TRACE: PARAMETER-EFFICIENT TIME SERIES FORECASTING VIA TRANSIENT CHAOTIC SYSTEM

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series forecasting plays a pivotal role in numerous real-world domains, ranging from climate prediction to financial engineering. Over the past decades, neural networks, as a data-driven approach, have dominated the field and have yielded significant performance. However, how these black-box architectures internalize temporal structures within latent representations remains a fundamental mystery. In this work, we observe that trained time-series forecasters propagate input perturbations with positive finite-time sensitivity, reminiscent of transient chaos. These observations suggest that the transient chaos emerging in trained forecasters can be explicitly used as a fixed, non-trainable transient chaotic system to process information. Motivated by this insight, we propose a TRAnsient Chaotic Evolution network (TRACE), which embeds temporal representations into the phase space of continuous-time chaotic oscillators and constrains their evolution through a causal coupling topology among time steps. In TRACE, the nonlinear temporal transformation is therefore performed by the fixed autonomous flow rather than by additional learnable layers, substantially reducing the number of trainable parameters. We provide rigorous geometric derivations that theoretically substantiate TRACE's effectiveness, and extensive empirical analyses further confirm that this framework naturally satisfies the geometric proposition. Furthermore, TRACE delivers competitive forecasting performance against current state-of-the-art models while providing physical interpretability. The code is available at `https://anonymous.4open.science/r/TRACE-6776`.

1 INTRODUCTION

Time Series Forecasting (TSF) is crucial for strategic decision-making in complex systems, such as quantitative finance and intelligent transportation (Wang et al., 2026). Methodologically, the field has evolved from classical statistical models to Deep Neural Networks (DNNs), including Recurrent Neural Networks (RNNs), Transformers, and Graph Neural Networks (GNNs) (Wang et al., 2026; Zhao et al., 2026). Although recent paradigms favor lightweight structures like Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) for computational efficiency, these models still fundamentally rely on purely data-driven, black-box nonlinear transformations (Wang et al., 2026). Consequently, existing frameworks inevitably face a severe trade-off between parametric compactness and representational capacity, a bottleneck that complicates their highly volatile environments and deployment in resource-constrained settings(Wang et al., 2026; Zhao et al., 2026).

A unifying bottleneck behind these limitations is that existing architectures must ***learn*** the required nonlinear temporal transformation from scratch through extensive parameterization. Rather than learning this transformation, we wonder what dynamical property emerges in the intermediate representations of a well-trained forecaster. A natural way to probe such properties is to examine how sensitive the hidden-state trajectory is to small input perturbations. This sensitivity is traditionally quantified by Maximum Lyapunov Exponent (MLE), which measures the asymptotic exponential divergence of nearby trajectories (Wolf et al., 1985). However, a deep network performs only a finite number of layer-wise updates; its forward propagation is therefore more appropriately characterized by the Finite-Time Maximum Lyapunov Exponent (FTMLE) (Storm et al., 2024).

Accordingly, we investigate the hidden-state dynamics of mainstream backbones for TSF tasks, including MLP-, Transformer-, and CNN-based forecasters, by tracking the propagation of small

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.

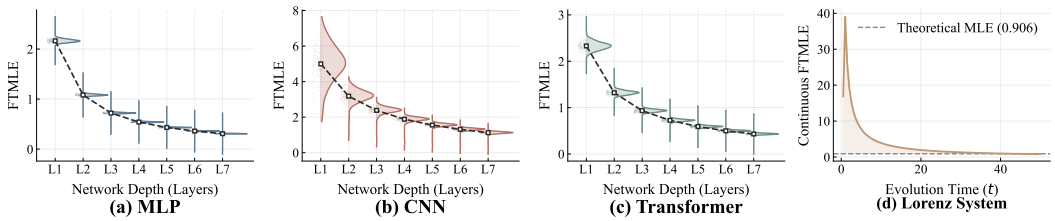


Figure 1: Layer-wise probability distributions of the Finite-Time Maximum Lyapunov Exponent (FTMLE) across mainstream TSF backbones. The ridges track the evolution of FTMLE from Layer 1 to Layer 7 under infinitesimal input perturbations for (a) MLP, (b) CNN, and (c) Transformer architectures, showing a tendency toward a positive finite-depth regime across the examined layers. (d) FTMLE trajectory of the classical Lorenz system along the time, exhibiting a similar pattern.

input perturbations. As shown in Fig. 1, the FTMLE values remain consistently positive across all architectures and converge asymptotically to a stable limit in deeper layers, a pattern similar to that of chaotic systems. This indicates that a well-trained model propagates infinitesimal differences in a manner analogous to chaotic systems, which are intrinsically sensitive to initial conditions. Moreover, since the forward pass has finite depth and such property may not persist indefinitely, this behavior is more accurately described as ***transient chaotic dynamics***. These observations suggest a different route: instead of learning nonlinear temporal transformations, we can directly employ a fixed dynamical system that intrinsically exhibits transient chaotic dynamics to perform nonlinear representation processing.

Given that chaos-like dynamics naturally emerge as an implicit mechanism for deep temporal processing, it is surprising that existing parameter-efficient architectures fail to explicitly utilize this property. The closest paradigm, Reservoir Computing (RC), offloads temporal processing to a fixed dynamical system (Kohei, 2018; Jaeger & Haas, 2004). However, its foundational Echo State Property (ESP) requires the system to asymptotically forget initial conditions, fundamentally clashing with the long-term sensitivity of chaos (Gallicchio, 2018).

To this end, we propose TRACE (TRAnsient Chaotic Evolution network), a paradigm shift from learning nonlinearity to leveraging deterministic transient chaotic Ordinary Differential Equations (ODEs). TRACE embeds a controlled, fixed transient chaotic system into the network as the primary engine of temporal processing. Inputs are embedded as the initial states, which then evolve autonomously according to the system's ODEs with a fixed inter-node coupling structure, where no learnable parameters are required. TRACE queries the system states only at a specific evolution time. Only the input embedding and readout are trainable. The primary contributions are summarized as follows:

- To the best of our knowledge, TRACE is the first time-series forecasting framework to use finite-time states of a fixed transient chaotic ODE as the main nonlinear representation processing core. By capturing the physical system state at an optimal evolution time, TRACE drastically reduces parameter overhead while achieving competitive performance against mainstream deep learning backbones.
- We establish a robust empirical link between data processing capacity and dynamical phase transitions. By exploring evolution hyperparameters, the results suggest that forecasting benefits from an intermediate expansion regime. Insufficient expansion may limit feature separation, whereas excessive expansion can distort local temporal information. This finding provides a principled, physics-informed guideline for system configuration.
- We show that the effectiveness of TRACE stems from the underlying dynamical mechanism rather than any specific choice of chaotic equations through experiments. By instantiating with diverse chaotic oscillators (e.g., Lorenz, Rössler), TRACE consistently delivers significant performance improvements across all application scenarios.
- We provide a geometric interpretation of the chaotic core's working mechanism through the lens of stretch-and-fold dynamics. We elucidate how global volume contraction and

local state expansion jointly facilitate temporal feature separation and redundancy removal, empirically validating this dual mechanism.

2 RELATED WORK

Deep Learning for TSF. Early deep learning for TSF primarily relied on recurrent networks like LSTM (Hochreiter & Schmidhuber, 1997) and GRU (Gulcehre et al., 2022). Following the Transformer (Vaswani et al., 2017), numerous attention-based architectures emerged, including Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022), and Conformer (Li et al., 2023). Recent paradigms further advance via sub-series tokenization, such as patch segmentation in PatchTST (Nie et al., 2023) and phase-domain semantics in PhaseFormer (Niu et al., 2026). Although representationally powerful, Transformers inherently incur substantial parameter overhead and computational bottlenecks.

To alleviate the overhead, a prominent shift toward structural simplicity has emerged. Notably, DLinear (Zeng et al., 2023) demonstrated that a single linear layer can outperform complex Transformers, spurring ultra-lightweight architectures like FITS (Xu et al., 2024) and SparseTSF (Lin et al., 2026). CNN-based structures like FACT (Qi et al., 2026) and orthogonal shifts like OLinear (Yue et al., 2026) further achieve SOTA accuracy with minimal parameters as well. However, these streamlined frameworks remain data-driven black boxes, offering limited transparency into their internal temporal representations. Current interpretability efforts, such as attention map visualizations (Rai et al., 2024), remain post-hoc and lack a principled mechanistic account. TRACE addresses this gap by grounding computation directly in the transparent evolution of coupled chaotic oscillators, rendering information processing fully traceable in phase space.

Neural Networks with Chaos. The integration of chaos into Artificial Neural Networks (ANNs) generally spans three paradigms. First, *microscopic modification* embeds chaotic dynamics directly into individual neurons. Examples include Aihara's model (Aihara et al., 1990), which uses refractory mechanisms for associative memory, and Transiently chaotic Neural Networks (TcNN) (Chen, 2011), which leverage decaying self-feedback to escape local minima. Similarly, NeuroChaos Learning (NL) (Harikrishnan & Nagaraj, 2019; Balakrishnan et al., 2019; Harikrishnan & Nagaraj, 2020) employs chaotic maps to generate complex states, extracting statistical features (e.g., entropy, firing time) for downstream classification. Second, *stochastic regularization* utilizes chaotic noise injection (Hopfield, 1982) or chaotic weight initialization (Jia et al., 2024) to enhance global optimization and network security. Third, *specialized structural variants* employ chaos for domain-specific tasks, such as WinnerLess Competition Networks (WLCNs) (Rabinovich et al., 2001) for sequential patterns, fractional-order systems (Chen et al., 2020) for encryption, fault detection frameworks (Przystałka & Moczulski, 2015), and memristive architectures (Lin et al., 2022) for hardware-driven hyperchaotic communications.

In contrast to these paradigms that embed chaos deep inside neural operations, TRACE encapsulates chaotic dynamics as a modular, plug-in core that interacts with the surrounding network only through initial conditions and final evolved states. This decoupled design provides structural interpretability and flexibility, which is a strategy largely unexplored in TSF.

3 METHODOLOGY

3.1 PROBLEM DEFINITION

Let $\mathbf{X} \in \mathbb{R}^{N \times T}$ denote a time series with N variates and T time steps. When $N > 1$, the series is called Multivariate Time Series (MTS); otherwise, it is referred to as Univariate Time Series (UTS). Given a historical sequence $\mathbf{X}$ with T timestamps, the task is to predict the future τ steps $\mathbf{Y} = \{X_{T+1}, \ldots, X_{T+\tau}\} \in \mathbb{R}^{N \times \tau}$. Our goal is to approximate the ground truth $\mathbf{Y}$ as closely as possible with predictions $\hat{\mathbf{Y}}$.

3.2 THE FRAMEWORK OF TRACE

As shown in Fig. 2, TRACE is a minimalist architecture that offloads temporal modeling to a fixed, physically-informed chaotic dynamical system. The framework consists of four stages: (i) the input

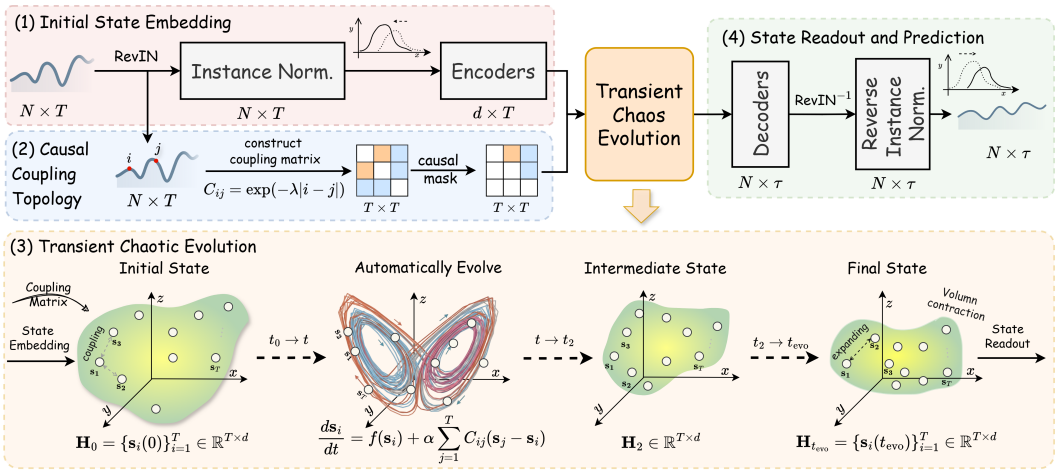


Figure 2: The overall framework of TRACE consists of four stages. (1) Initial State Embedding: the input series is normalized and encoded into the phase space of a chaotic oscillator as the initial state. (2) Causal Coupling Topology: a structured, time-decaying coupling matrix with a causal mask constrains inter-node interactions. (3) Transient Chaos Evolution: the coupled system evolves autonomously via the oscillator's ODE and coupling. (4) State Readout: the evolved state is decoded and de-normalized to produce the final forecast.

series is embedded into the phase space of a chosen chaotic oscillator, yielding an initial state $\mathbf{H}_0$; (ii) a causal coupling matrix is constructed to influence interactions among time steps/nodes; (iii) the coupled system evolves autonomously over a specified time interval, driven by the oscillator's ODE and the coupling topology; (iv) the evolved state is readout via a lightweight network to produce the final forecast. Only the embedding and readout stages are trainable.

