Causal discovery for time series with latent confounders

Christian Reiser

University of Stuttgart

Stuttgart, Germany

christian.reiser@insightme.org

Abstract

Reconstructing the causal relationships behind the phenomena we observe is a fundamental challenge in all areas of science. Discovering causal relationships through experiments is often infeasible, unethical, or expensive in complex systems. However, increases in computational power allow us to process the ever-growing amount of data that modern science generates. This development has led to an emerging interest in the problem of causal discovery from observational data. This work evaluates the LPCMCI algorithm, which aims to find generators compatible with a multi-dimensional, highly autocorrelated time series while some variables are unobserved. Our numerical results find that LPCMCI performs best on auto-dependencies, then contemporaneous dependencies, and struggles most with lagged dependencies. The source code of this project is available online[1].

Index Terms

causal discovery, statistics

I. INTRODUCTION

Reconstructing the causal relationships behind the phenomena we observe is a fundamental challenge in all areas of science. It helps to build physical models and predict the effect of interventions [1]. The goal is to distinguish direct and indirect dependencies between variables, determine the dependencies' directionality, and identify common drivers. The standard approach is to learn causal relationships through conducting experiments. However, interventions are often infeasible, unethical, or expensive in complex systems such as earth system science or healthcare [2]. Meanwhile, modern science generates a growing amount of data, which we can process due to increases in computational power. The availability of data and abundance of computing power to process it has led to an emerging interest in the problem of causal discovery from observational data [3].

In our work, we generate highly auto-correlated multi-dimensional time-series with latent variables that have similar properties to a dataset used in a study for mood prediction [4]. We briefly review methods that aim to reconstruct generators compatible with the observed data and elaborate on and evaluate a conditional independence constraint-based algorithm called LPCMCI [5] in more detail.

II. PROBLEM DESCRIPTION AND NOTATION

A. Preliminaries

Consider a multivariate state generator that we want to reconstruct from data. We assume it is stationary and follows a vector-auto-regressive process[2], similar as used in [5], and described by the structural causal model (SCM)

$$V_t^j = f_j \left(pa \left(V_t^j \right), \eta_t^j \right) \quad \text{with } j = 1, \ldots, \tilde{N}, \tag{1}$$

generating a multivariate time series $\mathbf{V}^j = \left(V_t^j, V_{t-1}^j, \ldots \right)$ for $j = 1, \ldots, \tilde{N}$. The functions f_i describe mechanisms in the physical world which depend on a set of causal parents $pa \left(V_t^j \right) \subseteq (\mathbf{V}_t, \mathbf{V}_{t-1}, \ldots, \mathbf{V}_{t-p_{ts}})$ and jointly independent noise variables η_t^j. p_{ts} represents the maximal time lag between cause and effect. The causal dependency of a pair of variables $\left(V_{t-\tau}^i, V_t^j \right)$, with the time lag $\tau \geq 0$ is the same as for all time shifted pairs $\left(V_{t'-\tau}^i, V_{t'}^j \right)$ due to *stationarity*. We allow for *contemporaneous* ($\tau = 0$) effects because, in many applications, the measurement frequency is often slower than the time scale of the causal processes. Although there are contemporaneous effects, we still assume *acyclicity*. The SCM has *latent variables*, because many real-world applications often only observe a subset $\mathbf{X} = \left\{ \mathbf{X}^1, \ldots, \mathbf{X}^N \right\} \subseteq \mathbf{V} = \left\{ \mathbf{V}^1, \mathbf{V}^2, \ldots \right\}$ of

[1]Sourcecode: https://github.com/christianreiser/correlate/blob/master/causal_discovery/LPCMCI/compute_experiments.py

[2]Auto-regressive process is a model, where the output variable depends only on its own lagged values and on a stochastic noise term. A *vector*-auto-regressive process extends the auto-regressive model to multiple dimensions. As explained later, the output of our vector-auto-regressive process additionally depends on its current values.

the time series with $N \leq \tilde{N}$. However, we assume that no *selection variables* determine which measurements to include in or exclude from the data sample. Breaking this assumption can lead to *selection bias*. Furthermore, we assume *faithfulness*, meaning that conditional independence (CI) in the observed distribution $P(\mathbf{X})$ is due to the causal structure of the underlying process instead of mere chance [5].

