CAUSALSR: SCALE-RESOLVED CAUSAL DISCOVERY FROM TIME SERIES WITH HETEROGENEOUS RESPONSES

Anonymous authors
Paper under double-blind review

ABSTRACT

Causal discovery from time series has become increasingly important across a wide range of application domains. In practice, multivariate time series often exhibit heterogeneous responses across short, medium, and long time scales, causing causal effects at different temporal scales to overlap in the observed signals and making scale-specific causal relations difficult to separate, orient, and recover reliably. To address this issue, we propose CausalSR, a scale-resolved causal discovery framework that reorganizes multivariate time series in the time-frequency domain and learns causal structures at their corresponding temporal scales, instead of directly operating on the original time-domain observations. Specifically, the framework separates scale-specific temporal components, estimates interaction strength and directional evidence within each scale, suppresses indirect relations through regularized conditional-association screening, and jointly refines the resulting graphs using reconstruction consistency and acyclicity constraints, thereby enabling the recovery of more reliable scale-resolved structural information from complex time series for downstream temporal analysis. Theoretically, we prove that coherence-weighted phase aggregation preserves directional temporal ordering and enables scale-specific directed edge inference, thereby providing theoretical support for scale-resolved causal discovery. Experiments on synthetic benchmarks and real-world temporal datasets demonstrate the effectiveness of our method.

1 INTRODUCTION

Causal discovery from time series is an important tool for understanding the underlying data-generating relations of complex dynamic systems (Pearl, 2009; Glymour et al., 2019; Takens, 1981). Beyond identifying statistical associations, causal discovery aims to reveal how variables directly influence each other over time, thereby providing interpretable structural knowledge for downstream temporal analysis (Spirtes et al., 2000). Such knowledge is useful for a wide range of data engineering tasks, including forecasting, anomaly diagnosis, intervention analysis, and knowledge-driven decision making (Runge et al., 2019a; Peters et al., 2017). A representative example arises in river and environmental monitoring systems (Blöschl & Sivapalan, 1995). Such systems are often described by multiple temporal signals that reflect meteorological forcing, upstream conditions, local hydrological states, downstream responses, and seasonal background variations (Beven, 2012). These signals may exhibit heterogeneous responses. For instance, local hydrological variables may react to external forcing over a short temporal range, upstream conditions may affect downstream observations after a delay, and seasonal or catchment-level factors may contribute to long-term evolving dependencies. For downstream tasks such as forecasting, early warning, and resource management, it is not sufficient to only predict future values of target variables. It is also important to understand which factors are more directly related to the observed changes, whether the influence appears as a short-term response, a delayed effect, or a long-term dependency, and how such structural relations can support interpretable analysis and decision making (Stein et al., 2025).

Most existing temporal causal discovery methods still analyze observed signals at a single time resolution or rely on a fixed lag structure (Runge et al., 2019b; Eichler, 2012). Constraint-based methods, Granger-style methods, optimization-based methods, and neural temporal models differ

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

in their assumptions and estimation procedures, but many of them directly operate on the original time-domain observations (Granger, 1969; Pamfil et al., 2020; Tank et al., 2022).

This design is effective when the dominant causal interactions are expressed within a relatively uniform temporal pattern. However, real-world multivariate time series often contain heterogeneous responses, where different interactions may appear as short-term responses, delayed influences, or long-term evolving dependencies. When all such effects are forced into a single temporal representation, short-term and long-term dependencies may interfere with each other during structure learning. As a result, the recovered causal structures may become less accurate, less stable, and less useful as structural knowledge for downstream temporal analysis.

A straightforward strategy is to enlarge the lag window, but this does not fully solve the problem. A longer lag window may capture delayed or persistent effects, but it also increases the dimensionality of the search space and may introduce redundant or spurious dependencies. A shorter lag window, on the other hand, may miss relations that are expressed over longer temporal ranges. More importantly, a fixed lag representation does not explicitly separate causal effects according to their response patterns. Thus, improving time-series causal discovery requires not only better graph learning algorithms, but also a temporal data representation that is better aligned with the heterogeneous responses of real-world dynamic systems.

To address this issue, we propose CausalSR, a scale-resolved causal discovery framework for time series with heterogeneous responses. Instead of directly learning causal structures from the original time-domain observations, the proposed framework reorganizes temporal data into a time-frequency representation, where temporal locality and scale information can be examined together (Mallat, 1999; Torrence & Compo, 1998). The input series are then separated into scale-specific components, allowing interactions with different response patterns to be analyzed within their corresponding temporal ranges (Grinsted et al., 2004). Within each scale, candidate causal relations are estimated by combining scale-specific dependence strength with directional temporal information. The resulting scale-specific structures are further refined through reconstruction consistency and acyclicity constraints, so that the learned causal relations remain connected to the original observations while better reflecting the multi-scale dynamics of the data.

2 Related Work

2.1 Temporal Data Representation

Temporal data representation plays an important role in time-series mining, forecasting, anomaly detection, and other data engineering tasks. For multivariate time series, a common strategy is to analyze observations directly in the original time domain, either as raw sequences or as lagged feature vectors (Lütkepohl, 2005). Such representations are simple and widely used, but they implicitly assume that the relevant temporal dependencies can be captured within a fixed observation window or a fixed lag structure. When the underlying system contains heterogeneous responses, this assumption may be restrictive because short-term responses, delayed influences, and long-term dependencies can be mixed in the same observed signals.

Frequency-domain and time-frequency representations have been widely studied to describe temporal patterns beyond the original time domain (Daubechies, 1992; Percival & Walden, 2000). Fourier-based representations characterize global frequency patterns, while wavelet-based representations preserve both temporal locality and scale information (Addison, 2017).

2.2 Time-Series Causal Discovery

Time-series causal discovery aims to recover directed relations among temporal variables from observational data. Classical approaches are closely related to Granger causality, where a variable is considered causal if its past values improve the prediction of another variable (Granger, 1969). Vector autoregressive models and pairwise Granger-style tests are representative examples of this line of work (Sims, 1980; Lütkepohl, 2005). Other methods use conditional independence tests to identify causal parents under temporal assumptions, such as PCMCI and related constraint-based approaches (Runge et al., 2019b; Runge, 2018; Kalisch & Bühlmann, 2007; Hyttinen et al., 2016). Optimization-based methods, including NOTEARS-style extensions for temporal data, formulate

causal discovery as a continuous constrained optimization problem (Zheng et al., 2018; Pamfil et al., 2020; Yu et al., 2019). There are also methods based on non-Gaussian assumptions, information-theoretic criteria, and neural temporal models (Shimizu et al., 2006; Hyvärinen et al., 2010; Sun et al., 2015; Nauta et al., 2019).

2.3 Problem Setting

Let $\mathbf{X}_t = [x_1(t), x_2(t), \ldots, x_d(t)]^T \in \mathbb{R}^d$ denote a d-dimensional multivariate time series with $T > 1$ observed time points and node set V. The objective is not only to determine whether a directed relation exists between two variables, but also to identify the temporal scale at which the relation is expressed. Throughout this work, we consider three temporal scales indexed by $\mathcal{K} = \{\mathrm{S}, \mathrm{M}, \mathrm{L}\}$, where S, M, and L denote the short, mid, and long temporal scales, respectively. Instead of producing one global graph that summarizes all temporal effects, the desired output is the scale-resolved graph collection $\mathcal{G}_{\mathrm{SR}} = \{\mathcal{G}^{(k)}\}_{k\in\mathcal{K}} = \{\mathcal{G}^{(\mathrm{S})}, \mathcal{G}^{(\mathrm{M})}, \mathcal{G}^{(\mathrm{L})}\}$. The associated adjacency matrix is denoted by $A^{(k)} \in \mathbb{R}^{d\times d}$, with $A^{(k)}_{ij} \neq 0 \iff i \to j$ in $\mathcal{G}^{(k)}$, $k \in \mathcal{K}$.

3 Method

This section presents CausalSR, a scale-resolved causal discovery framework for time series with heterogeneous responses. The framework consists of three stages: multi-scale time-series representation, scale-specific causal discovery, and joint refinement across scales. It first separates temporal components associated with different response ranges, then estimates directed relations within each scale, and finally refines the resulting graphs through reconstruction consistency and scale-wise structural constraints. The final output is the collection of refined short-, mid-, and long-scale causal graphs.

3.1 Workflow Overview

Given the observed multivariate time series $\mathbf{X}_t$, CausalSR first transforms each variable into a continuous wavelet representation and partitions the resulting coefficients into the short, mid, and long temporal bands. Within each of the three bands, interaction strength and directional temporal evidence are estimated from cross-spectral and phase information. Regularized conditional-association screening is then applied to suppress weak indirect relations. Finally, all scale-specific graphs are jointly refined so that they remain connected to the original observations and satisfy scale-wise acyclicity.

Fig. 1 illustrates the overall workflow.

