Beyond Stationarity in Time Series: Discovering Causal Structures and Latent Regimes via Markov Blankets

Lei Zan[1], Charles K. Assaad[2], Emilie Devijver[1], and Eric Gaussier[1]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France
xdzanlei@gmail.com, {emilie.devijver, eric.gaussier}@univ-grenoble-alpes.fr
[2] Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F75012, Paris, France charles.assaad@inserm.fr

Abstract. This paper introduces Regime-aware Constraint-Based and Noise-Based causal discovery with Markov Blankets (RCBNB-MB), a novel causal discovery algorithm for time series that relaxes the common assumption of a single, time-consistent causal structure. Time series are typically observed at discrete time points and often exhibit regime changes that challenge the assumption of a static causal structure, a limitation in many real-world dynamic systems. To address this challenge, RCBNB-MB identifies latent causal regimes, defined as subsets of time points within which a stable causal structure holds. The algorithm follows an iterative strategy that segments the time series into regimes and discovers the causal graph within each regime. By leveraging the Markov blanket rather than direct parents, RCBNB-MB gains robustness to errors in causal discovery and preserves predictive information. We provide theoretical guarantees for RCBNB-MB's ability to recover both regime transitions and causal graphs under reasonable assumptions. Furthermore, we validate its effectiveness through extensive experiments on simulated datasets with known ground truth and real-world IT monitoring data, where taking into account regime shifts is critical. Empirical results show that RCBNB-MB systematically outperforms baseline approaches in accurately detecting regime changes and their associated causal graphs, positioning it as a robust and versatile framework for non-stationary time series analysis.

Keywords: causal discovery · regime shift · non-stationarity.

1 Introduction

Causal discovery for time series is a fundamental problem in many domains, such as IT monitoring systems, climate science, economics, epidemiology, and neuroscience. Understanding causal relationships over time enables better prediction, diagnosis, and intervention in complex dynamic systems. However, most causal discovery methods assume stationarity, i.e., that causal dependencies remain constant. In practice, this assumption is often unrealistic: many systems undergo regime shifts, where causal influences change due to unobserved factors.

To address this, we propose a method that automatically segments time series into distinct regimes, within which stationarity holds. By jointly learning the segmentation and the causal structure in each regime, our approach provides a more flexible and robust framework for causal discovery in heterogeneous time series. A key challenge in this setting is that regime assignment must be inferred from data. Standard methods typically rely on the causal parent set (PA) to group time points into regimes, but errors in causal graph estimation can lead to incorrect assignments. To mitigate this, we instead use the Markov blanket [21] (MB), a set comprising a variable's parents, children, and spouses that renders the variable independent of all others when conditioned upon. As a theoretically optimal feature set for prediction and classification [33], the MB provides a more stable and informative representation for regime assignment. This is particularly relevant in applications where regime changes are implicit and unknown. For instance, in IT monitoring data [4], system behavior differs between normal and anomalous states, but the transition times are often unknown. Similar shifts can arise from hidden external factors in neuroscience, epidemiology, and climate science. Our method learns these regimes directly from data while recovering their causal structures.

Our main contributions are:

– Identifying a key challenge in heterogeneous time series: when instantaneous effects exist (e.g., due to low sampling rates), standard causal discovery methods struggle to correctly assign timestamps to the right regime.
– Reframing the regime assignment problem as a prediction task, leveraging the Markov blanket to improve robustness against errors in graph estimation.
– Introducing RCBNB-MB, a novel method that simultaneously identifies regimes and reconstructs regime-specific causal graphs, including instantaneous connections.
– Demonstrating the effectiveness of our approach in uncovering regime-dependent causal structures through an empirical validation on synthetic and real-world IT monitoring data.

The paper is organized as follows. Section 2 reviews existing causal discovery methods for heterogeneous time series and their limitations. Section 3 introduces our causal model. Section 4 presents our proposed method, RCBNB-MB. Section 5 provides experimental results on synthetic and real-world data. Finally, Section 6 concludes the paper.

2 Related work

Causal discovery for time series has been extensively reviewed, with a focus on consistency over time [2,10]. In this paper, we focus on cases where consistency

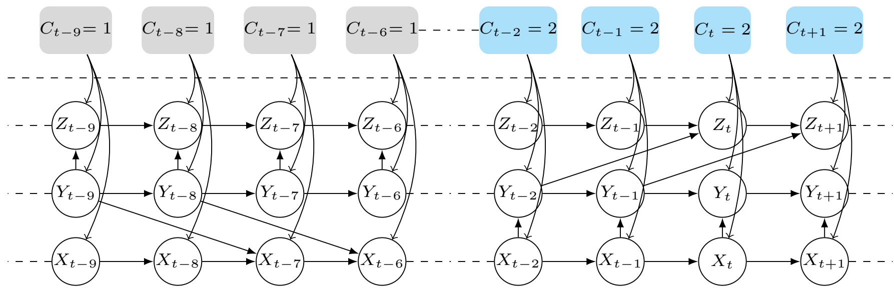
Table 1: Comparison of causal discovery methods for non-stationary time series. PA and MB denote the parent set and Markov blanket, respectively.

Method	**Regimes**	**Temporal Lags**	**Contemp. Effects**	**Fixed Causal Order**	**Prediction**
CD-NOD [14]	Segmentation	✓	✓	×	×
J(oint)-PCMCI$^+$ [11]	Input	✓	✓	×	×
LoSST [16]	Mahalanobis distance	×	×	×	×
SPACETIME [18]	Learnt	✓	✓	✓	PA
SDCI [27]	Learnt	✓	×	×	PA
RPCMCI [32]	Learnt	✓	×	×	PA
CASTOR [25]	Learnt	✓	✓	×	PA
RCBNB-MB	Learnt	✓	✓	×	MB

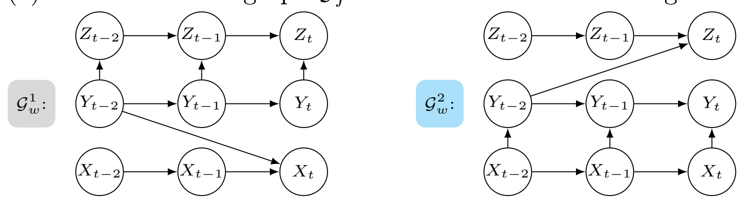
is violated, i.e., when causal relationships evolve over time or across regimes. We summarize the main characteristics in Table 1.

