Survey and Evaluation of Causal Discovery Methods for Time Series (Extended Abstract)*

Charles K. Assaad[1,2] , **Emilie Devijver**[1] , **Eric Gaussier**[1]
[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
[2]EasyVista
cassaad@easyvista, emilie.devijver@univ-grenoble-alpes.fr, eric.gaussier@imag.fr

Abstract

We introduce in this survey the major concepts, models, and algorithms proposed so far to infer causal relations from observational time series, a task usually referred to as *causal discovery in time series*. To do so, after a description of the underlying concepts and modelling assumptions, we present different methods according to the family of approaches they belong to: Granger causality, constraint-based approaches, noise-based approaches, score-based approaches, logic-based approaches, topology-based approaches, and difference-based approaches. We then evaluate several representative methods to illustrate the behaviour of different families of approaches. This illustration is conducted on both artificial and real datasets, with different characteristics. The main conclusions one can draw from this survey is that causal discovery in times series is an active research field in which new methods (in every family of approaches) are regularly proposed, and that no family or method stands out in all situations. Indeed, they all rely on assumptions that may or may not be appropriate for a particular dataset.

1 Introduction

Causality plays a central role in science and has been the subject of many debates among philosophers, biologists, mathematicians and physicists, to name but a few. Causality is implicit in the logic and structure of ordinary language and is embedded in our understanding mechanism that pushes humans to invoke questions. Why is it dark? Why is the sea salty? What is the effect of exercise on heart rate, of a vaccine on a particular disease? What is the effect of industrial pollution on the environment? And so, as already advocated by Spirtes, Glymour and Scheines, in attempting to answer such questions, both the baby and the scientist try to turn observations into causal knowledge [Spirtes *et al.*, 2001]. Causality is indeed crucial for explanatory purposes since an effect can be explained by its causes, regardless of the correlations it may have with other variables.

The recent decades have seen the development, from philosophers, mathematicians, and computer scientists, of different models and methods to infer causal relations from data and to reason on the basis of these relations (to, *e.g.*, predict the effect of changing a particular medication). If the first studies were dedicated to non temporal data, more and more studies now focus on time series. Indeed, time series arise as soon as observations, from sensors or experiments, for example, are collected over time. They are present in various forms in many different domains, as healthcare (through, *e.g.*, monitoring systems), Industry 4.0 (through, *e.g.*, predictive maintenance and industrial monitoring systems), surveillance systems (from images, acoustic signals, seismic waves, etc.) or energy management (through, *e.g.* energy consumption data). The number of scientific publications dedicated to causality in time series as well as the number of tools developed in this context have steadily increased to a point that it is difficult for non specialists to grasp the most important approaches proposed so far.

The goal of our survey is twofold: On the one hand, we want to introduce the major concepts, models, methods, and associated algorithms proposed so far to infer causal relations from observational time series, a task usually referred to as *causal discovery*; on the other hand, we want to assess how different methods for causal discovery in time series behave in practice. Several surveys on causal discovery have recently been proposed [Guo *et al.*, 2020; Nogueira *et al.*, 2021; Glymour *et al.*, 2019; Vowels *et al.*, 2021]. However, most of them do not discuss time series and when they do, they focus on Granger causality. In contrast, our survey is dedicated to causal discovery in time series and reviews all families of approaches proposed in this area.

The remainder of this extended abstract is organized as follows. After a brief description of the main concepts and assumptions in Section 2, we briefly present some of the different algorithms that we presented in our survey according to the family of approaches they belong to and summarize the main characteristics of these algorithms. Section 4 points out some aspect of causal discovery from time series that are not included in the survey. Lastly, Section 5 concludes this extended abstract.

*The full paper is published in Journal of Artificial Intelligence Research, 73:767–820, August 2022.

