

FreKoo++: Learning Continuous Spectral Dynamics for Temporal Domain Generalization

En Yu, Xiaoyu Yang, Wei Duan, Guangquan Zhang, Jie Lu *IEEE Fellow*

Abstract—Temporal Domain Generalization (TDG) aims to learn from historical domains and generalize to unseen future distributions under concept drift. Nevertheless, prevailing TDG methods struggle with complex real-world streaming scenarios involving both multi-scale drift patterns (e.g., long-term periodicity intertwined with short-term incremental changes) and local uncertainties, especially in continuous settings where observations arrive irregularly. To address this limitation, we propose FreKoo++, a novel continuous spectral-dynamical framework that pioneers the unification of continuous Koopman modal dynamics with adaptive spectral disentanglement. Specifically, FreKoo++ maps source-domain parameters into a compact latent space, modeling their evolution as a superposition of learnable continuous modes where complex eigenvalues jointly encode oscillatory frequency and temporal growth or decay. This formulation naturally accommodates irregular timestamps and supports arbitrary horizon extrapolation without rigid discrete stepping. Furthermore, we propose a new adaptive soft spectral weighting mechanism backed by stability and spectral regularization, which automatically isolates persistent dominant dynamics from transient noise without relying on manual frequency thresholds. We derive modal approximation and generalization bounds that characterize how amplitude and eigenvalue estimation errors propagate with the prediction horizon. Extensive experiments on both discrete and continuous TDG benchmarks demonstrate that FreKoo++ achieves state-of-the-art performance under complex multi-scale drifts and irregular sampling.

Index Terms—Domain Generalization, Concept Drift, Dynamic modeling, Frequency Analysis.

I. INTRODUCTION

MODERN machine learning models face significant challenges in dynamic environments where data distributions evolve over time [1]–[3]. Unlike static settings that assume an IID relationship between training and test data, real-world scenarios often involve continuously generated data streams (e.g., user activity logs, sensor readings, financial transactions) exhibiting temporal dependencies and non-stationarity due to factors like shifting user behavior, environmental variations, or system changes [4]–[6]. These temporal dynamics lead to distributional shifts over time, known as *concept drift*, which invalidates the alignment between historical and future out-of-distribution (OOD) data [7]–[9]. Consequently, models trained on past data frequently fail to generalize to future instances, leading to performance degradation and reduced reliability. This has spurred increasing interest in *Temporal Domain Generalization (TDG)*, which

En Yu, Xiaoyu Yang, Wei Duan, Guangquan Zhang and Jie Lu are with the Australian Artificial Intelligence Institute (AAIL), Faculty of Engineering and Information Technology, University of Technology Sydney (UTS), Australia (e-mail: en.yu-1@uts.edu.au; xiaoyu.yang-3@student.uts.edu.au; wei.duan@uts.edu.au; guangquan.zhang@uts.edu.au; jie.lu@uts.edu.au.)

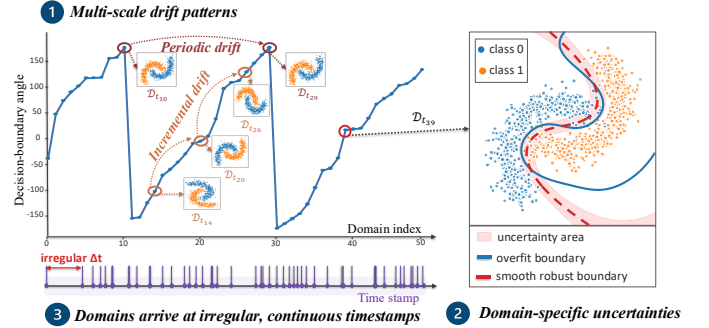


Fig. 1: Illustration of the three TDG challenges on a 2-Moons stream. The underlying distribution exhibits multi-scale periodicity, while empirical parameter estimates are corrupted by domain-specific uncertainties. Additionally, the irregularly sampled observations violate the time-invariant discrete transition assumed by standard sequence models.

seeks to learn from chronologically ordered source domains and generalize to unseen future distributions under evolving temporal shifts [10].

Recent progress in TDG generally follows two temporal modeling paradigms: *discrete* and *continuous* approaches. *Discrete TDG (DTDG)* methods operate under the rigid assumption of uniformly spaced temporal grids, enforcing step-wise transitions or generating synthetic OOD samples to model temporal evolution [10]–[13]. However, in real-world applications such as social media streams and clinical healthcare, observations unfold asynchronously and unpredictably [14]. This causes temporal domains to be irregularly distributed over continuous time [15]. Binding dynamics to fixed discrete steps prevents discrete methods from characterizing the continuous dynamics of irregular time-distributed domains. To solve this, *continuous TDG (CTDG)* approaches have recently emerged, leveraging continuous domain indices or neural ODEs to enable prediction at arbitrary timestamps [15]–[17].

Nevertheless, most methods implicitly assume that concept drift is primarily incremental or focus on ensuring local smoothness between adjacent domains, which causes two challenges as visualized in Fig. 1: 1) *Difficulty in multi-scale drift patterns*: real-world concept drift frequently exhibits long-term periodicity (e.g., seasonality, weekly/daily user patterns, economic cycles), not just smoothly incremental changes [18], [19]. Current approaches struggle to capture these recurring patterns spanning longer time horizons [12]. Their effectiveness often diminishes, particularly near phase transitions of periodic cycles where the inability to anticipate

pattern recurrence hinders generalization. 2) *Vulnerability to Overfitting Domain-Specific Uncertainties*: current TDG methods typically segment continuous data streams into discrete temporal domains. However, the complex drift patterns inherent in real-world streams make it difficult to guarantee that data within each domain remains IID. Consequently, localized uncertainties and non-IID characteristics within domains can complicate optimization and increase the risk of overfitting to domain-specific artifacts, ultimately undermining stable cross-temporal generalization [20].

These challenges underscore the necessity for principled frameworks that can capture persistent multi-scale dynamics while filtering local uncertainties. Since model parameters optimized on successive temporal domains provide a compact proxy for distributional evolution, we model parameter trajectories from a spectral perspective [15], [21], [22]. In the spectral domain, long-term trends and periodic components are typically concentrated in low-frequency structures, whereas transient domain-specific uncertainties often appear as high-frequency variations [23]–[26]. This separation offers a useful inductive bias for isolating predictable temporal dynamics from disruptive local fluctuations. Building upon this intuition, our preliminary work **FreKoo** [21] integrated Discrete Fourier Transform (DFT) with Koopman operator theory for discrete TDG. It disentangled parameter trajectories into distinct spectral bands and leveraged Koopman dynamics to extrapolate low-frequency dominant patterns while regularizing high-frequency uncertainties. However, FreKoo is strictly bound to uniform temporal grids and relies on manually pre-defined, non-adaptive frequency thresholds. This rigid design prevents dynamic spectral disentanglement and renders it fundamentally incompatible with continuous-time data streams where observations arrive asynchronously.

In this paper, we propose **FreKoo++**, a unified framework for continuous temporal domain generalization. FreKoo++ is motivated by the observation that the solution of a continuous-time linear dynamical system admits an exponential form that can be interpreted as a spectral decomposition in the Laplace domain. Specifically, each eigenvalue of the Koopman operator encodes both the growth/decay rate and oscillatory frequency of a latent dynamical mode. This connection enables us to directly parameterize temporal evolution through learnable Koopman eigenvalues, bypassing the need for discrete Fourier transforms and naturally supporting irregularly sampled observations. Concretely, FreKoo++ is implemented as a continuous-time parameterized module that learns Koopman eigenvalues from irregularly sampled parameter trajectories. This formulation supports prediction at arbitrary future timestamps without relying on discrete time stepping. Furthermore, leveraging the continuous spectral representation, we introduce a learnable soft spectral weighting mechanism together with a spectral energy regularization strategy, which adaptively emphasizes dominant low-frequency structures while suppressing high-frequency noise. This design removes the need for manually defined frequency thresholds and allows end-to-end optimization. To provide theoretical grounding, we analyze the approximation behavior of the proposed continuous-time formulation, bounding its modal approximation error and

deriving a generalization bound whose growth is controlled by the learned spectrum.

Our main contributions are summarized as follows:

- We propose **FreKoo++**, a pioneering continuous-time TDG framework that shifts the paradigm of parameter evolution from discrete-step transitions to learnable Koopman-inspired modal dynamics. By encoding frequency and temporal growth/decay into continuous eigenvalues, FreKoo++ naturally accommodates irregularly sampled source domains and arbitrary-horizon prediction.
- We introduce an adaptive spectral disentanglement mechanism with spectral and stability regularization, which automatically disentangles persistent dynamics from transient fluctuations without requiring manual frequency selection.
- We provide comprehensive theoretical and empirical validation for FreKoo++. Our theoretical analysis establishes rigorous approximation and generalization guarantees, while extensive experiments across diverse TDG benchmarks demonstrate its superior performance and robust extrapolation capabilities.

II. RELATED WORK

A. Temporal Domain Generalization

1) Discrete Temporal Domain Generalization (DTDG):

DTDG focuses on learning from chronologically ordered source domains to generalize to unseen future distributions [10]. The core objective is to effectively leverage temporal evolutionary cues without accessing target domain data. Existing approaches fall broadly into *data driven* and *model centric* paradigms. *Data driven* methods enhance temporal robustness by synthesizing temporally shifted distributions or enforcing invariant features. For instance, CODA constructs concept drift simulators to generate future-like samples [13], Foresee synthesizes out-of-distribution instances for non-stationary environments [11], and recent methods model instance-level evolution or learn temporally invariant representations to improve robustness to future shifts [27]–[29]. *Model centric* approaches instead integrate temporal dynamics directly into optimization objectives or model parameters. Gradient interpolation regularization encourages decision boundaries to evolve smoothly over time [10], DRAIN forecasts future neural network parameters via drift-aware recurrent architectures [12], and sequential autoencoders capture evolving latent domain structures [8], [30]. Despite these advances, most DTDG methods implicitly rely on uniformly spaced temporal grids and assume smooth incremental shifts.

2) Continuous Temporal Domain Generalization (CTDG):

Continuous temporal learning relaxes the restrictive assumption of rigid, uniformly spaced temporal grids. Early continuous domain adaptation frameworks treat time as a continuous domain index to learn smoothly transferable representations across evolving distributions [16], [31], [32]. More recently, Continuous Temporal Domain Generalization (CTDG) has emerged to address future-domain extrapolation under continuous, irregular observation timestamps [15], [33]. By supporting asynchronous observation arrivals and arbitrary-horizon

forecasting, CTDG represents a critical paradigm shift toward realistic streaming environments.

However, existing CTDG methods still exhibit two fundamental limitations. First, prevailing formulations rely heavily on monotonic or locally smooth evolution assumptions, rendering them inadequate for complex, multi-scale, or recurring concept drifts. Second, methods driven by standard continuous-time dynamics—such as Neural ODEs or continuous-time recurrent units—typically model parameter or feature transitions directly in the time domain [14], [34], [35]. As a result, they lack an explicit mechanism to disentangle persistent macro-trends from transient domain-specific fluctuations. This research overcomes these bottlenecks by establishing a continuous Koopman spectral representation, where complex eigenvalues explicitly quantify frequency and temporal persistence, enabling adaptive spectral disentanglement under irregular temporal sampling.

B. Concept Drift

Concept drift signifies a change in the underlying data distribution over time, formally occurring between time t and $t+1$ if the joint distribution $P_{t+1}(X, y) \neq P_t(X, y)$. This non-stationarity poses a fundamental challenge, requiring models to adapt dynamically to maintain predictive performance and reliability [7], [18], [36], [37]. Prior research treats it as unpredictable and often employing detect-then-adapt strategies [3], [38]. In contrast, the Temporal Domain Generalization (TDG) setting frequently focuses on leveraging more predictable evolutionary dynamics for proactive generalization [12], [39]. However, existing TDG methods often oversimplify these dynamics, primarily modeling smooth, incremental drifts while struggling to capture complex long-range temporal structures like periodicity which are prevalent in many real-world data streams yet demand different approaches than simple monotonic progression.

Furthermore, current TDG involves segmenting the continuous data stream into temporal domains $\{D_1, D_2, \dots, D_T\}$, assuming concept drift occurs primarily between these sequential domains [10]. This often struggles to perfectly align with real-world data streams, potentially violating the implicit assumption that data within each domain D_t is Independent and Identically Distributed (IID). Consequently, individual domains may harbor internal non-IID structures or localized noise patterns [40], [41]. The presence of intra-domain complexities leads to sensitivity to domain-specific noise and artifacts, hindering robust generalization, particularly for TDG models designed predominantly to address shifts occurring only at the domain boundaries. Effectively handling both the diverse inter-domain evolutionary dynamics (including periodicity) and these intra-domain data characteristics is therefore crucial for robust temporal generalization.

C. Frequency Learning and Koopman Dynamics

Frequency-domain learning has become an effective tool for modeling non-stationary temporal data. By transforming signals from the time domain to the spectral domain, frequency-aware methods can capture periodicity, seasonality, and multi-scale temporal dependencies that may be difficult to identify

through local time-domain statistics. For example, frequency-enhanced architectures have been used in long-term time-series forecasting [42], [43], adaptive time-frequency modeling [44], and frequency normalization for non-stationary forecasting [23]. Classical frequency analysis has also been widely used to detect periodic structures in temporal signals [45]–[47]. These studies suggest that spectral representations provide a useful inductive bias for separating stable temporal patterns from high-frequency disturbances.

Koopman theory provides another principled view of temporal dynamics by representing nonlinear evolution through a linear operator in an appropriate lifted space [48], [49]. Recent machine learning methods have combined Koopman operators with neural representations for forecasting dynamical systems and non-stationary time series [4], [25]. In temporal domain generalization, Koopman-based methods have also been explored to model evolving domains [28]. Our preliminary FreKoo framework first applies DFT-based decomposition to uniformly sampled parameter trajectories and then extrapolates selected components through discrete Koopman dynamics [21]. FreKoo++ instead directly learns a continuous modal representation, enabling spectral allocation and arbitrary-time extrapolation within a unified latent space.

III. PRELIMINARY

A. Temporal Domain Generalization

Given a chronologically ordered sequence of temporal domains extracted from historical data streams, Temporal Domain Generalization (TDG) aims to train a model that generalizes to unseen future data distributions. Let $\mathcal{T} = \{t_1, t_2, \dots, t_T\}$ denote the set of historical timestamps, where $t_i \in \mathbb{R}^+$ and $t_1 < t_2 < \dots < t_T$. At each timestamp t_i , we observe a source domain $\mathcal{D}_{t_i} = \{(\mathbf{x}_{t_i}^{(j)}, y_{t_i}^{(j)})\}_{j=1}^{N_i}$, drawn from a time-dependent joint distribution $P_{t_i}(X, Y)$ over a common input space $\mathcal{X}$ and label space $\mathcal{Y}$. Due to *concept drift*, $P_{t_i}(X, Y)$ may differ from $P_{t_j}(X, Y)$ for $i \neq j$. Depending on the temporal structure of the observed domains, TDG can be categorized into two paradigms:

Discrete TDG (DTDG) assumes uniformly spaced timestamps, i.e., $\Delta t_i = t_{i+1} - t_i \equiv C$. Given historical source domains $\{\mathcal{D}_{t_i}\}_{i=1}^T$, the objective is to learn a model $g(\cdot; \theta) : \mathcal{X} \rightarrow \mathcal{Y}$ that generalizes to the next unseen target domain $\mathcal{D}_{t_{T+1}} \sim P_{t_{T+1}}(X, Y)$ without accessing $\mathcal{D}_{t_{T+1}}$ during training [10].

Continuous TDG (CTDG) relaxes the uniform-spacing assumption and allows Δt_i to vary across domains. The objective is to learn a temporal prediction mechanism that generates model parameters $\theta(t_s)$ for an arbitrary future timestamp $t_s > t_T$, such that $g(\cdot; \theta(t_s))$ generalizes to the unseen target domain $\mathcal{D}_{t_s} \sim P_{t_s}(X, Y)$ without access to future-domain data during training [15].

B. Challenges

To characterize temporal distribution shifts from the perspective of model evolution, we consider the trajectory of optimal parameters induced by the time-varying distribution $P_t(X, Y)$. Specifically, let $\theta^*(t) =$

$\arg \min_{\theta} \mathbb{E}_{P_t} [\ell(g(X; \theta), Y)]$ denote the optimal parameter configuration at time t . In practice, this underlying trajectory cannot be directly observed. Instead, at each historical timestamp $t_i \in \mathcal{T}$, we obtain an empirical estimate $\bar{\theta}_{t_i}$ by optimizing the model on the finite temporal domain $\mathcal{D}_{t_i}$. Due to finite samples, intra-domain non-stationarity, and domain-specific variations, these empirical estimates inevitably deviate from the underlying trajectory. We model this observation process as,

$$\bar{\theta}_{t_i} = \theta^*(t_i) + \epsilon(t_i), \quad (1)$$

where $\epsilon(t_i)$ summarizes estimation error and local domain-specific fluctuations. Based on this parameter-space formulation, continuous temporal domain generalization confronts three fundamental challenges:

Challenge 1 (Multi-scale drift patterns in parameter space): Prevailing TDG methods often model temporal evolution through incremental changes or local smoothness between adjacent parameter configurations. While such formulations handle gradual shifts, they fail to capture long-range recurring dynamics. In many real-world streams, the underlying parameter trajectory $\theta^*(t)$ exhibits periodic or multi-scale patterns, where similar states reappear after a temporal lag—i.e., $\exists L > 0$ such that $\theta^*(t) \approx \theta^*(t + kL)$ for $k \in \mathbb{Z}^+$, despite short-term variations where $\theta^*(t) \neq \theta^*(t + \Delta)$. This non-monotonic trajectory cannot be adequately captured by step-wise local smoothness alone.

