Water Resources Research®

Bayesian Optimization With Transfer Learning: A Study on Spatial Variability of Rock Properties Using NMR Relaxometry

Rupeng Li[1,2] , Igor Shikhov[1], and Christoph H. Arns[1]

[1]School of Minerals and Energy Resources, University of New South Wales, Sydney, NSW, Australia, [2]College of Science, China University of Petroleum (Beijing), Beijing, China

Abstract Nuclear magnetic resonance measurements of sedimentary rocks are used to extract various transport properties including hydraulic conductivity and water retention curves. These estimates are controlled by intrinsic physical quantities like surface relaxivities, and effective relaxation time and restricted self-diffusion coefficient of water in clay. Sampling these properties on a set of core plugs presents a series of inverse problems where some of the extracted parameters are expected to be similar. To leverage such valuable information, we extend a previously developed single-task inverse solution workflow (ISW) to the multi-task case, transferring the knowledge gained from previous optimization tasks. Two multi-task kernels: intrinsic model of coregionalization (ICM) and linear model of coregionalization are compared to capture the underlying correlations. We consider three micro-CT images of Bentheimer sandstone from two different blocks imaged at different resolutions, following different segmentation pathways, and demonstrate our approach for the case of low and high task similarity. In both scenarios the multi-task ISW finds lower fitting residuals and uses only one-third to one-half of the function evaluations required by the single-task ISW. The scalability of the multi-task ISW is demonstrated by transferring knowledge of two completed optimization tasks to a third task, which outperforms the single-task ISW, with ICM showing faster convergence. The observed 4% difference for the values identified for samples from the same block and around 28% difference across blocks indicates significant spatial variability in surface relaxivity of the main mineral component, while effective clay parameters show a significantly higher variability.

1. Introduction

In groundwater hydrology and reservoir characterization, the primary sources of rock transport properties are laboratory core analysis and well-logging. Nuclear Magnetic Resonance (NMR) measurements, in particular transverse relaxation time measurements, have been used extensively in both disciplines. In reservoir characterization they were introduced as a sensitive probe of micro-structure (Kleinberg, 1996) and have been shown to correlate to permeability (Arns et al., 2005; Sen et al., 1990), and to predict pore size distributions (Arns, 2004; Ausbrooks et al., 1999; Borgia et al., 1995; Kenyon et al., 1988). In hydrogeology, NMR has become a quick and reliable tool for the estimation of permeability and thus hydraulic conductivity for groundwater applications, for both consolidated and unconsolidated aquifers using either the Schlumberger-Doll-Research (SDR) or Timur-Coates equation (Coates et al., 1991; Dlubac et al., 2013; Maliva et al., 2009) as well as water retention curves of partially saturated rocks (Costabel & Yaramanci, 2013; Jaeger et al., 2009; Mohnke, 2014). These applications utilize the relationship between transverse relaxation time T_2 and surface-to-volume ratio S/V

$$\frac{1}{T_2} = \frac{1}{T_{2b}} + \rho_2 \frac{S}{V} + \frac{D\gamma^2 G^2 t_E^2}{12},$$ (1)

where T_{2b} is the bulk transverse relaxation time, ρ_2 is transverse surface relaxivity, D is the diffusion coefficient of the fluid, γ is the gyromagnetic ratio, G is the average or effective internal gradient, and t_E is the echo-time spacing in the CPMG acquisition sequence. Under the simplifying assumptions of weak coupling and negligible bulk relaxation and dephasing Equation 1 reduces to

$$\frac{1}{T_2} = \rho_2 \frac{S}{V}.$$ (2)

Assuming a pore shape and constant ρ_2 then allows conversion of the T_2 distribution to a pore size distribution given surface relaxivity parameters calculated on the basis of T_2 and S/V available from additional measurements, for example, Brunauer-Emmett-Teller (BET) (Howard & Kenyon, 1992; Hürlimann et al., 1994; Stingaciu et al., 2009). In this interpretation the weak coupling assumption ensures that each pore corresponds to a particular T_2 relaxation time and the magnetization decay can be written as a sum of exponentials. It is clear from above that the heuristic parameter ρ_2 incorporates an ensemble average of all relaxation processes in the pore space; at an individual pore level it includes an average over different mineral surfaces as well as relaxation in clays, which may fill part of a pore. Consequently, if the assumptions leading to Equation 2 are violated and, for example, there is higher susceptibility contrast around a pore resulting in larger internal gradients, larger values of ρ_2 will compensate for the resultant faster relaxation (Duschl et al., 2015; Stingaciu et al., 2010). Alternate 1D and 2D NMR methods based on the Decay due to Diffusion in the Internal Field (DDIF) technique (Song et al., 2002), or similar in nature, were proposed to measure pore-size distributions of sedimentary rock, though constant pore shape and uniform constant surface relaxivity again are limiting assumptions (Afrough et al., 2021; Liu et al., 2014; Muncaci & Ardelean, 2013; Zhang et al., 2016).

Clays in sandstone may lead to diffusional coupling of water between micro- and macro-pore environments, complicating the interpretation of NMR relaxation responses. Since clay exhibits high surface area and small pores, the presence of clay has a strong effect on the T_2 distribution, shifting the T_2 distribution to shorter T_2 values with increasing clay content for both water-saturated sand-clay mixtures and natural soils (Anand et al., 2008; Osterman et al., 2019; Pohlmeier et al., 2009), which might lead to a failure of estimating the water retention curve for clay and silt contents higher than 10% (Costabel & Yaramanci, 2013). The sensitivity of the NMR responses toward the presence of clay is an interesting aspect in the context of pollution remediation efforts, since clay pockets may be difficult to remediate and thus release contaminants for an extended time. The associated diffusive transport depends on the restricted self-diffusion or effective diffusion coefficient of water in clay, which is also an important parameter controlling the effectiveness of magnetization exchange between clay regions and open pores. Determining clay properties experimentally, for example, the effective diffusion coefficient of water in clay, is usually done using bulk clay samples or pellets (Matteson et al., 2000; Sánchez et al., 2008). These measurements are unlikely to represent in-situ conditions in sandstones, where clay pockets would have undergone arbitrary compaction and result from various sedimentary processes.

Approaching this problem by introducing image-based computational methods, which by nature include local S/V information, utilizing intrinsic phase specific surface properties, and effective transport and relaxation parameters of clay regions, may allow determining these parameters by solving an inverse problem. The numerical simulation of NMR responses for rock samples has been challenging and typically simplified micro-structures were considered to analyze NMR responses qualitatively, for example, Toumelin et al. (2003). Arns (2004) introduced the simulation of NMR responses on micro-CT images, acquired on a first-generation instrument. Recently, the combination of fifth-generation high resolution micro-CT imaging with more sophisticated solvers allowed the simulation of realistic NMR responses mimicking experiments over a larger range of relaxation times (Arns et al., 2011; Connolly et al., 2019; Cui et al., 2021, 2022; Mitchell et al., 2012) and including partial saturation responses (Shikhov & Arns, 2015; Shikhov et al., 2017); NMR measurements on core plugs were reproduced by simulations based on high-resolution micro-CT images. Given the large number of possible NMR measurements by adaption of different pulse sequences encoding diffusion and relaxation, and associated sensitivity to physical properties, the availability of such solvers is a potential game changer.

A calibrated NMR response simulation may offer unique ways of analyzing NMR responses in terms of intrinsic physical properties of the system if the latter can be determined. Two major problems arise in such inverse problems: computational burden due to expensive NMR simulation and non-uniqueness of the solution due to ill-posedness. Towards the solution of the computational problem, Bayesian optimization has emerged at the forefront due to its high sample efficiency, flexibility, and reproducibility. It has been applied in recovering physical, geometry or structural parameters, for example, in quantum reaction dynamics (Deng et al., 2020; Vargas-Hernández et al., 2019), optical scatterometry (Schneider et al., 2019) and material design (Kalinin et al., 2020; Kitanidis & Vomvoris, 1983). In earlier literature the surrogate is known as Kriging, and it has been applied in the spatial modeling of physical or experimental parameters in groundwater hydrology (Kitanidis & Vomvoris, 1983), nondestructive testing (Bilicz et al., 2012), etc. The key advantage of Bayesian optimization is that it models the objective function using a Gaussian process (GP) surrogate which is computationally cheap to

evaluate, whereas both enumeration methods and evolutionary strategies require tens of thousands of expensive function evaluations. In addition to the computational burden, it is acknowledged that the solution of the inverse problem may be non-unique and various sets of physical properties may provide mathematically good fits. In the previous work on identifying unknown physical quantities appearing in T_2 relaxation in porous media, we assumed no prior knowledge about the physical quantities and only one observable is involved to constrain the quantities. We approached this problem using the single-task inverse solution workflow (single-task ISW) for the ill-conditioned inverse problems featuring a multi-modal search strategy for the proposal of promising candidates and a multi-modal solution analysis for categorizing solutions into unique partitions (UPs) dominated by local minima (Li et al., 2021). Specifically, we apply both a gradient-based optimizer—the multi-start L-BFGS-B—as well as an evolutionary strategy—the social learning-particle swarm optimizer—ensuring that the multi-modal nature of the severely ill-posed objective function is captured. All identified local minima are checked against external knowledge to determine the optimum solution. The problem is severely ill-conditioned because the objective function is the fitting residuals of T_2 distributions. These are themselves output from inverse problems, namely retrieving the T_2 distribution from the multi-exponential decay. The multi-modal search strategy and solution analysis clearly identifies two solution sets leading to visually satisfactory data fits.

Application of the single-task ISW is straightforward and requires neither prior knowledge nor manual tuning. However, the single-task ISW is non-ideal when we study the spatial variability of the physical quantities of the multiple core samples from the same or different blocks. Given identified parameter values of physical quantities for a similar core sample, we still have to start from scratch as if we still have no knowledge about the core sample. In practice, coring, CT-imaging, image processing, and subsequent simulations and experiments involving one sample require a significant effort, and overlooking such useful information may result in enormous waste in experimental effort and computational resources during the re-exploration of overlapping higher-dimensional solution domains. This may be particularly relevant if a series of core plugs from a cored section is considered. One way of leveraging data from the previously finished procedure of optimization is to take the best solution from the finished task and apply it directly to the new task. This strategy might perform well in some cases, but the success boils down to luck. Another way is to propose a prior distribution of model hyperparameters with their mean equaling the best settings from the finished task. However, this strategy requires excessive manual tuning on the covariances for the hyperparameters and the effect of Bayesian optimization might be compromised.

One of the earliest works describing transferring knowledge gained from known problems to a related problem is by Goovaerts (1997), who proposed the co-kriging approach that accounts for the correlations between variables. Co-kriging has been extensively studied and applied in the scope of geology describing spatial relationships between variables (Gelfand et al., 2004; Goulard & Voltz, 1992; Pelletier et al., 2004). Such techniques gained popularity in machine learning and are largely developed, among which two popular multi-task kernels are intrinsic coregionalization model (ICM) and linear coregionalization model (LCM) (Álvarez et al., 2012; Bonilla et al., 2007; Swersky et al., 2013). Over the years, multi-task Bayesian optimization has been applied in a *forward* manner in a variety of contexts, for example, aircraft design (Min et al., 2017), or machine learning algorithms tuning (Zhou et al., 2020). The idea is similar: exploiting training signals from secondary tasks to improve the learning on the related primary task, which is also known as transfer learning. In addition to GPs, transfer learning using deep neural networks is another active area of research, and there have been successful applications of deep learning frameworks for transfer learning in environmental applications, (Hu et al., 2016; Willard et al., 2021), geological applications (Cunha et al., 2020; de Lima et al., 2019), biological (Imrie et al., 2018) and chemical applications (Cai et al., 2020).

Presented with limited data and very costly function evaluations a compromise between quality and required time to find the solution sets needs to be made. As a result, the application of deep learning methods requiring a large number of function evaluations is undesirable, and we seek to find the optimal solution using a limited computational budget. Furthermore, we are interested in developing a method extracting parameters with direct physical meaning. As a consequence, we incorporate transfer learning capabilities into the single-task ISW using a vector-valued GP to solve *inverse* problems that arise while studying the spatial variability of the physical quantities controlling NMR relaxation in porous media. The multi-task inverse solution workflow (MT-ISW) automatically learns the correlation between tasks, leveraging information from related inverse problems to improve the search in the new inverse problem. Since the solutions sets are considered non-unique, we apply a multi-modal search strategy, which partitions the solution space into several unique regions and strategically
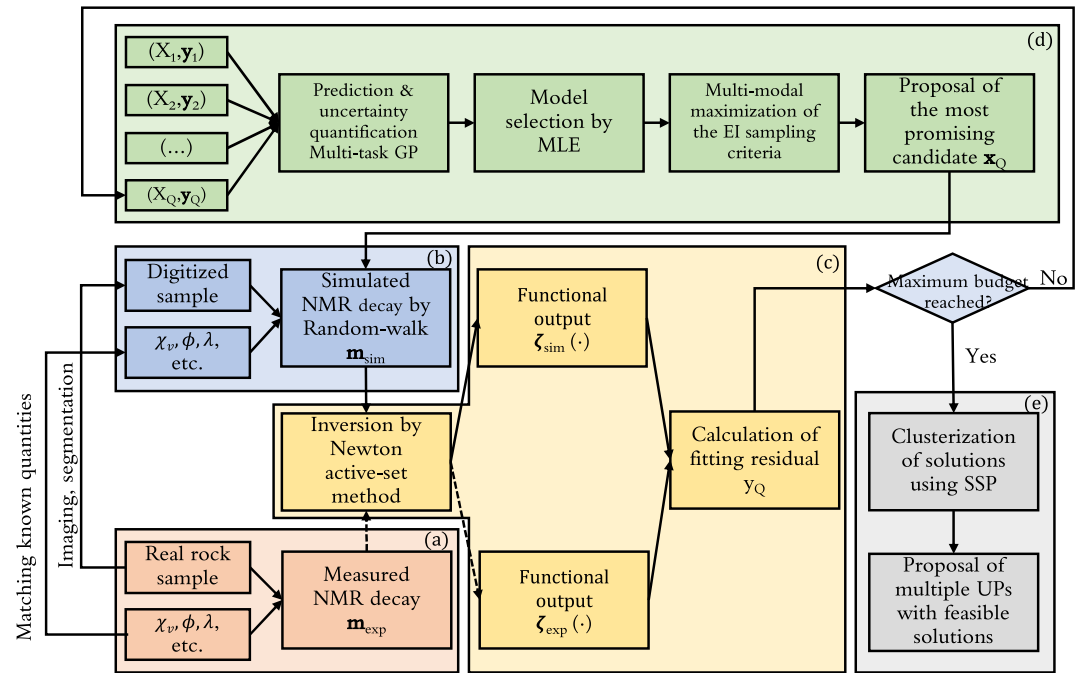
Figure 1. The multi-task inverse solution workflow. The target task (task Q) is started given $Q − 1$ source tasks (task 1 to task $Q − 1$).

balances exploration and exploitation by identifying both the global optimal solution and major local optimal solutions. Once those local optimum solutions are identified, the multi-task ISW selects candidates for evaluation where the associated partition has not been thoroughly explored. This paper is organized as follows: In Section 2 we provide an overview of the multi-task ISW using the inverse problem in the scope of NMR relaxation in porous media as an example. In Section 3, the two major components: vector-valued GP model and the multi-modal search strategy are developed. Results and discussion on identification of three physical parameters for Bentheimer sandstone given one or two completed tasks as well as the choice of the multi-task kernel are shown in Sections 4 and 5. Finally, we conclude our work in Section 6.

2. The Multi-Task ISW

The developed multi-task ISW contains five parts, namely (1) the actual data, (2) the forward solver, (3) the cost-function, (4) the multi-task Bayesian optimization and (5) the solution analysis. The multi-task ISW is presented in Figure 1 using the inverse problem in the scope of NMR relaxation in porous media as an example. Specifically, the complete procedure can be summarized as follows:

a) Measure T_2 relaxation decay $\mathbf{m}_{exp}$.
b) Obtain the digitized representation of the sample via CT-imaging (first iteration only), numerically derive the decays $\mathbf{m}_{sim}$ for the chosen sets of unknown intrinsic parameters $\mathbf{x}_Q$ (example: surface relaxivity of the resolved phase: quartz, effective diffusion coefficient of water in clay, effective relaxation time of clay regions), while honoring known quantities like porosity ϕ, volumetric magnetic susceptibility χ_v, bulk physical properties of water, and matching the physical kernel (see Equation 3) and regularization parameter λ.
c) Acquire the probability density distribution for T_2 using both simulated and measured decays, that is, $\zeta_{sim}(T_2)$ and $\zeta_{exp}(T_2)$. Evaluate the cost-function and calculate the L^2-norm of the fitting residuals for T_2 distributions.
d) Model the target task that is, task Q ($Q \in \mathbb{Z}^+$) using the vector-valued GP with knowledge transferred from the source tasks, that is, task 1 to task $Q − 1$, followed by determination of the most probable model by global maximization of the likelihood of data; then we maximize the multi-task expected improvement (EI) acquisition function to find the promising candidates for evaluation. Finally, we preferably propose a candidate for function evaluation, for which the solution region has not been extensively explored.

