A Logic-Based Approach to Causal Discovery: Signal Temporal Logic Perspective

Nasim Baharisangari[1], **Yucheng Ruan**[2], **Chengcheng Zhao**[2] and **Zhe Xu**[1]
[1]School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ 85287.
[2]College of Control and Engineering, Zhejiang University, 310027 Hangzhou, China.
{nbaharis, xzhe1}@asu.edu, {yuchengruan, chengchengzhao}@zju.edu.cn

Abstract

Causal discovery in time-series datasets is critical for understanding complex systems, especially when the *effectiveness* of causal relationships depends on both the *duration* and *magnitude* of the cause. We introduce a novel framework for causal discovery based on **Signal Temporal Logic (STL)**, enabling the extraction of interpretable causal diagrams (STL-CD) that explicitly capture these temporal dynamics. Our method first identifies statistically meaningful time intervals, then infers STL formulas that classify system behaviors, and finally employs transfer entropy to determine direct causal relationships among the formulas. This approach not only uncovers causal structure but also identifies the temporal persistence required for causal influence—an insight missed by existing methods. Experimental results on synthetic and real-world datasets demonstrate that our method achieves superior structural accuracy over state-of-the-art baselines, providing more informative and temporally precise causal models.

1 Introduction

In the realm of causal inference, learning the true underlying causal diagram using observational data is referred to as *causal discovery*. Information-theoretic causal discovery infers causal relations using entropy-based measures. Information theory provides measures such as entropy [Duan *et al.*, 2015; Yang *et al.*, 2024], mutual information [Hao *et al.*, 2015], and conditional mutual information [Runge, 2018], which can be employed to infer causal relationships. It applies to both static and time-series datasets. For example, [Shimizu *et al.*, 2006] proposes a causal discovery approach that relies on independent component analysis and does not require any pre-specified time-ordering of the variables.

The current causal discovery approaches mostly ignore the *temporal* aspect of the causal knowledge. For example, in certain time-series datasets, there might be some dominant cause-and-effect relationships among the variables where the effectiveness of the casual relationship depends on the duration of the cause. Consider, for instance, the scenario where for a fever-reducer drug to take effect, a patient must take the drug for four consecutive days. Such temporal patterns can be expressed through *signal temporal logic* (STL) formulas, a variant of temporal logic designed for real-valued variables. STL formulas strike a balance by being both human-interpretable and adhering to the precision of formal logics [Baharisangari *et al.*, 2022]. This enriches causal understanding when time critically shapes cause-effect dynamics.

Contribution: In this work, we present a framework for learning signal-temporal-logic-based causal diagrams (Signal Temporal Logic-based Causal Diagram (STL-CD)) from observational time-series datasets. (1) We define *signal-temporal-logic-based causality* in which we introduce STL-CDs and *temporal precedence* for STL formulas. (2) We present a novel algorithm for extracting time intervals with *statistically meaningful* trends in the given time-series dataset. (3) We present an algorithm for inferring STL formulas in the extracted time intervals such that they are sufficiently satisfied and violated by the *trajectories* in a given dataset. In this way, we can use information-theory-based methods such as transfer entropy to determine causal relationships among the STL formulas. (4) Finally, we introduce an algorithm for constructing STL-CDs using the inferred formulas. We apply the proposed framework on a car transmitter case study and a drug administration case study and compare the results with three baselines. To the best of our knowledge, this is the first work proposing a framework for signal-temporal-logic-based causal discovery.

1.1 Related Work

Causal discovery on time-series datasets has been extensively studied using constraint-based, score-based, and information-theoretic approaches. We focus on the latter in this work. Duan et al. [Duan *et al.*, 2015] introduced transfer 0-entropy (T0E) for causality analysis based on 0-entropy and 0-information, explicitly considering temporal aspects. Zhou et al. [Zhou *et al.*, 2022] proposed matrix-based conditional and high-order transfer entropy (CTEM, HTEM) for time-series causal discovery. Assad et al. [Assaad *et al.*, 2022] developed a framework for learning causal graphs from data with varying sampling rates. Su et al.

This work addresses a key gap in existing methods: the disregard of duration and logical relationships among variables. We introduce the first data-driven framework for learning *Signal Temporal Logic-based Causal Diagrams* (STL-CDs)

from time-series data. While prior studies, e.g., [Kleinberg and Mishra, 2009; Kleinberg, 2011], explored integrating temporal logic and causality, they do not construct temporal-logic-based diagrams from data. Sun et al. [Sun *et al.*, 2015] proposed a model-free causation entropy method to overcome the pairwise limitations of transfer entropy. While we employ transfer entropy to highlight STL-CD's capabilities, our framework is flexible and can integrate alternative metrics such as causation entropy.

2 Preliminaries

2.1 $(\mathbf{F}, \mathbf{G})$-Fragment Signal Temporal Logic

In this section, we briefly review $(\mathbf{F}, \mathbf{G})$-fragment signal temporal logic ($(\mathbf{F}, \mathbf{G})$-STL). First, we define *trajectory*.

Definition 1. *Given an D-dimensional domain $\mathcal{Y} \subset \mathbb{R}^D$ and the finite discrete time domain $\mathbb{I} = \{0, 1, \ldots, T-1\}$, we define a finite trajectory as a function $\xi : \mathbb{I} \to \mathcal{Y}$. We denote the length of the trajectory ξ by T. For a given labelled set of trajectories $\mathcal{D} = \{\xi_i, g_i\}_{i=1}^N$, we use $\xi_{i,d}$ to denote the d-th dimension of the i-th trajectory. Here, $g_i = +1$ represents the desired behavior and $g_i = -1$ represents the undesired behavior of the underlying system.*

$(\mathbf{F}, \mathbf{G})$-STL is a form of signal temporal logic that only incorporates the temporal operators $\mathbf{F}$ and $\mathbf{G}$. The syntax of $(\mathbf{F}, \mathbf{G})$-STL is defined recursively as follows.

$$\phi := \top \mid \pi \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \mathbf{G}_{[a,b]}\, \phi \mid \mathbf{F}_{[a,b]}\, \phi$$

where $\top$ stands for the Boolean constant *True*, π is an atomic predicate in the form of an inequality $f(\xi) > 0$ with f being some real-valued function and $\pi \in \mathcal{P}$ with $\mathcal{P} = \{\pi_1, \pi_2, ..., \pi_n\}$ is a finite set of atomic predicates with $n \in \mathbb{N} = \{1, 2, ...\}$. $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction) are standard Boolean connectives, "$\mathbf{F}$" and "$\mathbf{G}$" represent "eventually" and "always" temporal operators, respectively. In time interval $[a, b]$, we have $a \leq b$, and they are non-integers. We also define $L(\phi) \in \mathbb{B} = \{0, 1\}$ as a Boolean random variable for the truth value of the formula ϕ. The randomness of $L(\phi)$ stems from the fact we can use any random trajectory to evaluate the truth value of ϕ.

