Solving Multiphysics, Multiparameter, Multimodal Inverse Problems: An Application to NMR Relaxation in Porous Media

Rupeng Li, Igor Shikhov, and Christoph H. Arns*
School of Minerals and Energy Resources University of New South Wales, Sydney, NSW 2052, Australia

(Received 5 November 2020; revised 27 March 2021; accepted 5 April 2021; published 3 May 2021)

A general and robust Bayesian optimization framework for the extraction of intrinsic physical properties from an integration of pore-scale forward modeling and experimental measurements of macroscopic system responses is developed. The efficiency of the scheme, which utilizes Gaussian process regression, enables the simultaneous extraction of multiple intrinsic physical properties with a minimal number of function evaluations. Here it is applied to nuclear magnetic resonance (NMR) relaxation responses, paving the way for inverse problem approaches to digital rock physics given its general nature. NMR relaxation responses of fluids in porous media may be described by sums of multiexponential decays resulting in a relaxation time distribution. The shape of this distribution is dependent on intrinsic physical system properties, but also effects like diffusion coupling between different relaxation regimes in heterogeneous porous materials. Forward models based on high-resolution images are employed to naturally incorporate structural heterogeneity and diffusive motion without limiting assumptions. Extracting the required multiple intrinsic parameters of the system poses an ill-conditioned multiphysics multiparameter inverse problem where multiple scales are covered by the underlying microstructure. Exploration of the multidimensional search space given an expensive cost function makes an efficient solution strategy mandatory. We propose a workflow to match experimental measurements with simulations via Bayesian optimization, with special attention paid to the multimodal nature of the topography of the objective function using solution space partitioning. A multimodal search strategy using state-of-the-art evolutionary algorithms and gradient-based optimization algorithms guarantees that the multimodal nature is captured. The workflow is demonstrated on T_2 relaxation responses of Bentheimer sandstone, extracting three physical parameters simultaneously: the surface relaxivity of quartz grains, the effective transverse relaxation time, and the effective diffusion coefficient in clay regions. Multiple mathematically sound and physically plausible solutions corresponding to global minimum and multiple local minima of the objective function are identified within a limited number of function evaluations. Importantly, the shape of the experimental T_2 distribution is recovered almost perfectly, enabling the use of classical interpretation techniques and local analysis of responses based on numerical simulation.

DOI: 10.1103/PhysRevApplied.15.054003

I. INTRODUCTION

The characterization of materials, identification of chemical, physical, and structural properties of matter as well as governing processes and phenomena from known observations poses inverse problems in applied sciences. Given that the same observation may be the result of different combinations of system properties and governing phenomena, specific response mechanisms or unique property values may be difficult to establish. The fundamental ill-posedness of inverse problems, usually accompanied with nonlinearity and lack of stability of associated equations, dramatically complicates their solutions [1,2].

An important example is the interpretation of nuclear magnetic resonance (NMR) relaxation responses in porous media. The observed NMR signal is an ensemble average of spin states brought into nonequilibrium by a series of rf pulses. Spins return to equilibrium due to interactions with various surfaces, intrinsic processes in the fluids themselves, and diffusion in internal fields causing dephasing. The associated relaxation processes can be described by a sum of multiexponential decays, each having a specific relaxation rate. The determination of a fractional contribution from a given relaxation time (spectrum) to a total composite response is a nontrivial task and constitutes an ill-posed and ill-conditioned inverse problem. The spectrum can be inferred following one of the mathematical techniques commonly known as the inverse Laplace transform [3,4]. NMR relaxation responses from saturated porous systems provide important yet limited amounts of explicit information, e.g., pore-size estimates when a reference length scale is available—often chosen

*c.arns@unsw.edu.au

to be a mode of a pore-size distribution derived from mercury intrusion capillary pressure measurements despite the fact that it actually rather defines pore accessibility, and the application of a constant surface relaxivity [5,6]. Fluid saturations can be inferred from relaxation time or diffusion contrast between different fluids utilizing customized acquisition techniques relying on, e.g., partial polarization, T_1 weighting, relaxation-diffusion correlations and are subject to knowledge of wetting conditions or micro- and macroporosity relationships [7].

The identification of intrinsic fluid and surface properties of rock samples is problematic. For example, values of surface relaxivity of fully water-saturated Bentheimer sandstone reported in the literature fall into a broad interval ranging between 7.6 [8] and 15.4 μm/s [9]. Hardly any value has ever been reported on the effective diffusion coefficient of water-saturated kaolinite (as a rock constituent). Relaxation of water in kaolinite has been reported on artificially made samples with the important (and expected) qualitative conclusion that relaxation strongly depends on compaction [10] and iron content [11].

Increasingly, simulation tools are utilized for NMR interpretation, such as molecular dynamics simulations [12] and relaxation simulations on digitized representations of porous systems [13–15]. The calculations based on the forward calculation of NMR responses potentially may provide a wide range of physical properties impossible to measure directly or evaluate with any certainty otherwise: effective diffusion in microporous regions and connectivity to porosity, surface relaxivity of specific mineral phases, magnetic properties of selected regions, and more. However, deriving conclusions about intrinsic physical properties by forward simulations requires an extensive search across the multidimensional space of possible parameters to determine the correct configuration of inputs, posing an ill-conditioned multiphysics multiparameter inverse problem where multiple scales are covered by the underlying microstructure. This presents a significant challenge, since numerical simulations of NMR responses in porous media like rocks are computationally expensive, even with random walk methods, if inhomogeneous magnetic fields arising from susceptibility contrast are considered. In this general case a high time resolution is required to capture the evolution of magnetization, leading to a simulation time of several hours. This precludes exhaustive search strategies for higher-dimensional inverse problems.

Bayesian optimization (BO) techniques are some of the most efficient approaches for constrained optimization problems for which the number of function evaluations is severely limited by time or cost. BO has been successfully applied in reinforcement learning [16], computer graphics design [17], robotics [18], etc., in a forward manner to tune the performance of the complex, expensive black-box function. The key to its efficiency is that

(a) under the continuity assumption, BO often interpolates and extrapolates quite accurately over large distances in the search space [19];

(b) the decision making of BO balances not only the need to exploit the regression surface (by sampling where it is minimized) but also the need to improve the approximation (by sampling where prediction error may be high), making every search step well calculated and statistically optimum.

There have been several attempts to solve inverse problems in engineering using Bayesian statistics, e.g., in the geological and hydrogeological characterization of the subsurface structure, where the ill-posedness presents a significant challenge to the determination of the spatially dependent variable conditioned on the observations since a posterior ensemble of models fit the data equally well given limited observations. Mariethoz *et al.* [20] presented the iterative spatial resampling method to perturb realizations of a spatially dependent variable preserving the specified spatial dependence, which can be used as either a proposal distribution with Bayesian Markov chain Monte Carlo (Bayesian MCMC), or a perturbation strategy to reduce the model misfit to the measured data with simulated annealing. However, such methods require Bayesian inference directly on the variable of interest and that the "good-enough fit" be assessed using exact forward simulation, leading to tens of thousands of function evaluations, which preclude problems involving high-dimensional solution space or expensive cost functions; moreover, the ill-posedness is not fully addressed since the solutions are starting-guess dependent such that only local optimal solutions are found. In our work, we take BO commonly used in forward problems and apply it to inverse problems.

The contribution of this paper is threefold. First, we propose the inverse solution workflow (ISW) based on BO to identify values of multiple parameters simultaneously for which the multiscale, multiphysics expensive forward calculation is manipulated to perform nearly identical to observations within a limited computational budget. An overview of the ISW is displayed in Fig. 1. Using this approach, the computational burden is eased by establishing a cheap proxy for the objective function using Gaussian process regression (GPR), so that Bayesian MCMC is performed on GPR hyperparameters that control the topography of the solution space, after which both values and confidence intervals of the actual unknown variables are predicted using such posterior estimates. Next, the specific value of unknown variables with globally the highest potential to reduce the model misfit is identified using expected improvement (EI) criteria, and the new misfit is then calculated using forward NMR simulation once. Bayesian MCMC is then reapplied to accommodate for

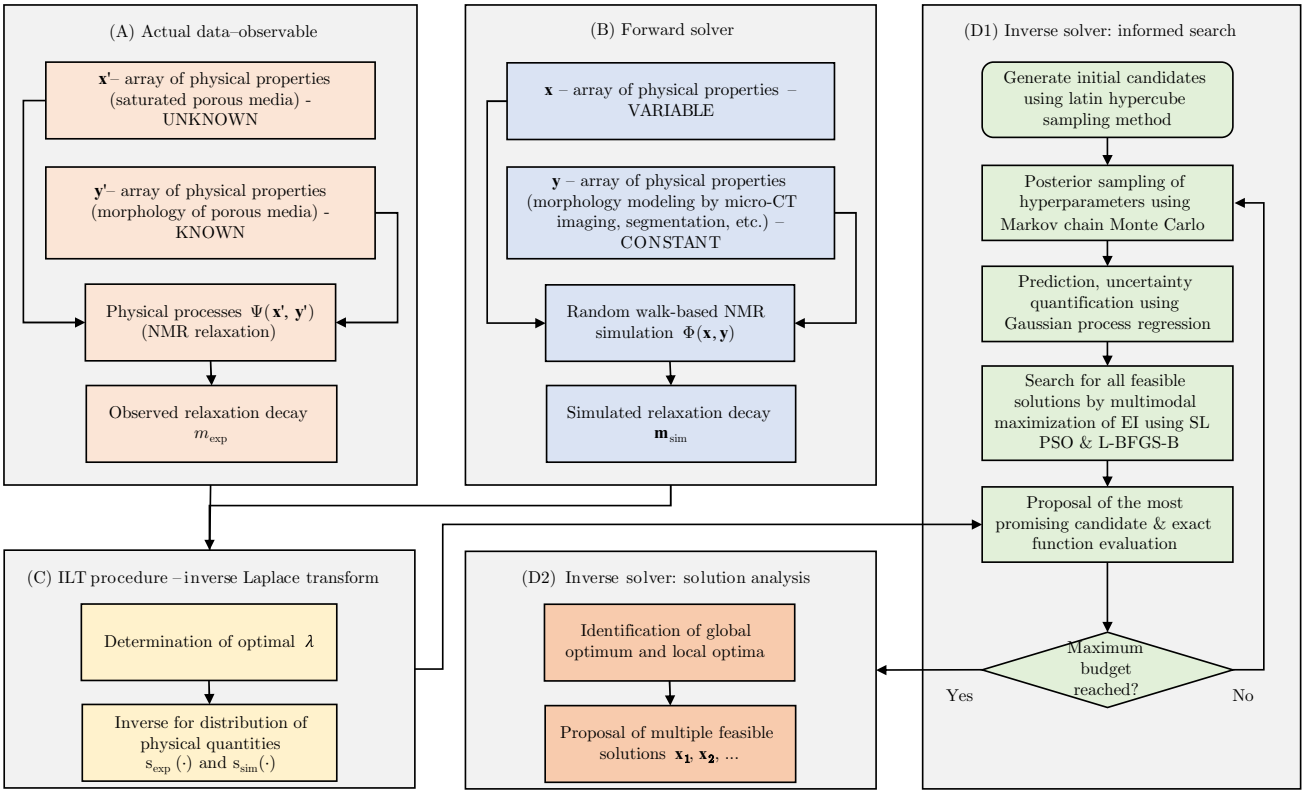

FIG. 1. The detailed workflow to solve an inverse problem of finding feasible combinations of physical properties of saturated sandstone resulting in NMR T_2 distributions identical to the experiment.

the enlarged observation set. This approach is advantageous since (1) locating where the objective function is minimized and where the prediction and uncertainty quantification error may be high using GPR is significantly cheaper than the exact evaluation, (2) pairwise correlations between physical parameters can be identified, (3) inference on hyperparameters unveils sensitivity of the objective function against each variable, and (4) the whole optimization process is reproducible.

Second, we propose a multimodal search strategy using state-of-the-art evolutionary algorithms and gradient-based optimization algorithms that guarantees that both the global optimal solution and (major) local optimal solutions are identified. Since real-world observations, e.g., transverse relaxation time (T_2) measurements, are always subject to noise, the topography of the noise-corrupted, observed objective function is different from that of the true objective function so that the optimality of the solutions might be altered. The issue is alleviated by identification of both the global optimal solution and major local optimal solutions by (1) formulation of a candidate pool comprising the global optimal solution identified using social-learning particle swarm optimization (SL PSO) algorithms [21], as well as local optimal solutions identified using box-constrained limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) [22], followed by (2) a proposal of the most promising candidate from this pool using solution space partitioning (SSP).

Third, we investigate how solutions obtained from the ISW are affected by realizations and levels of noise. Since the computation of the T_2 distribution, $\mathbf{s}(T_2)$, requires transforming the time-domain representation of the signal into a spectrum, which itself poses as an ill-posed inverse problem, the $\mathbf{s}(T_2)$ are subject to realizations of noise, and so is the topography of the objective function and optimality of the solutions. In this paper we identify solutions at seven logarithmically spaced signal-to-noise ratios (SNRs) and study the dependency of identified intrinsic parameters on SNRs.

The workflow is demonstrated on NMR T_2 relaxation responses of Bentheimer sandstone, identifying three physical parameters: the surface relaxivity of quartz grains, the *effective* transverse relaxation time, and the *effective* diffusion coefficient in clay regions. However, it is emphasized that the ISW is a general framework and applies to many other expensive simulations, whether being multiparameter, multiscale, or multiphysics.

This paper is organized as follows. In Sec. II we provide a summary of the ISW and its four main components, followed by a dedicated section detailing the optimization framework. Results and discussion on identification of

three physical parameters, together with a study on the effect of SNRs and the corresponding regularization parameter, are given in Sec. IV. This is followed by a sensitivity analysis in Sec. V and a discussion on practical issues in Sec. VI. We summarize our work in Sec. VII. Finally, for clarity of notation, we provide nomenclature lists in the Appendix.

II. INVERSE SOLUTION WORKFLOW

Inverse methods are powerful mathematical tools, employed, e.g., in astrophysics, thermophysics, gravitational lensing, seismological exploration (geophysics), and underwater acoustic tomography [23–25]. In abstract form the inverse problem can be represented as

$$\Theta(\mathbf{x}) = \mathbf{d} \quad \text{for } \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{d} \in \mathcal{D}, \tag{1}$$

where the operator Θ maps parameters $\mathbf{x}$ in the *parameter or solution space* $\mathcal{X}$ to observables $\mathbf{d}$ in the *data space* $\mathcal{D}$. Here $\mathcal{X} = \{\mathbf{x} \mid \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u\}$, a subset of $\mathbb{R}^N$, is a nonempty domain constrained by lower bounds ($\mathbf{x}_l$) and upper bounds ($\mathbf{x}_u$) defining the possible range of physical parameters. The solution of Eq. (1) can be obtained by minimization of the least-squares residual, i.e.,

$$\begin{aligned} &\text{minimize } ||\Theta(\mathbf{x}) - \mathbf{d}||_F^2 \\ &\text{subject to } \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{d} \in \mathcal{D}, \end{aligned} \tag{2}$$

where $||\cdot||_F$ denotes the Frobenius norm.

Consider now the NMR T_2 relaxation response of a natural porous system, which, without loss of generality, may be considered to be a sandstone for concreteness; there are directly observable physical quantities, e.g., the bulk relaxation time of water $T_{2b,w}$, diffusion coefficient of water D_w, and volumetric susceptibility of minerals χ_v. We are interested in the unknown parameter(s) $\mathbf{x}$ that are difficult to measure or disputable, e.g., surface relaxivity of quartz grains $\rho_q^\star$ ($\rho_q^\star$ is the continuum surface relaxivity that is proportionally larger than the discrete surface relaxivity ρ_q that is used for optimization, as will be described later), *effective* transverse relaxation time $T_{2e,c}$, or *effective* diffusion coefficient in clay regions $D_{e,c}$. Following the principles used in effective medium theory [26], the two key properties of interest are named "effective" to stress that these properties are not just straight volumetric averages, but also weighted by time-dependent processes, describing the behavior of the system. For instance, diffusion averaging is true for miscible fluids and also applicable to diffusional exchange at the imaginary interface between different regions. However, an effective diffusion process is the one experiencing sufficiently complete tortuosity effects observed at long observation times. It is "an average diffusion" if only applicable time and spatial scales are considered. Such values can be measured directly only if we have abundant clay with identical structure to that from the rock sample, which is practically impossible.

Given known or directly measured values of some key physical parameters, the remaining key physical quantities can be determined by matching measured and simulated T_2 distributions, by minimization of the squared difference between the two T_2 distributions using the ISW, with weights for all data points of the T_2 distribution being equally 1, which may be expressed as

$$\begin{aligned} &\text{minimize } \|\mathbf{s}_{\text{sim}}(T_2 \mid \mathbf{x}) - \mathbf{s}_{\text{exp}}(T_2)\|_F^2, \\ &\text{for } \mathbf{x} = [\rho_q, T_{2e,c}, D_{e,c}]^\top, \\ &\text{subject to } \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{3}$$

Here $\mathbf{s}_{\text{sim}}(T_2 \mid \mathbf{x})$ and $\mathbf{s}_{\text{exp}}(T_2)$ denote the discretized T_2 distributions inverted from simulated and measured NMR decays, respectively, whereas $\mathbf{s}_{\text{sim}}$ is conditioned on the values of the unknowns.

Solving Eq. (3) is problematic, since it requires evaluation of the feasibility of a candidate using the expensive physical simulator, making traditional enumeration methods, e.g., grid search, impractical especially if the dimension of $\mathcal{X}$ is high. Beyond that, since the topography of the noise-corrupted, observed objective function is different from that of the true but unknown objective function, the optimality of the solutions might be altered, and typical global optimization algorithms might not satisfy our need. For such considerations, we propose a workflow to identify multiple feasible combinations of physical quantities simultaneously by reproducing nearly identical experimental T_2 distributions, which is demonstrated in Fig. 1.