Table 1
Measured and Estimated Physical Properties of Bentheimer Sandstone

Block	Sample	χ_ν (μSI)	ϕ	Weight percentage (%)				
				Q	K	K-F	Na-F	Fe(III) oxide
I	BHG0	−1.80	0.238	95.61	1.31	1.72	0.72	0.028
II	BHG1, BHG2	−7.45	0.239	95.67	1.35	1.80	0.59	0.029

Note. The main mineral fractions derived from XRF oxides composition is given in wt%: Q: quartz, F: feldspar, K: kaolinite; χ_ν: volumetric magnetic susceptibility.

e. When run out of budget (total number of allowed function evaluations), a surrogate model for the fitting residuals is built using the vector-valued GP, followed by dividing the solution space into various UPs associated with different modes of the objective function using solution space partitioning (SSP), and finally report the top three solutions associated with each UP.

## 2.1. Experimental Data: Rock Properties and T_2 Relaxation

Three rock core plugs, hereafter labeled BHG0, BHG1, and BHG2, of 50 mm length and 25 mm diameter, were cored from two different blocks quarried from outcrops of Bentheimer Sandstone, an Early Cretaceous formation in North Western Germany. The cores are relatively similar to each other in appearance (light gray), mineralogy and porosity, and may be considered as a sample of a major formation of interest showing small variations between cores due to spatial heterogeneity. We report average measures for the two blocks in Table 1. X-ray fluorescence (XRF) was used to calculate fractions of major mineral constituents for each block. The magnetic susceptibility was measured on finely powdered samples using a Sherwood AUTO magnetic susceptibility balance.

NMR T_2 relaxation responses were measured with a Magritek Rock Core Analyzer operating at 2 MHz (^{1}H) resonance frequency. The magnetization decays were acquired via the CPMG pulse sequence (Carr & Purcell, 1954; Meiboom & Gill, 1958) summing the signal over four scans and following the standard four-step phase cycling sequence (Stejskal & Schaefer, 1974). The resulting decays were cleaned from systematic noise by subtracting the background response. The CPMG experiment was performed with an echo spacing of $t_E = 200$ µs and a signal-to-noise ratio (SNR) of 100 was achieved.

2.2. T_2 Simulations

NMR simulations are carried out on high-resolution images utilizing a random walk approach detailed previously (Arns et al., 2011). In addition to the sample (BHG0) for which a series of physical quantities are determined using the three-phase segmentation as in Li et al. (2021), we consider two additional Bentheimer gray samples as detailed above (BHG1 and BHG2) using the improved procedure of five-phase segmentation as shown below. The Bentheimer gray samples were imaged under the same conditions and an image processing workflow consistent in image processing steps and actual parameter choices was followed. The five-phase segmentation process is illustrated in Figure 2 via central slices through the 1240^3 voxel domain of BHG1. The noise standard deviation in the image was about 50 intensity units after limited neighborhood filtering and using a relatively long acquisition time of 12 hr in double-helix mode (UNSW Tyree X-ray CT facility, GE phoenix source). First, a more standard three-phase segmentation via a converging active contour (CAC) method introduced by Sheppard et al. (2004) is carried out, see Figures 2a–2c; this methods was also followed in Li et al. (2021). An additional two CAC segmentations allow the separation of feldspar and high-density minerals from the quartz phase (Figures 2d and 2e). Small clay clusters were removed, as well as feldspar clusters resulting from bleeding effects around high-density minerals. A snapshot of the three tomograms and their corresponding final segmentations is given in Figure 3. The corresponding intensity histograms depicted in Figure 4 illustrate the intensity ranges of the individual phases. The main variation in intensity distributions between the samples occurs in the more rare high-density inclusions. Respective sample statistics including phase fractions are reported in Table 2. We use the subscripts of p, c, q, f, h for resolved porosity, clay region, quartz, feldspar, and high-density phase, respectively. While there exists even smaller pores, for example, as in grain contact which is believed to exhibit very short relaxation time, this phase is not segmented since the effect is very small which is not observable on
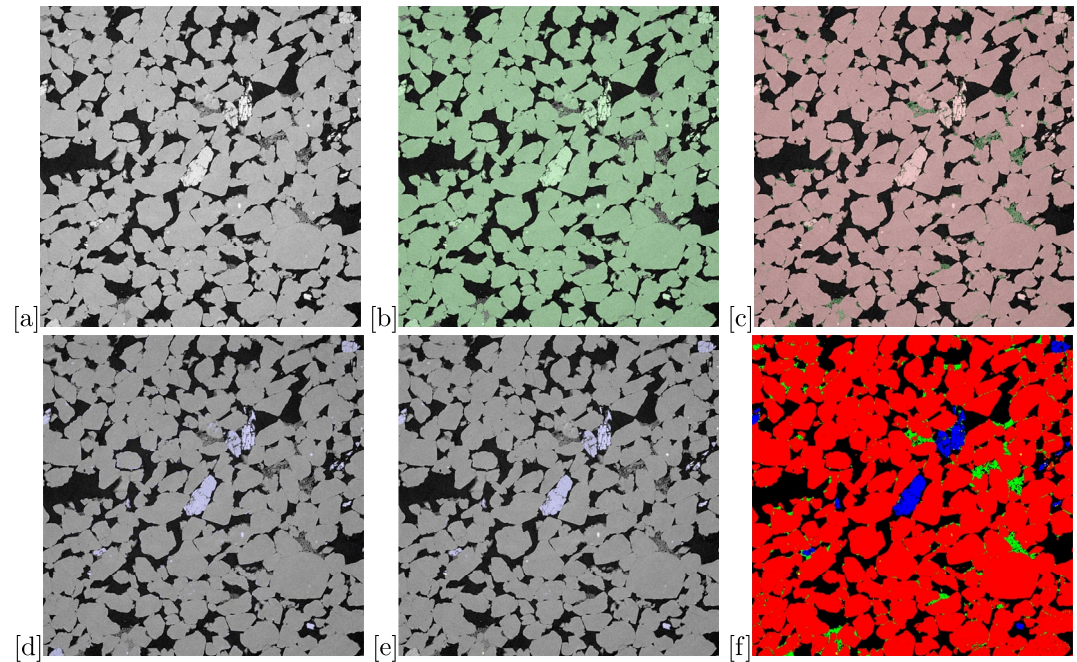
Figure 2. Illustration of the five-phase segmentation process for Bentheimer sandstone (BHG1), displaying central slices through the simulation domain (1240^3 voxel with voxelsize of $\epsilon = 0.216\,\mu m$): (a) gray-scale tomogram, linear intensity scale with $I \in (10{,}000, 15{,}000)$, (b) two-phase segmentation into pore + clay and higher density phases, (c) three-phase segmentation into pore/clay/higher density phases, (d) high-density segmentation extracting all minerals denser than quartz, (e) segmentation of high-density cluster by density average, (f) resultant five-phase segmentation (black: resolved porosity, green: effective clay, red: quartz, blue: feldspar, yellow: high density inclusions).
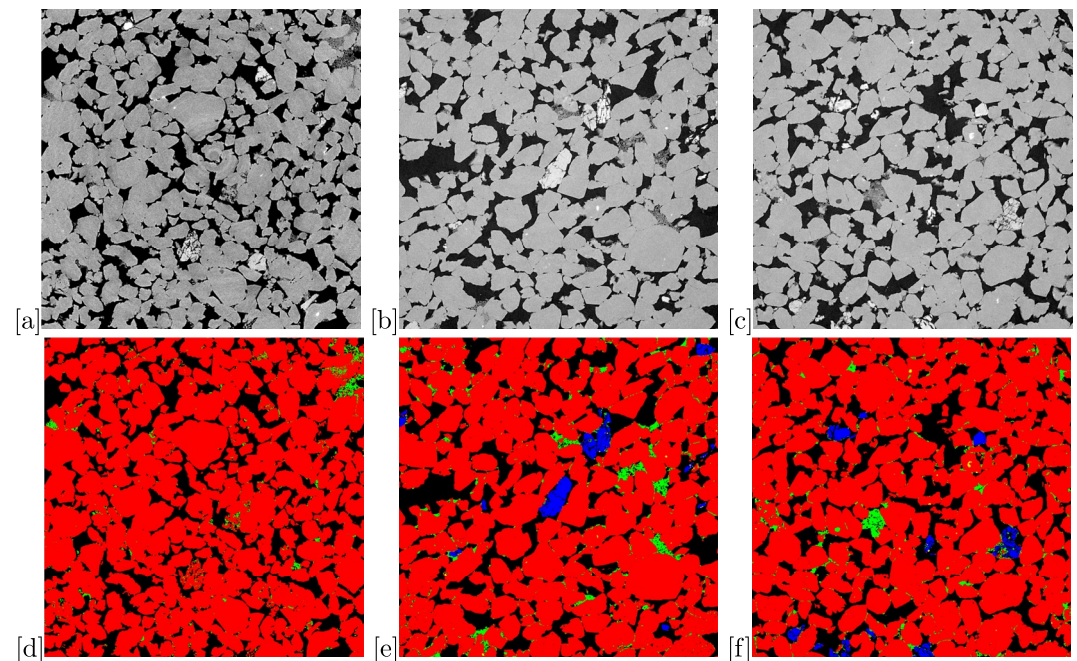
Figure 3. Central slices through the tomogram and resultant segmentation for the simulation domains of the Bentheimer samples. Gray-scale tomograms of (a) BHG0, 1120^3 voxel, $\epsilon = 2.89\,\mu m$, (b) BHG1, 1240^3 voxel, $\epsilon = 2.16\,\mu m$, (c) BHG2, 1240^3 voxel, $\epsilon = 2.26\,\mu m$, and corresponding segmentations: (d) BHG0, three-phase, (e) BHG1, five-phase, (f) BHG2, five-phase.
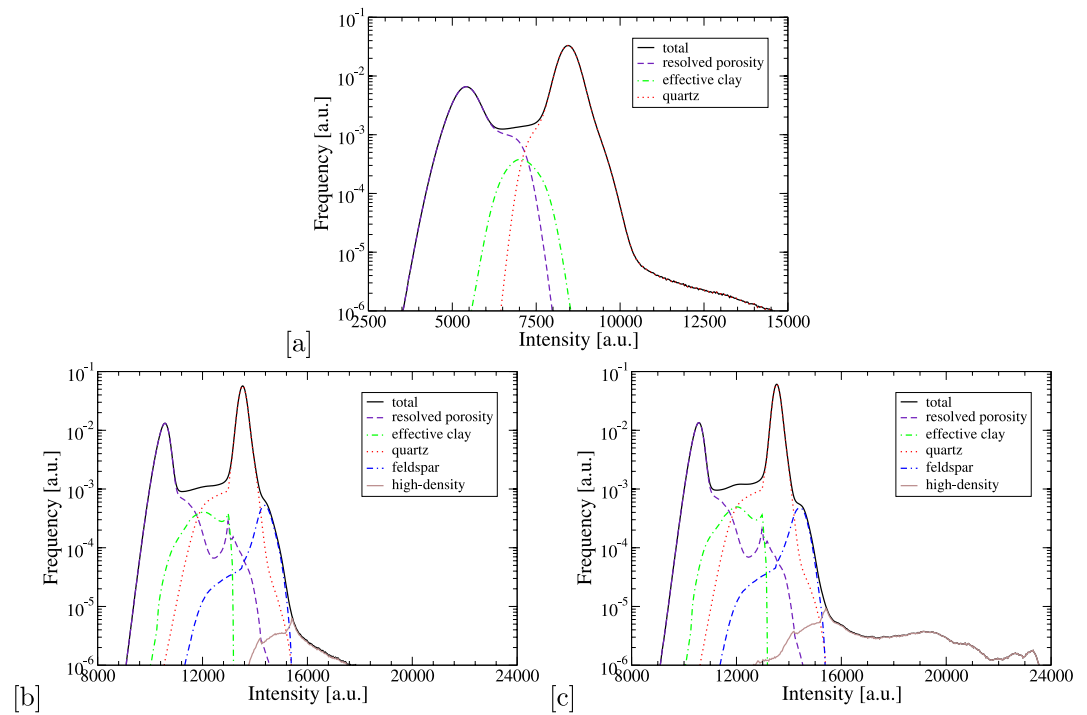
Figure 4. Intensity histograms of Bentheimer sandstone for the resultant segmentations shown in Figure 3. (a) BHG0, (b) BHG1, (c) BHG2.

the T_2 distribution (see reference T_2 distributions in Section 4). Very small pores, for example, grain contact, will be explicitly modeled if, by analysis they have non-negligible contribution.

Given the segmentation of the samples into different phases, the transverse relaxation decay is modeled following the approach of Li et al. (2021), which we quickly summarize here in order to introduce the intrinsic parameters, including those which are to be determined by the multi-task ISW. The internal magnetic field $\mathbf{B}_i(\mathbf{r})$ is calculated in the dipole approximation. The latter is calculated by convolving the susceptibility field with the field of a dipole. We use as volume susceptibilities χ_ν of the phases $\chi_{\nu,p} = -9.0\,\mu\mathrm{SI}$, $\chi_{\nu,q} = -12.0\,\mu\mathrm{SI}$, $\chi_{\nu,c} = 60.0\,\mu\mathrm{SI}$ for BHG0 (three-phase segmentation), and $\chi_{\nu,c} = -10.5\,\mu\mathrm{SI}$ for BHG1 and BHG2 (five-phase segmentation). Since we focus on NMR relaxation at low external field strength of 0.047 T and given the amount of feldspar is small, we assume $\rho_{2,f} = \rho_{2,q}$ and $\chi_{\nu,f} = \chi_{\nu,q}$. From known bulk volumetric susceptibility of the particular sample as given in Table 1 and using literature values for $\chi_{\nu,q}$ and $\chi_{\nu,c}$, the only remaining unknown susceptibility of dense phase $\chi_{\nu,h}$ can be calculated as following: $5{,}300\,\mu\mathrm{SI}$ for BHG1 and $1{,}900\,\mu\mathrm{SI}$ for BHG2.

The samples are considered fully saturated and random walks start in water-containing voxels (open pore space and clay regions) proportional to voxel porosity and with the diffusion coefficient of water ($D_0 = 2.15 \times 10^{-5}$ cm^2/s) in the open pore space, and a value to be determined in clay region voxels $D_{e,c}$. The other two parameters considered unknown are the surface relaxivity of quartz $\rho_{2,q}$ and the effective transverse relaxation time of the clay region $T_{2e,c}$. Here the surface relaxivity of clay is not needed since the clay is modeled using effective medium theory where bulk relaxation and surface relaxation are all lumped into $T_{2e,c}$.

We note that the random walks are carried out on a regular sub-grid conforming to the micro-CT image with resolution of $\epsilon_w = \epsilon/l_s$, where ϵ is the resolution of the tomogram and $l_s \in (5; 20)$ is a grid refinement factor depending on the unknown effective diffusion coefficient of water in the clay region, with $D_{e,c} \in (10^{-7}$ cm^2/s; 10^{-4} cm^2/s) and time-step $\tau_i = \epsilon_w^2/[6D(\mathbf{r})]$.

Table 2
Statistics of the Bentheimer Sandstone Segmentations

| Sample | ϕ | Domain | ϵ (µm) | Phase fraction (%) | | | | |
				f_p	f_c	f_q	f_f	f_h
BHG0	0.2306	1120³	2.888	22.35	1.40	76.24	–	–
BHG1	0.2398	1240³	2.159	22.92	2.12	73.61	1.31	0.04
BHG2	0.2373	1240³	2.226	22.52	2.41	73.69	1.26	0.11

Note. Porosity (ϕ) is calculated as $f_p + 0.5f_c$, domain size is in voxel, ϵ denotes voxel size, and f_i are phase fractions (f_p: resolved porosity, f_c: clay region, f_q: quartz, f_f: feldspar, f_h: high density).

