buildings

MDPI

Review

Physics-Informed Machine Learning in Subsurface Multiphysics Flow Modeling: Integrating Physical Constraints for Accelerated Simulation

Linchao Wang [1], Fei Xiong [1], Faning Dang [1], Lin Zhu [1] and Yi Xue [1,2,*]

[1] School of Civil Engineering and Architecture, Xi'an University of Technology, Xi'an 710048, China
[2] School of Energy and Mining Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China
* Correspondence: xueyi@xaut.edu.cn

Abstract

Subsurface thermo-hydro-mechanical (THM) coupled processes are fundamental to geomechanics, yet conventional mesh-based methods face high computational costs and limited efficiency in strongly nonlinear simulations and inverse problems. This review examines two representative physics-informed machine learning paradigms for THM modeling: physics-informed neural networks (PINNs) and neural operators (NOs). Relevant studies were identified through iterative keyword-based searches and citation tracking and were comparatively analyzed in terms of physical embedding, data dependence, computational efficiency, inverse capability, generalization, and engineering applications. The analysis shows that PINNs are well suited to physics-constrained simulation and parameter inversion from sparse data but are limited by training instability and loss imbalance. NOs enable rapid repeated forward predictions but depend strongly on representative training data and may perform poorly under out-of-distribution conditions. This review clarifies the complementary roles, trade-offs, and application boundaries of PINNs and NOs and highlights their hybrid integration as a promising route toward efficient and physically consistent subsurface THM simulation.

Keywords: PIML; THM coupling; PINNs; NOs; subsurface engineering simulation; parameter inversion; digital twin

1. Introduction: Subsurface THM-Coupled Processes—Frontier Challenges and Simulation Bottlenecks

1.1. The Importance and Complexity of THM-Coupled Processes

Subsurface THM-coupled processes serve as the fundamental physical basis for numerous energy development and geotechnical engineering systems. The accurate simulation of these processes is directly critical to the reliability of engineering design, operational efficiency, and long-term safety.

The theoretical development of THM coupling has progressed through a long period of refinement. In its early stage, research mainly addressed simplified single-field or two-field coupling problems. Typical examples include hydro-mechanical analyses constructed based on Terzaghi's effective stress principle and Darcy's law, which provided the basic conceptual basis for later multi-field coupling theories. In the 1940s and 1950s, Biot further formulated the theory of poroelasticity by linking pore fluid pressure with deformation of

Check for updates

Academic Editor: Eugeniusz Koda

Received: 6 August 2026
Revised: 29 August 2026
Accepted: 2 September 2026
Published: 4 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the solid skeleton, thereby offering a systematic description of fluid–solid interaction in saturated porous media [1]. Early in the development of nuclear waste geological disposal, the increasing demand for performance and safety assessment promoted the incorporation of thermal effects into coupled models [2]. In particular, McTigue proposed a linear theory for fluid-saturated porous thermoelastic media in 1986 [3]. In the 1990s, international collaborative programs such as DECOVALEX advanced the verification of THM models through benchmark studies and experimental comparisons, which strengthened their application to the geological disposal of high-level radioactive waste [4].

Nevertheless, for deep geotechnical engineering under high-temperature conditions, hydro-mechanical coupling alone is insufficient for reliable engineering design. Thermal stress can substantially influence the mechanical stability and failure response of fractured rock masses. As shown in Figure 1, temperature-induced stress variations change the effective normal stress and shear stress acting on fractures, driving the fracture stress state closer to the Mohr–Coulomb failure envelope. As a result, fracture slip may occur over a wider range of orientations, indicating that fully coupled THM models are indispensable for modern deep underground engineering design.

With the rapid development of unconventional oil and gas resources and enhanced geothermal systems (EGS), THM research has been extended to more complicated conditions, including fractured media and non-isothermal multiphase flow. Since the early 2000s, numerous THM numerical models have been developed to describe multi-field interactions in fractured rock masses. For example, Zhou et al. constructed a fully coupled THM numerical model for EGS and revealed how thermo-mechanical–hydrological interactions control fracture flow behavior and heat transfer efficiency [5]. Improvements in computational capacity and the growing availability of multi-scale observational data further promoted the evolution of THM simulation. The modeling framework has gradually expanded from continuum-based descriptions to discrete fracture networks, dual-porosity models, and multi-scale coupling approaches. These advances have provided important tools for representing complex subsurface flow systems and have also created favorable conditions for the introduction of emerging methods such as physics-informed machine learning (PIML).

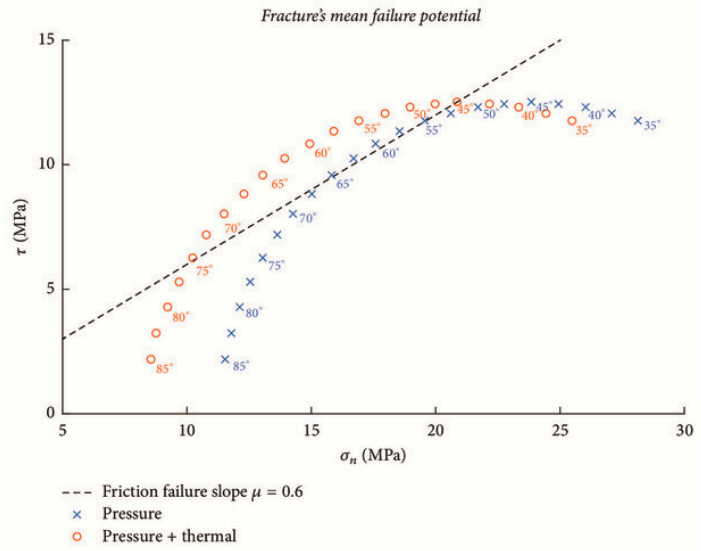


Figure 1. Thermal-stress-induced failure potential of rotated fractures shown in a classical Mohr–Coulomb diagram. Reprinted with permission from Ref. [6]. Copyright 2017 Wiley.

The engineering relevance of THM coupling is particularly prominent across multiple fields. In EGS, the thermal contraction induced by injecting low-temperature fluids alters the stress state of the rock mass, thereby dynamically regulating fracture conductivity and

heat exchange efficiency [7]. In hydraulic fracturing for unconventional oil and gas reservoirs, fluid pressure perturbations and thermo-mechanical responses jointly govern fracture propagation paths and proppant transport distribution. Recent numerical studies have further demonstrated that hydraulic fracture propagation is strongly affected by geological layering and heterogeneity. In layered formations, contrasts in interlayer stiffness, fracture toughness, and in situ stress can lead to distinct fracture-height growth and cross-layer propagation regimes [8], while inherent rock heterogeneity and grain-scale characteristics can alter fracture initiation pressure, propagation tortuosity, and fracture symmetry [9]. More complex stimulation configurations, such as cross-layer fracturing with directional perforation, further introduce strong dependencies on wellbore position, perforation depth, and perforation spacing, resulting in highly variable fracture geometries [10]. In coalbed methane (CBM) extraction, the shrinkage of the coal matrix caused by gas desorption leads to the time-dependent evolution of fracture permeability [11]. Furthermore, in heavy oil thermal recovery, the intense thermo-hydrological coupling driven by steam injection not only reduces oil viscosity but also significantly affects reservoir mechanical stability through associated thermal expansion [12]. Additionally, the decay heat generated during the geological disposal of high-level radioactive waste influences the long-term sealing performance of deep host rocks via THM coupling pathways. All these examples demonstrate that neglecting multi-field coupling effects can lead to severe misjudgments regarding system behavior, engineering productivity, and long-term safety.

From a physical mechanism perspective, THM coupling originates from the nonlinear interactions among temperature, fluid flow, and mechanical deformation. Among these interactions, the feedback of the mechanical state on hydraulic properties is particularly important. Changes in fluid pressure or temperature alter the effective stress state of porous or fractured media, whereas the resulting deformation modifies porosity, fracture aperture, and permeability, thereby affecting fluid and heat transport. In more complex thermo-hydro-mechanical-chemical (THMC) systems, chemical processes such as adsorption and desorption may further influence these coupled responses, increasing the complexity and predictive difficulty of subsurface systems.

From a mathematical perspective, THM processes can be described by multi-scale partial differential equation systems with strong nonlinear coupling. The interactions among thermal, hydraulic, and mechanical fields usually introduce pronounced stiffness and make the solution space highly complex. Nevertheless, in practical engineering cases involving strong nonlinearity, pronounced heterogeneity, or large contrasts in material properties, conventional numerical approaches often suffer from convergence difficulties and substantial computational burdens. Consequently, they are not always able to satisfy the practical requirements of multi-scenario optimization, fast parameter inversion, and real-time prediction.

1.2. Inherent Bottlenecks of Traditional Numerical Simulation and the Demand for Acceleration

Traditional numerical methods, represented by the finite element method, the finite difference method, and the finite volume method, have long served as core tools for solving underground THM coupling problems. Although these methods are highly mature in terms of theoretical rigor and engineering applicability, their inherent bottlenecks are becoming increasingly prominent when addressing modern engineering demands for high timeliness, strong adaptability, and intelligent decision-making, primarily manifesting in the following three aspects:

First, there is a major limitation in computational efficiency. THM coupling problems typically involve three-dimensional unstructured meshes, highly nonlinear constitutive relationships, and long-term time-step iterations, resulting in a single forward simula-

tion taking several hours or even days [13]. In unconventional oil and gas fracturing simulations, fluid flow, solid deformation, heat conduction, and damage evolution must be coupled synchronously, with each nonlinear iteration requiring the repeated solution of large-scale sparse linear equation systems. Such high computational costs make it difficult for traditional methods to support tasks that require hundreds or thousands of forward simulations, such as uncertainty quantification, real-time monitoring feedback, or online optimization.

