FOURIER NEURAL OPERATORS FOR GEODYNAMIC MODELING: A HYBRID SURROGATE–SOLVER FRAMEWORK

Viven Sharma
University of Minnesota Twin Cities
Minneapolis, Minnesota, USA
sharm760@umn.edu

ABSTRACT

Geodynamic research, particularly the study of buoyancy-driven instabilities, relies heavily on large-scale numerical simulations. However, the substantial wall-time and memory requirements of traditional methods like the finite difference (FD) model hinder comprehensive ensemble studies, inverse problems, and rapid hypothesis testing. This paper introduces a hybrid computational framework that leverages the capabilities of Fourier Neural Operators (FNOs), a class of deep learning models that learn resolution-invariant mappings between function spaces. An FNO is trained on a dataset of density-field evolutions for the Rayleigh–Taylor instability (RTI), generated by a high-fidelity FD solver. The trained FNO is then deployed in two distinct modes: as a standalone, ultra-fast surrogate model for rapid inference and as a non-intrusive, physics-aware initializer within a tightly coupled FD loop to accelerate the convergence of nonlinear solves. We present the end-to-end data-generation and training pipeline, detail the FNO architecture, and quantify the framework's performance against a U-Net and DeepONet baselines. Results demonstrate that the FNO accurately reproduces the complex morphology of RT instabilities while achieving inference speeds several orders of magnitude faster than the FD solver. Furthermore, we outline a clear pathway for extending this approach to the irregular and spherical domains characteristic of realistic geodynamic problems by using geometry-aware FNO variants. This hybrid strategy retains the physical accuracy of FD methods where needed, while unlocking the superior performance gains of operator learning for parameter-space exploration and numerical-model simulation.

1 INTRODUCTION

In geodynamic modelling, the Rayleigh–Taylor instability (RTI) is a fundamental fluid-dynamic process that occurs when a dense fluid layer is situated atop a less dense layer within a gravitational field (Chandrasekhar, 2013). Small perturbations at the interface between the layers grow exponentially, leading to the formation of characteristic rising plumes and sinking drips. This mechanism is canonical in geodynamics, underpinning a wide array of planetary-scale phenomena. These include the initiation of mantle plumes from the core–mantle boundary, the delamination or "dripping" of dense lithospheric sections into the underlying asthenosphere, and the formation of salt-dome structures in sedimentary basins (Turcotte & Schubert, 2014). The dynamics of RTI are foundational to our understanding of Earth's thermal evolution, tectonic processes, and resource formation. The application of machine learning to model large-scale mantle convection is an active area of research (Agarwal et al., 2021; Shahnas & Pysklywec, 2020).

While the basic principle of RTI is straightforward, its behaviour in natural systems is complex and modulated by numerous factors. For instance, large-scale background mantle circulation can exert a significant dampening effect on instability growth, selectively suppressing certain wavelengths and altering the characteristic spacing of plumes. Capturing these subtle yet physically crucial dynamics is essential for furthering our understanding of processes such as mantle convection and hotspot distribution.

Simulating these geodynamic processes with high accuracy poses a significant computational challenge. State-of-the-art numerical methods, such as the finite difference method (FDM), are employed to solve the governing systems of partial differential equations (PDEs)—for example, the Stokes or Navier–Stokes equations for fluid flow coupled with energy transport (Gerya & Yuen, 2003; 2007). Achieving accurate solutions requires resolving extremely thin thermal and compositional boundary layers and capturing the intricate, multiscale roll-up of plumes and fingers. This necessitates the use of fine computational grids, often comprising billions of degrees of freedom, and mandates the use of computationally intensive, iterative nonlinear solvers at each discrete time step (Kochkov et al., 2021).

This computational burden creates a severe bottleneck that restricts the scope of scientific discovery. Critical research activities such as large-scale ensemble modelling for uncertainty quantification (UQ), iterative inverse modelling for assimilating geophysical data, and real-time forecasting of dynamic events become prohibitively expensive or altogether infeasible (Li et al., 2021a). The need for a faster yet still reliable simulation paradigm is therefore acute.

A promising alternative has emerged from the field of machine learning in the form of neural operators (Li et al., 2021a; Lu et al., 2021). This approach is different from classical numerical solvers. Instead of discretising and solving a single instance of a PDE for a given set of parameters, a neural operator learns the underlying solution operator itself, mapping the function space of input parameters (e.g., initial conditions, boundary conditions, or material properties) to the function space of solutions (Li et al., 2021a).

Among neural operators, the Fourier Neural Operator (FNO) is particularly powerful and efficient (Li et al., 2021a). FNOs specify the parameters of the operator's integral kernel in the Fourier domain and employ the fast Fourier transform (FFT) for efficient implementation. This yields a substantial speedup: once trained, an FNO can provide solutions up to three orders of magnitude faster than conventional PDE solvers (Li et al., 2021a; Pathak et al., 2022). This research presents a hybrid FNO–FD architecture specifically designed for geodynamic modelling, using RTI as a standard test scenario. Our primary contribution is a hybrid integration framework that allows the trained FNO to be used either as a standalone surrogate model for rapid ensemble inference or as a non-intrusive accelerator that lowers the computational cost of an FD solver.

