GRAPH-SPECTRAL FLOW MATCHING FOR MULTIVARIATE TIME SERIES ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time series anomaly detection typically relies on evaluating discrepancies between observations and outputs produced by models trained on normal data. An alternative perspective is to characterize the distribution of normal data through the generative dynamics, *i.e.*, the velocity field, of flow matching models. However, standard flow matching typically adopts linear probability paths that overlook dependencies among variables, leading to a misalignment with the structured data distribution. To address this issue, we propose GRASP, a flow matching framework with a graph-spectral path for multivariate time series anomaly detection. GRASP incorporates graph structure into the probability path by minimizing a fixed-endpoint action that combines kinetic energy with graph Dirichlet energy. This formulation yields a closed-form path based on graph-frequency-dependent hyperbolic interpolation. A velocity predictor trained on normal data then detects anomalies using weighted velocity discrepancies aggregated across source samples, flow times, and graph frequencies. Theoretically, we establish that GRASP is invariant to the choice of Laplacian eigenbasis and decompose its expected oracle anomaly score into bounded endpoint uncertainty and graph-frequency-weighted Fisher discrepancy. Experiments on four benchmarks demonstrate the superior anomaly detection performance of GRASP and validate the effectiveness of its graph-spectral path and weighting mechanism.

1 INTRODUCTION

Industrial systems are increasingly monitored by sensor networks that generate large volumes of multivariate time series (MTS) data. Detecting anomalies in these data is critical for preventing failures, reducing downtime, and ensuring reliable operation (Wu et al., 2021; Deng & Hooi, 2021; Fink et al., 2026). However, fault labels are typically scarce and incomplete, as many fault types occur rarely or are not observed during data collection, motivating unsupervised approaches to MTS anomaly detection (Zhang et al., 2019; Audibert et al., 2020; Belay et al., 2023). Existing methods typically learn normal patterns and identify anomalies through prediction errors, reconstruction errors, or representation discrepancies (Jin et al., 2024; Chen & Eldardiry, 2024; Ho et al., 2025). These approaches are effective when abnormal behavior produces clear discrepancies between observations and model outputs. However, in practice, anomalous patterns may remain partially predictable or reconstructible, resulting in weak anomaly signals (Liu et al., 2025; Cho et al., 2025; Zhang et al., 2026b). This motivates anomaly detection methods that go beyond measuring only endpoint-based prediction or reconstruction discrepancies.

Flow matching offers an alternative paradigm by characterizing observations through generative dynamics rather than solely through endpoint outputs, which has been shown to be effective in image anomaly detection (Chen et al., 2026). During training time, a velocity predictor learns to match conditional target velocities along prescribed probability paths connecting source samples to normal observations. During inference time, anomalies are detected by measuring discrepancies between the predicted velocities and conditional target velocities along the entire probability path. The probability path determines both the states at which an observation is evaluated and the conditional target velocity against which it is compared (Lipman et al., 2023; Liu et al., 2023; Albergo et al., 2024). Therefore, the design of the probability path is crucial for flow-matching-based anomaly detection.

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

MTS variables often exhibit structured dependencies that can be represented by a sensor graph, providing an informative inductive bias for modeling cross-variable interactions (Bronstein et al., 2021; Jin et al., 2024). The graph Laplacian decomposes multivariate signals into graph-frequency modes that describe different patterns of variation across connected nodes (Shuman et al., 2013). Recent work further shows that anomalous behavior can induce heterogeneous energy shifts across graph frequencies, suggesting that different frequency modes carry distinct anomaly-relevant information (Liu et al., 2026). However, existing flow matching models typically adopt a standard linear conditional path, which interpolates all directions identically without accounting for the underlying graph structure (Kollovieh et al., 2025; Albergo et al., 2024). For structured data, this structural mismatch may produce intermediate states and target velocities that are inconsistent with the topology and geometry of the monitored system (Rozada et al., 2026; Fang et al., 2026; Zhang et al., 2026a). This limitation is more pronounced for MTS anomaly detection because the anomaly score depends not only on the observed endpoint, but also on the velocity discrepancies evaluated along the entire path. These observations motivate the design of a graph-informed probability path for flow matching.

In this work, we introduce flow matching with a GRAph-SPectral Path (GRASP), which explicitly incorporates structural information into the conditional probability path. When graph edges encode similarity between normal sensor signals, graph Dirichlet energy provides a natural measure for regularizing variation across connected nodes. We therefore formulate path construction as a fixed-endpoint variational problem that balances kinetic energy and graph Dirichlet energy. The resulting closed-form solution assigns a graph-frequency-dependent hyperbolic interpolation schedule to each frequency mode. Specifically, the zero-frequency modes recover standard linear interpolation, whereas higher-frequency modes undergo progressively stronger contraction at intermediate flow times. In this way, GRASP preserves the prescribed source and data endpoints while imposing a graph-dependent smoothness prior along the probability path. The velocity predictor trained on normal data then detects anomalies by measuring velocity discrepancies along this structured path.

The resulting path induces different conditional uncertainty of the intermediate state and velocity-residual scales across flow times and graph frequencies. To account for these differences, GRASP uses flow-time- and graph-frequency-dependent weights when aggregating velocity discrepancies into an anomaly score. Theoretically, we establish that the node-domain path, target velocity, and resulting anomaly score are invariant to the choice of orthonormal basis within repeated Laplacian eigenspaces. Under the population-optimal velocity predictor, we further show that the expected anomaly score decomposes into an endpoint-uncertainty term and a graph-frequency-weighted Fisher discrepancy. Experiments on four MTS benchmarks demonstrate superior anomaly detection performance of GRASP and the effectiveness of the graph-spectral path and the weighting schedule.

Our contributions are summarized below:

- We derive a closed-form graph-spectral probability path from a fixed-endpoint variational problem combining kinetic energy and graph Dirichlet energy, yielding frequency-dependent hyperbolic interpolation. Based on the graph-spectral path, we introduce an anomaly score that aggregates weighted velocity discrepancies across source samples, flow times, and graph frequencies.
- We establish that GRASP is invariant to different Laplacian eigenbases, and decompose the expected oracle anomaly score into bounded endpoint uncertainty and weighted Fisher discrepancy.
- We conduct experiments on four MTS benchmarks, demonstrating that GRASP achieves competitive or superior performance compared with the baselines. The ablation and sensitivity studies validate the effectiveness of the graph-spectral path design and anomaly scoring mechanism.

Further discussion of related work is provided in Appendix A.

2 PRELIMINARIES

Notations and Problem Definition. We use calligraphic letters, such as $\mathcal{X}$, to represent sets, uppercase bold letters, such as $\mathbf{X}$, to represent matrices, lowercase bold letters, such as $\mathbf{x}$, to represent vectors, and lowercase letters, such as x, to represent scalars. A complete summary of the notation is provided in Appendix B. We consider an MTS observation over N variables. A time-series window is represented as $\mathbf{X} \in \mathbb{R}^{N\times R}$, where its r-th column $\mathbf{x}_r \in \mathbb{R}^N$ contains the observations of all N variables at timestep r, and R denotes the window length. The relationships among the variables are captured by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ and $\mathcal{E}$ denote the sets of nodes and edges,

respectively. We denote by $\mathbf{L} \in \mathbb{R}^{N\times N}$ the symmetric normalized graph Laplacian matrix. Since $\mathbf{L}$ is real, symmetric, and positive semidefinite, it admits an eigendecomposition $\mathbf{L} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^\top$, where $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N]$ is an orthogonal matrix whose columns are the Laplacian eigenvectors and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ contains the corresponding eigenvalues ordered as $0 \le \lambda_1 \le \cdots \le \lambda_N \le 2$.

MTS Anomaly Detection. During training, we observe only normal time-series windows sampled from an unknown normal data distribution p. At test time, observations are drawn from a potentially anomalous distribution q. Let p_t and q_t denote the corresponding path marginals at flow time t. Our goal is to learn the generative dynamics of normal data without anomaly samples and construct an anomaly score that quantifies the deviation of a test observation from the learned normal dynamics.

Conditional Flow Matching. Flow matching learns a time-dependent velocity field that transports $\mathbf{X}_0 \sim p_0$ to $\mathbf{X}_1 \sim p_1$ along a prescribed probability path (Lipman et al., 2023; Albergo et al., 2024). In this paper, the entries of $\mathbf{X}_0$ are sampled independently from a standard Gaussian distribution, while p_1 corresponds to the normal data distribution p. The conditional path between an endpoint pair $(\mathbf{X}_0, \mathbf{X}_1)$ is specified by an interpolation map and the corresponding conditional target velocity:

$$\mathbf{X}_t = \phi_t(\mathbf{X}_0, \mathbf{X}_1), \qquad \mathbf{U}_t = \frac{\partial}{\partial t}\phi_t(\mathbf{X}_0, \mathbf{X}_1), \qquad t \in [0, 1], \tag{1}$$

which satisfies the boundary conditions $\phi_0(\mathbf{X}_0, \mathbf{X}_1) = \mathbf{X}_0$ and $\phi_1(\mathbf{X}_0, \mathbf{X}_1) = \mathbf{X}_1$. Conditional flow matching trains a parameterized velocity field $\mathbf{V}_t(\mathbf{X}; \boldsymbol{\theta})$ to approximate the conditional target velocity by minimizing

$$\mathcal{L}_{\mathrm{CFM}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],\mathbf{X}_0,\mathbf{X}_1}\left[\|\mathbf{V}_t(\mathbf{X}_t; \boldsymbol{\theta}) - \mathbf{U}_t\|_F^2\right], \tag{2}$$

where $\mathcal{U}[0, 1]$ is the uniform distribution between 0 and 1. A commonly used probability path is the linear interpolation, where the intermediate state and the conditional target velocity are defined as

$$\mathbf{X}_t = \phi_t(\mathbf{X}_0, \mathbf{X}_1) = (1 - t)\mathbf{X}_0 + t\mathbf{X}_1 \quad \text{and} \quad \mathbf{U}_t = \mathbf{X}_1 - \mathbf{X}_0. \tag{3}$$

This path construction applies the same linear interpolation schedule to all directions in the ambient data space and therefore does not explicitly account for structural dependencies among variables.

Graph Fourier Transform. The eigendecomposition of the graph aplacian matrix provides a graph Fourier basis for signals defined on the corresponding graph $\mathcal{G}$. Given an MTS graph signal $\mathbf{X} \in \mathbb{R}^{N\times R}$, its graph Fourier transform and the inverse graph Fourier transform are defined as follows:

$$\hat{\mathbf{X}} = \mathbf{\Psi}^\top\mathbf{X} \quad \text{and} \quad \mathbf{X} = \mathbf{\Psi}\hat{\mathbf{X}}. \tag{4}$$

The k-th row $\hat{\mathbf{x}}_k \in \mathbb{R}^R$ of $\hat{\mathbf{X}}$ contains the coefficients of the k-th graph-frequency mode across the R timesteps. The Laplacian eigenvalue λ_k characterizes the graph-frequency associated with eigenvector $\boldsymbol{\psi}_k$. Modes associated with smaller eigenvalues vary more smoothly over the graph, whereas modes associated with larger eigenvalues exhibit stronger variation across connected nodes. The graph Dirichlet energy admits a spectral decomposition that separates graph frequency modes:

$$\mathrm{Tr}\left(\mathbf{X}^\top\mathbf{L}\mathbf{X}\right) = \mathrm{Tr}\left(\hat{\mathbf{X}}^\top\mathbf{\Lambda}\hat{\mathbf{X}}\right) = \sum_{k=1}^{N}\lambda_k\|\hat{\mathbf{x}}_k\|_2^2. \tag{5}$$

This decomposition shows that the graph Dirichlet energy penalizes high-frequency components more strongly, thereby providing a natural measure of signal variation across connected nodes.

3 Flow Matching with Graph-Spectral Path

The standard linear path used in flow matching applies the same interpolation schedule to all directions, implicitly imposing an isotropic transport geometry. Ignoring the structural information in MTS data may produce intermediate states and target velocities that are poorly aligned with the underlying graph geometry. In this section, we develop GRASP, which explicitly incorporates structural information into probability-path design through a graph-informed variational formulation. A schematic comparison between GRASP and standard linear-path flow matching is given in Figure 1.

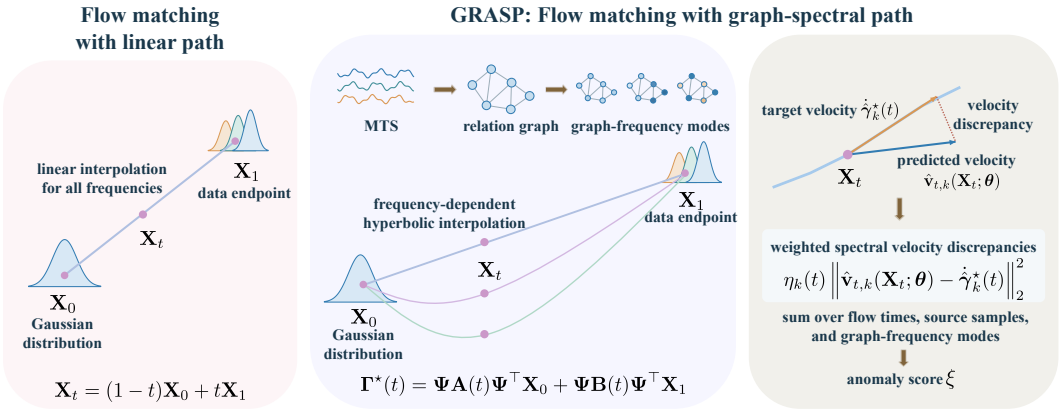


Figure 1: Overview of GRASP for MTS anomaly detection. Left: flow matching with the same linear interpolation for all graph-frequency modes. Middle: GRASP with the graph-spectral path that applies frequency-dependent hyperbolic interpolation. Right: anomaly scoring strategy of GRASP.

