



**HAL**  
open science

# Auto-weighted Bayesian Physics-Informed Neural Networks and robust estimations for multitask inverse problems in pore-scale imaging of dissolution

Sarah Perez, Philippe Poncet

► **To cite this version:**

Sarah Perez, Philippe Poncet. Auto-weighted Bayesian Physics-Informed Neural Networks and robust estimations for multitask inverse problems in pore-scale imaging of dissolution. *Computational Geosciences*, 2024, 28, pp.1175-1215. 10.1007/s10596-024-10313-x . hal-04191804v2

**HAL Id: hal-04191804**

**<https://hal.science/hal-04191804v2>**

Submitted on 20 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Auto-weighted Bayesian Physics-Informed Neural Networks and robust estimations for multitask inverse problems in pore-scale imaging of dissolution\*

Sarah Perez<sup>†,\*</sup> and Philippe Poncet<sup>†</sup>

<sup>†</sup>Universite de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP UMR CNRS-UPPA 5142, Pau, France.

\*The Lyell Centre, Heriot-Watt University, Edinburgh, United Kingdom.

Received: August 28, 2023 / Accepted: July 25, 2024

## Abstract

In this article, we present a novel data assimilation strategy in pore-scale imaging and demonstrate that this makes it possible to robustly address reactive inverse problems incorporating Uncertainty Quantification (UQ). Pore-scale modeling of reactive flow offers a valuable opportunity to investigate the evolution of macro-scale properties subject to dynamic processes in the context of Carbon Capture and Storage (CCS). Yet, they suffer from imaging limitations arising from the associated X-ray microtomography (X-ray  $\mu$ CT) process, which induces discrepancies in the properties estimates. Assessment of the kinetic parameters also raises challenges, as reactive coefficients are critical parameters that can cover a wide range of values. We account for these two issues and ensure reliable calibration of pore-scale modeling, based on dynamical  $\mu$ CT images, by integrating uncertainty quantification in the workflow.

The present method is based on a multitasking formulation of reactive inverse problems combining data-driven and physics-informed techniques in calcite dissolution. This allows quantifying morphological uncertainties on the porosity field and estimating reactive parameter ranges through prescribed PDE models, with a latent concentration field, and dynamical  $\mu$ CT observations. The data assimilation strategy relies on sequential reinforcement incorporating successively additional PDE constraints and suitable formulation of the heterogeneous diffusion differential operator leading to enhanced computational efficiency. We provide a robust and unbiased uncertainty quantification by straightforward adaptive weighting of Bayesian Physics-Informed Neural Networks (BPINNs), ensuring reliable micro-porosity changes during geochemical transformations.

We demonstrate successful Bayesian Inference in 1D+Time calcite dissolution based on synthetic  $\mu$ CT images with meaningful posterior distribution on the reactive parameters and dimensionless numbers. We eventually apply this framework to a more realistic 2D+Time data assimilation problem involving heterogeneous porosity levels and synthetic  $\mu$ CT dynamical observations.

**Keywords:** Hamiltonian Monte Carlo, Uncertainty Quantification, Multi-objective training, Imaging inverse problem, Pore-scale porous media, Artificial Intelligence, Bayesian Physics-Informed Neural Networks

**Funding:** This work was partially supported by E2S UPPA and by the Carnot Institute ISIFoR project MicroMineral under Grant Agreement P450902ISI, and by ANR project 20-CE45-0022 MucoReaDy.

## Article Highlights:

- The method presented can carry out the porosity and concentration fields (observable or latent) in pore-scale reactive flows with evolving fluid-solid interface, together with characteristic numbers (Damköhler, molecular diffusion, chemical coefficients and tortuosity index).
- The multitask formulation of such inverse problems is based on an automatic weighting of Bayesian Physics-Informed Neural Networks (BPINNs) and involves naturally uncertainty quantification.
- The inverse problems of dissolution 1D+Time and 2D+Time at the pore scale are provided.

\*This version of the article has been accepted for publication in Computational Geosciences, after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://dx.doi.org/10.1007/s10596-024-10313-x>

# 1 Introduction

Studying reactive flows in porous media is essential to manage the geochemical effects arising from CO<sub>2</sub> capture and storage in natural underground reservoirs. While long-term predictions are commonly modeled at the field scale [20], pore-scale approaches meanwhile provide insights into local geochemical interactions between the injected CO<sub>2</sub> and the aquifer structure [54]. Through mathematical homogenization of the sub-micrometer porous medium and appropriate modeling, one can simulate the reactive processes that occur at the pore scale and predict their impact on the macro-scale properties [4, 5]. Geochemical processes are critical components for understanding the mineral trapping mechanisms and local evolving interfaces, either due to precipitation, crystallization, or dissolution within the porous environment. In this sense, investigating the impact of such reactive processes provides insight into reservoir safety submitted to chemical interactions that may compromise the aquifer structure. Pore-scale modeling of reactive flow hence appears as a complementary mean to field scale studies wherein homogenization theory bridges the gap between these scales.

Pore-scale modeling in porous media is intrinsically related to X-ray microtomography (X-ray  $\mu$ CT) experiments. Advances in this imaging technique coupled with efficient numerical simulation offer a valuable opportunity to investigate dynamical processes and study the evolving macro-scale properties, such as the upscaled porosity and permeability [10, 49]. This is of great importance in the risk management perspective of CO<sub>2</sub> storage, and therefore ensuring the reliability of pore-scale modeling and simulation appears as crucial. Uncertainties, however, arise from the microtomography imaging process where artifacts, noise, and unresolved morphological features are intrinsic limitations inducing important deviations in the estimation of petrophysical properties [56, 19]. In particular, quantifying the impact of sub-resolved porosity in  $\mu$ CT images is identified as critical for geosciences applications [35, 65]. This limiting factor arises from the compromise between the field of view being investigated and the image resolution. For multi-scale porous media such as carbonate rocks, this trade-off can readily result in scan resolutions that do not fully resolve morphological features of the pore space. Intrinsic limiting factors remain in the X-ray  $\mu$ CT imaging process, and investigating their effects and related uncertainties is fundamental to developing more accurate predictive models at the pore scale.

In addition to these imaging uncertainties, proper assessment of the kinetic parameters raises challenges in the pore-scale modeling of reactive flows. Mineral reactivities, including reactive surface area, are critical parameters to account though they commonly suffer from discrepancies of several orders of magnitude [51]. Providing uncertainty estimates on these kinetic parameters is essential to ensure reliable calibration of pore-scale models for CO<sub>2</sub> mineral storage assessment. Unsuitable characterization of the reactive surface area, for instance, will considerably affect the numerical model generating highly distinct behaviors that can become inconsistent with experimental investigations. Such concern is widely known, and several experimental works have developed potential solutions that address dynamical  $\mu$ CT imaging processes of carbonate dissolution [52, 43]. This relies on 4D  $\mu$ CT and differential imaging techniques to derive averaged reaction rates and provide local maps of mineral reactivity at the porous medium surface. However, dynamical  $\mu$ CT scans also suffer from trade-off issues that may disrupt the identification of these parameters [81]. In addition to potential sub-resolved porosity, one needs to consider the compromise between the acquisition time capturing the dynamical process and the image quality. This may result in noisy observation data or non-physical variations leading to misleading estimations of the kinetic parameters. Querying the reliability of reactive parameters involved in pore-scale modeling is crucial, and time-resolved experiments of dynamical processes offer such an opportunity while suffering from imaging limitations.

Overall, we identify two current challenges to address reliable pore-scale modeling of reactive flows based on  $\mu$ CT images and ensure trustable evolutions of the macro-scale properties. The first challenge aims at quantifying morphological uncertainties on the porous medium sample due to unresolved features resulting from X-ray  $\mu$ CT. Investigating the uncertainties in the micro-porosity field is a major concern, and neglecting these uncertainty effects can bias the determination of the evolving petrophysical properties in geological applications. The second challenge concerns the uncertainty quantification of the kinetic parameters for reactive processes. In this sense, providing reliable mineral reactivity from dynamical  $\mu$ CT remains critical in order to perform relevant direct numerical simulation at the pore scale. The present article addresses these two challenges and incorporates Uncertainty Quantification (UQ) concerns in the workflow

of pore-scale modeling.

Accounting for these concerns, however, requires developing efficient data assimilation techniques to perform extensive parameter estimation studies, uncertainty quantification assessments, and improve model reliability. In fact, uncertainty quantification is commonly achieved through stochastic PDE models [29, 22] or probabilistic Markov Chain Monte Carlo (MCMC) methods embedding Bayesian inference [48, 64]. The main drawback being this requires numerous evaluations of the PDE model and can thus quickly become computationally expensive. To overcome such computational constraints, machine learning methods have appeared as a popular framework in geosciences and have shown effectiveness in building efficient surrogate models in PDE-based data assimilation problems [72, 23]. This offers alternatives and complementary means to traditional numerical methods to improve predictive modeling based on observation data and investigate uncertainty quantification within a Bayesian context. The development of machine-learning surrogate modeling incorporating uncertainty has, therefore, garnered increasing interest for a wide range of scientific applications [24, 22].

A popular framework combining physics-based techniques, data-driven methodology, and intrinsic uncertainty quantification are Bayesian Physics-Informed Neural Networks (BPINNs) [79]. This benefits from the advantages of neural network structures in building parameterized surrogate models and Bayesian inference standards in estimating probabilistic posterior distribution. BPINNs can, however, be prone to a range of pathological behaviors, especially in multi-objective and multiscale inverse problems. This is because their training amounts to sample from a weighted multitask posterior distribution for which the setting of the weights parameters is challenging. Ensuring robust Bayesian inference, indeed, hinges on properly estimating these distinct task weights. We thus rely on the efficient BPINNs framework developed in [55], which robustly addresses multi-objective and multiscale Bayesian inverse problems including latent field reconstruction. The strategy relies on an adaptive and automatic weighting of the target distribution parameters and objectives, which is an unlocking advance in bringing robustness and addressing more challenging problems within the BPINNs framework. Indeed, it benefits from enhanced convergence and stability compared to conventional formulations and reduces sampling bias by avoiding manual tuning of critical weighting parameters [39]. The adjusted weights bring information on the task uncertainties, improve the reliability of the noise-related and model adequacy estimates and ensure unbiased uncertainty quantification. All these characteristics are crucial to address reliable reactive inverse problems of calcite dissolution, and we thus built the present methodology upon this efficient data-assimilation framework.

In this article, we focus on a multitask inverse problem for reactive flows at the pore scale through data assimilation that incorporates uncertainty quantification by means of the Bayesian Physics-Informed Neural Networks framework presented in [55]. We intend to develop a novel approach for pore-scale imaging problems that combines dynamical microtomography and physics-based regularization induced by the PDE model of dissolution processes, for which the images are substantially noisy. To the best of our knowledge, investigating morphological and mineral reactivity uncertainties from the perspective of coupling physics-based models with data-driven techniques is the main novelty of this work. This formulation presents the joint ability to infer altogether kinetic parameters and quantify the residual micro-porosity generated by unresolved features in the  $\mu$ CT images. Overall, we aim to ensure reliable calibration of the PDE model and account for the morphological imaging uncertainty to provide meaningful evolution of the petrophysical properties due to the reactive process.

The present methodology relies on sequential reinforcement of the target posterior distribution, which successively incorporates additional constraints from the PDE model into the data assimilation process. This sequential splitting formulation arises from the strong coupling, in the reactive model, between the micro-porosity field related to the  $\mu$ CT observations and the solute concentration, which is a latent field. In particular, achieving the reconstruction of an adequate latent concentration for the reactive fluid field is crucial for the identification of the inverse parameters along with their uncertainty ranges, and requires the development of a dedicated sequential reinforcement approach. Therefore, we consider successive sampling steps dedicated to 1) preconditioning the micro-porosity surrogate model with pure regression on the dynamical  $\mu$ CT images, 2) preconditioning the latent reactive fluid and inferring a first reactive parameter through PDE-constrained tasks given a first modeling approximation of the coupled PDE problem, and 3) considering the overall data assimilation problem given the exact PDE model with two inverse parameters, a predictive posterior distribution on the micro-porosity and insight on the latent concentration field. In this

sense, while the current strategy draws significant inspiration from the PINNs framework, in which the regression of physical fields typically precedes the incorporation of physics-based constraints, the sequential reinforcement approach presented in this article also introduces a modeling reinforcement which is essential for inferring latent fields in a strongly coupled PDE system. On top of that, we propose a differentiation strategy wherein we consider a reformulation of the heterogeneous diffusion differential operator involved in the PDE model. This enhances the computational efficiency of the BPINN surrogate model and shows that suitable differential operator expressions considerably improve the computational cost, especially when dealing with complex non-linear operators.

The main contributions of this article are summarized below:

1. We infer reactive inverse parameter uncertainty ranges in prescribed PDE models through suitable dimensionless formulations in inverse problems, for which we identify and define the corresponding dimensionless numbers.
2. We quantify morphological uncertainties from a pore-scale perspective, coupling image-based and physics-informed techniques in dynamical dissolution processes.
3. We improve the relevance and reliability of predictions in dynamical systems through data-driven approaches and robust Bayesian Inference methodology.
4. We provide reliable quantification of the micro-porosity changes during geochemical transformations, with a focus on calcite dissolution processes.
5. We built an intrinsic data assimilation strategy for pore-scale imaging inverse problems relying on a sequential reinforcement approach and suitable formulation of the heterogeneous diffusion differential operator.

The remainder of this manuscript is organized as follows: In Sect. 2, we review the current challenges arising from uncertainty quantification concerns in pore-scale modeling of reactive flows, with a focus on  $\mu$ CT limitations and model reliability issues. We focus in Sect. 2.1 on the formulation of pore-scale modeling of reactive flows that we consider to study the dynamical dissolution of calcite. Sect. 3 describes the dimensionless expressions of the dissolution PDE model for direct and inverse problems. We identify the main differences in their formulations and we establish in Sect. 3.3 the dimensionless inverse problem on calcite dissolution that we address in the data assimilation approach, which ends up with equation (19). Sect. 4 is dedicated to presenting the efficient adaptive framework for Bayesian Physics-Informed Neural Networks, which has been developed in our previous work. In Sect. 5, we describe the proposed data assimilation strategy for pore-scale imaging inverse problems, with sequential reinforcement of the target posterior distribution and computational strategy for the differential operator expressions. We validate this strategy in Sect. 6 on several 1D+Time test cases of calcite dissolution based on synthetic  $\mu$ CT images. This particularly demonstrates successful Bayesian inference of the reactive parameters with posterior distributions on the dimensionless numbers. This also highlights consistent UQ on the micro-porosity field with uncertainty ranges on the residual micro-porosity, potentially unresolved, arising from the  $\mu$ CT dynamical images. Finally, we apply in Sect. 7 our methodology to a more realistic 2D+Time data assimilation problem of calcite dissolution with heterogeneous porosity levels and synthetic  $\mu$ CT dynamical observations.

## 2 Uncertainty Quantification in pore-scale modeling of reactive flows: context and motivation

Pore-scale modeling of reactive flows plays a crucial role in the long-term management of CO<sub>2</sub> capture and storage in natural underground reservoirs. Understanding the local geochemical interactions between the injected CO<sub>2</sub> and the aquifer structure and how it impacts the reservoir macro-scale properties is an active field in porous media research [34, 41, 70, 8]. These geochemical effects include mineral trapping through precipitation and crystallization but also dissolution reactions associated with flow, and transport mechanisms [54, 2]. Mathematical models of such processes, at the pore scale, are usually combined with Direct Numerical Simulations (DNS) of highly coupled and non-linear Partial Differential Equations (PDE). Such

PDE systems characterize the local evolving interfaces and provide insight into reservoir safety submitted to chemical interactions [43, 45, 53]. In this risk management perspective, ensuring the reliability of pore-scale modeling and simulation of reactive flows is therefore essential, and this requires embedding Uncertainty Quantification (UQ) concerns.

## 2.1 Modeling of pore-scale dissolution

The study of geochemical processes related to CO<sub>2</sub> capture and storage is crucial in the context of risk management and investigation of the coupled mechanisms occurring within aquifers. In particular, the dissolution of the carbonate rock architecture by the injected CO<sub>2</sub> may compromise the integrity of the geological reservoir. Pore-scale modeling of dissolution phenomena in porous media, therefore, remains an extensive research area [68, 44]. These mathematical models require a thin description of the highly heterogeneous pore structure in order to account for local interactions. The present article focus on the pore-scale dissolution of calcite subject to acidic transport in the subsoil. We, therefore, target the following irreversible chemical reaction with uniform stoichiometric coefficients:



In this section, we present the mathematical model used to simulate the calcite dissolution process (1) at the pore scale. We introduce a spatial domain  $\Omega \subset \mathbb{R}^n$ ,  $n = 1, 2, 3$  which corresponds to the porous medium described at its pore scale. This sample description involves a pure fluid region  $\Omega_F$ , also called void-space and assumed to be a smooth connected open set, and a surrounding solid matrix  $\Omega_S$  itself considered as a porous region. This region  $\Omega_S = \Omega \setminus \Omega_F$  is seen as complementing the full domain  $\Omega$ , which in practice represents the computational box of the numerical simulations, and the internal fluid/solid interface is denoted  $\Sigma$ . We denote by  $\varepsilon = \varepsilon_f = 1 - \varepsilon_s$  the micro-porosity field defined on  $\Omega$ , given  $\varepsilon_f$  and  $\varepsilon_s$  respectively the volume fractions of void and solid according to usual notations [67, 26]. This defines a micro continuum description of the porous medium such that  $\varepsilon = 1$  in the pure fluid region  $\Omega_F$  and takes a small value in the surrounding matrix  $\Omega_S$ . In fact, the local micro-porosity  $\varepsilon$  is assumed to have a strictly positive lower bound  $\varepsilon(x, t) \geq \varepsilon_0 > 0$  for all  $(x, t)$  in the spatiotemporal domain  $\Omega \times (0, T_f)$ . This lower bound  $\varepsilon_0$  characterizes the residual, potentially unresolved, porosity of the porous matrix. In practice, we set throughout this article  $\varepsilon_0 = 5\%$ . This micro-continuum formulation relies on a two-scale representation of the sample characterized by its micro-porosity field  $\varepsilon$ .

Such a two-scale description of the local heterogeneities in the carbonate rocks is appropriate to simulate the pore-scale physics and establish the governing flow and transport equations in each distinct region. Indeed, we consider the model on superficial velocity  $u$  introduced and derived rigorously by Quintard and Whitaker in the late 80s [58] and commonly used until nowadays [32, 76, 67, 44]:

$$\varepsilon^{-1} \frac{\partial \rho u}{\partial t} + \varepsilon^{-1} \nabla \cdot (\varepsilon^{-1} \rho u \otimes u) - \varepsilon^{-1} \nabla \cdot (2\mu D(u)) + \mu^* K_\varepsilon^{-1} u = f - \nabla p \quad (2)$$

along with the divergence-free condition  $\nabla \cdot u = 0$ . In this equation,  $D(u) = (\nabla u + \nabla u^T)/2$  is the shear-rate tensor,  $\mu$  is the dynamic viscosity,  $p$  is the volumic pressure,  $f$  the volumic driving force and  $\rho$  the fluid density. The related viscosity  $\mu^*$  coincides usually with the fluid viscosity  $\mu$  but may be different in order to account for viscous deviations. The quantities  $\rho$ ,  $\mu$ ,  $\mu^*$  and  $f$  are assumed to be constant. In contrast, the permeability  $K_\varepsilon$  refers to the micro-scale permeability and depends on the local micro-porosity field  $\varepsilon$ . In fact, the permeability of the micro-porous domain is modeled by the empirical Kozeny-Carman relationship [31, 17, 18]:

$$K_\varepsilon^{-1} = \kappa_0^{-1} \frac{(1 - \varepsilon)^2}{\varepsilon^3} \quad (3)$$

where  $\kappa_0$  is a coarse estimation of the reference macro-scale permeability. In this article, we consider both  $K_\varepsilon$  and  $\kappa_0$  as scalars, meaning we restrict ourselves to the isotropic case although this formalism can be extended to anisotropic porous media. The superficial velocity formulation (2) defines a two-scale model that can be solved on the overall domain  $\Omega$  — using for instance penalization principles — and retrieves the usual Navier-Stokes equation in the pure fluid region  $\Omega_F$  (since  $K_\varepsilon^{-1} = 0$  for  $\varepsilon = 1$ ). At low Reynolds

numbers and for highly viscous Darcian flows, equation (2) reduces to the following Darcy-Brinkman Stokes (DBS) model:

$$-\nabla \cdot (2\mu D(u)) + \mu\kappa_0^{-1} \frac{(1-\varepsilon)^2}{\varepsilon^2} u = \varepsilon(f - \nabla p), \quad \text{in } \Omega \quad (4)$$

where  $\mu^* = \mu$  for sake of readability. In the present work, we consider this DBS equation (4), which is adequate in the flow regime hypothesis of low Reynolds number representative in pore-scale modeling. The DBS equation based on the superficial velocity is an efficient formalism to model the hydrodynamic in multi-scale porous media.

The flow model (4) needs to be complemented by transport-reaction-diffusion equations of the different species involved in the geochemical processes. These equations are derived from the mass balance of the chemical species [67], and can be written under the form:

$$\frac{\partial \varepsilon \tilde{C}_k}{\partial t} + \nabla \cdot (u \tilde{C}_k) - \nabla \cdot (\alpha_k(\varepsilon) \varepsilon \nabla \tilde{C}_k) = \tilde{R}(\tilde{C}_k), \quad (5)$$

where  $\tilde{C}_k = \rho_f \bar{\omega}_{f,k} / M_k$  is a concentration per unit of fluid (following the notations introduced by Quintard and Whitaker in [59], and afterward by Soulaïne and al. in [67]) with  $M_k$  the molar mass of the  $k^{\text{th}}$  specie. The term  $\alpha_k(\varepsilon)$  is a space-variable effective diffusion coefficient and accounts for a reduced diffusion in the surrounding porous matrix due to the tortuosity effect, which is usually quantified using Archie's law [11]:

$$\alpha_k(\varepsilon) = D_{m,k} \varepsilon^\beta. \quad (6)$$

In this empirical relationship,  $\beta$  refers to the tortuosity index and  $D_{m,k}$  to the molecular diffusion of the considered species [71]. We finally introduce the concentration per unit of volume defined by  $C_k = \varepsilon \tilde{C}_k$ , so that the equation (5) is written

$$\frac{\partial C_k}{\partial t} + \nabla \cdot (\varepsilon^{-1} u C_k) - \nabla \cdot (D_{m,k} \varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C_k)) = R(C_k), \quad (7)$$

which is no more than superficial modeling of the chemistry. In the context of the current article, we are not interested in monitoring the dissolution products of (1) (*i.e.* the  $\text{Ca}^{2+}$  and  $\text{HCO}_3^-$  ions), hence we mainly focus on the concentrations of the acid and calcium carbonate species respectively denoted  $C(x, t) = [\text{H}^+]$  and  $C_s(x, t) = [\text{CaCO}_3]$ . The solid concentration is linked to the porosity through the molar volume of calcite  $v$  by the relation  $C_s = (1 - \varepsilon)/v$  with  $v = 36.93 \text{ cm}^3 \cdot \text{mol}^{-1}$ . In this configuration, the evolution of the acid phase (*i.e.* the concentration field  $C$ ) follows the equation (7), and the evolution of the solid phase with superficial concentration  $C_s$  is given by the same equation without transport nor diffusion:

$$\frac{\partial C_s}{\partial t} = R(C). \quad (8)$$

This reaction rate — related to the chemical reaction (1) — is written [46]:

$$R(C) = -K_s A_s \gamma_{\text{H}^+} C \mathbb{1}_{\{(1-\varepsilon)>0\}} \quad (9)$$

where  $K_s$  is the dissolution rate constant,  $A_s$  the specific reactive area, and  $\gamma_{\text{H}^+}$  the activity coefficient of the acid, whose physical units are respectively  $\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ,  $\text{m}^{-1}$  and  $\text{m}^3 \cdot \text{mol}^{-1}$  (such that the chemical activity  $a_{\text{H}^+} = \gamma_{\text{H}^+} C$  is dimensionless). The notation  $\mathbb{1}$  refers to a characteristic or activation function and ensures the rate of the chemical reaction is non-zero only in the presence of solid minerals. Along with its boundary and initial conditions, this defines a set of partial differential equations modeling reactive flows at the pore scale [69, 46]:

$$\left\{ \begin{array}{ll} -\nabla \cdot (2\mu D(u)) + \mu\kappa_0^{-1} \frac{(1-\varepsilon)^2}{\varepsilon^2} u = \varepsilon(f - \nabla p), & \text{in } \Omega \times (0, T_f) \\ \frac{\partial C}{\partial t} + \nabla \cdot (\varepsilon^{-1} u C) - \nabla \cdot (D_m \varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C)) = -K_s A_s \gamma_{\text{H}^+} C \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega \times (0, T_f) \\ \frac{\partial \varepsilon}{\partial t} = v K_s A_s \gamma_{\text{H}^+} C \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega \times (0, T_f) \\ + \text{adequate boundary and initial conditions, along with } \nabla \cdot u = 0 & \end{array} \right. \quad (10)$$

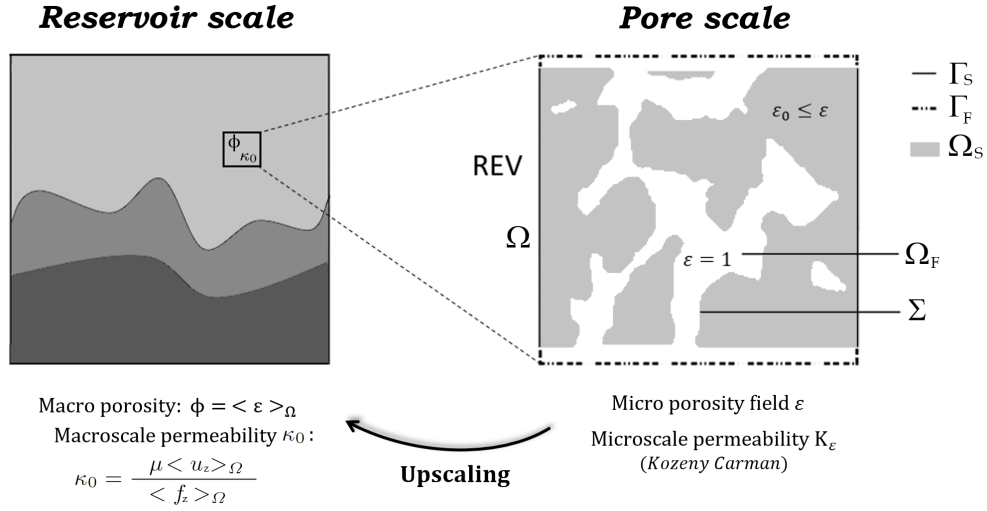


Figure 1: **From the pore-scale to the reservoir scale: an upscaling principle.** Schematic representation of a reservoir scale structure, on the left, with its inherent averaged isotropic macro-properties  $\phi$  and  $\kappa_0$  computed on a representative elementary volume (REV). Local description of the pore-scale heterogeneities in this REV, on the right, along with its intrinsic micro-scale properties. These are the local micro porosity field  $\varepsilon$  (bounded by the least physically possible porosity  $\varepsilon_0 > 0$ ) and the microscale permeability  $K_{\varepsilon}$ , based on the Kozeny-Carman relationship from equation (3).

which is strongly coupled, since  $u$  and  $C$  (by means of  $\varepsilon$ ) depend on each other. Finally, one can notice that the reactive system (10) is valid on the whole domain  $\Omega$ , whether the local state is fluid or not. In the pure fluid region, this system indeed converges toward a Stokes hydrodynamic model coupled with a standard transport-diffusion equation for the acid, with its molecular diffusion  $D_m$ . The overall system (10) defines the direct formulation of the calcite dissolution problem, following the chemical equation (1), at the pore-scale.

