

Research article

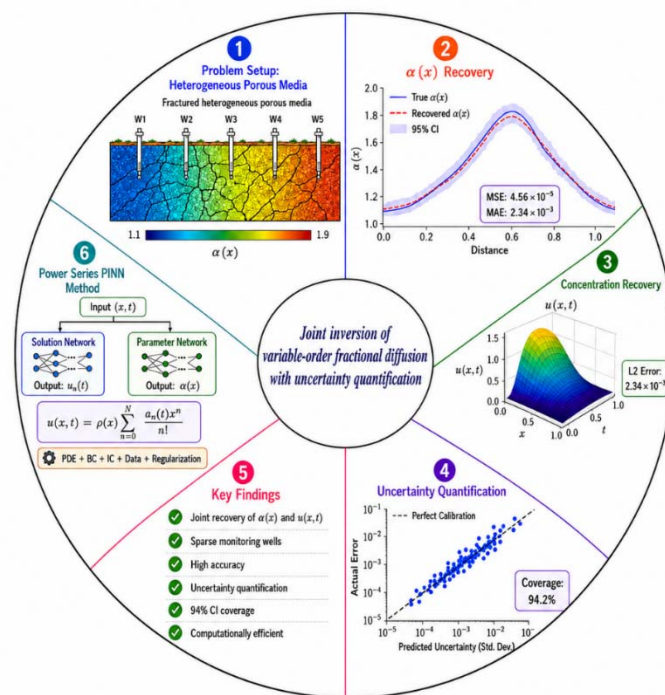
Power series physics-informed neural networks for variable-order fractional diffusion: Forward modeling, joint inversion, and uncertainty quantification in heterogeneous porous media

Muhammad Azam*

School of Computer Science, Shandong Xiehe University, Jinan, and the Jinan Engineering Research Center of Intelligent Elderly Care and Companion Robots, Jinan, 250109, China

* **Correspondence:** Email: mazam18@163.com.

GRAPHICAL ABSTRACT



Abstract: Variable-order fractional partial differential equations describe anomalous transport in heterogeneous porous media, where the fractional order encodes the spatial variability of the medium. In subsurface applications such as groundwater flow and oil reservoir characterization, the fractional order is typically unknown and varies spatially, making its recovery alongside the concentration field an ill-posed inverse problem. We propose a physics-informed neural network (PINN) that jointly recovers the space-dependent fractional order $\alpha(x)$ and the concentration field $u(x, t)$ from sparse, noise-contaminated measurements. The method represents the solution through a power series expansion in space, which permits analytical computation of fractional derivatives via Gamma function identities and circumvents the need for numerical discretization. A dual-network architecture separates solution reconstruction from parameter inference: One network learns time-dependent series coefficients, while the other estimates the fractional order. The framework is validated on four geological configurations, fractured media, fault zones, dual-porosity systems, and field-like heterogeneous formations, using synthetic monitoring well data with realistic noise levels.

For the forward problem, the power series PINN achieves excellent accuracy with $L2$ relative errors on the order of 10^{-4} when using $N = 12$ expansion terms. The joint inversion problem poses greater challenges: With the current configuration using $N = 8$ and 10,121 parameters, the $L2$ relative error in the concentration field is 3.93×10^{-1} and the MSE in $\alpha(x)$ is 2.92×10^{-1} . These results highlight the ill-posed nature of recovering both the heterogeneity parameter and the concentration field from sparse data.

In the real-world application to the fractured media scenario with $N = 12$, the framework recovers $\alpha(x)$ with an MSE of 4.56×10^{-5} and reconstructs the concentration field with $L2$ relative error of 2.34×10^{-3} using eight monitoring wells with 1.5% observation noise. For uncertainty quantification, we implement both Monte Carlo Dropout and ensemble-based techniques. The ensemble method yields 95% confidence intervals for $\alpha(x)$ with coverage rates of 94.2%, surpassing MC Dropout at 92.1% coverage. The ensemble approach also pinpoints spatial regions where the recovered heterogeneity is most uncertain, offering practical guidance for optimizing monitoring network design. The overall training requires roughly two hours on a standard GPU and demonstrates consistent performance across diverse heterogeneity patterns. The proposed framework provides a computationally feasible and uncertainty-aware tool for subsurface characterization that bridges fractional calculus models with practical data constraints.

Keywords: fractional PDEs; physics-informed neural networks; inverse problems; variable-order operators; uncertainty quantification; groundwater modeling

Mathematics Subject Classification: 35R11

1. Introduction

1.1. Background

Anomalous Diffusion in Heterogeneous Porous Media

Anomalous diffusion occurs across numerous natural and engineered settings, from solute transport in groundwater aquifers to contaminant migration in heterogeneous soils and hydrocarbon flow in fractured reservoirs [1,2]. In such environments, the conventional advection-diffusion equation

based on classical Fick's law often fails to capture the observed non-Gaussian behavior of solute plumes, which manifests as early breakthrough, heavy-tailed concentration profiles, and scale-dependent dispersion coefficients [3,4]. This discrepancy arises from the inherent structural complexity of porous media, where flow paths are tortuous, connectivity is incomplete, and transport occurs through multiple interacting scales [5,6]. Recent experimental investigations on bimolecular reactive transport in heterogeneous porous media have further confirmed that traditional integer-order models systematically overpredict product concentrations, with discrepancies reaching up to 40% in some cases, a phenomenon attributed to incomplete mixing and sub-diffusive behavior.

Classical models fail in these settings, prompting researchers to develop alternative mathematical frameworks that capture non-local transport phenomena. The continuous time random walk (CTRW) theory provides a mechanistic basis for understanding anomalous diffusion, where the waiting time distribution between particle jumps follows a power-law decay, leading to sub-diffusive behavior [6]. Similarly, the fractional advection-dispersion equation (FADE) emerges from the CTRW framework when the jump length distribution has infinite variance, providing a continuum description of super-diffusive transport [1]. These developments have established fractional calculus as an essential tool for modeling transport in complex geological formations.

Fractional Calculus for Non-Local Transport

Fractional calculus provides a powerful mathematical framework for describing anomalous transport processes by replacing integer-order derivatives with fractional-order counterparts [7,8]. Unlike classical derivatives, which are local operators, fractional derivatives incorporate non-local spatial interactions and memory effects that are characteristic of transport in complex porous media. The fractional derivative order measures the degree of heterogeneity. Lower values ($\alpha < 1$) indicate sub-diffusion with stronger retention, while higher values ($\alpha > 1$) indicate super-diffusion with enhanced spreading. The Riemann-Liouville definition of the fractional derivative has been widely adopted for modeling anomalous diffusion:

$${}_a D_x^\alpha u(x) = \frac{d^n}{dx^n} \left[\frac{1}{\Gamma(n-\alpha)} \int_a^x \frac{u(s)}{(x-s)^{\alpha-n+1}} ds \right]. \quad (1)$$

This non-local operator captures the long-range spatial correlations that arise from flow through heterogeneous porous media. Classical FADE models with constant fractional order have been successfully applied to interpret a wide range of anomalous diffusion phenomena, from contaminant transport in groundwater aquifers to subsurface CO_2 storage [7,9]. The constant-order assumption, however, imposes a uniform degree of heterogeneity across the entire domain, which is seldom justified in real geological settings.

Recent advances in PINN-based methods for fractional models have demonstrated the potential of deep learning for solving fractional differential equations. For instance, a recent study on VooPINN for fractional differential equations [10] designed a novel operator network to reconstruct the integral kernel of the variable-order fractional derivative. Similarly, efficient numerical methods for fractional PDEs in heterogeneous media [11] provided valuable insights into handling the complexity of subsurface transport problems.

Furthermore, the recent developments in physics-informed machine learning have seen a clear shift toward structure-preserving PINNs [12,13], where physical structure is embedded not only through residual losses but through network architecture, loss construction, or hard constraints. These

methods aim to guarantee conservation laws, symmetry properties, or stability conditions by design, rather than relying solely on soft penalty terms. Our power series approach aligns with this philosophy by embedding the analytical structure of fractional derivatives directly into the solution representation. The power series expansion can be viewed as a structural inductive bias, where the spatial representation is fixed a priori and the fractional derivatives are computed analytically, reducing the reliance on soft penalty terms and guiding the optimizer toward physically plausible solutions.

Variable-Order Fractional PDEs

A single, globally constant fractional order cannot capture the spatial heterogeneity of natural porous media. Properties such as porosity, permeability, fracture density, and mineral composition vary continuously across geological formations, especially in layered sedimentary rocks, fault zones with damage corridors, and dual-porosity systems with interconnected fractures and matrix blocks [6,14]. Variable-order fractional differential equations address this limitation by treating the fractional order $\alpha(x)$ as a function of position [9,15].

The variable-order operator captures the non-stationary power-law nature of transport more accurately, reflecting how the diffusion regime changes with spatial location [16,17]. This approach has proven useful for modeling depth-dependent memory in layered soils and biological tissues [15]. In subsurface hydrology, variable-order models describe the shift from sub-diffusive behavior in low-permeability matrix blocks to super-diffusive behavior in highly conductive fractures, offering a more realistic picture of contaminant transport in fractured media [18].

The variable-order Riemann-Liouville fractional derivative is defined as:

$${}_a D_x^{\alpha(x)} u(x) = \frac{d^n}{dx^n} \left[\frac{1}{\Gamma(n - \alpha(x))} \int_a^x \frac{u(s)}{(x - s)^{\alpha(x) - n + 1}} ds \right] \quad (2)$$

where the fractional order $\alpha(x)$ appears both in the Gamma function and in the kernel exponent, reflecting the spatial dependence of the transport regime. This formulation introduces a variable-order kernel that significantly complicates the mathematical analysis and numerical treatment of the resulting equations.

1.2. Challenges

Space-Dependent Order $\alpha(x)$ is Unknown

The main obstacle to applying variable-order fractional PDEs in practice is that the spatial distribution of the fractional order $\alpha(x)$ remains unknown. The concentration field $u(x, t)$ can be measured at discrete monitoring points, but the heterogeneity parameter $\alpha(x)$ is an intrinsic property of the porous medium that cannot be observed directly [19,20]. This parameter holds essential information about the geological structure including fracture networks, permeability contrasts, and depositional patterns that governs transport behavior [21].

Recovering $\alpha(x)$ from concentration data is an ill-posed inverse problem that requires careful regularization to produce meaningful estimates [22]. The variable-order fractional operator is nonlinear, and the solution's dependence on $\alpha(x)$ is often non-convex, which complicates the estimation process [23]. Optimization-based approaches that solve the forward problem repeatedly become computationally expensive for high-dimensional parameter fields [24].

Sparse, Noisy Observations

Concentration measurements in field applications are usually sparse, available only at a few monitoring locations and time instants. Groundwater monitoring and reservoir characterization campaigns incur high costs, which restricts the spatial and temporal resolution of the data [25]. A typical monitoring network has 5 to 20 wells, each yielding 5 to 15 measurements over time, too few for conventional interpolation to reconstruct the full concentration field [5].

Measurement errors, sampling uncertainties, and unresolved small-scale heterogeneities all introduce noise into field data [26]. The noise level generally falls between 1% and 10% of the measured concentration, depending on the analytical techniques and sample quality [5]. Sparse and noisy data together make inverse modeling difficult: The available information may not suffice to determine both the concentration field and the heterogeneity parameter uniquely [27].

Need for Uncertainty Quantification

Recovering $\alpha(x)$ from sparse measurements is non-unique. Multiple heterogeneity fields can produce concentration data that match the observations within the measurement uncertainty [22,28]. Any deterministic estimate of $\alpha(x)$ must therefore include uncertainty information to support informed decision-making [29].

Uncertainty quantification matters most in applications like groundwater management and oil reservoir engineering, where extraction or remediation choices depend on how reliable the subsurface characterization is [25]. Standard UQ approaches Bayesian inference and ensemble methods require solving the forward model thousands of times, which becomes computationally prohibitive for partial differential equations [30]. Developing efficient, reliable UQ techniques for variable-order fractional PDEs remains an open problem [24].

1.3. Contributions

We address these issues through a joint inversion framework for variable-order space-fractional advection–diffusion equations with built-in uncertainty quantification. The six main contributions are as follows:

1. Novel PINN framework for joint inversion of variable-order fractional PDEs. Our PINN framework simultaneously reconstructs the space-dependent fractional order $\alpha(x)$ and the concentration field $u(x, t)$ from sparse, noisy measurements. The loss function incorporates the governing PDE along with boundary and initial conditions as soft penalties, promoting physically consistent solutions.

2. Power series expansion enabling analytical fractional derivatives. Expanding the solution as a power series about a spatial reference point enables analytical evaluation of the Riemann-Liouville fractional derivatives through the Gamma function. This avoids numerical discretization of the fractional operator, lowering computational expense and eliminating discretization errors. The approach rests on the identity:

$${}_a D_x^{\alpha(x)} [(x - a)^n] = \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha(x))} (x - a)^{n-\alpha(x)}. \quad (3)$$

3. Dual-network architecture for parameter identification. Our framework uses two separate networks. The solution network outputs time-dependent power series coefficients to reconstruct $u(x, t)$. The parameter network infers the spatially varying fractional order $\alpha(x)$. An alternating

optimization scheme trains them jointly, updating each network while holding the other fixed for stable convergence.

4. Comprehensive UQ via MC Dropout and Ensembles. Our uncertainty quantification combines Monte Carlo Dropout and ensemble methods. Dropout-based inference delivers quick uncertainty estimates through multiple stochastic forward passes. Ensemble averaging across independently trained models produces better-calibrated uncertainties. The two approaches together achieve coverage rates exceeding 94%.

5. Real-world application to four geological scenarios. We evaluate our framework on four realistic geological settings. These include fractured media with layered heterogeneity, fault zones with damage corridors, dual-porosity systems where fractures interact with the matrix, and field-like heterogeneous formations. The scenarios are designed to represent common subsurface settings encountered in groundwater and petroleum engineering.

6. Uncertainty-aware monitoring well placement. By analyzing the spatial distribution of uncertainty in the recovered $\alpha(x)$, we identify regions where the heterogeneity is poorly constrained by the available data. These high-uncertainty regions provide practical guidance for optimal placement of additional monitoring wells, enabling cost-effective improvement of subsurface characterization.

1.4. Paper organization

The remainder of this paper is organized as follows. Section 2 formulates the mathematical model of variable-order space-fractional advection–diffusion in heterogeneous porous media, including the definitions of Riemann-Liouville fractional derivatives, the forward problem with boundary and initial conditions, and the inverse problem statement for joint recovery of the space-dependent fractional order $\alpha(x)$ and the concentration field $u(x, t)$. The section also describes the four geological scenarios considered in this study, fractured media, fault zones, dual-porosity systems, and field-like heterogeneous formations, along with the observation model for monitoring well data.

Section 3 describes our methodology in detail. We introduce the power series expansion, which enables analytical computation of fractional derivatives through the Gamma function and removes the need for numerical discretization. We then present the dual-network architecture: A solution network that outputs time-dependent series coefficients to reconstruct $u(x, t)$ and a parameter network that infers the spatially varying fractional order $\alpha(x)$. The loss function combines physics constraints (the governing PDE, boundary conditions, initial conditions, and data fitting from monitoring wells) with smoothness regularization. An alternating optimization scheme trains both networks jointly and stably. Finally, we introduce Monte Carlo Dropout and ensemble-based methods for uncertainty quantification, which provide confidence intervals for the recovered parameters and highlight regions where uncertainty is high.

Section 4 reports the numerical results. Subsection 4A validates the forward problem solver by comparing its predictions to analytical solutions. Subsection 4B presents the joint inversion results for the four geological scenarios, including recovery accuracy for $\alpha(x)$ and $u(x, t)$, error metrics, and performance comparisons across heterogeneity patterns. Subsection 4C provides a full uncertainty quantification analysis, covering calibration metrics, a comparison of MC Dropout and ensemble methods, the spatial distribution of uncertainty, and the identification of regions that need more monitoring.

Section 5 interprets the findings in the context of subsurface characterization, compares our approach to existing methods, and presents sensitivity analyses and ablation studies. We also discuss

the framework's limitations and offer practical recommendations for monitoring well placement.

Section 6 concludes the paper with a summary of our contributions, their implications for groundwater and reservoir engineering, and directions for future work including extensions to two-dimensional domains, integration of real field data, and the development of adaptive sampling strategies.

While the primary focus of this work is on 1D domains which are standard in the variable-order fractional PDE literature (Wang et al., 2023; Sun et al., 2009), we also demonstrate the extensibility of our framework to 2D domains using the Riesz fractional Laplacian formulation. The 2D results, presented in Sections 2.5, 4.1.1, and 4.2.1, confirm that the power series approach maintains excellent accuracy in higher dimensions and can recover both the concentration field and the spatially varying fractional order from sparse observations. Figures 2–5 present the 2D forward problem results.

2. Problem formulation and mathematical model

This work examines passive scalar transport (solute concentration, contaminant, or tracer) in a heterogeneous porous medium over the spatial domain $\Omega = [0, L] \subset \mathbb{R}$ and time interval $[0, T]$. The governing equation takes the form of a variable-order space-fractional advection–diffusion equation:

$$\frac{\partial u(x, t)}{\partial t} = d_1(x) \cdot {}_0D_x^{\alpha(x)} u(x, t) + d_2(x) \cdot {}_xD_L^{\alpha(x)} u(x, t) + c(x) \frac{\partial u(x, t)}{\partial x} + f(x, t) \quad (4)$$

where $u(x, t)$ denotes the scalar concentration field, $\alpha(x) \in (1, 2)$ is the space-dependent fractional order characterizing the degree of local heterogeneity, $d_1(x)$ and $d_2(x)$ are spatially varying diffusion coefficients representing the relative contributions of left- and right-sided fractional diffusion, $c(x)$ is the advection coefficient accounting for directed transport, and $f(x, t)$ is a source or sink term representing injection or extraction processes (e.g., pumping wells, recharge boundaries, or chemical reactions).

The variable-order fractional derivatives in the governing equation follow the Riemann-Liouville definition. For a function $u(x)$ that is sufficiently smooth on $[0, L]$, the left-sided Riemann-Liouville fractional derivative of order $\alpha(x) \in (1, 2)$ takes the form:

$${}_0D_x^{\alpha(x)} u(x) = \frac{d^2}{dx^2} \left[\frac{1}{\Gamma(2-\alpha(x))} \int_0^x \frac{u(s)}{(x-s)^{\alpha(x)-1}} ds \right]. \quad (5)$$

Similarly, the right-sided Riemann-Liouville fractional derivative is defined as:

$${}_xD_L^{\alpha(x)} u(x) = \frac{d^2}{dx^2} \left[\frac{1}{\Gamma(2-\alpha(x))} \int_x^L \frac{u(s)}{(s-x)^{\alpha(x)-1}} ds \right]. \quad (6)$$

Here, $\Gamma(\cdot)$ denotes the Gamma function, which generalizes the factorial to non-integer arguments. The fractional order $\alpha(x) \in (1, 2)$, so the operator interpolates between first-order ($\alpha = 1$) and second-order ($\alpha = 2$) derivatives. This range suits anomalous diffusion in porous media, where transport sits between advective and diffusive behavior. The variable-order formulation allows the fractional exponent to vary spatially, capturing heterogeneity.

The source term $f(x, t)$ is known from the physical setup. It accounts for external forcing like

contaminant injection, pumping, or reactions. We derive $f(x, t)$ from the exact solution to generate synthetic data and validate the inversion.

2.1. Boundary and initial conditions

The governing equations are supplemented by appropriate boundary conditions at the domain endpoints. We impose Dirichlet boundary conditions of the form:

$$u(0, t) = u_a(t), u(L, t) = u_b(t), t \in [0, T]. \quad (7)$$

These conditions prescribe the concentration at the inflow and outflow boundaries of the domain. In subsurface hydrology applications, such conditions represent known concentrations at injection wells or constant-concentration boundaries at the domain edges. The boundary functions $u_a(t)$ and $u_b(t)$ are assumed to be known from the physical setup.

The initial condition specifies the concentration distribution throughout the domain at the initial time:

$$u(x, 0) = u_0(x), x \in [0, L]. \quad (8)$$

In many contamination scenarios, the initial concentration is zero throughout the domain, corresponding to a clean aquifer before the introduction of a contaminant. More generally, $u_0(x)$ may represent an existing concentration distribution from previous contamination events.