(i) Initial State Embedding. To mitigate distribution shift, TRACE first applies Reversible Instance Normalization (RevIN) (Kim et al., 2022). The normalized sequence is mapped by a trainable MLP to initial conditions in the phase space of the selected oscillator

$$\mathbf{H}_0 = \text{Encoder}(\text{RevIN}(\mathbf{X})) \in \mathbb{R}^{d \times T}, \tag{1}$$

where d is the intrinsic dimension of the chosen chaotic oscillator (e.g., $d = 3$ for Lorenz system).These encoded vectors are initial conditions. And the encoder is an MLP with L_{enc} layers and d_{enc} hidden dimension. Each time step is thus embedded as a point in the oscillator's phase space, serving as the initial condition for evolution. We refer to each time step as a node (totally T nodes), and the dynamics system instantiated at each node as a chaotic oscillator.

(ii) Causal Coupling Topology. A key inductive bias in time series modeling is temporal causality. Past time steps can influence future ones, but not vice versa. To encode this principle, we introduce a structured coupling matrix $C \in \mathbb{R}^{T \times T}$ that impacts how nodes interact during evolution. Each entry C_{ij} specifies the influence strength from node j to node i, defined as

$$C_{ij} = M_{ij} \cdot \exp(-\lambda|i-j|), \tag{2}$$

where λ is the decay rate, and $M_{ij} \in \{0, 1\}$ is a causal mask that enforces the temporal direction

$$M_{ij} = \begin{cases} 1, & \text{if } j < i, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

(iii) Transient Chaotic Evolution. The dynamic system evolves autonomously in continuous time according to the coupled ODE, once the initial state $\mathbf{H}_0 = \{\mathbf{s}_i(0)\}_{i=1}^T$ and the coupling matrix C are in place. The dynamics of each node $\mathbf{s}_i \in \mathbb{R}^d$ are governed by its dynamic ODE f and the coupling from other nodes

$$\frac{d\mathbf{s}_i}{dt} = f(\mathbf{s}_i) + \alpha \sum_{j=1}^{T} C_{ij}(\mathbf{s}_j - \mathbf{s}_i), \tag{4}$$

where α controls the global coupling strength. The system is integrated numerically over a predefined evolution time t_{evo} such as RK4 (Kutta, 1901) or Euler (Euler, 1824), to obtain

$$\mathbf{H}_{t_{\text{evo}}} = \Phi_{t_{\text{evo}}}(\mathbf{H}_0), \tag{5}$$

where $\Phi_{t_{\text{evo}}}$ is the flow of the dynamical system to t_{evo} and $\mathbf{H}_{t_{\text{evo}}} = \{\mathbf{s}_i(t_{\text{evo}})\}_{i=1}^{T}$ is the final state.

This autonomous evolution acts as a computational core. The expansion of some nodes and contraction of phase space naturally separate or merge representations, analogous to what deep networks learn to do through stacked layers, but achieved without any learned parameters. We have provided relative proofs in Proposition 4.1

(iv) State Readout and Prediction. After evolution, final state $\mathbf{H}_{t_{\text{evo}}}$ is the result of nonlinear temporal processing. Then, the prediction is obtained through simple readout

$$\hat{\mathbf{Y}} = \text{RevIN}^{-1}(\text{Decoder}(\mathbf{H}_{t_{\text{evo}}})). \tag{6}$$

The decoder is an MLP with L_{dec} layers and d_{dec} hidden dimension.

Only the encoder and decoder are trained while heavy computation has been offloaded to the fixed chaotic dynamics. This design yields a highly parameter-efficient model.

4 THEORETICAL ANALYSIS

This section establishes the formal theoretical foundations of TRACE. First, we elucidate the underlying efficacy of the framework from a geometric perspective, revealing how chaotic dynamics yield robust feature representations through simultaneous global dissipation and local expansion. The comparative analysis about parameter complexity and the formal derivation of empirical risk bounds are deferred to the Appendix A.2.

4.1 Geometric Manifold Contraction and Expansion

The representational power of TRACE is fundamentally rooted in the dual geometric properties of the dissipative chaotic system. The global volume contraction of the entire phase space and the expansion of inter-node distances. This dynamic creates a stretch-and-fold mechanism, effectively separating temporal features among nodes while globally suppressing irrelevant noise.

Proposition 4.1 (Stretch-and-Fold Mechanism) *During a finite evolution time t_{evo}, the physical dynamics flow $\Phi_{t_{\text{evo}}}$ simultaneously imposes a volume contraction in the Td-dimensional global phase space and a localized spatial separation between node representations, thereby optimizing the feature distribution for readout.*

Proof of Proposition 4.1. When convenient, we flatten the global state and denote as a vector $\mathbf{H} \in \mathbb{R}^{Td}$ in evolution stage, governed by the continuous vector field

$$\mathbf{F}(\mathbf{H}) = \mathbf{f}(\mathbf{H}) - \alpha(\mathbf{L} \otimes \mathbf{I}_d)\mathbf{H}, \tag{7}$$

where $\mathbf{f}(\mathbf{H}) = \{f(\mathbf{s}_i)\}_{i=1}^{T}$ denotes the stacked vector field of uncoupled intrinsic chaotic dynamics across all individual nodes and $\mathbf{L}$ is the graph Laplacian constructed from C, $\otimes$ denotes the Kronecker product, and $\mathbf{I}_d \in \mathbb{R}^{d\times d}$ is the identity matrix.

Global Contraction (Redundancy Removal). The phase space volume under $\Phi_{t_{\text{evo}}}$ is determined by the determinant of its global Jacobian matrix. According to Jacobi's formula and the variational equation, the volume change rate is linked to the divergence of the vector field

$$\frac{d}{dt}\log|\det \mathcal{J}_t| = \text{tr}(\mathcal{J}_t^{-1}\dot{\mathcal{J}}_t) = \text{tr}(D\mathbf{F}(\mathbf{H}_t)) = \nabla\cdot\mathbf{F}(\mathbf{H}_t), \tag{8}$$

where $\mathcal{J}_t = D\Phi_t(\mathbf{H}_0)$ is the Jacobian matrix of the flow map at time t, $\dot{\mathcal{J}}_t$ denotes its time derivative, $\text{tr}(\cdot)$ is the trace operator, $D\mathbf{F}(\mathbf{H}_t)$ is the Jacobian matrix of the global vector field $\mathbf{F}$ evaluated at $\mathbf{H}_t$, and $\nabla\cdot\mathbf{F}$ denotes its divergence.

Integrating over t_{evo} yields

$$|\det \mathcal{J}_{t_{\text{evo}}}| = \exp\left(\int_0^{t_{\text{evo}}} \nabla\cdot\mathbf{F}(\mathbf{H}_t)\,dt\right), \tag{9}$$

where $\det$ denotes the determinant of a matrix and $\mathcal{J}_{t_{\text{evo}}} \in \mathbb{R}^{Td\times Td}$ is a square matrix whose determinant is well-defined.

For globally dissipative systems (i.e. Lorenz), the intrinsic divergence is strictly negative (Callier, 2007). Furthermore, the trace of the Laplacian matrix $\operatorname{tr}(L) > 0$ contributes an additional negative constant $-\alpha d \operatorname{tr}(L)$ to the global divergence. Consequently, there exists a constant $c > 0$ such that

$$\nabla \cdot \mathbf{F}(\mathbf{H}_t) \leq -c < 0, \tag{10}$$

and therefore $|\det \mathcal{J}_{t_{\text{evo}}}| \leq e^{-ct_{\text{evo}}} < 1$, guaranteeing exponential volume contraction. This contraction may suppress certain global variations, although volume contraction alone does not guarantee task-relevant denoising.

Local node Expansion (Feature Separation). Despite the global collapse of the phase space volume, the relative distances among individual node states are not necessarily compressed.

To analyze inter-node distances, we utilize the Master Stability Function (MSF) framework (Pecora & Carroll, 1998). We project the inter-node differences onto the eigenvectors of the Laplacian matrix $\mathbf{L}$, obtaining decoupled transverse modes ξ_k. The evolution of the k-th mode (associated with eigenvalue $\lambda_k > 0$) can be approximated by its Transverse Lyapunov Exponent (TLE):

$$\Lambda_k(t_{\text{evo}}) \approx \lambda_{\max}^{\text{local}}(t_{\text{evo}}) - \alpha\lambda_k, \tag{11}$$

where $\lambda_{\max}^{\text{local}}(t_{\text{evo}})$ is the maximum Lyapunov exponent of an uncoupled chaotic node.

Crucially, chaotic systems intrinsically possess stretching directions characterized by $\lambda_{\max}^{\text{local}} > 0$, which drive the exponential divergence of nearby trajectories. The coupling term $-\alpha\lambda_k$ acts as a mode-dependent damping that counteracts this expansion. For modes where the damping is weaker than the intrinsic stretching, i.e., $\alpha\lambda_k < \lambda_{\max}^{\text{local}}$, the net transverse exponent remains positive ($\Lambda_k > 0$), actively expanding the corresponding inter-node differences $\|\mathbf{h}_i - \mathbf{h}_j\|$ rather than collapsing. This selective preservation prevents the system from settling into a trivial synchronized state, ensuring a diverse set of node representations throughout the evolution window.

In addition, we emphasize that global volume contraction and local distance expansion are not contradictory, but rather complementary consequences of anisotropic dissipation inherent to chaotic flows. Volume contraction is governed by the product of expansion rates across all Td directions of phase space, while the stretching of inter-node distances is dominated by a small subset of unstable transverse modes that individually experience positive growth, thereby achieving the dual geometric objectives of global compression and local separation. The geometric prediction is empirically evaluated in Section 5.3.

5 EXPERIMENT

5.1 EXPERIMENTAL SETUP

Datasets. We evaluate TRACE on 8 widely adopted MTS datasets across sampling frequencies, and dimensions. For long-term forecasting, we use the ETT benchmarks (ETTh1, ETTh2, ETTm1, ETTm2), PEMS08 and Weather. For short-term forecasting, we adopt ILI and Wiki. These dataset description details are presented in Appendix A.3.

Baselines. We compare TRACE against 11 baselines including DLinear (Zeng et al., 2023), FITS (Xu et al., 2024), SparseTSF/MLP (SparseTSF) (Lin et al., 2026), and MixLinear (Ma et al., 2026) for linear-based models; ModernTCN (Luo & Wang, 2024), FACT (Qi et al., 2026), and TimesNet (Wu et al., 2023) for CNN-based architectures; PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024), and TimeMixer++ (Wang et al., 2025) for Transformer-based methods; and interPDN (Kong & Hong, 2026) as a hybrid baseline.

Chaotic Oscillators. TRACE supports a variety of continuous-time chaotic oscillators as its computational core. In this work, we instantiate with the Lorenz system (Lorenz, 2004), a prototypical examples of low-dimensional deterministic chaos. Lorenz system describes a simplified model of atmospheric convection. Following standard practice, we set $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$, placing the system in its classic chaotic regime. Under these parameters, the Lorenz attractor exhibits two unstable saddle foci, causing trajectories to spiral around one lobe before intermittently switching to

Table 1: Average experimental results in Long-Term Forecasting

Categories	Linear-Based										Transformer-Based								CNN-Based						Hybrid	
Models	**TRACE** (Ours)		DLinear (2023)		FITS (2024)		SparseTSF (2026)		MixLinear (2026)		PatchTST (2023)		iTrans. (2024)		TimeMixer++ (2025)		TimesNet (2023)		ModernTCN (2024)		FACT (2026)		interPDN (2026)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.3844**	**0.3819**	0.3960	0.3879	0.4925	0.4520	0.4160	0.3993	0.4850	0.4383	0.3878	0.3891	0.3976	0.3928	0.3850	0.3821	0.3920	0.3983	0.3906	0.3978	0.3970	0.3926	0.3872	0.3843
ETTm2	0.2719	0.3140	0.2853	0.3307	0.2910	0.3333	0.2941	0.3287	0.2940	0.3304	0.2850	0.3225	0.2837	0.3230	0.2766	0.3176	0.2932	0.3274	**0.2696**	0.3151	0.2733	**0.3138**	0.2772	0.3180
ETTh1	**0.4344**	0.4251	0.4522	0.4477	0.4400	0.4305	0.4476	**0.4242**	0.4426	0.4249	0.4484	0.4368	0.4507	0.4444	0.4498	0.4310	0.4707	0.4533	0.4406	0.4261	0.4527	0.4341	0.4422	0.4304
ETTh2	0.3690	0.3926	0.3861	0.4144	0.3758	0.3983	0.3783	0.3954	0.3874	0.4026	0.3828	0.4015	0.3790	0.4002	**0.3667**	0.3929	0.3980	0.4127	0.3760	0.3949	0.3723	0.3936	0.3693	**0.3913**
Weather	0.2425	0.2649	0.2694	0.2905	0.2733	0.2920	0.2609	0.2780	0.2921	0.3079	0.2585	0.2741	0.2632	0.2780	**0.2263**	0.2623	0.2585	0.2799	0.2362	0.2612	0.2358	**0.2580**	0.2463	0.2642
PEMS08	0.2012	0.2740	0.3712	0.3978	0.5340	0.4873	0.3559	0.3702	0.6510	0.5503	0.3350	0.3737	0.2344	0.3048	0.2013	0.2795	0.2788	0.3475	0.2650	0.3688	**0.1419**	**0.2333**	0.3520	0.3704
Counts	**5**	**4**	0	0	1	0	0	1	0	1	0	0	0	0	3	1	0	0	1	1	2	3	0	0

Table 2: Average experimental results in Short-Term Forecasting

Categories		Linear-Based										Transformer-Based								CNN-Based						Hybrid	
Models		**TRACE** (Ours)		DLinear (2023)		FITS (2024)		SparseTSF (2026)		MixLinear (2026)		PatchTST (2023)		iTrans. (2024)		TimeMixer++ (2025)		TimesNet (2023)		ModernTCN (2024)		FACT (2026)		interPDN (2026)	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ILI	@12	**1.3363**	0.6611	2.6494	1.1033	2.8598	1.0688	2.1553	0.8502	1.6381	0.6825	1.6810	0.7230	1.3894	0.6527	1.6031	0.6941	1.5460	0.7167	1.3930	**0.6441**	1.6357	0.7821	1.8943	0.8035
	@36	**1.6697**	**0.7953**	3.2282	1.2446	4.1300	1.4655	4.4683	1.3580	3.0650	1.1840	2.1285	0.8848	2.2499	0.9568	2.0477	0.8530	2.4535	0.9322	2.0335	0.8552	2.1623	0.9023	2.3265	0.9695
Wiki	@12	**6.7652**	**0.4024**	7.5032	0.4670	8.5148	0.5530	7.4778	0.4130	7.8025	0.4995	7.2728	0.4122	7.2218	0.4055	7.2511	0.4029	7.5664	0.4284	7.2747	0.4145	7.2832	0.4053	7.1790	0.4347
	@36	6.4498	0.4728	**6.3075**	0.4889	7.8133	0.6813	6.3831	**0.4426**	6.8597	0.5787	6.3605	0.4456	6.3938	0.4595	6.8257	0.4430	6.5543	0.4892	6.4683	0.4707	6.3609	0.4437	6.3871	0.4468
Counts		**3**	**2**	1	0	0	0	0	1	0	1	1	0	1	1	0	**3**	0	0	1	1	0	0	1	0

the other. Crucially, the system satisfies the global dissipativity condition $\nabla \cdot f = -(\sigma + 1 + \beta) < 0$, guaranteeing the phase space volume contraction essential for bounded neural representations.