Challenge 2 (Parameter-space uncertainties and noise corruption): The empirical parameter sequence $\{\bar{\theta}_{t_i}\}_{i=1}^T$ is corrupted by the perturbation term $\epsilon(t_i)$ in Eq. (1), which encapsulates finite-sample effects and domain-specific variations. Directly treating this noisy sequence as a clean dynamical signal causes models to overfit these transient fluctuations rather than tracking the true, persistent parameter trajectory $\theta^*(t)$, thereby severely degrading future-domain generalization.

Challenge 3 (Irregular timestamps and arbitrary-time parameter extrapolation): CTDG requires predicting the parameter configuration $\theta(t_s)$ at an arbitrary future timestamp $t_s > t_T$, while historical domains arrive at non-uniform intervals $\Delta t_i = t_{i+1} - t_i$. Existing discrete methods rely on time-invariant one-step parameter transitions (e.g., $\bar{\theta}_{t_{i+1}} = F(\bar{\theta}_{t_i})$) designed for uniform grids ($\Delta t_i \equiv C$). Under irregular sampling, such discrete transitions fail to scale with varying temporal gaps, leading to compounding error accumulation when recursively projected to arbitrary future times.

IV. FREKOO: FREQUENCY-KOOPMAN FRAMEWORK

For the DTDG setting, our preliminary framework **FreKoo** [21] models parameter trajectories through a two-stage spectral-dynamical approach (as shown in Fig. 2), i.e., Fourier-based spectral decomposition and frequency-specific dynamic modeling. Specifically, given a time-varying parameter trajectory $\Theta = [\theta_1, \theta_2, \dots, \theta_T]^\top \in \mathbb{R}^{T \times D}$ with T uniform timesteps and D dimensions, FreKoo first transforms it into frequency space via Discrete Fourier Transform (DFT) along the temporal axis: $\Theta^f = \mathcal{F}(\Theta)$, where $\Theta^f[f, d] = \sum_{t=1}^T \Theta[t, d] \cdot e^{-j2\pi ft/T}$. To isolate dominant dynamics, it computes the average spectral magnitude across dimensions

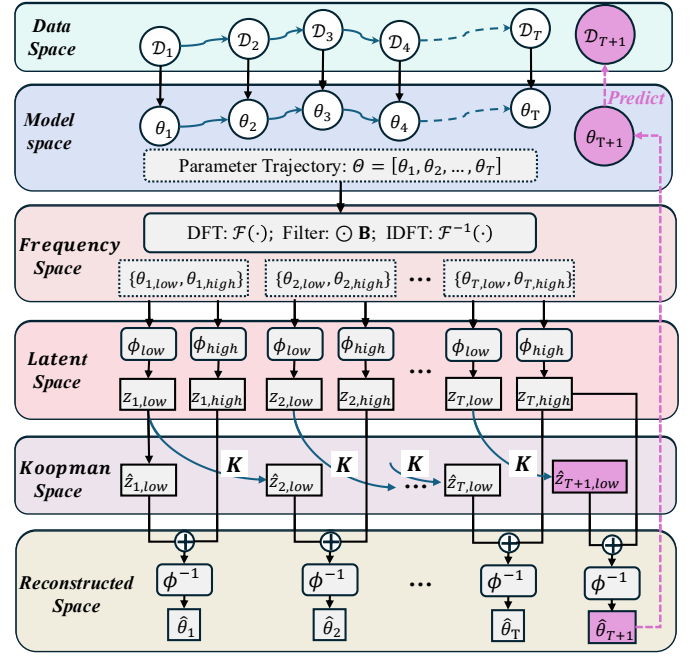


Fig. 2: FreKoo framework. It decomposes model parameter trajectories into low- and high-frequency components via the Fourier transform. Koopman dynamics extrapolate the low-frequency component, while temporal-difference regularization suppresses unstable high-frequency variations.

as an energy proxy, $M_f = \frac{1}{D} \sum_{d=1}^D |\Theta^f[f, d]|$. Based on a specified energy preservation ratio $\tau \in [0, 1]$, a binary mask $\mathbf{B} \in \{0, 1\}^{N_{\text{freq}} \times D}$ selects the top- Q frequency indices with the largest energy magnitudes. Reversing the transformation via Inverse DFT ($\mathcal{F}^{-1}$) yields a trajectory decomposition:

$$\Theta_{\text{low}} = \mathcal{F}^{-1}(\Theta^f \odot \mathbf{B}), \Theta_{\text{high}} = \mathcal{F}^{-1}(\Theta^f \odot (\mathbf{1} - \mathbf{B})), \quad (2)$$

where $\Theta = \Theta_{\text{low}} + \Theta_{\text{high}}$. Here, Θ_{low} encapsulates dominant trends and long-term periodic patterns (*Challenge 1*), whereas Θ_{high} captures transient domain-specific uncertainties (*Challenge 2*). To model the persistent low-frequency dynamics $\Theta_{\text{low}} = \{\theta_{1,\text{low}}, \dots, \theta_{T,\text{low}}\}$, FreKoo maps them into a higher-dimensional Hilbert space via an encoder $\phi_{\text{low}} : \mathbb{R}^D \rightarrow \mathbb{R}^m$, assuming near-linear evolution governed by a learnable Koopman operator $K \in \mathbb{R}^{m \times m}$. The low-frequency transition is expressed as $z_{t,\text{low}} = \phi_{\text{low}}(\theta_{t,\text{low}})$ and $\hat{z}_{t+1,\text{low}} = K z_{t,\text{low}}$, supervised by a Koopman consistency loss:

$$\mathcal{L}_{\text{koop}} = \sum_{t=1}^{T-1} \|z_{t+1,\text{low}} - \hat{z}_{t+1,\text{low}}\|_2^2. \quad (3)$$

Conversely, for the volatile high-frequency sequence $\Theta_{\text{high}} = \{\theta_{1,\text{high}}, \dots, \theta_{T,\text{high}}\}$, forward extrapolation is avoided to prevent noise amplification. Instead, FreKoo maps it via an encoder ϕ_{high} and imposes a temporal difference smoothness regularization:

$$\mathcal{R}_{\text{high}} = \sum_{t=1}^{T-1} \|z_{t+1,\text{high}} - z_{t,\text{high}}\|_2^2. \quad (4)$$

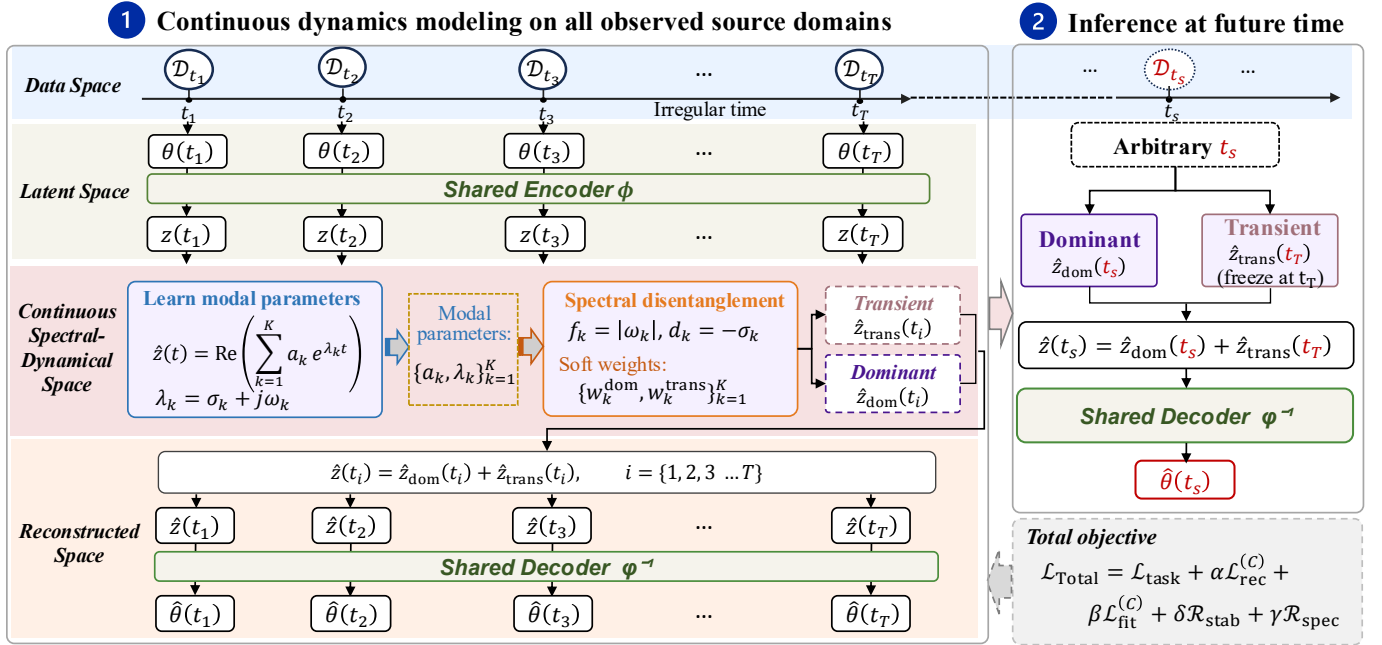


Fig. 3: FreKoo++ framework. **(1) Training Phase:** Model parameters from irregular timestamps are encoded into a continuous latent trajectory and fitted by Koopman modal dynamics. Adaptive soft weighting disentangles persistent dominant dynamics from transient noise, yielding an optimized continuous spectral representation via joint optimization. **(2) Inference Phase:** For an arbitrary target timestamp $t_s > t_T$, FreKoo++ analytically extrapolates the persistent dominant component to t_s , freezes the transient residual at t_T , and decodes their superposition to obtain target parameters $\hat{\theta}(t_s)$.

To predict parameters for the next timestamp $\hat{\theta}_{t+1}$, FreKoo fuses the extrapolated low-frequency latent with the regularized high-frequency context through a shared decoder ϕ^{-1} :

$$\hat{\theta}_{t+1} = \phi^{-1}(K\phi_{\text{low}}(\theta_{t,\text{low}}) + \phi_{\text{high}}(\theta_{t,\text{high}})). \quad (5)$$

The overall framework is trained end-to-end under the joint objective:

$$\mathcal{L}_{\text{frekoo}} = \mathcal{L}_{\text{task}} + \alpha \mathcal{L}_{\text{rec}} + \beta \mathcal{L}_{\text{koop}} + \gamma \mathcal{R}_{\text{high}}, \quad (6)$$

where $\mathcal{L}_{\text{task}}$ is the empirical source risk, and $\mathcal{L}_{\text{rec}} = \sum_{t=1}^{T-1} \|\theta_{t+1} - \hat{\theta}_{t+1}\|_2^2$ enforces overall parameter reconstruction fidelity.

Discussion 1 (Limitations of FreKoo): While FreKoo effectively addresses Challenges 1 and 2 in DTDG, it exhibits two inherent limitations: 1) Its reliance on DFT requires uniformly spaced domains ($\Delta t_i \equiv C$), making it fundamentally incompatible with continuous streams where observations arrive asynchronously (*Challenge 3*); and 2) It separates spectral decomposition from dynamical modeling, requiring distinct encoders ($\phi_{\text{low}}, \phi_{\text{high}}$) and a manually pre-defined threshold τ , which prevents dynamic, end-to-end spectral disentanglement. These constraints necessitate a unified continuous time framework with adaptive spectral selection.

V. FREKOO++: CONTINUOUS FREQUENCY-KOOPMAN FRAMEWORK

To overcome the limitations of discrete pipelines outlined in Discussion 1, we introduce FreKoo++, which establishes a unified continuous-time formulation that seamlessly bridges

spectral analysis and dynamic operator modeling. As shown in Fig. 3, FreKoo++ maps temporal parameter trajectories into a compact latent space and models their evolution via learnable continuous modal dynamics. Driven by these modal characteristics, an adaptive soft spectral weighting mechanism disentangles persistent dominant dynamics from domain-specific fluctuations. By operating directly in continuous time, FreKoo++ naturally handles irregular temporal observations and enables analytical prediction at arbitrary future horizons without requiring step-wise numerical integration.

A. Latent Trajectory Construction

For each observed timestamp t_i , let $\theta(t_i) \in \mathbb{R}^D$ denote the model parameters associated with the source domain $\mathcal{D}_{t_i}$. We collect these parameters across all timestamps to form a parameter trajectory $\Theta = [\theta(t_1), \theta(t_2), \dots, \theta(t_T)]^\top \in \mathbb{R}^{T \times D}$, which forms a time-stamped parameter trajectory and provides a task-dependent representation of all evolving source domains. Directly modeling temporal dynamics in the high-dimensional parameter space $\mathbb{R}^D$ is computationally expensive and may retain domain-specific parameter variations that are irrelevant to future prediction. FreKoo++ addresses this by projecting the parameter trajectory into a compact latent dynamical space $\mathbb{R}^m$ ($m \ll D$) using a shared encoder $\phi: \mathbb{R}^D \rightarrow \mathbb{R}^m$:

$$z(t_i) = \phi(\theta(t_i)), \quad i = 1, \dots, T, \quad (7)$$

where $z(t_i) \in \mathbb{R}^m$ represents the latent dynamical state at time t_i . The sequence $\{z(t_i)\}_{i=1}^T$ defines the observed continuous latent trajectory across all source domains.

B. Continuous Spectral-Dynamical Modeling

Given the observed latent trajectory $\{z(t_i)\}_{i=1}^T$, our primary goal is to establish a continuous-time parameterization that explicitly exposes its underlying spectral and dynamical properties. We draw theoretical motivation from continuous-time Koopman spectral theory [48], [49]. When observables span a finite-dimensional Koopman-invariant subspace, the infinitesimal generator of the continuous-time Koopman operator admits exponential modal decompositions whose eigenvalues govern both frequency and temporal growth or decay.

Proposition 1 (Koopman Spectral-Dynamical Relation): Let $z(t) \in \mathbb{R}^m$ denote the continuous latent state sampled at historical timestamps $\{t_i\}_{i=1}^T$ according to Eq. (7). Suppose the continuous-time evolution of $z(t)$ is governed by a finite K -dimensional Koopman-invariant subspace. Let $\mathcal{K}_c \in \mathbb{C}^{K \times K}$ be the matrix representation of the infinitesimal generator of the continuous Koopman operator restricted to this subspace. If $\mathcal{K}_c$ is diagonalizable, then the continuous latent trajectory admits the exact modal expansion:

$$z(t) = \text{Re} \left(\sum_{k=1}^K c_k v_k e^{\lambda_k t} \right), \quad (8)$$

where $\{(\lambda_k, v_k)\}_{k=1}^K$ denote the eigenvalue-mode pairs with eigenvalues $\lambda_k \in \mathbb{C}$ and Koopman mode vectors $v_k \in \mathbb{C}^m$, and $c_k \in \mathbb{C}$ is the modal coefficient determined by initial conditions. Each eigenvalue $\lambda_k = \sigma_k + j\omega_k$ jointly characterizes the continuous temporal behavior: the imaginary part $\omega_k = \text{Im}(\lambda_k)$ defines the oscillatory frequency, while the real part $\sigma_k = \text{Re}(\lambda_k)$ dictates the temporal growth ($\sigma_k > 0$) or decay ($\sigma_k < 0$) rate.

Proposition 1 provides the theoretical justification for our architectural design. In practice, finding an exact Koopman-invariant subspace via neural encoders is unfeasible. Instead, FreKoo++ directly employs the closed-form modal solution in Eq. (8) as a *parametric dynamical prior* to approximate the latent trajectory.

To eliminate scalar unidentifiability between c_k and v_k , we absorb them into a unified complex modal amplitude vector $a_k = c_k v_k \in \mathbb{C}^m$. Collecting all amplitudes as $A = [a_1, \dots, a_K] \in \mathbb{C}^{m \times K}$, the continuous modal approximation is parameterized as:

$$\tilde{z}(t) = \text{Re} \left(\sum_{k=1}^K a_k e^{\lambda_k t} \right). \quad (9)$$

Evaluating Eq. (9) at any timestamp t yields a continuous latent state without requiring step-wise transition rollouts. To bypass complex-valued optimization during backpropagation, we implement Eq. (9) through its real-valued trigonometric equivalent:

$$\tilde{z}(t) = \sum_{k=1}^K e^{\sigma_k t} [\text{Re}(a_k) \cos(\omega_k t) - \text{Im}(a_k) \sin(\omega_k t)]. \quad (10)$$

This reformulation guarantees that $\tilde{z}(t) \in \mathbb{R}^m$ is evaluated natively in real arithmetic. Non-oscillatory secular trends are naturally recovered when $\omega_k = 0$. Consequently, the learnable spectral parameters $\Lambda = \{\lambda_k\}_{k=1}^K$ and amplitudes $A \in \mathbb{C}^{m \times K}$

construct a continuous spectral coordinate system, replacing rigid DFT frequency bins with data-driven eigenvalue-indexed continuous modes.