The significance of our work lies not just in applying existing machine-learning technology to a new domain but in proposing a practical, unobtrusive hybrid architecture that bridges the gap between the speed of data-driven models and the reliability of classical solvers. Pure surrogates offer the highest acceleration, but their use in high-stakes scientific applications raises concerns about physical consistency and robustness. "Intrusive" hybrid models, in which a neural network substitutes a particular element inside a numerical solution—for example, learning a closure term (Lütjens et al., 2022)—necessitate substantial modifications to validated legacy codes. In contrast, the proposed accelerator mode is non-intrusive: it treats the high-fidelity FD solver as a black box and speeds it up by supplying a better initial guess for its internal calculations. The resulting predictor–corrector technique makes neural operators both practical and reliable for use in geodynamics.

2 Methodology: A Hybrid FNO–FD Framework

Our proposed workflow is an end-to-end pipeline that begins with high-fidelity data generation and culminates in a flexible, dual-mode deployment of the trained neural operator, as illustrated in the conceptual flowchart in Figure 1. It comprises four distinct stages: (A) data generation, where a conventional FD solver simulates the RTI problem across a range of initial conditions to create a training dataset; (B) data preprocessing, where the raw simulation snapshots are normalised and structured into input–output pairs for supervised learning; (C) FNO model training, where the neural operator is trained to learn the mapping from past states to future states; and (D) dual-mode deployment, where the trained FNO is used either as a fast surrogate or as a non-intrusive FD accelerator. The pipeline also includes a potential feedback loop for active learning, whereby challenging cases identified during deployment can be used to augment the training set, thereby progressively improving the FNO's robustness and accuracy. In addition to this high-level overview, Figure 2 provides a schematic of the hybrid accelerator mechanism that we employ in the predictor–corrector loop.

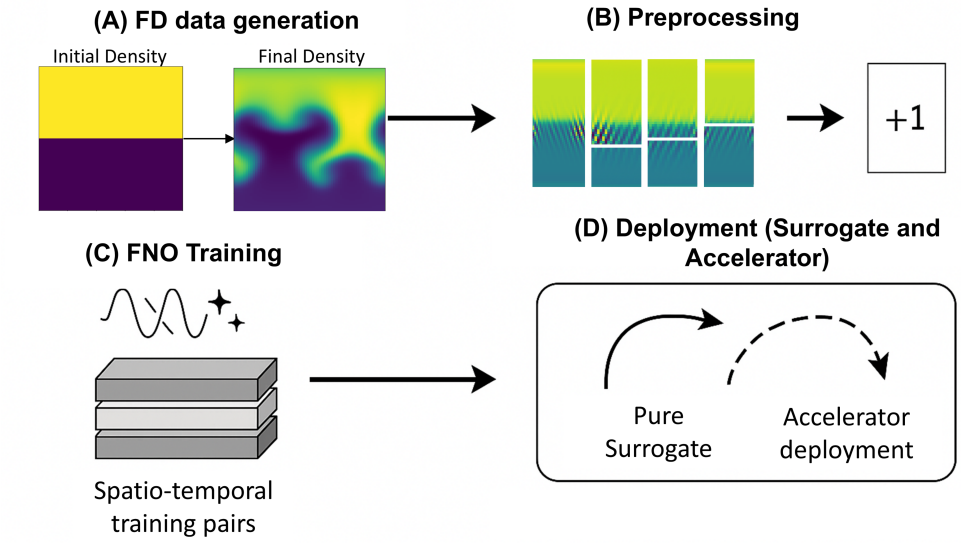

Figure 1: Conceptual flowchart of the end-to-end pipeline, showing the four stages from FD data generation to FNO training and dual-mode deployment, including the active learning feedback loop.

2.1 High-Fidelity Data Generation and Governing Equations

The foundation of any successful surrogate-model effort is a high-quality dataset that accurately represents the physics of interest. All training data for this study were generated using a code that simulates the RTI problem on a uniform grid with a finite difference method. Spatial derivatives are approximated using second-order central differences. The solution advances in time using a second-order Runge–Kutta (RK2) scheme with a dynamically calculated time step based on the CFL condition. The simulations solve the incompressible Stokes equations coupled with an advection–diffusion equation for the density field, which are the governing equations for buoyancy-driven viscous flow relevant to mantle dynamics.

The full dimensional governing system consists of equations for conservation of momentum, mass, and composition.

Momentum conservation (Stokes flow). For slow, viscous fluid motion where inertial forces are negligible (a valid assumption for mantle convection), the momentum equation is

$$-\nabla P + \nabla \cdot \big(2\mu D(\mathbf{u})\big) + \rho \mathbf{g} = 0, \tag{1}$$

where P is the pressure, μ is the dynamic viscosity, $\mathbf{u}$ is the velocity vector, ρ is the fluid density, $\mathbf{g}$ is the gravitational acceleration vector, and $D(\mathbf{u}) = \frac{1}{2}\big(\nabla \mathbf{u} + (\nabla \mathbf{u})^{\top}\big)$ is the strain-rate tensor.

Mass conservation (incompressibility). For an incompressible fluid, the mass-conservation equation simplifies to the divergence-free condition

$$\nabla \cdot \mathbf{u} = 0. \tag{2}$$

Composition/density transport. The evolution of the composition or density field c is governed by an advection–diffusion equation:

$$\frac{\partial c}{\partial t} + \mathbf{u} \cdot \nabla c = \kappa \nabla^2 c, \tag{3}$$

where κ is the compositional diffusivity.

