Machine Learning for Graphs

Reconstructing time-varying causal graphs through temporal networks

Michele Vannucci

studentnumber: 2819493

The code used to obtain the results of this report can be found here.

Abstract. In this report, after introducing the field of causal discovery and its core concepts, I provide an overview of the state of the art of causal discovery, addressing how the latter can be achieved in complex systems in which the causal dynamics change through time. The pretext to understands these concepts deeper is reproducing a recent research in the field [14], which focuses on stationary causal regimes. This research is then extended introducing a new algorithm for causal discovery, MCI-RNN, that uses temporal networks to represent the non-linear causal links between time-dependent observational variables.

Keywords: AI · Graphs · Causality · Dynamic systems

1 Introduction

The quest for causality has deeply characterized the scientific journey of humanity, that strove to understand the causes and the principles behind natural phenomena. This has been driven by the scientific method, which finds its roots in controlling the experimental variables separately to isolate their effects and causal links. But in some cases this is not feasible or ethical. This can be seen in the meteorological domain, where it's impossible to control variables like temperature of pressure in a determinate region to discern it's effects on other parts of the atmosphere. Another example are biological systems, like the human brain or the cardiovascular apparatus[12]. Moreover, these systems are too complex to be observed in action in their entirety as they are constituted by millions of chaotic interactions. Usually we measure values that describe and *summarize* parts of them as humidity, wind speed and precipitation in a geographic area in the meteorological domain, or BOLD signals in fMRI brain imaging[4]. This is where methods for causal discovery from observational time-series become necessary, they can discern when the statistical correlation between two variables entails causality between the two, by accounting for the possible confounding variables available.

For example let's say that we have two variables X and Y that are linearly correlated as $X = c \cdot Y + b$; In epidemiological research these can be the amount of physical activity (X) and certain health outcomes (Y). While one might conclude that physical activity directly influences these outcomes, lifestyle factors such as diet or overall health awareness (Z) could be confounding variables that affect both physical activity and health.

The true causality relationship between these links could be defined as $X = 2 \cdot Z + \epsilon_X$ and $Y = 3 \cdot Z + \epsilon_Y$, where Z, ϵ_X and ϵ_Y are i.i.d. random variables $\sim N(0, 1)$.

Within this research the causality links can always be modeled as a graph structure, Figure 4 represents the causal graph of these three variables.

We can easily prove that X and Y are linearly correlated even if they are causally independent by isolating Z in the second equation and substituting it in the first. To assess if there is a real underlying causality between our two variables without knowing the true equations that define them we can condition them on Z and see if they are still independent, more formally we test if $X \perp\!\!\!\perp Y \mid Z$, where $\perp\!\!\!\perp$ indicates independence. These can be done by computing the two linear regressions $Z \to X$ and $Z \to Y$, and the residuals of each, r_X and r_Y. Finally we check if these are still correlated. If the regression has enough samples it easily approximate the true coefficients of Z, 2 and 3 respectively, and in the residuals it remains the information of X and Y that is not explained by Z which in this case corresponds to ϵ_X and ϵ_Y, which are not correlated, proving the absence of a causal link between X and Y.

In the next section we'll see how this concept has been used to work with time-dependent data, and how recent research deals with the main and common assumption that has been used in the past: the stationarity of causal relationships, meaning that they don't change through time. The main focus of this report is on how we can account for time-varying causal relationships, as these are omnipresent in large variety of domains, as the ones already cited or follower-leader animal dynamics[9] and macroeconomics[8].

1.1 Contribution

- Researched and understood the topic through scientific articles of the field.
- Developed a Python implenentation to procedurally generate the synthetic data.
- Developed a Python implenentation to fit the linear coefficients between the nodes of a given causal graphs.
- Reproduced part of the experiments of the original paper.
- Developed the structure and Python implementation for the MCI-NN algorithm introduced here.
- Run more experiments to test the MCI-NN algorithm.
- Written this report.

2 Related Work

Different methods for causal discovery have been proposed in the history of the field. Runge et al. [13] give an overview of all the causal discovery frameworks. Back in 1969 Granger[2] addressed causal discovery quantitatively using prediction, while more recently this has been explored also from the perspective of nonlinear dynamics too [15]. Nonetheless, one of the most frequently adopted framework is constraint-based causal discovery. The latter is rooted on the concept of conditional independence that we introduced earlier, and the PC algorithm[1] that iteratively checks conditional-independence between all pairs of variables.