Definition 2. *The Boolean semantics of an $(\mathbf{F}, \mathbf{G})$-STL formula ϕ, for a trajectory ξ with the time length of T at time-step t is defined recursively as follows.*

$$\begin{aligned}
(\xi, t) &\models \pi \text{ iff } t \leq T \text{ and } f(\xi(t)) > 0 \\
(\xi, t) &\models \neg\phi \text{ iff } (\xi, t) \not\models \phi, \\
(\xi, t) &\models \phi_1 \wedge \phi_2 \text{ iff } (\xi, t) \models \phi_1 \text{ and } (\xi, t) \models \phi_2, \\
(\xi, t) &\models \mathbf{G}_{[a,b]}\, \phi \text{ iff } \forall t' \in [t+a, t+b],\ (\xi, t) \models \phi, \\
(\xi, t) &\models \mathbf{F}_{[a,b]}\, \phi \text{ iff } \exists t' \in [t+a, t+b],\ (\xi, t) \models \phi.
\end{aligned}$$

Robust semantics quantifies the margin at which a certain trajectory satisfies or violates an STL formula ϕ at time-step t. The robustness margin of a trajectory ξ with respect to an STL formula ϕ at time-step t is given by $r(\xi, \phi, t)$, where $r(\xi, \phi, t)$ can be calculated recursively via the robust semantics [Fainekos and Pappas, 2009]. For simplicity, we use STL to refer to $(\mathbf{F}, \mathbf{G})$-STL hereafter.

$$\begin{aligned}
r(\xi, \pi, t) &= f(\xi(t)), \\
r(\xi, \neg\phi, t) &= -r(\xi, \phi, t), \\
r(\xi, \phi_1 \wedge \phi_2, t) &= \min(r(\xi, \phi_1, t), r(\xi, \phi_2, t)), \\
r(\xi, \mathbf{G}_{[a,b]}\, \phi, t) &= \min_{t' \in [t+a, t+b]} r(\xi, \phi, t'), \\
r(\xi, \mathbf{F}_{[a,b]}\, \phi, t) &= \max_{t' \in [t+a, t+b]} r(\xi, \phi, t').
\end{aligned}$$

2.2 Information-Theory-Based Causal Discovery

In this section, we review the basics of information theory and its based method. Consider various events each with a probability of occurring denoted by $p_1, p_2, \ldots, p_n$, the entropy aiming to measure the uncertainty of information is defined as $H = -\sum_{1 \leq i \leq n} p_i \log p_i$. When targeting two variables, the joint entropy is given.

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \tag{1}$$

Except for joint entropy facing the bi-variate situation, the Transfer Entropy (TE), first proposed by Schreiber [Schreiber, 2000], measures directed information flows between variables with the minimum knowledge of system's dynamic [Mao and Shang, 2017]. We define TE from Y to X as follows.

$$TE_{Y,X} = \sum p(x_{i+\delta}, x_i, y_i) \log \frac{p(x_{i+\delta} | x_i, y_i)}{p(x_{i+\delta} | x_i)}, \tag{2}$$

where δ denotes the time delay, aiming to eliminate uncertainty decreased by the past of X.

3 Signal-Temporal-Logic-Based Causality

In this section, we introduce *signal-temporal-logic-based causality* (STL-CD) for STL formulas. We define STL-CD as follows.

Definition 3. *For a given set of predicates $\mathcal{P}$, STL-CD is a directed acyclic graph $\mathcal{G}$ where (1) each node represents an STL formula over $\mathcal{P}$, and (2) each edge ($\rightarrow$) represents a causal link between two nodes.*

Figure 1 represents two examples of STL-CDs with STL formulas as nodes. In STL-CDs in Figure 1, each pair of STL formulas is connected with a causal link $\phi_i \rightarrow \phi_j$, where ϕ_i is the cause and ϕ_j is the effect. STL-CD is an extension of the *temporal-logic-based causal diagram* (TL-CD) introduced in [Paliwal *et al.*, 2023][1] where the differences are as follows. (1) TL-CD uses *linear temporal logic* formulas as nodes in the DAG while STL-CD uses STL formulas. (2) STL-CD is non-deterministic while TL-CD is deterministic.

For a given STL-CD, if we evaluate the truth values of the STL formulas on the nodes of the STL-CD using a given trajectory ξ and replace the STL formulas with their corresponding truth values, then STL-CD is reduced to the standard causal diagram defined in Pearl causality [Pearl, 2009].

[1] [Paliwal *et al.*, 2023] is related to reinforcement learning, and it is assumed that the causal diagram is given while we discover it in our work.

Figure 1: Examples of STL-CDs.

For a STL-CD to be casually sound, we impose that the occurrence of the cause ϕ_i must precede that of the effect ϕ_j. In order to formalize this requirement, we first define *worst-case satisfaction time* $w_{\mathrm{s}}(\phi)$, the *worst-case violation time* $w_{\mathrm{v}}(\phi)$, the *best-case satisfaction time* $b_{\mathrm{s}}(\phi)$ and the *best-case violation time* $b_{\mathrm{v}}(\phi)$ of an STL formula.

Definition 4. *The **worst-case satisfaction** time $w_{\mathrm{s}}(\phi)$ (resp. **worst-case violation** $w_{\mathrm{v}}(\phi)$) is the maximum timesteps required to satisfy (resp. violate) the STL formulas ϕ using a trajectory ρ. The **best-case satisfaction** time $b_{\mathrm{s}}(\phi)$ (resp. **best-case violation** $b_{\mathrm{v}}(\phi)$) is the minimum timesteps required to satisfy (resp. violate) the STL formulas ϕ. The formulas to inductively calculate these parameters are shown in Table 1.*

Now, we formalize the precedence between the cause formula ϕ_i and the effect formula ϕ_j as follows.

Definition 5. ***(Temporal Precedence of STL Formulas)** We define that the STL formula ϕ_i precedes the STL formula ϕ_j if $\max\{w_v(\phi_i), w_s(\phi_i)\} < \min\{b_v(\phi_j), b_s(\phi_j)\}$.*

Example 1. *Suppose we have a drug that is administered to a patient, and we are interested in monitoring the effect of the drug on the patient's temperature. We can express this scenario using STL formulas. Our observations show that the patients who took the drug for four consecutive days eventually experienced temperature drops in the next two days. This observation can be expressed as $\mathbf{G}_{[0,3]}(\textit{taking the drug}) \rightarrow \mathbf{F}_{[4,5]}(\textit{patient's fever stops})$ as shown in Fig. 1 (Left).*

4 Signal-Temporal-Logic-Based Causal Discovery

In this section, we explain signal-temporal-logic-based causal discovery for a given time-series dataset $\mathcal{D}$. When dealing with a time-series dataset, the degree of causal influence among the attributes of the given dataset might vary at different time windows due to different factors. Hence, it is important to find the time intervals that might give us *statistically meaningful* information [Bryhn and Dimberg, 2011].

