

Inferring the time-varying coupling of dynamical systems with temporal convolutional autoencoders

Reviewed Preprint

v1 • November 12, 2024

Not revised

Josuan Calderon, Gordon J Berman 

Department of Physics, Emory University, Atlanta, GA, 30322, United States • Department of Biology, Emory University, Atlanta, GA, 30322, United States

 https://en.wikipedia.org/wiki/Open_access Copyright information

eLife Assessment

This **important** work presents a novel approach to infer causal relations in non-stationary time series data. To do so, the authors introduce a novel machine-learning model of Temporal Autoencoders for Causal Inference to identify and measure the direction and strength of time-varying causal interactions. The authors provide **solid** evidence for their claims through thorough numerical validation and comprehensive exploration of the method on both synthetic and real-world datasets. This is a timely contribution that may have theoretical and practical implications for diverse real-life applications.

<https://doi.org/10.7554/eLife.100692.1.sa2>

Abstract

Most approaches for assessing causality in complex dynamical systems fail when the interactions between variables are inherently non-linear and non-stationary. Here we introduce Temporal Autoencoders for Causal Inference (TACI), a methodology that combines a new surrogate data metric for assessing causal interactions with a novel two-headed machine learning architecture to identify and measure the direction and strength of time-varying causal interactions. Through tests on both synthetic and real-world datasets, we demonstrate TACI's ability to accurately quantify dynamic causal interactions across a variety of systems. Our findings display the method's effectiveness compared to existing approaches and also highlight our approach's potential to build a deeper understanding of the mechanisms that underlie time-varying interactions in physical and biological systems.

Introduction

Rather than being static in time, interactions between parts of a complex system continuously ebb and flow, with one variable driving another at one point, just to see the relationship reverse or lessen or disappear at a later point in time. Real-world signals are seldom stationary and well-

behaved, and their causal linkages and interactions frequently appear, disappear, and reappear, possibly changing in strength over time. Examples of such systems abound in neuroscience [1], ecology [3], finance, [4–6], and climate [7–9].

Despite the ubiquity of these dynamically altering interactions, however, most methods for assessing causality in complex dynamical systems have difficulty measuring how the direction and extent of interactions between variables in a system alter in time. This difficulty arises from several inherent characteristics of complex systems and the limitations of existing causal assessment methodologies. Most of these methods, including Granger Causality (GC) [10], often assume that the dynamical system should be approximately stationary, meaning their statistical properties do not change over time. Other common assumptions, such as linearity and time invariance are also often violated in real-world complex systems. These constraints significantly limit the applicability and accuracy of these approaches in many scenarios [11–16].

In addition, in systems where variables are strongly coupled and synchronized, some of these causality inference methods struggle to accurately infer the coupling strength and direction of causality [16]. This issue extends to scenarios of intermediate coupling, where the variables are neither weakly nor strongly linked. Additionally, the presence of noise in the system leads to a decrease in cross-mapping fidelity, revealing further limitations [17, 18]. Although the lack of correlation is neither necessary nor sufficient to demonstrate causation [19, 20], correlation does play an important role in many statistical methods as the basis for hypothesis tests for causality. Mirage correlations can appear in the simplest nonlinear systems [21]. Variables that may be positively correlated at some point in time can become anti-correlated some moments after or even lose all coherence. However, most causality methods do not adequately account for the fact that sudden changes in correlation over time between variables may indicate a change in the underlying temporal causal relationships.

In an attempt to overcome some of these limitations, here we introduce a new methodology for probing timevarying causal interactions using a new metric for assessing causal interactions combined with a novel machine learning architecture for causal inference, which we call Temporal Autoencoders for Causal Inference (TACI). We show the method's effectiveness on synthetic and real-world data sets, both in an absolute sense and in comparison to extant methods, particularly focused on how to find time-varying causal structure in complex dynamical systems.

Overview of Methodology

In our methodology, we adopt a two-fold approach towards developing a causal inference method that accurately assesses causality between variables $x(t)$ and $y(t)$ in the Granger sense for nonlinear systems with timevarying interactions. The first aspect of our approach is to use a novel surrogate data comparison metric - the Comparative Surrogate Granger Index (CSGI) – that measures the relative improvement in prediction accuracy when including both variables vs. one of them and a randomized version of the other. The other aspect is to use a two-headed Temporal Convolutional Network architecture to robustly capture the space of potential nonlinear mappings between variables across the entire time series. As will be observed, the CSGI with linear autoregressive models works well to identify causal interactions in situations where relatively straightforward mappings exist between variables, but the more complicated neural network model is more effective when the mappings are more non-linear.

Comparative Surrogate Granger Index (CSGI)

Informally, Granger Causality (GC) defines a causal interaction from $x(t)$ to $y(t)$ to be when knowing the full history of both $x(t)$ and $y(t)$ provides a better prediction about the future of $y(t)$ than knowing just the history of $y(t)$ alone. While there are many variations on this methodology

[22–24], the typical form used is to compare two models of similar type (e.g., linear autoregressive models, feedforward neural networks, etc.) based on their ability to predict the future state of $y(t)$. More explicitly, the comparison is between

$$y(t) \approx f(y_{t-1}, y_{t-2}, \dots) \quad (1)$$

and

$$y(t) \approx g(x_{t-1}, x_{t-2}, \dots, y_{t-1}, y_{t-2}, \dots). \quad (2)$$

Usually, an F-test is used to determine whether the latter model is preferred over the former.

This form, however, suffers from two limitations. First, the comparison is a binary one – the second model is “significantly” better than the first or it is not – thus, differences in the strength of coupling can not be detected, just the presence or absence. Second, because the F-test and similar methods incorporate strong assumptions about the underlying dynamics of the system, statistical statements deriving from these tests are often not robust under resampling or re-parameterization. In addition, because the model complexities for the two models being compared are inevitably quite different, with one typically having twice as many parameters as the other, the F-test often fails to detect causal interactions properly.

A common strategy for ameliorating these limitations is to compare not

$$f(y_{t-1}, y_{t-2}, \dots)$$

and

$$g(x_{t-1}, x_{t-2}, \dots, y_{t-1}, y_{t-2}, \dots),$$

but rather

$$f(x_{t-1}, x_{t-2}, \dots, y_{t-1}, y_{t-2}, \dots)$$

and

$$g(x_{t-1}^{(s)}, x_{t-2}^{(s)}, \dots, y_{t-1}, y_{t-2}, \dots),$$

where $x^{(s)}$ is a *surrogate* data set that shares similar statistical properties to $x(t)$ but is shuffled in some manner (e.g., shuffling values to preserve the distribution of values or shuffling phases of the time series’ Fourier Transform to preserve the frequency profile), or even a completely randomized time series. Typically, this comparison is accomplished through the Extended Granger Causality Index (EGCI) [22, 25]. If $\epsilon_y(t)$ are the residuals for fitting the future of $y(t)$ on the past of $y(t)$ and $\epsilon_{xy}(t)$ are the residuals for fitting the future of $y(t)$ on the pasts of both $x(t)$ and $y(t)$, then the EGCI is given by ratio between the relative reduction in residual variance when the past of $x(t)$ is included in the model:

$$\text{EGCI} = 1 - \frac{\text{var}(\epsilon_{xy}(t))}{\text{var}(\epsilon_y(t))}. \quad (3)$$

$y(t)$ is thus said to cause $x(t)$ if the EGCI using the actual values of $x(t)$ is significantly higher than the EGCI found substituting $x(t)$ for $x^{(s)}(t)$.

Our approach attempts to assess directly the relative increase in variance explained for the predictive model when using $x(t)$ vs. $x^{(s)}(t)$. Specifically, if R_{xy}^2 is the fraction of variance explained about the future of $y(t)$ using the pasts of $x(t)$ and $y(t)$ in the model and $R_{x^{(s)}y}^2$ is the fraction of variance explained using $x^{(s)}(t)$ and $y(t)$, then we define the Comparative Surrogate Granger Index (CSGI), $\chi_{x \rightarrow y}$, to be defined via

$$\chi_{x \rightarrow y} = \frac{R_{xy}^2 - R_{x^{(s)}y}^2}{\frac{1}{2}(R_{xy}^2 + R_{x^{(s)}y}^2)}. \quad (4)$$

This metric's advantages over the EGCI are that it is able to measure small changes in causal interactions and that it explicitly measures the difference in predictive power between using actual data and using surrogate data to predict the future. In this article, we will be measuring $\chi_{x \rightarrow y}$ and $\chi_{y \rightarrow x}$ for all pairs of variables to assess whether there is causal coupling between two variables, whether it is uni- or bi-directional, and the relative strength of the coupling.

Temporal Autoencoders for Causal Inference

While one advance in our methodology is the use of the CSGI in the previous section, the other novel contribution is the use of a new artificial neural network architecture to calculate the functions f and g that are used to predict the future of $y(t)$. The original (and still most common) models ([10]) for f and g are auto-regressive linear models of the form

$$y(t) = \sum_{i=1}^k (a_i x(t-i) + b_i y(t-i)) + \epsilon_t, \quad (5)$$

where $x^{(s)}(t)$ can be substituted for $x(t)$ when using the surrogate approach. While this relatively simple approach shows impressive performance in a variety of scenarios, these models fail to accurately predict known causal interactions for couplings that have weak to moderate coupling and are governed by nonlinear dynamics that are not well approximated by linear models [26], [27]. This inability is typically because the systems fail to satisfy separability. In other words, all information about a causative factor has to be inherent to that specific variable and can be omitted by removing that variable from the model, as is the case for purely stochastic or linear systems [16]. For systems with strongly nonlinear deterministic components, however, this assumption fails, and, accordingly, so do the predictions from auto-regressive linear GC [13], [27]. In addition, because these linear models have difficulty predicting information across multiple timescales, they often have difficulty detecting subtle shifts in causality as a function of time.

In recent years, a solution has been to replace the linear model in (5) with deep neural networks of varying architectures that, due to their expressive nature, excel in approximating complex functions [28]. These methods include the use of Variational Autoencoders to estimate causal effects [29], Causal Generative Neural Networks to learn functional causal models [30], Neural Granger to estimate non-linearly dependencies based on Granger causality principles [24], and the Temporal Causal Discovery Framework (TCDF) to address time delay causal relationships [31]. These methods, however, are often unwieldy to train, are prone to overfitting, and are susceptible to inaccuracies in the presence of a significant amount of noise.

In this paper, we introduce a novel neural network architecture for causality using a two-headed Temporal Convolutional Network (TCN) autoencoder (Fig. 1). TCNs, the primary building block of our approach, are a specialized type of neural network that integrates causal inference into convolutional architectures. First introduced for video-based action segmentation [32], TCNs quickly became popular due to their ability to extend most of the convolutional advantages of

regular CNNs – including sparsity and translational equivariance – into the time domain through a series of dilated and causal convolutional layers. A key characteristic of this framework is its simplicity, relatively long memory, and ability to outperform most convolutional architectures in autoregressive prediction tasks [33].

