Chaos, Solitons and Fractals 208 (2026) 118347

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Physics-informed continuous state space models for probabilistic imaging in complex nonlinear media: A 3d magnetotelluric application

Xingran Guo [a,b,c], Tiaojie Xiao [a,b,c,*], Tian Shu [a,b,c], Jincai Huang [d], Jie Liu [a,b,c], Keqin Li [e,f]

[a] College of Computer Science and Technology, National University of Defense Technology, Changsha 410000, Hunan, China
[b] Science and Technology and Parallel and Distributed Processing Laboratory, National University of Defense Technology, Changsha 410073, Hunan, China
[c] Laboratory of Software Engineering for Complex Systems, National University of Defense Technology, Changsha 410073, Hunan, China
[d] Big Data Institute, Central South University, Changsha 410083, Hunan, China
[e] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA
[f] College of Information Science and Engineering, Hunan University, Changsha 410000, Hunan, China

ARTICLE INFO

Keywords:
3D magnetotelluric inversion
State space models
Physics-informed deep learning
Uncertainty quantification
Computational geophysics

ABSTRACT

Three-Dimensional (3D) magnetotelluric (MT) inversion is inherently ill-posed and computationally expensive due to the non-linear diffusive nature of electromagnetic fields. Current Transformer-based deep learning models struggle with quadratic complexity and multi-scale fractal heterogeneity. We present MT-Mamba, a physics-informed probabilistic framework that redefines 3D inversion by treating depth-wise field evolution as a continuous-time dynamical system via State Space Models (SSMs). Leveraging the linear-complexity modeling of Dual-Path Mamba, our method effectively captures electromagnetic diffusion while reducing RMSE by 26.2% and accelerating inference by 3.44× compared to state-of-the-art architectures. Notably, the framework demonstrates robust zero-shot domain transfer, which is validated on the Mount St. Helens field dataset. Without site-specific fine-tuning, MT-Mamba successfully reconstructs the deep magmatic plumbing system with sharper structural boundaries than established 3D deterministic models. By reconciling physical fidelity with computational efficiency, MT-Mamba provides a robust tool for real-time, high-fidelity imaging of complex non-linear systems in real-world environments.

1. Introduction

The reconstruction of internal structures within complex non-linear systems from boundary measurements constitutes a grand challenge in applied physics and interdisciplinary sciences. In the context of geophysics, the high-resolution imaging of the Earth's subsurface electrical structure represents a specific instance of this inverse problem, paramount for addressing challenges spanning from strategic resource exploration [1] to unraveling lithospheric dynamics. The Magnetotelluric (MT) method, utilizing natural electromagnetic field variations to probe conductivity distributions [2], serves as a primary tool for these deep exploration tasks.

* Corresponding author at: Science and Technology and Parallel and Distributed Processing Laboratory, National University of Defense Technology, Changsha 410073, Hunan, China.
E-mail addresses: gxrnudt@nudt.edu.cn (X. Guo), xiaotiaojie@nudt.edu.cn (T. Xiao), wenbenst@163.com (T. Shu), huangjincaicsu@csu.edu.cn (J. Huang), liujie@nudt.edu.cn (J. Liu), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.chaos.2026.118347
Received 4 January 2026; Received in revised form 24 March 2026; Accepted 10 April 2026

However, the 3D MT inverse problem is inherently ill-posed and non-linear [3]. While classical deterministic algorithms, such as the seminal Occam's inversion [4] and modular frameworks like ModEM [5], have established a rigorous baseline, they are plagued by prohibitive computational costs and a strong dependence on initial models [6]. Although recent strategies have attempted to mitigate non-uniqueness through joint inversion [7], the fundamental trade-off between resolution and computational feasibility remains an unresolved bottleneck in analyzing these complex systems.

The advent of Deep Learning (DL) has precipitated a paradigm shift. Early approaches, leveraging Convolutional Neural Networks (CNNs), demonstrated the potential to accelerate parameter estimation by orders of magnitude [8,9]. To constrain the "black-box" nature of pure DL, the community has increasingly integrated physical laws into network training via Physics-Informed Neural Networks (PINNs) [10,11]. However, a critical limitation persists: standard CNNs, bounded by local convolutional kernels, struggle to capture the global, long-range diffusion characteristics of electromagnetic fields in heterogeneous media [12]. This "local receptive field" acts as a fundamental ceiling, preventing the accurate recovery of structures in environments with extreme resistivity contrasts or multi-scale fractal heterogeneity.

To overcome the locality bottleneck, recent research has pivoted towards Transformers, leveraging Self-Attention to model global context [13,14]. Pioneering works in seismic and EM inversion [15] have set new benchmarks for recovering complex structures. Nevertheless, this performance comes at a prohibitive cost: the attention mechanism scales quadratically with sequence length ($O(N^2)$), rendering it mathematically intractable for high-resolution 3D volumes without severe downsampling. Furthermore, standard attention mechanisms often exhibit a "low-pass filtering" tendency, averaging out high-frequency details essential for resolving sharp boundaries in chaotic media [16]. Consequently, researchers face a dilemma: choosing between the global context of Transformers and the computational feasibility of CNNs.

Moreover, a profound gap remains in reliability. Most existing methods are deterministic, obscuring the non-uniqueness inherent to geophysical inverse problems. In the era of "Trustworthy AI", scientific rigor demands uncertainty quantification (UQ) [17]. While Bayesian networks [18] and ensembles have begun to address this, achieving high-fidelity UQ alongside efficiency remains challenging, particularly for complex geological facies where data constraints are weak [19].

In this work, we propose MT-Mamba, a transformative architecture that reconciles global modeling with computational efficiency while enforcing rigorous physical constraints. Our approach is grounded in State Space Models (SSMs) [20], leveraging the linear complexity ($O(N)$) of the Mamba architecture to process high-resolution 3D volumes previously inaccessible to Transformers [21,22]. Critically, we treat the depth-wise evolution of electromagnetic fields not merely as a sequence, but as a continuous-time dynamical system governed by Ordinary Differential Equations (ODEs). Just as EM fields evolve continuously through the subsurface governed by Maxwell's equations, Mamba's hidden state evolves through the sequence, effectively modeling the depth-dependent attenuation (skin effect) as a continuous state transition. This unique mathematical isomorphism allows MT-Mamba to capture long-range physical dependencies without the computational overhead of attention maps.

MT-Mamba introduces three novel contributions. First, we devise a Physics-Informed Dual-Path architecture. Unlike generic vision backbones [23], we explicitly decouple spatial geometry from physical diffusion: one path captures 3D structural continuity, while the other treats the depth/frequency dimension as a temporal sequence [24]. This design enforces a "frequency scan" that mimics the causal skin effect. Second, we advance beyond deterministic regression by reformulating the inversion as a probabilistic task, outputting dense uncertainty estimates for risk assessment. Finally, validated against high-fidelity 3D simulations featuring fractal media characteristics (Von Karman noise), MT-Mamba achieves superior resolution and reliability [25] at a fraction of the computational cost of Transformers, marking a new era of efficient, physics-embedded imaging for complex systems [26].

The main contributions of this work are summarized as follows:

- **A Novel Physics-Informed Dual-Path Architecture:** We pioneer the integration of State Space Models (SSMs) into 3D electromagnetic imaging, proposing a linear-complexity ($\mathcal{O}(N)$) framework that treats depth-wise field evolution as a continuous-time dynamical system. By decoupling spatial geometry from physical diffusion, our Dual-Path Mamba architecture explicitly embeds the electromagnetic skin effect and causal attenuation into the network's inductive bias, transcending the receptive field limitations of CNNs and the quadratic bottlenecks of Transformers.
- **Probabilistic Paradigm for Trustworthy Inversion:** Moving beyond traditional deterministic regression, we reformulate 3D inversion as a probabilistic task to address the inherent ill-posedness of geophysical problems. By simultaneously outputting resistivity predictions (μ) and aleatoric uncertainty estimates (σ^2), the framework provides a critical reliability metric that quantifies confidence risks in deep or complex geological zones, establishing a new standard for trustworthy and self-aware AI in geophysics.
- **Structure-Aware Performance on Complex Synthetic Geology:** We introduce a composite objective function combining Negative Log-Likelihood (NLL) with spatial structural gradient constraints to enforce boundary sharpness and mitigate the smoothing artifacts common in traditional Tikhonov regularization. Validated against high-fidelity synthetic datasets featuring dipping dikes, stair-case faults, and fractal noise, MT-Mamba demonstrates State-of-the-Art (SOTA) performance in both numerical accuracy and geometric fidelity.
- **Field-Scale Zero-Shot Generalization and 3D Fidelity:** Crucially, we validate the framework's practical utility using the high-complexity Mount St. Helens field dataset [27]. Without any site-specific fine-tuning, the pre-trained MT-Mamba model successfully retrieves the deep magmatic plumbing system. It natively overcomes the geometric distortion artifacts inherent to traditional 2D inversions and provides ultra-fast 3D reconstructions that are highly consistent with established 3D NLCG models and independent regional seismicity data. This proves the model's robust generalization to real-world, highly 3D tectonic environments.

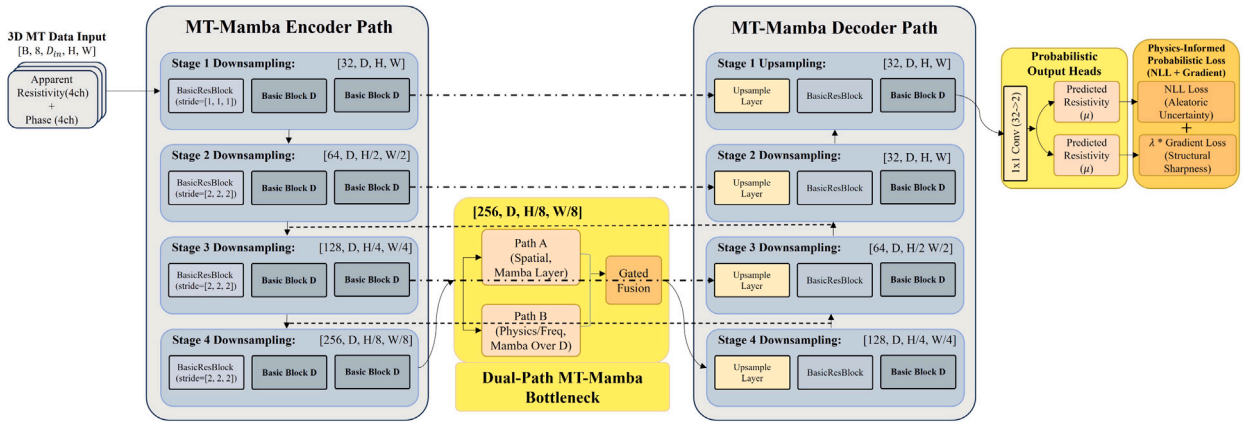

Fig. 1. The Probabilistic MT-Mamba architecture for 3D magnetotelluric imaging. The framework integrates a U-shaped encoder–decoder backbone (blue/gray shaded blocks, inherited structure) with novel physics-informed modules (yellow/orange shaded blocks). The network processes 8-channel 3D MT data (apparent resistivity and phase) through four downsampling stages to extract multi-scale features. The core innovation is the Dual-Path MT-Mamba Bottleneck situated at the lowest resolution, designed to capture both spatial geometry and frequency-dependent physical processes. The decoder reconstructs high-resolution features via upsampling and skip connections, leading to Probabilistic Output Heads that simultaneously predict the log-resistivity mean (μ) and aleatoric uncertainty variance (σ^2). The entire model is trained end-to-end using a composite Physics-Informed Probabilistic Loss combining Negative Log-Likelihood (NLL) and structural gradient terms.

2. Methodology

This section details the proposed Probabilistic MT-Mamba network. As illustrated in Fig. 1, our framework integrates a physics-aware dual-path design into a U-shaped architecture to simultaneously achieve high-fidelity 3D resistivity recovery and reliable uncertainty quantification.

To overcome the inherent scarcity of labeled 3D field data and strictly evaluate the generalization capability of deep learning models in heterogeneous geological environments, we propose a comprehensive framework that integrates high-fidelity physics-based simulation with a novel probabilistic deep learning architecture. The methodology is divided into two primary components: the construction of a rigorous forward modeling engine for data synthesis, and the development of the physics-informed MT-Mamba inversion network.