3.1 Graph-Spectral Path Design

The linear path widely adopted by flow matching models can be characterized as the minimizer of a variational problem with the kinetic energy Lagrangian function defined as follows (Du et al., 2026):

$$\mathbf{\Gamma}^\star = \underset{\mathbf{\Gamma}(0)=\mathbf{X}_0,\mathbf{\Gamma}(1)=\mathbf{X}_1}{\arg\min} \int_0^1 \mathcal{L}_{\text{lin}}(\mathbf{\Gamma}, \dot{\mathbf{\Gamma}}, t)dt \quad \text{with} \quad \mathcal{L}_{\text{lin}}(\mathbf{\Gamma}, \dot{\mathbf{\Gamma}}, t) = \frac{1}{2}\|\dot{\mathbf{\Gamma}}(t)\|_F^2, \tag{6}$$

where $\mathbf{\Gamma} : [0,1] \to \mathbb{R}^{N\times R}$ is the path function. Because this objective penalizes only kinetic energy, it treats all directions in the ambient data space isotropically. Although such a path is suitable for Euclidean data (Liu et al., 2023; Lipman et al., 2023; Albergo et al., 2024), it may produce intermediate states that are poorly aligned with the structure of data supported on non-Euclidean domains (Fang et al., 2026; Wyrwal et al., 2026). In MTS anomaly detection, since velocity discrepancies are evaluated throughout the entire path, such structural misalignment can lead to less informative anomaly signals. Under the assumption that time-varying graph signals vary smoothly over the relation graph, connected nodes tend to exhibit coherent behavior under normal operation (Kalofolias, 2016; Dong et al., 2016; Giraldo et al., 2022). We therefore propose to construct a graph-informed probability path by considering a Lagrangian that combines kinetic energy with graph Dirichlet energy:

$$\mathcal{L}_{\text{graph}}(\mathbf{\Gamma}, \dot{\mathbf{\Gamma}}, t) = \frac{1}{2}\|\dot{\mathbf{\Gamma}}(t)\|^2 + \frac{\tau}{2}\text{Tr}\left(\mathbf{\Gamma}(t)^\top \mathbf{L}\mathbf{\Gamma}(t)\right), \tag{7}$$

where the graph Dirichlet energy weight parameter $\tau \geq 0$ controls the strength of the graph regularization. The graph Dirichlet energy term penalizes abrupt variations across connected nodes in the graph at intermediate states. As a result, higher graph-frequency modes undergo stronger contraction while preserving the endpoints. The resulting frequency-dependent interpolation therefore treats different graph-frequency modes differently, respecting the phenomenon observed in Liu et al. (2026) that anomalies induce heterogeneous behavior across different graph frequency modes.

To solve the resulting graph-informed variational problem, we first transform the path to the graph Fourier domain as $\hat{\mathbf{\Gamma}}(t) = \mathbf{\Psi}^\top\mathbf{\Gamma}(t)$. Since $\mathbf{\Psi}$ is orthonormal and independent of t, we have $\|\dot{\mathbf{\Gamma}}(t)\|_F^2 = \|\dot{\hat{\mathbf{\Gamma}}}(t)\|_F^2$. The graph-informed variational problem can therefore be written as follows:

$$\hat{\mathbf{\Gamma}}^\star = \underset{\hat{\mathbf{\Gamma}}(0)=\hat{\mathbf{X}}_0,\hat{\mathbf{\Gamma}}(1)=\hat{\mathbf{X}}_1}{\arg\min} \int_0^1 \frac{1}{2}\sum_k \|\dot{\hat{\gamma}}_k(t)\|_2^2 + \frac{\tau}{2}\sum_k \lambda_k\|\hat{\gamma}_k(t)\|_2^2 dt, \tag{8}$$

where $\hat{\gamma}_k(t)$ denote the k-th row of $\hat{\mathbf{\Gamma}}(t)$. Since both the objective and the boundary conditions are separable across different k, *i.e.*, different graph frequencies, the problem in Equation (8) decomposes into N independent vector-valued variational problems:

$$\hat{\gamma}_k^\star = \underset{\hat{\gamma}_k(0)=\hat{\mathbf{x}}_{0,k},\hat{\gamma}_k(1)=\hat{\mathbf{x}}_{1,k}}{\arg\min} \int_0^1 \frac{1}{2}\|\dot{\hat{\gamma}}_k(t)\|_2^2 + \frac{\tau}{2}\lambda_k\|\hat{\gamma}_k(t)\|_2^2 dt, \tag{9}$$

where $\hat{\mathbf{x}}_{t,k}$ represents the k-th row of $\hat{\mathbf{X}}_t$. Let $\omega_k = \sqrt{\tau\lambda_k}$, we have the following result.

Theorem 1 (Graph-spectral path)**.** *The variational problem in Equation (9) admits a unique global minimizer given by*

$$\hat{\boldsymbol{\gamma}}_k^\star(t) = \alpha_k(t)\hat{\mathbf{x}}_{0,k} + \beta_k(t)\hat{\mathbf{x}}_{1,k}, \tag{10}$$

where

$$\alpha_k(t) = \frac{\sinh(\omega_k(1-t))}{\sinh(\omega_k)} \quad \textit{and} \quad \beta_k(t) = \frac{\sinh(\omega_k t)}{\sinh(\omega_k)}. \tag{11}$$

The proof of Theorem 1 is provided in Appendix D.1. Figure 2 visualizes the source coefficient $\alpha_k(t)$ and target coefficient $\beta_k(t)$ for different values of ω_k. The source coefficient $\alpha_k(t)$ decreases from one to zero, whereas the target coefficient $\beta_k(t)$ increases from zero to one. Unlike the standard linear path, the graph-spectral path assigns different interpolation schedules to different graph frequencies. Specifically, larger values of ω_k, corresponding to higher graph frequencies or stronger graph regularization, produce stronger contraction at intermediate flow times. The resulting path therefore introduces a graph-dependent structural inductive bias into flow matching. Based on Theorem 1, we obtain the corresponding graph-spectral velocity field:

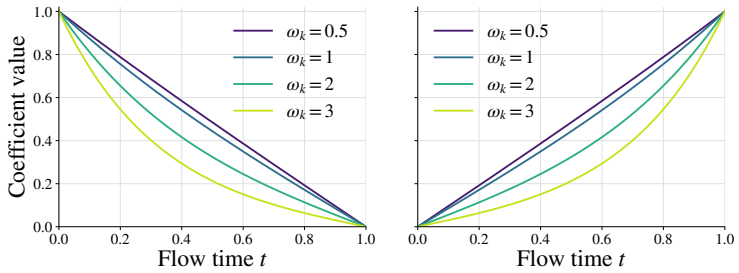


Figure 2: Coefficients $\alpha_k(t)$ (left) and $\beta_k(t)$ (right) with different ω_k.

$$\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) = \dot{\alpha}_k(t)\hat{\mathbf{x}}_{0,k} + \dot{\beta}_k(t)\hat{\mathbf{x}}_{1,k}, \tag{12}$$

where

$$\dot{\alpha}_k(t) = -\frac{\omega_k \cosh(\omega_k(1-t))}{\sinh(\omega_k)} \quad \text{and} \quad \dot{\beta}_k(t) = \frac{\omega_k \cosh(\omega_k t)}{\sinh(\omega_k)}. \tag{13}$$

Remark 1. *When $\omega_k = 0$, $\alpha_k(t)$ and $\beta_k(t)$ are defined by their continuous limit, i.e., $\alpha_k(t)$ becomes $1-t$ and $\beta_k(t)$ becomes t. The corresponding graph-frequency mode therefore degenerates to the standard linear path. Since the normalized graph Laplacian always has at least one zero eigenvalue with its multiplicity equal to the number of connected components (Chung, 1997; Shuman et al., 2013), the case for $\lambda_k = 0$, i.e., $\omega_k = 0$, always occurs for at least one graph-frequency mode.*

The graph-spectral path obtained in Theorem 1 can be written in matrix form as

$$\hat{\boldsymbol{\Gamma}}^\star(t) = \mathbf{A}(t)\hat{\boldsymbol{\Gamma}}(0) + \mathbf{B}(t)\hat{\boldsymbol{\Gamma}}(1) \tag{14}$$

with

$$\mathbf{A}(t) = \operatorname{diag}\left(\alpha_1(t), \ldots, \alpha_N(t)\right) \quad \text{and} \quad \mathbf{B}(t) = \operatorname{diag}\left(\beta_1(t), \ldots, \beta_N(t)\right). \tag{15}$$

Performing the inverse graph Fourier transform gives the corresponding node-domain path:

$$\boldsymbol{\Gamma}^\star(t) = \boldsymbol{\Psi}\hat{\boldsymbol{\Gamma}}^\star(t) = \boldsymbol{\Psi}\mathbf{A}(t)\boldsymbol{\Psi}^\top\boldsymbol{\Gamma}(0) + \boldsymbol{\Psi}\mathbf{B}(t)\boldsymbol{\Psi}^\top\boldsymbol{\Gamma}(1) = \boldsymbol{\Psi}\mathbf{A}(t)\boldsymbol{\Psi}^\top\mathbf{X}_0 + \boldsymbol{\Psi}\mathbf{B}(t)\boldsymbol{\Psi}^\top\mathbf{X}_1. \tag{16}$$

Then the corresponding node-domain velocity field is defined by

$$\dot{\boldsymbol{\Gamma}}^\star(t) = \boldsymbol{\Psi}\dot{\mathbf{A}}(t)\boldsymbol{\Psi}^\top\mathbf{X}_0 + \boldsymbol{\Psi}\dot{\mathbf{B}}(t)\boldsymbol{\Psi}^\top\mathbf{X}_1, \tag{17}$$

where

$$\dot{\mathbf{A}}(t) = \operatorname{diag}\left(\dot{\alpha}_1(t), \ldots, \dot{\alpha}_N(t)\right) \quad \text{and} \quad \dot{\mathbf{B}}(t) = \operatorname{diag}\left(\dot{\beta}_1(t), \ldots, \dot{\beta}_N(t)\right). \tag{18}$$

3.2 Model Training and Anomaly Scoring

We parameterize the velocity field associated with the graph-spectral path using a lightweight MLP architecture, similar to TSMixer as introduced in (Chen et al., 2023). The details of the model are provided in Appendix C. We train the velocity model in node domain with the regression objective:

$$\mathcal{L}_{\mathsf{GRASP}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],\mathbf{X}_0\sim p_0,\mathbf{X}_1\sim p}\left[\left\|\mathbf{V}_t(\mathbf{X}_t;\boldsymbol{\theta}) - \dot{\boldsymbol{\Gamma}}^\star(t)\right\|^2\right]. \tag{19}$$

Suppose that the normal data follows the distribution p, we define the oracle marginal velocity in the node and graph Fourier domains as

$$\mathbf{V}_t^p(\mathbf{X}) = \mathbb{E}[\dot{\mathbf{\Gamma}}^\star(t) \mid \mathbf{X}_t = \mathbf{X}] \quad \text{and} \quad \hat{\mathbf{V}}_t^p(\mathbf{X}) = \mathbb{E}[\dot{\hat{\mathbf{\Gamma}}}^\star(t) \mid \mathbf{X}_t = \mathbf{X}]. \tag{20}$$

Similarly, the posterior means of the data endpoint in the node and graph Fourier domains are

$$\mathbf{M}_t^p(\mathbf{X}) = \mathbb{E}[\mathbf{\Gamma}^\star(1) \mid \mathbf{X}_t = \mathbf{X}] \quad \text{and} \quad \hat{\mathbf{M}}_t^p(\mathbf{X}) = \mathbb{E}[\hat{\mathbf{\Gamma}}^\star(1) \mid \mathbf{X}_t = \mathbf{X}]. \tag{21}$$

Let $\hat{\mathbf{v}}_{t,k}^p$ and $\hat{\boldsymbol{\mu}}_{t,k}^p$ denote the k-th row of $\hat{\mathbf{V}}_t^p(\mathbf{X})$ and $\hat{\mathbf{M}}_t^p(\mathbf{X})$. Then we have the following result.

Lemma 1 (Graph-spectral residual scaling)**.** *For every $t \in (0,1)$ and graph-frequency mode k, the conditional velocity residual and the endpoint posterior residual in the graph Fourier domain satisfy*

$$\dot{\hat{\gamma}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) = \rho_k(t)\left(\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X})\right) \quad \textit{with} \quad \rho_k(t) = \frac{\omega_k}{\sinh(\omega_k(1-t))}. \tag{22}$$

The proof of Lemma 1 is provided in Appendix D.2. Note that when $\omega_k = 0$, $\rho_k(t)$ is defined by its continuous limit $\frac{1}{1-t}$. Lemma 1 shows that the velocity residual rescales the endpoint residual by a factor that varies across both flow times and graph frequencies.

The conditional velocity residual in the graph Fourier domain can be written in matrix form as

$$\dot{\hat{\mathbf{\Gamma}}}^\star(t) - \hat{\mathbf{V}}_t^p(\mathbf{X}) = \mathbf{P}(t)\left(\hat{\mathbf{X}}_1 - \hat{\mathbf{M}}_t^p(\mathbf{X})\right) \quad \text{with} \quad \mathbf{P}(t) = \operatorname{diag}\left(\rho_1(t), \ldots, \rho_N(t)\right). \tag{23}$$

Performing the inverse graph Fourier transform gives the node-domain conditional velocity residual:

$$\dot{\mathbf{\Gamma}}^\star(t) - \mathbf{V}_t^p(\mathbf{X}) = \mathbf{\Psi}\mathbf{P}(t)\mathbf{\Psi}^\top\left(\mathbf{X}_1 - \mathbf{M}_t^p(\mathbf{X})\right). \tag{24}$$

According to Theorem 1, we can see that the amount of endpoint information contained in an intermediate state also varies across flow times and graph frequencies. Consider a rescaled observation:

$$\frac{1}{\beta_k(t)}\hat{\gamma}_k^\star(t) = \frac{\alpha_k(t)}{\beta_k(t)}\hat{\mathbf{x}}_{0,k} + \hat{\mathbf{x}}_{1,k}. \tag{25}$$

Since the prior follows a standard Gaussian distribution, conditioning on the endpoint gives

$$\frac{1}{\beta_k(t)}\hat{\gamma}_k^\star(t) \mid \hat{\mathbf{x}}_{1,k} \sim \mathcal{N}\left(\hat{\mathbf{x}}_{1,k}, \frac{\alpha_k^2(t)}{\beta_k^2(t)}\mathbf{I}\right). \tag{26}$$

The conditional variance in Equation (26) quantifies the effective noise level of the rescaled intermediate state with respect to the endpoint. To account for both this noise level and the residual scaling elaborated in Lemma 1, we define the anomaly score weight as the inverse effective noise variance multiplied by the inverse squared residual-scaling factor as follows:

$$\eta_k(t) = \frac{\beta_k^2(t)}{\alpha_k^2(t)}\frac{1}{\rho_k^2(t)} = \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))}\frac{\sinh^2(\omega_k(1-t))}{\omega_k^2} = \frac{\sinh^2(\omega_k t)}{\omega_k^2}. \tag{27}$$

When $\omega_k = 0$, the weight $\eta_k(t)$ is defined by its continuous limits t^2. Figure 3 illustrates the weighting schedule for different values of ω. From the figure, we observe that the weighting schedule assigns larger coefficients to velocity discrepancies evaluated closer to the data endpoint and at higher graph frequencies. Given a test time window $\mathbf{X}_1$, we independently draw source samples $\mathbf{X}_0$ from the standard Gaussian distribution to form a finite prior set $\mathcal{X}_{\text{ad}}$. Let $\mathcal{R}_{\text{ad}} \subset (0,1)$ denote the set of flow times used for evaluation. For each source sample and evaluation time, we construct the intermediate state and conditional target velocity using the graph-spectral path. We then aggregate the weighted squared velocity discrepancies across source samples, flow times, and graph-frequency modes to obtain the anomaly score as follows:

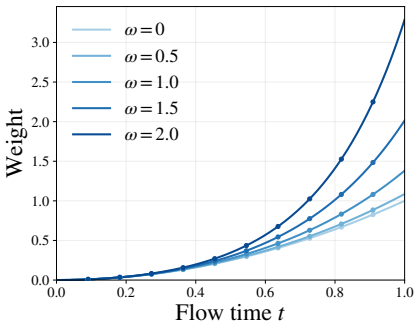


Figure 3: Weight schedule of $\eta_k(t)$.

$$\xi(\mathcal{X}_{\text{ad}}, \mathcal{R}_{\text{ad}}, \mathbf{X}_1) = \sum_{k=1}^{N}\sum_{t\in\mathcal{R}_{\text{ad}}}\sum_{\mathbf{X}_0\in\mathcal{X}_{\text{ad}}} \eta_k(t)\left\|\hat{\mathbf{v}}_{t,k}(\mathbf{X}_t;\boldsymbol{\theta}) - \dot{\hat{\gamma}}_k^\star(t)\right\|_2^2. \tag{28}$$

A larger value of ξ indicates a greater deviation from the velocity field learned from normal data and hence provide stronger evidence of anomalous behavior.

