DCD: Decomposition-based Causal Discovery from Autocorrelated and Non-Stationary Temporal Data

Anonymous authors
Paper under double-blind review

Abstract

Multivariate time series in domains such as finance, climate science, and healthcare often exhibit long-term trends, seasonal patterns, and short-term fluctuations, complicating causal inference under non-stationarity and autocorrelation. Existing causal discovery methods typically operate on raw observations, making them vulnerable to spurious edges and misattributed temporal dependencies. We introduce a decomposition-based causal discovery framework that separates each time series into trend, seasonal, and residual components and performs component-specific causal analysis. Trend components are assessed using stationarity tests, seasonal components using kernel-based dependence measures, and residual components using constraint-based causal discovery. The resulting component-level graphs are integrated into a unified multi-scale causal structure. This approach isolates long- and short-range causal effects, reduces spurious associations, and improves interpretability. Across extensive synthetic benchmarks and real-world climate data, our framework more accurately recovers ground-truth causal structure than state-of-the-art baselines, particularly under strong non-stationarity and temporal autocorrelation. Code is available at `https://github.com/noname2122/DCD-TMLR-2025`.

1 Introduction

Multivariate time series are fundamental across domains such as economics, environmental science, finance, and healthcare, where understanding causal relationships is crucial for forecasting, decision-making, and policy design. These datasets often contain long-term trends, seasonal patterns, and short-term residuals. Identifying causal relationships within these components provides insights into economic indicators, climate variability, and health outcomes. However, existing causal discovery (CD) methods struggle with the intricate temporal dependencies in multivariate time series, failing to disentangle long-term trends, seasonal variations, and short-term effects. This limitation leads to inaccurate causal attributions and spurious relationships.

Many algorithms cannot reliably distinguish between long-term causal effects, seasonal dependencies, and transient interactions, limiting their ability to identify accurate and meaningful causal structure (Granger, 1969; Spirtes et al., 2001; Runge, 2020). While deep learning architectures have significantly advanced time series forecasting by capturing long-term dependencies, they primarily focus on correlation (Liu et al., 2024; Piao et al., 2024; Wen et al., 2022). Consequently, the gap between predictive modeling and causal understanding remains a major challenge, particularly in distinguishing spurious correlations from causal influences (Vaswani et al., 2017; Wu et al., 2021; Dhaou et al., 2021).

To address these challenges, we propose a novel **decomposition-based causal discovery (DCD) framework**. By systematically decomposing time series into its fundamental component into trend, seasonality, and residuals, our approach enables targeted causal discovery at multiple temporal scales. This structured decomposition reduces spurious correlations and enhances the ability to identify complex causal relationships. Specifically, our approach applies **stationarity tests** (Kwiatkowski et al., 1992b) to the trend component to assess long-term dependencies, **kernel-based dependence measures** (Scholkopf & Smola, 2018) to the seasonal component to identify cyclic dependencies, and **constraint-based causal discovery** (Runge, 2020) to the residual component to identify short-term causal effects. The individual causal graphs derived from

these components are then integrated into a unified causal graph, capturing both long-term and short-term causal relationships. Our key contributions are as follows:

- **Decomposition-based Causal Discovery:** We introduce a structured framework that systematically separates and analyzes trend, seasonal, and residual components, enabling precise causal discovery.
- **Component-Specific Causal Discovery:** Most CD methods operate on raw time series where trend and seasonal structure are mixed with short-term dynamics, creating spurious dependencies and violating stationarity assumptions. We separate these effects and apply targeted CD approaches to each component, enabling accurate identification of long-term, periodic, and short-term causal relationships.
- **Empirical Validation:** We conduct extensive evaluations on synthetic and real-world datasets, demonstrating superior CD performance compared to state-of-the-art methods such as PCMCI+(Runge, 2020) and CD-NOD (Huang et al., 2020).

Experimental results confirm that our decomposition-based framework effectively disentangles causal relationships, reducing spurious associations while preserving essential causal dependencies. The approach provides actionable insights across domains such as finance, climate science, and healthcare. This decomposition-based approach bridges the gap between time series forecasting and CD, offering a more nuanced understanding of temporal causality in complex multivariate time series.

2 Related Work

This work intersects with several key areas of research: time series decomposition for isolating distinct temporal components, CD methods in time series, and emerging approaches for integrating these methods with deep learning models.

Time Series Decomposition and Representation Learning. Classical decomposition techniques aim to separate a time series into trend, seasonal, and irregular components to simplify analysis (Cleveland et al., 1990). Methods such as STL and moving averages (Cleveland et al., 1990; Slutzky, 1937) assume relatively smooth or linear behavior and can struggle in multivariate or nonlinear settings. More flexible signal-processing approaches, including Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) (Huang et al., 1998; Dragomiretskiy & Zosso, 2013), extract intrinsic mode functions that capture nonstationary and oscillatory patterns. However, these methods are primarily oriented toward frequency analysis rather than uncovering underlying causal structure.

Recent deep learning architectures further advance decomposition by learning multiscale temporal representations. Attention-based models such as Transformers (Vaswani et al., 2017), Autoformer (Wu et al., 2021), and Temporal Fusion Transformers (Lim et al., 2021) leverage self-attention or gating mechanisms to model long-range patterns, while N-BEATS (Oreshkin et al., 2019) explicitly decomposes series into trend and seasonal blocks. Although such architectures achieve state-of-the-art forecasting performance, they are not designed to identify causal relations: attention weights capture correlations or importance scores, but do not guarantee causal effects or distinguish spurious dependencies induced by trends and seasonal cycles.

Causal Discovery in Time Series Data. Causal discovery methods seek to uncover causal structures within observational time series data, which is challenging due to autocorrelation, temporal dependencies, and non-stationarity (Runge et al., 2019a). Traditional approaches such as Granger Causality (Granger, 1969; Zhang et al., 2024) assess causality by determining whether one time series can improve predictions of another. Although widely used, Granger causality assumes linear relationships and stationarity, leading to inaccuracies when applied to nonlinear or non-stationary time series data.

Constraint-based methods, including the PC Algorithm (Peters et al., 2017) and its variants, such as Fast Causal Inference (FCI) (Peters et al., 2017), and RFCI (Colombo et al., 2012), and aim to identify causal structures using conditional independence tests. While effective in certain cases, these methods often struggle

with high-dimensionality and autocorrelation in time series data, which can lead to spurious edges or missed causal links in complex datasets. Greedy search algorithms like GES and FGES (Chickering, 2002; Ramsey et al., 2017) have also been applied to explore possible causal structures, though they face similar challenges in handling dynamic dependencies within time series. Recent methods tailored for time series include PCMCI+ (Runge et al., 2019a), DYNOTEARS(Pamfil et al., 2020) and CDANs (Ferdous et al., 2023) which improve causal inference by distinguishing between lagged and instantaneous effects. PCMCI+ has been particularly effective in high-dimensional data, yet it does not incorporate a decomposition of the time series, meaning it may still overlook causal relationships if the data contains long-term trends or seasonal components that affect its temporal structure. This limitation suggests the need for approaches that explicitly consider component-specific causal relationships before constructing the final causal graph.

A key gap remains in integrating decomposition with advanced causal inference for multivariate time series. By separating each variable into trend, seasonal, and residual components and inferring causal relationships within each, spurious links induced by strong seasonality or drifting trends are reduced. Our work addresses this by proposing a unified pipeline that leverages the interpretability of decomposition while integrating CD across all timescales and frequency components, yielding a comprehensive structure that captures both short- and long-term dependencies.

3 Theoretical Framework and Methodology

In this section, we first formalize the problem of causal discovery in non-stationary multivariate time series and establish the theoretical conditions under which the true causal structure is identifiable. We then present the Decomposition-based Causal Discovery (DCD) framework, which is designed to satisfy these conditions through spectral separation and component-specific inference.

3.1 Theoretical Guarantees and Identifiability

We model the multivariate time series $\mathbf{X}(t) \in \mathbb{R}^d$ as a superposition of distinct stochastic processes operating at separable timescales. We formally define the conditions under which component-wise causal discovery recovers the true edges of $G^\star$ and quantify the error induced by imperfect decomposition (leakage).

3.1.1 Formal Problem Setup

Let $\mathbf{X}(t) = \{X_1(t), \dots, X_d(t)\}$ be a multivariate stochastic process indexed by $t \in \mathbb{Z}$. We assume the data generating process follows a structural causal model (SCM) adapted to the additive decomposition:

$$X_i(t) = T_i(t) + S_i(t) + R_i(t), \quad \forall i \in \{1, \dots, d\} \tag{1}$$

where T_i, S_i, R_i denote the trend, seasonal, and residual components, respectively.

Let $G^\star = (V, E^\star)$ be the directed acyclic graph representing the causal structure of the system across all time steps. We define the *Multi-Scale Causal Modularity* hypothesis, which posits that the edge set $E^\star$ can be partitioned into disjoint subsets based on the frequency domain of the causal mechanisms:

$$E^\star = E_T^\star \cup E_S^\star \cup E_R^\star \tag{2}$$

where $E_T^\star$ governs long-term dependencies (low-frequency), $E_S^\star$ governs cyclical dependencies (bounded band-pass), and $E_R^\star$ governs short-term interactions (high-frequency).

3.1.2 Assumptions

To guarantee the recoverability of $E^\star$, we rely on the following measure-theoretic and structural assumptions.

