CRGSTA: CROSS-DOMAIN ROOT CAUSAL GRAPH SPATIAL-TEMPORAL ATTENTION NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern monitoring systems generate massive, high-dimensional time series where failures rarely remain isolated but cascade across interdependent components. Identifying their true origins requires more than anomaly detection; it requires interpretable models that disentangle causal structure from noisy signals. While Granger causality has gained traction for root cause analysis (RCA), existing neural methods often rely on multilayer perceptrons applied independently at each time step, which increases parameter counts, struggles with long-range dependencies, and overlooks seasonal and periodic patterns. We introduce CrGSTA (Cross-domain Root causal Graph Spatial-Temporal Attention Network), a scalable and interpretable framework that unifies time- and frequency-domain representations through cross-domain attention. CrGSTA employs graph-based spatio-temporal attention to capture directional dependencies, while frequency-aware features recover periodic structure. A lightweight self-attention decoder reconstructs dynamics, ensuring deviations are attributed to true root causes rather than propagated effects. We conduct experiments along three dimensions: temporal scalability, spatial scalability, and ablations on domain contributions and fusion strategies. On both the Lotka–Volterra benchmark and the SWaT industrial dataset, CrGSTA new state of the art achieving up to 13% Avg@10 improvement by leveraging wider temporal windows with only 8.5M parameters compared to (200M+) of other baselines. By explicitly coupling temporal and frequency cues, CrGSTA balances accuracy, interpretability, and efficiency for RCA in complex monitoring environments, providing a foundation for more resilient and transparent analysis in real-world systems. `https://github.com/crgsta2025/CrGSTA`

1 INTRODUCTION

As digital infrastructures grow in scale and complexity, system failures are no longer isolated incidents but often trigger cascades of anomalies that spread across tightly coupled components Altenbernd et al. (2025). These anomalies, while infrequent, can severely disrupt application availability and compromise service reliability Nagalapatti et al. (2025). Traditional anomaly detection methods provide early warning signals, yet they fall short in answering the critical question of *why* the anomaly occurred Chen et al. (2019). Without this capability, operators face significant delays in recovery, leading to higher downtime and operational costs. Root cause analysis (RCA) addresses this gap by uncovering the underlying drivers of observed anomalies, disentangling direct causes from secondary effects, and enabling more targeted remediation Liu et al. (2023); Han et al. (2025). In complex cloud Nedelkoski et al. (2020) and cyber-physical environments Mathur & Tippenhauer (2016), where human monitoring alone is infeasible, automated RCA is essential for ensuring resilience and sustainable system management.

Root cause analysis (RCA) can be formally described as identifying, given a set of anomalous metrics, the top-K metrics most likely responsible for the anomaly Liu et al. (2023). Unlike anomaly detection, which merely signals abnormal behavior, RCA requires interpretability: models must reveal how components influence one another and propagate faults across the system. Achieving this using statisical methods Ikram et al. (2022); Shan et al. (2019)is particularly challenging in modern infrastructures, where a single incident may involve thousands of KPIs, rendering manual tracing

or heuristic correlations ineffective. Recent research has therefore shifted toward data-driven methods. Among them, neural Granger causality Granger (1969) has emerged as a principled tool for uncovering temporal dependencies between variables, offering a systematic way to infer directional relationships. However, contemporary neural Granger causality methods Han et al. (2025) typically rely on MLPs applied independently at each time step. Such architectures prevent the model from capturing spatial dependencies across metrics, limiting its explainability across system components. Moreover, the per-time-step design also constrains the temporal horizon the model can consider and causes a parameter explosion as system dimensionality grows. Additionally, these approaches fail to account for seasonal and periodic patterns, which are crucial for understanding recurring system behaviors. These limitations highlight the need for more advanced RCA frameworks that can jointly model spatial and temporal dependencies while remaining interpretable and scalable to high-dimensional, real-world datasets.

A promising direction for RCA is to represent time series from multiple perspectives. Frequency-domain transformations have been shown to reveal latent structures that remain obscured in the raw time domain Xu et al. (2024); Yi et al. (2025; 2023). Hybrid approaches that jointly leverage temporal and frequency representations have demonstrated strong performance in anomaly detection Dou et al. (2025); Bai et al. (2023a). Despite these advances, integrating interpretability, a critical requirement for RCA, into multi-domain representations remains largely unexplored. We posit that combining time and frequency perspectives while explicitly enforcing interpretability can significantly enhance RCA. By moving beyond single-domain limitations, such approaches are better equipped to uncover the underlying mechanisms of complex anomalies in high-dimensional, large-scale monitoring systems.

In this work, we propose CrGSTA (Cross-domain Root causal Graph Spatial-Temporal Attention Network), a scalable and interpretable framework for root cause analysis in multivariate time series. CrGSTA is grounded in Granger causality Han et al. (2025); Fu et al. (2024), enabling unsupervised modeling of normal system behavior and the identification of exogenous factors that drive anomalies. Inspired by prior work on neural Granger causality Han et al. (2025) and cross-domain time- and frequency representations Dou et al. (2025); Bai et al. (2023a), CrGSTA captures complementary patterns across domains while enhancing the interpretability of detected anomalies. CrGSTA employs a spatio-temporal encoder–decoder architecture. The encoder features parallel time- and frequency-domain paths, each applying spatial graph attention across time lags followed by temporal attention. Their outputs are integrated via cross-attention, producing interpretable latent representations that reveal exogenous influences. A lightweight self-attention decoder reconstructs the series, and deviations from the learned normal distribution during inference are flagged as potential root causes, distinguished from downstream effects. Overall, CrGSTA offers a principled and scalable framework for multi-domain RCA in complex, high-dimensional systems by unifying cross-domain representation learning, spatio-temporal attention, and Granger causal reasoning.

Our experiments demonstrate that CrGSTA establishes a new state of the art for root cause analysis in multivariate time series by jointly modeling temporal and frequency domains through a graph-based encoder–decoder. Across both synthetic and real-world datasets, CrGSTA consistently outperforms statistical, non-causal, and causal deep learning baselines, while preserving parameter efficiency. For instance, it achieves 0.782 Avg@10 on Lotka–Volterra and 0.426 on SWaT, surpassing prior methods by wide margins despite operating under a fixed budget of only 8M parameters—more than two orders of magnitude fewer than AERCA's 200M+. Ablation studies further highlight the indispensability of cross-domain integration and attention mechanisms, which together enable CrGSTA to capture complex spatio-temporal dependencies without the prohibitive computational overhead observed in existing causal models. These findings not only validate the effectiveness of CrGSTA's architectural design but also underscore its practicality for large-scale monitoring systems where efficiency and interpretability are critical. In doing so, CrGSTA advances root cause analysis beyond current trade-offs between accuracy and scalability, pointing toward a new generation of resource-efficient causal modeling frameworks for modern infrastructures.

This work is guided by the following research questions: **RQ1:** How does CrGSTA perform as the temporal window size increases, and how does it compare to statistical and deep learning baselines in terms of accuracy and parameter efficiency? **RQ2:** How does CrGSTA scale with the number of interacting variables, and how does its performance and parameter growth compare to other deep learning approaches? **RQ3:** What are the contributions of CrGSTA's architectural components and fusion strategies to its overall performance, and how do they impact parameter efficiency?

Our contributions are threefold: (1) We introduce CrGSTA, a novel unsupervised framework for root cause detection in multivariate time series that achieves a balance between scalability and interpretability, making it suitable for large-scale, complex real-world datasets. (2) We design a multi-path encoder–decoder architecture grounded in Granger causal reasoning, featuring parallel time- and frequency-domain paths. Spatial graph attention captures inter-variable dependencies, temporal self-attention models historical dynamics, and cross-attention fuses time- and frequency-domain representations, enabling the model to capture seasonality and periodic patterns. A lightweight self-attention decoder replaces conventional autoregressive stacks, resulting in substantial efficiency gains. (3) We perform extensive empirical evaluations on both synthetic and real-world datasets, systematically analyzing the impact of temporal and spatial dimensions as well as architectural choices, demonstrating the effectiveness and flexibility of CrGSTA in capturing complex causal relationships.

2 Related Work

Root cause analysis (RCA) in multivariate systems intersects with performance engineering, where the goal extends beyond anomaly detection to scalable, interpretable, and robust diagnostics.

2.1 Root Cause Analysis

RCA methods are broadly categorized into topology-driven, statistical, and causal inference–based approaches (Table 1). Topology-driven methods infer dependencies among variables and localize anomalies via graph traversal. For instance, MonitorRank Kim et al. (2013) scores service-level correlations using personalized PageRank Brin & Page (1998). While effective in structured environments, these methods often scale poorly in dynamic systems. Statistical techniques, in contrast, detect significant deviations in system metrics. ϵ-Diagnosis Shan et al. (2019) employs two-sample tests, whereas RCD Ikram et al. (2022) applies conditional independence tests to infer causal structures. Although efficient and interpretable, these methods struggle with complex anomalies. Data-driven approaches directly learn temporal and spatial dependencies from multivariate observations Han et al. (2025); Tuli et al. (2022), and causal inference–based methods treat anomalies as interventions in structural causal models Assaad et al. (2022). For example, AERCA Han et al. (2025) leverages autoencoders to capture Granger causal dependencies. However, many existing designs rely on shallow parameterizations (e.g., MLP-based causal coefficients), limiting robustness in complex systems.