The workflow consists of four main modules, namely the generation of actual data (A), the forward solver (B), the inverse Laplace transform procedure for the NMR inverse problem (C), and the inverse solver for the determination of the unknown physical parameters, which carries out the informed search (D1) and solution analysis (D2). These modules are detailed in the following subsections.

A. Reference data—observable

As example porous medium we use an outcrop Bentheimer sandstone quarried in Germany's North West—one of the most common reference rocks. This sandstone is a consolidated quartz arenite made of well-sorted subrounded to rounded grains of about 200 μm mean diameter. It exhibits 23.9% porosity and a low iron content of 3 wt % (expressed in Fe_2O_3), resulting in the rock's weak paramagnetism. Measurement of powdered specimens using a magnetic susceptibility balance (Sherwood Scientific) provided a volumetric susceptibility value of 1.8 μSI. We estimated the specific surface area of Bentheimer being $0.20 \sim 0.35$ m^2/g, of which quartz contributes approximately 20% and a minor quantity of

kaolinite provides approximately 75% [27]. A cylindrical core sample of 50 mm length and 25 mm in diameter is fully saturated with 3-wt % NaCl brine prior to NMR measurements.

Reference proton NMR T_2 relaxation measurements are performed using a 2 MHz Magritek Rock Core Analyzer equipped with a P54 probe. The magnet temperature is kept stable at 34.2 °C, while the sample is kept at 25.0 °C. Transverse magnetization relaxation decays are acquired using the standard Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence [28,29], composed by an initial $\pi/2$ pulse placing the magnetization vector into the transverse plane, followed by an echo train of refocusing π pulses (50 000 in total). All pulses are of 20 μs length. The time interval between π pulses (echo time) is set to $t_E = 200$ μs. Experiments are performed as a sum of quadruple scans following the standard four-step phase cycling sequence CYCLOPS (cyclically ordered phase sequence) [30]. Altogether we acquired 320 individual relaxation decays. Combining these allowed us to obtain a desired SNR over a fairly broad range. A similar procedure is utilized to obtain matching background noise signals.

B. Forward solver—simulated NMR response

The NMR relaxation response is simulated by a lattice random walk method on the segmented image. It generally follows Refs. [31–33], with the added capability of capturing dephasing in a general inhomogeneous magnetic field. For the latter, it follows Ref. [28]. The method is explained in detail in Ref. [13]; a brief account of the method is given here.

A 5 mm diameter sister sample of above Bentheimer sandstone is imaged via micro-x-ray-computed tomography (micro-CT) at a resolution of 2.89 μm. This resolution provides a sufficient field of view and is comparable to the ruler length (or diffusional averaging length) intrinsic to typical NMR measurements of surface to volume ratio, which is of the order of 2 μm [34,35]. The tomogram is corrected for beam hardening, filtered to reduce noise, and segmented into three phases with an active converging contour method [36], namely pore space (w), quartz (q), and clay region (c), as is shown in Fig. 2. The clay region is an unresolved mixed-phase containing mainly kaolinite and porosity.

Given the measured average volumetric susceptibility of dry Bentheimer sandstone $\chi_{v,\mathrm{ave}} = -7.6$ μSI, the NaCl brine $\chi_{v,w} = -9.0$ μSI, and assuming that the susceptibility of quartz is $\chi_{v,q} = -10.0$ μSI, we assigned the *effective* susceptibility for the clay regions (including 50% of brine saturated voids) of $\chi_{v,c} = 60.0$ μSI. The internal magnetic field at lattice resolution ϵ is derived in the dipole approximation [13,37] by carrying out the convolution operation between the susceptibility field $\chi(\mathbf{r})$ and the dipole field

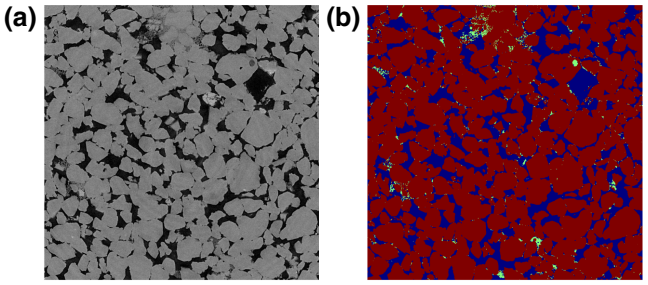

FIG. 2. Cross section through a 1120^3 voxel subdomain of Bentheimer sandstone; (a) tomogram and (b) phase segmented image with resolution $\epsilon = 2.89$ μm (blue represents the pore space, red represents the quartz, green represents the clay region).

$\mathbf{B}_{\mathrm{dip}}(\mathbf{r})$ as

$$\mathbf{B}_{\mathrm{int}}(\mathbf{r}) = (\chi_v * \mathbf{B}_{\mathrm{dip}})(\mathbf{r}) = \int \chi_v(\mathbf{r})\mathbf{B}_{\mathrm{dip}}(\mathbf{r}-\mathbf{r'})\,d\mathbf{r'} \quad (4)$$

with

$$\mathbf{B}_{\mathrm{dip}}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0}{4\pi}\left[\dfrac{3(\mathbf{m}\,\mathbf{r})\mathbf{r} - \mathbf{m}r^2}{r^5}\right] & \text{for } r > a, \\ \dfrac{2}{3}\mu_0\mathbf{m} & \text{for } r \le a, \end{cases} \quad (5)$$

$$\mathbf{m} = \mu_0 \frac{\chi_v}{1+\chi_v}\mathbf{B}.$$

Here a is the radius of the dipole, r is the distance from the dipole center, $\mu_0 = 4\pi \times 10^{-7}$ N A^{-2} is the magnetic permeability of the vacuum, and $\mathbf{m}$ is the magnetic dipole moment for a unit volume. For a given structure and internal magnetic field, the NMR response is calculated as the ensemble average of random walks on a subgrid with resolution of $\epsilon_w = \epsilon/l_s$ and time step $\tau_i = \epsilon_w^2/[6D(\mathbf{r})]$; in each time step i walkers are attempting to move to a neighboring side. We assign D_w to the open pore space, while the *effective* diffusion coefficient of the clay region $D_{e,c}$ is considered as unknown and varied in a broad range $[10^{-7}\ \mathrm{cm}^2/s; 10^{-4}\ \mathrm{cm}^2/\mathrm{s}]$. All diffusion terms involving the solid are set zero, pore-space clay-region cross terms are set to the geometrical average, and the probability to cross into the clay regions is scaled by porosity. The spatial resolution of the random walk is chosen by setting $l_s \in [5, 20]$ as a function of $\min(D_w, D_{e,c})$. Lattice random walks are started proportionally to voxel porosity and $\mathbf{B}_{\mathrm{int}}(\mathbf{r})$ of subgrid positions is derived by trilinear interpolation.

Treating coherent and incoherent parts of the magnetization decay separately, the transverse magnetization decay can be written as

$$M_{xy}(t_j) = \langle M_w(t_j)\rangle = \left\langle M_w(0)\cos[\phi_D(t_j)]\prod_i S_i \right\rangle, \quad (6)$$

where $M_w(t_j)$ is the magnetization decay of an individual walker. Here $\phi_D(t)$ is the accumulated phase at time t_j,

$$\phi_D = \phi - \phi_0 = \sum_{j=1}^{N} \gamma \tau_j [B_z(t_j) - B_z(0)], \qquad t_j = \sum_{i=1}^{j} \tau_i, \tag{7}$$

for a random walk starting with Lamor frequency $\omega_0 = \gamma B_z(0)$ and γ the gyromagnetic ratio. Here $\prod_i S_i$ incorporates surface and bulk or effective relaxation with $S_i = S_b S_s$, where $S_b = \exp(-\tau_i/T_{2b,w})$ for bulk relaxation, $S_b = \exp(-\tau_i/T_{2e,c})$ for effective relaxation in clay regions, and $S_s = 1 - 6\rho_2\tau_i/(\epsilon_w A)$ for surface relaxation ($S_s = 1$ for steps within the same phase), where A is a correction factor of order 1, which accounts for the details of the random walk implementation [32]. For imaged structures, this value is somewhat arbitrary and close to $A = 3/2$ for our random walk implementation; we report the discrete surface relaxivity ρ_q in the results for the inverse problem section, which is subject to a respective surface area correction factor. The continuous surface relaxivity of quartz is then the product of A and ρ_q. The CPMG sequence is implemented by recording M_{xy} at all integer multiples of the echo time $t_E = 2\tau$ and switching the sign of the accumulated phase at times $\tau + n_t t_E$, $n_t = 0, 1, \ldots, N_t$, with $N_t t_E$ the duration of the simulated relaxation decay. This implements a hard pulse approximation. Unknown parameters are the discrete surface relaxivity of quartz ρ_q, the effective transverse relaxation time of the clay region $T_{2e,c}$, and the *effective* diffusion coefficient of the clay region $D_{e,c}$. We set $\rho_c = 0$ given that surface relaxation within clay regions is already incorporated into $T_{2e,c}$.

C. Inverse Laplace transform procedure

It is widely accepted and considered certainly true for the broad range of artificial and natural porous media, including saturated rock samples, that due to the distribution of pore sizes the measured NMR relaxation signal is a combination of individual signal contributions represented by variously weighted monoexponential decays:

$$m_t(t) = \int s(T_2)\, e^{-t/T_2}\, dT_2 + \epsilon_n(t). \tag{8}$$

Here $m_t(t)$ is the measured data, $\epsilon_n(t)$ is the noise term assumed Gaussian, white and additive, and our objective is to estimate the probability density function of the transverse relaxation time $s(T_2)$.

Equation (8) is a Fredholm integral of the first kind known to be an ill-posed problem, and the density function $s(T_2)$ can be obtained following a procedure commonly known as the inverse Laplace transform (ILT). Given the observed time-domain transverse relaxation decay $\mathbf{m}_t$, Eq. (8) can be expressed in matrix notation after discretization as

$$\mathbf{m}_t = K\mathbf{s} + \mathbf{e}_n, \tag{9}$$

where $\mathbf{m}_t \in \mathbb{R}^{m\times 1}$ is the time-domain data, $\mathbf{s} \in \mathbb{R}^{N\times 1}$ is the discretized probability density function of T_2, $K \in \mathbb{R}^{m\times N}$ is the inversion kernel, and $\mathbf{e}_n \in \mathbb{R}^{m\times 1}$ is the noise usually assumed zero mean Gaussian. Seeking a solution in the least square sense and using a Tikhonov regularizer, Eq. (8) becomes an inverse problem, i.e.,

$$\begin{aligned} &\text{minimize } \|K\mathbf{s} - \mathbf{m}_t\|_F^2 + \lambda\|\mathbf{s}\|_F^2, \\ &\text{subject to } \mathbf{s} \geq \mathbf{0}, \end{aligned} \tag{10}$$

where λ is the regularization parameter. Equation (10) is a linearly constrained convex quadratic programming problem. Inversion from decay to T_2 distribution is achieved by minimization of the squared difference between decay and its reconstruction together with the second norm of the regularization term. The weights for all observation points on the decay are equally 1. For notational clarity, we refer to the solvers for Eq. (10), i.e., the Newton active-set method, as the ILT procedure [3,38].

Selection of λ_{opt} is nontrivial because (1) a smaller λ makes $\mathbf{s}(T_2)$ exhibit more features, but it should not be too small that $\mathbf{s}(T_2)$ exposes false features that are not reproducible for different noise realizations under the same SNR, and (2) a larger λ makes $\mathbf{s}(T_2)$ less informative but much more stable with respect to different noise realizations, but it should not be too large that not much information is revealed.

In this context, we propose a strategy for determination of λ_{opt} trading off stability (smoothness) and informativeness (resolution), which we discuss in Sec. IV.

D. Inverse solver

1. Informed search

Bayesian optimization enables an informed search such that the majority of function evaluations (FEs) are spent in promising regions. We give a brief description of the workflow that is detailed in Sec. III.

As is shown in Fig. 1, the informed search starts by evaluation of a few candidates randomly generated using the Latin hypercube sampling method. In the next step, a computationally cheap proxy is constructed using GPR with the distribution of model hyperparameters inferred by Bayesian MCMC.

To search within the multimodal regression model for proposing candidates, expected improvement, the amount of fitness value reduction a candidate is expected to induce, is calculated given the predicted mean values and variances, followed by the search for all promising candidates

using SL PSO and L-BFGS-B. The most promising candidate is then proposed for evaluation according to distance and fitness value metrics using SSP, followed by exact evaluation using the forward NMR simulator.

As soon as the simulated NMR responses are available, we use the Newton-active method with Tikhonov regularization to obtain the corresponding T_2 distribution. Candidates are added to the evaluated list upon completion of inversion, followed by a new round of inference, regression, multimodal maximization, candidate proposal, and exact evaluation. The informed search repeats itself until the optimization budget is depleted.

2. Solution analysis

After the optimization budget is used up, all solutions are clustered to the nearest local minima; the top 20 candidates are checked against mathematical and physical plausibility to identify the optimal solution.

III. BAYESIAN OPTIMIZATION FOR MULTIMODAL OBJECTIVE FUNCTIONS

In this section, we describe the construction of the computationally cheap proxy for NMR forward simulation, followed by identification of locations of both global maximum and local maxima (global maximizer and local maximizers) on the multimodal acquisition function, and finally the procedure of solution analysis.

A. A computationally cheap proxy for NMR forward simulation

1. Gaussian process regression

GPR, also well known as kriging in geostatistics [39,40], and as surrogates in computer experiments [41,42], is a nonparametric statistic method that has been actively applied in regression. A Gaussian process (GP) is a distribution over functions fully specified by its mean and covariance functions, which can be expressed as [43]

$$f(\mathbf{x}) \sim \mathcal{GP}[m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')], \tag{11}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_D)^\top$ is a candidate in $\mathcal{X}$ of dimension $D \times 1$, $f(\mathbf{x})$ is the process evaluated at location $\mathbf{x}$, $m(\mathbf{x})$ is the mean function, and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function between pairs of candidates $\mathbf{x}$ and $\mathbf{x}'$. A constant mean $m(\mathbf{x}) = m_0$ is used throughout this paper. We assume no prior knowledge about our problem, and we adopt the smooth, stationary, aperiodic, and infinitely differentiable covariance function: parameterized squared exponential (k_{SE}) as our covariance function, i.e.,

$$k_{\mathrm{SE}}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left[-\tfrac{1}{2}(\mathbf{x}' - \mathbf{x})^\top \Lambda^{-1}(\mathbf{x}' - \mathbf{x})\right], \tag{12}$$

where Λ and σ_f^2 denote the matrix of characteristic length scales and signal amplitude, respectively. Here $\Lambda = \mathrm{diag}\,(\boldsymbol{\ell})^{-2}$ accommodates different levels of sensitivity against each component of $\mathbf{x}$, where $\boldsymbol{\ell} = (l_1, \ldots, l_D)^\top$ is a vector of positive values, referred to as length scales. Incorporation of such length scales implements automatic relevance determination (ARD) [43].

For clarity, we define the observation set $\mathcal{D}_{\mathrm{obs}}$ and the prediction set $\mathcal{D}_{\mathrm{pred}}$ as

$$\begin{aligned} \mathcal{D}_{\mathrm{obs}} &= (X, \mathbf{y}) = \{(\mathbf{x}_i, y_i) \mid i = 1, \ldots, n\}, \\ \mathcal{D}_{\mathrm{pred}} &= (X^*, \mathbf{y}^*) = \{(\mathbf{x}_i^*, y_i^*) \mid i = 1, \ldots, n^*\}, \end{aligned} \tag{13}$$

where X and $\mathbf{y}$ denote observed candidates and observations, respectively, and X^* and $\mathbf{y}^*$ denote predicted candidates and predictions, respectively.

In GPR, the latent variable $\mathbf{f}$ is assigned a Gaussian prior with constant mean $\mathbf{m_0}$, i.e.,

$$\mathbf{f} \mid X \sim \mathcal{N}[\mathbf{m_0}, K_f(X, X)], \tag{14}$$

where $\mathcal{N}$ denotes normal distribution, $\mathbf{m_0} \in \mathbb{R}^{n \times 1}$ denotes the mean vector for observations, and $K_f(X, X)$ denotes the $n \times n$ covariance matrix for the latent variable $\mathbf{f}$ evaluated at all pairs of observed candidates, with the (p, q) entry specified by $k_{\mathrm{SE}}(\mathbf{x}_p, \mathbf{x}_q)$. This is similar for the other entries $K_f(X, X^*)$, $K_f(X^*, X^*)$, and $K_f(X^*, X)$. Equation (14) shows that entries of latent variable $\mathbf{f}$ conditioned on observation X are jointly Gaussian distributed.