A1. Causal Invariance and Asymptotic Independence. We assume that the structural causal mechanism governing the residual process $\mathbf{R}(t)$ is time-invariant over the analysis window. To ensure the consistency of conditional independence estimators from finite samples, we further require $\mathbf{R}(t)$ to satisfy a weak dependence condition (β-mixing) such that the dependence between events decays as their temporal

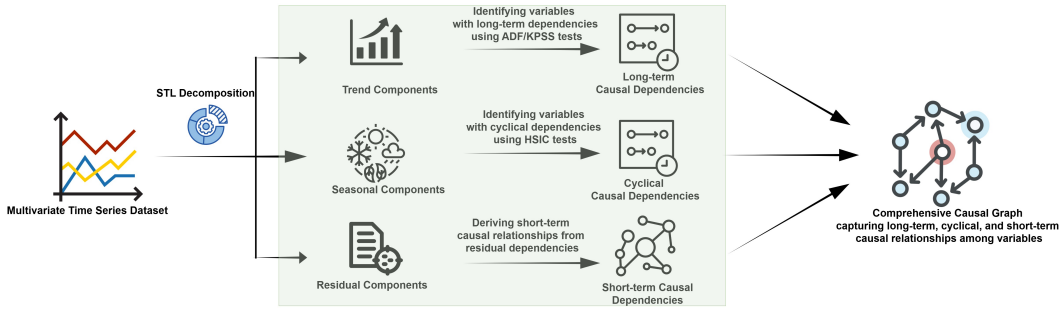

Figure 1: Overview of the proposed Decomposition-based Causal Discovery (DCD) framework. The multivariate time series dataset is decomposed into trend, seasonal, and residual components using STL decomposition. Each component undergoes a specific causal analysis—ADF/KPSS tests identify variables with long-term dependencies, HSIC tests identify variables with cyclical dependencies, and constraint-based search derives short-term causal relationships. The resulting causal graphs are then integrated into a comprehensive causal structure capturing multi-scale causal dependencies.

separation increases. Specifically, let $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{t+k}^{\infty}$ be the σ-algebras generated by $\{\mathbf{R}(s) : s \leq t\}$ and $\{\mathbf{R}(s) : s \geq t+k\}$. The mixing coefficient is defined as:

$$\beta(k) = \sup_{t} \mathbb{E}\left[\sup_{A \in \mathcal{F}_{t+k}^{\infty}} \left|P(A \mid \mathcal{F}_{-\infty}^{t}) - P(A)\right|\right]. \tag{3}$$

The condition that $\beta(k) \to 0$ as $k \to \infty$ guarantees that the effective sample size grows with n, permitting valid statistical inference of the causal structure without relying on strict stationarity assumptions (Rio et al., 2017; Merlevède et al., 2011).

A2. Spectral Separability. The components have disjoint support in the frequency domain. Let $f_Y(\omega)$ denote the spectral density of a process Y. We assume there exist cutoff frequencies $\omega_1 < \omega_2$ such that:

$$\operatorname{supp}(f_T) \subset [0, \omega_1), \tag{4}$$

$$\operatorname{supp}(f_S) \subset \bigcup_{k} \left[\frac{2\pi k}{P} - \delta, \frac{2\pi k}{P} + \delta\right], \tag{5}$$

$$\operatorname{supp}(f_R) \subset (\omega_1, \pi] \setminus \operatorname{supp}(f_S). \tag{6}$$

This assumption justifies the use of frequency-selective decomposition (like STL) to separate components (Cleveland et al., 1990).

A3. Linear Gaussian Dynamics and Bounded Leakage. To derive rigorous identifiability bounds where mutual information is upper-bounded by covariance, we assume for the theoretical analysis that the data-generating process follows a Linear Gaussian Structural Equation Model (LG-SEM). Under this assumption, we define the leakage parameter ε as the maximum cross-component correlation:

$$\lambda = \max_{i,j} \left\{ \frac{|\operatorname{Cov}(\widehat{T}_i, \widehat{R}_j)|}{\sigma_{\widehat{T}_i}\sigma_{\widehat{R}_j}}, \frac{|\operatorname{Cov}(\widehat{S}_i, \widehat{R}_j)|}{\sigma_{\widehat{S}_i}\sigma_{\widehat{R}_j}} \right\} \leq \varepsilon, \quad \text{with } \varepsilon \ll 1. \tag{7}$$

This ensures that spectral leakage corresponds directly to negligible mutual information between components.

A4. Causal Modularity. Cross-scale causal influences are negligible or effectively blocked by the decomposition. Formally, $R_i(t-\tau) \not\to T_j(t)$ and $R_i(t-\tau) \not\to S_j(t)$ for any lag τ.

Remark on Generality. While we invoke the Linear Gaussian assumption (A3) to establish the theoretical bounds in Theorem 1, the proposed DCD framework remains practically applicable to non-linear systems. The components of DCD are modular: STL decomposition effectively filters non-linear trends via LOESS smoothing, and the residual-level causal discovery employs non-parametric conditional independence tests (such as CMI-knn or HSIC) to detect non-linear dependencies. Thus, A3 serves as a sufficient condition for theoretical identifiability, while our experiments (Section 4) demonstrate robustness in complex, non-linear climate and industrial scenarios.

3.1.3 Identifiability Analysis

We now prove that under these assumptions, component-wise causal discovery yields a graph $\widehat{G}$ that converges to $G^\star$ up to an error term dominated by ε.

Lemma 1 (Projection of Seasonal Influence)**.** *Consider a causal pathway $S_X(t) \to Y(t)$ where S_X is a seasonal driver. Under Assumption A3, the decomposition of $Y(t)$ projects this influence onto $S_Y(t)$, orthogonal to the residual subspace $R_Y(t)$ up to order ε.*

$$R_Y(t) \perp\!\!\!\perp S_X(t) \mid S_Y(t) + \mathcal{O}(\varepsilon). \tag{8}$$

Proof. Let the causal mechanism be $Y(t) = f(S_X(t)) + \eta(t)$, where $\eta(t)$ represents other independent influences. Since $S_X(t)$ is a periodic function with fundamental frequency ω_0 (Assumption A2), the term $f(S_X(t))$ also resides in the seasonal frequency band (potentially including harmonics). The decomposition operator $\mathcal{D}$ partitions $Y(t)$ based on frequency. By Assumption A2 (Spectral Separability), the energy of $f(S_X(t))$ falls within the support of the seasonal component estimator $\widehat{S}_Y$. Ideally, $\widehat{S}_Y(t) \approx f(S_X(t))$ and $\widehat{R}_Y(t) \approx \eta(t)$.

However, due to finite-sample spectral leakage, a small fraction of the seasonal energy may spill into the residual band. Assumption A3 bounds this covariance by ε. Therefore, the information about $S_X(t)$ contained in $R_Y(t)$ is bounded by:

$$|\mathrm{Cov}(R_Y(t), S_X(t))| \leq |\mathrm{Cov}(R_Y(t), S_Y(t))| \cdot C \leq \varepsilon \cdot \sigma_{R_Y}\sigma_{S_Y}.$$

Consequently, the conditional mutual information $I(R_Y; S_X \mid S_Y) \to 0$ as $\varepsilon \to 0$. Thus, $R_Y(t) \perp\!\!\!\perp S_X(t)$ up to the error term $O(\varepsilon)$.

Lemma 2 (Weak Dependence of Residuals)**.** *Let the observed process be $X(t) = T(t) + S(t) + R(t)$. Under Assumptions A2 (Spectral Separability) and A3 (Bounded Leakage), the estimated residual component $\widehat{R}(t)$ satisfies the weak dependence (mixing) condition defined in Assumption A1, with deviation bounded by the leakage parameter ε.*

Proof. The raw process $X(t)$ typically violates the mixing assumption (A1) because $T(t)$ exhibits long memory (non-summable autocovariance) and $S(t)$ exhibits non-decaying periodic correlations. The residual estimator is given by $\widehat{R}(t) = X(t) - \widehat{T}(t) - \widehat{S}(t)$. By Assumption A2, the estimators $\widehat{T}$ and $\widehat{S}$ capture the spectral energy associated with the non-mixing components (frequencies $\omega \approx 0$ and harmonics). By Assumption A3, the variance of these components leaking into $\widehat{R}$ is bounded by ε. Consequently, the autocovariance of the estimated residual $\gamma_{\widehat{R}}(k)$ approximates the true residual autocovariance: $\gamma_{\widehat{R}}(k) \approx \gamma_R(k) + \mathcal{O}(\varepsilon)$. Since the true mechanism $R(t)$ is generated by short-term interactions (A1), its covariance decays exponentially. Thus, $\widehat{R}(t)$ behaves as a mixing process up to the noise floor ε, ensuring the validity of asymptotic CI tests.

Corollary 1 (Seasonal-to-Residual Removal)**.** *If the true system contains a causal dependency where a seasonal driver $S_X(t)$ influences a variable $Y(t)$, then under Assumptions A1–A4, this effect appears in the seasonal component $S_Y(t)$ rather than the residual $R_Y(t)$ after decomposition.*

Proof. By definition, $S_X(t)$ is a deterministic or cyclostationary process. Any causal effect $S_X(t) \to Y(t)$ induces a periodicity in $Y(t)$. The decomposition $Y(t) = T_Y(t) + S_Y(t) + R_Y(t)$ is designed to assign all periodic variability to $S_Y(t)$ (Assumption A2). Therefore, the residual $R_Y(t)$, defined as $Y(t) - T_Y(t) - S_Y(t)$, contains no periodic structure induced by $S_X(t)$ except for the bounded leakage term ε (Assumption A3). It follows that the causal link is structurally absorbed into the edge $S_X \to S_Y$ (or $t \to Y$ in the integrated graph), and no edge $S_X \to R_Y$ exists in the residual graph G_R.