2.1. 3D magnetotelluric forward modeling framework

To achieve high-fidelity simulation of complex geoelectric structures and generate the requisite training data, we employed the Finite Volume Method (FVM) on a staggered grid, implemented via the SimPEG open-source framework [8]. It is imperative to clarify that our data generation process is strictly based on rigorous numerical simulation, rather than being synthesized by generative AI models. This ensures that the mapping from resistivity models to electromagnetic responses adheres strictly to Maxwell's equations. All computations were executed on the Tianhe Supercomputer (Tianjin, China). The linear systems arising from the FVM discretization were solved using the MUMPS (MUltifrontal Massively Parallel sparse direct Solver). As a direct solver performing exact LU factorization, MUMPS provides a numerical ground truth that serves as a rigorous benchmark for evaluating the proposed deep learning inversion.

2.1.1. Governing equations and vector Helmholtz formulation

The propagation of time-harmonic electromagnetic fields (assuming a time dependence of $e^{i\omega t}$) is governed by Maxwell's equations. In the frequency domain, neglecting displacement currents for low-frequency geophysical applications (quasi-static approximation), the curl equations in a source-free region (since the MT source is a plane wave originating from the ionosphere) are given by:

$$\nabla \times \mathbf{E} = -i\omega\mu\mathbf{H}, \tag{1}$$

$$\nabla \times \mathbf{H} = \sigma\mathbf{E}, \tag{2}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic field vectors, ω is the angular frequency, μ is the magnetic permeability, and σ is the conductivity tensor.

By taking the curl of (1) and substituting it into (2), we derive the homogeneous second-order vector Helmholtz equation for the electric field:

$$\nabla \times \mu^{-1}\nabla \times \mathbf{E} + i\omega\sigma\mathbf{E} = 0. \tag{3}$$

This formulation serves as the foundational physics governing the magnetotelluric (MT) response in our simulation environment, where the source energy enters the domain via boundary conditions.

2.1.2. Primary–secondary field formulation

To ensure numerical stability and efficiently handle the plane-wave boundary source, we adopted the Primary–Secondary (P–S) formulation available in SimPEG. The total electric field $\mathbf{E}$ is decomposed into a primary field $\mathbf{E}_p$ (response of a 1-D layered background) and a secondary (scattered) field $\mathbf{E}_s$:

$$\mathbf{E} = \mathbf{E}_p + \mathbf{E}_s, \tag{4}$$

The primary field $\mathbf{E}_p$ is defined as the solution to Maxwell's equations for a simplified 1-D layered background model with conductivity σ_p, satisfying:

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E}_p + i\omega\sigma_p \mathbf{E}_p = 0, \tag{5}$$

Subtracting the primary field equation from the total field Eq. (3), we obtain the governing equation for the secondary field:

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E}_s + i\omega\sigma \mathbf{E}_s = \underbrace{-i\omega (\sigma - \sigma_p)\mathbf{E}_p}_{\mathbf{J}_{sc}}, \tag{6}$$

Here, we define $\mathbf{J}_{sc} = (\sigma - \sigma_p)\mathbf{E}_p$ as the equivalent "scattering source" current density. This virtual source term is driven purely by the conductivity contrast between the 3D anomalies and the 1-D background, effectively converting the boundary value problem into a volume source problem ($-i\omega\mathbf{J}_{sc}$) for the secondary field.

2.1.3. Finite volume discretization and MUMPS solver

The continuous Eq. (6) is discretized using the Finite Volume Method on a 3D TensorMesh (rectilinear grid). We utilized a staggered grid (Yee grid) scheme [28], where electric field components are defined on the cell edges and magnetic flux densities on the cell faces. This arrangement naturally preserves the divergence-free properties of the fields and the continuity of tangential components.

The discrete form of the curl operator is represented by the matrix $\mathbf{C}$. The material properties are handled via mass matrices: M_σ (conductivity) and $M_{\mu^{-1}}$ (inverse permeability). The discretized linear system for the secondary electric field $\mathbf{E}_s$ is:

$$(\mathbf{C}^T M_{\mu^{-1}} \mathbf{C} + i\omega M_\sigma)\mathbf{E}_s = \mathbf{q}, \tag{7}$$

where $\mathbf{A} = \mathbf{C}^T \mathbf{M}_{\mu^{-1}} \mathbf{C} + i\omega \mathbf{M}_\sigma$ is the large, sparse, complex-symmetric system matrix, and $\mathbf{q}$ is the discretized source term derived from the primary field interaction.

Given the ill-conditioned nature of $\mathbf{A}$ at low frequencies and high conductivity contrasts, iterative solvers often struggle with convergence. Therefore, we employed MUMPS (MUltifrontal Massively Parallel sparse direct Solver), a robust direct solver integrated via pymatsolver. MUMPS performs an LU factorization of the system matrix $\mathbf{A}$, offering superior stability and precision for full-spectrum MT modeling compared to iterative alternatives.

2.2. Survey configuration and dataset generation

Building upon this solver, we constructed a diverse synthetic dataset to train the neural network. The simulation setup was rigorously designed to reflect realistic geological settings and survey geometries.

2.2.1. Mesh discretization strategy

A distinction must be made between the forward simulation mesh and the inversion grid used by the neural network. The forward simulation domain was discretized into a fine-grained tensor mesh (e.g., $64 \times 64 \times 64$ within the $1000 \times 1000 \times 1000$ m core region) extending significantly beyond the survey area to satisfy boundary conditions and include air layers. However, for the deep learning inversion task, we focused exclusively on the core subsurface volume where exploration targets are located. Consequently, the high-resolution simulation outputs were cropped and downsampled to a uniform grid of $32 \times 32 \times 16$ cells. This z-dimension (16 layers) explicitly represents the subterranean conductive structures of interest, excluding the air layer and deep boundary padding used during the numerical forward modeling.

2.2.2. Stochastic geological modeling

Unlike prior studies that relied on simplified geometric primitives [8], we introduced a stochastic geometry generator to mimic the topological complexity of real-world crustal structures. The resistivity models $\mathbf{m}$ were synthesized by embedding three classes of anomalies into a background half-space ($\rho_{bg} = 100\,\Omega$ m):

1. *Dipping Dikes:* Modeled as thin, high-contrast tabular conductors with randomized strike (0°–180°) and dip angles (30°–90°), simulating fluid-filled fracture zones or magmatic intrusions.
2. *Ellipsoidal Bodies:* Defined by randomized semi-axes (a, b, c) and Euler rotation angles, representing discrete ore bodies or salt domes.

3. *Stair-case Faults:* Generated by cascading rectangular blocks with depth-dependent offsets, mimicking crustal displacement zones.

To emulate the randomness of geological media, we superimposed Von Karman fractal noise onto the background resistivity, enforcing a spatially correlated heterogeneity that challenges the network's ability to distinguish signal from geological noise.

2.2.3. Dataset characteristics

To test the network's robustness, the dataset includes structures with extreme aspect ratios, such as dipping thin dikes and rotated ellipsoids, which mimic the directional complexity often found in fault zones and mineral deposits. The dataset comprises 16 frequencies logarithmically spaced from 10^{-3} Hz to 10^{3} Hz. For each model, the impedance tensor $\mathbf{Z}$ was calculated from the simulated fields, and subsequently converted into apparent resistivity and phase, which serve as the primary inputs for our deep learning inversion model.

2.3. Theoretical framework: Physics-informed state space modeling

We reformulate the 3D MT inversion problem not merely as an image-to-image translation task, but as a sequence modeling problem governed by continuous latent dynamics. This perspective allows us to leverage State Space Models (SSMs), which combine the modeling power of recurrent neural networks (RNNs) with the computational efficiency of convolutional neural networks (CNNs) [23].

2.3.1. Continuous-time latent dynamics

In contrast to standard discrete sequence models (e.g., Transformers), we reconceptualize the spectral-depth evolution of the subsurface electromagnetic field as a continuous-time dynamical system. This perspective allows us to capture the diffusive nature of the physical process directly through a State Space Model (SSM) formalism. Let $h(t) \in \mathbb{R}^N$ denote the latent state of the medium at a generalized coordinate t (representing the continuous depth/frequency trajectory), and $x(t)$ denote the input impedance forcing function. The system dynamics are governed by the following Ordinary Differential Equation (ODE):

$$\begin{cases} \dot{h}(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) = \mathbf{C}h(t). \end{cases} \tag{8}$$

This continuous-time formulation represents a Linear Time-Invariant (LTI) system. The stability and dynamical properties of such LTI systems have been extensively investigated in the broader context of nonlinear science and control theory [29]. Here, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the state transition matrix that characterizes the intrinsic dynamics of the system. In the context of magnetotellurics, $\mathbf{A}$ governs the decay and phase shift of the electromagnetic field as it diffuses through conductive media, effectively encoding the "memory" of deep conductivity structures within the latent state trajectory. $\mathbf{B}$ and $\mathbf{C}$ serve as projection matrices linking the latent dynamics to the input control and output measurements, respectively.

2.3.2. Discretization via Zero-Order Hold (ZOH)

To implement this continuous system on digital hardware, we discretize the ODE using the ZOH principle. We introduce a learnable timescale parameter Δ, which represents the sampling step size in the latent space. The discrete system matrices are derived as:

$$\overline{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}. \tag{9}$$

2.3.3. The selective scan mechanism and physical isomorphism

The discrete recurrence $h_t = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t$ is computed using the parallel selective scan algorithm (S6), which enables global context modeling with linear complexity $\mathcal{O}(L)$ (where L is sequence length), avoiding the quadratic bottleneck of Transformer-based attention mechanisms [23].

Crucially, the data-dependent nature of this Selective Scan algorithm provides a rigorous mathematical isomorphism to the physics of electromagnetic diffusion. In conventional linear invariant systems, constant parameters fail to capture the extreme heterogeneity of the Earth's subsurface. By making the discretization step size Δ and the projection matrices $\mathbf{B}$ and $\mathbf{C}$ explicitly dependent on the input features x_t (representing local impedance and spatial gradients), the Selective Scan dynamically adjusts the hidden state transition matrix $\overline{\mathbf{A}}$.

Physical Interpretation: This data-dependent gating acts as an adaptive attenuation filter that elegantly mirrors the electromagnetic *skin effect* ($\delta \approx 503\sqrt{\rho/f}$). When the sequence traverses a highly conductive anomalous layer (low resistivity ρ), the network predicts a specific Δ configuration that drives $\overline{\mathbf{A}} \to \mathbf{0}$, thereby rapidly attenuating ("forgetting") the state history. This mathematically simulates the rapid energy dissipation and shallow penetration of EM waves in low-resistivity media. Conversely, when traversing a resistive background (high resistivity ρ), the gating mechanism preserves the latent state ($\overline{\mathbf{A}} \to \mathbf{I}$), enabling the deep propagation ("remembering") of low-frequency structural information. Thus, the Selective Scan does not merely process a generic spatial sequence; it explicitly simulates the causal, spatially-varying attenuation of EM fields in complex non-linear media, effectively addressing the resolution limits imposed by highly resistive shielding layers.

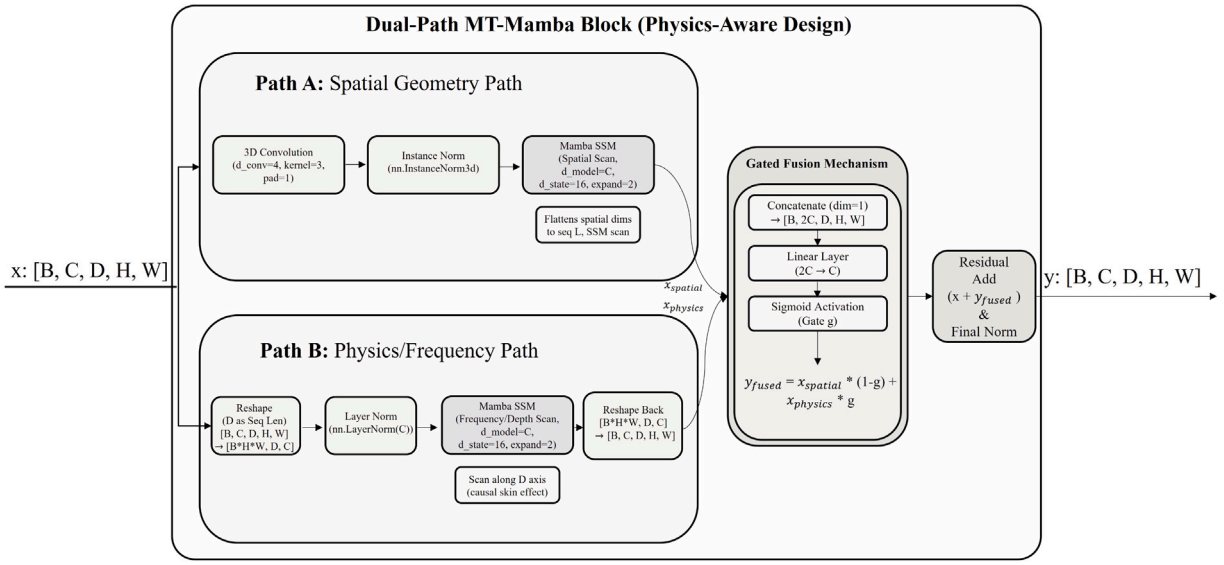

Fig. 2. Detailed mechanism of the physics-aware Dual-Path MT-Mamba block. This diagram expands the bottleneck module highlighted in Fig. 1. The input tensor $[B, C, D, H, W]$ splits into two parallel streams. Path A (Spatial Geometry Path) employs a 3D State Space Model (SSM) to model volumetric spatial correlations, utilizing 3D convolution and Instance Normalization. Path B (Physics/Frequency Path) introduces a novel frequency-scanning mechanism: the depth/frequency dimension D is reshaped into the sequence length for the SSM to explicitly model causal signal attenuation (skin effect) along the depth axis, using Layer Normalization. The outputs of both paths are merged via a learnable Gated Fusion Mechanism, where a sigmoid activation dynamically weights the contribution of spatial versus physical features ($y_{\text{fused}} = x_{\text{spatial}} \cdot (1-g) + x_{\text{physics}} \cdot g$) before final residual addition and normalization.