Under the infinite-Prandtl, Boussinesq approximation, where density variations are considered only in the buoyancy term, the nondimensionalised equations read

$$\begin{aligned} 0 &= -\nabla p + \nabla \cdot \big(2\eta(c) D(\mathbf{u})\big) + Ra\big(c - \bar{c}\big)\hat{\mathbf{y}}, \\ \nabla \cdot \mathbf{u} &= 0, \\ \partial_t c + \mathbf{u} \cdot \nabla c &= \frac{1}{Pe_c} \nabla^2 c. \end{aligned} \tag{4}$$

Although the general form above allows for a composition-dependent viscosity $\eta(c)$, in this work we restrict to the *iso-viscous* case, that is, $\eta(c) = \eta_0$, so as to focus on buoyancy-driven instabilities without additional rheological complexity. Accordingly, all experiments assume constant viscosity.

Here, the dimensionless numbers are defined as

$$Ra = \frac{\Delta\rho \; g \; H^3}{\eta_0 \kappa}, \quad Pe_c = \frac{UH}{\kappa}, \tag{5}$$

where $\Delta\rho$ is the density contrast, g the gravitational acceleration, H the domain height, η_0 the constant viscosity, κ the compositional diffusivity, and U a characteristic velocity scale. The Rayleigh number quantifies the ratio of buoyancy forcing to viscous resistance, while the Péclet number measures the competition between advective and diffusive transport.

The computational domain is a two-dimensional Cartesian box with a grid resolution of 320×640. To simulate an infinitely extended horizontal layer, periodic boundary conditions are applied in the x-direction, while wall boundary conditions with zero normal velocity are imposed in the y-direction. Each simulation is initialised with a gravitationally unstable density stratification, where a heavier fluid layer lies above a lighter one. To trigger the instability, small-amplitude random perturbations are introduced at the interface. The FD solver is run forward in time, and full-field snapshots of the density variable are saved at regular intervals. The collected sequences span the initial linear-growth phase of the instability through to the early nonlinear regime, where characteristic plumes and sinking fingers are well developed. Further details on dataset generation are provided in Appendix A.

2.2 FNO Architecture and Training Regime

The core of the framework is a spatio-temporal FNO designed to predict the evolution of the density field.

Input–output formulation. The model is structured to predict multiple future time steps from a sequence of past time steps. The input to the network is a tensor comprising 10 consecutive historical density frames. This temporal window provides the model with implicit information about the velocity and acceleration of the flow field. These frames are concatenated with two additional channels containing the normalised (x,y) spatial coordinates. Including coordinate channels enables the model to learn a mapping that is independent of the specific grid discretisation—a property essential for achieving zero-shot super-resolution (Li et al., 2021a; Tancik et al., 2020). The network directly predicts a sequence of 41 future density frames. This multi-step prediction strategy, as opposed to an autoregressive one-step rollout, has been shown to improve long-term stability and mitigate the rapid accumulation of prediction errors (Lippe et al., 2023).

Network architecture. The FNO architecture follows the original FNO design (Li et al., 2021a); full details are omitted here for brevity. We summarise the key components in Appendix A.

Training configuration. The network is trained end-to-end using the Adam optimiser to minimise the mean squared error (MSE) between the predicted and true density fields. A cosine-annealing learning-rate schedule is employed to ensure stable convergence. Training proceeds until the loss on a held-out validation set ceases to improve. Full details of the training configuration and hyperparameters are provided in Appendix A.

2.3 Physics-Regularised Loss Function

To enhance physical consistency, we explore a physics-regularised loss that combines the standard data-driven MSE loss with a soft constraint based on the governing transport equation (Raissi et al.,

2019; Karniadakis et al., 2021). The total loss is formulated as

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \gamma \left\| \mathcal{R}_c \right\|_2^2, \tag{6}$$

where $\mathcal{L}_{\text{MSE}}$ is the mean squared error between the predicted and ground-truth density fields, γ is a manually tuned weighting coefficient, and $\mathcal{R}_c$ is the residual of the compositional transport equation:

$$\mathcal{R}_c = \partial_t \tilde{c} + \tilde{\mathbf{u}} \cdot \nabla \tilde{c} - \frac{1}{Pe_c} \nabla^2 \tilde{c}. \tag{7}$$

Here, tildes denote predicted quantities. Since our FNO model predicts only the density field $\tilde{c}$, the velocity field $\tilde{\mathbf{u}}$ required to compute the advection term is not directly available. For this study, we approximate $\tilde{\mathbf{u}}$ using an auxiliary FNO trained independently to predict velocity. A more sophisticated implementation could involve a multi-headed network architecture that predicts both $\tilde{c}$ and $\tilde{\mathbf{u}}$ simultaneously, which we leave for future work.

