CONTEXT-FNODE: CONTEXT-AWARE FRACTIONAL NEURAL ORDINARY DIFFERENTIAL EQUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

While standard Neural Ordinary Differential Equations (ODEs) excel at modeling continuous dynamics, their strictly Markovian nature limits their capacity to capture the macroscopic long-tail memory inherent in physical kinematics. Existing fractional neural models address this historical inertia but restrictively assume isotropic scalar operators and fully autonomous, closed environments. To overcome these limitations, we propose the Context-aware Fractional Neural ODE (Context-fNODE), a unified mixed-order architecture. Context-fNODE replaces commensurate scalar operators with an incommensurate, vectorized Grünwald-Letnikov derivative to natively capture direction-specific anisotropic inertia. This framework is mathematically coupled with a co-evolving first-order latent environment to absorb unobservable exogenous perturbations. Theoretically, we prove this formulation bounds the spatial error manifold, strictly suppresses chaotic Lyapunov divergence, and preserves a linear numerical convergence rate. Extensive evaluations on chaotic dynamic systems and complex 3D kinematics demonstrate that Context-fNODE significantly outperforms state-of-the-art baselines, establishing superior long-horizon forecasting endurance in non-autonomous, anisotropic environments.

1 INTRODUCTION

Accurate trajectory forecasting is shifting from 2D planar tracking (Yao et al., 2024b; Peng et al., 2024; Mohamed et al., 2022; Liao et al., 2024) to unconstrained 3D kinematics for autonomous vehicles (Zhou et al., 2025; Lu & Qian, 2024). This exposes limitations in continuous-time Neural Ordinary Differential Equations (Neural ODEs) (Chen et al., 2018). Constrained by a Markovian assumption, first-order ODEs predict future states based solely on the present, failing to capture the physical inertia and long-tail memory critical to 3D dynamics (Strogatz, 2018). To incorporate non-local memory, recent architectures leverage Fractional Differential Equations (FDEs) (Podlubny, 1998), utilizing power-law integrals to integrate historical states, occasionally adding adaptive weighting (Kang et al., 2026; Cui et al., 2025).

However, applying existing fractional frameworks to 3D kinematics reveals three bottlenecks. First, an *isotropic limitation*: imposing a single scalar fractional order (Kang et al., 2026) deteriorates in anisotropic 3D dynamics (e.g., vertical drag vs. horizontal drift). Second, a *non-autonomous gap*: assuming a fully autonomous state space (Coelho et al., 2025) neglects unobservable exogenous forces (e.g., turbulence). This violates the ODE uniqueness theorem, triggering predictive divergence (Dupont et al., 2019; Rubanova et al., 2019). Finally, a *numerical discretization gap*: Caputo or Riemann-Liouville operators impose quadratic integration complexity (Kang et al., 2025) and lack native mapping to discrete sequential data, forcing unstable heuristic truncations that compound errors over long horizons.

To resolve these constraints, we introduce **Context-aware Fractional Neural ODE (Context-fNODE)**. It couples a vectorized fractional operator (capturing anisotropic inertia) with a co-evolving latent environment (absorbing exogenous perturbations). By leveraging the Grünwald-Letnikov (GL) derivative, Context-fNODE provides an exact, numerically stable mapping to discrete data, bypassing the compounding errors of standard operators. Our contributions are:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- We formulate a vectorized GL derivative that autonomously discovers distinct fractional orders, natively resolving 3D anisotropic kinematics without hardcoded biases.
- To bridge the non-autonomous gap, we develop a mixed-order dynamical system coupling the fractional-order spatial trajectory (endogenous inertia) with a first-order latent context that absorbs exogenous noise.
- We theoretically prove that this formulation bounds the spatial error manifold via the Mittag-Leffler function, suppresses chaotic Lyapunov divergence, and preserves a linear numerical convergence rate.
- Empirically, Context-fNODE demonstrates pronounced predictive advantages on chaotic dynamical systems, validating the mixed-order dynamics. Furthermore, it exhibits high robustness in noisy kinematic environments, achieving highly competitive long-horizon 3D predictions on UAV and UUV datasets against advanced Transformers, standard Neural ODEs, and scalar fractional variants.

2 Related Work

While discrete-time sequence models (e.g., Transformers, LSTMs) dominate general forecasting, their strict reliance on fixed temporal discretization introduces severe truncation errors when applied to continuous physical kinematics (Rubanova et al., 2019). Neural Ordinary Differential Equations (Neural ODEs) (Chen et al., 2018) elegantly resolve this discretization gap by modeling continuous vector fields. However, standard integer-order ODEs enforce a strict Markovian assumption, evaluating future states based solely on the present and thereby discarding the historical inertia critical to physical motion (Strogatz, 2018). To break this Markovian bottleneck, recent architectures integrate fractional calculus to accumulate historical states. Notably, frameworks like FADE (Kang et al., 2026) and variable-order networks (Cui et al., 2025) utilize fractional derivatives with adaptive weighting to capture long-tail memory. A comprehensive review of data-driven trajectory forecasting and physics-informed continuous dynamics is provided in Appendix A.

Despite these breakthroughs, existing fractional architectures exhibit two critical structural limitations in 3D environments: (1) an *isotropic constraint*, applying a scalar fractional order across all spatial dimensions that fails to capture direction-specific dynamics (e.g., horizontal momentum vs. vertical gravity), and (2) an *autonomous assumption*, relying solely on observable states and neglecting unobservable exogenous forces. Our proposed Context-fNODE directly addresses these gaps by intertwining a co-evolving latent context with a vectorized fractional operator.

3 Preliminaries and Motivation

Fractional calculus provides a framework for modeling continuous, non-Markovian dynamics. Standard Neural ODEs (Chen et al., 2018) govern a trajectory $\mathbf{p}(t) \in \mathbb{R}^3$ via a first-order derivative $\dfrac{d\mathbf{p}(t)}{dt} = f_\theta(\mathbf{p}(t))$, enforcing a memoryless Markovian assumption and discarding historical inertia. Fractional extensions address this by substituting the integer derivative with a fractional operator $D^\alpha \mathbf{p}(t) = f_\theta(\mathbf{p}(t))$. While integrating non-local memory, this restricts the system to an *isotropic* scalar order $\alpha \in \mathbb{R}$. Since physical forces act asymmetrically (e.g., vertical buoyancy vs. horizontal drag), a single scalar α assumes identical memory decay across all axes, failing to capture direction-specific dynamics. To resolve this anisotropic limitation, we advance the Grünwald-Letnikov (GL) derivative (Podlubny, 1998) into a *vectorized* formulation. Assigning an independent fractional order to each spatial axis via $\vec{\boldsymbol{\alpha}} = [\alpha_x, \alpha_y, \alpha_z]^\top$, we define:

$$D^{\vec{\boldsymbol{\alpha}}}\mathbf{p}(t) := \lim_{\Delta t \to 0} \Delta t^{-\vec{\boldsymbol{\alpha}}} \odot \sum_{k=0}^{\infty} \mathbf{w}_k(\vec{\boldsymbol{\alpha}}) \odot \mathbf{p}(t - k\Delta t) \tag{1}$$

where $\odot$ denotes the Hadamard product, and $\Delta t^{-\vec{\boldsymbol{\alpha}}} = [\Delta t^{-\alpha_x}, \Delta t^{-\alpha_y}, \Delta t^{-\alpha_z}]^\top$. The fading memory weights $\mathbf{w}_k(\vec{\boldsymbol{\alpha}}) \in \mathbb{R}^3$ are evaluated independently per axis to encapsulate anisotropic inertia:

$$\mathbf{w}_k(\vec{\boldsymbol{\alpha}}) = \left[(-1)^k \binom{\alpha_x}{k},\ (-1)^k \binom{\alpha_y}{k},\ (-1)^k \binom{\alpha_z}{k}\right]^\top \tag{2}$$

Proposition 1 (Universal Fading Memory Property)**.** *For any non-integer $\alpha > 0$, the scalar weight $w_k = (-1)^k \binom{\alpha}{k}$ acts as a power-law fading memory kernel with asymptotic decay $w_k = \mathcal{O}(k^{-\alpha-1})$ (explicitly $w_k \simeq \frac{1}{\Gamma(-\alpha)} k^{-\alpha-1}$) as $k \to \infty$. This ensures recent states exert dominant influence while historical states decay without vanishing across sub-diffusive ($0 < \alpha < 1$) and super-diffusive ($\alpha > 1$) regimes.*

The detailed mathematical proof is deferred to Appendix B.

Applying Proposition 1 to the vectorized operator decouples the inertial fading rates across axes. By independently evaluating the decay $\mathcal{O}(k^{-\alpha_c - 1})$ for each axis $c \in \{x, y, z\}$, the network models anisotropic memory (e.g., rapid vertical dampening vs. extended horizontal drift). Furthermore, this power-law decay justifies truncating the infinite series to a practical window W while preserving inertial fidelity, avoiding the rapid exponential decay ($\mathcal{O}(e^{-k})$) of standard recurrent networks.

While the vectorized GL operator resolves the memory deficit, relying solely on observable coordinates $\mathbf{p}(t)$ assumes an autonomous system, neglecting unobservable exogenous forces. Under the Picard-Lindelöf theorem, autonomous ODEs prohibit trajectory intersections in phase space. Forcing real-world kinematics - which intersect due to latent variables - into an autonomous formulation causes predictive divergence. This motivates our methodology: co-evolving a dynamic latent context alongside the spatial trajectory to establish a complete, context-aware dynamical system.

4 Mixed-Order Context-Aware Fractional Neural ODE

We propose the Context-aware Fractional Neural ODE (Context-fNODE) for 3D trajectory modeling. It bridges latent exogenous forces and endogenous anisotropic memory via a mixed-order coupled system using a vectorized fractional operator.

4.1 Continuous-Time Problem Formulation and Mixed-Order Dynamics

Let $s \in \mathbb{N}$ denote a discrete sequence index with timestamp $t_s = s \cdot \Delta t$. Given a historical trajectory $\mathcal{X} = \{\mathbf{p}_s \mid s = S - W + 1, \ldots, S\}$ over a look-back window W, we forecast the continuous future trajectory $\mathbf{p}(t_{>S})$ for $t_{>S} > t_S$. Unlike standard integer-order Initial Value Problems (IVPs) (Chen et al., 2018), physical trajectories interact with unobservable exogenous factors. Representing the state solely via spatial coordinates yields an incomplete state-space, triggering predictive divergence. To compensate, we extract a latent context $\lambda(t_S) \in \mathbb{R}^{d_{context}}$ from the kinematic history, acting as the initial condition for the unobservable environment: $\lambda(t_S) = \text{MLP}_{\phi_2}\big(\text{GRU}_{\phi_1}(\mathcal{X})\big)$.

We construct an augmented continuous state vector $\mathbf{z}(t_{>S}) = [\mathbf{p}(t_{>S}), \lambda(t_{>S})]^\top \in \mathbb{R}^{3+d_{context}}$ for $t_{>S} \in (t_S, \infty)$. Its co-evolution is governed by a mixed-order coupled differential equation, formulated as an IVP initialized at $\mathbf{z}(t_S)$:

$$\left[D^{\vec{\boldsymbol{\alpha}}} \mathbf{p}(t_{>S}),\ \frac{d\lambda(t_{>S})}{dt_{>S}} \right]^\top = \mathcal{N}_\theta(\mathbf{z}(t_{>S})) \tag{3}$$

where $\mathcal{N}_\theta(\mathbf{z}) = [\mathcal{N}_\theta^{(p)}(\mathbf{z}), \mathcal{N}_\theta^{(\lambda)}(\mathbf{z})]^\top$ decomposes the neural vector field. $D^{\vec{\boldsymbol{\alpha}}}$ is the vectorized GL derivative preserving spatial inertia, and $\frac{d}{dt_{>S}}$ is the first-order derivative capturing instantaneous environmental turbulence.

4.2 Numerical discretization and Autoregressive solver

To solve the IVP (Eq. 3), we transition to a discrete autoregressive framework with step size Δt. For step $s \geq S$, we adopt an explicit numerical solver, causally evaluating the neural vector field at $\mathbf{z}_s$ (Kang et al., 2026). The latent environment is integrated via a forward Euler step for contextual adaptation: $\lambda_{s+1} = \lambda_s + \Delta t \cdot \mathcal{N}_\theta^{(\lambda)}(\mathbf{z}_s)$. For spatial dynamics, we approximate the infinite-limit GL derivative via its discrete finite-difference analogue, truncated over memory window W. Equating the neural vector field's spatial projection with this approximation establishes the balance equation:

$$\mathcal{N}_\theta^{(p)}(\mathbf{z}_s) = \Delta t^{-\vec{\boldsymbol{\alpha}}} \odot \sum_{k=0}^{W} \mathbf{w}_k(\vec{\boldsymbol{\alpha}}) \odot \mathbf{p}_{s+1-k} \tag{4}$$

Since the zero-th order weight is the identity vector ($\mathbf{w}_0(\vec{\boldsymbol{\alpha}}) = \mathbf{1}$), isolating the $k = 0$ term resolves the single-step spatial state $\mathbf{p}_{s+1}$ via an autoregressive update:

$$\mathbf{p}_{s+1} = \underbrace{\Delta t^{\vec{\boldsymbol{\alpha}}} \odot \mathcal{N}_\theta^{(p)}(\mathbf{z}_s)}_{\text{Neural Vector Field Push}} - \underbrace{\sum_{k=1}^{W} \mathbf{w}_k(\vec{\boldsymbol{\alpha}}) \odot \mathbf{p}_{s+1-k}}_{\text{Endogenous Memory Damping}} \tag{5}$$

Equation 5 decouples trajectory evolution into two forces: the 'Neural Vector Field Push' provides instantaneous propulsion evaluated at state s, while 'Endogenous Memory Damping' exerts historical inertial pull from the past $(s, \ldots, s - W + 1)$. The future state $\mathbf{p}_{s+1}$ emerges from balancing active neural driving and fading inertial momentum.

To guarantee inference stability, directly evaluating $\mathbf{w}_k(\vec{\boldsymbol{\alpha}})$ via Gamma functions (which introduces numerical singularities) is avoided by pre-computing weights recursively:

$$\mathbf{w}_0(\vec{\boldsymbol{\alpha}}) = \mathbf{1}, \quad \mathbf{w}_k(\vec{\boldsymbol{\alpha}}) = \mathbf{w}_{k-1}(\vec{\boldsymbol{\alpha}}) \odot \left(\mathbf{1} - \frac{\vec{\boldsymbol{\alpha}} + \mathbf{1}}{k}\right) \quad \text{for } k \geq 1 \tag{6}$$

4.3 Optimization Strategy

Training fractional continuous models over extended horizons suffers from compounding truncation errors and exposure bias. To decouple trajectory forecasting fidelity from structural stability, we introduce a time-weighted robust objective and a topology-aware optimization regime.

4.3.1 Robust Loss Formulation with Structural Prior

Let M denote the prediction horizon length, and $\mathbf{e}_s = \hat{\mathbf{p}}_s - \mathbf{p}_s$ the spatial error at future step $s \in \{S + 1, \ldots, S + M\}$. To penalize long-term drift while bounding exploding gradients from outliers, we use the Huber loss (Huber, 1964). For scalar error e and threshold δ:

$$\mathcal{H}(e; \delta) = \begin{cases} \frac{1}{2}e^2 & \text{for } |e| \leq \delta \\ \delta(|e| - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \tag{7}$$

Applying an exponentially increasing temporal weight $w_s = \exp(\beta \cdot (s - S))$ $(\beta > 0)$ forces the network to prioritize global trajectory endurance over short-term curve fitting. Furthermore, optimizing the fractional order $\vec{\boldsymbol{\alpha}}$ over unconstrained $\mathbb{R}^3$ risks structural collapse. To prevent this, we introduce a structural prior penalty. The unified loss function is:

$$\mathcal{L}_{total}(\theta, \vec{\boldsymbol{\alpha}}) = \underbrace{\frac{1}{M} \sum_{s=S+1}^{S+M} w_s \sum_{c \in \{x,y,z\}} \mathcal{H}\left(e_{s,c}; \delta\right)}_{\text{Kinematic Drift Penalty}} + \underbrace{\lambda_{prior} \|\vec{\boldsymbol{\alpha}} - \vec{\boldsymbol{\alpha}}_{prior}\|_2^2}_{\text{Structural Prior Regularization}} \tag{8}$$

where $e_{s,c}$ is the error component along axis c at step s, and λ_{prior} dictates regularization strength.

Here, $\vec{\boldsymbol{\alpha}}_{prior}$ is a structural prior preventing fractional orders ($\vec{\boldsymbol{\alpha}}$) from drifting into unstable regions or deviating from expected baseline dynamics. Its selection depends on dataset properties:

- **Chaotic dynamical systems (Lorenz, Rössler, Thomas):** Since their theoretical governing equations are first-order ODEs, we set $\vec{\boldsymbol{\alpha}}_{prior} = [1.0, 1.0, 1.0]^\top$. The regularization weight (λ_{prior}) is tuned via Bayesian Optimization.
- **3D kinematic datasets (UAV and UUV):** Since exact fractional decay rates are unknown, we set $\lambda_{prior} = 0$. Thus, $\vec{\boldsymbol{\alpha}}_{prior}$ is not actively used, and fractional orders $\vec{\boldsymbol{\alpha}}$ are learned directly from data without structural constraints.

This separation ensures the network minimizes spatial drift while anchoring the memory operator to baseline physical bounds when applicable.

4.3.2 TOPOLOGY-AWARE SYNCHRONIZED PARAMETER UPDATE

Optimizing the framework requires decoupling the learning pace of the neural vector field from the fractional memory kernel, while aligning gradient descent with the physical topology.