Our approach, which we call Temporal Autoencoders for Causal Inference (TACI) is a neural network that consists of a two-headed TCN autoencoder, where two TCNs are used to encode time series $x(t)$ and $y(t)$, and a third is used for decoding an equivalently long time series describing the future trajectory of $y(t)$ (shifted by some time, T) from a relatively low-dimensional latent space that is derived from the outputs of the first two autoencoders. A more detailed description of our model and our training methodology can be found in *Materials and methods*. Code is available here: <https://github.com/bermanlabememory/Temporal-Autoencoders-For-Causal-Inference-TACI>.

For each comparison of interest, we train four versions of this network: one using $x(t)$ and $y(t)$ as input time series to predict the future of $x(t)$, another that is the same except for replacing $x(t)$ with the surrogate data $x^{(s)}(t)$, another pair of the networks that are structured the same except with x and y reversed in each case. Given these four trained networks, we can then make predictions for the future of the appropriate variable and calculate the fraction of variance explained over a moving window (i.e., $R_{xy}^2(t)$ or $R_{x^{(s)}y}^2(t)$). From these values, we can then apply Eqn. (4) to calculate the CGI values $\chi_{x \rightarrow y}$ and $\chi_{y \rightarrow x}$, which we will use to assess causal inference between these two variables.

Other Methods We Compare Against

Alongside comparing GC with linear autoregressive models, which we will refer to here as Surrogate Linear Granger Causality (SLGC), we will also test against two other commonly used methods: Convergent Cross Mapping and Transfer Entropy.

Convergent Cross Mapping (CCM)

Convergent Cross Mapping (CCM) was introduced to determine causation in systems that could be modeled as relatively noiseless deterministic dynamical systems [16]. The core concept of this approach is that according to Takens' Embedding Theorem, if $x(t)$ and $y(t)$ are two variables of a deterministic dynamical system, one can reconstruct $x(t)$ from a delay embedding of $y(t)$ if and only if the time derivative of $x(t)$ explicitly depends on $y(t)$ [34–36]. Thus, if it is possible to predict $x(t)$ from an embedding of $y(t)$ alone, we would say that y has a causal interaction with x . Practically, these predictions are calculated by predicting $x(t)$ from $y(t)$ (and vice versa) and computing the correlation coefficient between the actual and predicted values [16], with correlations near one implying a strong causal influence and correlations near zero implying no or little influence. Here, we used Scikit Convergent Cross Mapping (skccm), a Python-based-library implementation of CCM for causal discovery [37].

Transfer Entropy

Transfer entropy (TE) is a metric that quantifies a reduction in uncertainty in predicting the future of one variable given the past of another using formalism from information theory [38]. Specifically, we can measure the transfer entropy from $y(t)$ to $x(t)$ at a given distance in the future, τ , ($T_{Y \rightarrow X}(\tau)$) via

$$T_{Y \rightarrow X}(\tau) = H(x(t)|x(t-1), \dots, x(t-\tau)) - H(x(t)|x(t-1), \dots, x(t-\tau), y(t-1), \dots, y(t-\tau)), \quad (6)$$

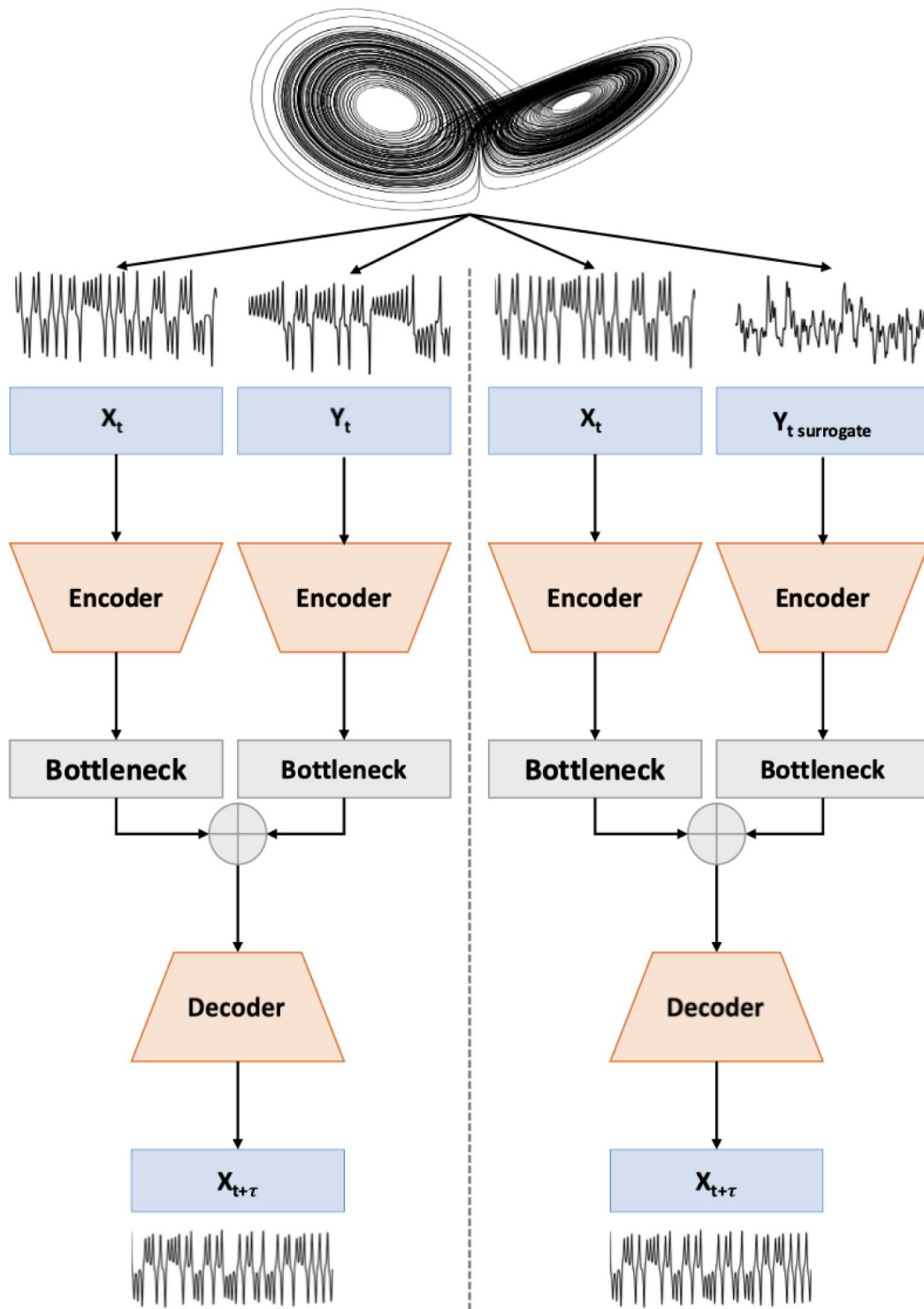


FIG.1.

Schematic of the Temporal Autoencoders for Causal Inference (TACI) Networks. We use a two-headed network consisting of Temporal Convolutional Networks that interact through a shared latent space to predict a time-shifted version of one of the two input time series. For each pair of variables we wish to examine (here, X and Y), we train two networks for each causal direction: one using X and Y as inputs and another using X and a randomized version of Y . We consider an interaction from $Y \rightarrow X$ to be causal if the network using the actual value of Y predicts the future of X better than the network using the surrogate version of Y . In this particular case, we show the approach applied to two different variables from the Lorenz system.

where $H(X|Y)$ is the Shannon entropy of the conditional probability distribution $p(X|Y)$. This quantity is zero if adding information about the past of $y(t)$ results in no reduction in our future guesses for $x(t)$, and if it is non-zero, the quantity can be interpreted as the rate of information flowing from Y to X . Practically, we calculate TE for our systems using the Java Information Dynamics Toolkit (JIDT or Infodynamics Toolkit) [39].

Results

Artificial Test Systems

To test the validity of our approach, we applied the methodology to a variety of different deterministic and stochastic dynamical system models with known causal interactions, finding that TACI performs well across all cases. In particular, we are interested in cases where the coupling changes in time, which we will explore in detail for the Coupled Hénon Maps system.

The Rössler-Lorenz System

Our first example case is a system of coupled chaotic attractors, where the Lorenz system ($\vec{y}$) [40] is driven by a Rossler oscillator ($\vec{x}$) [41]:

$$\begin{aligned}\dot{x}_1 &= -6(x_2 + x_3), \\ \dot{x}_2 &= 6(x_1 + 0.2x_2), \\ \dot{x}_3 &= 6(0.2 + x_3(x_1 - 5.7)), \\ \dot{y}_1 &= 10(y_2 - y_1), \\ \dot{y}_2 &= 28y_1 - y_2 - y_1y_3 + Cx_2^2, \\ \dot{y}_3 &= y_1y_2 - \frac{8}{3}y_3,\end{aligned}\tag{7}$$

where the constant C controls the coupling strength of the system, and the driving severely distorts the behavior of the Lorenz attractor as C increases [42]. Synchronization between $\vec{x}$ and $\vec{y}$ starts near $C = 2.14$, making the two systems' behavior effectively coupled above this point despite the lack of an explicit coupling term, making traditional formal notions of causality ill-posed (Fig. 2A) [43]. The solutions to the differential equations were generated by using a fourth-order Runge-Kutta method. C was chosen between 0 and 5, computing a time series of length 300,000 ($dt = 0.1$) after a burn-in time of 30,000 time points at each coupling strength.

As seen in Figure 2B-E, TACI is the only method of the four tried here that accurately predicts the unidirectional coupling from $\vec{x}$ to $\vec{y}$. SLGC (Fig. 2B) fails to predict any coupling whatsoever between the variables, and CCM and TE (Figs. 2C-D) predict bidirectional coupling (albeit with somewhat more information flowing from $\vec{x} \rightarrow \vec{y}$ than in the reverse). TACI, in contrast, predicts only unidirectional coupling until the point of synchronization ($C \approx 2.14$), after which, it predicts no effective causation in either direction.

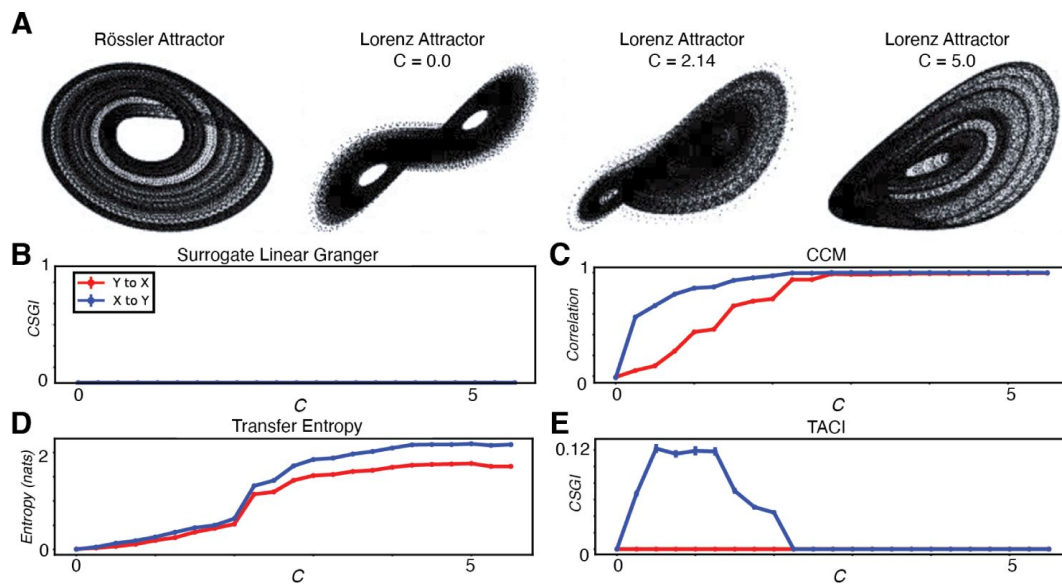


FIG.2.