Corollary 2 (Trend–Seasonal Leakage Bound)**.** *Let $X_i(t) = T_i(t) + S_i(t) + R_i(t)$ satisfy Assumption A3, and suppose that the true data-generating process contains a causal pathway $T_X(t) \longrightarrow S_Y(t)$. Then after decomposition, the induced effect from T_X onto Y appears predominantly in $T_Y(t)$, with at most $O(\varepsilon)$ contamination into $S_Y(t)$ and $R_Y(t)$.*

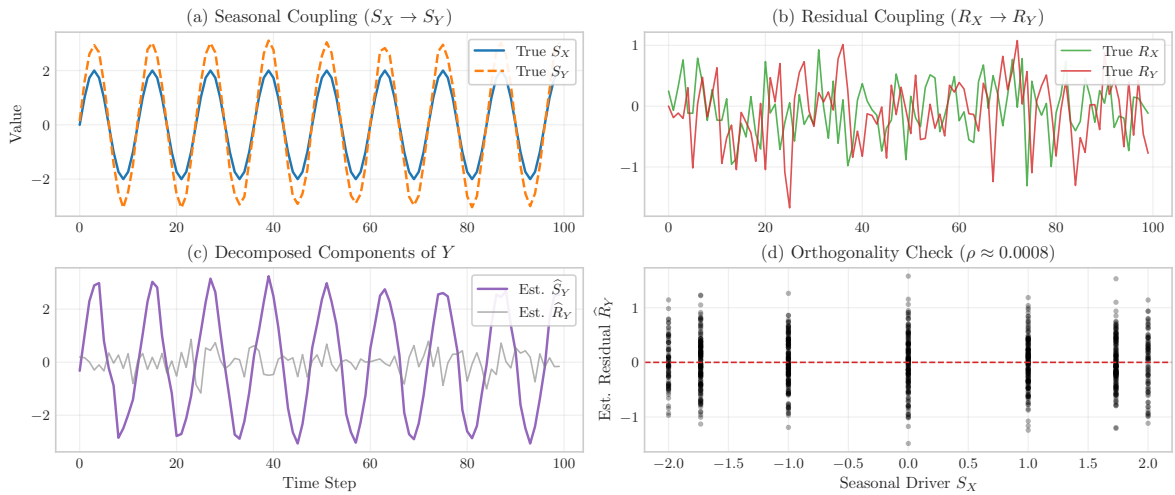

Figure 2: Empirical verification of Lemma 1. (a) The true seasonal component S_X drives S_Y. (b) The true residual R_X drives R_Y. (c) STL decomposition successfully separates the estimated seasonal $\widehat{S}_Y$ and residual $\widehat{R}_Y$ components. (d) Scatter plot confirming that the seasonal driver S_X is orthogonal to the estimated residual $\widehat{R}_Y$ ($\rho \approx 0.0008$), confirming that seasonal leakage is negligible.

Proof. Assume $Y(t) = \alpha T_X(t) + \xi(t)$. Since $T_X(t)$ is a low-frequency trend component (spectrum concentrated near zero frequency, $\omega \approx 0$), the induced term $\alpha T_X(t)$ is also low-frequency. The STL decomposition assigns low-frequency variations to the trend component $T_Y(t)$. The seasonal component $S_Y(t)$ captures distinct frequencies $\omega \in \{\frac{2\pi k}{P}\}$, and the residual $R_Y(t)$ captures high frequencies. Given the spectral separation (Assumption A2), the projection of the low-frequency signal $\alpha T_X(t)$ onto the seasonal or residual subspaces is negligible. Assumption A3 bounds this error, ensuring that $\mathrm{Cov}(T_X, S_Y) \leq O(\varepsilon)$ and $\mathrm{Cov}(T_X, R_Y) \leq O(\varepsilon)$. Thus, the causal relationship is preserved within the trend layer G_T.

Theorem 1 (Identifiability under Linear Gaussianity)**.** *Let the underlying system be a Linear Gaussian SEM satisfying Assumptions A1–A4. Let Ψ be a consistent constraint-based causal discovery algorithm (e.g., PCMCI+ (Runge, 2020)) with Type-I error α_n. The Structural Hamming Distance (SHD) between the estimated graph $\widehat{G} = \bigcup_{K \in \{T,S,R\}} \Psi(\widehat{K})$ and the true graph $G^\star$ is bounded by:*

$$\mathbb{E}[\mathrm{SHD}(\widehat{G}, G^\star)] \leq C_1 \cdot \varepsilon + C_2 \cdot \alpha_n \tag{9}$$

where C_1, C_2 are constants depending on the graph density.

Proof. The error decomposes into two sources:

1. **Leakage Error ($C_1 \cdot \varepsilon$):** Under the Linear Gaussian assumption (A3), zero covariance implies independence. By Lemma 1, any true seasonal or trend edge projected onto the residual space has a partial correlation bounded by $\mathcal{O}(\varepsilon)$. For a sufficiently high detection threshold $\tau > \varepsilon$, these spurious edges are rejected in the residual graph G_R.

2. **Estimation Error ($C_2 \cdot \alpha_n$):** For edges $e \in E_R^\star$, the residuals $\widehat{R}(t)$ satisfy the mixing condition (Assumption A1) once the non-stationary components are removed (Lemma 2). Consistent CI tests guarantee that the probability of incorrect edge recovery scales with $\alpha_n \to 0$ as $n \to \infty$ (Rio et al., 2017).

Thus, $\widehat{G}$ asymptotically recovers the union of component-specific true edges $E^\star$ as $\varepsilon \to 0$ and $n \to \infty$.

3.1.4 Empirical Verification of Seasonal Projection

To validate the projection properties utilized in Theorem 1 (specifically Lemma 1), we simulate a controlled bivariate system where a seasonal component causally drives another variable. We aim to confirm that standard decomposition (STL) correctly assigns this effect to the seasonal component (S_Y) rather than the residual (R_Y), satisfying the leakage bound assumption.

Data-Generating Process. We generate two scalar processes $X(t)$ and $Y(t)$ defined by the structural equations:

$$S_X(t) = A_X \sin\left(\tfrac{2\pi t}{P}\right) \tag{10}$$

$$T_X(t) = \beta_X t/n \tag{11}$$

$$R_X(t) = \varepsilon_X(t) \sim \mathcal{N}(0, \sigma_X^2) \tag{12}$$

$$X(t) = T_X(t) + S_X(t) + R_X(t) \tag{13}$$

$$S_Y(t) = \underbrace{g\big(S_X(t)\big)}_{\text{Seasonal Driver}} + \eta_S(t) \tag{14}$$

$$T_Y(t) = \beta_Y t/n \tag{15}$$

$$R_Y(t) = \underbrace{\gamma\, R_X(t-1)}_{\text{Residual Driver}} + \varepsilon_Y(t) \tag{16}$$

$$Y(t) = T_Y(t) + S_Y(t) + R_Y(t) \tag{17}$$

where $g(\cdot)$ is a linear transfer function ($1.5 \cdot S_X(t)$), $P = 12$, and $n = 1000$. Crucially, this system contains a direct causal path $S_X(t) \to S_Y(t)$ and a separate residual path $R_X(t-1) \to R_Y(t)$.

Validation Results. We apply STL decomposition to the raw observables $X(t)$ and $Y(t)$. As shown in Figure 2(d), the empirical cross-correlation between the recovered seasonal driver $\widehat{S}_X$ and the residual effect $\widehat{R}_Y$ is negligible ($|\rho| \approx 0.0008$). The causal influence of S_X is successfully projected onto $\widehat{S}_Y$ (Figure 2c), ensuring that the residual-level causal discovery identifies only the true short-term link $R_X \to R_Y$ without confounding from the seasonal signal.

3.2 The DCD Framework

Guided by Theorem 1, we propose the Decomposition-based Causal Discovery (DCD) framework. DCD is explicitly designed to satisfy Assumption A3 (Leakage Bound) by utilizing a robust decomposition method, thereby allowing for the accurate recovery of $G^\star$. The framework consists of three stages: (1) Time series decomposition, (2) Component-specific causal analysis, and (3) Graph integration. The overall procedure is summarized in Algorithm 1 and visualized in Figure 1.

Algorithm 1 Decomposition-based Causal Discovery (DCD)

Input: Multivariate time series $\{X_i(t)\}_{i=1}^N$ ▷ $X_i(t)$: value of variable i at time t
Output: Integrated causal graph G_{combined} ▷ G_{combined}: final multi-scale causal graph

Step 1: Decomposition of multivariate time series
  Call Algorithm 2 to obtain $\{T_i(t), S_i(t), R_i(t)\}$ ▷ T_i: trend, S_i: seasonal, R_i: residual components

Step 2: Component-level causal analysis
  Call Algorithm 3 on $\{T_i, S_i, R_i\}$
  Obtain $\{G_{T_i}, G_{S_i}, G_{R_i}\}_{i=1}^N$ ▷ G_{T_i}: trend graph, G_{S_i}: seasonal graph, G_{R_i}: residual causal graph

Step 3: Integration of causal graphs
  Call Algorithm 4 with $\{G_{T_i}, G_{S_i}, G_{R_i}\}$
  Obtain final G_{combined} ▷ Combines time-driven (trend/season) edges with residual causal edges

return G_{combined}

3.2.1 Decomposition (Satisfying Assumptions A2 & A3)

To effectively isolate the components $T_i(t)$, $S_i(t)$, and $R_i(t)$, we employ Seasonal-Trend decomposition using LOESS (STL) (Cleveland et al., 1990). We select STL over simple moving averages or rigid Fourier bases because it adapts to local non-linearities via robust LOESS smoothing, thereby minimizing spectral leakage (ε) between the seasonal and residual components. This separation is critical for satisfying the leakage bound defined in Assumption A3.

For each variable $X_i(t)$, we determine the optimal seasonal period P_i through a variance-maximization search. Let $\mathcal{P}$ be the set of candidate periods. For each $p \in \mathcal{P}$, we perform a trial decomposition and compute the variance of the extracted seasonal component $S_i^{(p)}$. The optimal period is selected as:

$$P_i^\star = \arg\max_{p \in \mathcal{P}} \left(\mathrm{Var}(S_i^{(p)}) \right), \tag{18}$$

subject to the constraint $\mathrm{Var}(S_i^{(p)}) \geq \tau$. If no period satisfies the variance threshold τ (e.g., $\tau = 0.1$), the variable is treated as aperiodic, and we set $S_i(t) = 0$. This step ensures that we do not force a seasonal decomposition on intrinsically non-seasonal data, preserving the integrity of the residual signal.

3.2.2 Component-Specific Causal Analysis

Trend Analysis (Long-term Dependency). The trend component $T_i(t)$ captures non-stationary, low-frequency evolution driven by the exogenous time index. To robustly identify these dependencies, we employ a confirmatory stationarity analysis using two complementary statistical tests (Hyndman, 2018):

1. **Augmented Dickey-Fuller (ADF):** Tests H_0: *Presence of a unit root* (non-stationary). Rejection implies stationarity (Dickey & Fuller, 1979).
2. **KPSS Test:** Tests H_0: *Trend-stationarity.* Rejection implies a non-stationary trend (Kwiatkowski et al., 1992a).