We use the Gaussian likelihood function to accommodate the noisy observations $\mathbf{y}$, i.e.,

$$\mathbf{y} \mid \mathbf{f} = \mathcal{N}(\mathbf{f}, \sigma_n^2 I). \tag{15}$$

Combining Eq. (14) with Eq. (15) yields

$$\mathbf{y} \mid X \sim \mathcal{N}[\mathbf{m_0}, K_n(X, X)], \tag{16}$$

where $K_n = K_f + \sigma_n^2 I$ is the covariance for the noisy targets $\mathbf{y}$ and $\mathbf{y}^*$. The joint distribution of the observations $\mathbf{y}$ and the predictions $\mathbf{y}^*$ is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m_0} \\ \mathbf{m_0^*} \end{bmatrix}, \begin{bmatrix} K_n(X, X) & K_f(X, X^*) \\ K_f(X^*, X) & K_n(X^*, X^*) \end{bmatrix}\right), \tag{17}$$

where $\mathbf{m_0^*} \in \mathbb{R}^{n^* \times 1}$ is the mean vector for predictions. By conditioning the predictions on observations, the posterior distribution of predictions can be expressed as

$$\mathbf{y}^* \mid X^*, X, \mathbf{y}, \boldsymbol{\theta} \sim \mathcal{N}[\boldsymbol{\mu}(X^*), \Sigma(X^*)], \tag{18}$$

where $\boldsymbol{\theta}$ is a vector of model hyperparameters, and

$$\begin{aligned} \boldsymbol{\mu}(X^*) &= \mathbf{m_0^*} + K_f(X^*, X) K_n^{-1}(\mathbf{y} - \mathbf{m_0}), \\ \Sigma(X^*) &= K_n^* - K_f(X^*, X) K_n^{-1} K_f(X, X^*). \end{aligned} \tag{19}$$

Specifically, for prediction y on a single predicted candidate $\mathbf{x}^*$, Eq. (18) becomes

$$y^* \mid \mathbf{x}^*, X, \mathbf{y}, \boldsymbol{\theta} \sim \mathcal{N}[\mu(\mathbf{x}^*), \sigma(\mathbf{x}^*)]. \tag{20}$$

2. Bayesian inference on model hyperparameters

There are $D+3$ hyperparameters in our model, assuming a constant mean term that is used for both predictions and observations, i.e.,

$$\boldsymbol{\theta}' = [\sigma_f^2, \sigma_n^2, m_0, l_1^2, \ldots, l_D^2], \tag{21}$$

and practically the base-10 log hyperparameters are used, except for m_0, i.e.,

$$\boldsymbol{\theta} = [\log_{10}(\sigma_f^2), \log_{10}(\sigma_n^2), m_0, \log_{10}(l_1^2), \ldots, \log_{10}(l_D^2)]. \tag{22}$$

It is appealing in terms of simplicity to adopt point estimates of hyperparameters either from maximum likelihood estimation (MLE) or from maximum a posterior (MAP) estimation. However, Bergstra *et al.* [44] pointed out that a fully Bayesian treatment, i.e., inference on the joint distribution of the hyperparameters using Bayes rule, is superior to optimization of hyperparameters. In our case, since a posterior ensemble of models fit the data equally well given limited observations, inference of the joint probability density function of the hyperparameters becomes necessary while a point estimate would rule out other possibilities. Moreover, $\boldsymbol{\theta}_{\mathrm{MLE}}$ or $\boldsymbol{\theta}_{\mathrm{MAP}}$ are initial-guess dependent and may actually converge to one of the local maxima. As a result, we use a fully Bayesian treatment of the hyperparameters with the posterior distribution of hyperparameters given as

$$p(\boldsymbol{\theta} \mid X, \mathbf{y}) = \frac{p(\mathbf{y} \mid X, \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y} \mid X)}, \tag{23}$$

where $p(\boldsymbol{\theta})$, $p(\mathbf{y} \mid X)$, and $p(\boldsymbol{\theta} \mid X, \mathbf{y})$ are the prior, likelihood, and posterior of the hyperparameters.

a. Prior distribution of hyperparameters. Treatment of hyperparameters is important since it is the hyperparameters that control the topography of the GPR proxy, explained in Sec. V. Assuming no prior knowledge on the optimization problem, it is appropriate to use noninformative priors such as a uniform prior $P(\boldsymbol{\theta}) \propto 1$ for the hyperparameters, whereas more specific and effective priors can be used when previous optimization results are available.

As suggested by Gelman [45], we apply an independent log-uniform prior to each hyperparameter,

$$P(\boldsymbol{\theta}_i) \sim \log \mathcal{U}(-\infty, +\infty), \qquad i = 1, 2, \ldots, D+3, \tag{24}$$

which is an improper prior, but practically leads to a proper posterior after checking that $\int p(\boldsymbol{\theta} \mid X, \mathbf{y})\, d\boldsymbol{\theta}$ is finite. How the optimization performance is assisted by or compromised by various hyperpriors is demonstrated in Sec. V.

b. Likelihood functions. In BO the likelihood is the joint probability of the observations given values of a set of hyperparameters, and it is also called marginal likelihood since we marginalize over the function values $\mathbf{f}$, i.e.,

$$\mathcal{L}(\boldsymbol{\theta} \mid X, \mathbf{y}) = p(\mathbf{y} \mid X, \boldsymbol{\theta}) = \int p(\mathbf{y} \mid \mathbf{f}, X, \boldsymbol{\theta}) p(\mathbf{f} \mid X, \boldsymbol{\theta})\, d\mathbf{f}, \tag{25}$$

where the prior $p(\mathbf{f} \mid X, \boldsymbol{\theta})$ and the likelihood $p(\mathbf{y} \mid \mathbf{f}, X, \boldsymbol{\theta})$ are defined in Eqs. (14) and (15), respectively. Performing the integration yields the likelihood in logarithmic form:

$$\log[p(\mathbf{y} \mid X, \boldsymbol{\theta})] = -\frac{1}{2}(\mathbf{y} - \mathbf{m_0})^\top K_n^{-1}(\mathbf{y} - \mathbf{m_0}) - \frac{1}{2}\log|K_n| - \frac{D}{2}\log 2\pi. \tag{26}$$

c. Posterior distribution of hyperparameters. Log transformation from both sides of Eq. (23) yields

$$\log p(\boldsymbol{\theta} \mid X, \mathbf{y}) = \log p(\mathbf{y} \mid X, \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) - \log p(\mathbf{y} \mid X), \tag{27}$$

where $p(\mathbf{y} \mid X) = \int p(\mathbf{y} \mid X, \boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ is a normalization constant that usually does not need to be determined.

d. Posterior sampling of hyperparameters using Bayesian MCMC. In statistics, Bayesian MCMC refers to the algorithm drawing samples from the joint posterior distribution of hyperparameters [in our case Eq. (27)] without explicit calculation of the distribution. Typical choices of MCMC samplers include the Metropolis-Hasting sampler [46], Gibbs sampler [47], etc., which can be very efficient to draw samples from most multivariate distributions, especially when Hamiltonian Monte Carlo is involved, despite tuning being required. Here we use the slice sampler proposed by Neal [48], which performs well with minimal tuning and can be easily extended for parallel chaining purposes.

B. Proposal of promising candidates for exact evaluation

1. Acquisition function

Given the GPR proxy and plausible hyperparameter settings inferred using Bayesian MCMC, we could naively query for the most promising candidate by identifying the global minimizer of the predictive mean, i.e., Eq. (20). The candidate selection criteria, also known as acquisition functions, take both value information and uncertainty information into consideration.

There are a variety of acquisition functions such as EI [49], upper confidence bound [50], entropy search [51] for sequential BO, and multistep EI [52], Kriging believer [53], local-penalization [54] for parallel BO. Since we evaluate one candidate at each iteration, we employ EI as our acquisition function since it exhibits good properties, such as continuity, integrability, and differentiability; meanwhile there is *no* parameter for manual tuning.

In minimization problems, EI acquisition function (a_{EI}) as a function of $\mathbf{x}^*$ is defined as the amount of improvement that the evaluation of a candidate is expected to induce, i.e.,

$$a_{\mathrm{EI}}(\mathbf{x}^*) = \int_{-\infty}^{y(\mathbf{x}^-)} [y(\mathbf{x}^-) - u] p(u)\, du, \tag{28}$$

where $\mathbf{x}^-$ is the current best observed candidate in $\mathcal{D}_{\mathrm{obs}}$ and the shorthand symbol, and the Gaussian-distributed variable $u = y^* \mid X, \mathbf{y}, \mathbf{x}^*, \boldsymbol{\theta}$ is given by Eq. (20). Equations (20) and (28) yield

$$\mathrm{EI}(\mathbf{x}^* \mid X, \mathbf{y}, \boldsymbol{\theta}) = [y(\mathbf{x}^-) - \mu(\mathbf{x}^*)]\Phi\left(\frac{y(\mathbf{x}^-) - \mu(\mathbf{x}^*)}{\sigma(\mathbf{x}^*)}\right) + \sigma(\mathbf{x}^*)\phi\left(\frac{y(\mathbf{x}^-) - \mu(\mathbf{x}^*)}{\sigma(\mathbf{x}^*)}\right), \tag{29}$$

where $\phi(x)$ and $\Phi(x)$ are the probability density function (PDF) and cumulative distribution function (CDF) of the standard normal distribution, respectively. For illustrative purposes, we show in Fig. 3 a one-dimensional example demonstrating the calculation of EI.

2. Multimodal optimization of integrated EI

Since we adopt a fully Bayesian treatment on hyperparameters, the prediction $\mathbf{y}$ is expressed as the expectation of $\mathbf{y}^*$ over the posterior distribution of $\boldsymbol{\theta}$, i.e., the weighted average of an ensemble of models,

$$\begin{aligned} p(\mathbf{y}^* \mid X, \mathbf{y}, X^*) &= \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{y}^*, \boldsymbol{\theta} \mid X, \mathbf{y}, X^*) \\ &= \int p(\mathbf{y}^* \mid X, \mathbf{y}, X^*, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid X, \mathbf{y})\, d\boldsymbol{\theta} \\ &= \sum_{i=1}^{n_s} p(\mathbf{y}^* \mid X, \mathbf{y}, X^*, \boldsymbol{\theta}_i), \end{aligned} \tag{30}$$

where n_s is the number of hyperparameter samples drawn using Bayesian MCMC. Similarly, Srinivas *et al.* [55] pointed out that EI should also be calculated as the expectation over the posterior distribution of hyperparameters,

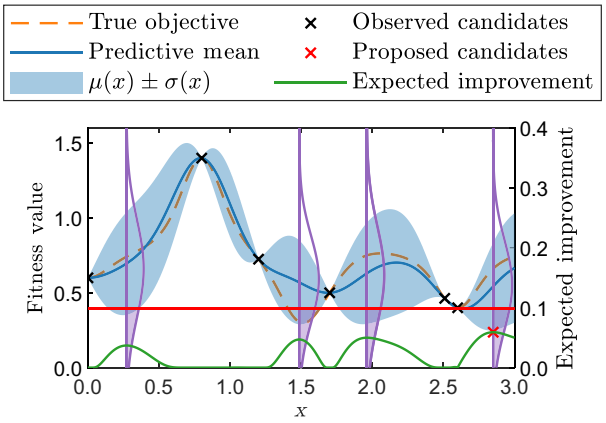

FIG. 3. Illustration of GPR (main axis) and EI (secondary axis) in a minimization problem. EI is high where the predictive value is low (exploitation) and where the uncertainty is high (exploration). The four peaks on the EI plot indicate the four local maxima of EI. How EI is calculated at the four locations is also displayed: the horizontal red line indicates the minimum value at the current iteration and the purple line indicates the probability density distribution of the predictive mean given $\mathbf{x}$, μ, and σ. The shaded purple region indicates the distribution of the possible improvement at that location, the integral of which is precisely EI.

i.e.,

$$\begin{aligned} a_{\mathrm{EI}}(\mathbf{x}^* \mid X, \mathbf{y}) &= \mathbb{E}_{\boldsymbol{\theta}}[a_{\mathrm{EI}}(\mathbf{x}^*, \boldsymbol{\theta} \mid X, \mathbf{y})] \\ &= \int a_{\mathrm{EI}}(\mathbf{x}^* \mid X, \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid X, \mathbf{y})\, d\boldsymbol{\theta} \\ &= \sum_{i=1}^{n_s} a_{\mathrm{EI}}(\mathbf{x}^* \mid X, \mathbf{y}, \boldsymbol{\theta}_i), \end{aligned} \tag{31}$$

which is also referred as integrated EI acquisition function (a_{iEI}). Mathematically, the maximization of the integrated EI can be written as

$$\text{maximize } a_{\mathrm{iEI}}(\mathbf{x}^* \mid X, \mathbf{y}), \text{ subject to } \mathbf{x}^* \in \mathcal{X}. \tag{32}$$

a. Partitioning the solution space. The expectation $a_{\mathrm{iEI}}(\mathbf{x}^* \mid X, \mathbf{y})$ incorporates an ensemble of regression models and its topography is rather complex and strongly multimodal, necessitating broader exploration in $\mathcal{X}$ so that the global optimal solution is covered. Ideally, we only need to identify the global maximizer of Eq. (31); however, in the topography of the noise-corrupted, observed objective function, a global solution may be obscured by multiple local solutions, such that not only the global optimal solution but also local optimal solutions need to be considered. Accordingly, we develop a subroutine for the proposal of the promising candidates for exact evaluation based on SSP; see Fig. 4.

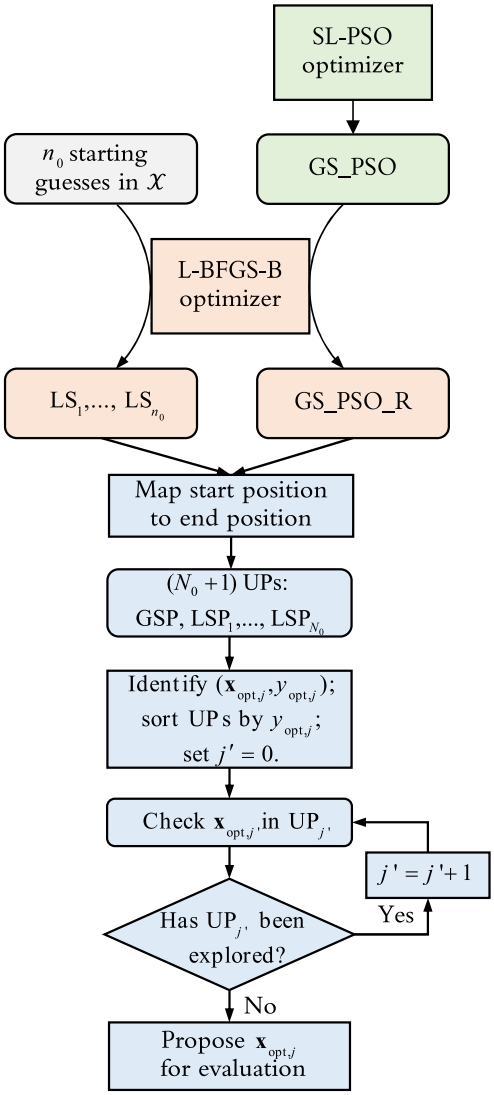

FIG. 4. Illustration of the subroutine searching for promising candidates by partitioning the solution space. GSP and LSP indicate the global solution partition and local solution partition, respectively, UP indicates the unique partition, _PSO indicates the solution is obtained using SL PSO, _R indicates the result is refined by L-BFGS-B.

Using the L-BFGS-B optimization algorithm and n_0 candidates in the solution space $\mathcal{X}$ as starting guesses, local optimal solutions, denoted as LS_i, $i = 1, 2, \ldots, n_0$, are identified; the global solution found by SL PSO but then refined by L-BFGS-B is denoted as GS_PSO_R.

Since we use a gradient-based optimizer, some starting guesses may end up with the same local optimizer of Eq. (32). By mapping the start position and end position of the candidates before and after the L-BFGS-B optimization algorithm is applied, we can partition the solution space $\mathcal{X}$ into various UPs, comprising a global solution partition and N_0 local solution partitions ($N_0 < n_0$). There is a local maximizer of integrated EI for each UP, referred to as $\mathbf{x}_{\text{opt}}$, altogether forming the proposal set

$$\mathcal{D}_{\text{prop}} = \{(\mathbf{x}_{\text{opt},j}, y_{\text{opt},j}) \mid j = 1, 2, \ldots, N_0 + 1\}, \tag{33}$$

where $y_{\text{opt},j}$ is the local optimum value for $\mathbf{x}_{\text{opt},j}$. The set $\mathcal{D}_{\text{prop}}$ is then sorted in descending order of $y_{\text{opt},j}$. Enumerating from the top of the list, we search for the first $\mathbf{x}_{\text{opt},j}$ that satisfies the condition that, for all $\mathbf{x}_i \in \mathcal{D}_{\text{obs}}$, $i = 1, 2, \ldots, n$,

$$\mathbb{E}_{\boldsymbol{\theta}}[\|\Lambda^{-1}(\mathbf{x}_{\text{opt},j} - \mathbf{x}_i)\|_2] \geq d_0, \tag{34}$$

where d_0 is a predefined distance threshold. Equation (34) is indeed a safe guard to prevent the proposal of a candidate that is too close to the previously evaluated candidates. Employing the L-BFGS-B optimization and SL-PSO algorithms together will guarantee that both the global optimal solution and major local optimal solutions are covered; see the detailed explanation given below.

b. The L-BGFS-B optimization algorithm. The L-BFGS-B is a quasi-Newton method of which the inverse Hessian matrix of the objective function is updated using the first-derivative information. We denote the global maximizer and multiple local maximizers of integrated EI obtained using L-BFGS-B as

$$\mathbf{x}_{\text{BFGS}} = \underset{\mathbf{x}^* \in \mathcal{X}}{\text{argmax}}\, a_{\text{iEI}}(\mathbf{x}^* | X, \mathbf{y}). \tag{35}$$

L-BFGS-B is adopted for the following reasons [56].

(a) It maintains the inverse Hessian implicitly using the first-derivative information as several vector pairs so that the Hessian vector product can be performed as inner products and vector summation, making it very efficient in terms of both storage and computation [22]. Computation of the first derivative is described below.

(b) Instead of saving the full $n \times n$ inverse Hessian approximation, it saves only the several most recent vectors in the interest of saving storage and speeding up computation while maintaining a good rate of convergence.

