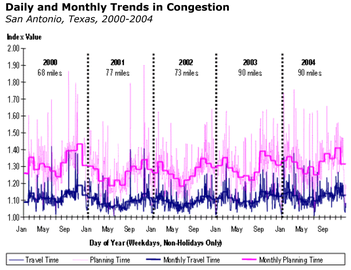
# Causal Discovery in Semi-Stationary Time Series

Shanyun Gao[1], Raghavendra Addanki[2], Tong Yu[2], Ryan A. Rossi[2], Murat Kocaoglu[1]

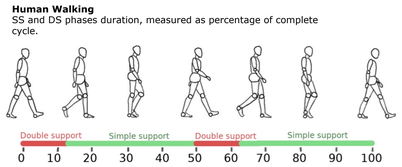
[1] Purdue University, West Lafayette, USA
[2] Adobe Research, San Jose, USA

## Motivation

- **Periodic** nature is commonly observed in many real-world time series data.
- Seasonality demonstrates a typical recurring pattern, particularly noticeable in climate variations, showcasing periodicity.
- Periodic changes in the causal relations are expected underlying this type of time series without assuming stationarity.



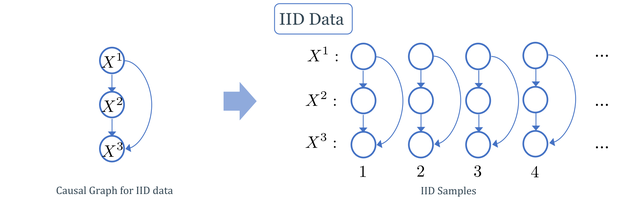

Source: Analysis of data from FHWA's *Mobility Monitoring Program*




Source: Kalouguine, Anne, et al. 2020

## PCMCI$_\Omega$: Capturing Periodicity of the Causal Structure

- **The goal**: Discover the causal structure from a specific type of non-stationary time series data, called the **semi-stationary** time series, where the underlying causal structures change periodically.
- **Challenge 1**: Autocorrelated Samples.
  - Most existing Conditional Independence tests assume Independent and Identically Distributed (IID) samples.
  - However, time series are non-IID samples.



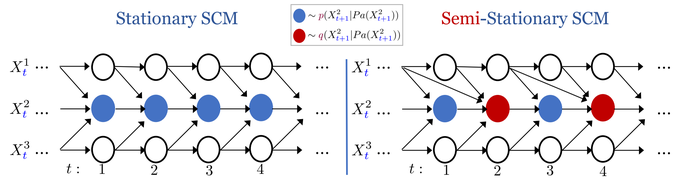

- **Q1:** How to deal with non-IID samples with constraint-based approaches under **stationary Structural Causal Model (SCM),** assuming all the causal relationships remain constant in direction throughout time?
  **A1:** In Runge et al. 2019, PCMCI applies the Momentary Conditional Independence test (MCI) to deal with the autocorrelations in samples.

  Testing $X^i_{t-\tau} \to X^j_t, \tau > 0$, for $CI$ any measure of dependence:

  CI tests: $CI(X^i_{t-\tau}, X^j_t | Pa(X^j_t) \setminus X^i_{t-\tau})$

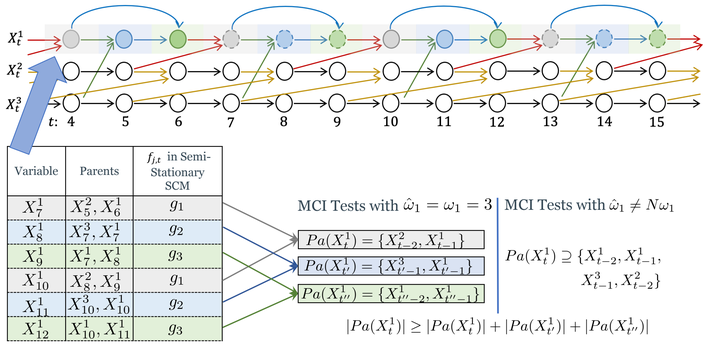
  MCI tests: $CI(X^i_{t-\tau}, X^j_t | Pa(X^j_t) \setminus X^i_{t-\tau}, Pa(X^i_{t-\tau}))$

- **Challenge 2**: Stationary SCM vs Semi-Stationary SCM.
  - Most existing causal discovery algorithms for time series assume stationarity (Assaad et al. [2022]).
  - However, alternating and recurring changes in the causal mechanisms may exist.




