Graph-Based Causal Discovery for Anomaly Detection in multivariate Time-Series Data

Anonymous Author(s) †

ABSTRACT

Multivariate time series causality analysis is crucial for tasks such as anomaly detection, forecasting, and system monitoring. Traditional Granger causality analysis focuses on the endogenous variables when identifying such causalities, and presents a limitation in the exclusion of exogenous variables in its analysis. In this paper, we propose a novel framework, named Gaussian Process Causal Discovery with Exogenous modeling (GPCDX), capable of inference from both Granger causality and exogenous variables. This framework includes a Gaussian Process-based module that estimates Granger causality distributions from the interactions between sensors, and includes a Variational Graph AutoEncoder (VGAE) that encodes exogenous information that is not explicitly explained by existing endogenous structures as latent variables. The inferred Granger causality distribution is incorporated into the decoder of the VGAE, allowing for a more robust time series reconstruction. The experiment results on various datasets show that the proposed architecture outperforms baseline models for anomaly detection. As such, this paper shows the significance of distributional representation of causal inferences and the importance of exogenous variables to complex time-series systems.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Online learning settings**; *Image processing*; • **Applied computing** → *Health informatics.*

KEYWORDS

Granger Causality, Exogenous Variable, Multivariate Time Series Anomaly Detection

ACM Reference Format:

Anonymous Author(s). 2025. Graph-Based Causal Discovery for Anomaly Detection in multivariate Time-Series Data. In *Proceedings of Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25) (CIKM '25).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Multivariate time series anomaly detection plays a critical role in various application domains such as industrial equipment, monitoring systems, medical diagnostics, and cyber-physical systems[41][16]. These systems often involve a growing number of sensors and variables, which are proportional to the size of the equipment[23][8], that interact in complex ways, and accurately understanding their temporal dependencies can significantly impact anomaly detection performance[3]. However, most existing approaches primarily focus on reconstruction or future value prediction, failing to fully exploit the underlying causal structure among variables[40]. This oversight limits the ability to identify how anomalies propagate through the system or to trace their root causes[11]. Traditional Granger causality analysis aims to infer directional relationships between time series by assessing whether the past values of one variable help predict the future of another[30]. While effective in certain scenarios, conventional Granger-based methods are constrained by linearity assumptions and typically apply only to observed internal variables, ignoring the influence of unobserved or external factors—known as exogenous variables[14]. These exogenous factors represent influences that cannot be explained by the historical values of internal variables alone and may stem from external conditions or latent sources[11]. For instance, in industrial machinery sensor data, sensor responses may vary under identical input conditions due to external factors such as temperature, weather, or scheduling. Likewise, component failures or cyberattacks can also be regarded as exogenous influences. Since these factors are not directly measured, they are often excluded from model training, despite their significant impact in real-world systems.

To address these challenges, we propose a novel framework called Gaussian Process Causal Discovery with Exogenous modeling (GPCDX). This framework extends conventional Granger causality analysis in two key ways. First, it employs Gaussian Process regression to model causal coefficients as distributions, enabling the capture of both nonlinear dependencies and uncertainty in causal relationships. Unlike prior work that typically represents Granger causal relationships using scalar-valued matrices, which may overlook complex temporal patterns such as trends and seasonality, our approach models inter-variable dependencies as distributions, allowing the representation of variability in causal influence over time. Second, we introduce Variational Graph AutoEncoders (VGAEs[18]) to infer latent exogenous variables that cannot be explained by endogenous structures alone. These latent variables are integrated into the time series reconstruction process to enhance the model's expressiveness. By combining Granger causal structures with exogenous variable representations, our framework enables more accurate reconstruction of normal time series patterns and more effective detection of anomalous patterns that deviate from the learned distributions. Extensive experiments on diverse time series datasets demonstrate that the proposed method outperforms existing approaches in anomaly detection tasks. These results highlight the importance of distributional representations in causal inference and the modeling of exogenous influences in complex time series systems.

CIKM '25, November 10–14, 2025, Seoul, Republic of Korea

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

2 RELATED WORK

2.1 Multivariate Time Series Anomaly Detection

Multivariate time series anomaly detection is a critical task for the early identification of abnormal states in complex systems[38]. The goal is to detect anomalous patterns by analyzing time series data composed of multiple interacting variables. Existing approaches are broadly categorized into reconstruction-based and prediction-based methods. Reconstruction-based approaches aim to learn the normal patterns of time series data by encoding and decoding the data using models such as Autoencoders[17] or Variational Autoencoders (VAEs)[19]. These models are trained on normal data and detect anomalies based on the reconstruction error between the input and output sequences. However, these methods often successfully reconstruct the abnormal data as well, and lack interpretability regarding the underlying causes of anomalies, as they rely solely on reconstruction performance[33]. Prediction-based approaches, on the other hand, utilize previous observations to directly predict the next time step or future window values[22]. Anomalies are detected based on high prediction errors. These methods naturally capture the temporal continuity of time series data and are effective for short-term forecasting. However, they can be vulnerable to long-term anomaly detection if the prediction horizon is limited, and they typically overlook external influences or latent factors that are not directly observable in the data, making them insufficient for modeling real-world systems where exogenous influences are significant[32]. Recent advances have introduced models based on Transformers and Temporal Convolutional Networks (TCNs) to better capture complex temporal dependencies[25]. Attention has also been given to representation learning-based approaches to address the limitations of reconstruction and prediction-based techniques[6]. These methods aim to learn expressive latent representations that capture high-level structures and patterns not easily observed or modeled by conventional techniques. Notable examples include frequency-domain representations (e.g., CATCH[37]), graph-based sensor dependency learning (e.g., GDN[9]), and patch-based structural modeling (e.g., TVOC, THOC), all of which attempt to uncover hidden patterns within time series data and leverage them for more effective anomaly detection.

2.2 Causal Discovery In Time Series

Causal discovery aims to uncover the causal relationships among variables, discovering how changes in one variable can affect others within a system. In the context of time series data, Granger causality is a widely adopted framework that defines causality based on the influence of past information on future outcomes[12]. Traditional Granger causality analysis relies on linear Vector Autoregressive (VAR) models[31], determining whether the past values of one variable significantly affect the prediction of another variable's future values[10]. However, such methods are inherently limited by their assumption of linearity, making them inadequate for capturing complex nonlinear interactions and uncertainty that commonly arise in real-world systems[36]. To overcome these limitations, recent research has explored nonlinear Granger causality approaches, utilizing models such as Multilayer Perceptrons (MLPs[21]) and attention-based temporal causal networks. These methods offer more flexible causal modeling capabilities and are better suited to represent diverse time series dynamics beyond the constraints of linear models. In spite of these advances, most existing studies still neglect to consider the influence of exogenous variables, despite its possible introduction of variability or explanation for anomalies that cannot be explained by endogenous variables alone. Therefore, to build causal models that are applicable to real-world settings, it is essential to explicitly account for these external or latent factors[40]. This study proposes a novel causal inference framework that estimates Granger causality as a probability distribution, while simultaneously inferring latent exogenous variables. By going beyond conventional deterministic causal structures, the proposed approach enables more interpretable and robust causal discovery, which in turn enhances the accuracy and reliability of anomaly detection in complex multivariate time series systems.