2.3. T_2 Inversion

NMR T_2 relaxation in water-saturated rock is mainly the result of interactions between spin-bearing water molecules and the surrounding environment, for example, surfaces of quartz, clay, and feldspar. The relaxation time is shorter if water is in a bounded region, or if internal field $\mathbf{B}_i(\mathbf{r})$ inhomogeneity due to susceptibility difference between water and minerals is present. Generally, the T_2 decay can be described as

$$m_{T_2}(\tau) = \int \zeta(T_2)\, e^{-\tau/T_2} \mathrm{d}T_2 + \eta_n(\tau),\tag{3}$$

where $m_{T_2}(\tau)$ is the transverse magnetization decay as function of time τ, $\eta_n(\tau)$ is additive white noise, and our objective is the estimation of the probability density function of the transverse relaxation time $\zeta(T_2)$. In matrix notation, Equation 3 can be written as $\mathbf{m}_{T_2} = K\zeta + \eta_n$, where $\mathbf{m}_{T_2}$ is the time-domain data, K is the inversion kernel, ζ is the discretized probability density function of T_2, and η_n is the noise usually assumed zero mean Gaussian. In an optimization framework, the problem is posed as

$$\text{minimize } \|K\zeta - \mathbf{m}_{T_2}\|_2^2 + \lambda\|\zeta\|_2^2,$$
$$\text{subject to } \zeta \geq \mathbf{0},\tag{4}$$

where $\|\cdot\|_2$ denotes the L^2-norm and λ is the regularization parameter. Equation 4 is a linearly constrained convex quadratic programming problem, or equivalently, the non-negative least squares (NNLS) problem, for which a variety of solvers are available (Bertsekas, 1982; Hansen, 1992; Lawson & Hanson, 1995). For consistency, all experimental and simulated decays are inverted using the Newton active-set method. In this study, decays are acquired at SNR = 100, and $\lambda_{\text{opt}} = 12.80$ is adopted as the optimal regularization parameter. Determination of λ_{opt} is non-trivial since it has to balance the trade-off between spectra resolution and solution stability (Song et al., 2002). λ_{opt} is determined using the approach as follows: First, acquire T_2 decays at the desired SNR and a higher SNR, and then invert the two decays using the same algorithm and the same λ. Finally, decrease the λ from a higher value with strong smoothing effect but low spectral resolution to a lower value with weak smoothing but high spectra resolution until the misfit between T_2 distributions inverted from high and low SNR reach a certain threshold. The procedure was proposed for inversion of Bentheimer sandstone T_2 magnetisation decays at SNR of [50,400] (Li et al., 2021), and was then applied to inversion of both T_1 and T_2 data at SNR = 100 (Li et al., 2022a).

3. Incorporating Transfer Learning Ability to the Single-Task ISW

We have demonstrated the ability of single-task ISW to efficiently sample promising candidates in the solution domain in Li et al. (2021). The key is to build a probabilistic model that can be used as a surrogate for the expensive function. In a single-task ISW framework a standard GP is often incorporated to specify the prior distribution over functions, given mean and covariance functions expressed as (Rasmussen & Williams, 2005),

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right),\tag{5}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_D)^\top$ is a candidate in the solution domain $\mathcal{X}$ 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 $\mathbf{x}'$. To implement automatic relevance determination (Rasmussen & Williams, 2005) accommodating different levels of sensitivity against each component of $\mathbf{x}$, we adopted a parameterized squared exponential covariance function k_{SE}, that is,

$$k_{\text{SE}}\left(\mathbf{x}, \mathbf{x}'\right) = \sigma_f^2 \exp\left(-\frac{1}{2}\left(\mathbf{x}' - \mathbf{x}\right)^\top \Lambda^{-1}\left(\mathbf{x}' - \mathbf{x}\right)\right),\tag{6}$$

where $\Lambda = \text{diag}\left(\ell\right)^{-2}$ denotes the matrix of characteristic length-scales with $\ell = (l_1, \ldots, l_D)^\top$, and σ_f^2 denotes the signal amplitude.

3.1. Gaussian Processes for Vector-Valued Functions

For multi-task ISW, the idea is to replace the standard GP kernel Equation 5 with a vector-valued GP kernel accommodating input-output pairs of Q tasks, where Q is a scalar. The Q objective functions can be treated as Q

outputs from a multi-output GP. Similarly to the single-output case, the vector-valued function $\mathbf{f} \in \mathfrak{R}^{Q \times 1}$ following a GP is also fully specified by its mean and covariance functions,

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}\big(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}')\big), \tag{7}$$

where $\{m_q(\mathbf{x}) | q = 1, \ldots, Q\}$ denotes the stacked mean function for each output and $\mathbf{K}$ denotes a matrix-valued function whose entries $(\mathbf{K}(\mathbf{x}, \mathbf{x}'))_{q,q'}$ express not only the correlation (similarity) between the input-pairs $\mathbf{x}$ and $\mathbf{x}'$, but also the correlation (similarity) between the outputs-pairs $f_q(\mathbf{x})$ and $f_{q'}(\mathbf{x}')$. The modeling of $\mathbf{K}$ will be discussed later in this section.

Assume that for each task there are n distinct observed candidates and observations, that is, $\mathcal{D}_{\text{obs}} = (X, \mathbf{y}) = \{(\mathbf{x}_{q,i}, y_{q,i}) | i = 1, \ldots, n, q = 1, \ldots, Q\}$. We use the Gaussian likelihood function to accommodate the noisy observations, that is, $\mathbf{y}|\mathbf{f} = \mathcal{N}(\mathbf{f}, \Sigma_n \otimes I)$, where $\Sigma_n = \text{diag}\left(\sigma_n^2\right)$ is a $Q \times Q$ diagonal matrix accommodating task-dependent noise levels, $\sigma_n^2 = \left(\sigma_{n,1}^2, \ldots, \sigma_{n,Q}^2\right)^\top$ is a vector of noise variances. The posterior distribution for predictions as well as for variances for Q tasks can be expressed as

$$\mathbf{y}^* | \mathbf{x}^*, X, \mathbf{y}, \theta \sim \mathcal{N}(\mu_{\mathbf{x}^*}, \Sigma_{\mathbf{x}^*}), \tag{8}$$

where $\mathbf{y} = \begin{bmatrix} y_1, \ldots, y_q \end{bmatrix}^\top$ concatenate Q outputs together, θ is a vector of model hyperparameters, and

$$\mu_{\mathbf{x}^*} = \mathbf{m}_0^* + \mathbf{K}_f(\mathbf{x}^*, X)\left[\mathbf{K}_f(X, X) + \Sigma_n \otimes I\right]^{-1}(\mathbf{y} - \mathbf{m}_0),$$
$$\Sigma_{\mathbf{x}^*} = \mathbf{K}_f(\mathbf{x}^*, \mathbf{x}^*) + \Sigma_n - \mathbf{K}_f(\mathbf{x}^*, X)\left[\mathbf{K}_f(X, X) + \Sigma_n \otimes I\right]^{-1}\mathbf{K}_f(X, \mathbf{x}^*). \tag{9}$$

Here $\otimes$ is the Kronecker product between matrices and $\mathbf{K}_f(X, X)$ and $\mathbf{K}_f(\mathbf{x}^*, \mathbf{x}^*)$ denote the $nQ \times nQ$ covariance matrix for the latent variable $\mathbf{f}$ evaluated at all pairs of observed candidates and predicted candidates, respectively; $\mu_{\mathbf{x}^*} = \begin{bmatrix} \mu_1(\mathbf{x}^*), \ldots, \mu_q(\mathbf{x}^*) \end{bmatrix}^\top$ concatenate Q predictions and $\Sigma_{\mathbf{x}^*} \in \mathfrak{R}^{Q \times Q}$ is the covariance matrix for Q tasks with variances on the diagonal.

3.2. Coregionalization Models

The key to model Q tasks (outputs) with built-in relationships is to define a multi-task kernel function $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ that utilizes a covariance function specifying the inter-task similarities. Álvarez et al. (2012) proposed a linear model of coregionalization (LCM) assuming that each function is a sum of Q' ($Q' \leq Q$) latent functions each associated with unique mean and covariance functions. Using LCM, the vector-valued GP kernel can be expressed as

$$\mathbf{K}(X, X) = \sum_{q'=1}^{Q'} \mathbf{B}_{q'} \otimes K_{q'}(X, X), \tag{10}$$

where $\mathbf{B}_{q'}$ is a $Q \times Q$ covariance matrix measuring the relationship between tasks, and $K(X,X)_{q'}$ is an $n \times n$ covariance matrix measuring the relationship between inputs, with the (i, j) entry specified by $k_{\text{SE}}(\mathbf{x}_i, \mathbf{x}_j)$. Both matrices are symmetric and positive semi-definite. Kernels composited in this way are called the sum of separable kernels since the covariance function $\mathbf{K}(X, X)$ is expressed as the sum of Q' covariance functions, which themselves are the product of two covariance functions. The important property of LCM is that for each latent function the correlation within the task and between tasks are decoupled. Meanwhile, the degree of correlation is adaptively learned by maximizing the likelihood, as discussed in the model selection. A complete derivation of Equation 10 can be found in Appendix A.

Practically, a simplified form of Equation 10 with $Q' = 1$ known as the intrinsic coregionalization model (ICM), is more frequently used (Bonilla et al., 2007; Swersky et al., 2013):

$$\mathbf{K}(X, X) = \mathbf{B} \otimes K(X, X). \tag{11}$$

ICM assumes that tasks are sampled from the same GP, and it has been extended to multi-task Bayesian optimization involving forward problems transferring knowledge for source tasks with medium to high similarities to the target task (Swersky et al., 2013). However, for our problem the similarities between tasks are not known a priori,

Table 3
Summary for Parameterization of Intrinsic Model of Coregionalization (ICM) or Linear Model of Coregionalization (LCM)

Task	$\boldsymbol{\theta}_L$	σ_n^2	$\mathbf{m}_0$	$\boldsymbol{\ell}$
1	$\theta_{L,1,1,1}, ..., \theta_{L,Q,Q,1}$	$\sigma_{n,1}^2$	$m_{0,1}$	$l_{1,1}^2, l_{1,2}^2, ... l_{1,D}^2$
2	$\theta_{L,1,1,2}, ..., \theta_{L,Q,Q,2}$	$\sigma_{n,2}^2$	$m_{0,2}$	$l_{2,1}^2, l_{2,2}^2, ... l_{2,D}^2$
...	...	...	...	...
Q	$\theta_{L,1,1,Q}, ..., \theta_{L,Q,Q,Q}$	$\sigma_{n,Q}^2$	$m_{0,Q}$	$l_{Q,1}^2, l_{Q,2}^2, ... l_{Q,D}^2$

Note. Parameterization of ICM only needs the first row in $\boldsymbol{\theta}_L$ and $\boldsymbol{\ell}$, whereas LCM needs additional $Q - 1$ rows.

and therefore we use LCM for multi-task learning and reduce Q' to one if the model is considered over-parameterized.

3.3. Model Selection and Infill Criterion

The key to parameterization of ICM or LCM lies in defining the symmetric and positive definite coregionalization matrix $\mathbf{B}$ or $\mathbf{B}_{q'}$. Suggested by Swersky et al. (2013), we represent $\mathbf{B}_{q'}$ using its Cholesky factor L, where $\mathbf{B}_q = LL^\top$, as

$$
L = \begin{bmatrix}
L_{11} & & & \\
L_{21} & L_{22} & & \\
... & ... & & \\
L_{Q1} & L_{Q1} & \cdots & L_{QQ}
\end{bmatrix},
\tag{12}
$$

whereas enforcing that the diagonal of L is non-negative. LCM simply involves Q times more model hyperparameters. Assuming constant mean functions for each task, we give the parameterization of hyperparameters in Table 3.

Both ICM and LCM call for Q mean hyperparameters and Q noise variance hyperparameters. ICM calls for $Q(Q + 1)/2$ hyperparameters in the Cholesky factor and D in length scales; LCM calls for $Q^2(Q + 1)/2$ hyperparameters in the Cholesky factor and DQ in length scales. In summary, the parameterization of the multi-task ISW using LCM or ICM can be expressed as $\theta_m' = \left[\boldsymbol{\theta}_L, \sigma_n^2, \mathbf{m}_0, \boldsymbol{\ell}\right]$, where practically the base-10 logarithm of noise variances and length scales squared are adopted, that is,

$$
\boldsymbol{\theta}_m = \left[\boldsymbol{\theta}_L, \log_{10}\left(\sigma_n^2\right), \mathbf{m}_0, \log_{10}\left(\boldsymbol{\ell}^2\right)\right].
\tag{13}
$$

It is natural and considered appropriate in single-task ISW to adopt a fully Bayesian treatment to infer the posterior distribution of hyperparameters (Li et al., 2021; Snoek et al., 2012). However, in the multi-task case such treatment soon becomes intractable due to the cubic growth of the computational burden with the number of observed candidates. As a result, we adopt the maximum likelihood estimation of hyperparameters by randomly generating hyperparameter samples using Latin hypercube sampling, followed by global maximization of the log-likelihood using the multi-start BFGS method. This ensures that the most probable model is selected with a very high chance.

The infill criterion, or acquisition function, is a mathematical technique that guides how the parameter space should be explored. In this study, the acquisition function of the multi-task ISW simply extends the EI acquisition function (a_{EI}) adopted in single-task ISW described in Li et al. (2021), but is more informative since the predictive mean and variance consider inter-task correlations. Derivation of the EI acquisition function can be found in Appendix B. For illustrative purposes, we give an example in Figure 5 demonstrating how the source task(s) can help guide the search on the target task. Since multi-task ISW utilizes other tasks to guide the search, prediction and uncertainty calculation for one task are much more targeted and efficient.

3.4. Solution Analysis

As discussed before, for inverse problems the solutions are non-unique. Noisy T_2 decays as well as the inverted T_2 distributions worsen matters since the inferred noise variance hyperparameter is larger and thus leads to higher uncertainties in the region where the global and local optimum solutions can be found. As a result, instead of simply choosing the candidate with the best fit as the final solution, we carry out a solution analysis using the following steps:

1. Cluster the solutions using the solution space partitioning proposed in Li et al. (2021) by "sliding" the solution to the nearest unique partition (UP, see Figure 6) using local derivatives information. Mathematically it can be expressed as

$$
\text{minimize } \mu\left(\mathbf{x}^* | X, \mathbf{y}, \boldsymbol{\theta}_m\right), \text{ subject to } \mathbf{x}^* \in \mathcal{X}.
\tag{14}
$$
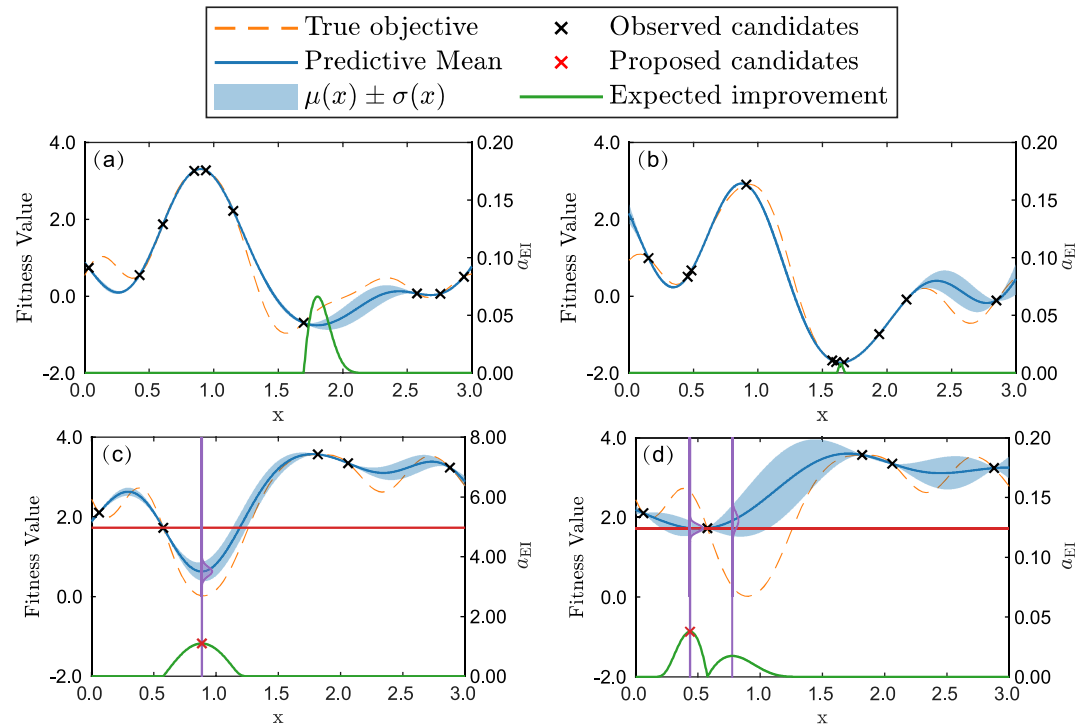
Figure 5. Illustration of the multi-task learning via linear model of coregionalization (LCM) for three correlated tasks in a minimization problem: (a) task 1, (b) task 2, and (c) task 3. Task 1 and task 2 are positively correlated ($\rho_{1-2} = 0.8$), task 1 and task 3 are negatively correlated ($\rho_{1-3} = -0.8$), task 2 and task 3 are negatively correlated ($\rho_{2-3} = -0.8$). Each plot displays prediction, uncertainty quantification (main axis), and expected improvement (EI) (secondary axis). There are 10 unique observations for both task 1 and 2, but only five observations for task 3. Prediction and uncertainty quantification for (a–c) use observations for all 3 tasks whereas for (d) this is accomplished via single-task Gaussian process (GP) using the five observations in task 3 only. The peaks of the green curve on the EI axis indicate the local maxima of $a_{\text{EI},r}$. It is clear how single-task GP missed the local minima given a limited number of observations, and how the LCM makes wiser predictions in task 3 aided by transfer learning.