Classically, methods for causal discovery in non-stationary time series assume the availability of predefined regime or context information. [13] introduced a causal discovery method for heterogeneous time series using nonlinear state-space models with linear causal modeling. [14] developed the CD-NOD framework, which models changes in causal mechanisms across regimes as being confounded by an unobserved pseudo-confounder. This confounding is addressed using a surrogate variable that captures these changes and reveals additional causal directions through the Independent Causal Mechanisms principle. However, CD-NOD does not preserve temporal information, providing only a summarized causal graph. CD-NOTS [31] assumes causal stationarity in the conditional distribution, resulting in a common causal graph for the entire time series. Extensions by [7] recover temporal relationships, while [34] reconstructs the causal structure using copula entropy, assuming the surrogate variable influences other variables at specific lags—a strong assumption that may not hold in many real-world scenarios. Similarly, J(oint)-PCMCI$^+$ [11] incorporates both observed and unobserved context variables, akin to surrogate variables, and preserves temporal relationships across regimes. JIT-LiNGAM [8] learns local, linear approximations for nonlinear, heterogeneous time series but does not account for temporal lags.

More useful for real-world datasets is the ability to infer regimes and recover causal structures without assuming predefined context variables. Some methods operate directly on the time series. [16] introduced a real-time method for updating causal graphs by identifying change points using the Mahalanobis distance. However, this approach assumes i.i.d. data, which is rarely true for time series. [18] introduced a method to detect regimes, contexts, and causal structures in multiple multivariate time series, but it assumes a fixed causal order and skeleton across contexts—a strong assumption that may not hold if causal relationships change direction between regimes. [9] addressed a special case of heterogeneity in time series, assuming that different causal mechanisms occur sequentially and periodically over time. This represents a strong assumption that may not hold in



(a) Full time causal graph $\mathcal{G}_f$ from two distinct latent regimes.



(b) Window causal graphs $\mathcal{G}_w^1$ (left) and $\mathcal{G}_w^2$ (right) from two distinct regimes.

Fig. 1: Two graphic causal relation representations for time series with two distinct regimes: (a) full time causal graphs from two distinct regimes (b) window causal graphs from two distinct regimes. The variable C_t is latent and has to be inferred.

many real-world scenarios. Another viewpoint is to construct regimes to improve forecasting. They are all (to our knowledge) based on forecasting with causal parents. [27,6] introduced a strategy to identify both causal relationships and states in state-dependent stationary time series. However, they do not account for instantaneous connections between observed time series. Similarly, [32] proposed RPCMCI, which detects regimes in heterogeneous time series and reconstructs causal graphs for each regime. However, this method focuses solely on time-lagged relationships, neglecting contemporaneous causal effects. [25] proposed CASTOR, which learns a DAG for each regime while determining the number of regimes and their sequential arrangement, based on an EM algorithm. If all those methods discover a causal graph per regime under different assumptions, they are discriminating regimes by using a forecasting model within the time series based on the causal parents. Our method, on the contrary, relies on the Markov blanket.

3 Causal modeling for heterogeneous time series

Consider a dynamic system represented by a multivariate time series of dimension d, observed over $\mathbf{T}$, and denoted as $\{\mathbf{X}_t\}_{t\in\mathbf{T}} = \{X_t^j\}_{t\in\mathbf{T},j\in\{1,\ldots,d\}}$. These data exhibit complex causal dependencies, naturally modeled using a

Directed Acyclic Graph (DAG), called the full-time causal graph, denoted as $\mathcal{G}_f = (\mathbf{V}_f, \mathbf{E}_f)$. Here, $\mathbf{V}_f$ represents vertices corresponding to time-indexed variables, and $\mathbf{E}_f$ captures directed causal relationships. A key aspect of our framework is the introduction of hidden context variables C_t, where $C_t \in \{1, \ldots, r\}$ indicates the regime to which each time point t belongs. The number of distinct regimes is denoted as r. We do not assume regimes are known a priori. Instead, C_t acts as an unobserved confounder, directly influencing all observed variables at time t and defining the regime-specific causal structure.

Assumption 1 (Semi-pseudo causal sufficiency)**.** *All hidden confounding between observed variables is captured by* $(C_t)_{t\in\mathbb{T}}$.

To infer the causal graph, we assume **consistency throughout time** [2] (also known as **causal stationarity** [28]) within each regime: causal relationships remain invariant over time, allowing us to sidestep the intractable full causal graph $\mathcal{G}_f$ and instead operate on window causal graphs $\mathcal{G}_w^k = (\mathbf{V}_w, \mathbf{E}_w^k)$ for $k \in \{1, \ldots, r\}$. While this assumption breaks down at regime boundaries (where structural shifts occur) it provides a principled way to balance model complexity and interpretability. These graphs capture causal dependencies within a finite window of length $\tau_{\max}$ which corresponds to the maximum lag between direct causes and effects. This leads to the following model.

Definition 1 (Regime-based multivariate time series). *A multivariate time series,* $\{\mathbf{X}_t\}_{t\in\mathbf{T}}$*, is called a* <u>*regime-based multivariate time series*</u> *if it can be partitioned into* $r > 1$ *disjoint regimes encoded by* $C_t \in \{1, \ldots, r\}$ *for* $t \in \mathbf{T}$*, with a specific window causal graph* $\mathcal{G}_w^k$ *in regime* k *s.t.* $\forall k_1, k_2 \in \{1, \ldots, r\}$, $k_1 \neq k_2 \Rightarrow \mathcal{G}_w^{k_1} \neq \mathcal{G}_w^{k_2}$.