C. Adaptive Spectral Disentanglement

Based on the spectral interpretation above, FreKoo++ assigns each learned mode a soft dominance weight. The imaginary part ω_k reflects the oscillatory frequency, while the real part σ_k controls the growth or decay behavior. These quantities make it possible to distinguish persistent structures from transient fluctuations, but the separation boundary should adapt to the temporal profile of each dataset.

For each mode $k \in \{1, \dots, K\}$, we define its frequency score as $f_k = |\omega_k|$ and its decay score as $d_k = -\sigma_k$ (for stable modes with $\sigma_k \leq 0$). We assign a soft dominance weight $w_k^{\text{dom}} \in (0, 1)$ using two learnable gating functions:

$$w_k^{\text{dom}} = \sigma(-\kappa_f(f_k - f_0)) \cdot \sigma(-\kappa_d(d_k - d_0)), \quad (11)$$

$$w_k^{\text{trans}} = 1 - w_k^{\text{dom}}, \quad (12)$$

where $\sigma(\cdot)$ is the Sigmoid function, $\kappa_f, \kappa_d > 0$ dictate gate sharpness, and f_0, d_0 are learnable thresholds optimized jointly with the network. A mode receives a high dominance weight ($w_k^{\text{dom}} \approx 1$) if it is both low-frequency and persistent ($\sigma_k \approx 0$). Conversely, modes exhibiting high frequencies or rapid attenuation are routed to the transient component (w_k^{trans}). Applying these weights continuously decomposes the latent trajectory at any time t :

$$\hat{z}_{\text{dom}}(t) = \text{Re} \left(\sum_{k=1}^K w_k^{\text{dom}} a_k e^{\lambda_k t} \right), \quad (13)$$

$$\hat{z}_{\text{trans}}(t) = \text{Re} \left(\sum_{k=1}^K w_k^{\text{trans}} a_k e^{\lambda_k t} \right). \quad (14)$$

Consequently, the full reconstructed latent state at any historical observation timestamp t_i ($i = 1, \dots, T$) is obtained by the superposition of both components:

$$\hat{z}(t_i) = \hat{z}_{\text{dom}}(t_i) + \hat{z}_{\text{trans}}(t_i). \quad (15)$$

D. Joint Optimization

Training FreKoo++ optimizes the source-domain parameters $\{\theta(t_i)\}_{i=1}^T$, the encoder ϕ and decoder ϕ^{-1} , the eigenvalues $\Lambda = \{\lambda_k\}_{k=1}^K$, the modal amplitudes $A \in \mathbb{C}^{m \times K}$, and the spectral thresholds f_0, d_0 under a unified objective. In practice, this is implemented via alternating gradient updates between source-domain task fitting and latent modal fitting, while all components remain coupled through a unified objective. The overall loss consists of three predictive fitting terms and two spectral regularization constraints. The fitting losses optimize source-domain task predictors, align the continuous modal trajectory with encoded latent states, and preserve parameter auto-encoding fidelity. The two spectral constraints govern dominant-mode stability and adaptive spectral disentanglement, respectively.

The task loss ensures predictive accuracy across the observed source domains. Denoting the training data from domain $\mathcal{D}_{t_i}$ as (X_{t_i}, Y_{t_i}) , the task loss is formulated as:

$$\mathcal{L}_{\text{task}} = \sum_{i=1}^T \ell(g(X_{t_i}; \theta(t_i)), Y_{t_i}). \quad (16)$$

To align spectral disentanglement with our continuous inference protocol, we formulate modal trajectory fitting as an asymmetric pseudo-future prediction problem across chronological source domains. Specifically, for each step $i = 2, \dots, T$, we treat t_{i-1} as the latest available observation and t_i as a pseudo-future timestamp. Following the asymmetric extrapolation strategy used at test time, the persistent dominant component is propagated forward to t_i , whereas the volatile transient component is held at the preceding observation t_{i-1} :

$$\hat{z}(t_i | t_{i-1}) = \hat{z}_{\text{dom}}(t_i) + \hat{z}_{\text{trans}}(t_{i-1}). \quad (17)$$

The resulting temporal predictive fitting loss is defined as:

$$\mathcal{L}_{\text{fit}}^{(C)} = \sum_{i=2}^T \|z(t_i) - \hat{z}(t_i | t_{i-1})\|_2^2, \quad (18)$$

where $z(t_i) = \phi(\theta(t_i)) \in \mathbb{R}^m$ is the encoded latent state of domain $\mathcal{D}_{t_i}$. Because the dominant and transient branches are evaluated at distinct timestamps, Eq. (18) depends explicitly on the gating weights $\{w_k^{\text{dom}}, w_k^{\text{trans}}\}_{k=1}^K$. It therefore directly supervises adaptive modal allocation while matching the asymmetric propagation rule executed during inference.

The reconstruction loss enforces auto-encoding consistency between the high-dimensional parameter space $\mathbb{R}^D$ and the compact latent space $\mathbb{R}^m$:

$$\mathcal{L}_{\text{rec}}^{(C)} = \sum_{i=1}^T \|\theta(t_i) - \phi^{-1}(\hat{z}(t_i))\|_2^2. \quad (19)$$

The first spectral constraint regularizes the dominant dynamics against unstable extrapolation. Dominant modes with positive growth rates ($\sigma_k > 0$) can explode exponentially when projected far beyond the historical observation window. We therefore penalize dominant modes exhibiting positive growth:

$$\mathcal{R}_{\text{stab}} = \sum_{k=1}^K w_k^{\text{dom}} (\max(0, \sigma_k))^2, \quad (20)$$

such that stable ($\sigma_k < 0$) or neutrally persistent ($\sigma_k = 0$) dominant modes incur no penalty. Compared with hard skew-symmetric constraints on an explicit generator matrix that force purely imaginary spectra and permit only undamped oscillations [15], this soft regularization accommodates natural temporal decay while explicitly suppressing exponential divergence.

The second spectral constraint suppresses noisy transient components while preventing the learned soft gates from collapsing into degenerate uniform distributions. Because transient modes are not extrapolated at inference time, FreKoo++ penalizes their modal energy and encourages decisive spectral separation:

$$\mathcal{R}_{\text{spec}} = \sum_{k=1}^K w_k^{\text{trans}} \|a_k\|_2^2 + (1 - \text{Var}_{k=1}^K(w_k^{\text{dom}})). \quad (21)$$

Algorithm 1 FreKoo++ Learning and Inference

Require: Source domains $\{\mathcal{D}_{t_1}, \dots, \mathcal{D}_{t_T}\}$ with timestamps $\mathcal{T}$; hyperparameters: $\alpha, \beta, \gamma, \delta$; number of modes K ; epochs N_{ep} .
Ensure: Predicted parameters $\hat{\theta}(t_s)$ for target time t_s .
1: Initialize base model $g(\cdot)$, encoder ϕ , decoder ϕ^{-1} , eigenvalues $\{\lambda_k\}_{k=1}^K$, amplitudes A , thresholds f_0, d_0 .
2: **for** epoch = 1 to N_{ep} **do**
3: **for** $i = 1$ to T **do**
4: Compute task loss for $g(\cdot; \theta(t_i))$ on $\mathcal{D}_{t_i}$.
5: $z(t_i) \leftarrow \phi(\theta(t_i))$ {encode parameters, Eq. (7)}
6: **end for**
7: Compute weights $w_k^{\text{dom}}, w_k^{\text{trans}}$ via Eqs. (11)–(12).
8: Compute reconstructed trajectory $\hat{z}(t_i)$ for $i = \{1, \dots, T\}$ via Eq.15.
9: Compute $\mathcal{L}_{\text{Total}}$ via Eq. (23).
10: Update $\phi, \phi^{-1}, \{\lambda_k\}, A, f_0, d_0$ and $\{\theta(t_i)\}$.
11: **end for**
12: **Inference at target time** $t_s > t_T$
13: $\hat{z}_{\text{dom}}(t_s) \leftarrow \text{Re}(\sum_k w_k^{\text{dom}} a_k e^{\lambda_k t_s})$
14: $\hat{z}_{\text{trans}}(t_T) \leftarrow \text{Re}(\sum_k w_k^{\text{trans}} a_k e^{\lambda_k t_T})$
15: $\hat{\theta}(t_s) \leftarrow \phi^{-1}(\hat{z}_{\text{dom}}(t_s) + \hat{z}_{\text{trans}}(t_T))$
16: **return** $\hat{\theta}(t_s)$

Proposition 2 (Spectral Energy Bound): Let mode $k \in \{1, \dots, K\}$ have amplitude vector $a_k = c_k v_k \in \mathbb{C}^m$ and decay rate $\sigma_k = \text{Re}(\lambda_k) < 0$. Its temporal energy over $[0, T_{\text{max}}]$ is

$$E_k = \|a_k\|_2^2 \cdot \frac{1 - e^{2\sigma_k T_{\text{max}}}}{-2\sigma_k} \leq \frac{\|a_k\|_2^2}{2|\sigma_k|}. \quad (22)$$

Proposition 2 confirms that for decaying modes ($\sigma_k < 0$), temporal energy over $[0, T_{\text{max}}]$ is strictly upper-bounded by the squared modal amplitude $\|a_k\|_2^2$ scaled by decay rates. For neutral modes ($\sigma_k = 0$), the energy scales as $T_{\text{max}} \|a_k\|_2^2$, establishing $\|a_k\|_2^2$ as a timestamp-independent surrogate for transient modal energy in Eq. (21). Meanwhile, the second term in Eq. (21) maximizes the variance of $\{w_k^{\text{dom}}\}_{k=1}^K$, preventing gating weights from collapsing into uninformative uniform assignments and ensuring crisp spectral disentanglement. Crucially, $\mathcal{R}_{\text{spec}}$ operates directly in the continuous spectral coordinate space rather than on adjacent time-domain finite differences, making it naturally invariant to non-uniform sampling intervals.

Overall, the full training objective is formulated as:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{task}} + \alpha \mathcal{L}_{\text{rec}}^{(C)} + \beta \mathcal{L}_{\text{fit}}^{(C)} + \delta \mathcal{R}_{\text{stab}} + \gamma \mathcal{R}_{\text{spec}}, \quad (23)$$

where $\alpha, \beta, \gamma, \delta > 0$ are balancing hyperparameters. All parameters are trained end-to-end via gradient descent.

E. Future Time Inference

Once trained, FreKoo++ executes an analytical extrapolation protocol to predict model parameters for an unseen target domain $\mathcal{D}_{t_s}$ at an arbitrary future timestamp $t_s > t_T$. To balance long-term trends with local temporal context, the decomposed components fulfill asymmetric inference roles. Persistent dominant modes, capturing macro-trends and recurring patterns,

are extrapolated directly to the future query timestamp t_s via $\hat{z}_{\text{dom}}(t_s)$ in Eq. (13). Conversely, transient modes capture volatile local fluctuations that are inherently unpredictable far into the future. Forward-extrapolating these modes risks noise amplification; FreKoo++ thus freezes their contribution at the last observed timestamp t_T , retaining $\hat{z}_{\text{trans}}(t_T)$ as stationary residual context.

Consequently, the predicted latent state $\hat{z}(t_s)$ and reconstructed target parameters $\hat{\theta}(t_s)$ are formulated as:

$$\hat{z}(t_s) = \hat{z}_{\text{dom}}(t_s) + \hat{z}_{\text{trans}}(t_T), \quad (24)$$

$$\hat{\theta}(t_s) = \phi^{-1}(\hat{z}(t_s)). \quad (25)$$

Evaluating closed-form exponential modal expressions at t_s and t_T bypasses recursive step-wise rollouts and numerical ODE solvers, securing constant-time $\mathcal{O}(1)$ query complexity for arbitrary prediction horizons. The complete procedure is detailed in Algorithm 1.

Discussion 2 (FreKoo versus FreKoo++): Compared to FreKoo [21], FreKoo++ achieves three core advances: 1) replacing rigid DFT grids with continuous Koopman eigenvalues ($\lambda_k = \sigma_k + j\omega_k$) to natively support irregular sampling; 2) substituting manual DFT cutoffs with learnable spectral gates (f_0, d_0) under a single encoder; and 3) superseding discrete step-wise transitions (Kz_t) with analytical modal extrapolation ($e^{\lambda_k t_s}$), enabling closed-form $\mathcal{O}(1)$ arbitrary-horizon forecasting.

VI. THEORETICAL ANALYSIS

This section provides rigorous theoretical grounding for FreKoo++ by addressing two fundamental claims: 1) the expressiveness of finite continuous Koopman modes in parameterizing latent trajectories, and 2) the explicit risk control provided by our stability ($\mathcal{R}_{\text{stab}}$) and spectral ($\mathcal{R}_{\text{spec}}$) regularizers. We first bound the continuous modal approximation error induced by finite spectrum truncation. We then derive a continuous-time generalization bound for unseen target domains, proving that long-horizon excess risk is strictly governed by individual modal growth rates.

A. Koopman Mode Approximation Error

We first bound the approximation error of the continuous modal parameterization (Eq. (9)). For notational clarity, theoretical statements use the complex modal form; the same bounds apply to real-valued states since $\text{Re}(\cdot)$ is non-expansive.

Assumption 1 (Finite Modal Dynamics and Transient Remainder): The true latent dynamics admit a finite K^* -order modal expansion $z(t) = \sum_{k=1}^{K^*} a_k e^{\lambda_k t}$ with $a_k \in \mathbb{C}^m$, $\lambda_k \in \mathbb{C}$. For a model retaining $K \leq K^*$ modes, the first K represent modeled dynamics; the remaining $K^* - K$ modes constitute a non-growing transient remainder: $\text{Re}(\lambda_k) \leq 0$ for all $k > K$.

Theorem 1 (Mode Approximation Bound): Under Assumption 1, let $\hat{z}(t) = \sum_{k=1}^K \hat{a}_k e^{\hat{\lambda}_k t}$. Then for any horizon $t_s > 0$:

$$\begin{aligned} \sup_{t \in [0, t_s]} \|z(t) - \hat{z}(t)\|_2 &\leq \underbrace{\sum_{k=1}^K \|a_k - \hat{a}_k\|_2 e^{\hat{\rho}_k t_s}}_{\text{amplitude error}} \\ &\quad + \underbrace{C_a \sum_{k=1}^K |\lambda_k - \hat{\lambda}_k| t_s e^{\varrho_k t_s}}_{\text{eigenvalue error}} \\ &\quad + \underbrace{\sum_{k=K+1}^{K^*} \|a_k\|_2}_{\text{truncation error}}, \end{aligned} \quad (26)$$

where $C_a = \max_{k \leq K} \|a_k\|_2$, $\hat{\rho}_k = \max(\text{Re}(\hat{\lambda}_k), 0)$, $\varrho_k = \max(\text{Re}(\lambda_k), \text{Re}(\hat{\lambda}_k), 0)$.

Proof 1 (Proof sketch): Adding and subtracting $\sum_{k=1}^K \hat{a}_k e^{\hat{\lambda}_k t}$ decomposes the error into amplitude, eigenvalue, and truncation terms. The amplitude term follows from the triangle inequality and $|e^{\lambda_k t}| \leq e^{\hat{\rho}_k t_s}$. For the eigenvalue term, the integral identity $e^{\lambda_k t} - e^{\hat{\lambda}_k t} = (\lambda_k - \hat{\lambda}_k) t \int_0^1 e^{[\hat{\lambda}_k + s(\lambda_k - \hat{\lambda}_k)]t} ds$ yields $|e^{\lambda_k t} - e^{\hat{\lambda}_k t}| \leq |\lambda_k - \hat{\lambda}_k| t e^{\varrho_k t}$. For the truncation term, $\text{Re}(\lambda_k) \leq 0$ implies $|e^{\lambda_k t}| \leq 1$. Full proof is detailed in Supplementary I-A.

Insights. Theorem 1 justifies the objective design in Eq. (23): $\mathcal{L}_{\text{fit}}^{(C)}$ and $\mathcal{L}_{\text{rec}}^{(C)}$ directly reduce the amplitude and eigenvalue estimation terms. The eigenvalue term further reveals that spectral errors are exponentially amplified when $\varrho_k > 0$; $\mathcal{R}_{\text{stab}}$ penalizes positive growth rates in dominant modes, encouraging $\hat{\rho}_k \rightarrow 0$ so that error growth scales at most linearly in t_s . The truncation term depends on the amplitudes of unmodeled modes; $\mathcal{R}_{\text{spec}}$ serves as an empirical surrogate by suppressing amplitudes of transient-weighted modes via w_k^{trans} , while its variance term prevents degenerate gate collapse.

B. Future Time Generalization Bound

We next translate the latent approximation analysis into a generalization bound on future-domain population risk. The derivation adheres to the inference protocol (Eq. (24)): dominant dynamics are extrapolated to $t_s > t_T$, while transient fluctuations are frozen at t_T . Our bound explicitly characterizes how spectral estimation errors compound over the prediction horizon $\Delta_s = t_s - t_T$.