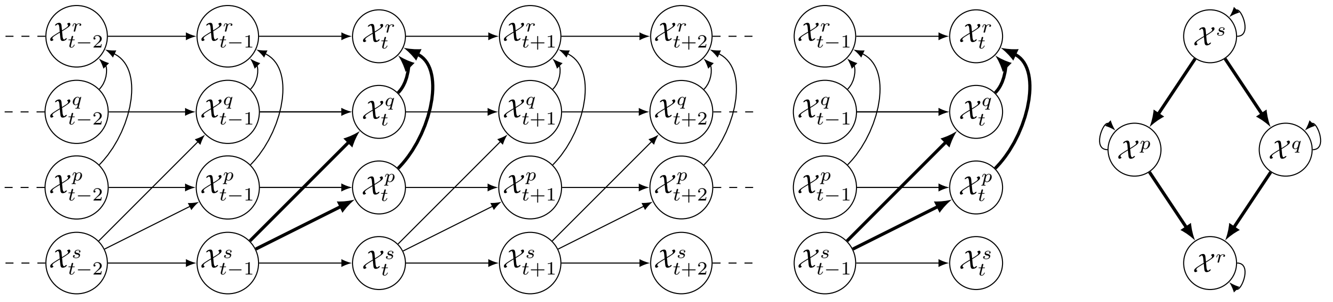

Figure 1: Different causal graphs to respresent a diamond structure with self causes: full time causal graph (5a), window causal graph (5b) and summary causal graph (5c). Note that the first one gives more information but cannot be inferred in practice, the second one is a schematic viewpoint of the full behavior, whereas the last one give an overview and can be deduced from the window causal graph.

2 Background

Causal discovery in time series aims at discovering, from observational data, causal relations within and between d-variate time series $\mathcal{X}$ where, for a fixed t, each $\mathcal{X}_t$ is a vector $(\mathcal{X}_t^1, \cdots, \mathcal{X}_t^d)$ in which each variable $\mathcal{X}_t^p$ represents a measurement of the p-th time series at time t. There are at least three ways to represent time series through a causal graph $\mathcal{G} = (V, E)$ with V the set of vertices and E the set of edges. The first is called a *full time causal graph* and represents a complete acyclic graph of the dynamic system, as illustrated in Figure 1a.

Definition 1 (Full Time Causal Graph)**.** *Let $\mathcal{X}$ be a multivariate discrete-time stochastic process and $\mathcal{G} = (V, E)$ the associated* full time causal graph*. The set of vertices in that graph consists of the set of components $\mathcal{X}^1, \ldots, \mathcal{X}^d$ at each time t. The edges E of the graph are defined as follows: variables $\mathcal{X}_{t-i}^p$ and $\mathcal{X}_t^q$ are connected by a lag-specific directed link $\mathcal{X}_{t-i}^p \to \mathcal{X}_t^q$ in $\mathcal{G}$ pointing forward in time if and only if $\mathcal{X}^p$ causes $\mathcal{X}^q$ at time t with a time lag of $i > 0$ for $p = q$ and with a time lag of $i \geq 0$ for $p \neq q$.*

It is usually not possible to infer general full time causal graphs as there usually is a single observation for each time series at each time instant and it is common to rely on the so-called *Consistency Throughout Time* assumption.

Definition 2 (Consistency Throughout Time)**.** *A causal graph $\mathcal{G} = (V, E)$ for a multivariate time series $\mathcal{X}$ is said to be* consistent throughout time *if all the causal relationships remain constant in direction throughout time.*

When assuming consistency throughout time, the full time causal graph can be contracted to give a finite graph which we call *window causal graph*. It is a representation of the causal graph through a time window, the size of which equals the maximum lag relating time series in the full time causal graph.

Definition 3 (Window Causal Graph)**.** *Let $\mathcal{X}$ be a multivariate discrete-time stochastic process and $\mathcal{G} = (V, E)$ the associated* window causal graph *for a window of size τ. The set of vertices in that graph consists of the set of components $\mathcal{X}^1, \ldots, \mathcal{X}^d$ at each time $t, \ldots, t + \tau$. The edges E of the graph are defined as follows: variables $\mathcal{X}_{t-i}^p$ and $\mathcal{X}_t^q$ are connected by a lag-specific directed link $\mathcal{X}_{t-i}^p \to \mathcal{X}_t^q$ in $\mathcal{G}$ pointing forward in time if and only if $\mathcal{X}^p$ causes $\mathcal{X}^q$ at time t with a time lag of $0 \leq i \leq \tau$ for $p \neq q$ and with a time lag of $0 < i \leq \tau$ for $p = q$.*

Figure 1b illustrates a window causal graph corresponding to the full time causal graph given in Figure 1a. In practice, it is often sufficient to know the causal relations between time series as a whole, without knowing precisely the relations between time instants, in addition, in some applications, an expert would like to validate a causal graph before using it, but validating a window causal graph can be difficult as it is difficult to determine the temporal lag between a cause and an effect. In these cases, one one can use a *summary causal graph*. An example of such a graph is given in Figure 1c. Note that since a summary causal graph is a abstraction of the full time causal graph, it can contain cycles.