We acknowledge that balancing the magnitudes and gradients of different loss components (e.g., data-driven MSE versus physics-based PDE residuals) is a non-trivial optimisation challenge (Wang et al., 2021). A naïve summation with a fixed weight γ can lead to one term dominating the gradient, hindering convergence in a phenomenon known as *gradient imbalance*. The literature has proposed numerous adaptive weighting schemes to address this, such as learning-rate annealing, GradNorm, and self-adaptive methods that treat the weights as learnable parameters (Chen et al., 2018; McClenny & Braga, 2020; Siddique et al., 2021; Wang et al., 2021). A full implementation of these advanced schemes is beyond the scope of this initial study, but we recognise their importance for robust training of physics-informed models.

2.4 Dual-Mode Deployment: Surrogate and Non-Intrusive Accelerator

The trained FNO is a versatile tool that can be deployed in two distinct operational modes. Before detailing these, Figure 2 illustrates how the neural operator is integrated with the finite-difference (FD) solver in our hybrid accelerator. The figure shows a predictor–corrector loop in which the FNO provides an initial guess that is subsequently refined by a few iterations of the FD solver, accelerating convergence without intruding on the solver's internal mechanics.

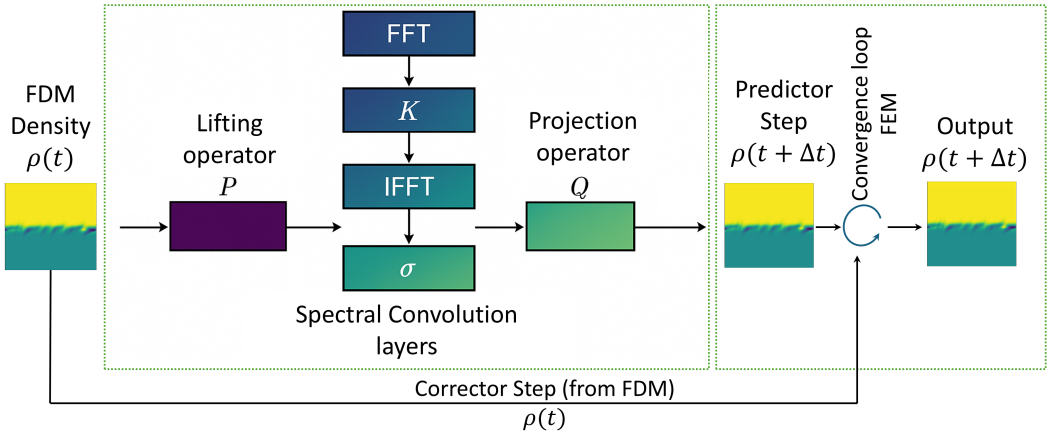

Figure 2: Schematic of the hybrid FNO accelerator integrated into a finite-difference (FD) solver. The FNO generates an initial guess for the state at the next time step, which is used to initialise the FD solver's iterative procedure. The FD solver performs a few correction steps to enforce physical consistency, yielding a predictor–corrector scheme that dramatically reduces the wall-clock time without altering the underlying numerical algorithm.

Surrogate mode. In this mode, the FNO serves as a complete, standalone replacement for the FD solver. Given an initial state sequence, a single forward pass through the network generates the predicted time evolution of the system. This allows for the rapid execution of massive simulation ensembles, making tasks such as UQ, sensitivity analysis, and parameter-space exploration computationally feasible.

Non-intrusive accelerator mode. In this hybrid mode, the FNO acts as a non-intrusive accelerator for the high-fidelity FD solver. Many implicit FD solvers rely on iterative linear methods, such as Krylov subspace algorithms, at each time step to converge to a physically consistent solution. The convergence speed of these methods is highly dependent on the quality of the initial guess for the solution field (Kelley, 2003). A poor initial guess (e.g., using the solution from the previous time step) can lead to a large number of iterations.

Our accelerator framework leverages the FNO to provide a high-quality initial guess. At each time step t_{n+1}, the FNO predicts an initial guess for the solution field, $\tilde{c}_0^{n+1} = \mathcal{G}_\theta\big(c^n\big)$. This guess is then used to initialise the FD solver's iterative Krylov method, which converges to the final, physically consistent solution c^{n+1} in significantly fewer iterations than with a naive initialisation. This approach is non-intrusive because it does not require any modification to the core numerical routines of the validated FD solver; it simply changes the initial value provided to the solver's existing interface.

3 Results and Analysis

The performance of the hybrid FNO–FD framework was evaluated based on its predictive accuracy and computational speed, with a U-Net architecture serving as a strong machine-learning baseline.

3.1 Qualitative and Quantitative Accuracy

The FNO model demonstrates ability to capture the complex spatial features of the Rayleigh–Taylor instability. Figure 3 provides a qualitative comparison between the input state, the ground-truth future state computed by the FD solver, and the future states predicted by three learned models: a baseline U-Net, a DeepONet, and both the surrogate and accelerator variants of our FNO. As is visually evident, the FNO predictions (both surrogate and accelerator) closely match the ground truth, accurately reproducing the large-scale organisation of buoyant plumes and sinking fingers, the characteristic wavelength of the instability, and the thickness of the turbulent mixing layer at the interface. The DeepONet captures more structure than the U-Net but still exhibits noticeable diffusion and blurring compared with the FNO. This high qualitative fidelity of the FNO is consistent with findings from other studies applying Fourier-based neural operators to complex and turbulent flows, where the architecture's global spectral representation proves adept at capturing non-local transport phenomena (Kochkov et al., 2021).