Evaluation Metrics and Loss Function. We evaluate performance via Mean Squared Error (MSE) and Mean Absolute Error (MAE). And L1 loss function from CARD (wang et al., 2024) is adopted as TRACE's loss function.

5.2 Main Results

Average results over four prediction lengths are reported in Table 1, where *Counts* denotes the number of top-2 results. We evaluate long-term forecasting with $\tau \in \{12, 24, 48, 96\}$ for PEMS and $\tau \in \{96, 192, 336, 720\}$ for remaining datasets, with a fixed lookback horizon $T = 96$. TRACE achieves highly competitive performance across the majority of benchmarks. Notably, on the ETT and PEMS datasets, TRACE can outperform both heavily parameterized Transformer and some lightweight models. This success validates that *chaotic evolution* serves as a powerful mechanism for representation learning. On Weather, while TRACE is slightly surpassed by multi-scale models such as TimeMixer++, it still maintains a highly competitive top-tier ranking. Weather is driven by high-dimensional meteorological information, where multi-scale models explicitly do well. In contrast, TRACE relies on a parameter-free, low-dimensional deterministic ODE system.

For short-term forecasting, two settings are applied: S1($T = 12$, $\tau \in \{3, 6, 9, 12\}$), and S2($T = 36$, $\tau \in \{24, 36, 48, 60\}$). Average results are presented in Table 1, where *Counts* is the number of top-2 results. TRACE exhibits good precision in short-term scenarios, particularly on ILI. In S1, TRACE achieves the lower MSE. TRACE uses the finite-time dynamical prior to transform short historical contexts into nonlinear latent representations. On Wiki, TRACE remains competitive in S1, while its S2 performance aligns with PatchTST. This reflects the event-driven nature of social media traffic compared to continuous physical processes. Nonetheless, TRACE's robustness across diverse domains underscores its applicability across datasets with different sampling frequencies, dimensions, and forecasting horizons.

While TRACE does not always provide the absolute best result for every single task, it consistently delivers highly competitive performance against much larger architectures. This reflects a natural trade-off, under lightweight and interpretable constraints, achieving superior performance across every scenario against heavily parameterized models is challenging. However, as an initial attempt, TRACE already secures good results on a subset of datasets and demonstrates robust overall competitiveness, proving that our physics-inspired approach is both feasible and practically effective. And detailed evaluations are provided in Appendix A.4.

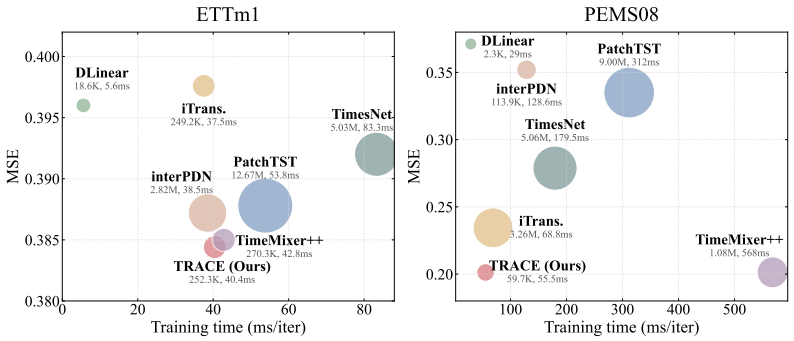


Figure 3: Model efficiency comparison. MSE and training time are averaged over four prediction horizons. Bubble size reflects the parameter count of each model.

5.3 More Analysis

Efficiency Advantages. Analyzed in Appendix A.2, TRACE offloads the core temporal processing to a parameter-free physical flow, resulting in substantially fewer trainable parameters and lower computational overhead. A detailed comparison of parameter counts and training efficiency is provided in Fig. 3. TRACE occupies the highly coveted bottom-left quadrant across both datasets, simultaneously delivering better predictive accuracy, highly competitive training speed, and a drastically minimized parameter footprint. This Pareto-efficiency empirically validates that offloading temporal dependency modeling to a fixed, non-trainable chaotic core can circumvent the heavy parameter scaling and severe computational bottlenecks inherent in conventional deep architectures.

Ablation Studies. The chaotic core and the coupling topology are the two defining components of TRACE. To empirically validate their contributions, we conduct ablation studies by systematically removing each component. As shown in Table 3, both are essential for achieving optimal forecasting accuracy. Notably, the variant that retains causal coupling without chaotic evolution is not evaluated, as the coupling mechanism is defined atop the evolving states and cannot operate independently.

Table 3: Ablation study on Dynamic Evolution and Causal Coupling

Chaos Evolution	Causal Coupling	ETTh1		ETTm1		ILI(S1)		ILI(S2)	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
×	×	0.4845	0.4594	0.4087	0.4062	1.6391	0.7259	2.0178	0.8705
✓	×	0.4556	0.4487	0.4022	0.4029	1.4380	0.6811	1.8929	0.8596
✓	✓	0.4344	0.4251	0.3844	0.3819	1.3363	0.6611	1.6697	0.7953

Impact of Evolution Time and Transient Chaos. In chaotic systems, FTMLE typically rises from a high value and converges to a positive asymptotic limit, reflecting varying degrees of chaoticity at different evolution stages. To investigate how this transient chaoticity affects TRACE's performance, we vary t_{evo} while fixing the integration steps $M = 20$, coupling strength $\alpha = 0.5$, and decay rate $\lambda = 1.0$. Results are reported in Fig.4.

As shown, MSE exhibits a V-shaped profile, with the best performance occurring in an intermediate finite-time expansion regime identified by the measured FTMLE ($t_{\text{evo}} \in [0.01, 0.5]$). At shorter evolution times, the elevated FTMLE drives rapid phase mixing that separates close-lying temporal patterns. As t_{evo} exceeds this optimal region, excessive transient expansion washes local features into the global attractor, steadily degrading accuracy. Longer prediction horizons (e.g., $\tau = 336$) slightly shift the optimal t_{evo} rightward, indicating that capturing extended dependencies benefits from deeper unrolling.Further analyses detailing the impact of other physical hyperparameters on the transient dynamical regimes are provided in Appendix A.5.

Geometric Dynamics Mechanisms Validation. Proposition 4.1 predicts that TRACE's chaotic flow simultaneously induces global volume contraction and local inter-node expansion during evolution. To empirically verify this dual mechanism, we track the evolution of the global Jacobian determinant and the pairwise distances between node states under the Lorenz oscillator. The re-

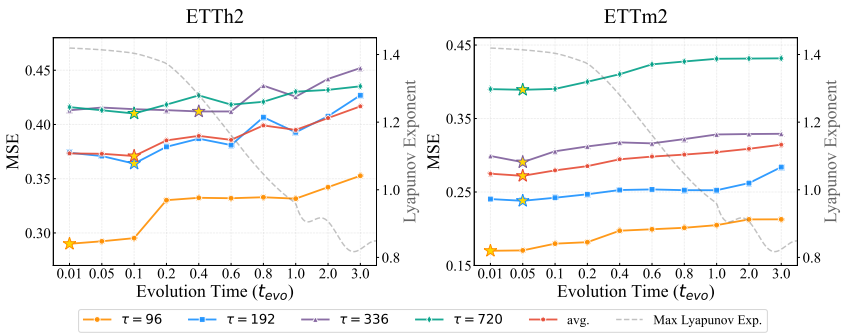


Figure 4: MSE (left axis) and the Finite-Time Maximum Lyapunov Exponent (right axis) versus t_{evo} for fixed $M = 20$, $\alpha = 0.5$, $\lambda = 1.0$.

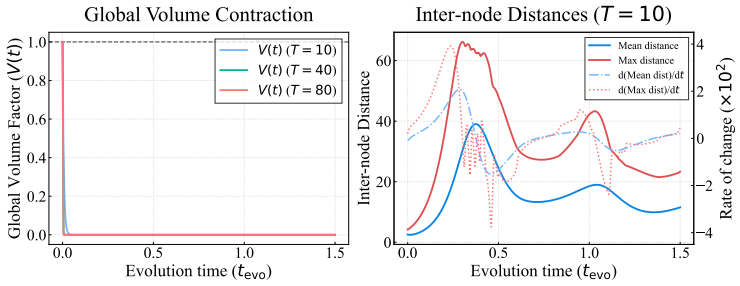


Figure 5: Empirical validation of the dual geometric dynamics mechanisms predicted by Proposition 4.1 under varying evolution times t_{evo}. Left: The global phase space volume contraction factor $V(t)$, quantified by the joint Jacobian determinant, across different system scales (node counts $T \in \{10, 40, 80\}$). Right: Trajectories of pairwise inter-node mean and maximum Euclidean distances alongside their corresponding rates of change $(\mathrm{d}(\cdot)/\mathrm{d}t)$ within a $T = 10$ node system.

sults are visualized in Fig. 5. The global phase space volume collapses rapidly while inter-node distances simultaneously expand during early evolution, as proposed in Proposition 4.1. This confirms the stretch-and-fold mechanism, where global contraction may filter noise, while local chaotic stretching enriches temporal representations.

6 Conclusion

We presented TRACE, a time series forecasting model with a fixed, non-trainable transient chaotic core for temporal processing. The finite-time evolution of a fixed ODE-based dynamical core performs the main nonlinear transformation of temporal representations, while a causal coupling topology shapes interactions among time-indexed states. This design not only avoids adding a deep stack of trainable nonlinear transformations and substantially reduces the trainable parameter count relative to the larger baselines, but also offers inherent interpretability. The entire temporal computation is physically grounded in the continuous phase-space flow of coupled chaotic oscillators, making the latent-state trajectory explicitly traceable in phase space from the learned initial embedding to the final readout. And we showed that this dynamical core naturally exhibits a geometric stretch-and-fold mechanism, simultaneously compressing global redundancy while expanding local features. Empirically, TRACE achieves competitive accuracy against state-of-the-art forecasters with substantially fewer parameters. Finally, a fully formalized understanding of how chaotic evolution reorganizes temporal representations remains an open question. Future work may explore automated dynamical regime selection and deeper mechanistic analysis of the chaotic properties for TSF.

AI USE STATEMENT

In this work, we used generative AI tools to assist in the writing of proofs and assist with translation. We have not used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results, and generating synthetic data sets is not applicable to this work as we exclusively utilized real-world datasets. Additionally, we used generative AI tools to summarize or analyse existing literature and for sourcing/searching for information. We have reviewed all AI-assisted work. Specifically, all theoretical propositions and mathematical derivations were originally formulated by the authors, with AI utilized solely to verify and polish the proofs. Furthermore, all AI-assisted translations and literature summaries were manually cross-checked for accuracy and contextual relevance. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ACKNOWLEDGMENTS

REFERENCES

K. Aihara, T. Takabe, and M. Toyoda. Chaotic neural networks. *Physics Letters A*, 144(6):333–340, 1990. ISSN 0375-9601. doi: https://doi.org/10.1016/0375-9601(90)90136-C. URL https://www.sciencedirect.com/science/article/pii/037596019090136C.

Harikrishnan Nellippallil Balakrishnan, Aditi Kathpalia, Snehanshu Saha, and Nithin Nagaraj. Chaosnet: A chaos based artificial neural network architecture for classification. *Chaos*, 29 11: 113125, 2019. URL https://api.semanticscholar.org/CorpusID:203837650.

Safieddine Bouali. Feedback loop in extended van der pol's equation applied to an economic model of cycles. *International Journal of Bifurcation and Chaos*, 09(04):745–756, 1999. doi: 10.1142/S0218127499000535. URL https://doi.org/10.1142/S0218127499000535.

Guoliang Cai and Juanjuan Huang. A new finance chaotic attractor. *International Journal of Nonlinear Science*, 3(3):213–220, 2007.

F. M. Callier. Dissipative systems analysis and control: Theory and applications (2nd edition)-[book review; b. brogliato, r. lozano, b. maschke]. *IEEE Transactions on Automatic Control*, 52 (7):1357–1358, 2007. doi: 10.1109/TAC.2007.900820.