2.2 Orthogonal Advances in Temporal Modeling

Recent progress emphasizes lightweight yet expressive architectures, ranging from linear attention blocks to compact Transformers Tan et al. (2024); Liu et al. (2024). Frequency-domain methods have also proven highly efficient; for example, a 10K-parameter Fourier model matched the performance of a 300M-parameter Transformer Zhou et al. (2022); Xu et al. (2024), inspiring models such as FilterNet Yi et al. (2025), FourierGNN Yi et al. (2023), and FreqTimeLoss Wang et al. (2025). Cross-domain architectures further enhance robustness by jointly leveraging temporal and spectral representations. CrossFuN Bai et al. (2023a) fuses temporal and spectral views, while DeAnomaly Dou et al. (2025) combines graph attention with time–frequency cross-attention to handle noisy multivariate data. These multi-domain approaches provide richer inductive biases than single-domain methods. Despite these advances, most anomaly detection models lack interpretability, and existing RCA approaches often rely on MLP-based Granger causality approximations that scale poorly and neglect temporal expressiveness. To address this gap, we propose CrGSTA, a spatio-temporal encoder–decoder that integrates time- and frequency-domain representations with graph-based causal reasoning, capturing long-range temporal dependencies and spatial interactions for scalable, interpretable RCA in complex multivariate systems.

3 Preliminaries and Problem Formulation

Root cause analysis (RCA) in multivariate time series aims to identify latent factors driving observed variables. Granger causality Granger (1969) formalizes this: for a d-dimensional series $\{x_t\}_{t=1}^{T}$, each component $x_t^{(j)}$ can be expressed as a function of past values plus an unexplained latent input

$z_t^{(j)}$,

$$x_t^{(j)} = f^{(j)}\big(x_{\le t-1}^{(1)}, \ldots, x_{\le t-1}^{(d)}\big) + z_t^{(j)}.$$

Here, $x^{(i)}$ Granger-causes $x^{(j)}$ if including its history improves prediction beyond $x^{(j)}$'s own past.

In an encoder–decoder view, the encoder extracts latent exogenous variables z_t by removing predictable components, producing an interpretable representation of unexpected influences. The decoder reconstructs observations from these latent variables, ensuring consistency with the generative process. Formally, with $\mathbf{z}_t \in \mathbb{R}^d$ and $\mathbf{x}_t \in \mathbb{R}^p$, the marginal likelihood is

$$\log P(\mathbf{x}_t) = \log \int P(\mathbf{x}_t \mid \mathbf{z}_t, A(t))\, P(\mathbf{z}_t)\, d\mathbf{z}_t,$$

where $A(t)$ encodes instantaneous causal structure. The intractable posterior $P(\mathbf{z}_t \mid \mathbf{x}_t)$ is approximated by a variational distribution $E_\phi(\mathbf{z}_t \mid \mathbf{x}_{\le t-1})$, yielding a VAE-like framework Kingma & Welling (2014). Graph attention captures cross-variable dependencies, temporal attention models sequential dynamics, and optional frequency-domain transformations reveal hidden patterns that improve interpretability.

RCA then identifies indices (j, t) where latent variables deviate due to anomalies, $\hat{z}_t^{(j)} = z_t^{(j)} + \epsilon_t^{(j)}$. Unlike standard anomaly detection, the focus is on the sources of abnormal behavior.

3.1 CrGSTA with Time-Frequency Cross-Attention

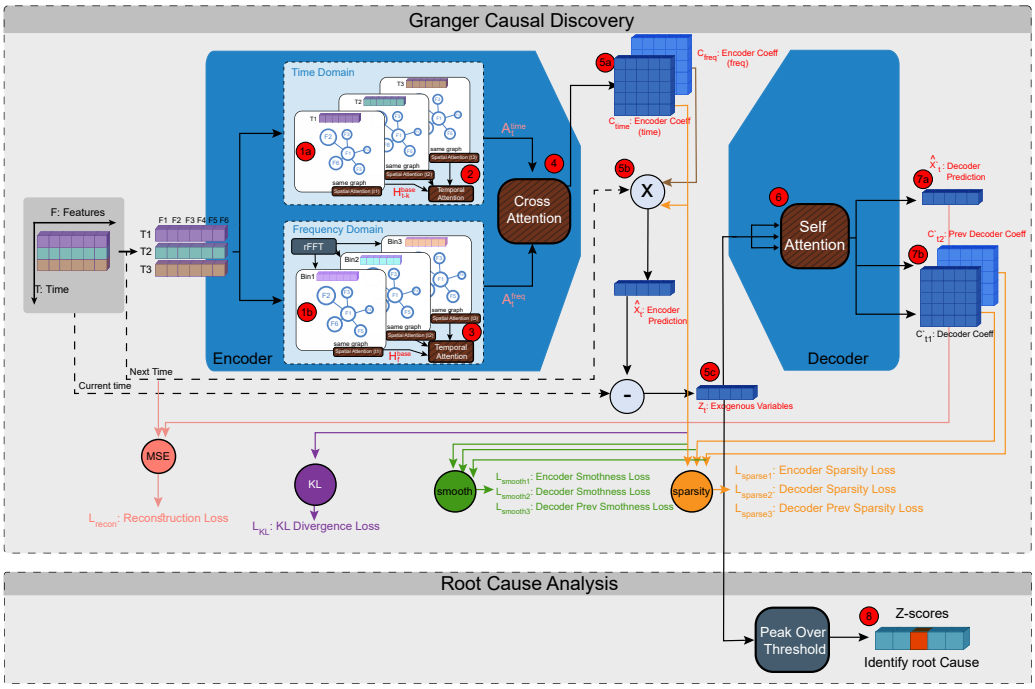


Figure 1: CrGSTA: Time-Frequency Cross-Attention Graph Spatio-Temporal Autoencoder

We present **CrGSTA** as a time-frequency cross-attention graph-based encoder-decoder for multivariate root cause identification, as illustrated in Fig. 1. The encoder estimates latent exogenous variables $E_\phi(\mathbf{z}_t \mid \mathbf{x}_{\le t})$, while the decoder reconstructs the observation $\mathbf{x}_t$ given past exogenous sequences $D_\theta(\mathbf{x}_t \mid \mathbf{z}_{\le t})$.

3.1.1 Encoder Structure

Windowing Time Series. Given $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ with d variables, we define sliding windows of length K: $\mathbf{W}_t = (\mathbf{x}_{t-K+1}, \ldots, \mathbf{x}_t)$, $\mathbf{W} = (\mathbf{W}_K, \ldots, \mathbf{W}_T)$, so each window is processed to capture both temporal and spatial dependencies.

Step 1: Base Spatial Graph (Shared Across Lags and Branches). We define a global, shared graph attention network (GNN) to compute pairwise influence between variables. Each variable in a time step forms a node in a fully-connected graph. This shared graph serves as the foundation for both the time-domain and frequency-domain branches:

$$\mathbf{H}_{t-k}^{\text{base}} = \text{GNN}(\mathbf{x}_{t-k}) \in \mathbb{R}^{d\times d}, \quad k = 1, \ldots, K \tag{1}$$

This design reduces parameter redundancy and ensures consistent modeling of interactions across domains.

Step 2: Time-Domain Branch. Using the shared base graph network, we apply temporal attention across lags to dynamically weight contributions of past observations:

$$\mathbf{A}_t^{\text{time}} = \text{TemporalAttn}([\mathbf{H}_{t-1}^{\text{base}}, \ldots, \mathbf{H}_{t-K}^{\text{base}}]) \in \mathbb{R}^{K\times d\times d}. \tag{2}$$

Step 3: Frequency-Domain Branch. The shared base graph is also leveraged to capture frequency-domain dependencies. First, a real FFT is applied along the temporal axis to extract periodic components, yielding $\mathbf{X}_f^{\text{freq}} = \text{rFFT}(\mathbf{W}_t)_f$ for $f = 1, \ldots, F$. The magnitudes of these frequency bins are then propagated through the shared graph network, followed by temporal attention across frequency bins:

$$\mathbf{H}_f^{\text{freq}} = \text{GNN}(|\mathbf{X}_f^{\text{freq}}|), \quad \mathbf{A}_t^{\text{freq}} = \text{TemporalAttn}([\mathbf{H}_1^{\text{freq}}, \ldots, \mathbf{H}_F^{\text{freq}}]) \in \mathbb{R}^{F\times d\times d}. \tag{3}$$

Step 4: Cross-Attention Fusion. After obtaining temporal and spectral representations, we introduce explicit information exchange between the two modalities. Two cross-attention modules are employed: one aligns frequency features with temporal context (time→freq), while the other aligns temporal features with spectral context (freq→time). This bi-directional interaction yields the enriched representations $\tilde{\mathbf{H}}^{\text{time}}$ and $\tilde{\mathbf{H}}^{\text{freq}}$:

$$\tilde{\mathbf{H}}^{\text{time}} = \text{CrossAttn}(\mathbf{A}_t^{\text{time}}, \mathbf{A}_t^{\text{freq}}), \quad \tilde{\mathbf{H}}^{\text{freq}} = \text{CrossAttn}(\mathbf{A}_t^{\text{freq}}, \mathbf{A}_t^{\text{time}}). \tag{4}$$

Step 5: Coefficient Projection and Prediction. The cross-attended representations from Step 4 are projected through linear layers into adjacency-like coefficient matrices (step 5a), yielding

$$\mathbf{C}_{\text{time}} = \text{Linear}(\tilde{\mathbf{H}}^{\text{time}}), \quad \mathbf{C}_{\text{freq}} = \text{Linear}(\tilde{\mathbf{H}}^{\text{freq}}),$$

which encode variable-to-variable dependencies across lags k. Empirically, we find that constraining the *time-domain* coefficients is sufficient for stable optimization of the loss functions. Nevertheless, both the time and frequency coefficients contribute to autoregressive prediction (step 5b):

$$\hat{\mathbf{x}}_{\text{time}} = \sum_{k=1}^{K} \mathbf{C}_{\text{time}}\, \mathbf{x}_{t-k}, \quad \hat{\mathbf{x}}_{\text{freq}} = \sum_{k=1}^{K} \mathbf{C}_{\text{freq}}\, \mathbf{x}_{t-k}, \tag{5}$$

where $\mathbf{x}_{t-k}$ represents the historical observations within the input window. These modality-specific predictions are then combined linearly to produce the next-step prediction, which is also used to compute the residual relative to the current observation (step 5c):

$$\hat{\mathbf{x}}_t = \omega_t \hat{\mathbf{x}}_{\text{time}} + \omega_f \hat{\mathbf{x}}_{\text{freq}}, \quad \mathbf{z}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t, \tag{6}$$

where ω_t and ω_f are the weights for combining both domains, and $\mathbf{z}_t$ is interpreted as a latent exogenous influence, capturing variability that is not explained by the temporal–spectral dynamics.