One of the practical challenges is the computation of derivatives of a_{EI} with respect to $\mathbf{x}^*$. Using the chain rule, the derivative of $a_{\text{EI}}[\mu(\mathbf{x}^*), \sigma(\mathbf{x}^*)]$ with respect to $\mathbf{x}^*$ is calculated as

$$\frac{da_{\text{EI}}}{d\mathbf{x}^*} = \frac{\partial a_{\text{EI}}}{\partial \mu}\frac{d\mu}{d\mathbf{x}^*} + \frac{\partial a_{\text{EI}}}{\partial \sigma}\frac{d\sigma}{d\mathbf{x}^*}, \tag{36}$$

where the partial derivatives of a_{EI} with respect to $\mu(\mathbf{x}^*)$ and $\sigma(\mathbf{x}^*)$ can be calculated using Eq. (29) as

$$\frac{\partial a_{\text{EI}}(\mu, \sigma)}{\partial \mu} = -\Phi(z_0), \tag{37}$$

$$\frac{\partial a_{\mathrm{EI}}(\mu,\sigma)}{\partial\sigma} = \phi(z_0), \tag{38}$$

with

$$z_0 = \frac{f_{\min} - \mu}{\sigma}. \tag{39}$$

The derivatives of μ and σ with respect to $\mathbf{x}^*$ can be derived using Eqs. (19) and (20):

$$\frac{d\mu}{\partial\mathbf{x}^*} = 2\Lambda^{-1}(X - \mathbf{x}^* \cdot \mathbf{1})\{K_f(X,\mathbf{x}^*) \odot [K_n(X,X)^{-1}(\mathbf{y} - \mathbf{m_0})]\}, \tag{40}$$

$$\frac{d\sigma}{\partial\mathbf{x}^*} = -2\Lambda^{-1}(X - \mathbf{x}^* \cdot \mathbf{1})\{K_f(X,\mathbf{x}^*) \odot [K_n(X,X)^{-1}K_f(X,\mathbf{x}^*)]\}. \tag{41}$$

A detailed description of the overall procedure of L-BFGS-B comprising determination of the active set and subspace maximization can be found in Refs. [56,57].

c. The social-learning particle swarm optimization. Originally proposed by Kennedy, the particle swarm optimization is a population-based evolutionary optimization technique that minimizes the objective function by iteratively improving candidate solutions [58]. The SL-PSO algorithm, in which particles learn and imitate the behaviors of the better individuals in the whole population instead of learning from a single best particle, possesses the following advantages.

(a) Early convergence to local optima is resolved with exploration ability improved.

(b) A sophisticated dimensional-dependent control of swarm size, learning rate, etc. makes this algorithm scalable to both low-dimensional and high-dimensional problems.

Similarly, we denote the global maximizer of integrated EI obtained using SL PSO as

$$\mathbf{x}_{\mathrm{SL\ PSO}} = \underset{\mathbf{x}^*\in\mathcal{X}}{\operatorname{argmax}}\, a_{\mathrm{iEI}}\left(\mathbf{x}^*|X,\mathbf{y}\right). \tag{42}$$

Description of the SL-PSO algorithm, as well as the detailed implementation such as termination, scalability, the trade-off between exploration and exploitation, can be found in Ref. [21].

The progression of BO for the multimodal objective function over two iterations is illustrated in Fig. 5 for an example problem with a single unknown parameter. BO starts with six observed candidates. At each iteration, a_{EI} is maximized to propose the most promising candidate. The GP proxy is then updated, getting closer to the true but unknown objective function, and the whole process is repeated. Note that the proposed candidate at step 2, also as the maximizer of a_{EI}, is extremely close to the global minimum of the true but unknown objective function.

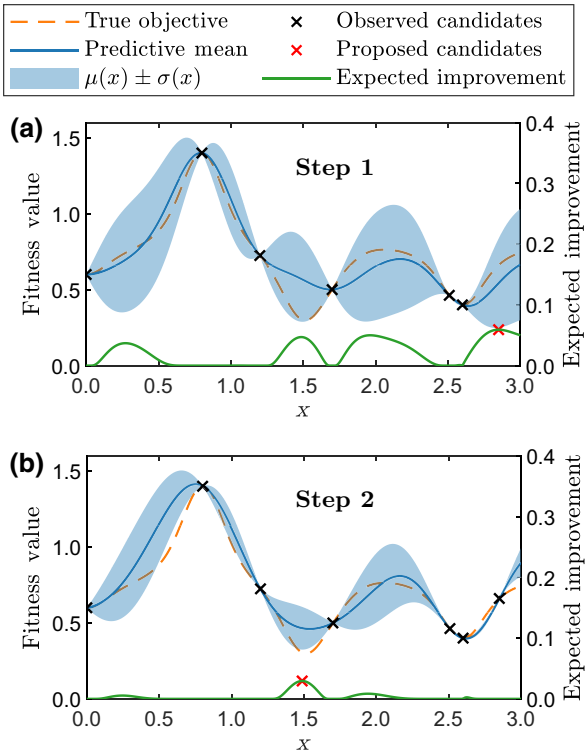

FIG. 5. Illustration of two consecutive steps using the ISW to find global and local minima of the designed one-dimensional objective function. GP mean and GP variance are calculated using Eq. (18). At each step a new candidate is proposed by maximization of EI (green line, right axis).

C. Solution analysis

In the ISW, new candidates are proposed and evaluated sequentially until the optimization budget is depleted. As mentioned in Sec. I, in the topography of the noise-corrupted, observed objective function a global solution may be obscured by multiple local solutions since relaxation decays $\mathbf{m}_{t,\mathrm{exp}}$ and corresponding distributions $\mathbf{s}_{\mathrm{exp}}(T_2)$ are intrinsically subject to noise. As a result, we create a pool of candidates to cover most feasible solutions, comprising a global optimal solution and major local optimal solutions, and then check for their physical plausibility. The multimodal GPR proxy for the objective function is analytically calculated with all UPs identified using SSP, with all candidates clustered to the nearest UP by checking which UP they end up with.

Figure 6 demonstrates the clustering results following successful identification of the global minimum; see

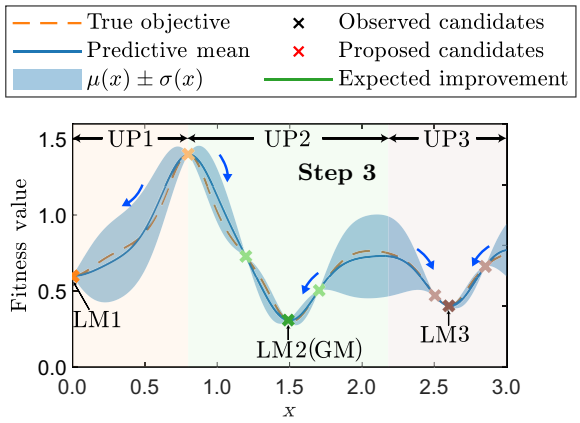

FIG. 6. In step 3, three UPs and the corresponding minima, i.e., local minimum 1 (LM 1) in UP 1 (LSP 1), global minimum (GM) in UP 2 (GSP), and local minimum 3 (LM 3) in UP 3 (LSP 2), colored orange, green, and brown, respectively, are identified using SSP. Of all observations, LMs are colored in denser colors. LM 2 will be proposed for evaluation in the next iteration. Hyperparameters and the GPR proxy are updated whenever a new candidate is evaluated.

Fig. 5(b). Eight candidates are clustered into three UPs. In this toy example, the ISW identifies all three UPs and three local optima (in this case, three local minima, or three LMs) using only four additional FEs. After clusterization, all LMs are checked against criteria of physical validity, and solutions that do not pass the plausibility test are removed.

IV. RESULTS FOR THE INVERSE PROBLEM

A. Determination of three physical quantities: ρ_q, $T_{2e,c}$, and $D_{e,c}$ at SNR = 100

Consider the optimization problem of NMR T_2 relaxation in Bentheimer sandstone with three unknown physical parameters ($D = 3$): the discrete surface relaxivity of quartz ρ_q, the effective transverse relaxation time $T_{2e,c}$, and the *effective* diffusion coefficient $D_{e,c}$ in clay regions. Since the latter two decision variables usually vary across many orders of magnitudes, we take the base-10 logarithm of T_{2e} and D_e as $\log_{10}(T_{2e,c}/\mathrm{s})$ and $\log_{10}[D_{e,c}/(\mu\mathrm{m}^2/\mathrm{s})]$. The search for feasible solutions is constrained by the solution space $\mathcal{X}$ specified in Table I. The solution space $\mathcal{X}$ is rather large and even a coarse discretization leads to $17 \times 11 \times 13 = 2431$ expensive NMR simulations. The budget of FEs is set to $40D = 120$, with the initial $4D = 12$ FEs spent on random sampling of the parameter space.

TABLE I. The three physical parameters of interest and the discrete three-dimensional solution space $\mathcal{X}$ resulting in 2431 combinations of physical parameters. The discretization is also used in Sec. V.

Parameters	Abbreviation	Min	Max	Discretization
ρ_{quartz} $/(\mu\text{m}^2/\text{s})$	ρ_q	0	16	0,1,...,16
$\log_{10}$ $(T_{2e,c}/\text{s})$	$\log_{10} T_{2e,c}$	-3	0	$-3.0, -2.7, \ldots, 0.0$
$\log_{10}[D_{e,c}/$ $(\mu\text{m}^2/\text{s})]$	$\log_{10} D_{e,c}$	-6.28	-4	$-6.4, -6.2, \ldots, -4.0$

1. Progression of optimization

Figure 7 illustrates the selection of the most promising candidate for evaluation at the 67th and 114th steps, balancing both exploration and exploitation. In each plot there are five isosurfaces, corresponding to the minimum (or maximum, speaking of integrated EI) 0.01%, 0.1%, 1%, 3%, and 5% of the values of that scalar field. The correlations among the three physical parameters are plotted in the ρ_q-$\log_{10} T_{2e,c}$, ρ_q-$\log_{10} D_{e,c}$, and $\log_{10} T_{2e,c}$-$\log_{10} D_{e,c}$ planes. The GP mean depicts the locations of UPs and the identified LMs together with other feasible solutions at that step. In the GP-mean plot, the long spindle-shaped correlations between each pair of parameter highlight a region of solutions providing a reasonably good fit. For better demonstration, for GP-mean, GP-variance plots, only the minimum 5% are shown on the isosurface plot, the correlation plot, and the colormap; for the integrated EI plot, only the maximum 5% are shown. Practically, good solutions are usually found within the region defined by the top 0.1% fitness value. Moreover, correlations between the ρ_q-$\log_{10} T_{2e,c}$ and ρ_q-$\log_{10} D_{e,c}$ pairs are positive, while the correlation between the $\log_{10} T_{2e,c}$-$\log_{10} D_{e,c}$ pair is negative. As expected, from the GP-variance plot we observe that frequently sampled domain regions exhibit smaller variance as compared to regions that are seldom sampled; the unsampled region with larger uncertainties is worthwhile exploring.

At step 67, from the integrated EI plot we observe that the proposed candidate $(5.774, -2.250, 5.316)$, the global maximizer of the integrated EI at the 67th step, turns out to be the global minimizer of the GP mean since both low GP mean and high GP variance are found in this region. Moreover, this candidate is found to be the solution with the lowest fitness value of all 120 evaluations.

At step 114, we observe from the GP-mean plot that the region in which good solutions are expected to be found narrowed; meanwhile, at this step the single feasible region shown in the GP mean and integrated EI of Fig. 7 starts to split into two narrower regions, respectively, revealing the multimodal nature of the topography of the true but unknown objective function; this is also clear from the correlation plot in the ρ_q-$\log_{10} D_{e,c}$ plane. Indeed, the split of the feasible region containing good solutions is more obvious at step 120, as will be displayed later. The most promising candidate determined at this step, i.e., $(6.032, -2.807, -4.796)$, is the solution

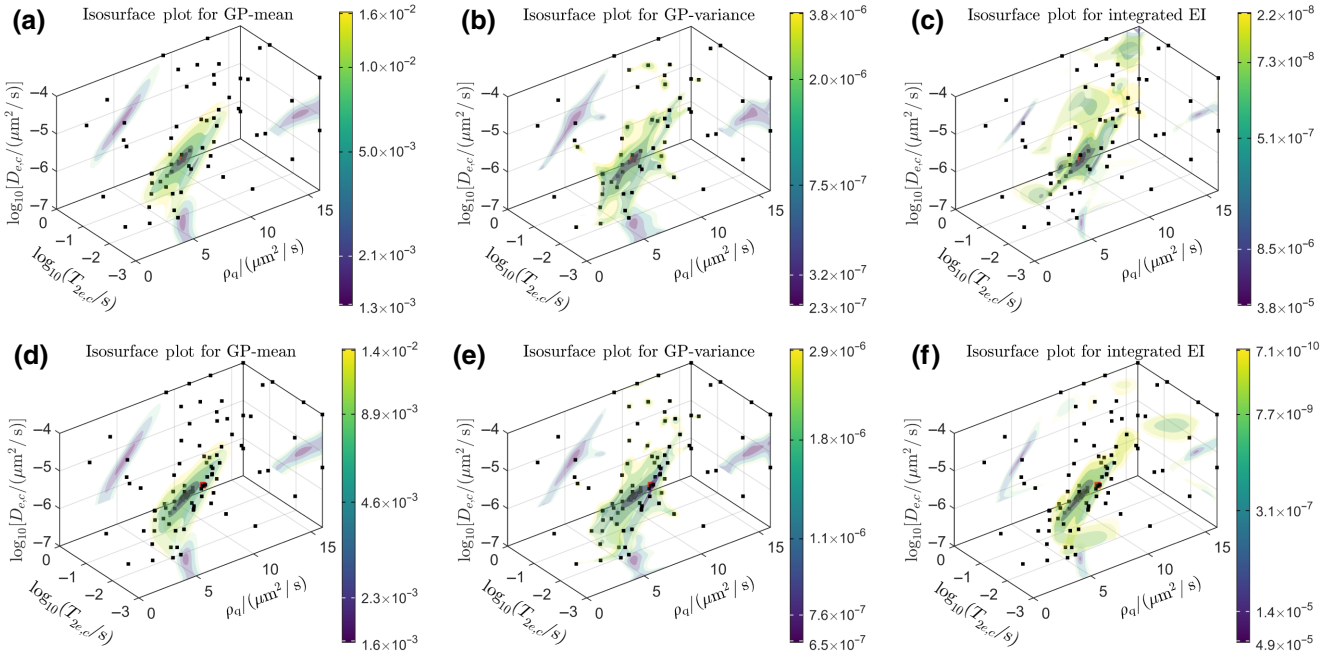

FIG. 7. Isosurfaces of the parameter space. (a) predictive GP mean, step 67; (b) predictive GP variance, step 67; (c) GP integrated EI, step 67; (d) predictive GP mean, step 114; (e) predictive GP variance, step 114; (f) GP integrated EI, step 114. Five levels of isosurfaces are displayed, corresponding to the minimum (mean and variance) or maximum (integrated EI) 0.01%, 0.1%, 1%, 3%, and 5% of the values of that scalar field. The absolute values corresponding to these isosurface levels are shown on the colorbar. The correlations among physical parameters evaluated at the best candidate are shown in the ρ_q-$\log_{10} T_{2e,c}$, ρ_q-$\log_{10} D_{e,c}$, and $\log_{10} T_{2e,c}$-$\log_{10} D_{e,c}$ planes. The contour plot and the isosurface plot share the same colorbar.

with the fourth lowest fitness value, as will be shown later.

Figure 8 displays the trace of the fitness value and minimum fitness value on the log scale. Huge fluctuations in the fitness value suggest that during the search process the optimization algorithm switches between exploration and exploitation. The minimum fitness value decreases significantly by two orders of magnitude during the first 70 FEs. Over the next 50 FEs there is little reduction in the minimum fitness value; however, the fitness value suggests that more candidates explored the promising region, indicating that the next evaluation step is increasingly wel l guided.

Figure 9 compares the posterior distribution of hyperparameters and the correlations among hyperparameters at the 67th and 114th steps. It is clear that inference of σ_n^2 is much more uncertain than other hyperparameters since inference of all hyperparameters is similar when comparing the two steps except for σ_n^2 whose PDF still undergoes big changes. At step 114 all inferred hyperparameters are approximately Gaussian distributed with new mean and standard deviation values compared with step 67. The search process in the later stage becomes well guided given more accurate inference on hyperparameters due to more FEs involved. Results at the final step—step 120—will be discussed in the following subsection.

2. Optimization results

We begin with the analysis of the inference on hyperparameters. Figure 10 shows the marginal distribution and the Gaussian fit for the six hyperparameters inferred at the final optimization step, i.e., step 120. All hyperparameters θ follow Gaussian distributions very well, indicating that variables θ_i' follow base-10 lognormal distributions,

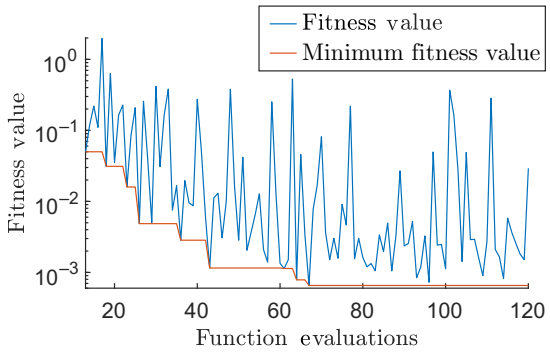

FIG. 8. Trace of the fitness value and minimum fitness value for the decay of interest at SNR = 100. Candidates with low fitness values are identified at step 67 (rankings first) and step 114 (rankings fourth). The horizontal axis starts at 13 since the first 12 candidates are preevaluated without using the ISW to avoid a cold start.

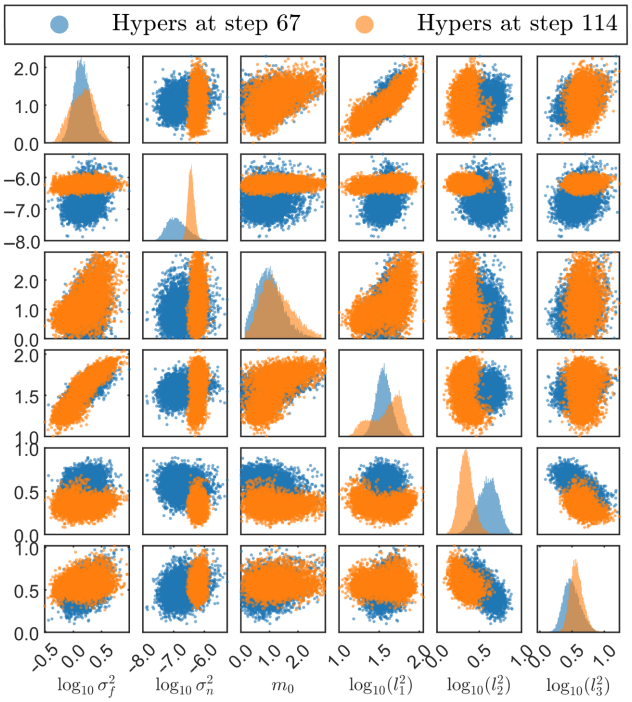

FIG. 9. The comparison of the correlation plots between each pair of hyperparameters at the 67th and 114th steps. The panels along the diagonal show histograms of 20 000 samples for the six hyperparameters $\boldsymbol{\theta}$ drawn using Bayesian MCMC after a burn-in period, and the off-diagonal panels display correlations between each pair of hyperparameters.

i.e.,

$$\theta_i' = 10^{\mu_{\theta_i}+\sigma_{\theta_i}Z}, \qquad i = 1, 2, D+1, \ldots, D+3, \tag{43}$$

where Z is the standard normal variable, such that the expectation value of the hyperparameters $\mathbb{E}(\theta_i')$ can be calculated using the equation

$$\mathbb{E}(\theta_i') = e^{\mu_{\theta_i}\ln(10)+\sigma_{\theta_i}^2[\ln(10)]^2/2}, \tag{44}$$

where μ_{θ_i} and σ_{θ_i} are the fitted mean and standard deviation of the variable's base-10 logarithm.