In this section, we first extract a set of time intervals $\mathcal{T}$ exhibiting *statistically meaningful trends* (Alg.1). For each dimension in $\mathcal{D}$, we then infer STL formulas over all intervals in $\mathcal{T}$ (Alg.2) and construct the STL-CD using these formulas (Alg. 3).

We use a time-series dataset $\mathcal{D}$ with T time steps, consisting of disjoint sets of *positive* trajectories $\mathcal{D}^{\text{pos}}$ (N^{pos} samples) and *negative* trajectories $\mathcal{D}^{\text{neg}}$ (N^{neg} samples). Positive trajectories exhibit desired system behavior, while negative trajectories represent undesired behaviors.

We infer STL formulas to classify positive and negative trajectories. The resulting STL-CD encodes one possible set of structural temporal causal relationships. By focusing on formulas that characterize desired behaviors, we obtain insights into the system dynamics. Further, instead of building the STL-CD directly from logical implications, we infer candidate formulas and identify causal relationships among them using transfer entropy. This guarantees that the final STL-CD represents only the strongest causal influences.

Example 2. *In the drug-administration example, the desired behavior represented by positive trajectories in $\mathcal{D}^{pos}$ can be when the patient's fever eventually stops. The undesired behavior which is represented by negative trajectories in $\mathcal{D}^{neg}$ can be when the patient has a fever at all times.*

The proposed framework does not use typical assumptions in causal discovery such as linearity, conditional independence, causal sufficiency, causal faithfulness, and non-Gaussianity since transfer entropy does not consider these assumptions [Lizier *et al.*, 2020; Vicente *et al.*, 2011; Barnett and Barrett, 2019]. Also, the proposed framework first converts the original time-series dataset into the Boolean dataset by evaluating the truth values of STL formulas to calculate the necessary probability density functions for entropy calculation.

4.1 Extracting Time Intervals with Statistically Meaningful Trends

Alg. 1 illustrates the steps we take to extract time intervals with *statistically meaningful* trends. We first take the following two steps to define *statistically meaningful* trends. (1) Given two trajectories ξ_1 with a 1D state x and ξ_2 with a 1D state y both with the length of T, we divide ξ_1 and ξ_2 into M sub-trajectories. (2) For each pair of sub-trajectories, we calculate the correlation value r^2 (Eq. (3)) and p-value between the values of x and y where the p-value is obtained by the two-sample Student's t statistic. In Eq. (3), T' is the length of an arbitrary sub-trajectory. Then, we define *statistically meaningful* trend between ξ_1 and ξ_2 as follows. "If one or several pairs of sub-trajectories yield $r^2 \geq 0.65$ and p-value ≤ 0.05, then a linear regression between trajectories ξ_1 and ξ_2 is deemed statistically meaningful".

$$r^2 = \frac{\left(\sum_{i=1}^{T'} \left((x_i - x_{avg})(y_i - y_{avg})\right)\right)^2}{\sum_{i=1}^{T'} (x_i - x_{avg})^2 \sum_{i=1}^{T'} (y_i - y_{avg})^2} \tag{3}$$

In Alg. 1, we apply this definition to extract the time intervals in which there are statistically meaningful trends between every two dimensions of a given dataset For a given dataset $\mathcal{D}$ with D dimensions (attributes) and N trajectories, we check for the existence of a statistically meaningful trend for every two different dimensions in all the trajectories (Lines 3-13). In doing so, we use the linear index of an upper triangular matrix to loop over the elements of the upper triangular matrix. We use this indexing to loop over all the unique pairs dimensions $d, j \in \{1, .., D\}$ (where $d \neq j$) for all the trajectories in $\mathcal{D}$. For a square matrix with D rows, the number of upper triangular elements is equal to $D' = \frac{D(D-1)}{2}$. For an arbitrary linear index $n \in \{1, ...D'\}$, we find the corresponding row index $d \in \{1, ..., D\}$ and column index $j \in \{1, ..., D\}$.

In Alg. 1, D' and N' are the numbers of elements of upper triangular matrices with D rows (number of dimensions in $\mathcal{D}$) and N rows (number of trajectories in $\mathcal{D}$) , respectively

The worst satisfaction time, worst violation time, best satisfaction time, and best violation time of an STL formula ϕ

$$b_{\mathrm{s}}(\pi) = w_{\mathrm{s}}(\pi) = b_{\mathrm{v}}(\pi) = w_{\mathrm{v}}(\pi) = 0,$$

$$\neg\phi : \begin{cases} b_{\mathrm{s}}(\neg\phi) = b_{\mathrm{v}}(\phi), \\ w_{\mathrm{s}}(\neg\phi) = w_{\mathrm{v}}(\phi), \\ b_{\mathrm{v}}(\neg\phi) = b_{\mathrm{s}}(\phi), \\ w_{\mathrm{v}}(\neg\phi) = w_{\mathrm{s}}(\phi); \end{cases}$$

$$\phi_1 \wedge \phi_2 : \begin{cases} b_{\mathrm{s}}(\phi_1 \wedge \phi_2) = \max\{b_{\mathrm{s}}(\phi_1), b_{\mathrm{s}}(\phi_2)\}, \\ w_{\mathrm{s}}(\phi_1 \wedge \phi_2) = \max\{w_{\mathrm{s}}(\phi_1), w_{\mathrm{s}}(\phi_2)\}, \\ b_{\mathrm{v}}(\phi_1 \wedge \phi_2) = \min\{b_{\mathrm{v}}(\phi_1), b_{\mathrm{v}}(\phi_2)\}, \\ w_{\mathrm{v}}(\phi_1 \wedge \phi_2) = \max\{w_{\mathrm{v}}(\phi_1), w_{\mathrm{v}}(\phi_2)\}; \end{cases}$$

$$\phi_1 \vee \phi_2 : \begin{cases} b_{\mathrm{s}}(\phi_1 \vee \phi_2) = \min\{b_{\mathrm{s}}(\phi_1), b_{\mathrm{s}}(\phi_2)\}, \\ w_{\mathrm{s}}(\phi_1 \vee \phi_2) = \max\{w_{\mathrm{s}}(\phi_1), w_{\mathrm{s}}(\phi_2)\}, \\ b_{\mathrm{v}}(\phi_1 \vee \phi_2) = \max\{b_{\mathrm{v}}(\phi_1), b_{\mathrm{v}}(\phi_2)\}, \\ w_{\mathrm{v}}(\phi_1 \vee \phi_2) = \max\{w_{\mathrm{v}}(\phi_1), w_{\mathrm{v}}(\phi_2)\}; \end{cases}$$