Further research tries to account for non-stationary time-series by modeling the varying continuous background variables that act as confounders. As instance, Malinsky and Spirtes[6] model non-stationarity as (continuous) stochastic trends in a linear autoregressive framework. Zhang et al.[17] have a similar approach in the context of constraint-based framework, detecting variables whose statistical relationship change through time. Further research addresses non-stationary time-series with a score-based framework, that approaches causal discovery as an unsupervised task of scoring each possible causal link to best represent the time series(minimizing the loss). Recently, Faruque et al.[1] developed a method in this direction that make us of Causal CNNs.

Finally, Saggioro et al. [14], model time-varying relationships as stationary regimes, which maximize expleinability of the uncovered regimes and model parsimony, making use of the PCMCI algorithm detailed below.

3 Background

The PCMCI algorithm, introduced by[12] has shown impressive potential in finding the causal connections of time-dependent variables from observational time-series. In fact, it deals effectively with one of the biggest enemy of causal discovery, high dimensionality, as shown in [12]. On the other hand, one of the main assumption of this method is that the causal structure between the variables is static, and that the strength, the sign, or the direction of the causal effects doesn't change through time. This is problematic since it has been shown in many domains that the causality between temporal observational variables doesn't stay fixed.

This is why Saggioro et. al [14], try to improve the PCMCI method to account for multiple stationary regimes with their research, which I will try here to reproduce and

[1] See [11] for a clear explanation.

extend further. Fig. 5 illustrates a simple motivating example of two variables with a different causal graph per regime.

In[14], the durations and distribution in time of the regimes are learned automatically, given a fixed number of regimes N_K within a two-step process.

At each iteration q:

- 1. the algorithm learns the causal networks given the current regime assignment.
- 2. The regimes are re-assigned to minimize the loss on the observational data given the the causal network generated in the step above. Since every regime is a binary mask on the time-steps with a constraint on the number of switches N_C, which is set based on expert knowledge, the regime assignment is optimized with linear programming.

This two-steps algorithm it's reported to reach convergences after 10-20 iterations[14]. Theorem 2.1 of [3] and Sec. II B in [7] show that these types of algorithms monotonously decrease the value of the cost functional.

N_K is usually set to low numbers as we want to be able to explain each different regime and contain the number total number of learned parameters of this method, as it has to learn a different causal graph for each regime).

The authors test their method proving its efficacy in a number of scenarios, with both synthetic and real data, but always assuming a linear link for each time lag. Meaning that a variable X could causally affect a variable Y with a time lag τ only as a linear regressor of the latter as: $Y_t = \theta X_{t-\tau} + \epsilon_X, \epsilon_X \sim N(0,1)$. After reproducing the results of the paper, I will propose a different method to account for different regimes representing the link with a non-linear temporal network such as RNNs or LTSMs. These can take into account the joint effect of multiple time lags simultaneously and represent different regimes through their non-linear nature with a single causal graph.

4 Research Reproduction

Within this work I'm going to reproduce part of the experiments carried over the synthetic dataset and the final experiment on the real meteorological data in Saggioro et al.[14].

The synthetic experiments for this paper were carried out to test the Regime-PCMCI capabilities to uncover the true causal graphs, link strengths and regimes that generated the data, which, on the other hand, can't be known for real data, but only estimated by experts.

The synthetic data was generated following the main assumptions of the Regime-PCMCI method: firstly, causal relationships are stationary within the same regime and secondly, the nature of the causal links is known, which in this case are generated with a linear stochastic process, as anticipated above. The latter is defined by the causal graph for every regime and, more formally, the value of a variable x^j is obtained as:

$$x_t^j = \sum_{k=1}^{N_K} \{\gamma_k(t)\}^{\text{ref}} \sum_{x_{t-\tau}^i \in \mathcal{P}_k^j} \left\{\Phi_k^j(i,\tau)\right\}^{\text{ref}} x_{t-\tau}^i + \eta_t^j, \eta_t^j \sim \mathcal{N}(0, \{\sigma^2\}^{\text{ref}}) \tag{1}$$

Where the summation on the left accounts for the ground-truth(ref) regime assignment $\{\gamma_k(r)\}^{\text{ref}}$, which has a value of 1 if regime k spans over the time step t and 0 otherwise. $\mathcal{P}_k^j$ is the set of the causal parents of j, meaning all the variables that have a direct causal link towards j, which can include j to account for auto-correlation. The ground-truth

coefficients of these linear links are $\left\{\Phi_k^j(i,\tau)\right\}^{\text{ref}}$, where i denotes the parent node of j, k the regime and τ the time lag of the link. Finally, η_t^j is Gaussian noise with $\sigma^{2^{\text{ref}}} = 1$.