Assumption 2 (Regularity, Realizability, and Modal Alignment): The task loss ℓ , predictive backbone g , and decoder ϕ^{-1} are respectively L_ℓ -, L_g -, and L_{dec} -Lipschitz. The optimal parameter trajectory is realizable in the latent space: there exists $z^*(t)$ such that $\theta^*(t) = \phi^{-1}(z^*(t))$, decomposable as $z^*(t) = z_{\text{dom}}^*(t) + z_{\text{trans}}^*(t)$, where the future transient remainder satisfies $\sup_{t \geq t_T} \|z_{\text{trans}}^*(t)\|_2 \leq B_{\text{trans}}$. The dominant components admit matched modal representations anchored at t_T : for any $\tau \geq 0$,

$$\begin{aligned} z_{\text{dom}}^*(t_T + \tau) &= \text{Re}\left(\sum_{k=1}^K b_k^* e^{\lambda_k^* \tau}\right), \\ \hat{z}_{\text{dom}}(t_T + \tau) &= \text{Re}\left(\sum_{k=1}^K \hat{b}_k e^{\hat{\lambda}_k \tau}\right), \end{aligned} \quad (27)$$

where $\hat{b}_k = w_k^{\text{dom}} \hat{a}_k e^{\hat{\lambda}_k t_T}$ and $\max_{1 \leq k \leq K} \|\hat{b}_k^*\|_2 \leq B_{\text{dom}}$. All learned modal parameters remain bounded after training convergence; in particular, the historical window factor $G_{\text{trans}}(t_T) = \left(\sum_{k=1}^K w_k^{\text{trans}} e^{2\hat{\rho}_k t_T}\right)^{1/2}$ is finite.

Theorem 2 (Continuous Time Generalization Bound): Let $\mathcal{R}_{t_s}(\theta) = \mathbb{E}_{P_{t_s}}[\ell(g(X; \theta), Y)]$. Under Assumptions 1–2, the excess risk $\mathcal{E}_{t_s} := \mathcal{R}_{t_s}(\hat{\theta}(t_s)) - \mathcal{R}_{t_s}(\theta^*(t_s))$ satisfies:

$$\mathcal{E}_{t_s} \leq L_\ell L_g L_{\text{dec}} (E_{\text{dom}}(t_s) + E_{\text{trans}}(t_s)), \quad (28)$$

where $E_{\text{dom}}(t_s) = \|\hat{z}_{\text{dom}}(t_s) - z_{\text{dom}}^*(t_s)\|_2$ and $E_{\text{trans}}(t_s) = \|\hat{z}_{\text{trans}}(t_s) - z_{\text{trans}}^*(t_s)\|_2$. With $\Delta_s = t_s - t_T$, $\hat{\rho}_k = \max(\text{Re}(\hat{\lambda}_k), 0)$, and $\varrho_k = \max(\text{Re}(\lambda_k^*), \text{Re}(\hat{\lambda}_k), 0)$:

$$\begin{aligned} E_{\text{dom}}(t_s) \leq & \underbrace{\sum_{k=1}^K \|\hat{b}_k - b_k^*\|_2 e^{\hat{\rho}_k \Delta_s}}_{\text{amplitude mismatch}} \\ & + \underbrace{B_{\text{dom}} \sum_{k=1}^K |\hat{\lambda}_k - \lambda_k^*| \Delta_s e^{\varrho_k \Delta_s}}_{\text{eigenvalue mismatch}}. \end{aligned} \quad (29)$$

The transient holding error satisfies:

$$E_{\text{trans}}(t_s) \leq B_{\text{trans}} + G_{\text{trans}}(t_T) \sqrt{\mathcal{R}_{\text{spec}}}. \quad (30)$$

Proof 2 (Proof sketch): Lipschitz continuity of ℓ , g , and ϕ^{-1} gives $\mathcal{E}_{t_s} \leq L_\ell L_g L_{\text{dec}} \|\hat{z}(t_s) - z^*(t_s)\|_2$; the triangle inequality with the inference rule yields Eq. (28). For the dominant branch, adding and subtracting $b_k^* e^{\hat{\lambda}_k \Delta_s}$ and bounding the exponential difference via the integral identity $e^{\hat{\lambda}_k \Delta_s} - e^{\lambda_k^* \Delta_s} = (\hat{\lambda}_k - \lambda_k^*) \Delta_s \int_0^1 e^{[\lambda_k^* + u(\hat{\lambda}_k - \lambda_k^*)] \Delta_s} du$ gives Eq. (29). For the transient branch, Cauchy–Schwarz yields $\|\hat{z}_{\text{trans}}(t_T)\|_2 \leq G_{\text{trans}}(t_T) \sqrt{\sum_k w_k^{\text{trans}} \|\hat{a}_k\|_2^2} \leq G_{\text{trans}}(t_T) \sqrt{\mathcal{R}_{\text{spec}}}$, where the last step uses $1 - \text{Var}(w_k^{\text{dom}}) \geq 0$. Full proofs are detailed in Supplementary I-B.

Insights. Theorem 2 shows that dominant extrapolation error is governed by anchor mismatch and spectral estimation error, both scaling with $e^{\hat{\rho}_k \Delta_s}$; $\mathcal{R}_{\text{stab}}$ drives $\hat{\rho}_k \rightarrow 0$, suppressing exponential divergence. For the transient branch, the proposed regularizers provide *dual coverage* over positive modal growth: a growing mode ($\sigma_k > 0$) with high w_k^{dom} is penalized by $\mathcal{R}_{\text{stab}}$, while one routed to the transient branch ($w_k^{\text{trans}} \approx 1$) has its amplitude suppressed by $\mathcal{R}_{\text{spec}}$. This joint effect keeps $G_{\text{trans}}(t_T)$ well-conditioned in practice; when all transient-weighted modes are non-growing, $G_{\text{trans}}(t_T) \leq \sqrt{K}$. Crucially, freezing transients at t_T renders E_{trans} independent of Δ_s , structurally insulating future predictions from horizon-dependent noise amplification.

Corollary 1 (Elimination of Exponential Error Amplification): If $\text{Re}(\lambda_k^*) \leq 0$ and $\text{Re}(\hat{\lambda}_k) \leq 0$ for all k , defining $\varepsilon_b = \sum_{k=1}^K \|\hat{b}_k - b_k^*\|_2$ and $\varepsilon_\lambda = \sum_{k=1}^K |\hat{\lambda}_k - \lambda_k^*|$, then:

$$E_{\text{dom}}(t_s) \leq \varepsilon_b + B_{\text{dom}}(t_s - t_T) \varepsilon_\lambda. \quad (31)$$

Stable dominant dynamics thus eliminate exponential error compounding; under exact spectral recovery ($\varepsilon_\lambda = 0$), the error is bounded by ε_b regardless of the horizon.

VII. EXPERIMENTS

A. Experimental Settings

1) *Datasets:* We evaluate FreKoo++ across both CTDG and DTDG settings. For CTDG, we benchmark on six irregularly sampled temporal datasets, including continuous variants of *2-Moons* and *Rot-MNIST* [54], *Twitter* [55], [56], *Yearbook* [57], *Cyclone* [58], and *House* [15]. For DTDG, we follow the evaluation protocol of FreKoo [21] and test on seven uniformly sampled temporal datasets, including *2-Moons*, *Rot-MNIST*, *ONP*, *Shuttle*, *Elec2*, *HousePrices*, and *ApplianceEnergy*. Because several benchmarks share similar names across the two paradigms but differ in their construction protocols, we append the suffix ‘-C’ to CTDG variants and ‘-D’ to DTDG variants for clarity. Detailed dataset statistics and construction procedures are provided in Supplementary II-A.

2) *Baselines and Implementations:* For CTDG, we compare against time-agnostic baselines, including *Offline*, *LastDomain*, and *IncFinetune*; domain generalization regularizers *IRM* [50] and *V-REx* [51]; continuous domain adaptation methods *CIDA* [31]; and temporal modeling baselines *TKNets* [28], *DRAIN* [12], *DRAIN-Δt* [12], *DeepODE* [52], *NeuralLio* [33] and *Koodos* [15]. For DTDG, we additionally compare with the full suite of baselines adopted in FreKoo [21]. Note that FreKoo is evaluated exclusively under DTDG, as its Discrete Fourier Transform (DFT) requires uniformly spaced timestamps. Detailed descriptions are provided in Supplementary II-B. All methods follow the same data splits, pre-processing, task backbones, and evaluation protocols within each benchmark setting. FreKoo++ employs lightweight MLPs as the encoder ϕ and decoder ϕ^{-1} , with the number of Koopman modes K set equal to the latent dimension m . Detailed implementation, hyperparameter, and training details are provided in Supplementary III.

B. Main Results

1) *Continuous Temporal Domain Generalization:* Table I presents the main performance evaluation across all CTDG benchmarks, where source domains arrive at irregular timestamps and target domains demand arbitrary-horizon forecasting. FreKoo++ consistently outperforms existing baselines in both classification and regression tasks, confirming that modeling parameter trajectories within a continuous spectral-dynamical space is more effective than either omitting temporal ordering or enforcing discrete transition rules.

Specifically, time-agnostic approaches (e.g., *Offline*, *LastDomain*, *IncFinetune*) flatten historical domains into static representations, thereby obscuring persistent temporal drift amidst localized, domain-specific fluctuations. Discrete temporal baselines (e.g., *DRAIN*, *DRAIN-Δt*) incorporate parameter transitions, yet their stepwise prediction mechanisms remain bound to local iterations and lack a global spectral description of the trajectory. While continuous-time architectures (e.g., *DeepODE*, *Koodos*) naturally handle irregular observation intervals, they do not utilize the learned spectrum to separate persistent and transient parameter modes. In contrast, FreKoo++ parameterizes parameter evolution as a superposition of eigenvalue-indexed Koopman modes, where the real

TABLE I: Performance comparison on **CTDG datasets**. The classification tasks report error rates (%) except for the AUC for the Twitter dataset. The regression tasks report MAE. ‘-’ implies that the method does not support the task.

Methods	Classification				Regression	
	2-Moons-C ↓	Rot-MNIST-C ↓	Twitter ↑	Yearbook ↓	Cyclone ↓	House-C ↓
Offline	13.5 ± 0.3	6.6 ± 0.2	0.54 ± 0.09	8.6 ± 1.0	18.7 ± 1.4	19.9 ± 0.1
LastDomain	55.7 ± 0.5	74.2 ± 0.9	0.54 ± 0.12	11.3 ± 1.3	22.3 ± 0.7	20.6 ± 0.7
IncFinetune	51.9 ± 0.7	57.1 ± 1.4	0.52 ± 0.01	11.0 ± 0.8	19.9 ± 0.7	20.6 ± 0.2
IRM [50]	15.6 ± 0.2	8.6 ± 0.4	0.53 ± 0.11	8.3 ± 0.5	18.0 ± 0.8	19.8 ± 0.2
V-REx [51]	12.8 ± 0.2	8.6 ± 0.3	0.58 ± 0.05	8.9 ± 0.5	17.7 ± 0.5	20.2 ± 0.1
CIDA [31]	18.7 ± 2.0	8.3 ± 0.7	0.63 ± 0.03	8.4 ± 0.8	17.0 ± 0.4	10.2 ± 1.0
TKNets [28]	39.6 ± 1.2	37.7 ± 2.0	0.57 ± 0.04	8.4 ± 0.3	-	-
DRAIN [12]	53.2 ± 0.9	59.1 ± 2.3	0.57 ± 0.04	10.5 ± 1.0	23.6 ± 0.5	9.8 ± 0.1
DRAIN- Δt [12]	46.2 ± 0.8	57.2 ± 1.8	0.59 ± 0.02	11.0 ± 1.2	26.2 ± 4.6	9.9 ± 0.1
DeepODE [52]	17.8 ± 5.6	48.6 ± 3.2	0.64 ± 0.02	13.0 ± 2.1	18.5 ± 3.3	10.7 ± 0.4
NeuralLio [33]	4.5 ± 1.3	5.4 ± 1.1	0.68 ± 0.01	6.1 ± 1.2	16.2 ± 0.2	9.0 ± 0.4
Koodos [15]	2.8 ± 0.7	4.6 ± 0.1	0.71 ± 0.02	6.6 ± 1.3	16.4 ± 0.3	9.0 ± 0.2
FreKoo++ (ours)	1.3 ± 0.2	3.8 ± 0.8	0.72 ± 0.01	5.4 ± 0.4	16.0 ± 0.2	8.8 ± 0.1

TABLE II: Performance comparison on **DTDG datasets**. Classification tasks report error rates (%), and regression tasks report MAE. “-” indicates that the method does not support the task.

Methods	Classification					Regression	
	2-Moons-D ↓	Rot-MNIST-D ↓	ONP ↓	Shuttle ↓	Elec2 ↓	House-D ↓	Appliance ↓
Offline	22.4 ± 4.6	18.6 ± 4.0	33.8 ± 0.6	0.77 ± 0.10	23.0 ± 3.1	11.0 ± 0.36	10.2 ± 1.1
LastDomain	14.9 ± 0.9	17.2 ± 3.1	36.0 ± 0.2	0.91 ± 0.18	25.8 ± 0.6	10.3 ± 0.16	9.1 ± 0.7
IncFinetune	16.7 ± 3.4	10.1 ± 0.8	34.0 ± 0.3	0.83 ± 0.07	27.3 ± 4.2	9.7 ± 0.01	8.9 ± 0.5
CDOT [16]	9.3 ± 1.0	14.2 ± 1.0	34.1 ± 0.0	0.94 ± 0.17	17.8 ± 0.6	-	-
CIDA [31]	10.8 ± 1.6	9.3 ± 0.7	34.7 ± 0.6	-	14.1 ± 0.2	9.7 ± 0.06	8.7 ± 0.2
GI [10]	3.5 ± 1.4	7.7 ± 1.3	36.4 ± 0.8	0.29 ± 0.05	16.9 ± 0.7	9.6 ± 0.02	8.2 ± 0.6
LSSAE [30]	9.9 ± 1.1	9.8 ± 3.6	38.8 ± 1.1	0.22 ± 0.01	16.1 ± 1.4	-	-
DDA [11]	9.7 ± 1.5	7.6 ± 0.7	34.0 ± 0.3	0.21 ± 0.02	12.8 ± 1.1	9.5 ± 0.12	6.1 ± 0.1
DRAIN [12]	3.2 ± 1.2	7.5 ± 1.1	38.3 ± 1.2	0.26 ± 0.05	12.7 ± 0.8	9.3 ± 0.14	6.4 ± 0.4
EvoS [53]	3.0 ± 0.4	7.3 ± 0.6	35.4 ± 0.2	0.23 ± 0.01	11.8 ± 0.5	9.8 ± 0.10	7.2 ± 0.1
Koodos [15]	1.3 ± 0.4	7.0 ± 0.3	33.5 ± 0.4	0.24 ± 0.04	-	8.8 ± 0.19	4.8 ± 0.3
FreKoo [21]	1.0 ± 0.3	6.9 ± 0.7	32.3 ± 0.3	0.20 ± 0.02	9.2 ± 0.7	9.0 ± 0.11	4.0 ± 0.1
FreKoo++ (ours)	0.9 ± 0.8	5.5 ± 0.7	31.1 ± 0.2	0.19 ± 0.01	10.1 ± 1.0	9.3 ± 0.10	3.9 ± 0.0

and imaginary components explicitly quantify temporal persistence and oscillatory frequency, respectively. This continuous spectral formulation enables the analytical extrapolation of stable dominant dynamics while insulating future predictions against volatile, non-generalizable noise. These consistent empirical gains validate that our proposed continuous-time dynamics and spectral disentanglement are exceptionally effective for continuous domains and long-horizon extrapolation.

2) *Discrete Temporal Domain Generalization*: Table II evaluates FreKoo++ under DTDG benchmarks with uniformly spaced temporal grids. Although designed for continuous irregularly sampled domains, FreKoo++ remains highly competitive under uniformly sampled DTDG benchmarks. It improves upon FreKoo on five of seven datasets, including 2-Moons-D, Rot-MNIST-D, ONP, Shuttle, and Appliance, while FreKoo remains stronger on Elec2 and House-D. These results demonstrate the strong backward compatibility and versatility of FreKoo++, proving that its continuous spectral-dynamical formulation remains exceptionally effective regardless of whether temporal domains are regularly or irregularly sampled.

C. Ablation Studies

To systematically validate the design rationale of FreKoo++, we conduct ablation studies on representative CTDG bench-

marks across two complementary dimensions: *method-level architectural variants* and *objective-level regularization variants*. As summarized in Table III, the empirical results demonstrate that every component in FreKoo++ plays an indispensable role in securing robust cross-temporal generalization.

The *method-level* variants confirm that our core architectural choices provide decisive performance guarantees. First, enforcing $\sigma_k \equiv 0$ (*Fixed spectrum*) restricts modes to purely unattenuated oscillations ($\lambda_k = j\omega_k$) and neglects temporal growth and decay dynamics. This causes catastrophic degradation across all tasks: classification error surges from 1.3% to 9.7% on 2-Moons-C, Twitter AUC drops sharply to 0.64, and Cyclone MAE degrades to 17.2. These drops prove that modeling growth and decay rates (σ_k) is as vital as modeling oscillatory frequencies (ω_k) for tracking parameter evolution. Second, replacing adaptive soft gates with *Hard gating* severely impairs generalization (e.g., Cyclone MAE worsens to 18.7), confirming that rigid frequency cutoffs introduce severe boundary artifacts. Similarly, locking gating thresholds via *Frozen gating* consistently degrades performance, validating the necessity of dynamically adapting spectral boundaries to data-specific temporal profiles.