In Table II, the $\mathbb{E}(\theta')$ column demonstrates different sensitivities of the fitness value against three physical quantities: in terms of variability in the fitness value, a change of 7.142 in ρ_q would be equivalent to a change of 1.473 in $\log_{10} T_{2e,c}$ or a change of 1.850 in $\log_{10} D_{e,c}$. This relationship is due to the application of ARD, as discussed in Sec. III. The pairwise correlation is readily identified by the isosurface plots given in Figs. 7 and 11. In consideration of Bayesian optimization stronger correlations lead to faster and more confident inference of hyperparameters, which makes the proposal of new candidates increasingly more effective. In consideration of physics, the correlation and underlying length scales derived using ARD (Table II) provide a confidence region within which the identified values are considered valid.

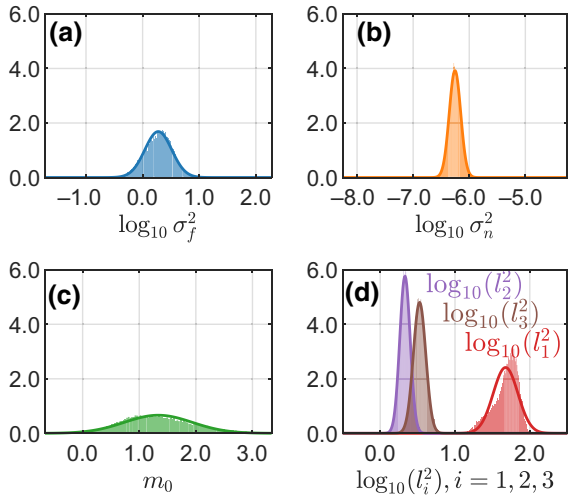

FIG. 10. Normal distribution fitted to the marginal distribution of all six hyperparameters inferred at step 120. The three length scales are plotted together for easier comparison of the kurtosis. Either horizontal axes or vertical axes are displayed in the same scale and the histograms are normalized so that the integral of the bar area is 1. The mean, standard deviation (SD), and expectation for each hyperparameter are given in Table II.

Although correlations between physical quantities are important, we do not need to manually feed such information to the ISW. The ISW tends to recover the correlation itself by constantly sampling and evaluating candidates in the solution domain to improve inference of the hyperparameters. As an example, we can see from Fig. 7 that the correlation gradually crystallizes from step 67 to step 114. By contrast, in Sec. V we show that the ISW fed with wrong correlation information completely failed to locate both local and global minima.

By using SSP (see Fig. 4), we identified two UPs in the given parameter space, as shown in Fig. 11. Although the two UPs appear very close to each other, they are independent UPs and cannot be merged. From each UP we propose the top three candidates; see Fig. 11 and Table III.

TABLE II. Details of normal distributions fitted to six hyperparameters corresponding to Fig. 10. All hyperparameters θ are inferred Gaussian distributed.

θ	mean(θ)	SD(θ)	θ'	$\mathbb{E}(\theta')$
$\log_{10}(\sigma_f^2)$	0.277	0.237	σ_f^2	2.195
$\log_{10}(\sigma_n^2)$	−6.254	0.102	σ_n^2	5.723×10^{-7}
m_0	1.336	0.601	m_0	1.336
$\log_{10}(l_1^2)$	1.676	0.165	l_1^2	7.142^2
$\log_{10}(l_2^2)$	0.331	0.069	l_2^2	1.473^2
$\log_{10}(l_3^2)$	0.526	0.083	l_3^2	1.850^2

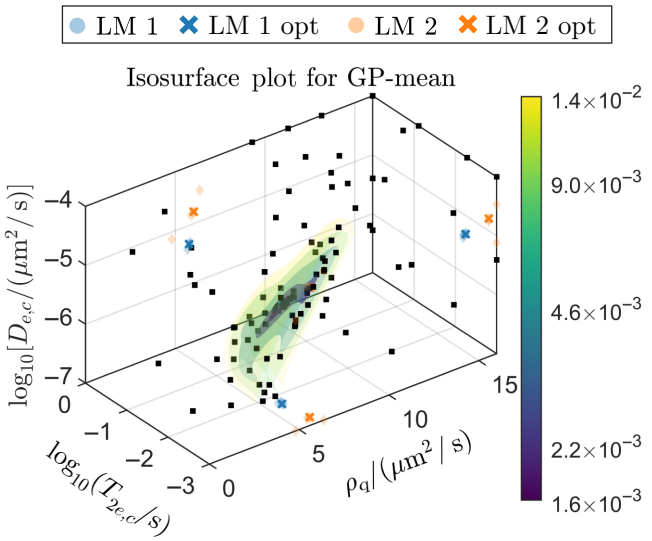

FIG. 11. Distribution of the two identified UPs at step 120 with the top three solutions from each UP. Five isosurface levels are displayed, corresponding to the minimum 0.01%, 0.1%, 1%, 3%, and 5% of the GP mean. Colored circles and crosses shown in the ρ-T_2, ρ-D_e, and T_2-D_e planes are projections of the candidates. Details of plotted candidates are shown in Table III.

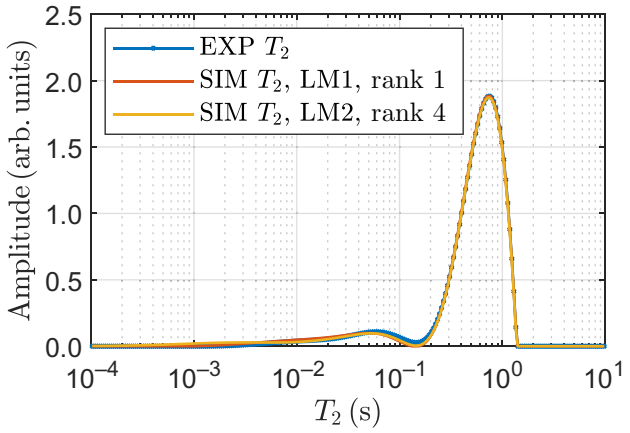

FIG. 12. The comparison of fit for the two LMs from the two UPs (see also Table III).

In each UP, the top three candidates with their projections on the three planes are noted as circles, with LMs plotted as crosses. As we have discussed before that the single UP splits into two UPs from step 114, it becomes more obvious at step 120. Note that ρ_q is the identified discrete surface relaxivity of quartz; actual continuum values would be a factor A larger, as discussed previously.

The corresponding T_2 distributions are shown in Fig. 12. While both predicted distributions are closely matching a reference, the application of physical constraints is required to determine the feasibility of each UP and each solution. Mathematical and physical validity will be checked against each solution listed in Table III.

TABLE III. The top three candidates in terms of fitness value (FV) within each UP, identified using the ISW. Listed candidates are displayed in Fig. 11. UPs are sorted in ascending order or values of LMs. Numbers in the *Rank* column indicate the rankings of the solutions out of 120 evaluated candidates in terms of their fitness values, whereas numbers in the *Step* column indicate at which step those solutions are proposed. The T_2 distributions corresponding to both LMs are shown in Fig. 12. Note that ρ_q is the lattice-based discrete surface relaxivity of quartz.

UP	Rank	ρ_q	$\log_{10} T_{2e,c}$	$\log_{10} D_{e,c}$	FV	Step
1	1	5.774	−2.250	−5.316	6.547×10^{-4}	67
	2	5.836	−2.193	−5.346	7.259×10^{-4}	96
	3	5.832	−2.261	−5.291	7.854×10^{-4}	64
2	4	6.032	−2.807	−4.796	8.209×10^{-4}	114
	20	5.892	−2.807	−4.842	1.515×10^{-3}	119
	28	4.805	−3.000	−5.123	2.052×10^{-3}	52

(a) *The active boundaries.* Solutions reaching the lower bound of the given solution space are indeed less practical as those values could be even more unrealistic if bounds are removed, not to mention we deliberately designed a large solution domain to overcome violation of the boundary. As a result, the solution at step 52 is excluded for subsequent assessment.

(b) *The diffusion coefficient.* The self-diffusion coefficient of water at 25 °C is 2.3×10^{-5} $\mu m^2/s^{-1}$), which corresponds to the base-10 logarithm −4.6021. As a result, physically valid candidates should satisfy $\log_{10} D_{e,c} < -4.6021$. No violation of this item is identified.

The optimality of UP 1 and UP 2 is considered in the following context.

(a) *The (mean) fitness value and rankings*. Theoretically, the solution with the lowest fitness value is the optimum solution in this minimization problem. Under such considerations, UP 1 is superior to UP 2 since both fitness values and rankings of the proposed solutions in UP 1 are better than solutions in UP 2. Meanwhile, out of the top 20 candidates only two are from UP 2.

(b) *The shape of the T_2 distribution*. The shape of the T_2 distribution is also a critical criterion in the identification of the optimal solution. Taking the 67th (best solution in UP 1) and the 114th (best solution in UP 2) for example, the 114th solution exhibits greater amplitude in the very short-time component, which is impractical as it displays features that the reference experimental T_2 distribution does not possess. Indeed, such a misfit is due to nonuniform sensitivity against the 256 T_2 bins: since we are minimizing the T_2 distribution difference in the least square

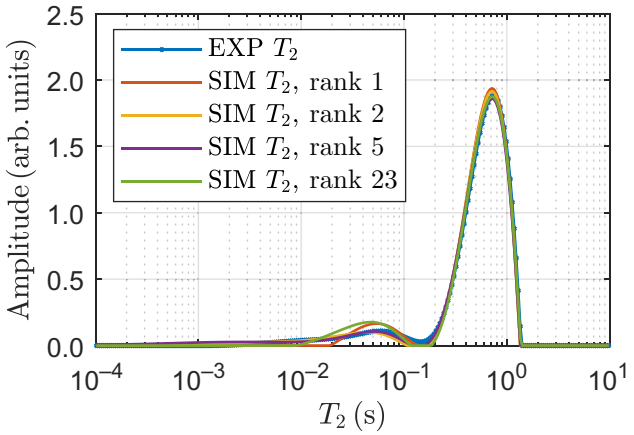

FIG. 13. The comparison of fit of the four LMs out of 2431 candidates (see also Table IV).

sense, not all differences are equally captured, e.g., the fitness value is more sensitive to a 10% change in the T_2 main (long-time component) peak than the T_2 minor (short-time component) peak.

After comparison of best candidates to the complete set of mathematical and physical criteria and constraints, candidate 67 is found to be the best overall solution. In situations when multiple UPs have solutions with similar rankings and pass plausibility tests, identification of the best solution requires a new round of optimization in a solution space further constrained by the domain covered by plausible UPs. Given the inferred values of hyperparameters, more informative prior distributions for hyperparameters can be used to properly guide the formulation of GPR, thus making the subsequent search more efficient. Comparative results of the noninformative priors, informative priors, and wrong priors for hyperparameters in terms of the global and local search ability can be found in Sec. V.

3. Comparison with grid search

To further understand the global and local search ability of the ISW, we performed an exhaustive search in the solution space $\mathcal{X}$ using grid search, with the discretization of $\mathcal{X}$ specified in Table I. Inversion of all simulated decays arising from exact evaluation of 2431 combinations of physical parameters is performed using the same settings. Four feasible UPs and LMs identified by performing GPR on all 2431 candidates on the grid are displayed in Table IV and Fig. 13. Comparing fitness values of the candidates in Tables IV and III we conclude that by using the ISW we are able to identify solutions with much lower fitness value but with significantly less computation resources.

By inference of hyperparameters using maximization of marginal likelihood, we find that the "ground truth" for the noise hyperparameter is $\sigma_n^2 = 3.223 \times 10^{-7}$. As a result, we confirm that the noise variance σ_n^2 inferred during optimization using the ISW is also inferred correctly, avoiding overfit or underfit regressions, an issue that will be discussed in Sec. V.

TABLE IV. Four feasible UPs and LMs identified by performing GPR on all 2431 candidates on the grid. The T_2 distributions corresponding to all four LMs are shown in Fig. 13. Note that ρ_q is the lattice-based discrete surface relaxivity of quartz.

UP	Rank	ρ_q	$\log_{10} T_{2e,c}$	$\log_{10} D_{e,c}$	FV
1	1	7	−1.5	−5.8	1.728×10^{-3}
2	2	7	−2.4	−5.0	1.846×10^{-3}
3	5	6	−3.0	−4.8	2.193×10^{-3}
4	23	5	−1.8	−6.2	4.136×10^{-3}

4. CPU time consumption

All simulations and optimizations have been carried out on the research supercomputer Gadi of the National Computational Infrastructure (Canberra, Australia) built on Intel Xeon Platinum 8274 3.20 GHz CPUs. The NMR T_2-CPMG simulation code is implemented in FORTRAN 95, and the ISW code is implemented in MATLAB® 2019b.

The NMR forward simulation is conducted on a 1000^3 domain with voxel size $\epsilon = 2.88$ μm and subgrid resolution $\epsilon_w = \epsilon/l_s$, paralleled on eight cores using OpenMPI with 5000 random walkers per core. Under such settings, the wall time of the simulation depends on the value of $D_{e,c}$ and typically takes between 10 to 15 min. For three-dimensional problems where we used 120 FEs, the ISW is expected to complete in 34 ± 2 h.

During the optimization using the ISW, more than 90% of the computational resources are consumed in the evaluation of expensive NMR simulations. From the remaining 10% required for the "next candidate" search, about half is spent on drawing samples of hyperparameters $\boldsymbol{\theta}$ using Bayesian MCMC and another half spent on the maximization of integrated EI using L-BFGS-B. Since both computations require Cholesky factorization of the covariance matrix K_n, of which the computational complexity is $O(N^3)$, where N is the ever-growing number of completed evaluations, we ease the burden of maximization of $a_{\mathrm{EI}}(\mathbf{x}^* \mid X, \mathbf{y})$ by precalculating and storing the factorization of K_n within each iteration.

B. Determination of three physical quantities at seven different SNRs

1. Determining λ_{opt} and the effect of noise realizations

As we established in Sec. II discussing the ILT procedure, λ_{opt} trades off stability (smoothness) and informativeness (resolution). The SNR-dependent value of λ_{opt} is crucial to the stability and correctness of the identified parameter values. Fortunately, values of λ_{opt} can be

TABLE V. Summary of λ_{opt} for seven log-spaced SNRs.

Number of scans	SNR	Acquisition time (min)	Number of decays	λ_{opt}
4	50	12	16	51.20
8	71	24	16	26.12
16	100	48	16	12.80
32	141	96	8	6.54
64	200	192	4	3.20
128	282	384	2	1.64
256	400	768	1	0.80

correctly determined if T_2 decays of the same sample over a range of SNRs are measured.

To determine λ_{opt} for the low-SNR decay, we consistently decrease λ until the monitored misfit between T_2 distributions of low SNR and high SNR becomes non-negligible. Table V displays λ_{opt}, in which λ_{opt} monotonically decreases with the SNR since ill-posedness of solving Eq. (10) decreases with increasing SNR, which also coincides with our belief that decays acquired at higher SNRs are more reliable and trustworthy.

Using this rule, T_2 distributions inverted from decays measured at seven SNRs are depicted in Fig. 14. For each SNR group, the mean value and standard deviation averaged over 2 to 16 distributions are displayed. Indeed, the confidence interval of the short-time T_2 component is

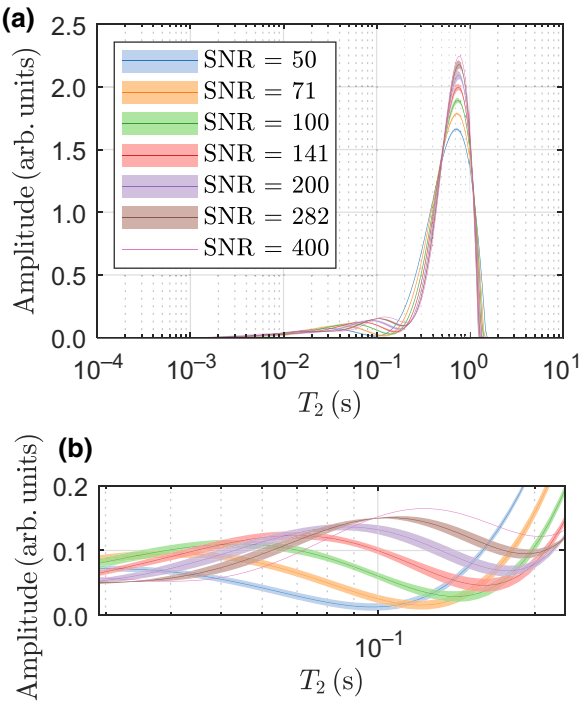

FIG. 14. Illustration of (a) the mean T_2 distributions at seven SNRs and (b) an enlarged view of the short-time peak. Within each SNR group, the solid line and shaded area represent the mean value and confidence interval, respectively, averaged over 2 to 16 distributions, indicated in Table V.