3.2 Multi-Scale Time-Series Representation

For each variable $x_i(t)$, the continuous wavelet coefficient at scale s and time location τ is

$$W_i(s,\tau) = \int_{-\infty}^{+\infty} x_i(t)\psi^*_{s,\tau}(t)\,dt, \tag{1}$$

where $\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right)$ is the scaled and shifted wavelet basis. Let Δt denote the sampling interval and f_c the center frequency of the mother wavelet; the pseudo-frequency associated with scale s is $f(s) = f_c/(s\Delta t)$.

The time-frequency plane is partitioned into three non-overlapping temporal bands:

$$B_k = \left[f^{(k)}_{\min}, f^{(k)}_{\max}\right], \qquad k \in \mathcal{K}. \tag{2}$$

The bands B_{S}, B_{M}, and B_{L} represent short-, mid-, and long-scale response ranges rather than individual lag values. For variable i, the coefficient retained within band B_k is

$$W^{(k)}_i(s,\tau) = \chi_{B_k}\big(f(s)\big)W_i(s,\tau), \tag{3}$$

where $\chi_{B_k}(\cdot)$ is the indicator function of B_k. The scale-specific time-domain component is reconstructed as $x^{(k)}_i(t) = \mathcal{W}^{-1}\left(W^{(k)}_i(s,\tau)\right)$.

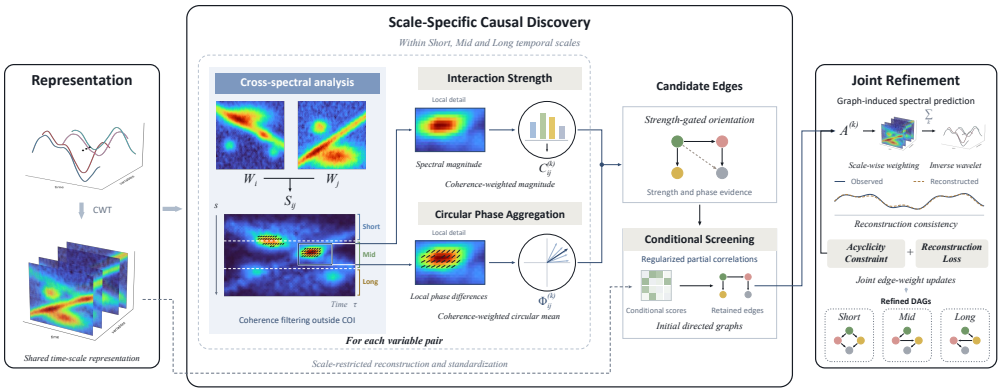


Figure 1: Overview of CausalSR. The observed multivariate time series is transformed into a shared time-scale representation. For each variable pair, cross-spectral analysis is restricted to short-, mid-, and long-scale bands; coherence-weighted interaction strength and circular phase aggregation provide edge strength and directional evidence. Candidate edges are then screened by regularized partial correlations to form initial scale-specific graphs. Joint refinement updates these graphs using reconstruction consistency and acyclicity constraints to produce refined short-, mid-, and long-scale causal structures.

3.3 Scale-Specific Causal Discovery

Within each temporal scale, CausalSR combines interaction strength, phase-based directional evidence, and multivariate conditional screening to estimate an initial directed graph. For each variable pair, the method first identifies reliable time-frequency regions, then aggregates interaction magnitude and phase ordering within those regions, and finally removes weak indirect relations through regularized screening.

For variables i and j, the local wavelet cross-spectrum is $S_{ij}(s,\tau) = W_i(s,\tau)W_j^*(s,\tau)$. The local phase of variable i and the phase difference between variables i and j are $\theta_i(s,\tau) = \arg(W_i(s,\tau))$ and $\phi_{ij}(s,\tau) = \theta_i(s,\tau) - \theta_j(s,\tau)$, respectively. The magnitude $|S_{ij}(s,\tau)|$ characterizes local interaction strength, whereas $\phi_{ij}(s,\tau)$ provides directional temporal evidence.

The normalized squared wavelet coherence is

$$R_{ij}^2(s,\tau) = \frac{\left|\mathcal{S}\left(s^{-1}S_{ij}(s,\tau)\right)\right|^2}{\mathcal{S}\left(s^{-1}|W_i(s,\tau)|^2\right)\mathcal{S}\left(s^{-1}|W_j(s,\tau)|^2\right)}, \tag{4}$$

where $\mathcal{S}(\cdot)$ denotes a smoothing operator.

Given the coherence threshold $\theta_c \in (0,1)$, the reliable coupling region is $M_{ij} = \{(s,\tau) : R_{ij}^2(s,\tau) \geq \theta_c, (s,\tau) \notin \text{COI}\}$, where COI denotes the cone of influence. For temporal band B_k, the scale-specific coherent region is

$$M_{ij}^{(k)} = \{(s,\tau) : f(s) \in B_k, \quad (s,\tau) \in M_{ij}\}. \tag{5}$$

Only coefficients within $M_{ij}^{(k)}$ are used to aggregate the interaction between variables i and j at scale k. This restriction reduces the influence of weak, unstable, and boundary-affected coefficients.

For $(s,\tau) \in M_{ij}^{(k)}$, the coherence weight is

$$w_{ij}^{(k)}(s,\tau) = \frac{R_{ij}^2(s,\tau)}{\sum_{(s',\tau')\in M_{ij}^{(k)}} R_{ij}^2(s',\tau')}. \tag{6}$$

The scale-specific interaction strength is

$$C_{ij}^{(k)} = \sum_{(s,\tau)\in M_{ij}^{(k)}} w_{ij}^{(k)}(s,\tau)\,|S_{ij}(s,\tau)|. \tag{7}$$

To aggregate the phase differences while preserving their circular structure, the weighted sine and cosine components are computed as

$$\begin{aligned} P_{ij}^{(k)} &= \sum_{(s,\tau)\in M_{ij}^{(k)}} w_{ij}^{(k)}(s,\tau)\sin\phi_{ij}(s,\tau), \\ Q_{ij}^{(k)} &= \sum_{(s,\tau)\in M_{ij}^{(k)}} w_{ij}^{(k)}(s,\tau)\cos\phi_{ij}(s,\tau). \end{aligned} \tag{8}$$

The aggregated phase is

$$\Phi_{ij}^{(k)} = \operatorname{atan2}\left(P_{ij}^{(k)}, Q_{ij}^{(k)}\right). \tag{9}$$

Using interaction-strength threshold τ_c, the adjacency entries for each unordered pair $\{i,j\}$ are assigned as

$$\left(A_{ij}^{(k)}, A_{ji}^{(k)}\right) = \begin{cases} \left(C_{ij}^{(k)}, 0\right), & C_{ij}^{(k)} \geq \tau_c,\ 0 < \Phi_{ij}^{(k)} < \dfrac{\pi}{2}, \\ \left(0, C_{ij}^{(k)}\right), & C_{ij}^{(k)} \geq \tau_c,\ -\dfrac{\pi}{2} < \Phi_{ij}^{(k)} < 0, \\ (0,0), & \text{otherwise.} \end{cases} \tag{10}$$

Self-loops are excluded by setting $A_{ii}^{(k)} = 0$.

For each scale k, the reconstructed component is standardized as $z_i^{(k)}(t) = \left(x_i^{(k)}(t) - \overline{x}_i^{(k)}\right)/\widehat{\sigma}_i^{(k)}$, where $\overline{x}_i^{(k)}$ and $\widehat{\sigma}_i^{(k)} > 0$ are the sample mean and standard deviation, and $\mathbf{z}^{(k)}(t) = [z_1^{(k)}(t), \ldots, z_d^{(k)}(t)]^T$. The empirical scale-specific correlation matrix is

$$E^{(k)} = \frac{1}{T-1}\sum_{t=1}^{T} \mathbf{z}^{(k)}(t)\mathbf{z}^{(k)}(t)^T. \tag{11}$$

To stabilize matrix inversion, the regularized precision matrix is computed as

$$\Omega^{(k)} = \left(E^{(k)} + \lambda I\right)^{-1}, \tag{12}$$

where $\lambda > 0$ and I is the $d \times d$ identity matrix. The normalized conditional-association score is

$$\rho_{ij|V\setminus\{i,j\}}^{(k)} = -\frac{\Omega_{ij}^{(k)}}{\sqrt{\Omega_{ii}^{(k)}\Omega_{jj}^{(k)}}}. \tag{13}$$

A retained candidate edge is removed when

$$\left|\rho_{ij|V\setminus\{i,j\}}^{(k)}\right| < \epsilon_\Omega. \tag{14}$$

This step suppresses relations that become weak after accounting for the remaining variables. It is used as a multivariate screening criterion, while the direction previously assigned from phase information is retained.

3.4 Joint Refinement Across Scales

The initial graphs are estimated independently within individual temporal scales. Although they encode scale-specific relations, their weights are primarily determined from local interaction evidence and may not jointly explain the original multivariate observations. CausalSR therefore refines the complete graph collection through reconstruction consistency and scale-wise structural constraints.