3 PRELIMINARY

3.1 Granger Causality Discovery

Granger causality is a statistical concept used to analyze causal relationships between variables in time series data. This approach is widely used in time series analysis as it identifies relationships based on predictive power while accounting for temporal delays[38]. The general definition is as follows: if the past values of a variable $x^{(i)}$ help predict the future values of another variable $x^{(j)}$, then $x^{(i)}$ is said to Granger-cause $x^{(j)}$. Traditionally, such relationships are inferred using linear Vector Autoregressive (VAR) models[5]. Let us consider a multivariate time series $x_t = (x_t^{(1)}, x_t^{(2)}, \dots x_t^{(N)})$, where N is the number of sensors and $x_t^{(i)}$ denotes the value of variable i at time t:

$$\exists\, x_{t-\tau:t-1}^{(i)\prime} \neq x_{t-\tau:t-1}^{(i)} \tag{1}$$

$$\begin{aligned} f_j(x_{t-\tau:t-1}^{(1)}, \dots x_{t-\tau:t-1}^{(i)\prime}, \dots x_{t-\tau:t-1}^{(N)}) \neq \\ f_j(x_{t-\tau:t-1}^{(1)}, \dots x_{t-\tau:t-1}^{(i)}, \dots x_{t-\tau:t-1}^{(N)}) \end{aligned} \tag{2}$$

Where, $x^{(i)\prime}$ denotes a perturbed version of $x^{(i)}$, indicating a change applied to the variable. If the prediction function f_j is sensitive to the past values of $x^{(i)}$, this suggests the existence of a Granger causal relationship from $x^{(i)}$ to $x^{(j)}$[7]. In traditional Granger analysis, the function f_j is assumed to be linear, and inference is typically performed using a Vector Autoregressive (VAR) model, which takes the following form:

$$x_t = \sum_{\tau=1}^{P} A_\tau x_{t-\tau} + \epsilon_t \tag{3}$$

Where, $A_\tau \in \mathbb{R}^{N\times N}$ represents the coefficient matrix at time lag τ, and $\epsilon_t \sim \mathcal{N}(0, \Sigma)$ denotes the noise term. If the element $[A_\tau]_{ji} \neq 0$, it is interpreted that variable $x^{(i)}$ has an influence on $x^{(j)}$ at lag τ. However, in real-world systems, the relationships between variables are often highly nonlinear, which limits the applicability of linear VAR models. To address this, recent studies have proposed deep learning-based Granger causality inference methods. These approaches approximate each prediction function f_j using neural architectures such as MLPs or RNNs, and assess causal relationships by perturbing or masking the inputs of specific variables. Thus, equation 3 is altered as such:

$$x_t^{(j)} = f_j(x_{t-\tau:t-1}^{(1)}, \dots x_{t-\tau:t-1}^{(N)}) + \epsilon_j^{(j)} \tag{4}$$

Where, f denotes a neural network. Granger causality, as such, identifies causal structures by evaluating the predictive contribution of past information. Unlike simple correlation-based methods, it provides a powerful means of interpreting directional relationships between variables within time series. As a result, it serves as a key concept that enhances both the interpretability and accuracy of time series modeling.

3.2 Gaussian Process Regression

A Gaussian Process (GP) is a non-parametric Bayesian approach for modeling distributions over functions[29]. It defines a prior over functions $f : \mathcal{X} \to \mathbb{R}$ such that for any finite set of inputs $\{x_1, \dots, x_n\} \subset \mathcal{X}$, the corresponding function values $\{f(x_1), \dots, f(x_n)\}$ follow a multivariate Gaussian distribution:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \tag{5}$$

where $m(x) = \mathbb{E}[f(x)]$ is the mean function (typically assumed to be zero), and $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ is the covariance (kernel) function. In this work, we use the Radial Basis Function (RBF) kernel defined as:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}\|x - x'\|^2\right) \tag{6}$$

where σ_f^2 controls the output variance and ℓ is the length-scale hyperparameter. Given observations $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, where $\mathbf{y} = [y_1, \dots, y_n]^\top \in \mathbb{R}^n$, we assume:

$$\mathbf{y} \sim \mathcal{N}(0, K_{XX} + \sigma_n^2 I) \tag{7}$$

where $K_{XX} \in \mathbb{R}^{n\times n}$ is the kernel matrix with entries $k(x_i, x_j)$, and σ_n^2 is the noise variance. For a test input x^*, the predictive distribution is:

$$\begin{aligned} p(f^*|x^*, X, \mathbf{y}) &= \mathcal{N}(\mu^*, \sigma^{*2}) \\ \mu^* &= k_{x^*X}(K_{XX} + \sigma_n^2 I)^{-1}\mathbf{y} \\ \sigma^{*2} &= k(x^*, x^*) - k_{x^*X}(K_{XX} + \sigma_n^2 I)^{-1}k_{Xx^*} \end{aligned} \tag{8}$$

Here, k_{x^*X} is the vector of covariances between the test point and the training points. GP hyperparameters are learned by maximizing the log marginal likelihood:

$$\begin{aligned} \log p(\mathbf{y}|X) = &-\frac{1}{2}\mathbf{y}^\top(K_{XX} + \sigma_n^2 I)^{-1}\mathbf{y} \\ &- \frac{1}{2}\log|K_{XX} + \sigma_n^2 I| - \frac{n}{2}\log(2\pi) \end{aligned} \tag{9}$$

This formulation naturally balances data fit and model complexity[28]. In our framework, each variable pair $(i \leftarrow j)$ is modeled using a GP to capture the dynamic causal influence of variable j on i. Instead of a fixed regression coefficient, our model infers the time-varying posterior mean $\mu_{ij}(t)$ and variance $\sigma_{ij}^2(t)$. These quantities are used to construct a distributional Granger causality structure, where structural deviations from the normal posterior are used to detect anomalies.

4 METHODOLOGY

This section describes the overall architecture and operational flow of the proposed framework, Gaussian Process Causal Discovery with Exogenous Modeling (GPCDX). The model aims to achieve more precise anomaly detection by probabilistically estimating Granger causal structures among variables, while simultaneously inferring latent exogenous factors that cannot be explained by the causal structure alone. The framework consists of two core components: (1) GrangerGPVAR, a module that estimates Granger causality using Gaussian Process regression, and (2) ExogenousVGAE, a Variational Graph Autoencoder module that infers latent exogenous variables from residual signals not captured by the causal model. The GrangerGPVAR module constructs an individual Gaussian Process regression model for each variable pair $(i \leftarrow j)$, estimating the conditional probability distribution that represents how the past values of variable j influence the focus value of variable i. The resulting causal distributions are aggregated into a causality matrix that reflects the overall causal dependencies among variables. This matrix is then used as an edge-weighted adjacency matrix for a graph where each sensor is treated as a node. The resulting causal graph is passed to the ExogenousVGAE module, which infers the latent distribution of exogenous factors and reconstructs the input time series. In this integrated design, the GrangerGPVAR module not only estimates causal coefficients but also provides a structured topological prior for ExogenousVGAE. This allows the model to disentangle unexplained signals as exogenous influences and reconstruct the original time series accordingly. By jointly modeling the learned causal structure and exogenous variables, the proposed framework enables fine-grained reconstruction of normal patterns and effective detection of anomalies.