Second, the capability and efficiency of parameter inversion are insufficient. Traditional inversion methods typically rely on gradient-based optimization algorithms or Bayesian inference frameworks, where each evaluation of the objective function requires calling a complete numerical simulation [14]. For high-dimensional spatially varying parameter fields, the inversion cost can increase rapidly with the parameter dimension. Even in conventional reservoir seepage inversion, gradient-based methods may converge to local optima because of the non-convexity of the objective function [15]. In THM multi-field strongly coupled systems, the nonlinearity of rock mechanics (such as contact and plastic damage) further exacerbates the complexity of the loss function landscape, making this issue even more severe; moreover, the mathematical derivation and code implementation for these systems are technically demanding. Furthermore, field observational data are often characterized by sparsity, noise, and multi-source heterogeneity. Traditional inversion frameworks struggle to effectively integrate such information, resulting in parameter sets with weak physical consistency and high uncertainty.

Third, there is a lack of support for multi-scenario optimization and rapid decision-making. Modern engineering design and operation often require evaluating system performance across numerous combinations of operational parameters to achieve optimal scheme selection. However, traditional high-fidelity simulators generally cannot provide "second-level" responses, which limits the practical application of advanced decision-making methods such as simulation-based optimization, stochastic programming, and reinforcement learning. For example, in heavy oil thermal recovery, the expansion morphology of the steam chamber is highly sensitive to injection strategies. To maximize recovery, thousands of schemes need to be evaluated rapidly, yet the runtime of a single traditional simulation far exceeds the practical decision-making time window. Although some studies have attempted to develop acceleration methods such as approximate physics-based discrete simulations [16], their accuracy in coupling the mechanical field is often limited, making it difficult to balance computational efficiency with multi-field fidelity.

Overall, the limitations of conventional numerical methods in computational efficiency, parameter inversion, and multi-scenario adaptability have become important barriers to the intelligent and precise management of underground THM systems. These limitations are fundamentally associated with the mathematical nature of THM processes. Specifically, such systems are described by strongly nonlinear coupled partial differential equations together with complex constitutive laws, which makes numerical solvers highly dependent on costly iterative procedures. Although scalable parallel THM simulators based on embedded discrete fracture models (EDFM) have been developed [17], their performance still faces challenges in practical applications involving high dimensionality, strong heterogeneity, and strict time requirements. In these cases, the interaction between mesh resolution and physical complexity may lead to persistent convergence difficulties, leaving traditional iterative strategies as a key computational bottleneck. Therefore, these challenges motivate the development of a new generation of simulation frameworks that can simultaneously maintain physical consistency and achieve high computational efficiency.

1.3. The Rise of PIML and the Review Pathway

PIML has recently attracted growing interest in computational science and engineering as a modeling paradigm that combines data-driven learning with physical principles. Its central idea is to incorporate established physical information, including governing equations, conservation constraints, and symmetry properties, into the design and optimization of machine learning models. By introducing such prior knowledge, PIML can reduce the dependence on large amounts of observational data while improving model generalization, physical consistency, and reliability in extrapolative predictions. As shown in Figure 2, physics-based prior information can be embedded in learning models in three main ways: at the data level by imposing physically constrained boundaries, at the training level by adding residual-based regularization terms to the loss function, and at the model level by encoding physical structures directly into the neural network architecture. For underground THM-coupled simulation, which represents a typical multiphysics problem, PIML provides a promising route to alleviate the limitations of conventional numerical methods in computational efficiency, parameter inversion, and multi-scenario optimization [18].

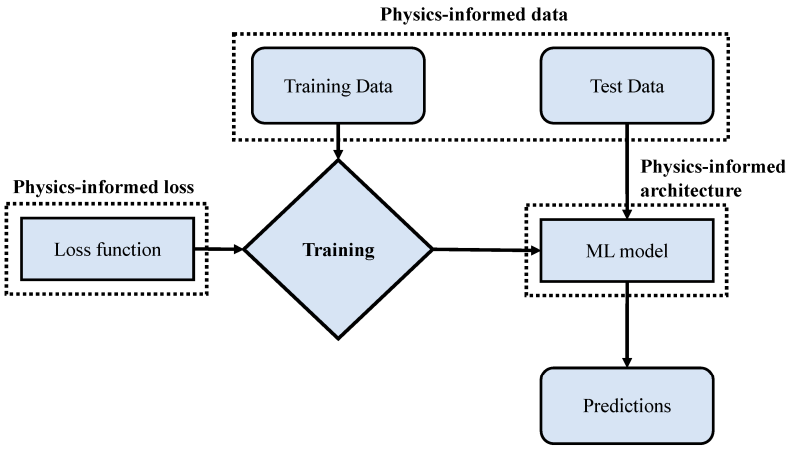


Figure 2. Representative approaches to integrating physics-domain knowledge within a supervised learning framework. Reprinted with permission from Ref. [19]. Copyright 2023 MDPI.

The origins of PIML are closely related to early attempts to combine neural network representations with differential equation solving in the 1990s. In 1998, Lagaris et al. [20] developed a general neural-network-based framework for solving ordinary and partial differential equations. In their formulation, the trial solution was represented by two components: one term designed to satisfy the boundary conditions and another term parameterized by a feedforward neural network. The network was then trained by reducing the residuals evaluated at collocation points [20]. Owing to the limited computational resources available at that time, this approach was not extensively developed. Later, in 2019, Raissi et al. [21] formally proposed PINNs by introducing PDE residuals into the loss function as physical constraints. This formulation provided a unified framework for forward and inverse problems and was applied successfully to representative cases in fluid mechanics, quantum mechanics, and reaction–diffusion systems [21], which greatly promoted the development of PIML. After that, PINNs were increasingly adopted in porous media flow and geoscience applications [22], while subsequent studies further extended this idea to geothermal system simulation and parameter estimation.

During the same period, operator learning theory also made important progress in learning mappings between function spaces. In 2021, Lu et al. [23] introduced the Deep Operator Network, termed DeepONet, which employs a two-part architecture composed of a branch network and a trunk network. This structure enables the approximation of nonlinear continuous operators and allows the model to learn the mapping from input parameter

functions to output solution functions [23]. Since then, neural operator architectures have continued to evolve, gradually improving their ability to handle high-dimensional parameter spaces and enabling faster solution prediction for complex systems. More recently, neural operators have been investigated in multiphysics coupling problems and have shown strong potential for complex-system modeling. Although neural operator models for fully coupled THM systems remain at an early stage, their successful use in simplified coupled settings, such as isothermal hydro-mechanical coupling and steady-state thermo-hydrological problems, has already demonstrated an important technical route for developing efficient underground flow simulators.

Among the existing PIML methods, PINNs and neural operators have attracted particular attention because they address THM simulations from complementary perspectives. In PINNs, the governing equations together with the initial and boundary conditions are incorporated into the training objective as physics-based constraint terms so that solving PDEs is reformulated as a mesh-free learning and optimization task. This formulation is especially suitable for inverse analysis, as it allows field variables and unknown system parameters to be inferred simultaneously from limited observational data. Therefore, PINNs can be interpreted as physics-constrained intelligent solvers.

By comparison, neural operators are designed to approximate operator mappings between input and output functions. For example, they can learn the relationship from permeability distributions and boundary conditions to solution fields such as pressure, temperature, and displacement. Although their training process generally depends on simulation datasets, a trained neural operator can generate predictions for new parameter settings within milliseconds. This feature makes neural operators well suited for rapid multi-scenario assessment and optimization, and they can therefore be regarded as surrogate models for simulation acceleration [24].

In addition to the two mainstream methods mentioned above, other technical branches have also emerged in the PIML field, such as physics-informed generative adversarial networks for uncertainty modeling, variational PINNs based on weak-form formulations, and operator learning methods with hard constraints. However, the application of these methods to THM coupling problems is still in the preliminary exploration stage, and their theoretical completeness and engineering practicality remain to be systematically validated. In contrast, PINNs and neural operators have accumulated a relatively rich body of case studies in underground energy engineering. With more explicit physics-embedding mechanisms, better reproducibility, and clearer engineering integration interfaces, they have become the two representative technologies currently driving the development of THM simulations within the PIML framework.

Although previous reviews have provided broad overviews of physics-informed deep learning and its applications in geoenergy development [25], they have generally emphasized the overall methodological development and representative applications of physics-informed approaches, rather than systematically examining the distinct roles of PINNs and neural operators in strongly coupled subsurface THM problems. In parallel, individual studies have demonstrated the potential of PINNs for physics-constrained forward and inverse modeling and of neural operators for accelerated forward prediction in porous-media and subsurface applications [22–24,26]. However, these two methodological routes are rarely discussed within a unified THM-oriented framework that explicitly considers their differences in physics embedding, data dependence, inverse-problem capability, computational efficiency, generalization, validation requirements, and engineering applicability. This gap is particularly important for THM systems, where strong nonlinear coupling, multi-scale behavior, heterogeneous geological properties, and sparse observations create challenges that are not fully reflected in single-field or weakly coupled problems.

This review (i) synthesizes the physical-embedding mechanisms and training strategies of PINNs and the operator learning mechanisms underlying neural operators; (ii) compares their data requirements, forward and inverse modeling capabilities, computational characteristics, and generalization limitations using evidence from representative THM and closely related subsurface studies; (iii) organizes reported engineering applications according to the specific computational demands of EGS, hydraulic fracturing, thermal recovery, and geological disposal; and (iv) identifies their complementary application boundaries and discusses potential hybrid PINN–neural-operator pathways for combining physical consistency with rapid multi-scenario prediction. By distinguishing established evidence from prospective integration strategies, the review aims to provide a more task-oriented basis for selecting and integrating PIML methods in subsurface THM applications.