B. Implications of latent variables

Since the causal model we want to reconstruct from observational data has unobserved variables, we would like to search over the space of directed acyclic graphs (DAGs) with latent variables. However, as we do not restrict the number of unobserved variables, we would search among an infeasible number of DAGs.

Instead, we represent each DAG with latent variables compatible with the observed data by a *maximal ancestral graph (MAG)* over the observed variables, illustrating confounded variables with bi-directed edges ($\leftrightarrow$) [6].

MAGs have different semantics than DAGs. In a MAG, the tail (-) of an arrow, as in A $\rightarrow$ B, says that A is an ancestor of B. Arrowheads ($<, >$), like the one pointing on B, specifies that B is not an ancestor of A. Furthermore, A and B might be confounded.

A MAG represents the uncertainty that a variable could have an arrowhead ($<$ or $>$) or tail (-) with a circle-head ($\circ$). For example, the link C $\circ\!\rightarrow$ D says that D is not an ancestor of C, but we do not know if A is an ancestor of B or not. This case leaves us with three possibilities. First, there could be a directed path from C to D, second, the dependency could only be due to confounding, or third, it could be a combination of both.

Since arrowheads declare the variable next to it as a non-ancestor, bidirectional edges $\leftrightarrow$ exclude both variables from being the ancestor of the other variable and thus propose confounding between the two. CI-based methods can not differentiate all MAGs. *Partially ancestral graphs (PAGs)* represent the Markov equivalence class that summarizes all different MAGs with the same set of conditional independencies [7]. This relationship between MAGs and a PAG is analogous to DAGs and a PDAG [7].

The goal is to find the PAG that satisfies the conditional independence information over the observed variables.

III. RELATED METHODS

Granger causality can infer the causal structure of time series. Intuitively we say that X Granger-causes Y when the prediction of Y from its past is improved by accounting for X's past. Then Y has to be dependent on the past of X given its past:

$$X \text{ Granger-causes } Y :\Longleftrightarrow Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \mid Y_{\text{past}(t)} \tag{2}$$

[3]. [8] extends Granger causality to a multivariate setting. However, it can be misleading with the presence of instantaneous effects [9] and requires that all variables are observed [8].

FullCI determines the absence of a link between one variable $X_{t-\tau}^i$ with time lag τ and another variable X_t^j by testing for their independence while conditioning on the past of all other variables $\mathbf{X}_t^- \setminus \{X_{t-\tau}^i\}$. Formally that is $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathbf{X}_t^- \setminus \{X_{t-\tau}^i\}$ [10]. A problem is that high dimensionality of the conditioning set decreases the detection power of FullCI [10]. Usually the researcher decreases dimensionality by restricting the maximal time-lag τ_{max} through background knowledge or sets it to the largest time-lag with significant unconditional dependence [10]. It is actually sufficient to condition only on the parents of the two variables [11], which allows decreasing the dimensionality of the conditioning set further, especially if the true graph is sparse.

The PC algorithm is well known to exploit sparsity by iteratively using small conditioning sets [10]. However, when the size of the conditioning set is very limited, chances increase that true parents are not in the conditioning sets and false-positive links tend to remain [10].

The PCMCI first runs a version of the PC algorithm adapted for time series, which removes most false links very efficiently, with the trade-off that some false positives remain [10]. In the second step, it tests for momentary conditional independence

$$\text{MCI} : X_{t-\tau}^i \not\perp\!\!\!\perp X_t^j \mid \widehat{\mathcal{P}}\left(X_t^j\right) \setminus \{X_{t-\tau}^i\}, \widehat{\mathcal{P}}\left(X_{t-\tau}^i\right) \tag{3}$$

to remove the remaining false links [10]. According to the author, this resulted in faster runtime and increased detection power than FullCI [10]. PCMCI satisfies all of our requirements, except it assumes causal sufficiency, which means the absence of latent confounders.

An algorithm that allows for latent confounders is the Fast Causal Inference algorithm (FCI) [12]. It is constraint-based and outputs a PAG which represents possible MAGs. The tsFCI algorithm adapts FCI to time-series [13] and satisfies all of our requirements. It extends FCI by additionally exploiting that causes precede effects. Furthermore, it assumes stationarity which allows applying a found momentary dependency between a pair of variables to be the same for all other shifted pairs.