2. Examine the top candidate in each UP, that is, local minimum (LM), for the physical plausibility test (see Figure 6); LMs that do not pass the test will be removed.
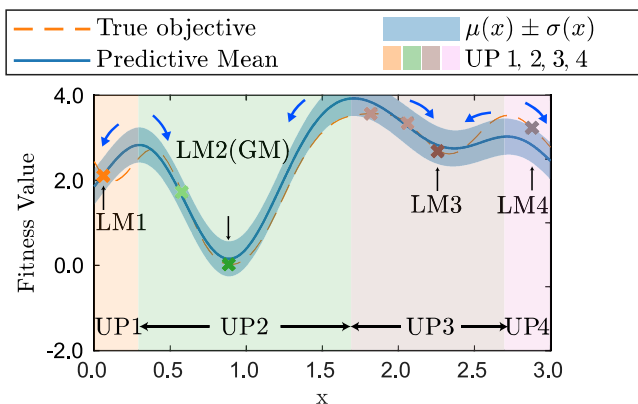

Figure 6. Two successive steps following Figure 5c. Four unique partitions (UPs) and the corresponding minima, that is, local minimum 1 (LM 1) in UP 1, global minimum in UP 2, LM 3 in UP 3, and LM 4 in UP 4, colored orange, green, brown and purple respectively, are identified using solution space partitioning. Of all observations, LMs are colored in denser colors.

We show in Figure 6 how the solutions are clustered into the nearest UP. Seven candidates are clustered into four UPs. In this 1D example, starting from Figure 5c, given the other two finished tasks, the multi-task ISW explores UP 2 and UP 3 efficiently using only two additional function evaluations. Most importantly, the multi-task ISW proposed candidate right at the LM of UP 2. The less important UP 3 and UP 4 will need more function evaluations to get fully explored. By contrast, starting from Figure 5d, the single-task ISW takes 6 more steps only to reach LM 2 (not displayed here). After clusterization using solution space partitioning, all solutions that do not pass the physical validity test are removed.

4. Results

In this section we apply the multi-task ISW to Bentheimer sandstone, using the completed optimization task for BHG0 as the source task. We identify values for the three unknown physical quantities for the samples BHG1 and BHG2: transverse relaxation time $\rho_{2,q}$, effective T_2 relaxation time $T_{2e,c}$ and effective diffusion coefficient of water in the clay region $D_{e,c}$, with the latter two considered in the form of base-10 logarithmic values as $\log_{10} T_{2e,c}$ and
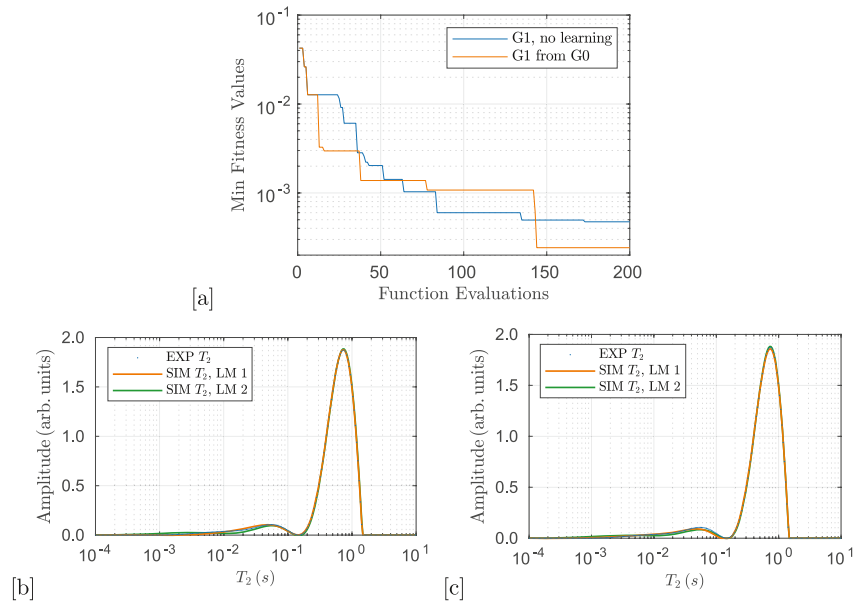
Figure 7. (a) Best fitness value plot of the BHG1 optimization problem with and without transfer learning. T_2 distributions for LM 1 and LM 2 identified with and without transfer learning are displayed in (b-c). Identified parameters corresponding to the local minimums (LMs) are listed in Table 4.

$\log_{10} D_{e,c}$. The performance of the multi-task ISW is first demonstrated on the optimization problem involving BHG1 as the target task, followed by the optimization problem for BHG2 with transfer learning from (a) BHG1 ($Q = 2$), and (b) from both BHG1 and BHG0 ($Q = 3$). Finally, we compare the performance of the two transfer learning kernels: LCM and ICM. The no-transfer case, that is, starting from scratch is provided as the baseline. While applied to a limited number of samples here, the approach would be applicable to the interpretation of an NMR log calibrated to a series of plugs or a measurement series resulting from a dynamic change in surface properties.

The objective of the optimization workflow is to minimize the L^2-norm of the fitting residuals of T_2 distributions between simulation and measurement, referred to as the fitness value. All source and target tasks share the same solution space $\mathcal{X}$, a region bounded by the upper and lower bounds assumed for the three unknowns: $\rho_{2,q}/$(μm/s) $\in (0, 24)$, $\log_{10}(T_{2e,c}/s) \in (-3, 0)$ and $\log_{10}(D_{e,c}/[cm^2/s]) \in (-6.28, -4)$. The Q' in Equation 10 is set to two when transfer learning is involved. With respect to the source task, only the first 120 simulated candidates are provided as the related observations for target tasks; with respect to the target task the initial 4D, that is, 12 solutions are pre-evaluated and kept identical for all target tasks to ensure a fair comparison. The optimization budget is extended to 200 function evaluations to ensure that major non-unique solution sets are identified, and that the optimization budget is sufficient for the single-task ISW to reduce fitness values to a level comparable to the multi-task ISW. We note that although this approach is also applicable to find solutions under a pre-specified tolerance, it is usually unknown and may never reachable so that we pose a limited optimization budget.

4.1. Transferring Knowledge From BHG0 to BHG1

We compare the efficacy of the single-task ISW versus the multi-task ISW for two scenarios: with and without transfer learning. The comparison reveals the higher efficacy of the multi-task ISW when observations from a related task is available, as shown in Figure 7a.

In the early stage, the multi-task ISW and the single-task ISW perform similarly before step 80, after which the single-task ISW gradually reduces the fitness value in the following 120 steps. The multi-task ISW does not minimize the fitness value until step 144. By analysis it is concluded that similar solutions as step 144 are proposed multiple times between step 85 and step 143; however, the proposed solutions are too close to the previously evaluated solutions which may result in computational instabilities so that they are not evaluated until

Table 4
Values for the Unknown Physical Quantities of BHG1 Identified With and Without Transfer Learning

UP	Step	$\rho_{2,q}$ (μm/s)	$\log_{10}(T_{2e,c}/$ [s])	$\log_{10}(D_{e,c}/$ [cm^2/s])	Fitness value	Rank
Multi-task ISW, BHG1 starting from BHG0						
1	**144**	**11.121**	**−1.921**	**−5.255**	**2.423 ×10⁻⁴**	**1**
	166	10.698	−1.937	−5.283	4.588×10^{-4}	2
	164	10.280	−1.875	−5.444	5.431×10^{-4}	3
2	**188**	**11.499**	**−2.702**	**−4.337**	**8.413 ×10⁻⁴**	**6**
	92	10.331	−3.000	−4.274	1.344×10^{-3}	17
	152	10.273	−2.815	−4.470	1.346×10^{-3}	18
Single-task ISW, BHG1 from scratch						
1	**173**	**10.404**	**−2.243**	**−5.070**	**4.737 ×10⁻⁴**	**1**
	135	10.360	−2.309	−4.984	4.957×10^{-4}	2
	84	11.059	−2.218	−4.981	5.995×10^{-4}	3
2	**162**	**10.975**	**−2.606**	**−4.559**	**7.033 ×10⁻⁴**	**4**
	155	10.707	−2.634	−4.567	8.621×10^{-4}	6
	130	11.368	−2.880	−4.170	9.068×10^{-4}	9

Note. Two major UPs with the top three solutions are listed for each method with LMs expressed in bold. *Step* indicate at which step the solution is found. *Rank* indicates the rank of the solution out of 200 evaluations in terms of the FV.

a solution in the less frequently sampled region is proposed at step 144. The multi-task ISW exploits the newly-identified promising domain until the optimization budget is depleted. We apply solution space partitioning and show the UPs and the associated best fit in Table 4 and Figures 7b and 7c respectively.

Table 4 shows the identified values for the three unknown quantities after solution space partitioning. Only two major UPs are displayed for each method. The multi-task ISW outperforms the single-task ISW by finding lower fitness values faster than the single-task ISW; the LM 1 identified by the multi-task ISW exhibit a 50% lower fitness value than the single-task ISW, and this is due to the fact that given finished optimization tasks involving BHG0, the multi-task ISW learns the topology of the objective function by recovering the underlying correlations between the source task and the target task, and thus makes wiser predictions and uncertainty quantifications. T_2 distributions corresponding to the two LMs are displayed in Figures 7b and 7c. It is acknowledged that estimation of the unknown quantities will lead to non-unique solutions, Gavalas et al. (1976) clearly stated the problem: "many different sets of property estimates may provide satisfactory and essentially indistinguishable data fits." In our case, the severe ill-posedness is well-catered by the multi-modal search strategy and the solution space partitioning so that both the single-task ISW and the multi-task ISW provide two sets of mathematically sound solutions. In fact, it is very difficult to distinguish LM 1 from LM 2 by appearance or by the rankings since both display comparatively low fitness values and good rankings, whether in the case of the single-task ISW or the multi-task ISW. The multi-modal search strategy clearly identifies the two unique regions and adaptively balances exploration and exploitation.

Table 5 further shows that the LM 2 in both scenarios is physically invalid since the identified values for restricted diffusion of water in clay are faster than unrestricted diffusion. Meanwhile, it can be observed from Figure 7 that in both cases the T_2 distribution associated with LM 2 exhibits a long tail to the left which does not exist in the measured T_2 distribution. As a result, LM 2 is considered invalid and we report the LM 1, for example, (7.414, −1.921, −5.255), as best parameter values for BHG1, which differs from the reported parameters for BHG0, that is, (5.774, −2.250, −5.316). The difference in the predicted parameters between samples is further compared in the discussion section. Since the source and the target task do not share the same solution set, it can be inferred that simply applying the best parameters identified to the target task is risky and the success of such "direct-transfer" entails certain elements of luck.

4.2. Transferring Knowledge From BHG0 and BHG1 to BHG2

Given completed optimization tasks involving BHG0 and BHG1, we identify values for the three unknown parameters associated with BHG2. Figure 8a illustrates the convergence of the best FV so far for the cost-function for four scenarios: transferring knowledge from BHG0 alone, from BHG1 alone, or from both BHG0 and BHG1. While the solution to the inverse problem with transfer learning from BHG0 is significantly faster than not utilizing prior information, a larger speedup can be noted when learning from BHG1 is included. This can be associated with the higher task similarity between BHG1 and BHG2, which are samples sourced from the same block. The multi-task ISW clearly benefits from such correlation and makes more precise predictions and decisions as to which set of parameters to evaluate. The gain in efficiency by transfer learning makes it more realistic to solve general inverse problems involving more expensive simulations, for example, measurement series including more samples or higher-dimensional D-T_2 or T_2-T_2 NMR responses. We performed solution space partitioning on the results obtained for the four scenarios, with UPs listed in Table 5 and the best fit corresponding to each LM displayed in Figures 8b–8e.

Table 5
Values for the Unknown Physical Quantities of BHG2 Identified With and Without Transfer Learning

UP	Step	$\rho_{2,q}$ (μm/s)	$\log_{10}(T_{2e,c}/$ [s])	$\log_{10}(D_{e,c}/$ [cm^2/s])	Fitness Value	Rank
Multi-task ISW, BHG2 from BHG0						
1	**156**	**12.201**	**−1.923**	**−4.822**	**2.267×10^{-4}**	**1**
	127	11.969	−1.894	−4.886	3.452×10^{-4}	2
	151	12.055	−1.831	−4.944	3.561×10^{-4}	3
Multi-task ISW, BHG2 from BHG1						
1	**87**	**11.932**	**−2.029**	**−4.782**	**2.375×10^{-4}**	**1**
	63	11.259	−1.914	−4.986	2.703×10^{-4}	2
	78	11.888	−1.877	−4.917	3.048×10^{-4}	3
Multi-task ISW, BHG2 from BHG0 and BHG1						
1	**141**	**11.894**	**−1.874**	**−4.930**	**1.965×10^{-4}**	**1**
	64	10.782	−1.976	−5.028	2.228×10^{-4}	2
	191	11.370	−1.856	−5.003	2.430×10^{-4}	3
2	**143**	**9.008**	**−2.997**	**−4.226**	**3.066×10^{-3}**	**82**
	167	10.638	−3.000	−4.000	3.686×10^{-3}	87
	175	9.885	−3.000	−4.147	3.690×10^{-3}	88
Single-task ISW, BHG2 from scratch						
1	**192**	**10.701**	**−2.018**	**−4.971**	**1.983×10^{-4}**	**1**
	193	10.630	−2.018	−4.978	3.034×10^{-4}	3
	198	10.656	−2.009	−4.984	5.172×10^{-4}	12
2	**175**	**11.520**	**−1.778**	**−5.151**	**2.730×10^{-4}**	**2**
	171	11.602	−1.750	−5.160	3.657×10^{-4}	4
	200	11.320	−1.743	−5.187	3.855×10^{-4}	5

Note. Two major unique partitions (UPs) with the top three solutions are listed for each method (LMs in bold; only one UP is displayed if there is only one UP after the application of solution space partitioning).

It is observed from Table 5 that in terms of UP 1, either single-task ISW or multi-task ISW proposes similar identified parameter values, though the no-transfer case spends almost twice the number of function evaluations required by learning from BHG0, and almost three times the function evaluations required for learning from BHG1 (and BHG0). The UP 2 identified for BHG2 with knowledge transferred from BHG0 and BHG1 is physically invalid since the identified value for $\log_{10}D_{e,c}$ is much larger than bulk water; meanwhile, the identified value for $\log_{10}T_{2e,c}$ is hitting the lower bound and also unrealistically low. However, LM 2 is a good solution by appearance alone as is shown in Figure 8d, though the fitness value is more than 10 times higher than for LM 1.