4.1 Problem Formulation

A multivariate time series $x_{1:T} = \{x_t \in \mathbb{R}^N | t = 1, \dots T\}$ represents a continuous temporal sequence composed of N interacting variables. The state at each time step t is denoted by $x_t = (x_t^{(1)}, \dots x_t^{(N)})$. The objective of the model is to accurately detect anomalous states by leveraging the structure learned from a normal time series. To this end, we define a sliding window of length P, where the window spanning from the timestamps $t - P$ to $t - 1$, denoted as $\mathbf{H}_{prev} \in \mathbb{R}^{N\times P}$, is used as input, and the window from time t to $t + P - 1$, denoted as $\mathbf{H}_{pred} \in \mathbb{R}^{N\times P}$ is used as the reconstruction target. Traditional Granger causality-based approaches estimate the causal structure among variables using linear regression coefficients $a_{ij}^\tau \in \mathbb{R}$, which can be expressed as:

$$x_t^{(j)} = \sum_{i=1}^{N}\sum_{\tau=1}^{P} a_{ij}^\tau x_{t-\tau}^{(i)} + \epsilon_t^{(j)} \tag{10}$$

While the approach is intuitive and computationally efficient, it has notable limitations. Specifically, it estimates each causal coefficient as a single scalar value, which prevents the model from capturing the variability of the causal effect and assumes a fixed, static structure—even though causal relationships may change over time or under different conditions. As a result, such assumptions can lead to inaccurate inference in dynamic systems. Moreover, real-world systems often involve influences that cannot be explained solely

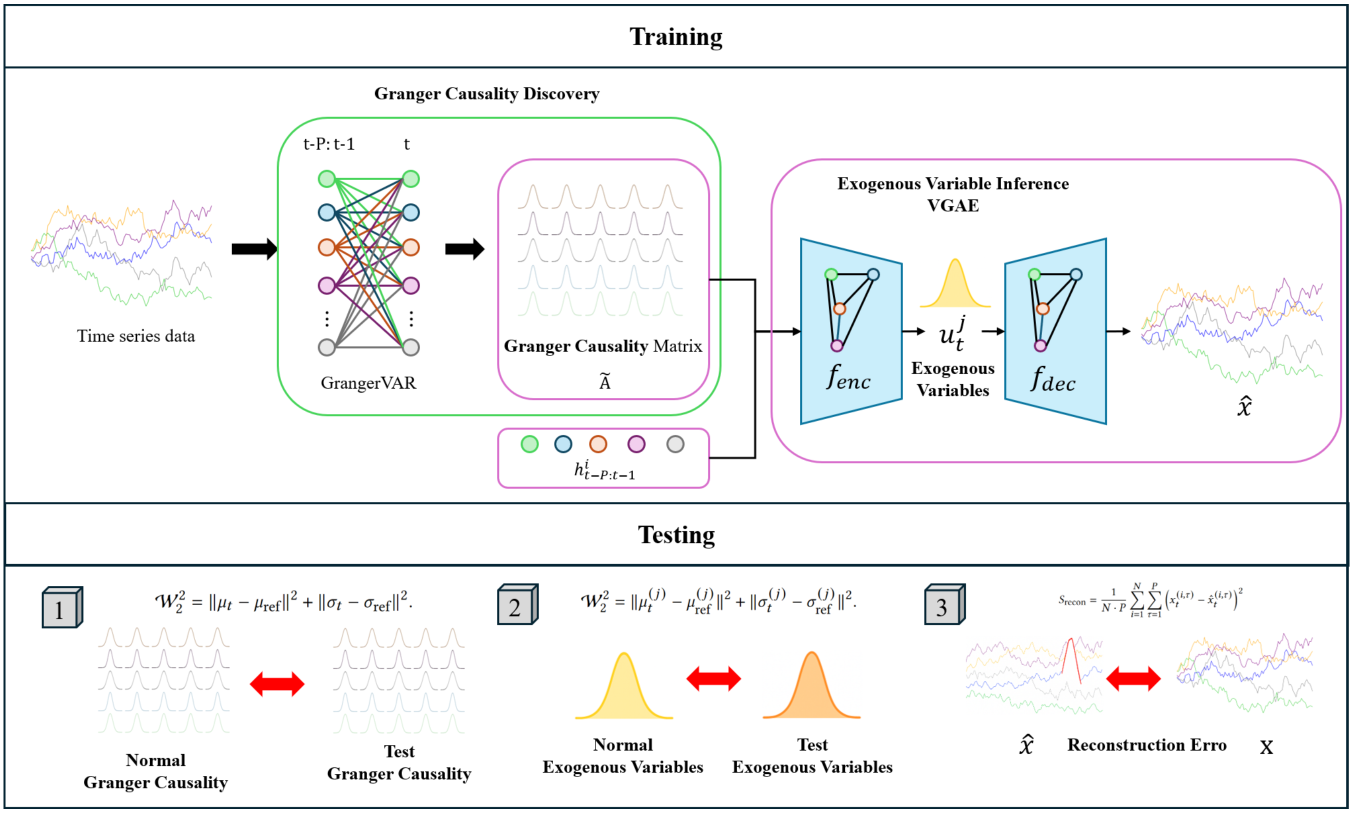

Figure 1: Overview of the proposed GPCDX architecture. During training (top), the model first infers the Granger causality matrix Â from historical time series data using a distributional GrangerVAR module. This matrix captures pairwise causal relationships with uncertainty. The inferred causality, along with temporal features, is then used to infer latent exogenous variables u_t^j via a variational graph autoencoder (VGAE), which is subsequently decoded to reconstruct the input sequence $\hat{x}_t$. During testing (bottom), the model processes anomalous sequences to extract current Granger causality and exogenous variables. These are used to reconstruct the input, and the final anomaly score is computed through the multi-score anomaly detection strategy.

by observed variables. These influences are referred to as exogenous variables, which represent external factors that lie outside the endogenous structure and cannot be captured through internal variable interactions alone. This can be formalized as follows:

$$x_t^{(j)} = f_j(\{x_{t-\tau}^{(i)}\}_{i=1,\ldots N;,\tau=1,\ldots P}) + u_t^{(j)} \tag{11}$$

Where, $u_t^{(j)}$ denotes a latent exogenous factor associated with variable j. Such exogenous factors may arise from externally injected information, abnormal interventions, or unobserved environmental conditions. Ignoring these influences that cannot be explained by the endogenous structure can compromise the interpretability of the model and potentially distort the underlying causal relationships. To overcome this limitation, the proposed approach estimates the Granger causal structure among variables in the form of Gaussian distributions, thereby capturing the variability of the coefficients and accommodating structural diversity. At the same time, it explicitly separates and infers information not explained by the causal structure as latent exogenous variables.