For temporal scale k, the graph-induced prediction of variable j is

$$\widehat{W}_j^{(k)}(s,\tau) = \sum_{\substack{i=1 \\ i\neq j}}^{d} A_{ij}^{(k)} W_i^{(k)}(s,\tau). \tag{15}$$

The predictions from all scales are combined as $\widehat{W}_j(s,\tau) = \sum_{k\in\mathcal{K}} \widehat{W}_j^{(k)}(s,\tau)$, and the corresponding time-domain prediction is $\widehat{x}_j(t) = \mathcal{W}^{-1}(\widehat{W}_j(s,\tau))$. The reconstruction loss, with the norm taken over the observed time points, is

$$L_{\mathrm{rec}} = \sum_{j=1}^{d} \|x_j(t) - \widehat{x}_j(t)\|_2^2 . \tag{16}$$

For every temporal scale, the structural function is

$$h\left(A^{(k)}\right) = \mathrm{tr}\left[\exp\left(A^{(k)} \circ A^{(k)}\right)\right] - d, \tag{17}$$

where $\circ$ denotes the Hadamard product and $\exp(\cdot)$ denotes the matrix exponential. The desired directed acyclic structure is imposed through $h(A^{(k)}) = 0$, $k \in \mathcal{K}$.

The final refinement objective combines reconstruction consistency and scale-wise structural validity:

$$\min_{\{A^{(k)}\}_{k\in\mathcal{K}}} \quad L_{\mathrm{rec}} + L_{\mathrm{DAG}}, \tag{18}$$

where

$$L_{\mathrm{DAG}} = \sum_{k\in\mathcal{K}} \left[\alpha_k h\left(A^{(k)}\right) + \frac{\rho}{2} h\left(A^{(k)}\right)^2\right]. \tag{19}$$

Here, α_k is the Lagrange multiplier associated with temporal scale k, and $\rho > 0$ is the penalty parameter. The graph weights, Lagrange multipliers, and penalty parameter are iteratively updated until the reconstruction objective stabilizes and the scale-wise structural violations become sufficiently small. Upon convergence, CausalSR returns $\mathcal{A}^* = \{A^{(\mathrm{S})*}, A^{(\mathrm{M})*}, A^{(\mathrm{L})*}\}$.

4 Theoretical Analysis

The following results concern scale-resolved reconstruction, coherent phase aggregation, regularized conditional screening, and scale-wise acyclicity. Complete proofs are provided in Appendix A; the acyclicity characterization follows Zheng et al. (2018).

Theorem 1 (Band-restricted reconstruction)**.** *Let the mother wavelet admit a valid inverse transform, and let $\{B_k\}_{k\in\mathcal{K}}$ denote the non-overlapping short, mid, and long temporal bands. Then*

$$\sum_{k\in\mathcal{K}} x_i^{(k)}(t) = \mathcal{W}^{-1}\left(\sum_{k\in\mathcal{K}} W_i^{(k)}(s,\tau)\right). \tag{20}$$

If the selected bands cover all retained frequencies, then

$$x_i(t) = \sum_{k\in\mathcal{K}} x_i^{(k)}(t) + r_i(t), \tag{21}$$

where $r_i(t)$ contains the unretained components and residual error.

Theorem 2 (Preservation of coherent phase ordering)**.** *Suppose that $M_{ij}^{(k)}$ is nonempty and*

$$0 < \phi_{ij}(s,\tau) < \frac{\pi}{2}, \qquad (s,\tau) \in M_{ij}^{(k)}. \tag{22}$$

Then

$$0 < \Phi_{ij}^{(k)} < \frac{\pi}{2}. \tag{23}$$

The negative-ordering case follows analogously. Thus, circular aggregation preserves a coherent local temporal ordering.

Theorem 3 (Well-posed conditional-association screening)**.** *For the empirical matrix $E^{(k)}$ defined in* (11) *and any $\lambda > 0$,*

$$E^{(k)} + \lambda I \tag{24}$$

is positive definite and invertible. Consequently, the score in (13) *is well defined.*

Theorem 4 (Scale-wise acyclicity)**.** *For a weighted adjacency matrix $A^{(k)}$,*

$$h\left(A^{(k)}\right) = \mathrm{tr}\left[\exp\left(A^{(k)} \circ A^{(k)}\right)\right] - d = 0 \tag{25}$$

if and only if the graph represented by $A^{(k)}$ is acyclic.

5 EXPERIMENTS

This section evaluates the proposed framework as a scale-resolved structural knowledge discovery tool for complex temporal data. The experiments examine overall structural recovery, real-world applicability, and generality under simpler uni-scale settings. We additionally visualize the recovered scale-specific relations in a real river system to assess whether the learned structures provide interpretable domain-level knowledge.

For controlled synthetic data, the data-generating process and scale-specific causal structures are known exactly. For CausalRivers and fMRI, the reported structural metrics measure consistency with the available reference graphs rather than absolute causal verification.

Table 1 summarizes the scale-dependent synthetic, CauseMe (Mooij et al., 2016), CausalRivers, and fMRI datasets (Stein et al., 2025; Smith et al., 2011). Experimental details and supplementary results are provided in Appendix F.

Table 1: Summary of datasets and their roles in the experimental evaluation.

Attribute	**Scale-dependent synthetic**	**CauseMe**	**CausalRivers**	**fMRI**
Controlled	Yes	Yes	No	No
Real-world	No	No	Yes	Yes
Scale-dependent	Yes	No	Yes	Yes
Uni-scale	No	Yes	No	No
Ground truth	Exact	Exact	Partial/domain	Reference
Main evaluation purpose	Structural recovery and reconstruction evaluation	Generality under simpler uni-scale settings	Evaluation on physical temporal systems	Evaluation on high-dimensional temporal dynamics

We compare the proposed framework with representative time-series causal discovery methods, including PCMCI (Runge et al., 2019b), DYNOTEARS (Pamfil et al., 2020), oCSE (Sun et al., 2015), Granger-pw (Granger, 1969), and VAR-LiNGAM (Hyvärinen et al., 2010). For the CauseMe uni-scale benchmark, we compare CausalSR with DirectLiNGAM (Shimizu et al., 2011), PCMCI (Runge et al., 2019b), and VAR (Lütkepohl, 2005). For structural recovery, we report F1, true positive rate (TPR), structural Hamming distance (SHD), and normalized structural Hamming distance (nSHD). Reported curves and table entries are obtained from repeated runs, with means used as the central estimates and displayed uncertainty representing variation across runs.

5.1 SCALE-DEPENDENT SYNTHETIC BENCHMARK

The scale-dependent synthetic benchmark evaluates causal discovery under heterogeneous response patterns. Because relations from different scales are superposed in the observed time domain, a single-resolution analysis may miss scale-specific edges or retain spurious relations induced by scale mixing. Fig. 2 reports F1, TPR, and nSHD as the number of variables increases.

CausalSR maintains the strongest F1 performance across all tested graph sizes, indicating a more favorable balance between false positive and false negative edges. Its TPR remains among the strongest methods as dimensionality increases, while several single-resolution baselines deteriorate more rapidly. CausalSR also produces substantially lower nSHD over most settings and remains competitive at the largest graph size. The simultaneous improvement in F1 and nSHD suggests that separating heterogeneous response components reduces both missing and structurally misplaced edges.

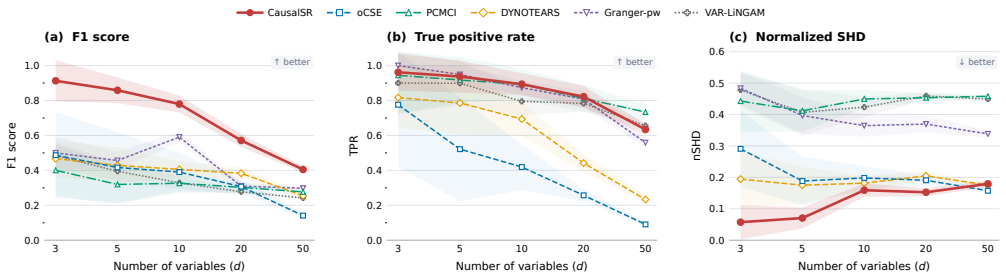


Figure 2: Structural recovery on the scale-dependent synthetic benchmark. Panels (a)–(c) show F1, TPR, and nSHD as graph size increases. Curves show means and shaded regions show variability across repeated runs.

5.2 Real-World Temporal Datasets

We next evaluate the framework on CausalRivers and fMRI. Compared with controlled synthetic systems, real temporal data contain stronger measurement noise, incomplete observations, and domain-specific external effects. The available graphs are therefore used as reference structures. CausalRivers represents a physical river system containing directional and scale-dependent propagation patterns (Stein et al., 2025). The fMRI data describe high-dimensional temporal interactions among brain regions, whose responses may also occur over different temporal ranges (Friston et al., 2003; Roebroeck et al., 2005; Seth et al., 2015).