Table 6 shows the inferred most probable hyperparameter settings at the step when LM 1 is identified for the three cases involving transfer learning. The LCM with $Q' = 2$ assumes that each multi-task kernel is decomposed into the sum of two separable kernels. The two Cholesky factors corresponding to $\mathbf{B}_1$ and $\mathbf{B}_2$ are given in L_1 and L_2, with the associated correlation matrices given in R_1 and R_2. The main component, that is, L with the higher value of its first entry on the diagonal is defined as L_1. For all three transfer learning scenarios, the multi-task ISW considers the multi-task kernel as a combination of the main component with larger length-scales and a secondary component with smaller length scales. It is considered that the main component captures the global trend of the correlated objective functions while the secondary component represents the task-dependent local variations. For all three scenarios, the amplitude of the secondary component is less than one-third of the main component. Furthermore, in the case of BHG2 with knowledge transferred from BHG1, the multi-task ISW considers that the main components making up the two objectives are sampled from the same GP, reflected in the L_1 being rank-one and all entries in R_1 being one, which is also true for BHG2 starting from BHG0 and BHG1. BHG2 from BHG0 is almost the same though it is the secondary components that are considered sampled from the same GP. In summary, despite being efficient and providing both mathematically and physically sound solutions, transfer learning using LCM is redundant given moderate correlation between BHG0 and BHG1 and high correlation between BHG1 and BHG2. In the following subsection, we will apply transfer learning using ICM, that is, with $Q' = 1$.

4.3. Transfer Learning Using ICM

The efficiency of transfer learning by ICM is examined by performing the same optimization routine but using ICM as the multi-task transfer learning kernel. It is expected that fewer function evaluations are required since much fewer model hyperparameters are involved for inference. The comparative results, that is, the plot of best FV so far in Figure 9a, and the identified parameter values in Table 7 demonstrate that ICM is more powerful for the inverse problem we study here since in all three scenarios the multi-task ISW using ICM finds lower fitness values faster than using LCM. Within 200 steps, the best FV for BHG2 with knowledge transferred from BHG1, from BHG1, and BHG0 both reach below 1×10^{-4}, as shown in Table 7. We allow for 50 additional function evaluations to demonstrate that all methods have converged, though it confirms that the fitness values are not improved for all four methods. This is also reflected in Table 7; there is no top solution found after step 200. For interested readers regarding all identified solutions with associated T_2 decays and distributions, we uploaded the complete optimization progress as supporting information (Figures S1–S8, Tables S1–S4).

The inferred hyperparameters and inter-task correlations for the models of transfer learning using ICM are displayed in Table 8. The total number of hyperparameters reduces from 16 to 10 for the case of two tasks, and from 24 to 15 for the case of three tasks. As shown in the correlation coefficient matrix R, the ICM recovers the
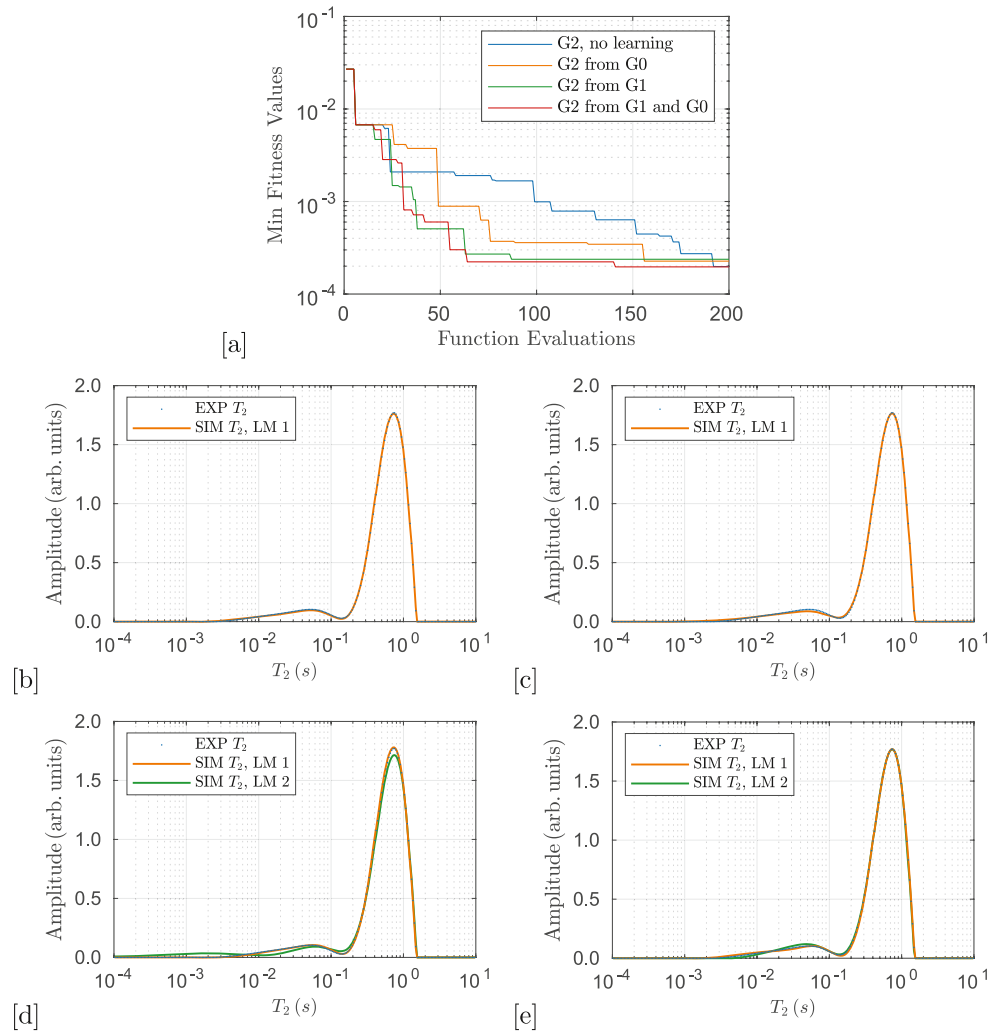
Figure 8. (a) Best fitness value plots of the BHG2 optimization problem with and without transfer learning. T_2 distributions for LM 1 and LM 2 identified with transfer learning from BHG0, from BHG1, from BHG0 and BHG1 are displayed in (b–d). T_2 distributions for LM 1 and LM 2 identified without learning are displayed in (e). Identified parameters corresponding to the local minimums (LMs) are listed in Table 5.

different similarities between tasks: around 70% similarity between BHG2 and BHG0, around 99% similarity between BHG2 and BHG1, and around 78% similarity between BHG1 and BHG2. It should be noted that these values indicate the correlation between the objective functions across the very large solution domain, and the correlation in the domain of interest, that is, where low fitness values are identified, is much lower than reported. Similar as transfer learning using LCM, the mean values $\mathbf{m}_0$ for all three samples are inferred as order of 1; the noise variances σ_n^2 for all three samples are inferred in the order of 10^{-7} (one exception with $\sigma_n^2 = 1.436 \times 10^{-6}$); the length scales inferred for ICM and LCM are all in the order of 1. The inference consistency between ICM and LCM indicates that for the inverse problem we study here, parameterization by LCM is redundant whereas ICM is more appropriate.

To better illustrate the goodness-of-fit, we show the T_2 distributions corresponding to only the LM 1 of the three scenarios in Figures 9b, 9d and 9f. The corresponding T_2 decays are displayed in Figures 9c, 9e and 9g. For fitness values at the level of 1×10^{-5}, there is no physically invalid solution producing a good fit, assuming that the measurement and the modeling procedure are reasonable. As a result, for BHG2 we report the LM 1 identified by the multi-task ISW with knowledge transferred from BHG0, that is, $(7.311, -1.908, -5.032)$ as the best solution.

Table 6
Hyperparameters Inferred for the Three Transfer Learning Scenarios Shown in Table 5, Each at the Step When the LM 1 Is Identified

$\boldsymbol{\theta}_s$	BHG2 from BHG0	BHG2 from BHG1	BHG2 from BHG0 and BHG1
L_1	$\begin{array}{c c c} & G0 & G2 \\ G0 & 1.107 & \\ G2 & 0.146 & 0.369 \end{array}$	$\begin{array}{c c c} & G1 & G2 \\ G1 & 1.185 & \\ G2 & 1.039 & 0 \end{array}$	$\begin{array}{c c c c} & G0 & G1 & G2 \\ G0 & 1.595 & & \\ G1 & 1.344 & 0 & \\ G2 & 1.183 & 0 & 0 \end{array}$
L_2	$\begin{array}{c c c} & G0 & G2 \\ G0 & 0.097 & \\ G2 & 0.174 & 0 \end{array}$	$\begin{array}{c c c} & G1 & G2 \\ G1 & 0.300 & \\ G2 & 0.414 & 0.008 \end{array}$	$\begin{array}{c c c c} & G0 & G1 & G2 \\ G0 & 0.147 & & \\ G1 & 0.306 & 0.156 & \\ G2 & 0.349 & 0.125 & 0.022 \end{array}$
R_1	$\begin{array}{c c c} & G0 & G2 \\ G0 & 1 & \\ G2 & 0.367 & 1 \end{array}$	$\begin{array}{c c c} & G1 & G2 \\ G1 & 1 & \\ G2 & 1 & 1 \end{array}$	$\begin{array}{c c c c} & G0 & G1 & G2 \\ G0 & 1 & & \\ G1 & 1 & 1 & \\ G2 & 1 & 1 & 1 \end{array}$
R_2	$\begin{array}{c c c} & G0 & G2 \\ G0 & 1 & \\ G2 & 1 & 1 \end{array}$	$\begin{array}{c c c} & G1 & G2 \\ G1 & 1 & \\ G2 & 1 & 1 \end{array}$	$\begin{array}{c c c c} & G0 & G1 & G2 \\ G0 & 1 & & \\ G1 & 0.891 & 1 & \\ G2 & 0.940 & 0.990 & 1 \end{array}$
$\mathbf{m}_0$	$\begin{bmatrix} 1.031 \\ 0.493 \end{bmatrix}$	$\begin{bmatrix} 0.995 \\ 0.937 \end{bmatrix}$	$\begin{bmatrix} 0.594 \\ 0.515 \\ 0.415 \end{bmatrix}$
σ_n^2	$\begin{bmatrix} 5.005 \times 10^{-7} \\ 5.673 \times 10^{-7} \end{bmatrix}$	$\begin{bmatrix} 1.627 \times 10^{-6} \\ 3.160 \times 10^{-7} \end{bmatrix}$	$\begin{bmatrix} 4.885 \times 10^{-7} \\ 1.281 \times 10^{-6} \\ 2.737 \times 10^{-7} \end{bmatrix}$
ℓ	$\begin{bmatrix} 4.830, 2.369, 3.172 \\ 3.712, 0.836, 0.943 \end{bmatrix}$	$\begin{bmatrix} 7.038, 0.728, 4.586 \\ 4.914, 1.903, 1.244 \end{bmatrix}$	$\begin{bmatrix} 6.779, 1.691, 3.449 \\ 3.178, 0.952, 1.377 \end{bmatrix}$

Note. Mathematical and physical definitions of the hyperparameters can be found in Equations 6, 9 and 12. R_1 and R_2 are the correlation coefficient derived from the Cholesky factor of the correlation matrix L_1 and L_2.

We show the isosurface plot for the objective function associated with BHG2 with knowledge transferred from BHG0 and BHG1 modeled by ICM in Figure 10. The five iso-levels in ascending order of fitness values represent the regions with high to medium probabilities of finding good solutions. We show the box-bounded solution subspace in Figure 10, which is approximately 5% of the complete solution space, for better illustration. Practically any candidate identified in the region enveloped by the second-lowest contour can be considered as a good solution, as shown in Table 7; it is observed that for parameter estimation problem involving BHG2, solutions with $\rho_{2,q}/(\mu m/s) \in [11.0, 12.0]$, $\log_{10}(T_{2e,c}/s) \in (-1.92, -1.85)$, and $\log_{10}(D_{e,c}/[cm^2/s]) \in [-4.9, -5.1]$ all
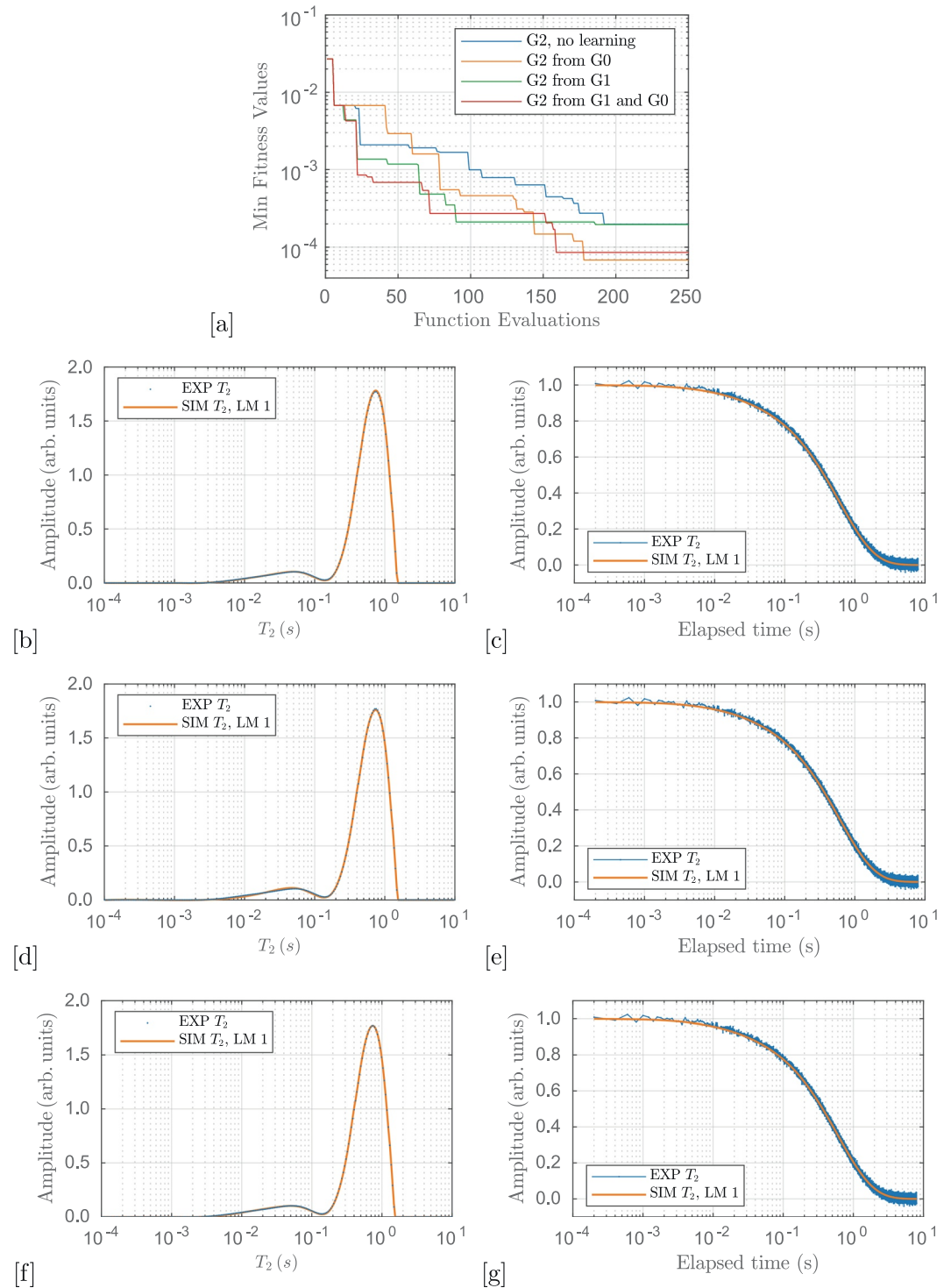
Figure 9. (a) Best fitness value plots of the BHG2 optimization problem with and without transfer learning using intrinsic model of coregionalization. T_2 distributions for LM 1 identified with transfer learning from BHG0, from BHG1, from BHG0 and BHG1 are displayed in (b, d, f). The corresponding T_2 decays are displayed in (c, e, g). Identified parameters corresponding to the local minimums (LMs) are listed in Table 7.

provide reasonable good fits and are physically valid. We also plot the top three solutions identified for BHG2 in the three scenarios using transfer learning, with solutions identified by ICM and LCM marked as ■ and × respectively. The majority of the solutions identified by ICM are in the very narrow most promising region, whereas solutions identified by LCM are slightly scattered.