Nonetheless, appropriate model calibration of the kinetic input parameters, such as the specific surface area  $A_s$  or the dissolution rate constant  $K_s$ , that compare with experimental results remains challenging. This comes from the observation these reactive constants can span over a wide range of orders of magnitude, inducing highly different behaviors in the system. Quantifying the uncertainties on these kinetic parameters thereby appears as a necessity to provide reliable reactive flow models at the pore scale.

## 2.2 X-ray microtomography limitations: toward the uncertainty quantification assessment

Independently of the modeling aspects developed in Sect. 2.1 and the choice of the numerical method used as a direct solver, pore-scale simulations are intrinsically related to X-ray microtomography (X-ray  $\mu$ CT). In fact, the latter provides, beforehand, scans of the complex shape geometry on a representative elementary volume (REV), defined as the characteristic minimal volume on which the microscopic variables can be averaged [12]. Pore-scale numerical simulations of reactive dynamical processes are then performed on this REV initial geometry — which defines the domain  $\Omega$  — tracking the dynamical interface evolutions and micro-properties changes. This REV concept also allows the passage from the pore scale to the Darcy scale by referring to representative criteria of the domain in terms of averaged properties, such as the macro-porosity  $\phi$ , bulk permeability  $\kappa_0$ , and the reactive surface area  $A_s$ . These bulk parameters are derived from the evolving micro-structures through homogenization principles, and upscaling of the governing equations [60, 73], as illustrated in Fig 1. Indeed, at the Darcy scale of an isotropic material, the upscaled porosity and the scalar absolute permeability  $\kappa_0$  are respectively defined by:

$$\phi = \langle \varepsilon \rangle_{\Omega} \quad \text{and} \quad \kappa_0 = \frac{\mu \langle u_z \rangle_{\Omega}}{\langle f_z \rangle_{\Omega}} \simeq \frac{\mu \phi \langle u_z \rangle_{\Omega_F}}{\langle f_z \rangle_{\Omega_F}}, \quad (11)$$

using the notations introduced in Sect. 2.1, where  $u_z$  and  $f_z$  represent the vertical components of the velocity and driving force and where  $\langle \cdot \rangle$  denotes the average on the corresponding domain. Therefore,



trustable measurement of the impact of the reactive processes on the porous medium macro-properties requires ensuring reliable quantification of the changes in micro-properties. This can be achieved under the constraint of having a fine description of the pore space, with correct knowledge of the surrounding solid matrix defined by the local micro-porosity field  $\varepsilon$ . An efficient representation of the porous sample at the pore scale is necessary to guarantee reliable estimation of the macro-properties evolutions along the reactive processes.

Advances in X-ray microtomography offer such an opportunity. X-ray  $\mu$ CT is regarded as a powerful high-resolution imaging technique able to non-destructively determine the inner structure of a porous sample up to a characteristic scale, which defines the voxel size. The voxels are small elementary volumes (of a few  $\mu\text{m}$ ) that compose the overall 3D reconstructed sample geometry and are identified by different grey levels characterizing the local attenuation of the material. The resulting dataset can either be segmented to separate the pore space (fluid phase) from the surrounding solid matrix or benefit from the information related to the greyscale values of the different voxels. The segmented images lend themselves to numerical simulations that require an explicit representation of the fluid-solid interfaces (*e.g.* Lattice-Boltzmann [1]), unlike the Darcy-Brinkman Stokes formulation presented in Sect. 2.1, which incorporates the voxel greyscale values. Indeed, these grey-level shades, depicting the material local attenuation, are correlated to the porosity field description  $\varepsilon$  and can be taken into account in the DBS model through the equation (4). This introduces Digital Rock Physics applications as the joint use of high-resolution X-ray computed microtomography and advanced simulation techniques to characterize, *inter alia*, the rock petrophysical properties and their evolutions [9, 10]. Pure imaging alternatives readily regard the resulting dataset to derive the sample's effective physical properties (porosity, permeability, dispersivity...) but also geochemical rates and mineral reactivity in dynamical processes [65, 51]. Therefore, X-ray microtomography is both a complementary means to numerical modeling at the pore scale and a fundamental imaging process on its own to study the CO<sub>2</sub> storage implications on porous material.

However, limitations in the  $\mu$ CT imaging process may affect the determination of medium effective properties, and query the reliability of the predictive models based on these inputs. In fact, several imaging artifacts exist and disrupt the efficient description of the pore space morphology. Firstly, the finite resolution of the  $\mu$ CT pipeline is challenging, as the interfaces appear blurry and do not manifest themselves as sharp intensity steps in the images, but rather as gradual intensity changes spanning over several voxels [62]. Actually, the local attenuation signal within a voxel is influenced by the material heterogeneity in its neighborhood, then the resulting grey scale value represents averaged properties: this is known as the partial volume effect [50]. This phenomenon is also involved when morphological features of interest are smaller than the characteristic voxel size, resulting in unresolved micro-porosity or roughness of the pore space walls. Quantifying sub-resolution porosity, which is a prevalent imaging artifact in  $\mu$ CT, and measuring its impact on numerical modeling and simulation is identified as critical for geosciences applications [35, 65, 19]. Such an issue is well-known and arises from a compromise between the sample volume being investigated and the scan resolution. For porous media covering a wide range of pore scales, this trade-off can readily result in voxel sizes that are not able to capture fully resolved morphological features of the pore space. Finally, in the presence of sharp density transitions, the different refraction index at either side of the interface furthermore leads to so-called edge enhancement which manifests itself as an over- and under-shoot of the grey level immediately next to the interface [14]. Consequently, the position of the material interface is prone to uncertainty, in addition to the roughness of the pore space walls, and therefore results in an approximation of the true morphology.

While the mentioned effects can be minimized, they cannot be eliminated and add uncertainties to the estimation of the effective properties, the characterization of the void/solid interfaces, and the reliability of the numerical models. In addition, the accuracy of X-ray  $\mu$ CT images is challenged by additional artifacts coming from both inherent physical and technical limitations [30]. It includes, among them, instrumental noise, beam hardening [74] which results in cupping (an underestimation of the attenuation at the center of the object compared to its edges) or drag/streak appearances (due to an underestimation between two areas of high attenuation), beam fluctuations along the scanning process and scatter radiations coming from the object and/or the detector. These variations can manifest as noise, ring or streak artifacts, and halos that are often hard to distinguish from real features and therefore hinder the identification of sample heterogeneities at multiple scales. Ubiquitous limiting factors remain in the X-ray  $\mu$ CT imaging process, and the assessment

of their related uncertainties is fundamental to developing more accurate predictive models.

### 2.3 Dynamical microtomography: mineral reactivity and imaging morphological uncertainties

Accounting for the  $\mu$ CT morphological uncertainties and sub-resolution porosity, introduced in Sect. 2.2, is essential in providing reliable pore-scale simulations of reactive flows. This is of primary importance when considering risk assessment and predicting meaningful evolutions of the rock macro-properties under geochemical effects. The study of these overall X-ray imaging limitations, therefore, raises concerns in the research community, and investigations are conducted on quantifying their implications on the effective properties. In fact, sub-resolution porosity may lead to a misleading estimation of the pore-space connectivity that disrupts the flow description within the REV and induces significant deviations in the computed permeability. Several modeling approaches, mainly based on upscaling principles, aim at quantifying these deviations. They cover DBS formulation altogether with the Kozeny-Carman equation (3), which estimates the permeability of the micro-porous domain through a heuristic relation with the residual micro-porosity [66]. However, in the absence of prior knowledge of this unresolved residual porosity, the setting of the micro-porous permeability becomes controversial. Alternatives rely on appropriate boundary conditions to model the unresolved features and wall roughness through slip-length formalism, and range from theoretical implications [32, 33] to the practical computation of the permeability deviations on real 3D  $\mu$ CT scans [56]. Apart from the modeling quantification of the effective properties uncertainties, experimental and imaging approaches are developed to resolve the sub-resolution porosity. This involves differential imaging techniques based on comparisons between several enhanced contrast scans [35], statistical studies based on  $\mu$ CT histograms [82], or deep learning methodologies such as Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) that provide super-resolved segmented images [6, 80]. Overall, uncertainty quantification of the X-ray  $\mu$ CT limitations either relies on appropriate mathematical modeling with the estimation of computed deviations or experimental approaches based on image treatment analysis of the  $\mu$ CT scans.

The reliability of pore-scale modeling related to X-ray  $\mu$ CT scans is questioned due to its inherent imaging limitations and morphological uncertainties. At the same time, proper assessment of the kinetic parameters in dynamic phenomena, including mineral reactivity and reactive surface area, also raise challenges. Actually, mineral reactivity is a critical parameter to account for in many geosciences applications though discrepancies of several orders of magnitude can be found in the literature [16, 25]. However, these parameters are usually regarded as input in the numerical models and eventually tuned to aggregate experimental results. Providing reliable uncertainty estimates on these kinetic parameters is, therefore, of great interest to provide trustable pore-scale reactive simulations. Such concern has received attention over the past decades, and considering dynamic imaging processes subsequently appears as a necessity. Several experimental works have already focused on 4D imaging techniques of carbonate dissolution to provide fundamental information on mineral reaction rates [52]. These kinetic characterization studies mostly rely on voxel-to-voxel subtraction of consecutive images in order to quantify the change of greyscale values, hence the evolution of the dissolution process and calcite retreat. This is referred to, in the literature, as differential imaging techniques and has been investigated for different imaging techniques such as X-ray  $\mu$ CT and atomic force microscopy (AFM) [35, 63]. These approaches enable to capture heterogeneous spatial distributions of calcite dissolution rates through successive real-time measurements. They provide local maps of mineral reactivity at the crystal surfaces and quantification of their morphological evolutions [63, 51]. Menke et al. [43] also performed *in situ* time-resolved experiments of carbonate dissolution under reservoir conditions (in terms of pressure and temperature) to derive averaged reaction rates and evaluate dynamical changes in the effective properties. Investigation of mineral surface reactivity is another challenging concern to ensure reliable calibration of pore-scale models for CO<sub>2</sub> dissolution, and is usually achieved through dynamical  $\mu$ CT experiments.

Nonetheless, dealing with dynamical  $\mu$ CT images brings its own challenges [81]. In addition to the unresolved features, dynamical imaging of chemical processes requires making a compromise between the acquisition time and the image quality. Indeed, capturing fast-dissolution processes, for instance, imposes short acquisition times and could result in highly noisy data since statistically, the number of photons reach-

ing the detector would be reduced. In such a case, differential imaging makes it difficult to distinguish between true morphological changes and the derivation of highly noisy data. On top of that, any additional movement in the sample, not related to the dissolution process but rather resulting from instrumentation artifacts, makes it challenging to work with dynamical samples only to characterize  $\mu$ CT errors and uncertainty. Indeed, Zhang and al. [81] identified on a Bentheimer sampler that about 32% of the voxels have at least a 2% difference in greyscale values between two consecutive fast scans. These differences are not physical-based variations but rather intrinsic uncertainties measurements. They also show that such artifacts' uncertainties can be reduced by using slower acquisition time, though this is not always feasible to capture fast-dynamical processes. Time-resolved experiments of dynamical processes can provide insights into kinetic dissolution rates, though this also suffers from imaging limitations that can lead to misleading estimations.

Inferring reliable mineral reactivity from dynamical microtomography and quantifying imaging morphological uncertainties are identified as the major issues that can bias the determination of evolving petro-physical properties in geological applications. Current methodologies addressing these problems mainly fall into two categories: on one side, purely model-related approaches based on static  $\mu$ CT scans and upscaling principles, and on the other side, image treatment analysis relying on experimental static or dynamical images. Nonetheless, neither morphological uncertainty nor reaction rate quantification has been investigated from the perspective of coupling physics-based models with data-driven techniques. To the best of our knowledge, the development of data assimilation approaches on pore-scale imaging problems that combine dynamical microtomography and physical regularization induced by the PDE model of reactive processes is the main novelty of the present manuscript. The motivation for this formulation lies in its joint ability to infer mineral reactivity parameters and quantify the residual micro-porosity generated by unresolved features in the microtomography imaging process. Therefore, we assert that a proper balance between dynamical microtomography imaging and their PDE-based physical formulation could provide insights into the uncertainty quantification issues related to reactive pore-scale modeling. In this direction, we propose a novel methodology that uses a physics-based dissolution model as regularization constraints to dynamical data-driven microtomography inference. This aim at quantifying both the uncertainties on kinetic parameters to perform reliable model calibration and the morphological imaging uncertainty on the unresolved micro-porosity field.

### 3 Direct and inverse problem setup

This section is dedicated to setup the dimensionless versions of the dissolution PDE model for direct and inverse problems. We establish the main differences in their dimensionless formulations and define the modeling assumptions used in the present article.

#### 3.1 Usual dimensionless formulation of the direct problem

The overall calcite dissolution PDE system, defined in equation (10), can model a wide range of dissolution regimes and patterns characterized by well-established dimensionless numbers. By setting  $x^* = x/L$  and  $t^* = tD_m/L^2$  and following the notations of Sect. 2.1, one can introduce the so-called Peclet and Reynolds numbers

$$\text{Pe} = \bar{u}L/D_m \quad \text{and} \quad \text{Re} = \rho\bar{u}L/\mu \quad (12)$$

where  $\bar{u}$  and  $L$  are respectively the characteristic velocity and length of the sample. In the context of pore-scale simulations, the inertial effects become negligible compared to viscous forces due to low Reynolds numbers — typically we have the assumption  $\text{Re} \ll 1$  throughout this article. Regarding the chemical reactions, two dimensionless numbers are defined: the catalytic Damköhler number denoted  $\text{Da}_{\text{II}}$  and its inherited convective number  $\text{Da}_{\text{I}}$ , expressed as

$$\text{Da}_{\text{II}} = \frac{K_s A_s \gamma_{\text{H}^+} L^2}{D_m} \quad \text{and} \quad \text{Da}_{\text{I}} = \frac{\text{Da}_{\text{II}}}{\text{Pe}} = \frac{K_s A_s \gamma_{\text{H}^+} L}{\bar{u}}. \quad (13)$$

The characteristic length  $L$  is usually related to average pore throat diameters or  $L^2$  can be set as the surface of a section divided by the average number of grains (*e.g.* see [28] for practical cases). Otherwise, it is

possible to set the characteristic length of the problem as  $L = \sqrt{\kappa_0}$ , provided an experimental or numerical estimation of  $\kappa_0$  [67]. All these dimensionless numbers are meaningful in direct dissolution problems to qualify the different dominant regimes in terms of diffusion, reaction, and advection.

Using the dimensionless variables  $(x^*, t^*)$ , the normalized concentration  $C^* = C/C_0$  and velocity  $u^* = u/\bar{u}$ , one finally gets the dimensionless formulation of the overall reactive flow system (10) on the dimensionless spatiotemporal domain  $\Omega^* \times (0, T_f^*)$ . This leads to the usual PDE model:

$$\left\{ \begin{array}{ll} -\Delta u^* + L^2 \kappa_0^{-1} \frac{(1-\varepsilon)^2}{\varepsilon^2} u^* = \varepsilon(f^* - \nabla p^*), & \text{in } \Omega^* \times (0, T_f^*) \\ \frac{\partial C^*}{\partial t^*} + \text{Pe} \nabla \cdot (\varepsilon^{-1} u^* C^*) - \nabla \cdot (\varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C^*)) = -\text{Da}_{\text{II}} C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ \frac{1}{v C_0} \frac{\partial \varepsilon}{\partial t^*} = \text{Da}_{\text{II}} C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ + \text{adequate boundary and initial conditions, along with } \nabla^* \cdot u^* = 0 & \end{array} \right. \quad (14)$$

obtained by means of multiplying the hydrodynamic DBS equation by  $L/\rho\bar{u}^2$  and the chemical equations in (10) by  $L^2/C_0 D_m$ . The notations  $f^*$  and  $p^*$  in the dimensionless DBS equation are defined by  $f^* = fL^2/(\mu\bar{u})$  and  $p^* = pL/(\mu\bar{u})$ , and finally  $C_0$  is a characteristic constant for the acid concentration field. This PDE system defines the overall dimensionless formulation of the direct problem of calcite dissolution.

### 3.2 Modeling assumptions on the direct and inverse problems

In the applications Sect. 6 and 7, we consider inverse problems in the dissolution process of calcite cores with heterogeneous porosity levels for 1D and 2D spatial configurations. Although the 1D+Time test case is purely synthetic and aims to validate the method developed in Sect. 5, the 2D+Time application addresses a more realistic problem that can be applied to isotropic porous samples. Several modeling assumptions are, though, made to address both the reactive direct and inverse problems. These modeling assumptions are detailed hereafter and determine the dissolution regime considered in the applications.

The dimensionless numbers Re and Pe are common to both the direct and inverse formulations and respectively establish the viscosity-dominated regime and the convective or diffusive transport regime. At the initial state in the dynamic imaging process, we assume that the porous medium is completely saturated with the acid by capillary effect. The amount of reactant at the pore interface is initially homogeneously distributed, and therefore we expect, at first, a cylindrical dissolution regime of the calcite core with spherical symmetry. Subsequently, the dissolution process may deviate from this cylindrical pattern due to local heterogeneities in the micro-porosity field  $\varepsilon$ . We also suppose a low Peclet hypothesis  $\text{Pe} \ll 1$ , so that the reactant diffusion is dominant over the advection phenomena resulting in more homogeneous dissolution rates at the interface (*e.g.* see [67] for the dissolution regimes characterization). In this sense, continuous acid injection is maintained at a given fluid flow rate to ensure a diffusive-dominated regime for the dissolution. Consequently, we neglect the advection effects in the present article and focus on the following reaction-diffusion system:

$$\left\{ \begin{array}{ll} \frac{\partial C^*}{\partial t^*} - \nabla \cdot (\varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C^*)) = -\text{Da}_{\text{II}} C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ \frac{1}{v C_0} \frac{\partial \varepsilon}{\partial t^*} = \text{Da}_{\text{II}} C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ C^* = 1, & \text{on } \partial\Omega^* \times (0, T_f^*) \\ C^*(x, 0) = C_{\text{init}}(x)/C_0 := C_{\text{init}}^*, & \text{in } \Omega^* \times \{0\} \end{array} \right. \quad (15)$$

written in its dimensionless form with the normalized concentration field  $C^* = C/C_0$ . The continuous acid injection is modeled through non-homogeneous Dirichlet boundary conditions on  $C^*$ , with the characteristic constant  $C_0$  chosen as the value of the Dirichlet boundary conditions on  $C$ . The initial condition on the micro-porosity field  $\varepsilon$  arises from the dry microtomography scan or the initial synthetic porous medium. Ultimately, we obtain a PDE model driven by one dimensionless number, namely the catalytic Damköhler

number, characterizing the ratio of the reaction rate over the diffusion effects. The system (15) is, therefore, consistent with the standard dimensionless formulation of the direct problem, subject to a diffusive-dominated transport regime.

However, we merely cannot consider the dimensionless temporal variable  $t^* = tD_m/L^2$  in a reactive inverse problem as it strongly depends on molecular diffusion  $D_m$ , which is among the unknown kinetic parameters to be estimated. In the next section, we focus on the challenge arising from the dimensionless formulation of a reactive inverse problem in the context of calcite dissolution.

### 3.3 Dimensionless inverse problem on calcite dissolution

In this article, we address pore-scale imaging inverse problems in dissolution processes. We aim to recover and quantify uncertainties both on the micro-porosity field description  $\varepsilon$  and the reactive parameters involved in the diffusion-reaction system. Among these inverse kinetic parameters, one can find the molecular diffusion  $D_m$ , the tortuosity index  $\beta$ , the dissolution rate constant  $K_s$ , and even the specific surface area  $A_s$  — usually estimated on the dry  $\mu$ CT scan. Consequently, these parameters — in particular  $D_m$  — cannot be used for the non-dimensionalization of the model since they are to be determined. Apart from special considerations of the tortuosity index of the sample, the other inverse parameters characterize the dissolution regime of the dynamical  $\mu$ CT experiment. In this sense, they provide insight into the physical catalytic Damköhler number  $\text{Da}_{\text{II}}$ , though the direct dimensionless formulation (15) is inappropriate for an inverse problem. The dimensionless temporal variable  $t^*$  in the PDE system (15) is, indeed, closely related to the unknown molecular diffusion, compromising its application to inverse modeling. Establishing the dimensionless formulation of the inverse dissolution problem is not straightforward and therefore requires a different dimensionless time.

In the inverse problem, we consequently introduce the new temporal variable

$$t^* = \frac{tD_{\text{ref}}}{L^2}, \quad (16)$$

where  $D_{\text{ref}}$  is a scaling factor of the dimensionless formulation for the chemical kinetics. This scaling factor can also be defined as  $D_{\text{ref}} = L^2/T$  introducing  $T$  the characteristic time for the dimensionless problem, which is not the physical characteristic time for the diffusion since the latter is unknown. In practice, we can rely on a rough estimation of physical dissolution time  $T_f$  — determining the dynamical process end — and a given dimensionless final time — usually  $T_f^* = 1$  — to set the factor  $D_{\text{ref}}$ . The estimations of this scaling parameter will be detailed on a case-by-case basis throughout the applications developed in Sect. 6 and 7. Using the new dimensionless variables  $(x^*, t^*)$ , the normalized concentration  $C^* = C/C_0$  along with the definition of  $D_{\text{ref}}$ , one can obtain the dimensionless formulation of the reaction-diffusion system in the context of inverse modeling, which leads to:

$$\begin{cases} \frac{\partial C^*}{\partial t^*} - D_m^* \nabla \cdot \left( \varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C^*) \right) = -\text{Da}_{\text{II}}^* C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ \frac{1}{vC_0} \frac{\partial \varepsilon}{\partial t^*} = \text{Da}_{\text{II}}^* C^* \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega^* \times (0, T_f^*) \\ C^* = 1, & \text{on } \partial\Omega^* \times (0, T_f^*) \\ C^*(x, 0) = C_{\text{init}}^*, & \text{in } \Omega^* \times \{0\}. \end{cases} \quad (17)$$

In reactive inverse problems, we thus obtain a PDE model driven by two dimensionless numbers denoted  $\text{Da}_{\text{II}}^*$  and  $D_m^*$  which are defined as:

$$\text{Da}_{\text{II}}^* := K_s A_s \gamma_{\text{H}^+} T = \frac{K_s A_s \gamma_{\text{H}^+} L^2}{D_{\text{ref}}} \quad \text{and} \quad D_m^* := \frac{D_m T}{L^2} = \frac{D_m}{D_{\text{ref}}}. \quad (18)$$

Finally, the physical Damköhler number corresponding to the dynamical  $\mu$ CT experiment is recovered as the a-posteriori ratio  $\text{Da}_{\text{II}} = \text{Da}_{\text{II}}^*/D_m^*$ . From now on, we consider this dimensionless formalism and forget

the star notation on the differential operator, domains, and field descriptions for the sake of readability. This results in the following inverse dimensionless PDE system:

$$\left\{ \begin{array}{ll} \frac{\partial C}{\partial t} - D_m^* \nabla \cdot (\varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C)) = -\text{Da}_{\text{II}}^* C \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega \times (0, T_f) \\ \frac{1}{v C_0} \frac{\partial \varepsilon}{\partial t} = \text{Da}_{\text{II}}^* C \mathbb{1}_{\{(1-\varepsilon)>0\}}, & \text{in } \Omega \times (0, T_f) \\ C = 1, & \text{on } \partial\Omega \times (0, T_f) \\ C(x, 0) = C_{\text{init}}, & \text{in } \Omega \times \{0\} \end{array} \right. \quad (19)$$

with  $\text{Da}_{\text{II}}^*$  and  $D_m^*$  the inverse parameters to estimate, and  $C_0$  and  $v$  constant parameters. The tortuosity index  $\beta$  is either set through a-priori estimation on the porous sample, modeling through the empirical Archie law, or regarded as an additional inverse parameter. Especially, the index  $\beta = 1$  is often considered for porous media with strong pore connections [71, 67] although practical upscaling of the diffusion can result in intermediate index values [28].

In addition to inferring the reactivity parameters  $\text{Da}_{\text{II}}^*$  and  $D_m^*$ , we aim to estimate the spatial variability on the porosity field  $\varepsilon$ . In this sense, we develop a data assimilation approach on pore-scale imaging that combine dynamical  $\mu\text{CT}$  experiments of calcite dissolution and physical regularization induced by the dimensionless PDE model (19). It benefits from the joint ability to quantify the ranges of mineral reactivity and the residual micro-porosity generated by unresolved features in the microtomography imaging process. This formulation also relevantly combines the advantages of experimental and modeling approaches and overcomes their own limitations. On the one hand, the dissolution process observation will bring insights into the unresolved morphological features and lead to a better characterization of the sample's initial state. On the other, the PDE model regularization can efficiently substitute the differential imaging approach, which is controversial for fast-dynamical processes subject to poor imaging quality. Therefore, one can address mineral reactivity inference for highly noisy dynamical  $\mu\text{CT}$  resulting from the compromise between scan quality and time resolution, as introduced in Sect. 2.3. The major challenge of this data assimilation formulation, though, relies on the PDE constraint for the concentration field as the  $\mu\text{CT}$  experiments do not provide information on the flow, transport, or diffusion of the chemical reactant. In the reactive inverse problem, the acid concentration is thus a latent field whose only the dimensionless boundary conditions are known in equation (19) through the normalizing constant  $C_0$ . In the next section, we will develop the methodology adopted to solve such a dissolution inverse problem, accounting for all the established modeling assumptions.

## 4 Bayesian Physics-Informed Neural Networks in pore-scale imaging: concepts and methods

Developing efficient data assimilation techniques is crucial to perform extensive parameter estimations, uncertainty quantification, and improving the reliability of direct pore-scale predictions. In particular, inverse problems are often subject to various sources of uncertainty that need to be quantified to ensure trustable estimations. This includes approximate model accuracy whose reliability can be questioned, with sparse or noisy data exhibiting measurement variability. Integrating physical principles, such as conservation laws or PDE models, in these inverse problems can though compensate for the lack of massive or accurate measurements through additional regularization constraints [42]. At the same time, embedding these physical regularizations allows addressing model accuracy in the total uncertainty quantification, especially when misleading a-priori uncertainty is assumed on the physical constraints [57]. Therefore, the combination of physics-based and data-driven methods offers an efficient alternative to overcome the limitations of both purely data-driven or purely modeling approaches. This has established data-driven inference as a complementary partner to theory-driven models in data assimilation and inverse modeling incorporating uncertainty quantification.