2. The Methodological Core of PIML

2.1. PINNs: Mesh-Free Solvers with Embedded Physics

Underground THM coupling processes can be mathematically described by a group of strongly nonlinear partial differential equations. Automatic differentiation enables the efficient and accurate evaluation of the required partial derivatives of neural network outputs with respect to input coordinates, including derivatives of high order when needed. Based on these derivatives, the residual terms of the governing PDEs can be formulated directly, without relying on spatial mesh discretization. In this way, PINNs provide a mesh-free framework for solving physical field problems.

By combining feature engineering, neural network architecture design, and loss-function regularization, physical prior knowledge can be effectively incorporated into PINN models, as illustrated in Figure 3. The concrete implementation strategy is closely related to the nature of the target physical problem and the selected network structure. With domain-specific governing equations and physical principles embedded in the learning process, PINNs can produce predictions that offer both high accuracy and physical interpretability in complex engineering applications.

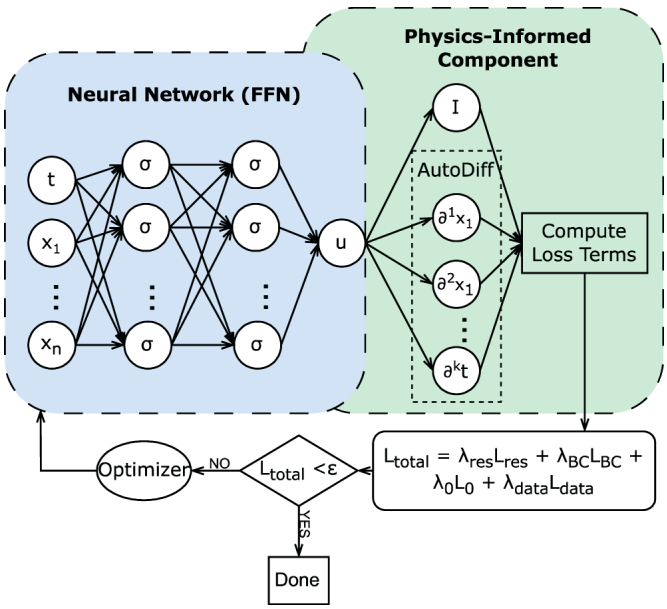


Figure 3. A typical architecture of a physics-informed neural network. Reprinted with permission from Ref. [27]. Copyright 2025 Springer.

In THM-coupled simulations, PINNs need to represent several strongly nonlinear physical fields at the same time. A common implementation is to use a multi-output neural network, or alternatively to construct several sub-networks that share latent features so that pore pressure $p(x,t)$, temperature $T(x,t)$, and displacement $u(x,t)$ can be predicted simultaneously. Joint physical constraints are then established by incorporating Biot's poroelasticity theory, Darcy–Fourier law, and the energy conservation equation into the learning framework. For example, in solving the classical hydro-mechanical Biot model, Cai et al. proposed a method that combines a multi-network structure with a fixed-stress split iterative scheme [28]. For more complex constitutive behaviors, including elastoplasticity, damage evolution, and thermo-mechanical–chemical coupling, the corresponding constitutive equations can be introduced into the loss function in the form of algebraic or differential constraints. When the Drucker–Prager yield criterion is used, for instance, the relation between the plastic strain increment and the stress state can be obtained through implicit integration algorithms in classical mechanics [29]. In addition, such constitutive information may also be embedded into an automatic-differentiation-based residual formulation or coupled with data-driven surrogate models for constitutive response prediction [30].

The architectural development of PINNs often focuses on improving the internal topology of the neural network. For instance, physics-informed nodes can be designed to incorporate governing equations into activation operations, while physics-informed layers can be used to enforce symmetry requirements and conservation laws during feature propagation. These network-level modifications provide a theoretical basis for improving the expressive capability of PINNs in multiphysics modeling. Nevertheless, THM-coupled problems introduce several difficulties for PINN training. First, the residual terms associated with different physical fields often differ significantly in both dimension and numerical scale, which may cause imbalance among the components of the loss function. As a result, the optimization process may be dominated by the major residual terms, making it difficult for secondary physical processes to converge properly [31]. Second, the strong nonlinearity caused by multiphase seepage and THM coupling, together with the large number of parameters involved in inversion, can intensify training oscillations and convergence problems [32]. Third, moving-boundary phenomena, such as saturation-front propagation in two-phase flow through porous media, require accurate prediction in regions that evolve dynamically. Fixed collocation point strategies often have limited ability to resolve such sharp front structures effectively [33,34].

2.2. Advanced Strategies for Enhancing the Robustness and Accuracy of PINNs

PINNs provide the possibility of mesh-free solutions for THM-coupled problems. Although the discontinuous boundaries caused by fracture propagation represent the greatest challenge in mechanical field simulations, variational/weak-form PINNs have been proven to effectively overcome the limitations of traditional methods in handling strongly singular problems. This solution strategy, based on the minimization of energy functionals, establishes a solid mathematical foundation for subsequent modeling of complex fracture evolution in multi-field coupling. Specifically, addressing the specific characteristics and complexities of THM systems, current research on model improvements mainly focuses on the following key strategies:

(1) Multi-scale Modeling and Multi-task Learning

THM field variables (e.g., pressure, temperature, and displacement) often exhibit significant scale differences in spatial gradients and temporal evolution (for instance, stress concentration at crack tips versus the slow variation in temperature at the regional scale). Traditional single-network architectures struggle to effectively capture such multi-scale features simultaneously. To address this, a multi-task learning architecture has

been introduced: it employs a shared underlying feature extractor and utilizes branch networks to output different physical fields separately, while designing independent loss weights for each task to achieve decoupled learning of physical processes at different scales. Furthermore, explicitly incorporating Fourier feature embeddings or wavelet basis functions into the input can effectively enhance the network's representation capability for high-frequency spatial variations, thereby significantly improving the resolution accuracy in high-gradient regions such as fractures and interfaces.

(2) Adaptive Collocation Point Sampling and Dynamic Loss Weighting

Fixed collocation point sampling strategies often lead to biased residual estimation when handling moving fronts (such as the steam chamber front in thermal recovery) or local high-gradient regions, thereby causing training instability. Adaptive sampling methods address this by iteratively identifying current high-residual regions (e.g., based on residual magnitudes or their gradient information) and dynamically increasing the density of collocation points in these areas, thus concentrating the network's optimization resources on sub-regions with complex physical processes [35]. On the other hand, due to the inherent differences in dimensions and numerical magnitudes among various governing equations, the weights of different terms in the loss function are prone to imbalance. Adopting adaptive weighting strategies (such as dynamic adjustments based on gradient statistics or uncertainty estimation) or primal–dual optimization frameworks can automatically balance multiple physical constraints, effectively preventing the training process from being overly dominated by a single governing equation [36].

To improve reproducibility in THM applications, the residuals of mass, momentum, and energy conservation should first be nondimensionalized and evaluated using consistent batch-wise statistics. Their weights can then be dynamically updated according to the relative residual magnitudes or gradient norms, with moving-average smoothing used to avoid abrupt oscillations. Gradient conflicts between different physical objectives can be detected through the cosine similarity of their gradients. When conflicting gradients occur, gradient-projection or multi-objective optimization strategies may be adopted to seek a Pareto-balanced solution rather than allowing a single physical field to dominate the training process. This procedure provides a practical workflow for balancing multiple THM constraints, although the optimal weighting rule remains problem-dependent.

(3) Variational and Weak-Form-Based PINNs

For THM problems involving solution discontinuities (such as displacement jumps across fracture surfaces or abrupt changes in material properties), the strong-form PDE residuals relied upon by classical PINNs require the solution to possess high-order differentiability, which is often physically unattainable [37]. Variational PINNs or weak-form PINNs transform the governing equations into their integral forms (e.g., the Galerkin weak form), requiring only first-order differentiability of the solution function. This significantly enhances their numerical adaptability to discontinuous interfaces and singularities. When simulating the evolution of natural fracture networks or hydraulic fracturing processes, this approach has demonstrated superior numerical stability compared to the strong-form formulation.

(4) Physics-Guided Network Architecture Design

Directly encoding prior physical knowledge (such as symmetries, conservation laws, and scale invariance) into the architecture of neural networks can substantially reduce the hypothesis space of the solution and guide the model to converge more rapidly toward physically plausible solutions. For instance, when dealing with physical fields that exhibit symmetric characteristics, symmetry group constraints can be directly embedded into

the network architecture [38]. This concept can also be applied when simulating axisymmetric EGS by adopting cylindrical coordinates as inputs and incorporating rotational symmetry constraints. In problems dominated by advection–diffusion or solute transport, replacing traditional tanh activation functions with periodic activation functions can more efficiently represent smooth solutions with oscillatory characteristics [39]. Furthermore, for conserved quantities that must be strictly satisfied (such as total mass or energy), specific conservation layers can be designed at the output end of the network to guarantee integral conservation via hard constraints. Such "physics-embedded" structural designs reduce the model's reliance on soft constraints in the loss function, fundamentally reinforcing physical consistency from the source.

Taken together, the aforementioned strategies work synergistically across four dimensions—representation capability, sampling mechanisms, mathematical formulations, and network architectures—to collectively enhance the robustness and practicality of PINNs in complex THM systems. However, the effective combination of these strategies currently still relies heavily on expert knowledge and manual hyperparameter tuning. There is an urgent need in the future to develop automated, adaptive, and unified training frameworks to lower the barrier to application and enable the broader, more reliable deployment of PINNs across a wide range of complex underground engineering scenarios.

2.3. NOs: Accelerating Engines for THM Forward Modeling

To address the efficiency limitations of conventional numerical simulators in tasks such as multi-scenario optimization, uncertainty quantification, and real-time decision support, neural operators have attracted increasing attention in recent years as a class of methods for learning the solution operators of partial differential equations [40]. In subsurface THM modeling, they have become an important research direction for accelerating forward simulations. Their main goal is to establish a parameterized operator mapping:

$$\mathcal{G}_\theta : \underbrace{\alpha(x)}_{\text{Input parameter field}} \mapsto \underbrace{\mu(x,t)}_{\text{Output physical field}} \quad (1)$$

where the input may include permeability distributions, thermal conductivity fields, injection settings, or boundary constraints, while the output corresponds to full-field physical responses, such as pressure, temperature, and displacement. After training, this operator can exploit the parallel computing capability of modern GPUs, including tensor cores, to generate high-fidelity predictions for unseen parameter configurations within milliseconds. As a result, the computational expense can be reduced by several orders of magnitude.

