

# Non-parametric causal inference for bivariate time series

James M. McCracken\* and Robert S. Weigel†

*Department of Physics and Astronomy*

*George Mason University*

*4400 University Drive MS 3F3, Fairfax, VA 22030-4444*

(Dated: August 23, 2015)

We introduce new quantities for exploratory causal inference between bivariate time series. The quantities, called *penchants* and *leanings*, are computationally straightforward to apply, follow directly from assumptions of probabilistic causality, do not depend on any assumed models for the time series generating process, and do not rely on any embedding procedures; these features may provide a clearer interpretation of the results than those from existing time series causality tools. The *penchant* and *leaning* are computed based on a structured method for computing probabilities.

## I. INTRODUCTION

Many scientific disciplines rely on observational data from systems in which it is difficult or impossible to implement controlled experiments or to control interventions. For example, there is no current technology that can control the interaction between the solar wind and the magnetic field measured at the surface of Earth, so space weather studies rely on data collected without performing controlled experiments. As a result, causal inference with observational data sets from such systems is difficult and the need to identify causal relationships given the weakness of correlation in doing so has lead to the development of several different time series causality tools [13, 21, 30, 34, 35].

Causal inference in time series involves finding “driving” relationships between different time series signals. Showing the existence, rather than the exact nature, of the driving relationship between the signals is often the primary goal. Thus, words like “driving”, “causality”, and related terms typically do not have straightforward analogs to the same terms used in other fields [14, 22, 33], e.g. theoretical quantum (e.g., [31]) or classical mechanics (e.g., [3]).

The development and study of causal inference techniques is often called *time series causality*. Most techniques fall into four broad categories related to either transfer entropy [35], Granger causality [13], state space reconstruction (SSR) [38], or lagged cross-correlation [2, 29]. These techniques have found application in a wide range of fields including neuroscience (e.g., [19]), economics (e.g., [6, 7]), and climatology (e.g., [25]).

In this article, we introduce a time series causality technique derived directly from the definition of probabilistic causality [41]. The technique is applied to synthetic and empirical bivariate time series data sets with known, or intuitive, causal relationships. We discuss the strengths and weaknesses of the technique and demonstrate how it may be useful for causal inference with empirical data

from systems in which it is difficult or impossible to implement controlled experiments or to control interventions.

## II. CAUSAL PENCHANT

We define the *causal penchant*  $\rho_{EC} \in [1, -1]$  as

$$\rho_{EC} := P(E|C) - P(E|\bar{C}). \quad (1)$$

The motivation for this expression is in the straightforward interpretation of  $\rho_{EC}$  as a causal indicator [26]; i.e., if  $C$  causes (or *drives*)  $E$ , then  $\rho_{EC} > 0$ , and if  $\rho_{EC} \leq 0$ , then the direction of causal influence is undetermined. If effect  $E$  is assumed to be measured in one time series and the cause  $C$  is assumed to be measured in a different time series, then the direction of causal influence can be determined by comparing various penchants when each time series is assigned to be the cause  $C$ .

Eqn. 1 can be rewritten using Bayes’ theorem

$$P(E|C) = P(C|E) \frac{P(E)}{P(C)} \quad (2)$$

and the definitions of probability complements

$$P(\bar{C}) = 1 - P(C) \quad (3)$$

$$P(\bar{C}|E) = 1 - P(C|E). \quad (4)$$

Using Eqn. 4 with Eqn. 2 gives

$$P(\bar{C}|E) = 1 - P(C|E) \frac{P(C)}{P(E)}$$

Inserting this into Eq. 2 written in terms of  $\bar{C}$ ,

$$P(E|\bar{C}) = P(\bar{C}|E) \frac{P(E)}{P(\bar{C})}$$

yields an alternative form of the second term in Eqn. 1

$$P(E|\bar{C}) = \left(1 - P(C|E) \frac{P(C)}{P(E)}\right) \frac{P(E)}{1 - P(C)},$$

---

\* jmcrcrac1@masonlive.gmu.edu

† rweigel@gmu.edu

This expression gives a penchant that requires only a single conditional probability estimate:

$$\rho_{EC} = P(E|C) \left( 1 + \frac{P(C)}{1 - P(C)} \right) - \frac{P(E)}{1 - P(C)} . \quad (5)$$

The penchant is non-parametric in the sense that there is no assumed functional relationship between the time series being investigated. This form of the penchant will be used along with a structured method for counting and notions from probabilistic causality to infer which time series in a given pair might be seen as “driving” the other. Our motivation for the penchant is the need for a time series causality quantity that is easily computed.

For the calculations in the following sections, the penchant is not defined if  $P(C)$  or  $P(\bar{C})$  are zero (because the conditionals in Eqn. 1 would be undefined). Thus, the penchant is not defined if  $P(C) = 0$  or if  $P(C) = 1$ . The former condition corresponds to an inability to determine causal influence between two time series when a cause does not appear in one of the series; the latter condition is interpreted as an inability to determine causal influence between two time series if one is constant. The use of Bayes’ theorem in the derivation of Eqn. 5 implies that the penchant is not defined if  $P(E)$  or  $P(\bar{E})$  are zero.

The method given in this work uses no *a priori* assignment of “cause” or “effect” to a given time series pair when using penchants for causal inference. So, operationally, the constraints on  $P(C)$  and  $P(E)$  only mean that the penchant is undefined between pairs of time series where one series is constant.

The penchant definition includes  $P(E|\bar{C})$ , which is the probability of an assumed effect occurring given a absence of the assumed cause. It has been argued that causality determination requires an intervention, and the absence of an assumed cause is unobservable, which implies the occurrence probability of the assumed effect should be conditioned on performing or not performing an action rather than on the presence or lack of an assumed cause [30]. Causal relations have been described as “a relation among events” [3], again implying the absence of an assumed cause cannot be used to identify causal relationships. These issues have been a part of probabilistic definitions of causality at least since the 1960s [41], and we do not attempt to solve them in this article. We circumvent these philosophical issues by using an expression that removes any conditioning on the absence of an assumed cause and the condition that the penchant is undefined when  $P(C) = 0$ ,  $P(C) = 1$ ,  $P(E) = 0$ , or  $P(E) = 1$ .

Although Eqn. 5 circumvents the issue of  $P(E|\bar{C})$  being unobservable, it does not account for confounding. The assumption that  $P(C)$  can be estimated from a scalar time series may be seen as an oversimplification of the dynamics. That is, it may be seen as an assumption that the assumed effect is only caused by the assumed cause. In this case, the penchant may be better interpreted as an indication of predictability rather than

causality (similar to arguments made regarding Granger causality [38]). This issue will not be addressed in this article; we emphasize, however, that we use terms such as cause, effect, causal inference, and related terms to specifically refer to the penchant and leaning quantities.

In this article, we seek to determine if the penchant is a useful quantity for the identification of causality relationships between time series. Our goal is identify the usefulness of the penchant (and leaning introduced below) for *exploratory causal inference*, i.e., inference intended to determine if (and what) causal structure may be present in a time series pair but not to *prove* or *confirm* such structure. There are scenarios in which any time series causality tool such as the penchant may incorrectly assign causal structure or may incorrectly not assign causal structure [16]. Furthermore, proof of causal relationships is often considered impossible with data alone [16, 17, 30]. The goal of this work is to draw as much information as possible from the given data to, e.g., guide the design of future experiments.

### III. CAUSAL LEANING

Consider the assignment of  $\mathbf{X}$  as the cause,  $C$ , and  $\mathbf{Y}$  as the effect,  $E$ . If  $\rho_{EC} > 0$ , then the probability that  $\mathbf{X}$  drives  $\mathbf{Y}$  is higher than the probability that it does not, which is stated more succinctly as  $\mathbf{X}$  has a penchant to drive  $\mathbf{Y}$  or  $\mathbf{X} \xrightarrow{pen} \mathbf{Y}$ .

It is possible, however, that the penchant could be positive when  $\mathbf{X}$  is assumed as the effect and  $\mathbf{Y}$  is assumed as the cause. (An example of this is given in Section IV C.) The *leaning* addresses this via

$$\lambda_{EC} := \rho_{EC} - \rho_{CE} \quad (6)$$

for which  $\lambda_{EC} \in [-2, 2]$ . A positive leaning implies the assumed cause  $C$  drives the assumed effect  $E$  more than the assumed effect drives the assumed cause, a negative leaning implies the effect  $E$  drives the assumed cause  $C$  more than the assumed cause drives the assumed effect, and a zero leaning yields no causal information.

The possible outcomes are notated as

$$\begin{aligned} \lambda_{EC} > 0 \quad \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} &\Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{Y} \\ \lambda_{EC} < 0 \quad \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} &\Rightarrow \mathbf{Y} \xrightarrow{lean} \mathbf{X} \\ \lambda_{EC} = 0 \quad \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} &\Rightarrow \text{no conclusion} \end{aligned}$$

with  $\{C, E\} = \{\mathbf{A}, \mathbf{B}\}$  meaning  $\mathbf{A}$  is the assumed cause and  $\mathbf{B}$  as the assumed effect.

If  $\lambda_{EC} > 0$  with  $\mathbf{X}$  as the assumed cause and  $\mathbf{Y}$  as the assumed effect, then  $\mathbf{X}$  has a larger penchant to drive  $\mathbf{Y}$  than  $\mathbf{Y}$  does to drive  $\mathbf{X}$ . That is,  $\lambda_{EC} > 0$  implies that the difference between the probability that  $\mathbf{X}$  drives  $\mathbf{Y}$  and the probability that it does not is higher than the difference between the probability that  $\mathbf{Y}$  drives  $\mathbf{X}$  and the probability that it does not.

The leaning is a function of four probabilities,  $P(C)$ ,  $P(E)$ ,  $P(C|E)$ , and  $P(E|C)$ . The usefulness of the leaning for causal inference will depend on an effective method for estimating these probabilities from times series and a more specific definition of the cause-effect assignment within the time series pair. An operational definition of  $C$  and  $E$  will need to be drawn directly from the time series data if the leaning is to be useful for causal inference. Such assignments, however, may be difficult to develop and may be considered arbitrary without some underlying theoretical support. For example, if the cause is  $x_{t-1}$  and the effect is  $y_t$ , then it may be considered unreasonable to provide a causal interpretation of the leaning without theoretical support that  $\mathbf{X}$  may be expected to drive  $\mathbf{Y}$  on the time scale of  $\Delta t = 1$ . This issue is, however, precisely one of the reasons for divorcing the causal inference proposed in this work (i.e., exploratory causal inference) from traditional ideas of causality, as was explained in the second paragraph of the introduction. Statistical tools are associational, and cannot be given formal causal interpretation without the use of assumptions and outside theories (see [16] for an in-depth discussion of these ideas). In practice, many different potential cause-effect assignments may be used to calculate different leanings, which may then be compared as part of the causal analysis of the data.

In this article, the probabilities required for the leaning calculation will be estimated from the data straightforwardly through a counting/binning procedure, and the cause-effect assignment may be varied but will always use a simple lag structure to avoid unnecessarily complex computations.

#### IV. MOTIVATING EXAMPLE

Consider a time series pair  $\{\mathbf{X}, \mathbf{Y}\}$  with

$$\begin{aligned}\mathbf{X} &= \{x_t \mid t = 0, 1, 2, \dots, 9\} \\ &= \{0, 0, 1, 0, 0, 1, 0, 0, 1, 0\} \\ \mathbf{Y} &= \{y_t \mid t = 0, 1, 2, \dots, 9\} \\ &= \{0, 0, 0, 1, 0, 0, 1, 0, 0, 1\}.\end{aligned}$$

Because  $y_t = x_{t-1}$ , one may conclude that  $\mathbf{X}$  drives  $\mathbf{Y}$ . However, to show this result using a leaning calculation requires first a calculation using the cause-effect assignment  $\{C, E\} = \{\mathbf{X}, \mathbf{Y}\}$ . For consistency with the intuitive definition of causality, we require that a cause must precede an effect. It follows that a natural assignment may be  $\{C, E\} = \{x_{t-l}, y_t\}$  for  $1 \leq l < t \leq 9$ . This cause-effect assignment will be referred to as the  $l$ -standard assignment.