The structural causal model (SCM) [20] for the j^{th} variable in a heterogeneous multivariate time series is given by:

$$X_t^j = \sum_{k=1}^{r} \mathbb{1}_{C_t=k} \times g^{j,C_t}(\mathbf{Pa}(X_t^j; \mathcal{G}_w^{C_t}), \epsilon_t^j) \text{ for } t \in \mathbf{T}. \quad (1)$$

Here, $\mathbb{1}_{C_t=k}$ is an indicator function that equals 1 if and only if $C_t = k$. The response function $g^{j,C_t}(\cdot)$ is deterministic within each regime and depends only on C_t. $\mathbf{Pa}(X_t^j; \mathcal{G}_w^{C_t})$ represents the causal parents of X_t^j in $\mathcal{G}_w^{C_t}$, including lagged and instantaneous variables. The noise terms $(\epsilon_t^j)_{t\in\mathbf{T},1\leq j\leq d}$ are assumed to be independent. Note that we allow for instantaneous causal relations in the window causal graph of each regime, which is often forgotten in the literature (except in methods such as CASTOR [25]). A regime does not necessarily correspond to a single contiguous time segment and can span multiple non-adjacent intervals. Figure 1(a) provides an example of a full-time causal graph for a multivariate time series with two distinct regimes, while Figure 1(b) depicts the corresponding window causal graphs $\mathcal{G}_w^1$ and $\mathcal{G}_w^2$.

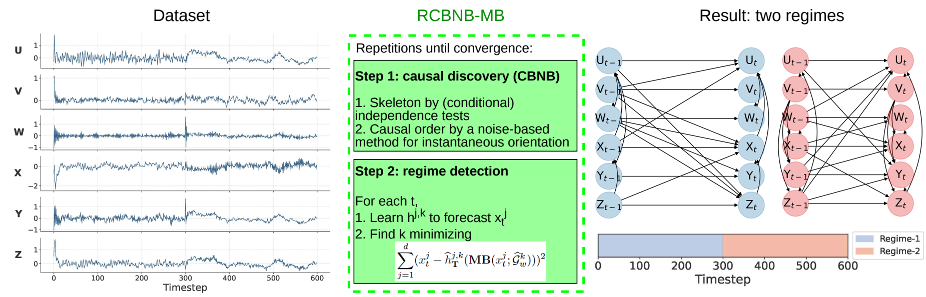


Fig. 2: **Illustration of our method.** Left: an example of a dataset with 6 time series, which violates the causal stationarity assumption. Middle: flowchart of our iterative method through the two steps: causal discovery and regime detection. Right: results of our method on this dataset: two regimes are considered, and we display the two causal graphs and the regime assignment.

4 Causal discovery from heterogeneous time series

The goal of causal discovery from heterogeneous, observational multivariate time series is twofold:

1. Reconstruct the causal graph $(\mathcal{G}_w^k)$ within each regime k,
2. Assign each timestamp to the correct regime.

We propose to address those two problems by an alternating approach, denoted as RCBNB-MB (Regime-aware Constraint-Based and Noise-Based causal discovery with Markov Blankets), which alternates between two steps by first constructing the window causal graph of each regime from given assignments of time instants to regimes, and second by re-assigning time instants to regimes according to the new window causal graphs. We detail these two steps below, and the full procedure is illustrated in Figure 2. In the following, we assume that the number of distinct regimes r is known.

4.1 Regime-Aware Window Causal Graph Discovery

RCBNB-MB begins by, given the regime index, discovering a causal graph, which reduces to time-series causal discovery within each regime. While any causal discovery algorithm could be integrated, we adopt CBNB [5], a hybrid method that prunes edges with a constraint-based approach and orients them with a noise-based one, thereby retaining the strengths of both families: the graph is fully oriented and the orientation search is restricted to the discovered skeleton, which improves its efficiency. More specifically, given the regime index for each time point, CBNB recovers the window causal graph $\mathcal{G}_w$ in two steps:

1. a constraint-based step, which infers the skeleton of the window causal graph using (conditional) independence tests and orients lagged edges using temporal ordering,

2. a noise-based step that orients the remaining instantaneous edges: within each group of variables linked by undirected edges, a restricted noise-based algorithm estimates a causal order among the instantaneous variables, including the lagged variables as covariates in the regressions to control for confounding.

CBNB's advantages are threefold: it requires only adjacency faithfulness [26] (not full faithfulness), recovers the true causal graph (not just its Markov class), and outperforms pure constraint/noise-based methods in small-sample settings [17,4]. CBNB's assumptions are formalized below.

Assumption 2 (Adjacency Faithfulness)**.** *Let $\mathcal{G}_w^k = (\boldsymbol{V}_w, \boldsymbol{E}_w^k)$ be the causal graph within any regime k. If two nodes X and Y in $\boldsymbol{V}_w$ are adjacent in $\mathcal{G}_w^k$, then they are dependent conditionally on any subset of $\boldsymbol{V}_w \backslash \{X, Y\}$.*

Assumption 3 (Identifiable functional model)**.** *Given a heterogeneous multivariate time series satisfying the SCM in Equation (1), each function $g^{j,k}(\cdot)$ in regime $k \in \{1, \dots, r\}$ for a component $j \in \{1, \dots, d\}$ belongs to an identifiable functional model class, as defined in [23].*

By [5, Theorem 2], CBNB correctly recovers the causal graph $\mathcal{G}_w^k$ for each regime k if regime indexes are correct, Assumptions 1, 2, Assumption 3, and consistency throughout time hold, and conditional independence tests are perfect. While CBNB assumes consecutive time points, we extend its use to gapped regimes (Section 5.2).