Table 7

Values for the Unknown Physical Quantities of BHG2 Identified Using Transfer Learning Using Intrinsic Model of Coregionalization

Unique partition (UP)	Step	$\rho_{2,q}$ (μm/s)	$\log_{10}$ ($T_{2e,c}$/[s])	$\log_{10}$ ($D_{e,c}$/ [cm^2/s])	Fitness value	Rank
Multi-task ISW, BHG2 from BHG0						
1	**178**	**10.967**	**−1.908**	**−5.032**	**6.832 × 10⁻⁵**	**1**
	171	10.967	−1.916	−5.021	1.199 × 10⁻⁴	2
	144	10.962	−1.866	−5.077	1.480 × 10⁻⁴	3
2	**104**	**11.627**	**−2.535**	**−4.247**	**1.632 × 10⁻³**	**48**
	179	12.046	−2.622	−4.094	1.715 × 10⁻³	50
	113	12.310	−2.668	−4.000	1.719 × 10⁻³	51
Multi-task ISW, BHG2 from BHG1						
1	**186**	**11.535**	**−1.848**	**−5.002**	**1.961 × 10⁻⁴**	**1**
	90	11.560	−1.992	−4.893	2.121 × 10⁻⁴	2
	94	11.477	−1.951	−4.968	2.906 × 10⁻⁴	3
2	**22**	**10.513**	**−2.486**	**−4.461**	**1.363 × 10⁻³**	**40**
	38	10.988	−2.710	−4.182	1.724 × 10⁻³	70
	58	11.414	−2.485	−4.356	1.909 × 10⁻³	75
Multi-task ISW, BHG2 from BHG0 and BHG1						
1	**159**	**11.498**	**−1.899**	**−4.967**	**8.563 × 10⁻⁵**	**1**
	157	11.521	−1.922	−4.941	1.698 × 10⁻⁴	2
	152	11.548	−1.959	−4.900	2.068 × 10⁻⁴	3
2	**38**	**12.541**	**−2.461**	**−4.212**	**9.301 × 10⁻⁴**	**42**
	48	12.721	−2.637	−4.000	1.415 × 10⁻³	59
	51	12.458	−2.701	−4.000	1.463 × 10⁻³	61

Note. Two major UPs with the top three solutions are listed for each method (LMs in bold).

5. Discussion

5.1. Predicted NMR Parameters

We summarize the identified values for physical quantities associated with BHG0, BHG1 and BHG2 under SNR = 100 in Table 9. The difference in results for the optimal values for $\rho_{2,q}$ between the two Bentheimer sandstone samples from the same and different blocks are 1.4% and 27.5% respectively; for $\log_{10} T_{2e,c}$ the differences are 0.7% and 17.5%; for $\log_{10} D_{e,c}$ the differences are 4.4% and 3.35%, whereas the maximum difference in the physical quantities in their original form, that is, $T_{2e,c}$ and $D_{e,c}$, between BHG0 and BHG1 (BHG2) is roughly a factor of 2. While susceptibility assignments were different for BHG0 versus BHG1 and BHG2, sensitivity tests showed that internal field effects are negligible (at 2 MHz), and not able to explain this difference. As a result, we attribute the variation between samples to sample heterogeneity. The 28% differences in values for surface relaxivity are not unusual for geological samples, see for example, Sen et al. (1990), and we attribute the difference to natural variations in $\rho_{2,q}$ for Bentheimer Sandstone.

For comparison, the identified values for surface relaxivity are close to the reported $\rho_2 = 8.0$ μm/s for Bentheimer sandstone determined using the Bayesian MR approach by comparing the T_2 distribution and pore size distribution (Holland et al., 2013), but are smaller than the reported $\rho_2 = 26$ μm/s for Bentheimer sandstone Afrough et al. (2021). The latter 2–3 times higher value is mainly due to the planar geometry assumed in their work, which provides less surface area per volume compared to spherical geometry.

Comparison of relaxation time $T_{2e,c}$ and effective diffusion coefficient $D_{e,c}$ in clay deduced in this study to literature values is complex due to several factors. All literature reported values are obtained from specially manipulated naturally occurring kaolinite deposits (kaolins). Such samples tend to be contaminated by other naturally occurring minerals, which is sometimes under-reported or ignored by authors. This is particularly important in regards to NMR measurements since kaolinite, even in the purest form, typically contains iron occurring as the most abundant structural impurity due to isomorphic substitution of Al atoms by Fe(III) in the octahedral sheet.

The two most frequently used kaolinite reference samples for research purposes from the Source Clay Repository (SCR), Georgia kaolinite KGa-1 and KGa-2, have an iron content expressed in Fe_2O_3 weight fraction varying between 0.1 wt% and 1 wt%, with corresponding impact on magnetic susceptibility (diamagnetic and paramagnetic respectively) and NMR surface relaxivity. Furthermore, kaolinite samples require preparation to take some sort of physical form prior to experimental measurements resulting in a very broad degree of compaction ranging from fluid-like slurry to specimens highly compacted using a variety of methods as opposing to the formation of authigenic kaolinite within a host sandstone rock. Noting these limitations, we attempt to define a realistic literature-based range of the effective T_2 relaxation time in saturated kaolinite for comparison.

Fleury et al. (2011) reported T_2 relaxation of kaolinite in powdered and compacted (41 wt% porosity) states, albeit using an NMR set operating at a higher Larmor frequency of 23.7 MHz and with partially saturated samples. Nevertheless their data provide a useful lower bound value of near surface T_2 relaxation of 2 ms. Matteson et al. (2000) reported T_2 relaxation times at 2 MHz Larmor frequency for variously compacted kaolinite clay samples. Regretfully, the authors reported only applied pressure (or/and centrifuge speed) used to obtain specimens, but not resulting porosity. The initial slurry exhibits a T_2 of 30 ms and the three samples of the smallest degree of compaction had T_2 values between 10 ms and 24 ms. Thus, the values determined by MT-ISW (BHG1 and BHG2) fall generally into the shorter relaxation time part of the interval reported in literature.

Comparison of MT-ISW derived water diffusion values in kaolinite with literature is complex for the same reason as T_2 relaxation – all studies focused on compacted systems. However, a reasonable estimate is possible

Table 8
Hyperparameters Inferred at the Step When the LM 1 Is Identified for the Three Transfer Learning Scenarios Shown in Table 7

$\boldsymbol{\theta}_s$	BHG2 from BHG0	BHG2 from BHG1	BHG2 from BHG0 and BHG1
L	$\begin{bmatrix} & G0 & G2 \\ G0 & 0.755 & \\ G2 & 0.883 & 0.880 \end{bmatrix}$	$\begin{bmatrix} & G1 & G2 \\ G1 & 1.161 & \\ G2 & 1.060 & 0.104 \end{bmatrix}$	$\begin{bmatrix} & G0 & G1 & G2 \\ G0 & 0.572 & & \\ G1 & 0.574 & 0.443 & \\ G2 & 0.413 & 0.357 & 0.108 \end{bmatrix}$
R	$\begin{bmatrix} & G0 & G2 \\ G0 & 1 & \\ G2 & 0.709 & 1 \end{bmatrix}$	$\begin{bmatrix} & G1 & G2 \\ G1 & 1 & \\ G2 & 0.995 & 1 \end{bmatrix}$	$\begin{bmatrix} & G0 & G1 & G2 \\ G0 & 1 & & \\ G1 & 0.778 & 1 & \\ G2 & 0.742 & 0.980 & 1 \end{bmatrix}$
$\mathbf{m}_0$	$\begin{bmatrix} 1.011 \\ 1.195 \end{bmatrix}$	$\begin{bmatrix} 1.161 \\ 1.037 \end{bmatrix}$	$\begin{bmatrix} 0.594 \\ 0.515 \\ 0.415 \end{bmatrix}$
$\boldsymbol{\sigma}_n^2$	$\begin{bmatrix} 5.339 \times 10^{-7} \\ 5.314 \times 10^{-7} \end{bmatrix}$	$\begin{bmatrix} 1.445 \times 10^{-7} \\ 4.045 \times 10^{-7} \end{bmatrix}$	$\begin{bmatrix} 4.780 \times 10^{-7} \\ 1.436 \times 10^{-6} \\ 5.143 \times 10^{-7} \end{bmatrix}$
ℓ	[6.038, 0.674, 2.414]	[5.070, 0.746, 2.188]	[3.954, 1.247, 1.837]

with the aid of diffusion-porosity scaling relationships and numerical simulation. Sánchez et al. (2008) reported effective diffusion values for highly compacted kaolinite samples KGa-2 (porosity 26%, bulk diffusion of water $D_0 = 2.3 \times 10^{-5}$ cm²/s at T = 25°C) as 0.17×10^{-5} cm²/s. This value can be taken as a lower bound estimate, though with excessive margin. For simplicity we use in this work the porosity value of kaolinite accumulations in Bentheimer sandstone of 50%, which in reality can be slightly higher and likely vary within 10% from the mean value. The experimental value from Sánchez et al. (2008) can be adjusted to conditions of our experiments, such as temperature (by using corresponding self-diffusion of water 2.15×10^{-5} cm²/s and porosity of 50% (by using one of many diffusion-porosity relationships). Assuming that effective to bulk diffusion ratios can be scaled with a porosity power law, the exponent is 1.93. It falls between 1.5 and 2, the two exponent values used by several common correlations, e.g., Bruggeman (1935). The resulting range of effective diffusion values is $0.54 \sim 0.76 \times 10^{-5}$ cm²/s. Dabat et al. (2020) reported effective diffusion values for SCR KGa-2 kaolinite samples of 50% porosity prepared using two methods of compaction (pressing the powder and centrifuge). They reported the range of effective water diffusion in these systems (depending on direction and compaction method) between 0.31 and 0.57 (in terms of effective diffusion to free diffusion ratio), which translates into absolute values as $0.67 \sim 1.23 \times 10^{-5}$ cm²/s (using $D_0 = 2.15 \times 10^{-5}$ cm²/s).

Mu et al. (2007) investigated the relationship between porosity and effective diffusion coefficient in random porous media (non-compacted tetrahedron and hexahedron packs) by numerical simulation. Porosity increases from 36% to 50% lead to an increase of effective diffusion to free diffusion ratio from $0.14 \sim 0.27$ to $0.29 \sim 0.41$, i.e., approximately factor 2 increase. The spread is due to the shape factor of the particles. In absolute terms (water, $D_0 = 2.15 \times 10^{-5}$ cm²/s) the effective diffusion values are: $0.30 \sim 0.58 \times 10^{-5}$ cm²/s for 36% porosity pack and $0.62 \sim 0.88 \times 10^{-5}$ cm²/s for 50% porosity system. Thus, MT-ISW predicted values are generally within the range of values reported (or rather derived from) literature, while the spread is likely due to the mismatch of kaolinite porosity in the micro-CT imaged cores and large cores used for NMR measurements, which we ignored here for simplicity using a constant kaolinite porosity of 50%.
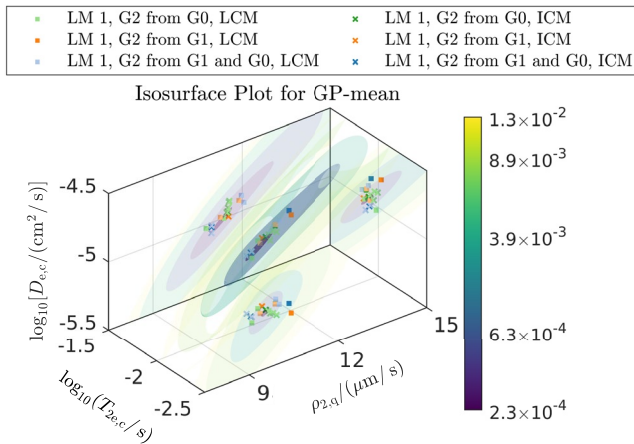
Figure 10. The isosurface plot of the objective function associated with BHG2 from BHG0 and BHG1 modeled by the multi-task inverse solution workflow using intrinsic model of coregionalization (zoomed view). We display five levels of isosurfaces, corresponding to the minimum 0.01%, 0.1%, 1%, 3%, 5% of the scalar field. The correlations between each pair of parameters evaluated at the step 159 are displayed in $\rho_{2,q} - \log_{10} T_{2e,c}$, $\rho_{2,q} - \log_{10} D_{e,c}$ and $\log_{10} T_{2e,c} - \log_{10} D_{e,c}$ planes. Both the isosurface plot and the correlation plots share the same colorbar. The displayed region is a small subset of the complete solution space; denser colors indicate regions with higher probability that mathematically good solutions can be found. Spatial distribution of the top three candidates for UP 1 associated with BHG2 identified in the three transfer learning scenarios listed in Tables 5 and 7 are plotted, with LM 1 in denser colors and other solutions in lighter colors. Solutions in the $\rho_{2,q} - \log_{10} T_{2e,c}$, $\rho_{2,q} - \log_{10} D_{e,c}$ and $\log_{10} T_{2e,c} - \log_{10} D_{e,c}$ planes are the projections.

We note that for the simulations here a micro-CT resolution of about 2.5 μm was utilized. This allowed imaging with significant field of view (5 mm diameter plugs). Reducing the uncertainty in effective clay parameters further may require a multi-scale characterization of the samples, for example, application of a super-resolution method. In our view this may only be necessary if one is interested in responses at higher field strength, where internal field effects are more pronounced, or for samples of lower porosity—for example, tight sandstones—where the porosity associated with grain contacts would be significantly larger. The alternative of using even smaller samples is less palatable due to associated issues with representative volume and coring damage.

5.2. Non-Unique Solution Sets

Since T_2 distributions were chosen as a reference (which are themselves the output of an inverse problem), the solution domain is likely to exhibit multiple local minima. We have shown in Section 4 that usually at least two UPs are identified by multi-modal maximization of the EI acquisition function and multi-modal solution analysis. This statement is true for this specific work, where only one type of NMR observable is utilized (T_2 relaxation). The number of UPs can be reduced and additional properties inferred by adding to the proposed method one or several complimentary NMR measurements including T_1 inversion recovery (Carr & Purcell, 1954) and T_1-T_2 correlation (Peemoeller et al., 1981; Song et al., 2002) to name but a few.

5.3. Comparison to Single-Task ISW

The advantage of multi-task ISW over the single-task ISW is mainly due to the incorporated correlation matrix for which the inter-task correlations are adaptively inferred. While already significant, the resultant speed-up of calculation by a factor of 2–3 will be more profound if more costly simulations (e.g., T_1-T_2 responses) or more unknowns are involved, which may require more function evaluations. Since the simulations need to be carried out sequentially, speed-up in both cases translates directly to time savings. Moreover, additional information from old samples is provided for free, making this approach more attractive if more finished tasks are available. Practically, model selection in the multi-task ISW is more time-consuming. Recall that the multi-task ISW deals with $2Q + Q^2(Q + 1)/2 + DQ$ hyperparameters for LCM and $2Q + Q(Q + 1)/2 + D$ for ICM, whereas the single-task ISW only calls for $D + 3$. The additional hyperparameters add more flexibility to the model and are relatively easy to infer given Q times more observations. In the single-task ISW,

Table 9
Summary Statistics for the Identified Values for the Three Samples: BHG0, BHG1, and BHG2, Expressed Using Mean, Standard Deviation ($\mu \pm \sigma$), and the Best Set of Parameters for the Unique Partition (UP) Within Which the Lowest Fitness Value Is Identified

Sample	Statistics	$\rho_{2,q}/(\mu m/s)$	$\log_{10}(T_{2e,c}/s)$	$\log_{10}(D_{e,c}/[cm^2/s])$
BHG0	$\mu \pm \sigma$	8.721 ± 0.053	-2.235 ± 0.037	-5.318 ± 0.028
	Best	8.661	$-2.250 \ (5.623 \times 10^{-3})$	$-5.316 \ (4.831 \times 10^{-6})$
BHG1	$\mu \pm \sigma$	10.670 ± 0.420	-1.911 ± 0.033	-5.327 ± 0.102
	Best	11.121	$-1.921 \ (1.199 \times 10^{-2})$	$-5.255 \ (5.559 \times 10^{-6})$
BHG2	$\mu \pm \sigma$	10.965 ± 0.003	-1.897 ± 0.027	-5.043 ± 0.030
	Best	10.967	$-1.908 \ (1.236 \times 10^{-2})$	$-5.032 \ (9.290 \times 10^{-6})$

Note. The BHG0 data is from Li et al. (2021), BHG1 data is from Table 4 and BHG2 data is from Table 7. Values shown in brackets are the base-10 exponentiation following the best value identified for $\log_{10}(T_{2e,c}/s)$ and $\log_{10}(D_{e,c}/[cm^2/s])$. All physically valid UPs are involved in this calculation.

we adopted full Bayesian treatment of the model hyperparameters since the joint distribution is usually multi-modal in the starting stage when only a small number of observations are made. However, in the multi-task ISW case, the cost for Bayesian inference of such many hyperparameters is prohibitive and the strategy is less superior given an adequate number of observations in the completed tasks. As a result, we choose the maximum likelihood estimate of the hyperparameters, following Swersky et al. (2013); Bonilla et al. (2007). We demonstrated in Section 4 that for problems with moderate to high correlations, the ICM is superior to LCM due to its simplicity.