If $T_i(t)$ is determined to be non-stationary (failure to reject H_0^{ADF} or rejection of H_0^{KPSS}), we infer that the variable is driven by long-term temporal factors. This is formally represented as a directed edge $t \to X_i$ in the trend graph G_T, effectively treating time as a latent confounder for the long-term dynamics.

Seasonal Analysis (Cyclic Dependency). To detect deterministic cyclic dependencies, we utilize the Hilbert–Schmidt Independence Criterion (HSIC), a kernel-based measure of dependence (Gretton et al., 2005). We test the independence between the seasonal component $S_i(t)$ and the time index t using Radial Basis Function (RBF) kernels. The hypothesis test is defined as:

$$H_0 : P_{S_i,t} = P_{S_i} P_t \quad \text{(Independence)}.$$

A statistically significant result ($p < \alpha$, estimated via permutation testing) implies that the seasonal component is deterministically coupled to the time index (Scholkopf & Smola, 2018). We represent this as a seasonal edge $t \to X_i$ in the graph G_S.

Residual Analysis (Short-term Causality). By removing non-stationary trends and deterministic seasonality, the residual term $R_i(t)$ approximates a stationary, weakly dependent process (satisfying Lemma 2). To recover the intrinsic causal topology, we apply a robust constraint-based causal discovery framework designed for high-dimensional time series (Runge et al., 2019a; Runge, 2020).

Rather than relying on a single specific algorithm, this stage generally employs iterative conditional independence testing to eliminate spurious edges while controlling for autocorrelation. The inference procedure specifically targets two types of short-term structures:

- **Lagged Dependencies:** It identifies persistent causal effects where past values of a variable influence the current state of another ($X_{t-\tau} \to Y_t$).

- **Contemporaneous Dependencies:** It detects instantaneous interactions occurring within the same time step ($X_t \to Y_t$), which are often masked by autocorrelated noise in raw data.

This process yields the residual causal graph G_R, capturing the essential short-term interactions freed from the confounding effects of shared trends or seasonality.

3.2.3 Integration of Multi-Scale Causal Graphs

The final step integrates the component-specific graphs into a unified structure $G_{\text{combined}} = (V \cup \{t\}, E_T \cup E_S \cup E_R)$. Since the components operate in effectively disjoint frequency bands (Assumption A2) and represent distinct physical mechanisms (Assumption A4), the edge sets are additive and conflict-free. The resulting multi-relational graph explicitly models the system at three levels:

- **Long-term edges (G_T):** Represent non-stationary drift driven by exogenous factors ($t \to X$).
- **Cyclic edges (G_S):** Represent deterministic periodic forcing ($t \to X$).
- **Short-term edges (G_R):** Represent the intrinsic, mechanistic causal web between variables ($X \to Y$).

This unified representation preserves the multi-scale nature of the system while ensuring the identifiability guarantees established in Theorem 1, preventing the spurious transitive edges that commonly arise when applying standard algorithms to raw, mixed-scale data.

4 Results and Discussion

In this section, we present the evaluation of our proposed framework using synthetic datasets and real-world climate dataset. We compare the performance of our approach against baseline methods, discuss its ability to identify causal relationships, and highlight the implications of the results.

4.1 Datasets

We conduct experiments on both synthetic and real-world datasets to comprehensively evaluate our decomposition-based causal discovery framework. For synthetic data, we generate multivariate time series with known ground-truth causal relationships. For real-world evaluation, we use monthly sea ice dataset that exhibits complex temporal dependencies.

4.1.1 Synthetic Datasets

We generate synthetic multivariate time series data with known ground-truth causal relationships through a structured autoregressive process with explicit lagged causal dependencies. For each configuration (n = 500, 1000, 1500 samples and d = 4, 6, 8, 10 variables), we construct the time series with varying lag orders (l = 2, 3, 4) Ferdous et al. (2025). The error terms, representing residual components as a combination of random noise and short-term causal relationships, are generated through:

$$e_i(t) = \sum_{j=1}^{i-1} \sum_{l=0}^{3} \alpha_{ij}^{l} e_j(t-l) + \epsilon_i(t) \tag{19}$$

where $\alpha_{ij}^{l} \in [0.3, 0.8]$ are the autoregressive coefficients for lag l, and $\epsilon_i(t) \sim \mathcal{N}(0, 0.1)$ represents Gaussian noise. This structure creates a mix of instantaneous (l = 0) and lagged (l = 1, 2, 3) causal relationships. The final time series combine trend, seasonal, and residual components:

$$X_i(t) = \mu_i + \beta_i t + A_i \sin(2\pi t/\omega_i) + e_i(t) \tag{20}$$

where μ_i is the baseline ($5 \leq \mu_i \leq 12$), β_i is the trend coefficient ($0 \leq \beta_i \leq 0.2$), A_i is the seasonal amplitude ($0 \leq A_i \leq 4$), and ω_i is the seasonal period ($12 \leq \omega_i \leq 25$). Variables are constructed with different combinations of these components to create a diverse set of temporal patterns. This design creates a complex causal structure with both contemporaneous and lagged relationships that closely mirrors real-world time series data.

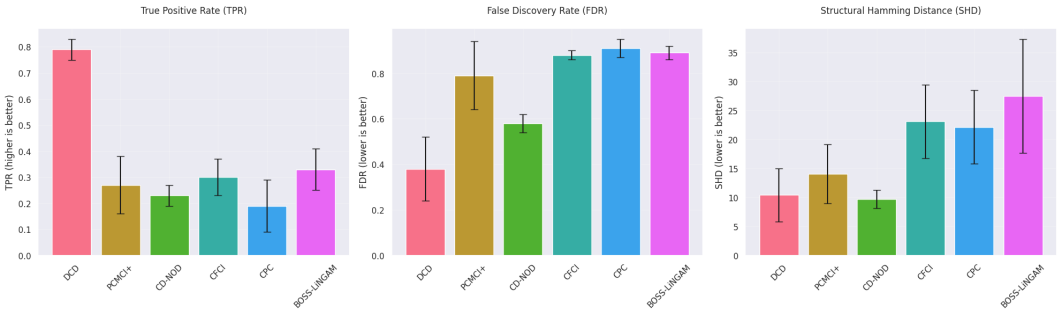

Figure 3: Performance evaluation of different causal discovery methods on synthetic datasets. The plots show True Positive Rate (TPR, higher is better), False Discovery Rate (FDR, lower is better), and Structural Hamming Distance (SHD, lower is better). Our method (DCD) demonstrates consistently superior performance across all metrics.

4.1.2 Real-world Datasets

Sea Ice Dataset: For real-world evaluation, we use the **monthly-averaged sea ice extent dataset** from the *ERA5 reanalysis product* and *NSIDC satellite observations* Cavalieri et al. (1996). This dataset spans **39 years (1979–2018)** and includes key atmospheric and oceanic variables that influence sea ice dynamics. The dataset is processed into **monthly-averaged time series** using an area-weighted method over the Arctic north of 25°N. This dataset is particularly suited for evaluating causal discovery methods due to its strong *seasonal* and *long-term dependencies*. The dataset provides monthly-averaged values of the variables summarized in Table 3.

ETTh1 Dataset: We also evaluate on the *ETTh1 electricity transformer dataset*, which records hourly measurements of transformer load and oil temperature from a real-world power station Zhou et al. (2021). The dataset captures **multi-resolution dynamics** driven by operational demand cycles and thermal diffusion effects. The ETTh1 dataset is well-suited for testing causal discovery under conditions of strong *seasonality*, *nonstationary trends*, and *nonlinear dependencies* in industrial systems.

4.2 Baselines

We compare our approach against several state-of-the-art causal discovery methods. A detailed description of the evaluation metrics used is provided in Appendix B. The baselines represent different approaches to temporal causal discovery:

PCMCI+ (Runge, 2020): A constraint-based method specifically designed for time series that handles both lagged and contemporaneous relationships. It employs momentary conditional independence tests to identify causal links while controlling for autocorrelation.

CD-NOD Huang et al. (2020): A causal discovery approach tailored for non-stationary/heterogeneous data. It combines invariance testing with constraint-based structure learning to identify causal relationships.

CFCI Peters et al. (2017): An adaptation of the Fast Causal Inference algorithm for time series data, capable of handling latent confounders and selection bias.

CPC Colombo et al. (2012): A modified version of the PC algorithm optimized for temporal data structures, using temporal priority to reduce the search space.

BOSS-LiNGAM Shimizu et al. (2006): A bootstrapping-based extension of LiNGAM designed for time series data, employing non-Gaussian models for causal discovery.

For fair comparison, we implement all baselines using their official code repositories and tune hyperparameters following authors' recommendations. Since CD-NOD does not account for lagged relationships, we evaluate its performance on a non-temporal version of the synthetic dataset, ensuring fairness.

