



Causal coupling inference from multivariate time series based on ordinal partition transition networks

Narayan Puthanmadam Subramaniyam ·
Reik V. Donner · Davide Caron ·
Gabriella Panuccio · Jari Hyttinen

Received: 5 October 2020 / Accepted: 4 June 2021 / Published online: 18 June 2021
© The Author(s) 2021

Abstract Identifying causal relationships is a challenging yet crucial problem in many fields of science like epidemiology, climatology, ecology, genomics, economics and neuroscience, to mention only a few. Recent studies have demonstrated that ordinal partition transition networks (OPTNs) allow inferring the coupling direction between two dynamical systems. In this work, we generalize this concept to the study of the interactions among multiple dynamical systems and we propose a new method to detect causality in multivariate observational data. By applying this method to numerical simulations of coupled linear stochastic processes as well as two examples of interacting nonlinear dynamical systems (coupled Lorenz systems and a network of neural mass models), we demonstrate that our approach can reliably identify the direction of interactions and the associated coupling delays. Finally, we study real-world observational microelectrode array electrophysiology data from rodent brain slices to identify the causal coupling structures underlying epilepti-

form activity. Our results, both from simulations and real-world data, suggest that OPTNs can provide a complementary and robust approach to infer causal effect networks from multivariate observational data.

Keywords Causality · Transition networks · Information theory · Nonlinear time series analysis · Ordinal patterns

1 Introduction

The detection of causal interactions is a fundamental problem in both natural and social sciences [1, 2]. Reliable statistical inference of causality can aid in making predictions, which may eventually allow to design proper intervention strategies [3]. For example, in neurological disorders such as epilepsy, predicting a seizure from electroencephalography (EEG) recordings and preventing its occurrence is a long-standing problem that has not yet been completely solved. In this context, reliable seizure prediction algorithms and consequent appropriate interventions to prevent seizure occurrence could be life-saving for the patient.

In the past few decades, several approaches have been developed to both identify and quantify the interdependence between observational time series. The Granger causality test can be used to infer causality between two time series [4], and its extension, partial directed coherence, allows to infer causality from multivariate data [5]. The Granger causality test is based on

N. P. Subramaniyam (✉) · J. Hyttinen
Faculty of Medicine and Health Technology and
BioMediTech Institute, Tampere University,
33520 Tampere, Finland
e-mail: narayan.subramaniyam@tuni.fi

R. V. Donner
Magdeburg–Stendal University of Applied Sciences and
Potsdam Institute for Climate Impact Research (PIK) –
Member of the Leibniz Association, Magdeburg, Germany

D. Caron · G. Panuccio
Enhanced Regenerative Medicine, Istituto Italiano di Tecnologia,
Genoa, Italy

linear regression. Under this framework, a causal relation of one system (or variable) X affecting another system Y (i.e., $X \rightarrow Y$) is statistically inferred if the variance of the prediction error of the behavior of Y can be reduced by including past information about observations of X in the regression model for Y . Since the classical Granger causality analysis is model-based and (in a strict sense) only valid for linear systems, more general bivariate approaches based on information theory have been proposed for identifying causality in applications to nonlinear dynamical systems [6]. These methods include, among others, transfer entropy [7], time-delayed mutual information [8], and the multivariate extension of transfer entropy [6]. In particular, transfer entropy can be considered as a generalization of Granger causality for nonlinear systems [9], while it has been shown to be equivalent to Granger causality for linear Gaussian models [10].

With the rising availability of powerful computer infrastructures and large observational data sets, the study of statistical interdependencies based on multivariate time series has found widespread applications, for instance, in the area of neuroscience. In the context of multi-channel EEG recordings, such interdependencies can be related to the concept of functional connectivity, which commonly refers to the statistical associations between pairs of signals that are measured in terms of simple linear correlations or variants thereof and, thus, cannot distinguish between direct and indirect connectivity [11]. Several statistical association measures have been employed to estimate functional connectivity from EEG data, including Pearson correlation, mutual information, synchronization likelihood, and phase locking value, all of which are symmetric measures and do not provide information regarding the direction of the associated information flow [8, 12–15].