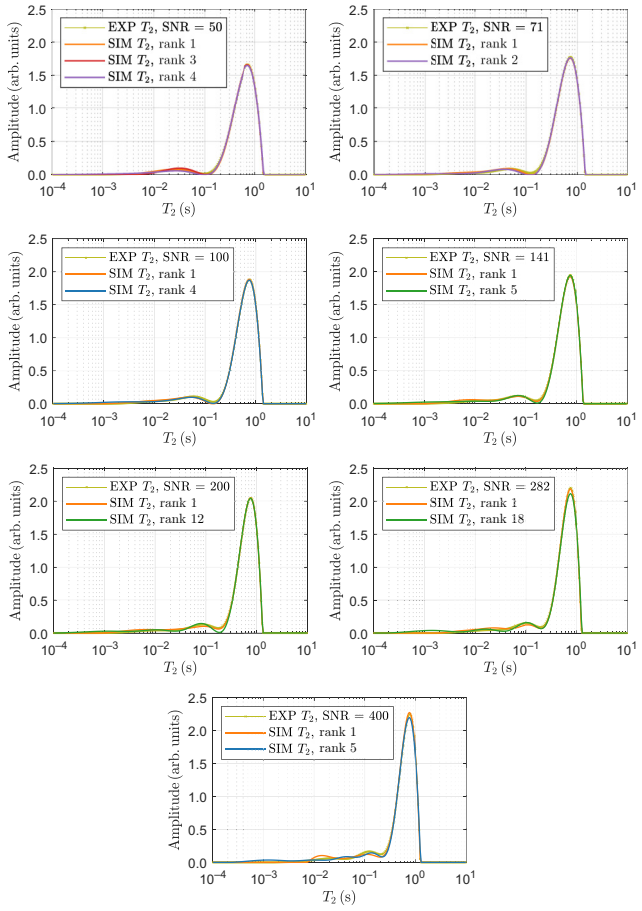

FIG. 15. Illustration of the representative fits in each SNR group. Simulated T_2 distributions are colored according to which generalized unique partition (GUP) they belong to. Fitness values and identified parameter values corresponding to each T_2 distribution are listed in Table VI.

around 0.5% of the maximum amplitude of the $\mathbf{s}(T_2)$ mean. Owing to such variations, a good fit to one decay is usually not a good fit to the other, so that identified parameter values are not only dependent on the SNR but also realizations of noise. As a result, there are still some uncertainties in the determination of the physical quantities from the optimization result of a single decay, while averaging results from multiple decays will help to reduce such uncertainties.

2. The effect of the SNR

Consider the effect of the SNR of the relaxation data on the parameter identification using the ISW. We minimize the effect of noise by reporting the mean identified parameter values. For each SNR, the number of repetitions and configurations of λ_{opt} are shown in Table V. For each SNR group, we show the optimization result for one decay, demonstrating all LMs and the corresponding physical parameter values in Fig. 15 and Table VI, respectively.

TABLE VI. The fitness values and identified physical parameter values corresponding to the fittings in Fig. 15.

SNR	GUP	R	FV	p1[a]	p2[a]	p3[a]	Ind
50	GUP 3	1	5.428×10^{-4}	6.510	−2.060	−5.393	68
	GUP 2	3	6.611×10^{-4}	6.647	−1.698	−5.749	99
	GUP 6	4	6.685×10^{-4}	6.430	−3.000	−4.537	77
71	GUP 3	1	8.828×10^{-4}	5.683	−2.468	−5.264	89
	GUP 6	2	1.075×10^{-3}	6.042	−3.000	−4.580	117
100	GUP 3	1	6.547×10^{-4}	5.773	−2.250	−5.3164	67
	GUP 5	4	8.209×10^{-4}	6.032	−2.807	−4.7960	114
141	GUP 3	1	7.880×10^{-4}	4.829	−2.398	−5.500	101
	GUP 4	5	1.201×10^{-3}	4.961	−3.000	−4.937	60
200	GUP 3	1	8.839×10^{-4}	4.928	−2.484	−5.394	110
	GUP 4	12	2.055×10^{-3}	5.063	−3.000	−4.885	84
282	GUP 3	1	1.345×10^{-3}	6.419	−2.254	−5.202	72
	GUP 4	18	3.748×10^{-3}	5.400	−3.000	−4.870	87
400	GUP 3	1	2.576×10^{-3}	6.264	−1.977	−5.426	108
	GUP 5	5	3.836×10^{-3}	5.415	−3.000	−4.799	76

[a]Parameter 1 (p1), ρ_q; p2, $\log_{10} T_{2e,c}$; p3, $\log_{10} D_{e,c}$.

We illustrate the resultant match of decays at SNR 50 and SNR 400 on both the linear scale and log-t scale in Fig. 16.

For each of the 63 decays, 120 candidates are clustered into multiple UPs using L-BFGS-B, resulting in 185 distinct UPs. UPs with LM rankings below 20 are excluded from the subsequent evaluation, leaving 139 UPs across seven SNR groups. Typically, the ISW identifies 2–4 UPs for each decay, which is advantageous to balance mathematically sound and physically plausible solutions. Using the linear correlation between each pair of physical parameters (Fig. 11), the 139 UPs are categorized as six GUPs in terms of the value of $\log_{10} D_{e,c}$, as shown in Fig. 17 and Table VII. GUP 1 is excluded for analysis due to its extremely low occurrence. GUP 6 is also excluded since its $\log_{10} D_{e,c}$ values are too high to be physically valid.

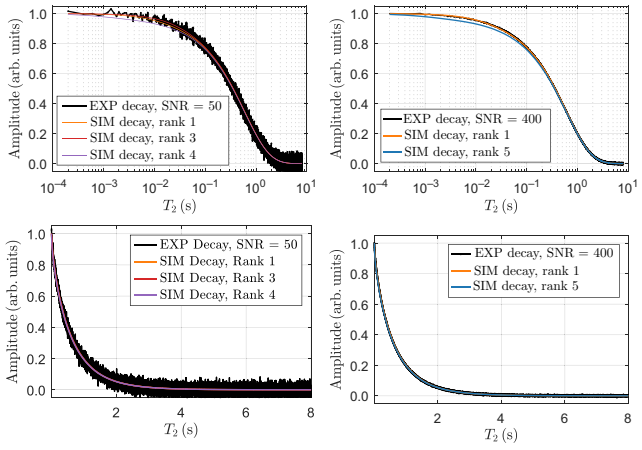

FIG. 16. Illustration of the decays on both the linear scale and log scale for low-SNR (50) and high-SNR (400) groups. Decays are colored according to which GUP they belong to.

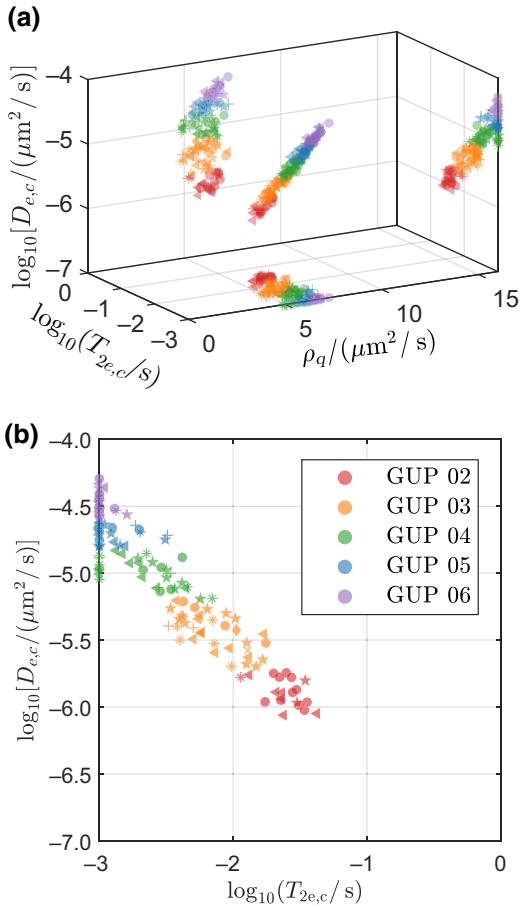

FIG. 17. (a) Illustration of the spatial distribution of the 139 UPs and LMs identified from 63 decays followed by categorization of all UPs into six GUPs. Potential correlations between pairs of physical parameters are shown as projections. (b) The strongest correlation is found between relaxation and diffusion in the clay effective phase. Categorization of the six GUPs is based on the value of $\log_{10} D_{e,c}$, which can be found in Table VII. Here SNR = 50 (•), 71 (⋆), 100 (∗), 141 (+), 200 (×), 282 (◀), and 400 (◇) are collectively demonstrated. The 139 LMs are colored according to which GUP they belong to.

For each SNR group, it is obvious that the six GUPs are scattered along a single line.

The ISW demonstrated capacity to reduce the fitness values across all SNR groups by two orders of magnitude at around 90 steps. Figure 18 displays the minimum fitness value for each SNR group, with confidence intervals shaded the same color as the lines. It is also evident from Table VI that the majority of the feasible solutions are found before step 90.

Under each SNR, the *mean* fitness values and the *mean* parameter values for each GUP averaged over 2 to 16 decays are reported in Table VIII; similarly, ρ_q is the identified discrete surface relaxivity of quartz, and actual continuum values $\rho_q^\star$ would be a factor A larger, as discussed previously. From Table VIII we observe that from

TABLE VII. Categorization of the 139 UPs into six GUPs. *Min* and *Max* indicate bin edges and each bin includes the left edge, but does not include the right edge, except for the last bin that includes both edges. The *Sum* indicates the number of UPs in that SNR group. For SNR groups 50, 71, 100, 141, GUPs with highest occurrences are colored in orange.

Generalized UP	Min	Max	Number of UPs						
			50	71	100	141	200	282	400
GUP 1	−6.3	−6.2	...	1	...	...	...	...	...
GUP 2	−6.2	−5.7	11	2	6	1	...	...	...
GUP 3	−5.7	−5.2	4	11	10	6	2	3	1
GUP 4	−5.2	−4.8	6	7	9	6	4	1	...
GUP 5	−4.8	−4.6	5	8	7	1	2	...	1
GUP 6	−4.6	−4	12	6	5	1	...	...	...
Sum			38	35	37	15	8	4	2

50 to 400, the SNR does affect the optimality of a single solution or UP since the topography of the objective function is affected. However, the optimality of GUP is maintained. It is also clear from Table VIII and Fig. 19 that both identified *mean* parameter and *mean* fitness values are SNR dependent, a feature that is expected since higher SNRs lead to smaller λ_{opt} and associated more versatile T_2 distributions.

Figure 19 displays comparisons of fitness values and identified physical parameter values of the four GUPs at seven SNRs. The correlations between the physical parameters we discovered in Sec. IV are much more obvious here: in all SNR groups, $\log_{10} T_{2e,c}$ is anticorrelated to $\log_{10} D_{e,c}$. Expectedly, within each GUP, identified values are dependent on the SNR: there is approximately a 7% decrease in ρ_q within the adjacent SNR groups in all GUPs, which is also true for $\log_{10} T_{2e,c}$ but with a smaller rate of decrease. By contrast, $\log_{10} D_{e,c}$ almost keeps constant

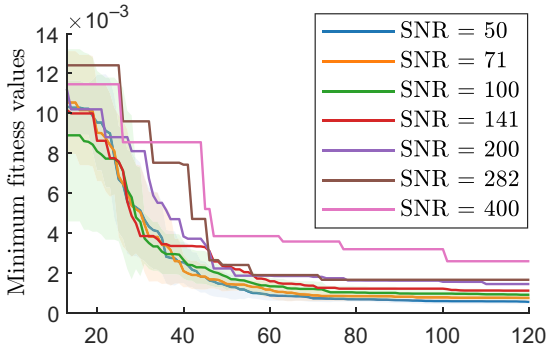

FIG. 18. Minimum fitness values across seven SNR groups. Confidence intervals for the first three SNR groups with 16 decays are displayed as shaded areas. The first 12 steps are not displayed (same fixed preevaluated candidates without using the ISW). The zig-zag shapes are due to a limited number of realizations of decay in high-SNR groups.

TABLE VIII. The results of the *mean* fitness value (MFV) and the *mean* parameter value (MPV) at the four physically valid GUPs for seven groups of decays. For each SNR, GUPs with minimum fitness values are colored in blue. Note that ρ_q is the lattice-based discrete surface relaxivity of quartz.

SNR	GUP	MFV	MPV		
			ρ_q	$\log_{10} T_{2e,c}$	$\log_{10} D_{e,c}$
50	GUP 2	6.143×10^{-4}	6.489	−1.584	−5.882
	GUP 3	5.150×10^{-4}	6.549	−2.119	−5.346
	GUP 4	7.308×10^{-4}	6.282	−2.600	−5.007
	GUP 5	7.938×10^{-4}	6.266	−2.921	−4.684
71	GUP 2	1.249×10^{-3}	6.532	−1.492	−5.886
	GUP 3	7.326×10^{-4}	6.151	−2.114	−5.432
	GUP 4	7.879×10^{-4}	6.427	−2.482	−5.029
	GUP 5	9.879×10^{-4}	6.125	−2.903	−4.727
100	GUP 2	1.217×10^{-3}	5.874	−1.639	−5.932
	GUP 3	1.088×10^{-3}	5.843	−2.146	−5.460
	GUP 4	1.195×10^{-3}	5.850	−2.153	−5.448
	GUP 5	1.306×10^{-3}	5.641	−2.753	−4.973
141	GUP 2	1.669×10^{-3}	5.832	−2.957	−4.700
	GUP 3	1.106×10^{-3}	5.400	−2.249	−5.466
	GUP 4	1.612×10^{-3}	5.774	−2.691	−5.017
	GUP 5	1.421×10^{-3}	6.699	−2.695	−4.692
200	GUP 3	1.249×10^{-3}	5.283	−2.391	−5.402
	GUP 4	1.934×10^{-3}	5.590	−2.757	−4.966
	GUP 5	1.479×10^{-3}	6.919	−2.611	−4.680
282	GUP 3	1.979×10^{-3}	5.957	−2.184	−5.378
	GUP 4	3.748×10^{-3}	5.400	−3.000	−4.870
400	GUP 3	2.576×10^{-3}	6.264	−1.977	−5.426
	GUP 5	3.836×10^{-3}	5.415	−3.000	−4.799

within each SNR group, which is as expected since this is how we defined the six GUPs.

Despite significant changes in the shape of T_2 distributions against SNRs, observing that, for all SNR groups, solutions from GUP 3 are physically plausible, and possess the lowest fitness value and the least false features, we conclude that, for the specific Bentheimer sandstone sample we use, GUP 3 is superior to other GUPs. This is also true for the optimization result of a single decay, as demonstrated in Table VI that the solutions in GUP 3 always possess the lowest fitness value and are most physically plausible.

V. SENSITIVITY ANALYSIS

As we established in Sec. IV when discussing the progress of optimization, practically, there is much more uncertainty in the inference of hyperparameter σ_n^2 than other hypers (Fig. 9). Moreover, due to the ill-posed nature of the ILT, small changes in physical parameter values may lead to small variations in T_2 decay but huge changes in the T_2 distribution after the ILT, making the observations noisy. As a result, successful identification of all feasible solutions is strongly dependent on whether the Gaussian

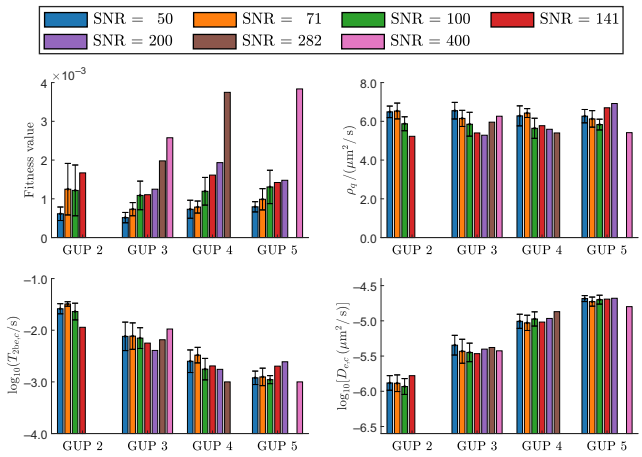

FIG. 19. The results of comparisons of fitness values and identified physical parameters of the four local minima at seven SNR values. Each plot shows four groups of bars, corresponding to the four GUPs that passed the physical validity test (GUP 1 and GUP 6 are not shown). Mean and standard deviations are shown for SNR groups with 16 decays.

process regression captures the true objective function, especially the noise hyperparameter.

A. Underfit and overfit

To understand what role is played by the noise hyperparameter σ_n^2, we display three types of regression model in Fig. 20: underfit, adequate, and overfit, induced by different hyperparameter configurations. With such a limited number of observations (i.e., five), the regression could be conducted in a variety of ways, but we found that the most likely, adequate configuration of hyperparameters is 20% more probable than the overfit configuration and more than twice as likely as the underfit configuration; see Table IX. The overfit model assumes noise-free data and the regression almost passes every observed candidate, which naturally leads to a smaller noise variance σ_n^2 and a smaller length scale l_1. By contrast, the underfit model assumes that all observations are noisy observations sampled from a smooth function and it makes no attempt to fit all candidates precisely. In this context, in the early stage when either an overfit or underfit model is possible, consideration of a posterior ensemble of models is advantageous since it is more versatile, adaptive, and accurate.

B. Global and local search ability

Since the topography of the regression is significantly affected by the choice of hyperparameters, we are curious about the following question: How does inference on the GPR hyperparameters affect the physical parameter identification of the ISW? In this subsection we analyze the sensitivity of the ISW performance against choices of the prior distribution on hyperparameters.