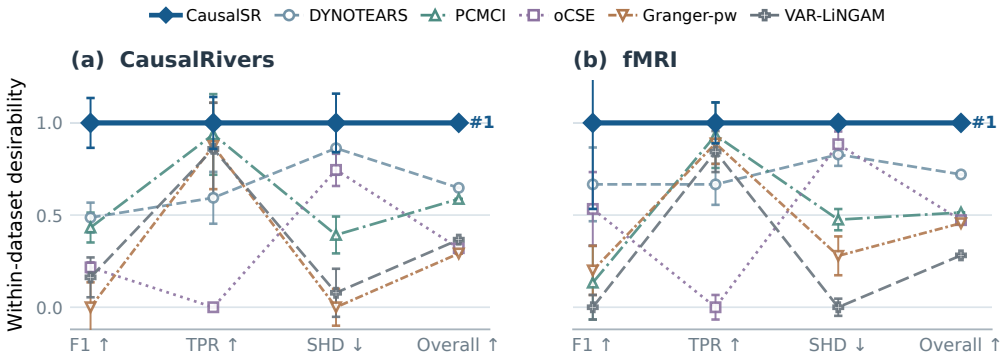


Figure 3: Performance profiles on (a) CausalRivers and (b) fMRI. The axes show F1, TPR, SHD, and Overall on a within-dataset desirability scale, with SHD reversed so that higher is better. Error bars show standard deviations; CausalSR is the thick dark-blue line.

As shown in Fig. 3, CausalSR exhibits the strongest and most consistent profile on both datasets. The profiles summarize cross-metric trade-offs, with Overall defined in Appendix F.3. The results should be interpreted as reference-based structural consistency because the available graphs may not represent complete causal truth.

5.3 Generality on Uni-Scale Settings

The preceding experiments focus on systems with heterogeneous response scales. We further evaluate whether the proposed framework remains useful when the data are closer to a uni-scale causal process. The CauseMe benchmark is evaluated with $d \in \{3, 5, 10, 20\}$.

Fig. 4 shows that all methods deteriorate as the number of variables increases, but their degradation profiles differ. Although DirectLiNGAM exhibits slightly stronger relative F1 retention, CausalSR maintains the strongest absolute endpoint profile. The result indicates that scale-resolved processing does not cause the method to collapse when scale heterogeneity is limited, although its relative advantage is naturally smaller than in the scale-dependent setting.

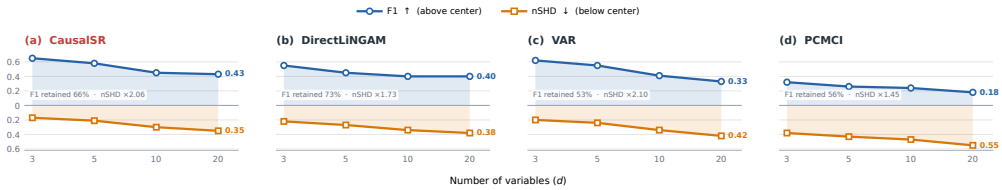


Figure 4: CauseMe results for (a) CausalSR, (b) DirectLiNGAM, (c) VAR, and (d) PCMCI using shared axis ranges. F1 is plotted above zero and nSHD below zero; annotations report endpoint values, F1 retention, and nSHD growth.

5.4 Scale-Resolved Recovery in CausalRivers

To examine the interpretability of the recovered structures, we further visualize a representative 10-node CausalRivers subnetwork. The three panels in Fig. 5 show the recovered relations associated with short-, mid-, and long-scale components. The displayed distance intervals summarize the river-network separation of the reference relations in each panel; the scale-specific graphs themselves are obtained from the corresponding temporal components.

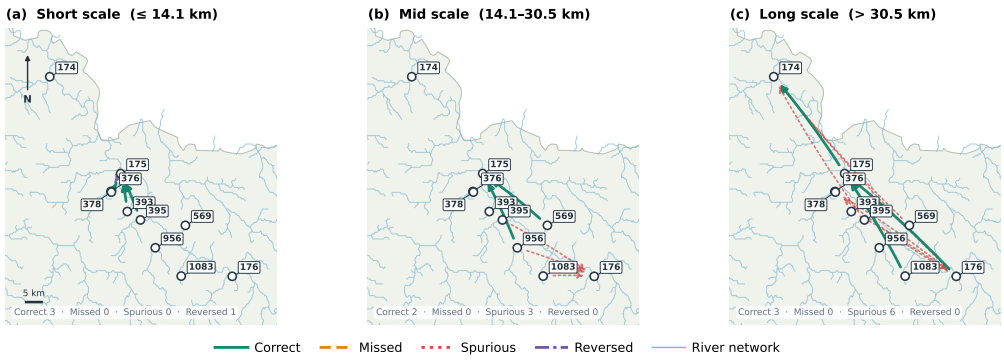


Figure 5: Scale-resolved recovery on a 10-node CausalRivers subnetwork. Panels (a)–(c) show short, mid, and long scale relations. Line styles distinguish correct, missed, spurious, and reversed edges; background lines show the river network.

This visualization indicates that the framework can preserve meaningful river-network propagation patterns across scales, while also revealing that longer-range structures are more difficult to distinguish because they involve broader and potentially more confounded dependencies.

6 Conclusion

This paper presented a scale-resolved causal discovery framework for time series with heterogeneous responses, aiming to reduce scale mixing and recover interpretable structural knowledge from complex temporal data. By organizing multivariate time series into scale-specific representations and refining the learned graphs through reconstruction consistency and structural constraints, the framework provides a way to distinguish causal relations operating at different temporal ranges. The experiments show that CausalSR achieves competitive structural recovery across the tested graph sizes while preserving competitive reconstruction fidelity. Several limitations remain. First, the current framework relies on predefined temporal bands, and inappropriate band partitions may weaken structural recovery. Second, the method mainly focuses on causal relations that can be reasonably separated across scales, while strong cross-scale modulation is not explicitly modeled. Third, as with many observational causal discovery methods, latent confounding may affect the recovered structures in real-world datasets. Future work will investigate adaptive band selection, explicit cross-scale causal modeling, and more scalable extensions for larger temporal systems.

AI USE STATEMENT

In this work, we used ChatGPT to assist with manuscript writing and editing, including improving language clarity and readability, organizing the main text and appendix, adapting the LaTeX formatting, and editing figure captions and the wording of accompanying descriptions. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Paul S Addison. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC Press, 2 edition, 2017.

Luís Aguiar-Conraria and Maria Joana Soares. The continuous wavelet transform: Moving beyond uni- and bivariate analysis. *Journal of Economic Surveys*, 28(2):344–375, 2014.

Lionel Barnett and Anil K Seth. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223:50–68, 2014.

Keith J Beven. *Rainfall-runoff modelling: the primer*. John Wiley & Sons, 2 edition, 2012.

Günter Blöschl and Murugesu Sivapalan. Scale issues in hydrological modelling: a review. *Hydrological Processes*, 9(3–4):251–290, 1995.

Bernard Cazelles, Mario Chavez, Dominique Berteaux, Frédéric Ménard, Jon Olav Vik, Stéphanie Jenouvrier, and Nils C Stenseth. Wavelet analysis of ecological time series. *Oecologia*, 156(2): 287–304, 2008.

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*, 40(1):294–321, 2012.

Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

Michael Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153:233–268, 2012.

Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54, 1996.

Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *NeuroImage*, 19(4): 1273–1302, 2003.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019.

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

Aslak Grinsted, John C Moore, and Svetlana Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11(5/6): 561–566, 2004.

Antti Hyttinen, Sergey Plis, Matti Järvisalo, Frederick Eberhardt, and David Danks. Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, volume 52 of *Proceedings of Machine Learning Research*, pp. 216–227. PMLR, 2016.

Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O Hoyer. Estimation of a structural vector autoregression model using non-Gaussianity. *Journal of Machine Learning Research*, 11 (56):1709–1731, 2010.

Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8(22):613–636, 2007.

Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, 2005. doi: 10.1007/978-3-540-27752-1.

Daniel Malinsky and Peter Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of the 2018 ACM SIGKDD Workshop on Causal Discovery*, volume 92 of *Proceedings of Machine Learning Research*, pp. 23–47. PMLR, 2018.

Stéphane Mallat. *A wavelet tour of signal processing*. Academic Press, 2 edition, 1999.

Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research*, 17(32):1–102, 2016.

Meike Nauta, Doina Bucur, and Christin Seifert. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, 2019.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. DYNOTEARS: Structure learning from time-series data. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1595–1605. PMLR, 2020.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2000.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT Press, 2017.

Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, 25(1):230–242, 2005.

Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in earth system sciences. *Nature Communications*, 10(1):2553, 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 Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12(33):1225–1248, 2011.