Encoder Output. In summary, the encoder produces two distinct outputs, each serving a specific purpose:

1. **Time-domain coefficients:** $\mathbf{C}_{\text{time}}$, which encode variable-to-variable dependencies and are directly used in the loss functions. These coefficients provide interpretability within the Granger-causal framework, as detailed in the subsequent sections.

2. **Latent exogenous variables:** $\mathbf{Z}_t \in \mathbb{R}^{d\times K}$, capturing influences not explained by the temporal–spectral dynamics, and serving as input to the decoder for reconstruction tasks.

3.1.2 DECODER STRUCTURE

The decoder reconstructs $\mathbf{x}_t$ from the exogenous sequence $\mathbf{Z}_t$ using a temporal-attention-based mechanism, avoiding fully autoregressive reconstruction.

Step 6: Projection and Windowed Attention. Each exogenous variable in the window is projected to a hidden representation $\mathbf{H}^{\text{enc}}_{t-K+\tau} = f_{\text{proj}}(\mathbf{z}_{t-K+\tau}), \quad \tau = 1, \ldots, K$, which are then aggregated via temporal attention across the window:

$$\mathbf{H}^{\text{temp}}_t = \text{TemporalAttn}(\mathbf{H}^{\text{enc}}_{t-K+1:t}), \tag{7}$$

producing a context-aware embedding for reconstruction.

Step 7: Output and Low-Rank Coefficients. The final prediction is obtained via a learnable output mapping $\hat{\mathbf{x}}_t = f_{\text{out}}(\mathbf{H}^{\text{temp}}_t)$, moreover generating low-rank coefficient matrices for interpretability:

$$\mathbf{C}_t = \mathbf{U}\mathbf{V}^\top, \quad \mathbf{C}_t \in \mathbb{R}^{d\times d}. \tag{8}$$

This structure efficiently captures temporal dependencies in the exogenous sequence while supporting interpretable causal attributions without maintaining separate decoders for past windows.

3.1.3 TRAINING OBJECTIVE

The encoder-decoder model is

$$\hat{\mathbf{x}}_t = \text{CrGSTA}_{\theta,\phi}(\mathbf{x}_{<t}),$$

with encoder parameters θ and decoder parameters ϕ. For a series of length T, the training objective combines reconstruction, regularization and independence:

Reconstruction Loss: encourages the model to reconstruct the current step from latent exogenous variables:

$$\mathcal{L}_{\text{recon}} = \sum_{t=K+1}^{T} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \tag{9}$$

Sparsity & Smoothness: promote interpretable coefficient matrices in encoder and decoder:

$$\mathcal{L}_{\text{sparse}} = \lambda_{\text{enc}} R(\mathbf{\Omega}_t) + \lambda_{\text{dec}}\big(R(\bar{\mathbf{\Omega}}_t) + R(\bar{\mathbf{\Omega}}'_t)\big), \tag{10}$$

$$\mathcal{L}_{\text{smooth}} = \gamma_{\text{enc}} S(\mathbf{\Omega}_{t+1}, \mathbf{\Omega}_t) + \gamma_{\text{dec}}\big(S(\bar{\mathbf{\Omega}}_{t+1}, \bar{\mathbf{\Omega}}_t) + S(\bar{\mathbf{\Omega}}'_{t+1}, \bar{\mathbf{\Omega}}'_t)\big) \tag{11}$$

where R denotes sparsity penalties and S encourages temporal smoothness of coefficients.

Exogenous Independence (KL): encourages the latent exogenous variables Z_t to be decorrelated and standardized:

$$\mathcal{L}_{\text{KL}} = \beta\, D_{\text{KL}}(P(\mathbf{Z}_t) \,\|\, Q) = \frac{1}{2}\big(\text{tr}(\Sigma_t) + \mu_t^\top \mu_t - d - \log\det \Sigma_t\big) \tag{12}$$

where Q is an isotropic Gaussian prior.

Total Objective: the sum of all components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{sparse}} + \mathcal{L}_{\text{smooth}} + \mathcal{L}_{\text{KL}} \tag{13}$$

This formulation preserves interpretability, enforces latent independence, and supports the single-decoder CrGSTA architecture in reconstructing the time series while highlighting causal attributions.

3.2 ROOT CAUSE LOCALIZATION

Step 8: Obtaining Root Causes: During deployment, for a new observation $\mathbf{x}_{t^*}$, its exogenous representation $\mathbf{z}_{t^*}$ is computed using the trained encoder. Standardized scores (z-scores) are calculated:

$$z^{(j)}_{t^*} = \frac{z^{(j)}_{t^*} - \mu^{(j)}}{\sigma^{(j)}}, \tag{14}$$

and variables exceeding an adaptive threshold (via SPOT) are flagged as potential root causes.

4 EXPERIMENTS

4.1 DATASETS

To evaluate the effectiveness of the CrGSTA framework, we conduct experiments on two datasets: a synthetic benchmark and a widely used real-world multivariate time series dataset (Table 2, more details in appendix A.2.1). We extend the Lotka–Volterra model Marcinkevičs & Vogt (2021) by increasing nonlinearity and stochastic variability, making anomaly detection more challenging; full details of the extended model are provided in the appendix A.2.1. Its controlled complexity allows for rigorous testing of root cause analysis methods under known causal structures. The SWaT (Secure Water Treatment) dataset Mathur & Tippenhauer (2016) is collected from a fully operational water treatment testbed, encompassing both normal operating conditions and attack scenarios. This dataset has become a standard benchmark for evaluating anomaly detection and root cause analysis in cyber-physical systems.

4.2 EXPERIMENTAL SETUP

Baselines and Comparison. We benchmark CrGSTA against statistical, non-causal, and causal deep learning approaches for root cause analysis. ϵ-Diagnosis Shan et al. (2019) detects root causes via pairwise significance tests, while RCD Ikram et al. (2022) constructs partial causal graphs to identify influential anomaly sources. Among non-causal models, FEDformer Zhou et al. (2022) and iTransformer Liu et al. (2024) leverage frequency-enhanced or dual-domain attention for forecasting, here adapted to root cause analysis by ranking variables via reconstruction errors. For causal deep learning, AERCA Han et al. (2025) employs lag-specific and stacked MLPs for autoregressive reconstruction. In contrast, CrGSTA integrates temporal dependencies through a recurrent attention-based GNN encoder and reconstructs causal dynamics via a self-attention decoder. We further ablate CrGSTA by varying domain inputs (temporal, frequency, or both), feature representations (magnitude vs. magnitude–phase), and fusion mechanisms (sum, concat, gated, attention).

Evaluation Metrics: We evaluate root cause identification using the *recall at top-k* metric (AC@k) and its average variant (Avg@k), following prior work Ikram et al. (2022); Li et al. (2022b). This measures the likelihood that true root causes appear among the top-k ranked variables. Sequences with multiple interventions are treated as single root cause sequences, consistent with point-adjust evaluation Koh et al. (2025); Bai et al. (2023b). Formal definitions are provided in the Appendix A.2.2. We also report the number of trainable parameters to assess efficiency, particularly for encoder–decoder models.

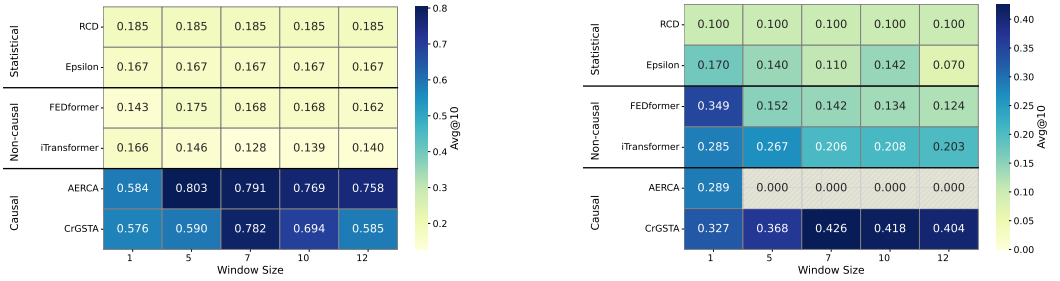
Implementation: We train two CrGSTA variants, differing only in spatial–temporal attention dimension (32 for Lotka–Volterra, 256 for SWaT), with 2 attention heads in both cases. The decoder is identical, using a lightweight self-attention layer with 64 hidden dimensions and 2 heads. Models are optimized with Adam (lr = 0.0001). Each experiment is repeated with multiple random seeds, and averages with standard deviations (reported in the appendix) ensure robustness. Experiments are run on a Linux workstation with an Intel i9-10900K CPU (20 cores, 3.70GHz), 32 GB RAM, and an NVIDIA RTX 3070 GPU (8 GB), using Python 3.10.12, PyTorch 2.7.1+cu126, and PyTorch Geometric 2.6.1. More details are in the Appendix A.3.