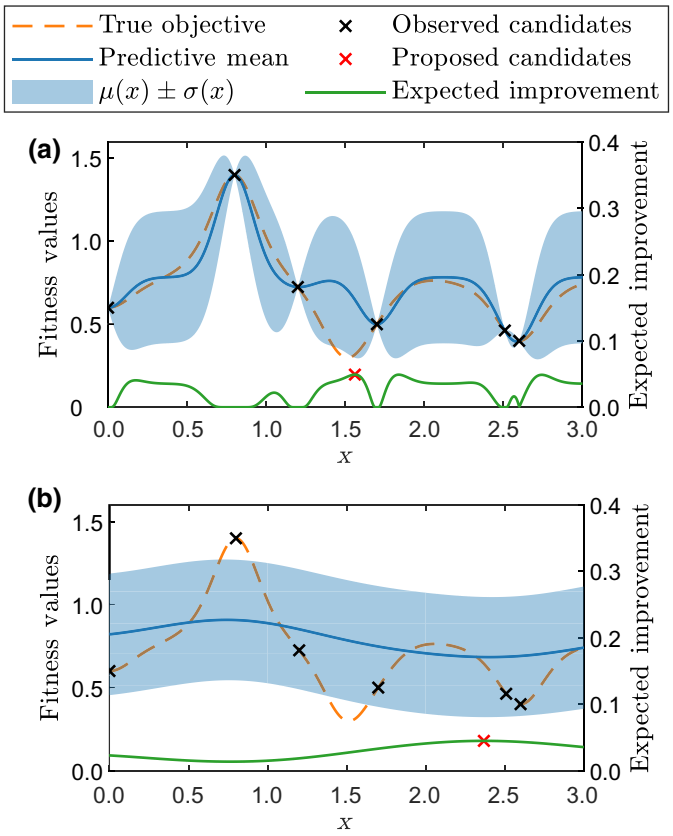

FIG. 20. (a) Overfit and (b) underfit regression for the same dataset as Fig. 5(a). The hyperparameter settings are shown in Table IX.

We empirically compare the search ability of the ISW against five variants of hyperpriors: H6-MCMC and H6-SX, where $X = 1, 2, 3, 4$. Detailed descriptions of all five variants are given in Table X. The four variants of H6-SX compare the different physical parameter identification capabilities given different hyperpriors adopted. As displayed in Table II, the inferred value is $\sigma_n^2 = 5.723 \times 10^{-7}$, and by comparing H6-S$X$ with H6-MCMC we know immediately if the correct prior could speed up the search process.

Recall from Table IV that after performing GPR using all 2431 candidates, we find that there are six UPs in the true objective function but only four are physically valid.

TABLE IX. Details of different configurations of hyperparameters shown in Figs. 20 and 5 with rankings and probabilities. The *Rank* column indicates the rankings out of 200 samples of hyperparameters.

Model	$\log_{10}(\sigma_f^2)$	$\log_{10}(\sigma_n^2)$	m_0	$\log_{10}(l_1^2)$	Prob	Rank
Adequate	−0.800	−5.280	0.784	−1.391	0.0116	1
Overfit	−0.851	−12.065	0.724	−1.882	0.0095	6
Underfit	−1.538	−0.952	0.839	−0.541	0.0055	37

TABLE X. Description of the ISW with five variants of hyperpriors.

Variants	Description
H6-MCMC	Uniform priors on all hyperparameters $\boldsymbol{\theta}$
H6-S1	Differs from H6-MCMC in that $\sigma_n^2 \sim \mathcal{N}(10^{-2}, 0)$
H6-S2	Differs from H6-MCMC in that $\sigma_n^2 \sim \mathcal{N}(10^{-4}, 0)$
H6-S3	Differs from H6-MCMC in that $\sigma_n^2 \sim \mathcal{N}(10^{-6}, 0)$
H6-S4	Differs from H6-MCMC in that $\sigma_n^2 \sim \mathcal{N}(10^{-8}, 0)$

Because of the existence of multiple local optima, the ability to identify not only the global optimal solution but also (major) local optimal solutions is crucial. To compare such abilities, we perform the ISW using the same settings as in Sec. IV testing the effectiveness of the hyperpriors except that the search is limited to the grid of size $17 \times 11 \times 13$ generated as a finite discretization of the parameter space; see Table I. During the optimization by the ISW, all proposed candidates are constrained to the nearest candidate on the grid. The decay under SNR = 100 used in Sec. IV is used throughout the comparison and each optimization problem using the ISW is repeated 128 times using different random streams.

As is shown in Figs. 21 and 22, due to overestimation in σ_n^2, the search in H6-S1 is completely biased, leading to misfit of T_2 distributions and very high fitness values as well as incorrectly identified values of physical quantities. Furthermore, from Table XI we observe that H6-S1 fails to identify either GUP 2 (containing the global optimal solution) or GUP 4 (containing local optimal solutions) in all 128 attempts. In fact, the GPR proxy is underfit owing to the relatively high value of σ_n^2 adopted, and thus providing incorrect insight into the topography of the true objective function causing the optimizer to consistently evaluate the nonpromising region. The poor search ability leads to 0 in both MC-G and MC-C. In H6-S2 the σ_n^2 is slightly better but is still underfit with compromised capability, reflected in the low MC-G and MC-C in Table XI and the high minimum fitness value in Fig. 21.

By contrast, the more appropriately configured H6-S3 and H6-S4 lead to significantly higher MC-G and MC-C. GUP 2 is identified by H6-S4 102 times and all best solutions from GUP 2 are the global optimal solution with rankings equaling 1, while GUP 4 is identified 29 times with mean rankings of 2.6. The MC-G and MC-C tell us that, by using the prior $\sigma_n^2 \sim \mathcal{N}(10^{-8}, 0)$, we expect to identify on average 0.80 global optimal solutions ranking at 1.0 or 1.20 optimal solutions, both global and local, ranking at 1.35. By comparison, $\sigma_n^2 \sim \mathcal{N}(10^{-6}, 0)$ is less effective with an inferior ability to identify both the global optimal solution and local optimal solutions.

Compared with the hyperprior $\sigma_n^2 \sim \mathcal{N}(10^{-8}, 0)$ that has a greater ability to identify the global optimal solution but not local optimal solutions, H6-MCMC identifies

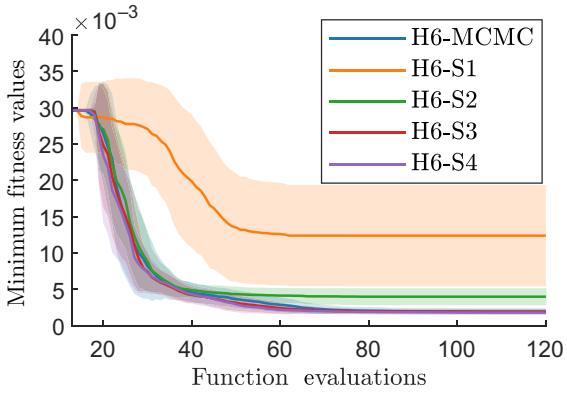

FIG. 21. The mean fitness values for five ISW variants. Means and standard deviations of the inverse problem using the ISW are displayed, repeated 128 times but using different random streams.

the GUP 2 at a slightly lower 89 times but GUP 4 at a significantly higher 112 times. This can be explained by H6-MCMC employing a probabilistic model involving a posterior ensemble of regression models, so that the topography of the GPR proxy constructed using H6-MCMC is much more complex than the H6-SX variants. In terms of the identification ability of the global optimal solution, $\sigma_n^2 \sim \mathcal{N}(10^{-8}, 0)$ is the most effective prior of all five variants.

In fact, we only tested five different priors and potentially there are more effective priors with which the MC-G could be very close to 1. With such priors, the regression could be adequate from the start, which reduces FEs spent on exploration and building up the joint distribution of hyperparameters since it knows the correct distribution before optimization starts. This suggests that the knowledge learnt in one optimization problem is reusable and

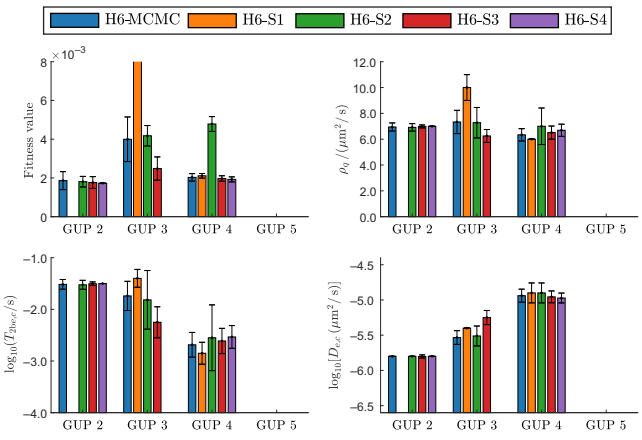

FIG. 22. The comparisons of fitness values and identified physical parameter values in the four physically valid GUPs for five ISW variants. Each plot shows four groups of bars (one for each GUP). Means and standard deviations for the 128 repetitions are displayed.

TABLE XI. Assessment of the five ISW variants. The *mean rankings* following the *counts* specifies the mean rankings of the solutions found within that GUP. MC-G indicates the mean counts of successful identification of the region containing the global optimal solution (GUP 2) and MC-C indicates the mean counts of successful identification of the complete region containing both the global optimal solution and local optimal solutions (GUP 2 and GUP 4), both calculated as GUP counts/number of repetitions (128). MR-G indicates the mean rankings of the solutions found in GUP 2 and MR-C indicates the mean rankings of the solutions found in GUP 2 and GUP 4.

			Pairs of solution counts and associated mean rankings									
GUP	Min	Max	H6-MCMC		H6-S1		H6-S2		H6-S3		H6-S4	
GUP 1	−Inf	−6.2	43	33.0	28	39.2	27	36.7	1	25.0	···	···
GUP 2	−6.2	−5.7	89	2.4	···	···	12	1.8	75	1.3	102	1.0
GUP 3	−5.7	−5.2	15	22.7	···	···	18	24.4	4	6.5	···	···
GUP 4	−5.2	−4.8	112	3.4	···	···	···	···	41	2.9	29	2.6
GUP 5	−4.8	−4.6	···	···	···	···	···	···	7	7.9	···	···
GUP 6	−4.6	Inf	41	15.5	97	144.4	69	30.4	···	···	···	···
Pairs of MC-G and MR-G			0.70	2.40	0	···	0.09	1.80	0.59	1.30	0.80	1.00
Pairs of MC-C and MR-C			1.57	2.96	0	···	0.09	1.80	0.91	1.87	1.02	1.35

transferable to similar problems, which could significantly save FEs for exploration of a similar parameter space to speed up the search.

In short, σ_n^2 plays a vital role in the global and local search ability of the ISW in that it controls the topography of the GPR proxy, which finally affects the physical parameter identification capability. Given finished optimization tasks, a more effective hyperprior can be derived to guide the search.

VI. DISCUSSION

Determining physical parameters using the ISW requires that the physical responses dependent on such parameters can be encoded into the time-domain data, i.e., a magnetization decay $\mathbf{m}_t$ for both measurement and simulation. In this work, $\mathbf{m}_t$ is sensitive to all three unknown parameters selected for optimization, i.e., all three parameters can significantly change observed NMR responses. This leads to practical considerations for the ISW regarding the forward solver, the ISW workflow, and associated SNR considerations for the input data.

A. Practical considerations for the ISW

The resolution of micro-CT imaging is approximately 1–2 μm at best. According to Hürlimann [35], for sandstones, pore size (or surface roughness) less than a critical length scale is considered small and falls into the motional averaging regime despite strong magnetic susceptibility differences. Depending on the sample, the critical length scale ranges from 1.3 to 3.6 μm. This suggests that, for voxels smaller than this, voxel-scale effective physical properties can be defined. In this work clay regions are represented by assigning a single homogeneous phase. In more complex cases a distribution of effective properties at tomogram resolution—or actual microstructure models to account for small-scale heterogeneity explicitly—could be considered.

Separation of internal field effects from surface relaxivity in NMR T_2-CPMG simulation is suggested. If that condition is not fulfilled, the extracted surface relaxivity value will incorporate additional dephasing induced by internal gradients. For stronger internal gradients, the surface relaxivity would then become a function of echo time and, for gases like methane, of pore pressure.

Since the underlying correlation between the physical parameters can be fully utilized to efficiently guide the search, it is advantageous if the physical parameters to be optimized have similar influence on the objective function so that no parameter would be considered "irrelevant." An example could be identifying effective transverse relaxation time constants and effective diffusion coefficients for sandstones with multiple clay types.

We adopted a SNR-dependent regularization parameter, which here requires SNR > 50. Below this SNR T_2 distributions are expected to be strongly biased. In a laboratory environment, the recommended range for the SNR is $[100; 280]$ considering both distribution and informativeness and acquisition time of decay.

B. Suggestions for the SNR choice

Experimentally acquired NMR relaxation responses are subject to various environmental effects, resulting in random and systematic signal fluctuations (noise). The SNR plays a key role in recovering the values of those physical quantities in two aspects. From the perspective of optimization, since the objective function, Eq. (3), involves $\mathbf{s}_{\mathrm{exp}}(T_2)$ inverted from noisy $\mathbf{m}_{t,\mathrm{exp}}$ and $\mathbf{s}_{\mathrm{sim}}(T_2)$ inverted from biased $\mathbf{m}_{t,\mathrm{sim}}$. For such reasons, the topography of the noise-corrupted, observed objective function is different from that of the true but unknown objective function, so

that the optimality of the solutions might be altered, necessitating identification of both the global optimal solution and local optimal solutions. From the perspective of the ILT, false features will arise due to inversion of relatively noisy decays. In fact, for T_2 measurements, a SNR of about 50 is the lowest value we observed for the given cores. Therefore, it would be best if the valid SNR range in which the recovered multiple physical parameters are deemed appropriate can be determined in advance.

For Bentheimer sandstone, the optimization results may be deemed acceptable even at a SNR as low as 50. Other rocks can make optimization more vulnerable to noise depending on the pore-size distribution, especially the micropore component (which is naturally a fast relaxation spot). By contrast, the long-time components of the T_2 distribution (macroporosity) typically associated with a relaxation time as long as hundreds of milliseconds could retain a signal well above the noise level for a few seconds of observation. For Bentheimer sandstone at $t_E = 200\ \mu$s, the inversion process for magnetization decay operates on approximately 70 times more data for the long-time component than the short-time component. As a result, the short-time components are very vulnerable to noise. One may consider more complex NMR sequences with variable t_E and initially higher CPMG pulse frequency. To design such a sequence optimally, prior knowledge of the pore-size distribution would be very desirable.

VII. CONCLUSIONS

In this work we have proposed an ISW to identify values of multiple parameters simultaneously for which the multiscale, multiphysics expensive forward calculation is manipulated to perform nearly identical to observations within a limited budget. The workflow is demonstrated on the inverse problem for identification of three physical parameters in the scope of NMR relaxation in porous media, based on minimization of residuals between simulated and experimentally measured transverse (T_2) relaxation distributions, where simulations are carried out directly on high-resolution tomographic images. To overcome the impact of noise arising in T_2 measurements, a search strategy using state-of-the-art evolutionary optimization algorithms —SL PSO and the gradient-based optimization algorithm L-BFGS-B—are chosen for the selection of plausible configurations of variables, which guarantees that the multimodal nature of the topography of the objective function is captured, and the global optimal set of unknown variables together with most of the local optimal set are identified using solution space partitioning. How solutions obtained from the ISW are affected by numerical realizations and levels of noise are investigated at seven logarithmically spaced SNRs comprising up to 16 uniquely measured decays in each SNR group.

We draw the following specific conclusions.

1. An excellent match between measured and simulated NMR T_2 distributions on Bentheimer sandstone is achieved, enabling the identification of the intrinsic surface relaxivity of the resolved solid (essentially ρ_q) as well as a specific *effective* diffusion coefficient and *effective* transverse relaxation time of clay regions.
2. The ISW uncovers correlations between pairs of parameters. The identified surface relaxivity of quartz is dependent on the *effective* clay transverse time and the *effective* clay diffusion coefficient, which affects standard NMR interpretations.
3. The identified surface relaxivity of quartz lies between $\rho_q = 8.1\ \mu$m/s and $\rho_q = 9.8\ \mu$m/s for a SNR ranging from 50 to 400, in agreement with estimates from mercury intrusion capillary pressure [59].
4. The identified surface relaxivity of quartz does not include the effect of internal fields, which is separately modeled. This is different from common approaches, where surface relaxivity plays the role of a fitting parameter. Consequently, it is to be expected that the presented approach is more sensitive to extracting information about physical surface interactions, in particular if considering different field strengths or combining longitudinal (T_1) and transverse (T_2) relaxation experiments.
5. For the practical application of the ISW, we recommend a micro-CT image resolution roughly in line with the NMR diffusion ruler length, which here is about 2 μm, and SNR > 100 to reduce bias sufficiently.

We note that once the solution domain has been explored, the work may be extended to similar optimization problems for other samples, or the same sample in a different state. An example would be tracking the change in surface relaxivity in dynamic experiments. This should not significantly increase the computational complexity as a full reexploration of the solution domain should not be required.

While the workflow is demonstrated using inverse problems in NMR relaxation in porous media, the principle is general and does not rely on the specific form of the simulator. Consequently, it may find wide applications, in particular in the area of digital rock physics or general porous media applications on the basis of tomographic images.

ACKNOWLEDGMENTS

C.H.A. acknowledges the Australian Research Council (DP190103744) for funding. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia) and an NCRIS enabled capability supported by the Australian Government, through merit grant m65. We thank Xiaoming Xue for valuable discussions.