Causal inference in the Rössler-Lorenz System. **A)** 2-dimensional projections of the Rössler attractor (left) and the Lorenz system (right three plots) as C increases. Mathematically, there is only coupling from $X \rightarrow Y$, but starting near $C = 2.14$, the two systems become synchronized, making finding the causal interactions an ill-posed problem. **B-E)** Results from applying the four methods to the system. Note that only TACI accurately predicts the unidirectional coupling in the regime above $C > 0$ and before synchronization occurs. Error bars are generated using a bootstrapping procedure (see Materials and Methods).

Coupled Bi-directional Two-Species Model

In contrast to the Rössler-Lorenz System, the bidirectional two-species model [44], is calculated in discrete time, and it exhibits (unsurprisingly) bi-directional coupling:

$$\begin{aligned}x(t+1) &= x(t)(3.8 - 3.8x(t) - Cy(t)) \\y(t+1) &= y(t)(3.5 - 3.5y(t) - 5Cx(t)).\end{aligned}\quad (8)$$

where C once again is the coupling strength, noting that the coupling strength is five times larger from $x \rightarrow y$ than in the reverse direction. In this system, separability is not satisfied (i.e., information about y is redundantly present in x and vice versa). Despite the fact this model is deterministic and dynamically coupled, it shows alternating periods of positive, negative, and zero correlation [16]. For values of $C \in [0, 0.35]$, we created a bivariate time series of length 300,000 (after a burn-in of 30,000 time points). The initial conditions were generated with random starting points drawn from the uniform distribution (0.01, 0.99).

Applying the four methods to these data (Fig. 3), we find that both TE and TACI correctly identify both the bi-directional aspect of the coupling and the increased causal link from $x \rightarrow y$ compared to $y \rightarrow x$. CCM identifies the bi-directionality correctly, but it does not identify the relative strength of the couplings, and SLGC is unable to identify any causal link from $y \rightarrow x$.

Coupled Autoregressive Models

Coupled autoregressive models are an extension of basic autoregressive models, intended to represent the dynamics of systems where multiple time series influence each other. In these models, the value of a variable at a given time point is not only a function of its own previous values but also depends on the past values of other variables in the system. Here, we study the following system consisting of two bidirectionally coupled autoregressive processes of the first order:

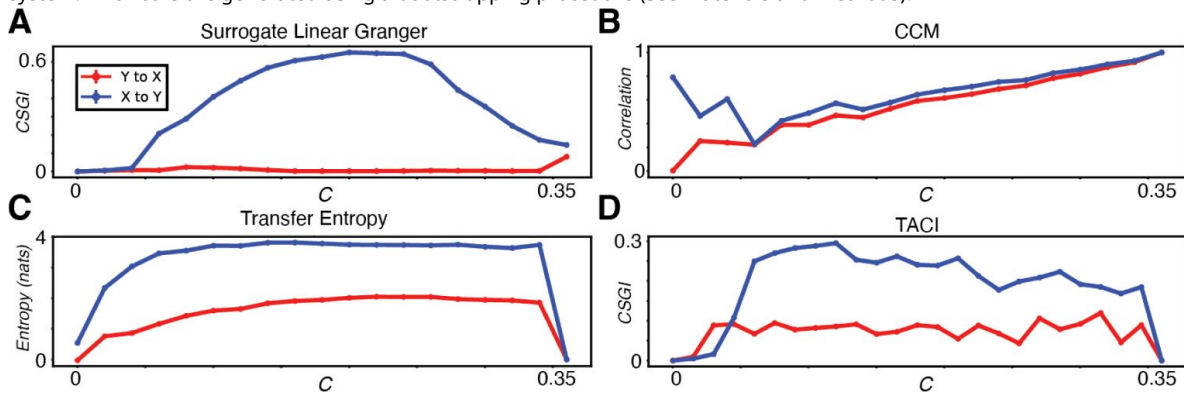
$$\begin{aligned}x(t+1) &= 0.5x(t) + 0.2y(t) + \epsilon_x(t), \\y(t+1) &= Cx(t) + 0.7y(t) + \epsilon_y(t),\end{aligned}\quad (9)$$

where C is the strength of the coupling between x and y , and $\epsilon_x(t)$ and $\epsilon_y(t)$ are drawn from a normal (Gaussian) distribution with a mean of 0 and $\sigma_x^2 = \sigma_y^2 = 0.1$. Higher values of C represent stronger couplings from $x \rightarrow y$, and for $C = 0$, the system is unidirectional (only the past of y has an impact on the future of x). We examined values of $C \in [0, 0.6]$ and created sets of bivariate time series of length $L = 300,000$ for each value of C (after a burnin time of 30,000 points). The initial conditions of the system were generated from the normal distribution with zero mean and unit variance.

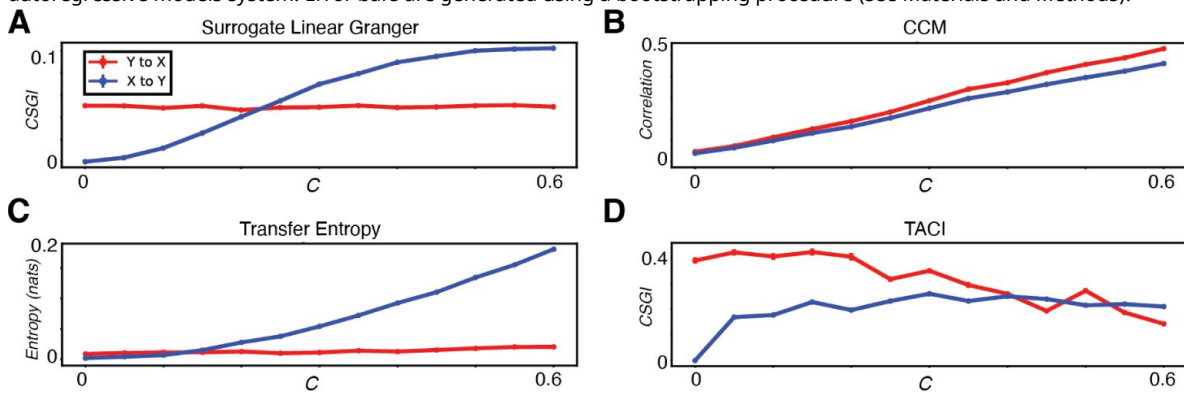
Fig. 4 shows that SLGC does very well at identifying the onset of bi-directionality for $C > 0$, with the coupling of $x \rightarrow y$ monotonically increasing with C . This fact is perhaps not surprising, as SLGC is based on precisely such linear systems. TACI also does a comparable job at detecting bi-directionality, even roughly predicting the switchover between $x \rightarrow y$ and $y \rightarrow x$ coupling strengths at $C = 0.5$. CCM, however, does not predict any coupling from $y \rightarrow x$ at $C = 0$, and TE does not predict any significant coupling from $y \rightarrow x$ across all values of C .

FIG.3.

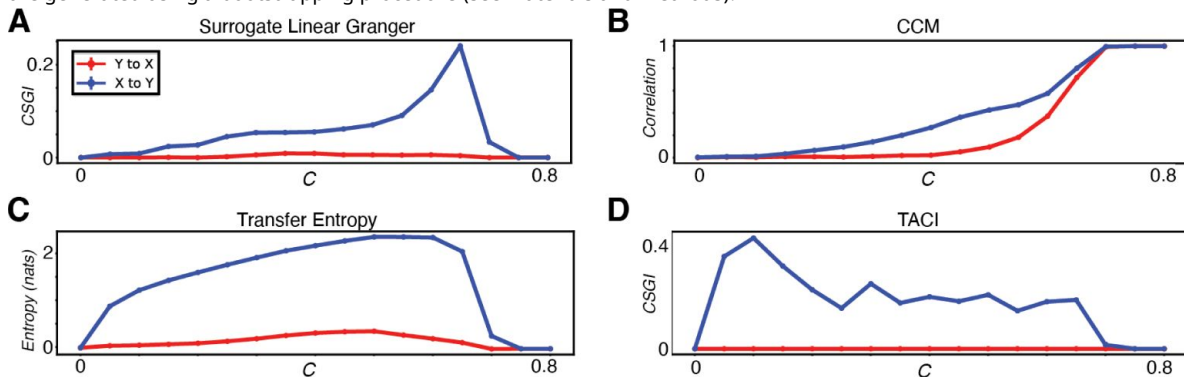
Causal inference in the bidirectional species system. **A-D**) Results from applying the four methods to the bidirectional species system. Error bars are generated using a bootstrapping procedure (see Materials and Methods).

**FIG.4.**

Causal inference in the coupled autoregressive models system. **A-D**) Results from applying the four methods to the coupled autoregressive models system. Error bars are generated using a bootstrapping procedure (see Materials and Methods).

**FIG.5.**

Causal inference in the coupled Hénon Maps system. **A-D**) Results from applying the four methods to the coupled Hénon Maps system. Here, only TACI accurately predicts univariate coupling across all values of C prior to synchronization. Error bars are generated using a bootstrapping procedure (see Materials and Methods).



Coupled Hénon Maps

Our last stationary example, the Hénon map, is a well-known example of a discrete-time dynamical system that exhibits chaotic behavior that was first developed as a simplified version of the Poincaré map of the Lorenz model [44], and in its chaotic regime, it is characterized by an attractor with a warped horseshoe shape. Here we consider a case of two Hénon maps, $\vec{x}$ and $\vec{y}$, with unidirectional coupling [45]:

$$\begin{aligned}x_1(t+1) &= 1.4 - x_1^2(t) + 0.3x_2(t), \\x_2(t+1) &= x_1(t), \\y_1(t+1) &= 1.4 - (Cx_1(t)y_1(t) + (1-C)y_1^2(t)) \\&\quad + 0.3y_2(t), \\y_2(t+1) &= y_1(t),\end{aligned}\tag{10}$$

where C controls the strength of the coupling from $\vec{x}$ to $\vec{y}$. For coupling strengths above $C > 0.65$, the systems start to show evidence of intermittent synchronizations. This on-off behavior becomes a fully synchronized state after $C > 0.7$ [46]. For $C \in [0, 0.9]$, we generated sequences of length 300,000 (after a burn-in period of 30,000) and analyzed data from x_1 and y_1 for each of the methods. TACI is the only method out of the four that correctly identifies the uni-directional coupling between from $\vec{x}$ to $\vec{y}$ (but not from $\vec{y}$ to $\vec{x}$), although SLGC is very close, it is statistically significantly different from zero at intermediate values of C . TE and CCM both predict bi-directional interactions, albeit with weaker coupling from $\vec{y}$ to $\vec{x}$ than in the reverse direction.