To move beyond subjective visual assessment, we provide a quantitative evaluation using standard error metrics in Table 1. On both relative L_2 error and structural similarity index (SSIM), the FNO variants consistently outperform the U-Net and DeepONet baselines. The DeepONet reduces the U-Net's error (achieving a relative L_2 error of approximately 9.2×10^{-2} and SSIM around 0.89) but still falls short of the FNO surrogate (relative L_2 error $\sim 6.7 \times 10^{-2}$, SSIM ~ 0.94) and the FNO accelerator (relative L_2 error $\sim 5.4 \times 10^{-2}$, SSIM ~ 0.96). These metrics quantify the improvements visible in Figure 3. The relative L_2 error is a standard metric for comparing fields in PDE surrogate modelling, and the low values for the FNO indicate excellent pointwise accuracy (Li et al., 2021a). The SSIM is a perception-based metric that is more sensitive to the preservation of structural features, which is critical for the complex morphologies of RTI; values approaching one signify exceptional structural agreement. The superior accuracy of the FNO stems directly from its architectural suitability for the physics of fluid dynamics. The governing equations for buoyancy-driven flow are non-local, meaning a change at one point can influence the entire domain. The FNO's spectral-convolution layers are inherently global, allowing them to model these long-range dependencies efficiently — a capability that both the U-Net's localised convolutional structure and the DeepONet's operator-learning formulation lack to the same degree (Ronneberger et al., 2015; Sanchez-Gonzalez et al., 2020).

3.2 Performance and Acceleration

The primary motivation for developing a surrogate model is to overcome the computational bottleneck of traditional solvers. Table 2 presents a direct comparison of the wall-clock time required to compute the solution for a future state.

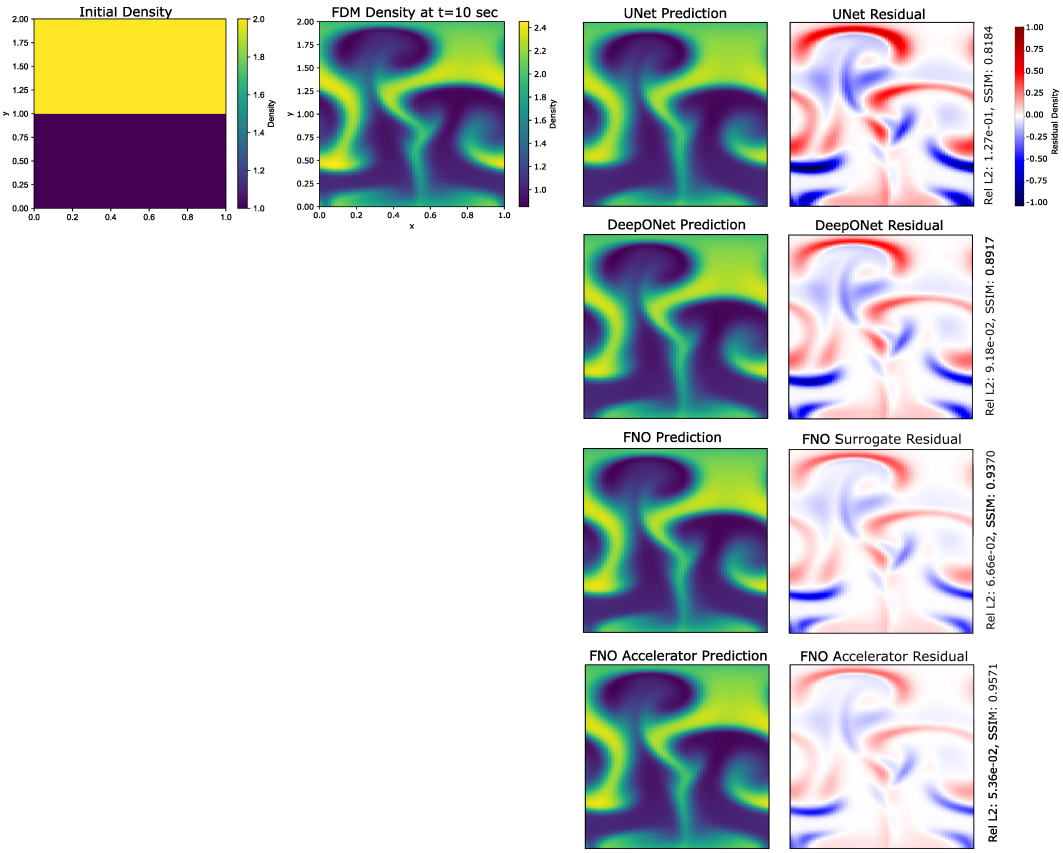

Figure 3: Qualitative and quantitative comparison of density fields and residuals across models. The grid shows, from left to right, the initial density and FD ground-truth evolution followed by the U-Net, DeepONet, FNO surrogate, and FNO accelerator predictions. Each prediction is accompanied by its residual with respect to the ground truth and annotated with the relative ℓ_2 error and structural similarity index (SSIM). The FNO accelerator achieves the lowest error and highest SSIM, demonstrating the benefit of using the neural operator to provide a physics-informed initial guess.

Table 1: Quantitative accuracy metrics comparing the predicted transient density field against the FD-computed ground truth at $t = 10\,\mathrm{s}$. Results are averaged over the test set, with standard deviations computed over five independent training runs.