Compared with conventional surrogate models, including fully connected neural networks and convolutional neural networks, neural operators do not simply learn mappings between finite-dimensional vectors defined on fixed grids. Instead, they approximate nonlinear mappings between input and output functions in infinite-dimensional function spaces [41]. This property gives neural operators mesh-independent behavior and stronger generalization ability. Once trained, they can be evaluated on different discretization schemes, geometric settings, or unstructured parameter fields. More specifically, because the learned relationship is formulated at the level of continuous function spaces, an operator trained on one fracture-network configuration may be extended to new simulation cases with different well layouts or previously unseen fracture morphologies, indicating considerable potential for engineering applications.

Among the major neural operator architectures, the Fourier neural operator (FNO) and DeepONet are two representative approaches.

(1) The Fourier neural operator (FNO) parameterizes the integral kernel in the Fourier domain via a stack of Fourier layers, enabling the model to capture global dependencies between input parameters and solution fields (Figure 4). Its architecture comprises a lifting layer (P) that projects the input function to a high-dimensional latent space, followed by iterative Fourier layers that perform spectral transformations ($F \rightarrow R \rightarrow F^{-1}$) combined with local linear operators (W), and a final projection layer (Q) that maps the latent representation back to the target output. Standard FNO is generally effective for elliptic and parabolic problems with relatively smooth solutions on regular grids. In THM simulations, it has been used to predict steady-state or quasi-steady-state pressure and temperature fields, including thermal breakthrough processes in EGS. This method has shown order-of-magnitude acceleration when solving parameterized PDE problems.

(2) DeepONet adopts a two-network structure composed of a branch net and a trunk net. The branch net is responsible for encoding the input function, whereas the trunk net takes the spatiotemporal coordinates as input. Through this design, DeepONet directly learns function-to-function mappings. Compared with architectures that rely more strongly on smooth field assumptions, DeepONet shows good adaptability to non-smooth solutions, such as fracture interfaces, as well as to complicated boundary conditions. Therefore, it is particularly useful for uncertainty quantification in highly heterogeneous reservoirs.

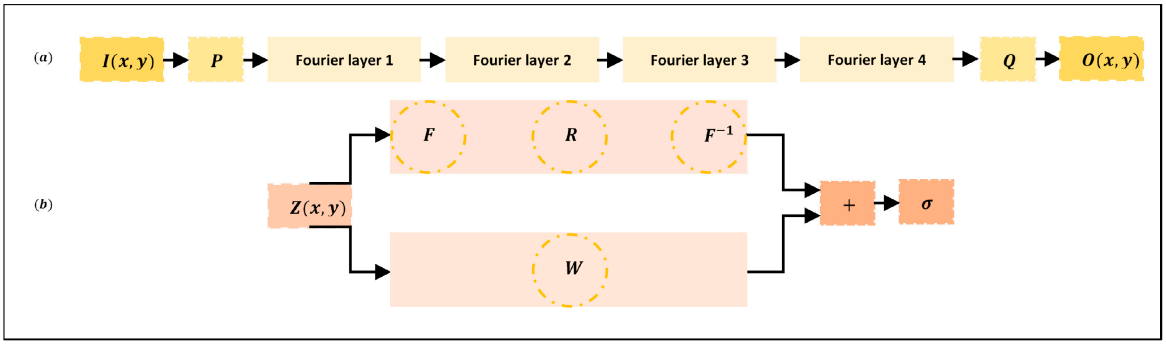


Figure 4. Architecture of the Fourier neural operator (FNO) (**a**) Overall architecture; (**b**) Fourier layer structure. Reprinted with permission from Ref. [42]. Copyright 2023 MDPI.

In the practice of underground energy engineering, neural operators have demonstrated remarkable acceleration effects while maintaining an accuracy comparable to that of high-fidelity numerical simulators [43]. It is worth noting that this speedup stems not only from the paradigm shift in neural operators—from point-by-point solving to direct mapping—but also from the hardware advantages of high-concurrency matrix operations on GPUs. This advancement makes previously computationally infeasible large-scale tasks viable, such as tens of thousands of Monte Carlo simulations, real-time optimization of fracturing schemes, and decades-long dynamic predictions for thermal recovery.

The training of neural operators relies on datasets consisting of input–output function pairs, which are primarily derived from the following sources:

(1) Synthetic data generated by high-fidelity numerical simulators, which are used to systematically explore the parameter space.

(2) Field observational data (such as distributed temperature sensing, InSAR deformation monitoring, and production history), which can be used to construct data assimilation-enhanced models.

(3) Physics-consistent solutions generated by PINNs—in scenarios where observational data are sparse or traditional simulations are prohibitively expensive, PINNs can

provide low-cost, highly physics-faithful "soft labels". These serve as a supplementary source of high-quality data for neural operator training, thereby establishing a synergistic paradigm of "physics-generated, data-driven acceleration".

Because the reviewed studies involve heterogeneous datasets and rarely report a standardized mixing ratio among these data sources, a universal proportion cannot currently be recommended. In practical data fusion, each source should first be evaluated against independent high-fidelity simulations or field observations. PINN-generated soft labels should be checked using both prediction discrepancies and governing-equation residuals. When different data sources conflict, greater weight should be assigned to samples with lower validation errors, lower predictive uncertainty, and stronger physical consistency. Samples satisfying application-specific tolerance thresholds may be retained, moderately inconsistent samples may be down-weighted, and samples exceeding the accepted tolerance should be excluded. These thresholds should be calibrated using an independent validation set and linked to the acceptable engineering error. Such confidence-aware screening and weighting can reduce the propagation and accumulation of systematic bias from PINN-generated labels during neural operator training.

Despite their significant advantages, the performance of neural operators remains constrained by the coverage of their training data. When input parameters fall outside the training distribution (e.g., under extreme physical property conditions), their extrapolation reliability may deteriorate sharply. Furthermore, for strongly nonlinear processes (such as multiphase flow and phase transitions), existing architectures still struggle to accurately capture abrupt structural changes in the solutions. Therefore, effectively integrating physical constraints to enhance extrapolation robustness remains a primary focus of current research. Recent emerging works, such as physics-constrained FNO and DeepONet with embedded conservation laws, attempt to incorporate THM governing equations as regularization terms into the loss functions, preliminarily demonstrating their potential to improve generalization capabilities.

2.4. Methodological Comparison and Integration Pathways

Although PINNs and neural operators are both important branches of PIML, they differ substantially in modeling philosophy, data requirements, computational behavior, and applicable problem settings [44]. Understanding these differences is essential for selecting appropriate technical routes in engineering applications and for improving the practical effectiveness of PIML methods. Therefore, this section compares the two approaches across several major aspects and further discusses possible strategies for their coordinated use.

PINNs are mainly developed for solving specific physical problems. They approximate the target solution by optimizing neural network parameters under the joint constraints of governing-equation residuals and available data. In this sense, PINNs can be viewed as mesh-free physics-based solvers. One of their main strengths is that they can address forward problems without relying on pre-generated training datasets. At the same time, they are naturally compatible with inverse tasks, including parameter-field reconstruction and source-term identification. This feature makes PINNs especially useful when observational data are limited, but the governing physical mechanisms are clearly defined. In addition, PINNs can flexibly incorporate different types of physical knowledge, such as partial differential equations, boundary conditions, failure-related physical constraints, and statistical information. Nevertheless, because PINNs are usually trained for individual problem settings, changes in boundary conditions, material parameters, or other configurations often require model retraining. As a result, their computational expense becomes a major limitation when a large number of scenarios need to be simulated rapidly.

By contrast, neural operators are intended to construct more general surrogate models by learning operator-level relationships between input and output functions. For example, they can establish mappings from permeability fields to pressure fields, thereby following a "train once, evaluate repeatedly" strategy. Their most prominent advantage is the ability to perform extremely fast inference, often reaching millisecond-level prediction speed. This makes neural operators highly attractive for applications involving large-scale forward evaluations, such as optimization, uncertainty quantification, and real-time prediction [45]. However, their accuracy strongly depends on the quality and representativeness of the training dataset. When input conditions fall outside the distribution covered during training, their extrapolation performance may be limited. Moreover, applying neural operators to inverse problems usually requires additional optimization procedures or Bayesian inference frameworks, which increases both implementation difficulty and computational cost compared with PINNs.

In inverse-PINN applications, the quality and distribution of observational data directly affect parameter identifiability and inversion robustness. Increased measurement noise and reduced spatial or temporal sampling generally increase the uncertainty of inferred permeability, thermal conductivity, and other physical parameters. Interpolation may smooth local gradients, whereas excessive downsampling may remove information associated with fractures, interfaces, or transient responses, thereby introducing bias into the inversion results. However, noise levels, spatial sampling densities, temporal resolutions, and preprocessing procedures were not consistently reported in the reviewed studies. The differences summarized in Table 1 should be interpreted as general tendencies rather than absolute characteristics, as their relative performance depends on the specific problem, data availability, model configuration, and computational conditions.

Table 1. Methodological comparison of PINNs and NOs in THM-coupled simulations.