Liping Chen, Hao Yin, Tingwen Huang, Liguo Yuan, Song Zheng, and Lisheng Yin. Chaos in fractional-order discrete neural networks with application to image encryption. *Neural Networks*, 125:174–184, 2020. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2020.02.008. URL https://www.sciencedirect.com/science/article/pii/S0893608020300538.

Shyan-Shiou Chen. Chaotic simulated annealing by a neural network with a variable delay: Design and application. *IEEE Transactions on Neural Networks*, 22(10):1557–1565, 2011. doi: 10.1109/TNN.2011.2163080.

Yuming Chen and Qigui Yang. Dynamics of a hyperchaotic lorenz-type system. *Nonlinear Dynamics*, 77(3):569–581, Aug 2014. ISSN 1573-269X. doi: 10.1007/s11071-014-1318-0. URL https://doi.org/10.1007/s11071-014-1318-0.

Xuewei Cheng, Ke Huang, and Shujie Ma. Generalization and risk bounds for recurrent neural networks. *Neurocomputing*, 616:128825, 2025.

Leonhard Euler. *Institutionum calculi integralis*, volume 1. Petropoli: impensis Academiae Imperialis Scientiarum, 3 edition, 1824.

Claudio Gallicchio. Chasing the echo state property. *arXiv preprint arXiv:1811.10892*, 2018.

Caglar Gulcehre, Kyunghyun Cho, Razvan Pascanu, and Yoshua Bengio. Learned-norm pooling for deep feedforward and recurrent neural networks. In *Machine Learning and Knowledge Discovery in Databases*, pp. 530–546, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-662-44847-2. doi: 10.1007/978-3-662-44848-9_34. URL https://doi.org/10.1007/978-3-662-44848-9_34.

N B Harikrishnan and Nithin Nagaraj. A novel chaos theory inspired neuronal architecture. In *2019 Global Conference for Advancement in Technology (GCAT)*, pp. 1–6, 2019. doi: 10.1109/GCAT47503.2019.8978360.

N.B. Harikrishnan and Nithin Nagaraj. Neurochaos inspired hybrid machine learning architecture for classification. In *2020 International Conference on Signal Processing and Communications (SPCOM)*, pp. 1–5, 2020. doi: 10.1109/SPCOM50965.2020.9179632.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554. URL https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554.

Herbert Jaeger. The"echo state"approach to analysing and training recurrent neural networks. 2001. URL https://api.semanticscholar.org/CorpusID:15467150.

Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004. doi: 10.1126/science.1091277. URL https://www.science.org/doi/abs/10.1126/science.1091277.

Bowen Jia, Huyu Wu, and Kaiyu Guo. Chaos theory meets deep learning: A new approach to time series forecasting. *Expert Systems with Applications*, 255:124533, 2024. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2024.124533. URL https://www.sciencedirect.com/science/article/pii/S0957417424014003.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jangho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022. URL https://api.semanticscholar.org/CorpusID:251647808.

Nakajima Kohei. Reservoir computing: Theory, physical implementations, and applications. *Reservoir Computing*, 2018.

Linghao Kong and Xiaopeng Hong. Time series forecasting via direct per-step probability distribution modeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(27): 22653–22661, Mar. 2026. doi: 10.1609/aaai.v40i27.39426. URL https://ojs.aaai.org/index.php/AAAI/article/view/39426.

W. Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Zeit. Math. Phys.*, 46:435–53, 1901.

M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. A Series of Modern Surveys in Mathematics Series. Springer, 1991. ISBN 9783540520139. URL https://books.google.co.jp/books?id=cyKYDfvxRjsC.

Xiaoming Li and Shijun Liao. The influence of numerical noises on statistics computation of chaotic dynamic systems. *arXiv preprint arXiv:1609.09354*, 2016.

Yan Li, Xinjiang Lu, Haoyi Xiong, Jian Tang, Jiantao Su, Bo Jin, and Dejing Dou. Towards long-term time-series forecasting: Feature, pattern, and distribution. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pp. 1611–1624, 2023. doi: 10.1109/ICDE55515.2023.00127.

Hairong Lin, Chunhua Wang, Li Cui, Yichuang Sun, Xin Zhang, and Wei Yao. Hyperchaotic memristive ring neural network and application in medical image encryption. *Nonlinear Dynamics*, 110(1):841–855, 2022. doi: 10.1007/s11071-022-07630-0. URL https://doi.org/10.1007/s11071-022-07630-0.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and C. L. Philip Chen. Sparsetsf: Lightweight and robust time series forecasting via sparse modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1):170–183, 2026. doi: 10.1109/TPAMI.2025.3602445.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 5816–5828. Curran Associates, Inc., 2022. doi: 10.52202/068431-0421. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/266983d0949aed78a16fa4782237dea7-Paper-Conference.pdf.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Edward N. Lorenz. *Deterministic Nonperiodic Flow*, pp. 25–36. Springer New York, New York, NY, 2004. ISBN 978-0-387-21830-4. doi: 10.1007/978-0-387-21830-4_2. URL https://doi.org/10.1007/978-0-387-21830-4_2.

Edward N. Lorenz. *Predictability – a problem partly solved*, pp. 40–58. Cambridge University Press, 2006.

DongHao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 31728–31770, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/86b1437c1e4c3b3c4debff98234a67e7-Paper-Conference.pdf.

Aitian Ma, Dongsheng Luo, and Mo Sha. Mixlinear: Extreme low resource multivariate time series forecasting with $0.1k$ parameters. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=QUj0KuCumD.

Filip Matzner and František Mráz. Locally connected echo state networks for time series forecasting. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 87101–87122, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/d8bd445c2abe1343cce0e14b361b2fb3-Paper-Conference.pdf.

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. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Yiming Niu, Jinliang Deng, and Yongxin Tong. Phaseformer: From patches to phases for efficient and effective time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Lk9SqMQzhX.

Louis M. Pecora and Thomas L. Carroll. Master stability functions for synchronized coupled systems. *Phys. Rev. Lett.*, 80:2109–2112, Mar 1998. doi: 10.1103/PhysRevLett.80.2109. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.2109.

Piotr Przystałka and Wojciech Moczulski. Methodology of neural modelling in fault detection with the use of chaos engineering. *Engineering Applications of Artificial Intelligence*, 41:25–40, 2015. ISSN 0952-1976. doi: https://doi.org/10.1016/j.engappai.2015.01.016. URL https://www.sciencedirect.com/science/article/pii/S0952197615000275.

Sibo Qi, Xiaozhu Jing, Liuyang Zhao, Wuman Luo, and Dian Zhang. FACT: Frequency-aware channel-guided multivariate time series forecasting, 2026. URL https://openreview.net/forum?id=73HAQ9K9Ar.

M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. I. Abarbanel, and G. Laurent. Dynamical encoding by networks of competing neuron groups: Winnerless competition. *Phys. Rev. Lett.*, 87:068102, Jul 2001. doi: 10.1103/PhysRevLett.87.068102. URL https://link.aps.org/doi/10.1103/PhysRevLett.87.068102.

Daking Rai, Yilun Zhou, Shi Feng, Abulhair Saparov, and Ziyu Yao. A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models. *arXiv e-prints*, art. arXiv:2407.02646, July 2024. doi: 10.48550/arXiv.2407.02646.

O.E. Rössler. An equation for continuous chaos. *Physics Letters A*, 57(5):397–398, 1976. ISSN 0375-9601. doi: https://doi.org/10.1016/0375-9601(76)90101-8. URL https://www.sciencedirect.com/science/article/pii/0375960176901018.

L Storm, Hampus Linander, J Bec, Kristian Gustavsson, and Bernhard Mehlig. Finite-time lyapunov exponents of deep neural networks. *Physical Review Letters*, 132(5):057301, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Shiyu Wang, Jiawei LI, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Ju Shengtong, Zhixuan Chu, and Ming Jin. Timemixer++: A general time series pattern machine for universal predictive analysis. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=1CLzLXSFNn.

xue wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. Card: Channel aligned robust blend transformer for time series forecasting. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 11245–11283, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/2f4d6f8e0f4f543db12260696b2a3551-Paper-Conference.pdf.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Chen Wang, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Alan Wolf, Jack B. Swift, Harry L. Swinney, and John A. Vastano. Determining lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):285–317, 1985. ISSN 0167-2789. doi: https://doi.org/10.1016/0167-2789(85)90011-9. URL https://www.sciencedirect.com/science/article/pii/0167278985900119.

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. URL https://openreview.net/forum?id=J4gRj6d5Qm.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=bWcnvZ3qMb.

Wenzhen Yue, Yong Liu, Hao Wang, Haoxuan Li, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. OLinear: A linear model for time series forecasting in orthogonally transformed domain. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=DAyKP1tvwI.

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*, volume 37, 2023.

Jingyuan Zhao, Fulin Chu, Lili Xie, Yunhong Che, Yuyan Wu, and Andrew F. Burke. A survey of transformer networks for time series forecasting. *Computer Science Review*, 60:100883, 2026. ISSN 1574-0137. doi: https://doi.org/10.1016/j.cosrev.2025.100883. URL https://www.sciencedirect.com/science/article/pii/S1574013725001595.

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 *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning (ICML 2022)*, 2022.

A Appendix

A.1 Optimization and Learning

Optimization of TRACE involves updating the parameters of the encoder ϕ and decoder ψ, both of which contain a few MLP layers. While the chaotic core contains no learnable weights, gradients must be reliably backpropagated through the discrete temporal steps of the numerical ODE solver.

A.1.1 Forward Numerical Integration

The continuous latent evolution is governed by the parameterized vector field $\frac{d\mathbf{H}}{dt} = \mathbf{F}(\mathbf{H}, C)$, aggregating the intrinsic chaotic ODEs of all nodes and the coupling from C. Concretely, by stacking the per-node dynamics defined in Eq. equation 4, the global evolution can be written as

$$\frac{d\mathbf{H}}{dt} = \mathbf{F}(\mathbf{H}, C) = \mathbf{f}(\mathbf{H}) - \alpha(\mathbf{L} \otimes \mathbf{I}_d)\mathbf{H}, \tag{12}$$

where $\mathbf{f}(\mathbf{H}) = [f(\mathbf{s}_1); \ldots; f(\mathbf{s}_T)]$ stacks the intrinsic chaotic ODE f applied independently to each node, $\mathbf{L} = \mathrm{diag}(C\mathbf{1}) - C$ is the graph Laplacian constructed from C, $\otimes$ denotes the Kronecker product, and $\mathbf{I}_d \in \mathbb{R}^{d\times d}$ is an identity matrix. The Laplacian coupling term captures the net diffusive influence exerted on each node by its neighbors, as prescribed by the coupling topology C.

To simulate this continuous flow over the interval $[0, t_{\mathrm{evo}}]$, we partition the domain into M discrete steps with a step size $\Delta t = t_{\mathrm{evo}}/M$.

Chaotic dynamics are notoriously sensitive to numerical truncation errors (Li & Liao, 2016). Thus, TRACE employs a fourth-order Runge-Kutta (RK4) integration scheme to suppress numerical truncation errors and ensure structural stability. The forward state transition from step m to $m+1$ is formulated as

$$\mathbf{H}_{m+1} = \mathbf{H}_m + \frac{\Delta t}{6}(\mathbf{K}_1 + 2\mathbf{K}_2 + 2\mathbf{K}_3 + \mathbf{K}_4), \tag{13}$$

where the intermediate vector field evaluations are defined recursively

$$\mathbf{K}_1 = \mathbf{F}(\mathbf{H}_m, C), \tag{14}$$

$$\mathbf{K}_2 = \mathbf{F}\left(\mathbf{H}_m + \frac{\Delta t}{2}\mathbf{K}_1, C\right), \tag{15}$$

$$\mathbf{K}_3 = \mathbf{F}\left(\mathbf{H}_m + \frac{\Delta t}{2}\mathbf{K}_2, C\right), \tag{16}$$

$$\mathbf{K}_4 = \mathbf{F}\left(\mathbf{H}_m + \Delta t\mathbf{K}_3, C\right). \tag{17}$$

After M iterations, the system yields the terminal state $\mathbf{H}_{t_{\mathrm{evo}}}$, which serves as the predictive representation for readout.

A.1.2 BACKWARD GRADIENT PROPAGATION

Given a task-specific loss function $\mathcal{L}(\hat{\mathbf{Y}}, \mathbf{Y})$, optimizing decoder parameter ψ is trivial

$$\nabla_{\psi}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{Y}}} \frac{\partial \hat{\mathbf{Y}}}{\partial \psi}. \tag{18}$$

For the encoder parameter ϕ, TRACE executes discrete backpropagation directly through the unrolled computational graph of the solver. By the chain rule, the gradient is expressed as

$$\nabla_{\phi}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial \mathbf{H}_{t_{\text{evo}}}} \cdot \mathcal{J}_{t_{\text{evo}}} \cdot \frac{\partial \mathbf{H}_0}{\partial \phi}, \tag{19}$$

where $\mathcal{J}_{t_{\text{evo}}} = \frac{\partial \mathbf{H}_{t_{\text{evo}}}}{\partial \mathbf{H}_0} = \prod_{m=0}^{M-1} \frac{\partial \mathbf{H}_{m+1}}{\partial \mathbf{H}_m}$ represents the global flow Jacobian matrix.