$$\mathbf{G}_{[a,b]}\phi : \begin{cases} b_{\mathrm{s}}(\mathbf{G}_{[a,b]}\phi) = b_{\mathrm{s}}(\phi) + b, \\ w_{\mathrm{s}}(\mathbf{G}_{[a,b]}\phi) = w_{\mathrm{s}}(\phi) + b, \\ w_{\mathrm{v}}(\mathbf{G}_{[a,b]}\phi) = w_{\mathrm{v}}(\phi) + b, \\ b_{\mathrm{v}}(\mathbf{G}_{[a,b]}\phi) = b_{\mathrm{v}}(\phi) + a; \end{cases}$$

$$\mathbf{F}_{[a,b]}\phi : \begin{cases} b_{\mathrm{s}}(\mathbf{F}_{[a,b]}\phi) = b_{\mathrm{s}}(\phi) + a, \\ w_{\mathrm{s}}(\mathbf{F}_{[a,b]}\phi) = w_{\mathrm{s}}(\phi) + b, \\ w_{\mathrm{v}}(\mathbf{F}_{[a,b]}\phi) = w_{\mathrm{v}}(\phi) + b, \\ b_{\mathrm{v}}(\mathbf{F}_{[a,b]}\phi) = b_{\mathrm{v}}(\phi) + a. \end{cases}$$

Table 1: The equations for calculating the worst satisfaction time $w_{\mathrm{s}}(\phi)$, worst violation time $w_{\mathrm{v}}(\phi)$, best satisfaction time $b_{\mathrm{s}}(\phi)$. and best violation time $b_{\mathrm{v}}(\phi)$ of an STL formula ϕ (Definition 4).

(Line 1). In Line 2, we initialize a 4D matrix of zeros denoted by $\boldsymbol{R} \in \mathbb{R}^{D' \times N' \times M \times M}$ where $M = M^{\max} - M^{\min} + 1$ and $M^{\max}$ and $M^{\min}$ are the given minimum and maximum number of intervals that we use to divide the trajectories into $m \in \{M^{\min}, ..., M^{\max}\}$ sub-trajectories. We use $\boldsymbol{R}$ to track the time intervals in which an arbitrary pair of dimensions meet the requirements of statistically meaningful trends. For a given number of intervals m and $\xi_{d',d}$, we denote the m-th sub-trajectory at dimension d by $\xi^m_{d',d}$. Alg. 1 loops over the unique pairs of dimensions d and j with $d \neq j$ (Lines 3-13) to extract the desired time intervals with statistically meaningful trends. In Lines 4 and 6, function `lin-up-itri()` calculates the corresponding row and column indices for upper triangular matrices with D' and N' elements, respectively.

In Lines 7-13, Alg. 1 divides a given pair of trajectories $\xi_{d',d}$ and $\xi_{j',j}$ into $m \in \{M^{\min}, ..., M^{\max}\}$ sub-trajectories (Line 8), respectively. Then, Alg. 1 calculates the p-value p (Line 10) and correlation value r using Eq. (3) (Line 11) between $\xi^m_{d',d}$ and $\xi^m_{j',j}$ and sets the element with the indices $n, m, k,$ and l of $\boldsymbol{R}$ to 1 if the requirements of the statistically meaningful trends are met (Line 13). $l \leq m$ is the index we use to loop over the sub-trajectories created by dividing $\xi_{d',d}$ into m sub-trajectories. In Lines 15-18, Alg. 1 extracts all the unique time intervals in which there is a statistically meaningful trend. In Line 17 of Alg. 1, $\boldsymbol{R}_{**ml}$ refers to the indices m and l of all the rows and columns in $\boldsymbol{R}$. In Line 17, Alg. 1 extracts the time intervals with statistically meaningful trends after dividing the time interval $[1, T]$ into m sub-intervals $\forall m \in \{M^{\min}, ..., M^{\max}\}$. Alg. 1 is designed for trajectories with two or more dimensions, but it can be simplified to work for 1D trajectories as well.

4.2 Signal Temporal Logic Inference

In this paper, we modify the STL inference method proposed in [Kong *et al.*, 2014] to infer STL formulas for a given dimension and time interval such that they are sufficiently satisfied and violated by trajectories in $\mathcal{D}$. Here, an optimization-based algorithm is used for inferring an STL formula ϕ from dataset $\mathcal{D}$ for a given dimension and time interval such that the inferred formula discriminates between the desired behavior and the undesired behavior of an underlying system at the given dimension and time interval. This optimization-based algorithm infers (1) the structure of ϕ including the temporal operators and Boolean connectives, and (2) the predicates such that he following objective function is minimized.

$$J(\phi_{\text{str}}, v, d, \tau, t) = \frac{1}{N}\sum_{i=1}^{N} \texttt{Loss}(g_i, r(\xi_{i,d}, \phi_{v,\tau}, t)) + \lambda\|\phi_{v,\tau}\|, \tag{4}$$

where ϕ_{str} is a candidate inferred STL structure, v is a candidate set of values to be used in the predicates in ϕ_{str}, r is the robustness degree, $\phi_{v,\tau}$ is the resulting STL formula after substituting the predicates in v and the time interval τ into ϕ_{str}, λ is a weighting parameter, $\|\phi_v\|$ is the number of predicates in ϕ_v, and `Loss` is a loss function.

Example 3. *In the drug administration example, for dimension $d_2 \in \{d_1, d_2\}$ where d_1 represents taking the drug and d_2 represents the temperature of the patient, $\phi_{str} := \mathbf{F}_{[?^1,?^2]}(\xi_{*,1} >?^3)$, $v = \{100\}$, and $\tau = [4, 5]$, $\phi_{v,\tau}$ can be written as $\phi_v := \mathbf{F}_{[4,5]}(\xi_{*,2} < 100)$ where $\xi_{*,2}$ denotes an arbitrary trajectory at dimension d_2. Here, we use "?" to denote an arbitrary unknown parameter including the value in a predicate or a bound in a time interval.*

Alg. 2 illustrates the steps we take to infer STL formulas for all the time intervals in $\mathcal{T}$ for each dimension d in $\mathcal{D}$. In Line 4, Alg. 2 randomly chooses $N_1 \leq N^{\text{pos}}$ trajectories from $\mathcal{D}^{\text{pos}}$ to form $\mathcal{D}_1$ and randomly chooses $N_2 \leq N^{\text{neg}}$ trajectories from $\mathcal{D}^{\text{neg}}$ to form $\mathcal{D}_2$. In Line 5, we cut the time interval τ out of the trajectories in $\mathcal{D}_1$ and $\mathcal{D}_2$ and form new trajectories with the length of $|\tau|$ and form $\mathcal{D}^{\text{red}}_1$ and $\mathcal{D}^{\text{red}}_2$, respectively. At Line 6, we infer an STL formula ϕ' using $\mathcal{D}^{\text{red}}_1$ and $\mathcal{D}^{\text{red}}_2$ at time interval $[0, |\tau|]$ such that the objective function (4) is minimized at time step $t = 0$. Then, at Line 7,

we map the time interval $[0, |\tau|]$ to the original time interval τ and form ϕ. The purpose of Lines 4-7 is to infer STL formulas that are sufficiently satisfied and violated by the trajectories in $\mathcal{D}$. Alg. 2 returns the sets $\Phi_1, ..., \Phi_D$ each containing $|\mathcal{T}|$ STL formulas where $|\mathcal{T}|$ is the number of time intervals in $|\mathcal{T}|$.