3.3 THEORETICAL ANALYSES

The graph Fourier basis is not unique when the graph Laplacian has repeated eigenvalues. More precisely, the eigenvectors within each repeated-eigenvalue eigenspace are identifiable only up to an orthogonal transformation (Agaskar & Lu, 2013; Sandryhaila & Moura, 2014; Deri & Moura, 2017). Let $\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\mathbf{Q}$ denote an alternative orthonormal eigenbasis, where $\mathbf{Q}$ is block orthogonal and acts only within eigenspaces associated with repeated eigenvalues. The same graph Laplacian then admits the eigendecomposition $\mathbf{L} = \tilde{\mathbf{\Psi}}\mathbf{\Lambda}\tilde{\mathbf{\Psi}}^\top$. The following result shows that this potential ambiguity of the graph Fourier basis does not affect the velocity field model, the target velocity, and the anomaly score used by GRASP.

Proposition 1 (Eigenbasis invariance)**.** *In GRASP, the node-domain graph-spectral path, its conditional target velocity, and the anomaly score in Equation (28) are all invariant to the choice of orthonormal Laplacian eigenbasis.*

The proof of Proposition 1 is provided in Appendix D.3. Under the population-optimal squared flow matching objective, the learned velocity predictor satisfies

$$\hat{\mathbf{V}}_t(\mathbf{X}_t; \boldsymbol{\theta}) = \hat{\mathbf{V}}_t^p(\mathbf{X}_t). \tag{29}$$

The trained velocity model approximates this population-optimal predictor. We therefore analyze the anomaly score obtained with this oracle normal data velocity $\hat{\mathbf{V}}_t^p$. Let $p_t(\hat{\mathbf{X}})$ and $q_t(\hat{\mathbf{X}})$ denote the graph-spectral path marginals induced by the normal endpoint distribution p and test endpoint distribution q, respectively. We define their spectral score matrices as

$$\hat{\mathbf{S}}_t^p(\hat{\mathbf{X}}) = \nabla_{\hat{\mathbf{X}}} \log p_t(\hat{\mathbf{X}}), \quad \hat{\mathbf{S}}_t^q(\hat{\mathbf{X}}) = \nabla_{\hat{\mathbf{X}}} \log q_t(\hat{\mathbf{X}}), \tag{30}$$

whose k-th rows are denoted as $\hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}})$ and $\hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}})$, respectively. We define $\hat{\boldsymbol{\mu}}_{t,k}^q$ analogously to $\hat{\boldsymbol{\mu}}_{t,k}^p$. In the following, we show that the expected oracle anomaly score of GRASP admits a decomposition that contains a term that measures deviations between normal and test distributions.

Theorem 2 (Decomposition of the expected oracle anomaly score)**.** *Let $\mathcal{X}_{\mathrm{ad}}$ contain a finite set of i.i.d. source samples from p_0, independent of $\mathbf{X}_1$, and let $\mathcal{R}_{\mathrm{ad}} \subset (0,1)$ be a finite set of evaluation times. Consider the anomaly score in Equation (28) evaluated using the oracle normal-data velocity predictor V_t^p. Its expectation over $\mathcal{X}_{\mathrm{ad}}$ and $\mathbf{X}_1 \sim q$ decomposes as*

$$\begin{aligned}\mathbb{E}_{\mathcal{X}_{\mathrm{ad}},q}\left[\xi(\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}, \mathbf{X}_1)\right] = \sum_{k=1}^{N} \sum_{t\in\mathcal{R}_{\mathrm{ad}}} |\mathcal{X}_{\mathrm{ad}}| \Bigg(& \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))} \mathbb{E}_{p_0,q}\left[\left\|\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X})\right\|_2^2\right] \\ & + \frac{\sinh^2(\omega_k(1-t))}{\sinh^2(\omega_k)} \mathbb{E}_{q_t}\left[\left\|\hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}}) - \hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}})\right\|_2^2\right]\Bigg),\end{aligned} \tag{31}$$

where $\mathbb{E}_{p_0,q}$ denotes expectation over independent $\mathbf{X}_0 \sim p_0$ and $\mathbf{X}_1 \sim q$ and the coefficients are defined through their continuous limits when $\omega_k = 0$.

Corollary 1 (Boundedness of the endpoint-uncertainty term)**.** *Under the setting of Theorem 2, assume that $\mathbb{E}_{\mathbf{X}_1\sim q}[\|\mathbf{X}_1\|_F^2] < \infty$. Then, for any finite set of evaluation times $\mathcal{R}_{\mathrm{ad}} \subset (0,1)$, the expected endpoint-uncertainty term in Equation (31) satisfies*

$$\sum_{k=1}^{N} \sum_{t\in\mathcal{R}_{\mathrm{ad}}} |\mathcal{X}_{\mathrm{ad}}| \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))} \mathbb{E}_{p_0,q}\left[\left\|\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X})\right\|_2^2\right] \le |\mathcal{X}_{\mathrm{ad}}||\mathcal{R}_{\mathrm{ad}}|NR. \tag{32}$$

The proof of Theorem 2 and Corollary 1 are provided in Appendix D.4 and Appendix D.5, respectively. The first term in Equation (31) captures graph-frequency-weighted irreducible endpoint uncertainty, which is bounded above as shown in Corollary 1. Whereas the second term in Equation (31) measures the graph-frequency-weighted Fisher discrepancy between the test and normal path marginals, which provides essential signals for anomaly detection. The decomposition therefore identifies a distribution-sensitive component of the proposed anomaly score that explicitly measures deviations between normal and test distributions.

Table 1: Anomaly detection performance comparison.

Model	SMAP			SMD			CICIDS			SWAN		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1	PRC	ROC	Best-F1	PRC	ROC	Best-F1
HBOS	0.1897	0.5650	0.2664	0.2678	0.7378	0.3556	0.2603	0.5586	0.3852	0.2792	0.5000	0.4365
COPOD	0.1927	0.5918	0.2841	0.2139	0.7193	0.2951	0.2214	0.5566	0.3733	0.2792	0.5000	0.4365
Autoencoder	0.2277	0.5859	0.2991	0.3613	0.7297	0.4214	0.3147	0.7276	0.3938	0.3294	0.4944	0.4371
USAD	0.2263	0.5850	0.3850	0.4105	0.8726	0.4810	0.2094	0.4619	0.3525	0.4405	0.5585	0.4365
CNN	0.2136	0.6577	0.3471	0.3730	0.7921	0.4313	0.2371	0.5543	0.3991	0.3760	0.4966	0.4365
OmniAnomaly	0.2420	0.5950	0.4065	0.4284	0.8822	0.5061	0.2138	0.4501	0.3548	0.4847	0.6130	0.4365
TranAD	0.2122	0.6104	0.3264	0.3304	0.7691	0.4181	0.2043	0.4704	0.3416	0.3360	0.4911	0.4365
A-Transformer	0.1340	0.5065	0.2060	0.0720	0.5013	0.1260	0.2396	0.4972	0.3060	0.2817	0.5021	0.4365
TimesNet	0.1792	0.5336	0.3161	0.2839	0.8087	0.3713	0.1905	0.4657	0.3121	0.3068	0.4505	0.4365
FITS	0.1455	0.5147	0.2798	0.2863	0.8299	0.4025	0.1849	0.4106	0.3126	0.3189	0.4607	0.4365
GDN	0.2197	0.6099	0.3394	0.4006	0.8014	0.4640	0.2926	0.7017	0.3805	0.6423	0.8356	0.6432
GCAD	0.2144	0.6348	0.3471	0.3569	0.8416	0.4535	0.3636	0.7191	0.3844	0.6445	0.8140	0.6231
CATCH	0.2450	0.6078	0.3664	0.4763	**0.8968**	0.5057	0.2866	0.6162	0.4023	0.4638	0.6218	0.4533
GRASP	**0.3149**	**0.7015**	**0.4178**	**0.4845**	0.8567	**0.5087**	**0.3996**	**0.8108**	**0.4615**	**0.7556**	**0.8758**	**0.6886**

4 EXPERIMENTS

We evaluate GRASP on four widely used MTS anomaly detection datasets covering different application domains: the spacecraft telemetry dataset SMAP (Hundman et al., 2018), the IT infrastructure dataset SMD (Su et al., 2019), the cybersecurity dataset CICIDS (Sharafaldin et al., 2018), and the astrophysics dataset SWAN (Angryk et al., 2020). For each dataset, we construct the sensor graph using normal training data. Specifically, we first compute a pairwise distance matrix $\boldsymbol{\Upsilon} \in \mathbb{R}^{N\times N}$ from the node features. The distances are converted into edge weights between 0 and 1 using the Gaussian kernel $\exp(-(\frac{\upsilon_{ij}}{\iota})^2)$, where the decay rate ι is set to the standard deviation of $\boldsymbol{\Upsilon}$. The resulting weight matrix is subsequently converted into a binary adjacency matrix with thresholding.

For each MTS, we use the first 80% of the normal training data for anomaly detector optimization and the remaining 20% for validation. We evaluate MTS anomaly detection performance with three metrics: the area under the receiver operating characteristic curve (ROC), the area under the precision-recall curve (PRC), and the Best-F1 score. ROC and PRC measure threshold-independent ranking performance, with PRC being particularly informative for highly imbalanced cases. The Best-F1 score measures point-wise detection performance with the threshold that maximizes the F1 score using test labels. It therefore represents retrospective detection potential rather than performance at a deployable threshold. For each dataset and random seed, we first average the results across its constituent series and then report the mean and standard deviation over ten random seeds.

We compare the proposed GRASP model with thirteen classical and deep-learning-based anomaly detection methods, including HBOS (Goldstein & Dengel, 2012), COPOD (Li et al., 2020), Autoencoder (Sakurada & Yairi, 2014), USAD (Audibert et al., 2020), CNN (Munir et al., 2018), OmniAnomaly (Su et al., 2019), TranAD (Tuli et al., 2022), A-Transformer (Xu et al., 2022), TimesNet (Wu et al., 2023), FITS (Xu et al., 2024), GDN (Deng & Hooi, 2021), GCAD (Liu et al., 2025), and CATCH (Wu et al., 2025). For these baselines, we use the implementations provided in mTSBench Zhou et al. (2026) whenever available. For the rest of models, we use their official implementations.

In the following, we first evaluate the anomaly detection performance of the GRASP model and the baselines in Section 4.1. Then we perform ablation studies to evaluate the individual contributions of the graph-spectral path, score weighting schedule, and data-dependent graph construction in Section 4.2. In Section 4.3, we further analyze how different flow times and graph frequencies contribute to the aggregated anomaly score. Additional implementation details and experimental results are provided in Appendix E. The effects of different velocity-field architectures are studied in Appendix E.6. Appendix E.8 analyzes how the model performs with different numbers of evaluated flow times and source samples, while Appendix E.9 analyzes how the model performs with different graph Dirichlet energy weight τ. The inference efficiency of the model is discussed in Appendix E.10.

4.1 ANOMALY DETECTION PERFORMANCE

Table 1 compares GRASP with the baselines on four datasets. The best results are highlighted in bold, and the second-best results are underlined. GRASP achieves the highest PRC and Best-F1 on all four datasets, as well as the highest ROC on three datasets. Its PRC gains are particularly

pronounced on SMAP and SWAN, indicating improved anomaly ranking across different application domains. On SMD, however, CATCH achieves a higher ROC, although GRASP retains the best PRC and Best-F1. Overall, these results show strong and consistent detection performance of GRASP.

4.2 Ablation Studies

We conduct ablation studies using three variants of GRASP: 1) GRASP-mean retains the graph-spectral path with a uniform weighting schedule across flow times and graph frequencies; 2) GRASP-ER replaces the data-dependent graph in GRASP-mean with an Erdős–Rényi graph that has the same sparsity level; 3) GRASP-lin replaces the graph-spectral path in GRASP-mean with the linear path by setting $\tau = 0$. The results on SMAP and SMD are presented in Table 2, while the results on CICIDS and SWAN are deferred to Appendix E.5. GRASP achieves the best results on all reported metrics. Its improvement over GRASP-mean supports the effectiveness of the proposed weighting schedule. GRASP-mean consistently outperforms GRASP-lin, indicating that the graph-spectral path provides more informative anomaly signals than the linear path. GRASP-mean also consistently outperforms GRASP-ER, suggesting that the data-dependent graph is more informative than a random graph.

Table 2: Ablation results on SMAP and SMD.

Model	SMAP			SMD		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1
GRASP-lin	0.2910	0.6879	0.3911	0.4762	0.8366	0.4968
GRASP-ER	0.2907	0.6909	0.3912	0.4797	0.8379	0.4993
GRASP-mean	0.2915	0.6941	0.3920	0.4801	0.8385	0.5005
GRASP	**0.3149**	**0.7015**	**0.4178**	**0.4845**	**0.8567**	**0.5087**

4.3 Flow-Time and Graph-Frequency Analysis

Figure 4 presents the unweighted and weighted anomaly signal attribution across flow times and graph-frequency quantiles on SMAP. Specifically, we equally divide the graph-frequency spectrum into eight quantiles and aggregate the squared velocity residuals within each flow-time and graph-frequency bin. Each heatmap is normalized separately to sum to 100%, so each cell reports its percentage of the total anomaly signal attribution. From the results, we observe that the anomaly evidence is concentrated toward the end of the flow trajectory and concentrated in a narrow graph-frequency band. Applying the weighting schedule further emphasizes the anomaly signals in later flow times and larger graph frequencies. This aligns with the fact that later states carry more information about the data endpoint than earlier states, which are more influenced by the Gaussian source. The dominant frequency band location differs across datasets (results on other datasets are deferred to Appendix E.7), suggesting that anomaly-relevant graph frequencies depend on the specific systems.