The *objective-level* variants demonstrate that all regularizers and loss terms contribute synergistically to trajectory

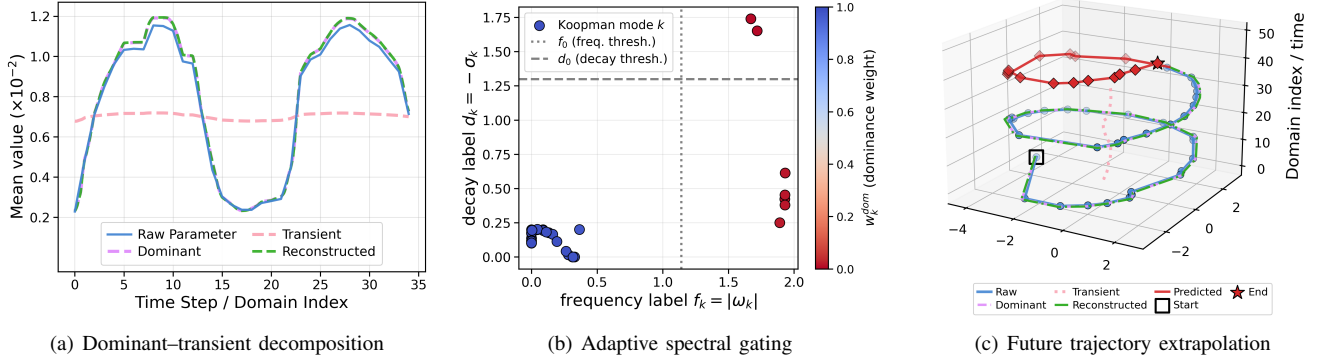


Fig. 4: Qualitative spectral analysis on 2-Moons-C. (a) The parameter trajectory is decomposed into a smooth dominant component and a short-lived transient residual. (b) Learned modes are organized in the frequency–decay plane, where adaptive gates separate persistent modes from transient ones without fixed DFT bins or manual frequency thresholds. (c) Continuous-time inference extrapolates the dominant dynamics while freezing the transient component, producing a coherent future trajectory.

TABLE III: Ablation studies on representative CTDG benchmarks. We report error rate (%), $\downarrow$ for 2-Moons-C, AUC ($\uparrow$) for Twitter, and MAE ($\downarrow$) for Cyclone. The best performance in each column is in **bold**.

Variant	2-Moons-C	Twitter	Cyclone
	Err. $\downarrow$	AUC $\uparrow$	MAE $\downarrow$
<i>Method-level variants</i>			
Fixed spectrum	9.7 ± 1.5	0.64 ± 0.11	17.2 ± 0.7
Hard gating	6.7 ± 2.8	0.67 ± 0.09	18.7 ± 0.3
Frozen gating	2.8 ± 0.3	0.70 ± 0.01	16.2 ± 0.2
<i>Objective-level variants</i>			
w/o $\mathcal{L}_{\text{rec}}^{(C)}$	1.6 ± 0.5	0.70 ± 0.02	16.4 ± 0.3
w/o $\mathcal{L}_{\text{fit}}^{(C)}$	2.0 ± 0.5	0.68 ± 0.02	16.2 ± 0.2
w/o $\mathcal{R}_{\text{stab}}$	2.5 ± 1.4	0.70 ± 0.01	16.1 ± 0.1
w/o $\mathcal{R}_{\text{spec}}$	6.7 ± 3.3	0.69 ± 0.01	16.8 ± 0.4
FreKoo++	1.3 ± 0.2	0.72 ± 0.01	16.0 ± 0.2

fidelity and extrapolation stability. Omitting transient spectral regularization ($\mathcal{R}_{\text{spec}}$) triggers the most severe performance drop (2-Moons-C error increases to 6.7%), directly validating Proposition 2 by confirming that penalizing transient amplitudes $\|a_k\|_2^2$ is essential for filtering domain-specific noise. Disabling stability regularization ($\mathcal{R}_{\text{stab}}$) impairs accuracy on 2-Moons-C (1.3% vs. 2.5%), supporting Theorem 2 and Corollary 1 by verifying that penalizing positive growth rates ($\sigma_k > 0$) prevents dominant extrapolation modes from diverging exponentially. Finally, removing either $\mathcal{L}_{\text{rec}}^{(C)}$ or $\mathcal{L}_{\text{fit}}^{(C)}$ degrades performance across all datasets, confirming their complementary roles in preserving auto-encoding fidelity and continuous modal trajectory alignment.

D. Qualitative Spectral Analysis

To provide qualitative intuition for our continuous spectral mechanism, Figure 4 visualizes the learned modal decomposition on 2-Moons-C across three complementary perspectives. Notably, all panels are generated from a single trained model to ensure consistency.

1) *Parameter-space trajectory decomposition*: Figure 4(a) illustrates how raw parameter evolution is cleanly disentangled into distinct spectral branches. The smooth dominant component tracks the persistent, recurring backbone of the parameter trajectory, whereas the transient component isolates localized, short-lived residual fluctuations. The high reconstruction fidelity confirms that our continuous modal parameterization faithfully preserves predictable long-term dynamics without overfitting domain-specific noise.

2) *Spectral distribution and adaptive gating*: Figure 4(b) plots the learned modal coordinates in the frequency–decay plane (f_k, d_k), where $f_k = |\omega_k|$ and $d_k = -\sigma_k$. Bypassing rigid DFT grids and manual frequency cutoffs, the learnable soft gates carve out a smooth, data-driven dominance boundary in the continuous spectral space. Persistent modes with low frequency and minimal decay ($\sigma_k \approx 0$) receive high dominance weights ($w_k^{\text{dom}} \approx 1$), whereas high-frequency or rapidly decaying modes are automatically routed to the transient branch.

3) *Extrapolated space-time trajectory*: Figure 4(c) projects continuous parameter evolution into a 3D space-time representation, highlighting FreKoo++’s exceptional long-horizon extrapolation capability governed by Eqs. (24)–(25). Far beyond the historical training window ($t_s \gg t_T$), the extrapolated dominant segment seamlessly extends the smooth cyclical trajectory into distant future domains without trajectory collapse or numerical divergence. Meanwhile, the transient residual is held constant at t_T as stationary local context. This visually validates both our long-term extrapolation stability and asymmetric inference protocol, as propagating persistent macro-dynamics secures reliable far-future forecasting while strictly preventing volatile local noise from corrupting long-term domain generalization.

E. Qualitative Analysis of Decision Boundaries

To evaluate how continuous parameter trajectory evolution translates into task-level generalization, Figure 5 visualizes the learned decision boundaries on 2-Moons-C across six

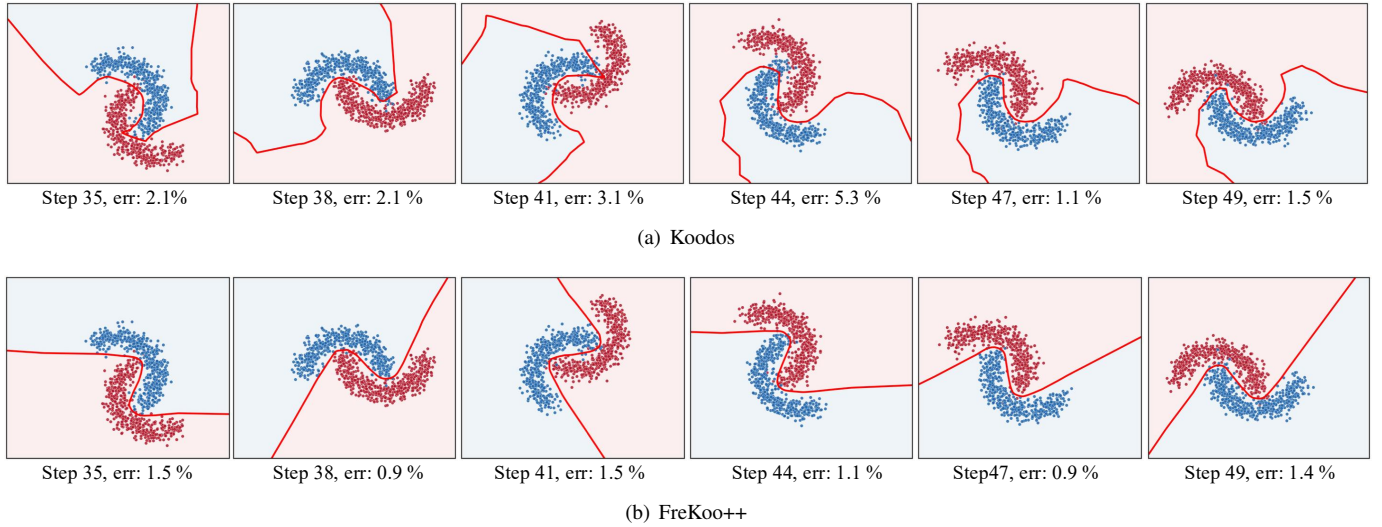


Fig. 5: Qualitative comparison of decision boundaries learned on 2-Moons-C across six unseen future domains. Red curves denote learned decision boundaries alongside the corresponding classification error rates.

unseen future temporal domains (steps 35 to 49). We compare FreKoo++ against the strongest continuous-time baseline Koodos [15]. As shown in Figure 5(a), Koodos produces convoluted and irregular decision boundaries. As time extends further beyond the historical training window (e.g., at step 44), Koodos suffers from severe boundary degradation, causing the classification error to surge to 5.3%. This instability stems from unconstrained continuous trajectory modeling, which overfits transient local noise and amplifies integration errors over longer extrapolation horizons. In contrast, FreKoo++ (Figure 5(b)) yields remarkably smooth, globally consistent, and physically intuitive decision boundaries that naturally rotate in synchronization with the underlying moon-cluster dynamics. Crucially, as the prediction horizon extends into distant future domains (steps 44 and 49), FreKoo++ maintains exceptional accuracy without experiencing performance degradation. These results are consistent with the stability predicted by Theorem 2 and Corollary 1, demonstrating that enforcing dominant mode stability ($\sigma_k^{\text{dom}} \leq 0$) effectively prevents boundary distortion and guarantees generalization robustness over continuous time. Additional decision boundary visualizations are provided in Supplementary IV-A.

VIII. CONCLUSION

In this paper, we presented FreKoo++, a continuous-time spectral-dynamical framework for temporal domain generalization. Motivated by the spectral-dynamical relation of continuous Koopman dynamics, FreKoo++ represents the evolution of task parameters through learnable modal components whose eigenvalues jointly encode oscillatory frequency and temporal growth or decay. This formulation removes the dependence on uniformly sampled temporal domains and fixed discrete transitions, allowing parameter evolution to be learned from irregular observations and evaluated at arbitrary future timestamps. In addition, FreKoo++ further introduces an adaptive soft spectral weighting mechanism that separates persistent dominant dynamics from transient domain-specific fluctuations,

together with spectral and stability regularization terms that suppress noisy modal components and control unstable long-horizon extrapolation. Theoretical analysis establishes modal approximation and continuous time generalization bounds, clarifying how the learned spectrum and the proposed regularizers affect future-domain risk. Experiments on both discrete and continuous TDG benchmarks demonstrate that FreKoo++ consistently improves temporal generalization. These results suggest that continuous spectral-dynamical modeling provides a principled and effective direction for learning robust models in evolving environments.

ACKNOWLEDGMENTS

The work was supported by the Australian Research Council (ARC) under Laureate project FL190100149 and Discovery Project DP220102635.

REFERENCES

- [1] E. Han, C. Huang, and K. Wang, “Model assessment and selection under temporal distribution shift,” in *International Conference on Machine Learning*. PMLR, 2024, pp. 17 374–17 392.
- [2] Z. Cao, Q. Xu, Z. Yang, X. Cao, and Q. Huang, “Cakge: Context-aware adaptive learning for dynamic knowledge graph embeddings,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 6, pp. 6225–6240, 2026.
- [3] E. Yu, J. Lu, K. Wang, X. Yang, and G. Zhang, “Drift-aware collaborative assistance mixture of experts for heterogeneous multistream learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 19, 2026, pp. 16 199–16 207.
- [4] R. Wang, Y. Dong, S. Arik, and R. Yu, “Koopman neural forecaster for time series with temporal distribution shifts,” in *International Conference on Learning Representations (ICLR)*, 2023.
- [5] X. Yang, E. Yu, W. Duan, and J. Lu, “Turning drift into constraint: Robust reasoning alignment in non-stationary multi-stream environments,” in *Forty-third International Conference on Machine Learning*, 2026.
- [6] A. Giarretta, M. Bisiacco, and G. Pillonetto, “Supervised learning in dynamic and non stationary environments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 10, pp. 9345–9351, 2025.
- [7] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, “Learning under concept drift: A review,” *IEEE transactions on knowledge and data engineering*, vol. 31, no. 12, pp. 2346–2363, 2018.

- [8] K. Xu, L. Chen, and S. Wang, "Coral: Concept drift representation learning for co-evolving time-series," *arXiv preprint arXiv:2501.01480*, 2025.
- [9] J. Wang, Z. Ren, R. Zhan, and Z. Zhou, "Distributionally robust policy learning under concept drifts," in *International Conference on Machine Learning*. PMLR, 2025, pp. 64 244–64 281.
- [10] A. Nasery, S. Thakur, V. Piratla, A. De, and S. Sarawagi, "Training for the future: A simple gradient interpolation loss to generalize along time," *Advances in Neural Information Processing Systems*, vol. 34, pp. 19 198–19 209, 2021.
- [11] Q. Zeng, W. Wang, F. Zhou, C. Ling, and B. Wang, "Foresee what you will learn: data augmentation for domain generalization in non-stationary environment," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 147–11 155.
- [12] G. Bai, C. Ling, and L. Zhao, "Temporal domain generalization with drift-aware dynamic neural networks," in *The Eleventh International Conference on Learning Representations*, 2023.
- [13] C.-Y. Chang, Y.-N. Chuang, Z. Jiang, K.-H. Lai, A. Jiang, and N. Zou, "Coda: Temporal domain generalization via concept drift simulator," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, 2025, pp. 131–142.
- [14] X. Yang, J. Lu, and E. Yu, "Walking the tightrope: Autonomous disentangling beneficial and detrimental drifts in non-stationary custom-tuning," *Advances in neural information processing systems*, vol. 38, pp. 116 167–116 193, 2026.
- [15] Z. Cai, G. Bai, R. Jiang, X. Song, and L. Zhao, "Continuous temporal domain generalization," *Advances in Neural Information Processing Systems*, vol. 37, pp. 127 987–128 014, 2024.
- [16] G. Ortiz-Jimenez, M. E. Gheche, E. Simou, H. P. Maretic, and P. Frossard, "Cdot: Continuous domain adaptation using optimal transport," *arXiv preprint arXiv:1909.11448*, 2019.
- [17] T. Qin, S. Wang, and H. Li, "Evolving domain generalization via latent structure-aware sequential autoencoder," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14 514–14 527, 2023.
- [18] E. Yu, J. Lu, B. Zhang, and G. Zhang, "Online boosting adaptive learning under concept drift for multistream classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 15, 2024, pp. 16 522–16 530.
- [19] B. Jiao, Y. Guo, D. Gong, and Q. Chen, "Dynamic ensemble selection for imbalanced data streams with concept drift," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 1, pp. 1278–1291, 2022.
- [20] S. Niu, J. Wu, Y. Zhang, Z. Wen, Y. Chen, P. Zhao, and M. Tan, "Towards stable test-time adaptation in dynamic wild world," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: <https://openreview.net/forum?id=g2YraF75Tj>
- [21] E. Yu, J. Lu, X. Yang, G. Zhang, and Z. Fang, "Learning robust spectral dynamics for temporal domain generalization," *Advances in Neural Information Processing Systems*, vol. 38, pp. 148 725–148 751, 2026.
- [22] Z. Liu, K. Zeng, Q. Ma, and J. T. Kwok, "Complematch: Boosting time-series semi-supervised classification with temporal-frequency complementarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 4, pp. 4440–4453, 2026.
- [23] W. Ye, S. Deng, Q. Zou, and N. Gui, "Frequency adaptive normalization for non-stationary time series forecasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 31 350–31 379, 2024.
- [24] J. Wang, M. Ye, Y. Kuang, R. Yang, W. Zhou, H. Li, and F. Wu, "Long-term feature extraction via frequency prediction for efficient reinforcement learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 3094–3110, 2025.
- [25] Y. Liu, C. Li, J. Wang, and M. Long, "Koopman: Learning non-stationary time series dynamics with koopman predictors," *Advances in neural information processing systems*, vol. 36, pp. 12 271–12 290, 2023.
- [26] Z. Liu, M. Cheng, Z. Li, Z. Huang, Q. Liu, Y. Xie, and E. Chen, "Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective," *Advances in Neural Information Processing Systems*, vol. 36, pp. 14 273–14 292, 2023.
- [27] Y. Jin, Z. Yang, X. Chu, and L. Ma, "Temporal domain generalization via learning instance-level evolving patterns," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024, pp. 4255–4263.
- [28] Q. Zeng, W. Wang, F. Zhou, G. Xu, R. Pu, C. Shui, C. Gagne, S. Yang, C. X. Ling, and B. Wang, "Generalizing across temporal domains with koopman operators," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 15, 2024, pp. 16 651–16 659.
- [29] D. Liang, J. Chen, X. Wang, Y. Wang, and S. Li, "Deepboots: Dual-stream residual boosting for drift-resilient time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 28, 2026, pp. 23 391–23 400.
- [30] T. Qin, S. Wang, and H. Li, "Generalizing to evolving domains with latent structure-aware sequential autoencoder," in *International Conference on Machine Learning*. PMLR, 2022, pp. 18 062–18 082.
- [31] H. Wang, H. He, and D. Katabi, "Continuously indexed domain adaptation," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 9898–9907.
- [32] Z. He, S. Li, W. Song, L. Yuan, J. Liang, H. Li, and K. Gai, "Learning time-aware causal representation for model generalization in evolving domains," in *Forty-second International Conference on Machine Learning*, 2025. [Online]. Available: <https://openreview.net/forum?id=KhCKypSaqx>
- [33] Z. Cai, Y. YAO, G. Bai, R. Jiang, X. Song, R. Shibasaki, and L. Zhao, "Continuous domain generalization," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
- [34] W. Luo, S. Chen, T. Zhou, and C. Gong, "A theoretical perspective on streaming noisy data with distribution shift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 4, pp. 4150–4166, 2026.
- [35] R. C. Hoover, J. James, P. May, and K. Caudle, "Koopman operator identification of model parameter trajectories for temporal domain generalization (komet)," *arXiv preprint arXiv:2603.26923*, 2026.
- [36] X. Li and L. Yao, "Distribution-conditioned adversarial variational autoencoder for valid instrumental variable generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 12, 2024, pp. 13 664–13 672.
- [37] P. Lu, J. Lu, A. Liu, and G. Zhang, "Early concept drift detection via prediction uncertainty," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 18, 2025, pp. 19 124–19 132.
- [38] X. Yang, J. Lu, and E. Yu, "Adapting multi-modal large language model to concept drift from pre-training onwards," in *The Thirteenth International Conference on Learning Representations*, 2025.
- [39] E. Yu, J. Lu, and G. Zhang, "Generalized incremental learning under concept drift across evolving data streams," in *Proceedings of the ACM Web Conference 2026*, 2026, pp. 3905–3916.
- [40] B. Peng, J. Lu, Y. Zhang, G. Zhang, and Z. Fang, "Distributional prototype learning for out-of-distribution detection," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 2025, pp. 1104–1114.
- [41] zhangkai wu, X. Fan, H. Wu, and L. Cao, "SCot: Unifying consistency models and rectified flows via straight-consistent trajectories," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. [Online]. Available: <https://openreview.net/forum?id=GV82iAD70j>
- [42] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International conference on machine learning*. PMLR, 2022, pp. 27 268–27 286.
- [43] K. Yi, J. Fei, Q. Zhang, H. He, S. Hao, D. Lian, and W. Fan, "Filtarnet: Harnessing frequency filters for time series forecasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 55 115–55 140, 2024.
- [44] H. Ye, J. Chen, S. Gong, F. Jiang, T. Zhang, J. Chen, and X. Gao, "Atfnet: Adaptive time-frequency ensembled network for long-term time series forecasting," *arXiv preprint arXiv:2404.05192*, 2024.
- [45] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 781–796, 2000.
- [46] N. Hurley-Walker, X. Zhang, A. Bahramian, S. McSweeney, T. O'Doherty, P. Hancock, J. Morgan, G. Anderson, G. Heald, and T. Galvin, "A radio transient with unusually slow periodic emission," *Nature*, vol. 601, no. 7894, pp. 526–530, 2022.
- [47] B. Wang, M. Tian, Y. Yu, Q. Dong, and D. Xu, "Enhanced adrc with quasi-resonant control for pmsm speed regulation considering aperiodic and periodic disturbances," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 3, pp. 3568–3577, 2021.
- [48] B. O. Koopman, "Hamiltonian systems and transformation in hilbert space," *Proceedings of the National Academy of Sciences*, vol. 17, no. 5, pp. 315–318, 1931.
- [49] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern koopman theory for dynamical systems," *arXiv preprint arXiv:2102.12086*, 2021.
- [50] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

