacturing Science and Engineering*, 132.

Massaro, F., Pastero, L., Rubbo, M., and Aquilano, D. (2008). Theoretical surface morphology of $\{01\bar{1}2\}$ acute rhombohedron of calcite a comparison with experiments and $\{01\bar{1}4\}$ cleavage rhombohedron. *Journal of Crystal Growth*, 310(3):706–715.

Mazzoli, A. and 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.

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics.

Meakin, P. and Rosso, K. M. (2008). Simple kinetic monte carlo models for dissolution pitting induced by crystal defects. *The Journal of chemical physics*, 129(20):204106.

Meerschaert, M. M., Kozubowski, T. J., Molz, F. J., and Lu, S. (2004). Fractional laplace model for hydraulic conductivity. *Geophysical Research Letters*, 31(8).

Nagy, K. and Lasaga, A. (1992). Dissolution and precipitation kinetics of gibbsite at 80 c and ph 3: The dependence on solution saturation state. *Geochimica et Cosmochimica Acta*, 56(8):3093–3111.

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

Noiriel, C. and Daval, D. (2017). Pore-scale geochemical reactivity associated with co2 storage: new frontiers at the fluid–solid interface. *Accounts of chemical research*, 50(4):759–768.

Noiriel, C., Madé, B., and Gouze, P. (2007). Impact of coating development on the hydraulic and transport properties in argillaceous limestone fracture. *Water resources research*, 43(9).

Noiriel, C., Oursin, M., Saldi, G., and Haberthür, D. (2018). Direct determination of dissolution rates at crystal surfaces using 3d x-ray microtomography. *ACS Earth and Space Chemistry*, 3(1):100–108.

Painter, S. (1996). Evidence for non-gaussian scaling behavior in heterogeneous sedimentary formations. *Water Resources Research*, 32(5):1183–1195.

Paquette, J. and Reeder, R. J. (1995). Relationship between surface structure, growth mechanism, and trace element incorporation in calcite. *Geochimica et Cosmochimica Acta*, 59(4):735–749.

Pollet-Villard, M., Daval, D., Fritz, B., Knauss, K. G., Schäfer, G., and Ackerer, P. (2016). Influence of etch pit development on the surface area and dissolution kinetics of the orthoclase (001) surface. *Chemical Geology*, 447:79–92.

Putnis, C. V. and Putnis, A. (2022). A mechanism of ion exchange by interface-coupled dissolution-precipitation in the presence of an aqueous fluid. *Journal of Crystal Growth*, 600:126840.

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

Recalcati, C., Siena, M., Riva, M., Bollani, M., and Guadagnini, A. (2024). Stochastic assessment of dissolution at fluid-mineral interfaces. *Geophysical Research Letters*, 51(7):e2023GL108080.

Renard, F., Putnis, C. V., Montes-Hernandez, G., King, H. E., Breedveld, G. D., and Okkenhaug, G. (2018). Sequestration of antimony on calcite observed by time-resolved nanoscale imaging. *Environmental science & technology*, 52(1):107–113.

Renard, F., Røyne, A., and Putnis, C. V. (2019). Timescales of interface-coupled dissolution-precipitation reactions on carbonates. *Geoscience Frontiers*, 10:17–27.

Ricci, D. and Braga, P. C. (2004). Recognizing and avoiding artifacts in afm imaging. *Atomic Force Microscopy: Biomedical Methods and Applications*, pages 25–37.

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

Riva, M., Neuman, S. P., Guadagnini, A., and Siena, M. (2013). Anisotropic scaling of berea sandstone log air permeability statistics. *Vadose Zone Journal*, 12(3):1–15.

Riva, M., Panzeri, M., Guadagnini, A., and Neuman, S. P. (2015b). Simulation and analysis of scalable non-gaussian statistically anisotropic random functions. *Journal of Hydrology*, 531:88–95.

Rohlfs, R. D., Fischer, C., Kurganskaya, I., and Luttge, A. (2018). Crystal dissolution kinetics studied by a combination of monte carlo and voronoi methods. *Minerals*, 8(4):133.

Rubin, Y. (1995). Flow and transport in bimodal heterogeneous formations. *Water Resources Research*, 31(10):2461–2468.

Rubin, Y. and Journel, A. G. (1991). Simulation of non-Gaussian space random functions for modeling transport in groundwater. *Water Resources Research*, 27:1711–1721.

Ruiz-Agudo, E., Putnis, C., Jiménez-López, C., and Rodriguez-Navarro, C. (2009). An atomic force microscopy study of calcite dissolution in saline solutions: The role of magnesium ions. *Geochimica et Cosmochimica Acta*, 73(11):3201–3217.

Ruiz-Agudo, E. and Putnis, C. V. (2012). Direct observations of mineral-fluid reactions using atomic force microscopy: the specific example of calcite. *Mineralogical Magazine*, 76(1):227–253.