4.2 Assigning Time Points to Regimes Using Markov-Blanket-based Prediction

Assuming the window causal graph for each regime is known, we assign each time point to the regime that minimizes the prediction error for all variables in $\mathbf{X}_t = \{\mathbf{X}_{t-\tau_{\max}}, \dots, \mathbf{X}_{t+\tau_{\max}}\}$. Predicting the value of a variable X_t^j given the other variables, $\mathbf{X}_{-t}^{-j} = \mathbf{X}_t \setminus X_t^j$, traditionally involves minimizing the expected squared error, and the set of potential covariates can be restricted to the Markov blanket of X_t^j in the causal graph $\mathcal{G}_w^k$, where $k \in \{1, \dots, r\}$, as established by [33]: for a class of functions $\mathcal{H}$,

$$\underset{h^{j,C_t} \in \mathcal{H}}{\operatorname{argmin}} \, \mathbb{E}\left[(X_t^j - h^{j,C_t}(\mathbf{X}_{-t}^{-j}))^2\right] = \underset{h^{j,C_t} \in \mathcal{H}}{\operatorname{argmin}} \, \mathbb{E}\left[(X_t^j - h^{j,C_t}(\mathbf{MB}(X_t^j; \mathcal{G}_w^{C_t})))^2\right]. \tag{2}$$

Since the Markov blanket is identical for all instances of X^j within a given regime, we leverage this structure to approximate the optimal predictor $h_*^{j,k}$, where $1 \leq k \leq r$, using empirical risk minimization:

$$\widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{MB}(x_t^j; \mathcal{G}_w^k)) = \underset{h^{j,C_t} \in \mathcal{H}}{\operatorname{argmin}} \sum_{t \in \mathbf{T}} \mathbb{1}_{C_t = k} \left(x_t^j - h^{j,C_t}(\mathbf{MB}(x_t^j; \mathcal{G}_w^{C_t}))\right)^2. \tag{3}$$

Using consistency results similar to the ones developed in [12] then leads to the following theorem, the proof of which is given in Appendix A, which provides a theoretical criterion for deciding the regime for a given variable at a particular time instant.

Theorem 1. *Suppose that $C_t = k$, that the function $h_*^{j,k}(\mathbf{MB}(x_t^j; \mathcal{G}_w^k))$ approximates x_t^j within a bounded error, i.e., $\exists M > 0, \forall x_t^j, |x_t^j - h_*^{j,k}(\mathbf{MB}(x_t^j; \mathcal{G}_w^k))| < M$, and that the estimate $\widehat{h}_{\mathbf{T}}^{j,k}$ converges to the "best" function $h_*^{j,k}$ in the following sense:*

$$\lim_{k(\mathbf{T})\to+\infty} \int_{\mathbf{S}} \left(\widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{S}) - h_*^{j,k}(\mathbf{S})\right)^2 dP(\mathbf{S}) = 0, \tag{4}$$

where $\mathbf{S}$ is any subset of $\mathbf{X}$. Then, for any other regime $\ell \neq k$:

$$\lim_{k(\mathbf{T})\to+\infty} \mathbb{E}\left[(X_t^j - \widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{MB}(X_t^j; \mathcal{G}_w^k)))^2\right] \leq \lim_{k(\mathbf{T})\to+\infty} \mathbb{E}\left[(X_t^j - \widehat{h}_{\mathbf{T}}^{j,\ell}(\mathbf{MB}(X_t^j; \mathcal{G}_w^\ell)))^2\right].$$

Note that Eq. 4 typically holds for universal approximators under conditions on the class of functions considered [12].

Under the assumption of sufficient sample size in each regime and bounded prediction error, Theorem 1 implies that time t should be assigned to regime k if, for all $\ell \neq k$:

$$\mathbb{E}\left[(X_t^j - \widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{MB}(X_t^j; \mathcal{G}_w^k)))^2\right] < \mathbb{E}\left[(X_t^j - \widehat{h}_{\mathbf{T}}^{j,\ell}(\mathbf{MB}(X_t^j; \mathcal{G}_w^\ell)))^2\right]. \tag{5}$$

Equality in Equation 5 can arise in three cases: (i) the Markov blankets are identical, (ii) one Markov blanket is a strict superset of the other, or (iii) alternative variable sets provide equivalent prediction accuracy. The following assumption[3] helps resolve such ambiguities:

Assumption 4. *Given a multivariate heterogeneous time series $\{\mathbf{X}_t\}_{t\in\mathbf{T}}$:*

- **(i)** *For any two distinct regimes $k_1, k_2 \in \{1, \ldots, r\}$, there exists a variable X_t^j such that $\mathbf{MB}(X_t^j; \mathcal{G}_w^{k_1}) \neq \mathbf{MB}(X_t^j; \mathcal{G}_w^{k_2})$.*
- **(ii)** *For any regime k and variable X_t^j, if a set of variables $\mathbf{S}$ is not a superset of the true Markov blanket, then the expectation in Theorem 1 using $\mathbf{S}$ differs from the one using $\mathbf{MB}(X_t^j; \mathcal{G}_w^k)$.*

Under Assumption 4, if multiple regimes yield the same prediction error, we break ties by selecting the regime with the smallest Markov blanket containing a variable satisfying (i). This leads to the final regime assignment criterion:
Criterion t2r (time to regime): *Assign C_t to the regime that minimizes:*

$$\sum_{j=1}^{d} (x_t^j - \widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{MB}(x_t^j; \widehat{\mathcal{G}}_w^k)))^2. \tag{6}$$

[3] This assumption is reasonable as it simply states that different regimes likely lead to different Markov blankets for at least some variables, and that if a set differs from the true Markov blanket without being a superset of it, it will likely yield a prediction different from the one of the true Markov blanket.

If multiple regimes minimize this criterion, select the one with the smallest Markov blanket over all variables satisfying Assumption 4(i).

The procedure we propose can be summarized as follows: for each regime k and each dimension j, a predictor is learned using the Markov blanket of X_t^j. Then, a sample $\{x_t^j\}_{j\in\{1,\dots,d\}}$ is assigned to the regime whose predictors yield the lowest overall prediction error. In the case of ties, the regime with the smallest Markov blanket is selected.