2.4. The MT-Mamba architecture

Based on this theoretical foundation, we propose MT-Mamba, a hierarchical probabilistic network. The architecture adopts a U-shaped encoder–decoder topology, distinguished by our novel Dual-Path Physics-Aware Bottleneck. The detailed internal structure of this novel dual-path block and its gated fusion mechanism is shown in Fig. 2.

2.4.1. Backbone: Hybrid CNN-SSM

The encoder extracts multi-scale features from the input impedance tensor $\mathbf{D} \in \mathbb{R}^{8\times16\times32\times32}$. Shallow layers employ residual convolutional blocks to capture high-frequency local texture information (e.g., sharp boundaries of small anomalies), while deeper layers transition to SSM-based processing. Critically, the input to the Mamba bottleneck is not the raw voxel data, but the high-level semantic feature maps extracted by the CNN encoder. This design avoids the need for explicit patch embedding layers typical in Vision Transformers, as the CNN implicitly tokenizes the 3D volume into feature vectors.

2.4.2. Dual-path physics-aware block

Standard 3D SSMs typically flatten volumetric data into a single 1-D sequence, confounding spatial correlations with spectral dependencies. To resolve this, we employ a Factorized Serialization Strategy. The latent feature volume $\mathbf{X} \in \mathbb{R}^{B\times C\times D\times H\times W}$ is processed through two orthogonal, task-specific pathways:

Path A: Spatial geometry scanning. This path is dedicated to reconstructing the 3D geometric morphology of subsurface structures. We treat the spatial dimensions (H, W) as the sequence domain. The feature maps are flattened using a row-major raster scan, converting the $H \times W$ spatial grid into a sequence of length $L = H \cdot W$. The Spatial SSM then performs a selective scan over this sequence, capturing long-range geometric continuity (e.g., the extension of a dike across the entire domain) that exceeds the local receptive field of conventional CNNs:

$$\mathbf{Y}_{spatial} = \text{SSM}_{spatial}(\text{Norm}(\mathbf{X})). \tag{10}$$

Path B: Physics-informed frequency scanning. This path explicitly enforces the causality of EM diffusion. We permute the input tensor to $\mathbf{X}_{perm} \in \mathbb{R}^{(B\cdot H\cdot W)\times D\times C}$, isolating the depth/frequency dimension D as the temporal sequence. We then perform a directed Mamba scan along this D-axis. By strictly ordering the sequence from high frequency (shallow) to low frequency (deep), the hidden states h_t integrate the attenuation history of the EM field layer by layer. This mechanism acts as a data-driven approximation of the 1-D Riccati equation used in analytical MT inversion, constraining the vertical resistivity profile to be physically consistent with the skin effect principle.

$$\mathbf{Y}_{physics} = \text{SSM}_{physics}(\text{Norm}(\mathbf{X}_{perm})). \tag{11}$$

Gated feature fusion. To dynamically integrate geometric cues with physical constraints, we employ a learnable gating mechanism inspired by multi-modal fusion networks [30]. A non-linear gate $\mathcal{G}$ estimates the reliability of each path at the voxel level:

$$\mathbf{Z} = \text{Concat}(\mathbf{Y}_{spatial}, \mathbf{Y}_{physics}), \tag{12}$$

$$\mathcal{G} = \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{Z})), \tag{13}$$

$$\mathbf{Y}_{out} = \mathcal{G} \odot \mathbf{Y}_{physics} + (1 - \mathcal{G}) \odot \mathbf{Y}_{spatial} + \mathbf{X}_{residual}, \tag{14}$$

where σ is the Sigmoid function, and $\odot$ denotes the Hadamard (element-wise) product. This allows the network to prioritize physical depth consistency in stratified media (where 1-D physics dominates) while focusing on spatial geometry at sharp lateral discontinuities (where 3D geometry dominates).

2.5. Probabilistic uncertainty quantification

To move beyond deterministic estimation and provide scientific guarantees of reliability, we formulate the inversion as a probabilistic regression task. To align with the robust L_1 loss function, we assume the predictive distribution of the resistivity y at each voxel follows a Laplacian distribution parameterized by the input $\mathbf{x}$:

$$p(y|\mathbf{x}) = \text{Laplace}(y; \mu(\mathbf{x}), b(\mathbf{x})). \tag{15}$$

Consequently, the final layer of MT-Mamba is modified to output two channels: the predicted mean log-resistivity μ and the aleatoric uncertainty parameter $s = \log(b)$, where b is the scale parameter of the Laplace distribution. This formulation allows the model to autonomously flag regions of low confidence — typically deep structures with weak data constraints or complex boundary zones — providing a critical metric for risk assessment in geological interpretation.

2.6. Structure-aware objective function

We propose a composite loss function $\mathcal{L}_{total}$ that balances data likelihood with structural regularization.

2.6.1. Negative Log-Likelihood (NLL) loss

In our probabilistic framework, we specifically model the predictive distribution $p(y|\mathbf{x})$ using a Laplace distribution rather than the conventional Gaussian distribution. Consequently, maximizing the log-likelihood of the target data is equivalent to minimizing the following Negative Log-Likelihood (NLL) loss:

$$\mathcal{L}_{NLL} = -\log p(y|\mathbf{x}) \propto \frac{1}{2}\exp(-s)\|y - \mu\|_1 + \frac{1}{2}s, \tag{16}$$

where $s = \log(\sigma)$ predicts the log-variance to ensure numerical stability. Here, the term $\exp(-s)$ acts as an adaptive attentional weight: for voxels where the model is uncertain (large s), the penalty for prediction error is attenuated. The $\exp(-s)$ term attenuates the penalty for prediction error in uncertain regions, which prevents the model from overfitting to ambiguous data or noise, while the term $\frac{1}{2}s$ serves as a regularizer to prevent the uncertainty estimate from diverging to infinity.

Crucially, the choice of the Laplace distribution is deeply rooted in geological reality. A Gaussian assumption corresponds to an L_2-norm penalty, which is highly sensitive to outliers and inherently tends to produce over-smoothed, blurry boundaries. In contrast, subsurface environments frequently feature sharp structural discontinuities (e.g., faults and intrusive dikes). The Laplace distribution inherently corresponds to an L_1-norm prior, which promotes sparsity in the residuals. This mathematical property is vital for preserving sharp, blocky boundaries of geological facies and makes the inversion significantly more robust against severe distortions and non-Gaussian noise artifacts.

2.6.2. Gradient consistency loss

To further mitigate the smoothing effect inherent in probabilistic models and explicitly recover sharp geological interfaces (e.g., fault planes), we enforce consistency in the 3D spatial gradients:

$$\mathcal{L}_{grad} = \sum_{i \in \{x,y,z\}} \|\nabla_i y - \nabla_i \mu\|_1, \tag{17}$$

where ∇_i represents the discrete gradient operator along axis i. The final objective is defined as $\mathcal{L}_{total} = \mathcal{L}_{NLL} + \lambda \mathcal{L}_{grad}$, with λ set to 0.2 empirically.

2.7. Computational complexity analysis

A critical advantage of MT-Mamba over Transformer-based counterparts (e.g., Swin-UNet) is its scalability. For a 3D input volume of size $V = D \times H \times W$, the self-attention mechanism in Transformers imposes a quadratic complexity of $O(V^2)$, rendering it computationally prohibitive for high-resolution geophysical grids. In contrast, our Dual-Path Mamba block operates with linear complexity. The Spatial Path scales as $O(V)$, and the Physics Path scales as $B \cdot H \cdot W \cdot O(D) = O(V)$. This linear scaling allows MT-Mamba to process high-resolution 3D volumes on standard GPU hardware without the need for patching or windowing, thereby preserving the global context that is indispensable for interpreting low-frequency electromagnetic signals.

3. Experiments and analysis

To rigorously validate the MT-Mamba framework as a transformative paradigm for 3D magnetotelluric (MT) inversion, we conducted a systematic and multi-layered experimental study. This section details the experimental setup, benchmarks, and a comprehensive analysis of the results. Our evaluation is structured to probe the model's capabilities across five critical dimensions: (1) the fidelity of resistivity reconstruction in geologically complex scenarios; (2) the volumetric coherence of 3D geometry recovery; (3) the reliability of probabilistic uncertainty quantification (UQ); (4) generalization to out-of-distribution structural topologies; and (5) computational tractability for high-performance computing (HPC) environments.

3.1. Physics-based data synthesis

Given the prohibitive cost of acquiring ground truth via extensive drilling and the scarcity of high-resolution 3D MT field data, we established a high-fidelity synthetic dataset generator, MT3D-HPC. This platform leverages the SimPEG framework to solve the frequency-domain Maxwell's equations on a staggered grid, ensuring that the training data strictly adheres to the physics of electromagnetic diffusion. In our simulations, we assume the subsurface is a non-magnetic isotropic medium, thus fixing the magnetic permeability μ to the vacuum permeability μ_0 ($4\pi \times 10^{-7}$ H/m).

3.1.1. Geological scenario design

As visualized in Fig. 3, the simulation domain is defined within a physical volume of $1000 \times 1000 \times 1000$ meters, discretized into a $64 \times 64 \times 64$ tensor mesh. The mesh design incorporates a core region with uniform cell spacing to capture fine-scale geological features, surrounded by padding cells that extend the boundaries to mitigate edge effects during forward modeling. To stress-test the inversion algorithms, we engineered four core geological scenarios (Types I-IV), each representing a distinct geophysical challenge:

1. **Type I: Rectangular Block (Baseline Calibration).** A canonical scenario featuring a single low-resistivity anomaly (10 Ω m) embedded in a homogeneous half-space (100 Ω m). This serves to calibrate the model's ability to localize centroids and estimate resistivity contrasts without geometric distortion.
2. **Type II: Stair-case Fault (Discontinuity Resolution).** A tilted, step-like conductive structure (10 Ω m) simulating listric faults. This scenario tests the network's capacity to resolve sharp, discontinuous interfaces and high-frequency vertical gradients, challenging the smoothness constraints typically imposed by Tikhonov regularization.
3. **Type III: Dipping Thin Dike (The "Resistive Barrier" Challenge).** This scenario features a high-resistivity dipping sheet (1000 Ω m) underlying a shallow, small low-resistivity target (1.5 Ω m). In classical EM theory, the high-resistivity basement acts as a barrier to vertical current flow, complicating the resolution of the conductor's depth extent and the interface between the two bodies. Accurately recovering the separation between the shallow anomaly and the dipping dike tests the model's ability to handle strong resistive contrasts.
4. **Type IV: Rotated Ellipsoid (Grid Independence).** A spatially directional conductive structure (10 Ω m) with arbitrary 3D rotation. This assesses the model's resilience against grid-imposed "staircase" artifacts when reconstructing smooth, curvilinear boundaries.

3.1.2. Simulation protocol

For each geological model, we simulated the impedance tensor $\mathbf{Z}$ across a broadband frequency range. Specifically, we selected 16 logarithmic frequencies spanning from 10^{-3} Hz to 10^{3} Hz. The frequency range was selected to ensure sensitivity coverage from the near-surface up to and beyond the base of the simulation domain (noting that the skin depth at the lowest frequency is approx. 10 km, ensuring the 1000 m target volume is fully resolved without boundary artifacts).