As already mentioned Regime-PCMCI is constitute by a two-step iterative optimization. In the first step, after we compute the causal structure $\{P_k\}^q$, where q is the iteration number, through PCMCI, we then estimate the linear coefficients with a linear regression between the causal parents $\{P_k^j\}$ of j and X^j itself:

$$\left\{\Phi_k^j(i,\tau)\right\}^{(q)} = \arg\min \left\| x_t^j - \sum_{x_{t-\tau}^i \in \mathcal{P}_k^{j,(q)}} \left\{\Phi_k^j(i,\tau)\right\} x_{t-\tau}^i \right\|_2^2 \tag{2}$$

This will reconstruct the coefficients $\left\{\Phi_k^j(i,\tau)\right\}^{\text{ref}}$ that generated the data in the first place, when $\{\mathcal{P}_k\}^q = \{\mathcal{P}_k\}^{\text{ref}}$

The second step of the algorithm minimizes the same loss but controlling for the regime assignment, this is defined by Eq.13 in Saggioro et al.[14] as:

$$\{\gamma_k(t)\}^{(q+1)} = \arg\min \sum_{k=1}^{N_K} \sum_{t=1}^{T} \gamma_k(t) \left\| \mathbf{x}_t - \left\{\hat{\mathbf{X}}_{k,t}\right\}^{(q)} \right\|_2^2 \tag{3}$$

Where $\left\{\hat{\mathbf{X}}_{k,t}\right\}^{(q)}$ is the data prediction[2] for every variable considering regime k at time-step t using the linear model with the coefficients computed by the first step(Eq. 4)

The paper reports synthetic experiments with two , three and ten variables. I will reproduce only two of the experiments in the bi-dimensional case, namely the ones that are referred to as *Sign* X^1X^2, and *Arrow Directiion* in Saggioro et al.[14].

Moreover, I reproduce the results obtained on the meteorological data. This experiment assesses if the Regime-PCMCI can infer the well studied relationship[16] between the ENSO monthly index, which represents the temperature and atmospheric pressure over the Pacific ocean, and the AIR(All-India Rainfall) index, which measures the monthly average precipitations over India.

4.1 Experiments & Results

Table 1 in the Appendix displays the configuration of the coefficients for the linear model that generates the data defined in Eq. 1.

Results on synthetic data We have two regimes: $k = 1$ and $k = 2$. In *Arrow direction* the causal link between the two variables X^1 and X^2 switches direction in the second regime, while the coefficient for the link $(0, 8)$ stays the same. On the other hand in *Sign* X^1X^2 the direction of the arrow doesn't change, but the sign of the link does. We also define auto-correlation of $0, 2$ for both variables and experiments. The data is generated for a total of $T = 3000$ time points, and each regime lasts a random value between 70 and 100 time-steps before we switch it, totaling to a number $N_C = 34$ of

[2] It's important to note that the linear model has already see this data-point as it is trained on the full dataset. Rather than validating and tuning the best model for prediction, the focus of this research is on finding the causal structure that best explains our data.

switches that is passed to Regime-PCMCI as a constraint for the regime optimization defined in Eq. 4. In the original paper the experiments were run for a number of $N_R = 100$ realizations of the synthetic datasets, here $N_R = 1$ due to time and computational constraints. Nonetheless, The two step iteration is capped to $N_Q = 20$ and the entire algorithm is re-run with $N_A = 100$ annealing steps as in the original paper, which allows to escape local optima, since the regime-optimization function defined in 4 is not convex.

Figure 1 and 2 show results for the *Sign* X^1X^2 and *Arrow direction* experiments respectively. We can see how the ground-truth graph is reconstructed in both experiments. This is reflected in the accuracy of the regime reconstruction. For the latter we obtain an error $\Delta\gamma\% = 1.86$ of and $\Delta\gamma\% = 4.36$ respectively, while we have a relative error of $\Delta\Phi\% = 4.61$ and $\Delta\Phi\% = 4.36$ on the coefficients for the sign and arrow direction experiments respectively. This is similar to the averaged results obtained over $N_R = 100$ realization in Saggioro et al. ,where we have $\Delta\gamma\% = 3$ for both experiments and $\Delta\Phi\% = 7$ and $\Delta\Phi\% = 9$ for the sign and arrow direction experiments respectively.