4.3 RQ1: PERFORMANCE IN TEMPORAL DIMENSION

We evaluate CrGSTA's temporal scalability by varying the input window size, fixing the number of interacting variables to 40 for Lotka–Volterra and using all 51 variables for SWaT. Results are presented in Fig. 2 and summarized in Tables 8, 9, with parameter efficiency shown in Fig. 5 in the Appendix. **Lotka–Volterra.** Statistical methods remain flat (Avg@10 $\approx$0.16–0.18), underscoring their inability to capture nonlinear dependencies. Non-causal deep models show mild temporal sensitivity but quickly saturate: FEDformer peaks at window size 5 (0.175), while iTransformer reaches a similar maximum at window size 1 (0.166). In contrast, causal modeling yields substantial gains: AERCA improves from 0.584 (window 1) to 0.803 (window 5), but this comes with near-linear parameter growth (0.3M $\rightarrow$ 3.1M), as shown in Fig. 5. CrGSTA achieves the best accuracy (0.782 at window 7) under a fixed parameter budget, with improvements attributable to cross-domain tempo-

ral modeling rather than sheer model size. **SWaT.** A similar pattern emerges. Statistical baselines remain below 0.2, while non-causal deep models reach only 0.315–0.334 without exhibiting scalability. AERCA again benefits from causal modeling but grows to over 100M parameters, making deployment impractical. CrGSTA reaches 0.426 at window 7—the best overall—while maintaining efficiency through cross-attention over medium-range dependencies. Notably, AERCA could in principle gain further performance with longer windows, but at the prohibitive cost of hundreds of millions of parameters. In contrast, CrGSTA preserves a stable parameter count across window sizes: for example, at window 7, CrGSTA uses only 8.5M parameters compared to AERCA's 200M+, a two-orders-of-magnitude reduction.

Summary. Statistical models fail to exploit temporal information; non-causal deep models capture limited temporal effects but saturate; causal models such as AERCA improve accuracy but incur prohibitive parameter costs. CrGSTA breaks this trade-off, achieving causal-level performance with stable parameterization, thereby highlighting the role of temporal–frequency interaction modeling in scalable root cause analysis.



(a) Lotka Volterra Avg@10 (b) SWAT Avg@10

Figure 2: Performance (Avg@10) for Lotka Volterra (left) and SWAT (right).

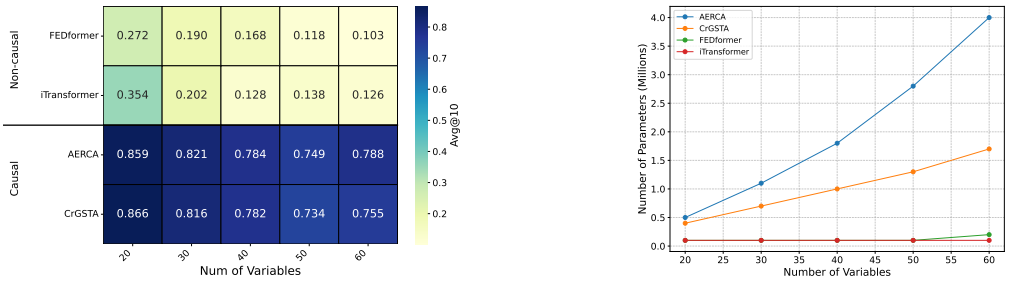
4.4 RQ2: Performance in the Spatial Dimension

To assess CrGSTA's spatial scalability, we evaluate its performance on the synthetic Lotka–Volterra dataset, fixing the temporal window to 7 and varying the number of variables from 20 to 60. Results are summarized in Fig. 3 and Table 7 in the appendix. **Causal vs. Non-Causal Models.** Across all variable counts, causal models (AERCA, CrGSTA) substantially outperform non-causal baselines (iTransformer, FEDformer). With 20 variables, CrGSTA achieves the highest Avg@10 of 0.866, followed by AERCA at 0.859, while non-causal models lag far behind (iTransformer 0.354, FEDformer 0.272). At 60 variables, CrGSTA maintains strong performance (0.755 Avg@10), whereas non-causal models degrade sharply (iTransformer 0.126, FEDformer 0.103), underscoring the importance of causal modeling in high-dimensional settings. **Parameter-Efficient Causal Performance.** CrGSTA delivers competitive accuracy with far fewer parameters than AERCA. At 20 variables, it reaches 0.866 Avg@10 with 0.4M parameters, slightly surpassing AERCA's 0.859 with 0.5M. At 50 variables, CrGSTA attains 0.734 with 1.3M, compared to AERCA's 0.749 with 2.8M. Even at 60 variables, it sustains 0.755 with 1.7M, while AERCA reaches 0.788 but requires 4.0M. Importantly, CrGSTA's parameter growth stems only from the incremental adapters added per variable, while the attention dimension remains fixed, ensuring scalability without architectural inflation.

Summary. On Lotka–Volterra, CrGSTA shows robust spatial scalability and parameter efficiency, maintaining over 75% Avg@10 with 60 variables while using less than half the parameters of AERCA, highlighting its effectiveness for complex, high-dimensional systems.

4.5 RQ3: Ablation Studies

We evaluate CrGSTA's components by varying spatial architectures and fusion strategies, fixing the temporal window to 7 and using 40 variables for Lotka–Volterra. To highlight architectural differences, we set the attention dimension to 32 on Lotka–Volterra and 256 on SWaT. Results and details on the ablation configurations are shown in Fig. 4 and Tables 8, 9 in the Appendix. **Spatial Architectures.** On Lotka–Volterra, temporal-only models (T) perform well (Avg@10=0.546). On SWaT,



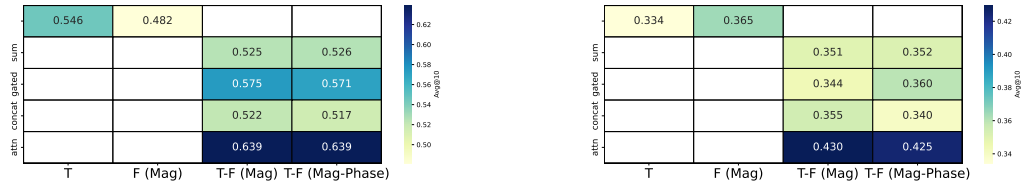
(a) Lotka Volterra Avg@10

(b) Lotka Volterra Num of Variables

Figure 3: Performance (Avg@10) and parameter scaling for Lotka Volterra (left) and SWAT (right). Heatmaps on top, parameters below.

however, frequency-only models (F) surpass temporal-only ones (0.365 vs. 0.334), highlighting the importance of frequency features in complex systems. **Fusion Strategies.** For synthetic data, combining temporal and frequency features (T-F) with simple fusion (sum, concat, gated) gives moderate gains (Avg@10=0.525–0.575). On SWaT, these methods underperform frequency-only models (0.334–0.360), suggesting naive fusion adds redundancy. By contrast, CrGSTA's cross-domain attention (T-F with attn) achieves the best results on both datasets (0.639 for Lotka–Volterra, 0.430 for SWaT), showing the effectiveness of adaptive integration. **Magnitude vs. Magnitude–Phase.** Magnitude-only features often match or outperform magnitude–phase. On Lotka–Volterra, both achieve Avg@10=0.639. On SWaT, magnitude-only slightly outperforms (0.430 vs. 0.425), suggesting phase may add noise slight in complex data. **Parameter Efficiency.** CrGSTA with attention fusion is compact (1.0M params on Lotka–Volterra, 8.5M on SWaT) compared to concat (5.5M and 21.5M+), confirming that gains stem from cross-domain design rather than size.

Summary. CrGSTA's strengths come from attention and cross-domain integration, enabling accurate and efficient root cause analysis.



(a) Lotka Volterra Ablations Avg@10

(b) SWAT Ablations Avg@10

Figure 4: Architectural and Combinatorial Ablations for Lotka Volterra (a) and SWAT (b).

5 Conclusion

We introduced CrGSTA, a novel framework for root cause analysis in multivariate time series that effectively integrates temporal and frequency domain information through a graph-based encoder-decoder architecture with cross-attention. Extensive experiments on both synthetic and real-world datasets demonstrate that CrGSTA consistently outperforms statistical methods, non-causal models, and other causal deep learning baselines in terms of accuracy and scalability. Ablation studies further highlight the critical role of attention mechanisms, cross-domain integration, and architectural design in enabling precise root cause identification. Importantly, CrGSTA achieves these gains while maintaining parameter efficiency, making it well-suited for practical deployment, whereas other causal models often entail prohibitive computational costs. For future work, we plan to explore extending CrGSTA with *state space models* such as Mamba to complement or replace attention mechanisms, which could enhance long-horizon temporal reasoning and mitigate quadratic scaling. We also aim to investigate integrating multimodal data sources, such as metrics and logs, and how to overcome the challenges of combining these heterogeneous signals for effective root cause analysis.

REFERENCES

Anton Altenbernd, Zhiyuan Wu, and Odej Kao. Amocrca: At most one change segmentation and relative correlation ranking for root cause analysis. In *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*, FSE Companion '25, pp. 1386–1393, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712760. doi: 10.1145/3696630.3731612. URL https://doi.org/10.1145/3696630.3731612.

C. K. Assaad, E. Devijver, and E. Gaussier. Discovery of extended summary graphs in time series. In J. Cussens and K. Zhang (eds.), *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI 2022)*, volume 180 of *Proceedings of Machine Learning Research*, pp. 96–106, Eindhoven, The Netherlands, Aug 1–5 2022. PMLR. URL https://proceedings.mlr.press/v180/assaad22a.html.