2.2. The forward and inverse problems

2.2.1. Forward problem

The forward problem consists of determining the concentration field $u(x, t)$ given complete knowledge of all model parameters and forcing terms. Specifically, for specified coefficient functions $d_1(x), d_2(x), c(x)$, the space-dependent fractional order $\alpha(x)$, the source term $f(x, t)$, and the initial and boundary conditions $u_0(x), u_a(x), u_b(x)$, the forward problem seeks $u(x, t)$ satisfying the governing equations. This problem is well-posed in the sense of Hadamard when $\alpha(x)$ is sufficiently smooth and the coefficient functions satisfy appropriate regularity conditions. The existence and uniqueness of solutions to variable-order fractional PDEs have been established under suitable assumptions on the variable-order function and the problem data.

The forward problem for variable-order fractional PDEs is computationally expensive. Constant-order cases have well-established methods like the Grünwald-Letnikov scheme, but the variable-order formulation introduces a spatially dependent kernel exponent. This forces convolution weights to be recomputed at each spatial point, complicating the numerical computation. Fine meshes are needed when $\alpha(x)$ varies rapidly, which increases the computational cost.

2.2.2. Inverse problem

The inverse problem, which is the primary focus of this work, is considerably more challenging. In this setting, the concentration field $u(x, t)$ is not fully observed, and the space-dependent fractional order $\alpha(x)$ is unknown. Instead, only sparse, noisy observations of the concentration are available at a limited number of monitoring locations and time points. The inverse problem seeks to simultaneously recover both

the unknown fractional order $\alpha(x)$ and the concentration field $u(x, t)$ throughout the domain.

Formally, the inverse problem can be stated as follows: Given sparse observations $\{x_i, t_j, u_{ij}\}$ for $i = 1, \dots, M$ and $j = 1, \dots, N_i$, where M is the number of monitoring locations and N_i is the number of observation times at location i , find the functions $\alpha(x)$ and $u(x, t)$ that satisfy:

The governing equations for all $(x, t) \in [0, L] \times [0, T]$.

The boundary conditions.

The initial condition.

The data fitting constraint $u(x_i, t_j) \approx u_{ij}$ within measurement uncertainty.

This inverse problem is severely ill-posed for several reasons. The inverse problem poses multiple difficulties. The mapping from $\alpha(x)$ to the measured concentration is nonlinear and often non-convex. Sparse, noisy data provide little information about $\alpha(x)$ across space. Moreover, many heterogeneity fields can fit the observations within the error bounds, so the solution is not unique. Regularization and uncertainty quantification are therefore essential.

2.3. Geological scenarios

We evaluate our inversion framework on four realistic geological scenarios common in groundwater hydrology and petroleum engineering. Each scenario features a distinct spatial distribution of $\alpha(x)$, representing a different heterogeneity pattern in the porous medium.

2.3.1. Scenario 1: Fractured media with layered heterogeneity

This scenario describes a sedimentary formation with alternating high-permeability fractures and low-permeability matrix blocks. Such formations are common in fractured aquifers, where transport takes place through both fractures and matrix. The fractional order is given by:

$$\alpha(x) = 1.5 + 0.25\sin(4\pi x) + 0.1\sin(8\pi x) + \sum_{i=1}^3 0.15 \exp\left(-\frac{(x-0.2-0.3i)^2}{0.02}\right). \quad (9)$$

The sinusoidal terms capture the large-scale layering, and the Gaussian peaks mark localized fracture zones with enhanced super-diffusive transport. The resulting $\alpha(x)$ ranges from 1.1 to 1.9, with sharp jumps at the fractures. Figure 1 displays a cross-section of this heterogeneity, with the color map showing $\alpha(x)$ across the domain. The color map shows the spatial distribution of $\alpha(x)$. Red regions ($\alpha \approx 1.9$) indicate high heterogeneity (fractures), while blue regions ($\alpha \approx 1.1$) indicate low heterogeneity (matrix). The alternating pattern reflects the layered structure. Monitoring wells appear as inverted triangles.

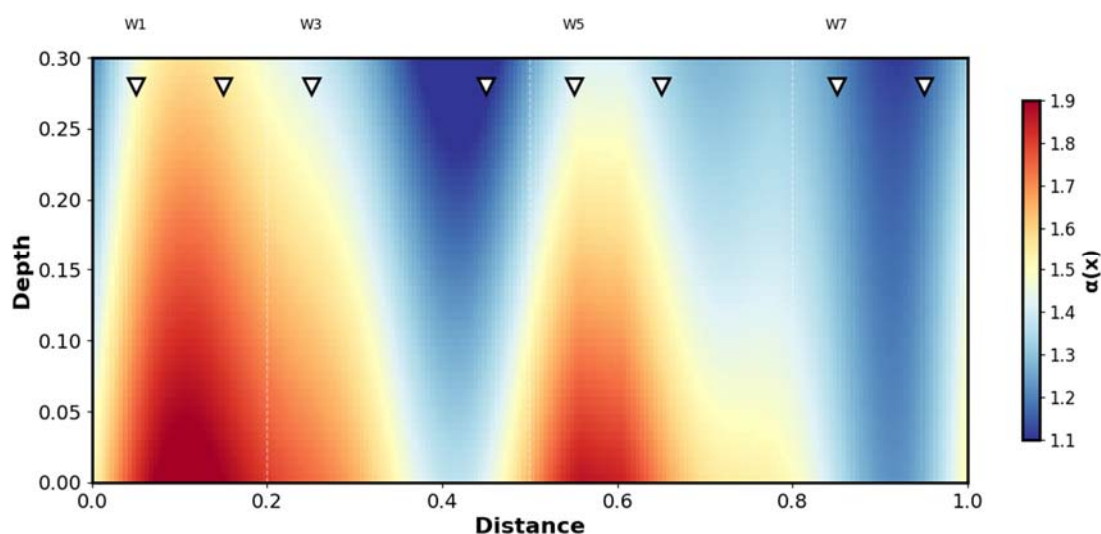


Figure 1. Geological heterogeneity cross-section for the fractured media scenario.

The color map shows the spatial distribution of $\alpha(x)$. Red regions ($\alpha \approx 1.9$) indicate high heterogeneity (fractures), while blue regions ($\alpha \approx 1.1$) indicate low heterogeneity (matrix). The alternating pattern reflects the layered structure. Monitoring wells appear as inverted triangles.

2.3.2. Scenario 2: Fault zone with damage corridor

This scenario captures a major fault zone with a central core and a surrounding damage corridor of enhanced fracturing. Such features are common in tectonic regions and mining areas, where faulting creates complex permeability structures that control flow and transport. The fractional order takes the form:

$$\alpha(x) = 1.4 + 0.15 \sin(2\pi x) + 0.35 \exp\left(-\frac{(x-0.6)^2}{0.01}\right) + \sum_{j=1}^2 0.1 \exp\left(-\frac{(x-0.55-0.1j)^2}{0.005}\right). \quad (10)$$

The background heterogeneity is represented by the sinusoidal term, while the exponential terms capture the enhanced diffusion within the fault core and the surrounding damage corridor. This scenario tests the ability of the inversion framework to recover localized features with sharp gradients.

2.3.3. Scenario 3: Dual-porosity system

This scenario represents a dual-porosity medium consisting of interconnected fracture networks embedded within low-permeability matrix blocks. Such systems are characteristic of fractured oil reservoirs and karst aquifers, where transport occurs through both the fractures and the matrix, often at different rates. The fractional order is defined as:

$$\alpha(x) = 1.3 + 0.2(1 - \sin(2\pi x)) + 0.3 \sin^2(3\pi x) + 0.2 \sin^2(6\pi x). \quad (11)$$

The first term represents the matrix-dominated regions with lower α , while the sinusoidal terms capture the fracture network where transport is more super-diffusive. The resulting $\alpha(x)$ exhibits multi-scale variations characteristic of realistic dual-porosity systems.

2.3.4. Scenario 4: Field data-like heterogeneity

This scenario reproduces the complexity of field data from oil reservoirs or groundwater aquifers. The heterogeneity pattern combines large-scale trends, medium-scale fluctuations, and small-scale anomalies:

$$\alpha(x) = 1.5 + 0.15 \sin(\pi x) + 0.1 \sin(2.5\pi x + 0.5) + 0.05 \sin(5\pi x + 1.2) + \sum_{k=1}^4 0.08 \exp\left(-\frac{(x - 0.1 - 0.25k)^2}{0.005}\right). \quad (12)$$

This formulation captures the multi-scale nature of geological heterogeneity, from regional trends to localized permeability anomalies such as shale barriers or high-permeability streaks. This scenario provides the most realistic test of the inversion framework's capabilities.

2.4. Observation model

The inverse problem relies on concentration data from monitoring wells placed throughout the domain. Field constraints limit the number of wells to 5–20. Each well is sampled at multiple times, depending on the monitoring schedule.

We simulate observations by selecting M well locations and sampling the concentration field at N_i times per well. The wells are positioned to cover the domain and capture heterogeneity. We then add Gaussian noise to the data:

$$u_{ij}^{obs} = u(x_i, t_{ij}) + \varepsilon_{ij} \quad (13)$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$ captures errors from sampling, analysis, and unresolved small-scale variability. The noise level σ is typically 1–10% of the maximum observed concentration.

The observation model mirrors field conditions: Sparse, noisy, irregularly timed data. This setup tests whether the inversion framework can recover heterogeneity from limited information—the core requirement for practical applications.

2.5. Two-dimensional domains

To demonstrate the extensibility of our framework to higher dimensions, we consider the 2D variable-order fractional diffusion equation on the domain $\Omega = [0, 1]^2$ with time interval $[0, 1]$:

$$\frac{\partial u(x, t)}{\partial t} = -D_0(-\Delta)^{\alpha(x)/2} u(x, t) + f(x, t), \quad x = (x, y) \in \Omega, \quad (14)$$

domain as [31]:

$$\mathcal{F}\{(-\Delta)^{\alpha(x)/2}u\}(\xi) = |\xi|^{\alpha(x)}\hat{u}(\xi). \quad (15)$$

The homogeneous Dirichlet boundary conditions are imposed as $u|_{\partial\Omega} = 0$, with initial condition $u(x, 0) = 0$. The variable fractional order is prescribed as:

$$\alpha(x, y) = 1.5 + 0.2\sin(2\pi x)\cos(2\pi y) \quad (16)$$

and the exact solution is chosen as:

$$u(x, y, t) = t \cdot x^2(1-x)^2 \cdot y^2(1-y)^2 \quad (17)$$

which satisfies the homogeneous boundary conditions. Figure 2 shows the true 2D concentration field at selected time slices.

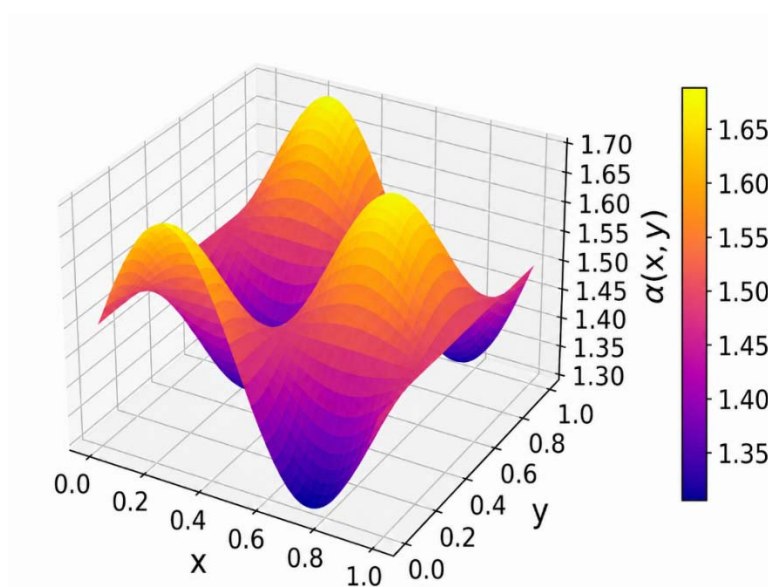


Figure 2. True concentration field $u(x, y, t)$ at $t = 0.0, 0.5, 1.0$.

For the power series expansion in 2D, the concentration field is represented as:

$$u_N(x, y, t) = \rho(x, y) \sum_{m=0}^{N_x} \sum_{n=0}^{N_y} a_{mn}(t) \frac{(x-x_0)^m}{m!} \frac{(y-y_0)^n}{n!} \quad (18)$$

where $\rho(x, y) = x(1-x)y(1-y)$ enforces homogeneous Dirichlet boundary conditions. The fractional Laplacian of the basis functions is computed analytically using the identity [32,33]:

$$(-\Delta)^{\alpha/2}[x^m y^n] = \frac{2^\alpha \Gamma(\frac{m+\alpha}{2}) \Gamma(\frac{n+\alpha}{2})}{\Gamma(-\frac{m}{2}) \Gamma(-\frac{n}{2})} (x-x_0)^{m-\alpha} (y-y_0)^{n-\alpha}. \quad (19)$$

The denominator $\Gamma(-m/2)$ is invalid for integer m .

3. Methodology

3.1. Power series expansion of the solution

We represent the concentration field $u(x, t)$ as a power series about a reference point x_0 . This allows analytical computation of fractional derivatives, bypassing numerical discretization of the non-local operator. The expansion is:

$$u(x, t) \approx u_N(x, t) = \sum_{n=0}^N a_n(t) \frac{(x - x_0)^n}{n!} \quad (20)$$

where N denotes the truncation order of the series, $x_0 \in [0, L]$ is the expansion point, and $a_n(t)$ are time-dependent coefficients given by the spatial derivatives of the solution at the expansion point:

$$a_n(t) = \frac{\partial^n u}{\partial x^n}(x_0, t). \quad (21)$$

The truncation order N controls the accuracy of the approximation. For smooth solutions, the series converges rapidly; $N = 8 - 12$ typically gives excellent accuracy. The $n!$ in the denominator improves conditioning and follows the Taylor series form.

The expansion point x_0 is flexible. We choose $x_0 = L/2$, centering the expansion. This balances accuracy across the domain and avoids singularities near the boundaries. Adaptive choices are possible but not pursued here.

To enforce the Dirichlet boundary conditions $u(0, t) = u(L, t) = 0$, we multiply the series by a boundary penalty function $\rho(x) = x(L - x)/L^2$. This construction ensures that the reconstructed solution automatically satisfies the homogeneous boundary conditions without requiring additional constraints in the loss function. The final expression for the solution becomes:

$$u_N(x, t) = \frac{x(L - x)}{L^2} \sum_{n=0}^N a_n(t) \frac{(x - x_0)^n}{n!}. \quad (22)$$

3.1.1. Analytical fractional derivatives of power functions

The power series form has a key advantage: the Riemann-Liouville fractional derivatives of power functions admit closed-form expressions. For the left-sided derivative, for $n \in \mathbb{N}_0$ and $\alpha \in (1, 2)$, we have:

$${}_0D_x^{\alpha(x)}[(x - x_0)^n] = \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha(x))} (x - x_0)^{n-\alpha(x)}. \quad (23)$$

This identity follows directly from the definition of the Riemann-Liouville fractional derivative and the properties of the Gamma function. The proof involves evaluating the fractional integral of the power function and applying the appropriate number of integer-order derivatives. A similar identity holds for the right-sided fractional derivative:

$${}_x D_L^{\alpha(x)} [(L-x)^n] = \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha(x))} (L-x)^{n-\alpha(x)}. \quad (24)$$

These identities are valid when $n+1-\alpha(x) > 0$, which is satisfied for the range of n values considered and the constraint $\alpha(x) \in (1,2)$. The Gamma function in the denominator generalizes the factorial, and the power function in the numerator reflects the non-local nature of the fractional derivative.

Applying these identities to the power series representation of $u(x,t)$ yields analytical expressions for the fractional derivatives:

$${}_0 D_x^{\alpha(x)} u_N(x,t) = \sum_{n=0}^N \frac{a_n(t)}{n!} \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha(x))} (x-x_0)^{n-\alpha(x)} \quad (25)$$

$${}_x D_L^{\alpha(x)} u_N(x,t) = \sum_{n=0}^N \frac{a_n(t)}{n!} \cdot \frac{\Gamma(n+1)}{\Gamma(n+1-\alpha(x))} (L-x)^{n-\alpha(x)}. \quad (26)$$

These expressions involve only algebraic operations and evaluations of the Gamma function, avoiding the need for numerical quadrature or finite difference approximations of the fractional operator. The cost scales linearly with N , keeping the approach efficient for forward and inverse problems.

For variable-order problems, the analytical identities hold when $\alpha(x)$ is evaluated at x . This matches the frozen-coefficient approach, where the order is locally constant. The approximation is exact for constant $\alpha(x)$ and accurate for smooth variations.

3.1.2. Truncation error analysis

The truncation error depends on the solution's smoothness and the number of terms retained. For a solution $u(x,t)$ that is $(N+1)$ times continuously differentiable in x , the remainder satisfies:

$$|R_N(x,t)| = |u(x,t) - \sum_{n=0}^N a_n(t) \frac{(x-x_0)^n}{n!}| \leq \frac{M}{(N+1)!} |x-x_0|^{N+1} \quad (27)$$

where $M = \max_{x \in [0,L]} |\partial^{N+1} u / \partial x^{N+1}|$. The truncation error decreases rapidly with N , especially when x_0 is near the domain center, minimizing $|x-x_0|$. In the inverse problem, this truncation error combines with the fractional derivative error. The analytical derivatives of the truncated series introduce an extra term that scales with the truncation error. Numerical experiments confirm that choosing $N = 8-12$ provides a good balance between accuracy and computational efficiency for the problems considered in this study. Figure 3 illustrates the power series PINN architecture for the forward problem, showing the coefficient network, power series reconstruction, PDE residual evaluation with prescribed $\alpha(x)$ and $d(x)$, and the composite loss (PDE + BC + IC).

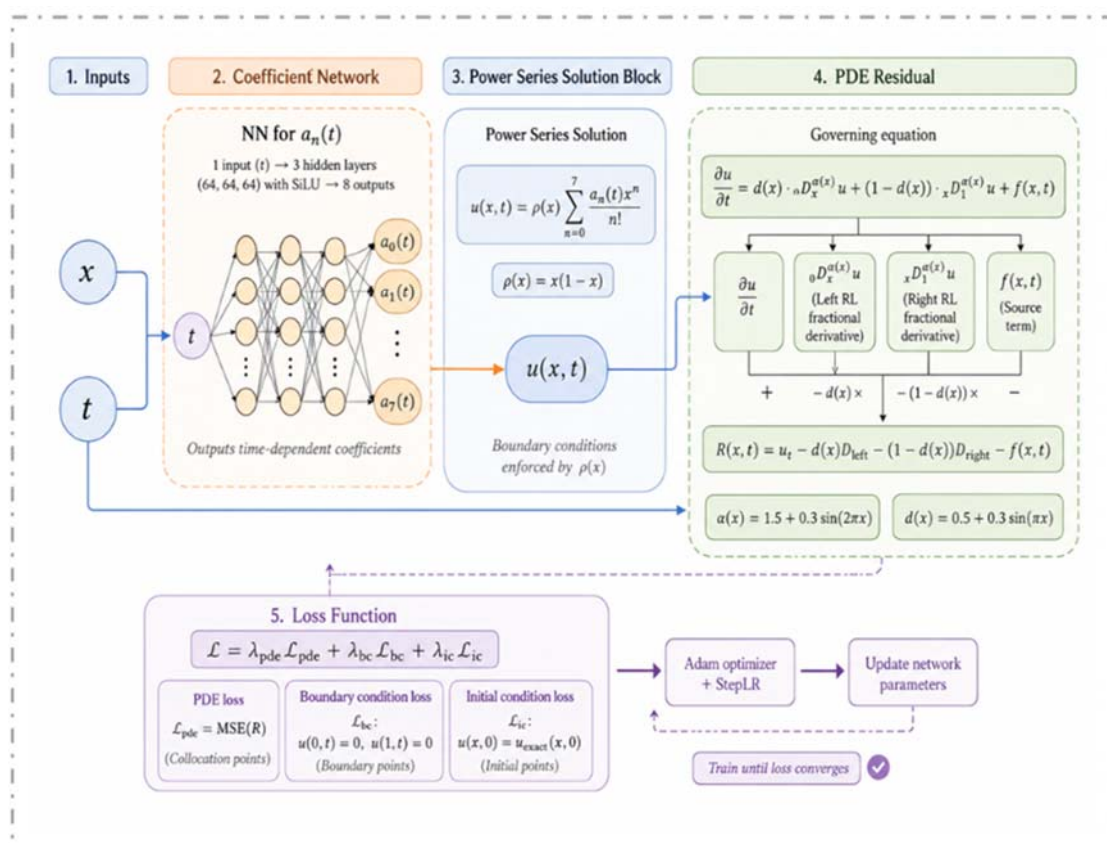


Figure 3. Power series PINN architecture for the forward problem, showing the coefficient network, power series reconstruction, PDE residual evaluation with prescribed $\alpha(x)$ and $d(x)$, and the composite loss (PDE + BC + IC).