Definition 4 (Summary Causal Graph)**.** *Let $\mathcal{X}$ be a multivariate discrete-time stochastic process and $\mathcal{G} = (V, E)$ the associated* summary causal graph*. The set of vertices in that graph consists of the set of time series $\mathcal{X}^1, \ldots, \mathcal{X}^d$. The edges E of the graph are defined as follows: variables $\mathcal{X}^p$ and $\mathcal{X}^q$ are connected if and only if there exists some time t and some time lag i such that $\mathcal{X}_{t-i}^p$ causes $\mathcal{X}_t^q$ at time t with a time lag of $0 \leq i$ for $p \neq q$ and with a time lag of $0 < i$ for $p = q$.*

The relations between a probability distribution and its causal graph are central to the construction of the graph. It is however not always possible to infer a causal graph solely from observational data on which one can only compute correlations and statistical independencies. For this reason, in addition to the acyclicity of the full time causal graph and consistency throughout time, all methods rely on at least two of the following assumptions:

- Stationarity, which states that the generative process does not change with respect to time;
- Causal Markov condition [Spirtes *et al.*, 2001; Pearl, 2000], which states that every variable is independent of all its nondescendants in the graph conditional on its parents;
- Causal sufficiency [Spirtes *et al.*, 2001; Pearl, 2000], which states that all common causes, i.e., confounders, of all observed variables are observed;

- Minimality [Spirtes *et al.*, 2001], which requires that all adjacent nodes are dependent;
- Faithfulness [Spirtes *et al.*, 2001; Pearl, 2000], which states that all conditional independencies are entailed from the causal Markov condition;
- Semi-parametric models [Peters *et al.*, 2017], which stipulates a general form for the underlying model, as linear models or nonlinear additive noise models;
- Temporal priority [Hume, 1738], which makes the process of causality asymmetric in time and is useful for orienting a causal relation when one knows that two variables are causally related, as well as its relaxed version which allows for instantaneous relations as the difference in time between two values of two time series may not be observed if the sampling frequencies of the time series are small.

3 Diffferent Families of Approaches for Causal Discovery

We now turn to the most widely used approaches used to infer causal graphs between time series. Additional approaches can be found in the complete survey.

Granger causality is one of the oldest concepts in causal inference, based on a statistical version of Hume's regularity theory [Hume, 1738] which states that causal relations can be inferred by the experience of constant conjunctions between causes and effects, a cause preceding its effects. Assuming stationary linear systems, one can assess whether $\mathcal{X}^p$ Granger-causes $\mathcal{X}^q$ by considering two autoregression models: an autoregressive restricted model that uses only past values of $\mathcal{X}^q$ to predict its current value and an augmented version of the autoregressive model that uses both past values of $\mathcal{X}^q$ and $\mathcal{X}^p$ to predict the current value of $\mathcal{X}^q$. If the augmented version is significantly more accurate than the restricted model, one can conclude that $\mathcal{X}^p$ Granger-causes $\mathcal{X}^q$. In a multivariate setting, a pairwise analysis can be performed using the bivariate approach denoted as PWGC. This approach does however not fully capture Granger's original ideas which assume that all relevant information is included in the analysis [Eichler, 2008]. To include all relevant information in the analysis, the multivariate Granger causality denoted as MVGC [Geweke, 1982; Chen *et al.*, 2004; Barrett *et al.*, 2010] was introduced. In MVGC, the restricted and augmented models are both based on a vector autoregressive instead of a simple autoregressive model, where the augmented model uses all observational time series whereas the restricted model uses all time series except $\mathcal{X}^p$. Analogously to the bivariate case, if the augmented model is significantly more accurate than the restricted model, one concludes that $\mathcal{X}^p$ Granger-causes $\mathcal{X}^q$. Note that several extensions of the above approach have been proposed, as the temporal causal discovery model, denoted as TCDF [Nauta *et al.*, 2019], which dispenses with the linear assumption made in the original proposal.