Christopher A Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for FMRI. *NeuroImage*, 54(2):875–891, 2011.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT Press, 2 edition, 2000.

Gideon Stein, Maha Shadaydeh, Jan Blunk, Niklas Penzel, and Joachim Denzler. CausalRivers - scaling up benchmarking of causal discovery for real-world time-series. In *The Thirteenth International Conference on Learning Representations*, 2025.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.

Jie Sun, Dane Taylor, and Erik M Bollt. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1):73–106, 2015.

Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young (eds.), *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pp. 366–381. Springer, 1981.

Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. Neural Granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2022.

Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.

Yue Yu, Jie Chen, Tian Gao, and Mo Yu. DAG-GNN: DAG structure learning with graph neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7154–7163. PMLR, 2019.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. DAGs with NO TEARS: Continuous optimization for structure learning. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

A Complete Proofs

For readability, each theorem from Section 4 is restated before its proof. All notation and assumptions are exactly those used in the main text.

A.1 Band-Restricted Reconstruction

Theorem 1 (Band-restricted reconstruction). Let the mother wavelet admit a valid inverse transform, and let $\{B_k\}_{k\in\mathcal{K}}$ denote the non-overlapping short, mid, and long temporal bands. Then

$$\sum_{k\in\mathcal{K}} x_i^{(k)}(t) = \mathcal{W}^{-1}\left(\sum_{k\in\mathcal{K}} W_i^{(k)}(s,\tau)\right).$$

If the selected bands cover all retained frequencies, then

$$x_i(t) = \sum_{k\in\mathcal{K}} x_i^{(k)}(t) + r_i(t),$$

where $r_i(t)$ contains the unretained components and residual error.

Proof of Theorem 1. From (3),

$$\sum_{k\in\mathcal{K}} W_i^{(k)}(s,\tau) = \left[\sum_{k\in\mathcal{K}} \chi_{B_k}\big(f(s)\big)\right] W_i(s,\tau). \tag{26}$$

By linearity of the inverse transform,

$$\begin{aligned}\mathcal{W}^{-1}\left(\sum_{k\in\mathcal{K}} W_i^{(k)}\right) &= \sum_{k\in\mathcal{K}} \mathcal{W}^{-1}\left(W_i^{(k)}\right)\\ &= \sum_{k\in\mathcal{K}} x_i^{(k)}.\end{aligned} \tag{27}$$

The frequencies outside the retained bands can be collected into $r_i(t)$, yielding (21). □

The result shows that the scale-specific components remain explicitly connected to the original observations.

A.2 Coherent Phase Aggregation

Theorem 2 (Preservation of coherent phase ordering). Suppose that $M_{ij}^{(k)}$ is nonempty and

$$0 < \phi_{ij}(s,\tau) < \frac{\pi}{2}, \qquad (s,\tau) \in M_{ij}^{(k)}.$$

Then

$$0 < \Phi_{ij}^{(k)} < \frac{\pi}{2}.$$

Proof of Theorem 2. Condition (22) implies

$$\sin\phi_{ij}(s,\tau) > 0, \qquad \cos\phi_{ij}(s,\tau) > 0. \tag{28}$$

Because the weights in (6) are nonnegative,

$$P_{ij}^{(k)} > 0, \qquad Q_{ij}^{(k)} > 0. \tag{29}$$

Therefore,

$$\Phi_{ij}^{(k)} = \operatorname{atan2}\left(P_{ij}^{(k)}, Q_{ij}^{(k)}\right) \tag{30}$$

lies in the first quadrant, proving (23). □

The negative-ordering case follows analogously. Thus, circular aggregation preserves a coherent local temporal ordering.

A.3 Regularized Conditional Screening

Theorem 3 (Well-posed conditional-association screening). For the empirical matrix $E^{(k)}$ defined in (11) and any $\lambda > 0$,

$$E^{(k)} + \lambda I$$

is positive definite and invertible. Consequently, the score in (13) is well defined.

Proof of Theorem 3. For every nonzero vector $\mathbf{v} \in \mathbb{R}^d$,

$$\begin{aligned} \mathbf{v}^T \left(E^{(k)} + \lambda I \right) \mathbf{v} &= \frac{1}{T-1} \sum_{t=1}^{T} \left[\mathbf{v}^T \mathbf{z}^{(k)}(t) \right]^2 + \lambda \|\mathbf{v}\|_2^2 \\ &> 0. \end{aligned} \tag{31}$$

Hence,

$$E^{(k)} + \lambda I \succ 0, \tag{32}$$

and its inverse exists. The diagonal entries of a positive definite precision matrix are positive, so

$$\sqrt{\Omega_{ii}^{(k)} \Omega_{jj}^{(k)}} > 0. \tag{33}$$

Therefore, the normalized conditional-association score is well defined. □

This result guarantees numerical validity of the screening step even when the empirical correlation matrix is singular or poorly conditioned.

A.4 Scale-Wise Acyclicity

Theorem 4 (Scale-wise acyclicity). For a weighted adjacency matrix $A^{(k)}$,

$$h\left(A^{(k)}\right) = \operatorname{tr}\left[\exp\left(A^{(k)} \circ A^{(k)}\right)\right] - d = 0$$

if and only if the graph represented by $A^{(k)}$ is acyclic.

Proof of Theorem 4. Let

$$B = A^{(k)} \circ A^{(k)}. \tag{34}$$

Since B is elementwise nonnegative,

$$h\left(A^{(k)}\right) = \sum_{\ell=1}^{\infty} \frac{\operatorname{tr}\left(B^{\ell}\right)}{\ell!}. \tag{35}$$

The diagonal entries of B^{ℓ} collect the weights of closed directed walks of length ℓ. If the graph is acyclic, no such walk exists and

$$\operatorname{tr}\left(B^{\ell}\right) = 0 \tag{36}$$

for every $\ell \geq 1$, yielding $h(A^{(k)}) = 0$.

Conversely, if a directed cycle exists, then

$$\operatorname{tr}\left(B^{\ell}\right) > 0 \tag{37}$$

for at least one cycle length ℓ. Equation (35) then gives $h(A^{(k)}) > 0$. □

Together, these results establish the reconstruction consistency, directional stability, numerical well-posedness, and structural validity required by the proposed scale-resolved causal discovery framework.

B Additional Motivation and Related Work

B.1 Additional Motivation

Knowing which variables directly affect a target signal can help reduce spurious correlations, improve the interpretability of predictive models, and support more reliable decisions under changing system conditions. Therefore, in complex temporal systems, causal discovery should be viewed not only as a graph recovery problem, but also as a knowledge discovery process that extracts useful structural information from temporal data.

The proposed framework adopts a scale-specific view of temporal causality. Rather than forcing all interactions into one global graph or one fixed lag structure, the framework assigns each temporal scale a corresponding causal graph, where edges represent direct structural relations among scale-specific components. This design enables the framework to distinguish short-term responses, delayed influences, and long-term dependencies, producing scale-resolved structural knowledge from complex time series. Such knowledge provides interpretable causal relations that can support downstream temporal analysis and knowledge-driven decision making.

B.2 Additional Temporal Data Representation Discussion

The time-frequency representations discussed in Section 2 have been used in signal processing, environmental monitoring, biomedical analysis, traffic forecasting, and other temporal data applications (Cazelles et al., 2008; Aguiar-Conraria & Soares, 2014). By decomposing temporal signals into components associated with different frequency or scale ranges, time-frequency methods provide a natural way to separate short-term fluctuations from long-term variations.

However, most existing temporal representations are designed mainly for prediction, classification, compression, or denoising. They do not directly address how representation affects causal structure learning. For causal discovery, this issue is important because mixed temporal effects may lead to spurious dependencies, unstable edge directions, or causal graphs that are difficult to interpret. Recent studies on causal representation learning also suggest that the choice of representation can strongly influence the quality of causal analysis (Schölkopf et al., 2021). Our work follows this perspective by treating scale-resolved temporal representation as a preprocessing and structural organization strategy for discovering reliable causal knowledge from complex time series.

B.3 Causality for Temporal Analysis

Causal discovery is often evaluated as a graph recovery problem, but in many data engineering applications it also serves as a tool for structural knowledge discovery (Fayyad et al., 1996). Recovered causal structures can help explain how variables interact, identify direct drivers of target signals, reduce spurious correlations, and provide interpretable evidence for forecasting, diagnosis, and decision making (Pearl et al., 2016). In complex systems such as hydrological monitoring, traffic networks, biomedical signals, and industrial processes, such structural knowledge can be as important as predictive accuracy itself.

Existing studies have used causal information to improve prediction, diagnosis, intervention analysis, and interpretability (Sugihara et al., 2012; Barnett & Seth, 2014). For example, causal relations can guide feature selection, constrain predictive models, or support counterfactual and intervention reasoning. However, the usefulness of these downstream applications depends on the reliability of the recovered causal structures. If causal discovery is affected by heterogeneous responses and scale mixing, the resulting graph may be unstable, difficult to interpret, or disconnected from the temporal patterns that generate the observations.