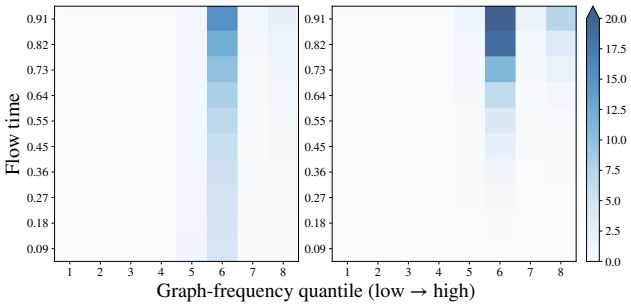


Figure 4: Unweighted (left) and weighted (right) anomaly signal attribution across time and frequency on SMAP.

5 Conclusion

In this work, we introduced a graph-informed flow matching approach for MTS anomaly detection named GRASP. It relies on a graph-spectral path which is constructed by minimizing a fixed-endpoint action that integrates kinetic energy with graph Dirichlet energy. GRASP detects anomalies by aggregating weighted velocity discrepancies across flow times, source samples, and graph frequencies. We proved that the model is invariant to different Laplacian eigenbasis and the expected oracle anomaly score can be decomposed into bounded endpoint uncertainty and graph-frequency-weighted Fisher discrepancy. Experiments on four datasets validated the effectiveness of GRASP.

Despite its promising results, GRASP has some limitations. First, it relies on a fixed graph, making it unsuitable when we have dynamically evolving graphs. Moreover, computing the graph Fourier basis can become costly as the number of variables increases. Future work could extend GRASP to be compatible with dynamic graphs and develop scalable approximations to graph spectral operations.

AI USE STATEMENT

In this work, we used OpenAI Codex for tasks such as proof reading, consistency checking, providing feedback, and implementing code. LLM-generated code was verified and tested for correctness. We have reviewed all AI-assisted work and take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

For the developed theoretical results, we have clearly mentioned the assumptions, and complete proofs are given in Appendix D. For the experiments, we use open-source data, and we provide implementation details in Appendix E.3. The code is available in the Supplementary Material.

REFERENCES

Ameya Agaskar and Yue M Lu. A spectral graph uncertainty principle. *IEEE Transactions on Information Theory*, 59(7):4338–4356, 2013.

Michael Samuel Albergo, Mark Goldstein, Nicholas Matthew Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *International Conference on Machine Learning*, 2024.

Rafal Angryk, Petrus Martens, Berkay Aydin, Dustin Kempton, Sushant Mahajan, Sunitha Basodi, Azim Ahmadzadeh, Xumin Cai, Soukaina Filali Boubrahimi, Shah Muhammad Hamdi, et al. SWAN-SF: Space weather analytics dataset-solar flares. *Harvard Dataverse dataset*, pp. 102, 2020.

Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A Zuluaga. USAD: Unsupervised anomaly detection on multivariate time series. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Mohammed Ayalew Belay, Sindre Stenen Blakseth, Adil Rasheed, and Pierluigi Salvo Rossi. Unsupervised anomaly detection for iot-based multivariate time series: Existing solutions, performance analysis and future directions. *Sensors*, 23(5):2844, 2023.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Hongjie Chen and Hoda Eldardiry. Graph time-series modeling in deep learning: A survey. *ACM Transactions on Knowledge Discovery from Data*, 18(5):1–35, 2024.

Shengzhe Chen, Mehrdad Moradi, Kamran Paynabar, and Hao Yan. Flow mismatching: Unsupervised anomaly detection via velocity discrepancies in flow matching models. *arXiv preprint arXiv:2605.23070*, 2026.

Si-An Chen, Chun-Liang Li, Sercan O Arik, Nathanael Christian Yoder, and Tomas Pfister. TSMixer: An all-MLP architecture for time series forecast-ing. *Transactions on Machine Learning Research*, 2023.

Dongchan Cho, Jiho Han, Keumyeong Kang, Minsang Kim, Honggyu Ryu, and Namsoon Jung. Structured temporal causality for interpretable multivariate time series anomaly detection. *Advances in Neural Information Processing Systems*, 2025.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *AAAI Conference on Artificial Intelligence*, 2021.

Joya A Deri and José MF Moura. Spectral projector-based graph fourier transforms. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):785–795, 2017.

Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst. Learning laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23): 6160–6173, 2016.

Shukai Du, Junzhe Zhang, and Yiming Li. Lagrangian flow matching: A least-action framework for principled path design. *arXiv preprint arXiv:2605.15419*, 2026.

Dengzhao Fang, Jingtong Gao, Yu Li, Xiangyu Zhao, and Yi Chang. Escaping the euclidean void: Manifold-informed flow matching for sequential recommendation. *arXiv preprint arXiv:2607.23762*, 2026.

Olga Fink, Vinay Sharma, Ismail Nejjar, Leandro Von Krannichfeldt, Sergei Garmaev, Zepeng Zhang, Amaury Wei, Gaetan Frusque, Florent Forest, Mengjie Zhao, et al. From physics to machine learning and back: Part I-learning with inductive biases in prognostics and health management. *Reliability Engineering & System Safety*, pp. 112213, 2026.

Jhony H Giraldo, Arif Mahmood, Belmar Garcia-Garcia, Dorina Thanou, and Thierry Bouwmans. Reconstruction of time-varying graph signals via sobolev smoothness. *IEEE Transactions on Signal and Information Processing over Networks*, 8:201–214, 2022.

Markus Goldstein and Andreas Dengel. Histogram-based outlier score (HBOS): A fast unsupervised anomaly detection algorithm. *KI-2012: poster and demo track*, 1:59–63, 2012.

Thi Kieu Khanh Ho, Ali Karami, and Narges Armanfard. Graph anomaly detection in time series: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018.

Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10466–10485, 2024.

Vassilis Kalofolias. How to learn a graph from smooth signals. In *International Conference on Artificial Intelligence and Statistics*, 2016.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Marcel Kollovieh, Marten Lienen, David Lüdke, Leo Schwinn, and Stephan Günnemann. Flow matching with gaussian process priors for probabilistic time series forecasting. In *International Conference on Learning Representations*, 2025.

Jinghan Li, Yuan Gao, Jinda Lu, Junfeng Fang, Congcong Wen, Hui Lin, and Xiang Wang. DiffGAD: A diffusion-based unsupervised graph anomaly detector. In *International Conference on Learning Representations*, 2025.

Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, and Xiyang Hu. COPOD: Copula-based outlier detection. In *IEEE international conference on data mining*, 2020.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Yilin Liu, Hongchao Zhang, Ahmad Taha, Taylor T Johnson, and Meiyi Ma. Modeling spectral energy shifts in spatio-temporal graph anomaly detection. In *International Conference on Machine Learning*, 2026.

Zehao Liu, Mengzhou Gao, and Pengfei Jiao. GCAD: Anomaly detection in multivariate time series from the perspective of granger causality. In *AAAI Conference on Artificial Intelligence*, 2025.

Mohsin Munir, Shoaib Ahmed Siddiqui, Andreas Dengel, and Sheraz Ahmed. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access*, 7:1991–2005, 2018.

Sergio Rozada, KB Vimal, Andrea Cavallo, Antonio G Marques, Hadi Jamali-Rad, and Elvin Isufi. Graph-aware diffusion for signal generation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2026.

Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Workshop on Machine Learning for Sensory Data Analysis*, 2014.

Aliaksei Sandryhaila and Jose MF Moura. Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on Signal Processing*, 62(12):3042–3054, 2014.

Iman Sharafaldin, Arash Habibi Lashkari, Ali A Ghorbani, et al. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSP*, 1(2018):108–116, 2018.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: Deep transformer networks for anomaly detection in multivariate time series data. *VLDB Endowment*, 15(6):1201–1214, 2022.

Yiğit Berkay Uslu, Samar Hadou, Sergio Rozada, Shirin Saeedi Bidokhti, and Alejandro Ribeiro. Graph signal generative diffusion models. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2026.

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

Cédric Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. CATCH: Channel-aware multivariate time series anomaly detection via frequency patching. In *International Conference on Learning Representations*, 2025.

Yulei Wu, Hong-Ning Dai, and Haina Tang. Graph neural networks for anomaly detection in industrial internet of things. *IEEE Internet of Things Journal*, 9(12):9214–9231, 2021.

Kacper Wyrwal, Ismail Ilkan Ceylan, and Alexander Tong. Topological flow matching. In *International Conference on Learning Representations*, 2026.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations*, 2024.

Maosheng Yang. Topological schrödinger bridge matching. In *International Conference on Learning Representations*, 2025.

Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.

Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, and Nitesh V Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In *AAAI Conference on Artificial Intelligence*, 2019.

Zepeng Zhang, Aref Einizade, Jhony H Giraldo, and Olga Fink. Spatiotemporal imputation with graph-informed flow matching. In *International Conference on Machine Learning*, 2026a.

Zepeng Zhang, Fuad Khuri, Keivan Faghih Niresi, and Olga Fink. GSLAD: Prototype-regularized graph structure learning for multivariate time series anomaly detection. *arXiv preprint arXiv:2609.15483*, 2026b.

Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *IEEE International Conference on Data Mining*, 2020.

Mengjie Zhao and Olga Fink. DyEdgeGAT: Dynamic edge via graph attention for early fault detection in IIoT systems. *IEEE Internet of Things Journal*, 11(13):22950–22965, 2024.

Xiaona Zhou, Constantin Brif, and Ismini Lourentzou. mTSBench: Benchmarking multivariate time series anomaly detection and model selection at scale. *Transactions on Machine Learning Research*, 2026.

Appendix

Table of Contents

A Related Work **15**

B Notation Summary **15**

C Velocity-Field Architecture **15**

D Proofs of Theoretical Results **17**
D.1 Proof of Theorem 1 17
D.2 Proof of Lemma 1 17
D.3 Proof of Proposition 1 18
D.4 Proof of Theorem 2 20
D.5 Proof of Corollary 1 21

E Experimental Details and Additional Results **22**
E.1 Datasets 22
E.2 Baselines 22
E.3 Implementation Details 23
E.4 Additional MTS Anomaly Detection Results 23
E.5 Additional Ablation Results 24
E.6 Comparison of Velocity-Field Architectures 24
E.7 Additional Flow-Time and Graph-Frequency Analysis 24
E.8 Sensitivity to Flow-Time Evaluations and Source Samples 25
E.9 Sensitivity to the Graph Dirichlet-Energy Weight 26
E.10 Inference Efficiency 27

A Related Work

Graph-based MTS anomaly detection. Unsupervised MTS anomaly detection methods learn patterns of normal system operation and identify observations that deviate from them. A common approach is to compute anomaly scores based on forecasting or reconstruction errors. For example, MTAD-GAT (Zhao et al., 2020) combines forecasting and reconstruction objectives during training and evaluates discrepancies between the observed signal and model outputs at inference. Graph-based approaches further incorporate dependencies among variables into these objectives. GDN (Deng & Hooi, 2021) learns sensor relationships among variables to support forecasting-based detection, whereas DyEdgeGAT (Zhao & Fink, 2024) constructs input-dependent graphs for reconstruction-based detection. Other approaches characterize anomalies through association discrepancies (Xu et al., 2022), differences in learned representations (Yang et al., 2023), deviations in the frequency domain (Wu et al., 2025), or shifts in graph-spectral energy Liu et al. (2026). More recently, diffusion-based methods have been introduced to reconstruct normal graph signals (Li et al., 2025). However, these forecasting-, reconstruction-, and diffusion-based methods primarily evaluate discrepancies at the predicted or reconstructed endpoint. Consequently, subtle or partially predictable anomalies may remain difficult to detect when they produce only weak endpoint discrepancies.

Flow matching and conditional path design. Flow matching learns a neural velocity field by regressing against conditional target velocities along prescribed probability paths that connect a simple source distribution to the data distribution. The choice of probability path is essential because it determines both the intermediate states and the target velocities used for training and evaluation (Du et al., 2026). Existing constructions, including rectified paths (Liu et al., 2023) and optimal-transport-based paths (Tong et al., 2024), typically use straight-line interpolation between coupled endpoints. From a variational perspective, the standard linear conditional path minimizes the action associated with a kinetic-energy Lagrangian under fixed-endpoint constraints (Benamou & Brenier, 2000; Villani, 2009), and it has become a standard choice in flow matching (Lipman et al., 2023). However, this construction depends only on the endpoints and does not explicitly incorporate the geometry or relational structure of the data. The resulting intermediate states may therefore be poorly aligned with data supported on non-Euclidean domains (Fang et al., 2026).

Graph-aware generative modeling. Several recent methods investigate incorporating graph structural information into diffusion or other generative frameworks. Graph-aware diffusion models introduce structural information through graph-based denoising architectures (Uslu et al., 2026) or graph-aware forward noising schedules (Rozada et al., 2026). Yang (2025) extends Schrödinger bridge matching to topological domains such as graphs and simplicial complexes, while Wyrwal et al. (2026) incorporates topological information into the reference process through a Laplacian-derived drift. These methods introduce graph structure into the denoising model or a predefined stochastic reference process. In contrast, GRASP derives the conditional probability path directly as the solution to a fixed-endpoint variational problem that combines kinetic energy with graph Dirichlet energy. This formulation yields a closed-form, graph-frequency-dependent path specifically designed for velocity-based MTS anomaly detection.

B Notation Summary

Table 3 summarizes the notation used throughout the paper.

C Velocity-Field Architecture

Given an intermediate state $\mathbf{X}_t \in \mathbb{R}^{N\times R}$ and its flow time t, we first encode t using a sinusoidal flow-time embedding function $\mathbf{e}(t)$. The intermediate state and time embedding are then concatenated and mapped to an initial hidden representation through an input projection:

$$\mathbf{H}^{(0)} = \phi_{\text{in}}\left(\mathbf{X}_t,\, e(t)\right), \tag{33}$$

where ϕ_{in} denote the input projection.

Following the TSMixer architecture (Chen et al., 2023), we process the hidden representation using L_{TSMixer} residual mixing blocks. Each block alternates between temporal mixing and cross-variable

Table 3: Summary of notation used throughout the paper.