Constraint-based approaches exploit conditional independencies to build a skeleton between variables. This skeleton is then oriented either according to temporal priority or according to a set of rules that define constraints on admissible orientations while assuming faithfulness. oCSE [Sun *et al.*, 2015] algorithm uses the *causation entropy* to find these conditional independencies under the assumption that all causal relations have a time-lag of size 1. Due to this assumption, temporal priority is sufficient to orient all edges and the summuary causal graph gives the same information as the window causal graph. However, time-lag of size 1 is not always satisfied in real world scenarios, so the PCMCI algorithm [Runge *et al.*, 2019] was introduced to detect time lagged causal relations in the form of a window causal graph. Note that PCMCI can be flexibly combined with any kind of conditional independence tests. Instantaneous causal relations, which were not supported in the initial algorithm, have been integrated to PCMCI [Runge, 2020] by conducting separately the edge removal for lagged conditioning sets and instantaneous conditioning sets. Lagged relations are treated as in PCMCI and instantaneous relations are oriented using the known PC-rules [Spirtes *et al.*, 2001] which were introduced for non temporal graphs. Both of the above algorithms assume causal sufficiency, however, there exists constraint-based algorithms, such as tsFCI [Entner and Hoyer, 2010], that do not need this assumption.

Noise-based approaches do not consider that statistical noise is as a nuisance that one has to live with. Instead, they consider it as a valuable source of insight to identify causal relations [Hoyer *et al.*, 2009; Climenhaga *et al.*, 2019]. Even though it has been shown that, in a non parametric setting, the noise does not help to distinguish between a cause and its effects [Peters *et al.*, 2017], this is no longer the case when using specific semi-parametric models as a a linear model with non-Gaussian noise [Shimizu *et al.*, 2006], or a nonlinear additive noise model [Hoyer *et al.*, 2009]. Therefore, the VarLiNGAM [Hyvärinen *et al.*, 2010] algorithm was proposed for uniquely identifying causal structures based on purely observational time series assuming a linear model with non-Gaussian noise. Similarly, TiMINo [Peters *et al.*, 2013] was introduced for nonlinear additive noise models. Both algorithms assume causal sufficiency and the minimality condition, which is a weaker version of faithfulness.

Score-based approaches aim to find the graph that best matches the data based on a score that typically strikes a balance between the likelihood of the data given the network and a penalty term related to the complexity of the network. Assuming causal sufficiency as well as linearity, the DYNOTEARS [Pamfil *et al.*, 2020] algorithm was proposed to simultaneously estimate instantaneous and time-lagged relationships between time series. This algorithm relies on minimizing a penalized loss based on the Frobenius norm of the residuals of a linear model.

Summary Table 1 displays the main characteristics of representative algorithms in the above families. As one can note, 4 algorithms infer a summary causal graph and 5 infer a window causal graph. It is of course possible to deduce the summary causal graph from a window causal graph and, in the case of oCSE, the summary causal graph is equivalent

Section	Algorithm	Causal graph	Faithfulness / Minimality	Causal Markov Condition	Instantaneous rel.	Lag > 1	Inference of self causes	Confounders	Inst. Hidden Conf	Lagged Hidden Conf.	Model based	Linear model	< 5 Hyper-parameters
Granger	PWGC	S			✗	✓	✗	✗	✗	✗	✓	✓	✓
	MVGC	S			✗	✓	✗	✓	✗	✗	✓	✓	✓
	TCDF	W			✓$^{\oplus}$	✓	✓$^{\oplus}$	✓	✓	✗	✓	✗	✗
Constraint-based	PCMCI	W	F	✓	✗†	✓	✓	✓	✗	✗	✗	✗	✓
	oCSE	S	F	✓	✗	✗	✓	✓	✗	✗	✗	✗	✓
	tsFCI	W	F	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓
Noise-based	VarLiNGAM	W	M	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓
	TiMINo	S	M	✓	✓	✓	✗	✓	✗	✗	✓	✗	✓
Score-based	DYNOTEARS	W			✓	✓	✓	✓	✗	✗	✓	✓	✓

Table 1: Summary of the main characteristics of representative algorithms in all the families discussed in this survey. For causal graphs, S means that the algorithm provides a summary causal graph whereas W means that the algorithm provides a window causal graph; F corresponds to faithfulness and M to minimality. For a fixed algorithm, check marks with $^{\oplus}$ are mutually exclusive. A cross mark with † indicates that the corresponding algorithm was recently extended to handle the information given in the corresponding column. An empty cell can either mean that the information given in the corresponding column was not discussed by the authors of the corresponding algorithm or that this information is not needed for the corresponding algorithm.