In RK4, the transition Jacobian at each step $\frac{\partial \mathbf{H}_{m+1}}{\partial \mathbf{H}_m}$ is a high-order approximation of the continuous evolution operator. It incorporates the Jacobians of the vector field at all four intermediate stages ($\mathcal{J}_m^i = \frac{\partial \mathbf{K}_i}{\partial \mathbf{H}_m}$). Thus, RK4 accurately tracks the geometric variations of the phase space, effectively modulating gradients along stable and unstable manifolds. This mathematical precision scales robustly under larger step sizes Δt, minimizing the recurrence depth M. Some relevant experiments are provided in Section A.7.4.

A.2 PARAMETER EFFICIENCY AND EMPIRICAL RISK BOUNDS

To further illustrate the advantages of TRACE, we analyze its parameter efficiency and generalization capability through the lens of statistical learning theory, supplementing the geometric analysis provided in the main text.

A.2.1 PARAMETER AND TIME COMPLEXITY

We compare TRACE with standard MLP, RNN and Transformer architectures. Let d_{hid} denote the hidden dimension of the baseline models, and let d_{enc}, d_{dec}, and d denote the embedding dimension, decoder hidden dimension, and chaotic oscillator dimension of TRACE, respectively. The theoretical complexities are summarized in Table 4.

Table 4: Parameter and Time Complexity Comparison.

Model	Parameter Complexity	Time Complexity
MLP	$\mathcal{O}(NTd_{hid} + d_{hid}\tau N)$	$\mathcal{O}(NTd_{hid} + d_{hid}\tau N)$
RNN	$\mathcal{O}(Nd_{hid} + d_{hid}^2 + N\tau d_{hid})$	$\mathcal{O}(T(Nd_{hid} + d_{hid}^2))$
Transformer	$\mathcal{O}(Nd_{hid} + d_{hid}^2 + d_{hid}\tau N)$	$\mathcal{O}(T^2 d_{hid} + Td_{hid}^2)$
TRACE (Ours)	$\mathcal{O}(Nd_{enc} + \tau N d_{dec})$	$\mathcal{O}(N(Td_{enc} + \tau d_{dec}) + MT^2 d)$

Standard architectures typically allocate $\mathcal{O}(d_{hid}^2)$ parameters. In contrast, TRACE treats historical time steps as interacting nodes in a $T \times T$ graph, using a *parameter-free* physical evolution for information processing. Trainable weights are limited to shallow embedding and readout, avoiding the $\mathcal{O}(d_{hid}^2)$ explosion. In standard RNN and Transformer, the dominant parameter term is the quadratic $\mathcal{O}(d_{hid}^2)$ arising from the recurrent or self-attention weight matrices.

In practice, d_{hid} must be set large enough (typically several hundred) to ensure sufficient representational capacity. In contrast, TRACE confines its trainable parameters to a shallow encoder and decoder. Even the dimensions are the same as d_{hid}, TRACE can avoid the $\mathcal{O}(d_{hid}^2)$ term. For time complexity, Transformers suffer an $\mathcal{O}(T^2 d_{hid})$ bottleneck. TRACE also models global dependencies via a $T \times T$ physical coupling matrix, but the $MT^2 d \ll T^2 d_{hid}$. Since d is intrinsically small (e.g., 3) and M is relative less (e.g., 20) while the hidden dimensions typically range into hundreds or more. Overall TRACE achieves significantly lower FLOPs.

A.2.2 Generalization Bounds and Sample Complexity

Beyond computational efficiency, TRACE can also yield tighter theoretical guarantees on generalization. Let m denote the number of training samples, $\hat{R}(\cdot)$ the empirical risk, and $R(\cdot)$ the expected risk. For a fully trainable RNN, the expected risk is bounded by the empirical risk plus a Rademacher complexity term that scales with both weight norms and input length (Cheng et al., 2025):

$$R(\hat{f}_{\mathrm{RNN}}) \leq \hat{R}(\hat{f}_{\mathrm{RNN}}) + \mathcal{O}\left(\frac{\|\mathbf{W}_\omega\|_F \sqrt{T}}{\sqrt{m}} + \sqrt{\frac{\log(1/\delta)}{m}}\right), \tag{20}$$

where $\|\mathbf{W}_\omega\|_F$ denotes the Frobenius norm of all trainable weights, and δ is the confidence level. As the network deepens or the input lengthens, the hypothesis space expands rapidly, requiring substantially more samples m to avoid overfitting.

For TRACE, the hypothesis class can be formally written as

$$\mathcal{F}_{\mathrm{TRACE}} = \{\psi \circ \Phi_{t_{\mathrm{evo}}} \circ \phi \mid \phi \in \mathcal{E}, \psi \in \mathcal{D}\}, \tag{21}$$

where $\mathcal{E}$ and $\mathcal{D}$ represent the function classes of the encoder and decoder, and $\Phi_{t_{\mathrm{evo}}}$ is the fixed deterministic flow. Let L_ϕ and L_ψ denote the Lipschitz constants of the encoder and decoder, and let $K_{\mathrm{chaos}}(t_{\mathrm{evo}})$ denote the Lipschitz constant of the fixed flow $\Phi_{t_{\mathrm{evo}}}$ on the bounded latent domain visited during evolution. Since $f = \psi \circ \Phi_{t_{\mathrm{evo}}} \circ \phi$, the Lipschitz constant of the induced predictor satisfies

$$L_f \leq L_\psi K_{\mathrm{chaos}}(t_{\mathrm{evo}}) L_\phi. \tag{22}$$

By the Ledoux–Talagrand contraction lemma (Ledoux & Talagrand, 1991) / covering-number argument for composite function classes, the Rademacher complexity satisfies

$$R_m(\mathcal{F}_{\mathrm{TRACE}}) \lesssim L_\psi K_{\mathrm{chaos}}(t_{\mathrm{evo}}) L_\phi \, R_m(\mathcal{H}). \tag{23}$$

For MLP encoder and decoder with 1-Lipschitz activations,

$$L_\phi \leq \prod_{l=1}^{L_{\mathrm{enc}}} \|W_l^\phi\|_2, \quad L_\psi \leq \prod_{l=1}^{L_{\mathrm{dec}}} \|W_l^\psi\|_2. \tag{24}$$

Therefore, a correct generalization bound is

$$R(\hat{f}_{\mathrm{TRACE}}) \leq \hat{R}(\hat{f}_{\mathrm{TRACE}}) + \mathcal{O}\left(\frac{K_{\mathrm{chaos}}(t_{\mathrm{evo}}) \prod_{l\in\mathrm{enc}} \|W_l^\phi\|_F \prod_{l\in\mathrm{dec}} \|W_l^\psi\|_F}{\sqrt{m}} + \sqrt{\frac{\log(1/\delta)}{m}}\right). \tag{25}$$

Importantly, $K_{\mathrm{chaos}}(t_{\mathrm{evo}})$ is not a universal constant. It is controlled by the finite-time growth of the flow Jacobian:

$$K_{\mathrm{chaos}}(t_{\mathrm{evo}}) \leq \exp\left(\int_0^{t_{\mathrm{evo}}} \|DF(H_s)\|_2 \, ds\right) \lesssim e^{\lambda_{\max} t_{\mathrm{evo}}}. \tag{26}$$

This exponential factor explains the V-shaped empirical trend in Fig. 4[TODO]. Short evolution times provide insufficient chaotic mixing, while overly long evolution times amplify sensitivity and loosen the generalization bound.

Comparing with the RNN bound in Eq. 20 should be done at the level of path complexity, not additive Frobenius norms. The RNN bound scales with the accumulated recurrent path norm and horizon T, whereas TRACE scales with

$$K_{\text{chaos}}(t_{\text{evo}}) \prod_{\text{enc}} \|W\|_2 \prod_{\text{dec}} \|W\|_2. \tag{27}$$

Since TRACE uses shallow encoder/decoder and operates in a controlled transient-chaos window, this product can be smaller than the deep recurrent path complexity. This advantage is conditional on keeping t_{evo} in the transient regime where $K_{\text{chaos}}(t_{\text{evo}})$ is moderate.

A.3 Dataset description

In this work, we use the following real-world datasets for performance evaluation. The dataset details are summarized in Table 5

- ETT (Zhou et al., 2021) datasets record seven channels related to electricity transformers from July 2016 to July 2018. It contains four datasets: ETTh1 and ETTh2, with hourly recordings, and ETTm1 and ETTm2, with 15-minute recordings.
- Weather (Wu et al., 2021) contains 21 meteorological variations (e.g., air temperature, humidity) recorded every 10 minutes in 2020.
- PEMS (Liu et al., 2022) provides public traffic sensor data from California, collected every 5 minutes. We use its subset (PEMS08) in this study.
- ILI[1] contains weekly records of influenza-like illness patient counts provided by the U.S. CDC from 2002 to 2021.
- Wiki[2] records daily page view counts for Wikipedia articles over two years (2018–2019). The first 99 channels are used in our work.

Table 5: Summary of datasets.

Dataset	Dim.	Frequency	Total len.	Split
ETTh1, ETTh2	7	Hourly	17,420	6:2:2
ETTm1, ETTm2	7	15 mins	69,680	6:2:2
Weather	21	10 mins	52,696	7:1:2
PEMS08	170	5 mins	17,856	6:2:2
ILI	7	Weekly	966	7:1:2
Wiki	99	Daily	730	7:1:2

A.4 Full results

The comprehensive experimental results detailed further validate the strong generalization capabilities of TRACE. And the *Counts* is the number of top-2 results.

In the long-term forecasting tasks (Table. 6), TRACE achieves the global results 18 times in the core metric of MSE and 17 times in MAE across various prediction lengths. It outperforms many strong baseline models, including Transformer-based (e.g., PatchTST), CNN-based (e.g., ModernTCN, TimeMixer++). Meanwhile, in the short-term forecasting tasks (Table. 7, TRACE maintains a impressive performance, securing 15 better MSE results and 11 better MAE results.

A.5 Transient Dynamical Regimes

Building upon the analysis of evolution time and transient chaos presented in the main text, this section further investigates how the structural hyperparameters of the physical coupling graph—specifically, the distance decay rate λ and coupling strength α—govern the overall chaoticity and predictive stability of the intermediate dynamical system.