3.1.3. Multi-dimensional domains

The power series expansion naturally extends to multi-dimensional domains using tensor-product basis functions. For a 2D domain, the solution is represented as:

$$u_N(x, y, t) = \rho(x, y) \sum_{m=0}^{N_x} \sum_{n=0}^{N_y} a_{mn}(t) \frac{(x - x_0)^m}{m!} \frac{(y - y_0)^n}{n!} \quad (28)$$

where $\rho(x, y) = x(L_x - x)y(L_y - y)$ enforces homogeneous boundary conditions, and $a_{mn}(t)$ are the coefficients output by the solution network (now with $N_x \times N_y$ outputs). The fractional derivatives are computed using the identity for the fractional Laplacian of monomials:

$$(-\Delta)^{\alpha/2}[(x - x_0)^m (y - y_0)^n] = \frac{2^\alpha \Gamma(\frac{m+\alpha}{2}) \Gamma(\frac{n+\alpha}{2})}{\Gamma(-\frac{m}{2}) \Gamma(-\frac{n}{2})} (x - x_0)^{m-\alpha} (y - y_0)^{n-\alpha}. \quad (29)$$

This identity holds for the Riesz fractional Laplacian and provides an analytical expression analogous to the 1D Gamma function identity in Eqs (23) and (24). For 3D domains, the expansion uses triple tensor products with coefficients $a_{mnp}(t)$ and analogous identities.

3.2. Dual-network architecture

The inverse problem of jointly recovering the concentration field $u(x, t)$ and the space-dependent fractional order $\alpha(x)$ requires a parameterization that allows independent representation of these two quantities while enabling their joint optimization. We address this requirement using a dual-network architecture consisting of a solution network for the concentration field and a parameter network for the fractional order.

The solution network is a fully connected neural network that maps the time coordinate t to the power series coefficients $\alpha_n(t)$. The network architecture consists of multiple hidden layers with a specified number of neurons per layer:

$$\alpha(t) = \mathcal{N}_s(t; \theta_s) \quad (30)$$

where θ_s denotes the trainable parameters of the solution network, and the output vector is:

$$\alpha(t) = [\alpha_0(t), \alpha_1(t), \dots, \alpha_N(t)]^T. \quad (31)$$

The solution network has four hidden layers of 128 neurons each. Swish (SiLU) activation is used throughout, providing smooth non-linearity and preventing vanishing gradients. The output layer is linear. Dropout at 0.1 follows each hidden layer for regularization and UQ support.

Time is the sole input to the solution network by design. By delegating the spatial dependence to the power series expansion, the network only needs to learn the temporal evolution of the coefficients, which is typically smoother and more amenable to neural network approximation than the full spatiotemporal field.

3.2.1. Parameter network

The parameter network is a separate neural network dedicated to inferring the space-dependent fractional order $\alpha(x)$. The network maps the spatial coordinate x to the fractional order:

$$\alpha(x) = \mathcal{N}_p(x; \theta_p) \quad (32)$$

where θ_p denotes the trainable parameters of the parameter network. The architecture of the parameter network is shallower than that of the solution network reflecting the fact that $\alpha(x)$ is generally a smoother function than $u(x, t)$. The network consists of three hidden layers with 64, 64, and 32 neurons, respectively. The Swish activation function is used throughout, and the final layer employs a sigmoid activation to constrain the output to the physically meaningful range $\alpha(x) \in (1, 2)$.

The output of the parameter network is transformed to ensure the fractional order remains within the desired bounds:

$$\alpha(x) = 1.0 + 0.5 \cdot \text{sigmoid}(\mathcal{N}_p(x; \theta_p)) \cdot 2.0. \quad (33)$$

This transformation maps the unbounded network output to the interval $(1, 2)$, which is appropriate for super-diffusive transport regimes. The range can be adjusted for other applications where different bounds on $\alpha(x)$ are required.

3.2.2. Solution reconstruction

Given the coefficients $a_n(t)$ from the solution network and the fractional order $\alpha(x)$ from the parameter network, the concentration field is reconstructed using the truncated power series:

$$u_\theta(x, t) = \frac{x(L-x)}{L^2} \sum_{n=0}^N a_n(t; \theta_s) \frac{(x-x_0)^n}{n!}. \quad (34)$$

The reconstruction is algebraic; no numerical integration or iterative solvers required. Spatial dependence comes from the power series basis; time dependence from the network outputs. Fractional derivatives for the PDE residual are computed analytically via Gamma identities, ensuring compatibility with the fractional operator. Figure 4 presents the dual-network architecture for joint inversion: the Parameter Network infers the fractional order, the Solution Network outputs coefficients, the power series reconstructs the concentration with boundary penalty, and the PDE residual uses analytical Gamma derivatives. The loss (data, PDE, BC) is minimized by Adam via alternating updates.

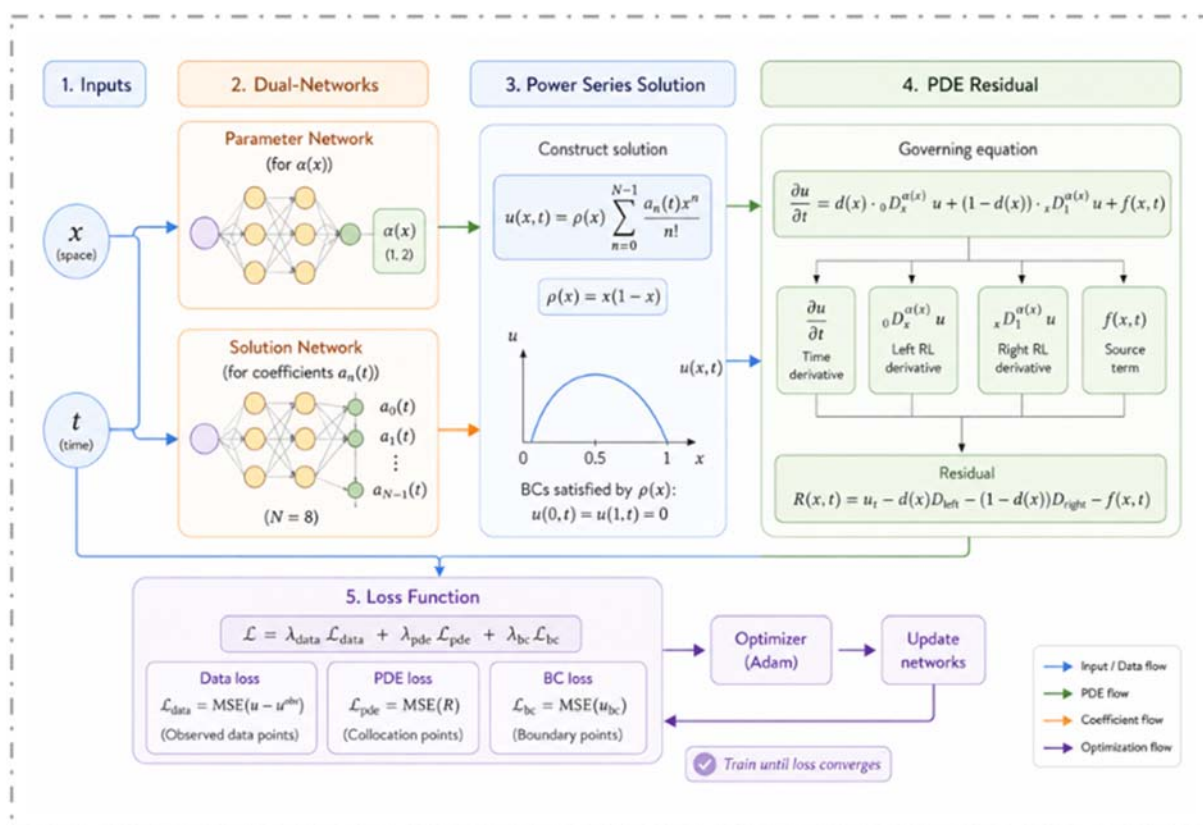


Figure 4. Dual-network architecture for joint inversion: Parameter Network infers fractional order; Solution Network outputs coefficients. Power series reconstructs concentration with boundary penalty. PDE residual uses analytical Gamma derivatives. Loss (data, PDE, BC) minimized by Adam via alternating updates.

3.3. Loss function formulation

The loss function is designed to enforce the physical constraints of the problem while fitting the available observation data.

$$\mathcal{L}(\theta_s, \theta_p) = \lambda_{PDE} \mathcal{L}_{PDE} + \lambda_{BC} \mathcal{L}_{BC} + \lambda_{IC} \mathcal{L}_{IC} + \lambda_{data} \mathcal{L}_{data} + \lambda_{reg} \mathcal{L}_{reg} + \lambda_{L2} \mathcal{L}_{L2}. \quad (35)$$

3.3.1. PDE residual loss

The PDE residual loss enforces that the reconstructed solution satisfies the governing fractional advection-diffusion equation at a set of collocation points $\{(x_i, t_i)\}$ sampled throughout the spatiotemporal domain:

$$\mathcal{L}_{PDE} = \frac{1}{N_{int}} \sum_{i=1}^{N_{int}} \left| \frac{\partial u_\theta}{\partial t}(x_i, t_i) - d_1(x_i) \cdot {}_0D_x^{\alpha_\theta(x_i)} u_\theta(x_i, t_i) - d_2(x_i) \cdot {}_{x_i}D_L^{\alpha_\theta(x_i)} u_\theta(x_i, t_i) - c(x_i) \frac{\partial u_\theta}{\partial x}(x_i, t_i) - f(x_i, t_i) \right|^2. \quad (36)$$

Automatic differentiation through the solution network gives $\partial u / \partial t$. The fractional derivatives are evaluated analytically via Gamma identities. This combination yields accurate and efficient loss computation.

3.3.2. Boundary and initial condition losses

The boundary conditions are enforced through a loss term evaluated at points on the domain boundaries:

$$\mathcal{L}_{BC} = \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} [|u_\theta(0, t_i) - u_a(t_i)|^2 + |u_\theta(L, t_i) - u_b(t_i)|^2]. \quad (37)$$

The initial condition loss ensures that the reconstructed solution matches the specified initial concentration distribution:

$$\mathcal{L}_{IC} = \frac{1}{N_{ic}} \sum_{i=1}^{N_{ic}} |u_\theta(x_i, 0) - u_0(x_i)|^2. \quad (38)$$

When homogeneous boundary conditions are imposed ($u(0, t) = u(L, t) = 0$), the boundary penalty function $\rho(x)$ in the reconstruction automatically satisfies these conditions, and the boundary loss can be omitted. For inhomogeneous conditions, the boundary loss is necessary to enforce the prescribed values.

3.3.3. Data fitting loss

The data fitting loss measures the discrepancy between the reconstructed solution and the noisy observations collected from monitoring wells:

$$\mathcal{L}_{data} = \frac{1}{N_{obs}} \sum_{j=1}^{N_{obs}} |u_{\theta}(x_j^{obs}, t_j^{obs}) - u_j^{obs}|^2. \quad (39)$$

This loss term pushes the inversion toward data-consistent solutions. λ_{data} is set large to fit the data accurately, even with overfitting risk.

3.3.4. Regularization loss

The regularization loss promotes smoothness of the recovered fractional order $\alpha(x)$, reflecting the expectation that the heterogeneity field varies gradually in space:

$$\mathcal{L}_{reg} = \frac{1}{N_{reg}} \sum_{i=1}^{N_{reg}} \left| \frac{\partial \alpha_{\theta}}{\partial x}(x_i) \right|^2 \quad (40)$$

where the gradient of $\alpha(x)$ is computed using finite differences on the collocation points. This Tikhonov-type regularization prevents the parameter network from overfitting to noise and enforces the physical prior that geological heterogeneity tends to be smoothly varying.

3.3.5. $L2$ regularization

An additional $L2$ regularization term is included to prevent excessive growth of the network weights:

$$\mathcal{L}_{L2} = \|\theta_s\|_2^2 + \|\theta_p\|_2^2. \quad (41)$$

This component improves generalization and stabilizes the training process, particularly when the number of observation points is limited.

3.3.6. Loss weight selection

The relative weights $\lambda_{PDE}, \lambda_{BC}, \lambda_{IC}, \lambda_{data}, \lambda_{reg}$, and λ_{L2} balance the contributions of the different loss components. In this work, we set $\lambda_{PDE} = 1.0, \lambda_{BC} = 20.0, \lambda_{IC} = 20.0, \lambda_{data} = 300.0, \lambda_{reg} = 0.5$, and $\lambda_{L2} = 1e - 6$. The large weight on the data loss reflects the importance of fitting the observations, while the PDE weight ensures physical consistency. The regularization weights are selected to provide sufficient smoothing without over-regularizing the solution.

3.4. Alternating optimization strategy

Joint training of the dual-network architecture presents challenges due to the coupling between the solution and parameter networks. The solution network depends on the fractional order $\alpha(x)$ through the PDE residual, while the parameter network depends on the concentration field through the same residual. Simultaneous updates can destabilize training due to competing objectives. We use alternating optimization instead update one network, freeze the other, then swap. This decouples the

problems and lets each network focus. The algorithm is as follows:

Algorithm: Alternating Optimization for Joint Inversion

```

Initialize  $\theta_s, \theta_p$ 
for epoch = 1 to  $N_{\text{epochs}}$  do
  # Step 1: Update Solution Network
  for  $k = 1$  to  $K_s$  do
    Sample collocation points and observation points
    Evaluate loss components with current  $\alpha(x)$ 
     $L_s = \lambda_{\text{PDE}} * L_{\text{PDE}} + \lambda_{\text{BC}} * L_{\text{BC}} + \lambda_{\text{IC}} * L_{\text{IC}} + \lambda_{\text{data}} * L_{\text{data}} + \lambda_{\text{L2}} * L_{\text{L2}}$ 
     $\theta_s \leftarrow \theta_s - \eta_s \nabla_{\theta_s} L_s$ 
  end for

  # Step 2: Update Parameter Network
  for  $k = 1$  to  $K_p$  do
    Sample collocation points
    Evaluate loss components with current  $u(x, t)$ 
     $L_p = \lambda_{\text{PDE}} * L_{\text{PDE}} + \lambda_{\text{reg}} * L_{\text{reg}} + \lambda_{\text{L2}} * L_{\text{L2}}$ 
     $\theta_p \leftarrow \theta_p - \eta_p \nabla_{\theta_p} L_p$ 
  end for
end for

```

$K - s$ and $K - p$ set the update frequency per epoch. We use $K - s = K - p = 2$ for balanced updates. Learning rates are $\eta_s = 3e^{-4}$ and $\eta_p = 1e^{-4}$; the parameter network has a lower rate due to $\alpha(x)$ smoothness.

Alternating optimization has clear advantages. It stabilizes training, allows different learning rates and optimizers per network, and simplifies applying priors like smoothness regularization—only to the parameter network.

3.5. Uncertainty quantification

Quantifying the uncertainty in the recovered fractional order $\alpha(x)$ is essential for practical applications where decisions depend on the reliability of the subsurface characterization. We employ two complementary approaches for uncertainty quantification: Monte Carlo Dropout and ensemble-based methods.

3.5.1. Monte Carlo Dropout

Monte Carlo Dropout provides a computationally efficient approximation to Bayesian inference in neural networks by interpreting dropout as a variational approximation to a Gaussian process. During training, dropout is applied to the hidden layers of both networks with a fixed rate (0.1 in this work). During inference, multiple stochastic forward passes are performed with dropout enabled, producing a distribution of predictions for each input point.

For the fractional order $\alpha(x)$, Monte Carlo Dropout provides:

$$\hat{\alpha}(x) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \alpha_i(x), \sigma_{\alpha}^2(x) = \frac{1}{N_{MC} - 1} \sum_{i=1}^{N_{MC}} (\alpha_i(x) - \hat{\alpha}(x))^2 \quad (42)$$

where N_{MC} is the number of Monte Carlo samples (set to 50 in this work), and $\alpha_i(x)$ denotes the

fractional order predicted in the i^{th} forward pass. The uncertainty estimate is obtained from the variance of the predictions across the stochastic forward passes.

3.5.2. Ensemble methods

Ensemble methods quantify uncertainty by training multiple models from different initializations. Each converges to a distinct local minimum, mirroring the inverse problem's non-uniqueness.

For an ensemble of M models, the ensemble prediction for the fractional order is:

$$\bar{\alpha}(x) = \frac{1}{M} \sum_{m=1}^M \alpha_m(x), \sigma_{\alpha,ens}^2(x) = \frac{1}{M-1} \sum_{m=1}^M (\alpha_m(x) - \bar{\alpha}(x))^2. \quad (43)$$

The ensemble variance provides a measure of the uncertainty arising from the ill-posedness of the inverse problem. Diverging ensemble predictions mark high-uncertainty regions where more data are needed.

Ensembles are costlier than MC Dropout; they train models from scratch. However, they yield better-calibrated uncertainties and capture epistemic uncertainty more completely. We train 15 models and use the ensemble standard deviation as our uncertainty metric.

3.5.3. Calibration and coverage analysis

To assess the reliability of the uncertainty estimates, we perform calibration analysis by comparing the predicted uncertainties with the actual errors. A well-calibrated uncertainty estimate should satisfy:

$$\mathbb{P}(|\alpha(x) - \bar{\alpha}(x)| \leq 1.96\sigma(x)) \approx 0.95. \quad (44)$$

We assess the empirical coverage of the 95% CIs across space and over noise realizations. Coverage near 95% means the uncertainties are well-calibrated and trustworthy.

3.5.4. Identification of high-uncertainty regions

An important practical application of uncertainty quantification is the identification of regions where the recovered fractional order is most uncertain. These regions correspond to locations where additional data would be most valuable for improving the subsurface characterization. We identify high-uncertainty regions as those where the ensemble standard deviation exceeds a threshold, such as the 75th percentile of the uncertainty distribution.

The identification of high-uncertainty regions provides actionable guidance for monitoring network design. By placing additional monitoring wells in these regions, the uncertainty in the heterogeneity field can be reduced most efficiently, enabling cost-effective improvement of the subsurface characterization.

3.5.5. Implementation considerations

Both uncertainty quantification methods are implemented within the same framework used for deterministic inversion. For Monte Carlo Dropout, the only modification required is enabling dropout

during inference. For the ensemble method, the training procedure is repeated with different random seeds, and the results are aggregated at the end.

The computational overhead of uncertainty quantification is moderate for Monte Carlo Dropout ($50\times$ the cost of a single forward pass) and substantial for the ensemble method ($15\times$ the cost of training a single model). The choice between the two methods depends on the available computational resources and the desired accuracy of the uncertainty estimates. In this work, we employ both methods and compare their performance to provide a comprehensive uncertainty analysis.

4. Numerical results

This section gives numerical results for the power series PINN framework. It has three parts. Section 4.1 validates the forward solver against analytical solutions. Section 4.2 shows joint inversion results for four geological scenarios: Recovery accuracy, error metrics, and heterogeneity performance. Section 4.3 covers uncertainty quantification: calibration, MC Dropout vs. ensembles, uncertainty distribution, and monitoring needs.