While studying complex systems such as neuronal networks, it is often important to identify not only the symmetric statistical associations, but also the causal relationships (i.e. driver-response relationships) between the involved sub-systems [2]. For example, effective connectivity [16] between individual neurons (or ensembles of neurons) is characterized by direct (causal) relationships. Hence, statistical inference of causality based on time series is a more informative approach to better understand the interplay between neuronal connectivity and dynamics, and is the key to study the structure-function relationship of neuronal networks [17]. To estimate effective connectiv-

ity, recent works using Granger causality-based methods such as directed coherence and transfer entropy have been proposed [18]. Although these concepts provide asymmetric measures, they cannot distinguish between causal (direct) and non-causal (indirect) interdependence among two subsystems. By contrast, partial directed coherence [5] (which has been often used in EEG studies) can distinguish between direct and indirect causal links but assumes a simple linear multivariate autoregressive model for describing the data. More recently, Bayesian filtering approaches have been suggested to estimate connectivity. However, such approaches also make strong assumptions about the underlying dynamical model [19–21]. Another approach known as dynamic causal modeling [22] has also been widely used to estimate causality among brain regions using EEG, magnetoencephalography or functional magnetic resonance imaging measurements. However, dynamic causal modeling is also a model-based approach and makes strong assumptions on the process by which the data are generated. In addition, it requires the pre-specification of several competing hypotheses, which may not always be available. For a detailed review on various methods for estimating neural connectivity and their applications in neuroscience, we further refer to [23, 24] and the references therein.

In the last years, a great deal of interest has emerged in characterizing dynamical systems using complex network-based time series analysis methods [25]. Those methods include, among others, recurrence networks [26–28], visibility graphs, [29] and transition networks [30], all of which feature different definitions of nodes and links in the resulting network representations of the time series under study. For instance, in the case of recurrence networks, the edges are defined based on the proximity of observed states in phase space, whereas for visibility graphs [26], mutually visible elements in a univariate time series are linked to form a network [29]. Finally, in the case of transition networks [30], certain discrete states or patterns are defined as nodes, and if one of these dynamical structures is followed by the other with nonzero probability along the observed (or reconstructed) trajectory, a directed edge is established between the corresponding nodes [31]. However, the majority of previous applications of the aforementioned methods have focused on univariate time series, while a generalization to multivariate time series would be a necessary step to allow

detecting signatures of causality. This aspect has not yet been systematically explored in full detail in the recent literature.

In this work, we consider a particular class of transition networks known as ordinal partition transition networks (OPTNs) [32] to infer causality from multivariate time series. OPTNs are based on the ordinal patterns embedded in a time series, the systematic analysis of which has originally been proposed in [33]. Such ordinal patterns reflect the respective rank order among a predefined sequence of univariate observation values. Identifying the series of subsequent ordinal patterns in a univariate time series results in a particular symbolic representation of the observed system's trajectory. It has been shown that under specific conditions, this ordinal partition exhibits the generating property, which makes it attractive for applications since it implies topological conjugacy between phase space and the ordinal symbolic dynamics [30]. For an unconstrained stochastic process, all possible ordinal patterns occur with equal probability. By contrast, for a time series produced by deterministic dynamics, certain ordinal patterns commonly do not appear, which are known as forbidden patterns [34] and provide a window of opportunity to test for possible determinism in a time series [35–37]. Furthermore, given the symbolic representation of the phase space trajectory (which controls the respective frequency of the different ordinal patterns), it is possible to compute a variety of dynamical characteristics, such as permutation entropy or a plethora of statistical complexity measures that can be defined based on the latter concept. Taken together, ordinal pattern-based analysis offers several advantages as compared to other more traditional nonlinear time series analysis techniques. The resulting methods are conceptually simple, computationally fast, and can capture information about the short-range temporal structure of the underlying time series [38]. In addition, they have been shown to be robust against additive noise. Finally, the calculation of ordinal patterns does not require any *a priori* knowledge of the data range, which is practical and advantageous in time series analysis [38].

An OPTN is based on the ordinal symbolic encoding of a time series and consists of [31] (1) nodes, which represent the individual ordinal patterns and (2) probability-weighted edges, which represent the transition frequencies between two successive ordinal pat-

terns. Previous applications of statistical complexity measures derived from OPTNs include the classification of cardiac dynamics based on electrocardiography data [30,39] and the analysis of EEG data from healthy and epileptic humans [40]. Although recent attempts have addressed bi- and multivariate extensions of OPTNs [34], often with the aim of characterizing different types of synchronization transitions, they have not yet provided thorough information about causal relationships among multivariate time series. Most recently, Ruan *et al.* [31] have proposed a strategy for the estimation of several complexity measures based upon bipartite OPTNs, which allows for statistical inference of the coupling direction among paired time series. However, their approach is limited by its bivariate nature. Hence, when applied to a multivariate data set, it cannot distinguish between direct and indirect causal connections among the individual time series.