## 4.1 Uncertainty Quantification in coupled physics-based and data-driven inverse problems

Several approaches were developed to address uncertainty concerns in the context of data assimilation. These uncertainty quantification problems either require stochastic PDE models [13, 29] — also used in sensitivity analysis — or probabilistic approaches such as Markov Chain Monte Carlo (MCMC) methods. The latter can be used in the Bayesian Inference framework to sample from a target posterior distribution, though this usually requires numerous evaluations of the forward PDE model. In this sense, developing efficient MCMC methodologies remains challenging since repeatedly solving a complex coupled PDE system is computationally expensive and therefore can quickly become prohibitive for uncertainty assessments. These computational concerns have motivated the emergence of surrogate models in Bayesian Inference to speed up the forward model evaluation. This covers methods ranging from Polynomial Chaos Expansions [40, 77] which rely on a representation of the physical model by a series of low-order polynomials of random variables, to neural network proxies [78, 3]. Both approaches present the advantage of creating a surrogate model that can be evaluated inexpensively compared to solving the forward problem through usual direct numerical simulations. Nonetheless, Polynomial Chaos expansions suffer from truncation errors due to the low order of the polynomials yielding inaccurate estimates of the posterior distributions [37]. On the contrary, deep learning methods have shown effectiveness in building surrogate models for a wide range of complex and non-linear PDEs encoding the underlying physical principles. Developing fast surrogate models based on machine learning has garnered increasing interest in accelerating Bayesian inference for a wide range of scientific applications [24, 22].

A popular framework in deep learning integrating both physics regularization, measurement data, and uncertainty estimates are Bayesian Physics-Informed Neural Networks (BPINNs) [79, 36]. BPINNs benefit from the combined advantages of neural network structures in building parameterized surrogate models based on physical principles and Bayesian inference standards in integrating uncertainty quantification. Introducing the Bayesian neural network parameters  $\theta \in \mathbb{R}^d$  building the surrogate model and the inverse parameters of the PDE model  $\mathcal{P}_{\text{inv}} \in \mathbb{R}^p$ , we define the joint set of unknown parameters as  $\Theta = \{\theta, \mathcal{P}_{\text{inv}}\}$ . The BPINN formulation aims to explore the posterior distribution of  $\Theta$

$$P(\Theta|\mathcal{D}, \mathcal{M}) \propto P(\mathcal{D}|\Theta)P(\mathcal{M}|\Theta)P(\Theta) \quad (20)$$

given some measurement data  $\mathcal{D}$  and a presumed model  $\mathcal{M}$  with unknown parameters. The posterior distribution expression (20) basically involves a likelihood term  $P(\mathcal{D}|\Theta)$  evaluating the distance to the experimental data, a PDE-likelihood term  $P(\mathcal{M}|\Theta)$  characterizing the potential modeling discrepancies, and a joint prior distribution  $P(\Theta)$ . Through a marginalization process, the posterior distribution (20) on the parameters  $\Theta$  then transfers into a posterior distribution of the predictions, also called a predictive Bayesian Model Average (BMA) distribution (*e.g.* see [75]):

$$P(y|x, \mathcal{D}, \mathcal{M}) = \int P(y|x, \Theta)P(\Theta|\mathcal{D}, \mathcal{M})d\Theta \quad (21)$$

where  $x$  and  $y$  respectively refer to the input (*e.g.* spatial and temporal points) and output (*e.g.* field prediction of the micro-porosity) of the neural network. The BPINN formulation hence provides a predictive distribution (21) of the quantities of interest (QoI), such as the output micro-porosity field, as well as posterior distributions over the model inverse parameters  $\mathcal{P}_{\text{inv}}$ . Sampling from the posterior distribution (20) is achieved through MCMC methods, which efficiently combine with fast surrogate models based on deep learning. In particular, one of the most popular MCMC schemes for BPINNs is Hamiltonian Monte Carlo (HMC), which provides a particularly efficient sampler for high-dimensional inference problems [15]. In addition to theoretical analyses, the HMC-BPINNs formulation also demonstrates numerical performances on both forward and inverse problems [79]. BPINNs with the HMC sampler appear as an efficient data-assimilation alternative coupling physics-based with data-driven approaches, and incorporating intrinsic uncertainty quantification.

The HMC sampler, in particular, introduces the dynamics of a fictive system composed of the unknown parameters  $\Theta$  — regarded as particle positions in the physical analogy — and auxiliary momentum variables  $r$  — regarded as particle velocities. It describes a conservative Hamiltonian system whose energy denoted  $H(\Theta, r)$  is the sum of a potential energy  $U(\Theta)$  which characterize the inverse problem formulation and

a kinetic energy  $K(r)$  accounting for momentum perturbations. The latter enables the sampler to diffuse across several energy levels and hence results in an efficient exploration of the joint posterior distribution  $\pi(\Theta, r)$  in the phase space, defined as follows:

$$\pi(\Theta, r) \sim e^{-H(\Theta, r)}. \quad (22)$$

The potential energy definition relies on a Bayesian probabilistic formulation of the inverse problem such that it depends on the posterior distribution (20) by the relation  $U(\Theta) = -\ln P(\Theta|\mathcal{D}, \mathcal{M})$ . Along with a Euclidean-Gaussian assumption for the kinetic energy (e.g. see [15] or [55]), this ensures that the marginal distribution of  $\Theta$  provides immediate samples of the target posterior distribution:

$$P(\Theta|\mathcal{D}, \mathcal{M}) \sim e^{-U(\Theta)}. \quad (23)$$

Efficient exploration of the joint distribution  $\pi(\Theta, r)$  in the phase space hence projects to samples of the target distribution (20) and then provides predictive BMA distributions on the QoI given by equation (21). The successive samples  $(\Theta, r)$  are generated by solving for the Hamiltonian dynamical system for the frictionless fictive particle of positions  $\Theta$

$$\begin{cases} d\Theta = \mathbf{M}^{-1}r dt \\ dr = -\nabla U(\Theta) dt, \end{cases} \quad (24)$$

through a symplectic integrator, such as the Störmer-Verlet also known as the leapfrog method. This account for a deterministic exploration of specific energy level sets — since the Hamiltonian energy is theoretically preserved by symplectic integrators — while the kinetic energy, through momentum sampling, enables a stochastic exploration between the energy levels. The HMC-BPINN formulation ensures efficient sampling of the target posterior distribution thanks to the description of a conservative Hamiltonian system related to the inverse problem description.

Overall, the potential energy term can be expressed under the general form (see [55] for detailed development of this weighted multi-potential energy):

$$U(\Theta) = \sum_{k=0}^K \lambda_k \mathcal{L}_k(\Theta) + \lambda_{K+1} \|\Theta\|^2 \quad (25)$$

where  $L_k = \lambda_k \mathcal{L}_k$  refers to the weighted  $k^{th}$  objective term, either corresponding to data-fitting log-likelihood or PDE regularization tasks. We assume here that the prior distribution on the set of parameter  $\Theta$  follows a Gaussian distribution such that  $P(\Theta) \sim \mathcal{N}(0, I_{p+d})$ . The weights  $\lambda_k$  are positive parameters integrating the various sources of uncertainties. Indeed, the deterministic PDE model is completed by stochastic representations of the model discrepancy, and the data-fitting likelihood is itself supplemented by stochastic modeling of the experimental noise, both affecting the weights  $\lambda_k$ . In this sense, a HMC-BPINN intends to capture and estimate the various sources of uncertainties whether aleatoric — arising from variability or randomness in the observations like sensor noise — or epistemic — caused by imperfect modeling hypothesis or ignorance in the model adequacy. Automatic management of these uncertainties, though, remains challenging as this relies on the appropriate setting of the critical weighting parameters  $\lambda_k$  arising from the expression of the multi-potential energy (25). Although some of these parameters — mainly the noise estimation — can be adjusted *offline* with pre-trained Generative Adversarial Networks (GAN) as proposed by Psaros et al. in [57], proper estimation of these weights is crucial to ensure robust uncertainty quantification. Unsuitable choices of these weights can lead to biased predictions and pathological behaviors of the HMC-BPINN sampler, especially in the context of complex real-world Bayesian inference involving multi-objective, multiscale, and stiffness issues. This needs the development of data assimilation strategies that robustly address these issues to achieve reliable uncertainty quantification in inverse problems.

## 4.2 A robust adaptive weighting sampling strategy for complex real-world Bayesian inference

The Bayesian Physics-Informed Neural Network paradigm offers the opportunity to query altogether the confidence in the predictions, the estimations of inverse parameters, and the model adequacy in inverse



problems incorporating uncertainty quantification. Despite their effectiveness, BPINNs can be difficult to use correctly in complex real-world Bayesian inference as they are prone to a range of pathological behaviors. These instabilities arise from the multi-potential energy term (25) in multi-objective inverse problems which likely involve conflicting tasks or multiscale issues. In particular, such a multi-potential energy directly translates to a weighted multitask posterior distribution for which achieving successful and unbiased sampling is challenging. Ensuring robust Bayesian inference in this context hinges on properly estimating the distinct task weights. Indeed, unsuitable choices of the weights  $\lambda_k$  in (25) can result in biased predictions, vanishing task behavior, or substantial instabilities in the Hamiltonian conservation. This can even prevent the sampler from identifying the highest posterior probability region, namely the Pareto front neighborhood, corresponding to predictions that correctly balance all the different tasks. While manual calibration of the critical  $\lambda_k$  weights is still commonplace [42, 36, 47], robust Bayesian inference strategies should not rely on a-priori hand-tuning or biased calibration of the posterior distribution. Indeed, appropriately setting these parameters is neither easy nor computationally efficient, especially for multi-objective inverse problems arising from real-world data. Developing an alternative that accounts for this multitask consideration becomes crucial to ensure robust sampling when dealing with coupled physics-based and data-driven inference.

We benefit from an efficient BPINN framework developed in our previous work [55], which robustly addressed multitask Bayesian inference problems with potential multiscale effects, stiffness issues, or competing tasks. This new strategy relied on an adaptive and automatic weighting of the target posterior distribution based on an Inverse Dirichlet control of the weights  $\lambda_k$  [39], which leverages gradient variances information of the different tasks:

$$\lambda_k = \left( \frac{\gamma^2}{\text{Var}\{\nabla_{\Theta}\mathcal{L}_k\}} \right)^{1/2}, \quad \text{with } \gamma^2 := \min_{t=0..K} (\text{Var}\{\nabla_{\Theta}\mathcal{L}_t\}), \quad \forall k = 0, \dots, K. \quad (26)$$

This results in an alternative sampler called Adaptively Weighted Hamiltonian Monte Carlo (AW-HMC) (see Appendix A for the AW-HMC algorithm and [55] for the detailed methodological development) that ensures the gradients distributions of the weighted potential energy terms  $L_k = \lambda_k\mathcal{L}_k$  in (25) have balanced variances. In this sense, the AW-HMC sampler avoids imbalanced conditions between the different tasks. Moreover, this enables us to concentrate the sampling on the Pareto front exploration after the adaptive procedure, as balancing the target distribution based on the minimum variance of the gradients  $\gamma^2$  in (26) contributes to adjusting the weights  $\lambda_k$  with respect to the most likely task in the multi-objective inverse problem. In particular, the most uncertain tasks will experience an automatic increase of their uncertainties. The weights  $\lambda_k$  are adjusted during the sampling for a number of adaptive steps  $\tau \leq N$ , characterizing the convergence of the weighted posterior distribution toward the high probability-density region and, therefore, allowing the efficient exploration of the Pareto front neighborhood after their adaptation. The stopping criterion of the adaptive weighting is based on the evolution of the weighted Hamiltonian energy  $H_{\lambda_{\tau}}(\Theta, r)$ , defined by

$$H_{\lambda_{\tau}}(\Theta, r) = \sum_{k=0}^{K+1} \lambda_k(\tau)\mathcal{L}_k(\Theta) + K(r) \quad (27)$$

for the adaptive sampling steps  $\tau$ . In particular, we rely on both a maximum number of adaptive steps  $N_{\max}$  and a threshold  $S_{\min}$  on the local variation of the Hamiltonian to set up the effective number of adaptive steps  $N$  (see Algorithm 1 in Appendix A).

The AW-HMC sampler benefits from enhanced convergence and stability compared to conventional samplers, such as HMC or NUTS [27], and reduces sampling bias by avoiding manual tuning of critical weighting parameters. In terms of computational efficiency, it retains comparable training costs to the previous alternatives while circumventing the issue of excessively reducing the leapfrog time step  $\delta t$  to ensure the Hamiltonian conservation, as can occur with NUTS. Indeed, when integrating multiple sources of uncertainties in a multi-objective inverse problem, the NUTS sampler will adjust  $\delta t$  to satisfy the stiffest constraint, which can require very small time steps and is thereby more likely to generate random walk behaviors. On the contrary, the AW-HMC sampler guarantees the Hamiltonian conservation with optimal leapfrog time step in terms of convergence rates for a similar training cost, therefore improving the sampler convergence and robustness. This new alternative also demonstrated efficiency in managing the scaling sensitivity of

the different terms either to noise distributions (homo- or hetero-scedastic) or multi-scale issues. In fact, the adjusted weights bring information on the distinct task uncertainties. This improves the reliability of the noise-related and model adequacy estimates as the uncertainties are quantified with minimal a-priori assumptions on their scaling. Our novel sampling strategy has demonstrated outstanding performances on several levels of complexity. This covers applications ranging from data-fitting predictions based on sparse measurements, physics-based data-assimilation problems, data-assimilation in inverse problems with unknown PDE model parameters, and data-assimilation in inverse problems with unknown parameters and latent fields [55]. Taken together, the AW-HMC sampler enhances BPINN robustness and offers a promising alternative as an overall data-assimilation strategy, extending its applications to more complex Bayesian inference problems. Indeed, this adaptive weighting sampling presents the ability to effectively address multiscale and multitask inverse problems, to couple UQ with physical priors, and to handle sparse noisy data. It also showed effectiveness in addressing stiff dynamics problems including latent field reconstruction and deriving unbiased uncertainty information from the measurement data. The AW-HMC strategy appears as an unlocking advance bringing robustness to the BPINNs and, therefore, provides a promising data assimilation framework to address robust and reliable Bayesian inference in multitask inverse problems.

## 5 Data assimilation strategy: sequential reinforcement and operator differentiation

In the present work, we focus on a multitask inverse problem for reactive flows at the pore scale involving the identification of two inverse parameters ( $Da_{\text{II}}^*$  and  $D_m^*$ ) and a latent concentration field  $C$ . This novel approach combines dynamical imaging data of calcite dissolution and physics-based regularization induced by the dimensionless PDE model (17). We build the current data assimilation approach upon the efficient AW-HMC framework for Bayesian Physics-Informed Neural Networks, presented in Sect. 4.2, to quantify both morphological and chemical parameter uncertainties. The challenges arising from reactive inverse problems in pore-scale imaging currently lie in the limited variety of available data, as the calibration of the coupled PDE system involves merely dynamical information obtained from the  $\mu$ CT scans. In particular, the lack of information on the concentration field evolution throughout dissolution necessitates identifying the latent field  $C$ , whose reconstruction is crucial for inferring the inverse parameters. Moreover, this work is positioned within a diffusive-dominated regime (see Sect. 3.2) and therefore neglects the advective effects, which would require estimating additional latent variables for the velocity field in the coupled PDE system (14). The shape complexity of 3D porous samples, along with the non-linear operators involved in the PDE model, are additional concerns that make such an application challenging, especially when targeting full 4D data assimilation on real-life materials. In this section, we present the data assimilation method developed to handle such pore-scale imaging inverse problems. Our methodology emphasizes the sequential reinforcement of the multi-potential energy and the efficient computation of the heterogeneous diffusion operator arising from Archie’s law. This first requires setting up a few dedicated notations.

### 5.1 Domain decomposition and sampling notation setup

Dynamical synthetic or experimental  $\mu$ CT images are available on the overall spatiotemporal domain  $\Omega \times (0, T_f)$ , and provide dissolution observations subject to noise and imaging limitations (see Sect. 2.2 and 2.3): we introduce the two-index set  $\mathfrak{I}m_{i,j}$  of image intensities (dissolution measurements) defined for the whole image voxels. Then, we define a subset of this image for sampling purposes, involving the positions  $\mathcal{D}$  as a subset of  $\bar{\Omega} \times (0, T_f)$  together with their image intensities  $Im$ , corresponding to  $N_{\text{obs}}$  partial and corrupted training observations:

$$\mathcal{D} = \{(x_k, t_k), \quad k = 1 \dots N_{\text{obs}}\} \quad (28)$$

On this set of discrete points, there exists mappings  $i$  and  $j$  such as the image intensity satisfies

$$Im_k := \mathfrak{I}m_{i(k),j(k)} = 1 - \varepsilon(x_k, t_k) + \xi(x_k, t_k), \quad k = 1 \dots N_{\text{obs}}$$

where the noise  $\xi \sim \mathcal{N}(0, \sigma^2 I)$  for which the standard deviation  $\sigma$  is automatically estimated in the AW-HMC sampler by means of the  $\lambda_k$  adjustment in equation (25). This relationship between the microtomogra-

phy images  $\text{Im}$  and  $\varepsilon$  comes from the correlation between the  $\mu\text{CT}$  values and the material local attenuation. Indeed, in a greyscale tomographic scan, the minimum signal corresponds to the least attenuating or the least dense areas (in  $\Omega_F$  where  $\varepsilon = 1$ ), while the maximum signal refers to the most attenuating areas (in  $\Omega_S$  where  $0 < \varepsilon_0 \leq \varepsilon < 1$ ). Due to the micro-continuum description of the medium based on the two-scale porosity assumption, each distinct region of the domain — namely  $\Omega_S$  and  $\Omega_F$  — is though regarded as a different term in the multi-potential energy definition. Such a distinction is prescribed since there is no guarantee that the data corruption is uniform: the measurement variability can differ locally when facing heteroscedastic noise. In particular, the artifact limitations tend to enhance the blurring effects at the material fluid/solid interface  $\Sigma$ . This motivates the special consideration of this interface neighborhood to account for the unresolved features and provide reliable morphological uncertainties.

In this sense, we introduce the Reactive Area of Interest (RAI) as the evolving fluid/mineral interface along the dissolution process that is defined by:

$$\text{RAI} = \{(x_k, t_k) \in \Omega \times (0, T_f) \text{ such that } \mathfrak{I}m_{i(k)+1, j(k)} < \mathfrak{I}m_{i(k), j(k)} \text{ and } 0.1 < \text{Im}_k < 0.9\} \quad (29)$$

where the imaging extreme values are ignored, as a correction criterion, to avoid integrating noise derivation artifacts into the definition of the RAI. These thresholds rely on the analysis of the  $\mu\text{CT}$  histogram of the initial porous medium dataset. We also define:

- The extended Reactive Area of Interest, denoted  $\text{RAI}^+$ , as the RAI augmented by a fluid tubular neighborhood of the RAI both in space and time, that is to say, the fluid region close to the evolving interface in  $\Omega \times (0, T_f)$ ,
- The reduced Reactive Area of Interest, denoted  $\text{RAI}^-$ , based on the acceptability criterion — positivity of the  $D_m^*$  estimations — defined thereafter and related to the relations (35)-(36).

Moreover, we introduce several discrete domains  $\mathcal{D}^\bullet$  defined as the intersection between the non-reactive part of  $\mathcal{D}$  and their respective time-dependent regions: fluid, solid, reactive, or boundary. For instance, one gets  $\mathcal{D}^F = (\mathcal{D} \cap Q_F) \setminus \text{RAI}$  with  $Q_F$  the evolving fluid region defined as

$$Q_F = \{(x, t) \in \Omega \times (0, T_f) \text{ such that } x \in \Omega_F(t)\}. \quad (30)$$

In the same way,  $\mathcal{D}^S = (\mathcal{D} \cap Q_S) \setminus \text{RAI}$  where  $Q_S$  is the evolving solid region and  $\mathcal{D}^\partial = \mathcal{D} \cap \{\partial\Omega \times (0, T_f)\}$ . The overall domain  $\Omega \times (0, T_f)$  is decomposed into several regions and respective training datasets that are involved in the sequential reinforcement of the multi-potential energy.

Finally, these different domains satisfy the following properties:

- $\text{RAI}^- \subset \text{RAI} \subset \text{RAI}^+$ ,
- $\Omega_F \cup \Omega_S = \Omega$ , where  $\Omega_F = \mathbb{1}_{(\varepsilon < 1)}$  in  $\Omega$  is an open set,
- $\mathcal{D} = \mathcal{D}^F \cup \mathcal{D}^S \cup \mathcal{D}^\partial \cup \text{RAI}$ .

## 5.2 Sequential reinforcement of the multi-potential energy

The data assimilation strategy developed in the present article relies on a sequential design of the multi-potential energy  $U(\Theta)$ , which will be reinforced to incorporate additional constraints and successive modeling approximations through dedicated sampling steps. While this sequential reinforcement approach draws significant inspiration from the PINNs framework, where the regression of physical fields typically precedes the incorporation of physics-based constraints, it also introduces a modeling reinforcement, crucial for inferring latent fields in strongly coupled PDE systems. In particular, this sequential splitting is necessary due to the coupling between the porosity field  $\varepsilon$ , related to the  $\mu\text{CT}$  imaging process, the latent concentration field  $C$ , and the two unknown inverse parameters  $\text{Da}_{\text{II}}^*$  and  $D_m^*$ . The first sampling step is, therefore, dedicated to providing a-priori estimations of the micro-porosity field through data-fitting terms only, while the following steps address the modeling reinforcement.

This sequence of tasks is then decomposed into three steps, from the one for which we have the most information to the set of tasks involving all the aspects and constraints we need to consider, as displayed in Fig 2 and developed thereafter:

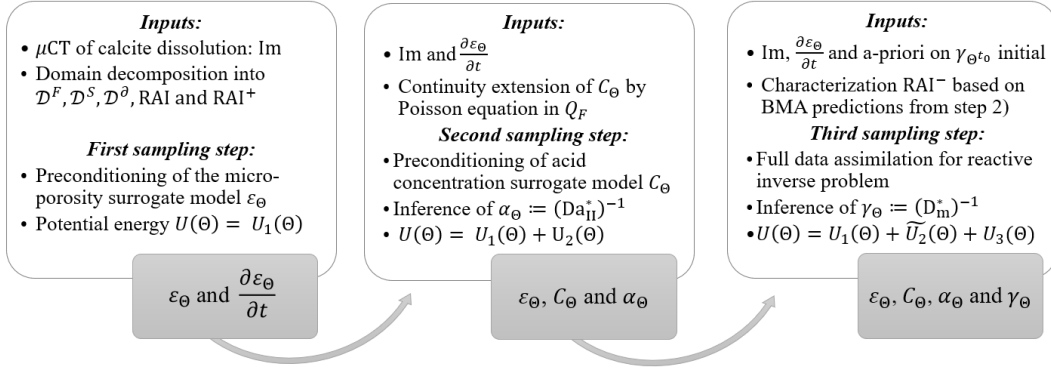


Figure 2: **Sequential graph of the potential energy reinforcement:** our data assimilation strategy incorporates additional physics-based constraints arising from the PDE model (19) through successive sampling steps. The notations are defined in Sect. 5.2.

- Step 1: Preconditioning on the micro-porosity by pure regression on image data,
- Step 2: Preconditioning of the latent reactive fluid with additional PDE constraint, given a first modeling approximation of the coupled PDE system
- Step 3: Overall data assimilation potential with the full and exact reactive model.

Extensions of the present methodology can be investigated for applications in other disciplines. However, the elaboration of the sequential reinforcement approach would rather be problem-dependent, and in this sense, a automatic formalisation of this process currently seems out of reach. It would essentially depend on the number of inverse parameters to identify, the presence of latent fields, the amount of data available, and would overall be affected by the under-determined or ill-conditioned nature of the problem. For instance, when facing numerous unknowns in a system with limited data, splitting the constraints and utilizing intermediate approximate modeling becomes increasingly necessary to infer the full dynamics. In this sense, the idea of sequential reinforcement, especially when dealing with complex and strongly coupled dynamic behaviors, can be applied to other fields.

For instance, calibrating a large number of inverse parameters in a system of chemical kinetic equations for an uncertainty-quantified pyrolysis model, with applications to the atmospheric entry of spacecraft, also requires sampling a Bayesian posterior in high dimensions with limited knowledge of various sources of uncertainty [22]. In biology, while may have access to a slightly larger variety of data, including measurements of concentrations and velocity fields in the context of kinetic reactions between proteins, the sparsity of the information and the absence of predictive models can, however, raise additional challenges [38].

Thereafter, we develop each of the sequential reinforcement steps within the framework of reactive inverse problems in pore-scale imaging.

### 5.2.1 Step 1: Preconditioning by pure regression on image data

The first sampling step 1 of the sequential splitting strategy aims to provide a preconditioning description of the surrogate micro-porosity field  $\varepsilon_\Theta$ . We consider a task differentiation between  $\mathcal{D}^S$  and  $\text{RAI}^+$  and hence, we discard from the training the fluid measurements which are far from the mineral interface — as they are of no interest in this first step to characterize morphological uncertainty on  $\varepsilon$ . The resulting potential energy term writes:

$$U_1(\Theta) = \frac{\lambda_0}{2\sigma_0^2} \|1 - \varepsilon_\Theta - \text{Im}\|_{\mathcal{D}^S}^2 + \frac{\lambda_1}{2\sigma_1^2} \|1 - \varepsilon_\Theta - \text{Im}\|_{\text{RAI}^+}^2 + \frac{1}{2\sigma_\Theta^2} \|\Theta\|^2 \quad (31)$$

where  $\sigma_k$  are unknown standard deviations characterizing the noise distributions on their respective areas, and we assume a prior distribution on  $\Theta$  given by  $P(\Theta) \sim \mathcal{N}(0, \sigma_\Theta^2 I_{p+d})$ . The notation  $\|\cdot\|$  refers to either the RMS (root mean square) norm — inherited from the functional  $\mathbb{L}^2$ -norm — for the two first

log-likelihood terms or to the usual Euclidean norm for the last log-prior term. In practice, we do not rely on a-priori manual calibration of the noise magnitudes — all  $\sigma_k$  are set to be equal — but rather use the AW-HMC sampler to automatically and adaptively estimate these uncertainties through adjustments of the  $\lambda_k$ . This is especially meaningful in the neighborhood of the evolving fluid/solid interface (corresponding to the RAI<sup>+</sup>) to study edge-enhancement implications and in the pure solid region to quantify the unresolved features. At the end of step 1, one gets a first a-priori estimation of the field  $\varepsilon$ , which presents the advantage of being denoised compared to the  $\mu$ CT images and hence is more suitable to differentiate. In this sense, we now have access to the time derivative of the surrogate porosity  $\varepsilon_\Theta$  that is subsequently used to provide some preconditioning of the latent concentration field  $C$ .