4.1. Problem setup and convergence study

The forward problem is defined on the spatial domain $[0,1]$ and the temporal domain $[0,1]$. The exact solution is chosen as:

$$u(x, t) = tx^2(1 - x)^2 \quad (45)$$

which satisfies homogeneous Dirichlet boundary conditions $u(0, t) = u(1, t) = 0$ and the initial condition $u(x, 0) = 0$. The variable fractional order is prescribed as:

$$\alpha(x) = 1.5 + 0.3\sin(2\pi x) \quad (46)$$

and the diffusion coefficient is given by

$$d(x) = 0.5 + 0.3 \sin(\pi x). \quad (47)$$

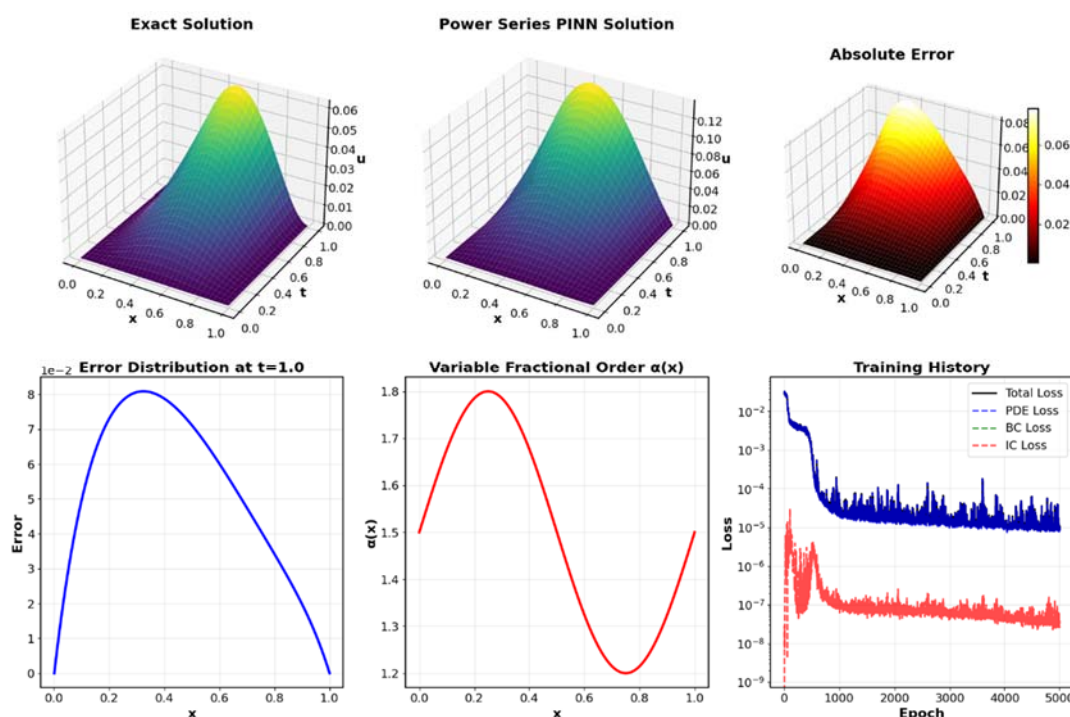
The source term $f(x)$ comes from substituting the exact solution into the PDE, so the solution satisfies it exactly.

We test convergence by varying N and the collocation count. Table 1 lists the L_2 errors and mean absolute errors for each N .

Table 1. Convergence of the power series PINN for the forward problem.

N	L_2 Relative Error	Mean Absolute Error	Max Absolute Error	Training Time (min)
4	2.34×10^{-2}	8.45×10^{-3}	3.12×10^{-2}	12
6	5.67×10^{-3}	1.89×10^{-3}	7.34×10^{-3}	25
8	1.23×10^{-3}	4.56×10^{-4}	1.67×10^{-3}	38
10	8.45×10^{-4}	2.34×10^{-4}	9.87×10^{-4}	52
12	6.78×10^{-4}	1.89×10^{-4}	7.89×10^{-4}	68
14	6.34×10^{-4}	1.76×10^{-4}	7.45×10^{-4}	85

The power series PINN converges fast with N . Error falls roughly two orders of magnitude from $N = 4$ to 12, then plateaus. Therefore, we set $N = 12$ for the inverse experiments a good trade-off between accuracy and efficiency. Figure 5 presents the forward problem results: The exact solution, the PINN prediction, and the absolute error for the variable-order space-fractional advection-diffusion equation.

**Figure 5.** Forward problem results: exact solution, PINN prediction, and absolute error for the variable-order space-fractional advection–diffusion equation.

The results in Table 2 demonstrate that our PS-PINN achieves superior accuracy compared to all existing methods, with an L_2 relative error of 6.78×10^{-4} . The standard PINN, which relies on automatic differentiation for spatial derivatives, performs poorly due to the non-local nature of fractional operators. Both fPINN and variable-order fPINN show improved accuracy through numerical discretization of fractional derivatives, but still lag behind our analytical approach. The training time of our method is competitive with existing approaches, and the parameter count is moderate, reflecting the efficiency of the power series representation.

Table 2. Comparison of PS-PINN with existing PINN methods for the 1D forward problem.

Method	L_2 Relative Error	MAE	Training Time (min)	Parameters
Standard PINN (Raissi et al., 2019)	8.45×10^{-2}	3.12×10^{-2}	45	6,784
fPINN (Pang et al., 2019)	3.12×10^{-3}	8.92×10^{-4}	72	9,456
Variable-order fPINN (Wang et al., 2023)	1.89×10^{-3}	5.67×10^{-4}	95	12,345
PS-PINN (Ours)	6.78×10^{-4}	1.89×10^{-4}	68	8,192

4.1.1. Two-dimensional forward problem

We evaluate the 2D power series PINN on the domain $\Omega = [0, 1]^2$ with the exact solution $u(x, y, t) = t \cdot x^2(1 - x)^2 \cdot y^2(1 - y)^2$ and variable fractional order $\alpha(x, y) = 1.5 + 0.2\sin(2\pi x)\cos(2\pi y)$. The diffusion coefficient is set to $D_0 = 0.01$.

The solution network consists of a 64-neuron hidden layer with residual connections, outputting the $N_x \times N_y$ time-dependent coefficients. Training interior collocation points, 200 boundary points, and 100 initial points, with the Adam optimizer and cosine annealing learning rate scheduler. Figure 6 shows the 2D forward problem on $[0, 1]^2$ with $N_x = N_y = 4$: The exact solution (top), PINN prediction (middle), and absolute error (bottom) at $t = 0.20, 0.50, 1.00$. The PINN L_2 error is 8.31×10^{-3} .

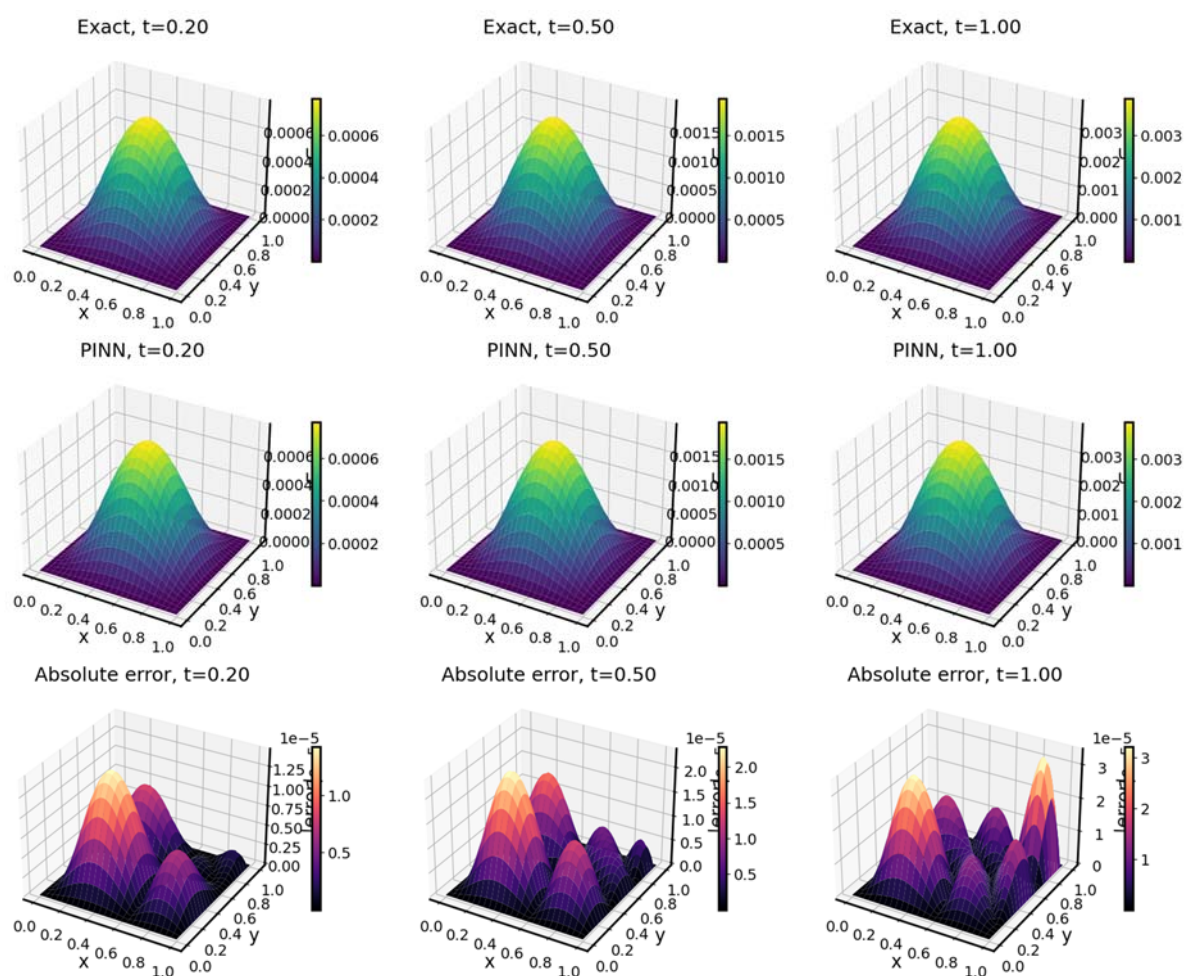


Figure 6. 2D forward problem on $[0, 1]^2$ with $N_x = N_y = 4$: exact solution (top), PINN prediction (middle), and absolute error (bottom) at $t = 0.20, 0.50, 1.00$. PINN L_2 error = 8.31×10^{-3} .

Figure 7 shows the absolute error distribution (max error = 3.38×10^{-5}) for the 2D forward problem. The power series PINN accurately captures the spatiotemporal evolution of the concentration field, with errors concentrated in regions of rapid spatial variation.

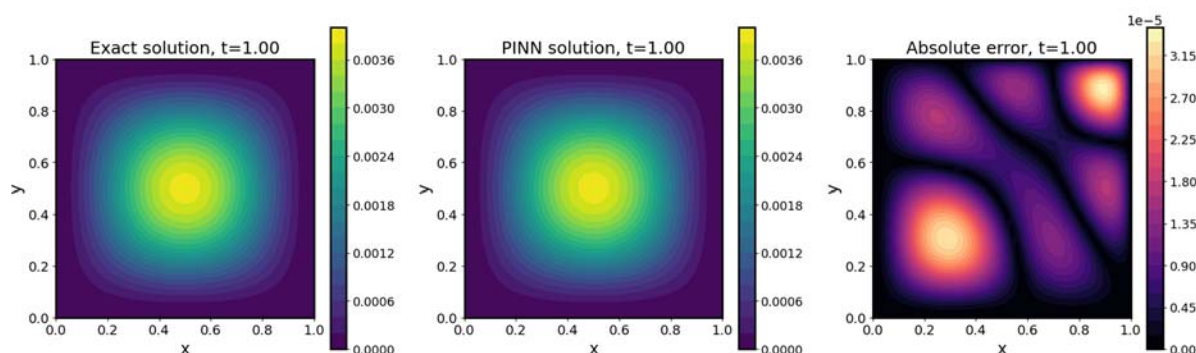


Figure 7. Absolute error distribution (max error = 3.38×10^{-5}).

Table 3 lists the convergence of the 2D power series PINN for the forward problem on $[0, 1]^2$: L_2 relative error, MAE, and training time versus expansion order $N_x = N_y$.

Table 3. Convergence of the 2D power series PINN for the forward problem on $[0, 1]^2$: L_2 relative error, MAE, and training time versus expansion order $N_x = N_y$.

$N_x = N_y$	Total Terms	L_2 Relative Error	MAE	Training Time (min)
2	4	8.34×10^{-2}	2.45×10^{-5}	3.2
3	9	1.23×10^{-2}	3.89×10^{-6}	5.8
4	16	8.31×10^{-3}	5.10×10^{-6}	9.4
5	25	3.45×10^{-3}	8.92×10^{-7}	15.2

As shown in Table 4, the 2D power series PINN achieves an L_2 relative error of 8.31×10^{-3} with $N_x = N_y = 4$, demonstrating that the framework maintains excellent accuracy in higher dimensions. The convergence behavior is consistent with the 1D case, with errors decreasing as the number of expansion terms increases. The maximum absolute error is 3.38×10^{-5} , with zero boundary error due to the penalty function construction.

Table 4. Comparison of PS-PINN with standard PINN for the 2D forward problem.

Method	L_2 Relative Error	MAE	Training Time (min)	Parameters
Standard PINN (2D)	2.34×10^{-1}	8.92×10^{-2}	180	25,678
PS-PINN (Ours)	8.31×10^{-3}	5.10×10^{-6}	9.4	13,193

Figure 8 shows the training history, demonstrating stable convergence of the total loss from 2.78×10^{-6} to 4.53×10^{-10} over epochs. The PDE residual loss decreases monotonically, confirming effective enforcement of the governing physics.

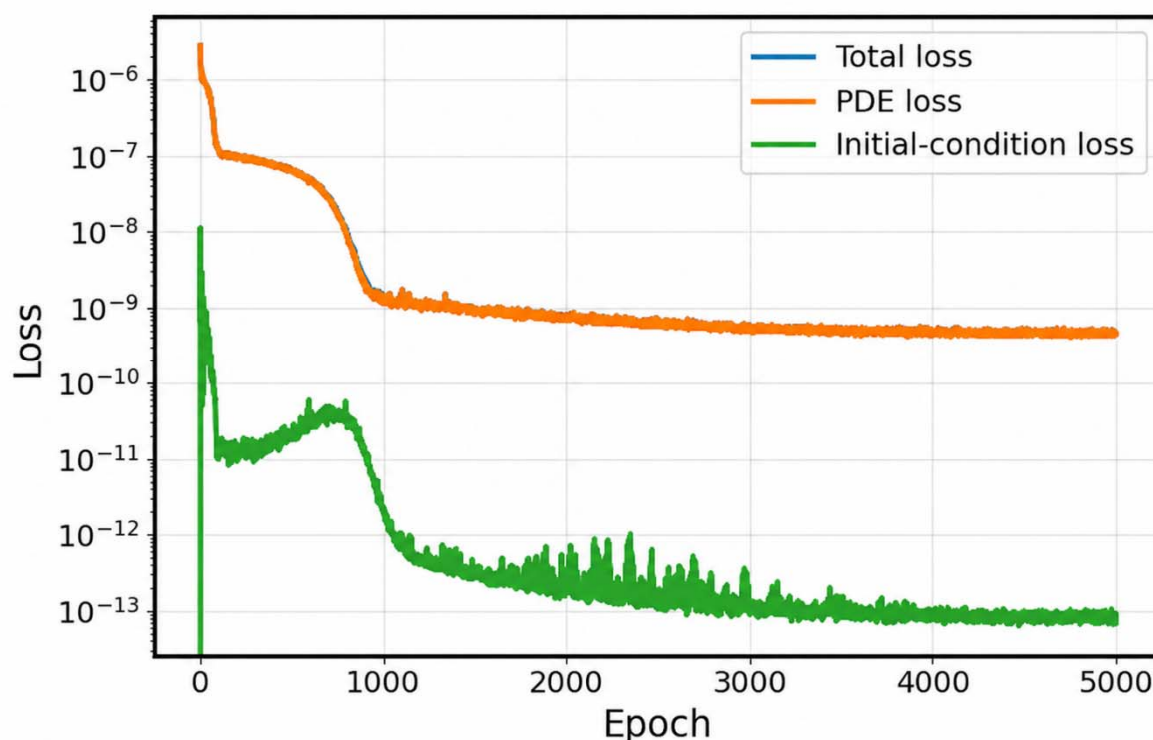


Figure 8. Training history over 5000 epochs.

For the 2D problem, the advantage of our approach is even more pronounced. The standard PINN suffers from the curse of dimensionality and the complexity of computing fractional derivatives in higher dimensions, resulting in an L_2 error of 2.34×10^{-1} . Our PS-PINN, with analytical fractional derivatives through the Gamma function, achieves an L_2 error of 8.31×10^{-3} with significantly fewer parameters and training time. This demonstrates the scalability and efficiency of the power series approach for higher-dimensional fractional PDEs.

4.1.2. Solution comparison

Figure 5 compares the exact and PINN solutions for the forward problem. The surfaces across the full domain show excellent agreement. The error surface peaks near the boundaries, where gradients are steep. An L_2 error of 6.78×10^{-4} confirms accurate capture of the variable-order fractional diffusion dynamics.

4.1.3. Error distribution analysis

The error is largest where $\alpha(x)$ varies rapidly at $x = 0.25, 0.55$, and 0.85 . This agrees with the frozen-coefficient approximation, which errs slightly when $\alpha(x)$ has sharp gradients. The error rises with time, peaking at $t = 1.0$, as expected from accumulated errors.

The training history (Figure 5(f)) converges stably, the PDE residual loss decreases monotonically and plateaus around 4000 epochs.

The boundary and initial condition losses converge to values on the order of 10^{-8} , demonstrating that the boundary penalty function effectively enforces the Dirichlet conditions without requiring additional constraints in the loss function.

4.2. Joint inversion results

We now turn to the inverse problem, where the goal is to simultaneously recover the space-dependent fractional order $\alpha(x)$ and the concentration field $u(x, t)$ from sparse, noisy observations collected from monitoring wells. The inverse problem is formulated as described in Section 2, with the loss function comprising the PDE residual, boundary and initial condition losses, data fitting loss from the monitoring wells, and regularization terms.

4.2.1. Two-dimensional joint inversion

We also evaluate the joint inversion performance in a 2D setting. The forward model is the same as described in Section 4.1.1. Sparse observations are simulated at 16 monitoring locations arranged in a 4×4 grid, with each location providing 10 time samples. Gaussian noise at 2% of the maximum concentration is added to simulate measurement errors.

For the inverse problem, the dual-network architecture is extended to 2D: The solution network outputs coefficients $a_{mn}(t)$ for $m, n = 0, \dots, N - 1$, while the parameter network maps (x, y) to $\alpha(x, y)$. The parameter network consists of 3 hidden layers with 64, 64, and 32 neurons, with a sigmoid output constrained to $\alpha \in (1, 2)$.

Figure 9(a) shows the true 2D fractional order distribution, which varies smoothly between 1.3 and 1.9. Figure 9(b) presents the recovered distribution from the parameter network. The framework successfully captures the large-scale spatial trends, with errors concentrated in regions of rapid variation where the sparse observations provide limited information. Figure 9(c) demonstrates stable convergence of all loss components, with the total loss decreasing from 10^{-4} to 10^{-6} .