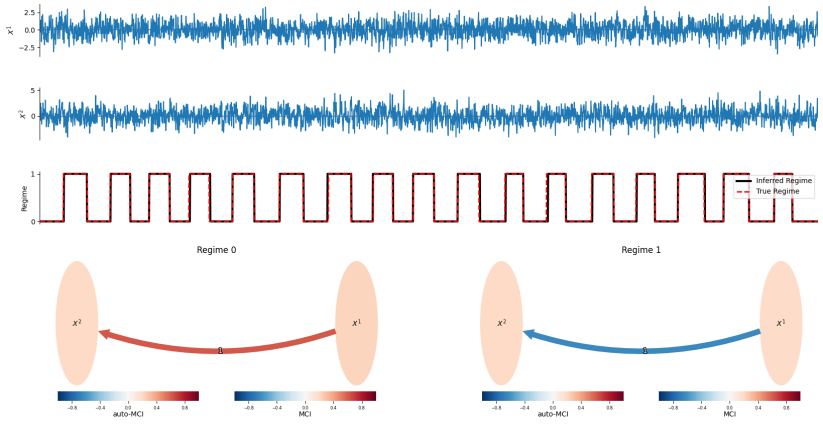


Fig. 1: Regimes assignments and causal graphs found for the *Sign* X^1X^2 experiment, the color of the links indicate the value of the MCI test statistic, while the color of the node represents the Auto-correlation MCI test statistic. The number on the arrow is the lag of the strongest link between the nodes.

Results on meteorological data Finally, I run Regime-PCMCI on the meteorological data[3] to recover the know interaction between the ENSO(El Niño and the Southern Oscillation) index and The AIR index. In particular we use the Nino3.4 index, which is a measure of the difference in Sea Surface Temperature(SST) in the longitude range of 170°W to 120°W [10]. In fact, there is a studied link with droughts during the summer monsoon season and the the ENSO index, more specifically they have a higher probability of happening in an *El Niño* event, during which sea temperatures are higher than normal for an extended period(Positive ENSO)[16, 5].

Unfortunately, using Regime-PCMCI I don't manage to recover this link using the data from 1871 to 2016 and $N_K = 2$, $N_C = 292$,$N_A = 100$ and $N_Q = 100$ as it's done in

[3] Openly available in the KNMI Climate Explorer at https://climexp.knmi.nl/.

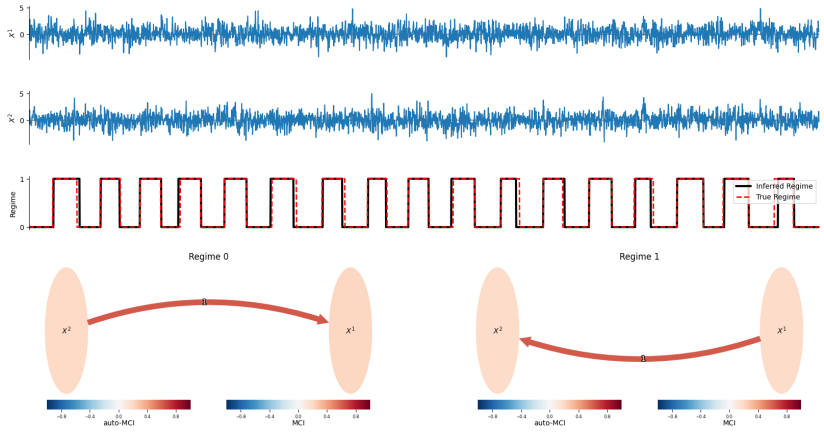
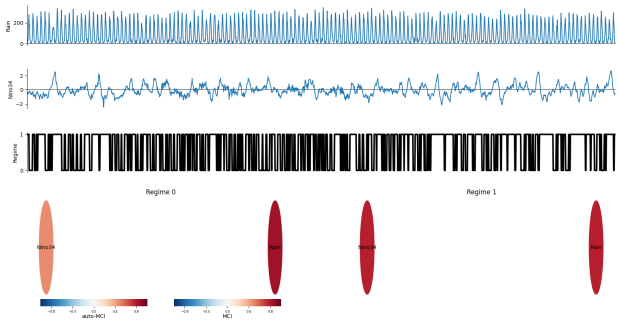
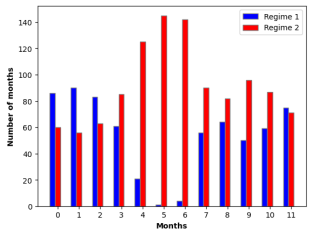


Fig. 2: Regimes assignments and causal graphs found for the *Arrow direction* experiment, the color of the links indicate the value of the MCI test statistic, while the color of the node represents the Auto-correlation MCI test statistic.