### 5.2.2 Step 2: Preconditioning of the latent reactive fluid

The second sampling step 2 relies on this first insight of the Bayesian neural network parameters obtained through step 1. We hence restart an adaptive weighting procedure with the AW-HMC sampler by adding additional constraints arising from the PDE model (19). As the acid concentration is a latent unknown field in our reactive inverse problem, we benefit from this second sampling step to provide a surrogate estimation  $C_\Theta$  of this field and identify a first reactive parameter, namely  $\text{Da}_{\text{II}}^*$ . In this sense, we impose a physics-based regularization linking the porosity derivative to the surrogate concentration field through the PDE equation:

$$\frac{1}{vC_0} \frac{\partial \varepsilon_\Theta}{\partial t} - \text{Da}_{\text{II}}^* C_\Theta \mathbb{1}_{\{(1-\varepsilon)>0\}} = 0 \quad (32)$$

where the calcite molar volume  $v$  and  $C_0$  are constant parameters — we assume the concentration  $C_0$  of continuous acid injection, defining the Dirichlet boundary conditions on  $C$ , to be known. In a direct formulation, equation (32) is imposed over the whole domain  $\Omega \times (0, T_f)$ , though the PDE constraint in inverse modeling mainly brings meaningful information on the RAI. Indeed, we have  $\frac{\partial \varepsilon_\Theta}{\partial t} \simeq -\frac{\partial \text{Im}}{\partial t} > 0$  on the reactive area of interest, which is useful to characterize the reaction regime through the  $\text{Da}_{\text{II}}^*$  dimensionless number. In the pure solid region where  $\frac{\partial \varepsilon_\Theta}{\partial t} = 0$ , the PDE constraint (32) translates into low acid penetration in  $Q_S$  that we impose through the condition  $C_\Theta = c_0 = 10^{-7}$ . In the fluid region  $Q_F$ , the latent acid concentration field is a solution of the following heat equation:

$$\frac{\partial C_\Theta}{\partial t} - D_m^* \Delta C_\Theta = 0 \quad (33)$$

with initial and boundary conditions, respectively unknown in the inverse formulation and non-homogeneous Dirichlet boundary conditions (see the dimensionless PDE model (19) from Sect. 3.3). Following the modeling assumptions of Sect. 3.2, especially on the diffusive dominated regime with  $\text{Pe} \ll 1$ , we though assume as a first approximation that the surrogate acid concentration field  $C_\Theta$  is driven by the quasy-stationary Poisson equation  $\Delta C_\Theta = 0$  in  $Q_F$ . This behaves as a continuity extension of the surrogate concentration field from the domain boundary to the mineral evolving interface defined by the RAI and furthermore serves as a first modeling approximation of the PDE system, which is necessary to infer the latent concentration field. Along with the PDE equation (32) in the RAI, this defines augmented multi-potential energy for the second sampling step:

$$\begin{aligned} U(\Theta) &= U_1(\Theta) + \frac{\lambda_2}{2\sigma_2^2} \left\| (vC_0)^{-1} \alpha_\Theta \frac{\partial \varepsilon_\Theta}{\partial t} - C_\Theta \right\|_{\text{RAI}}^2 + \frac{\lambda_3}{2\sigma_3^2} \|\Delta C_\Theta\|_{\mathcal{D}^F}^2 \\ &+ \frac{\lambda_4}{2\sigma_4^2} \left( \|1 - C_\Theta\|_{\mathcal{D}^\partial}^2 + \|c_0 - C_\Theta\|_{\mathcal{D}^S}^2 \right) + \frac{1}{2\sigma_\Theta^2} \|\Theta\|^2 \\ &:= U_1(\Theta) + U_2(\Theta) \end{aligned} \quad (34)$$

where the constant constraints on the boundary  $\mathcal{D}^\partial$  and solid  $\mathcal{D}^S$  datasets are gathered as a single term. The notation  $\alpha_\Theta$  here refers to the first inverse parameter effectively sampled, with  $\alpha_\Theta := (\text{Da}_{\text{II}}^*)^{-1}$ . This second step reinforces the sampling of the surrogate micro-porosity  $\varepsilon_\Theta$  by providing insight into the latent concentration field  $C_\Theta$  on the RAI and posterior distribution on the inverse parameter  $\text{Da}_{\text{II}}^*$ .

### 5.2.3 Step 3: Overall data assimilation potential with full reactive model

Finally, the third sampling step 3 will address the overall reactive inverse problem, to refine the micro-porosity and acid concentration predictions accounted for the exact and fully coupled PDE model (19) and provide uncertainty quantification on the inverse parameters  $D_{a_{II}}^*$  and  $D_m^*$ . The extension by continuity of the acid concentration — from the Poisson equation in  $Q_F$  — is replaced by its corresponding heat equation term (33) (see equation 37 bellow). We also use the diffusion-reaction PDE coupling  $\varepsilon_\Theta$  and  $C_\Theta$  to infer the dimensionless number  $D_m^*$ :

$$\frac{\partial C_\Theta}{\partial t} - D_m^* \nabla \cdot \left( \varepsilon_\Theta^{1+\beta} \nabla (\varepsilon_\Theta^{-1} C_\Theta) \right) + D_{a_{II}}^* C_\Theta = \frac{\partial C_\Theta}{\partial t} - D_m^* \nabla \cdot \left( \varepsilon_\Theta^{1+\beta} \nabla (\varepsilon_\Theta^{-1} C_\Theta) \right) + \frac{1}{vC_0} \frac{\partial \varepsilon_\Theta}{\partial t} = 0 \quad (35)$$

which is theoretically valid on the whole RAI for the inverse modeling. Nonetheless, the heterogeneous diffusion term  $\mathcal{D}_i(\varepsilon, C) := \nabla \cdot (\varepsilon^{1+\beta} \nabla (\varepsilon^{-1} C))$  arising from Archie's law becomes highly sensitive at the mineral boundary due to jumps in the porosity derivatives at the interface. This may disrupt the identification of the inverse parameter  $D_m^*$ . The PDE constraint (35) therefore needs to be imposed on a reduced neighborhood of the reactive area of interest, namely the  $\text{RAI}^-$  domain.

This restricted RAI is then defined by the eligible points of the RAI domain where  $D_m^*$  is predicted positive. From the overall samples of step 2, we compute the predictive BMA distributions of the two operators

$$\frac{\partial C_\Theta}{\partial t} + \frac{1}{vC_0} \frac{\partial \varepsilon_\Theta}{\partial t} \quad \text{and} \quad \mathcal{D}_i(\varepsilon_\Theta, C_\Theta) \quad (36)$$

on the domain  $\Omega \times (0, T_f)$  and then estimate  $D_m^*$  through equation (35) to define the  $\text{RAI}^-$  domain.

From this procedure, one also gets an estimate of the posterior distribution of  $D_m^*$  after sampling step 2 which is regarded as an initial a-priori on this inverse parameter in step 3. This will be further detailed in the applications (see Sect. 6 and 7). Taken together, the fully reinforced multi-potential energy for the third sampling step writes:

$$\begin{aligned} U(\Theta) &= U_1(\Theta) + \frac{\lambda_2}{2\sigma_2^2} \left\| (vC_0)^{-1} \alpha_\Theta \frac{\partial \varepsilon_\Theta}{\partial t} - C_\Theta \right\|_{\text{RAI}}^2 + \frac{\lambda_3}{2\sigma_3^2} \left\| \gamma_\Theta \frac{\partial C_\Theta}{\partial t} - \Delta C_\Theta \right\|_{\mathcal{D}^F}^2 \\ &+ \frac{\lambda_4}{2\sigma_4^2} \left( \|1 - C_\Theta\|_{\mathcal{D}^\theta}^2 + \|c_0 - C_\Theta\|_{\mathcal{D}^s}^2 \right) \\ &+ \frac{\lambda_5}{2\sigma_5^2} \left\| \gamma_\Theta \left( \frac{\partial C_\Theta}{\partial t} + (vC_0)^{-1} \frac{\partial \varepsilon_\Theta}{\partial t} \right) - \nabla \cdot \left( \varepsilon_\Theta^{1+\beta} \nabla (\varepsilon_\Theta^{-1} C_\Theta) \right) \right\|_{\text{RAI}^-}^2 + \frac{1}{2\sigma_\Theta^2} \|\Theta\|^2 \\ &:= U_1(\Theta) + \widetilde{U}_2(\Theta) + U_3(\Theta) \end{aligned} \quad (37)$$

where  $\gamma_\Theta := (D_m^*)^{-1}$  such that the set of inverse parameters we infer in practice is  $(\mathcal{P}_{\text{inv}})_\Theta = \{\alpha_\Theta, \gamma_\Theta\}$ . The data assimilation strategy developed in the present article incorporates successive physics-based constraints using a sequential reinforcement of the multi-potential energy  $U(\Theta)$ . This is achieved by splitting the sampling steps, which is required due to the strong coupling of the overall PDE system (19) involving latent field and unknown parameters. This overall algorithm is summarized in Fig 2.

### 5.3 Computational strategy for differential operator expression

This section is dedicated to the development of a differentiation strategy for efficient computation of the heterogeneous diffusion  $\mathcal{D}_i(\varepsilon, C)$  arising from Archie's law. Indeed, the third sampling step in the sequential reinforcement of the multi-potential energy (see Sect. 5.2) involves the computation of this diffusion operator through a neural network surrogate model. This implies the use of automatic differentiation (AD) which is a prevalent technique in deep-learning frameworks such as Physics-Informed Neural Networks (PINNs) and Bayesian Physics-Informed Neural Networks (BPINNs). Such an automatic differentiation relies on gradient backpropagation to compute the derivatives of the neural network functional outputs with respect to its inputs, by using the chain rule principle. AD is thus a fast computational technique when it comes to the evaluation of first and second-order derivatives of the output fields, namely the spatial gradient and

a) Comparison of diffusion operators on the 1D+Time

	$T_{\text{CPU}}$ (ms)	Speedup $S$	$T_{\text{GPU}}$ (ms)	Speedup $S$
Original operator $\mathcal{D}_i(\varepsilon, C)$	37.13	1	11.32	3.28
Developed operator (38c) with $\beta \neq 1$	24.78	1.49	6.985	5.32
Developed operator (38c) with $\beta = 1$	10.66	3.48	6.739	5.51

b) Comparison of diffusion operators on the 2D+Time

	$T_{\text{CPU}}$ (ms)	Speedup $S$	$T_{\text{GPU}}$ (ms)	Speedup $S$
Original operator $\mathcal{D}_i(\varepsilon, C)$	83.88	1	13.69	6.12
Developed operator (38c) with $\beta \neq 1$	72.63	1.16	10.63	7.89
Developed operator (38c) with $\beta = 1$	68.99	1.22	10.59	7.92

Table 1: **Computational times of the diffusion operators on the 1D+Time and 2D+Time inverse problem:** comparison between several expressions of the differential operator for CPU and GPU implementations. The Speedup is computed as  $S = T_0/T_\bullet$  with  $T_0 := T_{\text{CPU}}$  for the original expression  $\mathcal{D}_i(\varepsilon, C) = \nabla \cdot (\varepsilon^{\beta+1} \nabla(\varepsilon^{-1} C))$ . All the computational times are expressed in milliseconds (ms) and are averaged on several evaluations of the diffusion operator during the sampling procedure.

Laplacian operators, and the temporal partial derivatives. More complex non-linear operators resulting from two successive differentiation of non-trivial functional compositions — as this is the case for the  $\mathcal{D}_i(\varepsilon, C)$  operator — can though readily lead to high-computational cost. This observation leads to reconsidering the heterogeneous diffusion term as a succession of sum and product of first and second-order operators. Consequently, we consider the diffusion operator  $\mathcal{D}_i(\varepsilon, C)$  under two formulations: its compact form (38a) and its developed form (38c) reading

$$\mathcal{D}_i(\varepsilon, C) = \nabla \cdot (\varepsilon^{\beta+1} \nabla(\varepsilon^{-1} C)) \quad (38a)$$

$$= \nabla \cdot (\varepsilon^\beta \nabla C - \varepsilon^{\beta-1} C \nabla \varepsilon) = \nabla \cdot (\varepsilon^{\beta-1} (\varepsilon \nabla C - C \nabla \varepsilon)) \quad (38b)$$

$$= \varepsilon^{\beta-1} (\varepsilon \Delta C - C \Delta \varepsilon) + (\beta - 1) \varepsilon^{\beta-1} \nabla \varepsilon \cdot \nabla C + (\beta - 1) \varepsilon^{\beta-2} C \nabla \varepsilon \cdot \nabla \varepsilon \quad (38c)$$

Then we replace the expression of the diffusion in the multi-potential energy (37) with the novel operator formulation (38c). This makes possible to reduce the computational cost of evaluating this diffusion operator through merely the auto differentiation of the following terms:  $\nabla \varepsilon_\Theta$ ,  $\nabla C_\Theta$ ,  $\Delta \varepsilon_\Theta$ , and  $\Delta C_\Theta$ . Finally, we observe on the developed expression (38c) that the case  $\beta = 1$  even results in a more straightforward expression of Archie’s law which then writes  $\mathcal{D}_i(\varepsilon, C) = \varepsilon \Delta C - C \Delta \varepsilon$ . This is particularly convenient as the tortuosity index  $\beta = 1$  can be regarded as an approximation of the effective diffusivity in pore-scale models (*e.g.* see [67]). Furthermore, this reduced expression confirms the high sensitivity of the heterogeneous diffusion term at the mineral boundary  $\sigma$  due to the micro-porosity Laplacian involved in Archie’s law. Considering suitable differential operator expressions can enhance the surrogate model efficiency by reducing the automatic differentiation cost.

We perform some validations of this first insight through computational time measurements for the different expressions of the heterogeneous diffusion term. We hence compare the original formulation of  $\mathcal{D}_i(\varepsilon, C)$  with the developed operator (38c) for tortuosity coefficients  $\beta = 0.5$  — to consider the most general form — and  $\beta = 1$  that leads to the reduced evaluation of the heterogeneous diffusion. We account for computational times on both CPU and GPU devices and perform several evaluations of the diffusion operator expressions to provide averaged computational times along the sampling procedure of step 3. These results are summarized in Table 1 for the 1D+Time test case (Table 1.a) and 2D+Time application (Table 1.b) explicitly developed in the validation Sect. 6 and application Sect. 7. We define in each case the reference computational cost  $T_0$  as the CPU time necessary to evaluate the original operator expression  $\mathcal{D}_i(\varepsilon, C)$ .

This respectively leads to  $T_0 = 37.13$  ms and  $T_0 = 83.88$  ms for 1D+Time and 2D+Time applications. We then evaluate the speedup, denoted  $S$ , of the distinct operator formulations as  $S = T_0/T_\bullet$ , where  $T_\bullet$  are their respective computational times. It results from this comparison an effective improvement of the computational costs, either on CPU or GPU, when considering the developed operator (38c) from equation (38) — even in its general form for  $\beta \neq 1$  (see second rows of Table 1.a and 1.b). This highlights that the configuration that best optimizes the speedup is to use the operator (38c) on GPU devices. The improvement between the general and reduced form of the operator (38c) is, however, less significant especially in 2D+Time. This can be readily explained by the fact that most of the computational time is spent evaluating the gradient and Laplacian operators rather than their combination. In this sense, the general form (38c) can be used effectively regardless of the tortuosity index value  $\beta$ . Overall, the developed heterogeneous diffusion operator (38c) contributes to reducing the computational cost of its single evaluation.

Suitable choices in the differential operator expressions considerably improve the AD cost when considering complex non-linear operators with non-trivial functional compositions, as for Archie’s law. This, therefore, reduces the computational time spent in evaluating one instance of these operators. Such improvements are meaningful, accounting for the numerous surrogate model evaluations required for an appropriate sampling of the target posterior distribution (20) through MCMC samplers. Considering the developed formulation (38c) of the heterogeneous diffusion operator on GPU devices hence implies a significant enhancement of the overall computational time of the present data assimilation strategy.

## 6 Validation on synthetic 1D+Time calcite dissolution

In this article, we develop a novel data assimilation strategy to address reactive inverse problems in pore-scale imaging with uncertainty quantification. This aims to quantify morphological uncertainty on the micro-porosity field  $\varepsilon$  and estimate reliable ranges of chemical parameters through dynamical  $\mu$ CT noisy observations augmented with PDE models of dissolution. The present strategy is based on the robust Bayesian framework presented in [55] along with the AW-HMC sampler (see Sect. 4.2) and relies on sequential reinforcement of the multi-potential energy.

In this section, we validate the present methodology on inverse problems of calcite dissolution with heterogeneous porosity in artificial 1D spatial configurations. All the  $\mu$ CT measurements that we consider are synthetic observations resulting from direct numerical simulations of reactive flows with noise perturbation, in order to validate our methodology in well-established test cases. The validation test case is a purely synthetic 1D+Time problem, for which we check two configurations with distinct tortuosity indices with  $\beta = 1$  and  $\beta = 0.5$ .

### 6.1 Direct reactive model: problem set up

We consider two heterogeneous samples of 1D synthetic calcite cores whose initial geometries are characterized by the numerical  $\mu$ CT images presented in Fig 3 on a ‘physical’ spatial domain  $\Omega$  of width 0.3 mm. These initial images correspond to normalized greyscale tomographic scans, corrupted with noise that either accounts for sensor noise or unresolved morphological features. Direct numerical simulations (DNS) of reactive processes are then performed on these initial geometries to provide synthetic  $\mu$ CT dynamical images of dissolution. These observation data are generated by solving the reaction-diffusion system (15) by means of mesh-based and particle methods — namely a Backward Euler or Mid Point method for the time integration coupled with Particle Strength Exchanges scheme for the heterogeneous diffusion — on a Cartesian spatiotemporal grid of resolution  $N_x = 200$  and  $N_t = 240$ . Continuous acid injection is maintained through non-homogeneous Dirichlet boundary conditions on the domain  $\Omega$  to ensure a diffusive-dominated regime. Numerically, we consider a strong acid solution with pH = 0 such that the normalizing constant  $C_0$  equals 1. The characteristic length  $L$  of these porous samples is set to  $L = 0.1$  mm and the reactive parameters are respectively defined by  $K_s = 0.8913$  mol.m<sup>-2</sup>.s<sup>-1</sup>,  $D_m = 10^{-9}$  m<sup>2</sup>.s<sup>-1</sup>, and  $\gamma_{H^+} = 10^{-3}$  m<sup>3</sup>.mol<sup>-1</sup> — taken from the benchmark [44]. The reactive specific area  $A_s$  is set to  $A_s = 10^3$  m<sup>-1</sup>, and we do not account for the calcite molar volume  $v$  in these test cases — as such 1D+Time examples do not mean to be physically consistent but rather serve validation purposes. We also consider distinct tortuosity indexes, namely  $\beta = 1$  and  $\beta = 0.5$ , on the different geometries to address both the compact form (38a) and develop



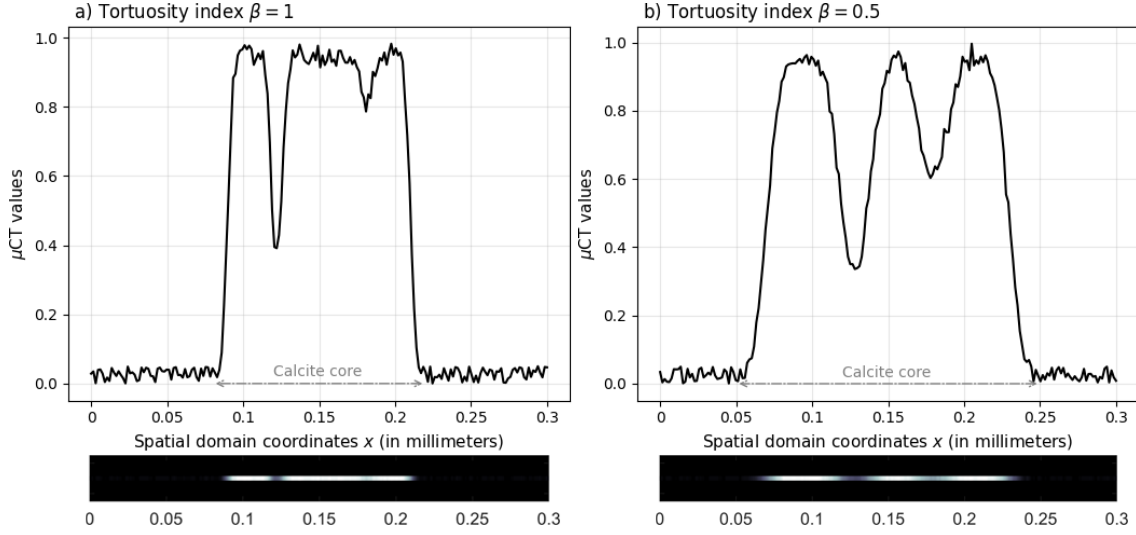


Figure 3: **Initial  $\mu$ CT images defining the porous sample geometries:** synthetic cases with tortuosity indices a)  $\beta = 1$  and b)  $\beta = 0.5$ . The  $\mu$ CT measurements are normalized, corrupted with noise, and provide the dataset  $\mathcal{I}_m$  before the dissolution process. The calcite core regions are identified by the double-headed arrows, and correspond to the maximum intensity in the greyscale tomographic scans displayed below.

form (38c) of the diffusion operator  $\mathcal{D}_i(\varepsilon, C)$  in the data assimilation problem. The DNS is performed until the overall calcite core is dissolved which corresponds to a characteristic final time  $T_f = 24$  s. Taken together, one gets a sequence of synthetic  $\mu$ CT images  $\mathcal{I}_{m_{i,j}}$ , similar to Fig 3, characterizing the dissolution process of the two calcite cores on the 'physical' spatiotemporal domain  $\Omega \times (0, T_f)$ .

## 6.2 Dimensionless inverse problem and dimensionless numbers

From the setting of these reactive parameters, one identifies the dissolution regime of these test cases given by the dimensionless catalytic Damköhler  $\text{Da}_{\text{II}} = 8.913$  from equation (13). The inference of this Damköhler number is though not straightforward in inverse problems, as developed in Sect. 3.3, and we define the dimensionless time  $t^*$  such that its related final time is  $T_f^* = 1$ . The dimensionless spatial variable  $x^*$  is computed as in the dimensionless formulation of the direct problem using  $x^* = x/L$ . For the data assimilation, we hence consider the dimensionless domain  $\Omega^* \times (0, T_f^*) = [0, 3] \times (0, 1)$  to extract the observation dataset

$$\mathcal{D} = \{(x_k, t_k) \in [0, 3] \times (0, 1), \quad k = 1 \dots N_{\text{obs}}\} \quad (39)$$

where the number of training points  $N_{\text{obs}} = 7725$  represents about 16% of the data required for the full field reconstructions on  $\Omega^* \times (0, T_f^*)$ . The dataset  $\mathcal{D}$  is divided into the corresponding datasets  $\mathcal{D}^S$ ,  $\mathcal{D}^F$ ,  $\mathcal{D}^{\text{RAI}}$ ,  $\text{RAI}^+$ ,  $\text{RAI}^-$  and  $\mathcal{D}^\partial$  that respectively cover around 50%, 4%, 15%, 13%, 10% and 8% of the  $N_{\text{obs}}$  training measurements. One also determines the scaling dimensionless factor  $D_{\text{ref}}$  which appears in the dimensionless inverse formulation (19):

$$D_{\text{ref}} = \frac{T_f^* L^2}{T_f} = 4.16 \times 10^{-10} \text{ m}^2 \cdot \text{s}^{-1} \quad (40)$$

according to the relation (16) and the estimations of  $T_f$ ,  $T_f^*$  and  $L$ . We characterize the dissolution regime in these reactive inverse problems by means of the two dimensionless numbers defined in (18), and one gets:

$$\text{Da}_{\text{II}}^* = 21.3912 \quad \text{and} \quad D_m^* = 2.4 \quad (41)$$

which are related to the inverse parameters to infer through  $(\mathcal{P}_{\text{inv}})_\Theta = \{\alpha_\Theta, \gamma_\Theta\} = \{(\text{Da}_{\text{II}}^*)^{-1}, (D_m^*)^{-1}\}$ . For such 1D+Time reactive inverse problems, the overall multi-potential energy finally writes in the case

$\beta = 1$ :

$$\begin{aligned}
U(\Theta) = & \frac{\lambda_0}{2\sigma_0^2} \|1 - \varepsilon_\Theta - \text{Im}\|_{\mathcal{D}^S}^2 + \frac{\lambda_1}{2\sigma_1^2} \|1 - \varepsilon_\Theta - \text{Im}\|_{\text{RAI}^+}^2 + \frac{\lambda_2}{2\sigma_2^2} \left\| \alpha_\Theta \frac{\partial \varepsilon_\Theta}{\partial t} - C_\Theta \right\|_{\text{RAI}}^2 \\
& + \frac{\lambda_3}{2\sigma_3^2} \left\| \gamma_\Theta \frac{\partial C_\Theta}{\partial t} - \frac{\partial^2 C_\Theta}{\partial x^2} \right\|_{\mathcal{D}^F}^2 + \frac{\lambda_4}{2\sigma_4^2} \left( \|1 - C_\Theta\|_{\mathcal{D}^\theta}^2 + \|10^{-7} - C_\Theta\|_{\mathcal{D}^S}^2 \right) \\
& + \frac{\lambda_5}{2\sigma_5^2} \left\| \gamma_\Theta \left( \frac{\partial C_\Theta}{\partial t} + \frac{\partial \varepsilon_\Theta}{\partial t} \right) - \left( \varepsilon_\Theta \frac{\partial^2 C_\Theta}{\partial x^2} - C_\Theta \frac{\partial^2 \varepsilon_\Theta}{\partial x^2} \right) \right\|_{\text{RAI}^-}^2 + \frac{1}{2\sigma_\Theta^2} \|\Theta\|^2
\end{aligned} \tag{42}$$

and is sequentially reinforced, as presented in Sect. 5.2, through three successive sampling steps using the AW-HMC sampler. The hyperparameters setting of the sequential AW-HMC samplers together with the neural network architecture are detailed hereafter.

### 6.3 Deep learning configuration

Regarding the deep learning strategy for the overall data assimilation problem, we use two distinct neural network architectures to define the micro-porosity and acid concentration surrogate models. Each surrogate model has, therefore, one single output corresponding to  $\varepsilon_\Theta$  and  $C_\Theta$ , respectively. This is preferred to merging the two outputs into a single neural network architecture to avoid a strong correlation between the output fields. Indeed, providing surrogate models not strongly correlated with a multiple-output neural network may require numerous hidden layers that straightforwardly impact the overall computational cost. On the contrary, using distinct neural networks makes it possible to build independent surrogate models while retaining few hidden layers and, therefore, a reasonable number of neural network parameters. Correlations between the two neural network architectures, and then the outputs fields  $\varepsilon_\Theta$  and  $C_\Theta$ , are merely achieved through the PDE model defining the multi-potential energy (42). This deep learning configuration is more meaningful for such a reactive data assimilation problem since a high correlation — due to the neural network architecture — with the latent field  $C_\Theta$  can highly disrupt the micro-porosity recovery. Relying on the PDE model to ensure relevant correlation hence appears as the most appropriate strategy.