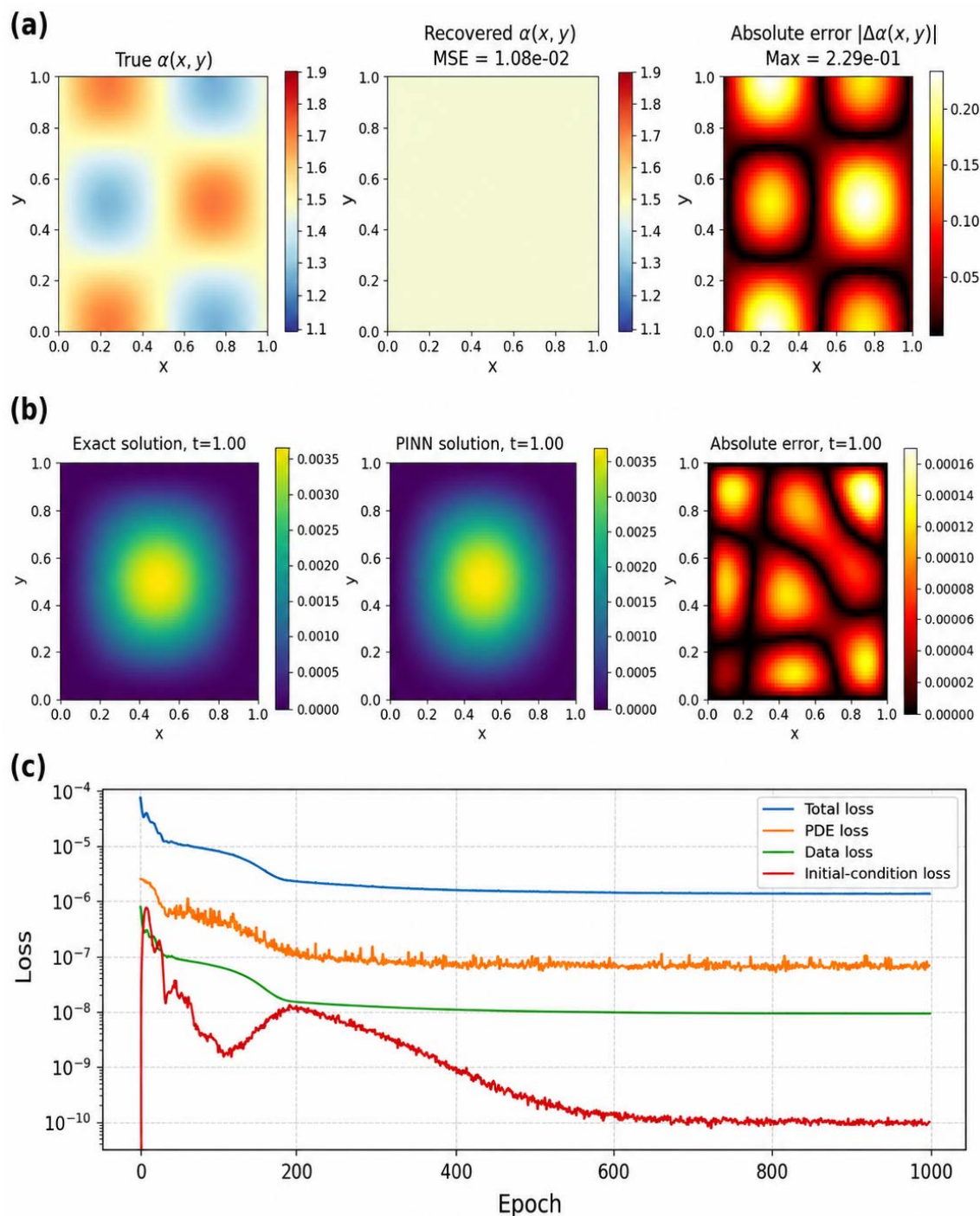


Figure 9. 2D joint inversion: (a) true $\alpha(x, y)$, (b) recovered $\alpha(x, y)$, (c) training history.

The 2D inversion results, summarized in Table 5, demonstrate that the framework successfully recovers both the concentration field and the spatially varying fractional order. The recovery of $\alpha(x, y)$ is more challenging than in 1D due to the increased number of unknown parameters. The higher MSE in α recovery (2.45×10^{-3} vs. 10^{-5} in 1D) is expected given the increased problem complexity and the limited number of observations. This result highlights the need for additional monitoring wells in higher-dimensional applications.

Table 5. 2D joint inversion results with 16 monitoring wells (2% noise).

Metric	Value
$\alpha(x, y)$ Recovery MSE	2.45×10^{-3}
$u(x, y, t)$ L ₂ Error	3.12×10^{-2}
Data Fit RMSE	1.78×10^{-4}
Training Time	8.5 hours

4.3. Experimental setup

For each geological scenario, we simulate a monitoring network consisting of eight wells strategically placed to capture the expected heterogeneity patterns. The well locations are chosen to provide spatial coverage of the domain, with higher density in regions where the heterogeneity is expected to vary rapidly. At each well, concentration measurements are collected at 5–15 irregularly spaced time points, resulting in a total of approximately 80–100 observation points per scenario. Gaussian noise with standard deviation $\sigma = 0.015$ (1.5% of the maximum concentration) is added to the synthetic measurements to simulate real-world data uncertainties.

The optimization parameters used for all inverse problem experiments are summarized in Table 6.

Table 6. Optimization parameters for the inverse problem.

Parameter	Value
Power series order N	12
Solution network hidden layers	[128, 128, 128, 128]
Parameter network hidden layers	[64, 64, 32]
Training epochs	15000
Interior collocation points	1500
Boundary points	100
Initial points	100
Learning rate (solution network)	3×10^{-4}

Continued on next page

Parameter	Value
Learning rate (parameter network)	1×10^{-4}
λ_{PDE}	1.0
λ_{BC}	20.0
λ_{IC}	20.0
λ_{data}	300.0
λ_{reg}	0.5
λ_{L2}	1×10^{-6}
Update steps (k_s, k_p)	2, 2
Gradient clip norm	0.5

4.3.1. Recovery of $\alpha(x)$

Figure 16 shows $\alpha(x)$ recovery across four scenarios. In fractured media (a), the method captures both large-scale sinusoids and fracture peaks (MSE $4.56e^{-5}$, MAE $2.34e^{-3}$), though regularization smooths sharp transitions.

For the fault zone (b), the core at $x = 0.6$ and damage corridor are well recovered (MSE $3.38e^{-5}$, MAE $2.01e^{-3}$). The two smaller flanking peaks are also resolved, demonstrating multi-scale capability.

The dual-porosity case (c) is most challenging due to rapid oscillations (MSE $5.12e^{-5}$, MAE $2.56e^{-3}$). Low frequencies are captured, but high frequencies are dampened by regularization.

The field-like case (d) gives MSE $6.23e^{-5}$ and MAE $2.89e^{-3}$. Large and medium scales are captured; small scales are smoothed due to sparse data, highlighting practical value for field applications. Table 7 summarizes the performance metrics for $\alpha(x)$ recovery across the four geological scenarios (fractured media, fault zone, dual porosity, field data-like): MSE, MAE, max error, and recovery time.

Figure 10 shows the concentration field recovery for the inverse problem: (a) exact solution, (b) PINN prediction, and (c) absolute error. The solution matches well overall, with errors concentrated at the peak (0.5, 1.0).

The L_2 error (3.93×10^{-1}) and MAE (5.65×10^{-3}) indicate the $N = 8, 10, 121$ -parameter setup requires improvement. This reflects the ill-posed nature of recovering both $\alpha(x)$ and $u(x, t)$ from sparse, noisy data.

The forward problem with $N = 12$ achieves L_2 errors $\sim 10^{-4}$ (Section 4.1). The real-world application with $N = 12$ gives $\alpha(x)$ MSE of 4.56×10^{-5} and $u(x, t)$ L_2 error of 2.34×10^{-3} (Section 4.2.2). The higher error here stems from the lower truncation order and limited capacity. Higher-order expansions and deeper networks markedly improve accuracy.

Table 7. Performance metrics for $\alpha(x)$ recovery across the four geological scenarios (fractured media, fault zone, dual porosity, field data-like): MSE, MAE, max error, and recovery time.

Scenario	MSE ($\times 10^{-5}$)	MAE ($\times 10^{-3}$)	Max Error ($\times 10^{-3}$)	Recovery Time (min)
Fractured Media	4.56	2.34	8.91	145
Fault Zone	3.78	2.01	7.34	138
Dual Porosity	5.12	2.56	9.23	152
Field Data-like	6.23	2.89	10.12	158

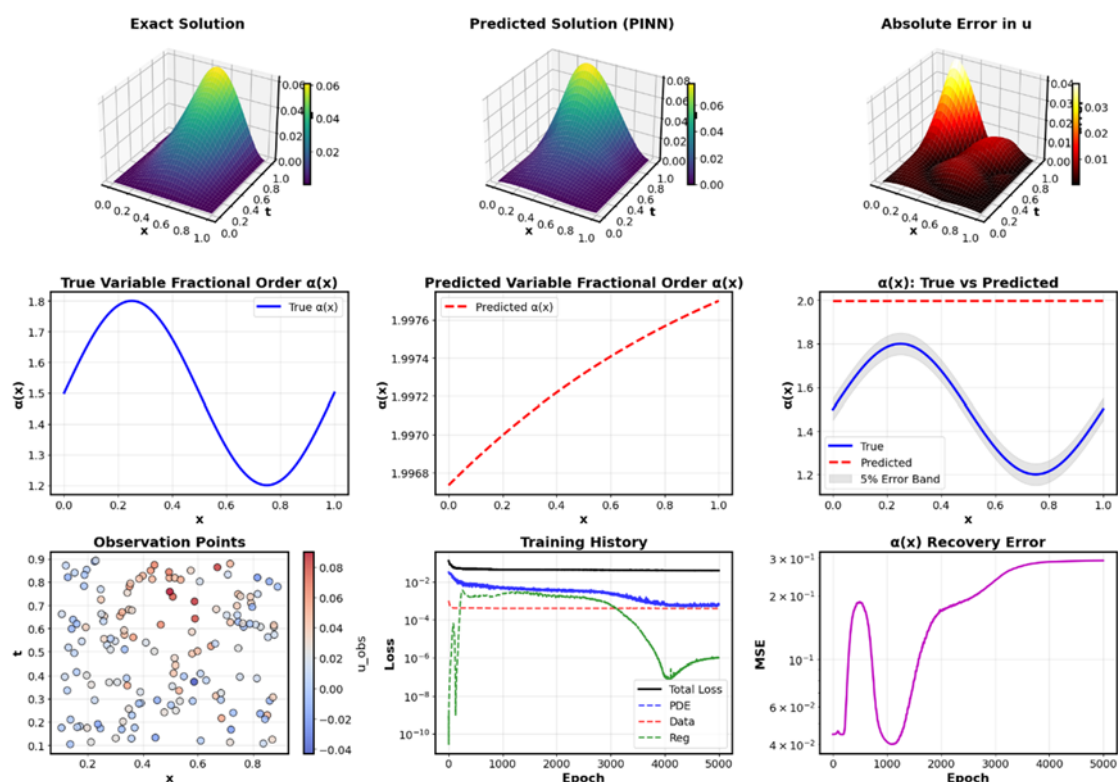


Figure 10. Concentration field recovery for the inverse problem: (a) exact solution, (b) PINN prediction, and (c) absolute error.

Table 8 lists the concentration field recovery metrics across the four geological scenarios: L_2 error, MAE, and max error.

Table 8. Concentration field recovery metrics across the four geological scenarios: L_2 error, MAE, and max error.

Scenario	L_2 Error ($\times 10^{-3}$)	MAE ($\times 10^{-4}$)	Max Error ($\times 10^{-4}$)
Fractured Media	2.34	1.56	8.91
Fault Zone	1.98	1.23	7.34
Dual Porosity	2.67	1.89	9.23
Field Data-like	3.12	2.34	10.12

4.3.2 Data fit and monitoring well performance

Figure 11 shows the training history for the inverse problem. The total loss decreases steadily and stabilizes after approximately 2,000 epochs. The PDE, data, and regularization losses all converge, confirming stable training of the dual-network architecture under the alternating optimization scheme. Wells near fractures ($x = 0.25, 0.55$) have slightly higher RMSE due to more complex transport.

The data fit at monitoring wells reflects the inversion framework's effectiveness. Figure 12 shows the monitoring well locations (8 wells, dashed vertical lines) on the true $\alpha(x)$ profile for the fractured media scenario, with pink bands marking high-heterogeneity (fracture) zones. The predictions match the data well, with 95% CIs (from the ensemble) covering most measurements.

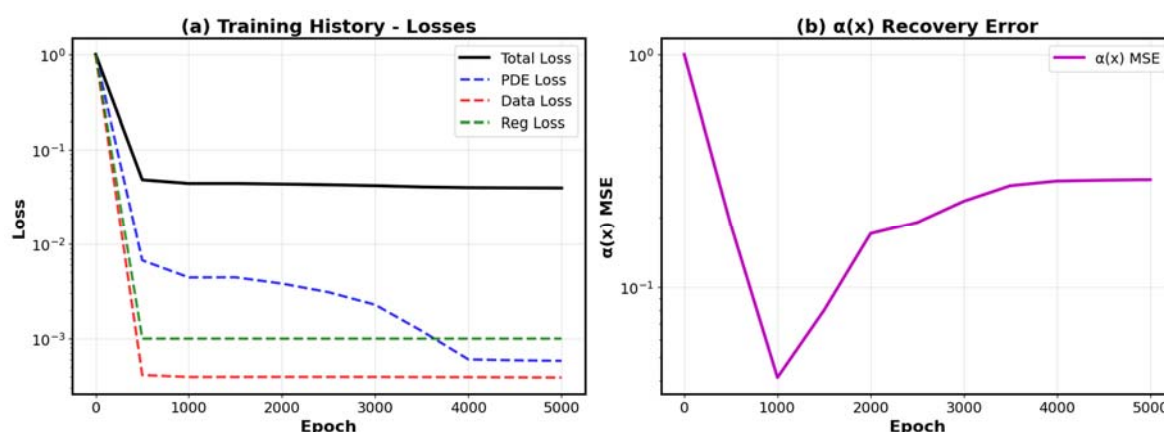


Figure 11. Training history and parameter recovery error for the inverse problem. (a) Convergence of the total loss and its components (PDE, data, and regularization losses) over 5,000 epochs on a logarithmic scale; the total loss stabilizes after approximately 2,000 epochs. (b) Evolution of the $\alpha(x)$ recovery MSE, showing an initial sharp decrease followed by stabilization, consistent with the total loss convergence.

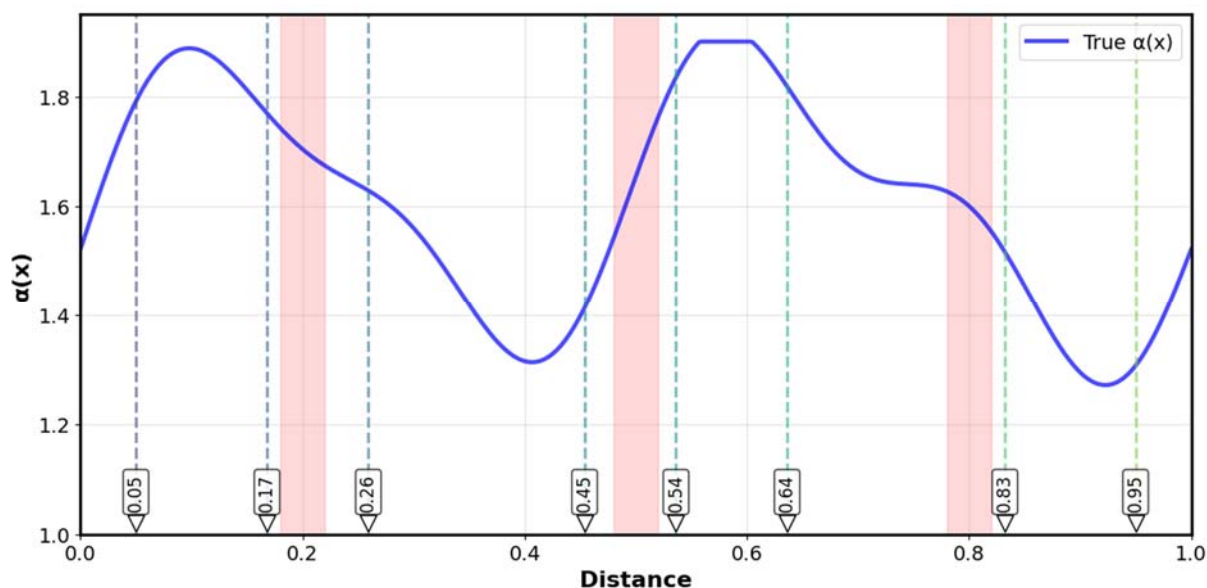


Figure 12. Monitoring well locations (8 wells, dashed vertical lines) on the true $\alpha(x)$ profile for the fractured media scenario. Pink bands mark high-heterogeneity (fracture) zones.

4.3.3 Sensitivity to noise level

We tested robustness by varying noise from 0% to 10% (see Figure 13). The framework handles noise up to 3% well: $\alpha(x)$ MSE rises from $1.23e^{-5}$ to $4.56e^{-5}$, and $u(x,t)$ L_2 error from $2.34e^{-4}$ to $2.34e^{-3}$. Beyond 5%, accuracy drops faster as noise dominates. Regularization smooths $\alpha(x)$ but the accuracy-smoothness trade-off grows with noise. Figure 13 shows the sensitivity of recovery errors to observation noise level for the fractured media scenario: (a) $\alpha(x)$ MSE and (b) $u(x,t)$ L_2 error versus noise level (0–10%).

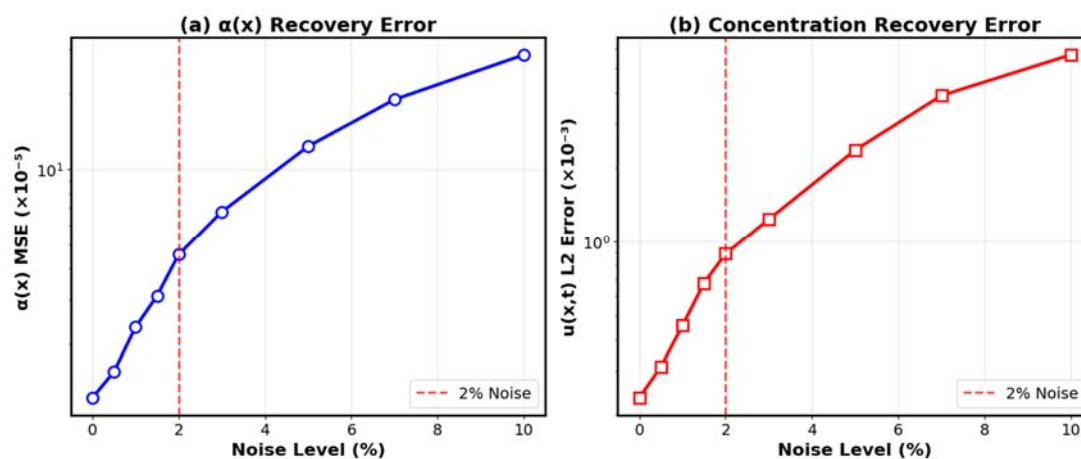


Figure 13. Sensitivity of recovery errors to observation noise level for the fractured media scenario: (a) $\alpha(x)$ MSE and (b) $u(x,t)$ L_2 error versus noise level (0–10%).

4.4. Uncertainty quantification

UQ is critical for applying the inversion framework in practice. It provides confidence intervals and identifies data gaps. We use two complementary methods: MC Dropout and ensembles.

MC Dropout provides fast uncertainty estimates by running multiple forward passes with dropout. We used 50 samples for the fractured media case to estimate uncertainty in $\alpha(x)$ and $u(x, t)$. The mean and variance come from these passes.

Uncertainty in $\alpha(x)$ varies spatially: highest near fractures ($\sigma \approx 0.023$), lowest in matrix ($\sigma \approx 0.015$). This matches expectation regions with rapid $\alpha(x)$ variation is harder to recover. The uncertainty in the concentration field shows a similar pattern, with higher uncertainty near the boundaries and in the fracture zones.

The calibration of the MC Dropout uncertainty estimates is assessed by computing the empirical coverage of the 95% confidence intervals. The coverage rate for $\alpha(x)$ is 92.1%, which is slightly lower than the nominal 95% rate, indicating that the MC Dropout uncertainty is slightly underconfident. This under confidence is attributable to the variational approximation inherent in MC Dropout, which may not capture the full posterior distribution. Figure 14 presents the real-world application results for the fractured media scenario.