Example 4. *Given $\tau = [2, 4]$, a positive 1D trajectory $\xi_1 := 1, 1, 2, 2, 2$, a negative 1D trajectory $\xi_2 := -1, -1, -2, -2, -2$. In Line 5 of Alg. 2, the function* `traj-reduced`(ξ_1, ξ_2, τ) *return $\xi_1^{\text{red}} := 2, 2, 2$ and $\xi_2^{\text{red}} := -2, -2, -2$. In Line 6,* `Infer-Formula`$(\xi_1^{\text{red}}, \xi_2^{\text{red}}, [0, 2], 1)$ *returns $\phi' := \mathbf{G}_{[0,2]}(d_1 > 0)$ and* `map-time`(ϕ', τ) *in Line 7 returns $\phi := \mathbf{G}_{[2,4]}(d_1 > 0)$.*

4.3 Constructing STL-CD

In this subsection, we introduce the algorithm we use to construct STL-CD using the STL formulas inferred in Subsection 4.2. Alg. 3 illustrates the steps we take to construct an STL-CD $\mathcal{G}$ for given sets of STL formulas $\Phi_1, ..., \Phi_D$. At Line 1, we initialize a $D \times D$ matrix denoted by $\boldsymbol{Dir}$ to track the values that determine causal directions among STL formulas that are used on the nodes of the obtained STL-CD. We utilize $RatioTE$ to evaluate the causal effect from ϕ and θ defined as $RatioTE_{\phi,\theta} = \frac{TE_{\phi,\theta}}{H(L(\theta),L(\phi))}$ where $TE_{\phi,\theta}$ is the transfer entropy between two random variables $L(\theta)$ and $L(\phi)$ (Eq. (2)) and $H(L(\theta), L(\phi))$ is the joint entropy between these two random variables (Eq. (1)). The reason we add $Ratio$ but not simply TE is that the entropy of different STL formulas inferred is different. For calculating $RatioTE$ for two different STL formulas ϕ and θ denoted by $RatioTE_{\phi,\theta}$, we generate the Boolean datasets $\mathcal{D}_\phi^{\text{bool}}$ and $\mathcal{D}_\theta^{\text{bool}}$ by evaluating the truth values of ϕ and θ using $\mathcal{D}$ and calculate $RatioTE_{\phi,\theta}$ using these Boolean datasets. For each two dimensions $d, j \in \{1, ..., D\}$ and $d \neq j$, we initialize a $|\mathcal{T}|\times|\mathcal{T}|$ matrix denoted by $\boldsymbol{RatioTEMat}$ (Line 4) to store the $RatioTE_{\phi,\theta}$ values between each two formulas $\phi \in \Phi_d$ and $\theta \in \Phi_j$ (Line 16). We only calculate the $RatioTE$ between each two ϕ and θ that satisfy the temporal precedence constraint (Line 12).

To fill the elements in $\boldsymbol{RatioTEMat}$, we calculate $RatioTE_{\phi,\theta}$ using each trajectory $\xi \in \mathcal{D}$ and then choose the maximum value (Line 16). In Lines 14-15, we calculate $RatioTE_{\phi,\theta}$ using $\xi_i^{\text{bool},\phi}$ and $\xi_i^{\text{bool},\theta}$ forall $i \in \{1, ..., N\}$ where $\xi_i^{\text{bool},\phi}$ is the Boolean 1D trajectory with T time-steps obtained by evaluating the truth values of ϕ using ξ_i at time-steps $t \in \{0, ..., T-1\}$. At Line 17, Alg. 3 fills the $\boldsymbol{Dir}_{dj}$ with the summation of all the $RatioTE$ values in $\boldsymbol{RatioTEMat}$ as the determinant of the causal direction between dimension d and j. In this way, we consider the cumulative $RatioTE$ values of all $|\mathcal{T}|$ formulas that we have inferred for each dimension d. At Lines 18 and 19, we extract the best pair of formulas for dimensions d and j with the corresponding $RatioTE$ value. In Line 20, we calculate the causal effect difference between dimension d and j and store it in $\boldsymbol{IndiMat}$. Finally, in Lines 21-28, we add a causal link from dimension d to j if $\boldsymbol{IndiaMat}_{dj} > 0$, else we add a causal link from dimension j to d.

For a given pair of dimensions j and d, Alg. 3 returns the pair of formulas ϕ_d and θ_j that has the highest $RatioTE_{\phi,\theta}$ along with the causal direction. If we have $\phi_d \rightarrow \theta_j$, then it is possible that the time interval of ϕ_d does not start from 0. The reason is that many of the statistically meaningful time intervals that Alg. 1 returns do not start from 0. For having pairs of formulas that are meaningful when converted to natural language, Alg. 3 maps the time intervals of ϕ_d and θ_j such that the time interval of ϕ_d starts from 0 (Lines 24 and 27). When the formulas are fixed, the direction obtained by transfer entropy converges to the ground truth causal direction as the amount of data increases.

Algorithm 1: Extracting time intervals with statistically meaningful trends.

Input: observational time-series dataset $\mathcal{D}$ with T time steps and N trajectories with D dimensions, maximum number of intervals M^{max}, minimum number of intervals M^{min}, significance level $\alpha = 0.05$

1 $N' \leftarrow \frac{N(N-1)}{2}, D' \leftarrow \frac{D(D-1)}{2}$
2 Initialize $\boldsymbol{R}$ as 4D matrix of $\mathbf{0} \in \mathbb{R}^{D'\times N'\times M\times M}$
3 **for** $n = 1, ..., D'$ **do**
4 $\quad d, j \leftarrow$ `lin-up-tri`(D, n)
5 $\quad$ **for** $k = 1, ..., N'$ **do**
6 $\quad\quad d', j' \leftarrow$ `lin-up-tri`(N, k)
7 $\quad\quad$ **for** $m = 1, ..., M$ **do**
8 $\quad\quad\quad$ Divide $\xi_{d',d}$ and $\xi_{j',j}$ into m sub-trajectories
9 $\quad\quad\quad$ **for** $l = 1, ..., m$ **do**
10 $\quad\quad\quad\quad p \leftarrow$ `ttest2`$(\xi^l_{d',d}, \xi^l_{j',j}, \alpha)$
11 $\quad\quad\quad\quad r^2 \leftarrow$ `r2corr`$(\xi^l_{d',d}, \xi^l_{j',j})$
12 $\quad\quad\quad\quad$ **if** $(p \leq \alpha) \wedge (r^2 \geq 0.65)$ **then**
13 $\quad\quad\quad\quad\quad \boldsymbol{R}_{nkml} \leftarrow 1$
14 $\mathcal{T} \leftarrow \emptyset$
15 **for** $m = 1, ..., M$ **do**
16 $\quad$ **for** $l = 1, ..., m$ **do**
17 $\quad\quad$ **if** $\exists m$ *and* $l, \boldsymbol{R}_{**ml} = 1$ **then**
18 $\quad\quad\quad$ Divide the time interval $[1, T]$ into m intervals and add the l-th time interval to $\mathcal{T}$
19 **return** $\mathcal{T}$