Non-stationary Coupled Hénon Maps

TACI is the only method that performed well across all four artificial test cases, but the challenge remains as to whether it can identify patterns in data that change over time. To test this idea, we generated time series from the coupled Hénon maps in (10) but with timevarying couplings, $C_{xy}(t)$ and $C_{yx}(t)$:

$$\begin{aligned}x_1(t+1) &= 1.4 - C_{yx}(t)y_1(t)x_1(t) \\&\quad + (1 - C_{yx}(t))x_1^2(t) + 0.3x_2(t), \\x_2(t+1) &= x_1(t), \\y_1(t+1) &= 1.4 - C_{xy}(t)x_1(t)y_1(t) \\&\quad + (1 - C_{xy}(t))y_1^2(t) + 0.3y_2(t), \\y_2(t+1) &= y_1(t).\end{aligned}\tag{11}$$

Here, the two coupling terms are similar to the coupling term, C , in (10) but with time-varying values and potentially allowing for coupling from y to x .

We performed four different tests to see how TACI performs when causal interactions alter with time: (i) setting $C_{yx}(t) = 0$ and toggling $C_{xy}(t)$ between 0 and 0.6, (ii) initially setting $C_{yx}(t) = 0$ and $C_{xy}(t) = 0.6$ and then switching the two half-way through the run, (iii) setting $C_{yx}(t) = 0$ and toggling $C_{xy}(t)$ between 0 and 0.6 but with pulses of $C_{xy}(t) = 0.6$ being set to different time widths, and (iv) setting $C_{yx}(t) = 0$ and stepping $C_{xy}(t)$ from 0 to 0.4 and back down again in steps of 0.1.

Other than the coupling changes, all time series were generated in an identical manner to the previous section. It is important to note that the network for TACI was only trained once on the entire time series, not specifically for each testing window. Thus, by creating a robust model, our network is able to identify complex causal dynamics that change in time without having to constantly fit new models, as would be the case for SLGC, CCM, and TE.

In **Fig. 6**, we show that TACI performs well in the first three of these scenarios, ably identifying when eliminations of causal interactions occur, as well as when $C_{yx}(t) = 0$ and $C_{xy}(t)$ flip. In addition, **Fig. 7** shows that the TACI network is able to identify how coupling strengths change with time.

a. Summary of Results on Artificial Test Systems Among the methods tested, only TACI is able to robustly infer known causal interactions between variables without incorrectly predicting non-existent interactions. TACI consistently differentiates between unidirectional and bidirectional coupling in low, moderate, and strong settings. Additionally, it accurately detects instances when the time series become synchronized in all tested scenarios. TACI excels in identifying complex causal dynamics that evolve over time, such as those observed in pulse systems with time-varying coupling. Given these successes in artificial systems, we will now apply the method to two real-world examples.

Jena Climate Dataset

The first data set we will test our model on is the “Jena Climate Dataset”, a detailed collection of weather measurements recorded by the Max Planck Institute for Biogeochemistry from a weather station located in Jena, Germany [47]. The dataset spans nearly eight years – from January 10, 2009, to December 31, 2016 – and includes 14 distinct meteorological features recorded every 10 minutes. These features include a wide range of atmospheric conditions, from temperature to relative humidity to vapor pressure deficit (see **Table 1** for details). Several example time series are shown in **Fig. 8**.

A key advantage of these data is that some of the interactions are known already due to empirical models of atmospheric dynamics, providing a good test case for our method on real data. One example is the relationship between relative humidity (R_H), the dew point (T_{dew}), and the temperature (T), which is given by

$$R_H = 100 \exp \frac{17.625 T_{dew}}{T_{dew} + 243.04} \bigg/ \exp \frac{17.625 T}{T + 243.04}, \quad (12)$$

where T_{dew} and T are in degrees Celsius and R_H is a percentage [48]. Calculating the partial derivative of R_H with respect to T (keeping T_{dew} fixed), we find that we should expect stronger interactions to occur from T to R_H at lower temperatures (**Fig. 9A**). After training our TACI model from each of the variables in the data set onto T , we indeed find that causal interactions peak during epochs when the temperature drops (**Fig. 9B**), showing that our method can accurately find temporal variations in causal interactions in messy real-world data.

Electrocorticography in Non-Human Primates

Lastly, we used electrocorticography (ECoG) data from non-human primates to test whether our methodology can detect time-varying interactions between brain regions from these electrophysiological signals. These data exhibit extraordinarily complex dynamics that shift in time as an animal changes its state: from sleep to wake, from satiated to hungry, from attending from one object to another, and so on [2]. These alterations are often subtle, and, thus, understanding how different regions of the brain drive one another’s activity requires a method that can detect how slight variations in the relationship between variables lead to changing interactions across time.

Here, we analyzed publicly available ECoG data from a single monkey (*Macaca fuscata*) [49–51]. These recordings consisted of 128 channels of data that recorded activity from a hemisphere of the monkey’s brain that covered the visual, temporal, parietal, motor, prefrontal,

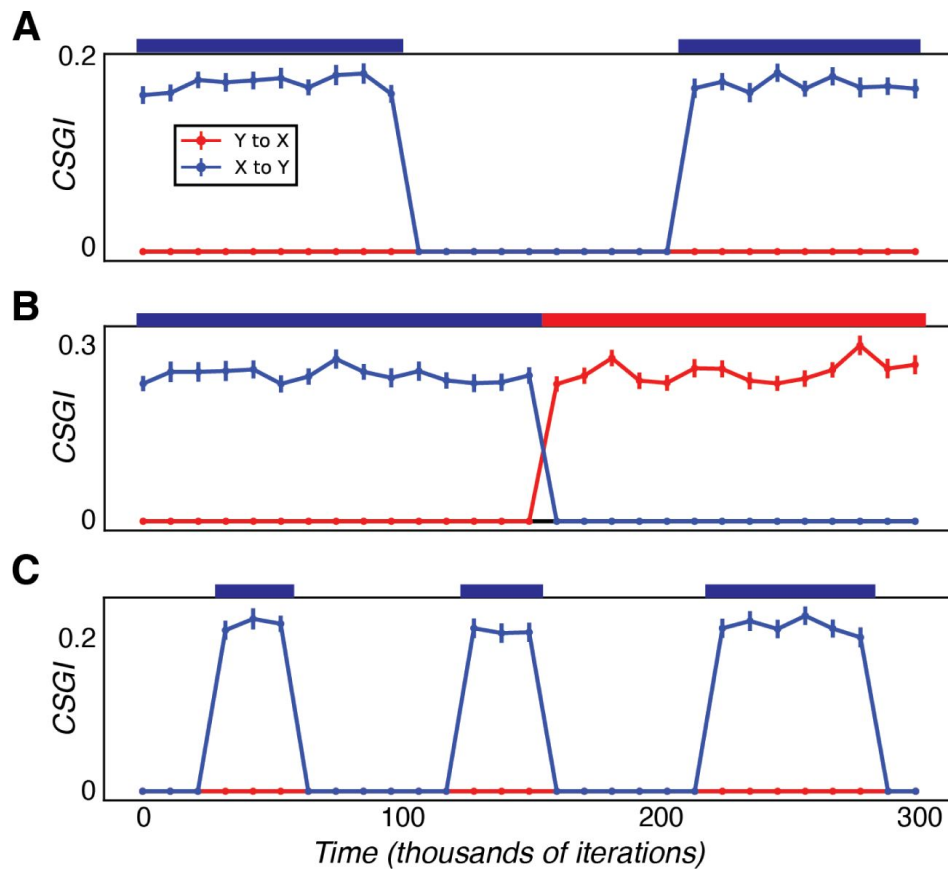


FIG.6.

TACI applied to coupled non-stationary Hénon Maps. **A)** A plot of the TACI inference when applied to the coupled Hénon Maps system where the coupling from $X \rightarrow Y$ is set to either $C_{xy} = 0.6$ (blue bar above the plot) or $C_{xy} = 0$ (no bar above the plot). **B)** Same as **A** but with a toggle from $C_{xy} = 0.6$ to $C_{yx} = 0.6$ (where the blue and red bars above the plot flip). **C)** Same as **A** but with multiple pulses of $C_{xy} = 0.6$ of varying sizes. Error bars are generated using a bootstrapping procedure (see Materials and Methods).

FIG.7.

TACI applied to coupled non-stationary Hénon Maps with ramped couplings. **A)** Inferred causal coupling as a function of time during the simulation. **B)** Time series of how the coupling from X to Y was stepped up and then down. Error bars are generated using a bootstrapping procedure (see Materials and Methods).

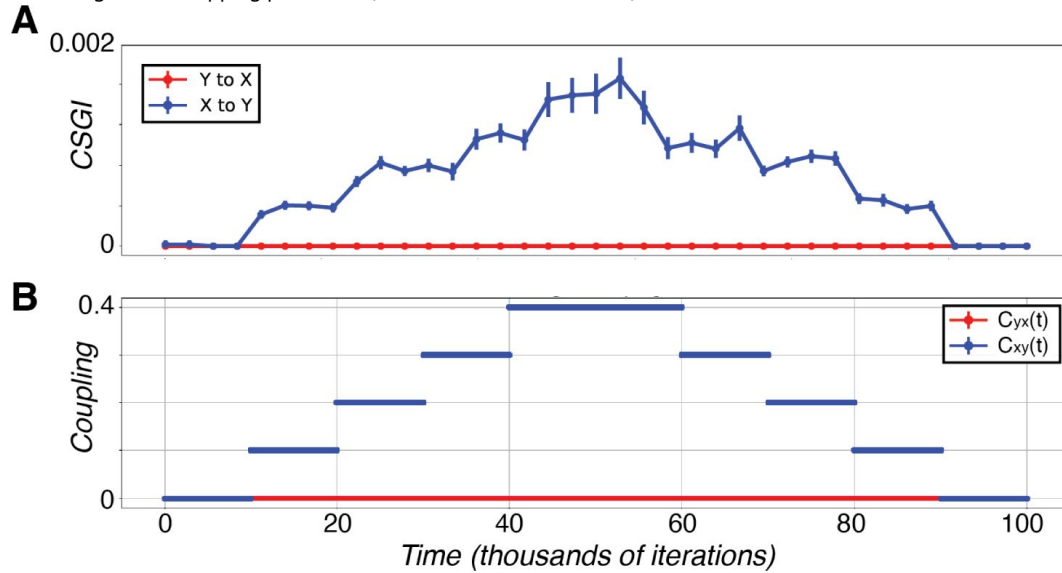


FIG.8.