[1] https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

[2] https://www.kaggle.com/datasets/sandeshbhat/wikipedia-web-traffic-201819

Table 6: Full experimental results in Long-Term Forecasting

Categories		Linear-Based										Transformer-Based						CNN-Based						Hybrid	
Models		**TRACE** (Ours)		DLinear (2023)		FITS (2024)		SparseTSF (2026)		MixLinear (2026)		PatchTST (2023)		iTrans. (2024)		TimeMixer++ (2025)		TimesNet (2023)		ModernTCN (2024)		FACT (2026)		interPDN (2026)	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	0.3211	**0.3403**	0.3315	0.3503	0.3530	0.3750	0.3597	0.3675	0.3717	0.3692	**0.3105**	0.3407	0.3272	0.3527	0.3185	0.3420	0.3252	0.3611	0.3258	0.3603	0.3161	0.3488	0.3192	0.3478
	192	**0.3638**	0.3716	0.3763	0.3725	0.4860	0.4450	0.3990	0.3870	0.5195	0.4541	0.3714	0.3779	0.3759	0.3770	0.3639	**0.3672**	0.3659	0.3834	0.3686	0.3876	0.3651	0.3729	0.3696	0.3701
	336	**0.3951**	**0.3900**	0.4070	0.3950	0.5310	0.4750	0.4249	0.4057	0.5039	0.4571	0.4050	0.4021	0.4119	0.4029	0.3973	0.3906	0.4056	0.4058	0.4018	0.4029	0.3983	0.3989	0.3974	0.3902
	720	**0.4574**	**0.4254**	0.4691	0.4339	0.6000	0.5130	0.4802	0.4370	0.5448	0.4727	0.4645	0.4358	0.4754	0.4384	0.4602	0.4287	0.4711	0.4429	0.4663	0.4402	0.5086	0.4499	0.4624	0.4292
	Avg.	**0.3844**	**0.3819**	0.3960	0.3879	0.4925	0.4520	0.4160	0.3993	0.4850	0.4383	0.3878	0.3891	0.3976	0.3928	0.3850	0.3821	0.3920	0.3983	0.3906	0.3978	0.3970	0.3926	0.3872	0.3843
ETTm2	96	0.1700	0.2485	0.1817	0.2581	0.1820	0.2660	0.1952	0.2680	0.1907	0.2658	0.1780	0.2535	0.1757	0.2540	0.1744	0.2521	0.1788	0.2577	**0.1676**	**0.2479**	0.1750	0.2487	0.1734	0.2515
	192	**0.2380**	**0.2907**	0.2450	0.3031	0.2530	0.3120	0.2561	0.3059	0.2542	0.3079	0.2457	0.2988	0.2430	0.2983	0.2382	0.2938	0.2545	0.3035	0.2387	0.2999	0.2399	0.2917	0.2384	0.2946
	336	0.2906	**0.3277**	0.3051	0.3451	0.3130	0.3490	0.3159	0.3443	0.3155	0.3462	0.3086	0.3395	0.3081	0.3409	0.2995	0.3345	0.3109	0.3406	0.2899	0.3281	**0.2886**	0.3280	0.2989	0.3339
	720	0.3891	0.3891	0.4095	0.4165	0.4160	0.4060	0.4091	0.3967	0.4156	0.4018	0.4076	0.3982	0.4081	0.3987	0.3944	0.3899	0.4288	0.4078	**0.3821**	**0.3844**	0.3896	0.3867	0.3983	0.3918
	Avg.	0.2719	0.3140	0.2853	0.3307	0.2910	0.3333	0.2941	0.3287	0.2940	0.3304	0.2850	0.3225	0.2837	0.3230	0.2766	0.3176	0.2932	0.3274	**0.2696**	0.3151	0.2733	**0.3138**	0.2772	0.3180
ETTh1	96	**0.3802**	0.3890	0.3883	0.4009	0.3850	0.3940	0.3875	**0.3883**	0.3903	0.3890	0.4008	0.4060	0.3887	0.4017	0.3940	0.3949	0.4251	0.4230	0.3803	0.3899	0.3844	0.3946	0.3827	0.3952
	192	0.4297	0.4202	0.4381	0.4309	0.4340	0.4220	0.4415	**0.4176**	0.4378	0.4180	**0.4275**	0.4204	0.4430	0.4357	0.4321	0.4200	0.4679	0.4490	0.4381	0.4198	0.4427	0.4257	0.4318	0.4206
	336	**0.4661**	0.4382	0.4835	0.4601	0.4760	0.4440	0.4812	0.4383	0.4786	0.4402	0.4814	0.4490	0.4824	0.4573	0.4912	0.4431	0.4827	0.4550	0.4758	**0.4381**	0.4827	0.4441	0.4763	0.4418
	720	**0.4618**	0.4531	0.4989	0.4988	0.4650	0.4620	0.4801	0.4528	0.4636	**0.4525**	0.4840	0.4720	0.4888	0.4830	0.4817	0.4662	0.5071	0.4862	0.4681	0.4568	0.5009	0.4720	0.4779	0.4639
	Avg.	**0.4344**	0.4251	0.4522	0.4477	0.4400	0.4305	0.4476	**0.4242**	0.4426	0.4249	0.4484	0.4368	0.4507	0.4444	0.4498	0.4310	0.4707	0.4533	0.4406	0.4261	0.4527	0.4341	0.4422	0.4304
ETTh2	96	0.2900	0.3408	0.2986	0.3484	0.2920	0.3400	0.2995	0.3388	0.3094	0.3451	0.3017	0.3417	0.2948	0.3419	0.2858	0.3338	0.3051	0.3502	0.2959	0.3371	**0.2824**	**0.3301**	0.2847	0.3330
	192	**0.3639**	**0.3806**	0.3875	0.4066	0.3770	0.3910	0.3743	0.3876	0.3880	0.3969	0.3734	0.3902	0.3747	0.3915	0.3648	0.3869	0.3832	0.3989	0.3642	0.3810	0.3691	0.3870	0.3642	0.3820
	336	0.4120	0.4188	0.4548	0.4570	0.4160	0.4250	0.4158	0.4193	0.4238	0.4281	0.4234	0.4292	0.4225	0.4274	**0.4096**	0.4219	0.4426	0.4382	0.4146	0.4205	0.4138	0.4209	0.4107	**0.4172**
	720	0.4101	0.4300	**0.4037**	0.4455	0.4180	0.4370	0.4236	0.4360	0.4285	0.4405	0.4328	0.4450	0.4242	0.4401	0.4067	**0.4291**	0.4613	0.4635	0.4292	0.4409	0.4238	0.4365	0.4174	0.4331
	Avg.	0.3690	0.3926	0.3861	0.4144	0.3758	0.3983	0.3783	0.3954	0.3874	0.4026	0.3828	0.4015	0.3790	0.4002	**0.3667**	0.3929	0.3980	0.4127	0.3760	0.3949	0.3723	0.3936	0.3693	**0.3913**
Weather	96	0.1601	0.2002	0.2056	0.2334	0.1960	0.2360	0.1818	0.2174	0.2146	0.2571	0.1778	0.2114	0.1813	0.2151	0.1550	0.2050	0.1712	0.2123	**0.1517**	0.1925	0.1521	**0.1859**	0.1652	0.2005
	192	0.2019	0.2405	0.2424	0.2688	0.2400	0.2710	0.2259	0.2558	0.2611	0.2888	0.2252	0.2525	0.2291	0.2560	0.2010	0.2450	0.2231	0.2569	0.1993	0.2377	**0.1976**	**0.2342**	0.2105	0.2416
	336	0.2603	0.2786	0.2851	0.3056	0.2920	0.3070	0.2799	0.2944	0.3126	0.3211	0.2784	0.2924	0.2835	0.2953	**0.2370**	**0.2650**	0.2835	0.3013	0.2579	0.2815	0.2558	0.2790	0.2671	0.2827
	720	0.3477	0.3405	0.3447	0.3543	0.3650	0.3540	0.3561	0.3443	0.3801	0.3647	0.3527	0.3400	0.3589	0.3457	**0.3120**	0.3340	0.3561	0.3491	0.3358	0.3331	0.3376	0.3330	0.3425	**0.3320**
	Avg.	0.2425	0.2649	0.2694	0.2905	0.2733	0.2920	0.2609	0.2780	0.2921	0.3079	0.2585	0.2741	0.2632	0.2780	**0.2263**	0.2623	0.2585	0.2799	0.2362	0.2612	0.2358	**0.2580**	0.2463	0.2642
PEMS08	12	0.0935	0.1968	0.1401	0.2530	0.1220	0.2330	0.1410	0.2432	0.4604	0.4540	0.1111	0.2230	0.0945	0.1979	0.1190	0.2220	0.1450	0.2479	0.1268	0.2680	**0.0856**	**0.1848**	0.1043	0.2075
	24	0.1486	0.2407	0.2350	0.3301	0.2360	0.3300	0.2120	0.3000	0.2039	0.3067	0.1760	0.2839	0.1381	0.2398	0.1490	0.2490	0.1923	0.2898	0.1727	0.3074	**0.0983**	**0.2001**	0.1825	0.2822
	48	0.2010	0.2889	0.4315	0.4538	0.5620	0.5400	0.3877	0.4041	0.4311	0.4792	0.3536	0.4110	0.2443	0.3276	0.2080	0.2920	0.2700	0.3512	0.2868	0.3965	**0.1458**	**0.2426**	0.3656	0.4090
	96	0.3617	0.3698	0.6781	0.5545	1.2160	0.8460	0.6828	0.5333	1.5089	0.9611	0.6992	0.5769	0.4604	0.4540	0.3290	0.3550	0.5079	0.5011	0.4735	0.5034	**0.2380**	**0.3057**	0.7555	0.5830
	avg.	0.2012	0.2740	0.3712	0.3978	0.5340	0.4873	0.3559	0.3702	0.6510	0.5503	0.3350	0.3737	0.2344	0.3048	0.2013	0.2795	0.2788	0.3475	0.2650	0.3688	**0.1419**	**0.2333**	0.3520	0.3704
Counts		**18**	**17**	1	0	1	0	0	4	1	4	2	1	1	1	12	6	0	0	9	7	12	14	3	6

Table 7: Full experimental results in Short-Term forecasting.

Categories		Linear-Based										Transformer-Based						CNN-Based						Hybrid	
Models		**TRACE** (Ours)		DLinear (2023)		FITS (2024)		SparseTSF (2026)		MixLinear (2026)		PatchTST (2023)		iTrans. (2024)		TimeMixer++ (2025)		TimesNet (2023)		ModernTCN (2024)		FACT (2026)		interPDN (2026)	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ILI	3	0.5600	0.4058	1.5320	0.8333	1.4610	0.7430	1.0342	0.5312	0.5510	0.3789	0.6460	0.4170	0.5827	0.3938	0.5934	0.3880	0.6815	0.4573	**0.4969**	**0.3590**	0.7957	0.5009	0.8991	0.5316
	6	1.0310	0.5724	2.3055	1.0385	2.3370	0.9740	1.6487	0.7228	1.2064	0.5993	1.2690	0.6290	1.0577	0.5937	1.2136	0.6052	1.1329	0.6275	**0.9848**	**0.5574**	1.1168	0.6270	1.5150	0.7256
	9	1.5413	0.7360	3.0110	1.2038	3.3970	1.1970	2.1979	0.8795	2.0535	0.7944	2.0210	0.8300	1.6194	0.7279	1.9029	0.7978	1.8018	0.8082	**1.4947**	**0.7150**	1.8686	0.9887	2.3751	0.9520
	12	**2.2130**	0.9302	3.7492	1.3378	4.2440	1.3610	3.7405	1.2672	2.7416	0.9575	2.7880	1.0160	2.2979	**0.8954**	2.7025	0.9855	2.5677	0.9738	2.5954	0.9452	2.7617	1.0118	2.7879	1.0049
	Avg.	**1.3363**	0.6611	2.6494	1.1033	2.8598	1.0688	2.1553	0.8502	1.6381	0.6825	1.6810	0.7230	1.3894	0.6527	1.6031	0.6941	1.5460	0.7167	1.3930	**0.6441**	1.6357	0.7821	1.8943	0.8035
Wiki	24	**1.5099**	**0.7701**	3.4856	1.3121	4.2650	1.5230	4.2110	1.3134	3.6528	1.2927	2.0460	0.8490	2.3279	0.9679	2.2099	0.8608	2.6411	0.9044	2.1376	0.8635	1.9899	0.8714	2.4248	0.9918
	36	**1.6427**	**0.7969**	3.1697	1.2279	3.7180	1.3630	4.2947	1.3108	3.0603	1.1661	2.3440	0.9120	2.1566	0.9303	1.9649	0.8335	2.3955	0.9461	2.1184	0.8761	2.4466	0.9833	2.2307	0.9420
	48	**1.7530**	**0.7985**	3.0238	1.1930	3.9940	1.4220	4.6571	1.3845	2.8241	1.1373	2.1230	0.8830	2.2264	0.9484	1.9861	0.8451	2.4328	0.9316	2.0388	0.8476	2.2373	0.9020	2.3825	0.9806
	60	**1.7734**	**0.8158**	3.2336	1.2456	4.5430	1.5540	4.7101	1.4233	2.7230	1.1399	2.0010	0.8950	2.2889	0.9804	2.0298	0.8727	2.3447	0.9469	1.8391	0.8334	1.9754	0.8525	2.2679	0.9634
	Avg.	**1.6697**	**0.7953**	3.2282	1.2446	4.1300	1.4655	4.4683	1.3580	3.0650	1.1840	2.1285	0.8848	2.2499	0.9568	2.0477	0.8530	2.4535	0.9322	2.0335	0.8552	2.1623	0.9023	2.3265	0.9695
	3	**6.5421**	0.4003	7.6719	0.4383	7.4700	0.4960	7.9576	0.4126	7.3390	0.4750	7.4934	**0.3947**	7.4636	0.4056	7.5562	0.4098	8.0588	0.4264	7.4852	0.3996	7.6416	0.4108	7.5563	0.4089
	6	**6.6888**	0.4026	7.5372	0.4549	8.3260	0.5440	7.5069	0.4058	7.8261	0.4932	7.4063	0.4269	7.2669	0.3963	7.2761	**0.3939**	7.6161	0.4279	7.3217	0.4111	7.2737	0.3948	6.8354	0.4320
	9	**6.8359**	**0.4068**	7.4300	0.4873	8.8690	0.5640	7.2826	0.4180	7.9709	0.5093	7.1002	0.4137	7.0802	0.4090	7.0990	0.4083	7.3251	0.4285	7.1425	0.4200	7.1546	0.4108	7.2612	0.4872
	12	**6.9941**	0.4000	7.3736	0.4873	9.3940	0.6080	7.1643	0.4157	8.0741	0.5204	7.0915	0.4136	7.0765	0.4110	7.0730	**0.3996**	7.2655	0.4309	7.1495	0.4271	7.0627	0.4049	7.0630	0.4105
	Avg.	**6.7652**	**0.4024**	7.5032	0.4670	8.5148	0.5530	7.4778	0.4130	7.8025	0.4995	7.2728	0.4122	7.2218	0.4055	7.2511	0.4029	7.5664	0.4284	7.2747	0.4145	7.2832	0.4053	7.1790	0.4347
	24	**7.0543**	**0.4307**	7.5198	0.4815	7.8870	0.6130	7.5666	0.4342	7.3755	0.5491	7.5138	0.4352	7.5573	0.4486	7.5663	0.4339	7.6972	0.4738	7.5861	0.4553	7.5290	0.4320	7.5472	0.4362
	36	6.6306	0.4735	**6.4021**	0.4874	7.7740	0.6600	6.4781	0.4407	7.0039	0.5734	6.4290	**0.4315**	6.4779	0.4566	7.5663	0.4339	6.6272	0.4849	6.5295	0.4654	6.4454	0.4410	6.4678	0.4429
	48	6.2085	0.4811	**5.7744**	0.4923	7.7370	0.7060	5.8571	0.4455	6.6578	0.5899	5.8320	**0.4437**	5.8712	0.4619	6.5135	0.4481	6.0179	0.4921	5.9506	0.4752	5.8460	0.4482	5.8702	0.4504
	60	5.9057	0.5059	**5.5338**	0.4943	7.8550	0.7460	5.6309	**0.4502**	6.4017	0.6026	5.6672	0.4720	5.6686	0.4710	5.6569	0.4563	5.8749	0.5060	5.8069	0.4869	5.6231	0.4535	5.6631	0.4577
	Avg.	6.4498	0.4728	**6.3075**	0.4889	7.8133	0.6813	6.3831	**0.4426**	6.8597	0.5787	6.3605	0.4456	6.3938	0.4595	6.8257	0.4430	6.5543	0.4892	6.4683	0.4707	6.3609	0.4437	6.3871	0.4468
Counts		**15**	**11**	4	0	0	0	0	3	3	1	3	4	3	3	2	9	0	0	5	6	3	3	2	0

A.5.1 Impact of Decay Rate

The decay rate λ controls how rapidly the coupling strength attenuates with temporal distance in C to modulates the overall chaoticity of the coupled system. We examine its effect by fixing $t_{\text{evo}} = 0.1$, $M = 20$, and $\alpha = 0.5$. As shown in Fig. 6, TRACE maintains stable performance across a broad range of λ, with optimal configurations clustering at higher decay rates ($\lambda \geq 0.5$), where the intermediate dynamical system attains a higher Maximal Lyapunov Exponent (as illustrated in left of Fig. 8). This insensitivity indicates strong parametric robustness.