So war we discussed methods that restrict the set of possible graphs via conditional independence and time ordering. The major downside of conditional independence-based methods is that they can not differentiate graphs within the Markov equivalence class. Linear models with non-gaussian additive noise (LiNGAMs) allow differentiation within the Markov equivalence class: If the true LiNGAM is

$$Y = \alpha X + N_Y, \quad N_Y \perp\!\!\!\perp X \tag{4}$$

with continuous random variables X, Y, and N_Y, then there exist no $\beta \in \mathbb{R}$ and random variable N_X for the LiNGAM

$$X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y \tag{5}$$

[3]. Note that this statement does not hold if the noise is Gaussian. One can find proof of this statement in Appendix C.1 of [3].

TS-LiNGAM is an algorithm based on LiNGAM but adapted for time series [14]. Like LiNGAM, it assumes linear dependencies and non-Gaussian noise. Since it can identify structure within the Markov equivalence class without relying only on time-ordering, it may have an advantage in handling contemporaneous links. One downside of TS-LiNGAM is that hidden confounders may lead to erroneous results [15].

IV. LPCMCI

Detection power quantifies the probability of finding a true link between two nodes. The detection power of conditional independence based methods can be increased by

- increasing the number of samples in the dataset, which results in more reliable CI information,
- decreasing the dimensionality of the problem [10],
- increasing the causal effect size between the two variables,
- increasing the statistical significance level of the conditional independence test, and
- decreasing the size of the conditioning set until it consists only of the parents of the two variables [10].

In real-world applications, it is often possible to take more samples, decrease the problem's dimensionality by excluding irrelevant variables through expert knowledge, and increase the effect size by enhancing the signal-to-noise ratio of the measurements. However, these properties are of less interest when developing methods and are usually fixed through the dataset.

Usually, the researcher sets the significance level to the false positive rate he allows [5]. However, one can relax the statistical significance level of the conditional independence tests, which usually increases detection power. Nevertheless, the downside is that it increases the risk of detecting more spurious links. Furthermore, these spurious links in later steps can also decrease detection power. The reason is that later CI tests of true links might condition on these false links and reduce the CI scores of the true links. Each score that falls below the significance level α leads to the erroneous removal of a true link. More intuitively, false links can "explain away" the effect of true links and thus remove them. In our test case, we are less interested in keeping the false positive rate below a specific value. We rather optimize alpha to maximize the harmonic mean of precision and recall through cross-validation.

LPCMCI improves its performance by decreasing the size of the conditioning sets by discarding conditioning sets containing known non-ancestors of the two variables but including their known parents [5]. It is challenging to do so because the conditioning sets have to contain the parents of the two variables before the CI-tests are completed. This order is not the case for predecessor algorithms. They first conduct the CI test and then orient the links. Only the latter yields parenthoods. To overcome the challenge that ancestries and parentships are unknown, LPCMCI entangles CI tests with edge orientation to identify parentships. Each time the algorithm gains new knowledge about parent- or ancestorship, it updates its conditioning sets. After the algorithm converges once, it re-initializes the graph but keeps the identified parentships. It applies the same process again but extends the conditioning sets with the previously identified parents. The procedure should remove most spurious links during the second run and assign edgemarks to the remaining links. However, the resulting graph tends to contain false links between confounded variables. To remove them, the algorithm applies one more round of CI test and edge orientations with a modified rule to select conditioning sets [3] to identify confounded links [5].

V. EXPERIMENTS

A. Data-generation

The synthetic data generating SCM is similar to the one used in [5] and given by

$$V_t^j = a_j V_{t-1}^j + \sum_i c_i f_i \left(V_{t-\tau_i}^i \right) + \eta_t^j \quad \text{for} \quad t \in \{1, \dots, T\}, \quad j \in \{1, \dots, \tilde{N}\}. \tag{6}$$

---

[3]For more information on these modified conditioning sets, read about the $napds_t$ sets in definition S5 of [5].

The model consists of $\tilde{N} = 11$ variables V_t^j. Three of these variables are selected at random and hidden later on. Every variable V_t^j has a linear auto-dependency $a_j V_{t-1}^j$, with $a_j \sim \mathcal{U}(0.3, 0.6)$, meaning that the strength of the auto-dependency is drawn from a uniform distribution with a lower bound of 0.3 and an upper bound of 0.6. In total there are $\tilde{N} = 11$ randomly chosen variable pairs $(V_t^j, V_{t-\tau_i}^i)$ with non-zero linear cross-dependencies $f_i \sim \pm\mathcal{U}(0.2, 0.5)$. Otherwise $f_i = 0$, meaning that all other variables have no cross-dependency. 60% of links are instantaneous, meaning they have a time-lag of $\tau = 0$. The other 40% have a time-lag of $\tau = 1$. The noises $\eta^j \sim \mathcal{N}(0, \mathcal{U}[0.5, 2])$. Every SCM generates a time-series of length T = 500, but if a model happens to be non-stationary, it is redrawn. We set the significance level to $\alpha = 0.26$, which we optimized via cross-validation. During this hyperparameter search, we used a reduced number of SCM realizations of 50 to increase speed. For the final experiment, we use data from 4000 different SCMs.