Symbol	Description
$N,\ R$	Number of variables and length of a time-series window.
$\mathcal{G} = (\mathcal{N}, \mathcal{E})$	Sensor graph with node set $\mathcal{N}$ and edge set $\mathcal{E}$.
$\mathbf{X} \in \mathbb{R}^{N \times R}$	Multivariate time-series window.
$\mathbf{L}$	Symmetric normalized graph Laplacian.
$\mathbf{\Psi}, \mathbf{\Lambda}$	Laplacian eigenvector and eigenvalue matrices satisfying $\mathbf{L} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}^\top$.
$\lambda_k, \boldsymbol{\psi}_k$	Laplacian eigenvalue and eigenvector associated with graph-frequency mode k.
$\hat{\mathbf{X}} = \mathbf{\Psi}^\top \mathbf{X}$	Graph Fourier transform of $\mathbf{X}$.
$\hat{\mathbf{x}}_{t,k}$	The k-th row of $\hat{\mathbf{X}}$, representing graph-frequency mode k across the time window.
$t \in [0, 1]$	Flow time.
$\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_t$	Source endpoint, data endpoint, and intermediate state.
p_0	Standard Gaussian source distribution.
$p,\ q$	Normal and test endpoint distributions.
$p_t,\ q_t$	Intermediate-state distributions induced by endpoint distributions p and q.
ϕ_t	Conditional interpolation map between $\mathbf{X}_0$ and $\mathbf{X}_1$.
$\mathbf{U}_t$	Conditional target velocity induced by ϕ_t.
$\mathbf{V}_t(\mathbf{X}; \boldsymbol{\theta})$	Learned time-dependent velocity field.
$\mathbf{\Gamma}^\star(t), \dot{\mathbf{\Gamma}}^\star(t)$	Optimal graph-spectral path and its velocity.
$\hat{\boldsymbol{\gamma}}_k^\star(t)$	The k-th graph-frequency component of $\Gamma^\star(t)$.
τ	Weight of the graph Dirichlet energy.
$\omega_k = \sqrt{\tau\lambda_k}$	Graph-spectral parameter associated with graph-frequency mode k.
$\alpha_k(t), \beta_k(t)$	Source and target interpolation coefficients for graph-frequency mode k.
$\mathbf{A}(t), \mathbf{B}(t)$	Diagonal matrices collecting $\alpha_k(t)$ and $\beta_k(t)$.
$\mathbf{V}_t^p(\mathbf{X})$	Oracle marginal velocity field induced by the normal endpoint distribution.
$\hat{\mathbf{v}}_{t,k}^p$	The k-th row of $\mathbf{V}_t^p(\mathbf{X})$.
$\mathbf{M}_t^p(\mathbf{X}), \mathbf{M}_t^q(\mathbf{X})$	Endpoint posterior means under p and q.
$\hat{\boldsymbol{\mu}}_{t,k}^p, \hat{\boldsymbol{\mu}}_{t,k}^q$	Endpoint posterior mean of graph-frequency mode k under distribution p and q.
$\rho_k(t)$	Scaling factor relating the conditional velocity residual to the endpoint posterior residual.
$\eta_k(t)$	Anomaly-score weight for graph-frequency mode k at flow-time t.
$\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}$	Sets of source samples and evaluation flow times used for anomaly scoring.
$\xi(\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}, \mathbf{X}_1)$	Anomaly score of test window $\mathbf{X}_1$.
$\hat{\mathbf{S}}_t^p, \hat{\mathbf{S}}_t^q$	Spectral score matrices of the normal and test path marginals.
$\mathbf{d}_{t,k}$	Endpoint posterior residual for graph-frequency mode k.
$\boldsymbol{\delta}_{t,k}$	Difference between the test and normal spectral scores of graph-frequency mode k.
$\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\mathbf{Q}$	Alternative Laplacian eigenbasis within repeated-eigenvalue eigenspaces.
$\mathbf{Q}$	Block-diagonal orthogonal transformation matrix acting within repeated-eigenvalue eigenspaces.

mixing as follows:

$$\tilde{\mathbf{H}}^{(\ell)} = \mathbf{H}^{(\ell-1)} + \mathcal{M}_{\mathrm{time}}^{(\ell)}\left(\mathrm{Norm}(\mathbf{H}^{(\ell-1)})\right), \tag{34}$$

$$\mathbf{H}^{(\ell)} = \tilde{\mathbf{H}}^{(\ell)} + \mathcal{M}_{\mathrm{var}}^{(\ell)}\left(\mathrm{Norm}(\tilde{\mathbf{H}}^{(\ell)})\right), \tag{35}$$

for $\ell = 1, \ldots, L_{\mathrm{TSMixer}}$. Here, Norm represents layer normalization (Ba et al., 2016), $\mathcal{M}_{\mathrm{time}}^{(\ell)}$ is an MLP applied along the temporal dimension and shared across variables, while $\mathcal{M}_{\mathrm{var}}^{(\ell)}$ is an MLP applied along the variable dimension and shared across time steps. Each mixing MLP consists of two linear layers with a nonlinear activation function ReLU and dropout.

Finally, an output projection maps the hidden representation back to the original data space:

$$\mathbf{V}_t(\mathbf{X}_t; \boldsymbol{\theta}) = \phi_{\mathrm{out}}\left(\mathbf{H}^{(L_{\mathrm{TSMixer}})}\right) \in \mathbb{R}^{N \times R}. \tag{36}$$

The alternating mixing operations enable the velocity predictor to capture temporal dependencies within individual variables and interactions across variables while retaining a lightweight, fully MLP-based architecture. Importantly, in GRASP, graph structure is incorporated through the probability path and conditional target velocity rather than through the velocity-field architecture itself.

D PROOFS OF THEORETICAL RESULTS

D.1 PROOF OF THEOREM 1

Proof. The Euler-Lagrange equation corresponds to the variational problem in Equation (9) is

$$\omega_k^2 \hat{\boldsymbol{\gamma}}_k(t) - \ddot{\hat{\boldsymbol{\gamma}}}_k(t) = \mathbf{0}, \tag{37}$$

whose general solution is

$$\hat{\boldsymbol{\gamma}}_k(t) = \mathbf{c}_{1,k} e^{\omega_k t} + \mathbf{c}_{2,k} e^{-\omega_k t} \tag{38}$$

with some $\mathbf{c}_{1,k}$ and $\mathbf{c}_{2,k}$. Since the hyperbolic functions satisfy

$$\sinh(\omega_k t) = \frac{e^{\omega_k t} - e^{-\omega_k t}}{2}, \quad \cosh(\omega_k t) = \frac{e^{\omega_k t} + e^{-\omega_k t}}{2}, \tag{39}$$

we can further transform it into

$$\hat{\boldsymbol{\gamma}}_k(t) = (\mathbf{c}_{1,k} - \mathbf{c}_{2,k}) \sinh(\omega_k t) + (\mathbf{c}_{1,k} + \mathbf{c}_{2,k}) \cosh(\omega_k t). \tag{40}$$

Considering the boundary condition, we have

$$\hat{\boldsymbol{\gamma}}_k^\star(t) = \frac{\hat{\mathbf{x}}_{1,k} - \cosh(\omega_k)\hat{\mathbf{x}}_{0,k}}{\sinh(\omega_k)} \sinh(\omega_k t) + \cosh(\omega_k t)\hat{\mathbf{x}}_{0,k}, \tag{41}$$

which can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{\gamma}}_k^\star(t) &= \left(\cosh(\omega_k t) - \frac{\cosh(\omega_k)\sinh(\omega_k t)}{\sinh(\omega_k)} \right) \hat{\mathbf{x}}_{0,k} + \frac{\sinh(\omega_k t)}{\sinh(\omega_k)} \hat{\mathbf{x}}_{1,k} \\ &= \frac{\sinh(\omega_k(1-t))}{\sinh(\omega_k)} \hat{\mathbf{x}}_{0,k} + \frac{\sinh(\omega_k t)}{\sinh(\omega_k)} \hat{\mathbf{x}}_{1,k}. \end{aligned} \tag{42}$$

When $\omega_k = 0$, the Euler-Lagrange equation reduces to

$$\ddot{\hat{\boldsymbol{\gamma}}}_k(t) = \mathbf{0}. \tag{43}$$

Under the same boundary conditions, its solution is

$$\hat{\boldsymbol{\gamma}}_k^\star(t) = (1-t)\hat{\mathbf{x}}_{0,k} + t\hat{\mathbf{x}}_{1,k}, \tag{44}$$

which coincides with the continuous limit of Equation (42) as $\omega_k \to 0$.

The kinetic-energy term in Equation (9) is strictly convex over the set of paths satisfying the fixed-endpoint constraints. Indeed, two admissible paths with identical derivatives can differ only by a constant, which must be zero because they share the same endpoints. Moreover, the graph-frequency regularization term is convex since $\omega_k^2 = \tau\lambda_k \geq 0$. Therefore, the complete objective in Equation (9) is strictly convex over its feasible set. Since the path in Equation (42) satisfies both the Euler–Lagrange equation and the boundary conditions, it is the unique global minimizer of the variational problem defined in Equation (9), through which the proof is completed. □

D.2 PROOF OF LEMMA 1

Proof. According to Theorem 1, the k-th graph-frequency component of the conditional path is

$$\hat{\boldsymbol{\gamma}}_k^\star(t) = \alpha_k(t)\hat{\mathbf{x}}_{0,k} + \beta_k(t)\hat{\mathbf{x}}_{1,k}. \tag{45}$$

For $t \in (0,1)$, we have $\alpha_k(t) > 0$. Then, we can obtain

$$\hat{\mathbf{x}}_{0,k} = \frac{1}{\alpha_k(t)} \left(\hat{\boldsymbol{\gamma}}_k^\star(t) - \beta_k(t)\hat{\mathbf{x}}_{1,k} \right). \tag{46}$$

Substituting Equation (46) into the conditional target velocity in Equation (12) gives

$$\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) = \frac{\dot{\alpha}_k(t)}{\alpha_k(t)} \left(\hat{\boldsymbol{\gamma}}_k^\star(t) - \beta_k(t)\hat{\mathbf{x}}_{1,k} \right) + \dot{\beta}_k(t)\hat{\mathbf{x}}_{1,k}. \tag{47}$$

which can be rewritten as

$$\dot{\hat{\gamma}}_k^\star(t) = \frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) + \rho_k(t)\hat{\mathbf{x}}_{1,k} \tag{48}$$

with

$$\rho_k(t) = \dot{\beta}_k(t) - \frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\beta_k(t). \tag{49}$$

For $\omega_k >$, substituting the expressions for $\alpha_k(t)$ and $\beta_k(t)$ yields

$$\begin{aligned}
\rho_k(t) &= \frac{\omega_k \cosh(\omega_k t)}{\sinh(\omega_k)} + \frac{\omega_k \cosh(\omega_k(1-t))}{\sinh(\omega_k)} \frac{\sinh(\omega_k)}{\sinh(\omega_k(1-t))} \frac{\sinh(\omega_k t)}{\sinh(\omega_k)} \\
&= \frac{\omega_k}{\sinh(\omega_k)} \frac{\cosh(\omega_k t)\sinh(\omega_k(1-t)) + \cosh(\omega_k(1-t))\sinh(\omega_k t)}{\sinh(\omega_k(1-t))} \\
&= \frac{\omega_k}{\sinh(\omega_k)} \frac{\sinh(\omega_k t + \omega_k(1-t))}{\sinh(\omega_k(1-t))} \\
&= \frac{\omega_k}{\sinh(\omega_k(1-t))}.
\end{aligned} \tag{50}$$

Based on the definition of the oracle marginal velocity, we have

$$\begin{aligned}
\hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) &= \mathbb{E}[\dot{\hat{\gamma}}_k^\star(t) \mid \mathbf{X}_t = \mathbf{X}] \\
&= \mathbb{E}[\frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) + \rho_k(t)\hat{\mathbf{x}}_{1,k} \mid \mathbf{X}_t = \mathbf{X}] \\
&= \mathbb{E}[\frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) \mid \mathbf{X}_t = \mathbf{X}] + \mathbb{E}[\rho_k(t)\hat{\mathbf{x}}_{1,k} \mid \mathbf{X}_t = \mathbf{X}] \\
&= \frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) + \rho_k(t)\hat{\boldsymbol{\mu}}_{t,k}^p.
\end{aligned} \tag{51}$$

Subtracting Equation (51) from Equation (48), we obtain

$$\begin{aligned}
\dot{\hat{\gamma}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) &= \frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) + \rho_k(t)\hat{\mathbf{x}}_{1,k} - \left(\frac{\dot{\alpha}_k(t)}{\alpha_k(t)}\hat{\gamma}_k^\star(t) + \rho_k(t)\hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X})\right) \\
&= \rho_k(t)\left(\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X})\right),
\end{aligned} \tag{52}$$

which completes the proof. □

D.3 Proof of Proposition 1

Proof. Let $\nu_1, \ldots, \nu_J$ denote the distinct eigenvalues of $\mathbf{L}$, and define

$$\mathcal{I}_j = \{k : \lambda_k = \nu_j\} \tag{53}$$

as the index set associated with eigenvalue ν_j. Any alternative orthonormal eigenbasis of $\mathbf{L}$ with the same eigenvalue ordering can be written as

$$\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi}\mathbf{Q}, \qquad \mathbf{Q} = \mathrm{blkdiag}\left(\mathbf{Q}_1, \ldots, \mathbf{Q}_J\right), \tag{54}$$

where each $\mathbf{Q}_j \in \mathbb{R}^{|\mathcal{I}_j| \times |\mathcal{I}_j|}$ is orthogonal. Hence, we have

$$\mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}, \qquad \mathbf{Q}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\mathbf{Q}. \tag{55}$$

All graph-spectral interpolation coefficients depend on the graph-frequency mode k only through λ_k. Therefore, their values are identical for all modes within the same repeated-eigenvalue eigenspace. It follows that

$$\mathbf{Q}\mathbf{A}(t) = \mathbf{A}(t)\mathbf{Q}, \qquad \mathbf{Q}\mathbf{B}(t) = \mathbf{B}(t)\mathbf{Q}, \tag{56}$$

and similarly,

$$\mathbf{Q}\dot{\mathbf{A}}(t) = \dot{\mathbf{A}}(t)\mathbf{Q}, \qquad \mathbf{Q}\dot{\mathbf{B}}(t) = \dot{\mathbf{B}}(t)\mathbf{Q}. \tag{57}$$

Under the alternative eigenbasis, the node-domain graph-spectral path satisfies

$$\begin{aligned}\dot{\mathbf{\Gamma}}^{\star}(t) &= \mathbf{\Psi}\dot{\mathbf{A}}(t)\mathbf{\Psi}^{\top}\mathbf{X}_0 + \mathbf{\Psi}\dot{\mathbf{B}}(t)\mathbf{\Psi}^{\top}\mathbf{X}_1 \\ &= \mathbf{\Psi}\mathbf{Q}\dot{\mathbf{A}}(t)\mathbf{Q}^{\top}\mathbf{\Psi}^{\top}\mathbf{X}_0 + \mathbf{\Psi}\mathbf{Q}\dot{\mathbf{B}}(t)\mathbf{Q}^{\top}\mathbf{\Psi}^{\top}\mathbf{X}_1 \\ &= \tilde{\mathbf{\Psi}}\dot{\mathbf{A}}(t)\tilde{\mathbf{\Psi}}^{\top}\mathbf{X}_0 + \tilde{\mathbf{\Psi}}\dot{\mathbf{B}}(t)\tilde{\mathbf{\Psi}}^{\top}\mathbf{X}_1.\end{aligned} \tag{58}$$

Thus, the node-domain conditional target velocity is invariant to the particular choice of eigenvectors within each degenerate eigenspace.