To bridge the domain gap between synthetic simulations and real-world noisy measurements, complex Gaussian noise was injected into the impedance tensor components ($Z_{xx}, Z_{xy}, Z_{yx}, Z_{yy}$). The noise levels were randomly sampled from a uniform distribution $U(0\%, 5\%)$, reflecting typical field data quality. The final dataset comprises 10,000 pairs of 3D resistivity models and their corresponding multi-frequency responses, split 8:1:1 for training, validation, and testing.

3.2. Implementation details and baselines

3.2.1. Network configuration and preprocessing

To handle the multi-scale nature of electromagnetic responses, we designed a 4-stage U-shaped architecture. The encoder channel dimensions were set to $[32, 64, 128, 256]$ to balance feature expressiveness with memory efficiency. The core Dual-Path Mamba bottleneck operates with a state expansion factor of $d_{state} = 16$ and a local convolution width of 4, ensuring robust sequence modeling. To align the variable-grid field measurements with the regular inversion mesh, we applied spline interpolation to the input impedance data, reshaping it to a uniform $16 \times 32 \times 32$ tensor (Frequency × Y × X). Input impedance phases were normalized by 90°, and apparent resistivities were log-transformed and min–max scaled to ensure numerical stability.

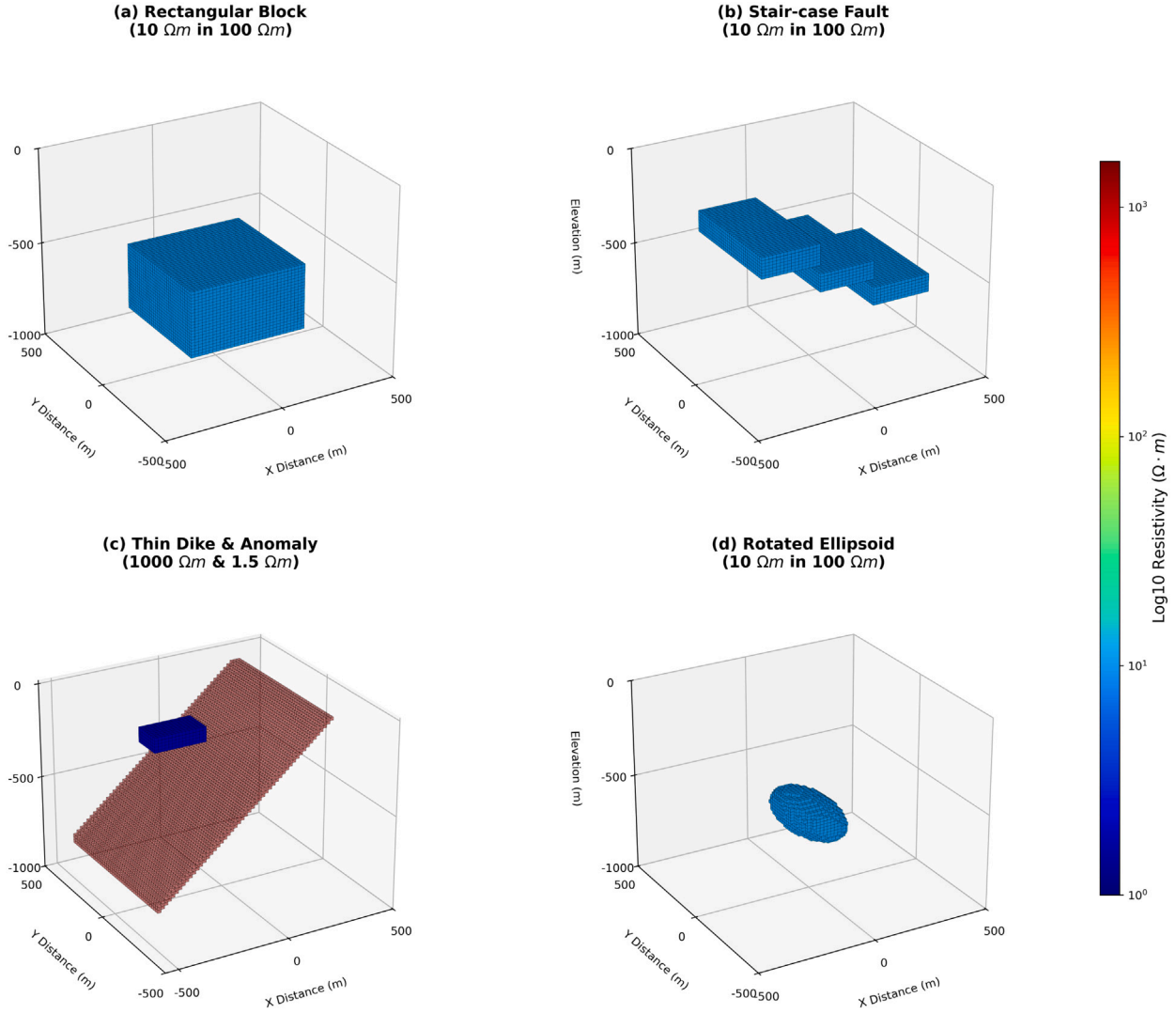

Fig. 3. 3D Ground Truth geological models used for synthetic data generation. The figure illustrates the four core 3D resistivity structures employed to evaluate the network's robustness: (a) Rectangular Block, (b) Stair-case Fault, (c) Dipping Thin Dike accompanied by a shallow small anomaly, and (d) Rotated Ellipsoid. The models are visualized in a physical coordinate system spanning $1000 \times 1000 \times 1000$ meters. Red voxels represent high-resistivity structures (specifically the Dike), while blue voxels indicate low-resistivity anomalies embedded in the background. These complex geometries serve as the rigorous ground truth labels for the 3D inversion tasks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.2. Training protocol

The MT-Mamba framework was implemented in PyTorch and trained on a high-performance computing cluster equipped with 1 NVIDIA A100 (80 GB) GPUs using distributed data parallelism. This hardware configuration allowed for a large effective batch size and accelerated convergence. We employed the AdamW optimizer with a base learning rate of 1×10^{-4} and a weight decay of 1×10^{-5} to prevent overfitting. A cosine annealing scheduler was applied to decay the learning rate over 500 epochs, with the total training process taking approximately 36 h. The composite Physics-Informed Probabilistic Loss ($\mathcal{L}_{total}$) was carefully balanced, combining Negative Log-Likelihood (NLL) for aleatoric uncertainty estimation and structural gradient terms for edge preservation.

3.2.3. Baselines

We benchmarked MT-Mamba against three distinct classes of methods to thoroughly demonstrate its superiority and added value:

Table 1
Main Quantitative Performance Comparison. **Note:** RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) are lower-is-better (↓), while the R^2 (Coefficient of Determination) is higher-is-better (↑). The proposed MT-Mamba achieves the best performance across all metrics.

Model	RMSE (↓)	MAE (↓)	R^2 (↑)
Traditional Inversion (Baseline)	0.0412	0.0285	0.910
U-Net (CNN Baseline)	0.0534	0.0351	0.852
SwinUNETR (Transformer SOTA)	0.0496	0.0322	0.880
MT-Mamba (Ours)	**0.0366**	**0.0258**	**0.955**

- **Traditional Inversion (e.g., ModEM/Occam):** The industry-standard deterministic optimization approach based on gradient descent. It serves as the ultimate physics-based baseline, though it typically requires carefully constructed initial models and extensive computational resources.
- **U-Net:** A standard 3D CNN representing local-receptive-field deep learning architectures, widely used as a geophysical segmentation and regression baseline.
- **SwinUNETR:** A hierarchical Vision Transformer (ViT) representing state-of-the-art medical and volumetric imaging models, utilizing shifted windows for global context.
- **MT-Mamba (Ours):** The proposed architecture integrating Dual-Path Mamba blocks, specifically designed to capture the diffusive physics and causal attenuation of EM fields.

3.3. Quantitative performance analysis

We first present a rigorous quantitative comparison using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). As detailed in Table 1, MT-Mamba establishes a new state-of-the-art benchmark, outperforming both purely data-driven deep learning architectures and traditional deterministic inversion.

The proposed model achieves an RMSE of 0.0366, representing a 26.2% reduction compared to the closest deep learning competitor, SwinUNETR (0.0496), and an 11.2% improvement over the traditional physics-based inversion baseline (0.0412). Furthermore, the R^2 score reaches 0.955, indicating that our model explains nearly all the variance in the complex 3D resistivity distribution.

Crucially, while traditional inversion achieves a respectable R^2 (0.910) by iteratively minimizing the data misfit, it fundamentally struggles to recover sharp structural boundaries due to inherent smoothing regularizations (e.g., L_2-norm penalties). In contrast, MT-Mamba transcends this limitation. This quantitative leap implies that the physics-informed State Space Model (SSM) mechanism captures long-range frequency-spatial dependencies more effectively than the fixed receptive fields of CNNs, the patch-based attention of Transformers, and the heavily regularized objective functions of traditional solvers.

3.4. High-fidelity reconstruction and physical consistency

To rigorously evaluate the model's performance in resolving complex 3D structures, we present a comprehensive visualization in Fig. 4, covering four geological scenarios with increasing complexity. Unlike previous studies that rely solely on single-frequency slices, we adopt a multi-view approach to verify both physical compliance and 3D geometric consistency.

- **Physics-Compliant Inputs:** Columns 1–4 of Fig. 4 display the full-frequency pseudo-sections for the input impedance tensor. These visualizations capture essential physical phenomena embedded in the data: the skin effect (blurring with depth/frequency), static shifts (vertical striping noise), and the polarization difference between TM (ρ_{xy}) and TE (ρ_{yx}) modes. Notably, in the Dike scenario (Row 3), the high-resistivity dike is clearly visible in the TM mode but suppressed in the TE mode due to current channeling—a fine-grained physical nuance that MT-Mamba successfully learns to interpret.
- **3D Structural Consistency:** We assess the inversion quality through two orthogonal cross-sections: **View A** (XZ-slice at $Y = 0$ m) and **View B** (YZ-slice at $X = 0$ m), ensuring that the recovered structures are spatially coherent.
- **Resolving High-Contrast Interfaces:** As shown in the **Dike scenario (Row 3)**, the shallow conductive block (Target) lies immediately above the underlying high-resistivity dike. While SwinUNETR produces a smeared result where the target blends into the background in both views, MT-Mamba successfully disentangles the shallow target from the resistive basement, recovering its sharp geometry and correct resistivity value in both orthogonal directions. This demonstrates the model's superior capability in handling non-linear EM coupling effects and preserving 3D continuity.

3.5. Volumetric 3D imaging and depth perception

Moving beyond 2D slices, Fig. 5 presents the full 3D volumetric rendering of the MT-Mamba inversion results, strictly corresponding to the ground truth geometries in Fig. 3.