Neural parameters θ (encompassing $\mathcal{N}_\theta^{(p)}$ and $\mathcal{N}_\theta^{(\lambda)}$) are updated via standard optimizers (e.g., Adam) with learning rate η_θ. Conversely, fractional orders $\vec{\boldsymbol{\alpha}}$ are updated via a synchronized gradient descent mechanism to avoid hardcoded biases:

$$\theta \leftarrow \theta - \eta_\theta \nabla_\theta \mathcal{L}_{total}, \quad \vec{\boldsymbol{\alpha}} \leftarrow \vec{\boldsymbol{\alpha}} - \eta_\alpha \mathcal{S}\big(\nabla_{\vec{\boldsymbol{\alpha}}} \mathcal{L}_{total}\big) \tag{9}$$

Here, η_α is the fractional learning rate, and $\mathcal{S}(\cdot)$ is a gradient synchronization operator enforcing spatial kinematic constraints. For symmetric systems, it averages directional gradients: $\mathcal{S}(\mathbf{g}) = \big(\frac{1}{3}\sum_c g_c\big)\,\mathbf{1}$. This couples the optimization of all axes, enforcing isotropic inertia. When disabled ($\mathcal{S}(\mathbf{g}) = \mathbf{g}$), gradients flow independently, enabling unconstrained anisotropic discovery where the network learns varying fading rates for distinct dimensions. By configuring the tuple $(\mathcal{S}, \lambda_{prior}, \eta_\alpha)$, the framework adapts to diverse physical topologies - from isotropic rigid-body kinematics to anisotropic underwater hydrodynamics and chaotic attractors.

4.4 ASYMPTOTIC STABILITY AND NUMERICAL CONVERGENCE

The detailed proofs for all theorems, lemmas, and propositions are deferred to Appendix B.

Before analyzing the prediction error, it is mathematically imperative to guarantee that the underlying mixed-order continuous system is fundamentally well-defined and solvable.

Proposition 2 (Existence and Uniqueness)**.** *Assuming a globally Lipschitz continuous neural vector field $\mathcal{N}_\theta$ with Lipschitz constant $L > 0$, the continuous mixed-order dynamical system admits a unique, well-posed solution for any initialized boundary state $\mathbf{z}(t_S)$.*

To establish rigorous error bounds for the vectorized fractional formulation, let $\alpha_{min} = \min(\alpha_x, \alpha_y, \alpha_z)$ denote the dominant fractional order governing the slowest inertial decay.

Theorem 1 (Fractional Error Bounding)**.** *Let $\mathbf{e}(t_{>S}) = \hat{\mathbf{z}}(t_{>S}) - \mathbf{z}(t_{>S})$ define the state prediction error over the continuous future horizon $t_{>S} > t_S$. Assuming the neural vector field $\mathcal{N}_\theta$ is globally Lipschitz continuous with constant $L > 0$, the error magnitude for the fractional system ($\mathbf{0} < \vec{\boldsymbol{\alpha}} \leq \mathbf{2}$) is bounded by the generalized Mittag-Leffler function $E_{\alpha_{min}}$:*

$$\|\mathbf{e}(t_{>S})\| \leq \|\mathbf{e}(t_S)\| E_{\alpha_{min}}\big(L(t_{>S} - t_S)^{\alpha_{min}}\big) \tag{10}$$

Building upon this bounded error manifold, we evaluate the framework's inherent robustness against chaotic divergence - a notorious failure point for strictly Markovian integer-order ODEs.

Theorem 2 (Asymptotic Stability against Lyapunov Divergence)**.** *In continuous dynamical systems, the maximal separation rate of nearby trajectories is structurally bounded by the Lipschitz constant L of the driving vector field. By substituting integer-order dynamics with the fractional integral operator, the standard linear exponential divergence ($\mathcal{O}(e^{L(t_{>S}-t_S)})$) is replaced by a sub-exponential generalized Mittag-Leffler bound. Specifically, under the regime where $L > 0$ and $0 < \alpha_{min} < 1$, this chaotic suppression holds analytically, modulating the asymptotic divergence envelope to* $\exp\big(L^{1/\alpha_{min}}(t_{>S} - t_S)\big)$.

Transitioning from continuous theoretical dynamics to the discrete numerical implementation, we must formally quantify the localized inaccuracies introduced by temporal discretization and finite-window memory constraints.

Lemma 1 (Operator and Memory Truncation Error)**.** *Assuming uniformly bounded spatial coordinates ($\|\mathbf{p}\| \leq P_{max}$) and C^2-continuous trajectories, the forward Euler integration of the latent environment yields a local truncation error of $\varepsilon_{\lambda,s} = \mathcal{O}(\Delta t^2)$. Concurrently, the truncated Grünwald-Letnikov operator yields a strictly bounded local spatial error $\varepsilon_{\mathbf{p},s} = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$.*

Remark 1 (Exact Mapping vs. Truncation Error)**.** *The term "exact mapping" refers to the structural property of the Grünwald-Letnikov (GL) derivative. Unlike Riemann-Liouville or Caputo operators that inherently require numerical quadrature approximations, GL is defined as a discrete infinite sum, allowing a natural, causal mapping to sequential data without discretization reformulation.*

The $\mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$ bounds in Lemma 1 reflect practical computational limits: the finite sampling rate ($\mathcal{O}(\Delta t)$) and the forced truncation of the infinite sum to a computationally tractable finite memory window ($\mathcal{O}(W^{-\alpha_{min}})$).

Finally, by accumulating these local step penalties across the entire forecasting sequence, we establish the global stability and convergence guarantees for the proposed autoregressive solver.

Theorem 3 (Global Convergence Criterion and Rate)**.** *Let M be the discrete forecasting horizon, corresponding to a maximum continuous integration time $T_f = M\Delta t$. The autoregressive solver globally converges to the true continuous trajectory as $\Delta t \to 0$ and $W \to \infty$. The global accumulated state error follows $\|\mathbf{E}_M\| = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$. Assuming a sufficiently large memory window W that strictly dominates the fractional truncation penalty ($W \gg \Delta t^{-1/\alpha_{min}}$), the framework achieves a linear global convergence rate of $\mathcal{O}(\Delta t)$ for any valid fractional order $\vec{\boldsymbol{\alpha}} > \mathbf{0}$.*

These results bridge continuous dynamical stability with discrete numerical reliability. Theoretically, our well-posed mixed-order system (Proposition 2) suppresses chaotic divergence (Theorems 1, 2), extending predictability horizons. Algorithmically, the finite-memory solver tracks this target (Lemma 1), guaranteeing linear global convergence without exponential error blowup (Theorem 3). Consequently, the framework's long-term forecasting endurance is a mathematically guaranteed consequence of our context-aware fractional formulation, not merely an empirical observation.

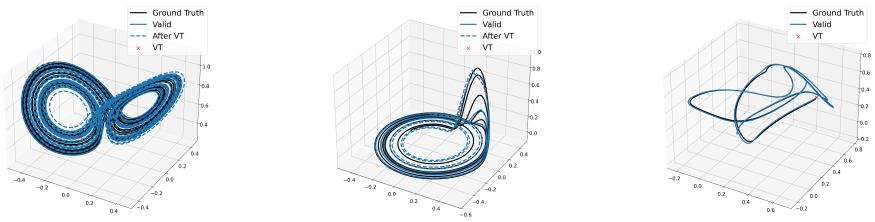
5 Experiment

We evaluate Context-fNODE across three distinct physical domains: rigid-body kinematics (UAVs), fluid dynamics (UUVs), and deterministic chaos. Our experiments highlight two core architectural capabilities: the co-evolving latent context's adaptability to unobservable perturbations, and the fractional derivative's capacity to preserve historical inertia and suppress long-horizon drift. Benchmarking against state-of-the-art baselines, we assess both quantitative accuracy and the topological preservation of dynamic trajectories. Extended qualitative rollouts are provided in Appendix E.

5.1 Chaotic Dynamics

Table 1: Performance across dynamical systems with Mean Squared Error (MSE) and Time Valid (T_v) (mean $\pm$ std). Mean and std values of T_v are rounded to the nearest integer.

	Lorenz63		Rössler		Thomas	
Model	MSE ↓	T_v ↑	MSE ↓	T_v ↑	MSE ↓	T_v ↑
DLinear (Zeng et al., 2023)	0.0358 ± 0.0016	24 ± 1	0.0012 ± 0.0012	498 ± 61	0.0198 ± 0.0139	330 ± 47
LSTM (Hochreiter & Schmidhuber, 1997)	0.0004 ± 2.9e-05	545 ± 112	0.0002 ± 0.0002	1127 ± 468	0.0112 ± 0.0017	466 ± 36
Informer (Zhou et al., 2021)	0.0883 ± 0.0277	117 ± 6	0.0004 ± 0.0003	497 ± 8	0.0026 ± 0.0017	1305 ± 126
NeuralODE (Chen et al., 2018)	0.0113 ± 0.0083	365 ± 128	2.4e-05 ± 1.4e-05	1720 ± 6	3.5e-05 ± 9.3e-06	3956 ± 83
Transformer (Vaswani et al., 2017)	0.0177 ± 0.0187	127 ± 4	0.0006 ± 0.0006	528 ± 22	0.0016 ± 0.0007	1137 ± 80
ESN (Jaeger & Haas, 2004)	0.0435 ± 0.0148	368 ± 34	7.7e-05 ± 2.1e-05	1414 ± 129	**1.1e-05** ± 6.0e-06	4353 ± 135
FADE (Kang et al., 2026)	0.0166 ± 0.0121	385 ± 33	8.9e-05 ± 5.5e-05	1695 ± 5	2.0e-05 ± 1.1e-05	3757 ± 226
Context-fNODE-scalar	0.0192 ± 0.0112	366 ± 3	6.7e-05 ± 1.8e-05	1700 ± 6	3.9e-05 ± 2.2e-05	4112 ± 169
Ours	**1.7e-05** ± 1.6e-05	**775** ± 115	**2.1e-05** ± 2.0e-05	**1727** ± 13	1.8e-05 ± 1.8e-05	**5000** ± 0



(a) Lorenz (b) Rössler (c) Thomas

Figure 1: 3D trajectory comparison between ground truth and predictions

Table 1 shows our architecture achieves superior performance across three chaotic benchmark systems. It attains the lowest MSE on Lorenz63 and Rössler, and the second-lowest on Thomas, indicating exceptional short-term predictive accuracy. Notably, it achieves the highest valid time T_v across all tasks, demonstrating its capability to maintain accurate trajectories over the longest horizons. Figure 1 visualizes this stability, showing tight alignment with the true attractors. Among baselines, ESNs and Neural ODEs perform strongly. This is expected: ESNs excel in chaotic forecasting (Lukoševičius & Jaeger, 2009), while Neural ODEs directly model the continuous vector field of these deterministic systems rather than approximating discrete transitions.

5.2 UAV Flight Trajectories dataset

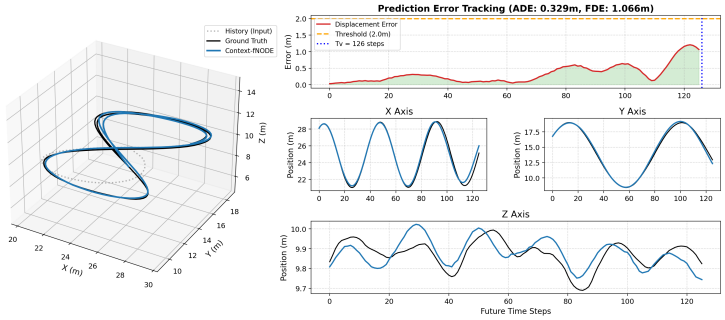
Table 2 benchmarks our model against temporal forecasters and trajectory predictors. Our architecture achieves the lowest ADE (**1.52m**) and RMSE (**1.73m**), reducing spatial deviations by nearly 50% compared to foundational baselines. Crucially, it delivers the highest forecasting endurance with a Valid Time of $T_v = \mathbf{84}$ **steps** (at a strict 2.0m divergence threshold). While SCA-LSTM yields a marginally lower FDE (2.10m vs. 2.17m), our dominant T_v proves that Context-fNODE maintains physical reliability for a significantly longer duration. These results validate our physics-encoded architecture over purely data-driven attention mechanisms. The fractional-order derivative captures long-term inertia, preventing the memory decay seen in integer-order ODEs. Concurrently, our structural prior ($\alpha_x = \alpha_y = \alpha_z$) ensures all 3D coordinates remain kinematically coupled, suppressing autoregressive divergence.

Table 2: Comparison on the UAV Flight Trajectories dataset (Mean $\pm$ Std across 5 seeds).

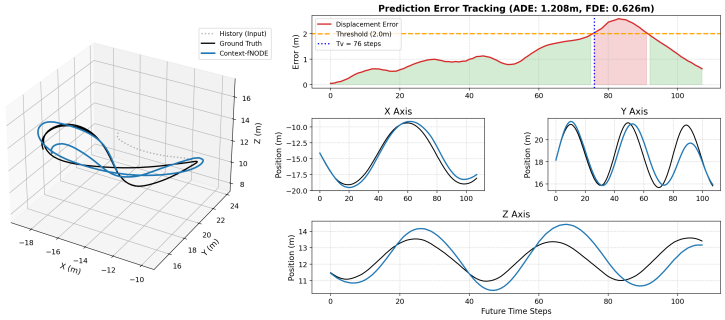
Model	ADE (m) ↓	RMSE (m) ↓	FDE (m) ↓	T_v (steps) ↑
LSTM (Hochreiter & Schmidhuber, 1997)	2.92 ± 0.04	3.41 ± 0.04	4.17 ± 0.05	41 ± 0
Informer (Zhou et al., 2021)	2.56 ± 0.03	2.89 ± 0.04	3.37 ± 0.04	42 ± 1
PatchTST (Nie et al., 2023)	2.71 ± 0.04	3.07 ± 0.04	3.42 ± 0.04	41 ± 0
DLinear (Zeng et al., 2023)	2.66 ± 0.04	2.99 ± 0.04	2.93 ± 0.04	29 ± 0
NeuralODE (Chen et al., 2018)	2.88 ± 0.04	3.09 ± 0.04	2.84 ± 0.04	49 ± 1
TrajCLIP (Yao et al., 2024a)	2.02 ± 0.10	2.34 ± 0.10	3.10 ± 0.13	67 ± 2
MS-TIP (singh chib et al., 2024)	1.84 ± 0.09	2.10 ± 0.09	2.51 ± 0.11	70 ± 3
MRGTraj (Peng et al., 2024)	1.80 ± 0.09	2.07 ± 0.09	2.43 ± 0.10	69 ± 2
S-Implicit (Mohamed et al., 2022)	1.63 ± 0.08	1.88 ± 0.08	2.46 ± 0.10	80 ± 3
SCA-LSTM (Zhou et al., 2025)	1.58 ± 0.07	1.80 ± 0.08	$\mathbf{2.10} \pm 0.09$	77 ± 3
FADE (Kang et al., 2026)	4.97 ± 0.21	5.65 ± 0.42	7.35 ± 0.77	21 ± 0
Context-fNODE-scalar	1.58 ± 0.09	1.80 ± 0.12	2.24 ± 0.16	80 ± 4
Context-fNODE (Ours)	$\mathbf{1.52} \pm 0.07$	$\mathbf{1.73} \pm 0.08$	$\underline{2.17} \pm 0.09$	$\mathbf{84} \pm 3$

Extrapolation and Exposure Bias: Noticeably, while the model is trained with a prediction horizon of $M = 20$ steps, it is evaluated on extended long-horizon rollouts of up to 175 steps. This capacity to extrapolate is an intentional feature demonstrating that Context-fNODE learns the true continuous underlying vector field rather than overfitting to a sequence mapping. Our model does not suffer from exposure bias because we avoid Teacher Forcing. During training, it operates in a purely closed-loop autoregressive manner, recursively feeding its own predicted states into the sliding window. Furthermore, our exponentially increasing temporal weight in the Huber loss heavily penalizes compounding autoregressive drift, forcing the system to stabilize early.

Figure 2 visually corroborates how the model's continuous fractional memory preserves momentum across diverse flight profiles. In Figure 2a, a smooth, near-planar figure-8 maneuver demonstrates exceptional long-horizon stability. Empowered by the co-evolving latent context that absorbs unobservable environmental dynamics, the predicted trajectory strictly follows the ground truth, securing a highly accurate ADE of 0.329m and completing the entire 126-step forecasting horizon without a single threshold breach. Figure 2b illustrates a more demanding scenario: a complex, multi-dimensional helical loop. Despite the coupled non-linear transitions across all three spatial axes, our rigid-body structural prior prevents individual coordinate decoupling. This allows the model to sustain highly reliable tracking for 76 steps. Even when uncertainty naturally compounds over this extended horizon, the eventual deviation manifests as a physically plausible, smooth curvature on the outer edge of a sharp turn, highlighting the framework's inherent kinematic stability.



(a) Stable, near-planar figure-8 flight profile demonstrating high spatial fidelity without divergence.



(b) Complex 3D helical maneuver exhibiting coupled multidimensional transitions.

Figure 2: Qualitative 3D trajectory rollouts of our proposed model on the UAV dataset.

5.3 3DMOTRAJ - UUV TRAJECTORIES

To further demonstrate the framework's adaptability across diverse physical domains, we evaluate Context-fNODE on the 3DMoTraj dataset, which captures the complex underwater kinematics of UUVs. Unlike aerial flights, aquatic environments impose severe, highly non-linear hydrodynamic drag and unpredictable currents that continuously perturb the vehicle's trajectory.

Table 3: Global performance on the 3DMoTraj UUV dataset (Mean ± Std across 5 seeds).

Model	**ADE** ↓	**FDE** ↓
LSTM (Hochreiter & Schmidhuber, 1997)	0.63 ± 0.134	1.02 ± 0.147
Informer (Zhou et al., 2021)	0.67 ± 0.015	1.05 ± 0.034
PatchTST (Nie et al., 2023)	0.65 ± 0.018	1.22 ± 0.013
DLinear (Zeng et al., 2023)	0.97 ± 0.012	1.85 ± 0.041
NeuralODE (Chen et al., 2018)	0.75 ± 0.009	1.26 ± 0.013
TrajCLIP (Yao et al., 2024a)	0.72 ± 0.021	1.36 ± 0.038
MS-TIP (singh chib et al., 2024)	0.70 ± 0.019	1.30 ± 0.029
MRGTraj (Peng et al., 2024)	0.68 ± 0.023	1.36 ± 0.045
S-Implicit (Mohamed et al., 2022)	0.66 ± 0.016	1.23 ± 0.021
SCA-LSTM (Zhou et al., 2025)	0.58 ± 0.014	1.01 ± 0.026
FADE (Kang et al., 2026)	0.92 ± 0.070	1.49 ± 0.164
Context-fNODE-scalar	0.58 ± 0.017	0.96 ± 0.022
Context-fNODE (Ours)	$\mathbf{0.54 \pm 0.008}$	$\mathbf{0.86 \pm 0.034}$

As detailed in Table 3, our proposed architecture establishes a new state-of-the-art on this domain-generalized benchmark, achieving a global mean ADE/FDE of **0.54/0.86 m**. It significantly outperforms the most competitive trajectory-specific baseline, SCA-LSTM, yielding a 15.7% reduction in Final Displacement Error (0.86m vs. 1.02m). Crucially, these empirical results explicitly validate the two core mathematical innovations of Context-fNODE:

- **The Necessity of Latent Context:** The performance gap between our model and standard fractional operators (e.g., FADE, which degrades to an FDE of 1.49m) underscores the critical importance of the co-evolving latent environment. While purely autonomous models are highly vulnerable to unobservable underwater turbulence, our context network dynamically abstracts generalizable fluid-dynamic perturbations, avoiding overfitting to localized noise and maintaining robust zero-shot transferability across novel aquatic topographies.