The learned velocity field model $\mathbf{V}_t(\mathbf{X}_t;\boldsymbol{\theta})$ is defined entirely in the node domain as a function of $\mathbf{X}_t$, which does not directly depend on $\mathbf{\Psi}$. Therefore, for fixed parameters $\boldsymbol{\theta}$, its output is unaffected by the choice of Laplacian eigenbasis. Moreover, because the node-domain conditional target velocity is invariant, the training objective in Equation (19) is also eigenbasis invariant.

It remains to establish the invariance of the anomaly score. Define the node-domain velocity residual as

$$\mathbf{R}_t = \mathbf{V}_t(\mathbf{X}_t;\boldsymbol{\theta}) - \dot{\mathbf{\Gamma}}^{\star}(t) \in \mathbb{R}^{N\times R}, \tag{59}$$

and its graph Fourier representation

$$\hat{\mathbf{R}}_t = \mathbf{\Psi}^{\top}\mathbf{R}_t. \tag{60}$$

The k-th row of $\hat{\mathbf{R}}_t$ is

$$\hat{\mathbf{r}}_{t,k} = \hat{\mathbf{v}}_{t,k} - \dot{\hat{\boldsymbol{\gamma}}}_k^{\star}(t). \tag{61}$$

Furthermore, define the diagonal weighting matrix

$$\mathbf{H}(t) = \operatorname{diag}\left(\eta_1(t),\ldots,\eta_N(t)\right). \tag{62}$$

Then, for a fixed t and $\mathbf{X}_0$, the corresponding contribution to the anomaly score can be written as

$$\begin{aligned}\sum_{k=1}^{N}\eta_k(t)\left\|\hat{\mathbf{v}}_{t,k} - \dot{\hat{\boldsymbol{\gamma}}}_k^{\star}(t)\right\|_2^2 &= \sum_{k=1}^{N}\eta_k(t)\left\|\hat{\mathbf{r}}_{t,k}\right\|_2^2 \\ &= \operatorname{Tr}\left(\hat{\mathbf{R}}_t^{\top}\mathbf{H}(t)\hat{\mathbf{R}}_t\right).\end{aligned} \tag{63}$$

Under the alternative eigenbasis $\tilde{\mathbf{\Psi}} = \mathbf{\Psi}\mathbf{Q}$, the corresponding spectral residual is

$$\tilde{\mathbf{R}}_t = \tilde{\mathbf{\Psi}}^{\top}\mathbf{R}_t = \mathbf{Q}^{\top}\hat{\mathbf{R}}_t. \tag{64}$$

Moreover, from the definition of the anomaly weight,

$$\eta_k(t) = \frac{\sinh^2(\omega_k t)}{\omega_k^2}, \qquad \omega_k = \sqrt{\tau\lambda_k}, \tag{65}$$

so $\eta_k(t)$ also depends on k only through λ_k. Consequently, all modes belonging to the same eigenspace have the same weight. Hence, $\mathbf{H}(t)$ is scalar within each degenerate eigenspace and satisfies

$$\mathbf{Q}\mathbf{H}(t) = \mathbf{H}(t)\mathbf{Q}, \qquad \mathbf{Q}\mathbf{H}(t)\mathbf{Q}^{\top} = \mathbf{H}(t). \tag{66}$$

Therefore,

$$\operatorname{Tr}\left(\tilde{\mathbf{R}}_t^{\top}\mathbf{H}(t)\tilde{\mathbf{R}}_t\right) = \operatorname{Tr}\left(\hat{\mathbf{R}}_t^{\top}\mathbf{Q}\mathbf{H}(t)\mathbf{Q}^{\top}\hat{\mathbf{R}}_t\right) \tag{67}$$

$$= \operatorname{Tr}\left(\hat{\mathbf{R}}_t^{\top}\mathbf{H}(t)\hat{\mathbf{R}}_t\right). \tag{68}$$

Thus, although individual spectral coordinates within a degenerate eigenspace depend on the choice of eigenbasis, their weighted aggregate contribution to the anomaly score is invariant.

Finally, summing the above equality over $t \in \mathcal{R}_{\mathrm{ad}}$ and $\mathbf{X}_0 \in \mathcal{X}_{\mathrm{ad}}$ shows that $\xi(\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}, \mathbf{X}_1)$ is unchanged under any orthogonal rotation within a repeated eigenspace. Therefore, the velocity field model, the conditional target velocity, and the resulting anomaly score are all invariant to the choice of Laplacian eigenbasis. □

D.4 Proof of Theorem 2

Proof. According to Theorem 1, we have

$$\hat{\boldsymbol{\gamma}}_k^\star(t) = \alpha_k(t)\hat{\mathbf{x}}_{0,k} + \beta_k(t)\hat{\mathbf{x}}_{1,k}. \tag{69}$$

Since the graph Fourier transform is orthonormal and $\mathbf{X}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, each source component satisfies $\hat{\mathbf{x}}_{0,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus, conditioning on the data endpoint gives

$$\hat{\boldsymbol{\gamma}}_k^\star(t) \mid \hat{\mathbf{x}}_{1,k} \sim \mathcal{N}\left(\beta_k(t)\hat{\mathbf{x}}_{1,k}, \alpha_k^2(t)\mathbf{I}\right). \tag{70}$$

The conditional density is

$$\phi_t(\hat{\mathbf{x}}_k(t) \mid \hat{\mathbf{x}}_{1,k}) = \frac{1}{(2\pi\alpha_k^2(t))^{R/2}} \exp\left[-\frac{\|\hat{\mathbf{x}}_k - \beta_k(t)\hat{\mathbf{x}}_{1,k}\|_2^2}{2\alpha_k^2(t)}\right]. \tag{71}$$

For $t \in (0,1)$, both $\alpha_k(t)$ and $\beta_k(t)$ are positive. The score over the k-th graph-frequency is

$$\begin{aligned}
\hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}}) &= \mathbb{E}_{p_0,p}\left[\nabla_{\hat{\mathbf{x}}_k} \log p_t(\hat{\mathbf{X}} \mid \hat{\mathbf{X}}_1) \mid \hat{\mathbf{X}}_t = \hat{\mathbf{X}}\right] \\
&= \mathbb{E}_{p_0,p}\left[-\frac{\hat{\mathbf{x}}_k - \beta_k(t)\hat{\mathbf{x}}_{1,k}}{\alpha_k^2(t)} \mid \hat{\mathbf{X}}_t = \hat{\mathbf{X}}\right] \\
&= -\frac{\hat{\mathbf{x}}_k}{\alpha_k^2(t)} + \frac{\beta_k(t)}{\alpha_k^2(t)}\hat{\boldsymbol{\mu}}_{t,k}^p.
\end{aligned} \tag{72}$$

Therefore, we have

$$\beta_k(t)\hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X}) = \hat{\mathbf{x}}_k + \alpha_k^2(t)\hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}}). \tag{73}$$

Applying the same identity under the test endpoint distribution q gives

$$\beta_k(t)\hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) = \hat{\mathbf{x}}_k + \alpha_k^2(t)\hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}}). \tag{74}$$

Thus, we have

$$\hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) - \hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X}) = \frac{\alpha_k^2(t)}{\beta_k(t)}\left(\hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}}) - \hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}})\right) \tag{75}$$

Based on Lemma 1, we know that

$$\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) = \rho_k(t)\left(\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X})\right). \tag{76}$$

Thus, we have

$$\begin{aligned}
\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) &= \rho_k(t)\left(\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) + \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) - \hat{\boldsymbol{\mu}}_{t,k}^p(\mathbf{X})\right) \\
&= \rho_k(t)\left(\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) + \frac{\alpha_k^2(t)}{\beta_k(t)}\left(\hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}}) - \hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}})\right)\right)
\end{aligned} \tag{77}$$

Define $\mathbf{d}_{t,k} = \hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X})$ and $\boldsymbol{\delta}_{t,k} = \hat{\mathbf{s}}_{t,k}^q(\hat{\mathbf{X}}) - \hat{\mathbf{s}}_{t,k}^p(\hat{\mathbf{X}})$, then we obtain

$$\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X}) = \rho_k(t)\left(\mathbf{d}_{t,k} + \frac{\alpha_k^2(t)}{\beta_k(t)}\boldsymbol{\delta}_{t,k}\right). \tag{78}$$

Thus,

$$\begin{aligned}
\|\dot{\hat{\boldsymbol{\gamma}}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X})\|_2^2 &= \rho_k^2(t)\left\|\mathbf{d}_{t,k} + \frac{\alpha_k^2(t)}{\beta_k(t)}\boldsymbol{\delta}_{t,k}\right\|_2^2 \\
&= \rho_k^2(t)\left(\|\mathbf{d}_{t,k}\|_2^2 + \frac{\alpha_k^4(t)}{\beta_k^2(t)}\|\boldsymbol{\delta}_{t,k}\|_2^2 + 2\frac{\alpha_k^2(t)}{\beta_k(t)}\mathbf{d}_{t,k}^\top\boldsymbol{\delta}_{t,k}\right).
\end{aligned} \tag{79}$$

Since

$$\hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) = \mathbb{E}_{p_0,q}[\hat{\boldsymbol{\gamma}}_k^\star(1) \mid \mathbf{X}_t = \mathbf{X}], \tag{80}$$

we have

$$\mathbb{E}_{p_0,q}\left[\mathbf{d}_{t,k} \mid \hat{\mathbf{X}}_t\right] = \mathbb{E}_{p_0,q}\left[\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}) \mid \hat{\mathbf{X}}_t\right] = \mathbf{0}. \tag{81}$$

As $\boldsymbol{\delta}_{t,k}$ only depends on $\hat{\mathbf{X}}_t$, the cross term vanishes:

$$\mathbb{E}_{p_0,q}\left[\langle \mathbf{d}_{t,k}, \boldsymbol{\delta}_{t,k}\rangle \mid \hat{\mathbf{X}}_t\right] = \langle \mathbb{E}_{p_0,q}\left[\mathbf{d}_{t,k} \mid \hat{\mathbf{X}}_t\right], \boldsymbol{\delta}_{t,k}(\hat{\mathbf{X}}_t)\rangle = \mathbf{0}. \tag{82}$$

Therefore

$$\mathbb{E}_{p_0,q}\left[\|\dot{\hat{\gamma}}_k^\star(t) - \hat{\mathbf{v}}_{t,k}^p(\mathbf{X})\|_2^2\right] = \rho_k^2(t)\mathbb{E}_{p_0,q}\left[\|\mathbf{d}_{t,k}\|_2^2\right] + \rho_k^2(t)\frac{\alpha_k^4(t)}{\beta_k^2(t)}\mathbb{E}_{q_t}\left[\|\boldsymbol{\delta}_{t,k}\|_2^2\right]. \tag{83}$$

Since the anomaly score sums the contributions from these identically distributed source samples in $\mathcal{X}_{\mathrm{ad}}$, linearity of expectation gives

$$\begin{aligned}&\mathbb{E}_{\mathcal{X}_{\mathrm{ad}},q}\left[\xi(\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}, \mathbf{X}_1)\right] \\ =&\sum_{k=1}^{N}\sum_{t\in\mathcal{R}_{\mathrm{ad}}} |\mathcal{X}_{\mathrm{ad}}|\left(\eta_k(t)\rho_k^2(t)\mathbb{E}_{p_0,q}\left[\|\mathbf{d}_{t,k}\|_2^2\right] + \eta_k(t)\rho_k^2(t)\frac{\alpha_k^4(t)}{\beta_k^2(t)}\mathbb{E}_{q_t}\left[\|\boldsymbol{\delta}_{t,k}\|_2^2\right]\right).\end{aligned} \tag{84}$$

Observe that

$$\eta_k(t)\rho_k^2(t) = \frac{\sinh^2(\omega_k t)}{\omega_k^2}\left(\frac{\omega_k}{\sinh(\omega_k(1-t))}\right)^2 = \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))} \tag{85}$$

and

$$\begin{aligned}\eta_k(t)\rho_k^2(t)\frac{\alpha_k^4(t)}{\beta_k^2(t)} &= \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))}\left(\frac{\sinh(\omega_k(1-t))}{\sinh(\omega_k)}\right)^4\left(\frac{\sinh(\omega_k)}{\sinh(\omega_k t)}\right)^2 \\ &= \frac{\sinh^2(\omega_k(1-t))}{\sinh^2(\omega_k)},\end{aligned} \tag{86}$$

we have the following result:

$$\begin{aligned}&\mathbb{E}_{\mathcal{X}_{\mathrm{ad}},q}\left[\xi(\mathcal{X}_{\mathrm{ad}}, \mathcal{R}_{\mathrm{ad}}, \mathbf{X}_1)\right] \\ =&\sum_{k=1}^{N}\sum_{t\in\mathcal{R}_{\mathrm{ad}}} |\mathcal{X}_{\mathrm{ad}}|\left(\frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))}\mathbb{E}_{p_0,q}\left[\|\mathbf{d}_{t,k}\|_2^2\right] + \frac{\sinh^2(\omega_k(1-t))}{\sinh^2(\omega_k)}\mathbb{E}_{q_t}\left[\|\boldsymbol{\delta}_{t,k}\|_2^2\right]\right),\end{aligned} \tag{87}$$

through which the proof is completed. □

D.5 Proof of Corollary 1

Proof. For $t \in (0,1)$, define the rescaled spectral observation

$$\mathbf{Z}_{t,k} = \frac{\hat{\gamma}_k^\star(t)}{\beta_k(t)} = \hat{\mathbf{x}}_{1,k} + \frac{\alpha_k(t)}{\beta_k(t)}\hat{\mathbf{x}}_{0,k}. \tag{88}$$

Since

$$\hat{\boldsymbol{\mu}}_{t,k}^q(\mathbf{X}_t) = \mathbb{E}_{p_0,q}[\hat{\mathbf{x}}_{1,k} \mid \mathbf{X}_t] \tag{89}$$

is the minimum mean-squared-error estimator of $\hat{\mathbf{x}}_{1,k}$, using $\mathbf{Z}_{t,k}$ as a candidate estimator gives

$$\begin{aligned}\mathbb{E}_{p_0,q}[\|\mathbf{d}_{t,k}\|_2^2] &\le \mathbb{E}_{p_0,q}[\|\hat{\mathbf{x}}_{1,k} - \mathbf{Z}_{t,k}\|_2^2] \\ &= \frac{\alpha_k^2(t)}{\beta_k^2(t)}\mathbb{E}_{p_0}[\|\hat{\mathbf{x}}_{0,k}\|_2^2] \\ &= R\frac{\alpha_k^2(t)}{\beta_k^2(t)}.\end{aligned} \tag{90}$$