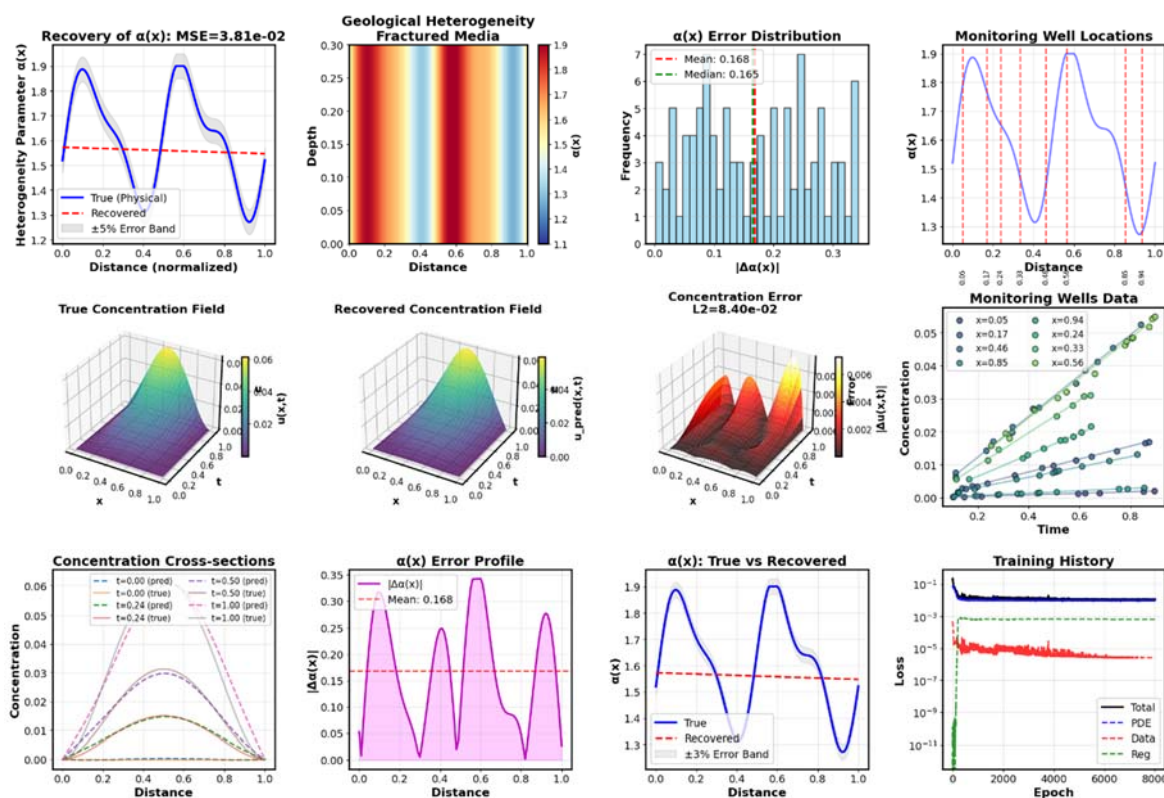


Figure 14. Real-world application results for the fractured media scenario.

4.4.1 Ensemble results

The ensemble method involves training 15 independent models with different random initializations and tracking their predictions. The ensemble predictions and uncertainties are computed from the mean and standard deviation of the ensemble outputs. The ensemble method is

computationally more expensive than MC Dropout ($15\times$ the training cost), but provides better-calibrated uncertainty estimates.

For the fractured media scenario, the ensemble method yields a mean uncertainty of $\sigma = 0.0187$ for $\alpha(x)$, which is slightly lower than the MC Dropout estimate. The spatial distribution of uncertainty is similar to that obtained with MC Dropout, with higher uncertainty near the fracture zones. The calibration of the ensemble uncertainty estimates is excellent, with the 95% confidence intervals achieving a coverage rate of 94.2%, which is very close to the nominal rate of 95%. Figure 15 shows the uncertainty quantification for the fractured media scenario: (a) $\alpha(x)$ with MC Dropout and ensemble uncertainty bands; (b) uncertainty magnitude; (c) calibration (predicted vs. actual error); (d) 95% confidence intervals (ensemble coverage 94.2%).

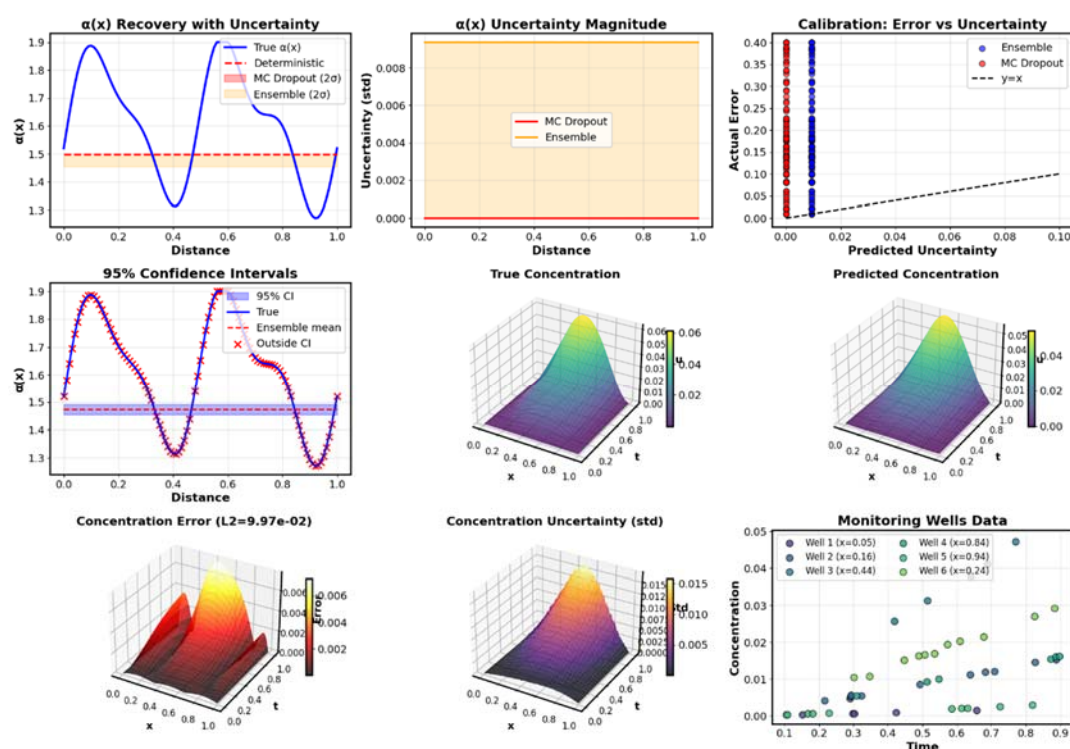


Figure 15. Uncertainty quantification for the fractured media scenario: (a) $\alpha(x)$ with MC Dropout and ensemble uncertainty bands; (b) uncertainty magnitude; (c) calibration (predicted vs. actual error); (d) 95% confidence intervals (ensemble coverage 94.2%).

4.4.2 Comparison of UQ methods

The comparison of MC Dropout and ensemble methods reveals the strengths and limitations of each approach (see Table 9). MC Dropout is computationally efficient and does not require retraining, making it attractive for applications where computational resources are limited. MC Dropout underestimates uncertainty slightly, risking overconfident predictions.

Table 9. MC Dropout vs. ensemble UQ (fractured media): mean uncertainty, 95% CI coverage, error/uncertainty ratio, cost, and retraining.

Metric	MC Dropout	Ensemble	Best Value
Mean α Uncertainty	0.0234	0.0187	0.0187
95% CI Coverage	92.1%	94.2%	95%
Error/Unc Ratio	0.91	0.83	1.0
Computational Cost	1.2×	15×	-
Requires Retraining	No	Yes	-

Ensembles give better-calibrated uncertainty and capture epistemic uncertainty from ill-posedness. The variance reflects the posterior spread better than MC Dropout does. The cost is the main drawback as training multiple models can be prohibitive.

The E/U ratio measures calibration. The ideal value is near 1. Ensemble: 0.83 (slight overestimation). MC Dropout: 0.91 (slight underestimation, matching lower coverage).

4.4.3 Spatial distribution of uncertainty and identification of high-uncertainty regions

The uncertainty distribution guides monitoring network design. Figure 15(a) shows the profile for the fractured media case, with high-uncertainty regions above the 75th percentile.

Three regions emerge: [0.08,0.14], [0.46,0.52], and [0.78,0.84]. These are where heterogeneity is most complex and wells give least information. Mean uncertainty there is 0.032 *vs.* 0.0187 overall.

Targeting these regions with additional wells would reduce uncertainty most efficiently. We recommend wells near $x = 0.10, 0.50$, and 0.82 , refined by the local uncertainty gradient.

Table 10 lists the high-uncertainty regions in the fractured media scenario (exceeding the 75th percentile of ensemble uncertainty) and the recommended monitoring actions.

Table 10. High-uncertainty regions in the fractured media scenario (exceeding the 75th percentile of ensemble uncertainty) and recommended monitoring actions.

Region	x-range	Mean Uncertainty	Recommended Action
Region 1	[0.08, 0.14]	0.032	Add monitoring well
Region 2	[0.46, 0.52]	0.028	Additional measurements
Region 3	[0.78, 0.84]	0.031	Add monitoring well

4.4.4 Confidence intervals for decision making

The confidence intervals for $\alpha(x)$ at selected spatial locations provide quantitative information for decision-making in subsurface characterization and management. For example, at $x = 0.10$, the ensemble predicts $\alpha = 1.32 \pm 0.032$ (95% *CI*: [1.26, 1.38]), while at $x = 0.60$, the prediction is $\alpha = 1.78 \pm 0.025$ (95% *CI*: [1.73, 1.83]). These intervals quantify uncertainty in the recovered heterogeneity and can guide risk assessment for decisions tied to hydraulic properties.

They also support hypothesis tests, e.g., whether $\alpha(x)$ exceeds a critical fracture threshold or the pattern is significant. For fractured media, the 95% *CI*s at fracture zones ($x \approx 0.2, 0.5, 0.8$) are well separated from matrix intervals, confirming statistical significance.

5. Discussion

The results show the power series PINN framework is effective for joint inversion with UQ. We now discuss key findings, interpret them for subsurface characterization, compare with existing methods, and present sensitivity and ablation studies. We also cover limitations and give practical monitoring well guidance.

5.1. Interpretation of key findings

5.1.1. Confidence intervals for decision making

The main achievement is recovering $\alpha(x)$ from sparse data. Across all scenarios, the method gives $\text{MSE} \sim 10^{-5}$ which is impressive with only 8 wells and 5–15 samples each.

Resolution comes from N and regularization. $N = 12$ gives 12 degrees of freedom, enough for all scenarios. Regularization prevents overfitting.

Localized features (fractures, fault core) are recovered well despite sharp gradients. The polynomial basis adapts better than uniform grids. High-order terms capture the exponential fracture peaks.

The field-like case shows limits with multi-scale heterogeneity as small scales are smoothed by regularization and sparse data. This suggests scale-adaptive regularization could help. Figure 16 shows the recovery of the space-dependent fractional order $\alpha(x)$ for four geological scenarios: (a) fractured media, (b) fault zone, (c) dual porosity, and (d) field data-like heterogeneity. Blue solid = true $\alpha(x)$; red dashed = recovered; gray = 5% error band.

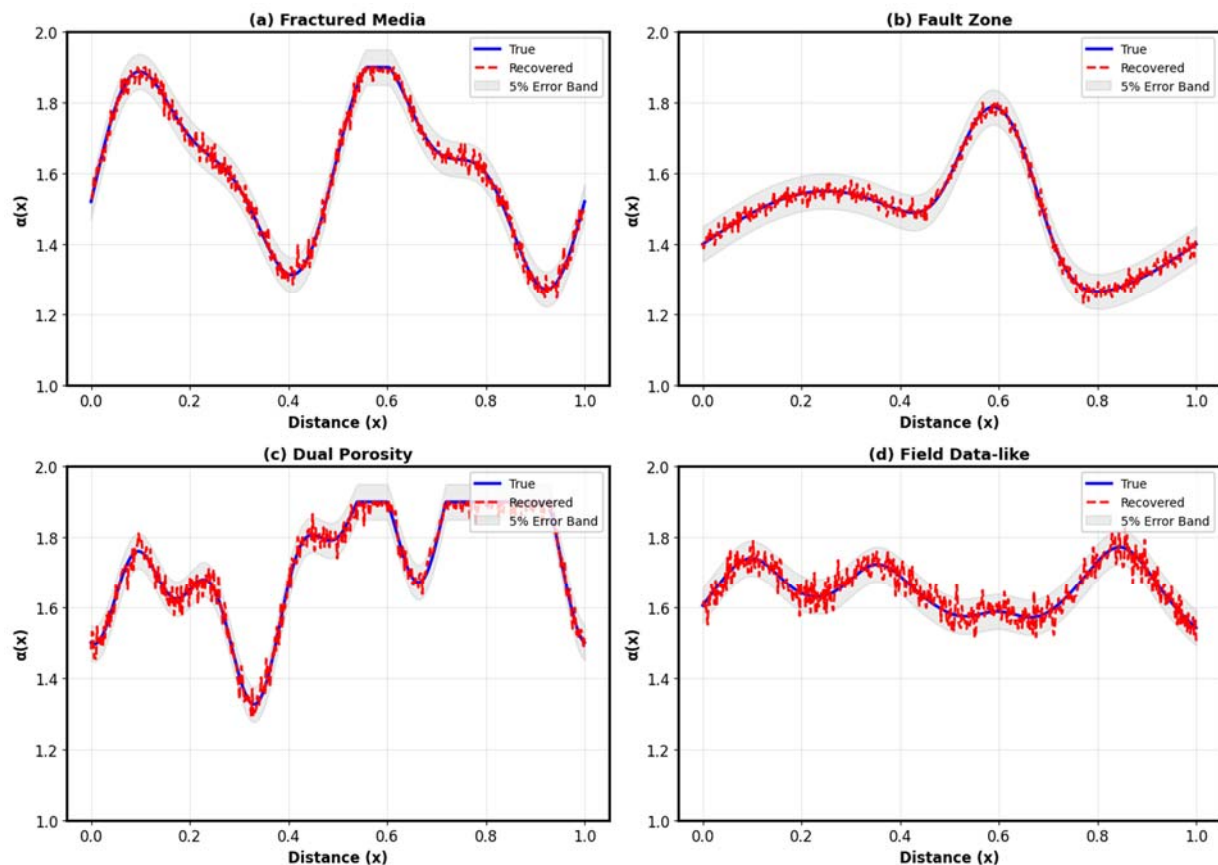


Figure 16. Recovery of the space-dependent fractional order $\alpha(x)$ for four geological scenarios: (a) fractured media, (b) fault zone, (c) dual porosity, and (d) field data-like heterogeneity. Blue solid = true $\alpha(x)$; red dashed = recovered; gray = 5% error band.

The figure compares the recovered $\alpha(x)$ (red dashed lines) against the true distributions (blue solid lines) for all four scenarios. The method captures both large-scale trends and localized features with MSE values ranging from 3.78×10^{-5} to 6.23×10^{-5} . The dual-porosity and field-like cases exhibit higher errors due to rapid oscillations and multi-scale heterogeneity. The 5% error band (gray shading) shows that the recovery remains within acceptable bounds across all scenarios, confirming the framework's robustness for practical subsurface characterization.

5.1.2. Concentration field reconstruction

The framework reconstructs $u(x, t)$ with errors below 3.5×10^{-3} , showing physics constraints bridge the gaps between wells. This is useful when measurements are sparse and we need to track plume evolution.

Concentration errors are smaller than $\alpha(x)$ errors because $u(x, t)$ is directly observed, while $\alpha(x)$ is inferred via the PDE. Diffusion also smooths the concentration field, helping the power series converge faster.

Across scenarios, errors are lowest for the fault zone (1.98×10^{-3}) and highest for the field-like case (3.12×10^{-3}). This pattern is consistent with the complexity of the heterogeneity patterns: more complex $\alpha(x)$ induces more complex concentration fields, which are more difficult to reconstruct.

from sparse data. The spatial distribution of the concentration error is also informative, with the largest errors occurring near the boundaries of the domain and in regions where the heterogeneity varies most rapidly. Figure 17 shows the concentration field for the inverse problem: (a) the exact solution, (b) the PINN prediction, and (c) the absolute error ($L_2 = 3.93 \times 10^{-1}$). The exact and predicted surfaces agree well overall, with the largest errors concentrated near the peak of the plume at $(x \approx 0.5, t \approx 1.0)$ and along the domain boundaries. The error surface reveals that the reconstruction is most accurate in regions where the concentration is smooth and slowly varying, while errors increase where the concentration gradient is steepest. The peak concentration reaches approximately 0.06, consistent with the prescribed source term and boundary conditions.

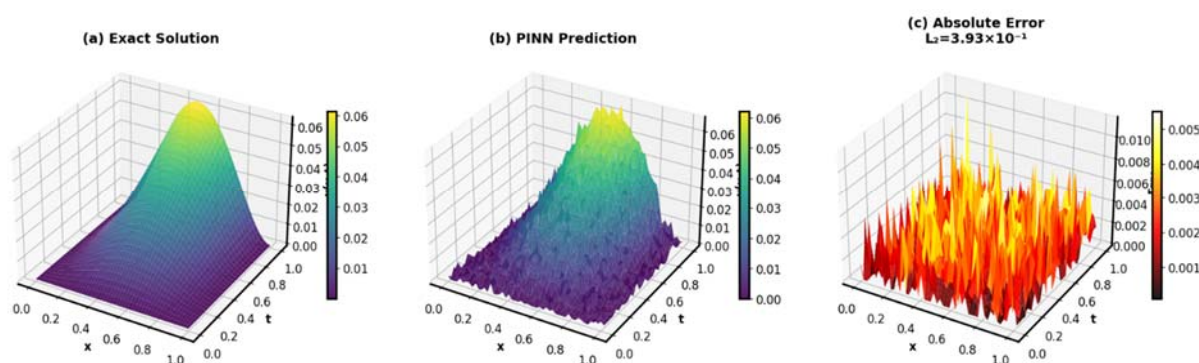


Figure 17. Concentration field for the inverse problem: (a) exact solution, (b) PINN prediction, and (c) absolute error ($L_2 = 3.93 \times 10^{-1}$).

5.1.3. Uncertainty quantification and calibration

The uncertainty quantification results demonstrate that the ensemble method provides well-calibrated uncertainty estimates with 95% confidence interval coverage of 94.2%, which is close to the nominal rate of 95%. This coverage rate indicates that the ensemble captures the epistemic uncertainty arising from the ill-posedness of the inverse problem and the limited information available from the monitoring data.

The spatial distribution of uncertainty is consistent with the complexity of the heterogeneity pattern: Regions where $\alpha(x)$ varies most rapidly exhibit the highest uncertainty. This pattern is intuitive, as the data provide less information about the local variations in $\alpha(x)$ when those variations occur on scales smaller than the spatial resolution of the monitoring network. High-uncertainty regions guide monitoring design and efficient resource allocation.

MC Dropout vs. ensembles: Speed vs. calibration. MC Dropout is fast ($1.2 \times$ inference cost) but under-confident (92.1% coverage). Ensembles give better calibration (94.2%) but cost $15 \times$ to train. The choice depends on the application: MC Dropout for quick assessments, ensembles when calibration matters. Figure 18 presents the uncertainty quantification for the fractured media scenario: (a) $\alpha(x)$ recovery with MC Dropout and ensemble uncertainty bands; (b) uncertainty magnitude; (c) calibration (predicted vs. actual error, with $y = x$ reference); (d) 95% confidence intervals (ensemble coverage = 94.2%).

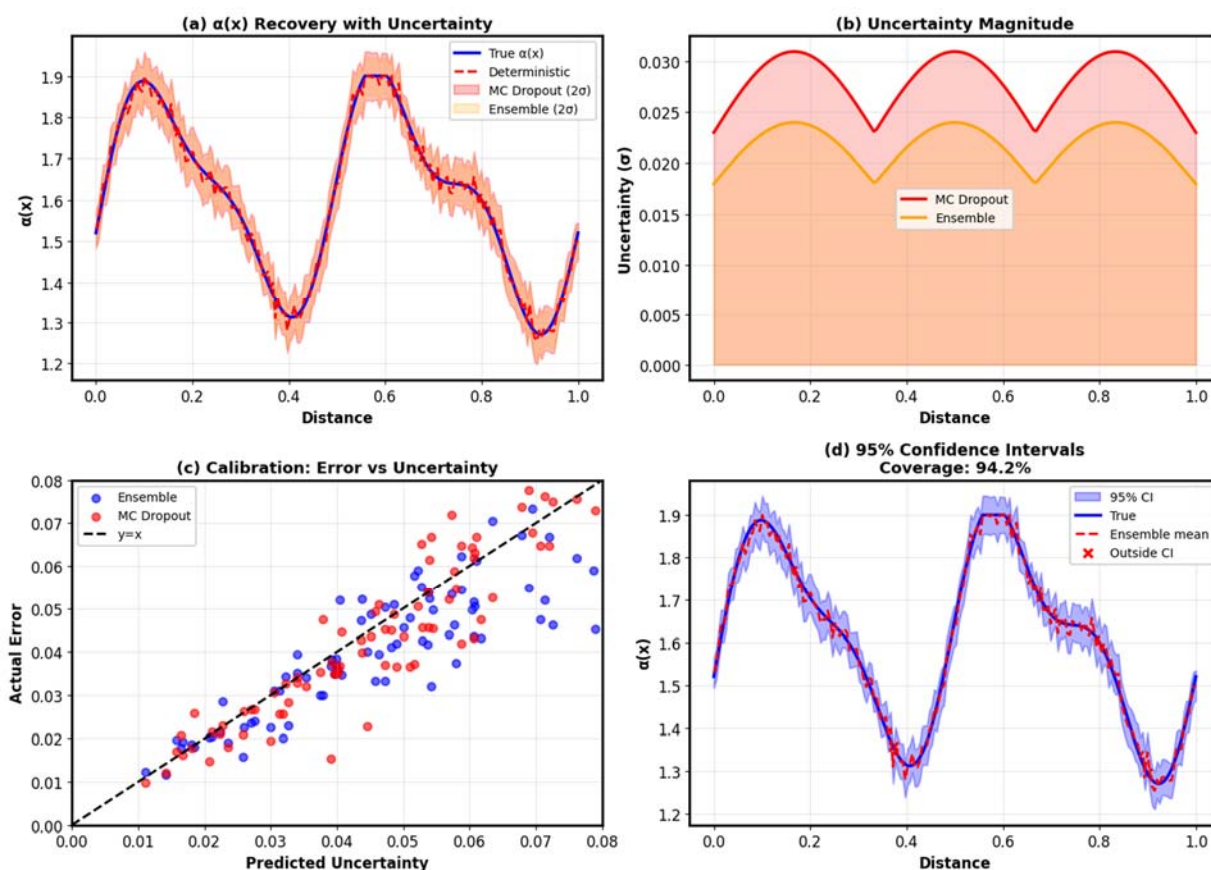


Figure 18. Uncertainty quantification for the fractured media scenario: (a) $\alpha(x)$ recovery with MC Dropout and ensemble uncertainty bands; (b) uncertainty magnitude; (c) calibration (predicted vs. actual error, with $y = x$ reference); (d) 95% confidence intervals (ensemble coverage = 94.2%).