Dimensions	**PINNs**	**NOs**
Modeling Paradigm	Physics-driven solver (single-scenario optimization)	Data-driven surrogate model (generalized mapping)
Training Data Requirements	Low; data mainly for inversion/enhancement	Moderate–high; depends on training coverage
Inverse-Problem Capability	Directly supported	Usually requires additional optimization
Solving Time for New Scenarios	Relatively high	Low after training
Extrapolation Capability	Physics-constrained but problem-dependent	Limited by training distribution
Typical THM Application Scenarios	Fracture parameter inversion (EGS), thermal–mechanical field reconstruction under sparse observations, moving-boundary problems (steam cavity)	Multi-scheme fracturing optimization, long-term geothermal production prediction, uncertainty propagation analysis
Sensitivity to Data Quality	In inverse problems, parameter identifiability and robustness are sensitive to observational noise, spatial and temporal sampling, and data preprocessing.	Prediction reliability depends on the quality, coverage, and consistency of the training data.

Based on the reviewed evidence, the following qualitative method-selection guidance can be summarized:

(1) Prioritize PINNs: When the problem is inversion-dominated, observational data is sparse, and the physical mechanism is clear. For example, using limited downhole temperature and pressure data to invert for reservoir thermal conductivity and permeability distributions.
(2) Prioritize NOs: When the task is forward-modeling-intensive, requires high-speed inference, and the parameter space is known and controllable. For example, in enhanced geothermal system design, rapidly evaluating long-term heat extraction efficiency under thousands of injection-production well pattern scenarios.
(3) Consider a fusion strategy: When the engineering problem involves both high-dimensional inversion and multi-scenario optimization requirements, or when it is necessary to simultaneously balance physical consistency and extremely high inference efficiency, the synergistic use of both methods should be considered.

Integrating PINNs and neural operators to leverage their complementary strengths represents a promising direction in PIML. The three conceptual pathways considered here are as follows:

(1) Data generation synergy: Leveraging PINNs to generate physically consistent synthetic datasets under data-free or sparse-data conditions, thereby training more robust neural operator surrogate models with enhanced generalization capabilities. This effectively overcomes the bottleneck of scarce high-quality simulation data.
(2) Solution acceleration synergy: A trained neural operator may provide a warm-start solution or a pre-trained representation for subsequent PINN optimization. A typical workflow consists of neural operator inference, initialization of the PINN solution field or network parameters, physics-constrained refinement using governing-equation and boundary-condition residuals, and final residual-based validation. The total computational cost includes the one-time training cost of the neural operator, its relatively low inference cost, and the remaining PINN refinement cost. The actual acceleration depends on the discrepancy between the neural operator prediction and the physically admissible solution. A sufficiently accurate initial prediction may reduce the number of PINN optimization iterations, whereas a poor initial prediction may provide little benefit or even bias the optimization process. Therefore, this strategy should currently be regarded as a promising conceptual pathway whose effectiveness still requires systematic validation on strongly coupled THM benchmark problems.
(3) Architecture fusion: A further strategy is to develop unified hybrid architectures by integrating physical residual constraints into neural operator frameworks. For example, the residual terms associated with governing physical laws can be embedded into the branch network or trunk network of DeepONet, leading to a physics-informed DeepONet, namely PI-DeepONet. As shown in Figure 5, this framework represents a general form of deep integration between neural operator learning and physical constraints, where geometric differential equations are used as an illustrative example of the constraint formulation. For THM applications, such physical constraints could be formulated using the residuals of the governing THM equations. With the aid of regularization, the predicted outputs are guided to satisfy the physical laws associated with THM multi-field coupling. In addition, fast prediction results generated by FNO can also be supplied to PINNs as initial or auxiliary inputs for local refinement, which helps improve simulation accuracy in key regions, such as fracture tips.

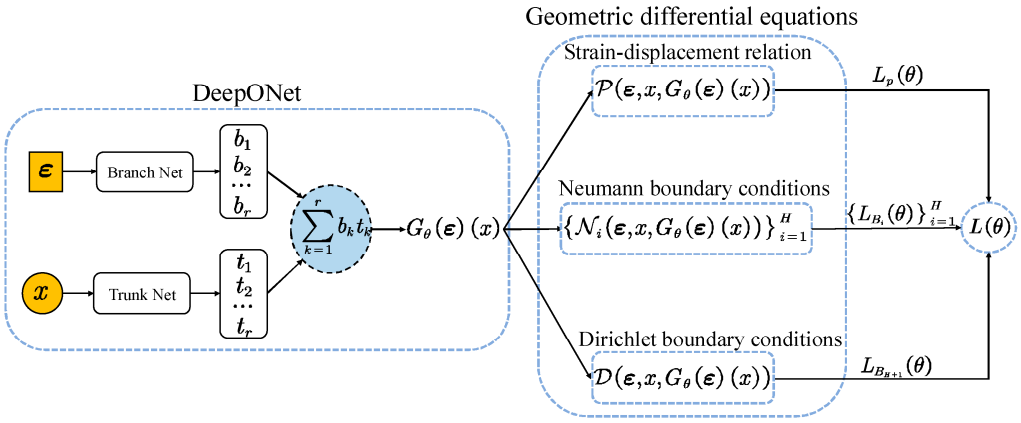


Figure 5. Physics-informed DeepONet architecture regularized by geometric differential equations, with geometric equations serving as representative physical constraints. Reprinted with permission from Ref. [46]. Copyright 2024 MDPI.

3. Application Scenarios: PIML in Earth Science Resource Development

3.1. EGS: Heat Extraction and Long-Term Management

EGS projects generally rely on techniques such as hydraulic fracturing to create artificial fracture networks in low-permeability hot, dry rock formations, thereby making the commercial development of deep geothermal resources possible. Their successful operation depends strongly on accurate simulation of THM-coupled processes, since these processes control several key engineering indicators, including the temporal evolution of fracture permeability, thermal breakthrough time, well-placement optimization, and long-term heat recovery efficiency [47]. Nevertheless, workflows based on conventional numerical simulations still face notable limitations. For reservoir parameter inversion, quantities such as spatially distributed permeability and thermal conductivity are usually estimated through computationally intensive iterative optimization, and a single inversion process may require several weeks. Such a time scale is difficult to match with the needs of field-scale decision-making. In addition, assessing different injection-production schemes often requires hundreds or even thousands of forward simulations. Traditional solvers are therefore unable to efficiently support real-time optimization, which forces engineering design to rely on more conservative strategies and may prevent actual heat recovery performance from approaching its theoretical potential.

PIML provides a potentially useful complementary approach for addressing these challenges. For parameter inversion, PINNs can improve the efficiency of EGS reservoir characterization by taking advantage of their mesh-free formulation, embedded physical constraints, and adaptability to sparse observational data. Amini et al. investigated the application of PINNs to THM-coupled processes in porous media, demonstrating their potential for complex multiphysics modeling [32]. Unlike methods that require predefined parameterization forms, this approach can directly predict continuous physical fields, which is valuable for high-resolution geological modeling. More specifically, by jointly representing steady-state limits and transient evolution, the model can make use of the connections among physical laws across different temporal scales. This enables more reliable prediction of critical engineering indicators under data-scarce conditions.

In the areas of long-term performance prediction and operational optimization, neural operators serve as highly efficient surrogate models, enabling multi-scenario evaluations at the second level. Although current geothermal surrogate models are predominantly based on convolutional neural networks or Gaussian processes and have utilized tens of thousands of TOUGH2 simulations to predict outlet temperatures, integrating highly efficient neural operators (such as FNO) with fully coupled THM simulations (e.g., TOUGH2–

FLAC3D) to support decade-scale dynamic predictions remains a critical direction for exploration [48–50]. It should be emphasized that most current PINN-based THM studies relevant to EGS remain at the benchmark or numerical-validation stage. Direct evidence of routine industrial-scale deployment in operational EGS projects is still limited. Therefore, the application of PINNs to field-scale EGS should presently be regarded as a promising research direction rather than an established industrial practice.

3.2. Unconventional Oil and Gas Development: From Reservoir Stimulation to Production Optimization

The economic and efficient development of unconventional oil and gas resources depends on accurate characterization and effective control of the complex fracture networks created by hydraulic fracturing. However, several key reservoir properties after fracturing, including the equivalent permeability field, the activation state of natural fractures, and the time-dependent evolution of mechanical parameters, are difficult to measure directly. This limitation makes the conventional "simulation–inversion–optimization" workflow inefficient in practical applications [51].

For example, in the fracturing development of unconventional reservoirs, traditional inversion approaches often combine discrete fracture models (DFM) or embedded discrete fracture models (EDFM) with geomechanical simulators, which leads to considerable computational expense. DFM can describe near-fracture fluid flow with high accuracy, but it usually requires extensive meshing and repeated remeshing whenever the fracture configuration or model parameters are updated. EDFM improves the computational efficiency to some extent, yet it still involves substantial computational overhead when applied to complex systems containing multiple fractures [52]. Similarly, the optimization of net present value for coordinated multi-well fracturing schemes is often simplified into heuristic decision rules, because a single numerical simulation may take several hours. As a result, the achievable economic benefits can be significantly reduced.

To improve efficiency, traditional purely data-driven surrogate models have been widely adopted. These methods establish statistical mapping relationships between fracturing parameters and production or NPV by training on thousands to tens of thousands of numerical simulation datasets. Although they enable rapid predictions, their purely data-driven, black-box nature results in weak physical consistency and limited extrapolation capabilities. Particularly when confronted with geological patterns or extreme operating conditions not covered in the training set, these models are prone to generating physically implausible interpretations, which restricts their credibility and practical value in high-risk investment decision-making.

To overcome these technical limitations, PIML provides an effective route for improving both the intelligence and physical credibility of reservoir characterization and production optimization in unconventional oil and gas development. This is achieved by combining physics-based prior knowledge with data-driven modeling strategies.