Because

$$\frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))} = \frac{\beta_k^2(t)}{\alpha_k^2(t)}, \tag{91}$$

multiplying both sides of Equation (90) by this factor gives

$$\frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k(1-t))}\mathbb{E}_{p_0,q}[\|\mathbf{d}_{t,k}\|_2^2] \le R. \tag{92}$$

For $\omega_k = 0$, the ratio in Equation (92) is defined by its continuous limit $t^2/(1-t)^2$, and the same bound holds. Finally, summing Equation (92) over all source samples, flow times, and graph-frequency modes yields

$$\sum_{k=1}^{N} \sum_{t \in \mathcal{R}_{\mathrm{ad}}} |\mathcal{X}_{\mathrm{ad}}| \frac{\sinh^2(\omega_k t)}{\sinh^2(\omega_k (1-t))} \mathbb{E}_{p_0,q} \left[\|\hat{\mathbf{x}}_{1,k} - \hat{\boldsymbol{\mu}}^q_{t,k}(\mathbf{X})\|_2^2 \right] \leq |\mathcal{X}_{\mathrm{ad}}| \, |\mathcal{R}_{\mathrm{ad}}| \, NR, \tag{93}$$

which completes the proof. □

E Experimental Details and Additional Results

E.1 Datasets

We evaluate GRASP on four MTS anomaly detection datasets from mTSBench (Zhou et al., 2026), covering spacecraft telemetry, server monitoring, network intrusion detection, and space-weather analysis. Each dataset contains separate training and test sequences, together with pointwise anomaly labels for evaluation. We use the data preprocessing and benchmark splits provided by mTSBench. We briefly introduce the four datasets below:

- SMAP (Hundman et al., 2018): The Soil Moisture Active Passive dataset contains telemetry collected from NASA's SMAP spacecraft. The benchmark subset comprises 51 multivariate sequences, each containing 26 variables. This dataset evaluates the ability to detect anomalous behavior in spacecraft telemetry.
- SMD (Su et al., 2019): The Server Machine Dataset contains monitoring measurements collected from server machines. We use 18 sequences with 39 variables each, where the labeled anomalies represent deviations from normal server operation.
- CICIDS (Sharafaldin et al., 2018): CICIDS2017 is a network intrusion detection dataset containing both benign traffic and multiple types of cyberattacks. The benchmark subset comprises six sequences, each represented by 73 network-flow features.
- SWAN (Angryk et al., 2020): The Space Weather Analytics for Solar Flares dataset contains multivariate time series of physical properties associated with solar active regions. We use the 39-variable sequence included in the mTSBench anomaly detection benchmark.

E.2 Baselines

We compare GRASP with 13 representative baselines spanning statistical outlier detection, forecasting- and reconstruction-based detection, graph-based modeling, Transformers, and frequency-domain methods. We briefly introduce these baselines below:

- **GDN** (Deng & Hooi, 2021) learns inter-sensor dependencies using node embeddings and graph attention, and detects anomalies using forecasting errors.
- **GCAD** (Liu et al., 2025) infers dynamic Granger-causal graphs from the gradients of a forecasting model and identifies anomalies through deviations in the learned causal patterns.
- **COPOD** (Li et al., 2020) estimates empirical copula-based tail probabilities and assigns larger anomaly scores to statistically extreme observations.
- **HBOS** (Goldstein & Dengel, 2012) constructs a histogram for each variable and combines the resulting density estimates under a feature-independence assumption.
- **TimesNet** (Wu et al., 2023) transforms one-dimensional time series into period-dependent two-dimensional representations to capture intra- and inter-period variations.
- **CNN** (Munir et al., 2018) uses temporal convolutions to forecast future observations and detects anomalies through prediction errors.
- **USAD** (Audibert et al., 2020) employs adversarially trained Autoencoders to learn normal temporal patterns and scores anomalies using reconstruction errors.

Table 4: Anomaly detection performance comparison on SMAP and SMD.

Model	SMAP			SMD		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1
A-Transformer	0.1340 ± 0.0072	0.5065 ± 0.0071	0.2060 ± 0.0110	0.0720 ± 0.0046	0.5013 ± 0.0049	0.1260 ± 0.0071
FITS	0.1455 ± 0.0047	0.5147 ± 0.0068	0.2798 ± 0.0063	0.2863 ± 0.0026	0.8299 ± 0.0022	0.4025 ± 0.0026
TimesNet	0.1792 ± 0.0071	0.5336 ± 0.0058	0.3161 ± 0.0089	0.2839 ± 0.0053	0.8087 ± 0.0021	0.3713 ± 0.0049
COPOD	0.1927 ± 0.0000	0.5918 ± 0.0000	0.2841 ± 0.0000	0.2139 ± 0.0000	0.7193 ± 0.0000	0.2951 ± 0.0000
HBOS	0.1897 ± 0.0000	0.5650 ± 0.0000	0.2664 ± 0.0000	0.2678 ± 0.0000	0.7378 ± 0.0000	0.3556 ± 0.0000
TranAD	0.2122 ± 0.0042	0.6104 ± 0.0056	0.3264 ± 0.0066	0.3304 ± 0.0022	0.7691 ± 0.0022	0.4181 ± 0.0018
CNN	0.2136 ± 0.0029	0.6577 ± 0.0088	0.3471 ± 0.0055	0.3730 ± 0.0049	0.7921 ± 0.0042	0.4313 ± 0.0046
Autoencoder	0.2277 ± 0.0006	0.5859 ± 0.0015	0.2991 ± 0.0006	0.3613 ± 0.0050	0.7297 ± 0.0034	0.4214 ± 0.0034
USAD	0.2263 ± 0.0049	0.5850 ± 0.0089	0.3850 ± 0.0075	0.4105 ± 0.0040	0.8726 ± 0.0019	0.4810 ± 0.0021
OmniAnomaly	0.2420 ± 0.0003	0.5950 ± 0.0002	0.4065 ± 0.0003	0.4284 ± 0.0001	0.8822 ± 0.0001	0.5061 ± 0.0001
CATCH	0.2450 ± 0.0028	0.6078 ± 0.0046	0.3664 ± 0.0041	0.4763 ± 0.0011	**0.8968 ± 0.0004**	0.5057 ± 0.0012
GDN	0.2197 ± 0.0115	0.6099 ± 0.0070	0.3394 ± 0.0087	0.4006 ± 0.0157	0.8014 ± 0.0104	0.4640 ± 0.0140
GCAD	0.2144 ± 0.0135	0.6348 ± 0.0119	0.3471 ± 0.0104	0.3569 ± 0.0089	0.8416 ± 0.0074	0.4535 ± 0.0109
GRASP	**0.3149 ± 0.0087**	**0.7015 ± 0.0130**	**0.4178 ± 0.0064**	**0.4845 ± 0.0043**	0.8567 ± 0.0038	**0.5087 ± 0.0034**

- **TranAD** (Tuli et al., 2022) combines Transformer-based sequence modeling, self-conditioning, and adversarial training for MTS anomaly detection.
- **OmniAnomaly** (Su et al., 2019) combines recurrent temporal modeling with a variational Autoencoder to capture stochastic temporal dependencies and detects anomalies using reconstruction probabilities.
- **Autoencoder** (Sakurada & Yairi, 2014) learns a compressed representation of normal observations and uses reconstruction errors as anomaly scores.
- **A-Transformer** (Xu et al., 2022) models prior and learned temporal associations through anomaly attention and detects anomalies using the discrepancy between these associations.
- **FITS** (Xu et al., 2024) models time series through learnable interpolation of low-frequency components in the complex frequency domain.
- **CATCH** (Wu et al., 2025) divides frequency-domain representations into patches and captures spectral patterns and inter-channel dependencies through masked channel fusion.

E.3 Implementation Details

We implement GRASP in PyTorch and evaluate it on SMAP, SMD, CICIDS, and SWAN. For each time series, the normal training data are chronologically divided into 80% for training and 20% for validation. We use a window length of 50 and set the graph Dirichlet energy weight parameter to $\tau = 2$ for all datasets. For graph construction, we binarize the edge-weight matrix using a fixed threshold of 0.5 across all datasets.

The velocity predictor used in GRASP consists of two time-conditioned TSMixer blocks with a hidden dimension of 128 and a dropout rate of 0.1. Flow time is encoded with a fixed 64-dimensional sinusoidal embedding. The model is trained with a batch size of 256 and a learning rate of 10^{-3}. Training is limited to 1500 epochs, and validation is performed every 50 epochs. No anomaly labels are used for training, validation, graph construction, or anomaly score calibration.

During inference, anomaly scores are computed from the weighted velocity discrepancies aggregated across graph-frequency modes, source samples, and flow times. We compute the average of five source samples, and each prior sample is evaluated at ten flow times sampled equally between 0 and 1. All experiments are conducted on a single 80GB NVIDIA A100 GPU. We repeat each experiment using ten random seeds.

E.4 Additional MTS Anomaly Detection Results

In Section 4.1, we have presented the mean MTS anomaly detection performance over 10 random seeds in Table 1. To further assess the stability of the compared methods, we further report the results with standard deviation on the SMAP and SMD datasets in Table 4, and on the CICIDS and SWAN datasets in Table 5.

Table 5: Anomaly detection performance comparison on CICIDS and SWAN.

Model	CICIDS			SWAN		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1
A-Transformer	0.2396 ± 0.0007	0.4972 ± 0.0025	0.3060 ± 0.0000	0.2817 ± 0.0018	0.5021 ± 0.0014	0.4365 ± 0.0000
FITS	0.1849 ± 0.0013	0.4106 ± 0.0039	0.3126 ± 0.0013	0.3189 ± 0.0120	0.4607 ± 0.0120	0.4365 ± 0.0000
TimesNet	0.1905 ± 0.0021	0.4657 ± 0.0052	0.3121 ± 0.0015	0.3068 ± 0.0080	0.4505 ± 0.0137	0.4365 ± 0.0000
COPOD	0.2214 ± 0.0000	0.5566 ± 0.0000	0.3733 ± 0.0000	0.2792 ± 0.0000	0.5000 ± 0.0000	0.4365 ± 0.0000
HBOS	0.2603 ± 0.0000	0.5586 ± 0.0000	0.3852 ± 0.0000	0.2792 ± 0.0000	0.5000 ± 0.0000	0.4365 ± 0.0000
TranAD	0.2043 ± 0.0021	0.4704 ± 0.0085	0.3416 ± 0.0007	0.3360 ± 0.0020	0.4911 ± 0.0039	0.4365 ± 0.0001
CNN	0.2371 ± 0.0019	0.5543 ± 0.0056	0.3991 ± 0.0009	0.3760 ± 0.0056	0.4966 ± 0.0086	0.4365 ± 0.0000
Autoencoder	0.3147 ± 0.0089	0.7276 ± 0.0408	0.3938 ± 0.0106	0.3294 ± 0.0025	0.4944 ± 0.0091	0.4371 ± 0.0006
USAD	0.2094 ± 0.0007	0.4619 ± 0.0019	0.3525 ± 0.0003	0.4405 ± 0.0088	0.5585 ± 0.0095	0.4365 ± 0.0000
OmniAnomaly	0.2138 ± 0.0000	0.4501 ± 0.0000	0.3548 ± 0.0000	0.4847 ± 0.0002	0.6130 ± 0.0004	0.4365 ± 0.0000
CATCH	0.2866 ± 0.0007	0.6162 ± 0.0006	0.4023 ± 0.0008	0.4638 ± 0.0017	0.6218 ± 0.0018	0.4533 ± 0.0009
GDN	0.2926 ± 0.0188	0.7017 ± 0.0239	0.3805 ± 0.0148	0.6423 ± 0.0231	0.8356 ± 0.0091	0.6432 ± 0.0155
GCAD	0.3636 ± 0.0194	0.7191 ± 0.0151	0.3844 ± 0.0055	0.6445 ± 0.0146	0.8140 ± 0.0070	0.6231 ± 0.0089
GRASP	**0.3996 ± 0.0083**	**0.8108 ± 0.0072**	**0.4615 ± 0.0037**	**0.7556 ± 0.0116**	**0.8758 ± 0.0101**	**0.6886 ± 0.0145**

Table 6: Ablation results on CICIDS and SWAN.

Variant	CICIDS			SWAN		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1
GRASP	**0.3996**	**0.8108**	**0.4615**	**0.7556**	**0.8758**	**0.6886**
GRASP-mean	0.3764	0.7593	0.4499	0.7552	0.8753	0.6881
GRASP-ER	0.3734	0.7546	0.4497	0.7534	0.8723	0.6857
GRASP-lin	0.3740	0.7543	0.4493	0.7486	0.8714	0.6841

GRASP achieves the highest PRC and Best-F1 on all four datasets, as well as the highest ROC on three of them. On SMD, GRASP obtains the best PRC and Best-F1, while CATCH achieves the highest ROC. The relatively small standard deviations of GRASP indicate that its performance is stable across random seeds.

E.5 Additional Ablation Results

Table 6 extends the ablation study in Table 2 to CICIDS and SWAN. The GRASP model consistently achieves the best performance across all metrics. The improvements are particularly clear on CICIDS, whereas the gains on SWAN are smaller but remain consistent. Overall, replacing the graph-spectral path with a linear path, removing the graph-spectral weighting scheme, or using a randomly generated graph leads to performance degradation. These results validate the individual contribution of the three design components.

E.6 Comparison of Velocity-Field Architectures

In this section, we examine the effect of the velocity-field architecture by replacing the default TSMixer backbone of GRASP with two time-conditioned alternatives based on a GNN and a Transformer. The GNN treats sensor channels as nodes and applies two residual GCN blocks for cross-channel propagation (Kipf & Welling, 2017). The Transformer treats sensor channels as tokens and employs two encoder layers with four-head self-attention (Vaswani et al., 2017). All three architectures use a hidden dimension of 128 and are trained and evaluated under the same experimental protocol.

As shown in Table 7, TSMixer achieves the highest PRC, ROC, and Best-F1 on all four datasets. The advantage is particularly pronounced on SWAN, where its PRC reaches 0.7552, compared with 0.5306 for the GNN and 0.5594 for the Transformer. These results support the use of the simple yet effective TSMixer model in GRASP. Under the considered setting, the tested GNN and Transformer alternatives provide no performance improvement.

E.7 Additional Flow-Time and Graph-Frequency Analysis

Figure 5 extends the anomaly attribution analysis in Section 4.3 to SMD, CICIDS, and SWAN datasets. We decompose the squared velocity residuals across flow time and graph-frequency quan-

Table 7: Ablation study of velocity-field architectures.