- **The Superiority of Anisotropic Inertia:** The underwater setting perfectly illustrates why scalar fractional assumptions fail in physical realities. Compared to our own isotropic variant (Context-fNODE-scalar, FDE 0.96m), the fully anisotropic Context-fNODE further reduces long-horizon drift by over 10%. This confirms that enforcing a single scalar memory order is fundamentally limiting when dealing with incommensurate physical forces - such as severe horizontal drag acting against vertical buoyancy. Allowing the autonomous discovery of axis-specific memory decay is therefore vital for real-world kinematic fidelity.

5.4 Window-size Sensitivity

Further detailed analyses (Appendix D) demonstrate the model's consistent robustness across all eight individual UUV scenes and validate our hyperparameter search. The choice of the memory window W dictates a critical trade-off. A smaller window (e.g., $W = 32$) prematurely truncates the Grünwald-Letnikov fractional operator, discarding the macroscopic historical inertia necessary to stabilize non-linear dynamics and resulting in earlier predictive divergence. Conversely, expanding the window excessively (e.g., $W = 72$) yields diminishing returns. Because the fractional fading memory weights follow a power-law decay, overly distant past states contribute minimal inertial signal; instead, they introduce irrelevant historical noise and unnecessarily exacerbate the linear $\mathcal{O}(W)$ computational overhead. Consequently, $W = 64$ emerges as the optimal truncation threshold, perfectly balancing the retention of essential physical momentum against computational efficiency and noise suppression.

6 Conclusion

In this paper, we propose the Context-aware Fractional Neural Ordinary Differential Equation (Context-fNODE), a continuous-depth framework integrating fractional calculus with dynamic structural priors. By adaptively modulating its disentangled fractional orders, the model seamlessly transitions from enforcing isotropic gradient synchronization for rigid-body aerodynamics to capturing physically incommensurate spatial dynamics in fluid environments. This adaptive mechanism ensures strict physical consistency and mitigates long-term trajectory drift.

Extensive evaluations confirm the efficacy of Context-fNODE across three complex domains. The framework exhibits its most pronounced performance gains on non-linear chaotic dynamics, theoretically validating the fractional operator's ability to minimize truncation error. Furthermore, it demonstrates robust applicability to noisy real-world kinematic environments, achieving highly competitive long-horizon predictive performance on 3D UAV flight trajectories and UUV underwater kinematics. Ultimately, this work establishes a robust, generalized paradigm for continuous-time trajectory forecasting in diverse dynamical systems.

Limitations Context-fNODE has two main limitations. First, standard Neural ODEs only need the current state ($\mathcal{O}(1)$ memory). Our fractional solver instead keeps a sliding window of past states, so memory and time are $\mathcal{O}(W)$ per integration step (Appendix C.5). The model needs this history to capture anisotropic inertia, but the extra latency makes real-time use harder on resource-constrained edge devices. Second, the Grünwald–Letnikov operator integrates over a fixed historical window, so it acts as a low-pass filter. This stabilizes long-horizon predictions against noise, but it also resists abrupt, unobservable control changes such as sudden UAV waypoint shifts or sharp UUV dives, and it causes an adaptive lag when a shock has no precursor in the history (Appendix E). Future work will look at attention-modulated fractional integration, which would shrink the effective window when the context changes quickly.

AI USE STATEMENT

In this work, we used generative AI tools solely to improve the manuscript's clarity, grammar, and overall readability. We have not used generative AI tools for conceptual or theoretical development, research design, implementation, experimentation, data analysis, or interpretation of results. The authors have reviewed all AI-assisted edits to ensure that they reflected the intended meaning and did not alter the scientific content, and take full responsibility for the final manuscript.

ETHICS STATEMENT

In this work, we used publicly available UAV/UUV trajectory datasets and datasets for chaotic dynamical systems. No human participants or personal or sensitive data were involved. We are not aware of any additional ethical concerns associated with this work.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide the necessary methodological and experimental details throughout the paper and supplementary material. Detailed descriptions of the model architectures are provided in the main text. Appendix C provides details on dataset generation, sources of publicly available datasets, data processing, and experimental setups. Detailed proofs of the theoretical analysis are provided in Appendix B. The complete implementation is provided as supplementary material.

REFERENCES

Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971, 2016. doi: 10.1109/CVPR.2016.110.

Ricardo Almeida. A caputo fractional derivative of a function with respect to another function. *Communications in Nonlinear Science and Numerical Simulation*, 44:460–481, 2017. ISSN 1007-5704. doi: 10.1016/j.cnsns.2016.09.006.

Suresh Bishnoi, Jayadeva Jayadeva, Sayan Ranu, and N M Anoop Krishnan. BroGNet: Momentum-conserving graph neural stochastic differential equation for learning brownian dynamics. In *The Twelfth International Conference on Learning Representations*, 2024.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 6572–6583, Red Hook, NY, USA, 2018.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Deep Learning and Representation Learning Workshop*, 2014.

C. Coelho, M. Fernanda P. Costa, and L.L. Ferrás. Neural fractional differential equations. *Applied Mathematical Modelling*, 144:116060, 2025. ISSN 0307-904X. doi: 10.1016/j.apm.2025.116060.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2019.

Wenjun Cui, Qiyu Kang, Xuhao Li, Kai Zhao, Wee Peng Tay, Weihua Deng, and Yidong Li. Neural variable-order fractional differential equation networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15):16109–16117, 2025. doi: 10.1609/aaai.v39i15.33769.

K. Diethelm. *The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2010. ISBN 9783642145742.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019.

Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11522–11530, 2020. doi: 10.1109/CVPR42600.2020.01154.

Rudolf Gorenflo, Anatoly A. Kilbas, Francesco Mainardi, and Sergei V. Rogosin. *Mittag-Leffler Functions, Related Topics and Applications*. Springer Monographs in Mathematics. Springer Berlin, Heidelberg, 2014. ISBN 978-3-662-43930-2.

Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2255–2264, 2018. doi: 10.1109/CVPR.2018.00240.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Alice Inbaraj and Jun Chen. Physics-informed neural models for uncertain trajectory prediction in urban air mobility. *Journal of Aerospace Information Systems*, 23(4):337–345, 2026. doi: 10.2514/1.I011720.

Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304:78–80, 2004.

Ruochen Jiao, Yixuan Wang, Xiangguo Liu, Simon Sinong Zhan, Chao Huang, and Qi Zhu. Kinematics-aware trajectory generation and prediction with latent stochastic differential modeling. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 565–572, 2024. doi: 10.1109/IROS58592.2024.10802438.

Qiyu Kang, Xuhao Li, Kai Zhao, Wenjun Cui, Yanan Zhao, Weihua Deng, and Wee Peng Tay. Efficient training of neural fractional-order differential equation via adjoint backpropagation. In *Proc. AAAI Conference on Artificial Intelligence*, Philadelphia, USA, 2025.

Qiyu Kang, Wenjun Cui, Xuhao Li, Yuxin Ma, Xueyang Fu, Wee Peng Tay, Yidong Li, and Zheng-Jun Zha. Neural fractional attention differential equations. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

Patrick Kidger, Ricky T. Q. Chen, and Terry Lyons. "Hey, that's not an ODE": Faster ODE Adjoints via Seminorms. *International Conference on Machine Learning*, 2021.

A.A.A. Kilbas, H.M. Srivastava, and J.J. Trujillo. *Theory And Applications of Fractional Differential Equations*. North-Holland Mathematics Studies. Elsevier Science & Tech, 2006. ISBN 9780444518323.

Zhilu Lai, Charilaos Mylonas, Satish Nagarajaiah, and Eleni Chatzi. Structural identification with physics-informed neural ordinary differential equations. *Journal of Sound and Vibration*, 508: 116196, 2021. ISSN 0022-460X. doi: 10.1016/j.jsv.2021.116196.

Haicheng Liao, Chengyue Wang, Zhenning Li, Yongkang Li, Bonan Wang, Guofa Li, and Chengzhong Xu. Physics-informed trajectory prediction for autonomous driving under missing observation. In Kate Larson (ed.), *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 6841–6849. International Joint Conferences on Artificial Intelligence Organization, 8 2024. doi: 10.24963/ijcai.2024/756. URL 10.24963/ijcai.2024/756. Main Track.

Shaoning Lu and Yue Qian. Enhanced trajectory forecasting for hypersonic glide vehicle via physics-embedded neural ode. *Drones*, 8(8), 2024. ISSN 2504-446X. doi: 10.3390/drones8080377.

Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer science review*, 3(3):127–149, 2009.

Ralf Metzler and Joseph Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77, 2000. ISSN 0370-1573. doi: 10.1016/S0370-1573(00)00070-3.

Nathaniel E. Michek, Piyush Mehta, and Wade W. Huebsch. Flight dynamic uncertainty quantification modeling using physics-informed neural networks. *AIAA Journal*, 62(11):4234–4246, 2024. doi: 10.2514/1.J063992.

Abduallah Mohamed, Deyao Zhu, Warren Vu, Mohamed Elhoseiny, and Christian Claudel. Social-implicit: Rethinking trajectory prediction evaluation and the effectiveness of implicit maximum likelihood estimation. In *European Conference on Computer Vision (ECCV)*, pp. 463–479. Springer, 2022. doi: 10.1007/978-3-031-20047-2_27.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

K.B. Oldham and J. Spanier. *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order*. Dover books on mathematics. Dover Publications, 2006. ISBN 9780486450018.

Guofei Pang, Lu Lu, and George Em Karniadakis. fPINNs: Fractional physics-informed neural networks. *SIAM Journal on Scientific Computing*, 41(4):A2603–A2626, 2019. doi: 10.1137/18M1229845.

Yusheng Peng, Gaofeng Zhang, Jun Shi, Xiangyu Li, and Liping Zheng. Mrgtraj: A novel non-autoregressive approach for human trajectory prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4):2318–2331, 2024.

Adolfo Perrusquía and Weisi Guo. Reservoir computing for drone trajectory intent prediction: A physics informed approach. *IEEE Transactions on Cybernetics*, 54(9):4939–4948, 2024. doi: 10.1109/TCYB.2024.3379381.

I. Podlubny. *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. Mathematics in Science and Engineering. Academic Press, 1998. ISBN 9780080531984.

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. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045.

Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. Latent odes for irregularly-sampled time series. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019.

Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *European Conference on Computer Vision (ECCV)*, 2020.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: a benchmark and a new model. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 5622–5632, Red Hook, NY, USA, 2017. ISBN 9781510860964.

Pranav singh chib, Achintya Nath, Paritosh Kabra, Ishu Gupta, and Pravendra Singh. MS-TIP: Imputation aware pedestrian trajectory prediction. In *Forty-first International Conference on Machine Learning*, 2024.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 2, pp. 2951–2959, 2012.

Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Madasamy Vellappandi and Sangmoon Lee. Physics-informed neural fractional differential equations. *Applied Mathematical Modelling*, 145:116127, 2025. ISSN 0307-904X. doi: 10.1016/j.apm.2025.116127.

Rui Wang and Rose Yu. Physics-guided deep learning for dynamical systems: A survey. *ACM Computing Surveys*, 58(5), 2025. ISSN 0360-0300. doi: 10.1145/3766887.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Pengfei Yao, Yinglong Zhu, Huikun Bi, Tianlu Mao, and Zhaoqi Wang. TrajCLIP: Pedestrian trajectory prediction method using contrastive learning and idempotent networks. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024a.

Pengfei Yao, Yinglong Zhu, Huikun Bi, Tianlu Mao, and Zhaoqi Wang. TrajCLIP: Pedestrian trajectory prediction method using contrastive learning and idempotent networks. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b.

Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII*, pp. 507–523, 2020. ISBN 978-3-030-58609-6. doi: 10.1007/978-3-030-58610-2_30.

Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9793–9803, 2021. doi: 10.1109/ICCV48922.2021.00967.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

Kai Zhao, Xuhao Li, Qiyu Kang, Feng Ji, Qinxu Ding, Yanan Zhao, Wenfei Liang, and Wee Peng Tay. Distributed-order fractional graph operating network. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Hao Zhou, Xu Yang, Mingyu Fan, Lu Qi, Xiangtai Li, Ming-Hsuan Yang, and Fei Luo. Three-dimensional trajectory prediction with 3DMotraj dataset. In *Forty-second International Conference on Machine Learning*, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

A Extended Literature Review

A.1 Data-Driven Trajectory Forecasting

Classical approaches to trajectory and time-series forecasting have heavily relied on discrete-time recurrent models, famously pioneered by Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997), Gated Recurrent Units (GRU) (Chung et al., 2014), and their spatio-temporal extensions such as Social-LSTM (Alahi et al., 2016) and TrajGRU (Shi et al., 2017). Recently, the paradigm has shifted significantly towards Transformer-based architectures. Models like Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), and PatchTST (Nie et al., 2023) have dominated general long-sequence time-series forecasting, while specialized architectures such as AgentFormer (Yuan et al., 2021) and STAR (Yu et al., 2020) have tailored attention mechanisms specifically for interacting multi-agent trajectories. Though highly expressive, their absolute effectiveness and computational efficiency over simpler linear models in certain contexts remain actively debated (Zeng et al., 2023).

In the specific domain of trajectory prediction, state-of-the-art methods have explored diverse discrete modeling strategies. Earlier breakthroughs leveraged Generative Adversarial Networks (GANs) (Gupta et al., 2018), Conditional Variational Autoencoders (CVAEs) (Salzmann et al., 2020), and Graph Neural Networks (GNNs) (Gao et al., 2020) to capture multi-modal human and vehicle behaviors. Building on this foundation, contemporary approaches have incorporated contrastive learning (Yao et al., 2024a), imputation-aware networks (singh chib et al., 2024), non-autoregressive generation (Peng et al., 2024), and implicit maximum likelihood estimation (Mohamed et al., 2022). Furthermore, the expansion of trajectory forecasting into complex 3D environments has been catalyzed by dynamic benchmarks such as 3DMoTraj (Zhou et al., 2025).

To address the physical implausibility often produced by purely data-driven models, recent efforts have begun integrating kinematic constraints and physics-informed losses to guide trajectory predictions (Liao et al., 2024; Jiao et al., 2024; Wang & Yu, 2025). However, a fundamental mathematical limitation shared by all these architectures is their strict reliance on fixed temporal discretization. These models fundamentally treat continuous physical motion as a disjointed sequence of temporal snapshots. Consequently, when applied to high-degree-of-freedom kinematics or irregularly sampled observations, they fail to smoothly interpolate spatial states between rigid step intervals. This inevitably leads to severe truncation errors and compounding structural drift over extended forecasting horizons (Rubanova et al., 2019; Kidger et al., 2021). Thus, these architectures fundamentally remain tied to discrete stepping and struggle to mathematically internalize the complex historical inertia inherent in real-world systems.

A.2 Continuous-Time and Physics-Informed Dynamics

To overcome the structural limitations of discrete stepping, Neural Ordinary Differential Equations (Neural ODEs) (Chen et al., 2018) pioneered the modeling of hidden states as continuous-time dynamic systems. By parameterizing the time derivative of the hidden state using a neural network and evaluating it via black-box ODE solvers, Neural ODEs natively support irregularly sampled time-series data and provide a mathematically rigorous bridge between deep learning and classical mechanics (Strogatz, 2018; Rubanova et al., 2019). Subsequent advancements, such as Augmented Neural ODEs (ANODEs) (Dupont et al., 2019) and memory-efficient adjoint backpropagation techniques (Kidger et al., 2021), have further enhanced their representational capacity and computational tractability, making them a cornerstone for continuous sequence modeling.

While highly successful in continuous domain mapping, standard Neural ODEs inherently operate as physically agnostic black boxes optimized solely to minimize statistical loss. The learned vector fields are often devoid of inductive physical biases. Consequently, in highly chaotic 3D environments subjected to severe non-Markovian noise (e.g., UAV aerodynamics or oceanic currents), relying exclusively on unconstrained neural representations causes these solvers to overfit to localized environmental disturbances and produce trajectory predictions that violate fundamental physical laws. Recognizing this critical blind spot, the paradigm of Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019) emerged to rigorously constrain neural manifolds using governing differential equations. This philosophy has deeply influenced continuous-time trajectory modeling, leading to architectures that embed Hamiltonian and Lagrangian mechanics directly into the learning

process to guarantee energy and momentum conservation (Greydanus et al., 2019; Cranmer et al., 2019).

In the specific context of dynamic systems and autonomous navigation, recent frameworks explicitly embed physical equations of motion to predict complex kinematics. Notable examples include PhysNODE (Lu & Qian, 2024) for hypersonic flight trajectories, Physics-Informed Neural ODEs (PINODEs) for structural identification (Lai et al., 2021), and momentum-conserving frameworks like BROGNET for Brownian dynamics (Bishnoi et al., 2024). Furthermore, recent uncertainty quantification models have heavily relied on physics-informed continuous architectures to predict flight dynamics and drone trajectory intent under missing observations and severe aerodynamic perturbations (Perrusquía & Guo, 2024; Michek et al., 2024; Inbaraj & Chen, 2026).