to the window causal graph. Roughly only half of the algorithms address the problem of discovering instantaneous relations. Most algorithms can detect confounders. However, only TCDF can detect instantaneous hidden confounders and only tsFCI can detect lagged hidden confounders. More generally, very few algorithms can deal with hidden variables, which violates the causal sufficiency assumption. Regarding the type of underlying models, almost all algorithms rely on a particular model (except PCMCI and oCSE). Among the algorithms relying on a model, roughly half of them rely on a linear model. Relying on a specific model can be an advantage when the data considered arises from a similar model. It can be of course a disadvantage when this is not the case and when the model family considered is not a universal approximator. Lastly, as one can note, most algorithms have few (less than 5) hyper-parameters, with the exception of TCDF which is based on deep neural networks.

4 Elaboration

Our survey reviews different causal discovery algorithms in the setting where consistency throughout time as well as other technical conditions are satisfied. However, it is important to note that there exists algorithms that relax these conditions: [Huang *et al.*, 2015; Huang *et al.*, 2020; Saggioro *et al.*, 2020] relax the consistency throughout time constraint, [Danks and Plis, 2013; Gong *et al.*, 2015] allow for temporal aggregation or subsampling, and [Kleinberg and Mishra, 2009; Kleinberg, 2011; Kleinberg, 2015] consider qualitative and mixed data. In addition, in each family of approaches, we presented what we think are the most known algorithms but one should bear in mind that there exists many extensions of these algorithms. For example, in the constraint-based family, the idea behind the oCSE algorithm was extended to handle lags of size different from 1, instantaneous relations as well as a method to directly infer the summary causal graph without going through a window causal graph [Assaad *et al.*, 2022a; Assaad *et al.*, 2022b]. In the noise based-family, identifiability was shown to be also possible under a post nonlinear additive noise model [Zhang and Hyvärinen, 2009]. Finally, it is worth noting that we did not include hybrid methods [Malinsky and Spirtes, 2018; Sanchez-Romero *et al.*, 2019; Assaad *et al.*, 2021] which sometimes yield better results than purebred methods.

5 Conclusion

Our survey presents different algorithms, pertaining to different families of approaches, for causal discovery in time series. The families we have retained here correspond to approaches à la Granger, constraint-based approaches, noise-based approaches and score-based approaches. Further details on those algorithms and their evaluation, as well as on other families of approaches (namely, logic-based approaches, topology-based approaches, and difference-based approaches) and associated algorithms can be found in the full paper [Assaad *et al.*, 2022c]. In a nutshell, one can draw from our survey that causal discovery in times series is an active research field in which new methods (in every family of approaches) are regularly proposed, and that no family or algorithm stands out in all situations. Indeed, they all rely on assumptions that may or may not be appropriate for a particular dataset.

References

[Assaad *et al.*, 2021] Charles K. Assaad, Emilie Devijver, Eric Gaussier, and Ali Ait-Bachir. A mixed noise and constraint-based approach to causal inference in time series. In Nuria Oliver, Fernando Pérez-Cruz, Stefan Kramer, Jesse Read, and Jose A. Lozano, editors, *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 453–468, Cham, 2021. Springer International Publishing.

[Assaad *et al.*, 2022a] Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Discovery of extended summary graphs in time series. In James Cussens and Kun Zhang, editors, *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pages 96–106. PMLR, 01–05 Aug 2022.

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

[Assaad *et al.*, 2022c] Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *J. Artif. Int. Res.*, 73, apr 2022.

[Barrett *et al.*, 2010] Adam B. Barrett, Lionel C. Barnett, and Anil K. Seth. Multivariate granger causality and generalized variance. *Physical review E*, 81:041907, 2010.

[Chen *et al.*, 2004] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. Analyzing multiple nonlinear time series with extended granger causality. *Physics Letters A*, 324:26–35, 2004.

[Climenhaga *et al.*, 2019] Nevin Climenhaga, Lane DesAutels, and Grant Ramsey. Causal inference from noise. *Noûs*, 2019.

[Danks and Plis, 2013] David Danks and Sergey Plis. Learning causal structure from undersampled time series. In *Advances in Neural Information Processing Systems*, Workshop on Causality, 2013.

[Eichler, 2008] Michael Eichler. Causal inference from time series: What can be learned from granger causality? *Proceedings from the 13th International Congress of Logic, Methodology and Philosophy of Science*, 2008.