Yunfei Bai, Jing Wang, Xueer Zhang, Xiangtai Miao, and Youfang Lin. Crossfun: Multiview joint cross-fusion network for time-series anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, 72:1–9, 2023a. doi: 10.1109/TIM.2023.3315420.

Yunfei Bai, Jing Wang, Xueer Zhang, Xiangtai Miao, and Youfang Lin. Crossfun: Multiview joint cross-fusion network for time-series anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, 72:1–9, 2023b. doi: 10.1109/TIM.2023.3315420.

Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

Junjie Chen, Xiaoting He, Qingwei Lin, Yong Xu, Hongyu Zhang, Dan Hao, Feng Gao, Zhangwei Xu, Yingnong Dang, and Dongmei Zhang. An empirical investigation of incident triage for online service systems. In *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pp. 111–120. IEEE, 2019.

Hui Dou, Pengcheng Shi, Yiwen Zhang, Pengfei Chen, and Zibin Zheng. Deanomaly: Anomaly detection for multivariate time series using robust decomposition and memory-augmented diffusion models. *IEEE Transactions on Instrumentation and Measurement*, 74:1–14, 2025. doi: 10.1109/TIM.2025.3570337.

Dongqi Fu, Yada Zhu, Hanghang Tong, Kommy Weldemariam, Onkar Bhardwaj, and Jingrui He. Generating fine-grained causality in climate time series data for forecasting and anomaly detection. In *ICML 2024 AI for Science Workshop*, 2024. URL https://openreview.net/forum?id=q6E14hueUt.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

Xiao Han, Saima Absar, Lu Zhang, and Shuhan Yuan. Root cause analysis of anomalies in multivariate time series through granger causal discovery. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=k38Th3x4d9.

Azam Ikram, Sarthak Chakraborty, Subrata Mitra, Shiv Saini, Saurabh Bagchi, and Murat Kocaoglu. Root cause analysis of failures in microservices through causal discovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 31158–31170, 2022.

Myunghwan Kim, Roshan Sumbaly, and Sam Shah. Root cause detection in a service-oriented architecture. *ACM SIGMETRICS Performance Evaluation Review*, 41(1):93–104, 2013.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, Conference Track Proceedings*, 2014. URL http://arxiv.org/abs/1312.6114.

Van Kwan Zhi Koh, Ye Li, Ehsan Shafiee, Zhiping Lin, and Bihan Wen. Harnessing forecast uncertainty in deep learning for time series anomaly detection with posterior distribution scoring. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2025. doi: 10.1109/ISCAS56072.2025.11043371.

Douglas Landsittel, Avantika Srivastava, and Kristin Kropf. A narrative review of methods for causal inference and associated educational resources. *Quality Management in Health Care*, 29 (4):260–269, 2020.

Mingjie Li, Zeyan Li, Kanglin Yin, Xiaohui Nie, Wenchi Zhang, Kaixin Sui, and Dan Pei. Causal inference-based root cause analysis for online service systems with intervention recognition. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22)*, pp. 3230–3240. ACM, 2022a.

Mingjie Li, Zeyan Li, Kanglin Yin, Xiaohui Nie, Wenchi Zhang, Kaixin Sui, and Dan Pei. Causal inference-based root cause analysis for online service systems with intervention recognition. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3230–3240. ACM, 2022b.

Chenghao Liu, Wenzhuo Yang, Himanshu Mittal, Manpreet Singh, Doyen Sahoo, and Steven CH Hoi. Pyrca: A library for metric-based root cause analysis. *arXiv preprint arXiv:2306.11417*, 2023.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.

Ricards Marcinkevičs and Julia E Vogt. Interpretable models for granger causality using self-explaining neural networks. *arXiv preprint arXiv:2101.07600*, 2021.

Aditya P Mathur and Nils O Tippenhauer. Swat: A water treatment testbed for research and training on ics security. In *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater)*, pp. 31–36. IEEE, 2016.

Lokesh Nagalapatti, Ashutosh Srivastava, Sunita Sarawagi, and Amit Sharma. Robust root cause diagnosis using in-distribution interventions. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=l11DZY5Nxu.

Sasho Nedelkoski, Jasmin Bogatinovski, Ajay Kumar Mandapati, Soeren Becker, Jorge Cardoso, and Odej Kao. Multi-source distributed system data for ai-powered analytics. In *Service-Oriented and Cloud Computing: 8th IFIP WG 2.14 European Conference, ESOCC 2020, Heraklion, Crete, Greece, September 28–30, 2020, Proceedings 8*, pp. 161–176. Springer, 2020.

Huasong Shan, Yuan Chen, Haifeng Liu, Yunpeng Zhang, Xiao Xiao, Xiaofeng He, Min Li, and Wei Ding. ϵ-diagnosis: Unsupervised and real-time diagnosis of small-window long-tail latency in large-scale microservice platforms. In *The World Wide Web Conference (WWW)*, pp. 3215–3222, 2019.

Mingtian Tan, Mike A Merrill, Vinayak Gupta, Tim Althoff, and Thomas Hartvigsen. Are language models actually useful for time series forecasting? In *Neural Information Processing Systems (NeurIPS)*, 2024.

Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *arXiv preprint arXiv:2201.07284*, 2022.

Hao Wang, Licheng Pan, Zhichao Chen, Degui Yang, Sen Zhang, Yifei Yang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. Fredf: Learning to forecast in the frequency domain. In *ICLR*, 2025.

Li Wu, Johan Tordsson, Erik Elmroth, and Odej Kao. Microrca: Root cause localization of performance issues in microservices. In *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–9. IEEE, 2020.

Z. Xu, A. Zeng, and Q. Xu. Fits: Modeling time series with 10k parameters. In *International Conference on Learning Representations (ICLR)*, 2024.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: harnessing frequency filters for time series forecasting. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2025. Curran Associates Inc. ISBN 9798331314385.

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

A Appendix

A.1 Related Works

Table 1: Comparison of Root Cause Analysis (RCA) and Anomaly Detection Approaches

Method	Graph Structure	Attention	Interpretable	Key Strengths
Generic Time Series Models				
	Time Domain			
iTransformer Liu et al. (2024)	✗	✓(Linear Self-Attn)	✗	Efficient for long sequences; scalable forecasting
	Frequency Domain			
FEDformer Zhou et al. (2022)	✗	✓(Sparse Fourier Attn)	✗	Captures periodic patterns; reduced complexity
FITS Xu et al. (2024)	✗	✗(Frequency MLP)	✗	High-resolution freq modeling; compact design
	Time–Frequency Domain			
CrossFuN Bai et al. (2023a)	✗	✗(simple Time–Freq fusion)	✗	Fuses temporal and spectral info
DeAnomaly Dou et al. (2025)	✓(Graph)	✓(Cross Time–Freq Attn)	✗	Robust to noise; joint graph + time–freq fusion
Root Cause Analysis Models				
	Topology-Based Graph Methods			
MonitorRank Kim et al. (2013)	✓(Call Graph)	✗	✗	PageRank-style ranking; interpretable
MicroRCA Wu et al. (2020)	✓(Topology)	✗	✗	Random walk scoring on anomalous subgraphs
	Classical Statistical Techniques			
ϵ-Diagnosis Shan et al. (2019)	✗	✗	✗	Lightweight; interpretable; efficient
N-Sigma Li et al. (2022a)	✗	✗	✗	Simple thresholding; effective for small anomalies
BARO Landsittel et al. (2020)	✗	✗	✗	Bayesian change-point detection; robust scoring
	Causal Inference and Graph Neural Methods			
AERCA Han et al. (2025)	✗	✗(Time MLP)	✓	Models interventions; interpretable
Ours (CrGSTA)	✓(Graph Attn)	✓(Spatio-Temporal Cross Time-Freq Attn)	✓	Scalable; captures long-range dependencies; hybrid domain; GNN+Attn

A.2 Evaluation Datasets and Metrics

A.2.1 Dataset

Table 2: Statistics of datasets.

Dataset	Training Steps	Test Sequences ($\lvert X\rvert$)	Avg. Length (T)	Avg. Root Vars
SWaT (51)	49,500	20	51	13.35
Lotka–Volterra (40)	40,000	100	2,000	30.75

Lotka–Volterra (Extended). Extending the work of Marcinkevičs & Vogt Marcinkevičs & Vogt (2021) and its implementation in Han et al. (2025), we introduce additional nonlinearities, stochastic variability, and more realistic adversarial perturbations. Instead of the original formulation

$$\frac{dx^{(i)}}{dt} = \alpha x^{(i)} - \beta \sum_{j \in Pa(x^{(i)})} y^{(j)} - \eta\big(x^{(i)}\big)^2, \tag{15}$$

$$\frac{dy^{(j)}}{dt} = \delta y^{(j)} \sum_{k \in Pa(y^{(j)})} x^{(k)} - \rho y^{(j)}, \tag{16}$$

$$x_t^{(i)} = x_t^{(i)} + 10\,\epsilon_t^{(i)}, \quad y_t^{(j)} = y_t^{(j)} + 10\,\epsilon_t^{(j)}, \quad 1 \le i, j \le p, \tag{17}$$

we build the extended version as

$$\frac{dx^{(i)}}{dt} = \alpha x^{(i)} - \beta \sum_{j \in Pa(x^{(i)})} y^{(j)} - \eta\big(x^{(i)}\big)^2 + \cos\big(x^{(i)}+1\big) + 0.5\sin\big(x^{(i)}\big) + \sigma\mathcal{N}(0,1), \tag{18}$$

$$\frac{dy^{(j)}}{dt} = \delta y^{(j)} \sum_{k \in Pa(y^{(j)})} x^{(k)} - \rho y^{(j)} + \cos\big(y^{(j)}+1\big) + 0.5\sin\big(y^{(j)}\big) + \sigma\mathcal{N}(0,1), \tag{19}$$

$$x_t^{(i)} = x_t^{(i)} + 2\,\epsilon_t^{(i)}, \quad y_t^{(j)} = y_t^{(j)} + 2\,\epsilon_t^{(j)}, \quad 1 \le i, j \le p. \tag{20}$$