The first neural network establishing the micro-porosity surrogate model  $\varepsilon_\Theta$  is composed of 4 hidden layers with 32 neurons per layer and a hyperbolic tangent activation function. The output layer is complemented by a rectified hyperbolic tangent  $\text{Tanh}^r(z) = 0.5(\text{Tanh}(z) + 1)$  to ensure output values between 0 and 1. This neural network complexity provides the best approximation of the micro-porosity during the first sampling step 1 while maintaining moderate computational costs. Indeed, we analyze in Fig 4 the impact of the neural network architecture both on the computational time spent on the sampling procedure and the Bayesian Model Average (BMA) accuracy, computed as:

$$\text{BMA-E}^\varepsilon = \|P(\varepsilon_\Theta | (x, t), \mathcal{D}, \mathcal{M}) - \varepsilon\|_{\Omega^* \times (0, T_f^*)}^2 = \|P(\varepsilon_\Theta | (x, t), \mathcal{D}, \mathcal{M}) - (1 - \text{Im})\|_{\Omega^* \times (0, T_f^*)}^2 \tag{43}$$

where the notation  $\|\cdot\|$  used here refers to the functional  $\mathbb{L}^2$ -norm and  $P(\varepsilon_\Theta | (x, t), \mathcal{D}, \mathcal{M})$  is the BMA approximation from equation (21) (e.g. see [75] or [55] for more details). For the optimal NN configuration, one estimates the sampling computational cost, providing the overall samples of the posterior distribution (20) based on the  $N_{\text{obs}}$  training measurements, to about 17 min. The BMA prediction obtained through equation (21), over the whole computational domain  $\Omega^* \times (0, T_f^*)$  is meanwhile immediate — less than 1 s on GPU and a few seconds on CPU. Recovering the latent concentration  $C_\Theta$  will require less neural network expressivity compared to the micro-porosity field which needs to integrate noisy data, and unresolved morphological features in its reconstruction. In this sense, we assume that the second neural network defining the acid surrogate model is composed of 3 hidden layers with 32 neurons per layer and a hyperbolic tangent activation function. This is only one layer less compared to the first NN for  $\varepsilon_\Theta$  but enables saving 1056 parameters. The number of network parameters is, therefore, 3297 for the first sampling and increases to 5538 for the second and third sampling steps with the two additional inverse parameters.

The other hyperparameters concerning the AW-HMC sampler are summarized in Table 2 for the sequential sampling steps. These sampler parameters involve, *inter alia*, several adaptive steps  $N$  during which

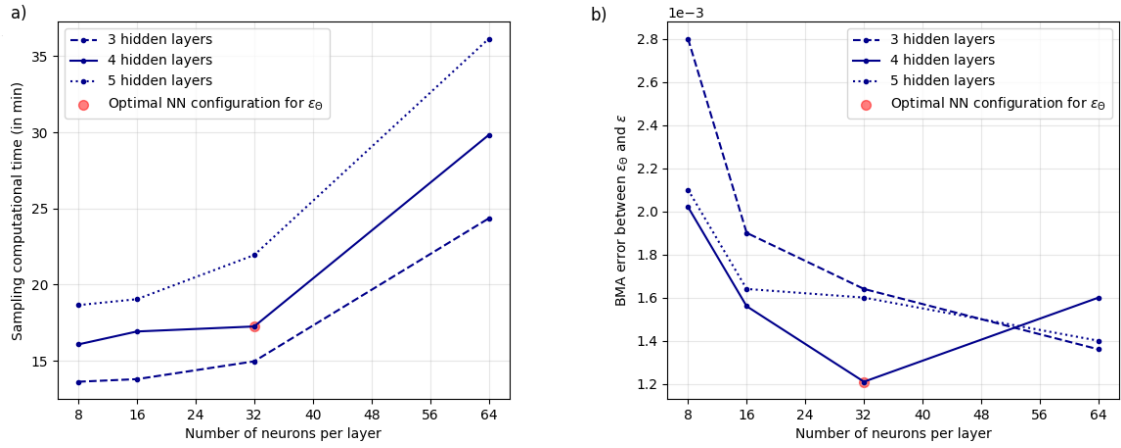


Figure 4: **Neural Network configuration choice for the surrogate model on the micro-porosity field:** a) Computational cost of sampling step 1 with respect to the neural network architectures both in terms of the number of hidden layers and neurons per layer. b) Bayesian Model Average (BMA) error between the surrogate model  $\varepsilon_\Theta$  and groundtruth  $\varepsilon$ , computed as in equation (43), for different neural network architectures.

Sampling step	Number of adaptive steps $N$	Number of samples $N_s$	Number of leapfrog steps $L$	Leapfrog time step $\delta t$
1) Preconditioning $\varepsilon_\Theta$	50	200	200	$1 \times 10^{-3}$
2) Preconditioning $C_\Theta$ + Inference $\alpha_\Theta$	20	200	200	$5 \times 10^{-4}$
3) Full data assimilation	4	200	200	$2 \times 10^{-4}$

Table 2: **AW-HMC hyperparameters on the 1D+Time reactive inverse problem:** Setting of the sampler hyperparameters for the three sequential sampling steps defined in Fig 2. The number of adaptive steps  $N$  along with the leapfrog parameters  $L$  and  $\delta t$  are AW-HMC sampler-specific parameters.

the critical weights  $\lambda_k$  are automatically adjusted through an Inverse Dirichlet basis using the equation (26), and a number of overall sampling steps  $N_s$ . The two other parameters, namely  $L$  and  $\delta t$ , are intrinsically related to the Hamiltonian Monte Carlo structure of the AW-HMC sampler. Indeed, they are involved in the deterministic step that relies on the leapfrog symplectic integrator to solve for the Hamiltonian dynamical system (24) (see Sect. 4.1). We also refer to our methodological article [55] for more details on these hyperparameters and especially to Algorithm 1 (Adaptively Weighted Hamiltonian Monte Carlo) for their respective role in the sampling phases.

## 6.4 Numerical results

We demonstrate the validity of our data assimilation approach on synthetic inverse problems of calcite dissolution whose initial core geometries are characterized in Fig 3, and for which the dynamical  $\mu$ CT images are provided through DNS (see Sect. 6.1). In the data assimilation Bayesian framework, we first select log-normal prior distributions on  $(\mathcal{P}_{\text{inv}})_\Theta = \{\alpha_\Theta, \gamma_\Theta\}$ , which ensures the positivity of the inverse parameters, and independent normal distribution for the neural network parameters  $\theta$ . Nonetheless, appropriate change of variables on the inverse parameters, namely  $\bullet_\Theta = e^{\bullet_\Theta}$ , makes it possible to consider Gaussian prior distributions on the newly defined set of parameter  $\Theta = \{\theta, \tilde{\mathcal{P}}_{\text{inv}}\}$  (e.g. see [79] or [55]). This is the underlying hypothesis considered when defining the log-prior term in the potential energy (42), where we assume  $P(\Theta) \sim \mathcal{N}(0, \sigma_\Theta^2 I_{p+d})$ . In practice, we use the standard deviation  $\sigma_\Theta = 10$  in the applications such that slightly diffuse distribution induces weakly informed priors on the  $\Theta$  parameters. We also impose weakly informed priors on the inverse parameters such that we do not rely on biased a-priori on their respective scaling. In this sense, we benefit from the AW-HMC sampler advantages to handle multiscale inverse problems with unknown informative priors. We also avoid hand-tuning of the distinct task uncertainties by setting all the  $\sigma_k$ ,  $k = 0 \dots 5$ , to be equal in equation (42). On the contrary, automatic adjustment of the

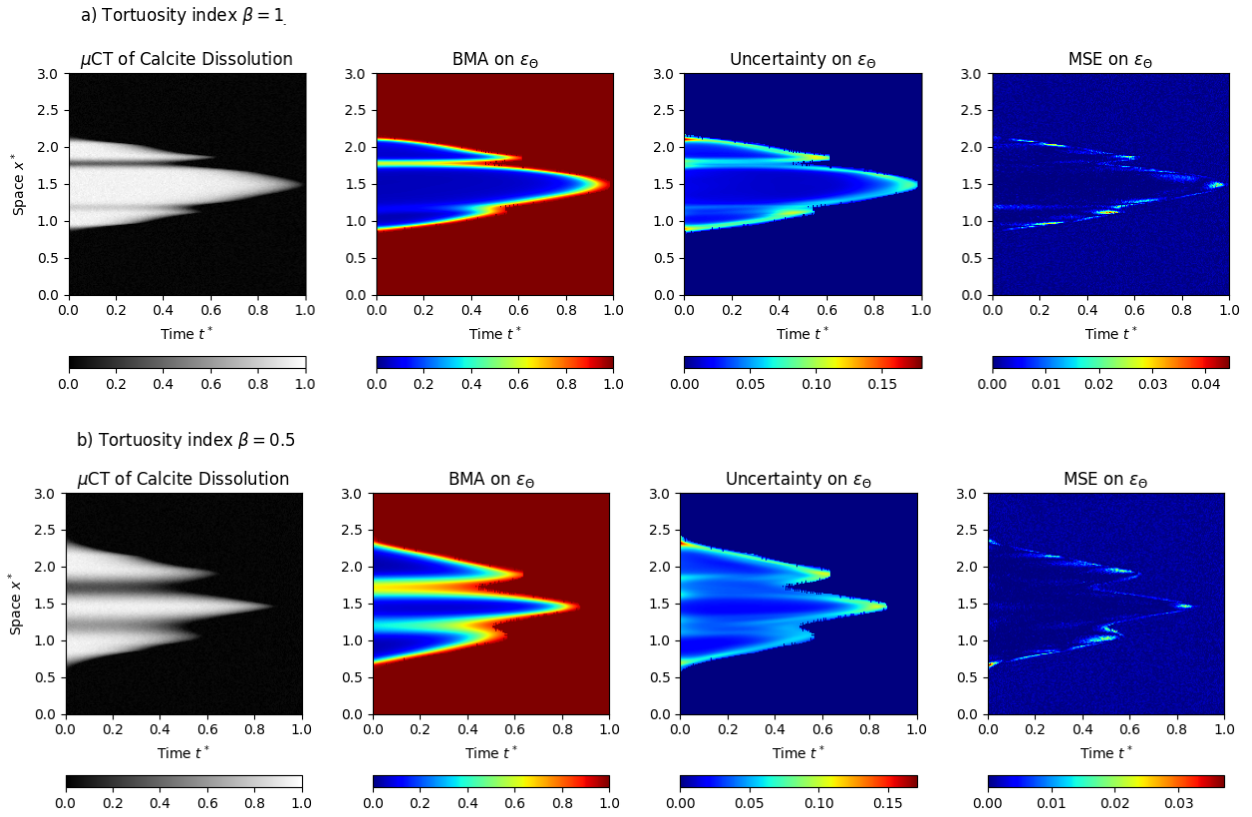


Figure 5: **Uncertainty Quantification on 1D+Time reactive inverse problem with data assimilation:** Bayesian Model Average (BMA) predictions on the micro-porosity field  $\varepsilon_{\Theta}$  with their local uncertainties — given by the standard deviation on the posterior distribution of the predictions — and mean squared errors (MSE). The top row corresponds to the initial geometry from Fig 3a with tortuosity index  $\beta = 1$ . The bottom row is related to the initial porous sample from Fig 3b with  $\beta = 0.5$ .

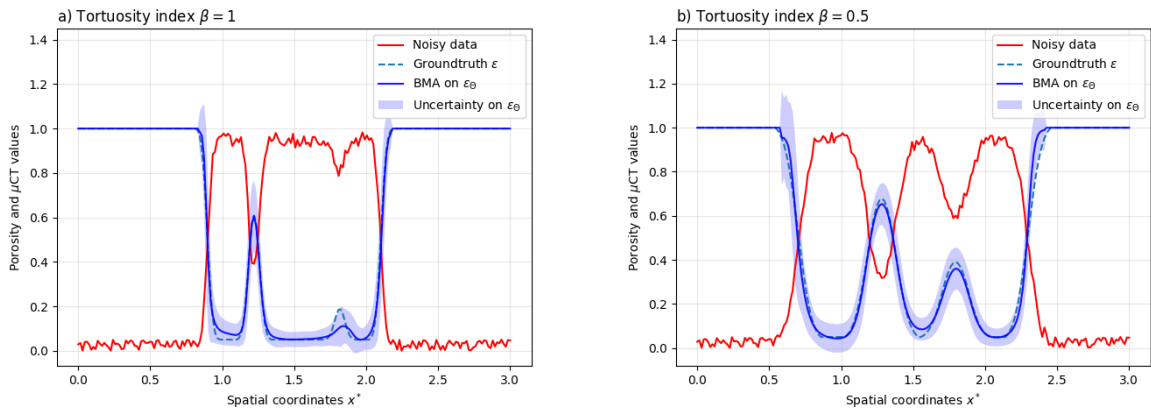


Figure 6: **Uncertainty Quantification on the micro-porosity field at the initial state ( $t^* = 0$ ) for 1D+Time reactive inverse problem:** Corrupted  $\mu$ CT image before dissolution, groundtruth on  $\varepsilon$ , BMA prediction, and uncertainty on  $\varepsilon_{\Theta}$  plotted along the spatial dimensionless coordinates  $x^*$ . Validation test cases with tortuosity indexes a)  $\beta = 1$  and b)  $\beta = 0.5$ .

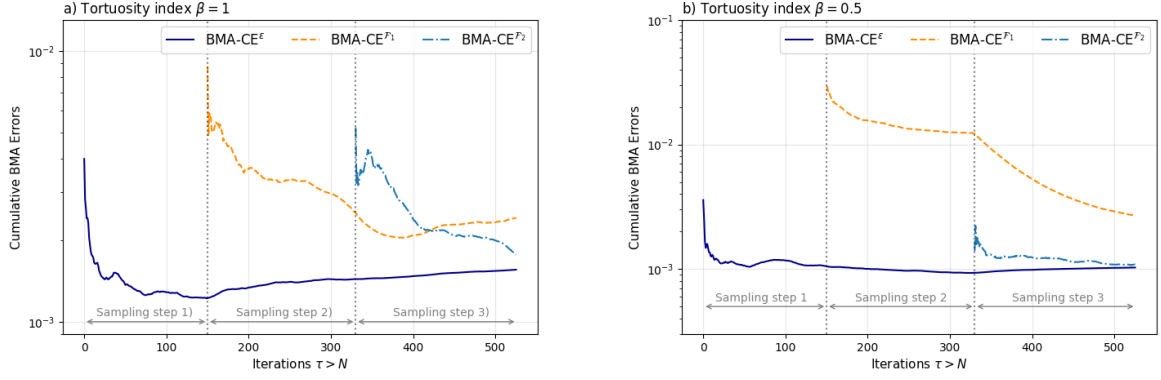


Figure 7: **Bayesian Model Average Cumulative Error diagnostics (BMA-CE)**, as defined in equations (45)-(47), for **1D+Time reactive inverse problem**: BMA-CE on the micro-porosity field prediction  $\varepsilon$  throughout the sampling iterations, and BMA-CE on the PDE constraint residuals  $\mathcal{F}_1$  and  $\mathcal{F}_2$  defined in (46) and introduced successively. The dotted vertical lines split the sampling steps in the sequential reinforcement of the multi-potential energy (42). Dissolution inverse problem on the initial geometries from a) Fig 3a with tortuosity index  $\beta = 1$  and b) Fig 3b with tortuosity index  $\beta = 0.5$ .

weighting parameters  $\lambda_k$  will provide intrinsic task uncertainties during the sampling procedure.

From the overall sampling procedure, we first obtain through equation (21) a Bayesian Model Average prediction on the porosity field  $\varepsilon_\Theta$  which is approximated by (e.g. see [75]):

$$P(\varepsilon_\Theta | (x, t), \mathcal{D}, \mathcal{M}) \simeq \frac{1}{N_s - N} \sum_{\tau=N}^{N_s} P(\varepsilon_\Theta | (x, t), \Theta^{t_\tau}) \quad (44)$$

where  $P(\varepsilon_\Theta | (x, t), \Theta^{t_\tau})$  is the surrogate model prediction of the micro-porosity resulting from the sampling iteration  $\tau$  for the set of parameters  $\Theta$  — including both the neural network and inverse parameters. Similarly, one can also compute local uncertainties on the output porosity field given the standard deviation metric on the posterior distribution of the predictions. These uncertainty quantification results are presented in Fig 5 along the whole dissolution time  $t^*$  for both the initial core geometries with distinct tortuosity indexes. We also compare the local uncertainties on the micro-porosity field with the traditional Mean Squared Errors (MSE) between the BMA surrogate prediction obtained by equation (44) and the groundtruth  $\varepsilon$ . This shows enhanced mean squared errors on the core edges during the calcite core dissolution which are, however, embedded in the local uncertainties. The latter uncertainties also tend to increase in these regions, characterizing the challenge of capturing reliable core interfaces from the dynamical  $\mu$ CT images. In this sense, one can query the confidence of mineral reactivity assessment using merely differential imaging techniques on the dynamical  $\mu$ CT scans. From the dynamical observation of the calcite core dissolution, we obtain uncertainties on the initial state geometry represented in Fig 6. This shows that the posterior prediction on  $\varepsilon_\Theta$  covers the groundtruth micro-porosity field  $\varepsilon$  and provides upper and lower bounds for the residual, potentially unresolved, micro-porosity  $\varepsilon_0$  estimation in the porous matrix — e.g.  $1.8\% \leq \varepsilon_0 \leq 9\%$  in the case  $\beta = 1$  for the 95% confidence interval, corresponding to approximately two standard deviations.

Moreover, we rely on the Bayesian Model Average Cumulative Error metric, denoted BMA-CE and introduced in [55], to quantify the sampling efficiency in terms of convergence along the marginalization process. We first compute the BMA-CE diagnostics for the porosity field  $\varepsilon$  based on:

$$\text{BMA-CE}^\varepsilon(\tau) = \left\| \frac{1}{\tau - N} \sum_{i=N}^{\tau} P(\varepsilon_\Theta | (x, t), \Theta^{t_i}) - (1 - \text{Im}) \right\|^2 \quad \forall \tau > N. \quad (45)$$

Equation (45) hence defines for each sampling step  $\tau$ , after the adaptive steps, a cumulative error characterizing the convergence of the BMA model toward the groundtruth  $\varepsilon$ . Such a diagnostic is computed during the overall sampling procedure, covering the three successive steps of sequential reinforcement. In the same manner, we extend this notion of convergence to the PDE constraints by computing the BMA-CE metric on

their respective residuals. We, therefore, introduce  $\mathcal{F}_1$  and  $\mathcal{F}_2$  the PDE constraints residuals arising from the reactive model (19) and involved in the multi-potential energy (42):

$$\begin{aligned}\mathcal{F}_1(\varepsilon_\Theta, C_\Theta) &:= \alpha_\Theta \frac{\partial \varepsilon_\Theta}{\partial t} - C_\Theta \\ \mathcal{F}_2(\varepsilon_\Theta, C_\Theta) &:= \gamma_\Theta \left( \frac{\partial C_\Theta}{\partial t} + \frac{\partial \varepsilon_\Theta}{\partial t} \right) - \left( \varepsilon_\Theta \frac{\partial^2 C_\Theta}{\partial x^2} - C_\Theta \frac{\partial^2 \varepsilon_\Theta}{\partial x^2} \right)\end{aligned}\quad (46)$$

to finally define their corresponding diagnostics BMA-CE $\mathcal{F}_\bullet$ :

$$\text{BMA-CE}^{\mathcal{F}_\bullet}(\tau) = \left\| \frac{1}{\tau - N} \sum_{i=N}^{\tau} P(\mathcal{F}_\bullet(\varepsilon_\Theta, C_\Theta) | (x, t), \Theta^{t_i}) \right\|^2 \quad \forall \tau > N. \quad (47)$$

These metrics are respectively computed on the sampling steps 2 and 3 for the residuals  $\mathcal{F}_1$  and  $\mathcal{F}_2$ , and the resulting convergence curves are presented in Fig 7 for both initial geometries. Successively introducing the additional PDE constraints results in deviations in the BMA-CE $^\varepsilon$  curve compared to the purely data-fitting sampling step 1. As we aim for a physics-based data assimilation of imperfect  $\mu$ CT data and seek to incorporate information on the static initial data through the combined use of observations and dissolution modeling, we can anticipate that introducing physics-based constraints will alter the BMA-CE predictions on the porosity field. The slight increase observed in the BMA-CE $^\varepsilon$  curves in Fig 7 highlights that merely considering the regression step 1 leads to underestimated uncertainties and tends to overfit the data. This means that the PDE model constraints, successively introduced in step 2 and 3, brings information and dynamical insights into the porosity field  $\varepsilon$  recovery instead of providing overfitting predictions. One also gets that the PDE constraints are satisfied by the convergence of their residual BMA-CE curves.

Regarding the inverse parameters inference, we represent in Fig 8a the histograms of the marginal posterior distributions of  $\alpha_\Theta := (\text{Da}_{\text{II}}^*)^{-1}$  for the two initial geometries from Fig 3. These distributions are obtained throughout the sampling steps 2 and 3, and provide intrinsic uncertainties on the parameter estimations. One also gets from Fig 8b the parameter trajectories when exploring the phase space distribution (22), where the convergence toward the mode during the adaptive steps is represented in blue. The final sampling, corresponding to the phase diagram trajectories for  $\tau > N$ , thus ensures an efficient exploration of the parameter mode neighborhood. Following the sequential reinforcement strategy detailed in Sect. 5, one gets at the end of the sampling step 2 a description of the restricted RAI $^-$  domain altogether with an initial estimate on the distribution of  $D_m^*$ .

The latter is regarded as an initial a-priori on the parameter  $\gamma_\Theta$  in step 3, and is obtained as follows:

- for the sampling iteration  $\tau = N \dots N_s$  in step 2, we compute the (cumulative) predictive BMA distributions of the two operators

$$\frac{\partial C_\Theta}{\partial t} + \frac{1}{vC_0} \frac{\partial \varepsilon_\Theta}{\partial t} \quad \text{and} \quad \mathcal{D}_i(\varepsilon_\Theta, C_\Theta),$$

- we establish the RAI $^-_\tau$  as the admissible points of the RAI where  $D_m^*$  is predicted positive for each sampling iteration  $\tau > N$ ,
- we compute a spatially averaged estimation of the parameter  $(\overline{D_m^*})_\tau$  on this eligible domain, where

$$(\overline{D_m^*})_\tau = \frac{1}{\#\text{RAI}^-_\tau} \sum_{k \in \text{RAI}^-_\tau} (D_m^*)_\tau(x_k, t_k) \quad \forall \tau > N,$$

- we can estimate a global distribution on  $D_m^*$  throughout the overall samples of step 2 where we discard the distribution tail after the 80<sup>th</sup> percentile,
- we finally evaluate the prior distribution on the inverse parameter  $\gamma_\Theta := (D_m^*)^{-1}$  with its mean value  $\overline{\gamma_\Theta}$  used as an initial a-priori in step 3,

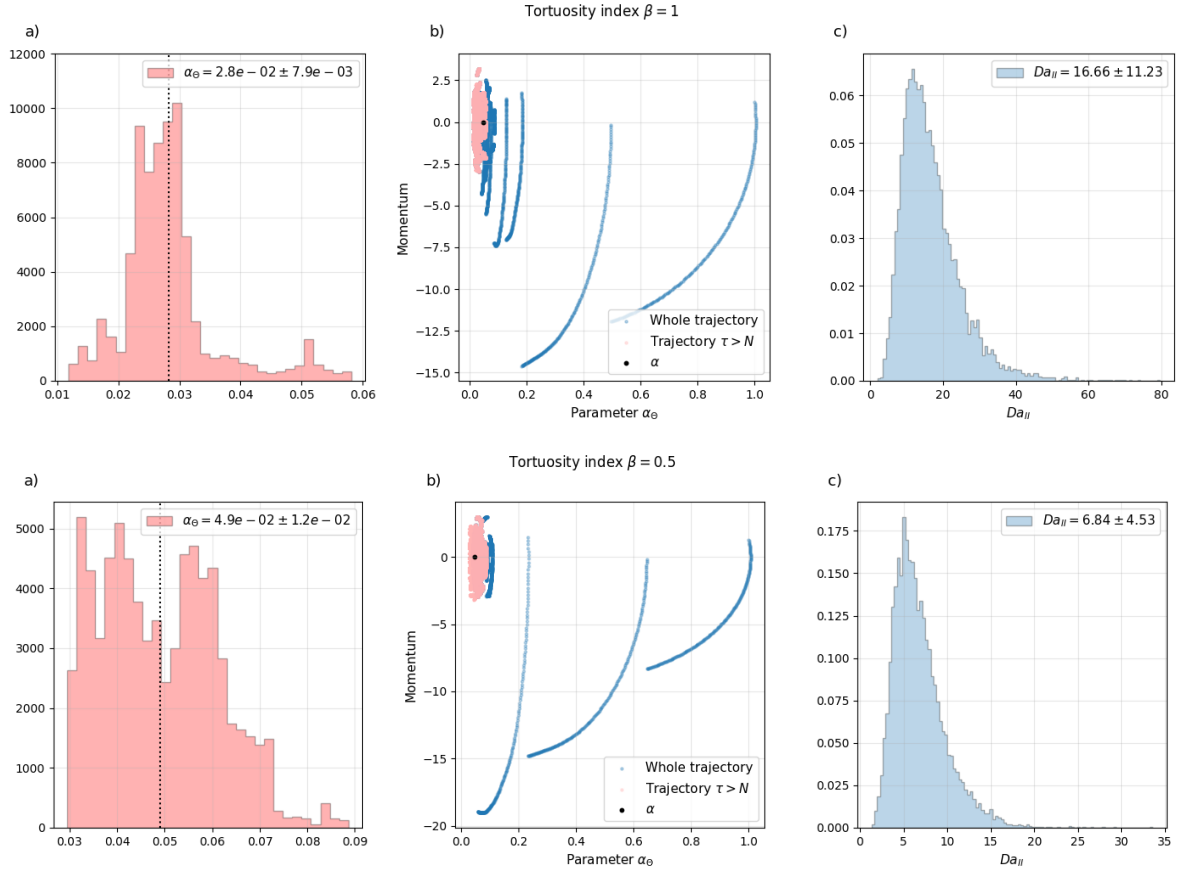


Figure 8: **Posterior distributions of 1D+Time reactive inverse problem:** a) Histogram of the marginal posterior distributions for the inverse parameter  $\alpha_\theta$ . b) Phase diagrams of its trajectory throughout the sampling, with the adaptive steps trajectories (in blue) and effective sampling (in red). The groundtruth values of  $\alpha$  are represented by the black dots. c) Resulting posterior distributions of the catalytic Damköhler number  $Da_{II}$ , determined through the log-normal distribution from relation (48). The top and bottom row respectively corresponds to the tortuosity indexes  $\beta = 1$  and  $\beta = 0.5$ .