Despite significantly improving predictive stability and physical plausibility, all aforementioned continuous models share a fundamental mathematical bottleneck: they rely exclusively on integer-order derivatives. This rigid mathematical formulation imposes a strict Markovian assumption, dictating that the system's instantaneous rate of change depends solely on its immediate local state. Consequently, integer-order architectures are mathematically unequipped to capture anomalous diffusion, historical inertia, and the fading memory effects that inherently characterize the long-term evolution of real-world physical systems (Metzler & Klafter, 2000).

A.3 Fractional Neural Architectures

Modeling the complex historical inertia and long-term fading memory of physical systems fundamentally requires non-local mathematical operators. This property is rigorously formulated by fractional calculus, which utilizes integral convolutions to accumulate historical states (Oldham & Spanier, 2006; Podlubny, 1998). Recognizing this intrinsic advantage, recent advances have systematically integrated fractional derivatives into continuous-time deep learning to break the Markovian bottleneck. Foundational works such as Fractional Neural ODEs demonstrated that fractional-order models can naturally embed long-range dependencies without requiring excessively deep recurrent networks. This integration has rapidly expanded into physics-informed domains, with fractional physics-informed neural networks (fPINNs) (Pang et al., 2019) and physics-informed neural fractional differential equations (PiNFDEs) (Vellappandi & Lee, 2025) seamlessly incorporating fractional operators into the loss functions and neural architectures to efficiently model complex, memory-dependent dynamical behaviors.

Most current fractional neural networks, however, heavily rely on mathematically rigid formulations. They typically employ fixed fractional kernels (such as stationary power-law kernels) (Zhao et al., 2024) or variable-order fractional derivatives (Almeida, 2017; Cui et al., 2025) that assign predefined, state-isolated decay weights to historical observations. This static memory weighting assumes a homogeneous temporal influence and cannot dynamically adapt to unpredictable, input-specific contextual shifts. To tackle this critical limitation, the Neural Fractional Attention Differential Equation (FADE) framework (Kang et al., 2026) recently introduced a paradigm shift by combining fractional calculus with contextual, learnable attention mechanisms. By replacing fixed kernel functions with neural attention kernels, FADE adaptively weights historical states based on their contextual relevance. This allows the model to selectively emphasize crucial temporal dependencies while filtering out obsolete or noisy historical information, demonstrating superior performance over both integer-order and traditional fractional models.

Despite these cutting-edge theoretical breakthroughs, existing state-of-the-art fractional architectures still exhibit two critical structural limitations when applied to real-world kinematics. First, they uniformly apply a scalar fractional order across all spatial dimensions, forcing an artificial isotropic constraint on dynamics that are fundamentally anisotropic in 3D space (e.g., horizontal momentum versus vertical gravity resistance). Second, they model the dynamic vector field relying solely on observable spatial states, neglecting the invisible but crucial external environmental forces that continuously drive the system. Our proposed Context-aware Fractional Neural ODE directly addresses this gap by intertwining a data-driven hidden physics extractor with a novel, overflow-safe vectorized fractional calculus mechanism, enabling independent, dimension-specific memory retention and robust physical interpolation.

B PROOFS OF THEORETICAL ANALYSIS

Restatement of Proposition 1 *(Universal Fading Memory Property): For any non-integer order $\alpha > 0$, the scalar weight $w_k = (-1)^k \binom{\alpha}{k}$ acts as a universal power-law fading memory kernel with an asymptotic decay $w_k = \mathcal{O}(k^{-\alpha-1})$ (or explicitly $w_k \simeq \frac{1}{\Gamma(-\alpha)} k^{-\alpha-1}$) as $k \to \infty$. This analytically ensures that recent states exert dominant inertial influence while historical states decay without vanishing, across both sub-diffusive ($0 < \alpha < 1$) and super-diffusive ($\alpha > 1$) regimes.*

Proof of Proposition 1. By mathematical definition, the fractional binomial coefficient evaluated via the Gamma function is explicitly given by:

$$w_k = (-1)^k \binom{\alpha}{k} = \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)}$$

As $k \to \infty$ for any given non-integer $\alpha > 0$, the discrete sequence index k will eventually surpass $\alpha + 1$, causing the argument $(\alpha - k + 1)$ to become strictly negative. To analytically resolve the Gamma function for negative arguments without undefined singularities, we apply Euler's reflection formula, $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$, with the complex argument $z = \alpha - k + 1$:

$$\Gamma(\alpha - k + 1) = \frac{\pi}{\Gamma(k-\alpha)\sin(\pi(\alpha - k + 1))} = \frac{\pi}{(-1)^{k-1}\Gamma(k-\alpha)\sin(\pi\alpha)}$$

Substituting this fundamental relation back into the expression for w_k, the alternating sign terms $(-1)^k$ and $(-1)^{k-1}$ perfectly cancel, leaving an overall negative sign:

$$w_k = -\frac{\Gamma(\alpha+1)\sin(\pi\alpha)}{\pi} \frac{\Gamma(k-\alpha)}{\Gamma(k+1)}$$

Applying the reflection formula once more to the constant coefficient portion, noting that $\Gamma(\alpha+1)\Gamma(-\alpha) = -\frac{\pi}{\sin(\pi\alpha)}$, the constant term simplifies strictly to $\frac{1}{\Gamma(-\alpha)}$. This yields the equivalent exact analytical formulation valid for all non-integer fractional regimes ($\alpha \notin \mathbb{N}$):

$$w_k = \frac{1}{\Gamma(-\alpha)} \frac{\Gamma(k-\alpha)}{\Gamma(k+1)}$$

To rigorously establish the asymptotic behavior as $k \to \infty$, we utilize the asymptotic expansion for the ratio of Gamma functions (Kilbas et al., 2006), which formally states that $\frac{\Gamma(k+a)}{\Gamma(k+b)} = k^{a-b}\left[1 + \mathcal{O}(k^{-1})\right]$. By substituting $a = -\alpha$ and $b = 1$, we obtain:

$$\frac{\Gamma(k-\alpha)}{\Gamma(k+1)} = k^{-\alpha-1}\left[1 + \mathcal{O}(k^{-1})\right] \implies w_k = \frac{1}{\Gamma(-\alpha)} k^{-\alpha-1}\left[1 + \mathcal{O}(k^{-1})\right]$$

Since $\alpha \notin \mathbb{N}$, the evaluation $\Gamma(-\alpha)$ is strictly isolated from the non-positive integer poles, rendering $\frac{1}{\Gamma(-\alpha)}$ a well-defined, finite non-zero constant. Thus, the asymptotic decay of the weights strictly follows $w_k = \mathcal{O}(k^{-\alpha-1})$. □

To maintain clarity and mathematical rigor in the subsequent proofs, we adhere to the conventions established in the main text: t_S denotes the physical timestamp of the current observation boundary, $t_{>S} \in (t_S, \infty)$ represents the future time domain, $s \geq S$ defines the discrete autoregressive sequence index, and $\alpha_{min} = \min(\alpha_x, \alpha_y, \alpha_z)$ denotes the dominant fractional order governing the slowest inertial decay.

Restatement of Proposition 2 *(Existence and Uniqueness): Assuming a globally Lipschitz continuous neural vector field $\mathcal{N}_\theta$ with Lipschitz constant $L > 0$, the continuous mixed-order dynamical system admits a unique, well-posed solution for any initialized boundary state $\mathbf{z}(t_S)$.*

Proof of Proposition 2. To establish existence and uniqueness, we bridge the discrete Grünwald-Letnikov (GL) definition with continuous integral operators. For sufficiently smooth trajectories, the continuous GL derivative converges exactly to the Riemann-Liouville and Caputo fractional

derivatives (Podlubny, 1998, Chapter 2). Because our autonomous state defines a standard Initial Value Problem (IVP) anchored at the observable boundary t_S, we operate under the equivalent Caputo continuous framework (Diethelm, 2010).

Let $\mathbf{A} = [\alpha_x, \alpha_y, \alpha_z, 1, \dots, 1]^\top \in (0, 2]^{3+d_{context}}$ denote the augmented unified order vector. By the fundamental theorem of fractional calculus, the coupled differential ODE system (Eq. equation 3) is equivalently transformed into a system of weakly singular Volterra integral equations of the second kind:

$$z_i(t_{>S}) = \mathbf{z}_{init}^{(i)}(t_{>S}) + \frac{1}{\Gamma(A_i)} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{A_i - 1} \mathcal{N}_\theta^{(i)}(\mathbf{z}(\xi))\, d\xi, \quad i = 1, \dots, 3 + d_{context}.$$

Here, the integral term constructs the Riemann-Liouville fractional integral operator acting component-wise, with ξ as the dummy variable of integration, and $\mathbf{z}_{init}^{(i)}(t_{>S})$ encapsulating the polynomial initial conditions.

We define the Picard integral operator $\mathcal{T} : X \to X$ on the complete Banach space $X := C([t_S, t_{>S}^{\max}], \mathbb{R}^{3+d_{context}})$. To facilitate a global bound, we endow X with the Bielecki-type norm $\|\mathbf{z}\|_\beta := \sup_{t_{>S} \in [t_S, t_{>S}^{\max}]} e^{-\beta(t_{>S} - t_S)} \|\mathbf{z}(t_{>S})\|_1$ for $\beta > 0$, where $\|\cdot\|_1$ denotes the standard L_1 vector norm:

$$(\mathcal{T}\mathbf{z})_i(t_{>S}) := \mathbf{z}_{init}^{(i)}(t_{>S}) + \frac{1}{\Gamma(A_i)} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{A_i - 1} \mathcal{N}_\theta^{(i)}(\mathbf{z}(\xi))\, d\xi.$$

Assuming $\mathcal{N}_\theta$ exhibits global Lipschitz continuity with constant $L > 0$, the initial condition polynomials cancel. Evaluating the L_1 norm of the difference for any $\mathbf{z}_1, \mathbf{z}_2 \in X$ yields:

$$\|(\mathcal{T}\mathbf{z}_1)(t_{>S}) - (\mathcal{T}\mathbf{z}_2)(t_{>S})\|_1 \le L \sum_{i=1}^{3+d_{context}} \frac{1}{\Gamma(A_i)} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{A_i - 1} \|\mathbf{z}_1(\xi) - \mathbf{z}_2(\xi)\|_1 d\xi.$$

Using the definition of the Bielecki-type norm, $\|\mathbf{z}_1(\xi) - \mathbf{z}_2(\xi)\|_1 \le e^{\beta(\xi - t_S)} \|\mathbf{z}_1 - \mathbf{z}_2\|_\beta$, we bound the operator difference:

$$e^{-\beta(t_{>S} - t_S)} \|(\mathcal{T}\mathbf{z}_1)(t_{>S}) - (\mathcal{T}\mathbf{z}_2)(t_{>S})\|_1 \le L \|\mathbf{z}_1 - \mathbf{z}_2\|_\beta \sum_{i=1}^{3+d_{context}} \frac{1}{\Gamma(A_i)} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{A_i - 1} e^{-\beta(t_{>S} - \xi)} d\xi.$$

By substituting $r = t_{>S} - \xi$ and extending the upper limit of integration to infinity, the resulting integral converges precisely to the Laplace transform of a power function (Diethelm, 2010). This evaluation provides the upper bound:

$$\frac{1}{\Gamma(A_i)} \int_0^{t_{>S} - t_S} r^{A_i - 1} e^{-\beta r} dr \le \beta^{-A_i}$$

which implies:

$$\|\mathcal{T}\mathbf{z}_1 - \mathcal{T}\mathbf{z}_2\|_\beta \le L \left(\sum_{i=1}^{3+d_{context}} \beta^{-A_i} \right) \|\mathbf{z}_1 - \mathbf{z}_2\|_\beta.$$

By choosing the parameter $\beta > 0$ to be sufficiently large, we ensure the scale factor $\kappa := L \sum_i \beta^{-A_i} < 1$. This concludes that $\mathcal{T}$ is a contraction mapping on X. By the Banach Fixed-Point Theorem, $\mathcal{T}$ admits a unique fixed point $\mathbf{z}^* = \mathcal{T}\mathbf{z}^*$, guaranteeing the existence of a well-posed mixed-order trajectory. □

Restatement of Theorem 1 *(Fractional Error Bounding): Let $\mathbf{e}(t_{>S}) = \hat{\mathbf{z}}(t_{>S}) - \mathbf{z}(t_{>S})$ define the state prediction error over the continuous future horizon $t_{>S} > t_S$. Assuming the neural vector field $\mathcal{N}_\theta$ is globally Lipschitz continuous with constant $L > 0$, the error magnitude for the fractional system ($\mathbf{0} < \vec{\boldsymbol{\alpha}} \le \mathbf{2}$) is bounded by the generalized Mittag-Leffler function $E_{\alpha_{min}}$:*

$$\|\mathbf{e}(t_{>S})\| \le \|\mathbf{e}(t_S)\| E_{\alpha_{min}}\big(L(t_{>S} - t_S)^{\alpha_{min}}\big) \tag{11}$$

Proof of Theorem 1. To bypass commutation complexities between vector norms and fractional differential operators, we analyze error accumulation strictly within the Volterra integral framework. The true and predicted trajectories, $\mathbf{z}(t_{>S})$ and $\hat{\mathbf{z}}(t_{>S})$, satisfy:

$$\mathbf{z}(t_{>S}) = \mathbf{z}_{init}(t_{>S}) + \frac{1}{\Gamma(\vec{\boldsymbol{\alpha}})} \odot \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{\vec{\boldsymbol{\alpha}}-\mathbf{1}} \odot \mathcal{N}_\theta(\mathbf{z}(\xi))\, d\xi$$

The prediction error $\mathbf{e}(t_{>S}) = \hat{\mathbf{z}}(t_{>S}) - \mathbf{z}(t_{>S})$ stems from discrepancies in both the initial conditions and the accumulated neural field predictions. Assuming the deviation in the initialization polynomials is uniformly bounded by the observable boundary error $\|\mathbf{e}(t_S)\|$, we have:

$$\mathbf{e}(t_{>S}) = \mathbf{e}(t_S) + \frac{1}{\Gamma(\vec{\boldsymbol{\alpha}})} \odot \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{\vec{\boldsymbol{\alpha}}-\mathbf{1}} \odot \Big(\mathcal{N}_\theta(\hat{\mathbf{z}}(\xi)) - \mathcal{N}_\theta(\mathbf{z}(\xi))\Big)\, d\xi$$

Applying the vector norm $\|\cdot\|$ and utilizing the triangle inequality for Bochner integrals, we bring the norm inside the integral. To resolve the vector-valued fractional orders into a tractable scalar bound, we isolate the dominant accumulation term. Since $\alpha_{min} = \min(\vec{\boldsymbol{\alpha}})$ corresponds to the most singular integration kernel and governs the worst-case historical inertia, it provides a rigorous upper bound for the component-wise integral:

$$\|\mathbf{e}(t_{>S})\| \leq \|\mathbf{e}(t_S)\| + \frac{1}{\Gamma(\alpha_{min})} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{\alpha_{min}-1} \|\mathcal{N}_\theta(\hat{\mathbf{z}}(\xi)) - \mathcal{N}_\theta(\mathbf{z}(\xi))\|\, d\xi$$

Subjecting the neural vector field to the global Lipschitz condition, $\|\mathcal{N}_\theta(\hat{\mathbf{z}}) - \mathcal{N}_\theta(\mathbf{z})\| \leq L\|\hat{\mathbf{z}} - \mathbf{z}\| = L\|\mathbf{e}\|$, we establish the following continuous fractional integral inequality:

$$\|\mathbf{e}(t_{>S})\| \leq \|\mathbf{e}(t_S)\| + \frac{L}{\Gamma(\alpha_{min})} \int_{t_S}^{t_{>S}} (t_{>S} - \xi)^{\alpha_{min}-1} \|\mathbf{e}(\xi)\|\, d\xi$$

This formulation fits precisely into the generalized fractional Grönwall's inequality (Diethelm, 2010, Theorem 6.19). Resolving this inequality maps the scalar Lipschitz constant L into the infinite series of the Mittag-Leffler function $E_{\alpha_{min}}(z) = \sum_{j=0}^{\infty} \frac{z^j}{\Gamma(\alpha_{min} j + 1)}$, yielding the closed-form upper bound. □

Restatement of Theorem 2 *(Asymptotic Stability against Lyapunov Divergence): In continuous dynamical systems, the maximal separation rate of nearby trajectories is structurally bounded by the Lipschitz constant L of the driving vector field. By substituting integer-order dynamics with the fractional integral operator, the standard linear exponential divergence ($\mathcal{O}(e^{L(t_{>S}-t_S)})$) is replaced by a sub-exponential generalized Mittag-Leffler bound. Specifically, under the regime where $L > 0$ and $0 < \alpha_{min} < 1$, this chaotic suppression holds analytically, modulating the asymptotic divergence envelope to* $\exp\left(L^{1/\alpha_{min}}(t_{>S} - t_S)\right)$.

Proof of Theorem 2. In continuous chaotic forecasting environments defined by a positive divergence rate ($L > 0$), integer-order ODEs ($\vec{\boldsymbol{\alpha}} = \mathbf{1}$) degrade the analytical error bound to the standard exponential function, $E_1(L(t_{>S} - t_S)) = e^{L(t_{>S}-t_S)}$. This identity inherently dictates an immediate, catastrophic compounding of numerical and initial errors over time.