Variant	SMAP			SMD			CICIDS			SWAN		
	PRC	ROC	Best-F1	PRC	ROC	Best-F1	PRC	ROC	Best-F1	PRC	ROC	Best-F1
TSMixer	**0.3149**	**0.7015**	**0.4178**	**0.4845**	**0.8567**	**0.5087**	**0.3996**	**0.8108**	**0.4615**	**0.7556**	**0.8758**	**0.6886**
GNN	0.2773	0.6758	0.3881	0.4659	0.8542	0.4926	0.3551	0.7232	0.4459	0.5306	0.6570	0.5188
Transformer	0.2639	0.6347	0.3795	0.4475	0.8272	0.4674	0.3283	0.6407	0.4331	0.5594	0.6758	0.5122

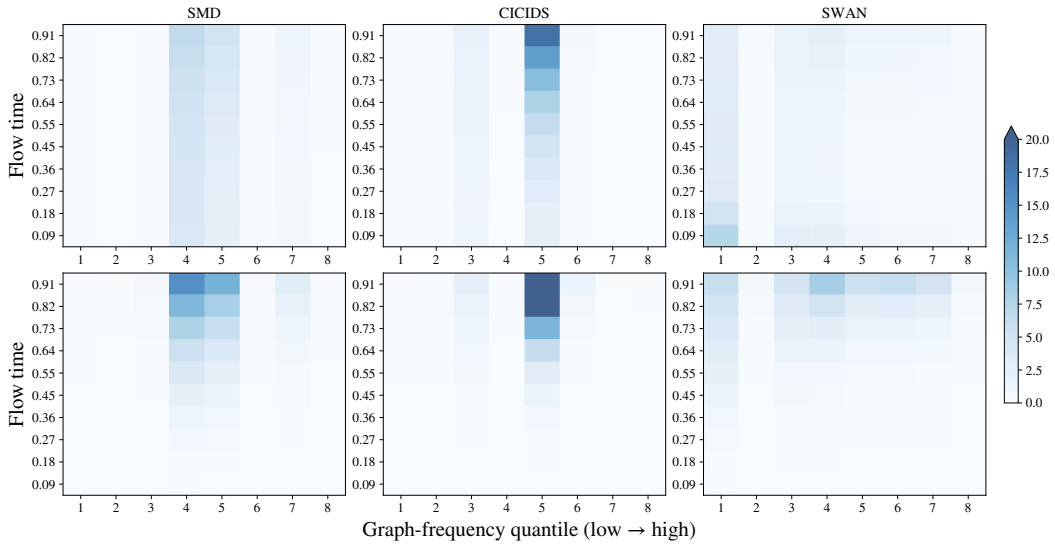


Figure 5: Anomaly score contributions across flow times and graph-frequency quantiles on SMD, CICIDS, and SWAN. Up: unweighted anomaly signal. Below: weighted anomaly signal.

tiles. The graph frequencies are ordered from low to high Laplacian eigenvalues, while larger flow times correspond to states closer to the data endpoint. The top row presents the unweighted residual contributions, and the bottom row presents the contributions after applying the weighting schedule.

On SMD and CICIDS, the anomaly evidence is primarily concentrated in intermediate graph-frequency quantiles at later flow times. Applying the score weights further shifts the attribution toward the data endpoint while preserving the dominant frequency regions. In contrast, SWAN exhibits a more diffuse attribution pattern across graph frequencies. These results suggest that the flow time toward the endpoint provides more informative anomaly signals for all datasets, while the informative graph-frequency components vary across datasets rather than being universally concentrated in the highest-frequency modes.

E.8 Sensitivity to Flow-Time Evaluations and Source Samples

In this section, we examine the sensitivity of GRASP to two inference-time hyperparameters: the number of flow-time evaluation points $|\mathcal{R}_{\text{ad}}|$ and the number of source samples $|\mathcal{X}_{\text{ad}}|$. We vary $|\mathcal{R}_{\text{ad}}| \in \{1, 2, 5, 10, 20, 50\}$ while fixing $|\mathcal{X}_{\text{ad}}| = 5$, and vary $|\mathcal{X}_{\text{ad}}| \in \{1, 2, 5, 10, 20\}$ while fixing $|\mathcal{R}_{\text{ad}}| = 10$. Figure 6 and Figure 7 report the mean and standard deviation over ten random seeds.

As shown in Figure 6, increasing the number of flow-time evaluation points generally improves performance on SMAP, SMD, and CICIDS, especially when $|\mathcal{R}_{\text{ad}}| < 10$. Performance on SWAN remains stable across different values of $|\mathcal{R}_{\text{ad}}|$. Figure 7 shows that GRASP is relatively insensitive to the number of source samples. Increasing $|\mathcal{X}_{\text{ad}}|$ yields modest improvements on SMAP and negligible changes on the other datasets. These results support the default choices of $|\mathcal{R}_{\text{ad}}| = 10$ and $|\mathcal{X}_{\text{ad}}| = 5$ as practical trade-offs between detection performance and inference cost.

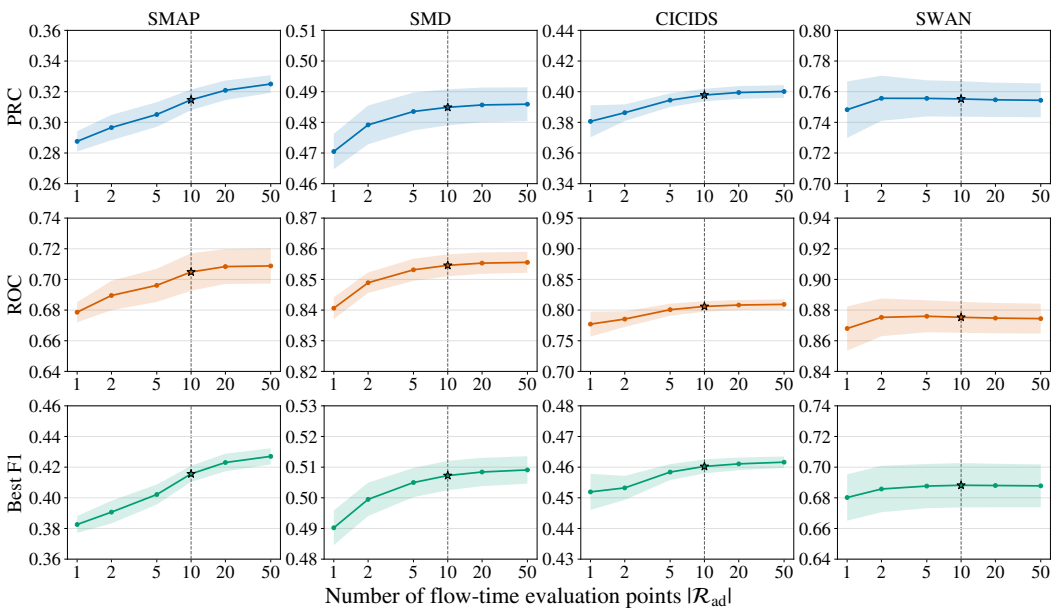


Figure 6: Sensitivity to the number of flow-time evaluation points $|\mathcal{R}_{\text{ad}}|$ with $|\mathcal{X}_{\text{ad}}| = 5$.

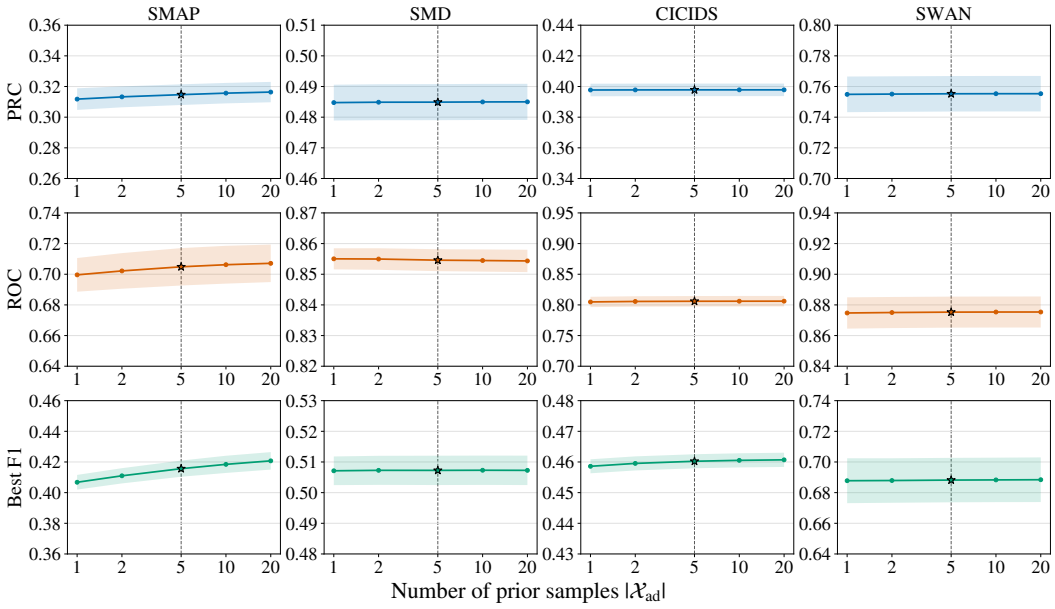


Figure 7: Sensitivity to the number of source samples $|\mathcal{X}_{\text{ad}}|$ with $|\mathcal{R}_{\text{ad}}| = 10$.

E.9 Sensitivity to the Graph Dirichlet-Energy Weight

The graph Dirichlet-energy weight τ controls the strength of the graph-structural regularization in the graph-spectral probability path, with $\tau = 0$ recovering the standard linear path. We evaluate $\tau \in \{0, 0.25, 0.5, 1, 2, 4, 8\}$ on all four datasets, retraining the model for each value while keeping all other hyperparameters fixed.

Figure 8 shows that the preferred value varies across datasets. SMD generally favors smaller values around $\tau = 0.5$, whereas SMAP and CICIDS attain their best PRC and Best-F1 near $\tau = 4$. SWAN performs best over an intermediate range around $\tau = 1$–2. Despite these dataset-specific

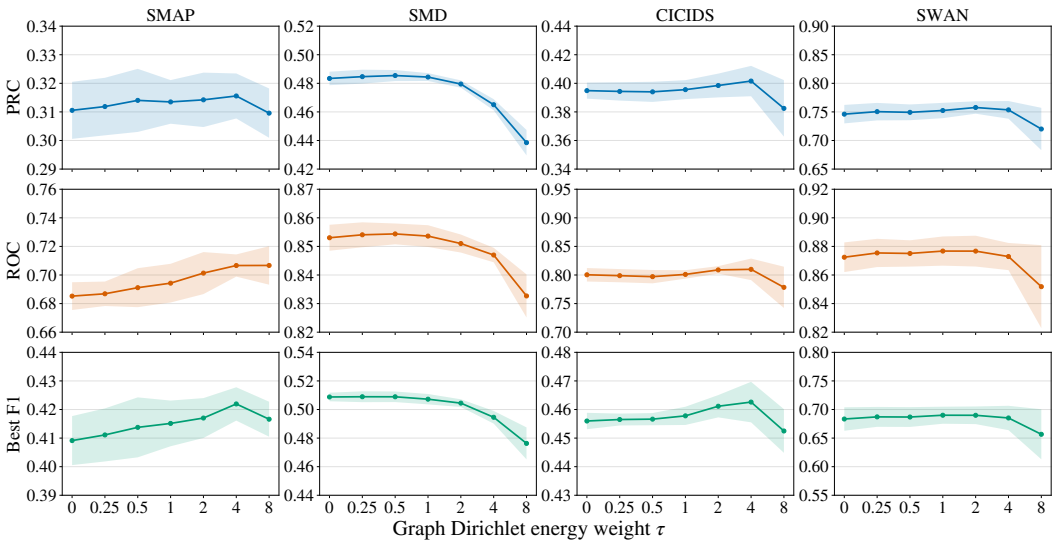


Figure 8: Sensitivity to the graph Dirichlet-energy weight τ.

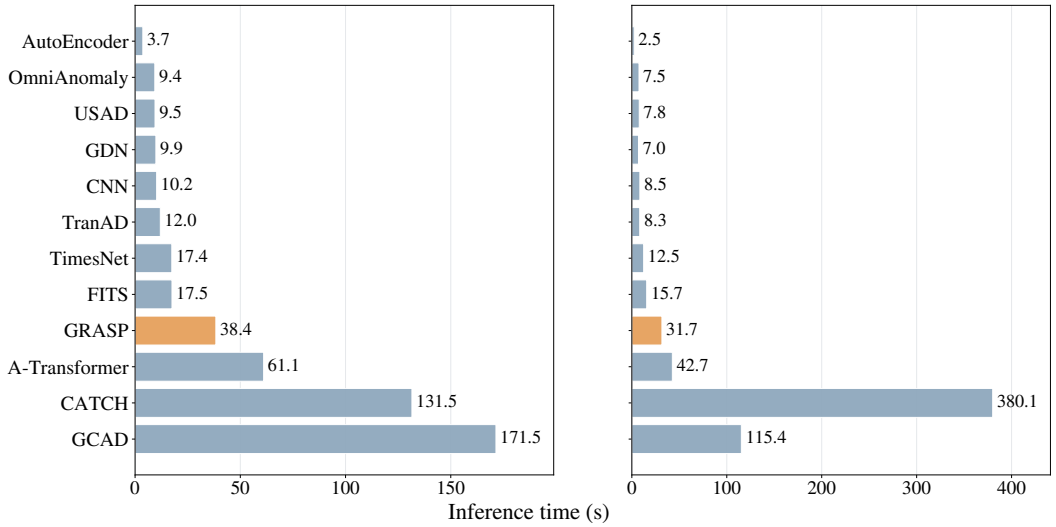


Figure 9: Inference time comparison on SMAP (left) and SMD (right).

optima, performance remains relatively stable for $\tau \in [0.5, 4]$. In contrast, $\tau = 8$ leads to a clear performance degradation, suggesting that excessively strong graph regularization could weaken the anomaly signal.

E.10 Inference Efficiency

In this section, we evaluate the inference efficiency of GRASP. Specifically, we compare the inference time of anomaly detectors on the test set of SMAP and SMD. For GRASP, we use the default setting with $|\mathcal{R}_{\text{ad}}| = 10$ flow-time evaluation points and $|\mathcal{X}_{\text{ad}}| = 5$ source samples. Note that the eigenvalue decomposition is not included in the inference time computation as it is only performed once for each dataset before training and reused during inference.

As shown in Figure 9, GRASP requires 38.4 seconds on SMAP and 31.7 seconds on SMD to score the full test sets. GRASP demonstrates faster inference compared with A-Transformer, CATCH, and GCAD on both datasets, but slower than the remaining baselines because its anomaly score ag-

gregates multiple flow-time evaluations and source samples. For application scenarios where faster inference is preferred, GRASP also provides a controllable trade-off between inference efficiency and detection performance. Specifically, as discussed in Section E.8, GRASP can be further accelerated by reducing the flow-time evaluation points $|\mathcal{R}_{\mathrm{ad}}|$ and source samples $|\mathcal{X}_{\mathrm{ad}}|$ with a bit of compromised performance.