[Entner and Hoyer, 2010] Doris Entner and Patrik O. Hoyer. On causal discovery from time series data using fci. In *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (PGM-2010)*, pages 121–128, Helsinki, Finland, 2010. HIIT Publications.

[Geweke, 1982] John Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.

[Glymour *et al.*, 2019] Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019.

[Gong *et al.*, 2015] Mingming Gong, Kun Zhang, Bernhard Schölkopf, Dacheng Tao, and Philipp Geiger. Discovering temporal causal relations from subsampled data. volume 37 of *Proceedings of Machine Learning Research*, pages 1898–1906, Lille, France, 2015. PMLR.

[Guo *et al.*, 2020] Ruocheng Guo, Lu Cheng, Jundong Li, P. Richard Hahn, and Huan Liu. A survey of learning causality with data: Problems and methods. *ACM Comput. Surv.*, 53(4), July 2020.

[Hoyer *et al.*, 2009] Patrik O. Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 689–696. Curran Associates, Inc., 2009.

[Huang *et al.*, 2015] Biwei Huang, Kun Zhang, and Bernhard Schölkopf. Identification of time-dependent causal model: A gaussian process treatment. In *24th International Joint Conference on Artificial Intelligence, Machine Learning Track*, pages 3561–3568, Palo Alto, California USA, 2015. AAAI Press.

[Huang *et al.*, 2020] Biwei Huang, Kun Zhang, Jiji Zhang, Joseph D Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *J. Mach. Learn. Res.*, 21(89):1–53, 2020.

[Hume, 1738] David Hume. *A Treatise of Human Nature*. Oxford University Press, 1738.

[Hyvärinen *et al.*, 2010] Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O. Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11:1709–1731, 2010.

[Kleinberg and Mishra, 2009] Samantha Kleinberg and Bud Mishra. The Temporal Logic of Causal Structures. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, Montreal, Quebec, 2009.

[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, pages 943–950. AAAI Press, 2011.

[Kleinberg, 2015] Samantha Kleinberg. *Why: A Guide to Finding and Using Causes*. O'Reilly Media, Incorporated, 2015.

[Malinsky and Spirtes, 2018] Daniel Malinsky and Peter Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Disocvery*, volume 92 of *Proceedings of Machine Learning Research*, pages 23–47, London, UK, 2018. PMLR.

[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):312–340, 2019.

[Nogueira *et al.*, 2021] A.R. Nogueira, J. Gama, and C.A. Ferreira. Causal discovery in machine learning: Theories and applications. *Journal of Dynamics & Games*, 8(3):203–231, 2021.

[Pamfil *et al.*, 2020] Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1595–1605. PMLR, 26–28 Aug 2020.

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000.

[Peters *et al.*, 2013] Jonas Peters, D. Janzing, and B. Schölkopf. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems 26*, pages 154–162, 2013.

[Peters *et al.*, 2017] Jonas Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, Cambridge, MA, USA, 2017.

[Runge *et al.*, 2019] Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), 2019.

[Runge, 2020] Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Jonas Peters and David Sontag, editors, *Proceedings of Machine Learning Research*, volume 124, pages 1388–1397. PMLR, 2020.

[Saggioro *et al.*, 2020] Elena Saggioro, Jana de Wiljes, Marlene Kretschmer, and Jakob Runge. Reconstructing regime-dependent causal relationships from observational time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(11):113115, 2020.

[Sanchez-Romero *et al.*, 2019] Ruben Sanchez-Romero, Joseph Ramsey, Kun Zhang, Madelyn Glymour, Biwei Huang, and Clark Glymour. Estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. *Network Neuroscience*, 3:274 – 306, 2019.

[Shimizu *et al.*, 2006] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

[Spirtes *et al.*, 2001] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2001.

[Sun *et al.*, 2015] Jie Sun, Dane Taylor, and Erik Bollt. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1):73–106, 2015.

[Vowels *et al.*, 2021] Matthew J. Vowels, Necati Cihan Camgöz, and Richard Bowden. D'ya like dags? A survey on structure learning and causal discovery. *CoRR*, abs/2103.02582, 2021.

[Zhang and Hyvärinen, 2009] Kun Zhang and Aapo Hyvärinen. On the identifiability of the post-nonlinear causal model. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, page 647–655, Arlington, Virginia, USA, 2009. AUAI Press.