- [51] D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, and A. Courville, “Out-of-distribution generalization via risk extrapolation (rex),” in *International conference on machine learning*. PMLR, 2021, pp. 5815–5826.
- [52] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, “Neural ordinary differential equations,” in *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [53] M. Xie, S. Li, L. Yuan, C. Liu, and Z. Dai, “Evolving standardization for continual domain generalization over temporal drift,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [54] L. Deng, “The MNIST database of handwritten digit images for machine learning research,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [55] L. Zhao, Q. Sun, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, “Feature constrained multi-task learning models for spatiotemporal event forecasting,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1059–1072, 2017.
- [56] G. Bai, J. Torres, J. Wang, L. Zhao, C. Abad, and C. Vaca, “Sign-regularized multi-task learning,” in *Proceedings of the 2023 SIAM International Conference on Data Mining*. SIAM, 2023, pp. 793–801.
- [57] H. Yao, C. Choi, B. Cao, Y. Lee, P. W. Koh, and C. Finn, “Wild-time: A benchmark of in-the-wild distribution shift over time,” in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 10 309–10 324.
- [58] B. Chen, B.-F. Chen, and H.-T. Lin, “Rotation-blended CNNs on a new open dataset for tropical cyclone image-to-intensity regression,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018, pp. 90–99.
- [59] M. Schirmer, M. Eltayeb, S. Lessmann, and M. Rudolph, “Modeling irregular time series with continuous recurrent units,” in *International Conference on Machine Learning*. PMLR, 2022, pp. 19 388–19 405.

Supplementary Material for FreKoo++: Learning Continuous Spectral Dynamics for Temporal Domain Generalization

SECTION I THEORETICAL ANALYSIS DETAILS

A. Proof of Theorem 1

We adopt two conventions throughout this supplementary section. **(i)** The omitted transient remainder in Assumption 1 is non-growing ($\text{Re}(\lambda_k) \leq 0$ for $k > K$); the learned transient-weighted dynamics in Theorem 2 are evaluated over the historical window $[0, t_T]$, with growth controlled by the window factor $G_{\text{trans}}(t_T)$, which is finite by Assumption 2. **(ii)** Since $\text{Re}(\cdot)$ is 1-Lipschitz ($\|\text{Re}(u) - \text{Re}(v)\|_2 \leq \|u - v\|_2$ for $u, v \in \mathbb{C}^m$), every bound derived for the complex expansion $\sum_k a_k e^{\lambda_k t}$ applies directly to the real-valued parameterization $\hat{z}(t) = \text{Re}(\sum_k a_k e^{\lambda_k t})$; we therefore suppress $\text{Re}(\cdot)$ in intermediate steps.

Restatement of Theorem 1 (Mode Approximation Bound). Under Assumption 1, let $\hat{z}(t) = \sum_{k=1}^K \hat{a}_k e^{\hat{\lambda}_k t}$. Then for any horizon $t_s > 0$:

$$\sup_{t \in [0, t_s]} \|z(t) - \hat{z}(t)\|_2 \leq \sum_{k=1}^K \|a_k - \hat{a}_k\|_2 e^{\hat{\rho}_k t_s} + C_a \sum_{k=1}^K |\lambda_k - \hat{\lambda}_k| t_s e^{\varrho_k t_s} + \sum_{k=K+1}^{K^*} \|a_k\|_2,$$

where $\hat{\rho}_k = \max(\text{Re}(\hat{\lambda}_k), 0)$, $\varrho_k = \max(\text{Re}(\lambda_k), \text{Re}(\hat{\lambda}_k), 0)$, and $C_a = \max_{1 \leq k \leq K} \|a_k\|_2$.

Proof 3: Adding and subtracting $\sum_{k=1}^K a_k e^{\hat{\lambda}_k t}$ decomposes the error as:

$$z(t) - \hat{z}(t) = \underbrace{\sum_{k=1}^K (a_k - \hat{a}_k) e^{\hat{\lambda}_k t}}_{E_{\text{amp}}(t)} + \underbrace{\sum_{k=1}^K a_k (e^{\lambda_k t} - e^{\hat{\lambda}_k t})}_{E_{\text{est}}(t)} + \underbrace{\sum_{k=K+1}^{K^*} a_k e^{\lambda_k t}}_{E_{\text{trunc}}(t)}. \quad (32)$$

The triangle inequality gives $\|z(t) - \hat{z}(t)\|_2 \leq \|E_{\text{amp}}(t)\|_2 + \|E_{\text{est}}(t)\|_2 + \|E_{\text{trunc}}(t)\|_2$.

Step 1: Amplitude estimation error. For $t \in [0, t_s]$:

$$\|E_{\text{amp}}(t)\|_2 \leq \sum_{k=1}^K \|a_k - \hat{a}_k\|_2 |e^{\hat{\lambda}_k t}| = \sum_{k=1}^K \|a_k - \hat{a}_k\|_2 e^{\text{Re}(\hat{\lambda}_k) t} \leq \sum_{k=1}^K \|a_k - \hat{a}_k\|_2 e^{\hat{\rho}_k t_s}, \quad (33)$$

where the last inequality uses $\hat{\rho}_k \geq \text{Re}(\hat{\lambda}_k)$ and $t \leq t_s$.

Step 2: Eigenvalue estimation error. Fix $t \in [0, t_s]$ and $k \in \{1, \dots, K\}$. By the fundamental theorem of calculus applied to $s \mapsto e^{(\hat{\lambda}_k + s(\lambda_k - \hat{\lambda}_k))t}$ on $[0, 1]$:

$$e^{\lambda_k t} - e^{\hat{\lambda}_k t} = (\lambda_k - \hat{\lambda}_k) t \int_0^1 e^{(\hat{\lambda}_k + s(\lambda_k - \hat{\lambda}_k))t} ds. \quad (34)$$

Since $|e^{\mu t}| = e^{\text{Re}(\mu)t}$ and $\text{Re}(\hat{\lambda}_k + s(\lambda_k - \hat{\lambda}_k))$ is linear in s , its maximum over $[0, 1]$ is attained at an endpoint:

$$|e^{\lambda_k t} - e^{\hat{\lambda}_k t}| \leq |\lambda_k - \hat{\lambda}_k| t e^{\max(\text{Re}(\lambda_k), \text{Re}(\hat{\lambda}_k))t} \leq |\lambda_k - \hat{\lambda}_k| t e^{\varrho_k t}, \quad (35)$$

where $\varrho_k = \max(\text{Re}(\lambda_k), \text{Re}(\hat{\lambda}_k), 0) \geq 0$. Because $\varrho_k \geq 0$, the map $t \mapsto t e^{\varrho_k t}$ is nondecreasing on $[0, t_s]$, so the supremum is attained at $t = t_s$. Combining with $C_a = \max_{k \leq K} \|a_k\|_2$:

$$\sup_{t \in [0, t_s]} \|E_{\text{est}}(t)\|_2 \leq \sum_{k=1}^K \|a_k\|_2 |e^{\lambda_k t} - e^{\hat{\lambda}_k t}| \leq C_a \sum_{k=1}^K |\lambda_k - \hat{\lambda}_k| t_s e^{\varrho_k t_s}. \quad (36)$$

Step 3: Modal truncation error. By Assumption 1, $\text{Re}(\lambda_k) \leq 0$ for $k > K$, so $|e^{\lambda_k t}| = e^{\text{Re}(\lambda_k)t} \leq 1$ for all $t \geq 0$. Hence:

$$\sup_{t \in [0, t_s]} \|E_{\text{trunc}}(t)\|_2 \leq \sum_{k=K+1}^{K^*} \|a_k\|_2 |e^{\lambda_k t}| \leq \sum_{k=K+1}^{K^*} \|a_k\|_2. \quad (37)$$

Summing Eqs. (33), (36), and (37) completes the proof.

B. Proof of Theorem 2

Theorem 3 (Restatement of Theorem 2): Under Assumptions 1–2, the future-domain excess risk at $t_s > t_T$ satisfies:

$$\mathcal{E}_{t_s} \leq L_\ell L_g L_{\text{dec}} (E_{\text{dom}}(t_s) + E_{\text{trans}}(t_s)), \quad (38)$$

where $E_{\text{dom}}(t_s) = \|\hat{z}_{\text{dom}}(t_s) - z_{\text{dom}}^*(t_s)\|_2$ and $E_{\text{trans}}(t_s) = \|\hat{z}_{\text{trans}}(t_T) - z_{\text{trans}}^*(t_s)\|_2$. With $\Delta_s = t_s - t_T$, $\hat{\rho}_k = \max(\text{Re}(\hat{\lambda}_k), 0)$, $\varrho_k = \max(\text{Re}(\lambda_k^*), \text{Re}(\hat{\lambda}_k), 0)$:

$$E_{\text{dom}}(t_s) \leq \sum_{k=1}^K \|\hat{b}_k - b_k^*\|_2 e^{\hat{\rho}_k \Delta_s} + B_{\text{dom}} \sum_{k=1}^K |\hat{\lambda}_k - \lambda_k^*| \Delta_s e^{\varrho_k \Delta_s}, \quad (39)$$

and the transient holding error satisfies:

$$E_{\text{trans}}(t_s) \leq B_{\text{trans}} + G_{\text{trans}}(t_T) \sqrt{\mathcal{R}_{\text{spec}}}, \quad (40)$$

where $G_{\text{trans}}(t_T) = (\sum_{k=1}^K w_k^{\text{trans}} e^{2\hat{\rho}_k t_T})^{1/2}$.

Proof 4: By Assumption 2, $\theta^*(t_s) = \phi^{-1}(z^*(t_s))$ and $\hat{\theta}(t_s) = \phi^{-1}(\hat{z}(t_s))$.

Step 1: Excess risk to latent error. By L_ℓ -Lipschitz continuity of ℓ and L_g -Lipschitz continuity of $g(X; \cdot)$:

$$\mathcal{E}_{t_s} \leq L_\ell L_g \|\hat{\theta}(t_s) - \theta^*(t_s)\|_2. \quad (41)$$

By L_{dec} -Lipschitz continuity of ϕ^{-1} :

$$\|\hat{\theta}(t_s) - \theta^*(t_s)\|_2 = \|\phi^{-1}(\hat{z}(t_s)) - \phi^{-1}(z^*(t_s))\|_2 \leq L_{\text{dec}} \|\hat{z}(t_s) - z^*(t_s)\|_2. \quad (42)$$

Combining Eqs. (41)–(42):

$$\mathcal{E}_{t_s} \leq L_\ell L_g L_{\text{dec}} \|\hat{z}(t_s) - z^*(t_s)\|_2. \quad (43)$$

Step 2: Latent error decomposition. By the inference rule $\hat{z}(t_s) = \hat{z}_{\text{dom}}(t_s) + \hat{z}_{\text{trans}}(t_T)$ (Eq. (24)) and the decomposition $z^*(t_s) = z_{\text{dom}}^*(t_s) + z_{\text{trans}}^*(t_s)$:

$$\|\hat{z}(t_s) - z^*(t_s)\|_2 \leq \underbrace{\|\hat{z}_{\text{dom}}(t_s) - z_{\text{dom}}^*(t_s)\|_2}_{E_{\text{dom}}(t_s)} + \underbrace{\|\hat{z}_{\text{trans}}(t_T) - z_{\text{trans}}^*(t_s)\|_2}_{E_{\text{trans}}(t_s)}. \quad (44)$$

Substituting into Eq. (43) proves Eq. (38).

Step 3: Dominant extrapolation error. Setting $\tau = \Delta_s$ in the anchored representations of Assumption 2:

$$z_{\text{dom}}^*(t_s) = \text{Re}\left(\sum_{k=1}^K b_k^* e^{\lambda_k^* \Delta_s}\right), \quad \hat{z}_{\text{dom}}(t_s) = \text{Re}\left(\sum_{k=1}^K \hat{b}_k e^{\hat{\lambda}_k \Delta_s}\right). \quad (45)$$

By the non-expansive property of $\text{Re}(\cdot)$:

$$E_{\text{dom}}(t_s) \leq \left\| \sum_{k=1}^K (\hat{b}_k e^{\hat{\lambda}_k \Delta_s} - b_k^* e^{\lambda_k^* \Delta_s}) \right\|_2. \quad (46)$$

Adding and subtracting $b_k^* e^{\hat{\lambda}_k \Delta_s}$:

$$\hat{b}_k e^{\hat{\lambda}_k \Delta_s} - b_k^* e^{\lambda_k^* \Delta_s} = (\hat{b}_k - b_k^*) e^{\hat{\lambda}_k \Delta_s} + b_k^* (e^{\hat{\lambda}_k \Delta_s} - e^{\lambda_k^* \Delta_s}). \quad (47)$$

The triangle inequality yields:

$$E_{\text{dom}}(t_s) \leq \sum_{k=1}^K \|\hat{b}_k - b_k^*\|_2 |e^{\hat{\lambda}_k \Delta_s}| + \sum_{k=1}^K \|b_k^*\|_2 |e^{\hat{\lambda}_k \Delta_s} - e^{\lambda_k^* \Delta_s}|. \quad (48)$$

For the first sum, $|e^{\hat{\lambda}_k \Delta_s}| = e^{\text{Re}(\hat{\lambda}_k) \Delta_s} \leq e^{\hat{\rho}_k \Delta_s}$. For the second sum, applying the integral identity to $u \mapsto e^{[\lambda_k^* + u(\hat{\lambda}_k - \lambda_k^*)] \Delta_s}$ on $[0, 1]$:

$$e^{\hat{\lambda}_k \Delta_s} - e^{\lambda_k^* \Delta_s} = (\hat{\lambda}_k - \lambda_k^*) \Delta_s \int_0^1 e^{[\lambda_k^* + u(\hat{\lambda}_k - \lambda_k^*)] \Delta_s} du. \quad (49)$$

Since $\text{Re}(\lambda_k^* + u(\hat{\lambda}_k - \lambda_k^*))$ is linear in u , its maximum over $[0, 1]$ is attained at an endpoint. Therefore:

$$|e^{\hat{\lambda}_k \Delta_s} - e^{\lambda_k^* \Delta_s}| \leq |\hat{\lambda}_k - \lambda_k^*| \Delta_s e^{\max(\text{Re}(\lambda_k^*), \text{Re}(\hat{\lambda}_k)) \Delta_s} \leq |\hat{\lambda}_k - \lambda_k^*| \Delta_s e^{\varrho_k \Delta_s}. \quad (50)$$

Substituting $\max_k \|b_k^*\|_2 \leq B_{\text{dom}}$ and Eq. (50) into Eq. (48) establishes Eq. (39).