Saggioro et al. This can be observed in Figure 3, both regimes do not have a link between the two observational variables. In the original paper the algorithm finds a link between the two variables ($ENSO \rightarrow AIR$) with a strength of -0.4 for the regime that is most frequent in summer and no link in the other regime. On the other, hand the two inferred regimes in my reproduction are seasonal as shown in Fig. 3(b), meaning that one regime is more frequent in summer (Monsoon season).



(a) Recovered regime assignments and graphs structure between the two regimes, in this case no links are found between the nodes for both regimes.



(b) Bar plot that shows the seasonality of the regimes distribution over the twelve months.

Fig. 3: Results for the real data experiment, studying the causal interaction between the ENSO(Nino3.4) index and the AIR(All-India Rainfall) index

5 Research extension

In this section I propose an alternative model, to implement a regime-dependent evolving causal network.

The main intuition behind this model is that non-linear causal link that consider the joint effect of multiple time-lags could already learn a representation of the changing regime within the given link, similarly to what is done in [1]. In this way the regime is not modeled as a latent variable as in [6, 17], but the time varying nature of the link is learned through gradient descent. The non-linear links between the variables are represented as temporal neural networks, which are implemented in the form of simple RNNs or LSTM, more generally, the type of causal model used $\mathbf{M}$ is seen as a Hyper-parameter of the method. This model takes as input the the last τ time steps of a variable X, and learns to predict the value of the next time step of a variable Y. More formally, $\mathbf{M}([X_{t-\tau}, ..., X_{t-2}, X_{t-1}]) = \hat{Y}_t$.

Since NNs, as opposed to Linear regression, are very prone to over-fitting we set these networks to have few hidden nodes and we test the statistical correlation between X and Y given the model $\mathbf{M}$ as the R-squared: $R^2_{X,Y} = 1 - \frac{\sum_{t=s}^{T}(Y_t - \hat{Y}_t)^2}{\sum_{t=s}^{T}(Y_t - \hat{\mu}_Y)^2}$, on the validation set. Where s indicates the first time-step of the latter and T the total amount of time-steps considered, $\hat{\mu}_Y$ is the mean of Y computed on the training data. This is approach is adjacent to Granger causality[2], but one of the main difference is that before we test the correlation we condition on the causal parents of X and Y, $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ respectively. This is similar to what is done within the already mentioned PCMCI algorithm, and more generally, in the constraint-based causal discovery framework. In fact, we apply an algorithm analogue to PCMCI, the only difference is on the correlation measure on how the temporal graph is represented. In fact, if the regular PCMCI temporal graph comprehends a "copy" of every variable corresponding to each considered time-lag, here we consider only two causal nodes per variable: $X^j_{[t-\tau,..,t-1]}$ and $X^j_{[t]}$ which represent the past and the future of X^j respectively. With this structure we can apply the PC step iteratively conditioning, for every node j, $X^j_{[t]}$ on $X^i_{[t-\tau,..,t-1]} \in \mathcal{P}_c(X^j_{[t]}) \setminus \mathcal{P}_c(X^l_{[t-\tau,..,t-1]})$, where i belongs to the c highest correlated parents of the previous iteration. $c = 0$ in the first iteration(empty conditioning set) and increases by one on every iteration. The conditioning is done by substituting$X^j_{[t]}$ with the residuals generated with the multivariate model: $r^j_{[t]} = X^j_{[t]} - \mathbf{M}(\mathcal{P}_c(X^j_{[t]}) \setminus \mathcal{P}_c(X^l_{[t]})$. Finally, the strength of the link between the conditioned node j and l is calculated as the R-squared $R^2_{X^l_{[t-\tau,..,t-1]}, r^j_{[t]}}$, if this value is $> \alpha_R$, we'll keep l in the considered parents of the next iteration ($\mathcal{P}_{c+1}(X^j_{[t]})$), otherwise if $R^2 \leq \alpha_R$ we consider j and l as conditionally independent. The MCI, refines the discovered graph and work by conditioning both l and j on their parents before testing for the correlation.

