Reducing Causal Illusions through Deliberate Undersampling

Anonymous Author(s)
Affiliation
Address
email

Abstract

Domain scientists interested in the causal mechanisms are usually limited by the frequency at which they can collect the measurements of social, physical, or biological systems. It is a reasonable assumption that higher frequency is more informative of the causal structure. This assumption is a strong driver for designing new, faster instruments. A task that is expensive and often impossible at the current state of technology. In this work, we show that counter to the intuition it is possible for causal systems to improve the estimation of causal graphs from undersampled time-series by augmenting the measurements with those collected at a rate slower than currently available. We present an algorithm able to take advantage of measurement time-scale graphs estimated from data at various sampling rates and lower the underdeterminacy of the system by reducing the equivalence size. We investigate the probability of cases in which deliberate undersampling yields a gain and the size of this gain.

1 Introduction

Domain scientists often study systems of interest through the lens of complex multivariate time series collected at fixed intervals as measurements of the interacting parts of the system. The search for causal structures, namely the search for causal relationships between several random variables, has been widely studied at least since Pearl et al. (2000) and is used in the social sciences (Blalock, 2017), climatology (Runge et al., 2019) and neuroscience (Zhang et al., 2020). Often, the goal is identification of a form of causal relationship among the parts of the system. The subject of our work is dynamic causal structures that arise as a result of these studies. It has been noted in a number of works that the initial causal structures derived from the data may differ depending on the frequency with which measurements are made (Danks and Plis, 2013; Rhea et al., 2015; Gong et al., 2017).

In multivariate dynamical systems there is a minimal characteristic time scale at which variables of interest exchange information. We call it the causal timescale as opposed to the measurement timescale at which the data is collected. Typically, the measurement timescale is significantly slower than the causal time scale leading to undersampling – unmeasured intermediate states. For example, the resolution of an fMRI scanner is 2000 ms (Constable, 2006), while the causal interactions of groups of neurons are at least 10-20 orders of magnitude faster (Perri et al., 2014; Freiwald et al., 1999). In many cases, we do not know the extent of undersampling, nor do all measurement methods undersample to the same degree. Thus, there can be significant changes in the apparent causal connections (Danks and Plis, 2013; Breitung and Swanson, 2002; Gong et al., 2015). Similarly, to a specific viewing angle that may lead to an optical illusion, undersampling may often lead to causal illusions.

There are algorithms for searching for a causal structure during undersampling (when the measurement is performed k times less than the time of the causal interaction) with known k (Gong et al., 2017;

Submitted to *A causal view on dynamical systems* workshop at NeurIPS 2022. Do not distribute.

Silvestrini and Veredas, 2008) and unknown k (Hyttinen et al., 2017; Plis et al., 2015; Abavisani et al., 2022). In each of these cases, there remains an error in the estimation of the causal structure, which decreases with increasing frequency of measurement.

A natural response to the challenge of undersampling is to conclude that we simply should measure faster. If we could somehow measure at the causal time scale, then we could avoid the challenges of undersampling. More generally, a natural intuition is that faster data collection always provides more information about causal interactions in a system, and so we should invest in developing better instrumentation (Fan et al., 2022; Kalhan et al., 2022). But while faster measurements almost always reduce equivalence class size (or at least, do not increase it), faster measurements are typically more costly, either in time, resources, or personnel. In some cases, it may not even by possible to measure sufficiently quickly. For example, increasing the frequency of blood sample collections could have a significant negative impact on patients' quality of life. Moreover, measuring *too* fast can introduce additional noise into our system measurements (or perhaps impact the system itself), thereby increasing the data collection requirements. Finally, a single-minded focus on increasing measurement speed can lead researchers to disregard previously-collected, though slower, data when a faster instrument is available.

The central idea in this paper is that this focus on faster measurement is, in some cases, unwarranted. In particular, one can often achieve equal improvements in learning by using *slower* measurements, not only faster ones. The basic intuition is that the slower measurements can provide a second "perspective" on the causal system that potentially reveals insights that were hidden from the first perspective. The present paper provides the first algorithm for reducing the causal structure error using additional measurements at a lower (i.e., slower) measurement frequency.