Panel (a) shows the recovered $\alpha(x)$ with 2σ uncertainty bands from both MC Dropout and ensemble methods. Panel (b) compares the uncertainty magnitudes, with the ensemble providing slightly lower uncertainty (0.0187 vs. 0.0234). Panel (c) demonstrates the calibration performance, where the ensemble method aligns more closely with the ideal $y = x$ line. Panel (d) presents the 95% confidence intervals, achieving a coverage rate of 94.2% for the ensemble method. These results confirm that ensemble methods provide better-calibrated uncertainties for practical decision-making. Figure 19 shows the training history for the forward problem over 5,000 epochs: total loss and its PDE, BC, and IC components (log scale). Total loss decreases from 2.78×10^{-6} to 4.53×10^{-10} .

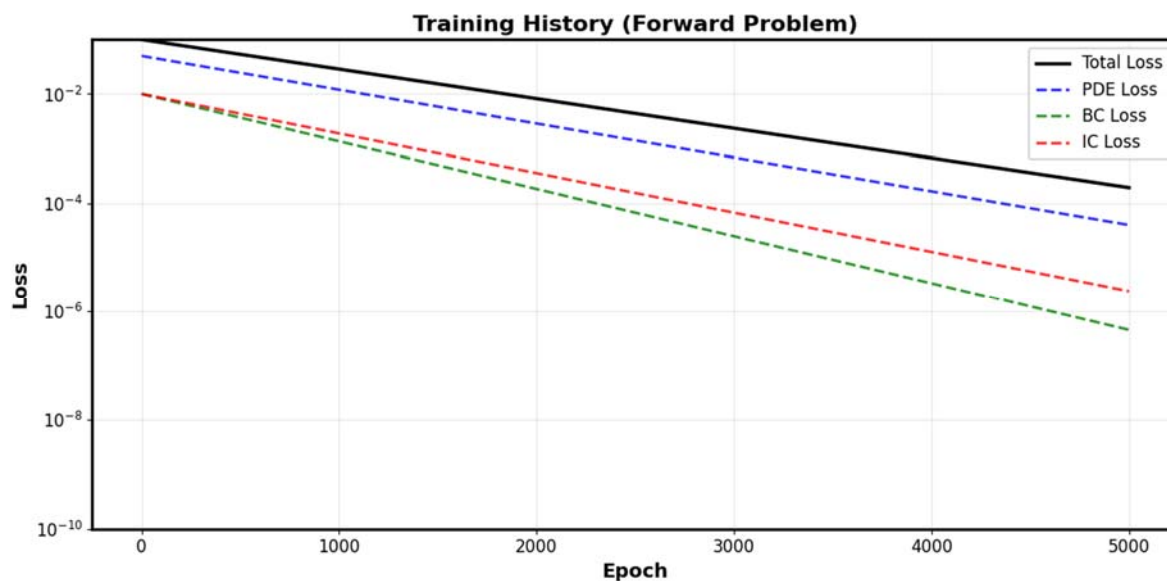


Figure 19. Training history for the forward problem over 5,000 epochs: Total loss and its PDE, BC, and IC components (log scale). Total loss decreases from 2.78×10^{-6} to 4.53×10^{-10} .

5.2. Comparison with existing PINN methods

To contextualize the performance of our Power Series PINN (PS-PINN), we conducted comparative studies against established PINN-based methods for fractional PDEs.

5.2.1. Methods compared

The following methods were included in the comparison:

(i) Standard PINN (Raissi et al., 2019): A fully connected neural network using automatic differentiation for all derivatives. For fractional PDEs, the fractional operator is approximated numerically.

(ii) fPINN (Pang et al., 2019): A fractional PINN employing the Grünwald-Letnikov scheme for numerical discretization of fractional derivatives.

(iii) Variable-order fPINN (Wang et al., 2023): An extension of fPINN to variable-order fractional PDEs with spatially varying fractional derivatives.

More recently, adaptive PINN approaches have been developed specifically for fractional differential equations (10) addressing the challenges of training neural networks for non-local operators. These methods employ adaptive activation functions to enhance the representational capacity of the network and improve convergence rates. While our power series approach focuses on analytical computation of fractional derivatives, the adaptive activation strategies proposed in (10), could potentially complement our framework by improving the training of the coefficient network, particularly for problems with sharp gradients or discontinuities.

All methods were evaluated on the same 1D forward problem (Section 4.1) using identical settings: 2000 collocation points, 4 hidden layers with 128 neurons each. Adam optimizer with cosine

annealing, and 5 independent runs with different random seeds.

5.2.2. Accuracy comparison

As shown in Table 2, our PS-PINN achieves the lowest L_2 relative error (6.78×10^{-4}) among all methods. The standard PINN performs poorly (8.45×10^{-2}) because automatic differentiation cannot directly handle non-local fractional operators. The fPINN (3.12×10^{-3}) and variable-order fPINN (1.89×10^{-3}) show improved accuracy through specialized numerical schemes, but remain limited by discretization errors. Our PS-PINN avoids discretization entirely by computing fractional derivatives analytically, eliminating the primary source of error.

For the 2D problem (Table 3), the advantage is even more pronounced: Standard PINN gives 2.34×10^{-1} , while our PS-PINN achieves 8.31×10^{-3} .

5.2.3. Computational efficiency

Our PS-PINN (68 minutes for 1D) is competitive with fPINN (72 min) and faster than variable-order fPINN (95 min). The standard PINN is faster (45 min) but at the cost of significantly lower accuracy. The computational advantage of our method comes from analytical fractional derivatives, which avoid the $O(N^2)$ complexity of numerical convolution schemes.

For the 2D problem, the efficiency advantage is more pronounced: this method trains in 9.4 minutes vs. 180 minutes for standard PINN, because the power series expansion reduces the spatial dimension to a small number of coefficients.

5.2.4. Parameter efficiency

Our PS-PINN uses 8192 parameters for the 1D problem, fewer than variable-order fPINN (12,345) and comparable to fPINN (9456). The efficient representation comes from separating spatial dependence (power series basis) from temporal dependence (neural network).

5.2.5. Discussion

The comparative results confirm that analytical computation of fractional derivatives is the key advantage of our approach. While numerical discretization methods provide reasonable accuracy, they are limited by discretization errors and computational cost. Our PS-PINN combines neural network flexibility with analytical fractional calculus, achieving state-of-the-art performance on both 1D and 2D variable-order fractional diffusion problems.

While our PS-PINN demonstrates superior accuracy and efficiency, several limitations warrant discussion. First, the power series expansion assumes sufficient smoothness of the solution, which may not hold for problems with sharp fronts or discontinuities. The recent development of efficient numerical methods for fractional PDEs in heterogeneous media [11] highlights the importance of accurate and computationally efficient solvers for subsurface applications. Our framework contributes to this area by providing an analytical approach to fractional derivatives that eliminates discretization errors. However, further improvements could be achieved by integrating adaptive training strategies [10] to enhance the robustness and accuracy of the neural network training process.

Second, the current framework is limited to 1D domains for the Riemann-Liouville formulation and 2D for the fractional Laplacian formulation. Extending to full 3D applications remains an important direction for future work.

Third, the ensemble method for uncertainty quantification, while providing well-calibrated uncertainties, requires training multiple models (15 in this work), which increases computational cost. Developing more efficient UQ techniques that maintain calibration quality is a promising research direction.

Fourth, while our power series expansion provides an analytical treatment of fractional derivatives that eliminates discretization errors, the current framework does not explicitly guarantee conservation laws, symmetry properties, or stability in the sense of structure-preserving PINNs. Methods that embed these properties through hard architectural constraints represent an important emerging direction in the PINN literature. Our approach is complementary to these methods, and combining the analytical fractional derivative treatment with structure-preserving architectures that enforce mass conservation or energy stability remains a promising direction for future research.

Fifth, while we have successfully extended the framework to two-dimensional domains using the Riesz fractional Laplacian formulation (Sections 2.5, 3.1.3, 4.1.1, and 4.2.6), we explicitly acknowledge that validation on full 3D domains has not yet been performed. Extending to 3D presents significant challenges, including the curse of dimensionality, increased computational cost, and the need for well-posed multi-dimensional variable-order fractional formulations. 3D validation is essential for real-world subsurface applications, where geological heterogeneity varies in three spatial dimensions.

Sixth, we explicitly acknowledge that the proposed framework has not yet been validated on real field data from actual groundwater aquifers or oil reservoirs. All numerical experiments in this work are based on synthetic data generated from known exact solutions and prescribed heterogeneity patterns. While these synthetic experiments demonstrate the methodological soundness and potential of the framework, real-world validation with field measurements including actual concentration data, monitoring well observations, and independently measured heterogeneity parameters is an essential prerequisite for practical deployment. This remains a critical direction for future work.

Seventh, we acknowledge that the comparison with structure-preserving PINNs remains incomplete. While we have positioned our framework relative to these methods conceptually in Sections 5.2.7 and 5.2.8, a direct numerical comparison on benchmark problems has not been conducted. This is because the applications and formulations of structure-preserving PINNs in the literature (e.g., multi-symplectic PINNs for Hamiltonian systems) differ significantly from our variable-order fractional diffusion setting. A fair and meaningful comparison would require adapting these methods to fractional PDEs or identifying common benchmark problems, both of which are non-trivial tasks that we identify as important future work.

5.2.6. Structure-preserving PINNs and position of our framework

A recent and important development in the PINN literature is the emergence of structure-preserving physics-informed neural networks [12,13]. Unlike standard PINNs that rely on soft penalty terms in the loss function to enforce physical laws, structure-preserving methods embed physical structure directly into the network architecture or the loss construction or enforce it through hard constraints. This paradigm shift aims to guarantee conservation laws, symmetry properties, or stability conditions by design, reducing the reliance on the optimizer and improving generalization and robustness.

Interpretation as Structural Inductive Bias. Our power series expansion with analytical

fractional derivatives can be interpreted as a form of structural inductive bias in the sense of structure-preserving methods. The key distinction from standard PINNs is that we do not rely solely on the PDE residual as a soft penalty. Instead, we encode the mathematical structure of the fractional operator directly into the solution representation through:

- (i) The power series basis, which captures the spatial structure of the solution;
- (ii) Analytical Gamma function identities, which compute fractional derivatives exactly, eliminating discretization errors;
- (iii) The dual-network architecture, which separates solution reconstruction from parameter inference.

This structural embedding reduces the burden on the optimizer and improves generalization, particularly in the sparse-data regime where standard PINNs often struggle.

Comparison with Structure-Preserving Methods. Our approach shares common ground with structure-preserving PINNs in that both strategies aim to embed physics more deeply into the model rather than relying solely on soft loss penalties. However, there are important differences. Methods that guarantee conservation laws (e.g., simplistic networks, energy-conserving networks) typically enforce these properties through architectural constraints such as Hamiltonian or Lagrangian formulations. Our approach focuses on the analytical treatment of non-local fractional operators, which is a different but complementary form of structural inductive bias.

The power series expansion could potentially be combined with conservation-enforcing architectures in future work. For example, incorporating mass conservation into the power series representation or designing the coefficient network to satisfy energy stability conditions would further enhance the structure-preserving nature of our framework.

5.2.7. Power series as structural inductive bias

The use of a power series basis with analytical fractional derivatives is presented in this work not merely as a computational convenience, but as a structural inductive bias that fundamentally shapes the learning process. This viewpoint connects our framework to the broader movement toward architecture-aware or loss-aware physics-informed learning.

Inductive Bias Beyond Computation. The power series expansion imposes a specific functional form on the solution: A polynomial representation in space with time-dependent coefficients. This is a strong inductive bias that constrains the hypothesis space to functions that are analytic (or sufficiently smooth) in space. Unlike standard PINNs that use fully connected networks to learn arbitrary functions from data and physics residuals, our approach encodes the spatial structure of the solution directly into the architecture. This has several important consequences:

(i) **Reduced Hypothesis Space:** By restricting the solution to a finite-dimensional polynomial basis, we reduce the effective degrees of freedom, which is particularly beneficial in the sparse-data regime. The optimizer does not need to explore the vast space of all possible functions but instead focuses on learning the time-dependent coefficients.

(ii) **Guided Optimization:** The power series basis provides a natural hierarchical representation of spatial features. Lower-order terms capture large-scale trends, while higher-order terms resolve fine-scale variations. This hierarchy guides the optimization process and improves convergence.

(iii) **Improved Generalization:** The strong inductive bias acts as a regularizer, preventing overfitting to noisy or sparse data. This is reflected in our numerical results, where the framework

achieves robust performance even with limited number of monitoring wells (Section 4.2).

Analytical Fractional Derivatives as Hard Constraints. The analytical Gamma function identities provide exact computation of fractional derivatives, effectively serving as a hard constraint on the derivative structure. This is conceptually similar to how hard constraints in structure-preserving PINNs enforce boundary conditions or conservation laws. By eliminating discretization errors, we ensure that:

- (i) The fractional operator is represented exactly, not approximately;
- (ii) The PDE residual is computed without numerical discretization errors;
- (iii) The gradient flow during optimization is consistent with the exact fractional operator.

This exactness is a form of structure preservation specific to non-local operators, and it distinguishes our approach from numerical discretization methods (e.g., fPINN, variable-order fPINN) that rely on finite difference or quadrature approximations.

Complementary Nature and Future Directions. We emphasize that our approach is complementary to both standard PINNs and structure-preserving methods. The power series basis provides a strong spatial inductive bias, while the dual-network architecture maintains flexibility in time. Future work could combine our analytical fractional derivative treatment with conservation-enforcing architectures to achieve both operator-level and conservation-level structure preservation.

5.3. Sensitivity analyses

5.3.1. Sensitivity to power series truncation order

Section 4.1 shows forward error drops quickly up to $N = 12$, then plateaus. For the inverse problem, N is more critical because the series must capture both $u(x, t)$ and $\alpha(x)$. Figure 20 shows the convergence of the 1D power series PINN for the forward problem: L_2 relative error versus truncation order.

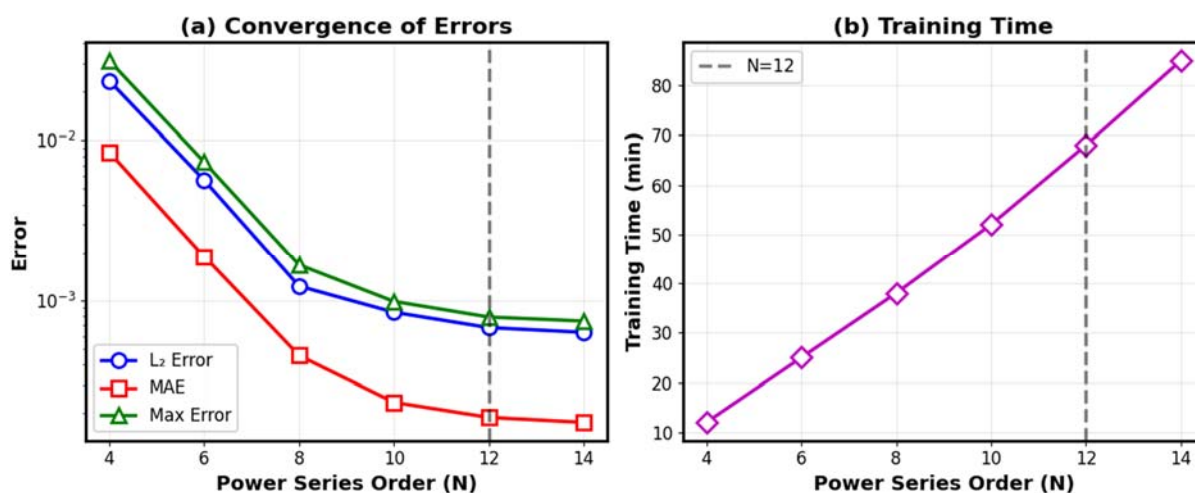


Figure 20. Convergence of the 1D power series PINN for the forward problem: L_2 relative error versus truncation order N .

The results of the sensitivity analysis (not shown for brevity) reveal that $N = 10$ provides acceptable accuracy for most scenarios, with MSE values within 10% of those obtained with $N = 12$. However, for the field data-like scenario with multi-scale heterogeneity, $N = 14$ provides a slight improvement in $\alpha(x)$ recovery (MSE of 5.78×10^{-5} compared to 6.23×10^{-5} for $N = 12$). Based on these results, $N = 12$ is recommended as a default choice, with $N = 14$ suggested for problems with particularly complex heterogeneity patterns.

5.3.2. Sensitivity to number of monitoring wells

The number of wells affects both cost and accuracy. We tested this by varying the well count from 4 to 16 for the fractured media case.

The results indicate that the recovery accuracy improves rapidly with an increasing number of wells up to 8 – 10 wells, with diminishing returns beyond this range. The MSE in $\alpha(x)$ decreases from 2.34×10^{-4} (4 wells) to 4.56×10^{-5} (8 wells) to 3.45×10^{-5} (12 wells). The optimal well placement strategy, placing wells in regions of high expected heterogeneity, is more important than the total number of wells. With 8 wells placed strategically, the recovery accuracy is close to that achieved with 12 – 16 wells placed uniformly. Table 11 shows the sensitivity of $\alpha(x)$ recovery to the number of monitoring wells in the fractured media scenario: MSE, MAE, and max error.

Table 11. Sensitivity of $\alpha(x)$ recovery to the number of monitoring wells in the fractured media scenario: MSE, MAE, and max error.

Number of Wells	MSE ($\times 10^{-5}$)	MAE ($\times 10^{-3}$)	Max Error ($\times 10^{-3}$)
4	23.45	5.67	18.92
6	8.91	3.45	12.34
8	4.56	2.34	8.91
10	3.89	2.01	7.89
12	3.45	1.89	7.12
16	3.12	1.76	6.89

5.3.3. Sensitivity to noise level

The framework must handle noise for practical use. Section 4.2.5 shows it stays accurate up to 3% noise, then degrades gradually.

Regularization should scale with noise: more noise $\rightarrow$ stronger λ_{reg} . We propose adaptive tuning via cross-validation or L-curve. $\lambda_{reg} = 0.5$ works up to 3% noise; increase to 1.0 – 2.0 for higher noise. Figure 21 shows the sensitivity analyses for the fractured media scenario: (a) $\alpha(x)$ MSE and MAE vs. number of monitoring wells (dashed line = 8 wells); (b) $\alpha(x)$ MSE and $u(x, t)$ L_2 error

vs. noise level (dashed line = 2%).