To overcome the aforementioned limitations of previous OPTN-based methods, in this work we propose an extension of OPTN-based time series analysis, which leverages the construction of multiple bipartite OPTNs (M-OPTN) to account for multivariate (i.e., comprising more than two components) time series. Specifically, we outline and thoroughly test an approach for distinguishing direct from indirect causal connections based on the conditional Shannon entropies of the bipartite constituents of the M-OPTN [31]. In order to demonstrate the effectiveness of our approach, we apply the proposed method to coupled linear stochastic processes, nonlinear dynamical systems (exemplified by three interacting Lorenz systems), and a network of coupled neural mass models. The latter type of system has been shown previously to mimic the dynamics exhibited by neurophysiological time series [41] and thus serves as a validation tool to test the applicability of our method to neuronal time series. Finally, as a real-world example, we study *in vitro* microelectrode array (MEA) recordings from an *in vitro* model of acute ictogenesis, i.e., rodent brain slices in which epileptiform discharges are induced by pharmacological treatment. The network interactions and the associated delays of epileptiform discharges propagation that occur in this *in vitro* model have been extensively characterized [45,46] and serve as a reliable reference to validate the causal network relationships and delays estimated by the proposed method.

2 Methodology

2.1 Ordinal partition transition networks

Given a univariate time series $X = \{x_t\}_{t=1}^T$, following Takens' embedding theorem, we can qualitatively reconstruct the underlying phase space trajectory by using M successively lagged replications of X , each separated by a lag d , yielding the vector

$$\mathbf{z}_t = [x_t, x_{t+d}, \dots, x_{t+(M-1)d}], \quad (1)$$

for $t = 1$ to $T - (M - 1)d$, where M and d are the embedding dimension and delay, respectively. Each embedding vector \mathbf{z}_t is mapped to a sequence of integers $(s_0, s_1, \dots, s_{M-1})$ that describes the rank order of its components (with 0 indicating the smallest value) and is a unique permutation of the set $\{0, 1, \dots, M-1\}$, thereby satisfying

$$x_{t+s_0d} \leq x_{t+s_1d} \leq x_{t+s_2d} \leq \dots \leq x_{t+s_{M-1}d} \quad (2)$$

and

$$s_{l-1} < s_l \quad \text{if} \quad x_{t-s_{l-1}} = x_{t-s_l}. \quad (3)$$

Note that there exist $M!$ different possible ordinal patterns when a time series is embedded in M dimensions, and we denote these patterns by $\pi_1, \pi_2, \dots, \pi_{M!}$.

As an example, consider a 5-dimensional embedding of a time series yielding an embedding vector

$$\{x_t, x_{t+d}, x_{t+2d}, x_{t+3d}, x_{t+4d}\} = \{3, 9, 10, 1, 6\}. \quad (4)$$

Here, $x_{t+3d} < x_t < x_{t+4d} < x_{t+d} < x_{t+2d}$, and thus this partition would be mapped to the ordinal pattern or symbol $\pi_k = \{3, 0, 4, 1, 2\}$. The exact numerical value of the resulting integer index $k \in \{1, \dots, M!\}$ depends on the specific sorting of the permutations, the default of which may differ among different algorithms and programming languages.

For a univariate time series, we can then construct an (unweighted or weighted) OPTN with $M!$ nodes by first repeating this encoding procedure for each embedding vector. A weighted OPTN is obtained by setting the weight of the edge between two nodes (permutations) to be equal to the empirical frequency of "transitions" (i.e., successive occurrences) between the corresponding possible ordinal patterns. An unweighted OPTN simply contains a directed edge of unit weight between the corresponding nodes if this frequency is nonzero.