In general, if we only have measurements at a single timescale, then using a slower measurement frequency will typically increase the size of the equivalence class of causal structures that could have produced the observed data. In this sense, the "natural intuition" is correct: if we can only have one dataset, then we should usually aim to have a faster timescale (up to the actual causal timescale). Of course, we are rarely limited in this way; in particular, we can often obtain a second dataset without substantial effort using the same equipment, at least if we are willing to use a slower timescale. The results in this paper show that we can gain significant information by taking this easier path, though the slower timescale must be "coprime" with respect to the original timescale.[1]

This result is arguably surprising—we can learn more by measuring more slowly—but analogical phenomena exist in other domains. Most notably, consider the case of image super-resolution. In this phenomenon, one often obtains substantial increases in image resolution not by using a higher-resolution imaging device, but rather by using a same-resolution device that is offset by a distance that is not an integer multiple of the pixel size. Alternately, one can simply use a *lower*-resolution imaging device; if the higher resolution is not an integer multiple of the lower resolution, then there will necessarily be many "offset" pixels, and so some super-resolution will be possible.

Measurement at a slower timescale does not necessarily enable super-resolution of the data itself, not least because the two data streams need not be collected simultaneously. However, the same intuition holds with regards to causal structure: by capturing an undersampled time series with a different "offset," we can learn a more detailed picture of the underlying causal structure. For example, if we use a slower measurement speed in fMRI, then although we do not improve the temporal resolution of the voxel-level data stream, we do improve the temporal resolution of the causal structures in the brain that generate those data streams. That is, we get a "resolution gain" in the causal system, rather than the independent variables.

The present paper thus (i) shows the existence of the counter-intuitive phenomenon that using slower measurements can help as much as faster ones; which thereby (ii) enables scientists to learn more about causal structure from undersampled data using existing measurement instruments, rather than waiting for the development of novel measurement methodologies.

[1]The exact conditions are explained below.

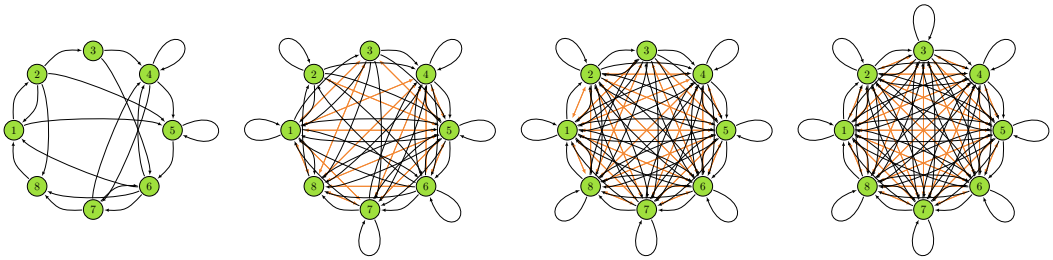

Figure 1: Example of causal graph (first) and effects of undersampling rates 2,3 and 4

2 Notation and definitions

2.1 Theoretical model

The generally accepted model of the dynamical causal system is the graph model (Shafer, 1995; Pearl, 1988; Lakshmikantham et al., 2010), we will refer to these as dynamic Bayesian networks or DBNs. DBN includes random variables $\mathbf{V}$ at the current time step t, as well as nodes for $\mathbf{V}$ at each previous times step in which there is a direct cause of the current values of $\mathbf{V}$. From a theoretical perspective, we assume that the underlying causal structure can be represented by a Markov order one graphical model in which no causal interactions happen instantaneously. This assumption is relatively innocuous since we are not assuming that we *measure* at this speed. For convenience, we present the DBN in the form of a compressed graph G in which time information is encded implicitly in the edges (i.e., $X \to Y$ in G means $X^{t-1} \to Y^t$ in the underlying DBN).

Let $t^0, t^1, t^2, \ldots, t^k, \ldots$ denote the time steps of the underlying causal structure and call the graph G or G^1 corresponding to this structure a *causal graph*. In the graph G, self-loops are possible, that is, edges leading from the vertex X to itself (i.e., $X^{t-1} \to X^t$). The data is undersampled with rate u if the measured time steps are $t^0, t^u, t^{2u}, \ldots, t^{ku}, \ldots$, then we call the graph G^u, corresponding to this structure *measured* or *undersampled* with rate u. For example, in Figure 1, the first graph is the *causal graph*, the second one is *measured* with $u = 2$, the third one is *measured* with $u = 3$, and for the fourth $u = 4$. It is worth noting that during undersampling, in addition to causal relationships (directed edges, which are indicated in black), correlations (bidirectional red edges) are added, which corresponds to the presence of common unobservable causes of variables in the current time step.