4.3 RCBNB-MB: an algorithm for causal discovery from multiple regimes

Finally, the overall problem can be framed as the following optimization task: determine the assignment vector $\mathbf{C} = [C_{t_1}, \dots, C_{t_2}]^T$, where each coordinate belongs to $\{1, \dots, r\}$, the regime-specific window causal graphs $\{\mathcal{G}_w^k\}_{1\leq k\leq r}$ and the functions $\mathbf{H} = \{\widehat{h}_{\mathbf{T}}^{j,k}\}$ for $1 \leq j \leq d$ and $1 \leq k \leq r$ that minimize:

$$L_{emp}(\mathbf{x}_t; \mathbf{C}, \mathbf{H}, \{\mathcal{G}_w^k\}_{1\leq k\leq r}) = \sum_{j=1}^{d} \sum_{t\in\mathbf{T}} \left(x_t^j - \sum_{k=1}^{r} \mathbb{1}_{C_t=k} \widehat{h}_{\mathbf{T}}^{j,k}(\mathbf{MB}(x_t^j; \mathcal{G}_w^k)) \right)^2 . \tag{7}$$

To ensure meaningful regime assignments, we furthermore impose the following constraints:

1. Each time point belongs to exactly one regime,
2. Each regime contains at least N_ℓ observations over the entire time span, with $N_\ell \geq (2\tau_{\max} + 1)d$, so as to be able to infer it using Markov blankets, and at most N_c transitions to prevent overly fragmented assignments.

To solve the optimization problem in Equation (7) under those constraints, RCBNB-MB proceeds as follows: first, it assigns time points randomly to regimes with $\mathbf{C}^{[0]}$, ensuring that each regime spans at least N_ℓ timestamps. Then, it alternates between the two steps:

Step 1 Given the current assignment $\mathbf{C}^{[\text{ite}]}$, estimate the causal graph $\mathcal{G}_w^k$ for each regime k using CBNB [5]. Then, for each variable X_t^j, determine its Markov blanket $\mathbf{MB}(X_t^j; \mathcal{G}_w^k)$ and update the function set $\mathbf{H}^{[ite]}$ by minimizing Equation (7).

Step 2 Keeping $\mathbf{H}^{[\text{ite}]}$ fixed, update the assignment vector $\mathbf{C}^{[\text{ite}+1]}$ using Criterion $t2r$, subject to constraints. This is a standard optimization problem with linear constraints that can be efficiently solved using existing solvers.

To avoid poor convergence, RCBNB-MB runs N_i random initializations and selects the solution with the lowest empirical risk after N_o iterations.

Assignments near regime transitions may be less accurate due to limited past data, but their impact is negligible when regime changes are sparse.

If regime assignments are correct, Step 1 recovers the true window causal graphs (and Markov blankets) as $k(\mathbb{T}) \to +\infty$ (Theorem 1, Assumption 4(ii)).

Step 2 reduces the squared error in L_{emp} if observed values are close to their expectations.

While there is no theoretical guarantee that RCBNB-MB consistently converges to the ground truth, empirical results in Section 5 show that its outputs are often close to the ground truth in the majority of cases. The pseudocode of the algorithm is presented in Appendix B, with the corresponding code available in the supplementary materials.

5 Experiments

We designed the following experiments to evaluate our method's accuracy in regime assignment and causal graph reconstruction.

Baselines Our framework, denoted RCBNB-MB, combines CBNB for causal discovery with the Markov blanket (MB) for regime detection. To assess the impact of these choices, we benchmark against alternative causal discovery methods, including VarLiNGAM [15], Dynotears [19], PCMCI [30], and PCMCI$^+$ [29], each paired with either the Markov blanket (MB) or direct parents (PA) as the predictive variable set, following the naming convention R{method}-{variable set}. Notably, RPCMCI-PA corresponds to the method in [32], while RPCMCI-MB is its MB-based variant. We also include CASTOR [25], the random baseline NegControl [24], CD-NOD [14], and J-PCMCI$^+$ [11] for comparison, while SPACETIME [18] is excluded due to its high computational cost. Since NegControl, CD-NOD, and J-PCMCI$^+$ cannot detect regimes within the time series, their MER values are not reported. All methods were implemented based on publicly available Python libraries, as described in Appendix E.

Evaluation metrics We evaluate two aspects of performance: (1) the accuracy of regime assignment and (2) the quality of causal graph reconstruction.

Regime assignment accuracy is measured using the Mean Error Rate (MER), which quantifies the proportion of incorrectly assigned timestamps, up to regime label mismatch. Formally, given the one-hot encoded assignment matrices $\mathbf{C}_{bina}$ (predicted) and $\mathbf{C}^*_{bina}$ (ground truth), MER is defined as $\text{MER} = \min_\pi |\pi(\mathbf{C}_{bina}) - \mathbf{C}^*_{bina}|_{r\times|\mathbf{T}|}/(2|\mathbf{T}|)$, where $|\cdot|_{r\times|\mathbf{T}|}$ denotes the element-wise L_1 norm and π is a permutation over rows.

Causal graph reconstruction is evaluated using two complementary metrics: the oriented F1-score, which compares the estimated window causal graphs with the ground-truth graphs within each regime (higher is better), and the normalized Structural Hamming Distance (lower is better), defined as $nSHD = \frac{1}{r}\sum_{k=1}^{r} SHD(\widehat{\mathcal{G}}^k_w, \mathcal{G}^k_w)/|\boldsymbol{E}^k_w|$, where $SHD(\cdot,\cdot)$ counts missing, spurious, and reversed instantaneous edges, and $|\boldsymbol{E}^k_w|$ is the number of edges in the ground-truth graph $\mathcal{G}^k_w$, nSHD can thus exceed 1. We compare with a random graph to get a negative control, as suggested in [24].

5.1 Simulated data

Simulation setup We consider linear causal relationships among six variables over a time series of length 600 (results for length 1,200 are provided in Appendix C.2, and results for a scenario with 15 variables in Appendix C.5), with 2 and 3 regimes. Each regime consists of consecutive time points, evenly dividing the time series (results for unequal regime sizes are provided in Appendix C.4). Time points at regime boundaries are independent, with the first point of each regime sampled from noise. For each time series length, we generate 50 datasets. We set $\tau_{\max} = 1$, allowing for contemporaneous dependencies. For each regime, the window causal graph is generated from the Erdos-Renyi [22] model: each variable has a self-causal link at lag 1, and all other edges appear with probability 0.3. To enforce regime differences, we constrain the number of shared edges between the causal graphs of two regimes to at most 7 and guarantee that at least one variable has a different Markov blanket across regimes.