From this perspective, the key issue is not only whether a causal graph can be learned, but whether the temporal data are represented in a way that allows meaningful causal structures to be recovered. A scale-resolved view is useful because it separates temporal components before structure learning and associates each recovered graph with its corresponding temporal range. This makes it possible to distinguish short-term responses, delayed influences, and long-term dependencies, while keeping the learned structures connected to the original observations through reconstruction and structural constraints. Different from methods that directly apply causal discovery to raw time-domain sig-

nals, the proposed framework aims to produce more reliable scale-resolved structural knowledge for downstream temporal analysis and knowledge-driven decision making.

C Additional Problem Details

C.1 Full Problem-Setting Context

This appendix expands the problem setting in Section 2.3, covering the conventional lag-based view, scale-specific structural formulation, and assumptions.

In contrast to the conventional formulation, we consider temporal systems with heterogeneous responses, where different causal interactions may be expressed through different temporal ranges. For example, some effects may appear as short-term responses, whereas others may emerge as delayed influences or long-term evolving dependencies. When these relations are estimated directly from the original observations under a single temporal resolution, their distinct response patterns may be mixed, making the resulting structure difficult to recover and interpret.

Each scale-specific causal graph $\mathcal{G}^{(k)} = (V, E^{(k)})$, $k \in \mathcal{K}$, describes the direct structural relations among variables within the corresponding temporal range.

Under this formulation, scale-resolved causal discovery addresses three related difficulties. First, heterogeneous responses may cause scale mixing, where short-term and long-term effects interfere with each other in the original time-domain observations. Second, relations operating over different temporal ranges are difficult to separate when they are represented only through one predefined lag structure or summarized into one global graph. Third, scale-specific graphs should remain connected to the original observations rather than becoming isolated artifacts of temporal decomposition.

C.2 Conventional Temporal Causal View

A common formulation of temporal causal discovery represents the current value of each variable as a function of the past observations of its causal parents. For variable $x_j(t)$, the corresponding temporal structural equation can be written as

$$x_j(t) = f_j\left(\{x_i(t-\ell) : i \in \mathrm{Pa}(j), \ell \in \mathcal{L}\}, \eta_j(t)\right), \tag{38}$$

where $\mathrm{Pa}(j)$ denotes the parent set of node j, $\mathcal{L} = \{1, \ldots, L\}$ is a predefined lag set, $\eta_j(t)$ is a disturbance term, and $f_j(\cdot)$ characterizes the temporal structural mechanism.

A commonly used linear representation is

$$x_j(t) = \sum_{i=1}^{d} \sum_{\ell=1}^{L} A_{ij}^{(\ell)} x_i(t-\ell) + \eta_j(t), \tag{39}$$

where $A_{ij}^{(\ell)}$ measures the lag-ℓ influence from variable i to variable j. The lag-specific coefficients are commonly summarized into a single variable-level causal graph

$$\mathcal{G}_{\mathrm{global}} = (V, E), i \to j \in E \iff \exists\, \ell \in \mathcal{L} \text{ such that } A_{ij}^{(\ell)} \neq 0. \tag{40}$$

This formulation is effective when causal effects can be adequately characterized by a predefined lag range under a common temporal resolution. However, although multiple lags may be considered, the relations are still estimated from the same time-domain observations and are generally summarized in one global graph. Consequently, short-term responses, delayed influences, and long-term dependencies may be superposed or merged when they are expressed through different temporal patterns.

C.3 Scale-Specific Causal View

To address these difficulties, we extend the conventional global temporal causal view to a scale-specific structural formulation (Pearl, 2009). The central premise is that causal interactions do not necessarily share one common temporal scale. Instead, different variables may influence each other

through mechanisms that are active over different temporal ranges. Under this view, the observed series of variable j is represented as the superposition of its short-, mid-, and long-scale components:

$$x_j(t) = \sum_{k \in \mathcal{K}} x_j^{(k)}(t) + \varepsilon_j(t), \tag{41}$$

where $x_j^{(k)}(t)$ denotes the component of variable j at scale $k \in \mathcal{K}$, and $\varepsilon_j(t)$ is an observation noise term.

For each $k \in \mathcal{K}$, the scale-specific causal relation is described by

$$x_j^{(k)}(t) = f_j^{(k)} \left(\mathbf{X}^{(k)}_{\mathrm{Pa}^{(k)}(j)}(t), \eta_j^{(k)}(t) \right), \tag{42}$$

where $\mathrm{Pa}^{(k)}(j)$ denotes the parent set of node j in $\mathcal{G}^{(k)}$, $\mathbf{X}^{(k)}_{\mathrm{Pa}^{(k)}(j)}(t)$ collects the scale-specific parent information within the temporal range represented by band B_k, $\eta_j^{(k)}(t)$ is a scale-specific disturbance term, and $f_j^{(k)}(\cdot)$ is the structural function at that scale.

An edge $i \to j$ in $\mathcal{G}^{(k)}$ represents a direct structural relation from the scale-k component of variable i to that of variable j. The same pair of variables need not behave identically across all scales. A relation may be strong at one temporal scale, weak or absent at another, or expressed through different response patterns across scales. This scale-specific view provides the basis for separating causal relations that would otherwise be mixed in a single time-domain graph.

C.4 Assumptions and Scope

The proposed framework is developed under the following assumptions.

Scale-resolved structure. The dominant temporal interactions can be approximated by a finite number of temporal scales. The main structural effects are therefore assumed to be at least partially separable in the time-frequency representation, although perfect separation is not required.

Sufficient temporal resolution. The sampling frequency is high enough to preserve the temporal variations associated with the scales of interest. If the observations are too sparse, short-term responses or delayed effects may not be reliably represented.

Causal sufficiency. The observed variables capture the main system components relevant to the discovered relations, and strong hidden confounding is limited. As in many causal discovery settings, severe latent confounding may distort the recovered graph structure (Colombo et al., 2012; Malinsky & Spirtes, 2018).

Local structural approximation. Within each temporal scale, the structural relations can be locally approximated for structure estimation and refinement. This assumption does not require the complete observed series to follow one global linear structural causal model.

Directional phase interpretation. Within a coherent temporal band, the sign of the aggregated phase is assumed to reflect the dominant lead–lag ordering of an active relation. Phase information is therefore used as directional temporal evidence together with interaction strength, multivariate screening, and acyclicity, rather than as a standalone causal identification condition.

Under these assumptions, the framework is most suitable for temporal systems with heterogeneous but reasonably separable response patterns. Its advantage may be limited in genuinely uni-scale settings. When strong cross-scale modulation dominates the system, additional mechanisms may be required to model interactions across scales explicitly.

D Additional Method Details

The multivariate wavelet representation is obtained by stacking the component-wise coefficients:

$$W(s,\tau) = \left[W_1(s,\tau), W_2(s,\tau), \ldots, W_d(s,\tau) \right]^T. \tag{43}$$

The representation $W_i(s,\tau)$ preserves both the temporal location and characteristic scale of local signal variations. It therefore provides the basis for separating interactions expressed through different temporal response ranges.

The corresponding scale-specific multivariate representation is

$$W^{(k)}(s,\tau) = \left[W_1^{(k)}(s,\tau), \ldots, W_d^{(k)}(s,\tau)\right]^T. \tag{44}$$

The resulting components preserve temporal locality while separating interactions associated with different response ranges. They are subsequently used for scale-specific interaction estimation, conditional-association screening, and graph refinement.

The interaction strength and aggregated phase provide complementary information. The former determines whether a sufficiently strong relation exists, whereas the latter summarizes the dominant lead–lag ordering within the coherent region. Using $\mathrm{atan2}$ preserves quadrant information and avoids the ambiguity of an ordinary arctangent.

The graph-induced reconstruction serves as a local linear consistency criterion for adjusting the scale-specific edge weights. It is not intended to reproduce the complete data-generating process or all exogenous innovations.

E PSEUDOCODE

Algorithms 1–3 describe the representation, scale-specific discovery, and joint refinement stages, respectively.

Algorithm 1 Multi-Scale Time-Series Representation.

Input: Multivariate time series $\mathbf{X}_t = [x_1(t), \ldots, x_d(t)]^T$, mother wavelet ψ, bands $\{B_k\}_{k\in\mathcal{K}}$
Output: Scale-specific coefficients $\{W^{(k)}\}_{k\in\mathcal{K}}$

for $i = 1, \ldots, d$ **do**
  Compute $W_i(s,\tau) \leftarrow \mathrm{CWT}(x_i(t); \psi)$
  Compute $\theta_i(s,\tau) \leftarrow \arg(W_i(s,\tau))$
end for
for each $k \in \mathcal{K}$ **do**
  for $i = 1, \ldots, d$ **do**
    Construct $W_i^{(k)}(s,\tau) \leftarrow \chi_{B_k}(f(s))W_i(s,\tau)$
  end for
  Stack $W^{(k)}(s,\tau) \leftarrow [W_1^{(k)}(s,\tau), \ldots, W_d^{(k)}(s,\tau)]^T$
end for
return $\{W^{(k)}\}_{k\in\mathcal{K}}$

Algorithm 2 Scale-Specific Causal Discovery.