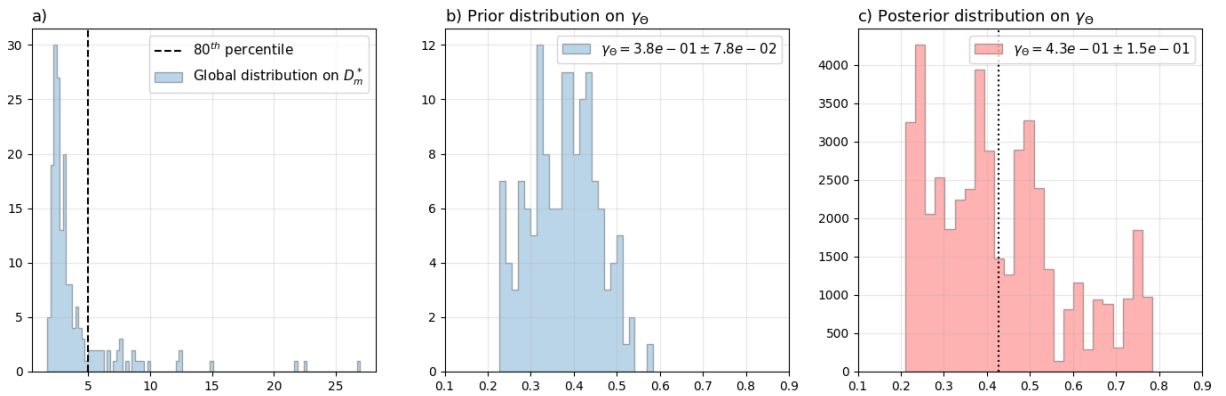


Figure 9: **Prior and posterior distributions on the inverse parameter  $\gamma_\theta$  for 1D+Time reactive inverse problem with tortuosity index  $\beta = 1$ :** a) Global distribution on the estimated parameter  $D_m^*$  arising from a-posteriori analysis on the sampling step 2. We discard the distribution tail after the vertical dashed line corresponding to the 80th percentile. b) Resulting prior distribution on the inverse parameter  $\gamma_\theta$ , used as an a-priori in step 3. c) Posterior distribution on  $\gamma_\theta$  obtained from the overall data assimilation problem throughout sampling step 3.

In the case  $\beta = 1$  for instance, one gets the global distribution on  $D_m^*$  shown in Fig 9a, which translates into the prior distribution on  $\gamma_\Theta$  given by Fig 9b with  $\overline{\gamma_\Theta} = 3.8 \times 10^{-1}$ . Finally, we obtain the posterior distribution from the overall data assimilation problem throughout sampling step 3. This results in the distribution on the inverse parameter  $\gamma_\Theta$  represented in Fig 9c, with the uncertainty range  $\gamma_\Theta \in [0.274, 0.582]$ .

Finally, one can estimate the posterior distribution on the catalytic Damköhler number  $\text{Da}_{\text{II}}$  resulting from the overall data assimilation problem on dynamical  $\mu\text{CT}$  images. This comes from the observation that each inverse parameter, namely  $\alpha_\Theta$  and  $\gamma_\Theta$ , is sought according to a log-normal distribution through the change of variable  $\bullet_\Theta = e^{\bullet_\Theta}$ . Hence, we obtain two normal posterior distributions on the random variables  $X_1$  and  $X_2$  respectively associated with  $\ln(\alpha_\Theta)$  and  $\ln(\gamma_\Theta)$  such that  $X_1 \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$  and  $X_2 \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$ . This combines into a normal distribution on  $\ln(\gamma_\Theta/\alpha_\Theta)$  given by  $(X_1 - X_2) \sim \mathcal{N}(\mu_\gamma - \mu_\alpha, \sigma_\gamma^2 + \sigma_\alpha^2)$ , which is nothing more than a log-normal posterior distribution on the Damköhler number  $\text{Da}_{\text{II}}$ . Indeed, one gets that the random variable  $X$  related to the dimensionless  $\text{Da}_{\text{II}}$  number follows

$$X \sim \text{Log-}\mathcal{N}(\mu_\gamma - \mu_\alpha, \sigma_\gamma^2 + \sigma_\alpha^2) := \text{Log-}\mathcal{N}(\mu, \sigma^2), \quad (48)$$

whose mean and variance are respectively computed as

$$\mathbb{E}[X] = e^{\mu + \sigma^2/2} \quad \text{and} \quad \text{Var}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1). \quad (49)$$

The resulting posterior distributions on the Damköhler number  $\text{Da}_{\text{II}}$  are represented in Fig 8c for both the initial calcite core geometries with their related tortuosity indexes. We finally obtain the uncertainty ranges  $\text{Da}_{\text{II}} \in [5.43, 27.89]$  and  $\text{Da}_{\text{II}} \in [2.30, 11.37]$  for the tortuosity indexes  $\beta = 1$  and  $\beta = 0.5$  respectively.

## 7 Pore-scale imaging inverse problem of calcite dissolution: 2D+Time application

In this section, we apply the data assimilation methodology developed in Sect. 5 to inverse problems for calcite dissolution with heterogeneous porosity levels. We consider a more realistic application involving the dissolution of a 2D calcite core following the configuration of the benchmark developed in [44]. This test case can provide a basis for reactive inverse problems in isotropic porous samples, although the  $\mu\text{CT}$  measurements are still synthetic observations resulting from DNS altered with noise.

### 7.1 Problem set up and dimensionless inverse formulation

We consider a 2D calcite crystal with a cylindrical shape, heterogeneous porosity levels, and two apertures, whose initial geometry is defined by the numerical  $\mu\text{CT}$  image from Fig 10, corrupted with Gaussian noise. We define the physical domain  $\Omega \subset \mathbb{R}^2$  of width 0.2 mm corresponding to the two-dimensional flow channel surrounding the calcite core. We first solve the direct formulation of the dissolution process for this initial geometry on a Cartesian spatiotemporal grid of resolution  $N_x = N_y = 100$  and  $N_t = 350$ . We assume a diffusive-dominated regime with continuous acid injection through non-homogeneous Dirichlet boundary conditions on  $\partial\Omega$ . We also consider, as in the 1D+Time validation test cases, a strong acid solution with  $\text{pH} = 0$  such that the normalizing constant is given by  $C_0 = 1 \text{ mol.L}^{-1}$ . The characteristic length  $L$  of these porous samples is set to  $L = 0.1 \text{ mm}$  and the reactive parameters are respectively defined by  $K_s = 0.8913 \text{ mol.m}^{-2}.\text{s}^{-1}$ ,  $D_m = 10^{-9} \text{ m}^2.\text{s}^{-1}$ , and  $\gamma_{\text{H}^+} = 10^{-3} \text{ m}^3.\text{mol}^{-1}$ , taken from the benchmark [44]. The reactive specific area  $A_s$  is set to  $A_s = 10^3 \text{ m}^{-1}$  — which is slightly underestimated compared to the computed value around  $7.4 \times 10^3 \text{ m}^{-1}$ . We account for the calcite molar volume  $v = 36.93 \times 10^{-3} \text{ L.mol}^{-1}$ , and set a tortuosity index  $\beta = 0.5$ . The DNS is performed until the overall calcite is dissolved which corresponds to a characteristic final time  $T_f = 175 \text{ s}$ . Taken together, one gets a sequence of synthetic  $\mu\text{CT}$  images  $\mathfrak{I}m_{i,j}$  characterizing the dissolution process of the calcite core on the spatiotemporal domain  $\Omega \times (0, T_f)$ .

Given these reactive parameters, we identify the same dissolution regime as in the 1D+Time test cases, with a catalytic Damköhler given by  $\text{Da}_{\text{II}} = 8.913$ . The inverse formulation, however, results in distinct



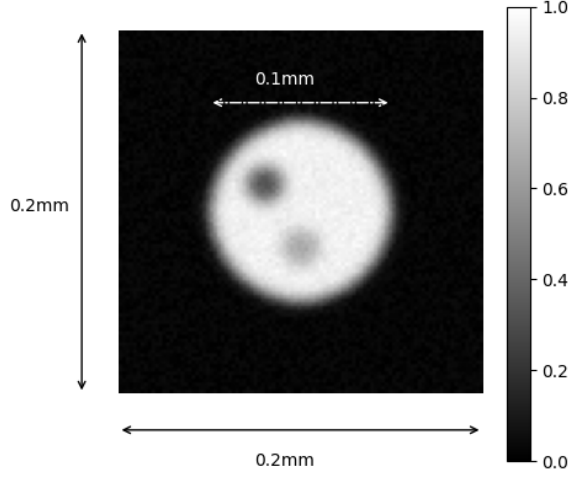


Figure 10: **Initial  $\mu$ CT image defining the 2D porous sample geometry:** Synthetic case with tortuosity index  $\beta = 0.5$ . The  $\mu$ CT measurements are normalized, corrupted with noise, and provide the observation dataset before the dissolution process. The cylindrical calcite core has a radius equal to 0.05 mm.

dimensionless numbers, namely  $Da_{II}^*$  and  $D_m^*$ , arising from the scaling dimensionless factor  $D_{ref}$ . This scaling factor is here determined by:

$$D_{ref} = \frac{T_f^* L^2}{T_f} = 5.714 \times 10^{-11} \text{ m}^2 \cdot \text{s}^{-1}, \quad (50)$$

where the final dimensionless time is  $T_f^* = 1$ . Therefore, one gets from equation (18) the following dimensionless numbers characterizing this 2D+Time reactive inverse problem:

$$Da_{II}^* = 155.9775 \quad \text{and} \quad D_m^* = 17.5 \quad (51)$$

which are related to the inverse parameters through  $(\mathcal{P}_{inv})_{\Theta} = \{\alpha_{\Theta}, \gamma_{\Theta}\} = \{(Da_{II}^*)^{-1}, (D_m^*)^{-1}\}$ . We consider the dimensionless domain  $\Omega^* \times (0, T_f^*) = [0, 2] \times [-1, 1] \times (0, 1)$ , given the characteristic length  $L$ , to extract the observation dataset

$$\mathcal{D} = \{(x_k, y_k, t_k) \in [0, 2] \times [-1, 1] \times (0, 1), \quad k = 1 \dots N_{obs}\} \quad (52)$$

where the number of training points  $N_{obs} = 15907$  represents less than 1% of the data required for the full field reconstructions on the spatiotemporal domain  $\Omega^* \times (0, T_f^*)$ . This dataset  $\mathcal{D}$  is then divided into  $\mathcal{D}^S$ ,  $\mathcal{D}^F$ ,  $\mathcal{D}^{RAI}$ ,  $RAI^+$ ,  $RAI^-$  and  $\mathcal{D}^{\partial}$  that respectively cover around 15.5%, 0.5%, 48%, 11%, 20% and 5% of the  $N_{obs}$  training measurements. Finally, the multi-potential energy writes on this dataset decomposition:

$$\begin{aligned} U(\Theta) = & \frac{\lambda_0}{2\sigma_0^2} \|1 - \varepsilon_{\Theta} - \text{Im}\|_{\mathcal{D}^S}^2 + \frac{\lambda_1}{2\sigma_1^2} \|1 - \varepsilon_{\Theta} - \text{Im}\|_{RAI^+}^2 + \frac{\lambda_2}{2\sigma_2^2} \left\| (vC_0)^{-1} \alpha_{\Theta} \frac{\partial \varepsilon_{\Theta}}{\partial t} - C_{\Theta} \right\|_{RAI}^2 \\ & + \frac{\lambda_3}{2\sigma_3^2} \left\| \gamma_{\Theta} \frac{\partial C_{\Theta}}{\partial t} - \Delta C_{\Theta} \right\|_{\mathcal{D}^F}^2 + \frac{\lambda_4}{2\sigma_4^2} \left( \|1 - C_{\Theta}\|_{\mathcal{D}^{\partial}}^2 + \|10^{-7} - C_{\Theta}\|_{\mathcal{D}^S}^2 \right) \\ & + \frac{\lambda_5}{2\sigma_5^2} \left\| \gamma_{\Theta} \left( \frac{\partial C_{\Theta}}{\partial t} + (vC_0)^{-1} \frac{\partial \varepsilon_{\Theta}}{\partial t} \right) - \mathcal{D}_i(\varepsilon_{\Theta}, C_{\Theta}) \right\|_{RAI^-}^2 + \frac{1}{2\sigma_{\Theta}^2} \|\Theta\|^2 \end{aligned} \quad (53)$$

where the heterogeneous diffusion operator  $\mathcal{D}_i(\varepsilon_{\Theta}, C_{\Theta})$  is computed in its developed form (38c) with  $\beta = 0.5$ . This overall potential energy is sequentially reinforced throughout three successive sampling steps, as detailed in Sect. 5.2 and validated in Sect. 6 on 1D+Time data assimilation problems.

Sampling step	Number of adaptive steps $N$	Number of samples $N_s$	Number of leapfrog steps $L$	Leapfrog time step $\delta t$
1) Preconditioning $\varepsilon_\Theta$	50	200	150	$1 \times 10^{-3}$
2) Preconditioning $C_\Theta$ + Inference $\alpha_\Theta$	40	200	150	$5 \times 10^{-4}$
3) Full data assimilation	10	200	150	$3 \times 10^{-4}$

Table 3: **AW-HMC hyperparameters on the 2D+Time reactive inverse problem:** Setting of the sampler hyperparameters for the three sequential sampling steps defined in Fig 2. The number of adaptive steps  $N$  along with the leapfrog parameters  $L$  and  $\delta t$  are AW-HMC sampler-specific parameters.

## 7.2 Deep learning framework and computational efficiency

Regarding the deep learning strategy, the framework is kept identical to the 1D+Time validation test cases (see Sect. 6.3). In this sense, we consider two distinct neural network architectures, for the micro-porosity and acid concentration surrogate models, which are respectively composed of 4 and 3 hidden layers with 32 neurons per layer. The number of network parameters is, therefore, 3297 for the first sampling and 5538 for the second and third sampling steps with the two additional inverse parameters. The setting of the AW-HMC sampler hyperparameters is also summarized in Table 3 for the successive sampling steps.

Besides, we investigate the impact of the problem dimensionality by analyzing the computational efficiency of the present data assimilation approach with sequential reinforcement process. In this sense, we compare the computational costs of the three successive sampling steps on the 1D+Time and 2D+Time reactive inverse problems. The results of these computational time measurements are presented in Table 4 for both configurations. The first columns compare the sampling times, which is the time required to provide the overall samples of the posterior distributions using the AW-HMC sampler. This training phase is performed on the  $N_{\text{obs}}$  observation data which are randomly selected and non-uniformly distributed on the whole Cartesian grids — respectively  $N_x \times N_t = 200 \times 240$  in 1D+Time and  $N_x \times N_y \times N_t = 100 \times 100 \times 350$  in 2D+Time. All the successive sampling steps are performed on GPU devices, and the computational times are expressed in hours, minutes and seconds (hh:mm:ss). In these sampling/training phases, the present methodology does not suffer from the curse of dimensionality. The 1D+Time and 2D+Time data assimilation problems present similar computational times, although the number of training observations is two times larger in 2D+Time. This establishes that most of the computational cost of the problem is correlated to the number of neural network parameters — as already confirmed in Sect. 6.3 and more specifically in Fig 4 — rather than the number of training points. Since it appears that the same neural network architecture as in 1D+Time is significant to describe the 2D+Time inverse problem, the computational efficiency of this 2D+Time data assimilation is significantly improved.

The second columns of Table 4 then compare the prediction time on CPU devices. This corresponds to the computational time necessary for the potential energy estimation and output field predictions on the whole domain  $\Omega^* \times (0, T_f^*)$ , along with the computation of the main differential operators required as additional outputs. Among the additional outputs, one finds the porosity time derivative at the end of step 1, and all the first-order derivatives and Laplacian operators for the porosity and concentration field at the end of step 2 — which are used to evaluate the initial a-priori on  $\gamma_\Theta$ . The first-order derivatives and Laplacian operators are also considered as output in step 3 to perform the a-posterior analysis based on the BMA-CE diagnostics. In contrast to the sampling phase, the computational time devoted to these predictions is larger when dimensionality increases. This comes from the observation that one needs to evaluate the output fields and additional differential operators on a Cartesian grid which is about 72 times larger in 2D+Time. The predictions of all these differential operators on the overall domain  $\Omega^* \times (0, T_f^*)$  are achieved through automatic differentiation and hence are consistent with their evaluations along the training phase. This can straightforwardly be replaced by other standard differential schemes — as finite differences or PSE schemes — to evaluate these operators on the Cartesian grid, using merely the predictive porosity and acid concentration fields. However, the computational improvement of the process may not be that significant, since one needs to evaluate these operators for all the  $N_s$  samplings steps — basically 600 times — to take into account their intrinsic uncertainties. This means that it only takes a few seconds per

a) Computational times for 1D+Time data assimilation

	Sampling time $T_{\text{GPU}}$ (in hh:mm:ss)	Prediction time $T_{\text{CPU}}$ (in hh:mm:ss)
Sequential step 1	00:17:13	00:00:04
Sequential step 2	01:38:01	00:00:13
Sequential step 3	02:38:58	00:00:14

b) Computational times for 2D+Time data assimilation

	Sampling time $T_{\text{GPU}}$ (in hh:mm:ss)	Prediction time $T_{\text{CPU}}$ (in hh:mm:ss)
Sequential step 1	00:16:26	00:04:59
Sequential step 2	01:21:01	00:34:34
Sequential step 3	02:41:10	00:41:18

**Table 4: Computational times of the successive sequential steps on the 1D+Time and 2D+Time inverse problem:** Comparison of the sampling times on GPU devices (first columns), and the prediction times on CPU devices (second columns) between 1D+Time and 2D+Time configurations. All the computational times are expressed under the form hh:mm:ss to ease the readability.

sample to evaluate these differential operators through automatic differentiation, which is thus comparable to other usual schemes. On top of that, the prediction phases here occur on CPU devices due to memory usage that is not marginal. Typically the micro-porosity field prediction on its own required the storage of  $N_s \times N_x \times N_y \times N_t$  floats in 2D+Time, which is equivalent to about 1.9GB. In this sense, both memory usage and computational efficiency of this prediction phase could be improved, and efforts must be made in this direction. In prospect, we would like to benefit from the parallel architecture of GPU devices by investigating and using appropriate domain decompositions of  $\Omega^* \times (0, T_f^*)$ .

### 7.3 Results and discussion

We apply our data assimilation approach with sequential reinforcement of the multi-potential energy on this 2D+Time reactive inverse problem of calcite dissolution, based on synthetic dynamical  $\mu\text{CT}$  observations generated by DNS. We guarantee the positivity of the inverse parameter inference by selecting log-normal prior distributions and applying the same change of variable  $\bullet_{\Theta} = e^{\bullet_{\tilde{\Theta}}}$  as in the validation Sect. 6. This is combined with independent normal distribution on the neural network parameters, such that we assume the overall prior distribution  $P(\Theta) \sim \mathcal{N}(0, \sigma_{\Theta}^2 I_{p+d})$ . We also impose weakly informed priors on the inverse parameters since we do not impose a-priori information on their respective scaling.

At the end of the sequential sampling, one gets the Bayesian Model Average prediction on the porosity field  $\varepsilon_{\Theta}$ , approximated as in equation (43), along with its local uncertainties during the whole dissolution process. These uncertainty quantification results are presented in Fig 12 for several dissolution times, including the initial condition for  $t^* = 0$  in Fig 12a. We compare these results with the synthetic  $\mu\text{CT}$  images and mean squared errors computed between the BMA surrogate prediction and the groundtruth  $\varepsilon$ . Similar to the 1D+Time application, we observe enhanced uncertainties on the calcite core interfaces, including the aperture edges, throughout the dissolution process. These regions of higher uncertainties mainly correspond to the reactive areas of interest, where the dynamic behavior is predominant, highlighting the challenge of capturing reliable core interfaces and asperities from dynamical  $\mu\text{CT}$ . This also establishes the physical effects of these edge uncertainties, as uncertainties in the interface evolution are driven by the dynamical process of dissolution, its regime, and therefore the inverse parameters identification. In particular, the 2D+Time application presents an asymmetry in the spatial distribution of the main asperities, which contributes to explain the slight asymmetry observed in the uncertainty along the edge of the core in Fig 12a at early times.

Indeed, when comparing the uncertainty at the initial time between the entire data assimilation process - resulting from the three sequential steps of reinforcement learning - and the first regression step, we observe

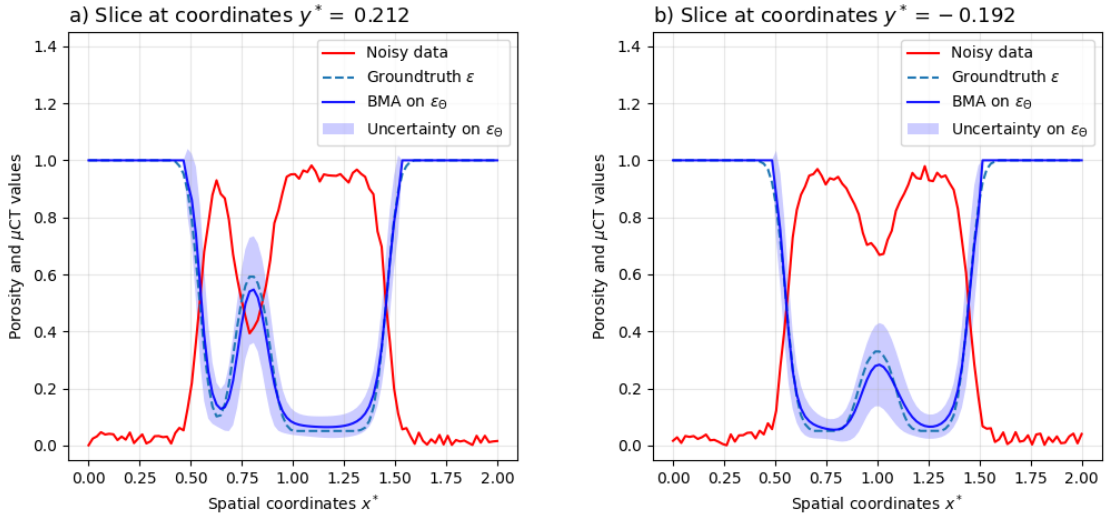


Figure 11: **Uncertainty Quantification on the micro-porosity field at the initial state ( $t^* = 0$ ) for 2D+Time reactive inverse problem:** Corrupted  $\mu$ CT image before dissolution, groundtruth on  $\varepsilon$ , BMA prediction, and uncertainty on  $\varepsilon_\Theta$ . The results are plotted along the horizontal white dashed lines from Fig 12, at spatial coordinates a)  $y^* = -0.192$  and b)  $y^* = 0.212$ .

that the uncertainty distribution along the edges differs. At the end of the first sequential step, considering the  $\mu$ CT data only, we tend to recover a homogeneous distribution of the uncertainty along the edge of the core (see Fig 13), which remains the area of higher uncertainties. The heterogeneous distribution of the uncertainty along the edges is enhanced by the addition of successive physics-based constraints, confirming the physical behavior of such an asymmetry due to geometrical effects in the dissolution process. In particular, the area of higher uncertainty in Figure 12a corresponds to the calcite core interface located in the neighborhood of the largest asperity, involving more complex local interactions throughout the dissolution. Overall, several factors can contribute to this asymmetry, including the influence of the spatial distribution of the asperities within the core and the effect of the dissolution regime driven by the dimensionless numbers. As a prospect, investigating the effects of different spatial configurations on real-life material scans would contribute to validate these observations.

The initial state also exhibits heterogeneous uncertainty distribution on the whole calcite, with lower uncertainties on the pure solid region. As this mineral interface decreases due to the dissolution, the uncertainties tend to become more homogeneously distributed (see Fig 12c and 12d for instance). One can, however, notice that the local mean squared errors are significantly embedded in the micro-porosity uncertainties, ensuring reliable predictions. In Fig 11, we detail these results on the initial state geometry — for  $t^* = 0$  — by plotting along the two dashed lines from Fig 12 the BMA and uncertainty on  $\varepsilon_\Theta$ , the groundtruth values, and the  $\mu$ CT observations. Considering the dynamical dissolution process also provide insight into the upper and lower bounds of the residual micro-porosity  $\varepsilon_0$  for the initial calcite core geometry. In the porous matrix, we obtain the estimations  $3\% \leq \varepsilon_0 \leq 10\%$  for the 95% confidence interval.