Table 1: Performance comparison of causal discovery methods at different lag values

Method	Lag 2			Lag 3			Lag 4		
	TPR ↑	FDR ↓	SHD ↓	TPR ↑	FDR ↓	SHD ↓	TPR ↑	FDR ↓	SHD ↓
DCD (Ours)	**0.78 ± 0.03**	**0.28 ± 0.13**	**7.83 ± 3.90**	**0.79 ± 0.04**	**0.38 ± 0.15**	10.42 ± 4.89	**0.80 ± 0.05**	**0.48 ± 0.06**	13.00 ± 3.64
PCMCI+	0.21 ± 0.11	0.81 ± 0.11	10.83 ± 3.07	0.30 ± 0.11	0.72 ± 0.22	13.50 ± 6.26	0.29 ± 0.08	0.86 ± 0.04	17.75 ± 2.70
CD-NOD	0.23 ± 0.04	0.57 ± 0.05	9.50 ± 1.73	0.23 ± 0.04	0.57 ± 0.05	**9.75 ± 1.54**	0.23 ± 0.04	0.57 ± 0.05	**9.75 ± 1.54**
CFCI	0.32 ± 0.06	0.88 ± 0.02	19.25 ± 6.18	0.30 ± 0.06	0.88 ± 0.01	24.00 ± 5.51	0.29 ± 0.08	0.88 ± 0.02	26.00 ± 5.70
CPC	0.18 ± 0.11	0.91 ± 0.06	17.08 ± 4.78	0.19 ± 0.06	0.91 ± 0.03	23.08 ± 5.26	0.20 ± 0.11	0.92 ± 0.03	26.25 ± 5.56
BOSS-LiNGAM	0.35 ± 0.08	0.89 ± 0.05	21.58 ± 5.21	0.32 ± 0.09	0.89 ± 0.02	28.42 ± 5.71	0.31 ± 0.08	0.90 ± 0.02	32.42 ± 13.50

4.3 Performance on Synthetic Datasets

We evaluate DCD across a broad range of synthetic scenarios varying in sample size, lag length, and dimensionality, using standard metrics—True Positive Rate (TPR ↑), False Discovery Rate (FDR ↓), and Structural Hamming Distance (SHD ↓), with formal definitions given in Appendix B. Aggregate results (Figure 3; Table **??**) show that DCD achieves the strongest overall performance, attaining a mean TPR of 0.79 ± 0.04 and an FDR of 0.38 ± 0.14. CD-NOD obtains the lowest SHD (9.67 ± 1.57) but achieves this by excluding all lagged edges, resulting in a substantially lower TPR of 0.23 ± 0.04 and thus limited causal coverage. Across sample sizes (Figure 4; Table 5), DCD remains highly stable: TPR remains near 0.79 while FDR increases only modestly from 0.36 to 0.40 as sample size grows from 500 to 1500. In contrast, PCMCI+ exhibits a sharp rise in FDR (from 0.79 to 0.82), and CFCI and CPC deteriorate even more severely, reflecting increasing overfitting and declining robustness with larger datasets. BOSS-LINGAM maintains moderate TPR but consistently suffers from high FDR, limiting its reliability. Varying the maximum temporal lag (Figure 5; Table 1) confirms that DCD maintains strong performance across temporal horizons, with TPR between 0.78–0.80 and a manageable increase in FDR (0.28–0.48). PCMCI+ shows substantial degradation at longer lags, while CFCI and CPC incur rapidly growing SHD, indicating difficulty in identifying both contemporaneous and lagged edges. BOSS-LINGAM also decays with increasing lag length, reflecting its sensitivity to extended temporal dependencies. As dimensionality increases (Figure 5; Table 2), all methods experience reduced accuracy due to the increased search space. Nevertheless, DCD retains the best balance between TPR and FDR and continues to recover meaningful lagged edges. CD-NOD maintains a comparatively low SHD but at the cost of discarding lagged structure, resulting in substantial loss in TPR. PCMCI+, CFCI, CPC, and BOSS-LINGAM scale poorly, showing rapidly rising FDR and SHD as variable count increases. Overall, these results demonstrate that DCD consistently outperforms existing causal discovery methods across a wide range of synthetic conditions. Its stability under increasing sample size, lag order, and dimensionality highlights the strength of its decomposition-based design and its ability to recover both contemporaneous and lagged causal relations in challenging multivariate time series settings.

4.4 Performance on Real-World Climate Dataset

Using the Arctic sea ice dataset, DCD uncovers multi-scale dependencies consistent with established climate dynamics. STL decomposition isolates the dominant 12-, 36-, and 48-month periodicities (except for snowfall) (Screen & Simmonds, 2012), and ADF/KPSS tests confirm stationarity in the extracted trends. Constraint-based search on residuals ($p < 0.05$, lag ≤ 3) reveals well-documented relationships including sea ice–SST feedback (Screen & Simmonds, 2010), temperature–humidity–radiation coupling (Francis & Hunter, 2006), and lagged responses such as sea ice reacting to SST and radiative forcing one month later. These results illustrate how decomposition enables DCD to separate long-term, seasonal, and short-term interactions in a physically coherent manner (Figure 6a, Appendix).

Directly applying PCMCI+ to the raw temporal dataset (Figure 6b, Appendix) yields numerous spurious or unoriented links, such as an ambiguous temperature–humidity relation that DCD correctly orients from temperature to humidity. PCMCI+ also fails to detect the documented SST → sea ice influence that DCD recovers, highlighting the importance of removing trends and cycles prior to CD. CD-NOD (Figure 6c, Appendix) identifies several known contemporaneous connections (e.g., radiation–temperature–sea ice) but captures temporal structure only minimally, often linking SST and sea ice to a time index rather than to each other.

Table 2: Performance comparison of causal discovery methods at different numbers of variables (i)

Method	$i = 4$			$i = 6$			$i = 8$			$i = 10$		
	TPR ↑	FDR ↓	SHD ↓	TPR ↑	FDR ↓	SHD ↓	TPR ↑	FDR ↓	SHD ↓	TPR ↑	FDR ↓	SHD ↓
DCD (Ours)	**0.84 ± 0.06**	**0.20 ± 0.15**	**4.11 ± 2.32**	**0.77 ± 0.00**	**0.42 ± 0.11**	**10.78 ± 3.49**	**0.79 ± 0.00**	**0.46 ± 0.06**	12.44 ± 2.30	**0.76 ± 0.00**	**0.44 ± 0.04**	14.33 ± 1.58
PCMCI+	0.34 ± 0.16	0.69 ± 0.26	8.44 ± 4.50	0.19 ± 0.10	0.90 ± 0.05	14.11 ± 3.72	0.25 ± 0.00	0.80 ± 0.05	14.78 ± 2.64	0.30 ± 0.00	0.79 ± 0.06	18.78 ± 3.35
CD-NOD	0.23 ± 0.01	0.50 ± 0.00	8.67 ± 0.50	0.18 ± 0.00	0.60 ± 0.00	12.00 ± 0.00	0.29 ± 0.00	0.60 ± 0.00	**8.00 ± 0.00**	0.22 ± 0.00	0.60 ± 0.00	**10.00 ± 0.00**
CFCI	0.26 ± 0.06	0.89 ± 0.01	18.56 ± 5.57	0.35 ± 0.10	0.87 ± 0.02	18.67 ± 1.66	0.29 ± 0.06	0.88 ± 0.02	24.56 ± 4.53	0.31 ± 0.03	0.88 ± 0.02	30.56 ± 3.24
CPC	0.07 ± 0.06	0.95 ± 0.03	18.44 ± 4.98	0.23 ± 0.12	0.90 ± 0.04	17.11 ± 3.14	0.24 ± 0.04	0.88 ± 0.04	24.67 ± 5.41	0.20 ± 0.00	0.91 ± 0.02	28.33 ± 4.36
BOSS-LiNGAM	0.26 ± 0.06	0.91 ± 0.02	21.22 ± 4.02	0.32 ± 0.08	0.89 ± 0.02	20.00 ± 2.50	0.31 ± 0.08	0.90 ± 0.03	31.67 ± 7.76	0.41 ± 0.03	0.87 ± 0.04	37.00 ± 10.56

By decomposing trend, seasonal, and residual components, DCD exposes multi-scale interactions including long-term warming signals and annual precipitation variability, producing a more physically interpretable network. BOSS-LiNGAM, CFCI, and CPC (Figures 6d–f, Appendix) tend to generate dense, overconnected graphs indicative of overfitting and weak discrimination between physical mechanisms and noise. In contrast, DCD yields a sparse, domain-consistent causal structure by explicitly accounting for non-stationarity and autocorrelation. Collectively, these comparisons demonstrate that decomposition is essential for reliable CD in climate systems characterized by strong periodicity, long-term trends, and autocorrelated variability.

4.5 Performance on Electricity Transformer Data (ETTh1)

DCD reveals characteristic multi-scale dependencies in the ETTh1 dataset. STL decomposition confirms strong periodicity (HSIC $p < 0.001$) and stationarity tests validate the extracted components. Residual-level search (lag ≤ 2, $\alpha = 0.05$) yields sparse, meaningful edges, including autoregressive effects (e.g., $\text{LUFL}_{-1} \to \text{LUFL}$), cross-load interactions ($\text{HUFL}_{-1} \to \text{MUFL}$), and meteorological persistence ($\text{OT}_{-1} \to \text{OT}$). Consistent with domain expectations, no causal edges from load to OT were identified, indicating independent short-term evolution of outdoor temperature.

Baselines behave differently. PCMCI+ recovers larger edge sets but produces numerous conflicting directions (e.g., LUFL $\leftrightarrow$ HUFL), reflecting orientation difficulties under nonstationarity. CFCI and CPC yield dense, noisy graphs, while BOSS-LiNGAM ignores lagged structure. In contrast, DCD's decomposition enables recovery of compact, physically coherent graphs, validating explicit multi-scale modeling for electrical load systems.

5 Limitations

Despite robust performance, DCD has theoretical and structural limitations. First, identifiability relies on spectral separability and Linear Gaussian dynamics; real-world cross-frequency coupling may induce leakage. Second, DCD conservatively models trends as exogenous ($t \to X_i$) to avoid spurious cointegration, potentially missing direct long-term mechanisms ($T_i \to T_j$). Finally, unlike time-varying methods such as CDANs (Ferdous et al., 2023), DCD assumes a stable residual structure, addressing non-stationarity via decomposition rather than modeling evolving edge weights.

6 Conclusion and Future Work

We introduced the Decomposition-based Causal Discovery (DCD) framework to address non-stationary confounding. By formalizing decomposition as spectral filtering, we established identifiability guarantees under separability and linear dynamics. DCD significantly outperforms baselines like PCMCI+ and CD-NOD, particularly under trend-driven confounding, while simulations confirm that standard decomposition satisfies leakage bounds for accurate inference.

Future work will focus on relaxing the theoretical assumptions identified in our limitations analysis, particularly by extending the framework to handle cross-frequency coupling where trends modulate seasonal amplitudes. We also aim to integrate DCD with time-varying causal discovery methods to simultaneously model structural shifts and trend-driven non-stationarity, while exploring deep learning-based decomposition to further reduce leakage in highly non-linear systems.