Ruan *et al.* recently extended the idea of OPTNs to bivariate time series, that may be interacting either

linearly or non-linearly [31]. In their framework, given the time series $\{x_{1,t}\}_{t=1}^T$ and $\{x_{2,t}\}_{t=1}^T$, derived from two dynamical systems X_1 and X_2 , we can derive the associated sequences of ordinal patterns underlying each time series as described above, containing the ordinal patterns $\pi_i^{x_1}$ and $\pi_j^{x_2}$ for X_1 and X_2 , respectively. One can now compute the (instantaneous or time-lagged) conditional co-occurrence frequencies $p(\pi_j^{x_2} | \pi_i^{x_1, \tau})$ by simply counting the number of cases in which $\pi_j^{x_2}$ occurs with a time-lag of τ following an occurrence of $\pi_i^{x_1}$. Note that we deviate here from the notation in [31] in order to allow a straightforward generalization of the concept of co-occurrence probabilities by including multiple variables X_i that may possibly cause variations of the given X_2 at different delays τ_i (see Section 2.2 below). Thus, $\tau = 0$ corresponds to looking at simultaneous co-occurrence while $\tau > 0$ would imply looking at lagged co-occurrence. Given these conditional co-occurrence frequencies, Ruan *et al.* proposed the estimation of the conditional entropy (CE) [31], given as

$$H_\tau(X_2|X_1) = - \sum_{i=1}^{M!} \sum_{j=1}^{M!} p(\pi_i^{x_1, \tau}, \pi_j^{x_2}) \log_2 p(\pi_j^{x_2} | \pi_i^{x_1, \tau}), \quad (5)$$

which gives the interaction $X_1 \rightarrow X_2$ at a delay of τ . The interaction in the other direction at lag τ , $X_2 \rightarrow X_1$ can be defined in an analogous way as follows,

$$H_\tau(X_1|X_2) = - \sum_{i=1}^{M!} \sum_{j=1}^{M!} p(\pi_j^{x_2, \tau}, \pi_i^{x_1}) \log_2 p(\pi_i^{x_1} | \pi_j^{x_2, \tau}), \quad (6)$$

If X_1 and X_2 are independent and their different ordinal patterns uniformly distributed, then $p(\pi_j^{x_2, \tau}, \pi_i^{x_1}) = \frac{1}{(M!)^2}$ and $p(\pi_i^{x_1} | \pi_j^{x_2, \tau}) = \frac{1}{M!}$ and thus $H_\tau(X_1|X_2) = \log_2 M!$, which is the upper bound for the CE value, denoted as H_{max} . On the other hand, if X_1 and X_2 are fully dependent, then $p(\pi_i^{x_1} | \pi_j^{x_2, \tau}) = 1$ and ideally $H_\tau(X_1|X_2) = 0$. Thus, as the strength of causal interaction from X_2 to X_1 at a given time lag τ increases, $H_\tau(X_1|X_2)$ decreases.

2.2 Causal inference strategy based on entropy measures from OPTNs

When dealing with multivariate data, i.e. data from three or more interacting systems, it is necessary to dis-

tinguish direct links from indirect ones. For example, consider a transitive chain $X_1 \rightarrow X_2 \rightarrow X_3$ (Fig. 1 (a)). Applying the bivariate OPTN derived entropy measure as described in [31] would lead to the detection of a non-existing connection from $X_1 \rightarrow X_3$, as shown in Fig. 2 (a). Similarly, consider a fork pattern, where $X_1 \rightarrow X_2$ and $X_1 \rightarrow X_3$ (Fig. 1 (b)). Here, X_1 is a common driver to both X_2 and X_3 and this will lead to a spurious connection between the X_2 and X_3 as shown in Fig. 2 (b). Note that in Fig. 2, an interaction $m \rightarrow n$ at negative delay is to be interpreted as $n \rightarrow m$.

In order to distinguish such direct from indirect links, a sophisticated way of conditioning has to be employed, thereby generalizing the previous strictly bivariate approach. In the following, we will detail a possible methodology to use OPTN-based entropy measures to infer causality from multivariate time series as outlined in Fig. 3. Note that this methodology rests upon certain general assumptions common to causal inference methods, most notably, the completeness of the set of variables analyzed (i.e. the absence of any possible hidden drivers).

STEP I: Pre-processing of observational data

Depending on the particular research problem, observational time series first have to be pre-processed, which includes standard procedures like band-pass filtering, resampling and removal of any noise or artifacts if present. If necessary, the data can be further divided into a number of overlapping windows to obtain a time-dependent estimate of the coupling measure.