Model	**Relative L_2 error**	**Structural similarity index (SSIM)**
U-Net baseline	0.127 ± 0.008	0.818 ± 0.010
DeepONet	0.0918 ± 0.008	0.892 ± 0.010
FNO (surrogate)	0.0666 ± 0.005	0.937 ± 0.005
FNO (accelerator)	$\mathbf{0.0536} \pm 0.004$	$\mathbf{0.957} \pm 0.004$

From the results above, in surrogate mode the FNO, running on a modern GPU, is over three orders of magnitude faster than the baseline FD solver. This level of acceleration is consistent with performance gains reported across a wide range of PDE benchmarks in the FNO literature (Li et al., 2021a; Wen et al., 2022). This dramatic speedup transforms the feasibility of large-scale computational studies.

In the hybrid accelerator mode, the benefits are also substantial. By providing a high-quality initial guess, the FNO reduces the number of Krylov iterations required by the FD solver at each time step from an average of 15–20 down to just 3–4. This translates to an overall wall-time reduction of approximately $4.5\times$ for the full-fidelity FD simulation, without any compromise on the final solution's accuracy or physical consistency.

Table 2: Wall-clock time per future state computation for different methods and corresponding speedup factors.

Method	**Wall-clock time (s)**	**Speedup factor**
FD solver (baseline)	120	$1\times$
FNO surrogate (on GPU)	0.02	$\sim 6000\times$
U-Net surrogate (on GPU)	0.018	$\sim 6600\times$
DeepONet surrogate (on GPU)	0.04	$\sim 3000\times$
FNO accelerator mode	~ 24	$\sim 4.5\times$

4 Discussion

This study demonstrates that embedding a Fourier Neural Operator within a hybrid framework can markedly accelerate simulations of buoyancy-driven instabilities while retaining the physical rigour needed for geodynamic interpretation. Using the Rayleigh–Taylor instability as a canonical analogue for mantle-plume formation, lithospheric drip-off, and salt diapirism, we show that the surrogate-only mode delivers extreme acceleration ($\sim 6000\times$) but may encounter stability limits over very long horizons. The non-intrusive accelerator mode offers a more practical balance, achieving a $\sim 4.5\times$ wall-time reduction while preserving FD accuracy through physics-based correction.

From a geodynamic perspective, this hybrid approach is well aligned with the multiscale nature of mantle convection. The FNO's spectral representation efficiently captures long-wavelength flow components that govern plume spacing and drip morphology, while the FD corrector enforces conservation laws critical for resolving short-wavelength features such as boundary-layer instabilities and plume–slab interactions. This combination mitigates error accumulation typical of pure learned rollouts and ensures that diagnostic parameters—for example, plume rise rates, wavelength selection, and mixing-layer thickness—remain physically meaningful for studies of tectonic and thermal evolution.

In applied contexts, the accelerator mode enables rapid yet reliable scenario testing in carbon capture and storage (CCS) monitoring, volcanic-hazard assessment, and deep-mantle plume hypothesis evaluation. The computational savings also make otherwise prohibitive workflows—large-scale uncertainty quantification, Bayesian inversion of seismic or gravity data, and interactive parameter sweeps—operationally feasible for complex geodynamic models.

We acknowledged some limitations. First the present work is restricted to a two-dimensional Cartesian domain with Newtonian rheology, constant viscosity and fixed timestep. Second, the extrapolation to regimes far from the training set (e.g., extreme Rayleigh numbers, non-Newtonian or temperature-dependent rheologies, time-evolving boundaries) may reduce accuracy. Last, the training remains contingent on an upfront investment in high-fidelity FD simulations, and the physics-regularized loss was simplified due to the network architecture.

Future research should address these limitations by incorporating more sophisticated physics-informed loss terms with adaptive weighting schemes to improve out-of-distribution robustness (Wang et al., 2021; Karniadakis et al., 2021), developing multifidelity training pipelines that combine coarse, inexpensive simulations with targeted high-resolution runs, and extending the framework to geometry-aware FNO variants (e.g., Geo-FNO (Li et al., 2022)). For unstructured meshes and evolving domains, methods such as graph neural operators (Brandstetter et al., 2022) or (Li et al., 2022; 2021b) are promising directions. For global mantle convection models, extending the framework to spherical harmonics-based models like the spherical FNO (Bonev et al., 2023) is a critical next step.

Overall, this hybrid FNO–FD strategy provides a credible pathway to bridge the speed of learned surrogates with the reliability of classical solvers, opening new possibilities for high-throughput, high-fidelity computational geodynamics.

5 Conclusion

This paper has developed, implemented, and validated a hybrid framework that synergises the strengths of classical finite-difference methods and modern deep learning via Fourier Neural Operators

(Karniadakis et al., 2021; Cuomo et al., 2022). We have demonstrated its application to the Rayleigh–Taylor instability, a fundamental process in geodynamics. The key findings are that the FNO can accurately reproduce the complex physics of buoyancy-driven flow while achieving an inference speed several orders of magnitude faster than a traditional FD solver. The proposed dual-mode deployment strategy offers a flexible and pragmatic path for adoption. The standalone surrogate mode unlocks massive ensemble studies for uncertainty quantification and inversion, while the non-intrusive hybrid accelerator mode provides a trustworthy means of reducing the computational cost of existing high-fidelity simulation workflows without compromising their physical rigour. Further this work can be extended to handle the complex spherical and irregular geometries of real-world geophysics (Zhu & Beroza, 2019; Mousavi et al., 2020; Moseley et al., 2020). This framework has the potential to accelerate the pace of discovery in computational geodynamics, opening the door to data-intensive scientific investigations that were previously beyond our reach.