References

D. J. Cavalieri, C. L. Parkinson, P. Gloersen, and H. J. Zwally. Sea ice concentrations from nimbus-7 smmr and dmsp ssm/i-ssmis passive microwave data, version 1. *National Snow and Ice Data Center*, 1996. doi: 10.5067/8GQ8LZQVL0VL.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. Seasonal decomposition of time series by loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp. 294–321, 2012.

Amin Dhaou, Antoine Bertoncello, Sébastien Gourvénec, Josselin Garnier, and Erwan Le Pennec. Causal and interpretable rules for time series analysis. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2764–2772, 2021.

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544, 2013.

Muhammad Hasan Ferdous, Uzma Hasan, and Md Osman Gani. Cdans: Temporal causal discovery from autocorrelated and non-stationary time series data. In *Machine Learning for Healthcare Conference*, pp. 186–207. PMLR, 2023.

Muhammad Hasan Ferdous, Emam Hossain, and Md Osman Gani. TimeGraph: Synthetic Benchmark Datasets for Robust Time-Series Causal Discovery, June 2025. URL `https://github.com/hferdous/TimeGraph`.

Jennifer A Francis and Elizabeth Hunter. New insight into the disappearing arctic sea ice. *Eos*, 87(46): 509–511, 2006.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pp. 63–77. Springer, 2005.

Biwei Huang, Kun Zhang, Jiji Zhang, Joseph D Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous nonstationary data. *J. Mach. Learn. Res.*, 21 (89):1–53, 2020.

Norden E Huang, Zheng Shen, Steven R Long, Manli Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

RJ Hyndman. *Forecasting: principles and practice*. OTexts, 2018.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178, 1992a.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3):159–178, 1992b.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

Zhiding Liu, Jiqian Yang, Mingyue Cheng, Yucong Luo, and Zhi Li. Generative pretrained hierarchical transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2003–2013, 2024.

Florence Merlevède, Magda Peligrad, and Emmanuel Rio. A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3):435–474, 2011.

Boris Oreshkin, Dmitrii Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics*, pp. 1595–1605. PMLR, 2020.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, Cambridge, MA, USA, 2017.

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2400–2410, 2024.

Joseph Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International journal of data science and analytics*, 3(2):121–129, 2017.

Emmanuel Rio et al. *Asymptotic theory of weakly dependent random processes*, volume 80. Springer, 2017.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1388–1397. Pmlr, 2020.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):1–13, 2019a.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, 5(11):eaau4996, 2019b.

Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2018.

James A Screen and Ian Simmonds. The central role of diminishing sea ice in recent arctic temperature amplification. *Nature*, 464(7293):1334–1337, 2010.

James A Screen and Ian Simmonds. Declining summer snowfall in the arctic: Causes, impacts and feedbacks. *Climate Dynamics*, 38(11):2243–2256, 2012.

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Eugen Slutzky. The summation of random causes as the source of cyclic processes. *Econometrica: Journal of the Econometric Society*, pp. 105–146, 1937.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, USA, 2nd edition, 2001.

Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.

A. Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

Qingsong Wen, Linxiao Yang, Tian Zhou, and Liang Sun. Robust time series analysis and applications: An industrial perspective. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4836–4837, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34: 22419–22430, 2021.

Ziyi Zhang, Shaogang Ren, Xiaoning Qian, and Nick Duffield. Learning flexible time-windowed granger causality integrating heterogeneous interventional time series data. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4408–4418, 2024.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

A Algorithmic Details

This section provides the full pseudocode for all internal components of the Decomposition-based Causal Discovery (DCD) framework. While the main manuscript presents only the high-level overview in Algorithm 1, the complete procedures underlying the decomposition, component-level causal analysis, and multi-scale graph integration are provided here for clarity, reproducibility, and implementation transparency.

Algorithm 2 Decomposition of Time Series Components

Input: Multivariate time series data $\{X_i(t)\}_{i=1}^{N}$
Output: Trend $T_i(t)$, seasonal $S_i(t)$, and residual $R_i(t)$ components for each $X_i(t)$

for each variable $X_i(t)$, $i = 1, \ldots, N$ **do**
    Estimate dominant period P_i via Fourier analysis or variance maximization
    Apply STL decomposition with period P_i:
        $X_i(t) = T_i(t) + S_i(t) + R_i(t)$
    if variance of $S_i(t) <$ threshold **then**
        Mark $S_i(t)$ as negligible and exclude from further analysis
    end if
end for

return $\{T_i(t), S_i(t), R_i(t)\}$ for each i

Algorithm 2 details the STL-based decomposition step in which each observed time series $X_i(t)$ is separated into its trend, seasonal, and residual components. This decomposition is central to DCD, as it isolates low-frequency long-term behavior, cyclical or periodic structure, and short-term fluctuations, allowing causal relationships to be inferred at the temporal scale where they are most meaningfully expressed.

Algorithm 3 Component-Level Causal Analysis

Input: $\{T_i(t), S_i(t), R_i(t)\}_{i=1}^N$
Output: Causal graphs $\{G_{T_i}, G_{S_i}, G_{R_i}\}_{i=1}^N$

Trend Component Analysis
for each $T_i(t)$ **do**
    Perform stationarity tests (ADF/KPSS)
    if $T_i(t)$ is non-stationary **then**
        Include $T_i(t)$ in trend causal graph G_{T_i}
    else
        Exclude $T_i(t)$ from time-dependent causal relationships
    end if
end for

Seasonal Component Analysis
for each $S_i(t)$ with variance above threshold **do**
    Compute HSIC statistic with RBF kernel
    Perform permutation test for p-value
    if p-value $< \alpha$ **then**
        Include $S_i(t)$ in seasonal causal graph G_{S_i}
    else
        Exclude $S_i(t)$ from time-dependent causal relationships
    end if
end for

Residual Component Analysis
for each $R_i(t)$ **do**
    Perform causal discovery using conditional independence tests
    Build G_{R_i} from identified lagged and contemporaneous dependencies
end for

return $\{G_{T_i}, G_{S_i}, G_{R_i}\}_{i=1}^N$

Algorithm 3 formalizes the component-level causal analysis stage. In this step, the trend components are evaluated using stationarity tests (ADF/KPSS) to determine whether they exhibit statistically significant temporal dependence; the seasonal components are assessed using kernel-based dependence measures such as HSIC to identify periodic or frequency-driven interactions; and the residual components undergo constraint-based causal discovery to extract lagged and contemporaneous causal dependencies. This separation enables DCD to avoid mixing heterogeneous temporal signals that typically lead to spurious or misleading causal edges in multivariate time series.

Algorithm 4 Integration of Component-Level Causal Graphs

Input: G_R (residual-level causal graph), $\mathcal{T}$ (variables with trend dependence), $\mathcal{S}$ (variables with seasonal dependence)
Output: Final multiscale causal graph G_{final}
Initialize $G_{\text{final}} \leftarrow G_R$
for each variable X_i **do**
    if $X_i \in \mathcal{T}$ **then**
        Add edge $T \to X_i$ to G_{final}
    end if
    if $X_i \in \mathcal{S}$ **then**
        Add edge $S \to X_i$ to G_{final}
    end if
end for
return G_{final}

Finally, Algorithm 4 describes how information from the trend, seasonal, and residual components is combined into a unified causal structure. Trend and seasonal dependence tests contribute edges from external temporal drivers (e.g., time or seasonal index) to the corresponding variables, while the residual-level graph provides the causal relationships among variables themselves. The resulting graph G_{combined} captures long-term trends, cyclical effects, and short-term interactions in a coherent multi-scale framework without requiring conflict resolution across components.

Together, these algorithms provide a complete and executable specification of DCD, enabling faithful implementation and facilitating extension to other domains or decomposition strategies.

B Evaluation Metrics

To assess the performance of our decomposition-based causal discovery framework, we employ multiple evaluation metrics commonly used in causal discovery and time-series analysis. These metrics quantify the accuracy of inferred causal graphs by comparing them against the ground truth.

B.0.1 True Positive Rate (TPR)

The **True Positive Rate (TPR)**, also known as *recall*, measures the fraction of correctly identified causal relationships relative to the total number of ground-truth causal edges. It is defined as:

$$\text{TPR} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{21}$$

A higher TPR indicates that the method successfully recovers more of the actual causal relationships present in the data Runge et al. (2019b).

B.0.2 False Discovery Rate (FDR)

The **False Discovery Rate (FDR)** quantifies the proportion of falsely identified causal edges among all inferred causal relationships. It is given by:

$$\text{FDR} = \frac{\text{False Positives (FP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{22}$$

A lower FDR indicates fewer spurious causal links, improving the reliability of the inferred causal graph Colombo et al. (2012).

Table 3: Summary of Experimental, Sea Ice, and ETTh1 Datasets

Summary of Datasets

(a) Synthetic: Sample Size

Size	Vars	Lag	Comp.
500	4–10	$l=2$	T,S,R
1000	4–10	$l=3$	T,S,R
1500	4–10	$l=4$	T,S,R

(b) Synthetic: Variable Size

Size	Vars	Lags	Comp.
500–1500	4	2,3,4	T,S,R
500–1500	6	2,3,4	T,S,R
500–1500	8	2,3,4	T,S,R
500–1500	10	2,3,4	T,S,R

(c) Real-world Datasets

Dataset	Size	Vars	Steps
Sea Ice	468	11	Monthly
ETTh1	1500	7	Hourly

Sea Ice dataset

Variable	Unit
Surface pressure	hPa
Wind velocity	m/s
Specific humidity	kg/kg
Air temperature	K
Shortwave rad.	W/m^2
Longwave rad.	W/m^2
Rain rate	mm/day
Snowfall rate	mm/day
Sea surf. temp.	K
Sea surf. salinity	PSU
Sea ice conc.	%

ETTh1 dataset

Variable	Description
HUFL	High-usage full load
MUFL	Medium-usage full load
LUFL	Low-usage full load
HULL	High-usage low load
MULL	Medium-usage low load
LULL	Low-usage low load
OT	Oil temp. (target)
Date	Timestamp (hourly)

Note. T = Trend, S = Seasonal, R = Residual components.