The cause-effect assignment is an assignment of a given structure or feature of the data in one time series as the “cause” and another structure or feature of the data in the other time series as the “effect”. For example, in the  $l$ -standard cause-effect assignment, the cause is the lag  $l$  time step in one time series and the effect is the

current time step in the other. The leaning compares the symmetric application of these cause-effect definitions to the time series pair. So, for the above example of  $\{C, E\} = \{x_{t-l}, y_t\}$ , the first penchant will be calculated using  $\{C, E\} = \{x_{t-l}, y_t\}$  and the second will be calculated using  $\{C, E\} = \{y_{t-l}, x_t\}$ . The second penchant is not the direct interchange of  $C \Leftrightarrow E$  from the first penchant because such an interchange would violate the assumption that a cause must precede an effect. For example, if the first penchant in the leaning calculation is calculated using  $\{C, E\} = \{x_{t-l}, y_t\}$ , then the second penchant is not calculated using  $\{C, E\} = \{y_t, x_{t-l}\}$  because the definition of the effect,  $x_{t-l}$ , precedes the definition of the cause,  $y_t$ .

#### A. Defining penchants

Given  $\{\mathbf{X}, \mathbf{Y}\}$ , one possible penchant (i.e., Eqn. 5) that can be defined using the 1-standard assignment is

$$\rho_{y_t=1, x_{t-1}=1} = \kappa \left( 1 + \frac{P(x_{t-1}=1)}{1 - P(x_{t-1}=1)} \right) - \frac{P(y_t=1)}{1 - P(x_{t-1}=1)},$$

with  $\kappa = P(y_t=1|x_{t-1}=1)$ . Another penchant defined using this assignment is  $\rho_{y_t=0, x_{t-1}=0}$  with  $\kappa = P(y_t=0|x_{t-1}=0)$ . These two penchants are called *observed* penchants because they correspond to conditions that were found in the measurements.

Two other penchants have  $\kappa = P(y_t=0|x_{t-1}=1)$  and  $\kappa = P(y_t=1|x_{t-1}=0)$ . These penchants are associated with unobserved conditions. Based on the values for these two penchants,  $\kappa = 0 \Rightarrow \rho_{y_t x_{t-1}} < 0$ , which is consistent with the claim that the effect,  $y_t = 0$  or  $1$  is not caused by the postulated cause,  $x_{t-1} = 1$  or  $0$ , respectively.

#### B. Computing penchants

The probabilities in the penchant calculations can be estimated from time series using counts, e.g.,

$$P(y_t=1|x_{t-1}=1) = \frac{n_{EC}}{n_C} = \frac{3}{3} = 1,$$

where  $n_{EC}$  is the number of times  $y_t = 1$  and  $x_{t-1} = 1$  appears in  $\{\mathbf{X}, \mathbf{Y}\}$ , and  $n_C$  is the number of times the assumed cause,  $x_{t-1} = 1$ , has appeared in  $\{\mathbf{X}, \mathbf{Y}\}$ .

Estimating the other two probabilities in this penchant calculation using frequency counts from  $\{\mathbf{X}, \mathbf{Y}\}$  requires accounting for the assumption that the cause must precede the effect by shifting  $\mathbf{X}$  and  $\mathbf{Y}$  into  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$  such that, for any given  $t$ ,  $\tilde{x}_t$  precedes  $\tilde{y}_t$ .

For this example, the shifted sequences are

$$\begin{aligned}\tilde{\mathbf{X}} &= \{0, 0, 1, 0, 0, 1, 0, 0, 1\} \\ \tilde{\mathbf{Y}} &= \{0, 0, 1, 0, 0, 1, 0, 0, 1\}\end{aligned}$$

which are both shorter than their counterparts above by a single value because the pendants are being calculated using the 1-standard cause-effect assignment. It follows that  $\tilde{x}_t = x_{t-1}$  and  $\tilde{y}_t = y_t$ .

The probabilities are then

$$P(y_t = 1) = \frac{n_E}{L} = \frac{3}{9} \quad (7)$$

and

$$P(x_{t-1} = 1) = \frac{n_C}{L} = \frac{3}{9}, \quad (8)$$

where  $n_C$  is the number of times  $\tilde{x}_t = 1$ ,  $n_E$  is the number of times  $\tilde{y}_t = 1$ , and  $L$  is the (“library”) length of  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$ .

### C. Mean observed leaning

The two observed pendants in this example under the assumption that  $\mathbf{X}$  causes  $\mathbf{Y}$  (with  $l = 1$ ) are

$$\rho_{y_t=1, x_{t-1}=1} = 1 \quad (9)$$

and

$$\rho_{y_t=0, x_{t-1}=0} = 1.$$

The observed pendants when  $\mathbf{Y}$  is assumed to cause  $\mathbf{X}$  are

$$\begin{aligned} \rho_{x_t=1, y_{t-1}=0} &= \frac{3}{7}, \\ \rho_{x_t=0, y_{t-1}=1} &= \frac{3}{7}, \end{aligned}$$

and

$$\rho_{x_t=0, y_{t-1}=0} = -\frac{3}{7}.$$

The *mean observed pendant* is the algebraic mean of the observed pendants. For  $\mathbf{X}$  causes  $\mathbf{Y}$ , it is

$$\begin{aligned} \langle \rho_{y_t, x_{t-1}} \rangle &= \frac{1}{2} (\rho_{y_t=1, x_{t-1}=1} + \rho_{y_t=0, x_{t-1}=0}) \\ &= 1 \end{aligned}$$

and for  $\mathbf{Y}$  causes  $\mathbf{X}$  is

$$\begin{aligned} \langle \rho_{x_t, y_{t-1}} \rangle &= \frac{1}{3} (\rho_{x_t=1, y_{t-1}=0} \\ &\quad + \rho_{x_t=0, y_{t-1}=1} + \rho_{x_t=0, y_{t-1}=0}) \\ &= \frac{1}{7}. \end{aligned}$$

The *mean observed leaning* that follows from the definition of the mean observed pendants is

$$\langle \lambda_{y_t, x_{t-1}} \rangle = \langle \rho_{y_t, x_{t-1}} \rangle - \langle \rho_{x_t, y_{t-1}} \rangle \quad (10)$$

$$= \frac{6}{7}. \quad (11)$$

The positive leaning implies the probability that  $x_{t-1}$  drives  $y_t$  is higher than the probability that  $y_{t-1}$  drives  $x_t$ ; i.e.,  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  given the 1-standard cause-effect assignment. This result is expected and agrees with the intuitive definition of causality in this example.

The *weighted mean observed pendant* is defined similarly to the mean observed pendant, but each pendant is weighted by the number of times it appears in the data; e.g.,

$$\begin{aligned} \langle \rho_{y_t, x_{t-1}} \rangle_w &= \frac{1}{L} (n_{y_t=1, x_{t-1}=1} \rho_{y_t=1, x_{t-1}=1} \\ &\quad + n_{y_t=0, x_{t-1}=0} \rho_{y_t=0, x_{t-1}=0}) \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} \langle \rho_{x_t, y_{t-1}} \rangle_w &= \frac{1}{L} (n_{x_t=1, y_{t-1}=0} \rho_{x_t=1, y_{t-1}=0} \\ &\quad + n_{x_t=0, y_{t-1}=1} \rho_{x_t=0, y_{t-1}=1} \\ &\quad + n_{x_t=0, y_{t-1}=0} \rho_{x_t=0, y_{t-1}=0}) \\ &= \frac{3}{63}, \end{aligned}$$

where  $n_{a,b}$  is the number of times the assumed cause  $a$  appears with the assumed effect  $b$  and  $L$  is the library length of  $\tilde{\mathbf{X}}$  (i.e.,  $L = N - l$  where  $N$  is the library length of  $\mathbf{X}$  and  $l$  is the lag used in the  $l$ -standard cause-effect assignment).

The *weighted mean observed leaning* follows naturally as

$$\begin{aligned} \langle \lambda_{y_t, x_{t-1}} \rangle_w &= \langle \rho_{y_t, x_{t-1}} \rangle_w - \langle \rho_{x_t, y_{t-1}} \rangle_w \\ &= \frac{60}{63}. \end{aligned}$$

For this example,  $\langle \lambda_{y_t, x_{t-1}} \rangle_w \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  as expected.

Conceptually, the weighted mean observed pendant is preferred to the mean pendant because it accounts for the frequency of observed cause-effect pairs within the data, which is assumed to be a predictor of causal influence. For example, given some pair  $\{\mathbf{A}, \mathbf{B}\}$ , if it is known that  $a_{t-1}$  causes  $b_t$  and both  $b_t = 0 \mid a_{t-1} = 0$  and  $b_t = 0 \mid a_{t-1} = 1$  are observed, then comparison of the frequencies of occurrence is used to determine which of the two pairs represents the cause-effect relationship.

For this example, the weighted mean observed leaning provides the same causal inference as the mean observed leaning. The weighted mean calculation will be used in the examples of the following sections.

### D. Unobserved pendants

The *unobserved* pendants for  $l = 1$  for  $\mathbf{X}$  causes  $\mathbf{Y}$  are

$$\begin{aligned} \rho_{y_t=1, x_{t-1}=0} &= -1 \\ \rho_{y_t=0, x_{t-1}=1} &= -1 \end{aligned}$$

and for  $\mathbf{Y}$  causes  $\mathbf{X}$  is

$$\rho_{x_t=1, y_{t-1}=1} = -\frac{3}{7}.$$

These values can be incorporated into the averaging calculation to yield a *mean total penchant*; i.e., for  $\mathbf{X}$  causes  $\mathbf{Y}$

$$\begin{aligned} \langle \langle \rho_{y_t, x_{t-1}} \rangle \rangle &= \frac{1}{4} (\rho_{y_t=1, x_{t-1}=1} + \rho_{y_t=0, x_{t-1}=0} \\ &\quad \rho_{y_t=1, x_{t-1}=0} + \rho_{y_t=0, x_{t-1}=1}) \\ &= 0 \end{aligned}$$

and for  $\mathbf{Y}$  causes  $\mathbf{X}$

$$\begin{aligned} \langle \langle \rho_{x_t, y_{t-1}} \rangle \rangle &= \frac{1}{4} (\rho_{x_t=1, y_{t-1}=1} + \rho_{x_t=0, y_{t-1}=0} \\ &\quad \rho_{x_t=1, y_{t-1}=0} + \rho_{x_t=0, y_{t-1}=1}) \\ &= 0. \end{aligned}$$

Thus, the *mean total leaning* (defined analogous to Eqn. 10) is  $\langle \langle \lambda_{y_t, x_{t-1}} \rangle \rangle = \langle \langle \rho_{y_t, x_{t-1}} \rangle \rangle - \langle \langle \rho_{x_t, y_{t-1}} \rangle \rangle = 0$ . No causal inference can be made with a leaning of zero because it implies  $\langle \langle \rho_{y_t, x_{t-1}} \rangle \rangle = \langle \langle \rho_{x_t, y_{t-1}} \rangle \rangle$ . Thus  $\mathbf{X}$  does not have a higher penchant to drive  $\mathbf{Y}$  than  $\mathbf{Y}$  does to drive  $\mathbf{X}$ , given the cause-effect assignment used in the leaning calculation. Such a conclusion would not be useful for causal inference, which implies the mean total leaning is not useful for causal inference in this example.