One evident issue with this formulation is that it could be computationally expensive as the number of dimensions N of our data increase, as its time-complexity is $O(C \cdot N^2 \cdot \frac{T}{B} \cdot E)$, where C are the number of PC iterations (max cond. set size), B is the batch size used for gradient descent and E is the number of epochs of each model training. Nonetheless, the number of variables usually considered in causal discovery is low, as too many can increase the False Positives of the conditional independence tests. Finally, the computational complexity of Saggioro et al. is arguably even higher as linear regression has a complexity very similar to gradient descent for simple networks (being the simplest case with no hidden layer a form linear-regression itself), and they re-execute their algorithm

a numberN_A of times to escape local optima on the regime learning part, the latter also adds computational overhead.

5.1 Experiments

I'm going to run the experiments for the proposete method,that I'll refer to as MCI-RNN, over the same datasets that we covered for the Regime-PCMCI to qualitatively compare the two methods. Due to time constraints I will report here only the combination of values that worked to reconstruct the true causal graphs. Further research can extend the analysis by isolating the effect of each parameter, while controlling for the others.

5.2 Experiments & Results

Table 2 of the appendix shows the parameters that correctly reconstructed the causal graph Both in the PC and MCI step of the algorithm for the $Sign\ X^1\ X^2$ dataset with a significance level $\alpha_R = 0.001$, while Tab. 4 and Tab. 5 displays the resulting R-squared for each combination of nodes after the PC phase and the MCI phase respectively.

Using the same parameters and significance level we manage to find the "true" graph in the *Arrow direction* experiment too. Table 6 and 7 show the resulting R^2, in this case we have a link in both direction, one of the two, should intuitively be activated by the LSTM model when we are in the regime in which that given link is present, and the other inhibited. Assuming that the model is able to distinguish the two regimes given the history passed for in a given time point.

Finally, we run the experiment on the meteorological data with the hyper-parameters defined in Tab. 3. I had to increase the epochs to 10 as the network couldn't learn the time series properly, this could be explained by the fact that the series it's shorter($T = 1752$) than the synthetic data case. I also augmented the hidden dimension to 100 as it was helping to get better performance on the held out validation sets even if one would expect the network to overfit with such a short time-series. Table 9 and 8 show the resulting R-squared tables.

6 Discussion & Conclusion

The difference in results for the real data experiment could be due to the fact that Saggioro et al. average the networks over the top 13 annealings, while I just get the causal structure for the best performing annealing, which is how Regime-PCMCI algorithm is defined. Additionally, it's not clear how they perform this averaging as they would need to match the regimes between the annealings, which makes this operation seem as a way to force their method to capture this observed causal phenomena. Furthermore, the seasonality in the regimes that both their experiment and mine observe could just be simply motivated by the fact that the AIR index is itself seasonal, as shown in Fig. 3(a).

With regards to the experiments on MCI-RNN we can argue that is unfair to compare this method on synthetic datasets generated specifically to be in accordance with Regime-PCMCI assumptions, first of all the stationary and linear regimes. Surely R-PCMCI has and advantage in terms of inductive bias. Despite that MCI-RNN manages to find the true graphs for the synthetic datasets, while we can see from the R-squared graphs that the PC step alone seems to have a higher detection power, while the MCI step could help to decrease false positives.

Unfortunately, over the meteorological data MCI-RNN doesn't manage to capture the observed link, but infers causality between AIR and ENSO in the opposite and theoretically wrong direction, this could be due to a myriad of factors, as well as some bug in the code or a too short time-series, further investigation is required.

I think MCI-RNN goes in a direction that could be interesting to explore further as it has an essential structure that bridges between different frameworks of causal discovery(e.g. Granger and constraint-based discovery) and machine learning. It would be interesting to test its predictive power and performances in a higher-dimensional case too.

To conclude, I think this research has opened up more questions than answers given, but the latter are the main motors of scientific studies in the quest for causality.