Within each regime, the data follow the structural causal model below:

$$X_t^j = \sum_{X_{t-\tau}^i \in \mathbf{Pa}(X_t^j)} b_{i,j,\tau} X_{t-\tau}^i + \epsilon_t^j, \tag{8}$$

where the coefficients are sampled as $b_{i,j,\tau} \sim \mathbf{U}((-0.9, -0.5) \cup (0.5, 0.9))$ and the noise term follows $\epsilon_t^j \sim \mathbf{U}(-0.1, 0.1)$. To prevent extreme values, if any observation within a regime exceeds 100, the dataset is regenerated.

Method configuration The regime assignment is initialized by randomly assigning each timestep to a regime with a probability of 0.95, initially creating overlaps. However, our constraints ensure that the final assignment is unique for each timestep. We set the number of initialization points to $N_i = 50$, the maximum optimization iterations to $N_o = 20$, the number of regimes to $r = 2$, and the minimum regime size to $N_\ell = 18$ $((2\tau_{\max}+1)d)$, ensuring estimator identifiability (see Appendix C.1 for a parameter robustness analysis of RCBNB-MB). Each regime allows up to $N_c = 3$ transitions. The function class $\mathbf{H}$ is chosen as linear. $\tau_{\min}$ is set to 0 and $\tau_{\max}$ to 1. For CBNB, PCMCI$^+$ is used in the constraint-based part and VarLiNGAM in the noise-based part. The significance threshold α is set to 0.1 for all (conditional) independence tests. VarLiNGAM and Dynotears are run with their default parameters. For CASTOR, CD-NOD, and J-PCMCI$^+$, the default parameters for the linear setting are adopted. For NegControl, the true number of nodes and edges for each dataset are provided as inputs.

Results The results are summarized in Table 2. First, we analyze the Mean Error Rate (MER), which remains low for most methods. For instance, our method, RCBNB-MB, achieves a MER of 1.27% in the *2 regimes* scenario, meaning that, on average, only about 8 out of 600 timestamps are misclassified. In the *3 regimes* scenario, this rate further decreases to 0.81%, demonstrating the method's reliability in correctly assigning timestamps to the appropriate regimes. In contrast, some methods, such as CASTOR, exhibit significantly higher MER values, exceeding 20%, indicating difficulties in detecting regime changes.

Table 2: **Results on generated data.** The Mean Error Rate (MER), the mean F1-score, and the mean normalized Structural Hamming Distance (nSHD) across two scenarios: *2 regimes* and *3 regimes*. The F1-score is evaluated under three conditions: *Total* (all edges considered), *Lagged* (only lagged edges considered), and *Instant* (only instantaneous edges considered). Reported values are means over 50 repetitions, with time series of length 600.

	2 regimes					*3 regimes*				
	MER	*Total*	*Lagged*	*Instant*	nSHD	MER	*Total*	*Lagged*	*Instant*	nSHD
RCBNB-MB	1.27%	**0.73**	**0.77**	**0.67**	**0.52**	0.81%	**0.66**	**0.71**	**0.58**	**0.55**
RCBNB-PA	1.27%	0.69	0.73	0.62	0.54	2.93%	0.65	0.69	0.57	0.57
RPCMCI-MB	0.24%	0.57	0.65	×	1.03	0.39%	0.54	0.62	×	1.07
RPCMCI-PA [32]	1.26%	0.57	0.66	×	1.00	0.35%	0.55	0.63	×	1.04
RPCMCI$^+$-MB	1.27%	0.64	0.70	0.46	0.60	2.95%	0.55	0.62	0.39	0.56
RPCMCI$^+$-PA	10.23%	0.55	0.61	0.36	0.66	7.79%	0.46	0.52	0.28	0.73
RVarLiNGAM-MB	0.28%	0.43	0.56	0.09	0.93	0.37%	0.42	0.55	0.08	0.93
RVarLiNGAM-PA	0.48%	0.43	0.53	0.17	0.96	1.78%	0.41	0.51	0.16	0.96
RDynotears-MB	16.91%	0.19	0.23	0.07	1.01	16.25%	0.19	0.23	0.09	1.02
RDynotears-PA	15.58%	0.18	0.23	0.07	0.99	15.03%	0.21	0.25	0.11	1.02
NegControl [24]	×	0.29	0.32	0.22	1.36	×	0.28	0.32	0.21	1.37
CD-NOD [14]	×	0.34	0.40	0.22	0.87	×	0.31	0.35	0.21	0.86
J-PCMCI$^+$ [11]	×	0.41	0.48	0.19	1.13	×	0.37	0.44	0.15	1.27
CASTOR [25]	20.07%	0.12	0.15	0.05	0.99	14.17%	0.11	0.13	0.06	1.00

Regarding the F1-score, RCBNB-MB achieves the highest overall performance, reaching 0.73 in the *2 regimes* scenario and 0.66 in the *3 regimes* scenario. This confirms its effectiveness in recovering causal structures under different conditions. The other variants of RCBNB also perform well, particularly RCBNB-PA, which maintains competitive scores. PCMCI-based methods show moderate performance, with RPCMCI-MB achieving scores around 0.64. The difference between the two PCMCI-based methods is significantly smaller than the difference between the two PCMCI$^+$-based methods. This can be attributed to the fact that, when the inferred graph does not include instantaneous relations, the estimated set of parents and the estimated Markov blanket are more similar to each other than the corresponding true sets of parents and Markov blanket in the ground-truth graph, which does contain instantaneous relations.

The nSHD results further confirm this trend. RCBNB-MB achieves the lowest nSHD in both scenarios, with values of 0.52 and 0.55, closely followed by RCBNB-PA. This indicates that both variants recover graph structures closest to the ground truth. In contrast, most competing methods obtain nSHD values close to or above 1, reflecting larger structural discrepancies.