APPENDIX: ABBREVIATIONS AND NOMENCLATURE

Abbreviations

ILT Inverse Laplace transform
NMR Nuclear magnetic resonance
BO Bayesian optimization
GPR Gaussian process regression
ISW Inverse solution workflow
MCMC Markov chain Monte Carlo
SSP Solution space partitioning
L-BFGS-B Box-constrained limited-memory Broyden-Fletcher-Goldfarb-Shanno
SL PSO Social-learning particle swarm optimization algorithm
SNR Signal-to-noise ratio
CPMG Carr-Purcell-Meiboom-Gill
micro-CT Micro-x-ray-computed tomography
ARD Automatic relevance determination
EI Expected improvement
FE Function evaluation
GP Gaussian process
MAP Maximum a posterior
MLE Maximum likelihood estimation
CDF Cumulative distribution function
PDF Probability density function
GSP Global solution partition
LSP Local solution partition
UP Unique partition
LM Local minimum
FV Fitness value
GUP Generalized unique partition
MFV Mean fitness value
MPV Mean parameter value
H6-MCMC Uniform priors on all hyperparameters
H6-SX H6-MCMC with wrong priors on σ_n^2
MC-C Mean counts of successful identification of the complete region containing both the global optimal solution and local optimal solutions (GUP 2 and GUP 4)
MC-G Mean counts of successful identification of the region containing the global optimal solution (GUP 2)

Math symbols

$a_{\mathrm{EI}}(\mathbf{x}^*)$ Expected improvement evaluated at $\mathbf{x}^*$
$a_{\mathrm{iEI}}(\mathbf{x}^*)$ Integrated expected improvement evaluated at $\mathbf{x}^*$
$\mathbf{d}$ Observables
$\mathrm{diag}(\boldsymbol{\ell})$ A diagonal matrix containing the elements of vector $\boldsymbol{\ell}$
D Dimension of input space
$\mathcal{D}$, $\mathcal{D}_{\mathrm{obs}}$, $\mathcal{D}_{\mathrm{pred}}$, $\mathcal{D}_{\mathrm{prop}}$ Data space, observation set, prediction set, proposal set
$\mathbb{E}_{q(x)}[z(x)]$ Expectation of $z(x)$ when $x \sim q(x)$
$\mathbf{f}$ Vector of Gaussian process latent function values
$\mathcal{GP}$ Gaussian process, $f(\mathbf{x}) \sim \mathcal{GP}[m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')]$; the function is distributed as a Gaussian process with mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$
I The identity matrix
$k(\mathbf{x}, \mathbf{x}')$ The covariance function between pairs of candidates $\mathbf{x}$ and $\mathbf{x}'$
k_{SE} Parameterized squared exponential kernel
K Inversion kernel for the ILT
K_f or $K_f(X, X)$ The $n \times n$ covariance matrix for the (noise-free) latent variable $\mathbf{f}$
K_n or $K_n(X, X)$ The $n \times n$ covariance matrix for the noisy targets $\mathbf{y}$ and $\mathbf{y}^*$, $K_n = K_f + \sigma_n^2 I$
ℓ_i Characteristic length-scale (for input dimension i)
$\boldsymbol{\ell}$ Vector of characteristic length scale
$\log_{10}$ Logarithm to the base 10
$\log \mathcal{U}(a, b)$ Log-uniform distribution defined on $[a, b]$
λ Regularization parameter
λ_{opt} Optimum regularization parameter
Λ Matrix of characteristic length scales, $\Lambda = \mathrm{diag}(\boldsymbol{\ell})^{-2}$
m_0 Constant mean
$\mathbf{m_0}$ Vector of constant mean
$m(\mathbf{x})$ The mean function of a Gaussian process
μ_{θ_i} Fitted mean of the variable's base-10 logarithm
$\mathcal{N}[\boldsymbol{\mu}(X^*), \Sigma(X^*)]$ Normal distribution: a Gaussian (normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ
$\phi(x)$ Probability density function
$\Phi(x)$ Cumulative distribution function
$\mathbb{R}^N$, $\mathbb{R}^{N\times 1}$, $\mathbb{R}^{m\times 1}$ Real spaces of dimensions N, $N \times 1$, $m \times 1$
σ_f^2 Variance of the signal
σ_n^2 Variance of the noise
σ_{θ_i} Standard deviation of the ith hyperparameter's base-10 logarithm
θ_i' The hyperparameter before taking the base-10 log
$\boldsymbol{\theta}$ Vector of hyperparameters
x_i The ith component of input $\mathbf{x}$
$\mathbf{x}$ Unknown parameter(s) of dimension $D \times 1$
$\mathbf{x}_l$ Lower bounds of $\mathbf{x}$
$\mathbf{x}_u$ Upper bounds of $\mathbf{x}$
$\mathbf{x}^-$ The global minimizer of the function
$\mathbf{x}_{\mathrm{BFGS}}$ Local maximizer of integrated EI identified using L-BFGS-B
$\mathbf{x}_{\mathrm{opt},j}$ The j th local maximizer of integrated EI
X $D \times n$ matrix of observed candidates
X^* $D \times n$ matrix of predicted candidates
$\mathcal{X}$ Solution space and also parameter space
y Observation
y^* Prediction

$\mathbf{y}$	$n \times 1$ vector of observations
$\mathbf{y}^*$	$n \times 1$ vector of predictions
$y_{\text{opt},j}$	The local optimum value evaluated at $\mathbf{x}_{\text{opt},j}$

Physics symbols

A	Correction factor for survival probability
$\mathbf{B}_{\text{dip}}(\mathbf{r})$	Dipole magnetic field
$\mathbf{B}_{\text{int}}(\mathbf{r})$	Internal magnetic field
ϕ	Phase of a spin-bearing random walker
ϕ_D	Accumulated phase of a spin-bearing random walker
ϕ_0	Initial phase of a spin-bearing random walker at $t = 0$
$B_z(t_j)$	z component of the internal gradient experienced by a spin-bearing walker at t_j
$D_{e,c}$	Effective diffusion coefficient of water in the clay region
D_w	Self-diffusion coefficient of water
$\mathbf{e}_n$	Random noise assumed zero mean Gaussian
$\epsilon_n(t)$	Random noise assumed zero mean Gaussian evaluated at time t
γ	Gyromagnetic ratio
$\mathbf{m}$	Magnetic dipole moment for a unit volume
$\mathbf{m}_t$	Transverse relaxation decay
$m_t(t)$	Transverse magnetization decay evaluated at time t
$\mathbf{m}_{t,\text{exp}}$	Measured transverse magnetization decay
$\mathbf{m}_{t,\text{sim}}$	Simulated transverse magnetization decay
$M_w(t_j)$	Magnetization decay of an individual walker
$M_{xy}(t_j)$	Transverse magnetization decay at t_j
μ_0	Magnetic permeability of the vacuum
ω_0	Larmor frequency
$\chi(\mathbf{r})$	Susceptibility field
χ_v	Volumetric susceptibility
$\chi_{v,\text{ave}}$	Average volumetric susceptibility
$\chi_{v,c}$	Volumetric susceptibility of the clay region
$\chi_{v,q}$	Volumetric susceptibility of quartz
$\chi_{v,w}$	Volumetric susceptibility of brine
ρ_c	Discrete surface relaxivity of clay
ρ_q	Discrete surface relaxivity of quartz
$\rho_q^\star$	Continuum surface relaxivity of quartz
$\mathbf{s}$	Discrete probability density function of physical quantities
$\mathbf{s}(T_2)$	Discrete probability density function of T_2
$\mathbf{s}_{\text{exp}}(T_2)$	$\mathbf{s}(T_2)$ inverted from $\mathbf{m}_{t,\text{exp}}$
$\mathbf{s}_{\text{sim}}(T_2)$	$\mathbf{s}(T_2)$ inverted from $\mathbf{m}_{t,\text{sim}}$
S_i	Survival probability during the ith echo time, $S_i = S_b S_s$
S_b	Survival probability due to bulk relaxation during the ith echo time
S_s	Survival probability due to surface relaxation during the ith echo time
T_{2e}	Effective transverse relaxation time
$T_{2e,c}$	Effective transverse relaxation time of the clay region
$T_{2b,w}$	Bulk relaxation time of water
t_E	Echo time
t_j	Time accumulated until step j
τ_i	Time step for random walker $\tau_i = \epsilon_w^2/[6D(\mathbf{r})]$

Other symbols

d_0	The predefined distance threshold
ϵ	Resolution of the CT image
ϵ_w	Voxel size of the CT image after subgriding
$n$and $n*$	Numbers of observation and prediction cases
n_s	Number of hyperparameter samples drawn using Bayesian MCMC
n_0	Number of candidates used as starting guesses for SSP, also number of LS
N_0	Number of local solution partitions
l_s	Subgriding factor
LS_i	The ith local optimal solution
$\Theta(\mathbf{x})$	Nonlinear, black-box function

[1] Robert L. Parker, Understanding inverse theory, Annu. Rev. Earth Planet. Sci. **5**, 35 (1977).

[2] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences* (Springer Fachmedien, Wiesbaden, 1993), Vol. 52.

[3] Per Christian Hansen, Numerical tools for analysis and solution of fredholm integral equations of the first kind, Inverse Probl. **8**, 849 (1992).

[4] Lalitha Venkataramanan, Yi-Qiao Song, and Martin D. Hürlimann, Solving fredholm integrals of the first kind with tensor product structure in 2 and 2.5 dimensions, IEEE Trans. Signal Process. **50**, 1017 (2002).

[5] Aytekin Timur, Nuclear magnetic resonance study of carbonate rocks, Log Analyst **13**, 3 (1972).

[6] W. E. Kenyon, J. J. Howard, A. Sezginer, C. Straley, A. Matteson, K. Horkowitz, and R. Ehrlich, in *30th SPWLA Annual Logging Symposium* (Denver, Colorado, USA, 1989), p. LL:1.

[7] Martin D. Hürlimann, Diffusion and relaxation effects in general stray field NMR experiments, J. Magn. Reson. **148**, 367 (2001).

[8] A. Souza, G. Carneiro, A. Boyd, M. Hürlimann, W. Trevizan, B. Coutinho, V. Machado, and R. Bagueira, in *31st International Symposium of the Society of Core Analysts*, SCA2016-047 (2016), p. 1.

[9] Q. Zhang, S.-W. Lo, C. C. Huang, G. J. Hirasaki, R. Kobayashi, and W. V. House, in *39th SPWLA Annual Logging Symposium* (Keystone, CO, USA, 1998), p. FF:1.

[10] A. Matteson, J. P. Tomanic, M. M. Herron, D. F. Allen, and W. E. Kenyon, NMR relaxation of clay/brine mixtures, SPE Reservoir Eval. Eng. **3**, 408 (2000).

[11] R. M. E Valckenborg, Leo Pel, and K. Kopinga, NMR relaxation and diffusion measurements on iron (iii)-doped kaolin clay, J. Magn. Reson. **151**, 291 (2001).

[12] Philip M. Singer, Dilip Asthagiri, Walter G. Chapman, and George J. Hirasaki, Molecular dynamics simulations of NMR relaxation and diffusion of bulk hydrocarbons and water, J. Magn. Reson. **277**, 15 (2017).
[13] Christoph H. Arns, Tariq AlGhamdi, and Ji-Youn Arns, Numerical analysis of nuclear magnetic resonance relaxation–diffusion responses of sedimentary rock, New J. Phys. **13**, 015004 (2011).
[14] Paul R. J. Connolly, Weichao Yan, Daniel Zhang, Mohamed Mahmoud, Michael Verrall, Maxim Lebedev, Stefan Iglauer, Peter J. Metaxas, Eric F. May, and Michael L. Johns, Simulation and experimental measurements of internal magnetic field gradients and NMR transverse relaxation times (T2) in sandstone rocks, J. Pet. Sci. Eng. **175**, 985 (2019).
[15] Yingzhi Cui, Igor Shikhov, Rupeng Li, Shitao Liu, and Christoph H. Arns, A numerical study of field strength and clay morphology impact on NMR transverse relaxation in sandstones, J. Petr. Sc. Eng. **202**, 108521 (2021).
[16] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, and Marc Lanctot *et al.*, Mastering the game of go with deep neural networks and tree search, Nature **529**, 484 (2016).
[17] Eric Brochu, Tyson Brochu, and Nando de Freitas, in *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Eurographics Association, Madrid, Spain, 2010), p. 103.
[18] Ruben Martinez-Cantin, Nando De Freitas, Eric Brochu, José Castellanos, and Arnaud Doucet, A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot, Auton. Robot. **27**, 93 (2009).
[19] Donald R. Jones, Matthias Schonlau, and William J. Welch, Efficient global optimization of expensive black-box functions, J. Global Optimization **13**, 455 (1998).
[20] Grégoire Mariethoz, Philippe Renard, and Jef Caers, Bayesian inverse problem and optimization with iterative spatial resampling, Water Resour. Res. **46**, 1 (2010).
[21] Ran Cheng and Yaochu Jin, A social learning particle swarm optimization algorithm for scalable optimization, Inf. Sci. **291**, 43 (2015).
[22] Jorge Nocedal and Stephen Wright, *Numerical Optimization* (Springer Science & Business Media, New York, 2006).
[23] R. G. Newton, Inverse problems in physics, SIAM Rev. **12**, 346 (1970).
[24] J. Trampert, and R. Snieder, *Wavefield Inversion*, edited by Armand Wirgin, CISM (Springer Verlag, New York, 1999), Vol. 398, p. 119.
[25] A. A. Oberai, Nachiket H. Gokhale, and Gonzalo R. Feijóo, Solution of inverse problems in elastisity imaging using the adjoint method, Inverse Probl. **19**, 297 (2003).
[26] K. R. McCall, D. L. Johnson, and R. A. Guyer, Magnetization evolution in connected pore systems, Phys. Rev. B **44**, 7344 (1991).
[27] Igor Shikhov, Donald S. Thomas, and Christoph H. Arns, On the optimum aging time: Magnetic resonance study of asphaltene adsorption dynamics in sandstone rock, Energy. Fuels. **33**, 8184 (2019).
[28] Herman Y. Carr and Edward M. Purcell, Effects of diffusion on free precession in nuclear magnetic resonance experiments, Phys. Rev. **94**, 630 (1954).
[29] S. Meiboom and D. Gill, Modified spin-echo method for measuring nuclear relaxation times, Rev. Sci. Instrum. **29**, 688 (1958).
[30] E. O. Stejskal and Jacob Schaefer, Data routing in quadrature FT NMR, J. Magn. Reson. **13**, 249 (1974).
[31] Kenneth S. Mendelson, Percolation model of nuclear magnetic relaxation in porous media, Phys. Rev. B **41**, 562 (1990).
[32] David J. Bergman, Keh-Jim Dunn, Lawrence M. Schwartz, and Partha P. Mitra, Self-diffusion in a periodic porous medium: A comparison of different approaches, Phys. Rev. E **51**, 3393 (1995).
[33] C. H. Arns, A. P. Sheppard, R. M. Sok, and M. A. Knackstedt, NMR petrophysical predictions on digitized core images, Petrophysics **48**, 202 (2007).
[34] M. D. Hürlimann, K. G. Helmer, L. L. Latour, and C. H. Sotak, Restricted diffusion in sedimentary rocks. determination of surface-area-to-volume ratio and surface relaxivity, J. Magn. Res. A **111**, 169 (1994).
[35] Martin D. Hürlimann, Effective gradients in porous media due to susceptibility differences, J. Magn. Res. **131**, 232 (1998).
[36] Adrian P. Sheppard, Robert M. Sok, and Holger Averdunk, Techniques for image enhancement and segmentation of tomographic images of porous materials, Physica A **339**, 145 (2004).
[37] Yi-Qiao Song, Novel NMR techniques for porous media research, Magn. Reson. Imaging **21**, 207 (2003).
[38] Charles L. Lawson and Richard J. Hanson, *Solving Least Squares Problems* (SIAM, Philadelphia, 1995).
[39] Georges Matheron, The intrinsic random functions and their applications, Adv. Appl. Probab. **5**, 439 (1973).
[40] Andre G. Journel and Charles J. Huijbregts, *Mining Geostatistics* (Academic press London, 1978), Vol. 600.
[41] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn, Design and analysis of computer experiments, Stat. Sci. **4**, 409 (1989).
[42] Thomas J. Santner, Brian J. Williams, William I. Notz, and Brain J. Williams, *The Design and Analysis of Computer Experiments* (Springer Science & Business Media, New York, 2003), Vol. 1.
[43] Carl Edward Rasmussen, in *Summer School on Machine Learning* (Springer, Berlin, 2003), p. 63.
[44] James S. Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl, in *25th Annual Conference on Neural Information Processing Systems (NIPS 2011)* (Neural Information Processing Systems Foundation, Granada, Spain, 2011), Vol. 24.
[45] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, Boca Raton, 2013).
[46] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, Biometrika **57**, 97 (1970).
[47] Stuart Geman and Donald Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of

images, IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721 (1984).

[48] Radford M. Neal *et al.*, Slice sampling, *Ann. Stat.* 31, 705 (2003).

[49] J. Mockus, in *Optimization techniques IFIP technical conference* (Springer, Berlin, 1975), p. 400.

[50] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger, in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10 (Omnipress, Madison, WI, USA, 2010), p. 1015.

[51] Philipp Hennig and Christian J. Schuler, Entropy search for information-efficient global optimization, J. Mach. Learn. Res. **13**, 1809 (2012).

[52] James T. Wilson, Frank Hutter, and Marc Peter Deisenroth, in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2018), p. 9906.

[53] David Ginsbourger, Rodolphe Le Riche, and Laurent Carraro, in *Computational Intelligence in Expensive Optimization Problems* (Springer Verlag, Berlin, 2010), p. 131.

[54] Javier González, Zhenwen Dai, Philipp Hennig, and Neil Lawrence, in *Proceedings of the Nineteenth International Workshop on Artificial Intelligence and Statistics* (PMLR, Cadiz, Spain, 2016), Vol. 51, p. 648.

[55] Jasper Snoek, Hugo Larochelle, and Ryan P Adams, in Advances in Neural Information Processing Systems (Curran Associates Inc., Red Hook, NY, USA, 2012), p. 2951.

[56] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[57] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Softw. **23**, 550 (1997).

[58] James Kennedy and Russell Eberhart, in *Proceedings of ICNN'95-International Conference on Neural Networks* (IEEE, Madison, WI, USA, 1995), Vol. 4, p. 1942.

[59] Kathyrn E. Washburn, Magdalena Sandor, and Yuesheng Cheng, Evaluation of sandstone surface relaxivity using laser-induced breakdown spectroscopy, J. Magn. Res. **275**, 80 (2017).