Example 5. *Assume for the Ex. 1, Alg. 3 returns $\mathbf{G}_{[4,7]}$(taking the drug)$\rightarrow\mathbf{F}_{[8,9]}$(patient's fever stops) for a set of trajectories with length $T = 20$. We map the time intervals in these formulas such that $\mathbf{G}_{[4,7]}$ starts from zero and we obtain $\mathbf{G}_{[0,3]}$(taking the drug)$\rightarrow\mathbf{F}_{[4,5]}$(patient's fever stops) which yields a meaningful natural language translation.*

5 Evaluation

We have applied the framework to both the synthesis and real-world vehicular sensory data sets. The former is constructed through simulation equations, while the latter is collected through vehicle onboard vehicular sensors on Controller Area Network (CAN) by reversing engineering the specification of the CAN protocol [Ruan *et al.*, 2025]. We also performed comparisons with the main causal discovery algorithms as follows:

Algorithm 2: Inferring $(\mathbf{F}, \mathbf{G})$-STL formulas

Input: observational time-series dataset $\mathcal{D}$ with T time steps and D dimensions, set of time bounds $\mathcal{T}$, Number of random positive and negative trajectories $N_1 \leq N^{\text{pos}}$ and $N_2 \leq N^{\text{neg}}$

$\Phi_1, ..., \Phi_D \leftarrow \emptyset$
for $d = 1, ..., D$ **do**
  for $\tau \in \mathcal{T}$ **do**
    $\mathcal{D}_1, \mathcal{D}_2 \leftarrow$ `rnd-traj`$(\mathcal{D}, N_1, N_2)$
    $\mathcal{D}_1^{\text{red}}, \mathcal{D}_2^{\text{red}} \leftarrow$ `traj-reduced`$(\mathcal{D}_1, \mathcal{D}_2, \tau)$
    $\phi' \leftarrow$ `Infer-Formula`$(\mathcal{D}_1^{\text{red}}, \mathcal{D}_2^{\text{red}}, [0, |\tau|], d)$
    $\phi \leftarrow$ `map-time`(ϕ', τ)
    Add ϕ to Φ_d
return $\Phi_1, ..., \Phi_D$

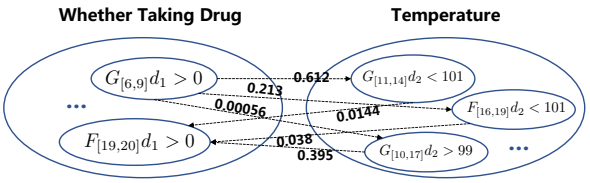


Figure 2: The primary calculation results of $RatioTE$ between STL formulas are partially displayed, where the direction is determined by time precedence and the weight of the edge is the value of $RatioTE$ between two STL formulas.

- Temporal Causal Discovery Framework (TCDF) [Nauta *et al.*, 2019] targets discovering non-linear causal relations between time series using deep learning networks with an attention mechanism within a dilated depthwise convolutional network.
- Multivariate Granger Causality (MVGC) [Barnett and Seth, 2014] is the representative methods for discovering Granger Causality using the vector autoregressive model.
- Symbolic Transfer Entropy [Staniek and Lehnertz, 2008] utilizes technique of symbolization to calculate transfer entropy getting rid of difficulties within probability estimation on non-stationary time-series.

In the results, the red arrow represents the wrong causal direction and the black arrow represents the correct one.

5.1 Drug-Administration Case Study

In this subsection, we apply the proposed method to a drug-administration case study where we use synthetic data where the ground truth of the desired behavior is *"taking the fever-reducer drug for four consecutive days causes the fever to eventually stop in the next days once the four consecutive days are finished"*. This statement can be as $\mathbf{G}_{[0,3]}$ *TakingDrug* $\rightarrow \mathbf{F}_{[4,?]}$ *FeverStop* where ? indicates an unknown time upper bound. The undesired behavior which is represented by negative trajectories is when the patient has a fever. For this case study, we generated the dataset $\mathcal{D}$ with 16 positive trajectories and 2 negative trajectories all with the length of $T = 25$ and two dimensions. The first dimension d_1 indicates whether the patient has taken the drug, i.e., $d_1 = -1$ indicates the drug has not been taken and $d_1 = -1$ indicates that the drug has been taken. The second dimension d_2 is the temperature value of the patient. $d_1 < 100F^{\circ}$ indicates the fever has stopped and $d_1 > 100F^{\circ}$ indicates that the fever has not been stopped. We also set $M^{\min} = 1$ and $M^{\max} = 20$. We have inferred 25 STL formulas for each dimension and then calculated the TE among pairs according to temporal precedence constraints. We partially display the result in Fig 2. Finally, the obtained STL-CD (with mapped formulas) is shown in Fig. 3a, unveiling the truth that taking the drug is the cause of falling temperature. The unmapped formulas are $\mathbf{G}_{[6,9]}$ for cause and $\mathbf{G}_{[11,15]}$ for effect where the duration of taking drug is four consecutive days as described by pairs of STL formulas with the maximum $RatioTE$ value. The sat-

Algorithm 3: STL-based causal discovery

Input: observational time-series dataset $\mathcal{D}$ with T time steps and D dimensions, sets of STL formulas $\Phi_1, ...\Phi_D$