VarLiNGAM-based methods exhibit performance comparable to J-PCMCI$^+$, while CD-NOD performs slightly better than the random baseline NegControl. In contrast, CASTOR and RDynotears-based methods fall below random baseline performance, highlighting their limitations in reconstructing causal graphs accurately.

Additionally, methods using the Markov blanket outperform in general those using only parents. This can be attributed to the fact that the Markov blanket offers a more robust set of variables for prediction, as it includes not only parents but also children and spouses. This representation provides MB-based methods with sufficient information to improve predictive accuracy and reduce sign errors, thereby enhancing the reliability of the subsequent causal discovery step.

The variances of the F1-scores are reported in Appendix C.3. They are generally small, around 0.01, except for RCBNB-PA, where they range between 0.02 and 0.04, and RPCMCI$^+$, which exhibits higher variance, reaching up to 0.09 for PA and 0.05 for MB. These results suggest that while most methods yield stable performance, some, particularly RPCMCI$^+$, show more variability across repetitions.

5.2 Real data

Data description For the real-world experiment, we use eight time series from an IT monitoring system provided by EasyVista[4], sampled at one-minute intervals, as introduced in [3]. These time series describe different system activities and are detailed in Appendix D. According to a system expert [3], an anomaly occurs between timesteps 46,683 and 46,783. To construct our dataset, we include an additional 1,000 timesteps before and after this interval, resulting in a total of 2,100 timesteps for analysis.

In [3], only a summary representation of the window causal graph for the normal regime was provided by the system expert. Therefore, to assess the performance of our method in this scenario, we first convert the obtained window causal graph into a summary graph [1] and then compute the F1-score. Furthermore, since we have access only to the ground truth causal structure of the normal regime, our evaluation is limited to the inferred causal graphs within the regimes that overlap with the true normal regime.

Method configuration Based on the method's performance in Section 5.1, RCBNB-MB stands out, so we focus on this method here. Due to the increased number of variables, we set the minimum regime size to $N_\ell = 24$ ($(2\tau_{\max} + 1)d$) with $\tau_{\max} = 1$. Additionally, we set the number of initialization points to $N_i = 200$, the number of regimes to $r = 2$ and 3, and allow up to $N_c = 4$ transitions per regime. Other parameters remain the same as those in Section 5.1.

Results In this experiment, the final value of the objective function, corresponding to Equation 7, is 3.56 when assuming two regimes and 3.02 when assuming three regimes. Figure 3 illustrates the detected change points and the corresponding regimes identified by RCBNB-MB when provided with prior knowledge of either two or three regimes. According to the system expert, the ground truth consists of only two regimes: normal and abnormal. The figure clearly demonstrates that RCBNB-MB effectively detects the regimes in both cases. Even when the number of regimes is set to three, the algorithm accurately identifies the normal and abnormal regimes, with only minor deviations. Additionally, it detects a third regime at the boundary between the two true regimes, suggesting

[4] https://www.easyvista.com/fr/produit/supervision-it/

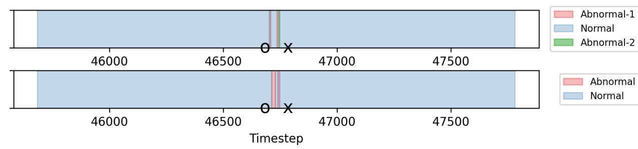


Fig. 3: RCBNB-MB assignments assuming 3 regimes (top) and 2 regimes (bottom). Blue denotes normal regime, while rose and green denote abnormal regimes. The circle (○) and cross (×) mark the expert identified anomaly start and end at timestamps 46,683 and 46,783, respectively.

Table 3: Results on real data. The F1-score for each regime is computed for our method RCBNB-MB, with 2 and 3 regimes.

	loss	F1-score in normal regimes
2 regimes	3.56	0.53
3 regimes	3.02	0.53

a possible transitional phase between them. Table 3 presents the F1-scores for detecting a summary representation of the window causal graph in the normal regime. Compared to previous studies, such as [4], where classical causal discovery methods were used to infer the graph, our method demonstrates superior performance.

6 Conclusion and Perspective

Understanding causal relationships in complex dynamic systems is important across many fields. Growing data availability creates opportunities but also challenges, particularly for heterogeneous time series with multiple regimes and changing causal mechanisms. We propose a method that identifies these regimes and recovers their causal structures by treating regime assignment as a prediction task based on Markov blankets. The use of Markov blankets in this setting is supported by both theoretical and experimental arguments. Lastly, we demonstrated the effectiveness of our approach in uncovering regime-dependent causal structures through an empirical validation on synthetic and real-world IT monitoring data.

There are several interesting aspects to explore in future work. In our experiments, we only considered linear relationships among variables and time series containing two or three distinct regimes. More complex scenarios involving non-linear relationships and more than three distinct regimes need to be tested. Additionally, timestamps at the boundaries of two distinct regimes are difficult to assign correctly due to the loss of information about the Markov blanket. Further research is needed to improve the assignment of these boundary timestamps. Lastly, we would also like to explore its potential for root cause analysis and anomaly detection.