Time series of Temperature, Dew Point, Relative Humidity, and Vapor Pressure Deficit from the Jena Climate Dataset.

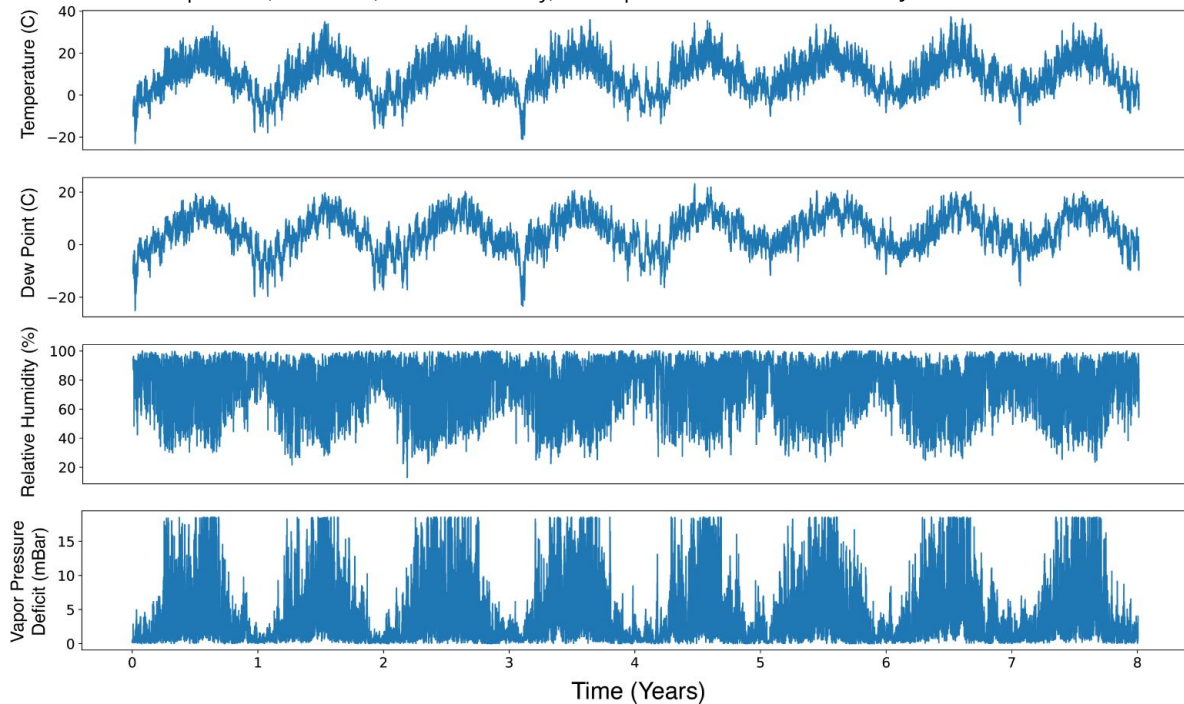


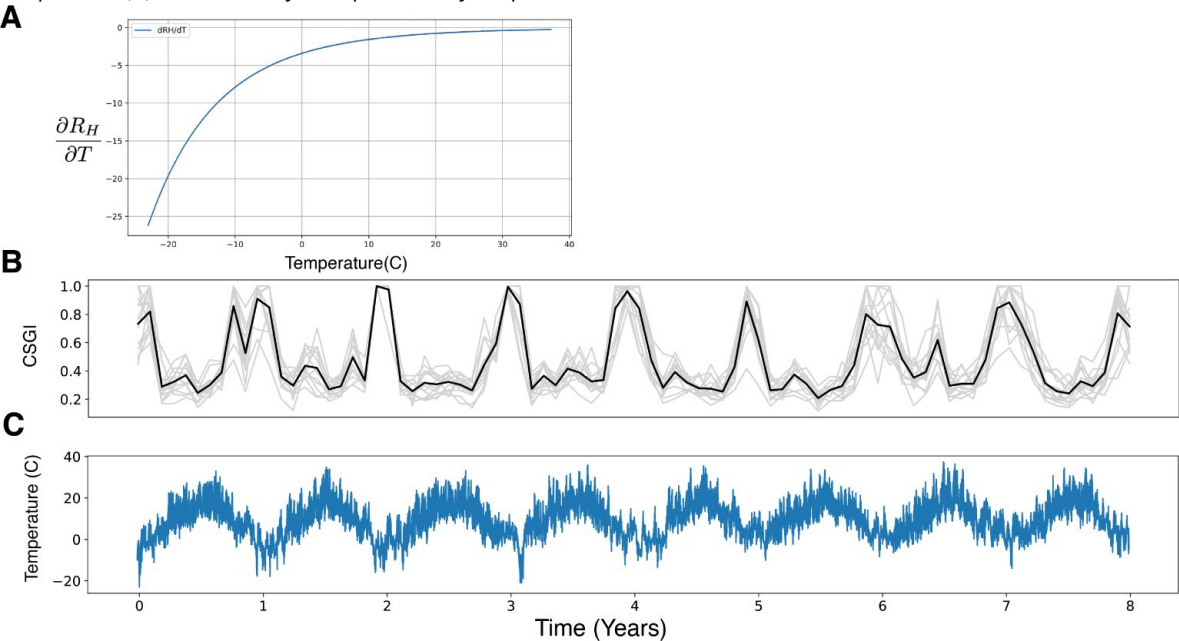
TABLE I.

Summary of Jena Climate Dataset Features

Feature	Description
Date/Time	Date-time reference
p (mbar)	Atmospheric pressure stated millibars
T (degC)	Temperature in Celsius
Tpot (K)	Temperature in Kelvin
T _{dew} (C)	Dew Point Temperature in Celsius
R _H (%)	Relative Humidity in percentage
VPmax (mbar)	Saturation vapor pressure
VPact (mbar)	Vapor pressure
VPdef (mbar)	Vapor pressure deficit
sh (g/kg)	Specific humidity
H ₂ O C (mmol/mol)	Water vapor concentration
rho (g/m ³)	Air density
wv (m/s)	Wind speed
max. wv (m/s)	Maximum wind speed
wd (deg)	Wind direction in degrees

FIG.9.

Causal interactions with relative humidity from the Jena Climate Dataset. **A)** Empirical relationship between relative humidity and air temperature (assumes $T_{dew} = 10$). Note the large negative partial derivative at low values of T . **B)** TACI predictions for causal interactions for how the other 13 variables in [Table I](#) affect relative humidity as a function of time across the eight years of the dataset (gray lines, mean trajectory is the black line). Note how causal influence peaks consistently when the temperature **(C)** is at its nadir, just as predicted by the plot in **A**.



and somatosensory cortices, sampling at 1kHz (details can be found in [49]). Data were collected during both awake and anesthetized states to examine neural activity across different consciousness levels. To generate an anesthetized state, the monkey was chair-restrained and propofol was injected intravenously. The recording sessions were structured into four distinct phases: an initial phase where the monkey is awake with eyes open, a subsequent phase where the monkey is awake but with its eyes covered, a phase where the monkey is under deep anesthesia, induced to reach a state of loss of consciousness, and a final stage where the monkey recovers from anesthesia with its eyes covered. The depth of anesthesia was assessed by monitoring the monkey's responsiveness to tactile stimulation and the presence of slow wave oscillations in the ECoG signal [49].

Previous studies analyzing these data for changes in causal interactions using Spectral Granger Causality [49] or CCM [50], but each was only able to analyze data at the level of the four phases described in the previous paragraph (each requiring training a separate model, as well). Specifically, we trained TACI on one monkey (George in [51]) with a sequence length of 50 to account for the extended autocorrelation time observed in the time series (average of 53). Approximately 53 minutes of data corresponding to the four previously outlined phases were utilized for this purpose. The training was conducted over 300 epochs or until the point of convergence. Further details of the parameters used can be found in **Table II**.

Finally, to compare with these previous studies, while we calculated the causal interactions between each pair of electrodes, we will present many of the results as the average result between pairs of electrodes assigned to the same region of the cortex. Here, we will be using the eight coarse-grained regions defined in [50]: the medial prefrontal cortex (mPFC), lateral prefrontal cortex (lPFC), pre-motor cortex (PMC), motor and somatosensory cortex (MSC), temporal cortex (TC), parietal cortex (PC), higher visual cortex (HVC), and lower visual cortex (LVC).

Fig. 10 shows time-averaged values of correlation (A), TACI-derived causal interactions (B), and Directionality (C), which we define as the difference in CGI values in one direction vs. the other, for epochs of time before, during, and after anaesthetization. For correlation, we measure the average Pearson correlation coefficient between all the electrodes assigned to the various regions. Note that the diagonal terms do not necessarily have to be equal to one here, as electrodes within a region are not perfectly correlated with one another. There are only minimal changes in brain region interactions across the three time windows when measuring correlation, but large differences emerge when analyzing the data using TACI. Specifically, we see that almost all interactions disappear during the anesthetized period, with the interactions beginning to re-emerge during the recovery period. These results differ from the results from CCM in [50], where they claimed that while some interactions decreased, others strengthened (this effect is seen in our Directionality measurements, however). Also interesting are the nearly vertical lines in **Fig. 10B**, implying that certain regions like the mPFC might be affected broadly by signals from various parts of the cortex - a finding that agrees with the commonly held notion that the mPFC's role often involves higher-level cognitive function [52]. Again, it should be noted that only one TACI network was trained per pair of interactions across all time epochs, unlike the other methods we describe, which must find interactions separately during each measurement period.

Lastly, taking advantage of the aforementioned property of TACI, we took a finer-grained look at how interactions between a pair of regions might change with time during the experiment, specifically the mPFC and the PC. In **Fig. 11**, we show how these regions' interactions alter with time. Using our approach, we observe how the coupling slowly decays upon administration of the propofol and how it rapidly increases a few minutes into the recovery period. Also interesting is that while during the awake periods, PC consistently has a casual interaction towards mPFC, the reverse interaction has significant temporal fluctuations whose study might lead to insights into how these brain regions drive each other during cognitive tasks.

Parameter	Description	Value
nb_filters	Number of filters in TCN layers.	32
kernel_size	Size of the kernel in TCN layers.	32
dilations	Dilation rates for TCN layers.	[1, 2, 4, ..., 32]
nb_stacks	Number of stacked TCN layers.	1
ts_dimension	Dimensionality of the time series.	1
dropout_rate_tcn	Dropout rate for TCN layers.	[0.0, ..., 0.5]
dropout_rate_hidden	Dropout rate for hidden layers.	[0.0, ..., 0.5]
conv_kernel_init	Kernel initializer for convolutional layers.	'he_normal'
latent_sample_rate	Downsampling rate in the latent space.	2
act_funct	Activation function used in layers.	'elu'
epochs	Number of training epochs.	300
batch_size	Batch size for training.	512
shuffle	Whether to shuffle the data during training.	[True or False]
scaling_method	Method used for scaling the input data.	Z score
loss_funct	Loss function used for training.	'mse'
noise	Standard deviation of Gaussian noise added.	[0.0, ..., 0.5]
window_len	Size of the rolling window for predictions.	value
seq_length	Length of sequences used for training/prediction.	[10, ..., 100]
lag	Lag between $x(t)$ and $y(t)$ for prediction.	[10, ..., 100]

TABLE II.