STEP II: Construction of M-OPTNs

Given a time series, its phase space is reconstructed following Takens' embedding theorem. The components of the resulting embedding vectors are then rank ordered to obtain the symbolic representation of the time series. For an N -channel multivariate time series, N such OPTNs are constructed, are referred to as M-OPTNs.

STEP III: CE from M-OPTN

After constructing the M-OPTN, we compute the bivariate information theoretic measure of CE as given in Eq. (5) for each pair of variables and define the matrix

Algorithm 1 Construct M-OPTN from multivariate time series

```

1: procedure COMPUTEM- OPTN
2:   Input: Multivariate time series  $X_1, X_2, \dots, X_N$ , embedding dimension  $M$ , and lag  $d$ .
3:   Output: M-OPTN  $\Pi$ 
4:   for  $n = 1$  to  $N$  do
5:     for  $t = 1$  to  $T - (M - 1)d + 1$  do
6:       Map  $\mathbf{z}_{n,t}$  to symbol  $(s_0, \dots, s_{M-1})$  using Equations (2) to (3).
7:       Assign ordinal patterns  $\Pi[t, n] = \pi_k^{X_n}$  based on the respective permutation for each  $t$  and  $n$ .
8:     end for
9:   end for
10: end procedure

```

$$\mathbf{H}_\tau = \begin{bmatrix} H_\tau(X_1) & H_\tau(X_1|X_2) & \dots & H_\tau(X_1|X_N) \\ H_\tau(X_2|X_1) & H_\tau(X_2) & \dots & H_\tau(X_2|X_N) \\ \vdots & \vdots & \ddots & \vdots \\ H_\tau(X_N|X_1) & H_\tau(X_N|X_2) & \dots & H_\tau(X_N) \end{bmatrix}, \tag{7}$$

where each off-diagonal term $H_\tau(X_n|X_m)$ represents a possible causal link from the $m - th$ time series to the $n - th$ time series, i.e., $X_m \rightarrow X_n$, at delay τ . Note that in what follows, we will not make use of the diagonal elements of \mathbf{H}_τ (representing the classical Shannon entropies – i.e. in our specific case the permutation entropies – of the individual processes), so that they could be safely ignored or just put to zero.

Let $\mathbf{H} = \{\mathbf{H}_{\tau_1}, \mathbf{H}_{\tau_2}, \dots, \mathbf{H}_{\tau_J}\}$ denote the CE matrices obtained from the M-OPTN for a range of J delays $\mathbf{T} = \{\tau_1, \dots, \tau_J\}$. Each of the matrices \mathbf{H}_τ obtained above is next thresholded using a hard threshold to obtain a new matrix $\hat{\mathbf{H}}_\tau$ with elements

$$\hat{H}_\tau(X_n|X_m) = h_{nm\tau} = \begin{cases} H_{\max}, & \text{if } H_\tau(X_n|X_m) \geq \lambda H_{\max} \\ H_\tau(X_n|X_m), & \text{otherwise} \end{cases} \tag{8}$$

where $H_{\max} = \log_2 M!$ and λ is usually set between 0.99 and 1. Due to the finite sample size, two independent processes will not have a CE value exactly equal to H_{\max} . To account for this numerical issue, we allow for some tolerance by setting the parameter λ . After this step, only dominant neighbors (both causal and non-causal) are retained for each node. The resulting matrix $\hat{\mathbf{H}}_\tau$ represents a weighted, directed network of N

Fig. 1 Example of spurious connections (shown as dotted red arrows) that would be inferred from a directed chain **a** and a fork pattern **b** if direct links are not distinguished from indirect links

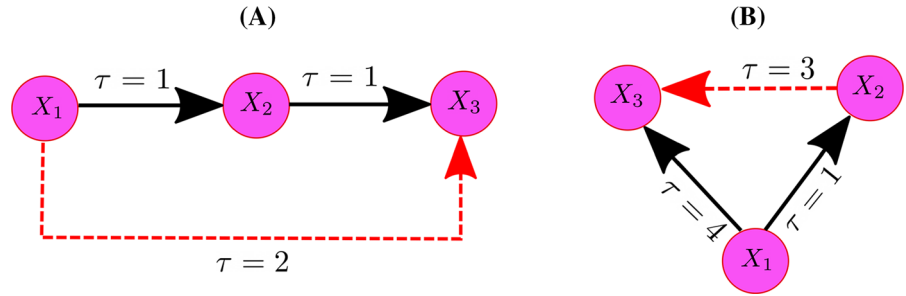


Fig. 2 Causal network inferred with the bivariate approach using OPTN-based CE for the directed chain **a** and fork **b** with a spurious connection $1 \rightarrow 2$ at $\tau = 2$ for the directed chain and $2 \rightarrow 3$ at $\tau = 3$ for the fork pattern