5.4. Decision on and Assessment of the Fitting

The choice of a reference for fitting (magnetization decay or T_2 distribution) is defined by the complexity of the latter. The problem stems from the nature of the T_2 kernel and the transformation of linearly acquired data into log-spaced solution components (so-called ILT transformation). Removing the ILT out of the optimization workflow undoubtedly simplifies it, but results in excessive errors especially in the short time T_2 components when present, which in turn affects the predicted amplitudes of long time components.

In our case the mode of the long-time peak and short-time peaks are around 750 and 50 ms, leading to 15 times more data points constraining the long-time component, so that in the decay-fitting scheme fitting the tail by modifying $\rho_{2,q}$ is almost 15 times as effective as fitting the first few hundred points by modifying $\log_{10} T_{2e,c}$ and $\log_{10} D_{e,c}$. What is worse, low sensitivities against the two unknowns in the decay-fitting scheme indicate that the solution is very unstable since it is difficult for the Bayesian optimizer to find their optimal value—solutions with similar values of the two unknowns are indistinguishable in terms of the L^2-norm of the fitting residual, leading to significantly more unwanted non-unique solution sets. By contrast, the distribution-fitting scheme does not suffer from these issues since by utilizing T_2 distributions it is automatically sensitive to both the short-time component and the long-time component regardless of how many data points are associated with the two peaks.

One observation of Figures 7–9 is that the minor peak (short relaxation times) is not fitted as well as the main peak (long relaxation times), which we conclude is due to the following reasons: (a) the fitting objective function we adopted here, that is, L^2-norm of the fitting residual, is sensitive to outliers or larger values since the squared error is very large, and (b) the smoothing effect applied to T_2 distributions is small, so that false features from random noise are explicit and cannot be reproduced from simulation. Figure 14 in Li et al. (2021) gives a clearer view that, when up to 16 T_2 distributions are inverted using the same λ_{opt} from decays measured under the same SNR, the variation (expressed in the standard deviation) of the minor peak is larger than that of the main peak. As a result, it is expected that fitting of the main peak is always better than the minor peak.

5.5. Limitation of the Transfer Learning Approach

In this study, we investigated the transfer learning approach between multiple core plugs from a long core of a larger formation with an overall similarity of higher than 70% from Table 7. The success of this approach is based on that (a) both functions are similarly sensitive to the unknown physical quantities, and (b) there are moderate overall inter-task similarities between source tasks and target tasks. For cores from different formations with decreased overall similarity, the approach might be less effective. Fortunately, for the general case where core samples with different mineralogy are sharing a subset of unknown quantities, for example, a sandstone with an additional clay introducing two unique unknowns, and a sandstone with feldspar as a standalone phase introducing one unique unknown, these issues can be addressed using a domain adaptation technique. Unknown parameters from the solution domain of both the source task and the target task will be mapped to a new (high-dimensional) solution domain using affine transformations (Min et al., 2021), where the transformation matrix is treated as hyperparameters that are inferred using maximum likelihood estimation during each iteration, giving rise to the potential for latent correlations between them. In extreme situations, the source and the target samples cored from lithologically different formations may share no unknown physical quantity in common. In this situation the introduced technique would behave equivalently to the single-task ISW, for example, little will fitting a carbonate benefit from fitting a sandstone.

6. Conclusions

A MT-ISW was introduced that incorporates transfer learning abilities into a Bayesian optimization framework. It expands on the single-task ISW for inverse problems arising from the identification of unknown physical quantities in the scope of NMR relaxation in porous media. The additional correlation matrix gradually captures the underlying correlations between tasks. A principled multi-modal search strategy comprising a gradient-based optimizer L-BFGS-B and the state-of-art social learning-particle swarm optimizer is applied to handle the severe ill-posedness of the objective function. We demonstrated our approach by studying the spatial variability of the three Bentheimer sandstones with both low and high task-similarities. Results show that in both scenarios the multi-task ISW finds a lower fitness residual faster than the single-task ISW, which proved significantly more effective than the single-task ISW. The scalability of the multi-task ISW was demonstrated by extending the workflow to more than one source task. We draw the following specific conclusions:

1. Incorporation of transfer learning abilities improves the prediction ability by using one-third to one-half of the function evaluations required by the single-task ISW attaining the same level of FV. Given the same function evaluations, the multi-task ISW usually achieves one-third to one-half lower FVs. Application of transfer learning using ICM will be significantly more effective when the source tasks and target task are correlated with 70% or higher inter-task correlation.
2. We adopted a multi-modal search strategy and the solution analysis to cater for the severe ill-posedness of the problem, through which the multi-task ISW is able to identify multiple solution sets for each sample that are visually indistinguishable by inspection of the relaxation spectra alone.
3. Extracted best estimates for surface relaxivity and effective clay relaxation time and diffusion coefficient of water show same-block variations of about 4%, while inter-block variations are about 28%, illustrating spatial variability quantitatively (see Table 9). Furthermore, the multi-task ISW is efficient and robust given this variation.

The multi-task learning approach was applied here to the extraction of intrinsic physical properties using a single objective function (simulating T_2 relaxation). It has direct applications to study the *spatial variability* of the geological and physical quantities using multiple core plugs from a long core of a larger formation, or to study the *temporal variability* of such quantities, for example, in the context of contaminant migration.

Expanding the approach to multiple measurements to constrain the solutions and thus reduce the number of non-unique solution sets should be straightforward and allows the simultaneous fitting of multiple NMR responses including higher-dimensional acquisition sequences. Applying the multi-task ISW to sets of T_1 and T_2 relaxation responses at higher magnetic field strengths, potentially including multiple T_2 measurements with different CPMG echo spacings, may increase the sensitivity to mineralogy and paramagnetic impurities due to the increased importance of internal magnetic fields.

Appendix A: Derivation of LCM

Give the data set $\mathcal{D}_{\text{obs}}$, the outputs are indeed expressed as the sum of the mean function and linear combinations of independent random functions, that is,

$$f_q(\mathbf{x}) = m_q + \sum_{s=1}^{S} a_{q,s}\, u_s(\mathbf{x}) \tag{A1}$$

where the latent functions $u_s(\mathbf{x})$ are independent samples from a specific Gaussian process (GP) following $\mathcal{N}(0, k_s(\mathbf{x}, \mathbf{x}'))$, $a_{q,s}$ is the scalar coefficient specifying relative contribution associating with each $u_s(\mathbf{x})$ for the qth task. Here we implement automatic relevance determination to the covariance function $k_s(\mathbf{x}, \mathbf{x}')$. Under independency assumption, the covariance between $u_s(\mathbf{x})$ is given by $\text{cov}\left[u_s(\mathbf{x}), u_{s'}(\mathbf{x}')\right] = k_s(\mathbf{x}, \mathbf{x}')$ if $s = s'$, and Equation A1 can be re-written by grouping together the latent functions $u_s(\mathbf{x})$ which share the same covariance, that is,

$$f_q(\mathbf{x}) = m_q + \sum_{s=1}^{S} \sum_{i=1}^{R_s} a_{q,s}^i u_s^i(\mathbf{x}). \tag{A2}$$

Equation A2 can be viewed as linear combinations of S groups of functions $u_s^i(\mathbf{x})$ within which the independent functions $u_s^i(\mathbf{x})$ share the same covariance. Under independency assumption, the covariance between any pair of $u_s^i(\mathbf{x})$ can expressed as

$$\text{cov}\left[f_q(\mathbf{x}) - m_q, f_{q'}\left(\mathbf{x}'\right) - m_{q'}\right] = \sum_{s=1}^{S} \sum_{i=1}^{R_s} a_{q,s}^i a_{q',s}^i k_s\left(\mathbf{x}, \mathbf{x}'\right), \tag{A3}$$

which exactly specifies the entries of $(\mathbf{K}\left(\mathbf{x}, \mathbf{x}'\right))_{q,q'}$. The multi-task kernel can be then expressed as

$$\left(\mathbf{K}\left(\mathbf{x}, \mathbf{x}'\right)\right)_{q,q'} = \sum_{s=1}^{S} b_{q,q'}^s k_s\left(\mathbf{x}, \mathbf{x}'\right) = \sum_{s=1}^{S} \mathbf{B}_s k_s\left(\mathbf{x}, \mathbf{x}'\right), \tag{A4}$$

where $b_{q,q'}^s = \sum_{i=1}^{R_s} a_{q,s}^i a_{q',s}^i$ specify entries of the coregionalization matrix $\mathbf{B}_s$.

Appendix B: Proposal of Promising Candidates for Exact Evaluation

B1. The Multi-Task Acquisition Function

In the case of the single-task inverse solution workflow (ISW), expected improvement, or expected improvement (EI) (Mockus, 1975), is one of the most effective criteria for candidate selection, balancing exploration and exploitation. For the multi-task ISW there are a variety of acquisition functions such as EI, entropy search (Hennig & Schuler, 2012), or max-value entropy search (Moss et al., 2021). Here we extend EI adopted in ISW to the multi-task case since it is continuous and analytically tractable, with *no* hyperparameter for manual tuning.

In the multi-task ISW, the EI acquisition function for task q ($a_{\text{EI},q}$) as a function of $\mathbf{x}^*$ is defined as the amount of improvement that the evaluation of a candidate is expected to induce from that task, that is,

$$a_{\text{EI},q}\left(\mathbf{x}^*\right) = \int_{-\infty}^{y_q\left(\mathbf{x}_q^-\right)} \left(y_q\left(\mathbf{x}_q^-\right) - u\right) p(u)\, \mathrm{d}u, \tag{B1}$$

where $\mathbf{x}_q^-$ and $y_q\left(\mathbf{x}_q^-\right)$ are the current best observed candidate and associated observation of the qth task, respectively, and the short-hand symbol $u = y_q^* | X, \mathbf{y}, \mathbf{x}^*, \boldsymbol{\theta}$ is given by Equation 9. Replacing u by Equation 9 yields

$$a_{\text{EI},q}\left(\mathbf{x}_m^* | X, \mathbf{y}, \boldsymbol{\theta}\right) = \left(y_q\left(\mathbf{x}_q^-\right) - \mu\left(\mathbf{x}^*\right)\right) \Phi\left(\frac{y_q\left(\mathbf{x}_q^-\right) - \mu\left(\mathbf{x}^*\right)}{\sigma\left(\mathbf{x}^*\right)}\right) \\ + \sigma\left(\mathbf{x}^*\right) \phi\left(\frac{y_q\left(\mathbf{x}_q^-\right) - \mu\left(\mathbf{x}^*\right)}{\sigma\left(\mathbf{x}^*\right)}\right), \tag{B2}$$

where $\phi(x)$ and $\Phi(x)$ denote the probability density function and cumulative density function of the standard normal distribution, respectively.

B2. Multi-Modal Maximization of EI

For the task q, we seek the solution of the following problem,

$$\text{maximize } a_{\text{EI},q}\left(\mathbf{x}^* | X, \mathbf{y}, \boldsymbol{\theta}_m\right), \text{ subject to } \mathbf{x}^* \in \mathcal{X}. \tag{B3}$$

We proposed the multi-modal optimization workflow in the (Li et al., 2021) using the single-task GP recovering multiple maxima of a_{EI}. Here we are using an identical approach except that the strategy of partitioning the solution space is only limited to the evaluated candidates associated with that task.

Data Availability Statement

The tomogram data and the NMR experimental data that support the findings of this study are made available online at the Digital Rocks Portal (project 425) under the ODC-BY 1.0 license (Li et al., 2022b). Raw and processed data for the complete optimization process are available as supporting information as Figures S1–S8

and Tables S1–S4. The numerical methods for the calculation of the NMR responses and the respective algorithms for the more basic aspects of the inverse solution workflow are introduced in detail in references (Li et al., 2021, 2022a). The NMR responses were calculated with an in-house software, namely the NMR package of morphy, version 1.9.40.

References

Afrough, A., Marica, F., MacMillan, B., & Balcom, B. J. (2021). Pore-size measurement from eigenvalues of magnetic resonance relaxation. *Physical Review Applied*, *16*(3), 034040. https://doi.org/10.1103/PhysRevApplied.16.034040

Álvarez, M. A., Rosasco, L., & Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, *4*(3), 195–266. https://doi.org/10.1561/2200000036

Anand, V., Hirasaki, G. J., & Fleury, M. (2008). NMR diffusional coupling: Effects of temperature and clay distribution. *Petrophysics*, *49*(04), 362–372.

Arns, C. H. (2004). A comparison of pore size distributions derived by NMR and X-ray-CT techniques. *Physica A: Statistical Mechanics and its Applications*, *339*(1), 159–165. https://doi.org/10.1016/j.physa.2004.03.033

Arns, C. H., AlGhamdi, T., & Arns, J.-Y. (2011). Numerical analysis of nuclear magnetic resonance relaxation–diffusion responses of sedimentary rock. *New Journal of Physics*, *13*(1), 015004. https://doi.org/10.1088/1367-2630/13/1/015004

Arns, C. H., Knackstedt, M. A., & Martys, N. S. (2005). Cross-property correlations and permeability estimation in sandstone. *Physical Review E*, *72*(4), 046304. https://doi.org/10.1103/PhysRevE.72.046304

Ausbrooks, R., Hurley, N. F., May, A., & Neese, D. G. (1999). Pore-size distributions in vuggy carbonates from core images, NMR, and capillary pressure. In *Paper Presented at the SPE Annual Technical Conference and Exhibition. Dallas, Texas*. https://doi.org/10.2118/56506-MS

Bertsekas, D. P. (1982). Projected Newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization*, *20*(2), 221–246. https://doi.org/10.1137/0320018

Bilicz, S., Lambert, M., Gyimothy, S., & Pavo, J. (2012). Solution of inverse problems in nondestructive testing by a kriging-based surrogate model. *IEEE Transactions on Magnetics*, *48*(2), 495–498. https://doi.org/10.1109/TMAG.2011.2172196

Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. (2007). Multi-task Gaussian process prediction. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS), Vancouver, British Columbia, Canada* (pp. 153–160). https://doi.org/10.5555/2981562.2981582

Borgia, G. C., Brown, R. J. S., & Fantazzini, P. (1995). Scaling of spin-echo amplitudes with frequency, diffusion coefficient, pore size, and susceptibility difference for the NMR of fluids in porous media and biological tissues. *Physical Review E*, *51*(3), 2104–2114. https://doi.org/10.1103/PhysRevE.51.2104

Bruggeman, D. A. G. (1935). Berechnung verschiedener physikalischer konstanten von heterogenen substanzen. i. dielektrizitätskonstanten und leitfähigkeiten der mischkörper aus isotropen substanzen. *Annalen der Physik*, *416*(7), 636–664. https://doi.org/10.1002/andp.19354160705

Cai, C., Wang, S., Xu, Y., Zhang, W., Tang, K., Ouyang, Q., et al. (2020). Transfer learning for drug discovery. *Journal of Medicinal Chemistry*, *63*(16), 8683–8694. https://doi.org/10.1021/acs.jmedchem.9b02147

Carr, H. Y., & Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, *94*(3), 630–638. https://doi.org/10.1103/PhysRev.94.630

Coates, G., Peveraro, R., Hardwick, A., & Roberts, D. (1991). The magnetic resonance imaging log characterized by comparison with petrophysical properties and laboratory core data. In *Paper Presented at the SPE Annual Technical Conference and Exhibition. Dallas, Texas*. https://doi.org/10.2118/22723-MS