References

1. Faruque, O. *et al.*: TS-CausalNN: Learning Temporal Causal Relations from Non-linear Non-stationary Time Series Data. arXiv preprint arXiv:2404.01466 (2024)
2. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society (1969)
3. Horenko, I.: Finite element approach to clustering of multidimensional time series. SIAM Journal on Scientific Computing **32**(1), 62–83 (2010)
4. Huettel, S., Song, A., McCarthy, G.: Functional Magnetic Resonance Imaging. Oxford University Press, Incorporated (2009)
5. International Research Institute for Climate and Society, ENSO and All-India Rainfall, (2025). `https://iridl.ldeo.columbia.edu/maproom/ENSO/Climate_Impacts/India_Rainfall.html`. Accessed: 2025-02-03.
6. Malinsky, D., Spirtes, P.: Learning the structure of a nonstationary vector autoregression. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 2986–2994 (2019)
7. Metzner, P., Putzig, L., Horenko, I.: Analysis of persistent nonstationary time series and applications. Communications in Applied Mathematics and Computational Science **7**(2), 175–229 (2012)
8. Muradoglu, G., Taskin, F., Bigan, I.: Causality between stock returns and macroeconomic variables in emerging markets. Russian & East European Finance and Trade **36**(6), 33–53 (2000)
9. Nakayama, S. *et al.*: Initiative, personality and leadership in pairs of foraging fish. PLoS One **7**(5), e36606 (2012)
10. National Centers for Environmental Information, El Niño / Southern Oscillation (ENSO) | Technical Discussion, (2025). `https://www.ncei.noaa.gov/access/monitoring/enso/technical-discussion`. Accessed: 2025-02-03.
11. Runge, J.: Introducing Conditional Independence and Causal Discovery, Medium article, accessed on 2025-02-02. (2023). `https://medium.com/causality-in-data-science/introducing-conditional-independence-and-causal-discovery-77919db6159c`
12. Runge, J. *et al.*: Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances **5**(11), eaau4996 (2019)
13. Runge, J. *et al.*: Inferring causation from time series in Earth system sciences. Nature communications **10**(1), 2553 (2019)
14. Saggioro, E. *et al.*: Reconstructing regime-dependent causal relationships from observational time series. Chaos: An Interdisciplinary Journal of Nonlinear Science **30**(11) (2020)
15. Sugihara, G. *et al.*: Detecting causality in complex ecosystems. science **338**(6106), 496–500 (2012)
16. Webster, P.J., Palmer, T.N.: The past and the future of El Niño. Nature **390**(6660), 562–564 (1997). `https://doi.org/10.1038/37499`

17. Zhang, K. *et al.*: Causal Discovery from Nonstationary/Heterogeneous Data: Skeleton Estimation and Orientation Determination. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 1347–1353 (2017). https://doi.org/10.24963/ijcai.2017/187

A Synthetic data generation

Table 1: Artificial model configurations for two low-dimensional experiments with $N_K = 2$ underlying regimes. $\{\Phi_k^j(i,\tau)\}^{\text{ref}}$ indicates the coefficient between node i and j at time lag τ for regime k.

Experiment	**Graphs**		**Linear coefficients**	
	$k=1$	$k=2$	$\{\Phi_1^j(i,\tau)\}^{\text{ref}}$	$\{\Phi_2^j(i,\tau)\}^{\text{ref}}$
Arrow direction	$X^1 \to X^2$	$X^1 \leftarrow X^2$	$\{\Phi_1^2(1,1)\}^{\text{ref}} = 0.8$	$\{\Phi_1^2(2,1)\}^{\text{ref}} = 0.8$
			$\{\Phi_1^1(1,1)\}^{\text{ref}} = 0.2$	$\{\Phi_1^1(1,1)\}^{\text{ref}} = 0.2$
			$\{\Phi_1^2(2,1)\}^{\text{ref}} = 0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}} = 0.2$
Sign X^1X^2	$X^1 \xrightarrow{\lvert a\rvert} X^2$	$X^1 \xrightarrow{-\lvert a\rvert} X^2$	$\{\Phi_1^2(1,1)\}^{\text{ref}} = 0.8$	$\{\Phi_2^1(1,1)\}^{\text{ref}} = -0.8$
			$\{\Phi_1^1(1,1)\}^{\text{ref}} = 0.2$	$\{\Phi_2^1(1,1)\}^{\text{ref}} = 0.2$
			$\{\Phi_1^2(2,1)\}^{\text{ref}} = 0.2$	$\{\Phi_2^2(2,1)\}^{\text{ref}} = 0.2$

B MCI-RNN experiments

B.1 Parameters

Symbol	**Name**	**Value**
τ	lookback	10
M	Network type type	`LSTM`
h	hidden dimension	10
α_R	significance threshold	0.001
s_p	training split (ratio)	0.9
C	max condition set size	2
E	number of epochs	5
B	batch size	16

Table 2: Parameters configuration for MCI-RNN model experiments on the synthetic data.