### E. Cause-effect assignment independence

The causal inference above assumed a cause-effect relationship was known to be correct. It can be shown, however, that causal inference is independent of the assumed cause-effect relationship. For example, consider the cause-effect assignment  $\{C, E\} = \{y_{t-l}, x_t\}$  with  $l = 1$ . The mean observed leaning is

$$\begin{aligned} \langle \lambda_{x_t, y_{t-1}} \rangle &= \langle \rho_{x_t, y_{t-1}} \rangle - \langle \rho_{y_t, x_{t-1}} \rangle \\ &= -\frac{6}{7}, \end{aligned}$$

which implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ , as expected for this example.

In general,  $\lambda_{AB} := \rho_{AB} - \rho_{BA} \Rightarrow -\lambda_{AB} = \rho_{BA} - \rho_{AB} := \lambda_{BA}$ . Thus, the causal inference is independent of which time series is initially assumed to be the cause (or effect).

### F. Tolerance domains

If the example time series contained noise, then a re-aliasation of the example time series  $\{\mathbf{X}', \mathbf{Y}'\}$  could be

$$\begin{aligned} \mathbf{X}' &= \{x'_t \mid t = 0, 1, 2, \dots, 9\} \\ &= \{0, 0, 1.1, 0, 0, 1, -0.1, 0, 0.9, 0\} \\ \mathbf{Y}' &= \{y'_t \mid t = 0, 1, 2, \dots, 9\} \\ &= \{0, -0.2, 0.1, 1.2, 0, 0.1, 0.9, -0.1, 0, 1\}. \end{aligned}$$

The previous time series pair,  $\{\mathbf{X}, \mathbf{Y}\}$  had only five observed penchants, but  $\{\mathbf{X}', \mathbf{Y}'\}$  has more due to the noise. It can be seen in the time series definitions that  $x'_t = x_t \pm 0.1 := x_t \pm \delta_x$  and  $y'_t = y_t \pm 0.2 := y_t \pm \delta_y$ . The weighted mean observed leaning for  $\{\mathbf{X}', \mathbf{Y}'\}$  is  $\langle \lambda_{y'_t, x'_{t-1}} \rangle_w \approx 0.19$ .

If the noise is not restricted to a small set of discrete values, then the effects of noise on the leaning calculations can be addressed by using the tolerances  $\delta_x$  and  $\delta_y$  in the probability estimations from the data. For example, the penchant calculation in Eqn. 9 relied on estimating  $P(y_t = 1 | x_{t-1} = 1)$  from the data, but if, instead, the data is known to be noisy, then the relevant probability estimate may be  $P(y_t \in [1 - \delta_y, 1 + \delta_y] | x_{t-1} \in [1 - \delta_x, 1 + \delta_x])$ .

If the tolerances,  $\delta_x$  and  $\delta_y$ , are made large enough, then the noisy system weighted mean observed leaning,  $\langle \lambda_{y'_t \pm \delta_y, x'_{t-1} \pm \delta_x} \rangle_w$ , can, at least in the simple examples considered here, be made equal to the noiseless system weighted mean observed leaning, i.e.,  $\langle \lambda_{y'_t \pm \delta_y, x'_{t-1} \pm \delta_x} \rangle_w = \langle \lambda_{y_t, x_{t-1}} \rangle_w$ .

Tolerance domains, however, can be set too large. If the tolerance domain is large enough to encompass every point in the time series, then the probability of the assumed cause becomes one, which leads to undefined penchants. For example, given the symmetric definition of the tolerance domain used in this section,  $\delta_x = 2$  implies  $P(x_{t-1} = 1 \pm \delta_x) = 1$ , which implies  $\langle \lambda_{y'_t, x_{t-1}} \rangle_w$  is undefined.

This example was used to motivate the need for an understanding of the noise in the measurements, which may not always be possible. If little is known about the noise, one strategy is to calculate the leanings with several different tolerances, increasing the size of the tolerance domains to the point where the penchants become undefined, and finding the tolerance domains for which the leaning changes sign. The sizes of these domains can then be compared to suspected noise levels. This strategy, and others, will be considered in more detail in the following sections. If the noise level is known, then the task becomes much simpler and the tolerances should just be set to the known (or estimated) noise levels for the individual time series.

As mentioned in Section III, the probabilities required for the leaning calculations are estimated in this example (and all the following ones) through straightforward counting of the data. As such, the tolerance domains may be thought of as bin widths for the probability estimations. Work has been done on optimal and data-driven bin width selection procedures (see, e.g., [36, 42]), usually in the context of finding histograms. Tolerance domains, however, may be thought of in terms of the causal inference for which the leaning is intended. The tolerance domain for the “cause” (or “effect”) is the domain in which an analyst considers the data may still reasonably be identified as a “cause” (or “effect”). It is not required to be symmetric, and the tolerance domain for one time series is not required to be equal to the tolerance domain

for the other (which is seen in the example above and most of the examples that follow).

### G. Stationarity dependence

Both  $\mathbf{X}$  and  $\mathbf{Y}$  are stationary in the original example time series pair  $\{\mathbf{X}, \mathbf{Y}\}$ . Suppose 1,000 zeros are appended to the end of each of these time series. The additional zeros in the times series may intuitively seem to make causal inference more difficult. The probabilities required for the penchant  $\rho_{y_t=1|x_{t-1}=1}$  become

$$P(y_t = 1|x_{t-1} = 1) = \frac{3}{3} = 1, \\ P(y_t = 1) = \frac{3}{1009},$$

and

$$P(x_{t-1} = 1) = \frac{3}{1009}.$$

These probabilities have become much smaller but the penchant remains the same. The same is true for  $\rho_{y_t=0|x_{t-1}=0}$ . Despite the additional zeros,  $\mathbf{Y}$  can still only take the values 1 or 0. This knowledge along with the above penchants implies  $n_{y_t=1, x_{t-1}=1} + n_{y_t=0, x_{t-1}=0} = L$ , which implies  $\langle \rho_{y_t, x_{t-1}} \rangle_w = 1$ . The other three observed penchants, however, do change as a result of the appended zeros. Previously,  $|\rho_{x_t=1, y_{t-1}=0}| = |\rho_{x_t=0, y_{t-1}=1}| = |\rho_{x_t=0, y_{t-1}=0}| = 3/7$ , but with the appended zeros,  $|\rho_{x_t=1, y_{t-1}=0}| = |\rho_{x_t=0, y_{t-1}=1}| = |\rho_{x_t=0, y_{t-1}=0}| = 3/1009$ . The weighted mean observed leaning,  $\langle \lambda_{y_t, x_{t-1}} \rangle_w$ , changes from 60/63 to approximately 1012/1009 because of the appended zeros. This value is higher than the previous value but yields the same causal inference.

Consider another non-stationary times series pair,  $\{\mathbf{X}_L, \mathbf{R}_L\}$ , where the non-stationary response signal is  $\mathbf{R}_L = \{0, 0, 0, 1, 1, 1, 2, 2, 2, 3\}$ . The weighted mean observed leaning calculated under the 1-standard assignment with no tolerance domains still leads to a causal inference that agrees with intuition; i.e.  $\langle \lambda_{r_t, x_{t-1}} \rangle_w \approx 0.11 \Rightarrow \mathbf{X}_L \xrightarrow{\text{lean}} \mathbf{R}_L$  as expected. This result, however, depends on the library length of the data.

$\{\mathbf{X}_L, \mathbf{R}_L\}$  is a specific instance of the following time series pair:

$$\{\mathbf{X}, \mathbf{R}\} = \{\{x_t\}, \{r_t\}\} \quad (12)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \begin{cases} 0 & \forall t \in \{t \mid t \bmod 3 \neq 0\} \\ 1 & \forall t \in \{t \mid t \bmod 3 = 0\} \end{cases} \quad (13)$$

and

$$r_t = x_{t-1} + r_{t-1} \quad (14)$$

with  $r_0 = 0$ . The weighted mean observed leaning, under the 1-standard assignment with no tolerance domains,

for  $\{\mathbf{X}, \mathbf{R}\}$  depends on  $L$ . As  $L$  is increased, the leaning calculation will eventually lead to causal inferences that do not agree with intuition; e.g.,  $L = 20 \Rightarrow \langle \lambda_{r_t, x_{t-1}} \rangle_w \approx 1.8 \times 10^{-3} \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{R}$  and  $L = 50 \Rightarrow \langle \lambda_{r_t, x_{t-1}} \rangle_w \approx -2.5 \times 10^{-3} \Rightarrow \mathbf{R} \xrightarrow{\text{lean}} \mathbf{X}$ .

As  $L$  is increased, the number of possible observed effects for a given observed cause increases. Thus, under the 1-standard assignment  $\{C, E\} = \{x_{t-1}, r_t\}$ ,  $x_{t-1} = 1$  precedes three different values,  $r_t = 1, 2$ , and  $3$ , if  $L = 10$ , but it precedes fifteen different values if  $L = 50$ . The leaning calculations are methods for counting (in a specific way) the number of times (and ways in which) an observed cause-effect pair appears in the data. The causal inference becomes more difficult for non-stationary time series pairs because repeated cause-effect pairs in the data may be more rare than in the stationary examples. This effect is very similar to the effect seen when the impulse signal was noiseless but the response was noisy. Unfortunately, it cannot be remedied with tolerance domains for the non-stationary case. For example, for  $\{\mathbf{X}, \mathbf{R}\}$ , the cardinality of the set  $\{r_t \mid x_{t-1} = 1\} \rightarrow \infty$  as  $L \rightarrow \infty$ , and penchants would not be defined given a tolerance domain for  $\mathbf{R}$  of  $\delta_r = \infty$ .

These shortcomings of the weighted mean observed leaning when applied to non-stationary data, however, do not imply that causal inference of non-stationary data cannot be done using a different application of the observed penchants. For example, replacing the weighted mean calculation in the weighted mean observed leaning calculation with a median calculation leads to a *median observed leaning*,  $[\lambda_{r_t, x_{t-1}}] \approx 5.3 \times 10^{-3} \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{R}$  for  $L = 50$  as expected, where  $[\cdot]$  is used to denote the median. Of course, even though the median leaning calculation agrees with intuition for a library length where the mean leaning calculation did not, there is no reason to believe the median leaning calculation will not also eventually provide counterintuitive causal inferences as  $L$  is increased.

A more basic strategy for dealing with non-stationary data would be to define the observed penchant using a different cause-effect assignment. For example, the  $l$ -standard assignment (with  $l = 1$ ) used above, i.e.,  $\{C, E\} = \{x_{t-1}, r_t\}$ , might be replaced with an  $l$ -AR (autoregressive) assignment with  $l = 1$  of  $\{C, E\} = \{(x_{t-1}, r_{t-1}), r_t\}$ . An observed penchant may be calculated with an assumed cause of  $(x_{t-1} = 1, r_{t-1} = 0)$  and an assumed effect of  $r_t = 1$ . The algorithms to compute the observed penchants from the data become more complicated as the cause-effect assignment becomes more complicated, but the basic definition of the penchant provides a very general conceptual framework for causal inference.

## V. SIMPLE EXAMPLE SYSTEMS

In this section the weighted mean observed leaning using the  $l$ -standard cause-effect assignment for various  $l$ , will be applied to dynamical systems and empirical data sets with known causal relationships. The usefulness of the leaning as a tool for causal inference is tested directly with synthetic and empirical time series data sets for which there is an intuitive understanding of the driving relationships within the system.