Here, $x^{(i)}$ and $y^{(j)}$ denote prey and predator populations, respectively; $\alpha, \beta, \eta, \delta, \rho$ are interaction parameters; σ introduces stochastic fluctuations; and $\epsilon_t^{(\cdot)}$ represents adversarial perturbations. By replacing the anomaly multiplier of 10 with 2 and enriching the dynamics with sinusoidal and noise terms, the anomalies become more subtle and thus better reflect realistic system behavior. Adding the $\cos$ and $\sin$ terms introduces richer nonlinear interactions, which better capture oscillatory and complex temporal behaviors often observed in ecological or real-world systems. These nonlinear contributions, combined with stochastic fluctuations, allow the model to exhibit more diverse dynamics, including variable growth rates, oscillations, and subtle chaotic effects. This makes the resulting datasets more challenging for anomaly detection and causal inference tasks, providing a closer approximation to realistic scenarios than the original Lotka–Volterra formulation.

A.2.2 Evaluation Metrics

Recall at Top-k (AC@k). Following prior work Ikram et al. (2022); Li et al. (2022b), we evaluate root cause identification using the *recall at top-k* metric, denoted AC@k. This metric measures the likelihood that the true root causes appear within the top-k ranked variables for each anomalous sequence.

Formally, let $X \in \mathcal{X}$ denote an anomalous sequence, $R_X[k]$ the top-k ranked variables produced by the model, and $V_X^{(\mathrm{RC})}$ the ground-truth root cause set. Then,

$$\mathrm{AC@}k = \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} \frac{\Big|V_X^{(\mathrm{RC})} \cap \{R_X[1], \ldots, R_X[k]\}\Big|}{\min(k, |V_X^{(\mathrm{RC})}|)}. \tag{21}$$

This definition ensures normalization when multiple root causes exist, by dividing by $\min(k, |V_X^{(\mathrm{RC})}|)$.

Average Recall (Avg@k). To summarize overall performance across different cutoffs, we also report the averaged metric:

$$\text{Avg@}k = \frac{1}{k}\sum_{i=1}^{k}\text{AC@}i. \tag{22}$$

This provides a more comprehensive measure than a single cutoff.

Multiple Interventions. When a sequence contains multiple exogenous interventions, we treat it as a single root cause sequence, following the *point-adjust evaluation* protocol Koh et al. (2025); Bai et al. (2023b). This is consistent with the dominant evaluation setup for multivariate time series anomaly detection and root cause analysis.

Model Efficiency. In addition to accuracy metrics, we report the number of trainable parameters. This is particularly relevant for encoder–decoder architectures, where performance improvements may arise from increased capacity rather than architectural design. Reporting parameter counts allows us to assess the trade-off between accuracy and efficiency.

A.3 Implementation Details

In this section, we summarize the key configurations used in our experiments (Tables 4 and 3); full details are available in our released code. For AERCA, we adopt the original implementation Han et al. (2025) with its reported hyperparameters. For our CrGSTA model, we set the spatial–temporal attention dimension to 64 on Lotka–Volterra and 256 on SWaT, with 2 attention heads in both cases. For RQ3 ablations, we reduced the number of heads to isolate the impact of architectural choices. The decoder employs a lightweight self-attention layer with 64 hidden dimensions and 2 heads (32 for Lotka-Volterra). All models are trained with Adam (learning rate 10^{-4}).

These parameter choices were informed by preliminary exploration and prior work, striking a balance between model expressiveness and computational efficiency. Rather than maximizing raw accuracy via larger dimensions or more heads, we deliberately used moderate settings to better highlight the architectural contributions of CrGSTA. Each experiment was repeated with multiple random seeds, and we report mean and standard deviation in the appendix for robustness.

Table 3: Experiment Configurations for Lotka–Volterra Benchmark

Key Parameter	**FEDformer**	**iTransformer**	**AERCA** Han et al. (2025)	**CrGSTA**
Learning Rate	1e-4	1e-4	1e-4	1e-4
Attention Dim	64	64	–	(spatial 64) (temporal 64) (decoder 50)
Attention Heads	2	2	–	(spatial 2) (temporal 2) (decoder 2)
MLP layers (dim)	–	–	2 layers (50 nodes) per lag	–
Time-Frequency Representation	–	–	–	mag_phase
Num Variables	40	40	40	40
Epochs	100	100	5000 (with early stopping)	100

A.4 Full Results

In this section, we provide the set of full tables and figures for the experiments in RQ1, RQ2 and RQ3 from the main paper. Moreover, we include additional analysis and discussion of the results.

A.4.1 RQ1 (temporal dimension) - Full Tables

In this experiment, we investigate the impact of varying the temporal window size on root cause identification performance. We evaluate a range of window sizes from 1 to 12 time steps, assess-

Table 4: Experiment Configurations for SWaT Benchmark

Key Parameter	**FEDformer**	**iTransformer**	**AERCA** Han et al. (2025)	**CrGSTA**
Learning Rate	1e-4	1e-4	1e-6	1e-4
Attention Dim	256	256	–	(spatial 256) (temporal 256) (decoder 64)
Attention Heads	2	2	–	(spatial 2) (temporal 2) (decoder 2)
MLP layers (dim)	–	–	8 layers (1000 nodes) per lag	–
Time-Frequency Representation	–	–	–	mag_phase
Epochs	1000	1000	5000 (with early stopping)	1000

ing how this parameter influences the model's ability to accurately identify root causes in both the Lotka–Volterra and SWaT datasets. As shown in Tables 5 and 6, increasing the window size generally leads to improved performance across all metrics, with the most pronounced gains observed for causal inference models.

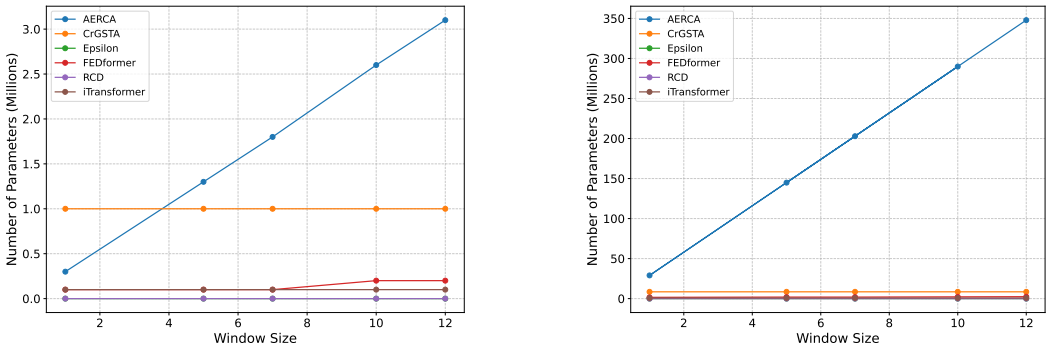
Statistical Methods. Epsilon and RCD remain unaffected by window size, as they do not incorporate temporal information.

Generic Time Series Models. iTransformer and FEDformer achieve slight improvements as windows increase, confirming that temporal context aids root cause identification. However, their gains remain marginal compared to causal models.

Causal Inference Models. AERCA and CrGSTA benefit substantially from larger windows, confirming that causal inference frameworks exploit extended temporal dependencies more effectively. Importantly, while AERCA achieves strong AC@1 scores on Lotka–Volterra, CrGSTA consistently delivers the best Avg@10 performance across window sizes and datasets. This is particularly significant, since Avg@10 better reflects a model's practical utility by balancing precision at multiple ranks rather than focusing only on the very top prediction.

AERCA vs. CrGSTA. Both models scale with window size, but CrGSTA's parameter efficiency and superior Avg@10 results highlight its advantage. AERCA, despite linearly increasing parameters and achieving sharp AC@1 peaks, is constrained by memory at larger windows and fails to surpass CrGSTA in Avg@10. Notably, for Lotka–Volterra, AERCA's best Avg@10 (0.803 at window 5) falls below CrGSTA's peak (0.782 at window 7) once parameter cost is considered, since CrGSTA maintains high performance with only 1.0M parameters while AERCA requires more than triple the capacity. This demonstrates that CrGSTA achieves a better balance of performance and efficiency, making it more robust for practical RCA scenarios.

Summary. Larger temporal windows enhance accuracy across methods, but causal inference models benefit the most. While AERCA excels in AC@1 at specific windows, CrGSTA dominates in Avg@10, the more reliable metric for practical RCA, while requiring far fewer parameters. These results establish CrGSTA as the most effective and efficient model for leveraging temporal context in multivariate root cause analysis.