4.2 Granger Causality Discovery: Gaussian Process Regression

Granger causality is a statistical framework used to determine causal relationships between variables in time series data. Specifically, if the past values of a variable $x^{(i)}$ provide significant predictive information for the future values of another variable $x^{(j)}$, then $x^{(i)}$ is said to Granger-cause $x^{(j)}$. Traditional Granger analysis is typically based on the linear Vector Autoregressive (VAR) model, where a multivariate time series $\mathbf{x}_t \in \mathbb{R}^N$ is modeled as equation 3. However, such approaches estimate regression coefficients as fixed linear values and thus cannot capture nonlinear, time-varying causal structures or predictive uncertainty.

To overcome these limitations, we propose a structure where each pair of variables (i, j) is modeled with an independent Gaussian Process (GP) regression. A GP is a nonparametric Bayesian regression model that defines a distribution over functions in function space. For an input $\mathbf{x}_{t-P:t-1}^{(i)} \in \mathbb{R}^P$, the conditional distribution of the output $x_t^{(j)}$ is inferred as:

$$P(x_t^{(j)} \mid \mathbf{x}_{t-P:t-1}^{(i)}) \sim \mathcal{N}(\mu_{ij}(t), \sigma_{ij}^2(t)) \tag{12}$$

Here, $\mu_{ij}(t)$ is the mean prediction and $\sigma_{ij}^2(t)$ denotes the predictive variance, which reflects the uncertainty in the causal influence. This formulation allows us to estimate Granger causality as a time-varying distribution, rather than as a static scalar coefficient as in traditional VAR-based models.

By performing such distributional inference across all variable pairs, we construct a Granger causality matrix $\mathbf{A}_t \in \mathbb{R}^{N\times N}$ at each time step t, defined as:

$$[\mathbf{A}_t]_{ij} = \mu_{ij}(t) \tag{13}$$

The theoretical foundation of GP-based inference lies in its construction of conditional distributions from a joint Gaussian prior. Given training data $(\mathbf{X}, \mathbf{y})$, the joint distribution over observed and test outputs is expressed as:

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{X}) \\ \mu(x^*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) + \sigma_n^2 I & K(\mathbf{X}, x^*) \\ K(x^*, \mathbf{X}) & K(x^*, x^*) \end{bmatrix}\right) \tag{14}$$

This leads to the following predictive distribution for a new input x^*:

$$P(y^* \mid x^*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\mu^*, \sigma^{*2}) \tag{15}$$

The resulting predictive mean and variance can be interpreted as dynamic causal coefficients, which are consistent with the Granger definition of conditional predictive dependence. The GP model is trained by maximizing the marginal log-likelihood, which is given by:

$$\log P(\mathbf{y} \mid \mathbf{X}) = -\frac{1}{2}\mathbf{y}^\top K^{-1}\mathbf{y} - \frac{1}{2}\log|K| - \frac{n}{2}\log(2\pi) \tag{16}$$

This objective incorporates both predictive accuracy and uncertainty calibration, and acts as a regularizer that prevents overfitting and encourages stable learning.

4.3 Exogenous Variable Inference: Variational Graph Autoencoder (VGAE)

In real-world multivariate time series systems, modeling based solely on endogenous Granger causal structures is often insufficient for fully explaining the observed dynamics[40]. Residual variations or irregular fluctuations that cannot be attributed to endogenous relationships may originate from latent external factors or hidden structural influences. In this work, we explicitly model such external influences as latent variables $u_t^{(j)} \in \mathbb{R}^d$ for each node j, and infer them using a Variational Graph Autoencoder (VGAE) architecture.

The VGAE consists of an encoder-decoder structure. The encoder estimates the posterior distribution of the exogenous latent variables, while the decoder utilizes these variables to reconstruct the time series, thereby quantifying the contribution of exogenous information. The goal of the encoder is to approximate the conditional distribution $q(u^{(j)}|\cdot)$ by estimating its mean and variance. For each node j, we use the local past window $h_t^{(j)} \in \mathbb{R}^P$ and summary statistics of the incoming Granger causal coefficients as input features.

Specifically, for a given target node j, the mean of the causal coefficients $\mu_{ij}^{(\tau)}$ is aggregated over all source nodes i and time lags τ to compute the following statistics:

$$\begin{aligned} \mu_{sum}^{(j)} &= \sum_{i=1}^{N}\sum_{\tau=1}^{P} \mu_{ij}^{(\tau)} \\ \log\sigma_{sum}^{(j)2} &= \log\left(\sum_{i=1}^{N}\sum_{\tau=1}^{P} \exp(\log\sigma_{ij}^{(\tau)2})\right) \end{aligned} \tag{17}$$

These values represent the overall strength and uncertainty of the incoming causal information. They are concatenated with the node's local history $h_t^{(j)}$ to form the encoder input $x^{(j)} = [h_t^{(j)}; \mu_{sum}^{(j)}; \log\sigma_{sum}^{(j)2}] \in \mathbb{R}^{P+2}$. Stacking all node features yields the input matrix $X \in \mathbb{R}^{N\times(P+2)}$ for the GCN encoder. The adjacency matrix $A \in \mathbb{R}^{N\times N}$, derived from the average Granger causality structure, is used to model structural dependencies and is symmetrically normalized as $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. The encoder computation is defined as:

For each node j :

$$\begin{aligned} \boldsymbol{\mu}_{:j}(t) &= [\mu_{1j}(t), \mu_{2j}(t), \ldots, \mu_{Nj}(t)]^T \\ \mathbf{x}_{\text{enc}}^{(j)} &= \left[\mathbf{h}_t^{(j)}; \boldsymbol{\mu}_{:j}(t); x_t^{(j)}\right] \in \mathbb{R}^{P+N+1} \\ \mathbf{H}_1 &= \text{ReLU}(\tilde{\mathbf{A}}\mathbf{X}_{\text{enc}}\mathbf{W}_1), \quad \boldsymbol{\mu}_u = \tilde{\mathbf{A}}\mathbf{X}_{\text{enc}}\mathbf{W}_\mu, \\ \log\boldsymbol{\sigma}_u^2 &= \tilde{\mathbf{A}}\mathbf{X}_{\text{enc}}\mathbf{W}_{\log\sigma} \end{aligned} \tag{18}$$

where $W_1, W_\mu, W_{\log\sigma}$ are learnable parameters for hidden representation, mean, and variance prediction, respectively. This results in the posterior distribution for the exogenous variable of each node:

$$q(u_t^{(j)}) = \mathcal{N}(\mu_u^{(j)}, \sigma_u^{(j)2}) \tag{19}$$

Latent variables are sampled using the reparameterization trick as:

$$u_t^{(j)} = \mu_u^{(j)} + \sigma_u^{(j)} \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{20}$$

The sampled latent variables are then passed to the decoder for reconstruction. The decoder input is similar in structure to the encoder input but includes the sampled exogenous variable $u_t^{(j)}$. Formally, the decoder input for each node is defined as:

$$z^{(j)} = [h_t^{(j)}; u_t^{(j)}; \mu_{sum}^{(j)}; \log\sigma_{sum}^{(j)2}] \in \mathbb{R}^{P+d+2} \tag{21}$$

Stacking all decoder inputs yields $Z \in \mathbb{R}^{N\times(P+d+2)}$, which is processed as follows:

$$H_2 = \text{ReLU}(\tilde{A}ZW_2), \quad \hat{H} = \tilde{A}H_2W_{\text{out}} \tag{22}$$

The output $\hat{H}$ represents the reconstructed time series, which is compared to the true values x_t using the Mean Squared Error (MSE) loss:

$$L_{\text{recon}} = \sum_{j=1}^{N} \left\|\hat{x}_t^{(j)} - x_t^{(j)}\right\|^2 \tag{23}$$

To regularize the latent space and constrain the information content of the exogenous variables, we apply a KL divergence term between the posterior $q(u_t^{(j)})$ and the prior $p(u_t^{(j)}) = \mathcal{N}(0, I)$:

$$L_{\text{exo\_KL}} = \sum_{j=1}^{N} \text{KL}\left[\mathcal{N}(\mu_u^{(j)}, \sigma_u^{(j)2}) \| \mathcal{N}(0, I)\right] \tag{24}$$

This KL term prevents the exogenous variable distribution from excessive collapse or dispersion, under the assumption that in normal conditions, exogenous factors should follow a stable prior with moderate uncertainty.