### A. Impulse with Noisy Response Linear Example

Consider the linear example dynamical system of

$$\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \quad (15)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \begin{cases} 2 & t = 1 \\ 0 & \forall t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

and

$$y_t = x_{t-1} + B\eta_t$$

with  $y_0 = 0$ ,  $B \in \mathbb{R} \geq 0$  and  $\eta_t \sim \mathcal{N}(0, 1)$ . Specifically, consider  $B \in [0, 1]$ . The driving system  $\mathbf{X}$  is a periodic impulse with a signal amplitude above the maximum noise level of the response system, and the response system  $\mathbf{Y}$  is a lagged version of the driving signal with  $\mathcal{N}(0, 1)$  of amplitude  $B$  applied at each time step.

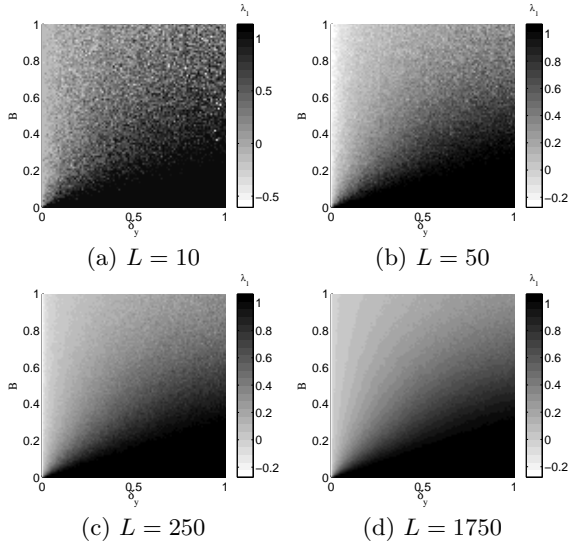


FIG. 1: The unitless leaning of Eqn. 15 is a function of the noise, the tolerance used for terms from  $\mathbf{Y}$ , and the length of the signals and both  $\delta_y$  and  $B$  are unitless. See the text for an explanation of the missing data for large  $\delta_y$ .

$B$	$\tilde{\lambda}$		
	method 1	method 2	method 3
0.0	1.0	1.0	1.0
0.1	0.40	1.0	0.48
0.5	0.39	0.79	0.26
0.8	0.30	0.44	0.10

TABLE I:  $\tilde{\lambda}$  using three different estimation methods for  $\delta_y$ : (1) lagged linear response deviation, (2) normalized standard deviation, and (3)  $n$ -bin mean standard deviation.

Figure 1 shows how the weighted mean observed leaning using the 1-standard cause-effect assignment,  $\tilde{\lambda}$ , changes as the noise amplitude  $B$  and tolerance  $\delta_y$  are increased in increments of 0.01. The synthetic data sets  $\mathbf{X}$  and  $\mathbf{Y}$  are constructed such that intuitively  $\mathbf{X}$  drives  $\mathbf{Y}$ . Thus, it is expected that  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  which implies  $\tilde{\lambda} > 0$ . Figure 1 shows that this expectation is met except when  $\delta_y > B$  even for a short library length of  $L = 10$ . Examples of undefined penchants due to large tolerance domains, as discussed in section IV F, are seen as  $\delta_y$  is increased in the  $L = 10$  example.

Figure 1 shows using the strategy of  $\delta_y = B$  always leads to causal inferences that agree with intuition for  $L > 10$  in this example. However, as discussed in section IV F, knowing  $B$  *a priori* may be unrealistic with empirical data sets. Consider the following three methods for estimating  $\delta_y$  from the data:

1. *lagged linear response deviation* -  $\delta_y$  is set to the mean absolute deviation of  $y_t$  from  $x_{t-1}$ ; i.e.,  $\delta_y = \langle |y_t - x_{t-1}| \rangle$ .
2. *normalized standard deviation* -  $\delta_y$  is set to the standard deviation of  $\{|\mathbf{Y} - \langle \mathbf{Y} \rangle|\}$  where  $\langle \mathbf{Y} \rangle$  is the mean of  $\mathbf{Y}$ ; i.e.,  $\delta_y = \sigma_{|y_t - \langle y_t \rangle|}$ .
3.  *$n$ -bin mean standard deviation* -  $\delta_y$  is set to the mean standard deviation of  $n$  bins of  $\mathbf{Y}$ ; i.e.,  $\delta_y = \langle \sigma_{B_i} \rangle$  where  $B_i$  is the  $i$ th bin of an  $n$ -bin histogram of  $\mathbf{Y}$ .

Table I shows  $\tilde{\lambda}$  for instances of Eqn. 15 with  $B = 0, 0.1, 0.5$ , and  $0.8$  and  $L = 100$  (and  $n = 5$  in method 3).

The three different methods yield different values for the leaning, but all the methods lead to the same causal inference,  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ , which agrees with intuition. These methods are meant to be examples of using the data to set  $\delta_y$  if  $B$  is not known. These methods are not expected to be reasonable estimates for  $\delta_x$  and  $\delta_y$  in general. For example, method 1 assumes a linear relationship between  $\mathbf{X}$  and  $\mathbf{Y}$  that may be unreasonable to assume in general. However, Table I shows different methods for setting  $\delta_y$  can lead to the same causal inference. Setting the tolerances requires an understanding of the noise in the time series data. The leaning is meant to be part of an exploratory causal analysis of the time series data and can-

not exist independently of other exploratory analysis of the data, including analysis of the noise levels.

### B. Cyclic Linear Example

The calculations in the previous subsection were only for the 1-standard assignment ( $l = 1$ ) and are expected to be useful for causal inference given Eqn. 15. However, deciding which  $l$ -standard assignment to use given empirical, rather than synthetic, data sets may be more difficult. It is expected that several different  $l$ -standard assignments would be used as part of any exploratory causal analysis using leaning. This section contains an example that plots the leanings for a set of different  $l$ -assignments and shows the maximum leaning in the set is near the expected value, i.e., near the lag value that appears explicitly in the dynamical system used to create the synthetic data sets.

Consider the linear example dynamical system of

$$\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \quad (16)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \sin(t)$$

and

$$y_t = x_{t-1} + B\eta_t$$

with  $y_0 = 0$ ,  $B \in [0, 1]$  in steps of 0.01 and  $\eta_t \sim \mathcal{N}(0, 1)$ . This example is very similar to the previous one, except that the driving system  $\mathbf{X}$  is sinusoidal.

Figures 2 and 3 were calculated for an instance of Eqn. 16 with  $L = 41$  generated by sampling one period of  $\mathbf{X}$  with  $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 2\pi\}$  and  $f = 1/20$ . Figure 2 shows the weighted mean observed leaning using the 1-standard assignment,  $\lambda$ , is always positive given  $\delta_y = B$ . So, as was seen in the previous example, the leaning implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ , which agrees with intuition for this example.

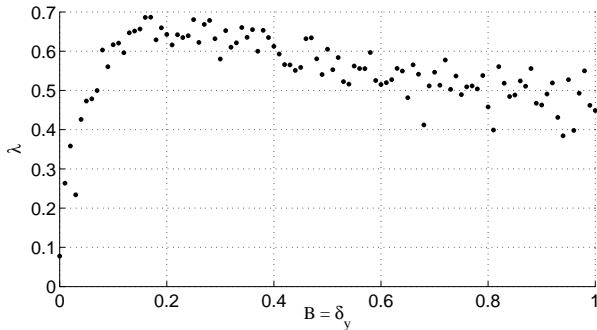


FIG. 2: Weighted mean observed leaning for Eqn. 16 and a tolerance for the leaning calculation set to  $\delta_y = B$ .  $\lambda$  is always positive, which implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ .

The driving relationship in this example can be difficult to discern using unmodified CCM techniques [24]. It has been argued that lagged cross-correlation techniques are the preferred causal inference tool in most situations because of their simplicity [10]. The lagged cross-correlation is defined as

$$\chi_{xy}^l = \frac{\mathbb{E}[(x_t - \mu_x)(y_{t-l} - \mu_y)]}{\sigma_x \sigma_y}, \quad (17)$$

where  $\mathbb{E}[z_t]$  is the expectation value of  $\{z_t\}$ ,  $\mu_{x(y)}$  is the mean of  $\mathbf{X}$  ( $\mathbf{Y}$ ), and  $\sigma_{x(y)}$  is the standard deviation of  $\mathbf{X}$  ( $\mathbf{Y}$ ). The cross-correlation is often used for causal inference by introducing a difference quantity [34]

$$\delta\chi_{xy}^l = \chi_{xy}^l - \chi_{yx}^l. \quad (18)$$

The sign of  $\delta\chi_{xy}^l$  is used, similar to the leaning approach, to determine the causal inference; i.e.,  $\delta\chi_{xy}^l > 0$  implies  $\mathbf{X}$  “causes”  $\mathbf{Y}$  and  $\delta\chi_{xy}^l < 0$  implies  $\mathbf{Y}$  “causes”  $\mathbf{X}$  [34].

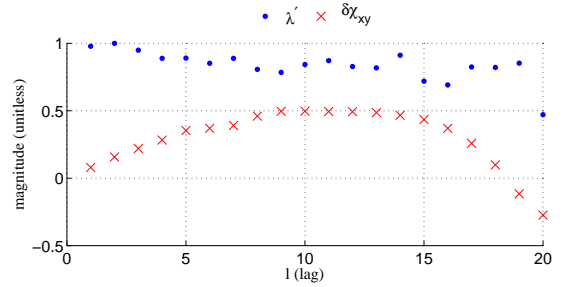


FIG. 3: (Color available online.) The unitless, normalized leaning,  $\lambda'$ , of Eqn. 16 can be plotted for different  $l$ -standard cause-effect assignments along with the cross correlation,  $\chi$  for the same lags,  $l$ , to show how the two values compare for this simple cyclic example.

Figure 3 shows how  $\delta\chi_{xy}^l$  compares to the leaning given  $l \in [1, 21]$  for an instance of Eqn. 16 with  $B = 0.5$ . In Figure 3, the leaning has been normalized for presentation clarity as

$$\lambda' = \frac{\lambda_l}{\max_{l \in [1, 21]} \lambda_l}, \quad (19)$$

where  $\lambda_l$  is the weighted mean observed leaning using the  $l$ -standard assignment ( $\lambda_1$  is plotted in Figure 2). The maximum leaning given  $l \in [1, 20]$  is approximately 0.625, so the normalized leanings shown in Figure 3 have a scaling factor of approximately 1.6.

Both  $\lambda'$  and  $\delta\chi_{xy}^l$  lead to the same causal inference, i.e.  $\mathbf{X}$  “drives”  $\mathbf{Y}$ , for  $l \in [1, 19]$ , although only the leaning agrees with intuition for  $l = 20$  and  $l = 21$  in this example. Thus, both tools agree with intuition for small lags in this simple cyclic example. The leaning, however, has its maximum values near the smallest lags, which is expected given Eqn. 16, while the cross-correlation difference has its maximum values at lags that do not explicitly appear in Eqn. 16.



### C. RL Circuit Example

The cross-correlation difference technique is also known to be unreliable given nonlinear dynamics [34]. Leanings of data sets generated from nonlinear dynamics will be discussed in Section V D. Neither of the previous examples has been physically motivated, so this section discusses exploratory causal inference of synthetic data sets generated from the well-known dynamics of a physical system.