- **Q2:** How to estimate the periodicity of causal mechanisms in $X^j$, denoted by $\omega_j$?
  **A2:** $\hat{\omega}_j \leftarrow \arg \min_{\text{all guesses of } \omega_j} \max_t |\widehat{Pa}(X^j_t)|$

## Theorem




| Variable | Parents | $f_{j,t}$ in Semi-Stationary SCM |
|---|---|---|
| $X^1_7$ | $X^2_5, X^1_6$ | $g_1$ |
| $X^1_8$ | $X^3_7, X^1_7$ | $g_2$ |
| $X^1_9$ | $X^1_7, X^1_8$ | $g_3$ |
| $X^1_{10}$ | $X^2_8, X^1_9$ | $g_1$ |
| $X^1_{11}$ | $X^3_{10}, X^1_{10}$ | $g_2$ |
| $X^1_{12}$ | $X^1_{10}, X^1_{11}$ | $g_3$ |

- Choosing $\hat{\omega}_j$ means collecting samples at every $\hat{\omega}_j$ steps.
- MCI tests are performed on these $\hat{\omega}_j$-grouped samples.
  - If $\hat{\omega}_j = \omega_j$ , samples from the same group share the same causal mechanism. $\Longrightarrow$ We are back to the Stationary SCM from Semi-Stationary SCM.
  - If $\hat{\omega}_j \neq N\omega_j$ , samples from the same group belong to a mixture conditional distribution. $\Longrightarrow$ By Lemma 3.2, all dependent relations within each component distribution remain. $\Longrightarrow$ With Hard Mechanism Change assumption, this leads to a denser graph.

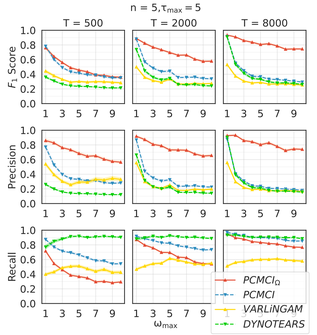
**Theorem 3.1** *Let $\hat{\mathcal{G}}$ be the estimated graph using the Algorithm PCMCI$_\Omega$, under assumptions A1-A7 and with an oracle (infinite sample size limit), we have that:*

$$\hat{\mathcal{G}} = \mathcal{G}$$

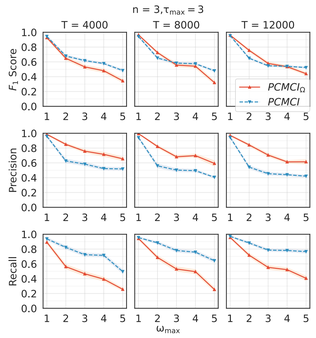
*almost surely.*

## Experiment Results

- **Simulations** with Continuous-valued Time Series and Discrete-valued Time Series.




4 algorithms are tested on 5−multivariate continuous-valued time series with Gaussian noise. Every line corresponds to a different algorithm. Every marker corresponds to the average performance over 100 trials.




2 algorithms are tested on 3−multivariate binary-valued time series with different time length.

- **Case study** with a real-world climate time series dataset, inspired by the application in Runge et al. 2019.
  - The data is from 1948-2022 with 900 months.
  - $X^{\text{wp}}_t$ denotes the monthly mean of surface pressure in the West Pacific; $X^{\text{cp}}_t$ , $X^{\text{ep}}_t$ and $X^{\text{ta}}_t$ denotes the monthly mean of air temperature in the Central Pacific, East Pacific, and tropical Atlantic, respectively.
  - We found that the causal effect from the tropical Atlantic air temperature to the Central Pacific air temperature would disappear every quarter of a year.

| $X$ | $\hat{\omega}$ | $\{\widehat{Pa}_k\}_{k\in[\hat{\omega}]}$ : True and illusory parent sets |
|---|---|---|
| $X^{\text{wp}}_t$ | 1 | $\{X^{\text{wp}}_{t-1}, X^{\text{wp}}_{t-2}, X^{\text{ep}}_{t-1}, X^{\text{ta}}_{t-1}\}$ |
| $X^{\text{cp}}_t$ | 3 | $\{X^{\text{cp}}_{t-1}\}; \{X^{\text{cp}}_{t-1}, X^{\text{cp}}_{t-2}, X^{\text{ta}}_{t-1}\}; \{X^{\text{cp}}_{t-1}, X^{\text{ta}}_{t-1}\}$ |
| $X^{\text{ep}}_t$ | 1 | $\{X^{\text{ep}}_{t-1}, X^{\text{ta}}_{t-1}, X^{\text{ta}}_{t-2}, X^{\text{cp}}_{t-1}\}$ |
| $X^{\text{ta}}_t$ | 1 | $\{X^{\text{ta}}_{t-1}, X^{\text{wp}}_{t-1}\}$ |

Table: Climate application results estimated from PCMCI$_\Omega$

## Conclusions

- We propose a non-parametric, constraint-based causal discovery algorithm PCMCI$_\Omega$ designed for semi-stationary time-series data, in which a finite number of causal mechanisms are repeated periodically.
- We establish the soundness of our algorithm and assess its effectiveness on continuous-valued and discrete-valued time series data.
- The algorithm PCMCI$_\Omega$ has the capacity to reveal the existence of periodicity of causal mechanisms in real-world datasets.

## References

Runge, Jakob, et al. "Detecting and quantifying causal associations in large nonlinear time series datasets." *Science advances* 5.11 (2019).

Assaad, Charles K., Emilie Devijver, and Eric Gaussier. "Survey and evaluation of causal discovery methods for time series." *Journal of Artificial Intelligence Research* 73 (2022).