B.0.3 Structural Hamming Distance (SHD)

The **Structural Hamming Distance (SHD)** measures the number of edge modifications (additions, deletions, or reversals) required to transform the inferred causal graph into the ground truth graph. It is expressed as:

$$\text{SHD} = \#(\text{added edges}) + \#(\text{deleted edges}) + \#(\text{reversed edges}) \tag{23}$$

A lower SHD indicates a closer match to the ground truth, making it a valuable metric for assessing causal discovery accuracy Tsamardinos et al. (2006). These evaluation metrics ensure a comprehensive assessment of our framework's ability to uncover causal relationships accurately while minimizing false discoveries.

C Synthetic Data Generation

To rigorously evaluate our decomposition-based causal discovery (DCD) framework, we generate synthetic multivariate time series data with known ground-truth causal structures. The synthetic datasets are designed to reflect real-world challenges such as autocorrelation, non-stationarity, and complex dependencies across variables.

C.0.1 Dataset Configuration

Each synthetic dataset consists of n time points and d variables, with the following configurations:

- Sample sizes: $n = 500, 1000, 1500$
- Number of variables: $d = 4, 6, 8, 10$
- Lag order: $l = 2, 3, 4$

Each variable in the multivariate time series is generated as a combination of three primary components: **trend**, **seasonality**, and **residual noise**. These components are constructed such that the resulting time series exhibit both long-term dependencies and short-term causal interactions.

C.0.2 Residual Component Generation

The residual component captures short-term dependencies between variables, including both instantaneous and lagged causal relationships. We define the residual error term $e_i(t)$ for each variable $X_i(t)$ as an autore-

Table 4: Hyperparameters used for each method. For synthetic datasets, the lag was set to the true generative lag. For ETTh1 and the climate dataset, lag was fixed at 2.

Method	Key Parameters
DCD (Ours)	STL decomposition; CI test: partial correlation; significance threshold $p < 0.05$.
PCMCI+	CI test: CMI-knn; $\alpha = 0.05$; majority collider rule; no FDR correction.
CD-NOD	CI test: partial correlation; adaptive thresholding; nonstationary surrogate tests.
CFCI	Conservative FCI variant; retains ambiguous colliders; respects temporal tiering.
CPC	Conservative PC; collider orientation only when unambiguous; outputs extended CPDAG (ambiguous triples underlined); CI test: partial correlation; $\alpha = 0.05$.
BOSS-LiNGAM	Two-step: CPDAG estimation via PC, followed by ICA-based orientation; uses Anderson–Darling test ($\alpha = 0.05$); assumes linear acyclic models.

gressive process:

$$e_i(t) = \sum_{j=1}^{i-1} \sum_{l=0}^{3} \alpha_{ij}^{l} e_j(t-l) + \epsilon_i(t), \tag{24}$$

where $\alpha_{ij}^{l} \in [0.3, 0.8]$ are autoregressive coefficients governing the strength of causal influence from variable X_j to X_i at lag l. $\epsilon_i(t) \sim \mathcal{N}(0, 0.1)$ is an independent Gaussian noise term. The summation ensures that $e_i(t)$ incorporates both instantaneous ($l = 0$) and lagged ($l = 1, 2, 3$) causal dependencies.

C.0.3 Construction of Time Series Components

Each variable $X_i(t)$ in the synthetic dataset is generated as a combination of its trend, seasonal, and residual components:

$$X_i(t) = \mu_i + \beta_i t + A_i \sin(2\pi t/\omega_i) + e_i(t), \tag{25}$$

where:

- **Trend Component:** $\mu_i + \beta_i t$ ($\mu_i \in [5, 12]$, $\beta_i \in [0, 0.2]$)
- **Seasonal Component:** $A_i \sin(2\pi t/\omega_i)$ ($A_i \in [0, 4]$, $\omega_i \in [12, 25]$)
- **Residual Component:** $e_i(t)$

This structured generation process creates a controlled environment where the true causal relationships are known, allowing precise evaluation of CD methods.

C.1 Detailed Analysis of Sample Size Sensitivity

Table 5 and Figure 4 presents a granular comparison of causal discovery performance across three distinct data availability regimes: limited ($n = 500$), moderate ($n = 1000$), and ample ($n = 1500$) sample sizes.

The results highlight the robustness of the proposed DCD framework. Even in the data-scarce setting ($n = 500$), DCD maintains a high True Positive Rate (TPR) of 0.79 ± 0.03, significantly outperforming the closest baseline, BOSS-LiNGAM, which achieves only 0.34. Furthermore, DCD consistently controls the False Discovery Rate (FDR) between 0.36 and 0.40. In contrast, constraint-based baselines such as PCMCI+, CFCI, and CPC exhibit high error rates, with FDRs consistently exceeding 0.75, suggesting a tendency to misidentify spurious correlations as causal links in the presence of trends.

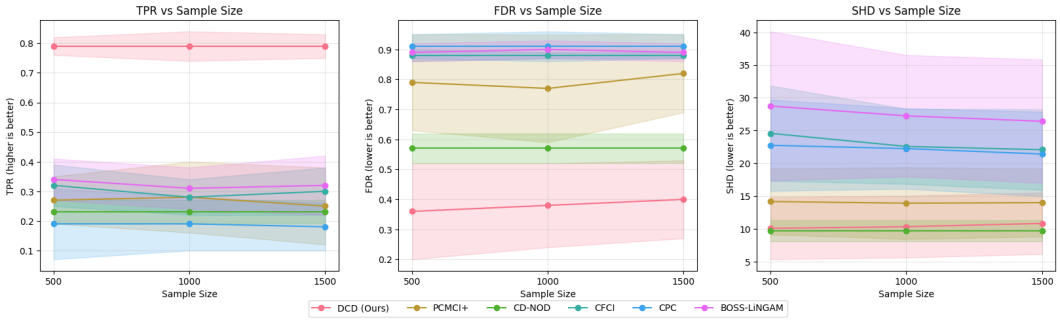

Figure 4: Impact of Sample Size on Causal Discovery Performance. The results indicate that our proposed method (DCD) achieves consistently high TPR, lower FDR, and lower SHD compared to baseline methods across different sample sizes. This demonstrates the robustness of our approach in varying data availability conditions.

Table 5: Performance comparison of causal discovery methods at different sample sizes

Method	**TPR** $\uparrow$	**FDR** $\downarrow$	**SHD** $\downarrow$
	$n = 500$		
DCD (Ours)	**0.79** $\pm$ **0.03**	**0.36** $\pm$ **0.16**	10.08 $\pm$ 4.74
PCMCI+	0.27 $\pm$ 0.08	0.79 $\pm$ 0.16	14.17 $\pm$ 5.04
CD-NOD	0.23 $\pm$ 0.04	0.57 $\pm$ 0.05	**9.67** $\pm$ **1.61**
CFCI	0.32 $\pm$ 0.07	0.88 $\pm$ 0.02	24.58 $\pm$ 7.27
CPC	0.19 $\pm$ 0.12	0.91 $\pm$ 0.04	22.75 $\pm$ 6.96
BOSS-LiNGAM	0.34 $\pm$ 0.07	0.89 $\pm$ 0.03	28.75 $\pm$ 11.39
	$n = 1000$		
DCD (Ours)	**0.79** $\pm$ **0.05**	**0.38** $\pm$ **0.14**	10.33 $\pm$ 4.72
PCMCI+	0.28 $\pm$ 0.12	0.77 $\pm$ 0.18	13.92 $\pm$ 5.47
CD-NOD	0.23 $\pm$ 0.04	0.57 $\pm$ 0.05	**9.67** $\pm$ **1.61**
CFCI	0.28 $\pm$ 0.06	0.88 $\pm$ 0.01	22.58 $\pm$ 5.73
CPC	0.19 $\pm$ 0.09	0.91 $\pm$ 0.05	22.25 $\pm$ 6.15
BOSS-LiNGAM	0.31 $\pm$ 0.07	0.90 $\pm$ 0.03	27.25 $\pm$ 9.27
	$n = 1500$		
DCD (Ours)	**0.79** $\pm$ **0.04**	**0.40** $\pm$ **0.13**	10.83 $\pm$ 4.69
PCMCI+	0.25 $\pm$ 0.13	0.82 $\pm$ 0.13	14.00 $\pm$ 5.20
CD-NOD	0.23 $\pm$ 0.04	0.57 $\pm$ 0.05	**9.67** $\pm$ **1.61**
CFCI	0.30 $\pm$ 0.08	0.88 $\pm$ 0.01	22.08 $\pm$ 6.16
CPC	0.18 $\pm$ 0.08	0.91 $\pm$ 0.04	21.42 $\pm$ 6.44
BOSS-LiNGAM	0.32 $\pm$ 0.10	0.89 $\pm$ 0.03	26.42 $\pm$ 9.42

It is worth noting that CD-NOD achieves the lowest Structural Hamming Distance (SHD) (≈ 9.67) across all settings. However, this metric must be interpreted alongside its low TPR (≈ 0.23). This discrepancy indicates that CD-NOD produces a highly sparse graph, failing to recover the majority of true lagged causal edges, which artificially lowers the distance metric without capturing the true underlying structure. DCD achieves the most balanced performance, combining high recall with moderate precision, regardless of sample size.