Consider a series circuit containing a resistor, inductor, and time varying voltage source related by

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R}{L}I(t), \quad (20)$$

where  $I(t)$  is the current at time  $t$ ,  $V(t) = \sin(t)$  is the voltage at time  $t$ ,  $R$  is the resistance, and  $L$  is the inductance. The time series pair for this example is then

$$\{\mathbf{V}, \mathbf{I}\} = \{\{V_t\}, \{I_t\}\} \quad (21)$$

where  $\mathbf{V}$  is the set of discrete values of  $V(t)$  evaluated using  $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 8\pi\}$  with  $f = 1/10$  and  $\mathbf{I}$  is the set of discrete values found either by solving Eqn. 20 numerically or by evaluating the analytical solution

$$I(t) = \frac{L}{D}e^{-\frac{t}{\tau}} + \frac{R}{D}\sin(t) - \frac{L}{D}\cos(t) \quad (22)$$

with  $D = L^2 + R^2$  and  $\tau = L/R$ , for the same time set used for  $\mathbf{V}$ .

Physical intuition is that  $V$  drives  $I$ , and so we expect to find that  $\mathbf{V} \xrightarrow{\text{lean}} \mathbf{I}$ . The weighted mean observed leaning using the 1-standard assignment,  $\lambda_1$ , can be used to test this expectation. Unlike the previous examples, however, there is no noise term in the dynamics (such as  $B$  in Eqn. 15 and 16), so setting the tolerance domains, e.g.,  $\delta_I$ , will not be as straightforward.

Table II shows  $\lambda_1$  for both the analytical solution and a numerical solution to Eqn. 20 using the *ode45* integration function in MATLAB. The time series  $\mathbf{V}$  is created by defining values at fixed points and using linear interpolation to find the time steps required by the ODE solver. Two different physical scenarios are considered in which  $L$  and  $R$  are constant,  $L = 10$  H and  $R = 5$   $\Omega$  and  $L = 5$  H and  $R = 20$   $\Omega$ .

The previously discussed strategy of increasing  $\delta_I$  until the leaning becomes undefined and then reporting the leaning calculated using the largest  $\delta_I$  for which it is defined would lead to a causal inference that agrees with intuition for this example. Specifically, from Table II(a)  $\delta_I = 10^{-2} \Rightarrow \lambda_1 \approx 0.7 \Rightarrow \mathbf{V} \xrightarrow{\text{lean}} \mathbf{I}$ , as expected.

Discussion on setting the tolerance domains has centered on understanding the noise in the system. This example illustrates that the “noise” being considered does not need to be a physical noise source in the system (there are no explicit noise terms in Eqn. 21). For example, the numerical tolerance of the ODE solver was set to  $10^{-3}$

for the results shown in both Table II, and for both examples setting  $\delta_I = 10^{-3}$  would lead to causal inferences that agree with intuition.

Consider, for example, the peak values of  $\mathbf{V}$ . The time steps of these peaks are  $\mathbf{T}_{\text{peak}} = \{t|V_t = 1\} = \{6, 26, 46, 66\}$ . The values of  $\mathbf{I}$  given  $\tau = 0.25$  that immediately follow these peaks are  $\mathbf{I}_{0.25}^{\text{peak}} = \{I_t|t \in \{7, 27, 47, 67\}\}$ . The peak values associated to  $\tau = 2$  can also be found. The standard deviation of the first set is  $\sigma_{0.25}^{\text{peak}} \approx 10^{-6}$  and the standard deviation of the second set is  $\sigma_2^{\text{peak}} \approx 10^{-2}$ . Table II(a) (for  $\sigma_{0.25}^{\text{peak}}$ ) and (b) (for  $\sigma_2^{\text{peak}}$ ) shows setting  $\delta_I$  to the appropriate standard deviation of the peaks would lead to causal inferences that agree with intuition. Rather than physical noise levels, the noise levels used to set the tolerance domains for the leaning calculations is better thought of as the spread in the possible values of an assumed effect that may reasonably be considered due to the same assumed cause.

This example can also illustrate the importance of sample frequency and sample length. The leaning calculation requires an assumed cause and effect pair to appear in the data enough times to provide a reliable estimates of probabilities. Thus, data that is sampled for too few periods or too sparsely can lead to counter-intuitive leanings. For example, if there is only a single peak in the assumed driving time series because of poor sampling, then there can only be a single response value, which would be insufficient to reliably provide the conditional probabilities in the leaning calculation for that assumed cause-effect pair. For Eqn. 21 with the analytical solution for  $\mathbf{I}$ , if  $\delta_I = 10^{-3}$  and  $\tau = 0.25$ , then  $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 2\pi\}$  with  $f = 1/10$  leads to  $\lambda_1 = -0.045$  and  $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 3\pi\}$  with  $f = 2/3$  leads to  $\lambda_1 = -0.167$ , both of which disagree with intuition.

### D. Nonlinear Example

The examples so far have all had a linear relationship between the driving signal and the response signal. Of the four broad categories of time series causality tools, transfer entropy [18] and SSR methods [38] are the two categories that can be applied to nonlinear data sets without modification. The conceptual framework of Granger causality is not restricted by the linearity of the data set [14], but the original formulation by Granger must be modified to do so [40]. Lagged cross-correlation techniques are known to be unreliable if the data sets are generated by nonlinear dynamics [34]. The next examples are for nonlinear systems.

Consider the nonlinear dynamical system of

$$\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \quad (23)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \sin(t)$$

$\delta_I$	$\lambda_1$ (ode45)	$\lambda_1$ (analytical)	$\delta_I$	$\lambda_1$ (ode45)	$\lambda_1$ (analytical)
0	-0.132	-0.089	0	-0.132	-0.132
$10^{-6}$	-0.132	0.493	$10^{-6}$	-0.132	-0.132
$10^{-5}$	-0.108	0.548	$10^{-5}$	-0.120	-0.096
$10^{-4}$	0.188	0.564	$10^{-4}$	0.011	0.098
$10^{-3}$	0.582	0.581	$10^{-3}$	0.398	0.386
$10^{-2}$	0.730	0.727	$10^{-2}$	0.676	0.675
$10^{-1}$	undefined	undefined	$10^{-1}$	0.314	0.315

(a)  $R = 20 \Omega$ ,  $L = 5 \text{ H}$                       (b)  $R = 5 \Omega$ ,  $L = 10 \text{ H}$

TABLE II: The leaning  $\lambda_1$  depends on both  $\delta_I$  and the method for computing  $\mathbf{I}$  in this example. These two cases show that the values of  $\delta_I$  for which the leaning starts to agree with intuition can also depend on the physical system parameters (e.g.,  $\tau$ ).

and

$$y_t = Ax_{t-1}(1 - Bx_{t-1}) + C\eta_t,$$

with  $y_0 = 0$ , with  $A, B, C \in [0, 1]$  and  $\eta_t \sim \mathcal{N}(0, 1)$  with  $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 6\pi\}$  and  $f = 1/30$  so that  $L = 181$ .

Figure 4 shows the weighted mean observed leaning using the 1-standard assignment, i.e.,  $\lambda_1$ , agrees with intuition over the considered domains of  $A$ ,  $B$ , and  $C$  if the tolerance domain for  $\mathbf{Y}$  is set to the noise level, i.e.,  $\delta_y = C$ . The result of  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  shows that causal inference using leanings on data sets generated from nonlinear dynamics can be performed similarly, and can lead to similarly intuitive results, as the data sets generated from linear dynamics.

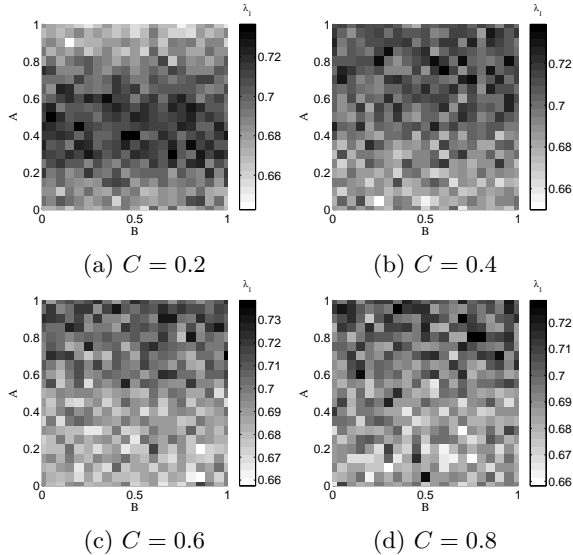


FIG. 4: Leaning,  $\lambda_1$ , of Eqn. 23 computed  $\delta_y = C$  as function of all three unitless parameters in Eqn. 23,  $A$ ,  $B$ , and  $C$ . The leaning agrees with intuition in this example for all the tested parameter values.

### E. Coupled Logistic Map Example

Proponents of SSR time series causality tools have pointed out the limitations of tools like lagged cross-correlation and Granger causality when the dynamics exhibit chaotic behavior [38]. A chaotic system is considered in this section.

Consider the nonlinear dynamical system of

$$\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \quad (24)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = x_{t-1}(r_x - r_x x_{t-1} - \beta_{xy} y_{t-1})$$

and

$$y_t = y_{t-1}(r_y - r_y y_{t-1} - \beta_{yx} x_{t-1})$$

where the parameters  $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbb{R} \geq 0$ . This pair of equations is a specific form of the two-dimensional coupled logistic map system often used to model population dynamics [23] and it was a system used in the introduction of cross convergent mapping (CCM) which is a SSR time series causality tool [38].

Sugihara et al. [38] note that  $\beta_{xy} > \beta_{yx}$  intuitively implies  $\mathbf{Y}$  “drives”  $\mathbf{X}$  more than  $\mathbf{X}$  “drives”  $\mathbf{Y}$ , and vice versa. Such intuition, however, can be difficult to justify for all instances of Eqn. 24. The  $x_{t-1}$  term that appears in  $y_t$  can be seen as a function of  $x_{t-2}$  with coefficients of  $\beta_{yx} r_x$ . These product coefficients suggest that if  $r_x > r_y$ , then  $\mathbf{X}$  may be seen as the stronger driver in the system even if  $\beta_{yx} < \beta_{xy}$ . The same argument can be made, with the appropriate substitutions, to show that  $\mathbf{Y}$  may be seen as the stronger driver in the system even if  $\beta_{xy} < \beta_{yx}$ . As such, there is no clear intuitive causal inference for this system. The conjectures presented in this paragraph, however, are supported by the leaning calculations (using the 1-standard assignment).

Figure 5 shows four instances of Eqn. 24 with different values for  $r_x$  and  $r_y$ . Each instance has a library length of  $L = 500$  and initial conditions of  $x_0 = 0.4$  and  $y_0 = 0.4$ . There is no clear, intuitive driver in this example, so

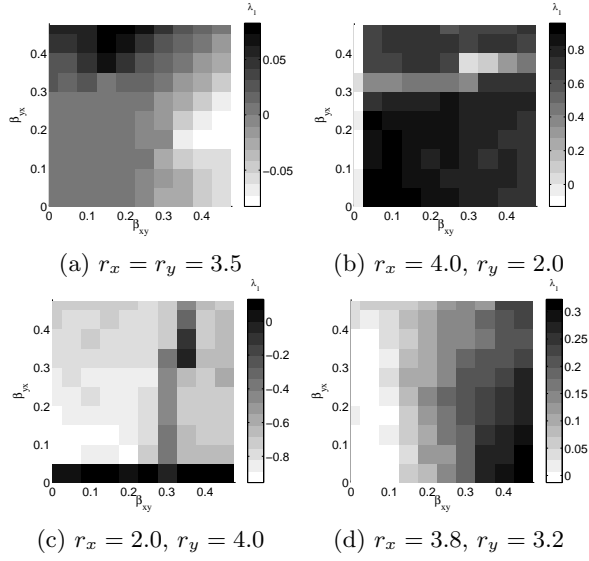


FIG. 5: Leaning,  $\lambda_1$ , of Eqn. 24 is a function of four unitless parameters in Eqn. 24,  $r_x$ ,  $r_y$ ,  $\beta_{xy}$ , and  $\beta_{yx}$  (along with the initial conditions  $x_0$  and  $y_0$ , which are fixed in this example). The tolerance domains are set as  $\delta_x = \sigma_{x_t - \langle x_t \rangle}$  and  $\delta_y = \sigma_{y_t - \langle y_t \rangle}$ . The leaning is defined using the 1-standard assignment, so

$$\lambda_1 > 0 \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}.$$

both tolerance domains must be set in the leaning calculation. The leaning is calculated using the 1-standard cause-effect assignment and estimated tolerance domains of  $\delta_x = \sigma_{x_t - \langle x_t \rangle}$  and  $\delta_y = \sigma_{y_t - \langle y_t \rangle}$ .