B. Evaluation metrics

Note that we aim to find the PAG that satisfies all conditional independence information instead of the data-generating DAG. Therefore, we evaluate against the oracle PAG.

The evaluation metrics distinguish between auto-dependencies and cross-dependencies. Furthermore, cross-dependencies are separated into contemporaneous links and lagged links, whereas auto-dependencies always have a lag, as contemporaneous auto-dependencies would violate the assumption of acyclicity.

Furthermore, we distinguished dependencies on whether adjacencies and their correct edgemarks are detected or not. An edgemark can be either a head ("<" or ">"), tail("-"), circle ("∘"), or conflict ("x").

A conflict during edge orientation occurs when orientation rules of two variable triples with a common edgemark propose to orient the common edgemark as both tail and head [16]. The underlying problem would be insufficient sample size or at least one assumption violation [17].

We evaluate the recall and precision of the predicted PAG against the oracle PAG: $Recall = \frac{tp}{tp+fn}$ and $Precision = \frac{tp}{tp+fp}$, where tp stands for true positives, which is the number of edgemarks and adjacencies that are in the oracle PAG as well as the predicted PAG. fn stands for false negatives, the number of adjacencies or edgemarks in the oracle PAG but not in the predicted PAG. fp stands for false positives, the number of adjacencies or edgemarks in the predicted PAG but not in the oracle PAG.

Note that if the algorithm predicts the original data-generating DAG, the evaluation would still return a non-perfect score if the predicted DAG differs from the oracle PAG, even though the oracle PAG includes the predicted DAG.

C. Numerical Results

TABLE I
NUMERICAL RESULTS

TABLE II
PRECISION AND RECALL OF EDGEMARKS AND ADJACENCIES, WHICH ARE FURTHER DISTINGUISHED BETWEEN LAGGED, AUTO-DEPENDENT, AND CONTEMPORANEOUS LINKS.

		precision	recall
adjacency	lagged	0.38	0.39
	auto	1.00	1.00
	contemp	0.60	0.77
	total	0.62	0.67
edgemark	lagged	0.27	0.26
	auto	0.74	0.73
	contemp	0.34	0.44
	total	0.42	0.46

TABLE III
RATE OF CONFLICTING AND UNORIENTED EDGEMARKS, WHICH ARE FURTHER DISTINGUISHED BETWEEN LAGGED, AUTO-DEPENDENT, AND CONTEMPORANEOUS LINKS.

		conflict rate	unorianted rate
conflict rate	lagged	0.00	0.14
	auto	0.00	0.17
	contemp	0.09	0.05
	total	0.03	0.13

We report the F_1-scores, which are the harmonic mean of precision and recall. The F_1-score of the total adjacency is 0.65. The F_1-score of the total edgemarks is 0.44. Recall and precision of edgemarks are lower than that of adjacencies. One reason for this result is that each adjacency-fn leads to two missing edgemarks, which increases the edgemark-fn by two. Also, vice versa, each adjacency-fp leads to an increase of edgemark-fp by two.

Auto-dependencies have the highest recall and precision. The reason is that the edgemark pointing to the node later in time always has an arrowhead because a variable of the future can never be an ancestor of the past. However, the prediction of auto-dependencies is not perfect due to the possibility that auto-dependencies are confounded. Therefore, false positives can occur when mismatching ∘ and "-" edgemarks at the other side of the arrow.

The algorithm struggles most to detect the lagged adjacencies and edgemarks. In our test case, conflicting edgemark proposals only occur when orienting contemporaneous dependencies. A higher conflict rate for contemporaneous links seems generally plausible, as time-order can not orient these links.

D. Example result

Figure 1(a) shows the data-generating DAG of eleven auto-dependent variables with eleven cross-dependencies. The number in the nodes represents the name of the variable. All auto-dependencies have a time delay of $\tau = 1$. A "1" written on a curved arrow indicates that the cross-dependency has a time-delay $\tau = 1$. Straight arrows without a number represent contemporaneous effects. The color of an edge or node represents the strength of the cross- or auto-dependency, respectively.