Step 4: Transient holding error. The triangle inequality gives:

$$E_{\text{trans}}(t_s) \leq \|\hat{z}_{\text{trans}}(t_T)\|_2 + \|z_{\text{trans}}^*(t_s)\|_2. \quad (51)$$

By Assumption 2, $\|z_{\text{trans}}^*(t_s)\|_2 \leq B_{\text{trans}}$.

For the learned transient contribution, the non-expansive property and the triangle inequality give:

$$\|\hat{z}_{\text{trans}}(t_T)\|_2 = \left\| \text{Re} \left(\sum_{k=1}^K w_k^{\text{trans}} \hat{a}_k e^{\hat{\lambda}_k t_T} \right) \right\|_2 \leq \sum_{k=1}^K w_k^{\text{trans}} \|\hat{a}_k\|_2 e^{\hat{\rho}_k t_T}, \quad (52)$$

where $|e^{\hat{\lambda}_k t_T}| = e^{\text{Re}(\hat{\lambda}_k) t_T} \leq e^{\hat{\rho}_k t_T}$. Writing $w_k^{\text{trans}} = \sqrt{w_k^{\text{trans}}} \cdot \sqrt{w_k^{\text{trans}}}$ and applying Cauchy–Schwarz:

$$\|\hat{z}_{\text{trans}}(t_T)\|_2 \leq \left(\sum_{k=1}^K w_k^{\text{trans}} e^{2\hat{\rho}_k t_T} \right)^{1/2} \left(\sum_{k=1}^K w_k^{\text{trans}} \|\hat{a}_k\|_2^2 \right)^{1/2}. \quad (53)$$

The first factor is $G_{\text{trans}}(t_T)$, which is finite by Assumption 2. For the second factor, since $1 - \text{Var}_{k=1}^K(w_k^{\text{dom}}) \geq 0$, the definition $\mathcal{R}_{\text{spec}} = \sum_k w_k^{\text{trans}} \|\hat{a}_k\|_2^2 + (1 - \text{Var}(w_k^{\text{dom}}))$ (Eq. (21)) implies:

$$\sum_{k=1}^K w_k^{\text{trans}} \|\hat{a}_k\|_2^2 \leq \mathcal{R}_{\text{spec}}. \quad (54)$$

Substituting into Eq. (53):

$$\|\hat{z}_{\text{trans}}(t_T)\|_2 \leq G_{\text{trans}}(t_T) \sqrt{\mathcal{R}_{\text{spec}}}. \quad (55)$$

Combining Eqs. (51) and (55) proves Eq. (40). Steps 1–4 together complete the proof.

C. Proof of Proposition 2

Proof 5: Define the cumulative temporal energy of mode k over $[0, T_{\text{max}}]$:

$$E_k = \int_0^{T_{\text{max}}} \|a_k e^{\lambda_k t}\|_2^2 dt. \quad (56)$$

Since $\|\text{Re}(u)\|_2 \leq \|u\|_2$, this upper-bounds the energy of the real-valued trajectory $\text{Re}(a_k e^{\lambda_k t})$.

Given $\lambda_k = \sigma_k + j\omega_k$ with $\sigma_k < 0$:

$$E_k = \|a_k\|_2^2 \int_0^{T_{\text{max}}} |e^{(\sigma_k + j\omega_k)t}|^2 dt = \|a_k\|_2^2 \int_0^{T_{\text{max}}} e^{2\sigma_k t} dt = \|a_k\|_2^2 \cdot \frac{1 - e^{2\sigma_k T_{\text{max}}}}{-2\sigma_k}. \quad (57)$$

Because $\sigma_k < 0$ and $T_{\text{max}} > 0$, we have $0 < e^{2\sigma_k T_{\text{max}}} < 1$, so $1 - e^{2\sigma_k T_{\text{max}}} \leq 1$. Therefore:

$$E_k \leq \frac{\|a_k\|_2^2}{-2\sigma_k} = \frac{\|a_k\|_2^2}{2|\sigma_k|}, \quad (58)$$

which establishes Eq. (22).

SECTION II EXPERIMENTAL DETAILS

A. Datasets

1) *CTDG Benchmarks:* We evaluate FreKoo++ on six continuous temporal domain generalization (CTDG) benchmarks characterized by asynchronous arrivals and irregularly sampled temporal gaps. For all CTDG datasets, timestamps are normalized to $[0, 1]$ while preserving their relative temporal intervals. Following the standard CTDG setup, the chronologically earliest 70% of domains serve as historical source training data, while the remaining latest 30% are reserved as unseen target domains for future extrapolation. The ‘-C’ suffix denotes the continuous, irregularly sampled variant to distinguish it from its discrete counterpart.

Rotated 2-Moons-C. This synthetic benchmark constructs a continuous variant of the two-moons problem to model rotational concept drift. It contains 1,000 two-dimensional samples per domain across two classes formed by upper and lower moon-shaped clusters. We sample 50 timestamps from $[0, 50]$ to construct 50 sequential domains. Domains 1–35 (35 source domains) are used for training, while domains 36–50 (15 target domains) serve as unseen test sets. Concept drift is induced by continuously rotating the dataset counter-clockwise at a constant rate of 18° per unit time.

Rotated MNIST-C. We extend the MNIST dataset [54] to continuous temporal shifts under irregular sampling. Fifty timestamps are sampled from $[0, 50]$, where each timestamp defines a domain containing 1,000 randomly selected digit images, yielding a total of 50 temporal domains. The first 35 rotated domains serve as training source data, while the remaining 15

domains are reserved for testing. Analogous to Rotated 2-Moons-C, progressive counter-clockwise rotation at 18° per unit time induces continuous visual domain drift.

Twitter. This benchmark evaluates influenza-risk prediction using streaming social media signals [55], [56]. Domains are constructed from tweet streams collected at 50 arbitrary starting timestamps across the 2010–2014 flu seasons, with each domain spanning a 7-day temporal window. We partition the data into 50 chronological domains, using the first 35 as training sources and the final 15 for testing. Features consist of disease-related term frequency statistics validated against CDC Influenza-Like Illness reports, where concept drift reflects both seasonal flu cycles and evolving user tweeting patterns.

Yearbook. Derived from the Wild-Time benchmark [57], this dataset contains frontal yearbook portraits collected from 1930 to 2013 across 128 high schools for gender classification. We sample 40 non-consecutive years across the 84-year span to construct 40 irregularly spaced domains, training on the first 28 source domains and evaluating on the final 12 future target domains. Long-term visual drift naturally emerges from non-uniform temporal intervals, reflecting shifts in fashion trends, photographic styles, and social context.

Cyclone. This benchmark addresses a regression task mapping satellite imagery of tropical cyclones to wind intensity [58]. Each cyclone event represents a distinct domain indexed by its occurrence date. Following Koodos [15], we extract West Pacific cyclone records from 2014 to 2016 to form 72 continuous temporal domains. We use the first 50 domains as training sources and the remaining 22 domains for future wind intensity prediction. Concept drift naturally arises from seasonal meteorological variations and event-driven irregular arrivals.

House-C. This real-estate regression benchmark predicts housing prices from property features based on transaction records from 2013 to 2019. Following the CTDG protocol, we sample 40 non-overlapping one-month temporal windows to create 40 chronological domains, using the first 28 domains for training and the remaining 12 domains for testing. Concept drift stems naturally from evolving local real-estate market conditions and non-uniform sampling intervals.

2) *DTDG benchmarks:* To verify backward compatibility under discrete settings, we evaluate FreKoo++ across seven temporal benchmarks following the experimental setup and evaluation protocol established in FreKoo [21]. Following standard DTDG practice, historical chronological domains serve as source training data, while the immediately succeeding unseen future domain is reserved as the target domain. This benchmark suite encompasses both synthetic and complex real-world temporal data streams with diverse drift patterns.

Rotated 2-Moons-D. This benchmark adapts the 2-Moons dataset to model concept drift via rotation. It contains 1,800 2-dimensional samples across two classes, divided into 10 sequential domains. Each domain is rotated 18° counter-clockwise relative to the previous one. We train on domains 0-8 and test on domain 9, where the drift is caused by the incremental rotation.

Rotated MNIST-D. We randomly sampled 1000 instances from the MNIST dataset and constructed a total of five domains by successively rotating them counter-clockwise by 15° , analogous to the Rotated 2-Moons-D setup. The first four rotated domains are used for training, while the fifth domain serves as the test set, creating incremental drift induced by progressive rotation transformations.

Online News Popularity (ONP)¹. This dataset aggregates heterogeneous features of articles published by Mashable over two years, aiming to predict social media shares (popularity). It comprises 39,797 samples with 58 features, where concept drift is characterized by temporal shifts in popularity patterns. We partition the data into 6 time-ordered domains, using the first five for training and the last for testing. The dataset undergoes slight real-world concept drift over the observed time period.

Shuttle². The Shuttle dataset contains 58,000 instances of multi-class flight status classification under severe class imbalance. It is partitioned into 8 time-stamped domains using a chronological split: domains spanning timestamps 30-70 serve as training data, while the most recent period (70-80) is reserved for testing. The dataset also has real-world concept drifts over the observed time period.

Electrical Demand³. This dataset records electricity demand in a province, addressing a binary classification task to predict whether 30-minute demand exceeds or falls below the daily average for that time period. After removing instances with missing values, it contains 28,222 samples with 8 features. It is partitioned into 30 two-week chronological domains, with the first 29 used for training and the 30th for testing. Seasonal variations in demand induce concept drift, making this a real-world benchmark capturing both periodic and incremental drift patterns.

HousePrices-D⁴. This dataset comprises housing price records from 2013 to 2019 for the regression task to predict property prices based on feature values. We treat each calendar year as a distinct domain, using 2013–2018 data for training and the final (2019) domain for testing. Concept drift emerges naturally from temporal economic shifts and market fluctuations over the years.

Appliances Energy Prediction⁵. This dataset addresses regression modeling for predicting appliance energy consumption in a low-energy building. Comprising 10-minute sensor readings over 4.5 months in 2016, it is partitioned into 9 chronological

¹<https://archive.ics.uci.edu/dataset/332/online+news+popularity>

²<https://archive.ics.uci.edu/dataset/148/statlog+shuttle>

³https://web.archive.org/web/20191121102533/http://www.inescporto.pt/jgama/ales/ales_5.html

⁴<https://www.kaggle.com/datasets/htagholdings/property-sales>

⁵<https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction>

TABLE IV: Detailed network architectures for the task predictive model and the uniform latent autoencoder (ϕ, ϕ^{-1}) across all benchmark datasets. All autoencoder hidden layers employ ReLU activations.

Dataset	Predictive model ($\theta \in \mathbb{R}^D$)	Latent encoder-decoder (ϕ, ϕ^{-1})
2-Moons-C	MLP: [2, 50, 50, 1] + ReLU + Sigmoid	ϕ : $[D \rightarrow 1024 \rightarrow 512 \rightarrow 128 \rightarrow m]$ ϕ^{-1} : $[m \rightarrow 128 \rightarrow 512 \rightarrow 1024 \rightarrow D]$
Rot-MNIST-C	$3 \times [\text{Conv}(32/32/64, k=3) + \text{ReLU} + \text{MaxPool}(2)] + \text{MLP}: [576, 128, 10] + \text{LogSoftmax}$	
Twitter	Embedding Network + MLP [32, 128, 32, 1] + ReLU + Sigmoid	
Yearbook	$3 \times [\text{Conv}(32/32/64, k=3) + \text{ReLU} + \text{MaxPool}(2)] + \text{MLP}: [1024, 128, 32, 1] + \text{ReLU} + \text{Sigmoid}$	
Cyclone	$4 \times [\text{Conv}(32/32/64/64, k=3) + \text{ReLU} + \text{MaxPool}(2)] + \text{MLP}: [1024, 128, 32, 1] + \text{ReLU}$	
House	MLP: [30, 400, 400, 1] + ReLU	

domains. We train on the first eight domains and evaluate on the final (most recent) ninth domain, with concept drift arising from temporal shifts in energy usage patterns across the observation period.

B. Baselines

1) *CTDG Baselines*: Following the continuous temporal domain generalization protocol in Koodos [15], we compare FreKoo++ against twelve representative CTDG baselines evaluated in Table I, categorized into four technical paradigms based on how they handle temporal information:

Time-agnostic baselines. These methods ignore explicit temporal ordering across domains. *Offline* aggregates all source domains into a single pool for empirical risk minimization. *LastDomain* trains exclusively on the chronologically latest source domain, adopting a naive recency-based heuristic. *IncFinetune* sequentially fine-tunes the network across chronological domains using a reduced learning rate.

Domain generalization and continuous adaptation baselines. *IRM* [50] and *V-REx* [51] penalize invariant predictor mismatches and risk variations across source domains, respectively. *CIDA* [31] conditions the model on continuous temporal indices, aligning representations via adversarial domain alignment with a domain-aware discriminator.

Temporal and parameter forecasting baselines. *TKNets* [28] leverages Koopman operators for discrete temporal transfer across evolving domains. *DRAIN* [12] forecasts future neural network parameters via a discrete recurrent network trained on source parameter trajectories. To accommodate irregular sampling intervals, *DRAIN- Δt* [15] replaces standard discrete recurrent units with continuous-time recurrent units [59].

Continuous-time dynamics baselines. *DeepODE* [15] maps model parameters into a latent space and evolves them via a Neural ODE solver [52]. *NeuralLio* [33] incorporates continuous-time Liouville operator dynamics to track evolving continuous distributions over time. *Koodos* [15] models the coupled evolution of data and parameter states using a continuous-time Koopman formulation under structural constraints.

To ensure a fair and rigorous comparison, all baseline methods share identical task backbone architectures with FreKoo++ across each benchmark dataset. Experimental results are either directly cited from their original publications under matched evaluation protocols or reproduced using their official open-source implementations under unified experimental settings. To ensure statistical reliability, all reported metrics represent the mean and standard deviation (mean $\pm$ std) computed over five independent runs with different random seeds.

2) *DTDG baselines*: For the DTDG compatibility evaluation, we follow the FreKoo protocol [21]. The comparison includes time-agnostic baselines *Offline*, *LastDomain*, and *IncFinetune*; continuous domain adaptation baselines *CDOT* [16] and *CIDA* [31]; discrete TDG methods *GI* [10], *LSSAE* [30], *Foresee* [11], and *DRAIN* [12]; and spectral-Koopman baselines *TKNets* [28], *Koodos* [15], and *FreKoo* [21].

SECTION III IMPLEMENTATION DETAILS

A. Network Architectures

To ensure rigorous comparability with state-of-the-art baselines, FreKoo++ integrates three primary modules: a dataset-specific task predictive model ($\theta \in \mathbb{R}^D$), a symmetric four-layer latent autoencoder (ϕ, ϕ^{-1}) , and the continuous Koopman spectral-dynamical module operating in $\mathbb{R}^m$. Detailed network architectures for the task backbones and the uniform autoencoder setup across all benchmark datasets are summarized in Table IV.