Input: Coefficients $\{W_i^{(k)}(s,\tau)\}$, bands $\{B_k\}_{k\in\mathcal{K}}$, thresholds $\theta_c, \tau_c, \epsilon_\Omega$, regularizer λ
Output: Initial scale-specific graphs $\mathcal{A} = \{A^{(k)}\}_{k\in\mathcal{K}}$

for each $k \in \mathcal{K}$ **do**
  Initialize $A^{(k)} \leftarrow 0$
  for each unordered variable pair $\{i, j\}$ **do**
    Compute $S_{ij}(s,\tau)$, $\phi_{ij}(s,\tau)$, and $R^2_{ij}(s,\tau)$
    Construct $M_{ij}^{(k)}$
    if $M_{ij}^{(k)} = \emptyset$ **then continue**
    Compute $C_{ij}^{(k)}$ and $\Phi_{ij}^{(k)}$
    if $C_{ij}^{(k)} \geq \tau_c$ and $0 < \Phi_{ij}^{(k)} < \pi/2$ **then**
      Set $A_{ij}^{(k)} \leftarrow C_{ij}^{(k)}$
    else if $C_{ij}^{(k)} \geq \tau_c$ and $-\pi/2 < \Phi_{ij}^{(k)} < 0$ **then**
      Set $A_{ji}^{(k)} \leftarrow C_{ij}^{(k)}$
    end if
  end for
  Reconstruct $\{x_i^{(k)}(t)\}_{i=1}^d$ from $\{W_i^{(k)}(s,\tau)\}_{i=1}^d$
  Construct $E^{(k)}$ and compute $\Omega^{(k)} \leftarrow (E^{(k)} + \lambda I)^{-1}$
  Remove retained edges satisfying (14)
end for
return $\mathcal{A} = \{A^{(k)}\}_{k\in\mathcal{K}}$

Algorithm 3 Joint Refinement Across Scales.

Input: Original series $\mathbf{X}_t$, coefficients $\{W^{(k)}\}_{k\in\mathcal{K}}$, initial graphs $\mathcal{A}$, optimization parameters
Output: Refined scale-specific graphs $\mathcal{A}^*$

Initialize $\{A^{(k)}\}_{k\in\mathcal{K}}$ from $\mathcal{A}$
Initialize multipliers $\{\alpha_k\}_{k\in\mathcal{K}}$ and penalty parameter ρ
repeat
  for each $k \in \mathcal{K}$ **do**
    for $j = 1, \ldots, d$ **do**
      Compute $\widehat{W}_j^{(k)}(s,\tau) \leftarrow \sum_{i\neq j} A_{ij}^{(k)} W_i^{(k)}(s,\tau)$
    end for
  end for
  Combine $\{\widehat{W}^{(k)}\}_{k\in\mathcal{K}}$ and recover $\widehat{\mathbf{X}}_t$ by inverse wavelet reconstruction
  Compute L_{rec} and L_{DAG}
  Update $\{A^{(k)}\}_{k\in\mathcal{K}}$ by gradient-based optimization
  Update $\{\alpha_k\}_{k\in\mathcal{K}}$ and ρ
until convergence
return $\mathcal{A}^* = \{A^{(\mathrm{S})*}, A^{(\mathrm{M})*}, A^{(\mathrm{L})*}\}$

F Experimental Setup and Additional Results

F.1 Dataset Context

We consider four types of datasets, as summarized in Table 1. The scale-dependent synthetic benchmark is the main controlled setting and is used to evaluate causal structure recovery when different relations operate at different temporal scales. The CauseMe benchmark provides a simpler uni-scale setting and is used to examine whether the framework remains competitive when temporal-scale heterogeneity is limited (Mooij et al., 2016). CausalRivers and fMRI are used to evaluate the framework on real temporal systems with more complex dynamic interactions (Stein et al., 2025; Smith et al., 2011). For these datasets, the available graphs are treated as reference structures.

F.2 Synthetic Data Generation

To generate multivariate time series with heterogeneous response scales, we construct one scale-specific directed acyclic graph for each of the short, mid, and long temporal bands:

$$\mathcal{G}^{(k)} = \left(V, E^{(k)}\right), \qquad k \in \mathcal{K}, \tag{45}$$

Different directed edges are assigned to the three graphs so that the resulting system contains short-, mid-, and long-scale temporal responses.

In this data-generation subsection, $E^{(k)}$ denotes an edge set, rather than the empirical correlation matrix in Section 3. Let $z_j^{(k)}(t)$ denote the latent component of variable j associated with scale k; here this notation refers to the generated component, not the standardized reconstructed component used in the method. It is generated from a stable scale-specific autoregressive structural process:

$$z_j^{(k)}(t) = \sum_{i \in \mathrm{Pa}^{(k)}(j)} a_{ij}^{(k)} z_i^{(k)}\left(t - \ell_{ij}^{(k)}\right) + u_j^{(k)}(t), \tag{46}$$

where $\mathrm{Pa}^{(k)}(j)$ is the parent set of node j at scale k, $a_{ij}^{(k)}$ is the corresponding structural coefficient, and $\ell_{ij}^{(k)}$ is the scale-dependent delay. The coefficients are selected so that the generated autoregressive system remains stable.

The scale-specific innovation is obtained by restricting an independent noise process to the corresponding frequency band:

$$u_j^{(k)}(t) = \mathcal{B}_k\left[\xi_j^{(k)}(t)\right], \qquad \xi_j^{(k)}(t) \sim \mathcal{N}\left(0, \sigma_k^2\right), \tag{47}$$

where $\mathcal{B}_k(\cdot)$ denotes the band-pass operator associated with B_k. The observed signal is constructed by superposing all scale-specific components:

$$x_j(t) = \sum_{k \in \mathcal{K}} z_j^{(k)}(t) + \varepsilon_j(t), \qquad \varepsilon_j(t) \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right). \tag{48}$$

Equations (46)–(48) provide exact ground-truth edge directions and scale assignments while allowing multiple response ranges to be mixed in the observed time-domain signals. We vary the number of variables over $d \in \{3, 5, 10, 20, 50\}$ and generate independent graph and noise realizations for each configuration.

F.3 Evaluation Metrics

Let TP, FP, and FN denote the numbers of true positive, false positive, and false negative edges. Precision, recall, and F1 are defined as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{49}$$

and

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{50}$$

TPR is equivalent to recall. SHD counts the edge additions, deletions, and reversals required to transform an estimated graph into its reference graph. The normalized value is

$$\mathrm{nSHD} = \frac{\mathrm{SHD}}{d(d-1)}. \tag{51}$$

For reconstruction evaluation on the controlled synthetic benchmark, we additionally report

$$\mathrm{MSE} = \frac{1}{dT}\sum_{i=1}^{d}\sum_{t=1}^{T}\left(x_i(t) - \widehat{x}_i(t)\right)^2 \tag{52}$$

and

$$\mathrm{MAE} = \frac{1}{dT}\sum_{i=1}^{d}\sum_{t=1}^{T}\left|x_i(t) - \widehat{x}_i(t)\right|. \tag{53}$$

For the robustness analysis, we define an integrated quality score

$$Q = \sqrt[3]{\mathrm{F1}\cdot\mathrm{TPR}\cdot(1-\mathrm{nSHD})}. \tag{54}$$

The geometric mean penalizes configurations that perform poorly on any one of the three structural criteria. A larger Q indicates a better balance between edge accuracy, true-edge recovery, and global structural error.

For the real-world comparison, we additionally report a within-dataset overall desirability score. For dataset D and method m, the F1 and TPR scores are normalized as

$$s_r^{(D,m)} = \frac{\overline{r}^{(D,m)} - \min_q \overline{r}^{(D,q)}}{\max_q \overline{r}^{(D,q)} - \min_q \overline{r}^{(D,q)}}, \qquad r \in \{\mathrm{F1}, \mathrm{TPR}\}, \tag{55}$$

where the bar denotes the mean over repeated runs and q indexes the compared methods. Because lower SHD is better, its desirability score is reversed:

$$s_{\mathrm{SHD}}^{(D,m)} = \frac{\max_q \overline{\mathrm{SHD}}^{(D,q)} - \overline{\mathrm{SHD}}^{(D,m)}}{\max_q \overline{\mathrm{SHD}}^{(D,q)} - \min_q \overline{\mathrm{SHD}}^{(D,q)}}. \tag{56}$$

The overall score is the equally weighted average

$$\mathrm{Overall}^{(D,m)} = \frac{s_{\mathrm{F1}}^{(D,m)} + s_{\mathrm{TPR}}^{(D,m)} + s_{\mathrm{SHD}}^{(D,m)}}{3}. \tag{57}$$

The score lies between zero and one and is used only to summarize the relative within-dataset performance of the compared methods.