Under the proposed fractional continuum ($\alpha_{min} \in (0, 1) \cup (1, 2]$), the heavy-tailed Mittag-Leffler function governs the stability manifold. As $t_{>S} \to \infty$, the asymptotic expansion of the Mittag-Leffler function for strictly positive arguments is dominated by its principal exponential component (Gorenflo et al., 2014):

$$E_{\alpha_{min}}(z) = \frac{1}{\alpha_{min}} \exp\left(z^{1/\alpha_{min}}\right) + \mathcal{O}(z^{-1}) \implies E_{\alpha_{min}}(z) \sim \frac{1}{\alpha_{min}} \exp\left(z^{1/\alpha_{min}}\right)$$

Substituting the dynamical argument $z = L(t_{>S} - t_S)^{\alpha_{min}}$ into this asymptotic expansion yields:

$$\begin{aligned} E_{\alpha_{min}}\left(L(t_{>S} - t_S)^{\alpha_{min}}\right) &\sim \frac{1}{\alpha_{min}} \exp\left(\left[L(t_{>S} - t_S)^{\alpha_{min}}\right]^{1/\alpha_{min}}\right) \\ &= \frac{1}{\alpha_{min}} \exp\left(L^{1/\alpha_{min}}(t_{>S} - t_S)\right) \end{aligned}$$

Recalling Theorem 1 and treating the initial state error magnitude $\|\mathbf{e}(t_S)\|$ as a constant multiplier, the overall distal prediction error asymptotically scales as:

$$\|\mathbf{e}(t_{>S})\| \sim \frac{\|\mathbf{e}(t_S)\|}{\alpha_{min}} \exp\left(L^{1/\alpha_{min}}(t_{>S} - t_S)\right)$$

This analytical deduction formalizes a crucial dynamical shift. While the strictly linear time-exponent $(t_{>S} - t_S)$ preserves the necessary topological chaos of the system, the effective Lyapunov growth rate driving that divergence is non-linearly transformed from L to $L^{1/\alpha_{min}}$. Consequently, under normalized divergence regimes (e.g., $L \in (0, 1)$ for sub-diffusive $\alpha_{min} < 1$, or $L > 1$ for super-diffusive $\alpha_{min} > 1$), this transformation strictly suppresses the exponential growth rate, systematically delaying catastrophic error accumulation compared to integer-order baselines. □

Restatement of Lemma 1 *(Operator and Memory Truncation Error): Assuming uniformly bounded spatial coordinates ($\|\mathbf{p}\| \leq P_{max}$) and C^2-continuous trajectories, the forward Euler integration of the latent environment yields a local truncation error of $\varepsilon_{\lambda,s} = \mathcal{O}(\Delta t^2)$. Concurrently, the truncated Grünwald-Letnikov operator yields a strictly bounded local spatial error $\varepsilon_{\mathbf{p},s} = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$.*

Proof of Lemma 1. For the latent environmental state, the local step error of the explicit first-order forward Euler update is a fundamental result derived directly from Taylor series expansions, analytically yielding the standard bound $\varepsilon_{\lambda,s} = \mathcal{O}(\Delta t^2)$.

For the physical space, the first-order GL finite difference operator natively bounds the continuous infinite history approximation, mathematically guaranteeing an operator truncation error of $\mathcal{O}(\Delta t)$ (Podlubny, 1998). However, deploying the autoregressive rollout strictly mandates truncating the infinite historical tail to a fixed practical window W. Because the fractional weights exhibit an asymptotic power-law decay, the vector magnitude of the weights is bounded by the slowest decaying component dictated by α_{min}. Thus, there exists a constant $C > 0$ such that $\|\mathbf{w}_k\|_\infty \leq Ck^{-\alpha_{min}-1}$.

Given the uniformly bounded spatial trajectory ($\|\mathbf{p}(t)\| \leq P_{max}$), the error introduced by the discarded sequence tail is analytically bounded by evaluating the continuous improper integral:

$$\begin{aligned}
\sum_{k=W+1}^{\infty} \|\mathbf{w}_k \odot \mathbf{p}(t_{s-k})\| &\leq P_{max} \sum_{k=W+1}^{\infty} \|\mathbf{w}_k\|_\infty \\
&\leq CP_{max} \int_W^{\infty} x^{-\alpha_{min}-1}\, dx \\
&= CP_{max} \left[\frac{x^{-\alpha_{min}}}{-\alpha_{min}}\right]_W^{\infty} = \frac{CP_{max}}{\alpha_{min}} W^{-\alpha_{min}}
\end{aligned}$$

Since $\frac{CP_{max}}{\alpha_{min}}$ acts as a fixed autonomous scalar multiplier, the memory truncation penalty strictly falls into the complexity class $\mathcal{O}(W^{-\alpha_{min}})$. Superimposing the operator discretization and memory restriction penalties establishes the total local spatial truncation error: $\varepsilon_{\mathbf{p},s} = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$. □

Restatement of Theorem 3 *(Global Convergence Criterion and Rate): Let M be the discrete forecasting horizon, corresponding to a maximum continuous integration time $T_f = M\Delta t$. The autoregressive solver globally converges to the true continuous trajectory as $\Delta t \to 0$ and $W \to \infty$. The global accumulated state error $\mathbf{E}_M$ follows:*

$$\|\mathbf{E}_M\| = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}}) \tag{12}$$

Assuming a sufficiently large memory window W that strictly dominates the fractional truncation penalty ($W \gg \Delta t^{-1/\alpha_{min}}$), the framework achieves a linear global convergence rate of $\mathcal{O}(\Delta t)$ for any valid fractional order $\vec{\boldsymbol{\alpha}} > \mathbf{0}$.

Proof of Theorem 3. A common theoretical pitfall in fractional numerics is assuming that global errors aggregate linearly over $M = T_f/\Delta t$ steps. Mathematically, discrete fractional errors propagate via fractional convolutions, inherently retaining memory of past perturbations.

Let $\mathbf{E}_s = \hat{\mathbf{z}}_s - \mathbf{z}(t_s)$ denote the global state prediction error accumulated at step s. The discrete residual dynamics, evaluated under the multi-order Grünwald-Letnikov difference operator $\nabla^{\vec{\boldsymbol{\alpha}}}_{\Delta t}$, satisfy:

$$\nabla^{\vec{\boldsymbol{\alpha}}}_{\Delta t}\mathbf{E}_s = \mathcal{N}_\theta(\hat{\mathbf{z}}_{s-1}) - \mathcal{N}_\theta(\mathbf{z}_{s-1}) + \varepsilon_s$$

where ε_s encapsulates the total local truncation error from Lemma 1. By taking the vector magnitude norm and enforcing the L-Lipschitz continuity of the neural field, the worst-case error accumulation is bounded by the most singular memory kernel corresponding to $\alpha_{min} = \min(\vec{\boldsymbol{\alpha}})$. This yields the nonlinear recursive bound:

$$\|\nabla^{\alpha_{min}}_{\Delta t}\mathbf{E}_s\| \leq L\|\mathbf{E}_{s-1}\| + \|\varepsilon_s\|$$

From Lemma 1, we analytically bound the local source error by explicit constants: $\|\varepsilon_s\| \leq C_1\Delta t + C_2 W^{-\alpha_{min}}$. Applying the discrete fractional Grönwall's inequality (Diethelm, 2010) to integrate this constant-order local bound over the finite terminal horizon T_f yields:

$$\|\mathbf{E}_M\| \leq \frac{T_f^{\alpha_{min}}}{\Gamma(\alpha_{min}+1)} E_{\alpha_{min}}(LT_f^{\alpha_{min}})\Big(C_1\Delta t + C_2 W^{-\alpha_{min}}\Big)$$

Because the continuous forecasting horizon T_f and the Lipschitz constant L are predefined, fixed scalar quantities for any given physical simulation, the leading fractional coefficient and the Mittag-Leffler function strictly evaluate to a finite constant amplification factor $\mathcal{K}$. This mathematical structure guarantees that the discrete autoregressive rollout restricts exponential blowups. Absorbing the constants into asymptotic notation, the global state convergence strictly follows:

$$\|\mathbf{E}_M\| = \mathcal{O}(\Delta t) + \mathcal{O}(W^{-\alpha_{min}})$$

Assuming the hyperparameter condition $W \gg \Delta t^{-1/\alpha_{min}}$ holds, the memory error scale $\mathcal{O}(W^{-\alpha_{min}})$ is mathematically dwarfed by the discretization term $\mathcal{O}(\Delta t)$ and becomes negligible. Under this regime, the global convergence rate is entirely governed by the numerical step size, securing the optimal linear bound $\mathcal{O}(\Delta t)$. □

C Datasets and Experiment Setup

C.1 General Problem Formulation

For the physical kinematic datasets (UAV and UUV), the forecasting task aims to predict the future 3D spatial coordinates $\hat{\mathbf{y}}_t \in \mathbb{R}^3$ over a prediction horizon M, given a historical observation window W. To explicitly mitigate accumulated long-term temporal drift across these tasks, the objective function is formulated as a time-weighted Huber loss. Let $\mathbf{y}_t$ and $\hat{\mathbf{y}}_t$ denote the ground truth and predicted spatial coordinates at future time step t. The objective function over the horizon M is defined as:

$$\mathcal{L} = \frac{1}{M} \sum_{t=1}^{M} \exp(\gamma t) \cdot \mathcal{H}_\delta\big(\hat{\mathbf{y}}_t - \mathbf{y}_t\big) \tag{13}$$

where γ is an exponential penalty rate designed to heavily weight the final displacement error (set to 0.1 in our experiments), and $\mathcal{H}_\delta(\cdot)$ is the robust Huber penalty function parameterized by a transition threshold $\delta = 0.1$ to maintain robustness against spatial outliers.

To optimize the network parameters $\boldsymbol{\theta}$, we eschew standard stochastic gradient descent in favor of an adaptive moment estimation algorithm coupled with decoupled L_2 regularization (weight decay). This optimizer dynamically adjusts the learning rates for each parameter based on first and second moments of the gradients, ensuring stable convergence in highly non-linear prediction landscapes.

We benchmark the predictive performance using standard trajectory forecasting metrics calculated in the un-normalized physical space (meters). Let N denote the total number of evaluated trajectories. The spatial accuracy is quantified using the following metrics:

- **Average Displacement Error (ADE)** measures the mean Euclidean distance between the predicted and ground truth positions across all predicted time steps:

$$\text{ADE} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{t=1}^{M} \left\| \hat{\mathbf{y}}_t^{(i)} - \mathbf{y}_t^{(i)} \right\|_2 \tag{14}$$

- **Final Displacement Error (FDE)** calculates the spatial discrepancy strictly at the last predicted step M, indicating the accumulation of long-term drift:

$$\text{FDE} = \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\mathbf{y}}_M^{(i)} - \mathbf{y}_M^{(i)} \right\|_2 \tag{15}$$

- **Root Mean Square Error (RMSE)** provides the square root of the average squared errors, heavily penalizing large deviations:

$$\text{RMSE} = \sqrt{\frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{t=1}^{M} \left\| \hat{\mathbf{y}}_t^{(i)} - \mathbf{y}_t^{(i)} \right\|_2^2} \tag{16}$$

Additionally, we introduce **Valid Time** (T_v), defined as the maximum number of consecutive steps the predicted trajectory remains within a strict spatial divergence threshold $d_{\max}$. For the i-th trajectory, this is formalized as:

$$T_v^{(i)} = \max \left\{ \tau \in \{1, \dots, M\} \mid \forall t \le \tau, \left\| \hat{\mathbf{y}}_t^{(i)} - \mathbf{y}_t^{(i)} \right\|_2 \le d_{\max} \right\} \tag{17}$$

The reported T_v is the global average across all trajectories, providing a highly practical measure of the temporal limits of reliable forecasting.

To ensure reproducibility and statistical robustness, all experiments were evaluated across five independent trials. Table 4 details the exact random seeds utilized for each dataset, including the deterministic chaotic attractors (Lorenz, Rössler, Thomas) and the 3D physical kinematic tasks (UAV, UUV). The mean and standard deviation metrics reported in our performance evaluations are derived directly from these configurations.

Table 4: Random seeds used for each dataset across 5 independent trials.

Lorenz	42	46	36	55	58
Rössler	42	30	36	38	58
Thomas	42	28	50	73	54
UAV	42	123	2024	8888	9999
UUV	42	135	77	888	9999

C.2 Chaotic Dynamical Systems

We consider three well-known 3D chaotic benchmark systems: the Lorenz63 system, the Rössler system, and the Thomas system (see Figure 3). These systems are deterministic, as they are governed by known equations. However, they remain challenging to model due to their extreme sensitivity to initial conditions, which makes them effective benchmarks for evaluating forecasting models that capture nonlinear and chaotic dynamics. Their governing equations are given below, together with the parameter settings used in our experiments.

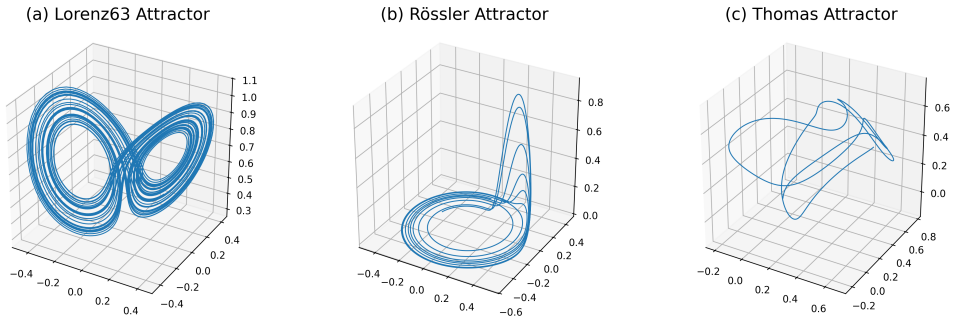


Figure 3: 3D Trajectories of Chaotic Benchmark Attractors

C.2.1 Lorenz63 system

The Lorenz63 system is governed by:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z. \end{aligned} \tag{18}$$

In this work, we use the standard chaotic parameter setting $(\sigma, \rho, \beta) = (10, 28, \frac{8}{3})$.

C.2.2 Rössler system

The Rössler system is governed by:

$$\begin{aligned} \frac{dx}{dt} &= -y - z, \\ \frac{dy}{dt} &= x + ay, \\ \frac{dz}{dt} &= b + z(x - c). \end{aligned} \tag{19}$$

We adopt the commonly-used chaotic configuration $(a, b, c) = (0.1, 0.1, 14)$.

C.2.3 THOMAS SYSTEM

The Thomas system is governed by:

$$\begin{aligned}\frac{dx}{dt} &= \sin(y) - bx,\\ \frac{dy}{dt} &= \sin(z) - by,\\ \frac{dz}{dt} &= \sin(x) - bz.\end{aligned} \tag{20}$$

We use $b = 0.208186$ to generate the system.

C.2.4 DATA GENERATION AND PREPROCESSING

These systems are numerically integrated using a fixed step size $\Delta t = 0.01$ with Wray's third-order method. Since they are autonomous, the dynamics can be expressed as $\frac{d\mathbf{u}}{dt} = f(\mathbf{u})$, where $\mathbf{u} = [x, y, z]^\top$ denotes the state. Given the state $\mathbf{u}_n$, one step to $\mathbf{u}_{n+1}$ can be computed via three stages:

$$\begin{aligned}\mathbf{u}^{(0)} &= \mathbf{u}_n,\\ \mathbf{k}_1 &= f(\mathbf{u}^{(0)}),\\ \mathbf{u}^{(1)} &= \mathbf{u}^{(0)} + \Delta t\gamma_1\mathbf{k}_1,\\ \mathbf{k}_2 &= f(\mathbf{u}^{(1)}),\\ \mathbf{u}^{(2)} &= \mathbf{u}^{(1)} + \Delta t(\gamma_2\mathbf{k}_2 + \eta_2\mathbf{k}_1),\\ \mathbf{k}_3 &= f(\mathbf{u}^{(2)}),\\ \mathbf{u}_{n+1} &= \mathbf{u}^{(2)} + \Delta t(\gamma_3\mathbf{k}_3 + \eta_3\mathbf{k}_2),\end{aligned}$$

where $\gamma = \left(\frac{8}{15},\ \frac{5}{12},\ \frac{3}{4}\right)$ and $\eta = \left(0,\ -\frac{17}{60},\ -\frac{5}{12}\right)$.

To ensure that the dynamics used are representative, the initial 5000 time steps are discarded as a burn-in period to remove transient behaviors. The dataset is subsequently split sequentially into a training set (6,000 steps), a validation set (3,000 steps), and a test set (5,000 steps). The data is scaled by the range of the training data, i.e., $1/(max_{train} - min_{train})$. The continuous trajectories are segmented into discrete sequences using a sliding window approach, where all three systems uniformly adopt a look-back window of length $W = 64$, a prediction horizon of 16, and a stride of 5.

C.2.5 EVALUATION METRICS FOR CHAOTIC SYSTEMS

To comprehensively evaluate the model, both short- and long-term performances are considered. For short-term accuracy, we compute the mean squared error (MSE) over the initial $N_{short} = 500$ steps of the test set:

$$MSE = \frac{1}{N_{short}} \sum_{t=1}^{N_{short}} \|\hat{\mathbf{u}}_t - \mathbf{u}_t\|_2^2 \tag{21}$$

For long-term predictability, we track the *Valid Time* (T_v) identically to the formulation in Eq. equation 17, but utilizing a dimensionless normalized error. Specifically, the normalized error $E(t)$ at time step t is formulated as the displacement divided by the root mean square (RMS) amplitude of the entire true test trajectory:

$$E(t) = \frac{\|\hat{\mathbf{u}}_t - \mathbf{u}_t\|_2}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{u}_i\|_2^2}} \tag{22}$$

where $N = 5,000$ represents the full length of the test set, and the Valid Time is recorded before the error crosses a chosen threshold $d_{\max} = 0.4$ for the first time. The valid time can be computed directly on the scaled data, as the scaling factor cancels out in the normalized prediction error.

C.2.6 Training and Optimization Details

Bayesian optimization Snoek et al. (2012) with 50 evaluation points was implemented, comprising 10 initial random samples followed by 40 sequential optimization steps using a Gaussian process surrogate model. The search space for our model is detailed in Table 5.

Table 5: Bayesian optimization search space

Hyperparameter	Lorenz63	Rössler	Thomas
Hidden dimension	[64, 128]	[64, 128]	[64, 128]
Context dimension	[16, 128]	[16, 128]	[16, 128]
η_θ	[1e-4, 2e-3]	[1e-4, 2e-3]	[1e-4, 2e-3]
η_α	[1e-3, 2e-2]	[1e-3, 2e-2]	[1e-3, 9e-3]
λ_{prior}	[1e-4, 5e-1]	[1e-3, 5e-1]	[1e-4, 5e-1]
Dropout	[0.0, 0.3]	[0.1, 0.3]	[0.1, 0.3]
Batch size	[64, 128]	[64, 128]	[64, 128]

To preserve the natural optimization landscape identified by the Bayesian search, no gradient clipping or manual clamping of the fractional orders is applied. Finally, to maximize data utilization and strictly prevent data leakage, the optimal model is retrained for a fixed 400 epochs on the concatenated training and validation sets - completely isolated from the test set - before the final evaluation.