4.4 Distribution-based Anomaly Scoring and Decision Rule

Our framework employs distributional representations to detect anomalies in multivariate time series. At each time step, the model infers three types of statistical quantities, and corresponding anomaly scores are computed based on deviations from the learned reference distributions:

- **Reconstruction Score:** The mean squared error (MSE) between the input sequence x_t and its reconstructed version $\hat{x}_t$, defined as
$$S_{\text{recon}} = \frac{1}{N \cdot P} \sum_{i=1}^{N} \sum_{\tau=1}^{P} \left(x_t^{(i,\tau)} - \hat{x}_t^{(i,\tau)}\right)^2 .$$
- **Granger Score:** The 2-Wasserstein distance between the inferred Granger causality distribution (μ_t, σ_t) and the reference distribution $(\mu_{\text{ref}}, \sigma_{\text{ref}})$, given by
$$\mathcal{W}_2^2 = \|\mu_t - \mu_{\text{ref}}\|^2 + \|\sigma_t - \sigma_{\text{ref}}\|^2.$$
- **Exogenous Score:** The 2-Wasserstein distance between the inferred latent exogenous variable distribution $(\mu_t^{(j)}, \sigma_t^{(j)})$ and the corresponding reference distribution $(\mu_{\text{ref}}^{(j)}, \sigma_{\text{ref}}^{(j)})$, computed as
$$\mathcal{W}_2^2 = \|\mu_t^{(j)} - \mu_{\text{ref}}^{(j)}\|^2 + \|\sigma_t^{(j)} - \sigma_{\text{ref}}^{(j)}\|^2.$$

Each score captures different aspects of the system's behavior: the reconstruction fidelity, the stability of causal relationships, and the behavior of unobserved latent influences. To determine whether a time window is anomalous, we adopt a rule-based decision mechanism. Specifically, we use a **2-out-of-3 rule**, where a sample is labeled as anomalous if at least two of the three scores exceed their respective thresholds. This strategy enhances robustness against false positives and enables detection of diverse anomaly types.

KL Divergence for Training vs. Wasserstein Distance for Inference. During training, the distributions over Granger coefficients and exogenous variables are regularized toward a standard normal prior $\mathcal{N}(0, I)$ using the Kullback-Leibler (KL) divergence. KL divergence is effective in encouraging compact and normalized latent representations by minimizing the divergence between the approximate posterior and the prior.

However, during inference, the goal shifts to quantifying the deviation of the inferred distribution from a reference distribution that characterizes the normal state. In this setting, KL divergence becomes less suitable due to its asymmetry and numerical instability when distributions have disjoint supports. In contrast, the **Wasserstein distance** offers a stable and interpretable metric that accounts for both the mean shift and the scale change between distributions. It always yields finite values and enables direct comparison even when distributions do not overlap. Therefore, this study employs KL divergence during the training phase to encourage convergence toward a standard normal distribution, and uses Wasserstein distance during the testing phase as a distance-based metric for anomaly detection.

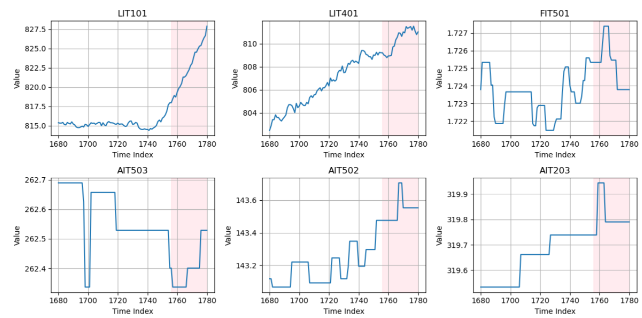

Figure 2: Visualization of sensor time series from the SWAT dataset during a real anomaly interval. Each subplot shows the readings from different sensors (e.g., LIT101, LIT401, FIT501), and the shaded pink regions indicate the ground-truth anomaly periods. Noticeable changes in trend, variance, or value range are observed within these intervals, indicating sensor responses to anomalous events.

5 EXPERIMENTS

To evaluate the structural anomaly detection performance of the proposed framework, we conducted experiments using various multivariate time series datasets. This evaluation focuses on verifying how effectively the distribution-based Granger causality approach can detect structural anomalies in real-world time series environments. The primary goal is to assess the interpretability, scalability, and sensitivity to high-dimensional relational structures. In this section, we sequentially present the used datasets, experimental settings, comparative results, and corresponding analysis.

5.1 Datasets

To evaluate the effectiveness and generalizability of the proposed framework, we conduct experiments on four widely used real-world multivariate time series anomaly detection benchmarks: SMD[34], MSL[15], PSM[1], and SWAT[24].

- **SMD (Server Machine Dataset)**: This dataset consists of sensor measurements collected from 28 different servers in a large Internet company. Each instance contains 38 variables including CPU usage, memory, disk I/O, and network traffic. Anomalies are injected based on realistic server failure patterns.
- **MSL (Mars Science Laboratory)**: Provided by NASA, this dataset contains telemetry data collected from the Curiosity Rover. It consists of 55-dimensional time series sequences with various system-level features and contains annotated

	SMD (F1 / AUC)	MSL (F1 / AUC)	PSM (F1 / AUC)	SWAT (F1 / AUC)
Our	0.417 / 0.786	0.801 / **0.870**	0.544 / **0.758**	0.769 / **0.823**
CATCH[37]	0.847 / **0.809**	0.693 / 0.663	0.781 / 0.650	0.736 / 0.344
MEMTO[33]	0.833 / 0.462	0.907 / 0.500	0.963 / 0.504	**0.928** / 0.741
DiffAD[39]	**0.904** / 0.656	**0.918** / 0.626	**0.979** / 0.451	0.874 / 0.612
iTrans[20]	0.812 / 0.781	0.710 / 0.611	0.854 / 0.592	0.718 / 0.242
DualTF[26]	0.679 / 0.631	0.855 / 0.576	0.725 / 0.600	0.695 / 0.567

Table 1: Performance Comparison (F1-score / AUROC) of Anomaly Detection Models on Four Benchmark Datasets

anomalies related to component failures or unusual environmental conditions.