F.4 Implementation Details

All methods are evaluated using the same input series and metric definitions. For methods requiring lag or regularization specifications, the default or commonly adopted settings are used unless otherwise stated.

The proposed framework uses the same default configuration across the main experiments unless otherwise specified. The coherence threshold θ_c determines the reliable time-frequency regions used for phase aggregation, while τ_c controls whether an aggregated relation is retained as a candidate edge. The conditional-association threshold ϵ_Ω removes weak multivariate associations, and λ regularizes the scale-specific correlation matrices. The band partition and the pair (θ_c, τ_c) are varied only in the robustness analysis.

For the overall real-world score, normalization is carried out separately within each dataset.

F.5 Detailed Structural Results

For the real-world comparison in Fig. 3, the original unnormalized metrics are retained here for reference. On CausalRivers, CausalSR achieves an F1 of 0.69 ± 0.05, a TPR of 0.64 ± 0.09, and an SHD of 5.4 ± 1.6. On fMRI, it achieves an F1 of 0.43 ± 0.07, a TPR of 0.71 ± 0.05, and an SHD of 51.2 ± 2.3.

On the CauseMe benchmark in Fig. 4, CausalSR retains 66% of its initial F1 and reaches an F1 of 0.43 at $d = 20$, while its nSHD increases to 0.35. At the same graph size, DirectLiNGAM obtains an F1 of 0.40 and an nSHD of 0.38, VAR obtains 0.33 and 0.42, and PCMCI obtains 0.18 and 0.55, respectively.

F.6 Reconstruction Evaluation

We further examine whether the learned scale-specific structures retain explanatory consistency with the generated observations. Table 2 reports MSE and MAE between the original standardized signals and their graph-induced reconstructions.

Table 2: Reconstruction error on the synthetic benchmark.

Method	**MSE** $\downarrow$	**MAE** $\downarrow$
oCSE	0.88 ± 0.05	0.76 ± 0.02
PCMCI	1.19 ± 0.16	0.86 ± 0.06
DYNOTEARS	$\mathbf{0.02 \pm 0.01}$	$\mathbf{0.08 \pm 0.01}$
Granger-pw	0.05 ± 0.01	0.18 ± 0.04
VAR-LiNGAM	0.45 ± 0.02	0.53 ± 0.01
CausalSR	$\underline{0.03 \pm 0.02}$	$\underline{0.12 \pm 0.04}$

DYNOTEARS obtains the lowest reconstruction error, whereas CausalSR achieves the second-lowest MSE and MAE. Together with the structural results in Fig. 2, this indicates that the improvement in graph recovery is achieved without substantially sacrificing reconstruction fidelity. Reconstruction error is interpreted as a consistency criterion rather than direct evidence of causal correctness.

F.7 Ablation Study

We conduct an ablation study to examine the contributions of scale partitioning and joint refinement. We also include band-wise DYNOTEARS and band-wise PCMCI to determine whether applying an existing method independently to each band is sufficient.

For a variant v, its relative F1 reduction from the full framework is

$$D_{\mathrm{F1}}^{(v)} = 100 \frac{\mathrm{F1}_{\mathrm{full}} - \mathrm{F1}_v}{\mathrm{F1}_{\mathrm{full}}}, \tag{58}$$

and its relative nSHD increase is

$$D_{\mathrm{nSHD}}^{(v)} = 100 \frac{\mathrm{nSHD}_v - \mathrm{nSHD}_{\mathrm{full}}}{\mathrm{nSHD}_{\mathrm{full}}}. \tag{59}$$

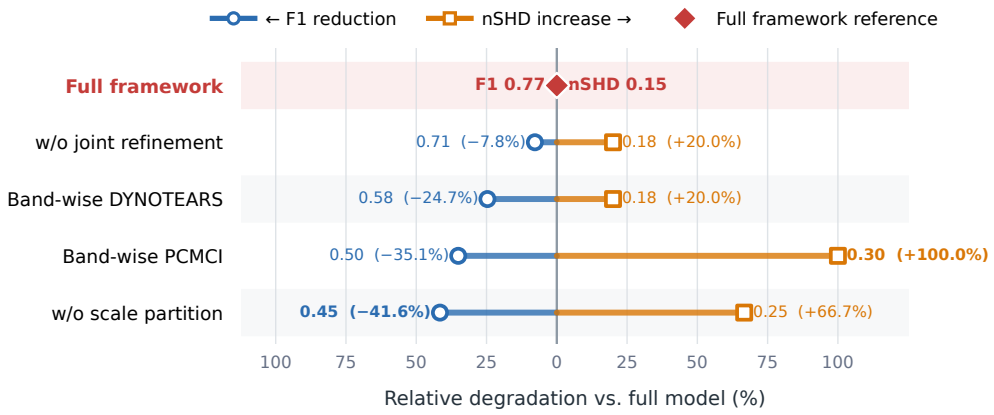


Figure 6: Relative degradation of the ablation variants compared with the full framework. The complete model is placed at the center as the reference. Values extending to the left represent the relative F1 reduction, while values extending to the right represent the relative nSHD increase. Labels beside the points report both the original metric value and the relative percentage change.

Fig. 6 shows that every simplified variant performs worse than the complete framework. Removing joint refinement reduces F1 from 0.77 to 0.71 and increases nSHD from 0.15 to 0.18, corresponding to a 7.8% F1 reduction and a 20.0% nSHD increase. Band-wise DYNOTEARS retains a relatively low nSHD but loses 24.7% of the F1 score, showing that independent band decomposition alone does not provide the same edge-recovery quality.

Band-wise PCMCI doubles nSHD and reduces F1 by 35.1%. Removing scale partitioning produces the largest F1 reduction, from 0.77 to 0.45, and increases nSHD by 66.7%. These comparisons confirm that scale separation is the main source of improvement, while conditional screening and joint refinement further stabilize the resulting graphs.

F.8 Robustness Analysis

We analyze robustness with respect to the frequency-band partition and the two principal selection thresholds. Rather than examining each metric independently, Fig. 7 jointly represents F1, TPR, and nSHD.

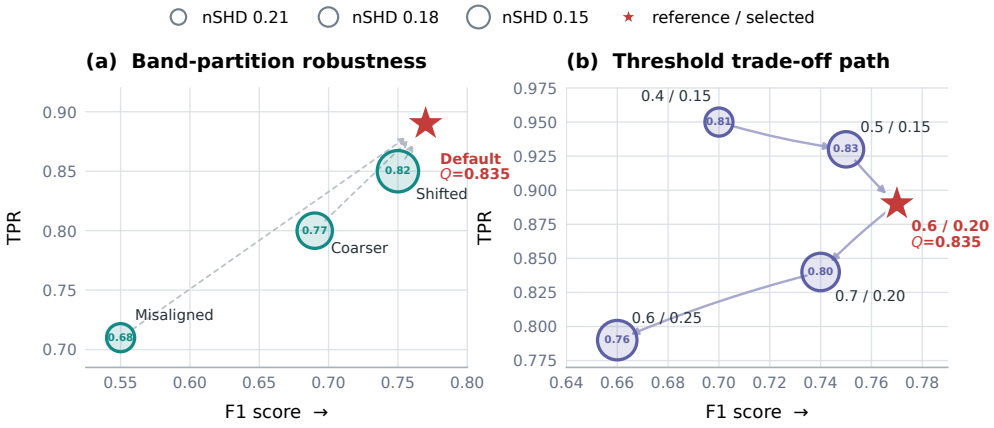


Figure 7: Robustness to alternative band partitions and threshold settings. F1 and TPR define the horizontal and vertical axes, and larger bubbles indicate lower nSHD. (a) Arrows point from alternative partitions to the default setting. (b) Performance across (θ_c, τ_c) combinations. Stars mark the default or selected setting, and Q denotes the quality score in (54).

In Fig. 7(a), the default partition occupies the most favorable upper-right region and obtains $Q = 0.835$. The shifted partition remains relatively close to the default configuration, showing that moderate perturbations of the band boundaries do not substantially change the recovered structure. Using coarser bands produces a larger reduction in both F1 and TPR, while the misaligned partition causes the strongest deterioration. The directional arrows make clear that performance improves consistently as the tested partitions approach the default scale organization.

Fig. 7(b) displays the performance trajectory obtained by varying the coherence threshold θ_c and detection threshold τ_c. Some threshold combinations produce high TPR but lower F1, indicating that they retain more true edges together with additional false positives. Other combinations increase sparsity but lose true relations. The selected pair $(\theta_c, \tau_c) = (0.6, 0.20)$ provides the most balanced trade-off among F1, TPR, and nSHD and achieves $Q = 0.835$. The trajectory also shows that the framework remains within a relatively stable performance region for several neighboring threshold combinations.