The validation of the inference is first performed using the Bayesian Model Average Cumulative Error (BMA-CE) on the micro-porosity field, which is computed along the three successive sampling steps of sequential reinforcement. We then introduce  $\mathcal{F}_1$  and  $\mathcal{F}_2$  the PDE constraints residuals arising from the reactive model (19) and involved in the multi-potential energy (53):

$$\begin{aligned} \mathcal{F}_1(\varepsilon_\Theta, C_\Theta) &:= (vC_0)^{-1} \alpha_\Theta \frac{\partial \varepsilon_\Theta}{\partial t} - C_\Theta \\ \mathcal{F}_2(\varepsilon_\Theta, C_\Theta) &:= \gamma_\Theta \left( \frac{\partial C_\Theta}{\partial t} + (vC_0)^{-1} \frac{\partial \varepsilon_\Theta}{\partial t} \right) - \mathcal{D}_i(\varepsilon_\Theta, C_\Theta) \end{aligned} \quad (54)$$

to estimate their BMA-CE $^{\mathcal{F}}$  diagnostics on sampling steps 2 and 3. For the 2D+time data assimilation problem, the BMA-CE metrics on the micro-porosity field  $\varepsilon$  and PDE residuals are straightforwardly extended from the formulae (45) and (47). The results are provided in Fig 14 and highlight the conver-

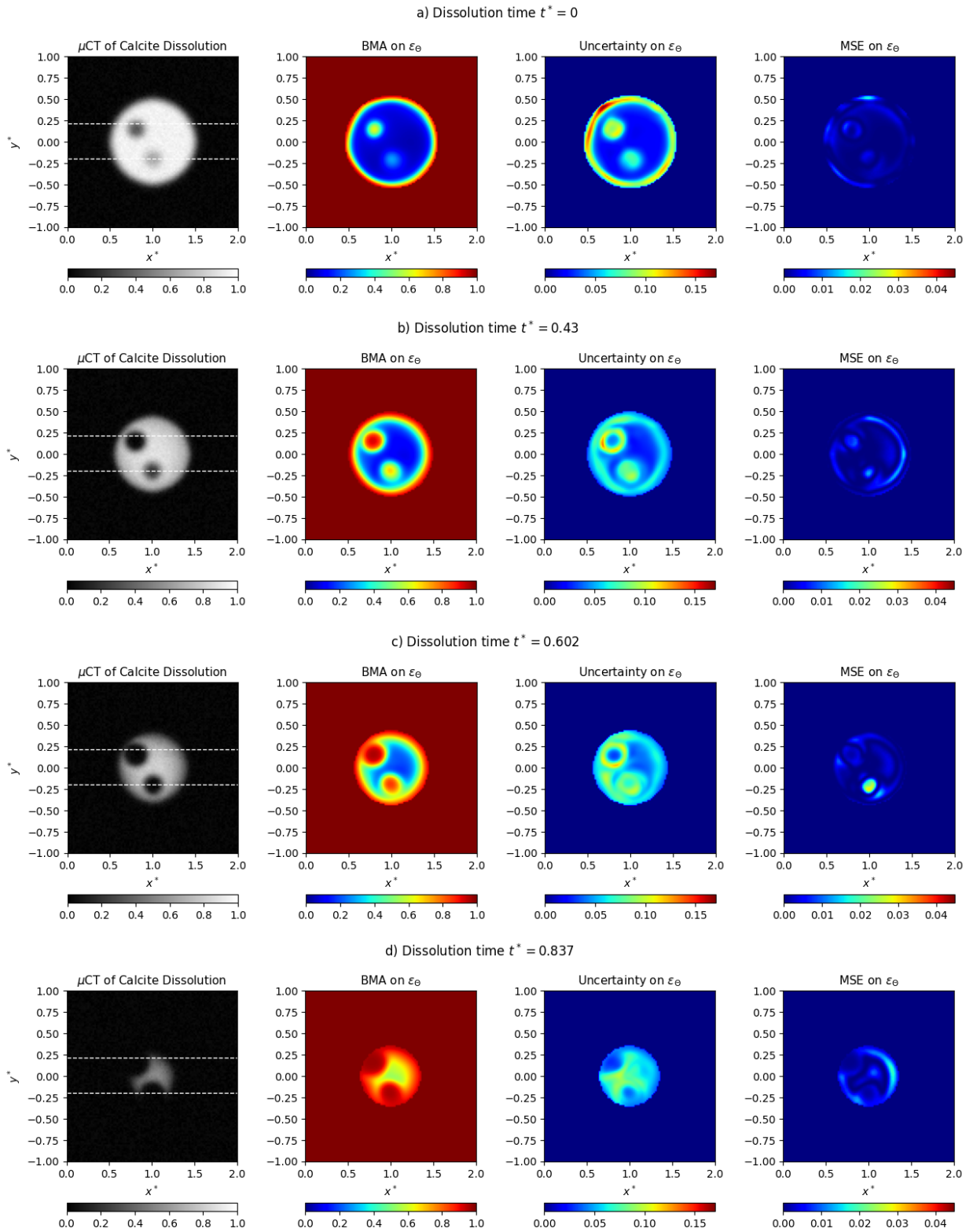


Figure 12: **Uncertainty Quantification on 2D+Time reactive inverse problem with data assimilation:** Bayesian Model Average (BMA) predictions on the micro-porosity field  $\varepsilon_{\Theta}$  with their local uncertainties and mean squared errors (MSE). Comparison with the  $\mu$ CT dynamical images at several dissolution times, in the dimensionless formulation: a) Initial condition at  $t^* = 0$ . Intermediate dissolution times at b)  $t^* = 0.43$ , c)  $t^* = 0.602$  and d)  $t^* = 0.837$ .

At dissolution time  $t^* = 0$

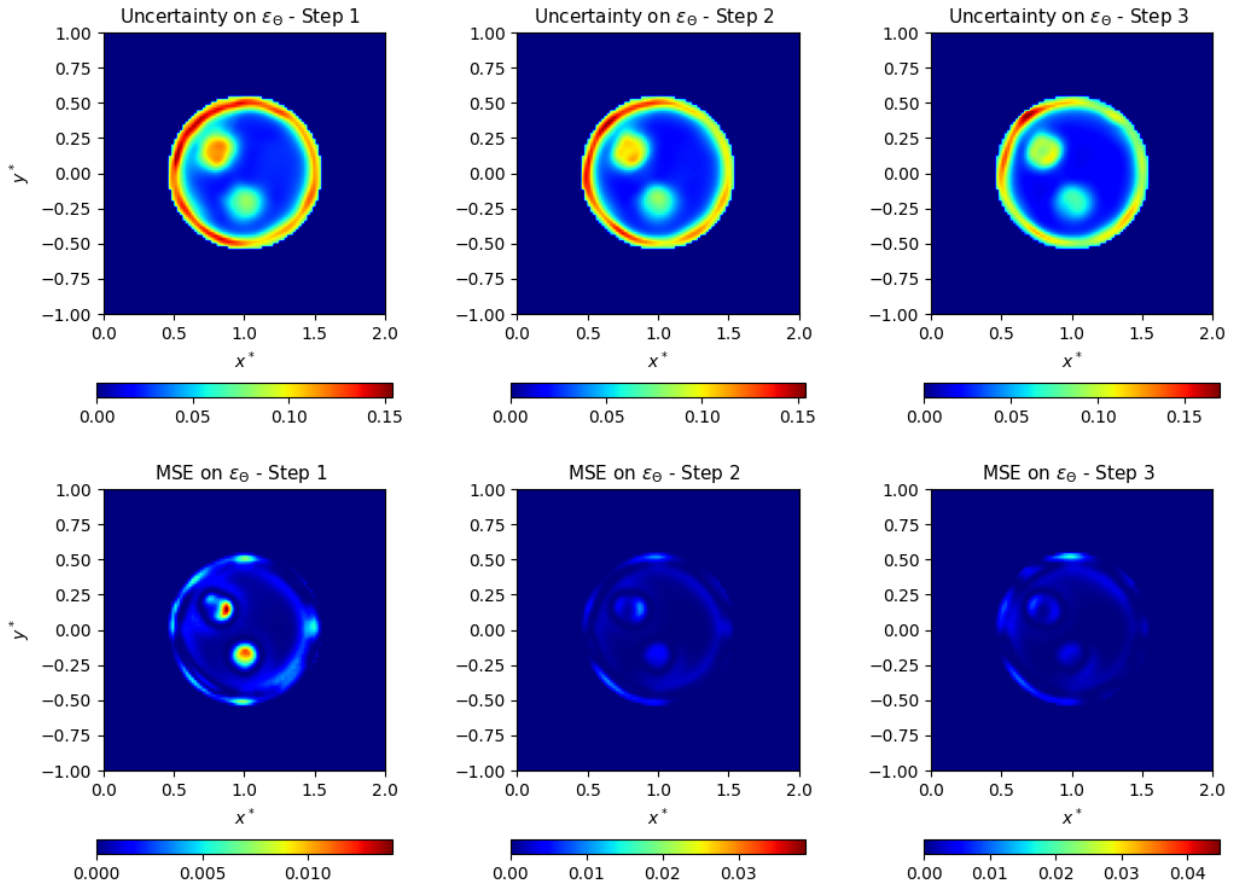


Figure 13: **Evolution of the Uncertainty Quantification on 2D+Time reactive inverse problem during sequential reinforcement:** Local uncertainty predictions on the micro-porosity field  $\varepsilon_\theta$  with their mean squared errors (MSE) at initial conditions  $t^* = 0$ . Comparison across the three steps of the sequential reinforcement approach, highlighting the evolution of uncertainties at the core edges and the effects of the spatial distribution of asperities in the physics-based data assimilation. Step 3 corresponds to the entire reactive inverse problem, incorporating all the physical constraints from the coupled PDE system (19).

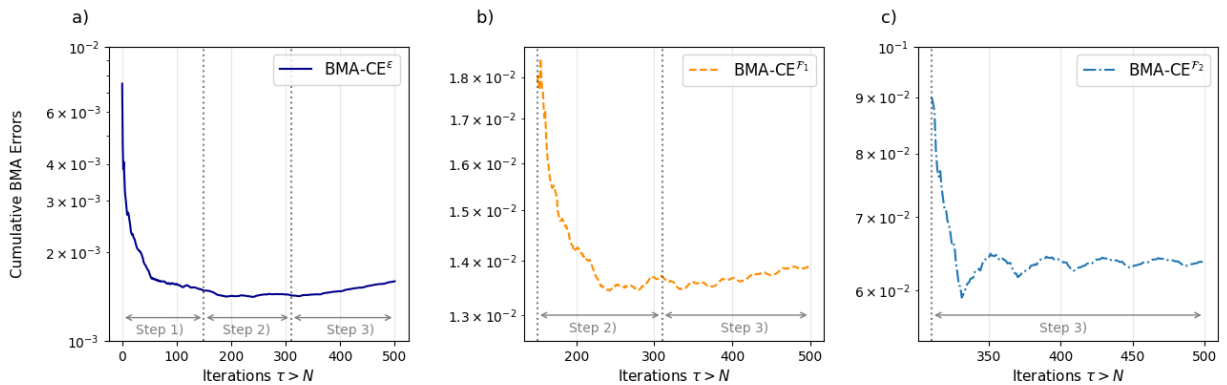
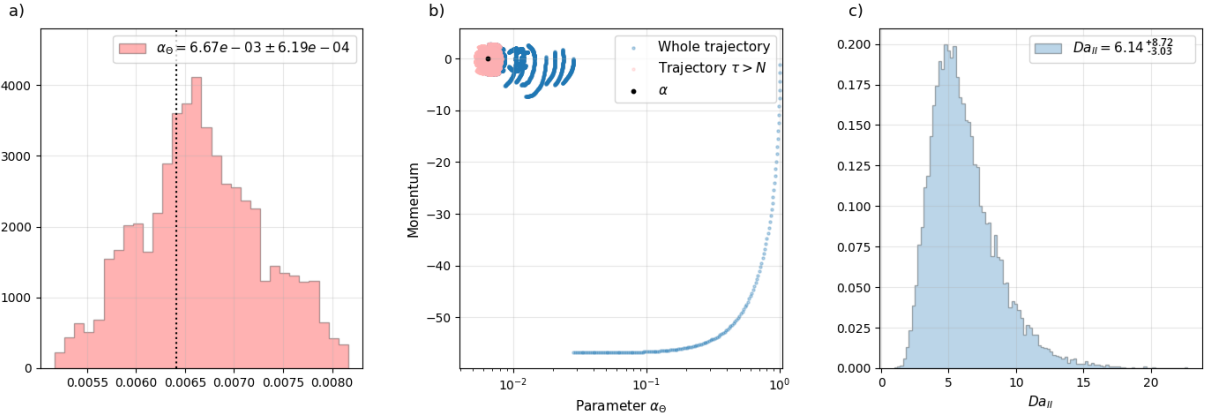


Figure 14: **Bayesian Model Average Cumulative Error diagnostics (BMA-CE) for 2D+Time reactive inverse problem:** BMA-CE on the micro-porosity field prediction  $\varepsilon$  throughout the sampling iterations, and BMA-CE on the PDE constraint residuals  $\mathcal{F}_1$  and  $\mathcal{F}_2$  defined by equations (54) and introduced successively. The dotted vertical lines mark the sequential reinforcement of the multi-potential energy (53).



**Figure 15: Posterior distributions of 2D+Time reactive inverse problem:** a) Histogram of the marginal posterior distribution for the inverse parameter  $\alpha_\Theta$ . b) Phase diagram of its trajectory throughout the sampling, with the adaptive steps trajectories (in blue) and effective sampling (in red). The groundtruth value of  $\alpha$  is represented by the black dot. c) Resulting posterior distribution of the catalytic Damköhler number  $Da_{II}$ , determined through the log-normal distribution from relation (48).

gence of each term toward final BMA errors, at the sampling iteration  $\tau = 500$ , scaling respectively about  $BMA-CE^\varepsilon(\tau) = 1.6 \times 10^{-3}$ ,  $BMA-CE^{\mathcal{F}_1}(\tau) = 1.4 \times 10^{-2}$  and  $BMA-CE^{\mathcal{F}_2}(\tau) = 6.4 \times 10^{-2}$ . From these convergence diagnostics, we observe a saturation of the PDE constraints that highlights the intrinsic uncertainties of their corresponding tasks in the multi-potential energy (53). In this sense, the PDE constraint  $\mathcal{F}_2$  involving the heterogeneous diffusion operator (38c) is the most uncertain term due to its high sensitivity to porosity variations. Nonetheless, we notice, as in the validation test cases from Sect. 6, that successive introduction of the PDE constraints leads to a slight increase in the BMA- $CE^\varepsilon$  curve, bringing information on the porosity field recovery by preventing overfitting issues. In particular, we observe in Fig 13 that successively incorporating the physics-based constraints affects the uncertainty distribution of the porosity, especially at the core edges, rendering them physically reliable. In this sense, such a bounce in the BMA-CE predictions provides more dynamical insights into the porosity field  $\varepsilon$ . The increase observed on BMA- $CE^{\mathcal{F}_1}$  in Fig 14 arises from both a deviation in the surrogate micro-porosity field  $\varepsilon_\Theta$  and a more physical approximation of the surrogate concentration field  $C_\Theta$  in step 3 compared to step 2. Indeed, step 2 considers a simplified coupled model for the reactive system based on the quasi-stationary assumption to provide a first approximation of  $C_\Theta$ . In step 3, the physics-based assimilation is fully constrained, with the complex coupled dynamical behavior given by the PDE system (19), and we therefore expect such deviations both in the BMA-CE curves and uncertainty estimates.

For the inverse parameters inference, we represent in Fig 15 the histograms of the marginal posterior distributions of  $\alpha_\Theta := (Da_{II}^*)^{-1}$  and its trajectory in the phase space illustrating the convergence toward its mode during the adaptive steps — represented in blue in Fig 15b. The latter shows that our data assimilation approach combined with the AW-HMC sampler from [55] makes it possible to capture the correct parameter range without prior knowledge of its scaling. Once the adaptive process ends, we effectively start sampling the inverse parameter mode neighborhood, represented by the phase space trajectories for  $\tau > N$  in red. We then follow the same process as in Sect. 6.4 to estimate the prior distribution on the inverse parameter  $\gamma_\Theta$ , and we obtain  $\overline{\gamma_\Theta} = 3.9 \times 10^{-2}$  which is used as an initial a-priori in the sampling step 3 (see Fig 16a and Fig 16b). The posterior distribution on the inverse parameter  $\gamma_\Theta$ , estimated throughout step 3, is represented in Fig 16c and provides the following predictive interval  $\gamma_\Theta \in [2.6 \times 10^{-2}, 5.4 \times 10^{-2}]$ . One can, finally, estimate the posterior distribution on the catalytic Damköhler number  $Da_{II}$  according to the log-normal distribution obtained by the relation (48). This results in the log-normal posterior distribution represented in Fig 15 whose mean and standard deviations are respectively given by  $\mathbb{E}[X] = 6.14$  and  $\sqrt{\text{Var}(X)} = 4.02$ . Finally, we obtain the 95% asymmetric confidence interval on the Damköhler number given by  $Da_{II} = 6.14^{+8.72}_{-3.03}$  for this data assimilation problem of calcite dissolution, which is consistent with the theoretical value.

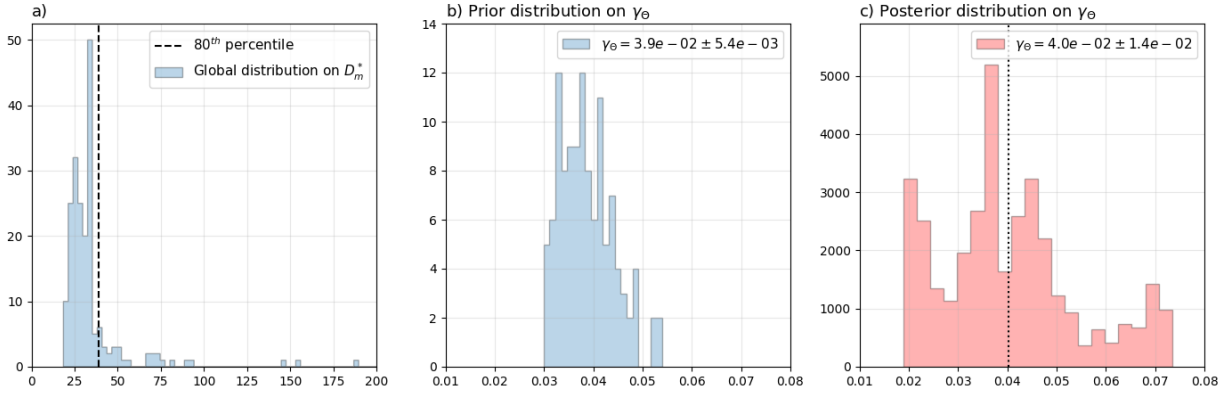


Figure 16: **Prior and posterior distributions on the inverse parameter  $\gamma_\Theta$  for 2D+Time reactive inverse problem:** a) Global distribution on the estimated parameter  $D_m^*$  arising from a-posteriori analysis on the sampling step 2. We discard the distribution tail after the vertical dashed line corresponding to the 80<sup>th</sup> percentile. b) Resulting prior distribution on the inverse parameter  $\gamma_\Theta$ , used as an a-priori in step 3. c) Posterior distribution on  $\gamma_\Theta$  obtained from the overall data assimilation problem throughout sampling step 3.

## 8 Concluding remarks

This work intended to address two major challenges related to uncertainty quantification in pore-scale modeling of reactive flows, which plays a crucial role in the long-term management of CO<sub>2</sub> capture and storage. Providing reliable macro-properties changes due to geochemical processes, such as CO<sub>2</sub> mineral trapping and dissolution within the porous environment, is essential to query reservoir safety. In this sense, we aim to ensure that the evolving petrophysical properties provide meaningful characterizations of these chemical processes instead of intrinsic deviations arising from imaging limitations. Some intrinsic limiting factors remain when modeling pore-scale dynamical systems based on  $\mu$ CT scans and lead to several trade-offs that can bias the predictions. In particular, this results in unresolved micro-porosity, especially when the scan resolution does not fully capture morphological features of the pore space under the constraint of having a representative elementary volume (REV) of the sample. Quantifying sub-resolution porosity, which is a prevalent imaging artifact, was the first challenge we identified, and therefore we focused on quantifying morphological uncertainties in the micro-porosity field. Our second concern was to investigate the reliability of kinetic parameters, such as mineral reactivity, in the context of reactive processes. Indeed, these are critical parameters to account for in pore-scale modeling, though their experimental estimations can suffer from wide discrepancies. Estimating proper order of magnitude and uncertainty ranges appears essential to ensure reliable calibration of pore-scale models, and afterward trustable management of CO<sub>2</sub> mineral storage. The present article investigated both these issues by integrating uncertainty quantification concerns in the workflow of pore-scale modeling.

Current methodologies investigating these problems regard them independently and fall into pure image treatment analysis related to experimental static or dynamical  $\mu$ CT images. Deep learning methodologies, for instance, are used to address sub-resolution porosity quantification. Multi-scale image reconstruction, which extrapolates the latent information of the porous structure, is obtained with Generative Adversarial Networks [80]. One also gets super-resolved segmented images from static  $\mu$ CT scans through Convolutional Neural Networks [6], which compensate for the unresolved morphological features. Regarding the mineral reactivity assessment, experimental works have been conducted on dynamical 4D  $\mu$ CT of carbonate dissolution. This provides, through differential imaging techniques, insight into local reaction rates at the mineral interfaces [52, 43]. However, all these approaches do not consider incorporating uncertainty quantification in the estimates. In contrast, alternatives accounting for deviations in the petrophysical properties due to imaging limitations are mainly purely model-related approaches [66, 56].

The main novelty of our work, therefore, lay in its ability to address both morphological uncertainty and reaction rate quantification from the perspective of coupling physics-based models with data-driven tech-



niques. In this sense, we have developed a data assimilation approach for pore-scale imaging problems that combine dynamical microtomography and physical regularization induced by PDE models of reactive processes. We integrated this novel data assimilation strategy into the Bayesian inference context through our efficient AW-HMC framework for BPINNs [55]. This also confirmed the great potential of this adaptive and self-balancing methodology and rendered BPINNs a promising approach to address complex data assimilation. In the pore-scale imaging context, in particular, we have focused on multitask inverse problems of calcite dissolution based on dynamical  $\mu$ CT images, along with two dimensionless inverse parameters and a latent concentration field. We have also assumed unknown informative priors on the different tasks scaling and relied on automatic adjustment of the uncertainties including noise-related estimations and model adequacy. In this sense, we provided reliable uncertainty quantification on the micro-porosity field description and reactive parameters. We also built our data assimilation upon a sequential reinforcement strategy of the multi-potential energy and thus the target posterior distribution. This involved successively integrating additional PDE constraints into the overall data assimilation process through dedicated sampling steps. Finally, we have also addressed computational concerns and have shown that suitable formulation of complex non-linear differential operators, especially the heterogeneous diffusion arising from Archie’s law, can significantly reduce the computational costs of these operators. Taken together, we presented an intrinsic data assimilation strategy for pore-scale imaging inverse problem and demonstrated its efficiency on several 1D+Time and 2D+time calcite dissolution problems.

Overall, our results confirmed enhanced morphological uncertainties localized on the calcite core edges throughout the dissolution process. This characterized the challenge of capturing reliable mineral interfaces from the dynamical  $\mu$ CT images, and therefore query the confidence of mineral reactivity assessment using merely differential imaging techniques on the  $\mu$ CT scans. Combining data-driven and physics-based approaches thus offers a promising alternative to overcome the limitations of each approach individually, and alleviate biased predictions. We also obtained reliable insight into the upper and lower bounds for the residual, potentially unresolved, micro-porosity  $\varepsilon_0$  in the porous matrix. These estimations can then be incorporated into direct numerical simulation solvers to measure the impact of these micro-porosity variations on the other petrophysical properties, such as permeability. This can also ensure that the macro-scale porosity evolutions due to the reactive processes are significant compared to these intrinsic morphological uncertainties at the pore scale. Finally, we have obtained posterior distribution on the dimensionless reactive parameters characterizing the dissolution inverse problems. We have shown that our data assimilation approach combined with the AW-HMC sampler made it possible to capture the correct parameter ranges without prior knowledge of their scaling, which confirmed the robustness and reliability of the inferences. Last but not least, we have identified uncertainty ranges on the usual catalytic Damköhler number  $Da_{II}$  resulting from the prescribed PDE model and dynamical observations of the dissolution process. This is of great interest to aggregate experimental investigations and direct numerical simulations, and therefore guarantee the reliability of pore-scale modeling and simulation of reactive flows. We now have the potential to effectively address robust and reliable uncertainty quantification in pore-scale imaging and to manage the impact of  $\mu$ CT limitations on the petrophysical properties and reactive parameters.

As future prospects, it would be interesting to apply the current data assimilation framework on real 3D samples obtained from experimental  $\mu$ CT dissolution scans. This could provide deeper insights into the relationship between experiments and mathematical modeling theory, thereby enhancing the reliability of predictive model calibration and the confidence in computational approaches for real-life reactive materials. However, the shape complexity of porous samples, along with the non-linear operators involved in the advection-diffusion-reaction PDE model, are additional concerns rendering these applications challenging. Performing 4D data assimilation of real-life reactive material will especially demand the development of efficient computational strategies to speed up the performance. The framework proposed in this article could benefit from enhanced parallelization on GPU devices, although managing the increased computational demand on CPU and GPU resources will be critical. For example, task parallelization across different processors during the sampling phase of a multi-objective data assimilation problem should be investigated, along with potential coupling with the symmetric splitting approach of the standard Hamiltonian Monte Carlo suggested by Cobb *et al.* [21], allowing parallelization on data subsets during the training phase. Regarding the prediction step, one can also investigate standard domain decomposition distributed across multiple devices to enhance the computational efficiency. These strategies aim to optimize the method ef-

efficiency while ensuring sustainable computational costs, which becomes crucial when addressing complex real-world inverse problems and data assimilation. Another prospect is to extend the inference to different kinds of dissolution regime, including advection dominant phenomena. This would require insights into the velocity field to integrate the overall reactive hydrodynamics PDE system within the data assimilation framework. Nonetheless, this implies the recovery of a additional latent fields and inverse parameter, namely the velocity and a modified Peclet number  $Pe^*$ . A potential approach is to leverage priors for hydrodynamics through fast surrogate modeling and fluid flow model proxy [7, 61] based on the morphological a-priori on the porous medium evolution. Subsequently, these priors could be incorporated and adjusted into the data assimilation of a full reactive flow problem. In this context, the estimation of the bulk permeability in the Kozeny-Carman correlation remains challenging since it drives the amount of concentration heading inside the porous matrix. On the experimental side, 4D data assimilation of real porous samples could eventually suffer from temporal deformations and collapse due to variations in the mechanical stress during the dissolution, that will require data preprocessing to minimize the noise arising from these mechanical deformations. Overall, considering these additional challenges, the present methodology has already demonstrated sufficient robustness to handle heavily noised data and will offer the opportunity to quantify uncertainties in pore-scale imaging obtained from real  $\mu$ CT data.

## A Review on the AW-HMC algorithm and stopping criterion for the weight adaptation

We briefly review the AW-HMC methodology previously developed in [55], along with the stopping criterion used for the weight adaptation. In Algorithm 1,  $\Theta^{t_0}$  refers to the initial state for both the neural network and inverse parameters and characterises the particle positions in the physical analogy.  $N_s$  is the total number of samples collected during the training and  $N_{\text{burn}}$  the burn-in steps, usually taken as the effective number of adaptive steps  $N$ .  $L$  and  $\delta t$  respectively represents the number of iterations and the step size used for in the leapfrog method to solve the Hamiltonian system (24), where  $\mathbf{M}$  is the covariance or mass matrix for the distribution on the particle momenta  $r$ .  $N_{\text{max}}$  and  $S_{\text{min}}$  are the maximum number of adaptive iterations and threshold used as stopping criterion for for the adaptive weighting.

We consider the weighted multi-potential energy

$$U(\Theta) = \sum_{k=0}^K \lambda_k \mathcal{L}_k(\Theta) + \lambda_{K+1} \|\Theta\|^2 := \sum_{k=0}^{K+1} \lambda_k \mathcal{L}_k(\Theta), \quad (55)$$

introduced in Sect. 4.1 and associated to the weighted Hamiltonian

$$H_{\lambda_\tau}(\Theta, r) = \sum_{k=0}^{K+1} \lambda_k(\tau) \mathcal{L}_k(\Theta) + K(r) = \sum_{k=0}^{K+1} \lambda_k(\tau) \mathcal{L}_k(\Theta) + \frac{1}{2} r^T \mathbf{M}^{-1} r, \quad (56)$$

to define the  $\lambda_k(\tau)$  along the sampling steps  $\tau$  of automatic and adaptive weighting. One should notice that the weight on the prior term, corresponding to  $\lambda_{K+1}$ , is not adjusted since it rather acts as regularization term than a specific task in the multi-objective inverse problem (see [55] for detailed development). The algorithm then alternates between deterministic steps by solving the Hamiltonian dynamical system for the fictive particle of position  $\Theta$  and momentum  $r$  during the leapfrog iterations (lines 11 to 16 in Algorithm 1), and stochastic steps resulting from the momentum sampling (lines 3 and 4) and allowing the exploration of various energy levels. The Metropolis-Hastings step, finally, plays the role of an acceptance criterion through a transition probability based on the weighted Hamiltonian and ensures energy preservation by rejecting samples that lead to divergent trajectories during the leapfrog numerical integration. Once all the samples are collected, predictive distributions on the main functional fields are built based on the Bayesian Model Average approximation (21).