For reservoir characterization, the PINN-based multi-source data fusion framework offers a promising basis for high-resolution inversion of post-fracturing reservoir properties [53]. By integrating different types of field observations, including microseismic data, pressure drawdown records, and InSAR measurements, it becomes possible to obtain a more detailed description of reservoir responses after hydraulic fracturing. To achieve high-resolution characterization of the fracture-induced stimulated reservoir volume (SRV), physics-informed neural network models can be deployed to capture coupled THM processes. These models aim to jointly use microseismic events, early-time pressure drawdown data, and surface InSAR deformation observations to infer the fracture-induced anisotropic permeability tensor field and localized zones of mechanical-parameter degradation.

Furthermore, to address the severe shortage of high-resolution field data, multi-fidelity PINNs have been introduced as a feasible methodological framework [54]. By combining low-cost analytical solutions with information from high-fidelity numerical simulations, this type of model can robustly estimate the permeability enhancement factor near the wellbore even when observational data are extremely limited. Therefore, it can provide important support for evaluating the effectiveness of hydraulic fracturing and reservoir stimulation.

At the production optimization level, surrogate models based on neural operators are gradually replacing traditional numerical simulators to undertake the tasks of rapid forward modeling and scheme optimization. Researchers have attempted to utilize neural operators to construct ultra-fast reservoir simulation surrogate models that account for multi-well interference and complex permeability fields [49], demonstrating the immense potential of operator learning for integration into on-site oilfield digital-twin platforms. To enhance the predictive capability for multiphase transient flow in fractured reservoirs post-hydraulic fracturing, researchers are exploring the extension of deep operator learning methods (such as FNO) to multi-stage dynamic modeling. By explicitly characterizing the physical evolution across the fracturing, flowback, and production stages, this approach provides data-driven support for real-time field analysis and rapid decision-making [26]. Furthermore, to support robust fracturing design under risk constraints, researchers are investigating the embedding of uncertainty quantification modules into DeepONet to output the probability distribution of Estimated Ultimate Recovery (EUR), thereby enabling the quantitative assessment of reserve prediction reliability.

3.3. PIML Practices in Diverse THM Scenarios

Beyond EGS and unconventional oil and gas development, PIML also demonstrates significant application potential in diverse THM-coupled scenarios, including CBM extraction, heavy oil thermal recovery, CO_2 geological storage, and high-level radioactive waste geological disposal. These advancements are driving the evolution of traditional numerical simulation paradigms toward more efficient and intelligent directions [25].

In the context of CBM extraction, research efforts have been dedicated to constructing numerical models that couple THM processes. By integrating the Langmuir adsorption effect and Biot's poroelasticity theory, these studies have systematically revealed the controlling mechanisms of multi-field coupling on permeability evolution during CBM production in the Qinshui Basin [55].

In heavy oil thermal recovery processes (such as Steam-Assisted Gravity Drainage, SAGD), the expansion of the steam chamber, phase-change heat transfer, and the non-Newtonian flow of heavy oil constitute a highly nonlinear thermo-hydro coupled system. Traditional numerical simulations require the fine-scale characterization of the steam–oil–water three-phase interface, resulting in extremely high computational costs. Cutting-edge research has successfully introduced FNO into multiphase flow modeling for CO_2 storage [56].

In the field of geological CO_2 storage, THM-coupled processes directly govern caprock integrity, CO_2 plume migration, and long-term storage security. Traditional simulation methods struggle to efficiently evaluate thousands of monitoring, verification, and injection scenarios. To achieve rapid joint assessment of storage capacity and induced fracturing risk in geological CO_2 storage, researchers are exploring the use of DeepONet to learn the end-to-end mapping from injection history and geological models to pressure, saturation, and strain fields, thereby enabling multiphysics prediction and risk quantification to be completed within seconds [57].

In the safety-sensitive context of high-level radioactive waste geological disposal, THM processes play a decisive role in the swelling behavior of buffer materials, stress redistribution within the host rock, and radionuclide migration pathways. Because repository experiments usually involve extremely long time scales and in situ monitoring data are rarely available, conventional simulation studies often depend strongly on theoretical assumptions. Neural-network-based methods can be applied to analyze fully coupled THMC responses of bentonite barriers in repository environments, thereby supporting sensitivity analysis of key failure modes [58]. At the same time, neural operators provide a rapid tool for evaluating how different disposal pit arrangements influence peak thermal and mechanical responses, which can assist the optimization of conceptual repository designs.

To improve the long-term safety assessment of high-level radioactive waste repositories, researchers have also begun to incorporate efficient neural operators into rapid risk-assessment frameworks for radionuclide transport. By quickly estimating plume diffusion extents and dose distributions under different disposal scenarios, these methods can support emergency response, disposal scheme comparison, and repository optimization [59]. Nevertheless, this application field requires exceptionally high model credibility and reliability. Existing PIML approaches still lack rigorous error-bound estimation and probabilistic failure guarantees, meaning that a considerable gap remains before they can satisfy strict engineering certification requirements [60].

Across these application scenarios, PINNs and neural operators exhibit different advantages and limitations. PINNs are more suitable for parameter inversion and field reconstruction when observations are sparse and governing equations are available, but they generally require problem-specific training and may face convergence difficulties in strongly nonlinear THM systems. Neural operators are more efficient for repeated forward predictions, multi-scenario evaluation, and operational optimization, but their performance depends strongly on the coverage and quality of the training data, and their reliability may decrease for out-of-distribution conditions. Therefore, the selection of a PIML approach should consider the target task, data availability, computational requirements, and acceptable uncertainty. Hybrid strategies may combine the physical consistency of PINNs with the rapid inference capability of neural operators, although their implementation complexity and error propagation still require further investigation.

4. Discussion: Key Challenges, Limitations, and Future Prospects

Some limitations discussed here are general to physics-informed learning, including optimization instability, sensitivity to hyperparameters, dependence on training-data coverage, and uncertainty in out-of-distribution predictions. In THM applications, however, these general limitations are further intensified by problem-specific characteristics, including strong nonlinear coupling among thermal, hydraulic, and mechanical fields, large differences in physical scales and units, constitutive nonlinearities, moving interfaces or fractures, and highly heterogeneous geological properties. These THM-specific features make loss balancing, convergence, and reliable generalization particularly challenging.

PIML can reduce computational costs and support forward prediction and inverse analysis in subsurface THM problems. However, its performance remains strongly dependent on problem complexity, data availability, and validation conditions. The reviewed evidence indicates that PIML has demonstrated clear potential for accelerating forward simulation and supporting inverse analysis in subsurface multiphysics problems. However, the current evidence base remains fragmented and is dominated by benchmark problems, numerical experiments, and simplified coupling conditions. Fully coupled, strongly nonlinear THM applications with systematic field-scale validation remain comparatively limited.

In addition, differences in governing equations, datasets, error metrics, hardware platforms, and validation protocols make direct quantitative comparison across studies difficult. Therefore, the reported performance should be interpreted as problem-specific evidence rather than as a universal indication of readiness for THM engineering deployment.

Despite the significant efficiency advantages demonstrated by PIML in subsurface THM-coupled simulations, its large-scale deployment in engineering scenarios with high risks and stringent reliability requirements still faces three fundamental challenges. Based on these evidence gaps, three limitations are particularly important for future THM applications.

First, optimization convergence and training stability remain major factors limiting the practical use of PINNs in complex THM systems. This difficulty is closely associated with the strong nonlinearity and multi-scale characteristics of the THM governing equations, and it is mainly reflected in the following aspects:

(1) Loss-scale imbalance: The residuals associated with mass, momentum, and energy conservation often differ considerably in their physical units and numerical ranges. When constant loss weights are adopted, the training process may be biased toward the dominant governing equations. As a consequence, secondary but still essential physical mechanisms, such as thermo-mechanical coupling effects, may receive insufficient optimization attention and fail to converge effectively.

(2) Instability of high-order derivative gradients: High-order spatiotemporal derivatives (e.g., $\nabla^2 p$, $\frac{\partial^2 \mu}{\partial t^2}$) computed via automatic differentiation are prone to vanishing or exploding gradients in the presence of strongly nonlinear constitutive models (such as elastoplasticity or phase transitions) or during long-term integration. This makes it difficult for the model to simultaneously and accurately capture both macroscopic field evolution and local singular behaviors (such as stress concentration at crack tips).

(3) Non-convex optimization traps: The loss landscape contains numerous local minima. When network initialization is suboptimal or the collocation points are poorly sampled, optimization algorithms are highly prone to converging to physically unreasonable spurious solutions. Currently, there is a lack of universally effective mechanisms to automatically identify and escape these spurious solutions [61]. Although advanced optimizers such as L-BFGS and primal–dual algorithms can mitigate convergence difficulties to some extent, a robust training paradigm applicable to various THM problems has not yet been established.

Although Bayesian neural networks and related probabilistic methods can distinguish aleatoric uncertainty, which originates from inherent data or parameter variability, from epistemic uncertainty, which is associated with limited training data and model inadequacy, extending rigorous uncertainty quantification to high-dimensional and strongly coupled THM systems remains a formidable challenge. The generalization capability of neural operators is strongly constrained by the distribution range of their training data. In THM applications, once input parameters exceed the training domain, models may produce physically unreasonable solutions that are difficult to detect in a timely manner. This limitation is particularly critical in complex subsurface systems, where out-of-distribution predictions may violate physical laws without exhibiting obvious anomalous features [62]. Although Bayesian PINNs and deep ensembles can provide probabilistic predictions, their computational costs and calibration performance remain problematic in high-dimensional THM systems. Consequently, a robust uncertainty quantification framework suitable for high-risk subsurface engineering applications has not yet been established. For engineering deployment, a practical runtime warning mechanism may combine normalized physical residuals with the predictive variance obtained from an ensemble of neural operators. Thresholds calibrated using validation cases and application-specific tolerances could classify predictions as acceptable, uncertain, or requiring recalculation by a PINN or a

conventional high-fidelity solver. Although this strategy does not provide a rigorous upper bound on the generalization error, it offers a feasible means of identifying potentially unreliable out-of-distribution predictions.