B. Training Details and Hyperparameters

We optimize FreKoo++ end-to-end using the Adam optimizer across all benchmark datasets. Optimization is parameterized with three distinct learning rate groups: η_1 for the predictive task model, η_2 for the latent encoder and decoder, and η_3 for the continuous Koopman spectral module. The overall joint loss objective is regulated by four hyperparameter weights: α for

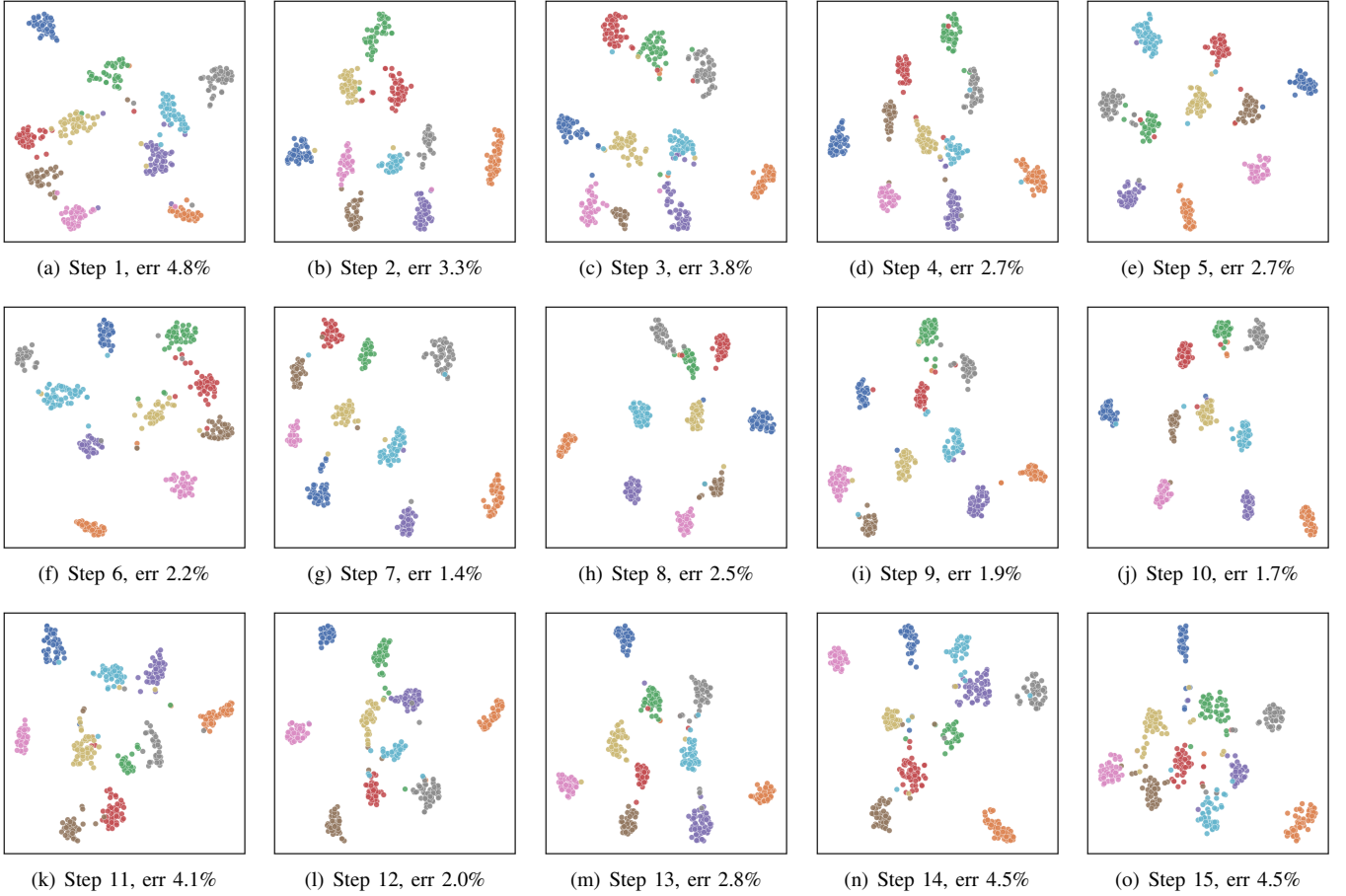


Fig. 6: t-SNE visualization on all 15 future domains of Rot-MNIST-C (colored by class). Each subfigure is one future domain with its step index and classification error.

parameter reconstruction ($\mathcal{L}_{\text{rec}}^{(C)}$), β for modal trajectory fitting ($\mathcal{L}_{\text{fit}}^{(C)}$), γ for transient spectral regularization ($\mathcal{R}_{\text{spec}}$), and δ for stability regularization ($\mathcal{R}_{\text{stab}}$).

For the **2-Moons-C** dataset, the model is trained for 300 epochs with learning rates $\eta_1 = 1 \times 10^{-2}$ and $\eta_2 = \eta_3 = 1 \times 10^{-3}$, regulated by loss weights $\alpha = 100$, $\beta = 1$, $\gamma = 1$, and $\delta = 10$. The **Rot-MNIST-C** configuration employs 400 training epochs with a uniform learning rate $\eta_1 = \eta_2 = \eta_3 = 1 \times 10^{-3}$ across all modules, governed by $\alpha = 100$, $\beta = 10$, $\gamma = 1$, and $\delta = 10$. For **Twitter**, training spans 200 epochs using $\eta_1 = 1 \times 10^{-2}$ for the task predictor and $\eta_2 = \eta_3 = 1 \times 10^{-3}$ for the autoencoder and spectral dynamics modules, combined with loss weights $\alpha = 10$, $\beta = 10$, $\gamma = 1$, and $\delta = 10$. The **Yearbook** dataset is optimized over 800 epochs with a uniform learning rate of 1×10^{-3} across all components, accompanied by loss coefficients $\alpha = 100$, $\beta = 10$, $\gamma = 1$, and $\delta = 10$. For **Cyclone**, we train the model for 500 epochs using a uniform learning rate $\eta_1 = \eta_2 = \eta_3 = 1 \times 10^{-3}$, with objective weights set to $\alpha = 100$, $\beta = 100$, $\gamma = 100$, and $\delta = 10$. Finally, the **House** dataset is trained over 600 epochs maintaining a uniform learning rate of 1×10^{-3} across all modules, regulated by loss coefficients $\alpha = 10$, $\beta = 10$, $\gamma = 1$, and $\delta = 10$.

SECTION IV SUPPLEMENTARY EXPERIMENTS

A. Qualitative Analysis of Extrapolated Test-Domain Representations

To inspect the geometric evolution of feature representations synthesized via continuous parameter extrapolation, we visualize the domain-adapted penultimate features across all unseen future target domains ($t_s > t_T$) using t-SNE. The resulting manifolds are colored by semantic class and annotated with per-domain error rates in Fig. 6 (Rot-MNIST-C) and Fig. 7 (Yearbook). Two key geometric properties validate the core design of FreKoo++:

a) Sustained long-horizon class separability: Feature discriminability remains well-preserved across the entire extrapolation window. Across all 15 future steps of Rot-MNIST-C and 12 future steps of Yearbook, the feature representations consistently form tight, well-isolated clusters corresponding to the respective semantic categories. This confirms that the target model configurations $\hat{\theta}(t_s)$ generated via closed-form Koopman modal superposition synthesize valid, discriminative decision

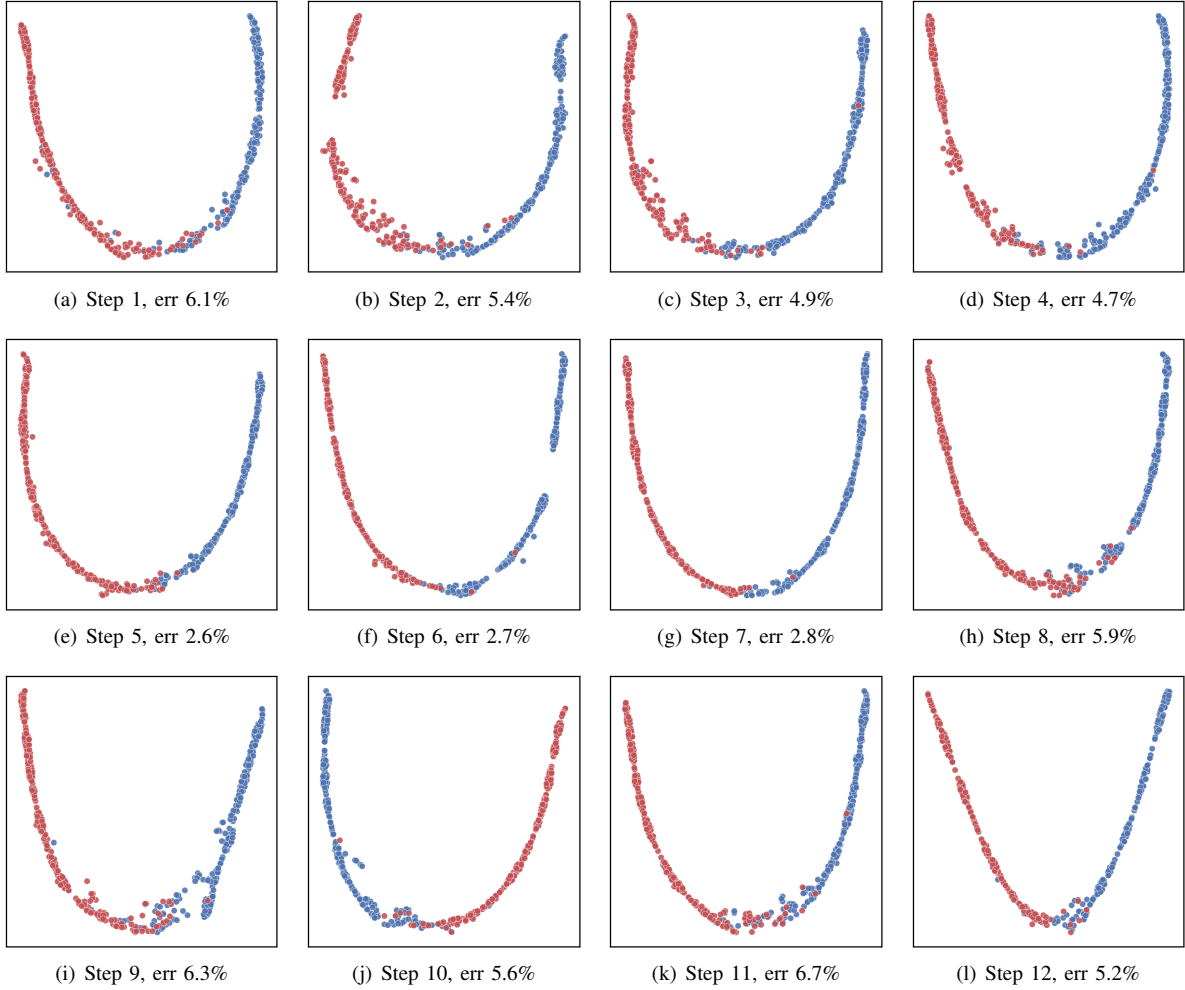


Fig. 7: t-SNE visualization on all 12 future domains of Yearbook (colored by class). Each subfigure is one future domain with its step index and classification error.

boundaries on the underlying model manifold, demonstrating that FreKoo++ captures true continuous distributional drift rather than overfitting to source-domain boundary states.

b) Horizon-invariant stability and error boundedness: The feature geometry exhibits stability over extended temporal horizons. The per-domain classification errors remain consistently low and tightly bounded (1.4%–4.8% on Rot-MNIST-C, 2.6%–6.7% on Yearbook), with no evidence of representation collapse or dispersion at distant time steps. This empirical trend provides direct qualitative evidence for Theorem 2 and Corollary 1: by analytically extrapolating non-growing dominant spectral dynamics while locking volatile transient components at t_T , FreKoo++ effectively eliminates the compounding numerical integration errors and boundary distortions that typically plague discrete or ODE-based autoregressive baselines.

Together, these visualizations demonstrate that FreKoo++ preserves a temporally coherent yet class-discriminative feature topology across arbitrary prediction horizons, providing qualitative corroboration for the quantitative superiority.

B. Hyperparameter Sensitivity Analysis

To evaluate the stability of FreKoo++ with respect to objective loss weights, we analyze performance sensitivity over a wide grid search range $\{0.01, 0.1, 1.0, 10, 100\}$ for α ($\mathcal{L}_{\text{rec}}^{(C)}$), β ($\mathcal{L}_{\text{fit}}^{(C)}$), γ ($\mathcal{R}_{\text{spec}}$), and δ ($\mathcal{R}_{\text{stab}}$) across six CTDG benchmark datasets. As illustrated in Fig. 8, our framework demonstrates remarkable robustness across several orders of magnitude. Across most real-world datasets (e.g., Cyclone, House, Twitter, Yearbook, and 2-Moons-C), the error curves remain remarkably flat across the entire hyperparameter spectrum, indicating that FreKoo++ does not require delicate parameter tuning to achieve superior generalization. Optimal performance is generally attained when loss weights are selected within the moderate range $[0.1, 10]$. Specifically, setting β ($\mathcal{L}_{\text{fit}}^{(C)}$) or α ($\mathcal{L}_{\text{rec}}^{(C)}$) excessively large (100) leads to an error increase on Rot-MNIST-C, as over-constraining parameter trajectory alignment forces the model to fit transient noise. Conversely, moderate values of stability (δ) and spectral regularization (γ) effectively constrain positive modal growth rates and filter transient noise without sacrificing

model capacity. Overall, these results confirm that FreKoo++ is numerically stable and consistently effective under standard loss weight configurations.

Fig. 9 illustrates the impact of modal dimension $K \in \{2, \dots, 64\}$ across all six CTDG datasets. When $K \leq 4$, insufficient spectral capacity causes severe underfitting and high variance (e.g., on 2-Moons-C and Rot-MNIST-C). Expanding K to the range of 16–32 consistently yields the best generalization, with $K = 32$ achieving peak AUC on Twitter and lowest MAE on House-C and 2-Moons-C. Beyond this point ($K = 64$), performance on noisy real-world datasets (e.g., House-C and Twitter) slightly deteriorates due to the over-parameterization of redundant modes. Thus, $K = 32$ serves as the global sweet spot balancing spectral expressiveness and noise suppression. For architectural regularity and implementation simplicity, we standardize $K = m = 32$ across all experiments.

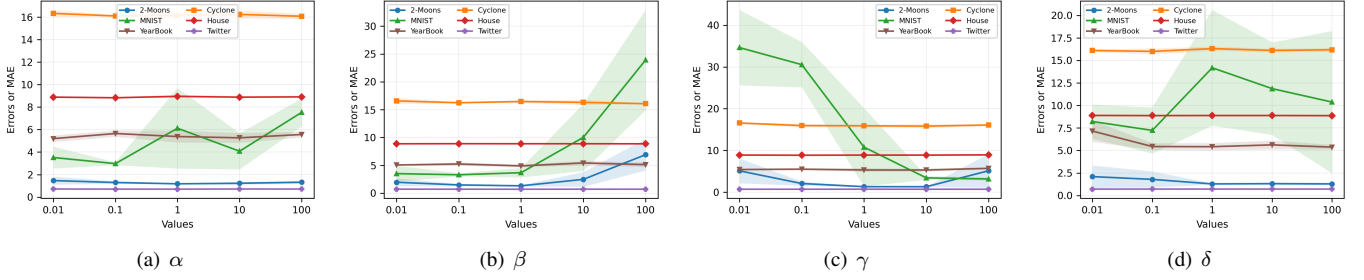


Fig. 8: Hyperparameter sensitivity analysis of FreKoo++ with respect to objective loss weights α ($\mathcal{L}_{\text{rec}}^{(C)}$), β ($\mathcal{L}_{\text{fit}}^{(C)}$), γ ($\mathcal{R}_{\text{spec}}$), and δ ($\mathcal{R}_{\text{stab}}$) swept over $\{0.01, 0.1, 1.0, 10, 100\}$ across six benchmark datasets.

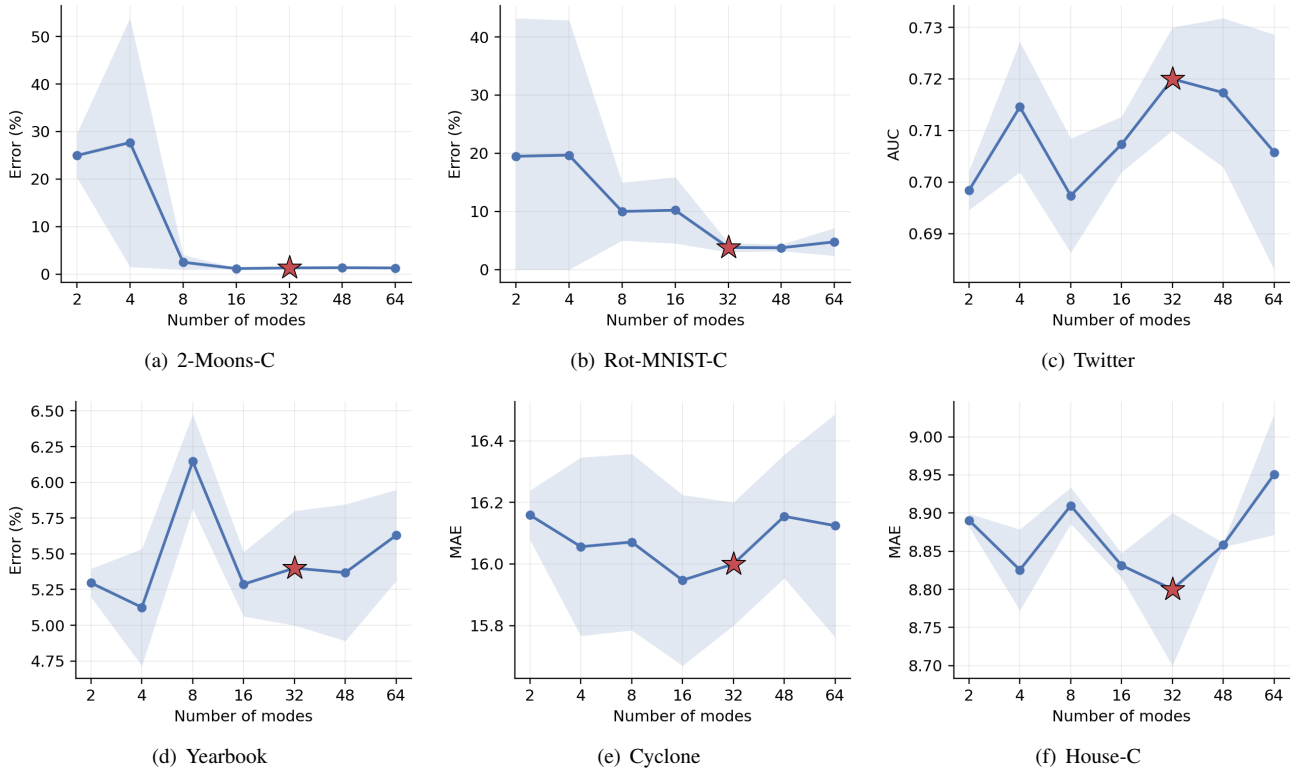


Fig. 9: Sensitivity of the number of Koopman modes K . The red star marks the adopted $K = 32$.