The stopping criterion for the weighting adaptation (line 6 in Algorithm 1) is based on the maximum number of iterations  $N_{\text{max}}$  combined with the convergence threshold  $S_{\text{min}}$  on the local variation of the Hamiltonian  $\bar{S}_\tau$ . Indeed, if there is no significant variation of the Hamiltonian, we have reached a well-fitted weighted posterior distribution targeting the high probability-density region given by the Pareto front

**Algorithm 1:** Adaptively Weighted Hamiltonian Monte Carlo (AW-HMC)

```

1 Sampling procedure:
2 for  $\tau = 1 \dots N_s$  do
3   Sample  $r^{t_{\tau-1}} \sim \mathcal{N}(0, \mathbf{M})$ ;
4   Set  $(\Theta_0, r_0) \leftarrow (\Theta^{t_{\tau-1}}, r^{t_{\tau-1}})$ ;
5   Weights adaptation:
6   if  $(\tau \leq N_{\max})$  and  $(\bar{S}_{\tau} \geq S_{\min})$  then
7     Compute  $\lambda_k(\tau) = \left( \frac{\min_{j=0..K} (\text{Var}\{\nabla_{\Theta} \mathcal{L}_j(\Theta_0)\})}{\text{Var}\{\nabla_{\Theta} \mathcal{L}_k(\Theta_0)\}} \right)^{1/2} \quad \forall k = 0..K$  and  $\lambda_{K+1}(\tau) = 1$ ;
8   else
9      $\lambda_k(\tau) = \lambda_k(\tau - 1) \quad \forall k = 0..K$  and  $\lambda_{K+1}(\tau) = 1$ 
10  end
11  Leapfrog:
12  for  $i = 0 \dots L - 1$  do
13     $r_i \leftarrow r_i - \frac{\delta t}{2} \sum_{k=0}^{K+1} \lambda_k(\tau) \nabla_{\Theta} \mathcal{L}_k(\Theta_i)$ ;
14     $\Theta_{i+1} \leftarrow \Theta_i + \delta t \mathbf{M}^{-1} r_i$ ;
15     $r_{i+1} \leftarrow r_i - \frac{\delta t}{2} \sum_{k=0}^{K+1} \lambda_k(\tau) \nabla_{\Theta} \mathcal{L}_k(\Theta_{i+1})$ ;
16  end
17  Metropolis-Hastings:
18  Sample  $p \sim \mathcal{U}(0, 1)$ ;
19  Compute  $\alpha = \min(1, \exp(H_{\lambda_{\tau}}(\Theta_0, r_0) - H_{\lambda_{\tau}}(\Theta_L, r_L)))$  using (56);
20  if  $p \leq \alpha$  then
21     $\Theta^{t_{\tau}} = \Theta_L$ ;
22  else
23     $\Theta^{t_{\tau}} = \Theta_0$ ;
24  end
25  Collect the samples after burn-in :  $\{\Theta^{t_i}\}_{i=N_{\text{burn}}}^{N_s}$ 
26 end

```

neighborhood, and the adaptation can be stopped. Given the stochastic nature of the Hamiltonian evolution, the variation  $\bar{S}_{\tau}$  is computed as the locally average, over the last  $p_{\text{mean}}$  values, of the linear regression  $S_{\tau}$  computing the Hamiltonian slopes over a short history — the last  $p_{\text{slope}}$  values of the Hamiltonian function. Denoting by  $H_i$  the statistical series of the Hamiltonian evolution for the sampling iterations  $X_i = i$ , we thereby define  $\bar{S}_{\tau}$  as:

$$\bar{S}_{\tau} = \frac{1}{p_{\text{mean}}} \sum_{i=\tau-p_{\text{mean}}+1}^{\tau} S_i \quad \text{where} \quad S_{\tau} = \frac{\text{Cov}\left(\{H_i\}_{i=\tau-p_{\text{slope}}+1}^{\tau}, \{X_i\}_{i=\tau-p_{\text{slope}}+1}^{\tau}\right)}{\text{Var}\left(\{X_i\}_{i=\tau-p_{\text{slope}}+1}^{\tau}\right)} \quad (57)$$

with  $p_{\text{slope}} = p_{\text{mean}}$  for each step of the sequential reinforcement process. In the 2D+Time application from Sect. 7, the previous stopping criterion leads to effective numbers of adaptive steps of  $N = 50, 40,$  and  $10$  for each sequential sampling step, respectively given the threshold  $S_{\min} = 1 \times 10^{-3}, 5 \times 10^{-3}$  and  $1 \times 10^{-2}$  (see Figure 17). Such an increase in the threshold value  $S_{\min}$  is expected as the potential energy, and thereby the Hamiltonian, is successively reinforced by additional physical constraints.

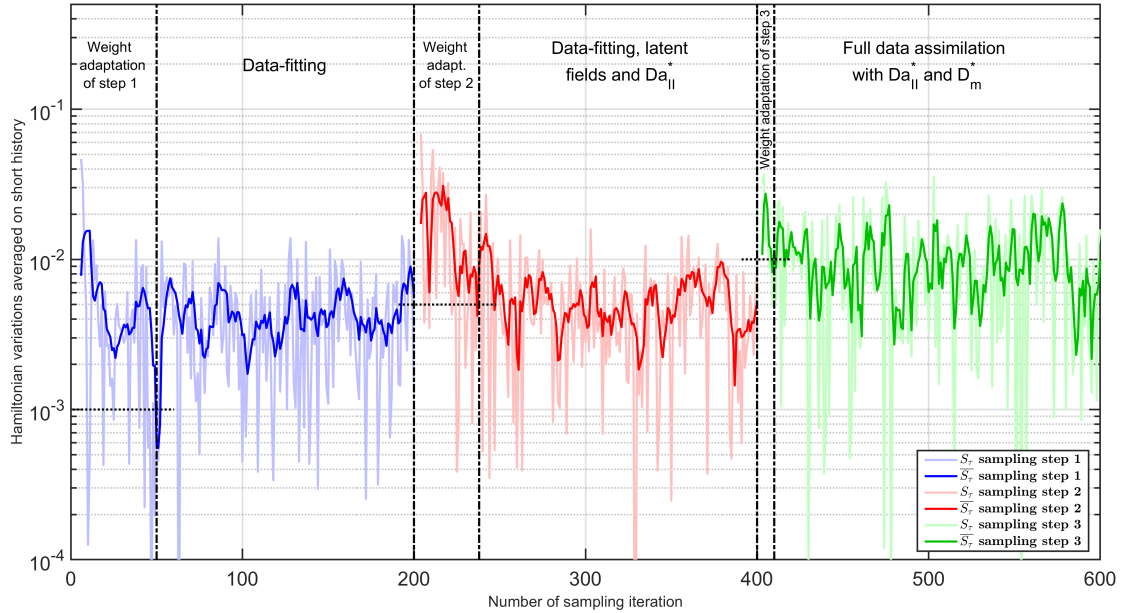


Figure 17: **Hamiltonian variations averaged on short history as stopping criterion for the weight adaptation:** The vertical dashed lines depict the start and the end of the weight adaptation processes for each step of the sequential reinforcement, developed in Sect. 5.2. The light-colored curves represent the best slopes  $S_\tau$  of the Hamiltonian, computed by linear regression over a short history through equation (57). The dark curves correspond to the variations  $\bar{S}_\tau$  along the sampling iterations, with the threshold values  $S_{\min}$  for each sequential step represented by the horizontal dashed lines.

## References

- [1] Mehrdad Ahkami, Andrea Parmigiani, Paolo Roberto Di Palma, Martin O. Saar, and Xiang-Zhao Kong. A lattice-boltzmann study of permeability-porosity relationships and mineral precipitation patterns in fractured porous media. *Computational Geosciences*, 24(5):1865–1882, 2020.
- [2] Fatima Al Hameli, Hadi Belhaj, and Mohammed Al Dhuhoori. CO2 sequestration overview in geological formations: Trapping mechanisms matrix assessment. *Energies*, 15(20), 2022.
- [3] Alex Alberts and Ilias Bilonis. Physics-informed information field theory for modeling physical systems with uncertainty quantification. *Journal of Computational Physics*, 486:112100, 2023.
- [4] G. Allaire and H. Hutridurga. Homogenization of reactive flows in porous media and competition between bulk and surface diffusion. *IMA Journal of Applied Mathematics*, 77(6):788–815, 2012.
- [5] Grégoire Allaire, Robert Brizzi, Andro Mikelić, and Andrey Piatnitski. Two-scale expansion with drift approach to the taylor dispersion for reactive transport through porous media. *Chemical Engineering Science*, 65(7):2292–2300, 2010. International Symposium on Mathematics in Chemical Kinetics and Engineering.
- [6] Naif J. Alqahtani, Yufu Niu, Ying Da Wang, Traiwit Chung, Zakhar Lanetc, Aleksandr Zhuravljov, Ryan T. Armstrong, and Peyman Mostaghimi. Super-Resolved Segmentation of X-ray Images of Carbonate Rocks Using Deep Learning. *Transport in Porous Media*, 143(2):497–525, 2022.
- [7] Shohreh Amini and Shahab Mohaghegh. Application of Machine Learning and Artificial Intelligence in proxy modeling for fluid flow in porous media. *Fluids*, 4(3), 2019.
- [8] Matthew Andrew, Branko Bijeljic, and Martin J. Blunt. Pore-scale imaging of trapped supercritical carbon dioxide in sandstones and carbonates. *International Journal of Greenhouse Gas Control*, 22:1–14, 2014.
- [9] Heiko Andrä, Nicolas Combaret, Jack Dvorkin, Erik Glatt, Junehee Han, Matthias Kabel, Youngseuk Keehm, Fabian Krzikalla, Minhui Lee, Claudio Madonna, Mike Marsh, Tapan Mukerji, Erik H.

- Saenger, Ratnanabha Sain, Nishank Saxena, Sarah Ricker, Andreas Wiegmann, and Xin Zhan. Digital rock physics benchmarks—part I: Imaging and segmentation. *Computers & Geosciences*, 50:25–32, 2013.
- [10] Heiko Andrä, Nicolas Combaret, Jack Dvorkin, Erik Glatt, Junehee Han, Matthias Kabel, Youngseuk Keehm, Fabian Krzikalla, Minhui Lee, Claudio Madonna, Mike Marsh, Tapan Mukerji, Erik H. Saenger, Ratnanabha Sain, Nishank Saxena, Sarah Ricker, Andreas Wiegmann, and Xin Zhan. Digital rock physics benchmarks—part II: Computing effective properties. *Computers & Geosciences*, 50:33–43, 2013.
- [11] G. E. Archie. The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Transactions of AIME*, 146:54–62, 1942.
- [12] Yehuda Bachmat and Jacob Bear. On the Concept and Size of a Representative Elementary Volume (REV). In *Advances in Transport Phenomena in Porous Media*, pages 3–20. Springer Netherlands, 1987.
- [13] Ioannis Baltas, Anastasios Xepapadeas, and Athanasios N. Yannacopoulos. Robust control of parabolic stochastic partial differential equations under model uncertainty. *European Journal of Control*, 46:1–13, 2019.
- [14] John Banhart. *Advanced Tomographic Methods in Materials Research and Engineering*. Monographs on the Physics and Chemistry of Materials. Oxford Univ. Press, Oxford, 2008.
- [15] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv*, 2018. arXiv:1701.02434.
- [16] Jay R. Black, Susan A. Carroll, and Ralf R. Haese. Rates of mineral dissolution under CO<sub>2</sub> storage conditions. *Chemical Geology*, 399:134–144, 2015.
- [17] P. C. Carman. Fluid flow through granular beds. *Transactions, Institution of Chemical Engineers, London*, 15:150–166, 1937.
- [18] W. D. Carrier. Goodbye, hazen; hello, kozeny-carman. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(11):1054–1056, November 2003.
- [19] Francisco J. Carrillo, Cyprien Soulaïne, and Ian C. Bourg. The impact of sub-resolution porosity on numerical simulations of multiphase flow. *Advances in Water Resources*, 161:104094, 2022.
- [20] Holger Class, Anozie Ebigbo, Rainer Helmig, Helge K. Dahle, Jan M. Nordbotten, Michael A. Celia, Pascal Audigane, Melanie Darcis, Jonathan Ennis-King, Yaqing Fan, Bernd Flemisch, Sarah E. Gasda, Min Jin, Stefanie Krug, Diane Labregere, Ali Naderi Beni, Rajesh J. Pawar, Adil Sbai, Sunil G. Thomas, Laurent Trenty, and Lingli Wei. A benchmark study on problems related to CO<sub>2</sub> storage in geologic formations. *Computational Geosciences*, 13(4):409–434, 2009.
- [21] Adam D. Cobb and Brian Jalaian. Scaling Hamiltonian Monte Carlo inference for Bayesian neural networks with symmetric splitting. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 675–685. PMLR, 2021.
- [22] Joffrey Coheur, Thierry E. Magin, Philippe Chatelain, and Maarten Arnst. Bayesian identification of pyrolysis model parameters for thermal protection materials using an adaptive gradient-informed sampling algorithm with application to a Mars atmospheric entry. *International Journal for Uncertainty Quantification*, 13(2), 2023. Publisher: Begel House Inc.
- [23] D. Degen, D. Caviedes Voullième, S. Buitter, H.-J. Hendriks Franssen, H. Vereecken, A. González-Nicolás, and F. Wellmann. Perspectives of physics-based machine learning for geoscientific applications governed by partial differential equations. *Geoscientific Model Development Discussions*, 2023:1–50, 2023.
- [24] Marta D’Elia, Hang Deng, Cedric Fraces, Krishna Garikipati, Lori Graham-Brady, Amanda Howard, et al. Machine Learning in Heterogeneous Porous Materials. *arXiv*, 2022. arXiv:2202.04137.
- [25] Filip Dutka, Vitaliy Starchenko, Florian Osselin, Silvana Magni, Piotr Szymczak, and Anthony J.C. Ladd. Time-dependent shapes of a dissolving mineral grain: Comparisons of simulations with microfluidic experiments. *Chemical Geology*, 540:119459, 2020.

- [26] Jean-Matthieu Etancelin, Peter Moonen, and Philippe Poncet. Improvement of remeshed lagrangian methods for the simulation of dissolution processes at pore-scale. *Advances in Water Resources*, 146:103780, 2020.
- [27] Matthew D. Hoffman and Andrew Gelman. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [28] L. Hume and P. Poncet. A velocity-vorticity method for highly viscous 3d flows with application to digital rock physics. *Journal of Computational Physics*, 425:109910, 2021.
- [29] Baasansuren Jadamba, Akhtar A. Khan, Miguel Sama, Hans-Jorg Starkloff, and Christiane Tammer. A convex optimization framework for the inverse problem of identifying a random parameter in a stochastic partial differential equation. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):922–952, 2021.
- [30] Richard A. Ketcham and William D. Carlson. Acquisition, optimization and interpretation of x-ray computed tomographic imagery: applications to the geosciences. *Computers & Geosciences*, 27(4):381–400, 2001. 3D reconstruction, modelling & visualization of geological materials.
- [31] J. Kozeny. Ueber kapillare leitung des wassers im boden. *Sitzungsber Akad. Wiss., Wien*, 136(2a):271–306, 1927.
- [32] Didier Lasseux, Michel Quintard, and Stephen Whitaker. Determination of permeability tensors for two-phase flow in homogeneous porous media: Theory. *Transport in Porous Media*, 24(2):107–137, 1996-08.
- [33] Didier Lasseux, Tony Zaouter, and Francisco J. Valdés-Parada. Determination of klinkenberg and higher-order correction tensors for slip flow in porous media. *Phys. Rev. Fluids*, 8:053401, May 2023.
- [34] Li Li, Catherine A. Peters, and Michael A. Celia. Upscaling geochemical reaction rates using pore-scale network modeling. *Advances in Water Resources*, 29(9):1351–1370, 2006.
- [35] Qingyang Lin, Yousef Al-Khulaifi, Martin J. Blunt, and Branko Bijeljic. Quantification of sub-resolution porosity in carbonate rocks by applying high-salinity contrast brine using x-ray microtomography differential imaging. *Advances in Water Resources*, 96:306–322, 2016.
- [36] Kevin Linka, Amelie Schäfer, Xuhui Meng, Zongren Zou, George Em Karniadakis, and Ellen Kuhl. Bayesian Physics Informed Neural Networks for real-world nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, page 115346, 2022.
- [37] Fei Lu, Matthias Morzfeld, Xuemin Tu, and Alexandre J. Chorin. Limitations of polynomial chaos expansions in the bayesian solution of inverse problems. *Journal of Computational Physics*, 282:138–147, 2015.
- [38] Suryanarayana Maddu, Bevan L. Cheeseman, Ivo F. Sbalzarini, and Christian L. Müller. Stability selection enables robust learning of differential equations from limited noisy data. *Proceedings of the Royal Society A*, 478(2262):20210916, 2022.
- [39] Suryanarayana Maddu, Dominik Sturm, Christian L Müller, and Ivo F Sbalzarini. Inverse Dirichlet weighting enables reliable training of physics informed neural networks. *Machine Learning: Science and Technology*, 3(1):015026, 2022.
- [40] Youssef M. Marzouk and Habib N. Najm. Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009.
- [41] Y. Mehmani, T. Sun, M. T. Balhoff, P. Eichhubl, and S. Bryant. Multiblock pore-scale modeling and upscaling of reactive transport: Application to carbon sequestration. *Transport in Porous Media*, 95(2):305–326, 2012.
- [42] Xuhui Meng, Hessam Babae, and George Em Karniadakis. Multi-fidelity Bayesian neural networks: Algorithms and applications. *Journal of Computational Physics*, 438:110361, 2021.

- [43] Hannah P. Menke, Branko Bijeljic, Matthew G. Andrew, and Martin J. Blunt. Dynamic Three-Dimensional Pore-Scale Imaging of Reaction in a Carbonate at Reservoir Conditions. *Environmental Science & Technology*, 49(7):4407–4414, 2015. Publisher: American Chemical Society.
- [44] Sergi Molins, Cyprien Soulaire, Nikolaos I. Prasianakis, Aida Abbasi, Philippe Poncet, Anthony J. C. Ladd, Vitalii Starchenko, Sophie Roman, David Trebotich, Hamdi A. Tchelepi, and Carl I. Steefel. Simulation of mineral dissolution at the pore scale with evolving fluid-solid interfaces: review of approaches and benchmark problem set. *Computational Geosciences*, 25(4):1285–1318, 2021.
- [45] Sergi Molins, David Trebotich, Gregory H. Miller, and Carl I. Steefel. Mineralogical and transport controls on the evolution of porous media texture using direct numerical simulation. *Water Resources Research*, 53(5):3645–3661, 2017.
- [46] Sergi Molins, David Trebotich, Carl I. Steefel, and Chaopeng Shen. An investigation of the effect of pore scale flow on average geochemical reaction rates using direct numerical simulation. *Water Resources Research*, 48(3), 2012.
- [47] Joseph P. Molnar and Samuel J. Grauer. Flow field tomography with uncertainty quantification using a Bayesian physics-informed neural network. *Measurement Science and Technology*, 33(6):065305, 2022.
- [48] A. Mondal, Y. Efendiev, B. Mallick, and A. Datta-Gupta. Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump markov chain monte carlo methods. *Advances in Water Resources*, 33(3):241–256, 2010.
- [49] Peyman Mostaghimi, Martin J. Blunt, and Branko Bijeljic. Computations of absolute permeability on micro-CT images. *Mathematical Geosciences*, 45(1):103–125, 2013.
- [50] Mathias Nehler, Ferdinand Stoeckhert, Anne Oelker, Jörg Renner, and Erik Saenger. Evaluating porosity estimates for sandstones based on x-ray micro-tomographic images. *Solid Earth Discussions*, pages 1–45, 03 2019.
- [51] Catherine Noiriél, Matthias Oursin, Giuseppe Saldi, and David Haberthür. Direct determination of dissolution rates at crystal surfaces using 3d x-ray microtomography. *ACS Earth and Space Chemistry*, 3(1):100–108, 2019. Publisher: American Chemical Society.
- [52] Catherine Noiriél and François Renard. Four-dimensional X-ray micro-tomography imaging of dynamic processes in geosciences. *Comptes Rendus. Géoscience*, 354:255–280, 2022.
- [53] Catherine Noiriél and Cyprien Soulaire. Pore-Scale Imaging and Modelling of Reactive Flow in Evolving Porous Media: Tracking the Dynamics of the Fluid–Rock Interface. *Transport in Porous Media*, 140(1):181–213, 2021.
- [54] Ryan L. Payton, Yizhuo Sun, Domenico Chiarella, and Andrew Kingdon. Pore scale numerical modelling of geological carbon storage through mineral trapping using true pore geometries. *Transport in Porous Media*, 141(3):667–693, 2022.
- [55] Sarah Perez, Suryanarayana Maddu, Ivo F. Sbalzarini, and Philippe Poncet. Adaptive weighting of Bayesian physics informed neural networks for multitask and multiscale forward and inverse problems. *Journal of Computational Physics*, 491:112342, 2023.
- [56] Sarah Perez, Peter Moonen, and Philippe Poncet. On the Deviation of Computed Permeability Induced by Unresolved Morphological Features of the Pore Space. *Transport in Porous Media*, 141(1):151–184, 2022.
- [57] Apostolos F. Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, 477:111902, 2023.
- [58] M. Quintard and S. Whitaker. Two phase flow in heterogeneous porous media: the method of large-scale averaging. *Transport in Porous Media*, 3:357–413, 1987.
- [59] Michel Quintard and Stephen Whitaker. Two-phase flow in heterogeneous porous media: The method of large-scale averaging. *Transport in Porous Media*, 3(4):357–413, 1988.

- [60] Michel Quintard and Stephen Whitaker. Transport in ordered and disordered porous media: volume-averaged equations, closure problems, and comparison with experiment. *Chemical Engineering Science*, 48(14):2537–2564, 1993.
- [61] Javier E. Santos, Duo Xu, Honggeun Jo, Christopher J. Landry, Maša Prodanović, and Michael J. Pyrcz. PoreFlow-Net: A 3D convolutional neural network to predict fluid flow through porous media. *Advances in Water Resources*, 138:103539, 2020.
- [62] Steffen Schlüter, Adrian Sheppard, Kendra Brown, and Dorthe Wildenschild. Image processing of multiphase images obtained via x-ray microtomography: A review. *Water Resources Research*, 50(4):3615–3639, 2014.
- [63] Martina Siena, Gianlorenzo Bussetti, Chiara Recalcati, Monica Riva, Lamberto Duò, and Alberto Guadagnini. Statistical characterization of heterogeneous dissolution rates of calcite from in situ and real-time AFM imaging. *Transport in Porous Media*, 140(1):291–312, 2021.
- [64] Martina Siena and Monica Riva. Impact of geostatistical reconstruction approaches on model calibration for flow in highly heterogeneous aquifers. *Stochastic Environmental Research and Risk Assessment*, 34(10):1591–1606, 2020.
- [65] Pavel Smal, Philippe Gouze, and Olivier Rodriguez. An automatic segmentation algorithm for retrieving sub-resolution porosity from x-ray tomography images. *Journal of Petroleum Science and Engineering*, 166:198–207, 2018.
- [66] C. Soulaïne, F. Gjetvaj, C. Garing, S. Roman, A. Russian, P. Gouze, and H. A. Tchelepi. The Impact of Sub-Resolution Porosity of X-ray Microtomography Images on the Permeability. *Transport in Porous Media*, 113(1):227–243, 2016.
- [67] Cyprien Soulaïne, Sophie Roman, Anthony Kavscek, and Hamdi A. Tchelepi. Mineral dissolution and wormholing from a pore-scale perspective. *Journal of Fluid Mechanics*, 827:457–483, 2017. Edition: 2017/08/24 Publisher: Cambridge University Press.
- [68] Cyprien Soulaïne, Sophie Roman, Anthony Kavscek, and Hamdi A. Tchelepi. Pore-scale modelling of multiphase reactive flow: application to mineral dissolution with production of CO<sub>2</sub>. *Journal of Fluid Mechanics*, 855:616–645, 2018.
- [69] Carl I. Steefel and Antonio C. Lasaga. A coupled model for transport of multiple chemical species and kinetic precipitation/dissolution reactions with application to reactive flow in single phase hydrothermal systems. *American Journal of Science*, 294(5):529–592, 1994.
- [70] Carl I. Steefel, Sergi Molins, and David Trebotich. Pore scale processes associated with subsurface CO<sub>2</sub> injection and sequestration. *Reviews in Mineralogy and Geochemistry*, 77(1):259–303, 2013.
- [71] N. Wakao and J.M. Smith. Diffusion in catalyst pellets. *Chemical Engineering Science*, 17(11):825 – 834, 1962.
- [72] Ying Da Wang, Martin J. Blunt, Ryan T. Armstrong, and Peyman Mostaghimi. Deep learning in pore scale imaging and modeling. *Earth-Science Reviews*, 215:103555, 2021.
- [73] Stephen Whitaker. Theory and applications of transport in porous media: The method of volume averaging. *The Netherlands: Kluwer Academic Publishers*, page 81, 1999.
- [74] D. Wildenschild and A. Sheppard. X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Advances in Water Resources*, 51:217–246, 2013.
- [75] Andrew G Wilson and Pavel Izmailov. Bayesian Deep Learning and a Probabilistic Perspective of Generalization. In *Advances in Neural Information Processing Systems*, volume 33, pages 4697–4708. Curran Associates, Inc., 2020.
- [76] B. D. Wood, K. Radakovich, and F. Golfier. Effective reaction at a fluid–solid interface: Applications to biotransformation in porous media. *Advances in water resources*, 30(6-7):1630–1647, 2007.
- [77] Liang Yan and Tao Zhou. Adaptive multi-fidelity polynomial chaos approach to bayesian inference in inverse problems. *Journal of Computational Physics*, 381:110–128, 2019.



- [78] Liang Yan and Tao Zhou. An adaptive surrogate modeling based on deep neural networks for large-scale bayesian inverse problems. *Communications in Computational Physics*, 28(5):2180–2205, 2020.
- [79] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.
- [80] Yongfei Yang, Fugui Liu, Jun Yao, Stefan Iglauer, Mozhdeh Sajjadi, Kai Zhang, Hai Sun, Lei Zhang, Junjie Zhong, and Vadim Lisitsa. Multi-scale reconstruction of porous media from low-resolution core images using conditional generative adversarial networks. *Journal of Natural Gas Science and Engineering*, 99:104411, 2022.
- [81] Yulai Zhang, Peyman Mostaghimi, and Ryan T. Armstrong. On the challenges of greyscale-based quantifications using X-ray computed microtomography. *Journal of Microscopy*, 275(2):82–96, 2019.
- [82] Li Zhuang, Hyu-Soung Shin, Sun Yeom, Chuyen Ngoc Pham, and Young-Jae Kim. A novel method for estimating subresolution porosity from CT images and its application to homogeneity evaluation of porous media. *Scientific Reports*, 12(1):16229, 2022.