Figure 1(b) shows data of the eight observed variables after dropping the information of the three latent variables. Note that the information about the number of hidden variables is also removed and can not be used in the next steps.

(a) Original DAG of the data-generating process.
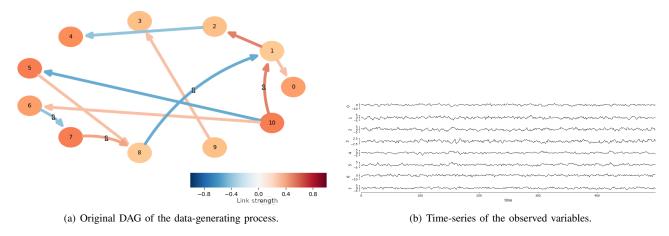

(b) Time-series of the observed variables.

Fig. 1. Screenshots of the data-generating DAG and its observed data.

Figure 2(a) shows the oracle PAG, which serves as ground truth. It satisfies all conditional independence information of the observed DAG perfectly. ○-edgemarks, e.g., as shown in the arrow from node 7 to node 8, represent the possibility that both a tail ("-") or arrowhead ("<",">") edgemark could exist. A bidirected, e.g., between nodes 1 and 8, indicates that neither of the two variables is a parent of the other, but they are confounded. The strength of all dependencies has a maximum value because generating the oracle PAG through conditional independence does not rely on the data but uses perfect conditional independence information of the data-generating DAG after removing the latent variables.

Figure 2(a) shows the predicted PAG. The colors represent the absolute link strength. It seems like there are many of the false positives have a weak absolute link strength.
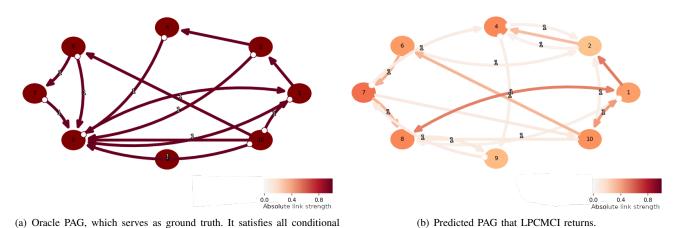

(a) Oracle PAG, which serves as ground truth. It satisfies all conditional independence information of the observed DAG perfectly.

(b) Predicted PAG that LPCMCI returns.

Fig. 2. Oracle and predicted PAG

VI. Limitations and future work

A measurement frequency slower than the causal processes could violate the assumption of acyclicity of the graph. Furthermore, the algorithm is limited to stationarity processes and assumes no selection bias in the data. In this work, we omitted background knowledge of (non-)ancestorships but planned to use it in future work as we expect that it increases performance.

The evaluation process uses the oracle PAG as ground truth, which in most cases differs from the data-generating DAG. One can see the difference between these two graphs by comparing Figure 1(a) and Figure 2(a). The downside of evaluating against the oracle PAG is that a perfect prediction of the data-generating DAG would usually not receive a perfect score.

Figure 3 shows that the predicted PAG (b) has many false positives with a weak link strength when evaluating against the data-generating DAG (a). Figure 3(c) shows the predicted PAG where weak links with a strength below 0.10 are removed. In this example of Figure 3 we observe that removing weak links decreases false positives without increasing false negatives. In future work, we want to investigate this phenomenon in more detail. Moreover, instead of removing weak links at the end, we also want to investigate if it would be helpful to iteratively lower the significance level α to remove false positives without inducing false negatives.

The algorithm only returns the absolute values of the predicted link strength, as one can see in Figure 2. The reason is that it computers the link strength via the p-value. We plan to differentiate between positive and negative link strengths in future work by using a distance correlation metric (e.g., Pearson correlation coefficient in the linear case).
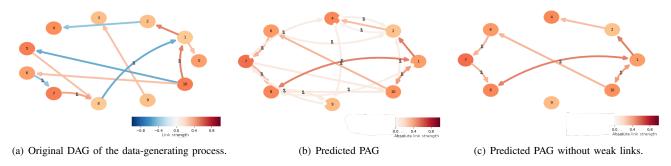

(a) Original DAG of the data-generating process. (b) Predicted PAG (c) Predicted PAG without weak links.