Parameters used in the TACI model training and prediction phases (ranges indicate the parameter range used across the examples in this chapter)

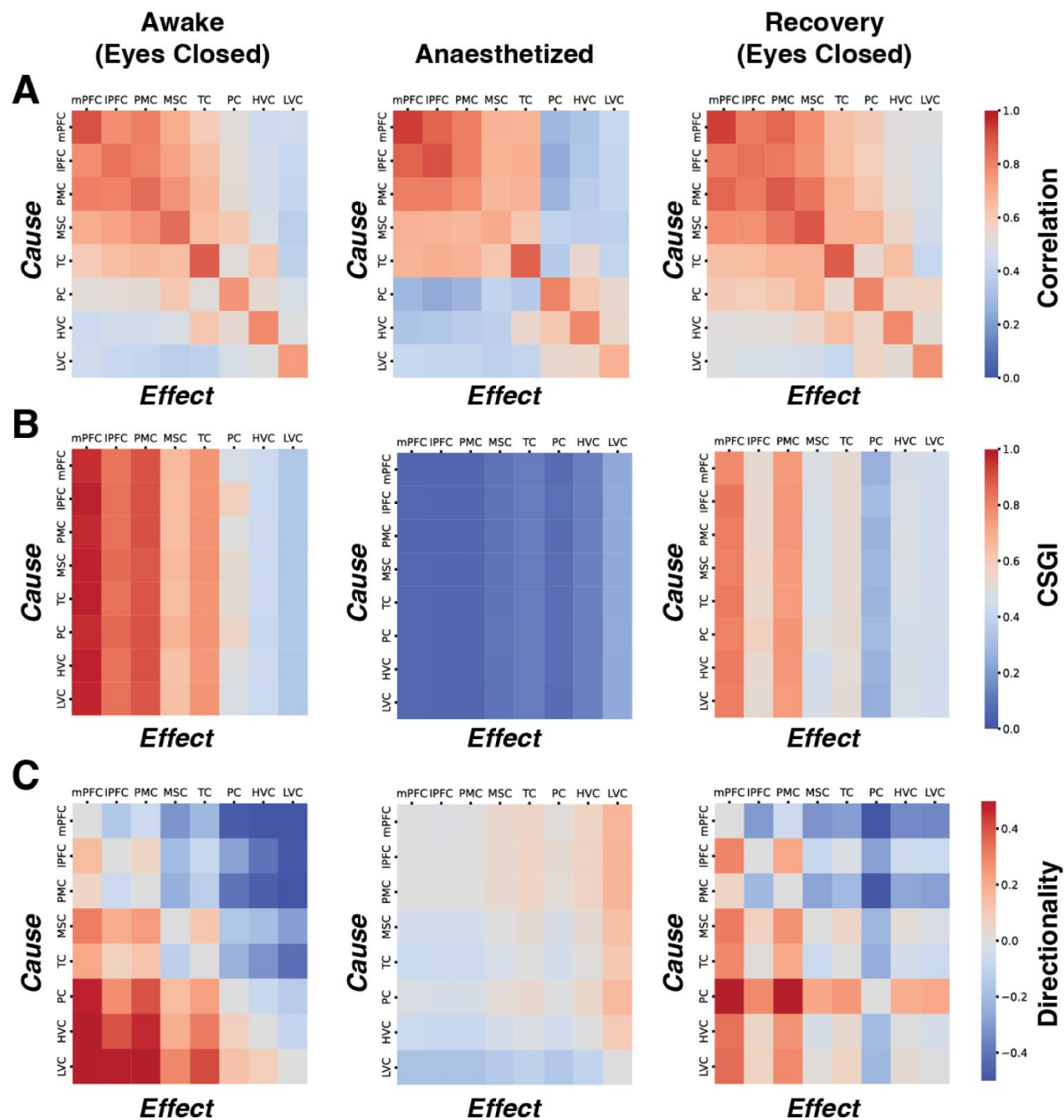


FIG.10.

Interactions between brain regions in ECoG data. Each plot here shows the average interaction between all electrodes within each of the 8 coarse-grained regions described in the text. The left matrices are from before the anesthesia was administered, the middle matrices are from when the monkey was anesthetized, and the right plots are from the recovery period. **A** is the Pearson correlation between the signals, **B** is the TACI-derived inference of causal interaction, and **C** displays the TACI Directionality – the difference between the CSGI score in one direction minus the CSGI score in the other direction.

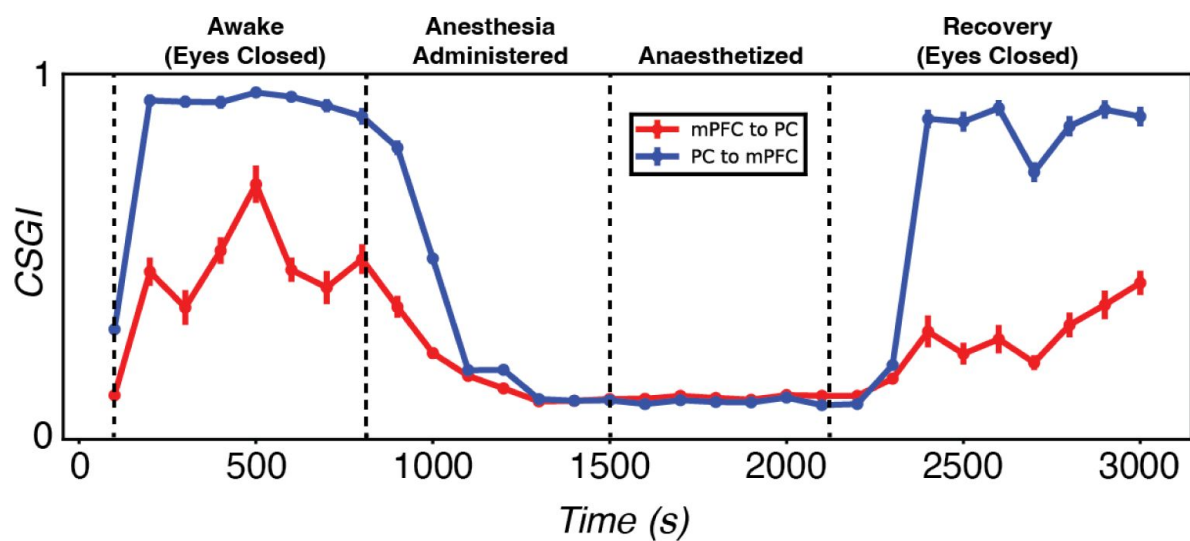


FIG.11.

Causal interactions across time between Parietal and medial Prefrontal Cortices. Plot of the average TACI-derived interactions between PC and mPFC over the course of the anesthesia experiment. Error bars are the standard errors of the mean across all electrode interactions, and the dashed lines represent change points in the experimental protocol (labeled above the axes).

Discussion

In this article, we introduce a new methodology for probing time-varying causal interactions in complex dynamical systems using a novel machine learning architecture for causal inference, Temporal Autoencoders for Causal Inference (TACI), combined with a novel metric for assessing causal interactions using surrogate data. A particular advantage of our approach is being able to train a single model that captures the dynamics of the time series across all points in time, allowing for timevarying interactions to be found without retraining, a computationally expensive endeavor for most artificial neural networks. We found that our method performed well compared to other methods in the field on synthetic data sets with known causal interactions, including those with time-varying couplings between variables. We also found that our method was able to identify known interactions between variables in a climate data set and was able to discover subtle temporal fluctuations in coupling in non-human primate ECoG data.

Our approach, while novel, is not without its limitations. One of the primary concerns is the extensive training time and the resource-intensive nature of the model. Implementing TACI, especially on large datasets, requires significant computational power and time. We envision that several technical improvements in the network architecture and training will allow for the method to be sped-up considerably, however. Another concern is the potential for overfitting due to TACI's considerable modeling capacity. While the framework is designed to capture the nuanced dynamics of causal relationships over time, like most other causal network models, this method can fit data too closely if not trained properly, resulting in models that perform exceptionally well on training data but generalize poorly to unseen data. Furthermore, TACI incorporates elements of the Granger causality approach, which means it also inherits some of its problems. Granger causality assumes that the causal variable contains unique information about the future values of the effect variable, which might not always hold true in complex systems where numerous latent factors influence outcomes. Lastly, but importantly, as our approach is based solely on observational data, TACI only attempts to provide hypotheses about causal relationships between variables or to infer important relationships between variables when perturbation experiments are impossible or unethical to perform.

These limitations withstanding, however, the results presented in this chapter provide evidence that our approach will be broadly applicable to complicated data sets with time-varying causal structure, with particular promise for neural data, where we hope to build our understanding of how parts of the brain shift their interactions as behavioral states and needs alter in the world.

Materials and Methods

At its core, TACI uses a two-headed autoencoder architecture implemented in a two-step process aiming to facilitate the prediction of future states and the inference of causal relationships between different time series datasets. In the first application, the two-headed autoencoder is utilized to process the original time series data, $x(t)$ and $y(t)$. The encoder segments of this autoencoder independently process $x(t)$ and $y(t)$, capturing and encoding their temporal dynamics and features into latent representations. These representations are then merged in the bottleneck, combining the distilled information from both time series into a unified latent space that encapsulates potential causal interactions. From this combined latent representation, the decoder works to reconstruct or predict the future trajectory of $y(t)$, shifted by a time τ . The second application involves replacing $x(t)$ with the surrogate data $x^{(s)}(t)$. This surrogate data is generated to mimic the statistical properties of $x(t)$ but is designed to break any potential causal link between $x(t)$ and $y(t)$.

This two-step process is essential for figuring out how these variables are linked to one another. The model can validate the presence of a causal relationship by comparing the predictive accuracy of the decoder when using the original $x(t)$ versus the surrogate $x^{(s)}(t)$. A significant drop in accuracy with the surrogate data suggests that the original $x(t)$ contains specific information causally linked to the future states of $y(t)$.

Architecture

In the TACI architecture, the concept of a two-headed encoder is employed to simultaneously process two distinct time series datasets, denoted as $x(t)$ and $y(t)$. This design allows for the independent yet parallel analysis of each time series, enabling the model to capture and encode their individual characteristics and temporal dynamics before merging their representations during the bottleneck process. The input sequences are selected to be greater in length than the autocorrelation time of each variable. This ensures that the sequences capture meaningful temporal dependencies and dynamics. A GaussianNoise layer is added to enhance the model's ability to generalize and prevent overfitting.