Symbol	Name	Value
τ	lookback	10
M	Network type type	`LSTM`
h	hidden dimension	100
α_R	significance threshold	0.001
s_p	training split (ratio)	0.9
C	max condition set size	2
E	number of epochs	10
B	batch size	16

Table 3: Parameters configuration for MCI-RNN model experiments on the meteorological data.

B.2 Resulting R-squared

PC parent discovery	**X**=X^1	**X**=X^2
Y=X^1	0.0638	0.0000
Y=X^2	0.0080	0.1264

Table 4: R^2 matrix from the final parent discovery iteration over the *Sign* X^1 X^2 dataset.

MCI Step	**X**=X^1	**X**=X^2
Y=X^1	0.060	0.000
Y=X^2	0.005	0.116

Table 5: Final R^2 for each $(X \to Y)$ link after the MCI step over the *Sign* X^1 X^2 dataset.

PC Parent discovery	**X**=X^1	**X**=X^2
Y=X^1	0.0355	0.1825
Y=X^2	0.0260	0.0399

Table 6: R^2 matrix from the final parent discovery iteration over the *Arrow direction* dataset.

MCI Step	**X**=X^1	**X**=X^2
Y=X^1	0.048	0.199
Y=X^2	0.020	0.017

Table 7: Final R^2 for each $(X \to Y)$ link after the MCI step over the *Arrow direction* dataset.

PC parent discover	**X=AIR**	**X=ENSO**
Y=AIR	0.26797	0.00000
Y=ENSO	0.01247	0.75167

Table 8: R^2 matrix after the final parent-discovery iteration over the meteorological data

MCI Step	**X=AIR**	**X=ENSO**
Y=AIR	0.305	0.000
Y=ENSO	0.007	0.568

Table 9: Final R^2 for each $(X$ß$Y)$ link during the MCI step over the meteorological data

C Additional figures

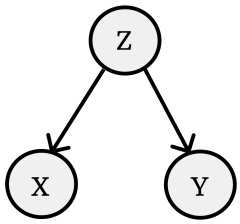


Fig. 4: A simple example of a causal graph. Z acts as a common driver of X and Y

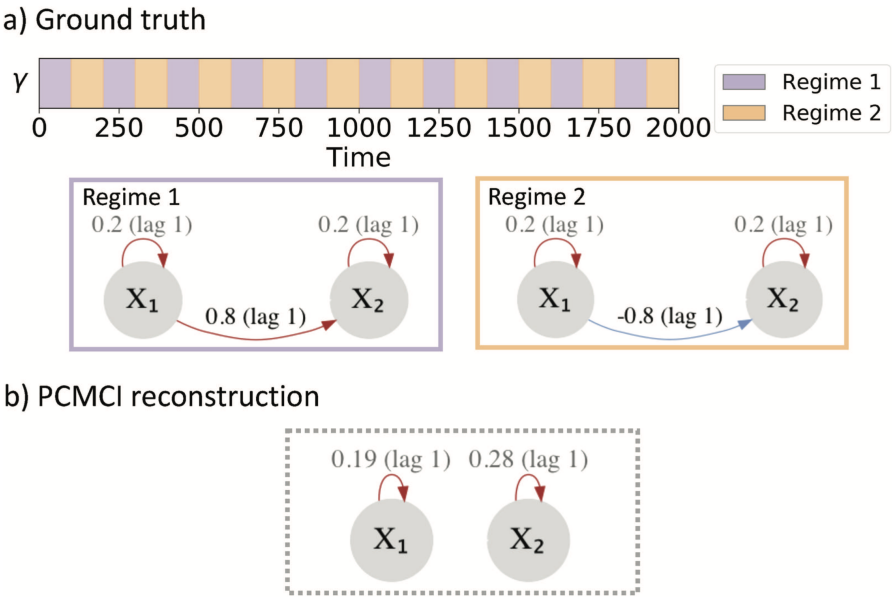


Fig. 5: Motivating example.(a)Ground through distribution of the two regimes, and corresponding causal graphs. The links are labeled with the associated linear coefficient $\Phi_k^j(i,\tau)$ and lag τ . The sign of the coefficient is highlighted by the color (red for positive, blue for negative) (b) Network reconstruction with PCMCI estimated from the whole time series, i.e., if links are wrongly assumed to be stationary. Reprinted from Saggioro et al.[14]