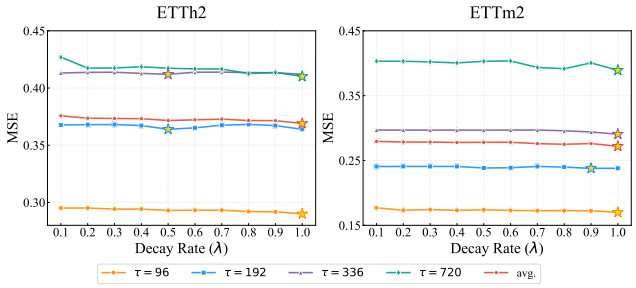


Figure 6: Effect of decay rate λ on MSE with $t_{\text{evo}} = 0.1$, $M = 20$, $\alpha = 0.5$.

A.5.2 Impact of Coupling Strength

Coupling strength α governs the intensity of inter-node interactions and directly affects the overall chaoticity. We evaluate its impact by fixing $t_{\text{evo}} = 0.1$, $M = 20$, and $\lambda = 1.0$. As delineated in Fig. 7, TRACE demonstrates high robustness, with MSE curves remaining remarkably stable across $\alpha \in [0.1, 1.0]$. Notably, the optimal configurations exhibit a horizon-dependent divergence between weak coupling ($\alpha = 0.1$) and moderate coupling ($\alpha = 0.5$).

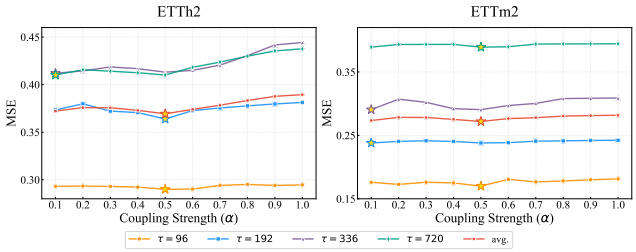


Figure 7: Effect of coupling strength α on MSE with $t_{\text{evo}} = 0.1$, $M = 20$, $\lambda = 1$.

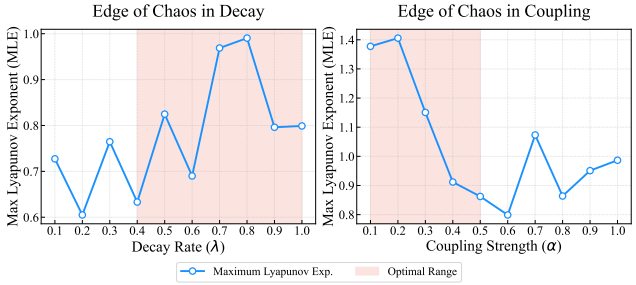


Figure 8: Asymptotic MLE of the coupled system with respect to decay rate λ (left) and coupling strength α (right). Shaded areas highlight the optimal parameter regimes for predictive performance.

To further understand these optimal configurations, we analyze the asymptotic MLE of the coupled system across λ and α, as shown in Fig. 8. The empirically optimal windows ($\lambda \geq 0.4$, $\alpha \in [0.1, 0.5]$) correspond to a positive MLE regime, confirming that TRACE operates near the edge of chaos, where dynamics are neither too dissipative nor excessively turbulent. Longer prediction horizons favor smaller α for sustained expansion, while shorter horizons benefit from stronger coupling that prioritizes structural stability.

A.6 Parameters Sensitivity

A.6.1 The Number of Encoders and Decoders

L_{enc} and L_{dec} denote the number of MLP layers in the encoder and decoder, respectively. We investigate their impact on TRACE, as summarized in Fig. 9. As shown, TRACE performs best with shallow encoder and decoder configurations, confirming that deep peripheral layers are unnecessary when paired with the chaotic core.

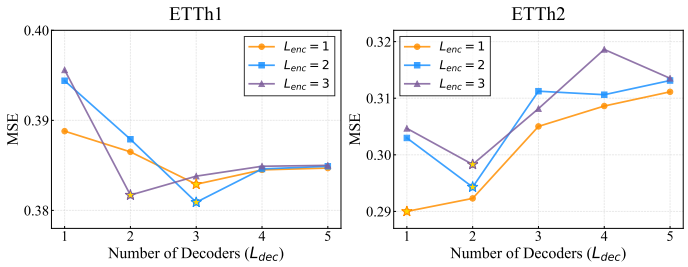


Figure 9: Effect of varying encoder and decoder depths. Input length $T = 96$ and prediction length $\tau = 96$.

A.6.2 Impact of Hidden Dimension

d_{enc} and d_{dec} control the hidden dimensions of the encoder and decoder, respectively. Their effects on TRACE are visualized in Fig. 10. TRACE favors an asymmetric dimensional configuration, a compact encoder paired with an expanded decoder to fully decode the rich representations generated by the chaotic core.

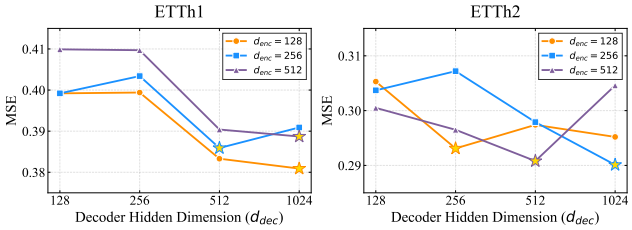


Figure 10: Effect of varying hidden dimensions in encoder and decoder. Input length $T = 96$ and prediction length $\tau = 96$.

A.7 More Analysis

A.7.1 Comparison with ESN/RC

As discussed in Section 1, RC is the closest existing paradigm to our motivation. However, its foundational ESP requires the system to asymptotically forget initial conditions, which fundamentally clashes with the long-term sensitivity of chaos. This theoretical mismatch makes RC/ESN a natural baseline for comparing two opposing design principles. To this end, we compare TRACE with the vanilla ESN (Jaeger, 2001) and its improved variant, LCESN (Matzner & Mráz, 2025).

Table A.7.1 compares TRACE with ESN and LCESN. LCESN is competitive only at short horizons; as the prediction length grows, its error increases sharply. This degradation reflects the core mismatch of reservoir computing: the ESP forces the system to asymptotically forget initial conditions, whereas chaotic temporal processing (as shown in Fig. 1) relies on sustained sensitivity to them. TRACE avoids this conflict by evolving a fixed transient chaotic core and reading out its finite-time state, thereby preserving perturbation sensitivity without any additional trainable nonlinear layers.

Table 8: Comparison with ESN and LCESN.

Dataset	Horizon	TRACE		ESN		LCESN	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	0.3211	0.3403	0.3603	0.3805	**0.3184**	0.3642
	192	0.3638	0.3716	0.4097	0.4095	0.3684	0.3907
	336	0.3951	0.3900	0.4377	0.4255	0.4250	0.4380
	720	**0.4574**	**0.4254**	0.4982	0.4559	0.5701	0.5224
	Avg.	**0.3844**	**0.3819**	0.4265	0.4179	0.4205	0.4288
ETTm2	96	**0.1700**	**0.2485**	0.1880	0.2664	0.1795	0.2780
	192	0.2380	**0.2907**	0.2941	0.3037	**0.2330**	0.3191
	336	**0.2906**	**0.3277**	0.3082	0.3399	0.2980	0.3290
	720	**0.3891**	**0.3891**	0.4074	0.3958	0.4390	0.4670
	Avg.	**0.2719**	**0.3140**	0.2994	0.3265	0.2874	0.3483
ETTh1	96	**0.3802**	**0.3890**	0.4055	0.4084	0.5148	0.5058
	192	**0.4297**	**0.4202**	0.4597	0.4306	0.7090	0.6190
	336	**0.4661**	**0.4382**	0.5250	0.4678	0.9305	0.7010
	720	**0.4618**	**0.4531**	0.5619	0.4998	1.0270	0.8008
	Avg.	**0.4344**	**0.4251**	0.4880	0.4516	0.7953	0.6567
ETTh2	96	**0.2900**	**0.3408**	0.3071	0.3513	0.3234	0.3868
	192	**0.3639**	**0.3806**	0.3801	0.3994	0.4045	0.4410
	336	**0.4120**	**0.4188**	0.4323	0.4321	0.4730	0.4880
	720	**0.4101**	**0.4300**	0.4498	0.4532	0.5563	0.5445
	Avg.	**0.3690**	**0.3926**	0.3924	0.4090	0.4393	0.4651
Weather	96	**0.1601**	**0.2002**	0.1687	0.2038	0.1695	0.2155
	192	**0.2019**	**0.2405**	0.2144	0.2488	0.2180	0.2690
	336	**0.2603**	**0.2786**	0.2804	0.2998	0.2640	0.3220
	720	0.3477	0.3405	**0.3466**	**0.3387**	0.3519	0.3409
	Avg.	**0.2425**	**0.2649**	0.2525	0.2728	0.2509	0.2869

A.7.2 Impact of Embedding Strategy

By default, TRACE embeds each time step as a node, mapping $\mathbb{R}^{N\times T} \to \mathbb{R}^{d\times T}$. An alternative strategy is to embed each variate as a node, mapping $\mathbb{R}^{N\times T} \to \mathbb{R}^{N\times d}$. We compare the two embedding strategies in Table 9. As shown in Table 9, embedding each time step as a node consistently outperforms the variate-wise alternative, since chaotic evolution amplifies temporal differences across nodes and is fully exploited only when each time step is mapped to an independent initial state.

Table 9: Comparison of different initial state embedding strategies

State Embedding	ETTh1		ETTh2		ILI(S1)		ILI(S2)	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$N \to d$	**0.4344**	**0.4251**	**0.3690**	**0.3926**	**1.3363**	**0.6611**	**1.6697**	**0.7953**
$T \to d$	0.4655	0.4532	0.3884	0.4105	1.6145	0.7015	2.0876	0.9545

A.7.3 Numerical Integration vs. Automatic ODE Solving

TRACE uses RK4 for numerical integration. And we compare it with Euler integrators and an adaptive ODE solver (e.g., `torchdiffeq`) in terms of forecasting accuracy and computational efficiency, and further examine the effect of the integration step count M. As shown in Figs. 11 to 13, RK4 achieves the lowest MSE at moderate step counts and closely tracks the adaptive solver's accuracy, while maintaining deterministic, lightweight execution. This establishes fixed-step RK4 as the optimal choice for integrating the chaotic core within deep learning pipelines.

A.7.4 Gradient Flow Analysis

We further examine the gradient flow through the chaotic core under RK4 and Euler numerical solvers. The gradient norm along the unrolled dynamics is visualized in Fig. 14 to assess stability and distinguish the propagation behaviors of the compared integrators. As shown in Fig. 14, RK4 maintains smooth and stable gradient propagation even under deep unrolling, whereas the Euler

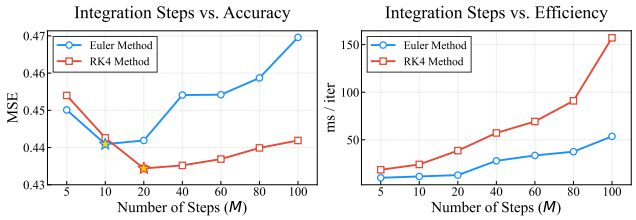


Figure 11: Impact of integration step count M with Euler and RK4 solvers (evaluated on ETTh1, $96 \rightarrow 96$, $t_{\text{evo}} = 0.1$).

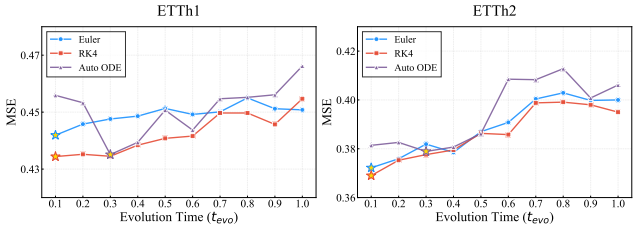


Figure 12: Comparison of forecasting MSE among Euler, RK4, and adaptive ODE solvers across varying t_{evo} (evaluated on ETTh1, average performance, $t_{\text{evo}} = 0.1$).

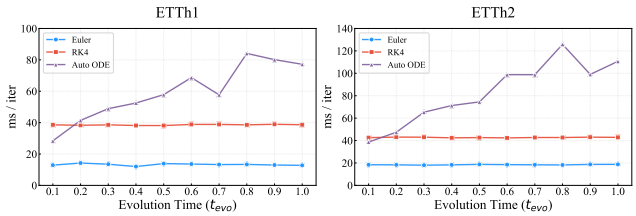


Figure 13: Computational efficiency comparison (ms/iter) among Euler, RK4, and adaptive ODE solvers with respect to t_{evo} (evaluated on ETTh1, $96 \rightarrow 96$, $t_{\text{evo}} = 0.1$).

method suffers from sharp gradient discontinuities due to accumulating truncation errors. The adaptive solver is excluded from this comparison because its adjoint-based backward integration does not share the same discrete-step gradient structure, making it hard to direct visualization.