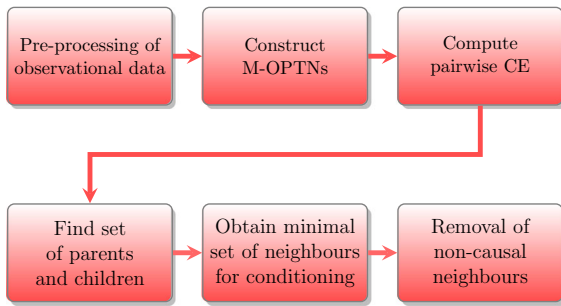
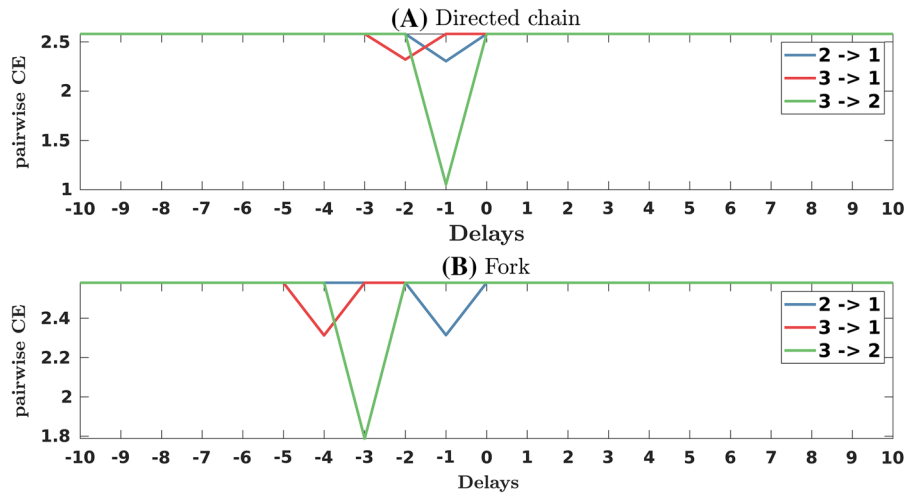


Fig. 3 Proposed methodology

nodes, where the strength of the link between the node at delay τ is inversely proportional to the CE value $H_\tau(X_n|X_m)$ and no link exists between two nodes if $H_\tau(X_n|X_m) \geq \lambda H_{max}$.

STEP IV: Find set of parents and children

The set of matrices $\hat{\mathbf{H}} = \{\hat{\mathbf{H}}_{\tau_1}, \hat{\mathbf{H}}_{\tau_2}, \dots, \hat{\mathbf{H}}_{\tau_j}\}$ represents a weighted multi-layer network $G = \{V, E\}$, where $V = \{X_1, \dots, X_N\}$ denotes a set of nodes (i.e.

Algorithm 2 Compute pairwise CE using M-OPTN

```

1: procedure COMPUTECE
2:   Input: M-OPTN  $\Pi$ , threshold  $\lambda$  and range of delays  $\mathbf{T} = \{\tau_1, \dots, \tau_j\}$ .
3:   Output: CE matrix  $\hat{\mathbf{H}}$ 
4:   for  $i = 1$  to  $J$  do
5:     for  $n = 1$  to  $N$  do
6:       for  $m = 1$  to  $N$  do
7:         Compute  $H_{\tau_i}(X_n|X_m)$  using Equation (5)
8:       end for
9:     end for
10:  end for
11:  Set  $\mathbf{H} = [\mathbf{H}_{\tau_1}, \mathbf{H}_{\tau_2}, \dots, \mathbf{H}_{\tau_j}]$ .
12:  Threshold  $\mathbf{H}$  using  $\lambda$  and  $H_{max}$  as shown in Equation (8).
13: end procedure
    
```