The most important part of the encoder includes the use of a Temporal Convolutional Network (TCN) layer. Thus, capturing the long-term dependencies within each time series. This layer utilizes several key parameters: “*nb filters*” sets the number of convolutional filters, “*kernelSize*” affects the temporal extent of each convolution, “*dilatiois*” allows the model to efficiently gather information across various temporal distances. Additionally, “*Dropout*” layers are used to decrease overfitting by randomly dropping units during the training phase. Following the TCN, a *Conv1D* layer continues to process the data for each series, allowing the network to change dimensionality while preserving temporal resolution. An *AveragePooling1D* layer may then downsample the *Conv1D* layer's output by pooling across the temporal dimension. This operation reduces the sequence length, emphasizing significant features and further decreasing data dimensionality. The data is subsequently processed by a series of *Dense* layers that compress it into a dense, lower-dimensional latent representation. The size of these layers decreases in each successive layer, concentrating the information into a more compact form.

The bottleneck stage starts once the two-headed encoder has finished processing and compressing the input sequences into a lower-dimensional latent space representation. The Bottleneck merges these latent representations through an element-wise multiplication operation. By combining the representations in this manner, the model effectively captures the potential interactions and dependencies between the time series, which are essential for uncovering causal relationships.

Once the latent representations are merged in the Bottleneck, this combined representation is forwarded to the Decoder. The Decoder's task is to predict the future trajectory of the target time series. The first step in the Decoder is to progressively upscale the combined latent representation. This is achieved through a series of *Dense* layers, where each layer aims to increase the dimensionality of the data. The number and size of these layers are determined by the complexity of the data and the level of compression achieved by the Encoder. After the initial upscaling, an *UpSampling1D* layer is used to increase the sequence length to its original size, effectively reversing the pooling operation performed in the Encoder. A TCN layer is used to ensure that the reconstructed data maintains its temporal integrity and dynamics. This layer mirrors the TCN configuration in the Encoder, utilizing the same parameters for “*nb filters*”, “*kernelSize*”, and “*dilations*” to capture the temporal dependencies and patterns necessary for accurate prediction. Lastly, a *Dense* output layer produces the final prediction of the future states of the target time series.

Training and Prediction

As discussed earlier, the training phase of the TACI model involves four distinct configurations of the network. Central to this phase is the use of the Mean Squared Error (MSE) as the loss function, which facilitates the optimization of predictions for future trajectories against actual observed values. The Adam optimizer [53] is employed for its adaptive learning rate capabilities. Training is performed across 300 epochs to give the model enough time for the parameters to adjust and converge toward optimal solutions. The parameters controlling the batch size and data shuffling are finely tuned to balance computational efficiency and the promotion of model generalization. Callbacks such as ReduceLROn-Plateau, EarlyStopping, and ModelCheckpoint are employed in this phase for optimizing the training process by adjusting learning rates, preventing overfitting, and preserving the best model state, respectively.

Surrogate data were created by drawing random values from a uniform distribution between zero and one until a time series the length of the original one was generated. An alternative method would have been to create a surrogate time series by first converting the original series into the frequency domain through a Fourier transform. Then, we could apply random phase shifts, making sure the amplitude spectrum remained unaltered. This randomness is crucial to breaking any specific temporal dependencies present in the original series. Following this process, an inverse Fourier transform could be employed to reconstruct the series back into the time domain. This step generates a new time series that mirrors the original in terms of its overall power distribution but only has random contingencies with its partner data set. In practice, however, we found that this latter methodology did not result in more accurate results in training TACI, so we focused on the initially described method for generating surrogate data in this study.

After training is completed, the model moves on to the prediction phase, where the focus shifts to evaluating the trained model. In the first step of the prediction phase, the pre-trained models are loaded, each representing a unique configuration designed in the training phase to capture and analyze the causal dynamics between the time series datasets $x(t)$ and $y(t)$. At the same time, the full original dataset is divided into sequences with the same length and structure as the models were trained on. The prediction process occurs over defined rolling windows to allow for a temporal exploration of the dataset, enabling the models to make predictions for future states of the time series within each window. The models' accuracy in forecasting future time series states is quantitatively evaluated for each rolling window using the R^2 metric. To enhance the reliability and confidence of these assessments, 100 bootstrap samples are generated for each window. The causal inference for each rolling window can be determined using the CSGI Eq. 4. Through this calculation, the model not only quantifies the strength and direction of the causal relationship but also shows its variation over time, providing a dynamic and temporal perspective on causal inference.

For each interval, a bootstrap strategy is implemented. This strategy involves creating a set number of surrogate samples by randomly resampling within the interval. These samples are then used to evaluate the model's predictions, which are generated under two conditions: one using the actual interactions between the time series and another using the surrogate data. By employing Equation 4, it's possible to derive scores from which we compute both the mean and standard deviation. These computations provide insight into the average performance and variability of the model's predictions across the bootstrap samples. The utilization of bootstrap methods significantly enhances the analytical depth by ensuring that the derived error bars and confidence intervals are supported by a solid statistical foundation. These statistics play a vital role in establishing the error bars in the plotted figures. By repeating this procedure across all intervals, the method provides a comprehensive view of how model performance fluctuates over time and under different conditions.

Acknowledgements

Both authors were supported by the Human Frontier Science Program (RGY0076/2018) and the Simons Foundation (707102 & 876207), and JC was supported by the NSF Physics of Living Systems Student Research Network (PHY-1806833). GJB would like to acknowledge the Aspen Center for Physics, where many of the initial ideas for this work were generated.

References

- [1] Zhao Y, Billings SA, Wei HL, Sarrigiannis PG (2013) **A Parametric Method to Measure Time-Varying Linear and Nonlinear Causality With Applications to EEG Data** *IEEE Transactions on Biomedical Engineering* **60**:3141–3148 <https://doi.org/10.1109/TBME.2013.2269766>
- [2] Buzsáki G (2006) **Rhythms of the Brain** Oxford, UK: Oxford University Press
- [3] of Scientific Unions Scientific Committee on Oceanic Research IC, of Scientific Unions Scientific Committee on Oceanic Research Canadian National Committee IC (1983) **Proceedings of the Joint Oceanographic Assembly 1982 General Symposia: Dalhousie University, Halifax, Nova Scotia, Canada** Canadian National Committee for the Scientific Committee on Oceanic Research
- [4] Jammazi R, Ferrer R, Jareño F, Shahzad SJH (2017) **Time-varying causality between crude oil and stock markets: What can we learn from a multiscale perspective?** *International Review of Economics & Finance* **49**:453–483 <https://doi.org/10.1016/j.iref.2017.03.007>
- [5] Emirmahmutoglu F, Denaux Z, Topcu M (2021) **Time-varying causality between renewable and non-renewable energy consumption and real output: Sectoral evidence from the United States** *Renewable and Sustainable Energy Reviews* **149** <https://doi.org/10.1016/j.rser.2021.111326>
- [6] Hammoudeh S, Ajmi AN, Mokni K (2020) **Relationship between green bonds and financial and environmental variables: A novel time-varying causality** *Energy Economics* **92** <https://doi.org/10.1016/j.eneco.2020.104941>
- [7] Latif M, Kleeman R, Eckert C (1997) **Greenhouse Warming, Decadal Variability, or El Niño? An Attempt to Understand the Anomalous 1990s** *Journal of Climate* **10**:2221–2239 [https://doi.org/10.1175/1520-0442\(1997\)010<2221:GWDVOE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<2221:GWDVOE>2.0.CO;2)
- [8] Wang X, Wang D, Zhou W (2009) **Decadal variability of twentieth-century El Niño and La Niña occurrence from observations and IPCC AR4 coupled models** *Geophysical Research Letters* **36** <https://doi.org/10.1029/2009GL037929>
- [9] Jajcay N, Kravtsov S, Sugihara G, Tsonis AA, Palus M (2018) **Synchronization and causality across time scales in El Niño Southern Oscillation** *NPJ Climate and Atmospheric Science* **1** <https://doi.org/10.1038/s41612-018-0043-7>
- [10] Granger CWJ (1969) **Investigating Causal Relations by Econometric Models and Cross-spectral Methods** *Econometrica* **37**:424–438
- [11] Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D (2019) **Detecting and quantifying causal associations in large nonlinear time series datasets** *Science Advances* **5** <https://doi.org/10.1126/sciadv.aau4996>
- [12] Nauta M, Bucur D, Seifert C (2019) **Causal Discovery with Attention-Based Convolutional Neural Networks** *Machine Learning and Knowledge Extraction* **1**:312–340 <https://doi.org/10.3390/make1010019>

- [13] Stokes PA, Purdon PL (2017) **A study of problems encountered in Granger causality analysis from a neuroscience perspective** *Proceedings of the National Academy of Sciences* **114** <https://doi.org/10.1073/pnas.1704663114>
- [14] McCrorie JR, Chambers MJ (2006) **Granger causality and the sampling of economic processes** *Journal of Econometrics* **132**:311–336 <https://doi.org/10.1016/j.jeconom.2005.02.002>
- [15] Yuan AE, Shou W (2022) **Data-driven causal analysis of observational biological time series** *eLife* **11** <https://doi.org/10.7554/eLife.72518>
- [16] Sugihara G, May R, Ye H, hao Hsieh C, Deyle E, Fogarty M, et al. (2012) **Detecting Causality in Complex Ecosystems** *Science* **338**:496–500 <https://doi.org/10.1126/science.1227079>
- [17] Monster Dan, Fusaroli Riccardo, Tylen Kristian, Roepstorff Andreas, Sherson Jacob F. (2017) **Causal inference from noisy time-series data — Testing the Convergent Cross-Mapping algorithm in the presence of noise and external influence** *Future Generation Computer Systems* **73**:52–62 <https://doi.org/10.1016/j.future.2016.12.009>
- [18] Ye H, Deyle ER, Gilarranz LJ, Sugihara G (2015) **Distinguishing time-delayed causal interactions using convergent cross mapping** *Scientific Reports* **5** <https://doi.org/10.1038/srep14750>
- [19] Hsieh Ch, Glaser SM, Lucas AJ, Sugihara G (2005) **Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean** *Nature* **435**:336–340
- [20] May RM, Levin SA, Sugihara G (2008) **Ecology for bankers** *Nature* **451**:893–894
- [21] Mysterud A, Stenseth NC, Yoccoz NG, Langvatn R, Steinheim G (2001) **Nonlinear effects of large-scale climatic variability on wild and domestic herbivores** *Nature* **410**:1096–1099
- [22] Chen Y, Rangarajan G, Feng J, Ding M (2004) **Analyzing multiple nonlinear time series with extended Granger causality** *Physics Letters A* **324**:26–35 <https://doi.org/10.1016/j.physleta.2004.02.032>
- [23] Marinazzo D, Pellicoro M, Stramaglia S (2008) **Kernel Method for Nonlinear Granger Causality** *Phys Rev Lett* **100** <https://doi.org/10.1103/PhysRevLett.100.144103>
- [24] Tank A, Covert I, Foti N, Shojaie A, Fox EB (2021) **Neural Granger Causality** *IEEE Transactions on Pattern Analysis and Machine Intelligence* :1–1 <https://doi.org/10.1109/tpami.2021.3065601>
- [25] Schiatti L, Nollo G, Rossato G, Faes L (2015) **Extended Granger causality: a new tool to identify the structure of physiological networks** *Physiological Measurement* **36** <https://doi.org/10.1088/0967-3334/36/4/827>
- [26] Granger CWJ (1969) **Investigating Causal Relations by Econometric Models and Cross-spectral Methods** *Econometrica* **37**:424–438
- [27] Pearl J (2009) **Causality: Models, Reasoning and Inference** USA: Cambridge University Press
- [28] Goodfellow I, Bengio Y, Courville A (2016) **Deep Learning** MIT Press
- [29] Louizos C, Shalit U, Mooij J, Sontag D, Zemel R, Welling M (2017) **Causal Effect Inference with Deep Latent-Variable Models**