Ruiz-Agudo, E., Putnis, C. V., Wang, L., and Putnis, A. (2011). Specific effects of background electrolytes on the kinetics of step propagation during calcite growth. *Geochimica et Cosmochimica Acta*, 75(13):3803–3814.

Russo, D. (2002). Stochastic analysis of macrodispersion in gravity-dominated flow through bimodal heterogeneous unsaturated formations. *Water resources research*, 38(7):19–1.

Salamon, P., Fernàndez-Garcia, D., and Gómez-Hernández, J. (2007). Modeling tracer transport at the made site: The importance of heterogeneity. *Water Resources Research*, 43(8).

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall.

Schabernack, J. and Fischer, C. (2022). Improved kinetics for mineral dissolution reactions in pore-scale reactive transport modeling. *Geochimica et Cosmochimica Acta*, 334:99–118.

Schabernack, J. and Fischer, C. (2024). Pulsating dissolution of crystalline matter: A surface-controlled process. *Geochimica et Cosmochimica Acta*.

Schertzer, D. and Lovejoy, S. (1988). Multifractal simulations and analysis of clouds by multiplicative processes. *Atmospheric research*, 21(3-4):337–361.

Schmidt, W. U. and Alkire, R. C. (1994). Use of atomic force microscopy to image surfaces during fluid flow. *Journal of the Electrochemical Society*, 141(7):L85.

Siena, M., Bussetti, G., Recalcati, C., Riva, M., Duò, L., and 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.

Siena, M., Guadagnini, A., Bouissonnié, A., Ackerer, P., Daval, D., and Riva, M. (2020). Generalized sub-Gaussian processes: Theory and application to hydrogeological and geochemical data. *Water Resources Research*, 56.

Siena, M., Recalcati, C., Guadagnini, A., and Riva, M. (2023). A gaussian-mixture based stochastic framework for the interpretation of spatial heterogeneity in multimodal fields. *Journal of Hydrology*, 617:128849.

Siena, M., Riva, M., Giamberini, M., Gouze, P., and Guadagnini, A. (2019). Statistical modeling of gas-permeability spatial variability along a limestone core. *Spatial Statistics*, 34:100249.

Smith, M. E., Knauss, K. G., and Higgins, S. R. (2013). Effects of crystal orientation on the dissolution of calcite by chemical and microscopic analysis. *Chemical Geology*, 360-361:10–21.

Teng, H. H. (2004). Controls by saturation state on etch pit formation during calcite dissolution. *Geochimica et Cosmochimica Acta*, 68(2):253–262.

Tidwell, V. C. and Wilson, J. L. (1999). Upscaling experiments conducted on a block of volcanic tuff: results for a bimodal permeability distribution. *Water Resources Research*, 35:3375–3387.

Trindade Pedrosa, E., Kurganskaya, I., Fischer, C., and Luttge, A. (2019). A statistical approach for analysis of dissolution rates including surface morphology. *Minerals*, 9(8):458.

Veneziano, D., Langousis, A., and Furcolo, P. (2006). Multifractality and rainfall extremes: A review. *Water resources research*, 42(6).

Veneziano, D. and Yoon, S. (2013). Rainfall extremes, excesses, and intensity-duration-frequency curves: A unified asymptotic framework and new nonasymptotic results based on multifractal measures. *Water Resources Research*, 49(7):4320–4334.

Walton, W. (1948). Feret‘s statistical diameter as a measure of particle size. *Nature*, 162(4113):329–330.

Weber, F., Kowarik, I., and Säumel, I. (2014). Herbaceous plants as filters: Immobilization of particulates along urban street corridors. *Environmental pollution*, 186:234–240.

White, A. F. and Brantley, S. L. (2003). The effect of time on the weathering of silicate minerals: why do weathering rates differ in the laboratory and field? *Chemical Geology*, 202(3-4):479–506.

Winter, C., Tartakovsky, D., and Guadagnini, A. (2003). Moment differential equations for flow in highly heterogeneous porous media. *Surveys in Geophysics*, 24:81–106.

Yang, C.-Y., Hsu, K.-C., and Chen, K.-C. (2009). The use of the levy-stable distribution for geophysical data analysis. *Hydrogeology journal*, 17(5):1265.

Ye, M., Meyer, P. D., and Neuman, S. P. (2008). On model selection criteria in multimodel analysis. *Water Resources Research*, 44(3).

Zhang, Y. (2009). Hierarchical geostatistical analysis of an experimental stratigraphy. *Mathematical Geosciences*, 41(2):145–162.

Zhang, Y., Person, M., Paola, C., Gable, C. W., Wen, X.-H., and Davis, J. M. (2005). Geostatistical analysis of an experimental stratigraphy. *Water resources research*, 41(11).