C.3 UAV Trajectories

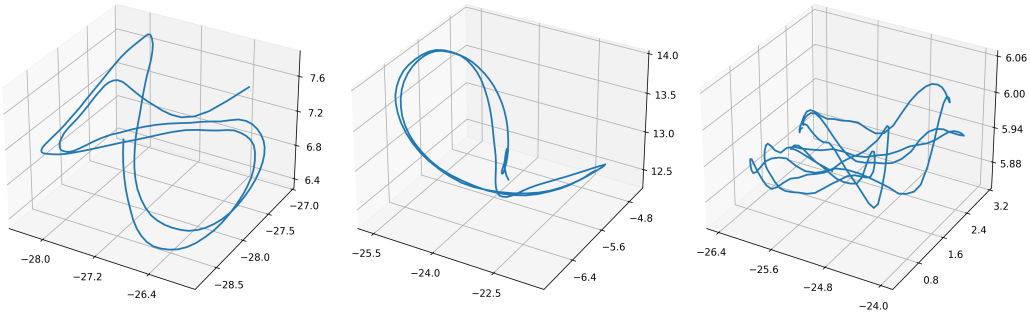


Figure 4: Examples of 3D UAV trajectories

C.3.1 Dataset and Significance

Unmanned Aerial Vehicles (UAVs) operate as highly complex physical systems governed by nonlinear aerodynamics, underactuated rigid body dynamics, and continuous 3D spatial transitions. Accurate long-term trajectory forecasting is therefore critical for downstream tasks such as autonomous collision avoidance and intent prediction. To rigorously evaluate our model's capacity to capture these underlying dynamics, we utilize the *Synthetic UAV Flight Trajectories* dataset [1]. This dataset provides high-fidelity, continuous 3D positional coordinates (x, y, z) over time. By leveraging synthetically generated flight profiles, the dataset ensures access to robust, noise-free ground truth that inherently embeds the precise kinematic constraints and varied maneuverability patterns of rotary-wing aircraft. This allows for an isolated and objective evaluation of our model's physical consistency and predictive accuracy.

C.3.2 Data Processing and Architecture

The trajectory sequences are sampled with a time interval of $\Delta t = 0.1$s. We frame the forecasting task with an observation window of $W = 50$ steps to predict the subsequent $M = 20$ steps. The

[1] https://huggingface.co/datasets/riotu-lab/Synthetic-UAV-Flight-Trajectories

training data is generated using a sliding window approach with a stride of 10. For the model architecture, we configure the Context-aware fNODE with a hidden dimension of 128, a context dimension of 64, and 2 layers, while applying a dropout rate of 0.2 to prevent overfitting on specific flight patterns.

C.3.3 TRAINING STRATEGY

The model parameters $\boldsymbol{\theta}$ are optimized via the adaptive moment estimation algorithm defined in Section C.1, utilizing an initial step size of $\eta_0 = 10^{-3}$ and a weight decay coefficient of $\lambda = 10^{-4}$. To ensure stable convergence, the step size is subject to an annealing schedule, decaying multiplicatively by a factor of 0.5 whenever the validation empirical risk plateaus. The objective function strictly follows the Huber loss parameterized in Eq. equation 13. Crucially, for inherently stable UAV rigid-body kinematics, the fractional derivative natively preserves momentum. Setting $\gamma = 0$ avoids over-regularization and ensures perfectly balanced gradient propagation across the prediction horizon.

Furthermore, the structural constraint operator $\mathcal{S}$ is activated without explicit prior regularization ($\lambda_{prior} = 0$). Because flying rigid bodies exhibit isotropic macroscopic inertia, spatial symmetry must be maintained across the fractional orders ($\alpha_x = \alpha_y = \alpha_z$). Rather than rigidly hardcoding a scalar order, this symmetry is dynamically enforced during backpropagation through an isotropic gradient projection $\nabla_{\vec{\boldsymbol{\alpha}}}\mathcal{L} \leftarrow \left(\frac{1}{3}\sum_{i\in\{x,y,z\}}\frac{\partial\mathcal{L}}{\partial\alpha_i}\right)\mathbf{1}$ where $\mathbf{1} \in \mathbb{R}^3$ is the vector of ones. This mechanism ensures that the parameter updates remain symmetric in the spatial domain while synchronously updating via the step size $\eta_\alpha = \eta_\theta$.

C.3.4 EVALUATION METRICS

The predictive performance on the UAV dataset is benchmarked utilizing the spatial metrics defined in Section C.1. Specifically, we report the Average Displacement Error equation 14, Final Displacement Error equation 15, and Root Mean Square Error equation 16. The Valid Time (T_v) is evaluated using Eq. equation 17 with a strict spatial divergence threshold of $d_{\max} = 2.0$ meters.

C.4 UUV TRAJECTORIES

C.4.1 DATASET AND SIGNIFICANCE

Unmanned Underwater Vehicles (UUVs) operate in highly non-linear fluid environments, subject to complex hydrodynamic forces such as variable drag, buoyancy, and unpredictable ocean currents. To evaluate our model's adaptability to these conditions, we utilize the 3DMoTraj dataset Zhou et al. (2025), encompassing 8 distinct underwater scenes (`3Dscene1` to `3Dscene8`). This dataset captures continuous 3D spatial transitions, providing a rigorous testbed for modeling multi-axis aquatic dynamics.

C.4.2 DATA PROCESSING AND ARCHITECTURE

We frame the forecasting task with an observation window of $W = 8$ steps to predict a future horizon of $M = 12$ steps. The trajectory sequences are extracted using a sliding window approach with a stride of 2. For the Context-aware fNODE architecture applied to UUVs, we utilize a hidden dimension of 64 and a context dimension of 64 across 2 layers. A dropout rate of 0.2 is maintained to enhance generalization across the varied topographical properties of the 8 scenes.

C.4.3 TRAINING STRATEGY

The network parameters are similarly optimized using the adaptive moment algorithm incorporating weight decay regularization ($\lambda = 10^{-4}$) with an initial learning rate of 10^{-3}. To ensure stable convergence, we employ a dynamic learning rate scheduling strategy that decays the rate by a factor of 0.5 upon observing a validation error plateau for 5 consecutive epochs. The optimization objective minimizes the time-weighted Huber loss as defined in Eq. equation 13.

Crucially, for UUV modeling, we configure the structural prior matrix $\mathcal{S}$ as the identity matrix $\mathbf{I}$, eliminate the isotropic penalty ($\lambda_{prior} = 0$), and synchronize the learning rates ($\eta_\alpha = \eta_\theta$). In underwater environments, spatial dynamics are inherently incommensurate due to competing asymmetric forces (e.g., vertical buoyancy opposing gravity versus horizontal hydrodynamic drag). By completely relaxing isotropic constraints, this formulation allows the network to organically discover independent, axis-specific fractional memory profiles.

C.4.4 Evaluation Metrics

Predictive performance for UUVs is analogously quantified in the un-normalized physical space. We compute the Average Displacement Error equation 14 and Final Displacement Error equation 15 independently for each of the $K = 8$ underwater scenes. To benchmark the overall robustness across highly varied topographies and hydrodynamic conditions, we report the Global Mean ADE and FDE:

$$\text{ADE}_{global} = \frac{1}{K}\sum_{k=1}^{K}\text{ADE}_k, \quad \text{FDE}_{global} = \frac{1}{K}\sum_{k=1}^{K}\text{FDE}_k \tag{23}$$

where ADE_k and FDE_k denote the spatial errors evaluated exclusively on the k-th scene.

Similar to the UAV evaluation, we track the Valid Time (T_v) as defined in Eq. equation 17. We maintain the identical spatial divergence threshold of $d_{\max} = 2.0$ meters to rigorously assess the temporal boundaries of reliable trajectory forecasting in complex fluid environments.

C.5 Hardware Setup and Runtime Analysis

To ensure full transparency regarding the computational footprint of our proposed architecture, we report the hardware specifications and empirical runtimes across all benchmarked datasets.

Regarding GPU type and hardware setup, computational resources were allocated based on task complexity. Evaluations on chaotic systems (Lorenz, Rössler, Thomas) were conducted on an Apple MacBook Pro M4 (10-core CPU, 16GB unified memory). The computationally intensive 3D physical kinematics tasks (UAV and UUV datasets) were executed on a dedicated workstation equipped with a 16-core CPU, 29GB RAM, and a single NVIDIA RTX A6000 GPU (24GB VRAM).

Table 6 compiles the average training time per epoch and the inference time per step for each dataset. As expected, the fractional integration introduces an $\mathcal{O}(W)$ computational overhead per step, which scales linearly with the historical sliding window size W.

Table 6: Computational complexity and empirical runtime analysis across all benchmarked datasets.

Task	**Training Time (min/epoch)**	**Inference Time (sec/step)**
Lorenz	0.0220 ± 0.0002	$0.0005 \pm$ 3.8e-06
Rössler	0.0207 ± 0.0002	$0.0005 \pm$ 1.4e-06
Thomas	0.0220 ± 0.0003	$0.0005 \pm$ 1.4e-06
UAV	1.93 ± 0.0429	0.651 ± 0.013
UUV	1.14 ± 0.0242	0.134 ± 0.016

D ABLATION STUDY

To systematically evaluate the individual contributions of each mechanism within the Context-fNODE architecture, we conducted a comprehensive ablation study. The model variants were tested across rigid-body kinematics (UAV), fluid dynamics (UUV), and three synthetic chaotic systems. The ablated baselines are defined as follows:

- **fNODE (w/o Context λ):** This variant completely removes the co-evolving latent context network. By forcing the system to rely solely on observable spatial history, it strips the model's capacity to dynamically absorb and adapt to unobservable exogenous perturbations.
- **Context-NODE:** This variant replaces the fractional-order derivative with a standard integer-order operator ($\vec{\alpha} = \mathbf{1}$). Although it retains the context network, it enforces a memoryless Markovian assumption, failing to capture the long-term historical inertia naturally preserved by the Grünwald-Letnikov formulation.
- **w/o Structural Prior:** This configuration disables the isotropic penalty ($\lambda_{prior} = 0$). It allows each spatial axis to independently learn its own decoupled, anisotropic fractional memory profile, removing any regularization that encourages macroscopic kinematic symmetry.
- **w/o Time-Weight:** This variant removes the exponential time-weighted penalty ($\beta = 0$), reverting to a standard uniform Smooth L_1 loss. Consequently, it treats short-term predictions and long-term horizon forecasting with equal importance, neutralizing the explicit mechanism designed to suppress long-term trajectory drift.

Table 7: Ablation study of Context-fNODE components. Predictive accuracy is reported as ADE/FDE for kinematic datasets and MSE/T_v for dynamic chaotic systems.

Model Variant	UAV Dataset	UUV Dataset	Dynamic Lorenz63	Dynamic Rössler	Dynamic Thomas
	ADE / FDE	ADE / FDE	MSE / T_v	MSE / T_v	MSE / T_v
fNODE	1.80 / 2.54	0.60 / 0.97	0.0006 / 529	0.0001 / 1702	7.1e-05 / 3898
Context-NODE	1.64 / 2.28	0.58 / 0.98	0.0362 / 279	9.0e-05 / 1705	7.0e-05 / 4051
w/o Structural Prior	**1.52 / 2.17**	**0.54 / 0.86**	0.0380 / 281	3.8e-05 / 1707	9.9e-05 / 1831
w/o Time-Weight	**1.52 / 2.17**	0.59 / 1.03	0.0231 / 126	0.0002 / 1703	0.0001 / 3879
Context-fNODE	**1.52 / 2.17**	**0.54 / 0.86**	**1.7e-05 / 775**	**2.1e-05 / 1727**	**1.8e-05 / 5000**

D.1 RIGID-BODY KINEMATICS (UAV)

In the rigid-body UAV dataset, the integrated Context-fNODE framework demonstrates its strict necessity for long-horizon stability. Removing the co-evolving latent environment (**fNODE**) inflates the ADE from 1.52m to 1.80m. This proves that observable spatial history alone is insufficient; without the latent context to dynamically infer unobservable exogenous factors (such as wind gusts or aerodynamic drag), the model cannot accurately extrapolate future states. Furthermore, reverting to an integer-order system (**Context-NODE**) strips the model of its physical momentum, increasing the Final Displacement Error (FDE) from 2.17m to 2.28m. Integer-order derivatives enforce a strict Markovian assumption, abruptly truncating the memory of past trajectories and failing to capture the continuous inertial forces acting on the rigid body. Notably, the **w/o Time-Weight** and **w/o Structural Prior** variants perform identically to the full model. This empirically validates our framework's domain adaptability: because UAV flight paths represent inherently stable, kinematically symmetric rigid-body motion, the full model optimally converges to a uniform loss profile ($\beta = 0$) to avoid over-regularization, and naturally learns an isotropic memory distribution without relying on forced structural constraints.

D.2 FLUID DYNAMICS (UUV)

For the fluid-dynamic UUV dataset, the ablation results underscore the extreme complexity of aquatic environments. Here, stripping away the context network (**fNODE**) severely degrades performance (FDE jumps to 0.97m), as the model becomes blind to unobservable underwater currents and

turbulent hydrodynamics. Similarly, removing fractional memory (**Context-NODE**) discards historical inertia, increasing FDE to 0.98m, highlighting that fluid resistance possesses a long-lasting memory effect that integer operators cannot model. Unlike the UAV setting, removing the exponential time-weight (**w/o Time-Weight**) noticeably degrades performance, pushing the FDE up to 1.03m. This confirms that heavily penalizing late-stage divergence is strictly required to suppress the compounding autoregressive drift caused by unpredictable fluid perturbations. Crucially, the **w/o Structural Prior** configuration yields performance perfectly identical to the full Context-fNODE (**0.54 / 0.86m**). By intentionally relaxing the isotropic structural prior ($\lambda_{prior} = 0$) in the optimal model, the network organically avoids the bottleneck of rigid assumptions. It successfully captures decoupled, anisotropic fluid forces - such as vertical buoyancy opposing gravity versus horizontal hydrodynamic drag - achieving superior stability without forced macroscopic symmetry.

D.3 Deterministic Chaotic Systems

Finally, the evaluation across the three synthetic chaotic datasets (Lorenz63, Rössler, Thomas) unequivocally highlights the necessity of the complete, unmodified Context-fNODE architecture. Predicting strange attractors demands extreme precision, and removing any component triggers severe destabilization. The most profound degradation is observed in the integer-order variant (**Context-NODE**), where the Mean Squared Error inflates by orders of magnitude (e.g., spiking from 1.7×10^{-5} to 0.0362 on Lorenz63), causing the Valid Time (T_v) to collapse prematurely. This definitively proves that integer-order ODEs suffer from rapid memory decay and simply cannot sustain the highly non-local, non-linear dynamics inherent to chaotic attractors.

Furthermore, unlike the kinematic datasets, disabling the isotropic penalty (**w/o Structural Prior**) severely cripples chaotic forecasting (e.g., T_v drops from 775 to 281 on Lorenz63). Chaotic differential equations are intrinsically coupled; allowing spatial axes to learn decoupled, independent memory profiles breaks the delicate mathematical symmetry required to accurately traverse the attractor's lobes. Finally, removing the temporal penalty (**w/o Time-Weight**) allows microscopic short-term errors to compound exponentially due to the system's positive Lyapunov exponents, causing the trajectory to rapidly diverge from the strange attractor. Integrating all context, fractional, and structural mechanisms empowers the full framework to achieve unprecedented predictive endurance, gracefully reaching the absolute 5000-step evaluation limit on the Thomas system.

D.4 Sensitivity Analysis for Window Size

The choice of the sliding window size W controls the strict trade-off between capturing long-tail historical inertia and the computational memory footprint ($\mathcal{O}(W)$). To study the effect of the window size and validate our hyperparameter search, we conducted a sensitivity analysis on the chaotic dynamical systems by varying $W \in \{32, 56, 64, 72\}$.

As shown in Table 8, the empirical results confirm the usefulness of Bayesian Optimization. The model achieves its peak performance (lowest MSE and highest Valid Time T_v) at $W = 64$ across all three systems. A smaller window (e.g., $W = 32$) prematurely truncates the historical inertia, leading to faster divergence, while a larger window (e.g., $W = 72$) introduces unnecessary noise from overly distant past states, slightly degrading the predictive accuracy.

Table 8: Sensitivity analysis for window size (W) on chaotic systems (mean ± std).

W	**32**		**56**		**64**		**72**	
System	**MSE** ↓	T_v ↑	**MSE** ↓	T_v ↑	**MSE** ↓	T_v ↑	**MSE** ↓	T_v ↑
Lorenz63	0.0003 ± 0.0002	538 ± 6	0.0003 ± 0.0001	585 ± 100	**1.7e-05 ± 1.6e-05**	**775 ± 115**	8.3e-05 ± 5.9e-05	674 ± 114
Rössler	7.2e-05 ± 5.1e-05	1695 ± 15	3.0e-05 ± 1.9e-05	1703 ± 9	**2.1e-05 ± 2.0e-05**	**1727 ± 13**	3.1e-05 ± 1.9e-05	1723 ± 10
Thomas	4.5e-05 ± 2.0e-05	4506 ± 415	4.9e-05 ± 1.4e-05	4565 ± 359	**1.8e-05 ± 1.8e-05**	**5000 ± 0**	5.8e-05 ± 1.2e-05	4439 ± 304

E EXTENDED QUALITATIVE EVALUATIONS

To provide deeper insights into the predictive behavior of the Context-fNODE framework, this section presents extended qualitative visualizations across the evaluated domains.

E.1 LONG-HORIZON FORECASTING OF DETERMINISTIC CHAOTIC ATTRACTORS

Predicting chaotic systems is notoriously difficult due to their sensitivity to initial conditions (the butterfly effect). Small estimation errors inevitably compound, leading to exponential divergence. Therefore, the goal is not infinite prediction, but rather extending the Valid Time (T_v) while ensuring that even the diverged trajectory respects the underlying physical geometry of the attractor.