$\boldsymbol{Dir} \leftarrow \mathbf{0} \in \mathbb{R}^{D \times D}$
for $d = 1, ..., D$ **do**
  for $j = 1, ..., D$ **do**
    $\boldsymbol{RatioTEMat} \leftarrow \mathbf{0} \in \mathbb{R}^{|\mathcal{T}| \times |\mathcal{T}|}$
    if $d \neq j$ **then**
      $n \leftarrow 0$
      for $\phi \in \Phi_d$ **do**
        $n \leftarrow n + 1$
        $m \leftarrow 0$
        for $\theta \in \Phi_j$ **do**
          $m \leftarrow m + 1$
          if $\big(\max\{w_v(\phi), w_s(\phi)\} < \min\{b_v(\theta), b_s(\theta)\}\big)$ **then**
            Evaluate the truth values $L(\phi)$ and $L(\theta)$ using trajectories from $\mathcal{D}$ separately and construct the Boolean datasets $\mathcal{D}_\phi^{bool}$ and $\mathcal{D}_\theta^{bool}$, respectively
            $\boldsymbol{TETraj} \leftarrow \mathbf{0}^{N \times 1}$
            for $i = 1, ..., N$ **do**
              $\boldsymbol{TETraj}_i \leftarrow$ `Compute-TE`$(\xi_i^{\texttt{bool},\phi}, \xi_i^{\texttt{bool},\theta})$
        $\boldsymbol{RatioTEMat}_{nm} \leftarrow \max \boldsymbol{TETraj}$
      $\boldsymbol{Dir}_{dj} \leftarrow$ `sum`$(\boldsymbol{TEMat})$
      TE-$dj \leftarrow \max \boldsymbol{RatioTEMat}$
      $\phi_d, \theta_j \leftarrow \arg\max_{\phi,\theta} (\boldsymbol{RatioTEMat})$
$\boldsymbol{IndiaMat} \leftarrow (\boldsymbol{Dir} - \boldsymbol{Dir}^{\text{tran}})./\boldsymbol{Dir}$
for $d = 1...D$ **do**
  for $j = d + 1, ..., D$ **do**
    if $(\boldsymbol{IndiaMat}_{dj} > 0)$ **then**
      $\phi'_d, \theta'_j \leftarrow$ `map0`(ϕ_d, θ_j)
      Add a casual arrow from ϕ'_d to θ'_j with the corresponding TE value of TE-dj
    else
      $\phi'_d, \theta'_j \leftarrow$ `map0`(θ_j, ϕ_d)
      Add a casual arrow from θ'_j to ϕ'_d with the corresponding TE value of TE-jd
return *STL-CD* $\mathcal{G}$

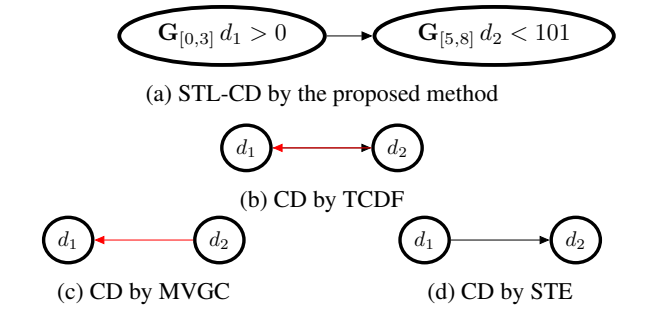
isfaction of the effect (mapped) formula $\mathbf{G}[5,8]d_2 < 101$ implies the satisfaction of the effect formula in the ground truth. This means that the temporal behavior specified by $\mathbf{G}[5,8]d_2 < 101$ encompasses the temporal behavior specified by $\mathbf{F}_{[4,?]}$*FeverStop*. As for the predicate $d_2 < 101$, the value $d_2 = 101$ indicates having a fever, and the value $d_2 = 99$ indicates not having a fever in the trajectories and the predicate $d_2 < 101$ captures "not having a fever" correctly. We also apply three other baseline methods on Drug synthesis dataset and their results are shown in Fig. 3b, Fig. 3c, Fig. 3d. As seen, only STE generated the correct CD. However, only the proposed method extracted the ground truth duration of taking drug that causes the fever to drop. We quantify structural accuracy by the percentage of correctly identified edges. In the drug case study, our method achieved **100% accuracy**, outperforming TCDF (0%) and MVGC (0%).

5.2 On-Board Vehicular Sensory Data

In this case study, we collect on-board vehicular sensory data from CAN bus when the vehicle is under normal operation. The dataset $\mathcal{D}$ has 9 normalized trajectories with the length of $T = 3000$ and each trajectory has four dimensions. d_1 is the "vehicle speed", d_2 is the "engine speed", d_3 is the "accelerator paddle position", and d_4 is the "throttle position". The underlying causal diagram originates from common Electronic Control Units (ECUs) within the vehicle including the Throttle Control Unit [Loh *et al.*, 2013], Transmission Control Unit [Tamada *et al.*, 2020], etc. According to the design of the control logic inside ECUs, we could figure out the ground truth for these four attributes (without any STL formulas).

$\mathcal{D}$ has nine positive trajectories and three negative trajectories where positive trajectories represent the situation where the vehicle is under operation (the vehicle is moving) and the negative trajectories represent the situation where the vehicle is not under operation (where the vehicle is static). Fig. 4a shows the obtained results for Vehicle dataset with $M^{\text{min}} = 90$ and $M^{\text{max}} = 100$. As seen, the proposed method discovered five edges correctly out of six possible edges among the four attributes. In Fig. 4a, it can be seen that three different formulas are inferred for d_1, three different formulas for d_3, two different formulas for d_2, and two different formulas for d_4 wherein each pair of formulas, the temporal precedence constraint is met and the causal link has a correct direction. For example, $\mathbf{G}_{[0,29]}(d_4 > 0.5936) \rightarrow \mathbf{F}_{[537,565]}(d_1 > 0.797)$ means "the normalized throttle position being greater than 0.5936 for 30 consecutive time units causes the normalized vehicle speed to eventually be greater than 0.797 in the time window of [537, 565]". Moreover, several such STL formulas in different time intervals with statically meaningful trends can be used in the abstraction of the temporal properties in the context of causality and the abstraction of systems dynamics. The formulas we get could be used as forming blocks for vehicle dynamics STL specifications benchmarks such as in [Hoxha *et al.*, 2014]. For discovering the ground truth STL formulas of this case study, we need to use more data and perform more extensive experiments.

Fig. 4b shows the obtained CD by MVGC where the correct discovered edges are $d_3 \rightarrow d_4$, $d_3 \rightarrow d_2$, and $d_4 \rightarrow d_2$.

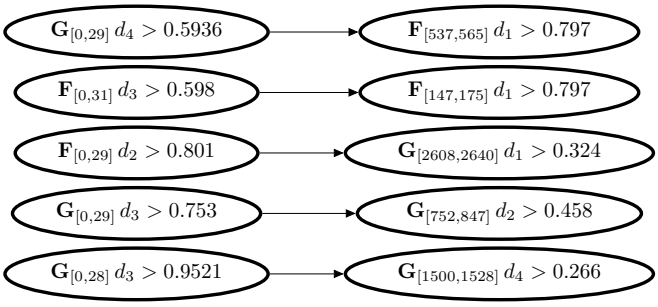


(a) STL-CD by the proposed method

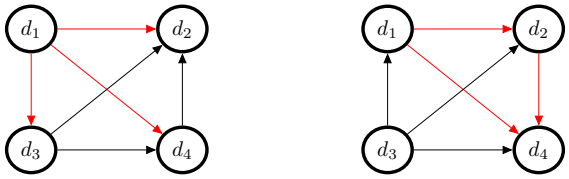
(b) CD by TCDF

(c) CD by MVGC (d) CD by STE

Figure 3: The obtained results for the drug-administration case study.