Let $[\![H]\!]$ denote the equivalence class of measured graph H: the set of all graphs G for which $\exists u$ such that $G^u = H$. And $[\![H_1, H_2, \ldots, H_k]\!]$ is the set of all G^1 for which $\exists u_1, u_2, \ldots, u_k$ such that $G^{u_1} = H, G^{u_2} = H, \ldots, G^{u_k} = H$.

We say that a graph G^1 exhibits *non-monotonic undersampling* if and only if $\exists u < v$ such that $[\![G^u, G^v]\!] \neq [\![G^u]\!]$. That is, our uncertainty about the possible G^1 does not necessarily monotonically increase as the undersampling increases. More practically, this property implies that measuring G^v enables us to further reduce the space of possible G^1, even though G^v is even more undersampled than G^u (see Figure 2). Graphs that exhibit *monotonic undersampling* are those where $\forall u < v [\![G^u]\!] \subseteq [\![G^v]\!]$.

2.2 Computational model

For our task is to show the existence of the phenomenon or reducing "causal illusions" and the utility of deliberate undersampling for that, it suffices to demonstrate it in computational experiments on specific class of graphs. We specifically focus on graphs that are strongly connected components (SCCs) with gcd = 1.

The whole set of graphs with directed and bidirected edges can be divided into those that are reachable versus unreachable by undersampling. That is, a graph H is reachable by undersampling if $\exists u > 1, G^1$ such that $G^u = H$ and unreachable otherwise. The proportion of reachable graphs among

(a) A graph in the monotonic undersampling step

(b) A non-monotonic undersamping step

Figure 2: Venn diagrams for equivalence classes of different measured graphs. In these examples, the graphs for which the equivalence classes for the blue and white ovals are presented are graphs with different undersampling rates values but the same causal graph. $[\![G]\!]$ denotes the equivalence class of the graph G. (a) Left graph is G^3 and right is G^5. The equivalence class $[\![G^5]\!]$ totally includes the equivalence class of $[\![G^3]\!]$ (b) Left graph is H^3 and right is H^5. In this example, the intersection of $[\![H^5[$ and $[\![H^3]\!]$ is not equal to $[\![H^3]\!]$.

```
%( * input graph 1 edge specifications here * e.g.: hdirected(1,5,1) ... )
%( * input graph 2 edge specifications here * e.g.: hdirected(1,5,2) ... )
#const n = 10, maxu = 20
node(1..n).
1 {u(1..maxu, 1)} 1.
1 {u(1..maxu, 2)} 1.
{edge1(X,Y)} :- node(X), node(Y).
directed(X, Y, 1) :- edge1(X, Y).
directed(X, Y, L) :- directed(X, Z, L-1), edge1(Z, Y), L <= U, u(U, _).
bidirected(X, Y, U) :- directed(Z, X, L), directed(Z, Y, L), node(X;Y;Z), X < Y, L
    < U, u(U, _).
:- directed(X, Y, L), not hdirected(X, Y, K), node(X;Y), u(L, K).
:- bidirected(X, Y, L), not hbidirected(X, Y, K), node(X;Y), u(L, K), X < Y.
:- not directed(X, Y, L), hdirected(X, Y, K), node(X;Y), u(L, K).
:- not bidirected(X, Y, L), hbidirected(X, Y, K), node(X;Y), u(L, K),
X < Y.
:- M = N, {u(M, 1..2); u(N, 1..2)} == 2, u(M, _), u(N, _).
```

Listing 1: Clingo code for sRASL

random graphs is small,[2] it makes sense to generate G^u from a certain class of graphs G to investigate statistical characteristics. In particular, if G^1 consists of a single strongly connected component (i.e., $\forall X, Y \exists \pi : X \to \ldots \to Y$) plus a small constraint, then all G^u are also a single strongly connected commponent Danks and Plis (2013). The easiest way to get such a system is to connect all the nodes in a simple cycle, and then add k random edges (possibly self-loops).