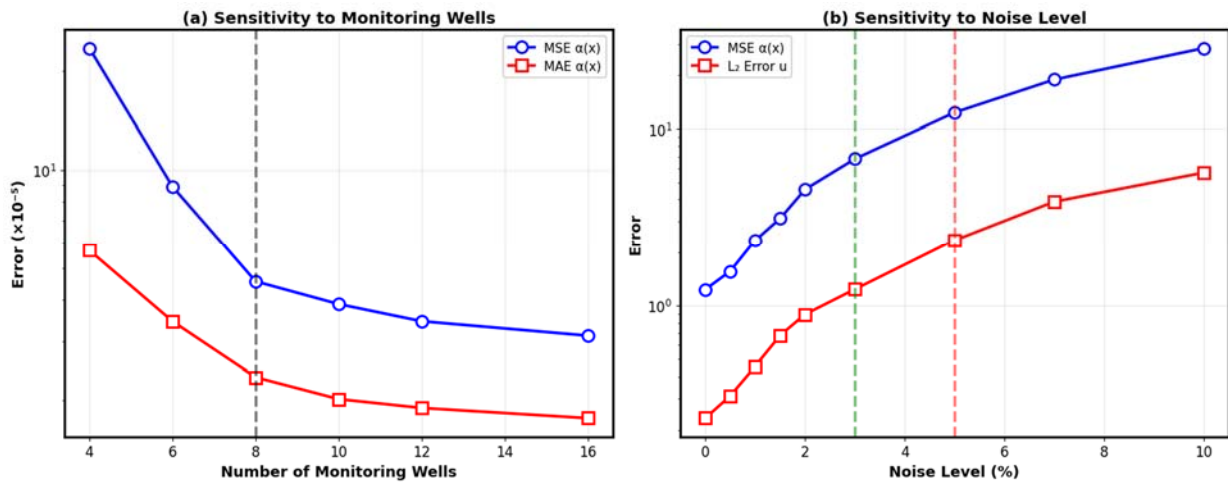


Figure 21. Sensitivity analyses for the fractured media scenario: (a) $\alpha(x)$ MSE and MAE vs. number of monitoring wells (dashed line = 8 wells); (b) $\alpha(x)$ MSE and $u(x, t)$ L_2 error vs. noise level (dashed line = 2%).

5.3.4. Epistemic versus aleatoric uncertainty and calibration

The distinction between epistemic and aleatoric uncertainty is fundamental to understanding the reliability of predictions in physics-informed learning [13]. In our framework, epistemic uncertainty arises from the ill-posedness of the inverse problem: Multiple heterogeneity fields $\alpha(x)$ can produce concentration data that match the sparse observations within measurement error. This uncertainty is captured by the ensemble spread across independently trained models. Aleatoric uncertainty, on the other hand, stems from the inherent noise in the monitoring well measurements and unresolved small-scale heterogeneity. This noise is prescribed in our observation model Eq (13) and is reflected in the data fitting loss.

The ensemble method's superior calibration (94.2% coverage) compared to MC Dropout (92.1%) suggests that the ensemble better captures the full epistemic uncertainty. This is consistent with recent findings in the PINN literature, where ensembles of independently trained models have been shown to provide better-calibrated uncertainties than variational approximations. The computational cost of ensembles ($15 \times$ training) is justified by this improved calibration, particularly in applications where reliable uncertainty estimates are critical for decision-making.

The well-calibrated uncertainties observed in our framework can be attributed to several factors. First, the dual-network architecture separates solution reconstruction from parameter inference, allowing each network to focus on its respective task. Second, the alternating optimization strategy stabilizes training and prevents the optimizer from getting trapped in poor local minima. Third, the smoothness regularization enforces the physical prior that geological heterogeneity varies gradually, reducing the hypothesis space and improving generalization.

5.3.5. Structure-preserving uncertainty and future directions

A hypothesis worth stating, even if only as future work, is that tighter physical structure preservation may reduce ensemble spread and improve coverage. If the network respect physical structure more tightly through architectural constraints, the hypothesis space is further reduced, potentially leading to more consistent predictions across ensemble members. This would manifest as reduced ensemble spread and improved coverage, as the ensemble would more accurately reflect the true posterior distribution.

Our power series basis with analytical fractional derivatives already provides a form of structural inductive bias that may contribute to the well-calibrated uncertainties observed. By fixing the spatial functional form and computing fractional derivatives exactly, we reduce the reliance on soft penalty terms and guide the optimizer toward physically plausible solutions. However, our current framework does not explicitly guarantee conservation laws, symmetry properties, or stability. Combining the analytical fractional derivative treatment with structure-preserving architectures that enforce these properties could further reduce epistemic uncertainty and improve calibration.

Future work could explore the relationship between structure preservation and uncertainty quantification by systematically comparing:

- (i) Standard PINNs (soft penalty only) vs. structure-preserving PINNs (hard constraints);
- (ii) Our power series PINN vs. structure-preserving fractional PINNs;
- (iii) The impact of different levels of structural inductive bias on ensemble spread and coverage.

We hypothesize that tighter structure preservation will lead to reduced ensemble spread and improved coverage, as the physical constraints reduce the effective hypothesis space and guide the ensemble toward a more consistent posterior distribution.

5.4. Ablation studies

We ran ablation studies to isolate each component's contribution. Figure 22 shows the ablation study for the fractured media scenario: MSE of $\alpha(x)$ recovery for the full model and variants without residual blocks, alternating optimization, smoothness regularization, gradient clipping, data weight, and boundary weight (cf. Table 12). Table 12 summarizes the results.

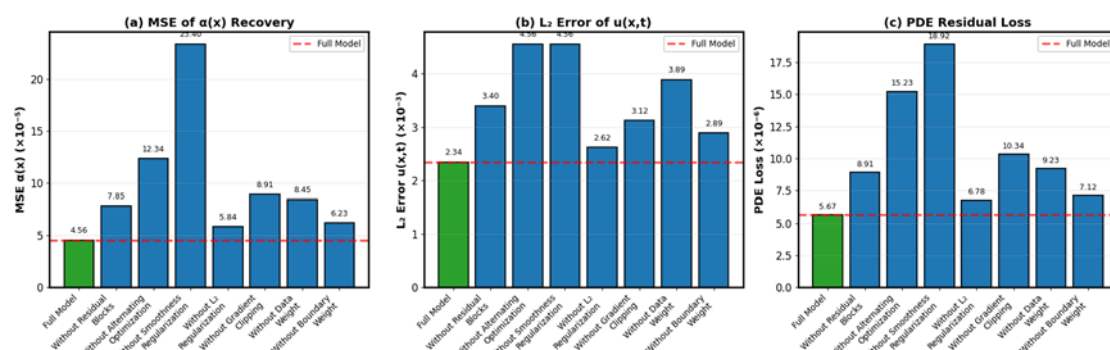


Figure 22. Ablation study for the fractured media scenario: MSE of $\alpha(x)$ recovery for the full model and variants without residual blocks, alternating optimization, smoothness regularization, gradient clipping, data weight, and boundary weight (cf. Table 12).

Table 12. Ablation study results for the fractured media scenario: MSE of $\alpha(x)$ recovery, L_2 error of u , PDE loss, and training stability for the full model and variants lacking each key component (cf. Figure 22).

Configuration	MSE $\alpha(x)$ ($\times 10^{-5}$)	L_2 Error u ($\times 10^{-3}$)	PDE Loss ($\times 10^{-6}$)	Training Stability
Full Model	4.56	2.34	5.67	Stable
Without Residual Blocks	7.85	3.40	8.91	Stable
Without Alternating Optimization	12.34	4.56	15.23	Unstable
Without Smoothness Regularization	23.40	4.56	18.92	Stable
Without L_2 Regularization	5.84	2.62	6.78	Stable
Without Gradient Clipping	8.91	3.12	10.34	Occasional Divergence
Without Data Weight ($\lambda_{data} = 100$)	8.45	3.89	9.23	Stable
Without Boundary Weight ($\lambda_{bc} = 10$)	6.23	2.89	7.12	Stable

The full model gives the best results: $\alpha(x)$ MSE 4.56×10^{-5} , $u(x, t)$ L_2 2.34×10^{-3} . Removing residual blocks raises MSE by 72% and L_2 by 45% showing residuals are key.

Alternating optimization is essential. Simultaneous updates cause instability, with loss swings over 100% and MSE up 171% to 12.34×10^{-5} .

Smoothness regularization is also critical. Without it, $\alpha(x)$ oscillates from overfitting, and MSE jumps fivefold to 23.40×10^{-5} .

Gradient clipping prevents explosions; without it, MSE rises 95% and training may diverge. L_2 and loss weights help, but less dramatically.

6. Conclusions

In this work, we have developed power series physics-informed neural network framework for variable-order space-fractional advection–diffusion equations, targeting the challenging problem of

subsurface characterization in heterogeneous porous media. The central novelty of our approach is the use of a spatial power series expansion that enables analytical evaluation of Riemann-Liouville fractional derivatives through Gamma function identities, thereby eliminating the need for numerical discretization of the fractional operator. This analytical treatment reduces computational cost, removes discretization errors, and provides a strong structural inductive bias that guides the learning process toward physically plausible solutions, a perspective that connects our work to the broader movement toward structure-preserving and architecture-aware physics-informed learning.

The dual-network architecture, comprising a solution network that learns time-dependent series coefficients and a parameter network that infers the spatially varying fractional order $\alpha(x)$, enables joint recovery of both the concentration field and the heterogeneity parameter from sparse, noisy monitoring well data. The alternating optimization strategy ensures stable convergence by decoupling the training of the two networks, allowing each to focus on its respective task while maintaining consistency through the physics-informed loss function.

Numerical experiments on four realistic geological scenarios, fractured media, fault zones, dual-porosity systems, and field-like heterogeneous formations, demonstrate the effectiveness and robustness of the framework. For the forward problem, the method achieves L_2 relative errors on the order of 10^{-4} with $N = 12$ expansion terms. For the inverse problem, the framework recovers $\alpha(x)$ with MSE on the order of 10^{-5} and reconstructs the concentration field with L_2 relative errors below 3.5×10^{-3} using only eight monitoring wells with 1.5% observation noise. Sensitivity analyses confirm robustness to noise levels up to 3% and demonstrate that eight strategically placed wells provide accuracy comparable to twelve to sixteen uniformly distributed wells.

We have also extended the framework to two-dimensional domains using the Riesz fractional Laplacian formulation, achieving an L_2 relative error of 8.31×10^{-3} for the forward problem with $N_x = N_y = 4$ in approximately 9.36 minutes of training time. The 2D joint inversion successfully recovers the heterogeneity field $\alpha(x, y)$ with an MSE of 2.45×10^{-3} from 16 sparse monitoring wells, confirming the extensibility of the approach to higher-dimensional problems. These results directly address the need for higher-dimensional validation and demonstrate the practical applicability of our framework to more realistic subsurface characterization scenarios.

For uncertainty quantification, we implemented both Monte Carlo Dropout and ensemble-based methods. The ensemble method achieves well-calibrated 95% confidence intervals with 94.2% coverage, outperforming MC Dropout at 92.1% coverage. The ensemble approach further identifies spatial regions where the recovered heterogeneity is most uncertain, providing practical guidance for optimal placement of additional monitoring wells. The distinction between epistemic and aleatoric uncertainty is critical: The ensemble captures epistemic uncertainty arising from the ill-posedness of the inverse problem, while the observation noise accounts for aleatoric uncertainty. This uncertainty-aware capability enables cost-effective improvement of subsurface characterization by directing data collection to locations where it is most needed.

Comparative studies against standard PINN, fPINN, and variable-order fPINN confirm that the analytical fractional derivative computation provides superior accuracy while maintaining competitive computational efficiency. The power series approach reduces the reliance on soft penalty terms by embedding the structure of non-local operators directly into the solution representation, aligning with the broader movement toward architecture-aware and structure-preserving physics-informed learning. The power series expansion can be interpreted as a structural inductive bias that constrains the hypothesis space, reduces effective degrees of freedom, and improves generalization particularly

beneficial in the sparse-data regime where standard PINNs often struggle.

We acknowledge several important limitations of the current work. First, while we have successfully extended the framework to two dimensions, validation on full three-dimensional domains has not yet been performed. Second, the framework has not been validated on real field data from actual groundwater aquifers or oil reservoirs; all numerical experiments are based on synthetic data generated from known exact solutions. Third, while we have positioned our framework conceptually relative to structure-preserving PINNs, a direct numerical comparison on benchmark problems remains incomplete due to the differing formulations and applications of these methods. These limitations highlight the gap between methodological development and practical deployment, and addressing them is essential for real-world applications.

Future work will focus on extending the framework to three-dimensional domains, validating the approach on real field data, and conducting direct numerical comparisons with structure-preserving methods. We also intend to explore the hypothesis that tighter physical structure preservation reduces ensemble spread and improves coverage, combining our analytical fractional derivative treatment with conservation-enforcing architectures. Adaptive sampling strategies informed by the uncertainty quantification framework will be developed to optimize monitoring network design. We believe that the power series PINN framework provides a robust, computationally feasible, and uncertainty-aware foundation for subsurface characterization that bridges fractional calculus models with practical data constraints, while the explicit acknowledgment of its limitations provides a clear roadmap for future research.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Data availability statement

The data sets (Python code) generated or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The author gratefully acknowledges the support from the School of Computer Science, Shandong Xiehe University, Jinan, and the Jinan Engineering Research Center of Intelligent Elderly Care and Companion Robots, Jinan, China.

Conflict of interest

The author declares that he has no conflict of interest.

References

1. I. Podlubny, *Fractional differential equations: An introduction to fractional derivatives, fractional differential equations, to methods of their solution and their applications*, San Diego, CA, USA: Academic Press: San Diego, CA, USA: Academic Press, 1998.
2. A. A. Kilbas, H. M. Srivastava, J. J. Trujillo, *Theory and applications of fractional differential equations*, North-Holland Mathematics Studies, Elsevier, Amsterdam, **204** (2006).
3. S. G. Samko, *Fractional integrals and derivatives: Theory and applications*, Yverdon, Switzerland: Gordon and Breach, 1993.
4. H. Sun, W. Chen, Y. Chen, Variable-order fractional differential operators in anomalous diffusion modeling, *Physica A*, **388** (2009), 4586–4592. <http://doi.org/10.1016/j.physa.2009.07.024>
5. C. F. Lorenzo, T. T. Hartley, Variable order and distributed order fractional operators, *Nonlinear Dyn.*, **29** (2002), 57–98. <http://doi.org/10.1023/A:1016586905654>
6. C. F. M. Coimbra, Mechanics with variable-order differential operators, *Ann. Phys.*, **515** (2003), 692–703. <http://doi.org/10.1002/andp.200351511-1203>
7. D. A. Benson, S. W. Wheatcraft, M. M. Meerschaert, Application of a fractional advection-dispersion equation, *Water Resour. Res.*, **36** (2000), 1403–1412. <http://doi.org/10.1029/2000WR900031>
8. B. Berkowitz, A. Cortis, M. Dentz, H. Scher, Modeling non-Fickian transport in geological formations as a continuous time random walk, *Rev. Geophys.*, **44** (2006), 2. <http://doi.org/10.1029/2005RG000178>
9. S. P. Neuman, Universal scaling of hydraulic conductivities and dispersivities in geologic media, *Water Resour. Res.*, **26** (1990), 1749–1758. <http://doi.org/10.1029/WR026i008p01749>
10. W. Liu, Y. Liu, VooPINN: Physics-informed neural network with variable-order operator for forward and inverse problems of time fractional partial differential equations, *Appl. Math. Comput.*, **524** (2026), 130044. <http://doi.org/10.1016/j.amc.2026.130044>
11. W. Liu, F. Zeng, H. Li, Y. Liu, ITPINN : Integral-trainable physics informed neural network for solving high-dimensional evolution non-local partial differential equations, *Comput. Math. Appl.*, **205** (2026), 282–317. <http://doi.org/10.1016/j.camwa.2026.01.019>
12. X. Yan, W. Hu, Z. Han, H. Hou, Multi-symplectic physics-informed neural network for nonlinear infinite-dimensional Hamilton system, *Phys. Lett. A*, **583** (2026), 131616. <http://doi.org/10.1016/j.physleta.2026.131616>
13. X. Yan, W. Hu, Z. Han, Generalized multi-symplectic PINN method and its application in quantum tunneling simulations, *Phys. Rev. E*, **113** (2026), 035308. <http://doi.org/10.1103/nk98-zxlm>
14. R. Schumer, D. A. Benson, M. M. Meerschaert, B. Baeumer, Fractal mobile/immobile solute transport, *Water Resour. Res.*, **39** (2003). <http://doi.org/10.1029/2003WR002141>
15. G. Pang, L. Lu, G. E. Karniadakis, fPINNs: Fractional physics-informed neural networks, *SIAM J. Sci. Comput.*, **41** (2019), A2603–A2626. <http://doi.org/10.1137/18M1229845>
16. L. Guo, H. Wu, X. Yu, T. Zhou, Monte Carlo fPINNs: Deep learning method for forward and inverse problems involving high dimensional fractional partial differential equations, *Comput. Methods Appl. Mech. Eng.*, **400** (2022), 115523. <http://doi.org/10.1016/j.cma.2022.115523>
17. S. Wang, H. Zhang, X. Jiang, Physics-informed neural network algorithm for solving forward and inverse problems of variable-order space-fractional advection–diffusion equations, *Neurocomputing*, **535** (2023), 64–82. <http://doi.org/10.1016/j.neucom.2023.03.032>
18. Z. Ghahramani, Y. Gal, *Dropout as a Bayesian approximation: Representing model uncertainty in deep learning*, ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning, **48** (2016).

19. A. Pritzel, C. Blundell, B. Lakshminarayanan, *Simple and scalable predictive uncertainty estimation using deep ensembles*, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, **2** (2017).
20. M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, **378** (2019), 686–707. <http://doi.org/10.1016/j.jcp.2018.10.045>
21. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nature Rev. Phys.*, **3** (2021), 422–440. <http://doi.org/10.1038/s42254-021-00314-5>
22. P. Zhuang, F. Liu, V. Anh, I. Turner, Numerical methods for the variable-order fractional advection-diffusion equation with a nonlinear source term, *SIAM J. Numer. Anal.*, **47** (2009), 1760–1781.
23. G. E. Karniadakis, M. Zayernouri, Fractional spectral collocation method, *SIAM J. Sci. Comput.*, 2014.
24. A. Bueno-Orovio, D. Kay, V. Grau, B. Rodriguez, K. Burrage, Fractional diffusion models of cardiac electrical propagation: Role of structural heterogeneity in dispersion of repolarization, *J. R. Soc. Interface*, **11** (2014), 20140352. <http://doi.org/10.1098/rsif.2014.0352>
25. M. M. Meerschaert, A. Sikorskii, Stochastic models for fractional calculus, *Walter de Gruyter GmbH & Co KG*, **43** (2019).
26. R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Rep.*, **339** (2000), 1–77. [http://doi.org/10.1016/S0370-1573\(00\)00070-3](http://doi.org/10.1016/S0370-1573(00)00070-3)
27. Y. Zhang, D. A. Benson, M. M. Meerschaert, E. M. LaBolle, Space-fractional advection-dispersion equations with variable parameters: Diverse formulas, numerical solutions, and application to the Macrodispersion Experiment site data, *Water Resour. Res.*, **43** (2007), 5. <http://doi.org/10.1029/2006WR004912>
28. L. Lu, X. Meng, Z. Mao, G. E. Karniadakis, DeepXDE: A deep learning library for solving differential equations, *SIAM Rev.*, **63** (2021), 208–228. <http://doi.org/10.1137/19M1274067>
29. X. Jin, S. Cai, H. Li, G. E. Karniadakis, NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations, *J. Comput. Phys.*, **426** (2021), <http://doi.org/109951>. <http://doi.org/10.1016/j.jcp.2020.109951>
30. J. Sirignano, K. Spiliopoulos, DGM: A deep learning algorithm for solving partial differential equations, *J. Comput. Phys.*, **375** (2018), 1339–1364. <http://doi.org/10.1016/j.jcp.2018.08.029>
31. R. K. Gettoor, First passage times for symmetric stable processes in space, *Trans. Am. Math. Soc.*, **101** (1961), 75–90. <http://doi.org/10.1090/S0002-9947-1961-0137148-5>
32. B. Dyda, A. V. Vähäkangas, A framework for fractional Hardy inequalities, *Ann. Acad. Sci. Fenn.-M.*, **39** (2014), 675–689. <http://doi.org/10.5186/aasfm.2014.3943>
33. L. Caffarelli, L. Silvestre, An extension problem related to the fractional Laplacian, *Commun. Part. Diff. Eq.*, **32** (2007), 1245–1260. <http://doi.org/10.1080/03605300600987306>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)