References

1. Assaad, C.K., Devijver, E., Gaussier, E.: Discovery of extended summary graphs in time series. In: Uncertainty in Artificial Intelligence. pp. 96–106. PMLR (2022)
2. Assaad, C.K., Devijver, E., Gaussier, E.: Survey and evaluation of causal discovery methods for time series. Journal of Artificial Intelligence Research **73**, 767–819 (2022)
3. Assaad, C.K., Ez-Zejjari, I., Zan, L.: Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In: Ruiz, F., Dy, J., van de Meent, J.W. (eds.) Proceedings of The 26th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 206, pp. 8395–8404. PMLR (25–27 Apr 2023)
4. Aït-Bachir, A., Assaad, C.K., de Bignicourt, C., Devijver, E., Ferreira, S., Gaussier, E., Mohanna, H., Zan, L.: Case studies of causal discovery from it monitoring time series (2023)
5. Bystrova, D., Assaad, C.K., Arbel, J., Devijver, E., Gaussier, E., Thuiller, W.: Causal discovery from time series with hybrids of constraint-based and noise-based algorithms. Transactions on Machine Learning Research (2024)
6. Cai, R., Huang, L., Chen, W., Qiao, J., Hao, Z.: Learning dynamic causal mechanisms from non-stationary data. Applied Intelligence **53**(5), 5437–5448 (Jun 2022)
7. Ferdous, M.H., Hasan, U., Gani, M.O.: Cdans: Temporal causal discovery from autocorrelated and non-stationary time series data. In: Machine Learning for Healthcare Conference. pp. 186–207. PMLR (2023)
8. Fujiwara, D., Koyama, K., Kiritoshi, K., Okawachi, T., Izumitani, T., Shimizu, S.: Causal discovery for non-stationary non-linear time series data using just-in-time modeling. In: Conference on Causal Learning and Reasoning. pp. 880–894. PMLR (2023)
9. Gao, S., Addanki, R., Yu, T., Rossi, R., Kocaoglu, M.: Causal discovery in semi-stationary time series. Advances in Neural Information Processing Systems **36** (2024)
10. Gong, C., Zhang, C., Yao, D., Bi, J., Li, W., Xu, Y.: Causal discovery from temporal data: An overview and new perspectives. ACM Comput. Surv. **57**(4) (Dec 2024)
11. Günther, W., Ninad, U., Runge, J.: Causal discovery for time series from multiple datasets with latent contexts. In: Evans, R.J., Shpitser, I. (eds.) Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence. Proceedings of Machine Learning Research, vol. 216, pp. 766–776. PMLR (31 Jul–04 Aug 2023)
12. Györfi, L., Kohler, M., Krzyzak, A., Walk, H., et al.: A distribution-free theory of nonparametric regression, vol. 1. Springer (2002)
13. Huang, B., Zhang, K., Gong, M., Glymour, C.: Causal discovery and forecasting in nonstationary environments with state-space models. In: International conference on machine learning. pp. 2901–2910. PMLR (2019)
14. Huang, B., Zhang, K., Zhang, J., Ramsey, J.D., Sanchez-Romero, R., Glymour, C., Schölkopf, B.: Causal discovery from heterogeneous/nonstationary data. J. Mach. Learn. Res. **21**(89), 1–53 (2020)
15. Hyvärinen, A., Zhang, K., Shimizu, S., Hoyer, P.O.: Estimation of a structural vector autoregression model using non-gaussianity. JMLR **11**(5) (2010)
16. Kummerfeld, E., Danks, D.: Tracking time-varying graphical structure. Advances in neural information processing systems **26** (2013)
17. Malinsky, D., Danks, D.: Causal discovery algorithms: A practical guide. Philosophy Compass **13**(1), e12470 (2018)

18. Mameche, S., Cornanguer, L., Ninad, U., Vreeken, J.: Spacetime: Causal discovery from non-stationary time series (2025)
19. Pamfil, R., Sriwattanaworachai, N., Desai, S., Pilgerstorfer, P., Georgatzis, K., Beaumont, P., Aragam, B.: Dynotears: Structure learning from time-series data. In: International Conference on Artificial Intelligence and Statistics. pp. 1595–1605. PMLR (2020)
20. Pearl, J.: Causality. Cambridge university press (2009)
21. Pellet, J.P., Elisseeff, A.: Using markov blankets for causal structure learning. JMLR **9**(7) (2008)
22. P.Erdos, A.Renyi: On random graphs i. Publ. math. debrecen **6**(290-297), 18 (1959)
23. Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B.: Identifiability of causal graphs using functional models. In: Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. p. 589–598. UAI'11, AUAI Press, Arlington, Virginia, USA (2011)
24. Petersen, A.H.: Are you doing better than random guessing? a call for using negative controls when evaluating causal discovery algorithms. In: Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence. UAI '25 (2025)
25. Rahmani, A., Frossard, P.: Causal temporal regime structure learning. In: Li, Y., Mandt, S., Agrawal, S., Khan, M.E. (eds.) International Conference on Artificial Intelligence and Statistics, AISTATS 2025, Mai Khao, Thailand, 3-5 May 2025. Proceedings of Machine Learning Research, vol. 258, pp. 4546–4554. PMLR (2025)
26. Ramsey, J., Spirtes, P., Zhang, J.: Adjacency-faithfulness and conservative causal inference. In: Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. p. 401–408. UAI'06, AUAI Press, Arlington, Virginia, USA (2006)
27. Rodas, C.B., Tu, R., Li, Y., Kjellstrom, H.: Causal discovery from conditionally stationary time series. In: UAI 2022 Workshop on Causal Representation Learning (2022)
28. Runge, J.: Causal network reconstruction from time series: From theoretical assumptions to practical estimation. Chaos: An Interdisciplinary Journal of Nonlinear Science **28**(7), 075310 (2018)
29. Runge, J.: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In: Conference on Uncertainty in Artificial Intelligence. pp. 1388–1397. PMLR (2020)
30. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances **5**(11), eaau4996 (2019)
31. Sadeghi, A., Gopal, A., Fesanghary, M.: Causal discovery from nonstationary time series. International Journal of Data Science and Analytics (2025)
32. Saggioro, E., de Wiljes, J., Kretschmer, M., Runge, J.: Reconstructing regime-dependent causal relationships from observational time series. Chaos: An Interdisciplinary Journal of Nonlinear Science **30**(11), 113115 (2020)
33. Tsamardinos, I., Aliferis, C.F.: Towards principled feature selection: Relevancy, filters and wrappers. In: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. pp. 300–307. PMLR (03–06 Jan 2003)
34. Yang, J., Rao, X.: Copula entropy based causal network discovery from non-stationary time series. In: Antonacopoulos, A., Chaudhuri, S., Chellappa, R., Liu, C.L., Bhattacharya, S., Pal, U. (eds.) Pattern Recognition. pp. 115–131. Springer Nature Switzerland, Cham (2025)