Connolly, P. R., Yan, W., Zhang, D., Mahmoud, M., Verrall, M., Lebedev, M., et al. (2019). Simulation and experimental measurements of internal magnetic field gradients and NMR transverse relaxation times (T_2) in sandstone rocks. *Journal of Petroleum Science and Engineering*, *175*, 985–997. https://doi.org/10.1016/j.petrol.2019.01.036

Costabel, S., & Yaramanci, U. (2013). Estimation of water retention parameters from nuclear magnetic resonance relaxation time distributions. *Water Resources Research*, *49*(4), 2068–2079. https://doi.org/10.1002/wrcr.20207

Cui, Y., Shikhov, I., Li, R., Liu, S., & Arns, C. H. (2021). A numerical study of field strength and clay morphology impact on NMR transverse relaxation in sandstones. *Journal of Petroleum Science and Engineering*, *202*(7), 108521. https://doi.org/10.1016/j.petrol.2021.108521

Cui, Y., Shikhov, I., Liu, S., & Arns, C. H. (2022). NMR relaxation modelling in porous media with dual-scale-resolved internal magnetic fields. *Transport in Porous Media*, *142*(3), 453–474. https://doi.org/10.1007/s11242-022-01752-0

Cunha, A., Pochet, A., Lopes, H., & Gattass, M. (2020). Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. *Computers & Geosciences*, *135*, 104344. https://doi.org/10.1016/j.cageo.2019.104344

Dabat, T., Porion, P., Hubert, F., Paineau, E., Dazas, B., Grégoire, B., et al. (2020). Influence of preferred orientation of clay particles on the diffusion of water in kaolinite porous media at constant porosity. *Applied Clay Science*, *184*, 105354. https://doi.org/10.1016/j.clay.2019.105354

de Lima, R. P., Bonar, A., Coronado, D. D., Marfurt, K., & Nicholson, C. (2019). Deep convolutional neural networks as a geological image classification tool. *The Sedimentary Record*, *17*(2), 4–9. https://doi.org/10.2110/sedred.2019.2.4

Deng, Z., Tutunnikov, I., Averbukh, I. S., Thachuk, M., & Krems, R. (2020). Bayesian optimization for inverse problems in time-dependent quantum dynamics. *The Journal of Chemical Physics*, *153*(16), 164111. https://doi.org/10.1063/5.0015896

Dlubac, K., Knight, R., Song, Y.-Q., Bachman, N., Grau, B., Cannia, J., & Williams, J. (2013). Use of NMR logging to obtain estimates of hydraulic conductivity in the High Plains aquifer, Nebraska, USA. *Water Resources Research*, *49*(4), 1871–1886. https://doi.org/10.1002/wrcr.20151

Duschl, M., Galvosas, P., Brox, T. I., Pohlmeier, A., & Vereecken, H. (2015). In situ determination of surface relaxivities for unconsolidated sediments. *Water Resources Research*, *51*(8), 6549–6563. https://doi.org/10.1002/2014WR016574

Fleury, M., Gautier, S., Norrant, F., & Kohler, E. (2011). Characterization of nanoporous systems with low field NMR: Application to kaolinite and smectite clays, Austin, Texas, SCA2011-23.

Gavalas, G., Shah, P., & Seinfeld, J. H. (1976). Reservoir history matching by Bayesian estimation. *Society of Petroleum Engineers Journal*, *16*(06), 337–350. https://doi.org/10.2118/5740-PA

Gelfand, A. E., Schmidt, A. M., Banerjee, S., & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, *13*(2), 263–312. https://doi.org/10.1007/BF02595775

Goovaerts, P. (1997). *Geostatistics for natural resource evaluation*. Oxford University Press.

Goulard, M., & Voltz, M. (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, *24*(3), 269–286. https://doi.org/10.1007/BF00893750

Hansen, P. C. (1992). Numerical tools for analysis and solution of Fredholm integral equations of the first kind. *Inverse Problems*, *8*(6), 849–872. https://doi.org/10.1088/0266-5611/8/6/005

Hennig, P., & Schuler, C. J. (2012). Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, *13*, 1809–1837. https://doi.org/10.5555/2188385.2343701

Holland, D. J., Mitchell, J., Blake, A., & Gladden, L. F. (2013). Grain sizing in porous media using Bayesian magnetic resonance. *Physical Review Letters*, *110*(1), 018001. https://doi.org/10.1103/PhysRevLetter110.018001

Howard, J. J., & Kenyon, W. E. (1992). Determination of pore size distribution in sedimentary rocks by proton nuclear magnetic resonance. *Marine and Petroleum Geology*, *9*(2), 139–145. https://doi.org/10.1016/0264-8172(92)90086-T

Hu, Q., Zhang, R., & Zhou, Y. (2016). Transfer learning for short-term wind speed prediction with deep neural networks. *Renewable Energy*, *85*, 83–95. https://doi.org/10.1016/j.renene.2015.06.034

Hürlimann, M., Helmer, K., Latour, L., & Sotak, C. (1994). Restricted diffusion in sedimentary rocks. Determination of surface-area-to-volume ratio and surface relaxivity. *Journal of Magnetic Resonance, Series A*, *111*(2), 169–178. https://doi.org/10.1006/jmra.1994.1243

Imrie, F., Bradley, A. R., van der Schaar, M., & Deane, C. M. (2018). Protein family-specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data. *Journal of Chemical Information and Modeling*, *58*(11), 2319–2330. https://doi.org/10.1021/acs.jcim.8b00350

Jaeger, F., Bowe, S., Van As, H., & Schaumann, G. E. (2009). Evaluation of ^{1}H NMR relaxometry for the assessment of pore-size distribution in soil samples. *European Journal of Soil Science*, *60*(6), 1052–1064. https://doi.org/10.1111/j.1365-2389.2009.01192.x

Kalinin, S. V., Ziatdinov, M., & Vasudevan, R. K. (2020). Guided search for desired functional responses via Bayesian optimization of generative model: Hysteresis loop shape engineering in ferroelectrics. *Journal of Applied Physics*, *128*(2), 024102. https://doi.org/10.1063/5.0011917

Kenyon, W. E., Day, P. I., Straley, C., & Willemsen, J. F. (1988). A three-part study of NMR longitudinal relaxation properties of water-saturated sandstones. *SPE Formation Evaluation*, *3*(03), 622–636. https://doi.org/10.2118/15643-PA

Kitanidis, P. K., & Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*, *19*(3), 677–690. https://doi.org/10.1029/WR019i003p00677

Kleinberg, R. (1996). Utility of NMR T_2 distributions, connection with capillary pressure, clay effect, and determination of the surface relaxivity parameter ρ_2. *Magnetic Resonance Imaging*, *14*(7), 761–767. https://doi.org/10.1016/S0730-725X(96)00161-0

Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems*. SIAM. https://doi.org/10.1137/1.9781611971217

Li, R., Shikhov, I., & Arns, C. (2022a). A Bayesian optimization approach to the simultaneous extraction of intrinsic physical parameters from t_1 and t_2 relaxation responses. *SPE Journal*, *1*, 1–23. https://doi.org/10.2118/210563-PA

Li, R., Shikhov, I., & Arns, C. (2022b). Bentheimer sandstone for extraction of phase-specific petrophysical quantities using low-field NMR [Dataset]. Digital Rocks Portal. https://doi.org/10.17612/1J6K-SH07

Li, R., Shikhov, I., & Arns, C. H. (2021). Solving multiphysics, multiparameter, multimodal inverse problems: An application to NMR relaxation in porous media. *Physical Review Applied*, *15*(5), 054003. https://doi.org/10.1103/PhysRevApplied.15.054003

Liu, H., Nogueira d'Eurydice, M., Obruchkov, S., & Galvosas, P. (2014). Determining pore length scales and pore surface relaxivity of rock cores by internal magnetic fields modulation at 2MHz NMR. *Journal of Magnetic Resonance*, *246*, 110–118. https://doi.org/10.1016/j.jmr.2014.07.005

Maliva, R. G., Clayton, E. A., & Missimer, T. M. (2009). Application of advanced borehole geophysical logging to managed aquifer recharge investigations. *Hydrogeology Journal*, *17*(6), 1547–1556. https://doi.org/10.1007/s10040-009-0437-z

Matteson, A., Tomanic, J. P., Herron, M. M., Allen, D. F., & Kenyon, W. E. (2000). NMR relaxation of clay/brine mixtures. *SPE Reservoir Evaluation and Engineering*, *3*(05), 408–413. https://doi.org/10.2118/66185-PA

Meiboom, S., & Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, *29*(8), 688–691. https://doi.org/10.1063/1.1716296

Min, A. T. W., Gupta, A., & Ong, Y.-S. (2021). Generalizing transfer Bayesian optimization to source-target heterogeneity. *IEEE Transactions on Automation Science and Engineering*, *18*(4), 1754–1765. https://doi.org/10.1109/TASE.2020.3017644

Min, A. T. W., Sagarna, R., Gupta, A., Ong, Y.-S., & Goh, C. K. (2017). Knowledge transfer through machine learning in aircraft design. *IEEE Computational Intelligence Magazine*, *12*(4), 48–60. https://doi.org/10.1109/MCI.2017.2742781

Mitchell, J., Chandrasekera, T., & Gladden, L. (2012). Numerical estimation of relaxation and diffusion distributions in two dimensions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *62*, 34–50. https://doi.org/10.1016/j.pnmrs.2011.07.002

Mockus, J. (1975). On Bayesian methods for seeking the extremum. In *Proceedings of the Optimization Techniques IFIP Technical Conference, Novosibirsk, Russia* (pp. 400–404). https://doi.org/10.1007/3-540-07165-2_55

Mohnke, O. (2014). Jointly deriving NMR surface relaxivity and pore size distributions by NMR relaxation experiments on partially desaturated rocks. *Water Resources Research*, *50*(6), 5309–5321. https://doi.org/10.1002/2014WR015282

Moss, H. B., Leslie, D. S., & Rayson, P. (2021). MUMBO: MUlti-task max-value Bayesian optimization. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases, Ghent, Belgium* (pp. 447–462). https://doi.org/10.1007/978-3-030-67664-3_27

Mu, D., Liu, Z.-S., Huang, C., & Djilali, N. (2007). Prediction of the effective diffusion coefficient in random porous media using the finite element method. *Journal of Porous Materials*, *14*(1), 49–54. https://doi.org/10.1007/s10934-006-9007-0

Muncaci, S., & Ardelean, I. (2013). The influence of the magnetic impurity content on the pore size distribution determination via the DDIF technique. *Applied Magnetic Resonance*, *44*(3), 365–373. https://doi.org/10.1007/s00723-012-0382-1

Osterman, G., Sugand, M., Keating, K., Binley, A., & Slater, L. (2019). Effect of clay content and distribution on hydraulic and geophysical properties of synthetic sand-clay mixtures. *Geophysics*, *84*(4), E239–E253. https://doi.org/10.1190/geo2018-0387.1

Peemoeller, H., Shenoy, R., & Pintar, M. (1981). Two-dimensional NMR time evolution correlation spectroscopy in wet lysozyme. *Journal of Magnetic Resonance*, *45*(2), 193–204. https://doi.org/10.1016/0022-2364(81)90116-5

Pelletier, B., Dutilleul, P., Larocque, G., & Fyles, J. W. (2004). Fitting the linear model of coregionalization by generalized least squares. *Mathematical Geology*, *36*(3), 323–343. https://doi.org/10.1023/B:MATG.0000028440.29965.2d

Pohlmeier, A., Haber-Pohlmeier, S., & Stapf, S. (2009). A fast field cycling nuclear magnetic resonance relaxometry study of natural soils. *Vadose Zone Journal*, *8*(3), 735–742. https://doi.org/10.2136/vzj2008.0030

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian processes for machine learning (adaptive computation and machine learning)*. The MIT Press.

Sánchez, F. G., Van Loon, L. R., Gimmi, T., Jakob, A., Glaus, M. A., & Diamond, L. W. (2008). Self-diffusion of water and its dependence on temperature and ionic strength in highly compacted montmorillonite, illite and kaolinite. *Applied Geochemistry*, *23*(12), 3840–3851. https://doi.org/10.1016/j.apgeochem.2008.08.008

Schneider, P.-I., Hammerschmidt, M., Zschiedrich, L., & Burger, S. (2019). Using Gaussian process regression for efficient parameter reconstruction. In *Paper Presented at Metrology, Inspection, and Process Control for Microlithography XXXIII, San Jose, California* (Vol. 10959).1095911. https://doi.org/10.1117/12.2513268

Sen, P. N., Straley, C., Kenyon, W. E., & Whittingham, M. S. (1990). Surface-to-volume ratio, charge density, nuclear magnetic relaxation, and permeability in clay-bearing sandstones. *Geophysics*, *55*(1), 61–69. https://doi.org/10.1190/1.1442772

Sheppard, A. P., Sok, R. M., & Averdunk, H. (2004). Techniques for image enhancement and segmentation of tomographic images of porous materials. *Physica A*, *339*(1–2), 145–151. https://doi.org/10.1016/j.physa.2004.03.057

Shikhov, I., & Arns, C. H. (2015). Evaluation of capillary pressure methods via digital rock simulations. *Transport in Porous Media*, *107*(2), 623–640. https://doi.org/10.1007/s11242-015-0459-z

Shikhov, I., d'Eurydice, M. N., Arns, J.-Y., & Arns, C. H. (2017). An experimental and numerical study of relative permeability estimates using spatially resolved T_1-z NMR. *Transport in Porous Media*, *118*(2), 225–250. https://doi.org/10.1007/s11242-017-0855-7

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada* (Vol. 2, pp. 2951–2959). https://doi.org/10.5555/2999325.2999464

Song, Y.-Q., Venkataramanan, L., Hürlimann, M., Flaum, M., Frulla, P., & Straley, C. (2002). T_1-T_2 correlation spectra obtained using a fast two-dimensional Laplace inversion. *Journal of Magnetic Resonance*, *154*(2), 261–268. https://doi.org/10.1006/jmre.2001.2474

Stejskal, E., & Schaefer, J. (1974). Data routing in quadrature FT NMR. *Journal of Magnetic Resonance*, *13*(2), 249–251. https://doi.org/10.1016/0022-2364(74)90014-6

Stingaciu, L. R., Pohlmeier, A., Blümler, P., Weihermüller, L., van Dusschoten, D., Stapf, S., & Vereecken, H. (2009). Characterization of unsaturated porous media by high-field and low-field NMR relaxometry. *Water Resources Research*, *45*(8), 1–11. https://doi.org/10.1029/2008WR007459

Stingaciu, L. R., Weihermüller, L., Haber-Pohlmeier, S., Stapf, S., Vereecken, H., & Pohlmeier, A. (2010). Determination of pore size distribution and hydraulic properties using nuclear magnetic resonance relaxometry: A comparative study of laboratory methods. *Water Resources Research*, *46*(11), W11510. https://doi.org/10.1029/2009WR008686

Swersky, K., Snoek, J., & Adams, R. P. (2013). Multi-task bayesian optimization. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada* (Vol. 2, pp. 2004–2012). https://doi.org/10.5555/2999792.2999836

Toumelin, E., Torres-Verdín, C., Chen, S., & Fischer, D. (2003). Reconciling NMR measurements and numerical simulations: Assessment of temperature and diffusive coupling effects on two-phase carbonate samples. *Petrophysics: The SPWLA Journal of Formation Evaluation and Reservoir Description*, *44*(02), 91–107.

Vargas-Hernández, R., Guan, Y., Zhang, D., & Krems, R. (2019). Bayesian optimization for the inverse scattering problem in quantum reaction dynamics. *New Journal of Physics*, *21*(2), 022001. https://doi.org/10.1088/1367-2630/ab0611

Willard, J. D., Read, J. S., Appling, A. P., Oliver, S. K., Jia, X., & Kumar, V. (2021). Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. *Water Resources Research*, *57*(7). https://doi.org/10.1029/2021wr029579

Zhang, Y., Xiao, L., Liao, G., & Song, Y.-Q. (2016). Direct correlation of diffusion and pore size distributions with low field NMR. *Journal of Magnetic Resonance*, *269*, 196–202. https://doi.org/10.1016/j.jmr.2016.06.013

Zhou, Y., Wang, T., & Peng, X. (2020). Paper presented at 2020 IEEE Congress on evolutionary computation (CEC). In *MFEA-IG: A multi-task algorithm for mobile agents path planning, Glasgow, United Kingdom* (pp. 1–7). https://doi.org/10.1109/CEC48606.2020.9185906