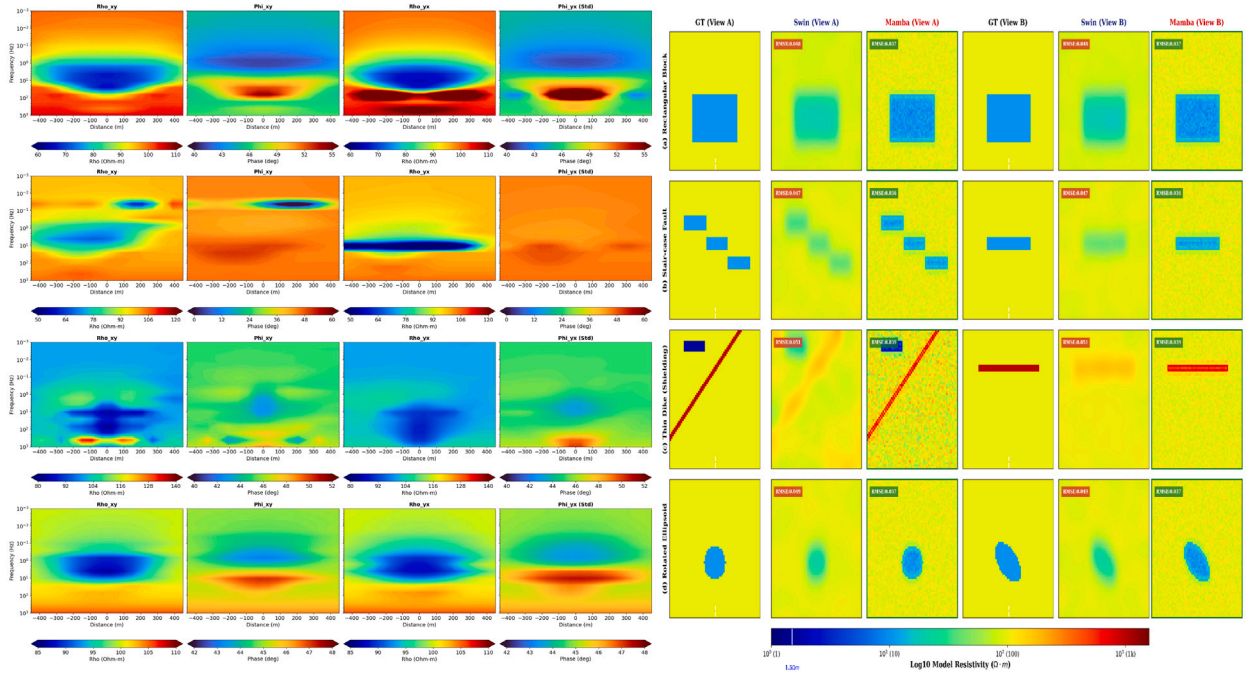

Fig. 4. Qualitative comparison of 3D inversion results via physics-compliant multi-view analysis. Columns 1–4 (Input Data): Apparent resistivity (ρ_{xy}, ρ_{yx}) and impedance phase (ϕ_{xy}, ϕ_{yx}) pseudo-sections derived from the central profile ($Y = 0$ m). The vertical axis represents decreasing frequency (from high to low), explicitly visualizing the skin effect where deeper structures appear more diffusive. Note the distinct physical characteristics: the TM mode (ρ_{xy}) preserves lateral boundaries of high-resistivity anomalies (e.g., the Dike), while the TE mode (ρ_{yx}) exhibits current channeling effects, suppressing resistive targets. Vertical striping in resistivity simulates realistic static shift noise. **Columns 5–7 (View A):** Vertical cross-sections of the inverted resistivity models along the XZ plane at $Y = 0$ m. **Columns 8–10 (View B):** Orthogonal vertical cross-sections along the YZ plane at $X = 0$ m. The comparison highlights MT-Mamba's ability to recover sharp boundaries and shallow targets situated above resistive basements (e.g., the conductive block in Row 3) compared to SwinUNETR. The consistency between View A and View B confirms the 3D structural integrity. (RMSE values are annotated in the top-left corner of each result.)

The 3D point cloud visualization reveals the model's exceptional depth perception. In the Dike scenario (Fig. 5c), we employed a differential transparency rendering technique: the high-resistivity dike is rendered as a semi-transparent red cloud, while the shallow low-resistivity anomaly is an opaque blue core. This visualization confirms that the spatial separation between these two bodies is strictly maintained in 3D space. Furthermore, the reconstructions exhibit realistic physical characteristics: the resolution naturally degrades with depth (manifesting the "skin effect"), resulting in slightly smoother edges for deeper anomalies compared to near-surface structures. This confirms that MT-Mamba respects the underlying physics of diffusive EM fields.

3.6. Probabilistic uncertainty quantification (UQ)

A critical advantage of our framework is the ability to quantify aleatoric uncertainty (σ). Fig. 6 illustrates the relationship between the predicted mean (μ), the uncertainty map (σ), and the absolute error ($|\mu - GT|$).

The results demonstrate a strong spatial correlation between high uncertainty and high reconstruction error. As expected, uncertainty peaks at geological interfaces (where resistivity contrasts are sharpest) and at greater depths (where the signal-to-noise ratio naturally decays). This property renders the model "self-aware": it flags regions where the inversion is ambiguous. In a real-world exploration context, this uncertainty map provides invaluable guidance for risk assessment—drilling decisions can be deferred in areas of high uncertainty until further data is acquired.

3.7. Generalization to unseen structures

To evaluate robustness against out-of-distribution data, we tested the model on a complex, unseen geological system featuring angular unconformities and fault networks, as shown in Fig. 7.

While SwinUNETR struggles to maintain the continuity of the fault lines, often creating disconnected artifacts, MT-Mamba preserves the structural integrity of the fault system. It correctly positions the deep anomaly without geometric distortion. This suggests that the model has learned the underlying physics mapping from EM response to resistivity structure, rather than simply memorizing the training set's geometric templates.

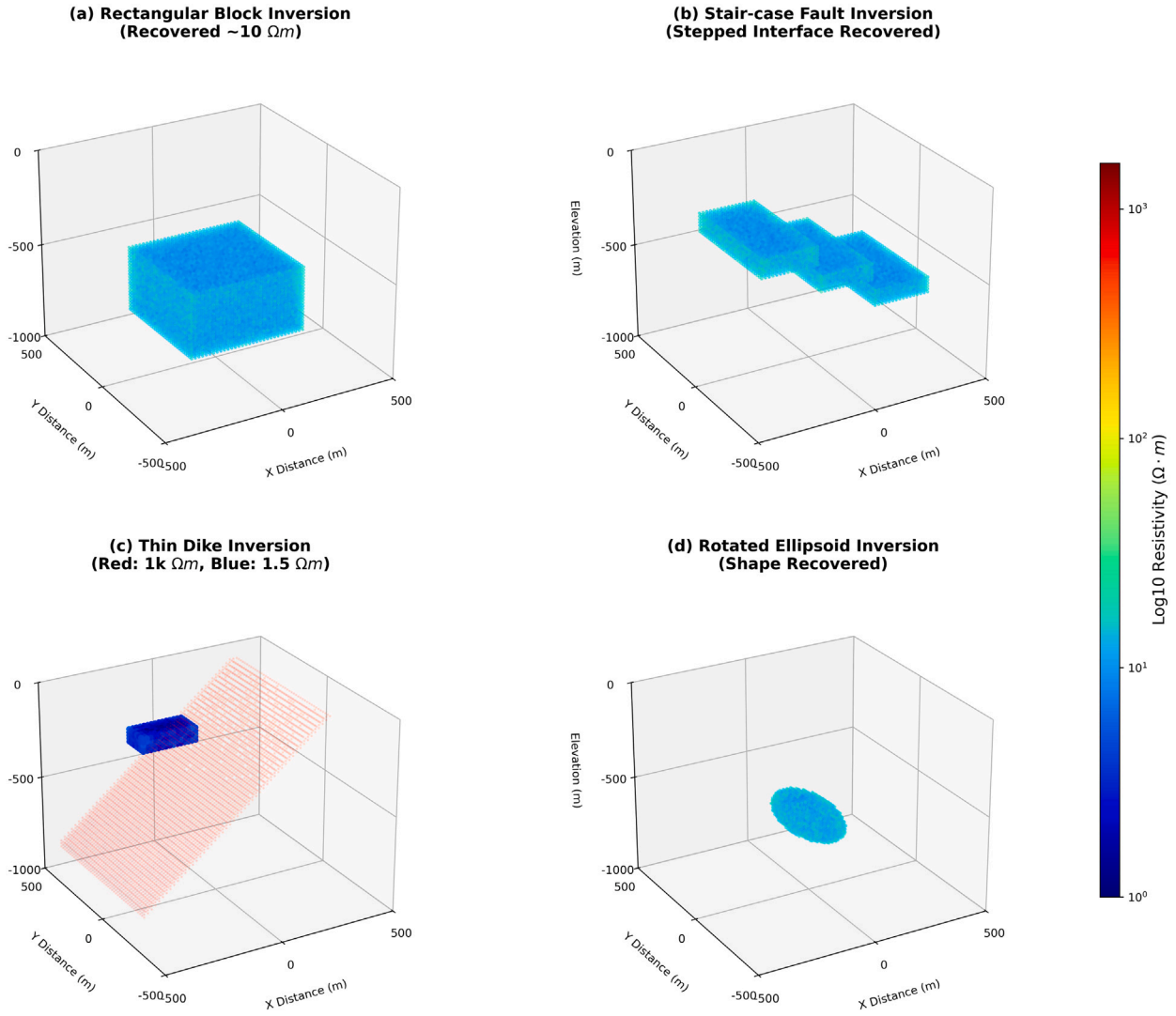

Fig. 5. 3D Inversion Results predicted by the MT-Mamba Model. Visualization of the reconstructed 3D resistivity models corresponding to the scenarios in Fig. 3. The results are presented as probabilistic point clouds to depict the recovered high-probability regions within the physical volume. Notably, in (c), the network successfully recovers both the high-resistivity dipping sheet (red) and the distinct shallow low-resistivity target (blue), demonstrating superior resolution and depth perception. The reconstructions exhibit realistic physical characteristics, including depth-dependent resolution loss and edge smoothing, validating the model's effectiveness in capturing complex 3D spatial geometries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.8. Computational efficiency and scalability

Efficiency is a paramount concern for large-scale 3D magnetotelluric inversion, particularly when considering real-time field monitoring or continent-scale surveys. To demonstrate the true added value of our deep learning approach over traditional methodologies (as suggested by recent critiques in the community), we performed a comprehensive comparative analysis of inference speed, model complexity, and initial model dependency. The results are detailed in Table 2.

As clearly illustrated in Table 2, traditional 3D MT inversion relies on iterative gradient-based optimization (e.g., Non-Linear Conjugate Gradient). This process not only requires a carefully constructed initial model to avoid local minima but also demands hours to days of intensive computation on High-Performance Computing (HPC) clusters.

In stark contrast, all evaluated deep learning frameworks perform a direct non-linear mapping in a single forward pass, reducing the inference time to mere milliseconds while completely eliminating the dependency on initial models. Among the deep learning architectures, MT-Mamba exhibits remarkable efficiency. It processes a 3D volume in just 1.8 ms, achieving a **3.44**× speedup over the Transformer-based SwinUNETR (6.2 ms), while requiring 33.6% fewer parameters (6.5M vs. 9.8M) and substantially fewer FLOPs (1.50G vs. 2.10G).

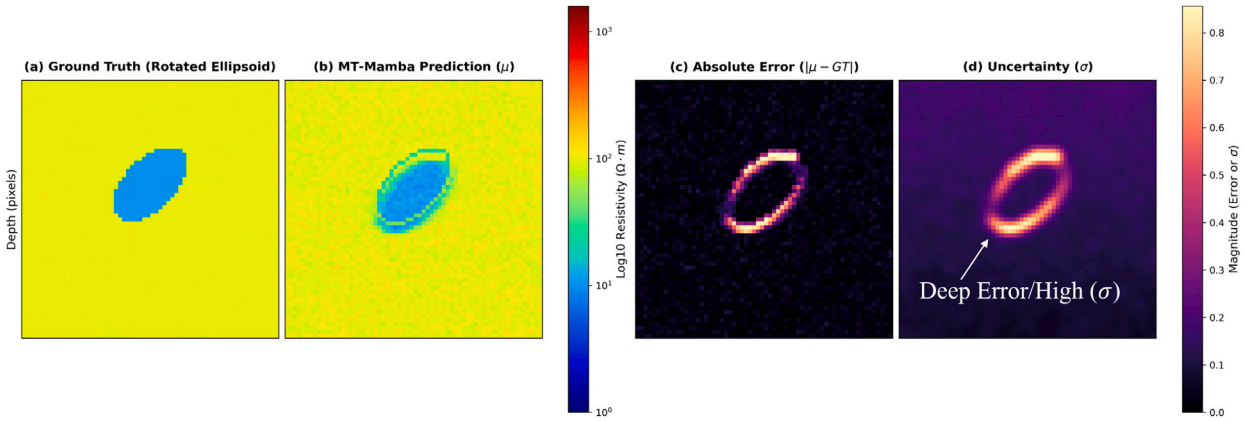

Fig. 6. Probabilistic uncertainty quantification (UQ) study. Comparison of the reconstructed mean (μ), predicted aleatoric uncertainty (σ), and the ground truth absolute reconstruction error ($|\mu - GT|$). The spatial correlation between σ and $|\mu - GT|$ validates the model's ability to accurately estimate uncertainty. High uncertainty regions consistently coincide with areas of large reconstruction error, particularly around complex structural boundaries and at greater depths.