Finally, the lack of rigorous mathematical theoretical guarantees represents a fundamental bottleneck for PIML to advance toward industrial-grade applications and certification. Currently, neither PINNs nor neural operators possess strict mathematical theories regarding the existence, uniqueness, stability, and convergence rates of their solutions. Particularly for strongly coupled, nonlinear, and non-smooth THM systems, several critical theoretical questions remain unanswered: Under what conditions can PINNs guarantee approximation to the true solution? How do network architecture capacity and collocation point density quantitatively affect error upper bounds? Can the generalization error of surrogate models be explicitly correlated with physical residuals? In contrast, traditional numerical methods (such as the finite element method) boast well-established a priori and a posteriori error estimation frameworks and stability analysis theories (e.g., the LBB condition for mixed formulations of multiphysics problems) [63]. This weak theoretical foundation not only hinders the scientific soundness and effectiveness of algorithm design but also makes it difficult for regulatory bodies and engineering practitioners to establish sufficient trust in the computational results of PIML.

Overall, although PIML has achieved breakthrough progress in computational efficiency and modeling flexibility, it still needs to make systematic leaps in algorithm robustness, uncertainty quantification, and mathematical theoretical foundations to truly evolve into a reliable, trustworthy, and certifiable engineering simulation tool. This current landscape provides clear guidance for the key future research directions outlined below.

Future research should focus on extending PIML to fully coupled THMC systems, developing hybrid PINN–neural-operator frameworks, improving uncertainty quantification and interpretability, and enhancing the representation of strongly heterogeneous and multi-scale geological structures. The integration of PIML with observational data, conventional simulators, and digital-twin systems may further improve its reliability and practical value in subsurface engineering. However, these developments require standardized benchmarks, rigorous validation, and reliable error-control mechanisms before large-scale engineering deployment.

5. Conclusions

This review examined the methodological characteristics and subsurface applications of physics-informed neural networks (PINNs) and neural operators (NOs) for thermo-hydro-mechanical (THM) modeling. The two approaches address complementary computational needs: PINNs are well suited to physics-constrained field reconstruction and parameter inversion under sparse observational conditions, whereas NOs excel in repeated forward simulations, multi-scenario evaluation, and near-real-time prediction. Neither method is universally superior; the choice should be guided by the target task, data availability, computational requirements, and acceptable uncertainty level.

Despite recent progress, several key gaps limit the broader deployment of PIML in subsurface THM problems. First, most reported applications address simplified or weakly coupled conditions; fully coupled, strongly nonlinear THM models with systematic field-scale validation remain scarce. Second, standardized benchmark problems, consistent evaluation protocols, and rigorous uncertainty quantification frameworks are still lacking. Third, the mathematical foundations concerning solution stability, convergence guarantees, and error bounds have not reached the maturity of conventional numerical methods. These limitations indicate that PIML should currently be regarded as a complementary acceleration and inference tool for high-fidelity simulation rather than a complete replacement.

Future research should prioritize field-scale validation of fully coupled THM models, the establishment of standardized benchmarks and evaluation protocols, robust uncertainty and error-control mechanisms, and hybrid PINN–neural-operator frameworks that combine physical consistency with efficient repeated prediction. Closer integration of physical modeling, numerical simulation, observational data, and learning-based methods will be essential to advance PIML toward reliable engineering deployment.

Author Contributions: L.W.: Writing—original draft, data curation. F.X.: Conceptualization, writing—review and editing. F.D.: Investigation. L.Z.: Writing—review and editing. Y.X.: Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Basic Research Program of Shaanxi (No. 2025JC-YBQN-742) and the Special Scientific Research Program of the Education Department of Shaanxi Provincial Government (No. 25JK0577).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

BFGS	Broyden–Fletcher–Goldfarb–Shanno
CBM	Coalbed Methane
DeepONet	Deep Operator Network
DFM	Discrete Fracture Model
EDFM	Embedded Discrete Fracture Model
EGS	Enhanced Geothermal System
FNO	Fourier Neural Operator
GPU	Graphics Processing Unit
InSAR	Interferometric Synthetic Aperture Radar
L-BFGS	Limited-Memory Broyden–Fletcher–Goldfarb–Shanno
NOs	Neural Operators
PDE	Partial Differential Equation
PI-DeepONet	Physics-Informed Deep Operator Network
PIML	Physics-Informed Machine Learning
PINNs	Physics-Informed Neural Networks
SRV	Stimulated Reservoir Volume
THM	Thermo-Hydro-Mechanical
THMC	Thermo-Hydro-Mechanical–Chemical

References

1. Biot, M.A. General theory of three-dimensional consolidation. *J. Appl. Phys.* **1941**, *12*, 155–164. [CrossRef]
2. Tsang, C.F.; Barnichon, J.D.; Birkholzer, J.; Li, X.L.; Liu, H.H.; Sillen, X. Coupled thermo-hydro-mechanical processes in the near field of a high-level radioactive waste repository in clay formations. *Int. J. Rock Mech. Min. Sci.* **2012**, *49*, 31–44. [CrossRef]
3. McTigue, D.F. Thermoelastic response of fluid-saturated porous rock. *J. Geophys. Res. Solid Earth* **1986**, *91*, 9533–9542. [CrossRef]
4. Jing, L.; Tsang, C.F.; Stephansson, O. DECOVALEX—An international co-operative research project on mathematical models of coupled THM processes for safety analysis of radioactive waste repositories. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1995**, *32*, 389–398. [CrossRef]
5. Zhou, D.; Tatomir, A.; Sauter, M. Thermo-hydro-mechanical modelling study of heat extraction and flow processes in enhanced geothermal systems. *Adv. Geosci.* **2021**, *54*, 229–240. [CrossRef]
6. Jansen, G.; Miller, S.A. On the Role of Thermal Stresses during Hydraulic Stimulation of Geothermal Reservoirs. *Geofluids* **2017**, *2017*, 4653278. [CrossRef]
7. Figueiredo, B.; Tsang, C.F.; Niemi, A. The influence of coupled thermomechanical processes on the pressure and temperature due to cold water injection into multiple fracture zones in deep rock formation. *Geofluids* **2020**, *2020*, 8947258. [CrossRef]

8. Huang, L.; Dontsov, E.; Fu, H.; Lei, Y.; Weng, D.; Zhang, F. Hydraulic fracture height growth in layered rocks: Perspective from DEM simulation of different propagation regimes. *Int. J. Solids Struct.* **2022**, *238*, 111395. [CrossRef]
9. Huang, L.; Liu, J.; Zhang, F.; Dontsov, E.; Damjanac, B. Exploring the influence of rock inherent heterogeneity and grain size on hydraulic fracturing using discrete element modeling. *Int. J. Solids Struct.* **2019**, *176–177*, 207–220. [CrossRef]
10. Wang, X.-H.; Huang, L.-K.; Li, C.-H.; Wu, A.-A.; Zhu, S.-R.; Qian, L.; Lu, K. Fracture propagation mechanisms of cross-layer fracturing with directional perforation in off-target horizontal wells. *Pet. Sci.* **2026**, *23*, 2622–2638. [CrossRef]
11. Palmer, I.; Mansoori, J. How permeability depends on stress and pore pressure in coalbeds: A new model. *SPE Reserv. Eval. Eng.* **1998**, *1*, 539–544. [CrossRef]
12. Taghizadeh, R.; Goshtasbi, K.; Manshad, A.K.; Ahangari, K. Geomechanical and thermal reservoir simulation during steam flooding. *Struct. Eng. Mech.* **2018**, *66*, 505–513. [CrossRef]
13. Li, L.; Guo, X.; Zhou, M.; Xiang, G.; Zhang, N.; Wang, Y.; Wang, S.; Landjobo Pagou, A. The investigation of fracture networks on heat extraction performance for an enhanced geothermal system. *Energies* **2021**, *14*, 1635. [CrossRef]
14. Li, J.; Marzouk, Y.M. Adaptive construction of surrogates for the Bayesian solution of inverse problems. *SIAM J. Sci. Comput.* **2014**, *36*, A1163–A1186. [CrossRef]
15. Oliver, D.S.; Chen, Y. Recent progress on reservoir history matching: A review. *Comput. Geosci.* **2010**, *15*, 185–221. [CrossRef]
16. Gallardo, E.; Deutsch, C.V. Approximate physics-discrete simulation of the steam-chamber evolution in steam-assisted gravity drainage. *SPE J.* **2019**, *24*, 477–491. [CrossRef]
17. Wang, T.; Sun, Z.; Sun, H.; Chen, Z.; Yao, J. Development of a scalable parallel compositional simulator for thermo-hydromechanical coupling in fractured rocks using an embedded discrete fracture model. *SPE J.* **2024**, *29*, 2545–2565. [CrossRef]
18. Yan, B.; Gudala, M.; Hoteit, H.; Sun, S.; Wang, W.; Jiang, L. Physics-informed machine learning for noniterative optimization in geothermal energy recovery. *Appl. Energy* **2024**, *365*, 123179. [CrossRef]
19. Sharma, P.; Chung, W.T.; Akoush, B.; Ihme, M. A review of physics-informed machine learning in fluid mechanics. *Energies* **2023**, *16*, 2343. [CrossRef]
20. Lagaris, I.E.; Likas, A.; Fotiadis, D.I. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Trans. Neural Netw.* **1998**, *9*, 987–1000. [CrossRef] [PubMed]
21. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
22. He, Q.; Tartakovsky, A.M. Physics-Informed Neural Network Method for Forward and Backward Advection-Dispersion Equations. *Water Resour. Res.* **2021**, *57*, e2020WR029479. [CrossRef]
23. Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G.E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229. [CrossRef]
24. Azizzadenesheli, K.; Kovachki, N.; Li, Z.; Liu-Schiaffini, M.; Kossaifi, J.; Anandkumar, A. Neural operators for accelerating scientific simulations and design. *Nat. Rev. Phys.* **2024**, *6*, 320–328. [CrossRef]
25. Wang, N.; Chen, Y.; Zhang, D. A comprehensive review of physics-informed deep learning and its applications in geoenergy development. *Innov. Energy* **2025**, *2*, 100087. [CrossRef]
26. Kuang, T.; Liu, J.; Yin, Z.; Jing, H.; Lan, Y.; Lan, Z.; Pan, H. Fast and robust prediction of multiphase flow in complex fractured reservoir using a Fourier neural operator. *Energies* **2023**, *16*, 3765. [CrossRef]
27. Michaloglou, A.; Papadimitriou, I.; Gialampoukidis, I.; Vrochidis, S.; Kompatsiaris, I. Physics-Informed Neural Networks in Materials Modeling and Design: A Review. *Arch. Comput. Methods Eng.* **2025**, *33*, 5223–5260. [CrossRef]
28. Cai, M.; Gu, H.; Hong, P.; Li, J. A combination of physics-informed neural networks with the fixed-stress splitting iteration for solving Biot's model. *Front. Appl. Math. Stat.* **2023**, *9*, 1206500. [CrossRef]
29. Szabó, L.; Kossa, A. A new exact integration method for the Drucker–Prager elastoplastic model with linear isotropic hardening. *Int. J. Solids Struct.* **2012**, *49*, 170–190. [CrossRef]
30. Roy, A.M.; Guha, S.; Sundararaghavan, V.; Arróyave, R. Physics-infused deep neural network for solution of non-associative Drucker–Prager elastoplastic constitutive model. *J. Mech. Phys. Solids* **2024**, *185*, 105570. [CrossRef]
31. Wang, S.; Teng, Y.; Perdikaris, P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM J. Sci. Comput.* **2021**, *43*, A3055–A3081. [CrossRef]
32. Amini, D.; Haghighat, E.; Juanes, R. Inverse modeling of nonisothermal multiphase poromechanics using physics-informed neural networks. *J. Comput. Phys.* **2023**, *490*, 112323. [CrossRef]
33. Hanna, J.; Aguado, J.V.; Comas-Cardona, S.; Askri, R.; Borzacchiello, D. Residual-based adaptivity for two-phase flow simulation in porous media using Physics-informed Neural Networks. *Comput. Methods Appl. Mech. Eng.* **2022**, *396*, 115086. [CrossRef]
34. Abbasi, J.; Jagtap, A.D.; Moseley, B.; Hiorth, A.; Andersen, P.Ø. Challenges and advancements in modeling shock fronts with physics-informed neural networks: A review and benchmarking study. *Neurocomputing* **2025**, *657*, 131440. [CrossRef]
35. Mao, Z.; Meng, X. Physics-informed neural networks with residual/gradient-based adaptive sampling methods for solving partial differential equations with sharp solutions. *Appl. Math. Mech.* **2023**, *44*, 1069–1084. [CrossRef]