3 dRASL

In this section, following Abavisani et al. (2022) we present and explain the `clingo` answer set programming (ASP) solver implementation of a solution for the rate agnostic causal learning (RASL) problem (Plis et al., 2015). We follow the conventions of Abavisani et al. (2022) and call the algorithm dRASL, were d stands for "deliberate undersampling". Listing 1 presents the essential parts of the `clingo` code for the problem. Lines 1 – 2 are generated automatically by Python given two input graphs estimated at "co-prime" measurement rates (or generated ensuring this property). These are the graph-specific predicates. The first line lists the directed and bidirected edges of both graphs H^u and H^{u+1}. For example, existence of edge from node 1 to node 5 in H^u is translated to `hdirected(1, 5, 1)`. Line 2 focuses on the same but for H^{u+1}. Line 4 sets the number of nodes in H and sets the maximum undersampling rate to consider. The undersampling rate needs to have a fixed upper bound. However, it is often possible to guess this upper bound from domain-specific considerations. For example, in functional MRI, the maximum undersampling rate can arguably be set to 20 as the

[2] Since H can have bidirected edges, the space of H is bigger than the space of G^1 by a factor of $2^{(n \text{ choose } 2)}$. Any particular G^1 can lead to multiple graphs in H-space (as u varies), but H-space only sparsely covered by various G^u.

difference between hundred milliseconds of the neural causal timescale and 2 second measurement time scale hardly can be more than 20. Lines 5, 6 set two different predicates for under sampling rates, one for each H. These undersamplings can be anywhere from 1 to `maxu`. Line 7 is a choice rule that considers that each edge in the G^1 can exist. Lines 8, 9, 10 generate predicates for directed and bidirected paths of lengths L, practically encoding the forward algorithm of undersampling. Lines $11 - 14$ ensure that the candidate G^1 agrees with both H^u and H^{u+1} in terms of having the same directed and bidirected paths of different lengths, as well as absence of the same directed and bidirected paths of different lengths. Finally, line 15 makes sure that H^u and H^{u+1} are indeed distinct graphs.

The main results of the correctness of this algorithm are based on the correctness of the problem formulation in Listing 1, which follows by construction, and validity and completeness of `clingo`'s ASP solver algorithms. The latter is proven elsewhere, including (Hyttinen et al., 2017; Abavisani et al., 2022)

4 Results

In this section, we present the results of our computational experiments on the potential learning benefits of deliberate undersampling. In particular, we empirically demonstrate that non-monotonic undersampling is quite common in the class of SCCs with $\gcd(G^1) = 1$, which forms the largest and, arguably, the most practically relevant class of SCCs.

4.1 Estimating prevalence of non-monotonic undersampling graphs

The purpose of this computational experiment is to find a lower bound on the proportion of *non-monotonic undersampling* causal graphs G. To estimate this bound, we generated 100 SCCs with n vertices connected in a large ring and a random arrangement of k additional edges for each value of $n \in \{5, 6, 7, 8, 9\}$ and $k \in [1, 30]$. Only graphs with $\gcd() = 1$ were considered as the most frequent and the most practically relevant. Presence of even a single auto-correlation in an SCC leads to $\gcd() = 1$ and auto-correlations are common in most studied systems.

Given a randomly generated SCC G^1, we produced a list of G^u graphs for all u values available for this graph before convergence. After that incrementing u by 1 we sequentially ran sRASL of Abavisani et al. (2022) to produce $[\![G^u]\!]$ and our dRASL modification on (G^u, G^{u+1}) to obtain $[\![G^u, G^{u+1}]\!]$ The SCC was considered as non-monotonic in the evaluation if there is such u for which $[\![G^u, G^{u+1}]\!] < [\![G^u]\!]$ or, in other words, there is a non-zero gain in using dRASL (see Figure 2b).

Notably, the density of graph with gain (Figure 3b) is a lower bound on the true density of such graphs, since it follows from the definition of *non-monotonic undersampling* graphs that v can be anything greater than u as long as they are co-prime, while we restrict ourselves to the case $v = u + 1$. Moreover, due to computational limitations, we used the dRASL algorithm to test u for the random SCCs only up to a limit of a certain maximum u, which is lower than the maximum value of u for graphs with k edges added to the initial ring, when $k < 4$. Therefore, the estimate for graphs with $k < 4$ could be much higher than that presented in the graph.