Figure 5(a) shows the intuition of  $\beta_{xy} < \beta_{yx} \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  can be true if  $r_x = r_y$ . However, Figure 5(b) and (c) shows  $r_x > r_y \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  and  $r_x < r_y \Rightarrow \mathbf{Y} \xrightarrow{\text{lean}} \mathbf{X}$  can be strong enough implications to make the values of  $\beta_{xy}$  and  $\beta_{yx}$  irrelevant over the considered domains. Figure 5(d) shows this effect can be pronounced even in instances of Eqn. 24 where  $r_x$  and  $r_y$  are close.

Sugihara et al. [38] also discuss how a naive [27] application of Granger causality to the system described in Eqn. 24 may lead to conclusions that do not agree with intuition, while CCM does. The causal inference suggested by the leaning calculations of this subsection implies both CCM and leanings may be useful time series causality tools in situations where Granger causality is not. It has also been shown that CCM may fail to agree with intuition in example systems for which it has already been shown that leaning calculation do, e.g., Section VC [24].

The complexity of determining causal relationships in this system may make the system less of a convincing example of the leaning calculation than the previous examples. However, Figure 5 shows the weighted mean observed leaning using the 1-standard cause-effect assignment can provide causal inferences that may be consid-

ered intuitively justifiable, even if the system does not have an unequivocal driver.

## F. Impulse with Multiple Noisy Responses Example

All of the previous examples have ignored possible causal confounders. The presence of confounders in the system is a serious problem for causal inference in general [17, 30]. Time series causality usually seeks to answer the less general causal inference question of “Given two time series, which may be considered the stronger driver?” Nevertheless, some bivariate time series causality tools consider causal inference in systems with potential confounders by trying to relate the estimated bivariate driving relationships within a collection of more than two time series data sets (e.g., see CCM [38]). The next example explores the use of leaning calculations in such a scenario.

Consider the multivariate system of

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} = \{\{x_t\}, \{y_t\}, \{z_t\}\} \quad (25)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \begin{cases} 2 & t = 1 \\ 0 & \forall t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

and

$$y_t = x_{t-1} + B\eta_t,$$

and either (case 1)

$$z_t = y_{t-1} \quad (26)$$

or (case 2)

$$z'_t = y_{t-1} + y_t = y_{t-1} + x_{t-1} + B\eta_t \quad (27)$$

or (case 3)

$$z''_t = y_{t-1} + x_{t-1} + z_{t-1} \quad (28)$$

with  $y_0 = 0$ ,  $B \in \mathbb{R} \geq 0$ ,  $\eta_t \sim \mathcal{N}(0, 1)$ , and  $L = 500$ .

In case 1,  $\mathbf{Z}$  depends directly on  $\mathbf{Y}$  and indirectly on  $\mathbf{X}$  (through  $\mathbf{Y}$ , which depends directly on  $\mathbf{X}$ ). The intuitive causal inference is then  $\mathbf{Y} \xrightarrow{\text{lean}} \mathbf{Z}$  and  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Z}$ . Case 2, despite the additional  $\mathbf{Y}$  dependence in  $\mathbf{Z}$ , has the same intuitive causal inference as case 0. In case 2,  $\mathbf{Z}$  depends directly on itself and both  $\mathbf{Y}$  and  $\mathbf{X}$ . Case 3 also has the same intuitive causal inference.

Figure 6 shows the weighted mean observed leaning using the 1-standard cause-effect assignment (with  $\delta_x = 0$  and  $\delta_y = \delta_z = B$ ),  $\lambda_1$ , may lead to causal inferences that do not agree with intuition for case 1 and case 3, even though case 2 agrees with intuition for all points within

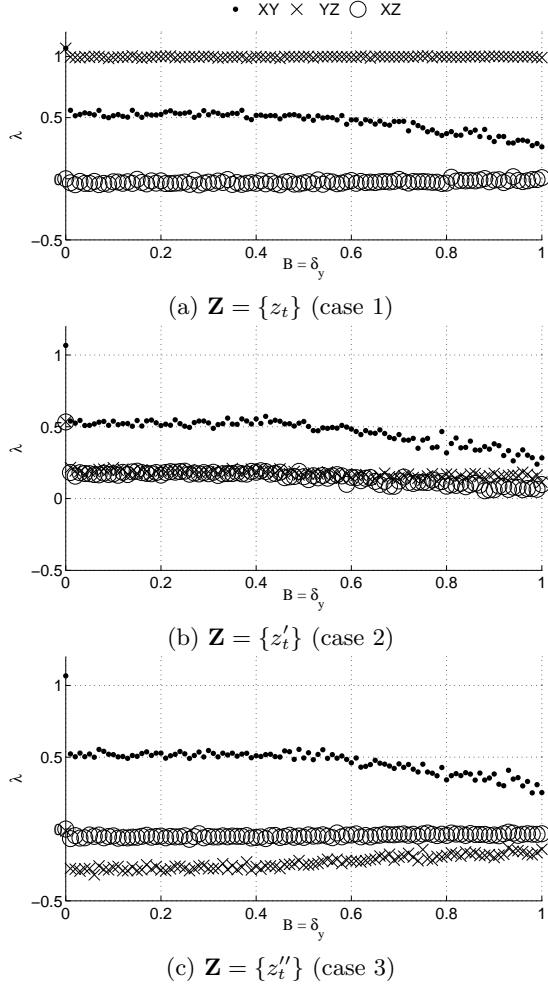


FIG. 6: The (unitless) weighted mean observed leaning using the 1-standard cause-effect assignment,  $\lambda_1$  only leads to causal inferences that agree with intuition for all points within the considered noise level domain,  $B$ , in this example for case 2.  $B$  is the unitless noise parameter found in Eqn. 25, and the symmetric tolerance domains for  $\mathbf{Y}$  and  $\mathbf{Z}$  are set to this value. All of the leanings are expected to be positive in each case.

the considered noise levels domain. For case 2, the leaning calculation implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  and  $\mathbf{Y} \xrightarrow{\text{lean}} \mathbf{Z}$  as expected, but it also seem to imply that no causal inference can be made about the relationship between  $\mathbf{X}$  and  $\mathbf{Z}$ . For case 3, the leaning calculation also implies no causal inference can be made about the relationship between  $\mathbf{X}$  and  $\mathbf{Z}$ , but, unlike case 1, it also implies  $\mathbf{Z} \xrightarrow{\text{lean}} \mathbf{Y}$ , which is counter-intuitive.

These results may imply that  $\lambda_1$  is unable to identify confounded driving (i.e., situations in which the effect of the driving variable is mediated by another variable). For example, in case 1, the driving of  $\mathbf{Z}$  by  $\mathbf{X}$  occurs through the interaction of  $\mathbf{Y}$  and  $\mathbf{Z}$ . For case 1,  $\lambda_1$  implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$  but does not imply  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Z}$ . For case 3,  $\lambda_1$

	case 1	case 2	case 3
$\lambda_{AR}^{xy}$	0.150	0.159	0.169
$\lambda_{AR}^{yz}$	-0.002	0.133	0.447
$\lambda_{AR}^{xz}$	0.691	0.030	0.735

TABLE III: The leaning calculation depends strongly on the cause-effect assignment. The table shows the weighted mean observed leaning using the 1-AR assignment and may be compared with Figure 6, which showed this leaning calculation using the 1-standard assignment.

implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y} \xleftarrow{\text{lean}} \mathbf{Z}$ , which may imply that  $\lambda_1$  is not a reliable causal inference tool in autoregressive systems.

The results of Figure 6 may also be considered an indication that the cause-effect assignment is insufficient. It was previously mentioned that exploratory causal analysis using the leaning would involve comparing several different cause-effect assignments. The set of tested cause-effect assignments need not only include  $l$ -standard assignments. Consider the weighted mean observed leaning,  $\lambda_{AR}^{xy}$  using the 1-AR cause-effect assignment, i.e.,  $\{C, E\} = \{x_{t-1} \text{ and } y_{t-1}, y_t\}$ . Table III shows this leaning calculation, using  $\delta_y = \delta_z = B = 0.6$ , for the same bivariate relationships shown in Figure 6.

Table III implies  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y} \xrightarrow{\text{lean}} \mathbf{Z}$  for case 2 and 3, as expected, but not for case 1 (which  $\lambda_1$  did imply). The leaning calculations are part of an exploratory causal analysis and must be considered using several different cause-effect assignments when trying to understand the potential causal structure of a set of times series data. The cause-effect assignments can also be expanded beyond the bivariate and autoregressive definitions, e.g.,  $\{C, E\} = \{x_{t-1} \text{ and } y_{t-1} \text{ and } z_{t-1}, y_t\}$ , but such extensions will not be considered in this article.

## VI. EMPIRICAL DATA

Empirical data sets with known (or assumed) causal relationships may be used to understand how exploratory causal inference using leanings might be done if the system dynamics are unknown (or sufficiently complicated to make first principle numerical comparisons cumbersome). The examples shown in this section are intended to demonstrate that causal inference using leaning calculations can agree with the causal “truth” in empirical data sets. The analysis shown here is not expected to illustrate how the leaning may be used for exploratory causal analysis of empirical data for which there is no causal “truth”. Such analysis is expected to be more complicated than that which is shown here (e.g., involving multiple tolerance domain calculations and the comparison of different cause-effect assignments).

Figure 7 shows a time series pair with causal “truth”

from the UCI Machine Learning Repository (MLR) [1]. This data repository is a collection of data sets (some of which are time series) with known, intuitive, or assumed causal relationships meant for use in the testing of causal discovery algorithms in machine learning [1].

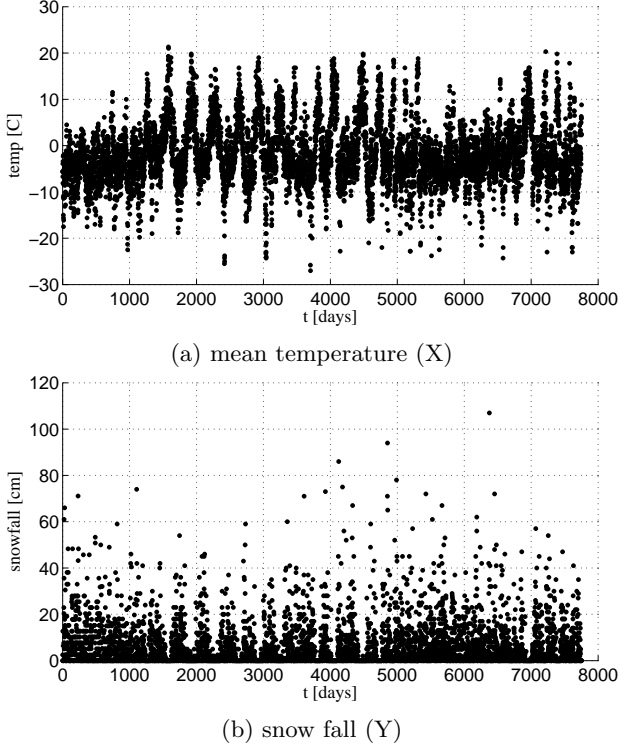


FIG. 7: This times series data pair is expected to have the causal relationship of  $X \rightarrow Y$ , where  $X$  or  $Y$  is marked in parenthesis for each time series.