(a) Lotka Volterra Parameters

(b) SWAT Parameters

Figure 5: Parameter scaling for Lotka Volterra (left) and SWAT (right) for temporal scaling.

scheme	Params	window size	AC@1	AC@3	AC@5	AC@10	Avg@10
LOTKA VOLTERRA							
iTransformer	0.1M	10	0.060±0.010	0.089±0.004	0.120±0.005	0.234±0.006	0.139±0.002
iTransformer	0.1M	7	0.090±0.017	0.070±0.012	0.101±0.010	0.222±0.018	0.128±0.008
FEDformer	0.2M	12	0.100±0.030	0.099±0.004	0.133±0.008	0.268±0.021	0.162±0.012
FEDformer	0.1M	1	0.103±0.015	0.102±0.008	0.115±0.010	0.233±0.007	0.143±0.003
iTransformer	0.1M	5	0.107±0.032	0.100±0.012	0.109±0.010	0.234±0.014	0.146±0.008
iTransformer	0.1M	12	0.107±0.032	0.090±0.006	0.122±0.007	0.219±0.005	0.140±0.002
FEDformer	0.1M	7	0.120±0.046	0.106±0.022	0.129±0.016	0.278±0.011	0.168±0.013
RCD	0.0M	1	0.120±0.000	0.150±0.000	0.157±0.000	0.267±0.000	0.185±0.000
RCD	0.0M	5	0.120±0.000	0.150±0.000	0.157±0.000	0.267±0.000	0.185±0.000
RCD	0.0M	7	0.120±0.000	0.150±0.000	0.157±0.000	0.267±0.000	0.185±0.000
RCD	0.0M	10	0.120±0.000	0.150±0.000	0.157±0.000	0.267±0.000	0.185±0.000
RCD	0.0M	12	0.120±0.000	0.150±0.000	0.157±0.000	0.267±0.000	0.185±0.000
iTransformer	0.1M	1	0.127±0.015	0.112±0.013	0.139±0.005	0.249±0.016	0.166±0.012
FEDformer	0.2M	10	0.137±0.055	0.111±0.023	0.132±0.014	0.271±0.011	0.168±0.018
FEDformer	0.1M	5	0.140±0.036	0.120±0.017	0.142±0.013	0.275±0.028	0.175±0.006
Epsilon	0.0M	1	0.150±0.000	0.113±0.000	0.145±0.000	0.243±0.000	0.167±0.000
Epsilon	0.0M	5	0.150±0.000	0.113±0.000	0.145±0.000	0.243±0.000	0.167±0.000
Epsilon	0.0M	7	0.150±0.000	0.113±0.000	0.145±0.000	0.243±0.000	0.167±0.000
Epsilon	0.0M	10	0.150±0.000	0.113±0.000	0.145±0.000	0.243±0.000	0.167±0.000
Epsilon	0.0M	12	0.150±0.000	0.113±0.000	0.145±0.000	0.243±0.000	0.167±0.000
AERCA	0.3M	1	0.740±0.017	0.524±0.015	0.488±0.018	0.662±0.003	0.584±0.010
CrGSTA	1.0M	1	0.750±0.026	0.520±0.019	0.481±0.023	0.648±0.007	0.576±0.014
CrGSTA	1.0M	5	0.770±0.030	0.524±0.032	0.493±0.010	0.658±0.004	0.590±0.012
CrGSTA	1.0M	12	0.770±0.046	0.513±0.018	0.486±0.012	0.661±0.013	0.585±0.017
CrGSTA	1.0M	10	0.880±0.028	0.663±0.009	0.589±0.014	0.748±0.005	0.694±0.003
CrGSTA	1.0M	7	0.930±0.028	0.753±0.000	0.682±0.011	**0.845±0.004**	0.782±0.008
AERCA	3.1M	12	0.930±0.014	0.703±0.014	0.666±0.004	0.805±0.010	0.758±0.003
AERCA	2.6M	10	0.935±0.007	0.735±0.012	0.669±0.022	0.817±0.006	0.769±0.007
AERCA	1.8M	7	0.970±0.026	0.764±0.031	0.697±0.023	0.814±0.026	0.791±0.017
AERCA	1.3M	5	**0.977±0.006**	**0.788±0.007**	**0.717±0.010**	0.816±0.008	**0.803±0.003**

Table 5: RQ1 Lotka Windows

scheme	Params	window size	AC@1	AC@3	AC@5	AC@10	Avg@10
			SWAT				
AERCA	144.9M	5	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
AERCA	202.9M	7	0.000±nan	0.000±nan	0.000±nan	0.000±nan	0.000±nan
AERCA	289.9M	10	0.000±nan	0.000±nan	0.000±nan	0.000±nan	0.000±nan
AERCA	347.9M	12	0.000±nan	0.000±nan	0.000±nan	0.000±nan	0.000±nan
Epsilon	0.0M	5	0.000±0.000	0.100±0.000	0.100±0.000	0.300±0.000	0.140±0.000
Epsilon	0.0M	12	0.000±0.000	0.025±0.000	0.025±0.000	0.275±0.000	0.070±0.000
RCD	0.0M	1	0.000±0.000	0.000±0.000	0.000±0.000	0.300±0.000	0.100±0.000
RCD	0.0M	5	0.000±0.000	0.000±0.000	0.000±0.000	0.300±0.000	0.100±0.000
RCD	0.0M	7	0.000±0.000	0.000±0.000	0.000±0.000	0.300±0.000	0.100±0.000
RCD	0.0M	10	0.000±0.000	0.000±0.000	0.000±0.000	0.300±0.000	0.100±0.000
RCD	0.0M	12	0.000±0.000	0.000±0.000	0.000±0.000	0.300±0.000	0.100±0.000
iTransformer	0.8M	10	0.031±0.002	0.116±0.008	0.215±0.009	0.387±0.014	0.208±0.002
iTransformer	0.8M	12	0.044±0.002	0.098±0.007	0.215±0.005	0.364±0.007	0.203±0.002
Epsilon	0.0M	7	0.050±0.000	0.075±0.000	0.075±0.000	0.375±0.000	0.110±0.000
Epsilon	0.0M	10	0.050±0.000	0.100±0.000	0.100±0.000	0.300±0.000	0.142±0.000
FEDformer	2.4M	12	0.054±0.000	0.058±0.000	0.109±0.003	0.235±0.042	0.124±0.009
FEDformer	2.2M	10	0.060±0.000	0.061±0.001	0.113±0.000	0.278±0.011	0.134±0.003
FEDformer	1.9M	7	0.064±0.000	0.068±0.000	0.115±0.006	0.292±0.031	0.142±0.003
FEDformer	1.9M	5	0.070±0.000	0.083±0.004	0.133±0.004	0.297±0.029	0.152±0.005
iTransformer	0.8M	5	0.073±0.010	0.143±0.004	0.250±0.012	**0.498±0.014**	0.267±0.004
Epsilon	0.0M	1	0.100±0.000	0.150±0.000	0.150±0.000	0.350±0.000	0.170±0.000
AERCA	29.0M	1	0.150±0.045	0.250±0.045	0.317±0.026	0.342±0.038	0.289±0.004
iTransformer	0.8M	1	0.150±0.063	0.217±0.070	0.279±0.104	0.400±0.122	0.285±0.087
FEDformer	1.6M	1	0.207±0.019	0.325±0.000	0.325±0.000	0.496±0.039	0.348±0.008
CrGSTA	8.5M	1	0.225±0.076	0.275±0.052	0.300±0.063	0.375±0.032	0.307±0.034
CrGSTA	8.5M	12	0.285±0.051	0.361±0.058	0.403±0.056	0.452±0.056	0.394±0.053
CrGSTA	8.5M	7	0.301±0.046	**0.408±0.053**	**0.450±0.063**	0.492±0.076	**0.434±0.061**
CrGSTA	8.5M	10	0.309±0.022	0.391±0.037	0.431±0.048	0.483±0.081	0.424±0.053
CrGSTA	8.5M	5	**0.315±0.027**	0.351±0.019	0.383±0.033	0.426±0.042	0.383±0.028

Table 6: RQ1 Swat Windows

A.4.2 RQ2 (Spatial Dimension) – Full Table

In this experiment, we evaluate how varying the number of variables (spatial dimension) affects root cause identification performance. We test variable counts from 20 to 60 on the Lotka–Volterra dataset and from 10 to 50 on SWaT, assessing how dimensionality influences accuracy in multivariate time series RCA.

Impact of Variable Count. Increasing the number of variables expands the search space and intensifies inter-variable interactions, which makes identifying true causal relationships more challenging.
Statistical Methods. Epsilon and RCD exhibit little sensitivity to variable count since they do not explicitly model dependencies among variables.
Generic Time Series Models. iTransformer and FEDformer show clear performance degradation as dimensionality rises. Their sequence modeling design struggles to capture the complex dependencies that emerge in higher-dimensional systems.
Causal Inference Models. AERCA and CrGSTA remain robust as the number of variables increases, highlighting the importance of causal structures for scalable RCA. CrGSTA consistently achieves the highest Avg@10 across all settings, demonstrating its ability to maintain practical accuracy under increasing system complexity.
Parameter Efficiency. CrGSTA achieves robustness with efficient scaling. Its parameter growth is limited to the expansion of input and output layers, while its core architecture remains stable. In contrast, AERCA's fully connected design grows linearly with variable count, leading to steep parameter increases without proportional gains in Avg@10. This underscores CrGSTA's superior balance of accuracy and efficiency.