As we can see from Figure 3, deliberate undersampling is most helpful for the sparse G^1 graphs. In Figure 3 the X-axis represents the density of the original SCC we generated (as described in the Section 2.2), or the graph on the causal time scale. Such graphs (Figure 3a) are not uniformly distributed from 0 to 1 in density, but are distributed from the minimum to the maximum values of the function $(n + k)/n^2$. The Y-axis shows how many different graphs can be obtained from the original graph G by the undersampling operation G^u. Figure 3b shows the kernel density estimate for those graphs that, according to our lower bound estimate, turned out to be *non-monotonic undersampling*.

From the difference in the normalized KDE distributions (Figure 3c) of all graphs and graphs for which deliberate undersampling leads to a gain, we see that SCCs with gain are most likely to be found among graphs with a lower density and higher maximum u.

4.2 Density of measured *non-monotonic undersampling* graphs

Above, we considered density of SCCs that have non-monotonic undersampling mode relative to total number of graphs. From a practical point of view, the properties of the measured G^u, for which

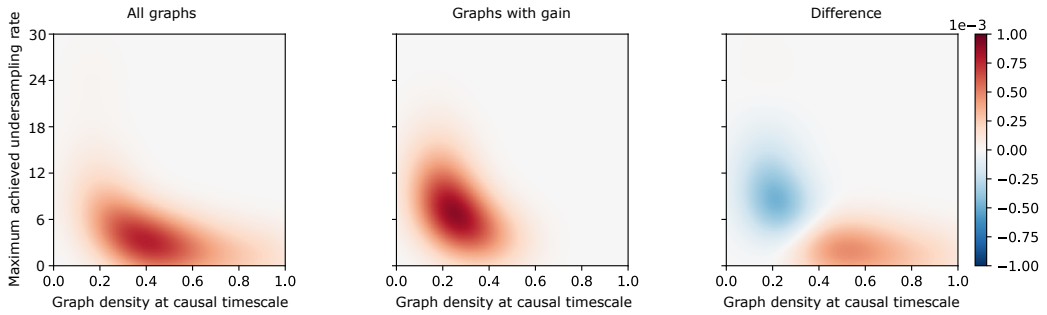

Figure 3: Kernel density estimation (KDE) of graphs with gcd() = 1. X-axis: The density of the causal graph G^1. Y-axis: The undersampling rate after which the graph G^1 converges to a super-clique or cyclically repeats. (a) Normalized KDE for all SCCs. (b) Normalized KDE for SCCs with *non-monotonic undersampling* mode. (c) Difference between KDE of all SCCs and SCCs with *non-monotonic undersampling* modes.

deliberate undersampling gives a gain, seem more interesting. On the histogram in Figure 4, the abscissa shows the density of the causal timescale graph G^1, and the ordinate shows the density of the measurement timescale graph for all *non-monotonic undersampling* graphs in the experiment. Density is computed as N/n^2, where N - is the number of all edges of the current graph and n - the number of nodes. As can be seen from the figure, the largest number of measured graphs is in the density range from 0.5 to 0.8, while the density of causal graphs does not exceed 0.6.

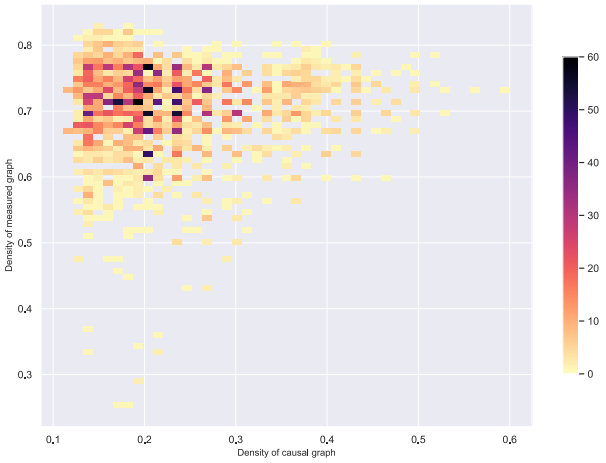

Figure 4: The histogram of *non-monotonic undersampling* graphs G. X-axis: The density of the graph in causal timescale G^1. Y-axis: The density of the graph in measured timescale G^u. Color map: number of graphs from uniformly generated in causal timescale 3000 n-cycled graphs with n=5,6,7,8,9 (total 15000)

.