Figure 7(a) and (b) are times series of the daily snowfall (the expected response) and mean temperature (the expected driver) from July 1 1972 to December 31 2009 at Whistler, BC, Canada (Latitude:  $50^{\circ}04'04.000''$  N, Longitude:  $122^{\circ}56'50.000''$  W, Elevation: 1835.00 meters). From [1], “Common sense tells us that  $X$  [mean temperature] causes  $Y$  [snow fall] (with maybe very small feedback of  $Y$  on  $X$ ). Confounders are present (e.g., day of the year).” These time series correspond to data set 87 of the MLR [1].

As noted previously, the primary difficulty in using the leaning for exploratory causal analysis is the determination of the cause-effect assignment and tolerance domains. The above data is meant only to illustrate the use of leanings, so while a thorough analysis of the noise in the system should precede the leaning calculations, such a step is avoided here for brevity.

The symmetric tolerance domains are estimated using the maximum standard deviations of the  $n$  sets of binned points of an  $n$ -bin histogram of the normalized time series data  $\mathbf{X}'$  and  $\mathbf{Y}'$ , where  $\mathbf{X}' = \frac{\mathbf{X} - \langle \mathbf{X} \rangle}{\sigma_{\mathbf{X}}}$ ,  $\mathbf{Y}' = \frac{\mathbf{Y} - \langle \mathbf{Y} \rangle}{\sigma_{\mathbf{Y}}}$ , and  $n = \lfloor 0.1L \rfloor$  (i.e.,  $n$  is the closest integer that is not larger

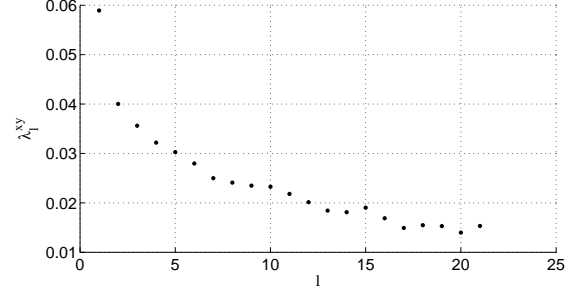


FIG. 8: The weighted mean observed leaning,  $\lambda_l^{xy}$ , using the 1-standard cause-effect assignment for  $l \in [0, 21]$ . The time series pair  $\tau_1 = (\mathbf{X}, \mathbf{Y})$  with  $\mathbf{X}$  and  $\mathbf{Y}$  shown in Figure 7(a) and (b), respectively. The expected causal inference is  $\mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ ; i.e., the expectation is  $\lambda_l^{xy} > 0$  for every point in the plotted domain.

than 10% of the library length  $L$ ). This estimation is similar to the  $n$ -bin mean standard deviation technique discussed in Sec. V A.

The cause-effect assignment will be set naively because, again, the purpose of this article is not to study these particular time series in detail. To reiterate the previous comment regarding tolerance domains, detailed study would be required to have confidence in using leanings for exploratory causal analysis. However, the convenience of having causal “truths” is that we can take the naive approach of simply testing many different cause-effect assignments and compare the results to the expected causal inference.

Figure 8 shows the weighted mean observed leaning,  $\lambda_l^{xy}$ , using the 1-standard cause-effect assignment with  $l \in [0, 21]$  and using  $\sigma_x$  and  $\sigma_y$  estimated in the manner described above.

Figure 8 shows the leanings, using the given tolerance domains, imply causal inferences that agree with the causal truths for the tested pair with 1-standard assignments.

This example also highlights the problem of determining which  $l$ -standard assignment to use for the causal inference. If it is decided that the causal inference depends on

$$\lambda_{max}^{xy} = \lambda_{l'}^{xy} \quad (29)$$

where

$$|\lambda_{l'}^{xy}| = \max_l |\lambda_l^{xy}|, \quad (30)$$

then  $\lambda_{max}^{xy} = 0.040 \Rightarrow \mathbf{X} \xrightarrow{\text{lean}} \mathbf{Y}$ , which agrees with the causal truth.

The NASA OMNI data set consists of hourly-averaged time series measurements of several different space weather parameters from 1963 to present, collected from more than twenty different satellites, along with sunspot number and several different geomagnetic indices, including  $D_{st}$ , collected from the NOAA National Geophysical

Data Center [20]. The disturbance storm time,  $D_{st}$ , is a measure of geomagnetic activity [39]. The magnetic field measurements in the OMNI data sets, specifically  $B_z$  in GSE coordinates [15], is believed to be a driver of  $D_{st}$  [11].

Let  $\mathbf{P}_z^L = \{\{B_z(t'), \{D_{st}(t')\} \mid t' \in [t'_0, L]\}$  be an ordered subset of the available time series data  $\{\{B_z(t), \{D_{st}(t)\} \mid t = 0, 1, 2, \dots, N\}$  where  $N$  is the number of hourly data points in the OMNI data set. If  $\lambda_1^z$  is the weighted mean observed leaning for  $\mathbf{P}_z^L$  using the 1-standard cause-effect assignment, then  $n$  samples of  $\mathbf{P}_z^L$ , each with a different  $t'_0$ , would produce a set of  $n$  leanings,  $\{\lambda_1^z\}$ , from which the causal inference could be drawn.

Let  $L = 500$  and  $n = 10^4$ . The symmetric tolerance domains are naively set with  $f\sigma_{|B'_z - \langle B'_z \rangle|}$  and  $f\sigma_{|D_{st} - \langle D_{st} \rangle|}$  for each sampled times series of length  $L$  with  $f=0.05$ . The starting points for each time series are sampled from a uniform distribution over  $[0, N - L]$ . Figure 9(a) shows the causal inference drawn from each set of leanings agrees with intuition, i.e.,  $B_z \xrightarrow{\text{lean}} D_{st}$ , if the causal inference is based on, e.g., the mean value from the set of leanings  $\langle \lambda_l^z \rangle$  with  $l = 1$ . The algebraic means  $\langle \lambda_l^z \rangle$  found using different  $l$ -standard assignments with  $l \in [1, 20]$  are shown in Figure 9(b). The causal inference is  $B_z \xrightarrow{\text{lean}} D_{st}$  for every  $l$  in the plotted domain.

This example is the first for which a set of leanings has been used for causal inference, which may imply statistical testing should be used. The sample mean was used for causal inference and happened to agree with intuition for this example, but would the same conclusion be drawn using a formal hypothesis test? How should the null hypothesis and test statistic be constructed? Such questions can be subtle (see, e.g., [17]). The sampling procedure used to produce Figure 9(a) produces 2,796 defined leanings, 95% of which are below  $4.2 \times 10^{-3}$  and 5% of which are below  $-4.0 \times 10^{-3}$ . A 90% confidence that the leaning falls in the interval  $[-4.0 \times 10^{-3}, 4.2 \times 10^{-3}]$ , however, is not a strong indication that the data supports the intuitive causal structure. The mean of the set is  $\mu = 3.5 \times 10^{-3}$ , and the variance is  $\sigma^2 = 7.0 \times 10^{-5}$ . If it is assumed that the leaning in this example is distributed as  $\mathcal{N}(\mu, \sigma^2)$ , then a 95% confidence interval may be  $[\mu - 2\sigma, \mu + 2\sigma] = [-4.9 \times 10^{-3}, 5.6 \times 10^{-3}]$ , which, again, does not strongly support the intuitive causal inference for this example. Approximately 40% of the leanings in this example are negative, which may imply that there is only a 60% confidence that this data supports the intuitive causal inference, given the tolerance domains and cause-effect assignments.

Suppose a null hypothesis is defined as  $\langle \lambda_1^z \rangle = 0$ . The standard error is  $\text{SE} = \sigma/\sqrt{n} = 5.0 \times 10^{-4}$ , from which the t-test statistic follows [5] as  $t = \mu/\text{SE} = 7.08$ . A two-tailed t-test (i.e., calculating the  $p$ -value with an alternative hypothesis of  $\langle \lambda_1^z \rangle \neq 0$ ) returns a  $p$ -value of approximately zero [28], which implies the null hypothesis should be rejected in favor of the alternative at any significance

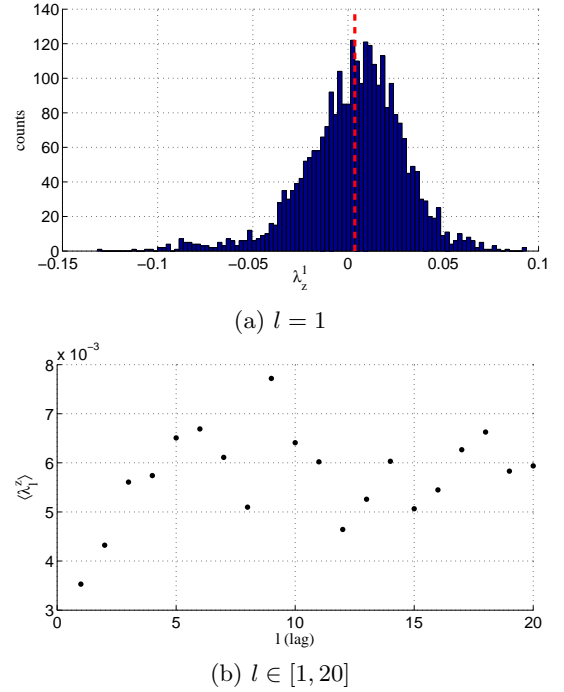


FIG. 9: (Color available online) (a) A histogram of the  $(B'_z, D_{st})$  set of  $n = 10^4$  (unitless) weighted mean observed leanings, using the 1-standard cause-effect assignment, i.e.,  $\lambda_z^1$ , show causal inferences that agree with intuition for the OMNI data set. The red dashed lines show the algebraic mean of the sets. The geomagnetic field component  $B'_z$  is calculated by taking the  $B_z$  times series in the OMNI data set and then setting  $B'_z = 0$  if  $B_z > 0$  [4, 8]. (b) The algebraic means of the aforementioned sets of weighted mean observed leanings, i.e.,  $\langle \lambda_l^z \rangle$ , are positive for all leanings calculated using  $l$ -standard assignment given  $l \in [1, 20]$ .

level. A right-tailed t-test (i.e., the alternative hypothesis is  $\langle \lambda_1^z \rangle > 0$ ) also returns a  $p$ -value of approximately zero. A left-tailed t-test (i.e., the alternative hypothesis is  $\langle \lambda_1^z \rangle < 0$ ) returns a  $p$ -value of approximately one, which implies the null hypothesis cannot be rejected in favor of the alternative at any significance level. These hypothesis tests seem to imply the population mean of the sampled leanings calculated in this example is likely not zero (which implies the time series pair has some causal structure) and is likely greater than zero (which implies the causal inference made with the leaning agrees with intuition). These conclusions, however, depend on whether or not the t-test is applicable to this example. For example, the assumption that the sample mean of the leanings can be assumed to follow a normal distribution based on the central limit theorem [5] may rely on the sampled time series from which the leaning were calculated being independent and identically distributed, which may not be true. The assumptions used in these statistical tests are intended to be illustrative. Such assumptions should be explored in depth to formally develop a statistical test

for causal inference using leanings. The sampling procedure used in this example may not be applicable to other data sets for which the leaning may still be a useful causal inference tool. Thus, it may not be possible that a single statistical test will be appropriate for all sets of leaning calculations.