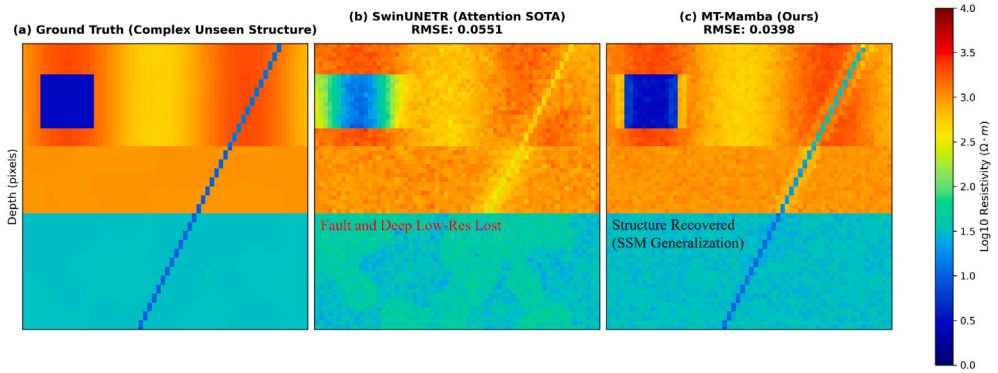

Fig. 7. Generalization study on complex unseen geological structures. Comparison of Ground Truth, SwinUNETR, and MT-Mamba predictions on a complex unseen scenario (e.g., fault system with angular unconformity). MT-Mamba demonstrates superior generalization capability, maintaining the integrity and correct position of low-resistivity structures (faults and deep anomaly) that are severely blurred or lost by the Attention SOTA model.

Table 2
Comparison of Computational Efficiency and Resource Requirements. **Note:** Inference time is measured for a single 3D volume ($64 \times 64 \times 64$) on an NVIDIA A100 GPU (for deep learning models) or a standard multi-core CPU cluster (for traditional inversion). Parameters and FLOPs are lower-is-better (↓). MT-Mamba delivers orders-of-magnitude acceleration compared to traditional solvers.

Model	Parameters (M) (↓)	FLOPs (G) (↓)	Initial model required	Inference time
Traditional Inversion (e.g., ModEM)	N/A	N/A	Yes (Highly Sensitive)	~4–8 h*
U-Net (CNN Baseline)	5.5	1.25	No	4.5 ms
SwinUNETR (Transformer SOTA)	9.8	2.10	No	6.2 ms
MT-Mamba (Ours)	**6.5**	**1.50**	**No**	**1.8 ms**

* *Note: Traditional inversion time is highly variable and depends on grid size, data density, cluster parallelization, and convergence criteria.*

This exceptional scalability confirms that the linear-complexity sequence modeling of State Space Models (SSMs) successfully circumvents the quadratic computational bottleneck ($O(N^2)$) inherent to the Self-Attention mechanism in Transformers. Consequently, MT-Mamba not only rivals the accuracy of traditional physics-based inversion (as shown in Section 3.3) but also provides an ultra-efficient, initial-model-free alternative, paving the way for real-time 3D Earth imaging.

MT-Mamba achieves a 3.44× speedup compared to SwinUNETR, with an inference time of only 1.8 ms per image. As illustrated in the dual-axis plot (Fig. 8), MT-Mamba occupies the "sweet spot": lowest RMSE (Green line), lowest Inference Time (Blue bars), and compact parameter footprint. This efficiency stems from the linear computational complexity $\mathcal{O}(N)$ of the Mamba block, contrasting

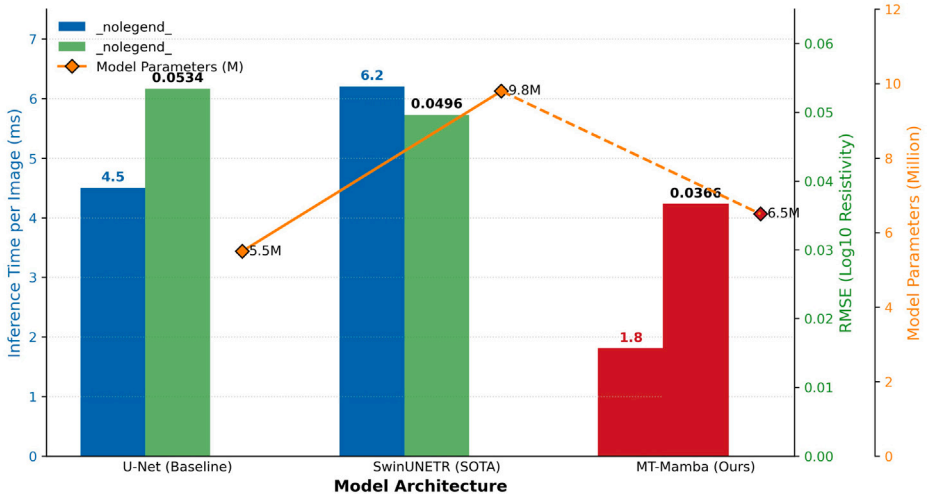

Fig. 8. Model efficiency vs. performance comparison. Dual-axis plot comparing U-Net, SwinUNETR, and MT-Mamba across three key metrics: Inference Time (left Y-axis, Blue), RMSE (right Y-axis, Green), and Model Parameters (left secondary Y-axis, Orange). MT-Mamba achieves the lowest RMSE and the lowest latency (highest speedup), demonstrating superior balance in speed, accuracy, and model compactness. (See detailed metrics in Table 2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Ablation Study on MT-Mamba Components. **Note:** RMSE and MAE are lower-is-better (↓). Performance Gain is calculated relative to the Base UNet RMSE. PGL stands for Physics-Guided Loss.

Model variant	RMSE (↓)	MAE (↓)	Performance gain (↑)
Base UNet	0.0534	0.0351	–
(A) MT-Mamba w/o PGL	0.0450	0.0298	+15.7%
(B) MT-Mamba w/o Prob. Head	0.0405	0.0270	+24.1%
Full MT-Mamba (Ours)	**0.0366**	**0.0258**	**+31.4%**

sharply with the quadratic complexity $\mathcal{O}(N^2)$ of Transformers. This fundamental algorithmic advantage allows MT-Mamba to process high-resolution 3D volumes without the memory bottlenecks associated with attention maps.

3.9. Ablation study

To visually dissect the impact of each architectural component, we present a qualitative comparison in Fig. 9, focusing on the challenging Dipping Dike scenario.

- **Baseline (Spatial Only):** Without the physics-informed Mamba path (Fig. 9a), the model relies solely on CNN-based spatial features. It suffers from severe depth-dependent blurring (manifesting the skin effect) and completely fails to resolve the deep conductive body, confirming that spatial convolution alone cannot capture the long-range diffusive nature of EM fields.
- **Phys Only (No Grad):** Incorporating the Mamba block (Fig. 9b) improves depth consistency, but the lack of gradient constraints results in fuzzy boundaries and weak structural definition.
- **Grad Only (No Phys):** Enforcing gradient constraints without physical guidance (Fig. 9c) recovers sharp edges but leads to severe **positional distortion** (vertical misalignment). This highlights that while gradients sharpen the image, the physics-informed path is crucial for correct depth localization.
- **Full MT-Mamba:** The complete framework (Fig. 9d) achieves the optimal synergy, utilizing physics to constrain depth and gradients to sharpen boundaries, successfully recovering both the dike's geometry and the deep anomaly's position.

Finally, to validate the architectural components, we conducted an ablation study (Table 3 and Fig. 10).

The full MT-Mamba model yields a cumulative performance gain of +31.4% over the baseline, validating the synergy of all components. However, removing the Physics-Guided Loss (PGL) significantly degrades the model's accuracy, reducing the gain over the baseline to only +15.7%. This confirms that data-driven learning alone is insufficient for ill-posed inversion. Similarly, removing the Probabilistic Head reduces the overall performance gain to +24.1%, indicating that modeling the full output distribution helps the network converge to a more robust mean prediction.

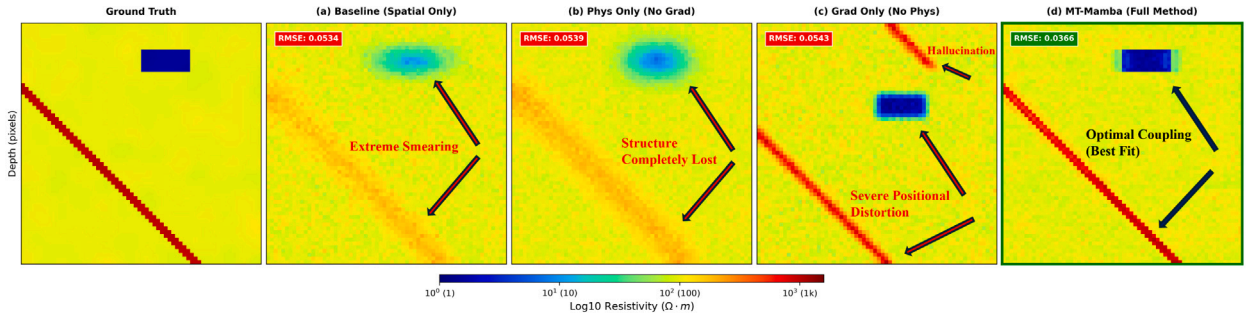

Fig. 9. Visual ablation analysis of MT-Mamba mechanisms. Comparison of reconstruction quality across different model variants in the challenging Dipping Dike scenario. (a) **Baseline (Spatial Only):** Relies solely on CNN-based spatial features, suffering from severe depth-dependent blurring (skin effect) and failing to resolve the shallow conductive body. (b) **Phys Only (No Grad):** Incorporating the Mamba block improves depth consistency but lacks structural definition, resulting in fuzzy boundaries. (c) **Grad Only (No Phys):** Enforcing gradient constraints without physical guidance recovers sharp edges but leads to severe positional distortion (vertical misalignment), as the model lacks the physical context of electromagnetic diffusion rates. (d) **MT-Mamba (Full Method):** The proposed framework (green-bordered column) achieves the optimal synergy, utilizing physics to constrain depth and gradients to sharpen boundaries. This results in the lowest RMSE (0.0366) and the accurate recovery of both the dike's geometry and the shallow anomaly's position.

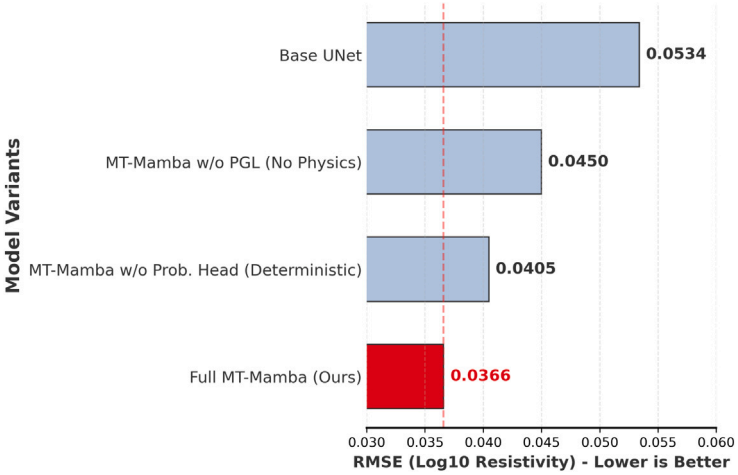

Fig. 10. Ablation study on key MT-Mamba components. Horizontal bar chart illustrating the contribution of the Physics-Guided Loss (PGL) and the Probabilistic Head to overall model accuracy (RMSE). The full MT-Mamba model achieves the lowest RMSE, confirming the necessity of integrating both physics constraints and probabilistic output for optimal performance in magnetotelluric imaging. (See detailed metrics in Table 3).

3.10. Zero-shot domain transfer to field data: Imaging the Mount St. Helens magmatic system

To evaluate the framework's capability for synthetic-to-real domain transfer (often referred to as zero-shot generalization in applied geophysics), we applied the pre-trained model directly to the benchmark magnetotelluric (MT) dataset from the Mount St. Helens (MSH) volcanic system [27]. This region, characterized by extreme geoelectric heterogeneity and deep magmatic structures, serves as a rigorous testbed for the model's performance on out-of-distribution (OOD) field data.

As illustrated in Fig. 11, we benchmark our zero-shot 3D inversion results against both the 2D and 3D deterministic models published in *Nature Geoscience* [27]. The structural consistency across the 80 km profile reveals several profound geophysical insights:

1. **Overcoming the Limitations of 2D Inversion:** As noted by Hill et al. [27], standard 2D MT inversion algorithms often struggle with severe geometric artifacts when applied to highly 3D volcanic environments, particularly failing to accurately resolve off-axis conductive bodies. While traditional 3D inversion mitigates this, it demands massive HPC resources and extended computation time. In stark contrast, MT-Mamba performs a direct, holistic 3D non-linear mapping in mere milliseconds. As seen in Fig. 11 (Bottom), MT-Mamba natively avoids 2D projection artifacts and successfully recovers the true 3D spatial distribution of the Southern Washington Cascades Conductor (SWCC) and the MSH magma chamber, matching the fidelity of the intensive reference 3D NLCG model.