- [30] Goudet O, Kalainathan D, Caillou P, Guyon I, Lopez-Paz D, Sebag M (2018) **Causal Generative Neural Networks**
- [31] Nauta M, Bucur D, Seifert C (2019) **Causal Discovery with Attention-Based Convolutional Neural Networks** *Machine Learning and Knowledge Extraction* **1**:312–340 <https://doi.org/10.3390/make1010019>
- [32] Lea C, Vidal R, Reiter A, Hager GD, Hua G, Jégou H (2016) **Temporal Convolutional Networks: A Unified Approach to Action Segmentation** *Computer Vision – ECCV 2016 Workshops Cham*: Springer International Publishing :47–54
- [33] Bai S, Kolter JZ, Koltun V (2018) **An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling** *arXiv*
- [34] Dixon PA, Milicich MJ, Sugihara G (1999) **Episodic Fluctuations in Larval Supply** *Science* **283**:1528–1530 <https://doi.org/10.1126/science.283.5407.1528>
- [35] Takens F, Rand D, Young LS (1981) **Detecting strange attractors in turbulence** *Dynamical Systems and Turbulence, Warwick 1980* Berlin, Heidelberg: Springer Berlin Heidelberg :366–381
- [36] Deyle ER, Sugihara G (2011) **Generalized Theorems for Nonlinear State Space Reconstruction** *PLOS ONE* **6**:1–8 <https://doi.org/10.1371/journal.pone.0018295>
- [37] NickC1 (2018) **skccm: State-space Reconstruction by k- Nearest Neighbors Convergent Cross Mapping** *GitHub*
- [38] Schreiber T (2000) **Measuring Information Transfer** *Phys Rev Lett* **85**:461–464 <https://doi.org/10.1103/PhysRevLett.85.461>
- [39] Lizier JT (2014) **JIDT: An information-theoretic toolkit for studying the dynamics of complex systems** *Frontiers in Robotics and AI* **1** <https://doi.org/10.3389/frobt.2014.00011>
- [40] Lorenz EN (1963) **Deterministic Nonperiodic Flow** *Journal of the Atmospheric Sciences* **20**:130–141 [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- [41] Rössler OE (1976) **An Equation for Continuous Chaos** *Physics Letters A* **57**:397–398 [https://doi.org/10.1016/0375-9601\(76\)90101-8](https://doi.org/10.1016/0375-9601(76)90101-8)
- [42] Quyen MLV, Martinerie J, Adam C, Varela FJ (1999) **Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy** *Physica D: Nonlinear Phenomena* **127**:250–266 [https://doi.org/10.1016/S0167-2789\(98\)00258-9](https://doi.org/10.1016/S0167-2789(98)00258-9)
- [43] Quiroga RQ, Arnhold J, Grassberger P (2000) **Learning driver-response relationships from synchronization patterns** *Phys Rev E* **61**:5142–5148 <https://doi.org/10.1103/PhysRevE.61.5142>
- [44] Hénon M (1976) **A two-dimensional mapping with a strange attractor** *Communications in Mathematical Physics* **50**:69–77 <https://doi.org/10.1007/BF01608556>
- [45] Paluš M, Vejmelka M (2007) **Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections** *Phys Rev E* **75** <https://doi.org/10.1103/PhysRevE.75.056211>

- [46] Schiff SJ, So P, Chang T, Burke RE, Sauer T (1996) **Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble** *Phys Rev E* **54**:6708–6724 <https://doi.org/10.1103/PhysRevE.54.6708>
- [47] (2024) **Jena Climate Dataset**
- [48] Lawrence MG (2005) **The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications** *Bulletin of the American Meteorological Society* **86**:225–234 <https://doi.org/10.1175/BAMS-86-2-225>
- [49] Yanagawa T, Chao ZC, Hasegawa N, Fujii N (2013) **Large-Scale Information Flow in Conscious and Unconscious States: an ECoG Study in Monkeys** *PLOS ONE* **8** <https://doi.org/10.1371/journal.pone.0080845>
- [50] Tajima S, Yanagawa T, Fujii N, Toyozumi T (2015) **Untangling Brain-Wide Dynamics in Consciousness by Cross-Embedding** *PLoS Comput Biol* **11** <https://doi.org/10.1371/journal.pcbi.1004537>
- [51] Nagasaka Y, Shimoda K, Fujii N (2011) **Multidimensional Recording (MDR) and Data Sharing: An Ecological Open Research and Educational Platform for Neuroscience** *PLOS ONE* **6**:1–7 <https://doi.org/10.1371/journal.pone.0022561>
- [52] Anastasiades PG, Carter AG (2021) **Circuit organization of the rodent medial prefrontal cortex** *Trends in Neurosciences* **44**:550–563 <https://doi.org/10.1016/j.tins.2021.03.006>
- [53] Kingma DP, Ba J (2014) **Adam: A method for stochastic optimization** *arXiv*

Author information

Josuan Calderon

Department of Physics, Emory University, Atlanta, GA, 30322, United States

Gordon J Berman

Department of Physics, Emory University, Atlanta, GA, 30322, United States, Department of Biology, Emory University, Atlanta, GA, 30322, United States

For correspondence: gordon.berman@emory.edu

Editors

Reviewing Editor

Armita Nourmohammad

University of Washington, Seattle, United States of America

Senior Editor

Aleksandra Walczak

École Normale Supérieure - PSL, Paris, France

Reviewer #1 (Public review):

Summary:

The authors make a new contribution with careful computational validation/exploration of their method on synthetic and real-world datasets. Overall, I find their results significant and their presentation compelling.

Strengths:

The authors provide extensive computational validation of their approach to synthetic and real-world datasets of increasing complexity.

Weaknesses:

The authors should provide a comparison of their approach to other state-of-the-art neural network-based methods. Without this, it is difficult to tell which aspects of their approach (novel coupling metric, or network architecture) are most important for their results.

<https://doi.org/10.7554/eLife.100692.1.sa1>

Reviewer #2 (Public review):

Summary:

This paper introduces a new methodology for probing time-varying causal interactions in complex dynamical systems using a novel machine-learning architecture of Temporal Autoencoders for Causal Inference (TACI) combined with a novel metric (CSGI) for assessing causal interactions using surrogate data. This is a timely contribution in the field of causal inference from temporal data which has been largely restricted to stationary time series so far. However, the benchmarking of the proposed methods could be improved.

Strength:

The method's capacity to uncover piecewise time-varying non-linear dynamic systems is demonstrated on synthetic datasets as well as on two real-world applications on climate and brain activity data. A particular advantage of the approach is to train a single model capturing the dynamics of the whole time series, thereby allowing for time-varying interactions to be found without retraining over different time periods.

Weaknesses:

(1) It is not clear why the new metric Comparative Surrogate Granger Index CSGI (Eq.6) should be better than the Extended Granger Causality Index EGCI (Eq.5), which can also be used to compare the information about $y(t)$ contained in the actual data $x(t)$ versus in a randomized surrogate $x^s(t)$, as implemented in the proposed metric (Eq.6).

(2) The benchmarking of the new approach TACI against earlier metrics (ie Surrogate Linear Granger, Convergent Cross Mapping, and Transfer Entropy) should be revised:

(a) The details of the computation should be provided to clarify how the different metrics are estimated notably between multidimensional variables [for instance to estimate $T_{y \rightarrow x}$ for $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$].

(b) Reliable implementations of the different metrics should be used, as some of the reported results do not seem right. In particular, the unidirectional examples, Eq.9 (Figure 2) and Eq.12 (Figure 5), are expected to lead to vanishing transfer entropies from Y to X, ie $T_{y \rightarrow x} = 0$, for all

values of the coupling parameter below the synchronization threshold. This can be verified by computing transfer entropies as conditional mutual information using MIIC R package, i.e. $T_{y \rightarrow x} = I(x(t); y(t-1) | x(t-1))$.

(c) Besides, some reported benchmarks focus on peculiar non-linear systems displaying somewhat "pathological" behaviors. For instance, the two Hénon maps with unidirectional coupling Eq.12 (Figure 5) lead to an equality between the two variables, i.e. $y(t)=x(t)$ for all t , above the synchronization threshold $C>0.7$. This leads mathematically to zero transfer entropy upon synchronization, as $I(x(t); y(t-1)) = 0$.

d) By contrast, Eq.9 (Fig.2) leads to strongly coupled, yet non-identical variables above the synchronization threshold. This strong coupling can be shown to yield non-vanishing transfer entropies in both directions, as observed in Figure 2c, and does not correspond to "incorrect prediction of non-existent interactions", as stated in the "Summary of Results on Artificial Test Systems". Clearly synchronized variables do interact and their bidirectional transfer entropies are actually consistent with a non-causal (or bidirectional) relationship. Only a vanishing transfer entropy in one direction implies a causal relation (in the opposite direction). Likewise, vanishing transfer entropies in both directions imply either independent variables or a spurious dependency between them due to an unobserved common cause L , i.e. $X \leftarrow (L) \rightarrow Y$. This is usually represented with a bidirected edge ($X \leftrightarrow Y$), which is different from a bidirectional relation corresponding to two opposite unidirectional edges (ie $X \rightarrow Y$ and $X \leftarrow Y$). It is therefore surprising that TACI metric vanishes in both directions upon synchronization in this case (Eq.9, Figure 2), as one would expect to learn variable $y(t)$ more reliably using the actual data $x(t)$.

e) In order to assess TACI performance on non-stationary time series, it might be more informative to benchmark it on datasets displaying intermittency rather than synchrony. In particular, the change of causal directions over time, presented as one of the motivations for the new approach, should be more thoroughly benchmarked in the paper. For instance, it would be nice to demonstrate the tracking of the spontaneous reversal of causal relation in a simple 'toggle switch' regulatory network between two mutually repressing genes + expression noise. This is something that causal inference methods assuming stationarity cannot do.

(3) Concerning the real-world applications, the analysis of the electrocorticography (ECoG) data does not seem to be in strong disagreement with the general trends of the original more detailed study by Tajima et al 2015. Could the authors better delineate what are the common versus conflicting findings between the two approaches? The main difference appears to be the near loss of interaction in the anesthetized state, which might be linked to TACI's tendency to report no interaction between synchronized variables as discussed in d) above. Does the anesthetized state correspond to a global synchrony of the brain regions? This could be easily validated by a more direct analysis of synchrony.

<https://doi.org/10.7554/eLife.100692.1.sa0>