(a) STL-CD by the proposed method



(b) CD by MVGC (c) CD by STE

Figure 4: The obtained results for the Vehicular Sensory Data case study

Fig. 4c shows the obtained CD by STE where the correct discovered edges are $d_3 \rightarrow d_1$, $d_3 \rightarrow d_2$ and $d_3 \rightarrow d_4$. TCDF did not return any causal link. The method we proposed not only discovers the closest causal diagram but also reveals a potential temporal logic within the controller. In the vehicular case, our method achieved **83% accuracy**, also surpassing MVGC (50%) and STE (50%).

6 Conclusion

In this paper, we introduced signal-temporal-logic-based causal discovery to explicitly capture the temporal aspects of causal relationships, enabling agents to reason about causality over time while considering the order and duration of events. This approach has the potential to improve decision-making in complex, dynamic environments. As future work, we aim to optimize the extraction of statistically meaningful time intervals to maximize their utility based on dataset size. Additionally, we plan to extend the method to construct a complete DAG rather than focusing on pairwise nodes.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grants CNS 2339774, CNS 2304863, and IIS 2332476. Yucheng Ruan and Chengcheng Zhao want to acknowledge the National Natural Science Foundation of China under Grant 62273305 and the Natural Science Foundation of Zhejiang Province under Grant LR25F030002.

References

[Assaad *et al.*, 2022] Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Entropy-based discovery of summary causal graphs in time series. *Entropy*, 24(8), 2022.

[Baharisangari *et al.*, 2022] Nasim Baharisangari, Jean-Raphaël Gaglione, Daniel Neider, Ufuk Topcu, and Zhe Xu. Uncertainty-aware signal temporal logic inference. In Roderick Bloem, Rayna Dimitrova, Chuchu Fan, and Natasha Sharygina, editors, *Software Verification*, pages 61–85, Cham, 2022. Springer International Publishing.

[Barnett and Barrett, 2019] Lionel Barnett and Adam B. Barrett. Quantifying 'causality' in complex systems: Understanding transfer entropy. In *Probabilistic Graphical Models*, pages 309–324. Springer, 2019.

[Barnett and Seth, 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.

[Bryhn and Dimberg, 2011] Andreas C Bryhn and Peter H Dimberg. An operational definition of a statistically meaningful trend. *PLoS One*, 6(4):e19241, 2011.

[Duan *et al.*, 2015] Ping Duan, Fan Yang, Sirish L. Shah, and Tongwen Chen. Transfer zero-entropy and its application for capturing cause and effect relationship between variables. *IEEE Transactions on Control Systems Technology*, 23(3):855–867, 2015.

[Fainekos and Pappas, 2009] Georgios E. Fainekos and George J. Pappas. Robustness of temporal logic specifications for continuous-time signals. *Theoretical Computer Science*, 410(42):4262–4291, 2009.

[Hao *et al.*, 2015] Zhifeng Hao, Hao Zhang, Ruichu Cai, Wen Wen, and Zhihao Li. Causal discovery on high dimensional data. *Applied Intelligence*, 42(3):594–607, April 1 2015.

[Hoxha *et al.*, 2014] Bardh Hoxha, Houssam Abbas, and Georgios Fainekos. Benchmarks for temporal logic requirements for automotive systems. *ARCH@ CPSWeek*, 34:25–30, 2014.

[Kleinberg and Mishra, 2009] Samantha Kleinberg and Bud Mishra. The temporal logic of causal structures. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, page 303–312, Arlington, Virginia, USA, 2009. AUAI Press.

[Kleinberg, 2011] Samantha Kleinberg. A logic for causal inference in time series with discrete and continuous variables. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, page 943–950. AAAI Press, 2011.

[Kong *et al.*, 2014] Zhaodan Kong, Austin Jones, Ana Medina Ayala, Ebru Aydin Gol, and Calin Belta. Temporal logic inference for classification and prediction from data. In *Proceedings of the 17th International Conference on Hybrid Systems: Computation and Control*, HSCC '14, page 273–282, New York, NY, USA, 2014. Association for Computing Machinery.

[Lizier *et al.*, 2020] Joseph T. Lizier, Mikhail Prokopenko, and Albert Y. Zomaya. Estimating transfer entropy in continuous time between neural spike trains or other event-based data. *PLOS Computational Biology*, 16(2):e1008054, 2020.

[Loh *et al.*, 2013] Robert NK Loh, Witt Thanom, Jan S Pyko, Anson Lee, et al. Electronic throttle control system: modeling, identification and model-based control designs. *Engineering*, 5(07):587, 2013.

[Mao and Shang, 2017] Xuegeng Mao and Pengjian Shang. Transfer entropy between multivariate time series. *Communications in Nonlinear Science and Numerical Simulation*, 47:338–347, 2017.

[Nauta *et al.*, 2019] Meike Nauta, Doina Bucur, and Christin Seifert. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):19, 2019.

[Paliwal *et al.*, 2023] Yash Paliwal, Rajarshi Roy, Jean-Raphaël Gaglione, Nasim Baharisangari, Daniel Neider, Xiaoming Duan, Ufuk Topcu, and Zhe Xu. Reinforcement learning with temporal-logic-based causal diagrams. *arXiv preprint arXiv:2306.13732*, 2023.

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

[Ruan *et al.*, 2025] Yucheng Ruan, Chengcheng Zhao, Zeyu Yang, Yuanchao Shu, Peng Cheng, and Jiming Chen. Picacan: Reverse engineering physical semantics of signals in can messages using physically-induced causalities. *IEEE Transactions on Mobile Computing*, 2025.

[Runge, 2018] Jakob Runge. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 938–947. PMLR, 09–11 Apr 2018.

[Schreiber, 2000] Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

[Shimizu *et al.*, 2006] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7:2003–2030, dec 2006.

[Staniek and Lehnertz, 2008] Matthäus Staniek and Klaus Lehnertz. Symbolic transfer entropy. *Physical review letters*, 100(15):158101, 2008.

[Sun *et al.*, 2015] 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.

[Tamada *et al.*, 2020] Sireesha Tamada, Debraj Bhattacharjee, and Pranab K Dan. Review on automatic transmission control in electric and non-electric automotive powertrain. *International Journal of Vehicle Performance*, 6(1):98–128, 2020.

[Vicente *et al.*, 2011] Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1):45–67, 2011.

[Yang *et al.*, 2024] Zeyu Yang, Liang He, Yucheng Ruan, Peng Cheng, and Jiming Chen. Unveiling physical semantics of plc variables using control invariants. *IEEE Transactions on Dependable and Secure Computing*, 2024.

[Zhou *et al.*, 2022] Wanqi Zhou, Shujian Yu, and Badong Chen. Causality detection with matrix-based transfer entropy. *Information Sciences*, 613:357–375, 2022.