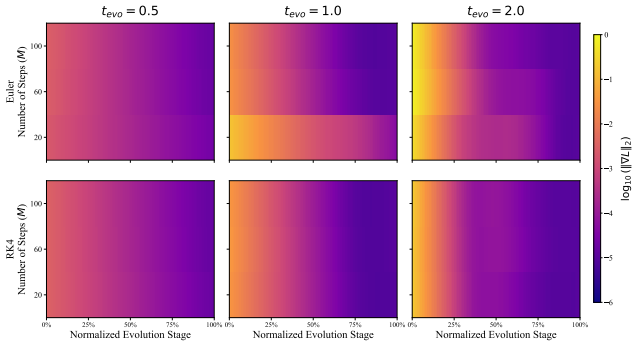


Figure 14: Evolutionary gradient flow analysis of the chaotic core under different numerical integration schemes (evaluated on ETTh1, $96 \rightarrow 96$).

A.7.5 NON-CHAOTIC ALTERNATIVES

To verify that the transient chaotic dynamics are indispensable to TRACE, we implement a non-chaotic alternative by replacing the default chaotic core with a heavily damped, predictable system. Specifically, we instantiate a 2D damped harmonic oscillator governed by the second-order linear differential equation: $\ddot{x} + \gamma\dot{x} + \omega^2 x = 0$. By introducing states x and $y = \dot{x}$, this system is unrolled as a coupled first-order ODE system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = y, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -\omega^2 x - \gamma y. \tag{28}$$

Results in Fig. 15 show that the non-chaotic variant consistently underperforms the original TRACE, and its forecasting error escalates as damping strengthens. This suggests that while a stable dissipative system may still retain some representational capacity for simple patterns, it lacks the continuous state expansion and mixing inherent to chaotic dynamics, which appear crucial for preserving the complex temporal dependencies required for accurate long-term forecasting.

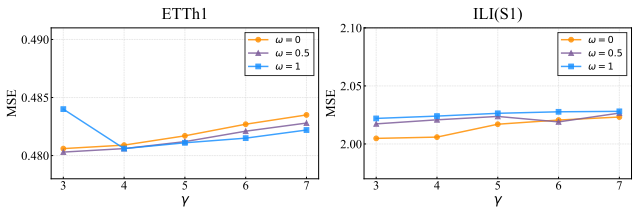


Figure 15: Ablation study on non-chaotic alternatives within the dynamical core (evaluated on ETTh1 and ILI(S1), $96 \to 96$, $t_{\text{evo}} = 0.1$).

A.7.6 MORE CHAOS OSCILLATORS

We further evaluate other low-dimensional chaotic systems beyond the default Lorenz, including Rössler (Rössler, 1976), Bouali (Bouali, 1999), Hyper-Lorenz (Chen & Yang, 2014), Hyper-Rössler (Rössler, 1976), and the finance-inspired (Cai & Huang, 2007) systems. Results are summarized in Table 10. TRACE exhibits strong compatibility across chaotic oscillators, with different priors suiting different data pattern. Higher-dimensional attractors (Hyper-Lorenz, Hyper-Rössler, $d = 4$) excel on power datasets, while the classic 3D Lorenz system achieves the best performance on epidemiological data, capturing volatile, sparse temporal patterns more effectively.

Table 10: Performance comparison across various chaos oscillators

Systems	Lorenz	Rössler	Hyperlorenz	Hyperrössler	Bouali	Finance
Dim.	3	3	4	4	3	3
ETTh1	0.4344	0.4398	**0.4338**	0.4350	0.4527	0.4629
ETTh2	0.3690	0.4001	0.4021	**0.3673**	0.3876	0.4026
ILI(S1)	**1.3363**	1.4043	1.3966	1.5796	1.4997	1.6997
ILI(S2)	**1.6697**	1.8565	1.6712	1.7627	1.8009	2.0169

A.7.7 HIGHER-DIMENSIONAL CHAOTIC CORES

We investigate if increasing the dimensionality of the chaotic oscillator can enhance the representational capacity of TRACE, as the common practice of expanding hidden dimension in nerual networks. To this end, we replace the default Lorenz system with a high dimensional Lorenz-96 mode (set $F = 8$) (Lorenz, 2006) and evaluate its performance under identical settings. As shown in Fig. 16, scaling the Lorenz-96 core dimension yields remarkably flat error curves. This contrasts sharply with the common sense in nerual networks, where expanding hidden dimensions typically boosts performance, and suggests that TRACE's capacity is governed primarily by chaotic mixing rather than the raw dimensionality of the embedding space. Once fully developed chaos is established, a low-dimensional attractor already suffices to encode complex temporal dependencies,

justifying the use of compact chaotic cores as a parameter-efficient alternative to over-parameterized architectures.

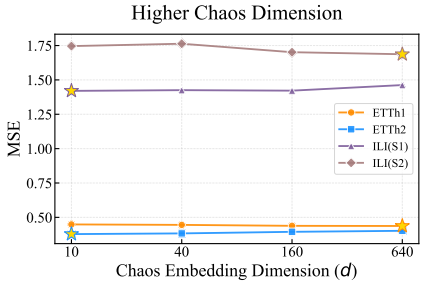


Figure 16: Impact of scaling the chaotic core dimensionality using the Lorenz-96 system (configured with forcing constant $F = 8$, evaluated across ETTh1, ETTh2, ILI(S1), and ILI(S2)).

A.7.8 Coupling Strategy

By default, TRACE constructs the coupling matrix based on temporal causality with an exponential decay. To further validate this choice, we compare it against alternative coupling strategies, including a fully-coupled variant (no mask), and linear decay. Results are reported in Table 11. The causal mask overall outperform the fully-coupled strategy, and exponential decay surpasses linear decay, together validating that a causal, exponentially fading coupling provides the effective prior for chaotic evolution.

Table 11: Performance comparison of different coupling strategies and decay types.

Strategy	Decay	ETTh1		ETTh2		ILI(S1)		ILI(S2)	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
all	linear	0.4565	0.4409	0.4026	0.4149	1.4145	0.7015	1.8665	0.9217
	exponent	**0.4342**	**0.4251**	0.3754	0.3964	1.5831	0.8020	1.6876	0.8145
temporal causal	linear	0.4454	0.4387	0.3804	0.4014	1.5034	0.7452	1.6794	0.8092
	exponent	0.4344	**0.4251**	**0.3690**	**0.3926**	**1.3363**	**0.6611**	**1.6697**	**0.7953**

A.7.9 Interpretability and Visualizations

To provide an intuitive understanding of the transient dynamics inside TRACE, we visualize the phase space trajectories projected on the x-y plane along with the state evolution of each node. Representative results under the Lorenz system are shown in Fig. 17 and Fig. 18.

Fig. 17 shows that the continuous vector field condenses arbitrary initial states onto a bounded, low-dimensional chaotic attractor, preventing divergence. Fig. 18 confirms that this geometric constraint produces rich, synchronized oscillatory features across components. Together, they demonstrate that TRACE leverages the expressiveness of chaotic attractors to construct a geometrically stable and interpretable latent space.

A.7.10 Numerical Validation of the Chaotic Sensitivity Constant

The generalization bound in Appendix A.2.2 depends on the sensitivity of the fixed flow,

$$K_{\text{chaos}}(t_{\text{evo}}) = \sup_{\mathbf{H}} \|D\Phi_{t_{\text{evo}}}(\mathbf{H})\|_2, \tag{29}$$

which is not an absolute constant but scales as $e^{\lambda_{\max} t_{\text{evo}}}$. We estimate it by integrating the variational equation $dJ/dt = D\mathbf{F}(\mathbf{s})J,\ J(0) = I$, along the Lorenz ODE from multiple attractor initial states and taking $K_{\text{chaos}}(t) = \sigma_{\max}(J(t))$. Here the standard chaotic Lorenz parameters are used ($\sigma = 10$, $\beta = 8/3$, $\rho = 28$), and the pointwise maximum over initial states is taken to approximate the supremum in the definition above.

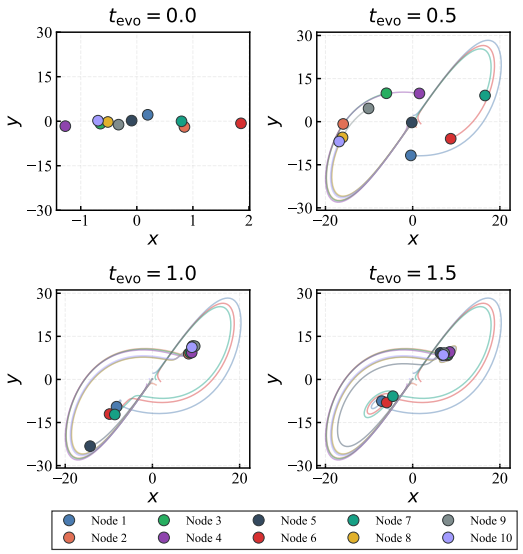


Figure 17: 2D phase space trajectory projection of a $T = 10$-node system across four transient evolution snapshots ($t_{\text{evo}} \in \{0.0, 0.5, 1.0, 1.5\}$).

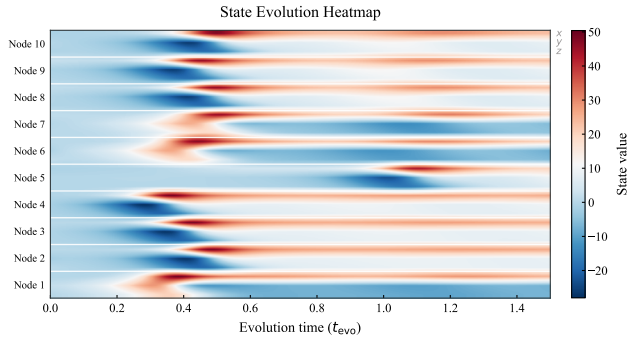


Figure 18: State evolution heatmap visualizing the continuous 3D coordinate transformation for $T = 10$ nodes over normalized evolution time t_{evo}.

Fig. 19 shows that K_{chaos} grows monotonically with t_{evo}, whereas MSE is V-shaped with the optimum at $t_{\text{evo}} \in [0.05, 0.2]$. The knee of K_{chaos} coincides with the onset of MSE degradation, confirming that the optimal evolution time is the transient window where chaotic mixing suffices for feature separation while flow sensitivity remains moderate. This also indicates that the degradation at large t_{evo} is intrinsic to the chaotic flow rather than an artifact of numerical integration.

We emphasize that the growth of K_{chaos} is not in conflict with the global volume contraction in Proposition 4.1. Volume contraction is governed by $|\det(D\Phi_{t_{\text{evo}}})|$, whereas K_{chaos} is governed by the largest singular value of $D\Phi_{t_{\text{evo}}}$. A dissipative chaotic flow can therefore contract phase-space volume globally while stretching selected directions locally—exactly the stretch-and-fold mechanism underlying TRACE. This provides an empirical justification for treating $K_{\text{chaos}}(t_{\text{evo}})$ as a time-dependent factor in the generalization bound, rather than an absolute constant, and explains why increasing t_{evo} indefinitely is harmful despite the parameter-free nature of the chaotic core.

A.7.11 Empirical Risk and Generalization Gap

Section A.2 establishes a *conditional* generalization advantage for TRACE over RNNs: when the evolution time t_{evo} remains in the transient regime where $K_{\text{chaos}}(t_{\text{evo}})$ is moderate, the complexity penalty of TRACE can be tighter than that of a deep recurrent stack, provided the empirical risk is

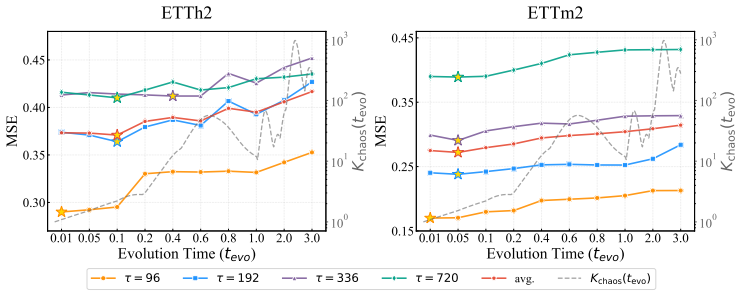


Figure 19: Numerical estimation of $K_{\text{chaos}}(t_{\text{evo}})$ (right axis) overlaid with MSE (left axis) on ETTh2 and ETTm2. K_{chaos} is the maximal singular value of the variational Jacobian along the Lorenz flow. The optimal MSE window $t_{\text{evo}} \in [0.05, 0.2]$ coincides with the transient regime where K_{chaos} remains moderate; beyond it, its rapid growth accompanies the MSE upturn.

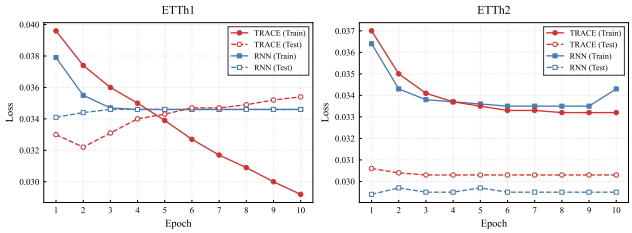


Figure 20: Empirical evaluation of learning risk and generalization dynamics.

comparable. This is a conditional statement, since the bound in Eq. 25 depends on $K_{\text{chaos}}(t_{\text{evo}})$ and thus on the chosen evolution regime, rather than holding uniformly over all configurations.

To examine this prediction empirically, we compare the training error, test error, and their gap between TRACE and an RNN baseline. The RNN consists of 5 layers with a hidden dimension of 256, and TRACE is evaluated at its default transient regime $t_{\text{evo}} = 0.1$. Results are summarized in Fig. 20. Since the theory predicts a *conditional* rather than an unconditional advantage, we do not expect the empirical gap of TRACE to be uniformly smaller; instead, we examine whether the test error remains competitive while the complexity penalty is substantially reduced.