36. Perez, S.; Maddu, S.; Sbalzarini, I.F.; Poncet, P. Adaptive weighting of Bayesian physics informed neural networks for multitask and multiscale forward and inverse problems. *J. Comput. Phys.* **2023**, *491*, 112342. [CrossRef]
37. Zhang, Z.; Lee, J.H.; Sun, L.; Gu, G.X. Weak-formulated physics-informed modeling and optimization for heterogeneous digital materials. *PNAS Nexus* **2024**, *3*, 186. [CrossRef] [PubMed]
38. Liu, Z.; Liu, Y.; Yan, X.; Liu, W.; Guo, S.; Zhang, C. AsPINN: Adaptive symmetry-recomposition physics-informed neural networks. *Comput. Methods Appl. Mech. Eng.* **2024**, *432*, 117405. [CrossRef]
39. Faroughi, S.A.; Soltanmohammadi, R.; Datta, P.; Mahjour, S.K.; Faroughi, S. Physics-informed neural networks with periodic activation functions for solute transport in heterogeneous porous media. *Mathematics* **2023**, *12*, 63. [CrossRef]
40. Kontolati, K.; Goswami, S.; Em Karniadakis, G.; Shields, M.D. Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems. *Nat. Commun.* **2024**, *15*, 5101. [CrossRef] [PubMed]
41. Hua, N.; Lu, W. Basis operator network: A neural network-based model for learning nonlinear operators via neural basis. *Neural Netw.* **2023**, *164*, 21–37. [CrossRef] [PubMed]
42. Choubineh, A.; Chen, J.; Wood, D.A.; Coenen, F.; Ma, F. Fourier Neural Operator for Fluid Flow in Small-Shape 2D Simulated Porous Media Dataset. *Algorithms* **2023**, *16*, 24. [CrossRef]
43. Ma, X.; Zhong, R.; Zhan, J.; Zhou, D. Enhancing subsurface multiphase flow simulation with Fourier neural operator. *Heliyon* **2024**, *10*, e38103. [CrossRef] [PubMed]
44. Faroughi, S.A.; Pawar, N.M.; Fernandes, C.; Raissi, M.; Das, S.; Kalantari, N.K.; Kourosh Mahjour, S. Physics-guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: Fluid and solid mechanics. *J. Comput. Inf. Sci. Eng.* **2024**, *24*, 040802. [CrossRef]
45. Lu, L.; Meng, X.; Cai, S.; Mao, Z.; Goswami, S.; Zhang, Z.; Karniadakis, G.E. A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data. *Comput. Methods Appl. Mech. Eng.* **2022**, *393*, 114778. [CrossRef]
46. Zhao, Z.; Yang, X.; Ding, D.; Wang, Q.; Zhang, F.; Hu, Z.; Xu, K.; Wang, X. Displacement Reconstruction Based on Physics-Informed DeepONet Regularizing Geometric Differential Equations of Beam or Plate. *Appl. Sci.* **2024**, *14*, 2615. [CrossRef]
47. Zhang, X.; Huang, Z.; Lei, Q.; Yao, J.; Gong, L.; Yang, W.; Yan, X.; Li, Y. Improving heat extraction performance of enhanced geothermal systems: Insights from critical fracture network parameter and multi-objective optimization method. *Appl. Therm. Eng.* **2022**, *213*, 118671. [CrossRef]
48. Qin, Z.; Jiang, A.; Faulder, D.; Cladouhos, T.T.; Jafarpour, B. Efficient optimization of energy recovery from geothermal reservoirs with recurrent neural network predictive models. *Water Resour. Res.* **2023**, *59*, e2022WR032653. [CrossRef]
49. Badawi, D.; Gildin, E. Neural operator-based proxy for reservoir simulations considering varying well settings, locations, and permeability fields. *Comput. Geosci.* **2025**, *196*, 105826. [CrossRef]
50. Chen, J.; Xu, T.; Liang, X.; Zhang, S. Evaluation and Optimization of Heat Extraction Strategies Based on Deep Neural Network in the Enhanced Geothermal System. *J. Energy Eng.* **2023**, *149*, 04022045. [CrossRef]
51. Wang, J.; Xie, H.P.; Matthai, S.K.; Hu, J.J.; Li, C.B. The role of natural fracture activation in hydraulic fracturing for deep unconventional geo-energy reservoir stimulation. *Pet. Sci.* **2023**, *20*, 2141–2164. [CrossRef]
52. Zhou, B.; Chen, Z.; Song, Z.; Tang, Z.; Wang, B.; Olorode, O. A new analytically modified embedded discrete fracture model for pressure transient analysis in fluid flow. *J. Hydrol.* **2024**, *636*, 131330. [CrossRef]
53. He, Q.; Barajas-Solano, D.; Tartakovsky, G.; Tartakovsky, A.M. Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Adv. Water Resour.* **2020**, *141*, 103610. [CrossRef]
54. Meng, X.; Karniadakis, G.E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *J. Comput. Phys.* **2020**, *401*, 109020. [CrossRef]
55. Fan, Z.; Fan, G.; Zhang, D.; Luo, T.; Han, X.; Xu, G.; Tong, H. Mechanism of Burial Depth Effect on Recovery Under Different Coupling Models: Response and Simplification. *Appl. Sci.* **2025**, *15*, 11657. [CrossRef]
56. Wen, G.; Li, Z.; Azizzadenesheli, K.; Anandkumar, A.; Benson, S.M. U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Adv. Water Resour.* **2022**, *163*, 104180. [CrossRef]
57. Lee, J.E.; Zhu, M.; Xi, Z.; Wang, K.; Yuan, Y.O.; Lu, L. Efficient and generalizable nested Fourier-DeepONet for three-dimensional geological carbon sequestration. *Eng. Appl. Comput. Fluid Mech.* **2024**, *18*, 2435457. [CrossRef]
58. Hu, G. Physics-based neural networks for coupled thermo-hydro-mechanical-chemical analysis of FEBEX bentonite. *Environ. Earth Sci.* **2026**, *85*, 172. [CrossRef]
59. Ayoub, A.; Wainwright, H.M.; Wang, L.; Sansavini, G. An enhanced fourier neural operator surrogate for radioactive plume transport forecasting. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 3165–3175. [CrossRef]
60. Xu, Y.; Kohtz, S.; Boakye, J.; Gardoni, P.; Wang, P. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108900. [CrossRef]
61. Vidal, R.; Zhu, Z.; Haeffele, B.D. Optimization landscape of neural networks. In *Mathematical Aspects of Deep Learning*; Cambridge University Press: Cambridge, UK, 2022; Volume 1, p. 200. [CrossRef]

62. Shriver, F.; Watson, J. Physically adversarial thermal hydraulics evaluation of deep learning models for pressurized water reactors. *Prog. Nucl. Energy* **2022**, *146*, 104149. [CrossRef]
63. Ainsworth, M.; Oden, J.T. A posteriori error estimation in finite element analysis. *Comput. Methods Appl. Mech. Eng.* **1997**, *142*, 1–88. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.