A bootstrapping [9] procedure can be set up with the sample of leaning calculations, whereby  $10^6$  means are calculated from new sets (of the same size as the original set) of leanings that have been sampled (with replacement) from the original set. This procedure yields no negative means; the null hypothesis that the mean leaning value is actually negative (i.e.,  $\langle \lambda_1^z \rangle < 0$ ) can be rejected with a  $p$ -value less than  $10^{-6}$ . The 90% confidence interval for the mean of the  $10^6$  bootstrapped means is  $[3.48 \times 10^{-3}, 3.57 \times 10^{-3}]$ , which, again, implies the mean leaning for this example is positive.

## VII. SPURIOUS LEANINGS

Consider the linear system of

$$\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \quad (31)$$

where  $t = 0, 1, 2, \dots, L$ ,

$$x_t = \begin{cases} 2 & t = 1 \\ 0 & \forall t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

and

$$y_t = \eta_t$$

with  $\eta_t \sim \mathcal{N}(0, 1)$ . The first time series,  $\mathbf{X}$ , is the periodic impulse that drove the example system in Eqn. 15. The second time series,  $\mathbf{Y}$ , is standard Gaussian noise applied at each time step.

There is no intuitive causal relationship in Eqn. 31. However, Figure 10 shows the weighted mean observed leaning using the 1-standard assignment may lead to spurious causal inferences for different symmetric tolerance domains  $\delta_y$ , given  $\delta_x = 0$ . The causal inference becomes inconclusive as the library length  $L$  is increased; i.e., the leaning moves towards zero for the tested tolerance domains as the library length of Eqn. 31 is increased. However, the use of leanings for causal inference with Eqn. 31 at smaller library length, e.g.,  $L = 10$ , may imply a spurious relationship.

The spurious leanings shown in this example may imply causal relationships that do not exist in the system. Leaning calculations may be part of an exploratory causal inference analysis, but care must be taken to ensure a causal relationship is actually present in the data, even if the directionality (or other features) of that relationship are unknown. The relationship between leanings and causality as it is typically understood in physics (i.e., involving interventions into the systems under investigation, e.g., through experiments [30]) is not currently

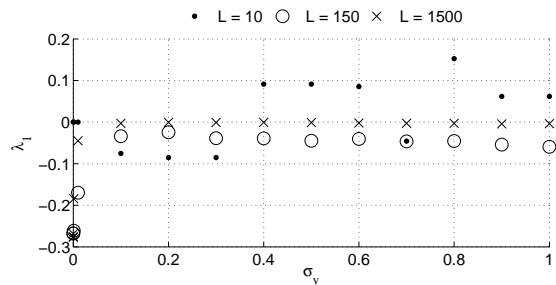


FIG. 10: Eqn. 31 leads to spurious leanings, i.e. weighted mean observed leanings using the 1-standard assignment,  $\lambda_1$ , that depend on both the tolerance domain  $\sigma_y$  and the library length  $L$ .

known. This article is exploring the use of leanings as part of an exploratory causal analysis in times series data, not as a definition or proof of causality in a dynamical system.

## VIII. CONCLUSION

Causal inference using observational data alone is a difficult task [21]. This problem is important in many fields, but in physics in particular, there are often subfields for which direct experimentation is not technologically feasible.

Exploratory causal analysis, as it has been described here, involves many different techniques, including Granger causality (GC), transfer entropy (TE), cross correlation (CC), and state space reconstruction (SSR). Each of these techniques has well-known shortcomings. For example, GC is parametric, TE can be computationally difficult [18], CC can be unreliable [34], and SSR relies on correctly setting lag times and embedding dimensions [37]. Causal leaning has been introduced to overcome many of these shortcomings: it is non-parametric, may be based on counting, and the only adjustable parameters are the tolerance domains and cause-effect assignments.

No attempt has been made to interpret causal leanings in terms of current philosophical causality studies. For example, there is no exploration of how causal leanings are associated with token or *prima facie* causality [21]. We have grouped the leaning method under the broad term of time series causality inference, which implies the technique is distinct from other data causality methods, including direct acyclic graph (DAG) [30] and temporal logic [21] techniques. Causal leanings have been introduced here as a practical tool and connections with the broader fields of data causality and causality foundations are left for future work. For example, leanings may be a subset of the more general temporal logic presented by Kleinberg [21] and may have interpretations within Good's probabilistic causal framework of propensities and weights of evidence [12].

There are many open questions regarding the use of leanings for causal inference that have not been explored in this article. For example, how should the magnitude of the leaning calculations be interpreted?; if there are two weighted mean observed leanings  $\lambda_1$  and  $\lambda_2$  with different cause-effect assignments such that  $\lambda_1 > \lambda_2$  – does the first cause-effect assignment represent a “stronger” driver than the second?; how should leanings of 0, 2, or  $-2$  be interpreted with respect to the cause-effect assignments?; and how should leanings calculated using different cause-effect assignments be compared?

There has also been no formal exploration of using leanings as part of statistical tests, as is often done with GC [32]. The use of histograms in Figure 9 may be con-

sidered a first step toward statistical interpretations of leanings.

Finally, this article has discussed the use of leanings as part of an exploratory causal analysis of time series data. Exactly how such an analysis should be conducted is, however, still an open question. For example, given a report of GC, TE, SSR, and leanings (all calculated in various ways), how should the results be interpreted holistically? There are many potentially confusing scenarios in which, e.g., two techniques lead to opposite causal inferences. The most reasonable time series causality techniques to use for a given exploratory causal analysis may depend strongly on the data itself, but general guidelines for such analysis is, as far as we know, unknown.

- 
- [1] K. Bache and M. Lichman. UCI machine learning repository, 2013.
  - [2] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2013.
  - [3] Mario Bunge. *Causality and modern science*. Courier Corporation, 1979.
  - [4] RK Burton, RL McPherron, and CT Russell. The terrestrial magnetosphere- a half-wave rectifier of the interplanetary electric field. *Science*, 189(4204):717–718, 1975.
  - [5] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.
  - [6] Jean-Marie Dufour, Denis Pelletier, and Éric Renault. Short run and long run causality in time series: inference. *Journal of Econometrics*, 132(2):337–362, 2006.
  - [7] Jean-Marie Dufour and Eric Renault. Short run and long run causality in time series: theory. *Econometrica*, pages 1099–1125, 1998.
  - [8] James W Dungey. Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, 6(2):47, 1961.
  - [9] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.
  - [10] Mahmoud El-Gohary and James McNames. Establishing causality with whitened cross-correlation analysis. *Biomedical Engineering, IEEE Transactions on*, 54(12):2214–2222, 2007.
  - [11] W. D. Gonzalez, J. A. Joselyn, Y. Kamide, H. W. Kroehl, G. Rostoker, B. T. Tsurutani, and V. M. Vasyliunas. What is a geomagnetic storm? *Journal of Geophysical Research: Space Physics*, 99(A4):5771–5792, 1994.
  - [12] I. J. Good. Causal propensity: A review. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1984:pp. 829–850, 1984.
  - [13] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
  - [14] C.W.J. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2(0):329 – 352, 1980.
  - [15] MA Hapgood. Space physics coordinate transformations: A user guide. *Planetary and Space Science*, 40(5):711–717, 1992.
  - [16] P. Illari and F. Russo. *Causality: Philosophical Theory Meets Scientific Practice*. Oxford University Press, 2014.
  - [17] G.W. Imbens and D.B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.
  - [18] A. Kaiser and T. Schreiber. Information transfer in continuous processes. *Physica D: Nonlinear Phenomena*, 166(12):43 – 62, 2002.
  - [19] Maciej Kaminski, Mingzhou Ding, Wilson A. Truccolo, and Steven L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 85(2):145–157, 2001.
  - [20] JH King and NE Papitashvili. Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *Journal of Geophysical Research: Space Physics (1978–2012)*, 110(A2), 2005.
  - [21] S. Kleinberg. *Causality, Probability, and Time*. Causality, Probability, and Time. Cambridge University Press, 2012.
  - [22] Yan Liu and TM Bahadori. A survey on granger causality: A computational view. *University of Southern California*, pages 1–13, 2012.
  - [23] Alun L. Lloyd. The coupled logistic map: a simple model for the effects of spatial heterogeneity on population dynamics. *Journal of Theoretical Biology*, 173(3):217 – 230, 1995.
  - [24] James M. McCracken and Robert S. Weigel. Convergent cross-mapping and pairwise asymmetric inference. *Phys. Rev. E*, 90:062903, Dec 2014.
  - [25] Timothy J Mosedale, David B Stephenson, Matthew Collins, and Terence C Mills. Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation. *Journal of climate*, 19(7), 2006.
  - [26] Eqn. 1 may appear related to the *weight of evidence* [12]. The two quantities, however, are fundamentally different, both mathematically and conceptually. A requirement of weight of evidence is that the posterior probability should be recoverable provided probability of the evidence alone; i.e.,  $P(C|E)$  should be derivable from  $\rho_{EC}$  and  $P(E)$  if  $\rho_{EC}$  can be considered weight of evidence as Good describes (see Eqn. (2) of [12]). This requirement is not met



- by the penchant, as the probability of the cause,  $P(C)$ , is also required to derive the posterior. The intended scope of the penchant is also much more restricted than that of the weight of evidence. Most of the examples presented by Good in [12] cannot be addressed directly with any times series causality tools. Eqn. 1 is, however, identical in form to the *Eells measure of causal strength* (also called the *probability contrast*) [? ]. Eqn. 1 will not be used as a measure of causal strength in this work. Rather, it will be used as a measure indicative of driving without any concern for notions of the “strength” of such driving. As such, we will use the term *penchant* to avoid confusion. Also see [21] for a discussion of similar expressions.
- [27] Sugihara et al. [38] do not explore any of the numerous non-linear extensions of Granger causality. The theoretical foundations of Granger causality are independent of its practical implementations, and failures of Granger causality may be failures of a specific implementation, e.g., using linear forecast models with non-linear data [14].
  - [28] This calculation, and all t-test calculations discussed in the section were performed with the MATLAB function *ttest*.
  - [29] Roberto D Pascual-Marqui, Rolando J Biscay, Jorge Bosch-Bayard, Dietrich Lehmann, Kieko Kochi, Toshihiko Kinoshita, Naoto Yamada, and Norihiro Sadato. Assessing direct paths of intracortical causal information flow of oscillatory activity with the isolated effective coherence (icoh). *Name: Frontiers in Human Neuroscience*, 8:448, 2014.
  - [30] Judea Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.
  - [31] A. Peres. *Quantum Theory: Concepts and Methods*. Fundamental Theories of Physics. Springer, 1995.
  - [32] David A. Pierce and Larry D. Haugh. Causality in temporal systems: Characterization and a survey. *Journal of Econometrics*, 5(3):265 – 293, 1977.
  - [33] David L. Roberts and Stephen Nord. Causality tests and functional form sensitivity. *Applied Economics*, 17(1):135–141, 1985.
  - [34] David Rogosa. A critique of cross-lagged correlation. *Psychological Bulletin*, 88(2):245, 1980.
  - [35] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.
  - [36] David W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.
  - [37] Michael Small and C.K. Tse. Optimal embedding parameters: a modelling paradigm. *Physica D: Nonlinear Phenomena*, 194(34):283 – 296, 2004.
  - [38] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.
  - [39] M Sugiura and T Kamei. *IAGA Bulletin N 40*. International Association of Geomagnetism and Aeronomy, 1991.
  - [40] Xiaohai Sun. Assessing nonlinear granger causality from multivariate time series. In Walter Daelemans, Bart Goethals, and Katharina Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 5212 of *Lecture Notes in Computer Science*, pages 440–455. Springer Berlin Heidelberg, 2008.
  - [41] Patrick Suppes. *A Probabilistic Theory of Causality*. North Holland Publishing Company, 1970.
  - [42] M. P. Wand. Data-based choice of histogram bin width. *The American Statistician*, 51(1):59–64, 1997.