- **PSM (Power System Monitoring)**: A large-scale dataset collected from a real power plant environment. It includes 25 sensor signals monitoring system-level electrical and mechanical states. The anomalies reflect faults and operational abnormalities under real deployment conditions.
- **SWAT (Secure Water Treatment)**: A cyber-physical system dataset from a water treatment testbed. It includes 51 physical and logical sensor values representing levels, flows, pressures, and valve/pump statuses. The dataset contains both natural and attack-induced anomalies reflecting diverse real-world failure modes.

Each dataset consists of multivariate time series data with distinct sensor counts, sampling rates, and anomaly characteristics. All datasets include ground-truth labels for anomaly intervals and are widely adopted as standard benchmarks for evaluating unsupervised anomaly detection models. Figure 2 visualizes time series from key sensors (e.g., LIT101, LIT401, FIT501) during a real anomaly interval in the SWAT dataset. Each plot illustrates the temporal evolution of sensor readings, with the shaded pink regions indicating the labeled anomaly periods. Within these intervals, noticeable changes in trend, variance, and value range are observed, indicating that the sensors are responding to anomalous events.

Dataset	Sensors	train	test	anomalies
SWaT	51	47,520	44,991	12.20%
SMD	38	28,479	28,479	9.46%
MSL	55	3,682	2,856	0.74%
PSM	25	132,481	87,841	27.76%

Table 2: Number of Sensors and Anomaly Ratio per Dataset

5.2 Experimental Setup

For each dataset, we apply a fixed sliding window approach with a window size of 100. During training, only normal samples are used to fit the model. The test phase involves both normal and anomalous sequences, where the model outputs anomaly scores at each window. The Granger inference module is instantiated using Gaussian Process regression with RBF kernels. To compare temporal dependencies across variables, we model each pair-wise interaction $(i \leftarrow j)$ as an independent GP. The exogenous inference module uses a graph-based variational autoencoder with adjacency weights dynamically constructed from the inferred Granger structure. All models are trained using the Adam optimizer with a learning rate of 1e-3, batch size of 256, and early stopping based on validation reconstruction loss. The model is implemented using PyTorch and trained on NVIDIA A5000. Anomaly scores are calculated using three metrics: reconstruction error, Granger-based distributional shift, and exogenous deviation.

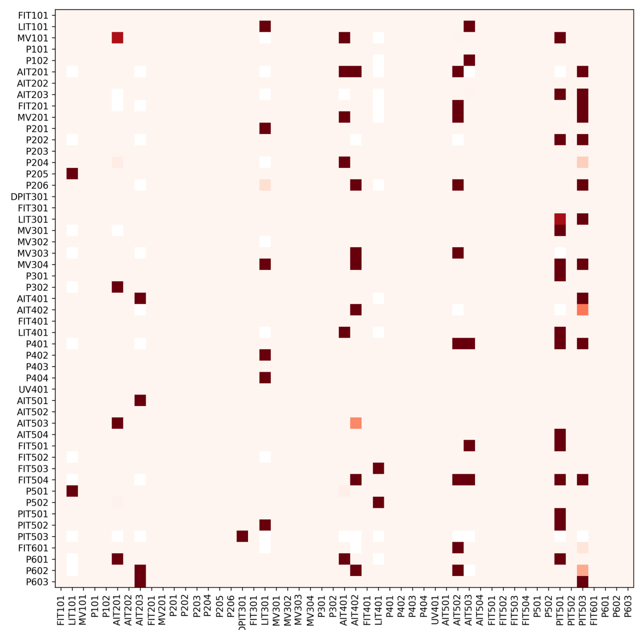

Figure 3: Deviation map showing the Wasserstein distance between the reference Granger causality distribution (from normal data) and the Granger distribution inferred from the test window shown in Figure 2. Each cell represents the deviation between a pairwise causality link $(i \leftarrow j)$. Darker colors indicate greater deviation, highlighting which causal relationships are disrupted during the anomaly period. This map provides interpretable insight into which sensor pairs experienced abnormal changes in temporal dependencies.

5.3 Experimental Results

To evaluate the effectiveness of the proposed distribution-based causal inference framework, we conducted experiments on four representative multivariate time series anomaly detection benchmark datasets. The performance of our model was compared against major baselines, using two key metrics: F1-score, which reflects both precision and recall, and AUROC (Area Under the Receiver Operating Characteristic Curve), which measures overall discrimination ability between normal and anomalous intervals. The experiment has yielded results, which can be explained in a three part analysis: analysis on performance comparison, analysis on sensor-performance relationship, and the analysis on the effects of distribution based representation.

5.3.1 ***Analysis on performance comparison.*** As shown in Table 1, the proposed model achieved notably high AUROC scores on the MSL (0.870) and SWAT (0.823) datasets, outperforming or closely matching existing methods. This indicates that the model is capable of sensitively capturing structural deviations between normal and abnormal system states. Particularly in the SWAT dataset, which involves complex interactions among numerous sensors, the model maintained stable and reliable detection performance. A key observation is that AUROC performance improved with the number of sensors (i.e., variables).

5.3.2 ***Analysis on sensor-performance relationship.*** According to Table 2, the number of sensors varies across datasets—MSL (55), SWAT (51), SMD (38), and PSM (25). The model achieved stronger performance on MSL and SWAT, while performance was comparatively lower on SMD (0.786) and PSM (0.704). This trend suggests that as the number of interacting variables increases, the model is able to learn richer Granger causal structures[2][4] and detect structural shifts more accurately. This result does not merely reflect increased computational effort due to higher dimensionality, but rather highlights that the expressive capacity of the learned causal distribution grows with dimensionality. As the latent structure becomes more complex, the model gains greater robustness against noise, achieving structural stability in its inference. In other words, in high-dimensional multivariate settings, the relational structure itself becomes the key signal for anomaly detection. The proposed model effectively tracks this structural movement in a probabilistic manner, thereby achieving stronger detection capability. Conversely, when the number of sensors is small, the model has access to fewer causal pathways, resulting in lower-dimensional distributions that are more susceptible to noise. As a result, structural anomalies become harder to isolate, and detection performance degrades accordingly. This performance drop, reflected in lower AUROC scores on datasets like PSM, empirically supports the theoretical limitations of low-dimensional causal inference in complex systems.[3][13]

5.3.3 ***Analysis on distribution based representations.*** Unlike prior approaches that rely solely on reconstruction error or attention-based representations, our model quantifies time-dependent causal relationships between variable pairs (i, j) in the form of probability distributions. Anomalies are detected not through direct deviations in raw values, but via meaningful deviations in the structural properties of these distributions[35]. This enables the model to capture even subtle and nuanced signs of abnormal behavior in complex systems, offering a higher level of sensitivity and interpretability. On the other hand, the F1-score was relatively low for some datasets (SMD: 0.417, PSM: 0.544). This can be attributed to the fixed-threshold binarization process[27][41], where short-lived anomalies or marginal anomalies near the decision boundary may not be adequately captured. Since the proposed model detects anomalies based on structural changes in probabilistic distributions, it does not operate within a fully optimized framework for conventional threshold-based decision-making. Specifically, structural deviations inferred by the model require carefully calibrated binarization rules to be interpreted as definitive anomalies. However, such mappings are not yet fully established in this study, resulting in occasional missed detections. This represents a clear performance limitation, not of the model architecture itself, but of the interpretational mismatch between the novel distribution-based scoring scheme and traditional binary classification systems. Future research may address this limitation by introducing adaptive thresholding, distribution-aware binarization, or context-sensitive post-processing techniques to enhance detection precision. In the contrast, this study departs from conventional Granger causality approaches that rely on scalar regression coefficients and proposes a novel framework that models causal structures as probability distributions. This allows the model to capture not only the strength of causal relationships but also their uncertainty, enabling it to detect structural transitions in the system with high sensitivity. By redefining the objective of anomaly detection from identifying abnormal behavior to changes in relationships between variables, the proposed method shifts the focus from value-level irregularities to underlying structural transitions and relational distortions. This conceptual advancement offers new opportunities for system diagnostics, root cause analysis, and causal graph-based interpretability.