4.3 Statistics of the gain of *non-monotonic undersampling* graphs

The purpose of this experiment is to estimate the gain given by the deliberate undersampling. The gain is calculated as $gain = \log_{10} r_1/r_{12}$. Where $r_1 = [\![G^u]\!]$ and $r_{12} = [\![G^u, G^{u+1}]\!]$. This means, that, if $gain = 2$, the deliberate undersampling reduces the number of solutions 100-fold. It is worth noting that the gain in this section determines how the number of solutions (the size of the equivalence class) changes when an additional, slower dimension is added. For clarity, we will give an example with fMRI. Assume that the causal time scale is 100 ms and the time resolution is 2000 ms. Then the undersampling rate will be 20. In this case, adding a measurement for $u + 1$ corresponds to a

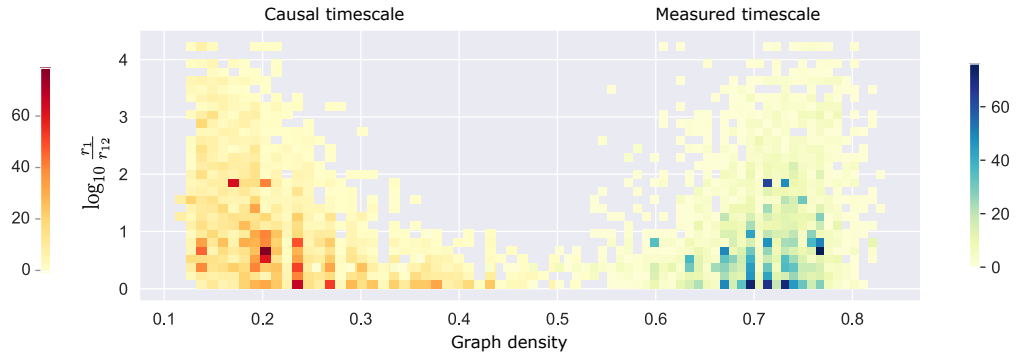

Figure 5: The histogram of the gain of *non-monotonic undersampling* graphs G. X-axis: The density of the graph. Y-axis: Decimal logarithm of r_1/r_{12} where $r_1 = [G^u]$ and $r_{12} = [G^u, G^{u+1}]$. Z(colour): number of graphs from uniformly generated in causal timescale 3000 n-cycled graphs with n=5,6,7,8,9 (total 15000). (left): The histogram in causal timescale. (right): The histogram in measured timescale.

measurement with a time resolution of 2100 ms. In this case, we measured the graph G^u and got r_1 possible graphs, from which we can get the graph G^u by undersampling. When we measured the system with the new temporal resolution, we got an additional constraint that the graphs in the causal temporal resolution must comply with, and the size of the equivalence class became r_{12}. Two histograms (Figure 5) show the same *non-monotonic undersampling* graphs in two different time scales - in the causal (or original G^1 graphs) and in the measured (or G^u graphs). Although there are individual cases for which the gain is 3-4 orders of magnitude, but most of them are concentrated in the area of gain from units to hundreds (0-2 orders of magnitude).

5 Conclusion

This work highlights the problem of distortion of the measured causal structure when measuring with a time resolution lower than the characteristic causal time. We have positively answered the question whether it is possible to use deliberate undersampling to reduce indeterminacy of the problem compared to the algorithms using data at a single sampling rate. Surprisingly, in many cases deliberate undersampling gives a desider effect of reducing the "causal illusion". Having considered a specific class of graphs for which the causal system does not degenerate under slow measurement, we have shown the statistical features of the density of such graphs and the gain that can be obtained using deliberate undersampling. Our work serves as a demonstration that there is much to gain when faster measurements are not an option for any of the possible reasons. One of the potentially larger impacts of this result is the potential of reuse of historical data collected at slower timescales, when datasets with somewhat faster sampling rates are already available. Limitations of our work include the missing theoretical investigation on the conditions when deliberate undersampling is helpful to get some information gain. Ideally, we would want a test on the measured graph to tell whether deliberate undersampling would help. However, even characterization of the properties conducive to the gain of G^1 graphs in the causal timescale could be insightful. This, however, is still work in progress.