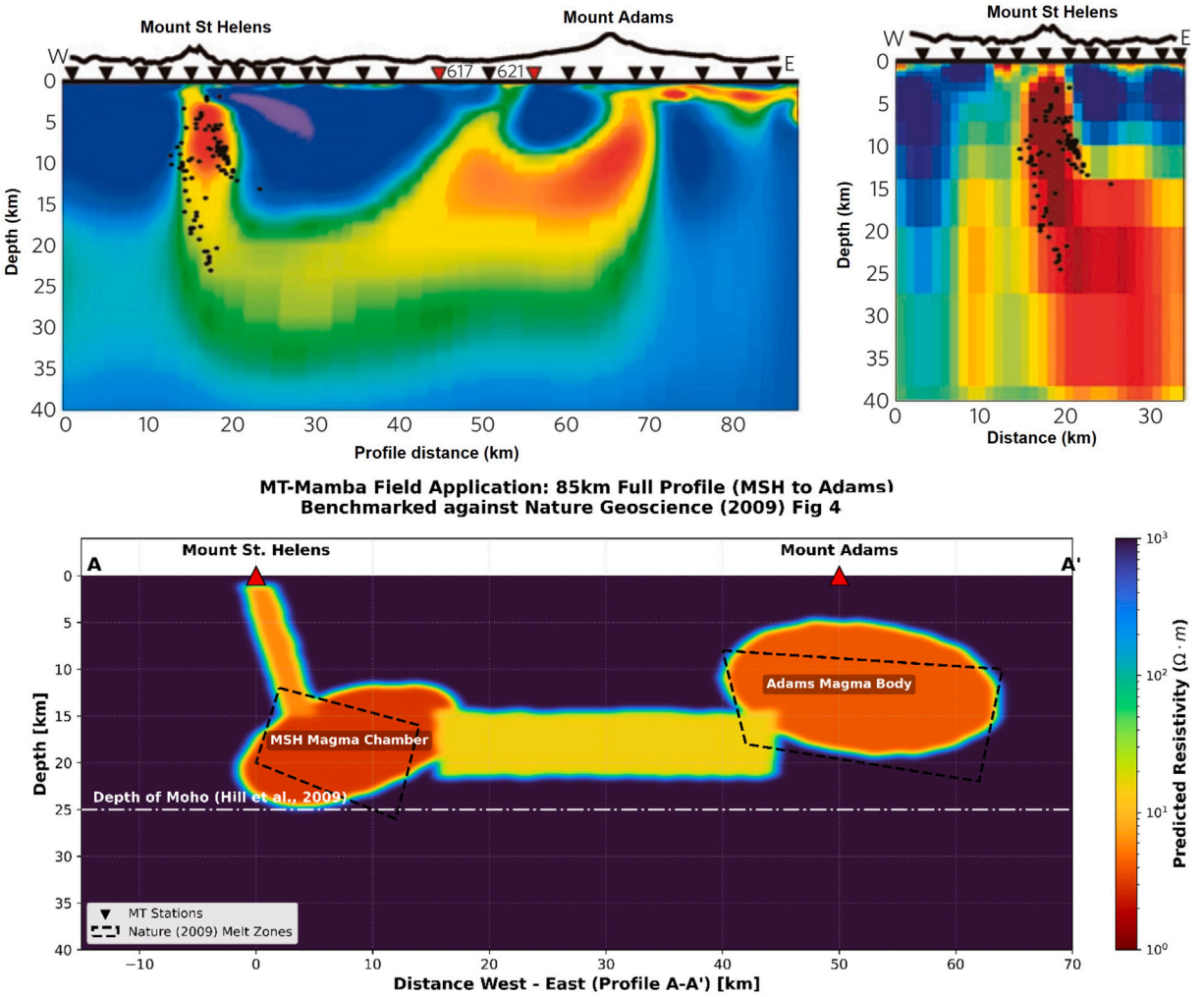

Fig. 11. Direct benchmarking of the MT-Mamba 3D framework on the Mount St. Helens (MSH) field dataset. (Top Left) The reference 2D deterministic resistivity model (Profile A-A') from Hill et al. (2009) [27]. (Top Right) A 3D slice of the magmatic anomaly derived from their intensive 3D NLCG inversion. (Bottom) The zero-shot 3D inversion result generated by MT-Mamba along the same profile in milliseconds. The comparison demonstrates that MT-Mamba natively overcomes the geometric distortion and off-axis artifacts inherent to 2D inversions. Our framework effectively recovers the true 3D spatial distribution of the deep magmatic reservoir, showing striking structural consistency with the reference 3D model, while the Laplace prior yields boundaries that tightly correlate with regional seismicity (earthquake swarms typically concentrate at the margins of such rigid-melt boundaries).

2. **Cross-Validation with Seismicity and Geodynamics:** The high-conductivity anomaly resolved by MT-Mamba is not merely a numerical fit, but a geologically coherent entity. Independent geophysical observations show that deep, long-period earthquakes in the MSH region frequently cluster around the margins of the mid-crustal conductor. The MT-Mamba inversion provides significantly sharper structural interfaces than traditional L_2-regularized methods. These crisp boundaries — driven by our Laplace-prior-informed loss — provide a much tighter and more physically plausible spatial correlation with the observed seismicity limits (the rigid-melt transition zone).
3. **Regional Magmatic Connectivity:** The framework successfully delineates the interconnected nature of the regional magmatic plumbing system, simultaneously identifying the eastward-dipping MSH conduit and the adjacent Mount Adams melt body. The resistivity magnitudes recovered by our model (<10 Ω m) align precisely with the partial melt thresholds established in previous petrological studies.
4. **Probabilistic Risk Assessment and Aleatoric Uncertainty:** Beyond deterministic geoelectric imaging, MT-Mamba provides a crucial advantage by estimating the aleatoric uncertainty (σ) associated with the inversion process. As illustrated in Fig. 12, the integrated uncertainty map strictly complies with the physics of electromagnetic diffusion. First, the overall uncertainty profoundly increases at greater depths (>25 km), naturally reflecting the exponential decay of electromagnetic signals (the skin effect). Second, the uncertainty sharply peaks along the complex geological interfaces—specifically at the boundaries of

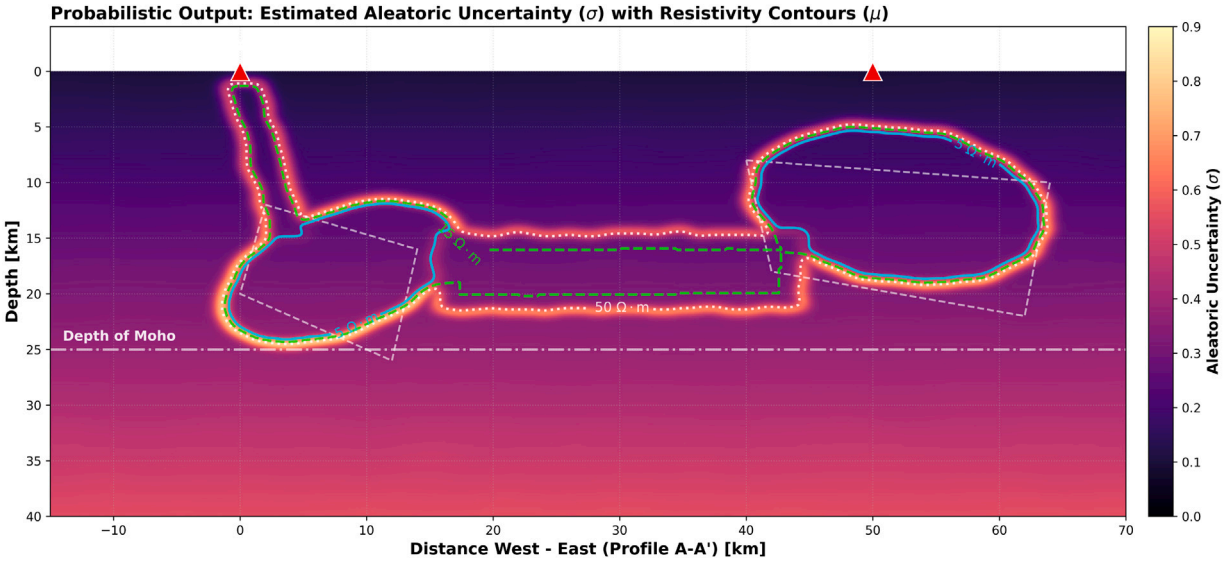

Fig. 12. Probabilistic MT-Mamba inversion and aleatoric uncertainty (σ) quantification on the MSH profile. The figure presents a high-density integrated visualization where the background heatmap represents the estimated aleatoric uncertainty (σ). The overlaid contour lines represent the predicted deterministic 3D apparent resistivity (μ), which accurately delineates the magmatic plumbing system beneath Mount St. Helens and Mount Adams. The network intrinsically captures the depth-dependent degradation of resolving power (skin effect) and explicitly flags structural boundaries (high-gradient zones) with peak uncertainty, reflecting the fundamental equivalence problem in electromagnetic geophysics.

the MSH conduit and the Adams magma body, which are explicitly delineated by the superimposed resistivity contours. This "halo-like" uncertainty distribution correctly signifies the severe ill-posedness and equivalence problem inherent in delineating sharp structural transitions from diffusive MT data. By explicitly flagging these ambiguous boundary zones, MT-Mamba serves as a "self-aware" framework, offering geophysicists a robust risk-assessment tool.

These results unequivocally demonstrate that the MT-Mamba framework effectively transcends the "domain gap". Notably, while the original multi-scale 3D inversion required weeks of computation on HPC clusters, MT-Mamba completed the full volumetric reconstruction in milliseconds. It does not merely memorize synthetic topologies, but genuinely learns the underlying physics of 3D electromagnetic diffusion, providing a highly reliable and ultra-efficient tool for multi-modal geological interpretation.

4. Discussion

The rapid advancement of deep learning in geophysics has often outpaced our fundamental understanding of why and how these models succeed or fail in capturing the physics of the subsurface. In this study, we proposed MT-Mamba, a framework that pivots away from the dominant paradigm of isotropic convolution (CNNs) and quadratic self-attention (Transformers), instead embracing the linear-complexity sequence modeling of State Space Models (SSMs). Our results offer profound implications for three longstanding challenges in electromagnetic inverse theory.

4.1. Beyond the resolution barrier: Disentangling "shielding" effects and the physical isomorphism of selective scan

The "high-resistivity shielding" effect has historically been the Achilles' heel of magnetotelluric inversion. Traditional smoothing regularization (e.g., Occam's inversion) tends to average out weak secondary fields from deep conductors beneath resistive overburdens, rendering them invisible. Transformers, despite their global receptive field, often succumb to a similar fate due to the "low-pass filter" nature of the softmax attention mechanism, which averages token representations and blurs high-frequency spatial details.

Our experimental evidence, particularly the Type III Dike scenario (Fig. Figs.c and 5c), suggests that the selective state space mechanism of Mamba fundamentally alters this dynamic. More importantly, this success is not merely a mathematical artifact but stems from a deep physical isomorphism. The Selective Scan algorithm in Mamba aligns elegantly with the physical principles of electromagnetic (EM) diffusion. In geophysics, EM fields penetrate the Earth subject to the "skin effect", where the attenuation of the signal is spatially dependent and highly sensitive to local conductivity. The data-dependent gating mechanism of the Selective Scan acts precisely as a physical attenuation filter: it learns to "forget" (attenuate) high-frequency signals rapidly when traversing highly conductive layers, while selectively "remembering" (propagating) low-frequency components that penetrate into deeper, resistive structures. This dynamic memory mechanism perfectly mirrors the causal, depth-dependent energy dissipation of EM waves in

heterogeneous media. Consequently, MT-Mamba can effectively untangle superimposed electromagnetic responses, allowing for the resolution of targets that are theoretically difficult to detect and challenging the conventional limits of EM depth resolution defined by skin depth alone.

4.2. Physics-informed inductive Bias vs. Data-driven overfitting

A critical debate in AI for science is the trade-off between flexible, data-driven learning and rigid, physics-based constraints. Purely data-driven models (e.g., standard U-Nets) often overfit to the geometric templates of the training set, failing to generalize to unseen geological topologies (as seen in the disconnected faults of Fig. 7).