Fig. 3. Weak links of the predicted PAG tend to be false positives compared to the original data-generating DAG.

A. Conclusion

In this work, we generated multi-dimensional time-series with high auto-correlation and latent variables and searched for methods to reconstruct the data-generating process from its observed data. We briefly explained the advantages and limitations of today's main methods and applied the LPCMCI algorithm. For our generated datasets, LPCMCI performs best on auto-dependencies, then contemporaneous dependencies, and struggles most with lagged dependencies.

VII. Acknowledgments

References

[1] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect.* New York: Basic Books, 2018.

[2] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler, "Inferring causation from time series in Earth system sciences," *Nature Communications*, vol. 10, no. 1, p. 13, Jun. 2019, bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Climate sciences;Computational science;Databases;Environmental sciences;Statistical physics, thermodynamics and nonlinear dynamics Subject_term_id: climate-sciences;computational-science;databases;environmental-sciences;statistical-physics-thermodynamics-and-nonlinear-dynamics. [Online]. Available: https://www.nature.com/articles/s41467-019-10105-3

[3] J. Peters, D. Janzig, and B. Schölkopf, *Elements of causal inference: foundations and learning algorithms.* The MIT Press, Nov. 2018, vol. 88. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00949655.2018.1505197

[4] C. Reiser, "Predicting and Visualizing Daily Mood of People Using Tracking Data of Consumer Devices and Services," *arXiv:2202.03721 [cs, stat]*, Feb. 2022, arXiv: 2202.03721. [Online]. Available: http://arxiv.org/abs/2202.03721

[5] A. Gerhardus and J. Runge, "High-recall causal discovery for autocorrelated time series with latent confounders," *arXiv:2007.01884 [cs, stat]*, Feb. 2021, arXiv: 2007.01884. [Online]. Available: http://arxiv.org/abs/2007.01884

[6] T. Richardson and P. Spirtes, "Ancestral graph Markov models," *The Annals of Statistics*, vol. 30, no. 4, Aug. 2002. [Online]. Available: https://projecteuclid.org/journals/annals-of-statistics/volume-30/issue-4/Ancestral-graph-Markov-models/10.1214/aos/1031689015.full

[7] J. Zhang, "Causal Reasoning with Ancestral Graphs," *Journal of Machine Learning Research*, p. 38, 2008.

[8] C. W. J. Granger, "Testing for causality: A personal viewpoint," *Journal of Economic Dynamics and Control*, vol. 2, pp. 329–352, Jan. 1980. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016518898090069X

[9] C. Granger, "Some recent development in a concept of causality," *Journal of Econometrics*, vol. 39, no. 1, pp. 199–211, Sep. 1988. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0304407688900450

[10] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "PCMCI - Detecting and quantifying causal associations in large nonlinear time series datasets," *Science Advances*, vol. 5, no. 11, p. 15, 2019, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/10.1126/sciadv.aau4996

[11] J. Pearl, *Causality: models, reasoning, and inference*. Cambridge, U.K. ; New York: Cambridge University Press, 2000.

[12] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, ser. Lecture Notes in Statistics, J. Berger, S. Fienberg, J. Gani, K. Krickeberg, I. Olkin, and B. Singer, Eds. New York, NY: Springer New York, 2000, vol. 81. [Online]. Available: http://link.springer.com/10.1007/978-1-4612-2748-9

[13] D. Entner and P. O. Hoyer, "On Causal Discovery from Time Series Data using FCI," 2010.

[14] A. Hyvarinen, S. Shimizu, and P. O. Hoyer, "Causal Modelling Combining Instantaneous and Lagged Effects: an Identifiable Model Based on Non-Gaussianity," *ICML*, p. 8, 2008.

[15] J. Peters, D. Janzing, and B. Schölkopf, "Causal Inference on Time Series using Restricted Structural Equation Models," in *Advances in Neural Information Processing Systems*, vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/hash/47d1e990583c9c67424d369f3414728e-Abstract.html

[16] J. Runge, "PCMCI+ - Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets," *PMLR Proceedings*, p. 46, 2020.

[17] "Tigramite – Causal discovery for time series datasets," Mar. 2022, original-date: 2017-05-15T10:20:26Z. [Online]. Available: https://github.com/jakobrunge/tigramite/blob/d9d06a1d865582c2f68ee71b313e93dcbd77abba/tutorials/tigramite_tutorial_pcmciplus.ipynb