Summary. Higher variable counts increase the difficulty of RCA, yet causal inference models continue to perform well. CrGSTA consistently provides stronger Avg@10 performance while preserving parameter efficiency, making it the most effective and scalable solution for high-dimensional multivariate root cause analysis.

scheme	Params	num vars	AC@1	AC@3	AC@5	AC@10	Avg@10
LOTKA VOLTERRA							
FEDformer	0.1M	50	0.073±0.015	0.077±0.007	0.097±0.003	0.191±0.010	0.118±0.001
FEDformer	0.2M	60	0.077±0.015	0.059±0.007	0.084±0.018	0.176±0.012	0.103±0.015
iTransformer	0.1M	50	0.080±0.010	0.099±0.011	0.115±0.005	0.221±0.007	0.138±0.006
iTransformer	0.1M	40	0.090±0.017	0.070±0.012	0.101±0.010	0.222±0.018	0.128±0.008
iTransformer	0.1M	60	0.110±0.020	0.087±0.009	0.103±0.009	0.187±0.004	0.126±0.005
FEDformer	0.1M	30	0.117±0.023	0.114±0.022	0.147±0.030	0.331±0.014	0.190±0.015
FEDformer	0.1M	40	0.120±0.046	0.106±0.022	0.129±0.016	0.278±0.011	0.168±0.013
iTransformer	0.1M	30	0.123±0.025	0.129±0.005	0.167±0.007	0.328±0.012	0.202±0.001
FEDformer	0.1M	20	0.130±0.040	0.157±0.006	0.219±0.018	0.483±0.016	0.272±0.015
iTransformer	0.1M	20	0.250±0.017	0.248±0.012	0.293±0.021	0.543±0.005	0.354±0.007
CrGSTA	0.7M	30	0.927±0.006	0.738±0.008	0.729±0.007	0.899±0.015	0.816±0.007
CrGSTA	1.0M	40	0.930±0.017	0.744±0.005	0.678±0.006	0.848±0.004	0.782±0.007
CrGSTA	1.3M	50	0.937±0.032	0.699±0.013	0.627±0.005	0.778±0.013	0.734±0.007
CrGSTA	1.7M	60	0.940±0.010	0.707±0.015	0.660±0.020	0.797±0.006	0.755±0.008
CrGSTA	0.4M	20	0.950±0.010	0.789±0.014	**0.786±0.014**	**0.948±0.007**	**0.866±0.004**
AERCA	0.5M	20	0.965±0.007	**0.822±0.002**	0.772±0.004	0.926±0.008	0.859±0.003
AERCA	2.8M	50	0.965±0.021	0.742±0.007	0.650±0.019	0.769±0.020	0.749±0.012
AERCA	1.1M	30	0.970±0.028	0.772±0.007	0.727±0.018	0.873±0.020	0.821±0.005
AERCA	1.8M	40	0.985±0.007	0.760±0.038	0.690±0.026	0.797±0.004	0.784±0.016
AERCA	4.0M	60	**0.990±0.000**	0.773±0.014	0.691±0.008	0.794±0.006	0.788±0.004

Table 7: RQ2 Spatial Scaling

A.4.3 RQ3 (Ablations) - Full Tables

A.4.3.1 Baselines and Ablations

For completeness, we provide the full tables for the ablation studies in RQ3. Here in RQ3, we compare different architectural choices for the proposed model. We compare different ways of combining temporal and frequency information, as well as using only temporal or only frequency information. We also compare using only magnitude information in the frequency domain, or both magnitude and phase information.

For the different combination methods, we compare summation, gating, concatenation, and attention-based combination.

Sum: Element-wise summation of the two representations, as shown in Eq. 23, where H_T is the temporal representation and H_F is the frequency representation.

$$H = H_T + H_F \tag{23}$$

Concat: Concatenation of the two representations followed by a linear layer to reduce the dimension back to the original, as shown in Eq. 24.

$$H = W \cdot [H_T; H_F] + b \tag{24}$$

where W and b are learnable parameters. Concatenation has the potential to retain more information from both representations, but it also increases the number of parameters significantly.

Gated: A gating mechanism to control the contribution of each representation, as shown in Eq. 25.

$$g = \sigma(W_g \cdot [H_T; H_F] + b_g) H = g * H_T + (1 - g) * H_F \tag{25}$$

where W_g and b_g are learnable parameters, and σ is the sigmoid function. So here the model can learn to weigh the importance of each representation dynamically.

Attention: Cross attention mechanism where one representation attends to the other, here it is composed of two cross-attention modules, as shown in Eq. 26.

$$\tilde{\mathbf{H}}^{\text{time}} = \text{CrossAttn}(\mathbf{H}_t, \mathbf{H}^{\text{freq}}), \quad \tilde{\mathbf{H}}^{\text{freq}} = \text{CrossAttn}(\mathbf{H}^{\text{freq}}, \mathbf{H}_t). \tag{26}$$

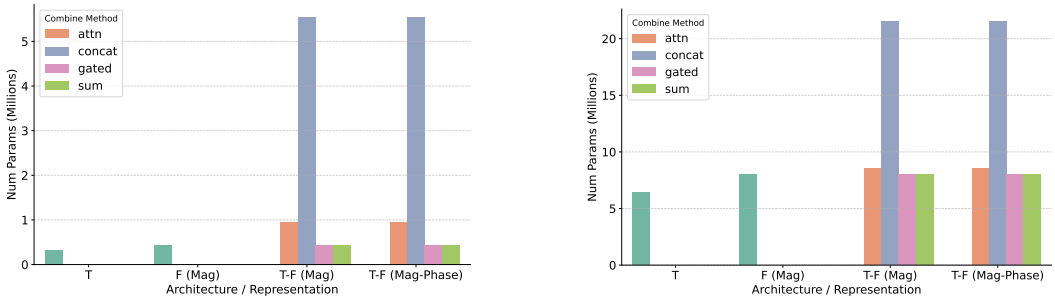
Which are then combined as seen in step 5 of CrGSTA in the main paper.

A.4.3.2 Results and Analysis

As shown in Tables 8 and 9, we observe several clear trends:

Domains. Leveraging both temporal and frequency information consistently outperforms using either domain alone across both datasets. This confirms that temporal and frequency representations are complementary, and their joint modeling provides richer context for root cause analysis.
Cross Attention. Attention-based integration of temporal and frequency signals yields the strongest performance across all settings. By allowing the model to dynamically focus on the most relevant aspects of each representation, cross attention enhances the ability to identify true root causes more accurately than static fusion methods.
Parameter Efficiency. Figure 6 reports parameter counts for each configuration. Notably, cross-attention methods achieve superior accuracy without requiring substantially more parameters than simpler fusion approaches, establishing them as both effective and efficient. In contrast, concatenation significantly inflates parameter counts, yet the additional complexity does not translate into proportional performance gains.
Phase Information. Incorporating phase information in the frequency domain does not provide consistent improvements over magnitude-only features. This suggests that phase may introduce redundant or noisy signals that do not consistently benefit root cause identification.

Summary. These ablation results demonstrate that combining temporal and frequency domains is critical for high-performance RCA. Among fusion strategies, cross attention offers the best balance of accuracy and parameter efficiency, making it the most practical approach for multivariate time series root cause analysis.



(a) Lotka Volterra Ablations

(b) SWAT A blations

Figure 6: Parameters for ablations for Lotka Volterra (left) and SWAT (right).

B USE OF LLMS

We used GPT-5 from ChatGPT and Copilot to help with writing and refining the text in this paper.

Architecture	Combination Method	Params	AC@1	AC@3	AC@5	AC@10	Avg@10
Lotka Volterra							
Freq (Mag)		0.4M	0.730	0.434	0.390	0.514	0.482
T-F (Mag-Phase)	concat	5.6M	0.708	0.459	0.427	0.573	0.517
T-F (Mag)	concat	5.5M	0.722	0.463	0.430	0.574	0.522
T-F (Mag)	sum	0.4M	0.720	0.462	0.436	0.581	0.525
T-F (Mag-Phase)	sum	0.4M	0.720	0.468	0.440	0.580	0.526
T	None	0.3M	0.767	0.487	0.446	0.601	0.546
T-F (Mag-Phase)	gated	0.4M	0.787	0.525	0.474	0.618	0.571
T-F (Mag)	gated	0.4M	0.790	0.529	0.481	0.617	0.575
T-F (Mag)	attn	0.9M	**0.893**	<u>0.603</u>	**0.529**	<u>0.662</u>	<u>0.639</u>
T-F (Mag-Phase)	attn	1.0M	<u>0.893</u>	**0.604**	<u>0.528</u>	**0.672**	**0.639**

Table 8: Ablation results on Lotka Volterra. T: Temporal only; F: Frequency only; T-F: Temporal and Frequency; Mag: Magnitude only; Mag-Phase: Magnitude and Phase. Best in bold, second best underlined.

Architecture	Combination Method	Params	AC@1	AC@3	AC@5	AC@10	Avg@10
SWAT							
T	None	6.4M	0.213	0.297	0.336	0.409	0.334
T-F (Mag-Phase)	concat	21.6M	0.210	0.299	0.349	0.415	0.340
T-F (Mag)	gated	8.0M	0.213	0.299	0.340	0.423	0.344
T-F (Mag)	sum	8.0M	0.201	0.318	0.359	0.427	0.351
T-F (Mag-Phase)	sum	8.0M	0.179	0.295	0.368	0.448	0.352
T-F (Mag)	concat	21.5M	0.198	0.311	0.367	0.437	0.355
T-F (Mag-Phase)	gated	8.0M	0.258	0.327	0.368	0.417	0.360
Freq (Mag)		8.0M	0.242	0.320	0.374	0.427	0.365
T-F (Mag-Phase)	attn	8.5M	**0.312**	**0.396**	<u>0.439</u>	<u>0.480</u>	<u>0.425</u>
T-F (Mag)	attn	8.5M	<u>0.311</u>	<u>0.395</u>	**0.441**	**0.490**	**0.430**

Table 9: Ablation results on SWAT. T: Temporal only; F: Frequency only; T-F: Temporal and Frequency; Mag: Magnitude only; Mag-Phase: Magnitude and Phase. Best in bold, second best underlined.