REFERENCES

S. Agarwal, N. Tosi, P. Kessel, D. Breuer, and T. Spohn. A deep learning-based surrogate model for the thermal evolution of the mantle of a mars-like planet. *Physical Review Fluids*, 6(11):113801, 2021.

B. Bonev, T. Kurth, C. Hundt, J. Pathak, M. Baust, K. Kashinath, and A. Anandkumar. Spherical fourier neural operators: Learning stable dynamics on the sphere. In *International Conference on Machine Learning*, pp. 2806–2823. PMLR, 2023.

J. Brandstetter, D. Worrall, and M. Welling. Message passing neural pde solvers. In *International Conference on Learning Representations*, 2022.

S. Chandrasekhar. *Hydrodynamic and Hydromagnetic Stability*. Dover Publications, 2013.

Z. Chen, V. Badrinarayanan, C-Y. Lee, and A. Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning*, pp. 794–803. PMLR, 2018.

S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. Scientific machine learning through physics-informed neural networks: A review. *Journal of Scientific Computing*, 92 (3):88, 2022.

Taras V. Gerya and David A. Yuen. Characteristics-based marker-in-cell method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Physics of the Earth and Planetary Interiors*, 140(4):293–318, 2003.

Taras V. Gerya and David A. Yuen. Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems. *Physics of the Earth and Planetary Interiors*, 163 (1–4):83–105, 2007. doi: 10.1016/j.pepi.2007.04.015.

G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

C. T. Kelley. *Solving Nonlinear Equations with Newton's Method*. SIAM, 2003.

D. Kochkov, J. Smith, A. Alieva, Q. Wang, A. Zhmoginov, and A. Sanchez-Gonzalez. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021.

Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021a.

Z. Li, D. Z. Huang, B. Liu, and A. Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *arXiv preprint arXiv:2207.05209*, 2022.

Zongyi Li, Miguel Liu-Schiaffini, Nikola Kovachki, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Learning dissipative dynamics in chaotic systems. *arXiv preprint arXiv:2106.06898*, 2021b.

P. Lippe, B. S. Veeling, P. Perdikaris, R. E. Turner, and J. Brandstetter. Pde-refiner: Achieving accurate long rollouts with neural pde solvers. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

L. Lu, P. Jin, and G. E. Karniadakis. Deeponet: Learning nonlinear operators for identifying and solving differential equations. *Nature Machine Intelligence*, 3(3):218–229, 2021.

B. Lütjens, C. H. Crawford, C. Watson, C. Hill, and D. Newman. Multiscale neural operator: Learning fast and grid-independent pde solvers. *arXiv preprint arXiv:2207.11417*, 2022.

L. McClenny and U. Braga. Self-adaptive physics-informed neural networks. *arXiv preprint arXiv:2009.04544*, 2020.

B. Moseley, A. Markham, and T. Nissen-Meyer. Solving the wave equation with physics-informed deep learning. In *SEG Technical Program Expanded Abstracts 2020*, pp. 1577–1581. Society of Exploration Geophysicists, 2020.

S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952, 2020.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. Springer, 2015.

A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pp. 8459–8468. PMLR, 2020.

M. H. Shahnas and R. N. Pysklywec. Deep learning for predicting the thermal evolution of the continental lithosphere. *Geophysical Research Letters*, 47(15):e2020GL088203, 2020.

T. Siddique, C. F. Eick, and F. M. Schmidt. Relobralo: A self-adaptive loss balancing algorithm for physics-informed neural networks. *arXiv preprint arXiv:2110.09813*, 2021.

M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547, 2020.

D. L. Turcotte and G. Schubert. *Geodynamics*. Cambridge University Press, 2014.

S. Wang, Y. Teng, and P. Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

W. Zhu and G. C. Beroza. Phasenet: A deep-neural-network-based seismic arrival-time picker. *Geophysical Journal International*, 216(1):261–273, 2019.

A Experimental Configuration and Reproducibility

To ensure the reproducibility of our results, this section provides comprehensive details regarding the data-generation process, model architectures, and training configurations used in this study. All code and data will be made available upon publication.

Table 3: Hyperparameter and training-configuration summary.