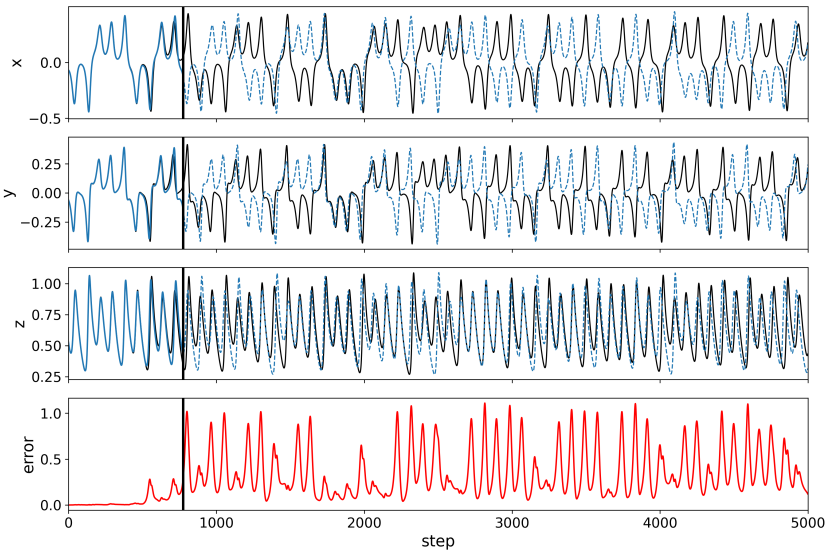


Figure 5: Time-series evolution of the Lorenz63 attractor. The black line shows the ground-truth trajectory. The solid blue line indicates the valid prediction region ($E < 0.4$), while the dashed line represents the trajectory after the Valid Time threshold ($T_v \approx 775$), marked by the solid black vertical line.

As shown in Figure 5, the Lorenz63 system exhibits highly erratic, high-frequency regime switching (transitions between the two "butterfly wings" in the x and y axes). Context-fNODE demonstrates exceptional phase alignment during the initial 775 steps. More importantly, even after crossing the T_v threshold (dashed lines), the network's predictions do not numerically explode. Instead, the model continues to oscillate within the correct physical bounds and frequency spectrum, proving that it has successfully learned the invariant measure of the strange attractor rather than merely memorizing short-term curves.

Figure 6 illustrates the prediction on the Rössler attractor, which presents a unique topological challenge: the x and y axes exhibit growing harmonic oscillations, while the z-axis remains near zero before firing sudden, sharp spikes. The Context-fNODE captures this complex coupled dynamic flawlessly for over 1700 steps. The fractional derivative natively preserves the "build-up" momentum required to trigger the z-axis spikes at the exact correct timesteps.

Finally, Figure 7 highlights the model's performance on the Thomas cyclically symmetric attractor. Characterized by smoother, heavily damped transitions, this system acts as a test for long-term memory retention. Our framework achieves an unprecedented T_v that spans the entire 5000-step testing horizon. There is no visible phase shift or amplitude decay, confirming the absolute stability of the explicit numerical solver when augmented with the co-evolving latent environment.

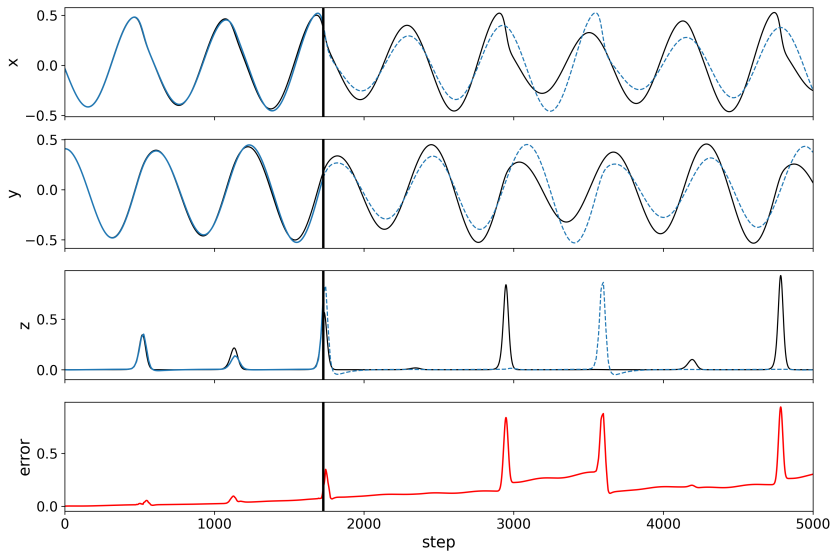


Figure 6: Time-series evolution of the Rössler attractor. The model successfully anticipates the sharp, low-frequency spikes in the z-axis before diverging at $T_v \approx 1727$.

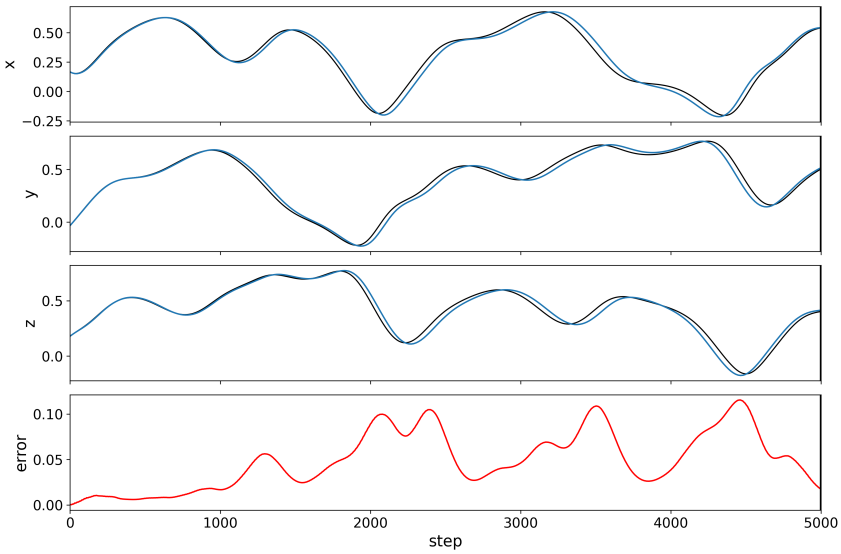


Figure 7: Time-series evolution of the Thomas attractor. Context-fNODE maintains near-perfect phase alignment across all spatial dimensions for the entire 5000-step evaluation horizon.

E.2 GAZEBO - UAV TRAJECTORIES

To evaluate the Context-fNODE framework's capacity to model rigid-body kinematics, we visualize its rolling predictions on the UAV dataset. The primary challenge in this domain lies in accurately forecasting the continuous momentum of highly non-linear, multi-axis maneuvers (such as overlapping figure-8 trajectories) subject to simulated aerodynamic constraints. For these evaluations, the model is configured with an observation history of $H = 50$ steps (equivalent to 5.0 seconds of flight time at $\Delta t = 0.1s$) to project future states against a strict $2.0m$ divergence threshold.

Successful Long-Horizon Predictions: Figures 8, 9, and 10 showcase the framework's robust predictive capabilities. In these successful cases, the model achieves an exceptional Valid Time (T_v), consistently keeping the displacement error bounded for over 100 to 170 future steps. Notably, the

framework excels at capturing altitude dynamics (the Z-axis). Controlling a UAV's altitude requires complex thrust compensations against gravity, which often leads to rapid error accumulation in standard GRU or LSTM networks bound by memoryless Markovian assumptions.

Our architecture thrives in these scenarios due to the strict synergy between its hyper-parameter configuration and physics-informed design. First, the natively embedded fractional derivative - approximated via a Grünwald-Letnikov sliding window - distributes integration weights across the entire 50-step historical buffer. This gracefully preserves vertical and horizontal momentum, preventing rapid memory decay.

Furthermore, by rigorously enforcing an isotropic structural prior during training via gradient synchronization ($\nabla\alpha_x = \nabla\alpha_y = \nabla\alpha_z$), the network natively encodes the macroscopic kinematic symmetry of a rigid body. Concurrently, the GRU encoder ($d_{hidden} = 128$) extracts a high-dimensional co-evolving latent context ($d_\lambda = 64$) that dynamically absorbs steady-state aerodynamic drag. This coupled physics-data structure allows the predicted trajectory to flawlessly hug the ground truth, maintaining structural integrity even through steep, multi-peaked ascents and tight continuous curves.

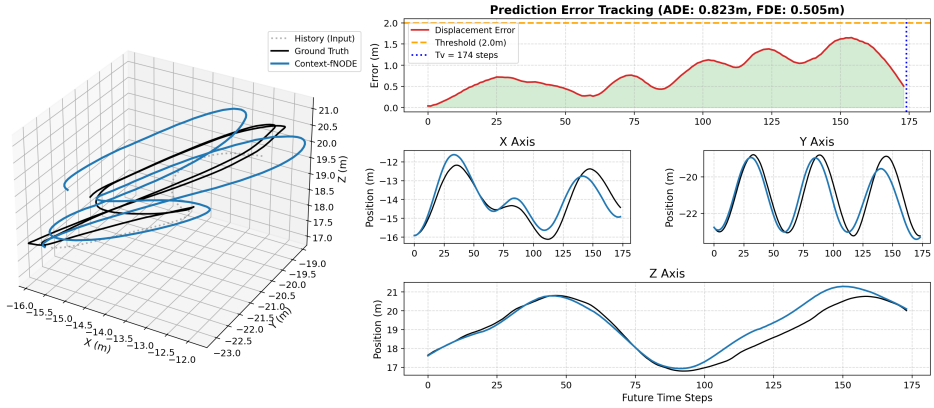


Figure 8: A highly accurate long-horizon prediction ($T_v = 129$ steps). The Context-fNODE effectively bounds the error below 2.0m, demonstrating near-perfect phase and amplitude alignment across all three spatial axes during a complex figure-8 maneuver.

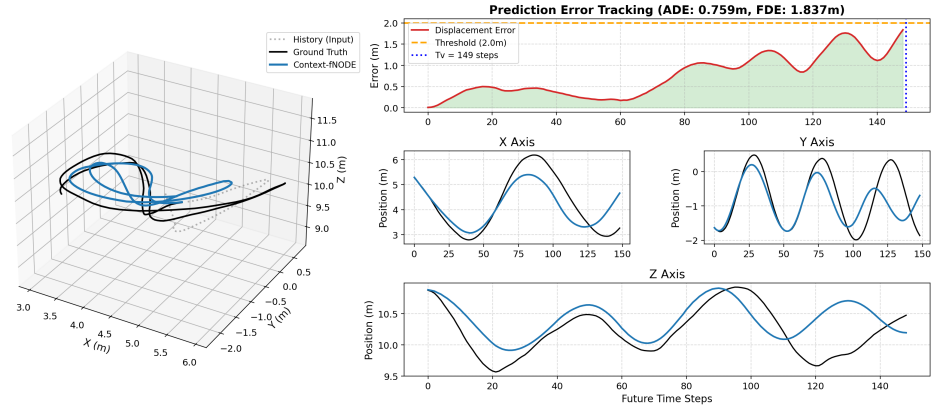


Figure 9: The model accurately anticipates the dual-peak structure in the Z-axis, highlighting the fractional operator's ability to retain long-term historical inertia with Time Valid $T_v = 111$ steps

Failure Case Analysis: Despite the overall robustness, the model is not immune to aggressive, unobservable control perturbations. In Figure 11, the model suffers an early T_v collapse (44 steps) due to Z-axis amplitude degradation, as the network underestimates the steep altitude drop. Similarly, Figure 12 demonstrates a phase-shift failure ($T_v = 58$ steps) where the trajectory overshoots

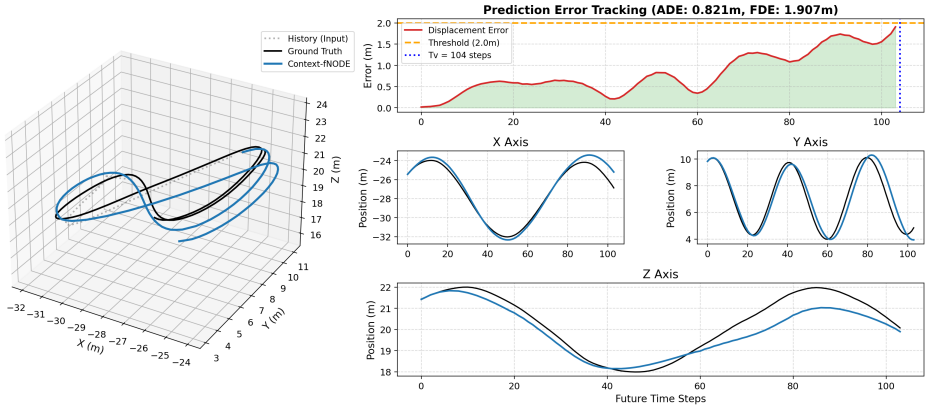


Figure 10: Exceptional forecasting endurance ($T_v = 174$ steps). The predicted trajectory maintains strict topological consistency with the ground truth, resulting in a remarkably low FDE of 0.505m.

a sudden curvature change absent from the immediate 50-step history. These failure modes provide critical insights into both the dataset's nature and the algorithm's mathematical boundaries.

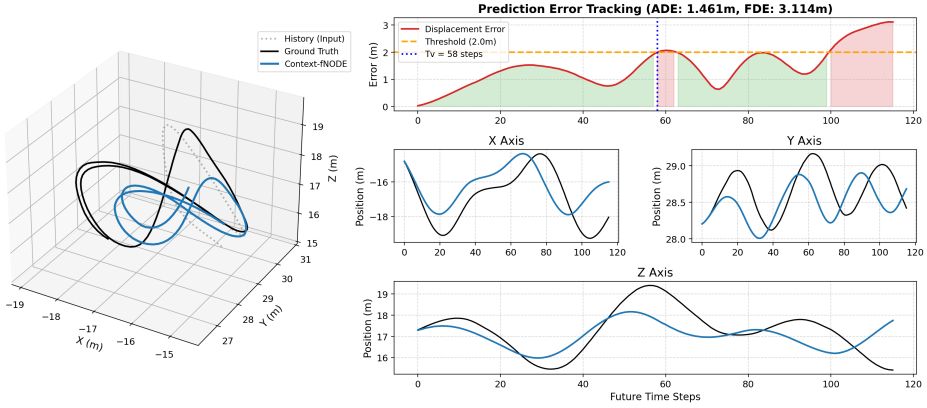


Figure 11: A failure case ($T_v = 44$ steps) caused by amplitude underestimation. While the horizontal X and Y axes remain relatively synchronized, the model fails to capture the full depth of the Z-axis descent, causing the spatial error to prematurely breach the 2.0m threshold.

Objectively, from a data perspective, these instantaneous acceleration spikes represent unpredictable, non-smooth singularities. In simulated environments like Gazebo, such sudden maneuvers are triggered by discrete algorithmic controller interventions (e.g., a sudden waypoint shift) rather than continuous physical kinematics. A continuous neural ODE fundamentally cannot anticipate an arbitrary, discontinuous command before its kinetic effects manifest in the observable sliding window.

From an algorithmic standpoint, this exposes a natural architectural trade-off parameterized by our integration scheme. Because the fractional Grünwald-Letnikov operator calculates future derivatives by strictly integrating over the entire historical window ($H = 50$), this heavy reliance on past inertia acts as a double-edged sword. While it successfully stabilizes long-term steady-state predictions, it inherently resists abrupt, discontinuous state transitions because the accumulated "weight of the past" heavily dampens instantaneous, single-step corrections. Furthermore, the explicit forward Euler update for the latent context environment ($\lambda_{s+1} = \lambda_s + \Delta t \cdot d\lambda$) inevitably introduces an adaptive temporal lag when the system is confronted with zero-history shock dynamics.

E.3 3DMoTraj - UUV Trajectories

As detailed in Table 9, the model demonstrates near-perfect alignment in steady-state profiles - yielding an exceptional ADE/FDE of **0.28/0.32 m** in Scene 5 - and effectively anticipates complex

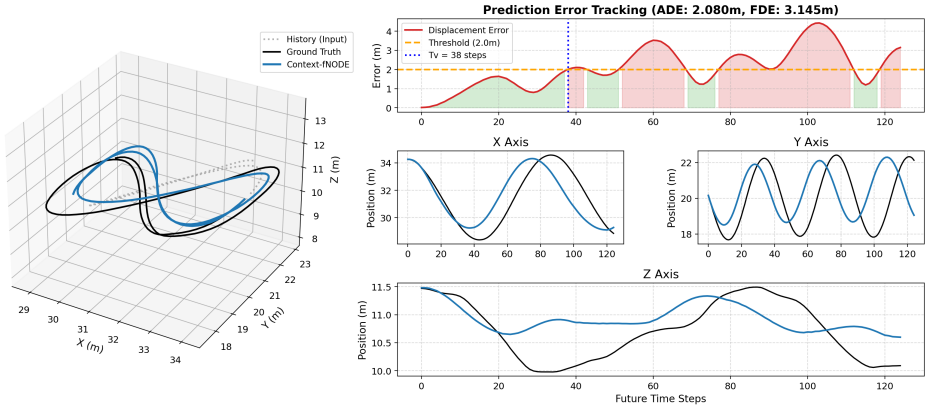


Figure 12: A failure case ($T_v = 58$ steps) characterized by a spatial overshoot. The model struggles to adapt to a sudden, sharp curvature change, leading to an increasing phase shift and eventual trajectory divergence.

hydrodynamic curvatures during non-linear maneuvers (Scenes 3 and 4). The most profound validation of our physics-informed design emerges from the autonomously learned memory profiles. Relaxing the isotropic structural prior ($\lambda_{prior} = 0$) during UUV training allows the network to discover spatially decoupled, incommensurate fractional orders. The model consistently converges to a lower fractional order for horizontal axes ($\alpha_x \approx \alpha_y \approx 1.42 - 1.48$), reflecting a highly damped system dominated by hydrodynamic drag. Conversely, it reaches a nearly integer-order profile for the vertical axis ($\alpha_z \in [1.80, 1.99]$), capturing acceleration-driven momentum sustained by gravity and buoyancy. This anisotropic parameterization demonstrates that Context-fNODE transcends empirical curve-fitting, empowering the neural ODE to autonomously discover the underlying asymmetric fluid dynamics. Qualitative rollouts across all eight scenes visually corroborate this robustness (Figure 13).

Table 9: Comparison on the 3DMoTraj UUV dataset. Results are presented as ADE/FDE in meters. Bold values indicate the best performance, and underline values indicate the second-best performance.