In summary, despite some limitations in threshold-sensitive metrics such as F1-score, the proposed model demonstrates strong potential and interpretability for structure-aware anomaly detection in complex multivariate time series. Its robust performance in datasets with high sensor density and strong inter-variable dependencies suggests promising applicability in large-scale sensor networks and industrial IoT systems, where structural changes among variables are critical to anomaly understanding.

	SWAT (F1 / AUC)	PSM (F1 / AUC)
1-out-of-3	0.729 / **0.823**	**0.544** / **0.758**
2-out-of-3	**0.769** / **0.823**	0.398 / **0.758**
All-3	0.050 / **0.823**	0.081 / **0.758**
Combine	0.720 / 0.796	0.504 / 0.576

Table 3: We compare four decision rules for anomaly detection: 1-out-of-3 and 2-out-of-3 flag anomalies if one or two of the three scores (Granger, Exogenous, Reconstruction) exceed thresholds. All-3 requires all three scores to exceed thresholds. Combine sums the deviations from all scores into a single value for detection.

5.4 Evaluation of Multi-Score Anomaly Detection Strategies

In this study, we experimentally compared four decision-making strategies for detecting anomalies based on three distribution-based scores: Reconstruction, Granger, and Exogenous. These strategies are defined as follows: (1) **1-out-of-3** determines an anomaly if at least one of the three scores exceeds its threshold; (2) **2-out-of-3** flags an anomaly if two or more scores exceed their thresholds, serving as a moderate rule; (3) **All-3** is the most conservative, requiring all three scores to exceed thresholds for anomaly detection; and (4) **Combine** computes the sum of deviations (shifts) from all three scores into a single composite score for binarization. As shown in Table 3, the 2-out-of-3 strategy achieved the highest F1-score in the SWAT dataset, which has a large number of sensors and complex inter-variable dependencies. This suggests that aggregating multiple structural signals is more effective in capturing anomalies in high-dimensional systems. In contrast, the 1-out-of-3 strategy yielded the best performance on PSM, which has fewer sensors, indicating that in simpler systems, a single strong anomaly signal may be sufficient for detection. This trend implies that the more complex the system structure and the higher the number of variables, the more effective multi-criteria strategies become. In such settings, models must capture multi-layered structural changes, making ensemble-based or conservative decision strategies more appropriate. Interestingly, the Combine strategy resulted in the lowest F1-score across datasets. This outcome highlights that simply summing the three scores disregards their distinct statistical meanings—such as causal strength, uncertainty, and exogeneity—leading to diluted information and reduced sensitivity. Especially in distribution-based anomaly detection, accurate results are not only dependent on the magnitude of the scores, but are also dependent on their movement (distribution shift) and spread (variance). The Combine strategy, which fails to preserve the relative structure between scores, is empirically shown to degrade detection performance in this context. Overall, the experiment demonstrates that decision strategies considering combinations of structural shifts are more effective than relying on a single unified score in distribution-based structural anomaly detection. These findings emphasize the need for adaptive decision mechanisms or context-aware rule selection, where the interplay between score semantics and anomaly types is taken into account. Ultimately, this suggests that anomaly detection should evolve beyond simple thresholding and adopt more refined reasoning frameworks that preserve structural interpretability and multi-level anomaly dynamics.

6 CONCLUSION

In this study, we proposed a novel anomaly detection framework, GPCDX, which models Granger causality not as a single scalar coefficient but as a probability distribution, while simultaneously inferring latent exogenous variables through a Variational Graph Autoencoder. The proposed framework focuses on detecting structural anomalies in multivariate time series data by shifting the perspective from value-based anomaly detection to one that emphasizes changes in inter-variable relational structures. By quantitatively capturing both the strength and uncertainty of causal relationships, the model is capable of sensitively detecting structural transitions within a system. Moreover, by inferring exogenous factors that cannot be explained through endogenous interactions, the model enables structural interpretation of anomalies that would otherwise remain undetected by conventional approaches. Experimental results on four widely-used benchmark datasets—SMD, MSL, PSM, and SWAT—demonstrated that the proposed model achieves high AUROC performance, particularly on high-dimensional datasets with a large number of sensors such as MSL and SWAT. This validates the model's scalability and its ability to detect structurally-driven anomalies. Ultimately, this work introduces a new direction for anomaly detection by modeling causal relations as probability distributions, thereby expanding both the interpretability and inferential capacity of structure-based detection. By focusing on changes in relational structures rather than merely detecting abnormal values, the proposed model lays a promising foundation for a wide range of applications including root cause analysis, causal inference, and intelligent monitoring of complex real-world sensor systems characterized by rich interactions.

7 GENAI USAGE DISCLOSURE

The overall code development and manuscript preparation were carried out by the authors; however, manuscript proofreading and translation were performed using OpenAI's ChatGPT, and code refactoring was assisted by Cursor.

REFERENCES

[1] Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki. 2021. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2485–2494.

[2] Sipan Aslan and Hernando Ombao. 2024. Granger Causality in High-Dimensional Networks of Time Series. *arXiv preprint arXiv:2406.02360* (2024).

[3] Ane Blázquez-García, Angel Conde, Usue Mori, and Jose A Lozano. 2021. A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR)* 54, 3 (2021), 1–33.

[4] Ruichu Cai, Yunjin Wu, Xiaokai Huang, Wei Chen, Tom ZJ Fu, and Zhifeng Hao. 2024. Granger causal representation learning for groups of time series. *Science China Information Sciences* 67, 5 (2024), 152103.

[5] Manuel Castro, Pedro Ribeiro Mendes Júnior, Aurea Soriano-Vargas, Rafael de Oliveira Werneck, Maiara Moreira Gonçalves, Leopoldo Lusquino Filho, Renato Moura, Marcelo Zampieri, Oscar Linares, Vitor Ferreira, et al. 2023. Time series causal relationships discovery through feature importance and ensemble models. *Scientific reports* 13, 1 (2023), 11402.

[6] Ching Chang, Chiao-Tung Chan, Wei-Yao Wang, Wen-Chih Peng, and Tien-Fu Chen. 2024. TimeDRL: Disentangled Representation Learning for Multivariate Time-Series. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 625–638.

[7] Yuxiao Cheng, Runzhao Yang, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. 2023. Cuts: Neural causal discovery from irregular time-series data. *arXiv preprint arXiv:2302.07458* (2023).