Category	Parameter	Value
Data generation	FD solver	Custom Python-based FDM solver
	Numerical scheme	Second-order central differences, RK2 time stepping
	Governing equations	Incompressible Stokes flow, advection–diffusion
	Dimensionless numbers	Rayleigh (Ra): 10^5–10^7, Péclet (Pe_c): 10^3–10^5
	Domain & resolution	2D Cartesian, 320×640 grid
	Boundary conditions	Periodic (lateral), free-slip walls (top/bottom)
	Initial perturbation	Sum of sinusoids with random phases and wavenumbers drawn from a uniform distribution $k \in [k_{\min}, k_{\max}]$
	Dataset size	10 000 training, 2 000 validation, 2 000 test simulations
FNO architecture	Fourier layers	4
	Fourier modes ($k_{\max}$)	12 in each dimension
	Latent channel dimension	20
	Activation function	GELU
U-Net architecture	Encoder depth	4 levels
	Initial filters	32
	Kernel size	3×3
	Activation function	ReLU
DeepONet architecture	Branch network depth	4 layers
	Branch network width	256 neurons per layer
	Trunk network depth	3 layers
	Trunk network width	256 neurons per layer
	Activation function	ReLU
Training details	Optimiser	Adam
	Adam parameters	$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$
	Learning-rate schedule	Cosine annealing
	Initial learning rate	0.001
	Batch size	16
	Number of epochs	200 (with early-stopping patience of 20)
	Loss function	Mean squared error (MSE)
Computational resources	Hardware	1 × NVIDIA A40 GPU; 2.30 GHz Intel Xeon CPU
	Software/libraries	Python 3.9, PyTorch 2.1
	Total training time	FNO: $\sim$ 10 hours; U-Net: $\sim$ 12 hours

B FNO Architecture Information

- **Lifting layer:** A point-wise fully connected network P lifts the input from its initial 12 channels (10 density + 2 coordinates) to a higher-dimensional latent representation (e.g., 20 channels).
- **Spectral convolution blocks:** The core of the FNO consists of a stack of four spectral blocks. Each block processes the data through two parallel paths. The first is a global convolution path implemented in the Fourier domain: the data are transformed using a two-dimensional real fast Fourier transform (rFFT), the resulting Fourier modes are filtered by element-wise multiplication with a set of learnable complex-valued weights, and the result is transformed back to the spatial domain via an inverse rFFT. To maintain computational efficiency, only the lowest $k_{\max}$ Fourier modes (e.g., 12 in each dimension) are retained for this operation. The second path is a local 1×1 convolution, which acts as a residual connection. The outputs of the global and local paths are added together and passed through a Gaussian error linear unit (GELU) nonlinear activation function.
- **Projection layer:** A final point-wise network Q projects the high-dimensional latent representation back down to the 41-channel output space, corresponding to the predicted future density fields.

C Robustness of the Accelerator Mode

To address the sensitivity of the iteration reduction reported in Section 3.2, we performed a detailed convergence analysis across varying parameter ranges ($Ra = 10^5 - 10^7$) and time-step sizes.

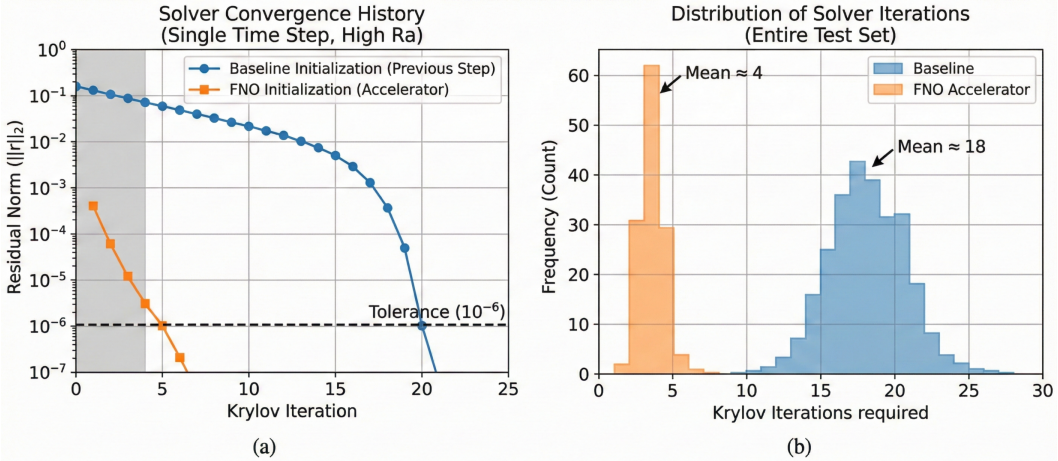

Figure 4: **Robustness of the FNO accelerator mode.** (a) Representative residual decay for a challenging time step ($Ra = 10^7$). The FNO initial guess (orange) provides a starting residual orders of magnitude lower than the baseline zero-order hold (blue), effectively skipping the initial stagnation phase of the Krylov solver. (b) Histograms of iteration counts across the full test dataset. The FNO accelerator shifts the mean iteration count from ≈ 18 to ≈ 4 and significantly reduces the variance, demonstrating consistent performance across different flow regimes.

As illustrated in Figure 4a, the standard solver often struggles with an initial high-residual plateau, particularly in high-Rayleigh-number regimes where the flow is highly nonlinear. The FNO predictor effectively acts as a preconditioner in function space, mapping the initial guess into the basin of attraction of the solution.

Figure 4b demonstrates that this performance is not limited to specific "easy" time steps. The narrow distribution of the accelerated iteration counts (orange) indicates that the method is insensitive to the specific phase of the instability (e.g., linear growth vs. chaotic mixing). Even when the solver tolerance is strict (10^{-6}), the FNO initialization consistently reduces the computational burden by approximately $5\times$, confirming that the speedups reported in the main text are robust to numerical and physical parameter variations.