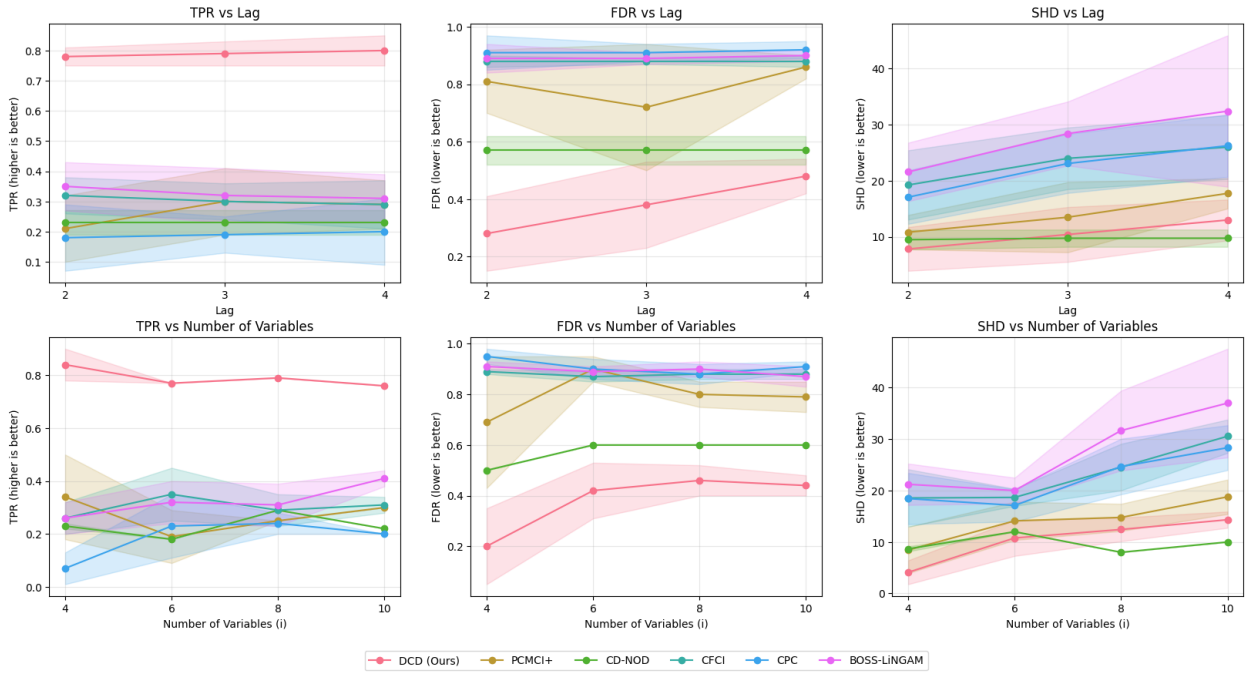

Figure 5: Impact of dataset characteristics on causal discovery performance. The first row evaluates the effect of temporal lag, and the second row assesses the impact of the number of variables. Our approach remains robust across different experimental conditions.

C.2 Detailed Analysis of Lag and Dimensionality Effects

In this section, we provide a granular analysis of how varying dataset characteristics such as temporal lag length and system dimensionality affect the performance of causal discovery algorithms. The visual trends are summarized in Figure 5, while Tables 1 and 2 present the quantitative results.

Impact of Temporal Lag

Table 1 reports the performance metrics across ground-truth lag orders $l \in \{2, 3, 4\}$.

- **Robustness of DCD:** Our proposed framework demonstrates high stability across increasing lags, maintaining a True Positive Rate (TPR) of approximately 0.78–0.80. While the False Discovery Rate (FDR) rises from 0.28 to 0.48 as dependencies become more distant, DCD consistently outperforms baselines, which generally suffer from FDRs exceeding 0.70.
- **Baseline Limitations:** Constraint-based methods like PCMCI+ and CFCI show significant performance degradation or consistently high error rates (FDR > 0.8) as lags increase. CD-NOD maintains a low Structural Hamming Distance (SHD) of ≈ 9.75, but this is driven by a consistently low TPR (≈ 0.23); because CD-NOD does not explicitly model lagged temporal dependencies, it fails to recover the majority of the causal structure, defaulting to a sparse graph.

Impact of Dimensionality

Table 2 evaluates scalability by varying the number of variables d from 4 to 10.

- **Scalability:** As the number of variables increases, the combinatorial search space for potential edges grows, typically leading to higher false positives. DCD effectively mitigates this through decomposition, starting with a high TPR of 0.84 at $d = 4$ and maintaining a competitive 0.76 at $d = 10$.

- **Comparison:** Baselines struggle to balance precision and recall in higher dimensions. For instance, CPC exhibits extremely low sensitivity in sparse settings (TPR of 0.07 at $d = 4$), while CFCI and BOSS-LiNGAM maintain high false discovery rates (> 0.85) across all dimensions. DCD offers the most balanced profile, successfully isolating true causal mechanisms even as the system complexity grows.

C.3 Qualitative Analysis of Inferred Causal Graphs

Figure 6 provides a visual comparison of the causal structures inferred by DCD and five baseline methods on the real-world Arctic Sea Ice dataset. This qualitative assessment highlights how different approaches handle the confounding effects of non-stationarity and autocorrelation.

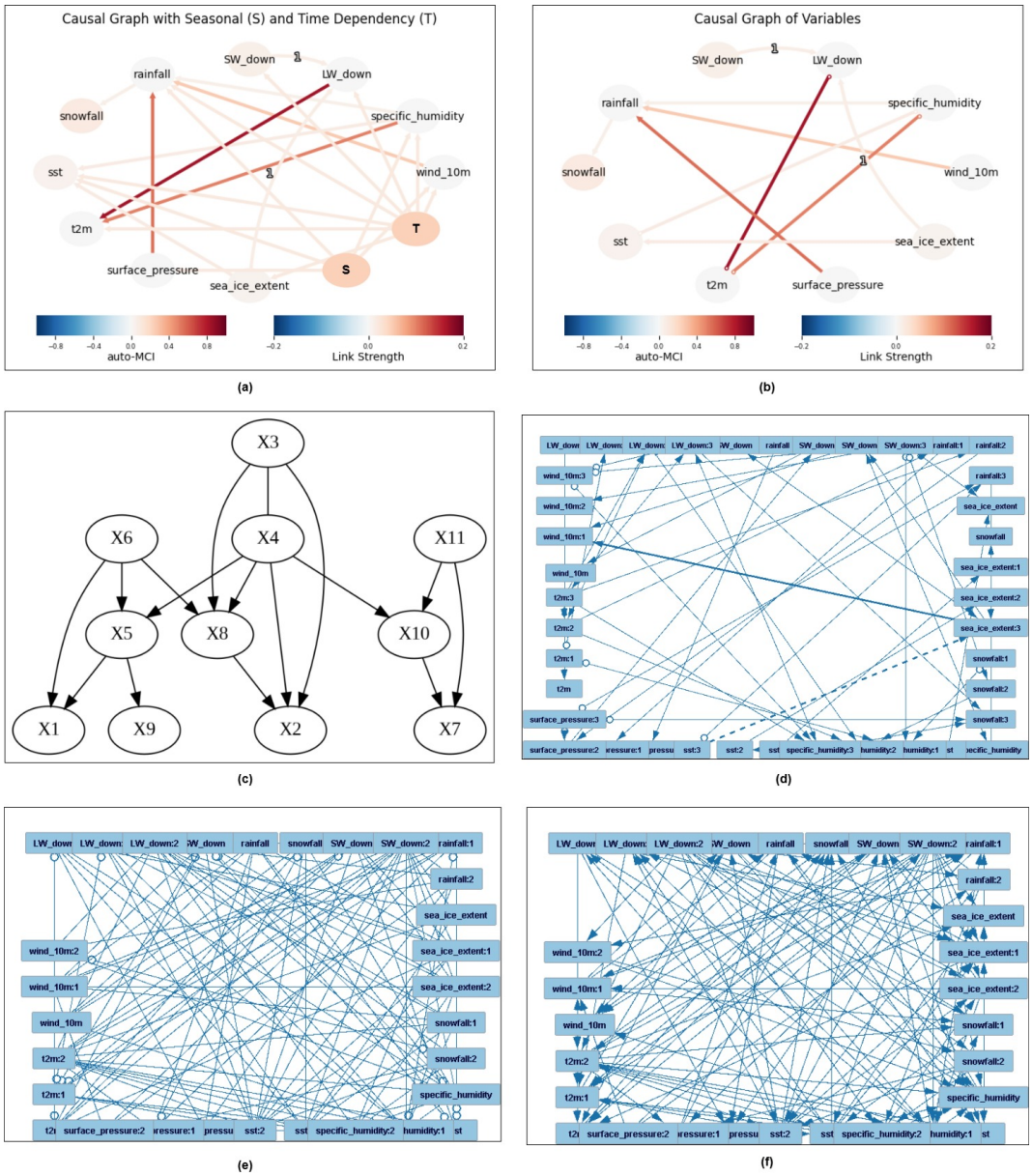

Figure 6: Comparison of causal discovery methods. (a) DCD (Ours), (b) PCMCI+, (c) CD-NOD, (d) BOSS-LiNGAM, (e) CFCI, and (f) CPC. The causal graphs illustrate the relationships identified by each method, where DCD effectively captures both seasonal (S) and time-dependent (T) influences.

DCD (Ours)

The graph produced by DCD (Figure 6a) is notably sparse and physically interpretable. By explicitly modeling trend (T) and seasonal (S) components as distinct drivers, DCD successfully isolates short-term mechanistic links in the residual layer. For instance, it recovers the well-documented feedback loop between Sea Surface Temperature (SST) and Sea Ice Concentration, oriented correctly in time. The separation prevents the "hairball" effect seen in other methods, where shared trends induce spurious all-to-all connectivity.

Baselines

- **PCMCI+ (Figure 6b):** While designed for time series, PCMCI+ applied to raw non-stationary data struggles to distinguish between causal influence and shared temporal drift. The resulting graph contains ambiguous or unoriented edges (e.g., between temperature and humidity) and fails to detect the critical lagged influence of SST on Sea Ice, likely because the strong seasonal signal masks the subtler residual interaction.
- **CD-NOD (Figure 6c):** This method correctly identifies non-stationarity but tends to link variables directly to the time index rather than to each other. Furthermore, because it does not explicitly model lagged dependencies, the graph captures contemporaneous correlations well (e.g., Radiation $\leftrightarrow$ Temperature) but misses the dynamic, time-delayed structure essential for climate process understanding.
- **Constraint-based Baselines (CFCI, CPC, BOSS-LiNGAM):** As shown in Figures 6d–f, these methods produce highly dense and overconnected graphs. This "dense" structure is a hallmark of false positives driven by autocorrelation; when a method cannot filter out the strong serial correlation inherent in climate data, it often misinterprets the memory of a variable as a causal link from another variable with similar memory. The resulting graphs offer limited interpretability for domain experts.