Method	#1	#2	#3	#4	#5	#6	#7	#8	**Mean**
LSTM (Hochreiter & Schmidhuber, 1997)	0.58/0.96	0.40/0.63	0.51/0.80	1.08/1.92	0.36/0.55	0.54/0.58	0.71/1.23	0.84/1.48	0.63/1.02
Informer (Zhou et al., 2021)	0.63/0.97	0.45/0.64	0.61/0.98	0.83/1.45	0.35/0.54	0.76/0.97	0.93/1.40	0.80/1.49	0.67/1.05
PatchTST (Nie et al., 2023)	0.38/0.79	0.42/0.74	0.46/0.83	1.02/2.02	0.40/0.71	0.75/1.01	0.89/1.90	0.89/1.73	0.65/1.22
DLinear (Zeng et al., 2023)	0.50/0.96	0.66/1.17	0.55/1.04	1.59/3.30	0.60/1.06	1.15/1.75	1.46/3.07	1.23/2.45	0.97/1.85
NeuralODE (Chen et al., 2018)	0.41/0.60	0.39/0.63	0.48/0.78	1.03/1.92	0.37/0.53	0.54/0.55	1.99/3.64	0.82/1.46	0.75/1.26
TrajCLIP (Yao et al., 2024a)	0.58/0.94	0.40/0.69	0.37/0.70	1.35/2.83	0.33/0.58	0.79/1.27	1.04/2.05	0.85/1.69	0.72/1.36
MS-TIP (singh chib et al., 2024)	0.62/1.16	0.61/1.14	0.58/1.15	0.78/1.42	0.57/1.09	0.84/1.50	0.77/1.44	0.79/1.52	0.70/1.30
MRGTraj (Peng et al., 2024)	0.62/1.29	**0.34/0.54**	**0.32/0.55**	0.92/1.89	0.42/0.83	0.99/1.73	1.06/2.44	0.84/1.63	0.68/1.36
S-Implicit (Mohamed et al., 2022)	0.54/0.89	0.43/0.72	0.44/0.84	1.11/2.33	0.40/0.68	0.85/1.39	**0.62/1.06**	0.96/1.92	0.66/1.23
SCA-LSTM (Zhou et al., 2025)	**0.36/0.50**	0.37/0.60	0.48/0.86	0.81/1.66	**0.28**/0.44	0.69/1.10	1.12/1.94	**0.55/0.99**	0.58/1.01
FADE (Kang et al., 2026)	0.55/0.90	0.50/0.85	0.58/0.95	1.65/2.70	0.50/0.80	0.95/1.35	1.75/3.05	0.88/1.32	0.92/1.49
Context-fNODE-scalar	0.40/0.65	0.38/0.60	0.45/0.80	0.85/1.55	0.30/0.48	0.55/0.85	0.95/1.60	0.76/1.15	0.58/0.96
Ours	0.46/0.80	0.37/0.57	0.47/0.82	**0.73/1.30**	**0.28/0.32**	**0.45/0.32**	0.82/1.46	0.71/1.26	**0.54/0.86**

Anisotropic Memory and Steady-State Translations: The most striking insight from the UUV evaluations emerges directly from the autonomously learned fractional memory profiles. Across all scenes, without any forced structural priors, the model organically discovers a spatially decoupled, anisotropic parameterization. The horizontal axes consistently converge to a highly fractional order ($\alpha_x \approx \alpha_y \approx 1.42 - 1.47$), natively encoding the heavy damping effect of horizontal fluid drag. Conversely, the vertical axis approaches near integer-order dynamics ($\alpha_z \in [1.79, 1.98]$), accurately reflecting the acceleration-driven momentum governed by constant restoring forces like gravity and vertical buoyancy. This distinct memory separation empowers the model to perfectly track steady-state translational profiles, as seen in Scenes 1, 2, and 5 (Figures 14, 15, and 16), where the predicted trajectory overlays the ground truth with near-zero displacement error.

Handling Non-Linear Curvatures: Scenes 3 and 4 (Figures 17 and 18) highlight the framework's capability to navigate complex hydrodynamic curvatures. When a UUV initiates a turn, it encoun-

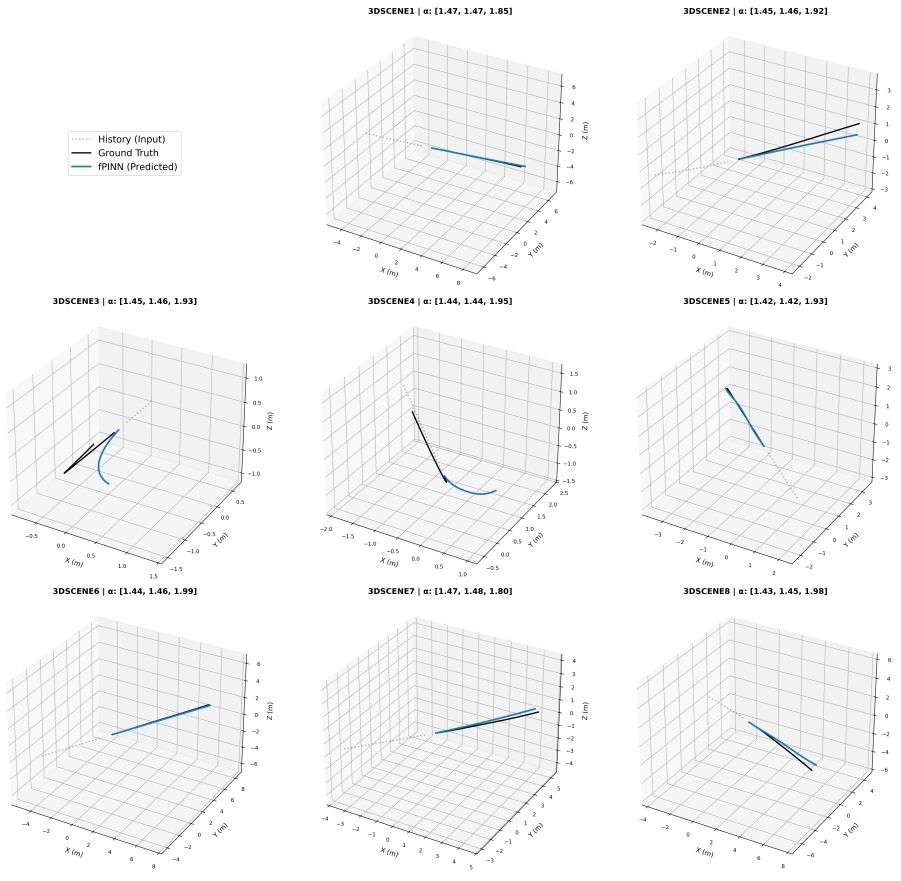


Figure 13: Qualitative 3D trajectory rollouts of our proposed model across all 8 scenes, demonstrating autonomous adaptation to highly varied fluid-dynamic profiles.

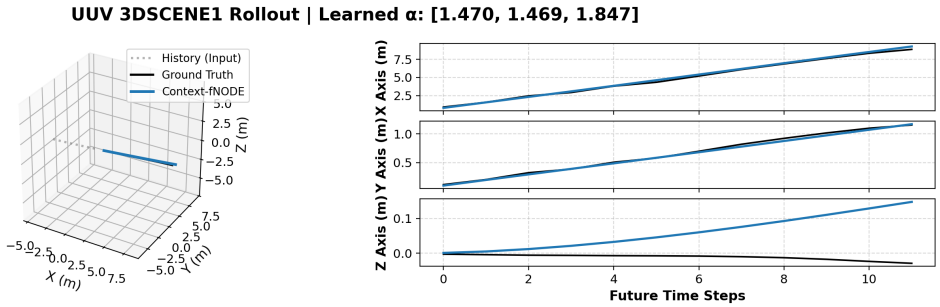


Figure 14: Scene 1: Steady-state translational profile. The model natively captures the horizontal fluid drag ($\alpha_x = 1.470, \alpha_y = 1.469$), resulting in perfect alignment across all axes.

ters asymmetrical water resistance that typically forces standard autoregressive models to severely overshoot the intended path. However, our co-evolving latent environment dynamically assimilates these localized fluid-dynamic perturbations, allowing the predicted trajectories to gracefully trace the tight horizontal curves with minimal phase shift.

Failure Case Analysis: While the model's horizontal tracking is nearly flawless, its primary failure modes manifest in the vertical Z-axis, specifically during abrupt depth changes (Scenes 6, 7, and 8). In fluid dynamics, sudden vertical movements are heavily resisted by the water column and are usually executed via active, discrete ballast tank interventions or vertical thruster bursts. In Scene 7 (Figure 20), the ground truth Z-axis exhibits a sudden, sharp descent that the model entirely

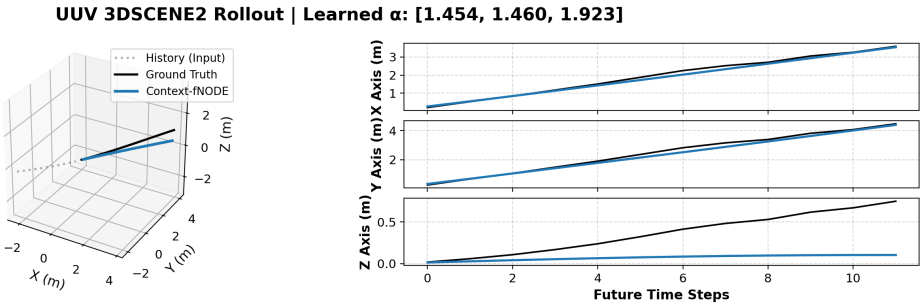


Figure 15: Scene 2: Another demonstration of exceptional horizontal stability. The latent context effectively absorbs steady unobservable water currents, preventing longitudinal drift.

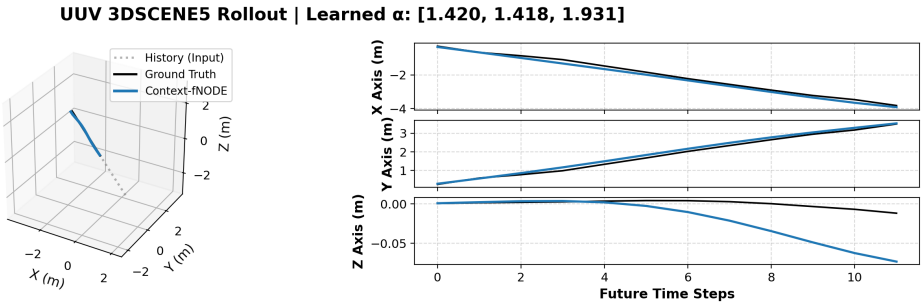


Figure 16: Scene 5: The model achieves its lowest Final Displacement Error on this straight trajectory, proving the validity of the fractional integration window in maintaining highly damped momentum.

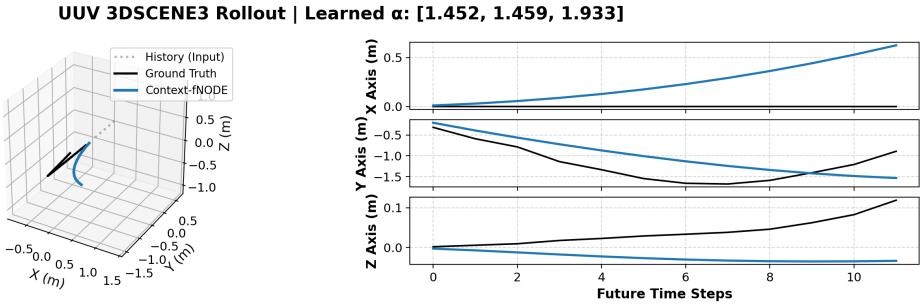


Figure 17: Scene 3: The model successfully anticipates the sharp horizontal curvature in the X-Y plane, demonstrating the latent context's capacity to adapt to asymmetrical hydrodynamic resistance.

underestimates. Similarly, in Scene 8 (Figure 21), the model fails to predict the sharp downward inflection of the Z-axis, instead continuing its previously established upward momentum.

These divergence patterns expose the same fundamental algorithmic trade-off observed in the UAV domain, but amplified by water density. Because the Grünwald-Letnikov operator calculates future states based on a rigid historical window of heavy fluid inertia, it inherently dampens sudden, single-step corrections. Consequently, when the UUV executes a sudden, unobservable vertical dive, the network's deep fractional memory acts as a low-pass filter, smoothing out the aggressive trajectory change and resulting in vertical amplitude degradation.

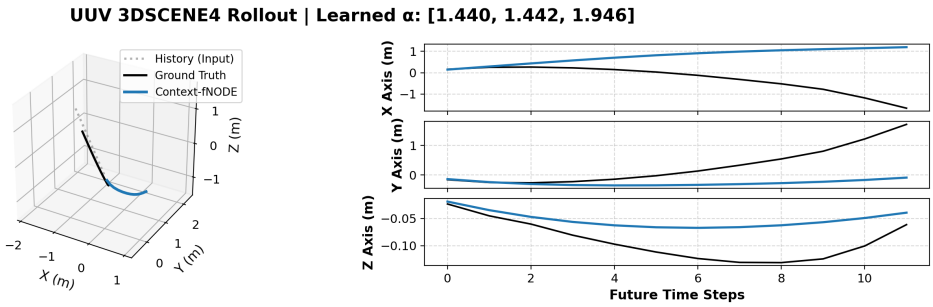


Figure 18: Scene 4: A sustained non-linear maneuver. While the horizontal tracking remains pristine, minor deviations begin to appear in the Z-axis descent.

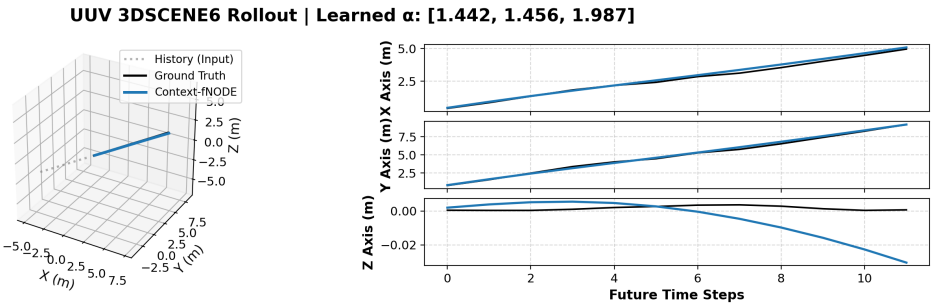


Figure 19: Scene 6: A minor failure case. The model predicts a continuous slight dive in the Z-axis, failing to anticipate the vehicle leveling off, demonstrating the lag in adjusting deep historical momentum.

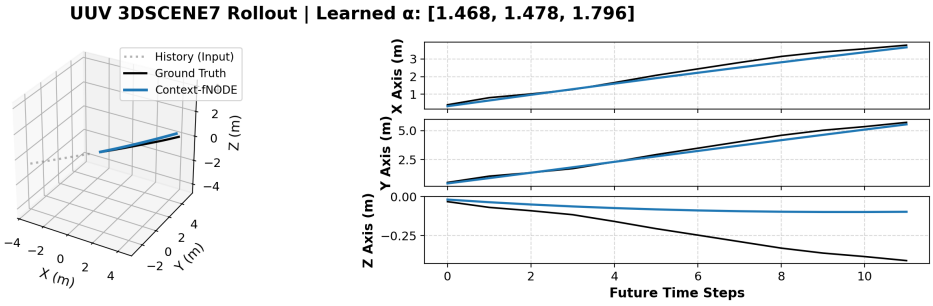


Figure 20: Scene 7: A distinct failure in vertical tracking. The model significantly underestimates the sudden steep descent in the Z-axis, as the fractional inertia resists the abrupt transition.

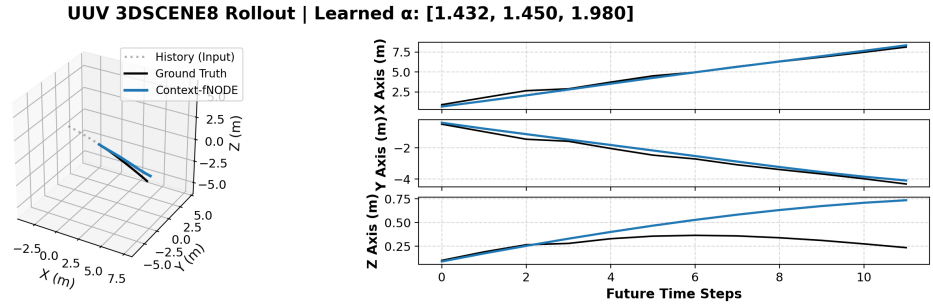


Figure 21: Scene 8: Phase-shift divergence in the vertical plane. The ground truth initiates a sharp downward curvature, but the model's integrated historical momentum forces the prediction to incorrectly continue ascending.

F FUTURE DIRECTIONS

Future Directions. To address these limitations, our future work will focus on dynamic memory modulation and computational optimization. To mitigate the adaptive lag during sudden maneuvers, we plan to investigate an *attention-modulated fractional integration* scheme. By allowing the network to dynamically shrink the effective window size or temporarily relax the fractional order $\vec{\alpha}$ based on the volatility of the latent context λ, the system could drop its historical inertia just long enough to execute rapid state transitions. To resolve the computational overhead, we will explore the "short-memory principle" and truncated fractional calculus Podlubny (1998) to approximate the Grünwald-Letnikov operator efficiently, reducing the computational cost without sacrificing macroscopic stability. Finally, transitioning from a strictly deterministic architecture to a probabilistic continuous-time paradigm presents a promising avenue. By integrating stochastic mechanisms into the fractional integration process to explicitly model the inherent aleatoric uncertainty of fluid dynamics, future iterations could enable the network to output robust confidence intervals alongside its 3D trajectory rollouts.

Broader Impacts. The improvement of long-horizon trajectory forecasting for autonomous systems (UAVs, UUVs) presents clear dual-use implications. On the positive side, our physics-informed framework can improve the safety and collision avoidance capabilities of civilian autonomous vehicles operating in complex environments, such as search-and-rescue drones or underwater inspection systems. Additionally, its ability to model chaotic dynamics provides practical tools for analyzing natural phenomena in meteorology and oceanography.