[8] Ivan Cherednikov. 2023. *Anomaly detection in industrial time series sensor data*. Master's thesis. Norwegian University of Life Sciences, Ås.

[9] Ailin Deng and Bryan Hooi. 2021. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4027–4035.

[10] Wiebke Günther, Urmi Ninad, and Jakob Runge. 2023. Causal discovery for time series from multiple datasets with latent contexts. In *Uncertainty in Artificial Intelligence*. PMLR, 766–776.

[11] Xiao Han, Saima Absar, Lu Zhang, and Shuhan Yuan. 2025. Root Cause Analysis of Anomalies in Multivariate Time Series through Granger Causal Discovery. In *The Thirteenth International Conference on Learning Representations*.

[12] Uzma Hasan, Emam Hossain, and Md Osman Gani. 2023. A survey on causal discovery methods for iid and time series data. *arXiv preprint arXiv:2303.15027* (2023).

[13] Lukas Heppel, Andreas Gerhardus, Ferdinand Rewicki, Jan Deeken, and Günther Waxenegger-Wilfing. 2024. Leveraging Causal Information for Multivariate Timeseries Anomaly Detection. In *35th International Conference on Principles of Diagnosis and Resilient Systems (DX 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 11–1.

[14] Sanqing Hu and Hualou Liang. 2012. Causality analysis of neural connectivity: New tool and limitations of spectral Granger causality. *Neurocomputing* 76, 1 (2012), 44–47.

[15] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. 2018. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 387–395.

[16] SHERIF KAMEL Hussein and MOHAMED A El-Dosuky. 2023. Anomaly detection in cyber-physical systems using explainable artificial intelligence and machine learning. *Journal of Theoretical and Applied Information Technology* 101, 8 (2023), 0–0.

[17] Tung Kieu, Bin Yang, Chenjuan Guo, and Christian S Jensen. 2019. Outlier detection for time series with recurrent autoencoder ensembles.. In *Ijcai*. 2725–2732.

[18] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).

[19] Shuyu Lin, Ronald Clark, Robert Birke, Sandro Schönborn, Niki Trigoni, and Stephen Roberts. 2020. Anomaly detection for time series using vae-lstm hybrid model. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 4322–4326.

[20] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. 2023. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625* (2023).

[21] Zehao Liu, Mengzhou Gao, and Pengfei Jiao. 2025. GCAD: Anomaly Detection in Multivariate Time Series from the Perspective of Granger Causality. *arXiv preprint arXiv:2501.13493* (2025).

[22] Zhe Liu, Xiang Huang, Jingyun Zhang, Zhifeng Hao, Li Sun, and Hao Peng. 2024. Multivariate time-series anomaly detection based on enhancing graph attention networks with topological analysis. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 1555–1564.

[23] Lauri Lovén, Virve Karsisto, Heikki Järvinen, Mikko J Sillanpää, Teemu Leppänen, Ella Peltonen, Susanna Pirttikangas, and Jukka Riekki. 2019. Mobile road weather sensor calibration by sensor fusion and linear mixed models. *PloS one* 14, 2 (2019), e0211702.

[24] Aditya P Mathur and Nils Ole Tippenhauer. 2016. 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)*. IEEE, 31–36.

[25] Muhammad Abdan Mulia, Muhammad Bintang Bahy, Muhammad Zain Fawwaz Nuruddin Siswantoro, Nur Rahmat Dwi Riyanto, Nella Rosa Sudianjaya, and Ary Mazharuddin Shiddiqi. 2024. Kbjnet: Kinematic bi-joint temporal convolutional network attention for anomaly detection in multivariate time series data. *Data Science Journal* 23, 1 (2024).

[26] Youngeun Nam, Susik Yoon, Yooju Shin, Minyoung Bae, Hwanjun Song, Jae-Gil Lee, and Byung Suk Lee. 2024. Breaking the time-frequency granularity discrepancy in time-series anomaly detection. In *Proceedings of the ACM Web Conference 2024*. 4204–4215.

[27] Minjae Ok, Simon Klüttermann, and Emmanuel Müller. 2024. Exploring the impact of outlier variability on anomaly detection evaluation metrics. *arXiv preprint arXiv:2409.15986* (2024).

[28] Huida Qiu, Yan Liu, Niranjan A Subrahmanya, and Weichang Li. 2012. Granger causality for time-series anomaly detection. In *2012 IEEE 12th international conference on data mining*. IEEE, 1074–1079.

[29] Matthias Seeger. 2004. Gaussian processes for machine learning. *International journal of neural systems* 14, 02 (2004), 69–106.

[30] Ali Shojaie and Emily B Fox. 2022. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application* 9, 1 (2022), 289–319.

[31] Ghazi Shukur and Panagiotis Mantalos. 2000. A simple investigation of the Granger-causality test in integrated-cointegrated VAR systems. *Journal of Applied Statistics* 27, 8 (2000), 1021–1031.

[32] Chathurangi Shyalika, Harleen Kaur Bagga, Ahan Bhatt, Renjith Prasad, Alaa Al Ghazo, and Amit Sheth. 2024. Time Series Foundational Models: Their Role in Anomaly Detection and Prediction. *arXiv preprint arXiv:2412.19286* (2024).

[33] Junho Song, Keonwoo Kim, Jeonglyul Oh, and Sungzoon Cho. 2023. Memto: Memory-guided transformer for multivariate time series anomaly detection. *Advances in Neural Information Processing Systems* 36 (2023), 57947–57963.

[34] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2828–2837.

[35] Kai Ming Ting, Zongyou Liu, Hang Zhang, and Ye Zhu. 2022. A new distributional treatment for time series and an anomaly detection investigation. *Proceedings of the VLDB Endowment* 15, 11 (2022), 2321–2333.

[36] Axel Wismüller, Adora M Dsouza, M Ali Vosoughi, and Anas Abidin. 2021. Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data. *Scientific reports* 11, 1 (2021), 7817.

[37] Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. 2024. Catch: Channel-aware multivariate time series anomaly detection via frequency patching. *arXiv preprint arXiv:2410.12261* (2024).

[38] Shengjie Xia, Wu Sun, Xiaofeng Zou, Panfeng Chen, Dan Ma, Huarong Xu, Mei Chen, and Hui Li. 2024. MFAM-AD: an anomaly detection model for multivariate time series using attention mechanism to fuse multi-scale features. *PeerJ Computer Science* 10 (2024), e2201.

[39] Chunjing Xiao, Zehua Gou, Wenxin Tai, Kunpeng Zhang, and Fan Zhou. 2023. Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 2742–2751.

[40] Wenzhuo Yang, Kun Zhang, and Steven CH Hoi. 2022. A causal approach to detecting multivariate time-series anomalies and root causes. *arXiv preprint arXiv:2206.15033* (2022).

[41] Chaoli Zhang, Yingying ZHANG, Qingsong Wen, Lanshu Peng, Yiyuan Yang, Chongjiong Fan, Minqi Jiang, Lunting Fan, and Liang Sun. [n. d.]. Benchmarking Multivariate Time Series Anomaly Detection with Large-Scale Real-World Datasets. ([n. d.]).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009