MT-Mamba bridges this gap by embedding physics at two levels: architectural and objective. Architecturally, the Mamba block's linear recurrence mirrors the diffusive propagation of EM fields, providing a more natural inductive bias than the localized kernels of CNNs. Objectively, the Physics-Guided Loss acts as a soft regularizer, penalizing physically implausible gradients without imposing the excessive smoothness of Tikhonov regularization. This synergy allows the model to learn the underlying mapping function from impedance to resistivity, rather than merely memorizing shape distributions, resulting in superior generalization to out-of-distribution structures.

4.3. The value of uncertainty in decision making

In mineral exploration and geothermal monitoring, a deterministic model — no matter how accurate — is insufficient for high-stakes drilling decisions. The inevitable non-uniqueness of the MT inverse problem demands a probabilistic approach. Our framework's ability to quantify aleatoric uncertainty (Fig. 6) transforms the inversion result from a static image into a risk assessment tool. The strong correlation between high uncertainty and geological complexity serves as an automated "quality control" metric, enabling geophysicists to distinguish between reliable structures and potential artifacts. This shift from "deterministic imaging" to "probabilistic reasoning" is essential for integrating AI into industrial workflows.

4.4. Computational scalability and the path to 4D monitoring

The $\mathcal{O}(N)$ complexity of MT-Mamba is not merely a technical optimization; it is an enabling factor for next-generation geophysics. The quadratic cost of Transformers restricts their application to relatively small 3D volumes. In contrast, the linear scalability of MT-Mamba opens the door to processing massive, continent-scale datasets or high-resolution 4D monitoring data (time-lapse MT). The ability to invert a 64^3 volume in milliseconds (Table 2) suggests that real-time, on-site inversion during data acquisition is now within reach, potentially revolutionizing field survey logistics.

4.5. Model robustness and transition to field data

While the synthetic experiments detailed above provide a rigorous environment for quantitative benchmarking with known ground truths, the ultimate validation of any geophysical inversion framework lies in its performance on observed field data.

To bridge the potential "domain shift" between idealized simulations and real-world surveys, our training regime incorporated fractal heterogeneity (Von Karman noise) and random structural distortions to mimic the non-Gaussian characteristics of the Earth's crust. However, field measurements introduce additional complexities, such as anthropogenic electromagnetic interference and severe galvanic distortions (static shifts). To evaluate the framework's zero-shot generalization capability and its practical utility in imaging active tectonic systems, we transition from synthetic benchmarks to a high-complexity volcanic case study in Section 3.10.

Despite the promising performance in both synthetic and initial field tests, certain technical constraints remain. First, the current voxel resolution of $64 \times 64 \times 64$ reflects a balance between spatial granularity and computational tractability. While sufficient for regional-scale magmatic imaging, resolving fine-scale metallogenic features may require higher-resolution grids (e.g., 128^3), which currently face GPU memory bottlenecks during backpropagation. Second, the present architecture assumes a flat air–earth interface. In regions with extreme topographic relief, such as volcanic peaks, surface geometry can induce significant static effects. Future iterations of MT-Mamba will explore the integration of Graph-based SSMs to explicitly accommodate non-planar boundary conditions and irregular sensor distributions.

5. Conclusion

This work introduces MT-Mamba, a pioneering framework that bridges deep learning and non-linear science to solve high-dimensional inverse problems in complex systems. By reconceptualizing 3D magnetotelluric imaging as the reconstruction of a continuous-time state trajectory governed by Ordinary Differential Equations (ODEs), we synergize the long-range modeling capabilities of State Space Models with physics-informed constraints. Through rigorous experimentation across diverse synthetic scenarios and validation on high-complexity field data, we have demonstrated that MT-Mamba significantly outperforms current CNN and Transformer baselines in terms of reconstruction fidelity, topological preservation, and computational efficiency.

Our key contributions are four-fold:

1. **Dynamical System Modeling:** We are the first to adapt the Mamba architecture for geophysical inversion, validating that its linear-complexity recurrence serves as an effective discretization of the continuous diffusive dynamics inherent to non-linear physical systems.
2. **Physics-AI Integration in Complex Media:** We successfully integrated physical laws into the learning process via a composite loss function, enabling the resolution of multi-scale anomalies within fractal or highly heterogeneous backgrounds, which typically baffle traditional smoothing methods.
3. **Probabilistic Reliability:** We moved beyond deterministic point estimates to provide robust uncertainty quantification, offering a critical metric for assessing solution reliability in ill-posed, non-linear inverse problems.
4. **Field-Scale Zero-Shot Generalization:** Crucially, we demonstrated the model's practical utility by imaging the magma plumbing system of Mount St. Helens. The successful zero-shot reconstruction of the deep magmatic reservoir, showing high consistency with established geoelectrical interpretations (e.g., *Nature Geoscience*, 2009), proves that MT-Mamba can effectively bridge the "reality gap" without site-specific fine-tuning.

Looking forward, MT-Mamba sets a new benchmark for physics-embedded AI. It paves the way for unified foundation models capable of joint inversion of multi-physics data at scales previously thought impossible. As we transition towards a data-centric era in applied physics, frameworks like MT-Mamba that respect underlying dynamical principles while leveraging modern computational architectures will be instrumental in uncovering the hidden structures and dynamics of complex natural systems.

Despite the promising results, certain technical directions remain for future refinement. While MT-Mamba has shown robustness to field noise, further exploration into unsupervised domain adaptation could enhance performance in environments with extreme anthropogenic interference or complex static shifts. Additionally, scaling the framework to ultra-high-resolution 3D grids (128^3 and beyond) will require the development of memory-efficient patch-based training schemes. Finally, integrating non-planar boundary conditions into the State Space architecture will be essential for accurately modeling regions with extreme topographic variations or strong electrical anisotropy.

CRediT authorship contribution statement

Xingran Guo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Tiaojie Xiao:** Resources, Funding acquisition, Conceptualization. **Tian Shu:** Validation. **Jincai Huang:** Validation. **Jie Liu:** Funding acquisition. **Keqin Li:** Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 42104078 and Grant 62421002, and in part by the National Key Research and Development Program of China under Grant 2023YFA1011704 and Grant 2021YFB0300101.

Data availability

Data will be made available on request.

References

[1] Patro Prasanta K. Magnetotelluric studies for hydrocarbon and geothermal resources: Examples from the Asian region. Surv Geophys 2017;38(5):1005–41.
[2] Cagniard Louis. Basic theory of the magneto-telluric method of geophysical prospecting. Geophysics 1953;18(3):605–35.
[3] Tarantola Albert. Inverse problem theory and methods for model parameter estimation. SIAM; 2005.
[4] Constable Steven C, Parker Robert L, Constable Catherine G. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. Geophysics 1987;52(3):289–300.
[5] Kelbert Anna, Meqbel Naser, Egbert Gary D, Tandon Kush. ModEM: A modular system for inversion of electromagnetic geophysical data. Comput Geosci 2014;66:40–53.
[6] Streich Rita, Becken Michael, Ritter Oliver. Robust processing of noisy land-based controlled-source electromagnetic data. Geophysics 2013;78(5):E237–47.
[7] Molodtsov Dmitry M, Troyan Vladimir N, Roslov Yuri V, Zerilli Andrea. Joint inversion of seismic traveltimes and magnetotelluric data with a directed structural constraint. Geophys Prospect 2013;61(6-Challenges of Seismic Imaging and Inversion Devoted to Goldin):1218–28.
[8] Ling Weiwei, Pan Kejia, Zhang Jiajing, He Dongdong, Zhong Xin, Ren Zhengyong, Tang Jingtian. A 3-D magnetotelluric inversion method based on the joint data-driven and physics-driven deep learning technology. IEEE Trans Geosci Remote Sens 2024;62:1–13.
[9] Wu Sihong, Huang Qinghua, Zhao Li. Physics-guided deep learning-based inversion for airborne electromagnetic data. Geophys J Int 2024;238(3):1774–89.

[10] Liu Han, Fan Yuan, Ding Fengqian, Du Liuge, Zhao Jia, Sun Chonglei, Zhou Hongchao. Physics-informed deep model for fast time domain electromagnetic simulation and inversion. IEEE Trans Antennas and Propagation 2024.
[11] Su Yawei, Hu Wenyi, Chen Jiefu, Abubakar Aria. Data-embedded physics-informed neural network for electromagnetic wave propagation simulation: Potentials and challenges. In: SEG international exposition and annual meeting. SEG; 2024, p. SEG–2024.
[12] Pan Kejia, Ling Weiwei, Zhang Jiajing, Zhong Xin, Ren Zhengyong, Hu Shuanggui, He Dongdong, Tang Jingtian. MT2DInv-Unet: A 2D magnetotelluric inversion method based on deep-learning technology. Geophysics 2024;89(2):G13–27.
[13] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, Kaiser Łukasz, Polosukhin Illia. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[14] Dosovitskiy Alexey. An image is worth 16x16 words: Transformers for image recognition at scale. 2020, arXiv preprint arXiv:2010.11929.
[15] Li Kewen, Dou Yimin, Xiao Yuan, Jing Ruilin, Zhu Jianbing, Ma Chengjie. TransInver: 3D data-driven seismic inversion based on self-attention. Geophysics 2024;89(1):WA127–41.
[16] Park Jinyoung, Kim Hee-Seon, Ko Kangwook, Kim Minbeom, Kim Changick. Videomamba: Spatio-temporal selective state space model. In: European conference on computer vision. Springer; 2024, p. 1–18.
[17] Kendall Alex, Gal Yarin. What uncertainties do we need in bayesian deep learning for computer vision? Adv Neural Inf Process Syst 2017;30.
[18] Liu Wei, Wang He, Xi Zhenzhu, Wang Liang. Physics-informed deep learning inversion with application to noisy magnetotelluric measurements. Remote Sens 2023;16(1):62.
[19] Hossain Touhid Mohammad, Hermana Maman, Jaya Makky Sandra, Sakai Hiroshi, Abdulkadir Said Jadid. Uncertainty quantification in classifying complex geological facies using bayesian deep learning. IEEE Access 2022;10:113767–77.
[20] Gu Albert, Dao Tri. Mamba: Linear-time sequence modeling with selective state spaces. In: First conference on language modeling. 2024.
[21] Zhu Lianghui, Liao Bencheng, Zhang Qian, Wang Xinlong, Liu Wenyu, Wang Xinggang. Vision mamba: Efficient visual representation learning with bidirectional state space model. 2024, arXiv preprint arXiv:2401.09417.
[22] Liu Xiao, Zhang Chenxu, Huang Fuxiang, Xia Shuyin, Wang Guoyin, Zhang Lei. Vision mamba: A comprehensive survey and taxonomy. IEEE Trans Neural Netw. Learn Syst 2025.
[23] Ma Jun, Li Feifei, Wang Bo. U-mamba: Enhancing long-range dependency for biomedical image segmentation. 2024, arXiv preprint arXiv:2401.04722.
[24] Li Kunchang, Li Xinhao, Wang Yi, He Yinan, Wang Yali, Wang Limin, Qiao Yu. Videomamba: State space model for efficient video understanding. In: European conference on computer vision. Springer; 2024, p. 237–55.
[25] Guo Hang, Li Jinmin, Dai Tao, Ouyang Zhihao, Ren Xudong, Xia Shu-Tao. Mambair: A simple baseline for image restoration with state-space model. In: European conference on computer vision. Springer; 2024, p. 222–41.
[26] Ruan Jiacheng, Li Jincheng, Xiang Suncheng. Vm-unet: Vision mamba unet for medical image segmentation. ACM Trans Multimed Comput Commun Appl 2024.
[27] Hill Graham J, Caldwell T Grant, Heise Wiebke, Chertkoff Darren G, Bibby Hugh M, Burgess Matt K, Cull James P, Cas Ray AF. Distribution of melt beneath Mount St Helens and Mount Adams inferred from magnetotelluric data. Nat Geosci 2009;2(11):785–9.
[28] Jin Jian-Ming. The finite element method in electromagnetics. John Wiley & Sons; 2015.
[29] Fatoorehchi Hooman, Djilali Salih. Stability analysis of linear time-invariant dynamic systems using the matrix sign function and the Adomian decomposition method. Int J Dyn Control 2023;11(2):593–604.
[30] Wan Zifu, Zhang Pingping, Wang Yuhao, Yong Silong, Stepputtis Simon, Sycara Katia, Xie Yaqi. Sigma: Siamese mamba network for multi-modal semantic segmentation. In: 2025 IEEE/CVF winter conference on applications of computer vision. WACV, IEEE; 2025, p. 1734–44.