References

Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2), 2000.

Jr Blalock. *Causal models in the social sciences*. Routledge, 2017.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):1–13, 2019.

Gemeng Zhang, Aiying Zhang, Biao Cai, Zhuozhuo Tu, Vince D Calhoun, and Yu-Ping Wang. Detecting abnormal connectivity in schizophrenia via a joint directed acyclic graph estimation model. *arXiv preprint arXiv:2010.13029*, 2020.

David Danks and Sergey Plis. Learning Causal Structure from Undersampled Time Series. 12 2013. doi: 10.1184/R1/6492101.v1. URL https://kilthub.cmu.edu/articles/journal_contribution/Learning_Causal_Structure_from_Undersampled_Time_Series/6492101.

Christopher K Rhea, Adam W Kiefer, W Geoffrey Wright, Louisa D Raisbeck, and F Jay Haran. Interpretation of postural control may change due to data processing techniques. *Gait & posture*, 41(2):731–735, 2015.

Mingming Gong, Kun Zhang, Bernhard Schölkopf, Clark Glymour, and Dacheng Tao. Causal discovery from temporally aggregated time series. In *Uncertainty in artificial intelligence: proceedings of the... conference. Conference on Uncertainty in Artificial Intelligence*, volume 2017. NIH Public Access, 2017.

R Todd Constable. Challenges in fmri and its limitations. In *Functional MRI*, pages 75–98. Springer, 2006.

Rinaldo Livio Perri, Marika Berchicci, Donatella Spinelli, and Francesco Di Russo. Individual differences in response speed and accuracy are associated to specific brain activities of two interacting systems. *Frontiers in behavioral neuroscience*, 8:251, 2014.

Winrich A Freiwald, Pedro Valdes, Jorge Bosch, Rolando Biscay, Juan Carlos Jimenez, Luis Manuel Rodriguez, Valia Rodriguez, Andreas K Kreiter, and Wolf Singer. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *Journal of neuroscience methods*, 94(1): 105–119, 1999.

Jörg Breitung and Norman R Swanson. Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis*, 23(6):651–665, 2002.

Mingming Gong, Kun Zhang, Bernhard Schoelkopf, Dacheng Tao, and Philipp Geiger. Discovering temporal causal relations from subsampled data. In *International Conference on Machine Learning*, pages 1898–1906. PMLR, 2015.

Andrea Silvestrini and David Veredas. Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22(3):458–497, 2008.

Antti Hyttinen, Sergey Plis, Matti Järvisalo, Frederick Eberhardt, and David Danks. A constraint optimization approach to causal discovery from subsampled time series data. *International Journal of Approximate Reasoning*, 90:208–225, 2017.

Sergey Plis, David Danks, Cynthia Freeman, and Vince Calhoun. Rate-agnostic (causal) structure learning. *Advances in neural information processing systems*, 28, 2015.

Mohammadsajad Abavisani, David Danks, and Sergey Plis. Constraint-based causal structure learning from undersampled graphs. *arXiv preprint arXiv:2205.09235*, 2022.

Jingjing Fan, Kevin Sitek, Bharath Chandrasekaran, and Abhra Sarkar. Bayesian tensor factorized mixed effects vector autoregressive processes for inferring granger causality patterns from high-dimensional neuroimage data. *arXiv preprint arXiv:2206.10757*, 2022.

Shivam Kalhan, Jessica McFadyen, Naotsugu Tsuchiya, and Marta I Garrido. Neural and computational processes of accelerated perceptual awareness and decisions: A 7t fmri study. *Human Brain Mapping*, 2022.

Glenn Shafer. P. spirtes, c. glymour and r. scheines," causation, prediction and search"(book review). *Synthese*, 104(1):161–176, 1995.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

Vangipuram Lakshmikantham, Srinivasa Leela, Zahia Drici, and FA McRae. *Theory of causal differential equations*, volume 5. Springer, 2010.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Section 3.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 2.2 and Section 4.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] The results of this work do not focus on the computation time of the solution
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 3.
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]