the different component processes) common to the different layers (which represent the different delays τ_j), while $E = \{E_1, \dots, E_J\}$ is a set of edges that will commonly differ among the layers. For every node X_m in the multi-layer network defined by $\hat{\mathbf{H}}$, we identify a set of k_m parents $\mathcal{P}_{X_m} = \{p_1, \dots, p_k\}$ at delays $\{\tau_{p_1}, \dots, \tau_{p_{k_m}}\}$ and l_m children $\mathcal{C}_{X_m} = \{c_1, \dots, c_l\}$ at delays $\{\tau_{c_1}, \dots, \tau_{c_{l_m}}\}$, respectively, which can span

across all possible layers. The set of parents for a node X_m is given by

$$\mathcal{P}_{X_m} = \{h_{mn\tau} | h_{mj\tau} < H_{max}, n \neq m, \tau \in \mathbf{T}\} \quad (9)$$

where $h_{mj\tau}$ describes the element in row m and column j of the single-layer adjacency matrix $\hat{\mathbf{H}}_\tau$, cf. Equation (8). In a completely analogous way, we define the set of children of X_n node as

$$\mathcal{C}_{X_n} = \{h_{mn\tau} | h_{in\tau} < H_{max}, m \neq n, \tau \in \mathbf{T}\} \quad (10)$$

where $h_{in\tau}$ represents the element in row i and column n of $\hat{\mathbf{H}}_\tau$. By defining those two sets for all nodes (variables) X_m (X_n), we collapse the information contained in the multi-layer adjacency matrix $\hat{\mathbf{H}}$ to the essential strong bivariate (time-lagged) linkages among the set of considered variables.

Algorithm 3 Find set of parents and children from pairwise CE matrix

```

1: procedure FINDPDC
2:   Input: Set of matrices  $\mathbf{H}$ ,  $H_{max}$ .
3:   Output:  $\mathcal{P}_{X_n}$  at delays  $\{\tau_{p_1}, \dots, \tau_{p_k}\}$  and  $\mathcal{C}_{X_n}$  at delays  $\{\tau_{c_1}, \dots, \tau_{c_l}\}$  for each node  $X_n$ .
4:   for  $n = 1$  to  $N$  do
5:     Compute  $\mathcal{P}_{X_n}$  and  $\mathcal{C}_{X_n}$  using Equation (9) and (10).
6:   end for
7: end procedure

```

STEP V: Identifying the minimal set of neighbors for conditioning

Given at set of parents for node X_m , $\mathcal{P}_{X_m} = \{p_1, \dots, p_k\}$ at delays $\{\tau_{p_1}, \dots, \tau_{p_k}\}$, to test for a causal connection from node X_n to node X_m , we seek to define a minimal conditioning set $\mathcal{P}_{X_m}^{min} \subset \mathcal{P}_{X_m}$, given as

$$\mathcal{P}_{X_m}^{min} = \mathcal{P}_{X_m} \cap \hat{\mathcal{C}}_{X_n} = \{p'_1, \dots, p'_r\} \quad (11)$$

where $\hat{\mathcal{C}}_{X_n}$ is the set of children of node X_n , which does not include X_m , i.e. $\hat{\mathcal{C}}_{X_n} = \mathcal{C}_{X_n} \setminus X_m$. The set $\hat{\mathcal{C}}_{X_n}$ will be an empty set if $\mathcal{C}_{X_n} = \{X_m\}$, i.e. the only child of node X_n is node X_m , and in this case we set $\mathcal{P}_{X_m}^{min} = \{X_m\}$. Also, it is possible that $\mathcal{P}_{X_m}^{min}$ will be an empty set due to no common elements between \mathcal{P}_{X_m} and $\hat{\mathcal{C}}_{X_n}$. In this case, we set $\mathcal{P}_{X_m}^{min} = \mathcal{P}_{X_m} \cap \mathcal{P}_{X_n}$. If $\mathcal{P}_{X_m}^{min}$ is still an empty set, then we set $\mathcal{P}_{X_m}^{min} = \{X_m\}$. Note that in each of these cases, along with the conditioning set $\mathcal{P}_{X_m}^{min}$, we also obtain the corresponding delays, $\{\tau'_1, \dots, \tau'_r\}$,

with $|\mathcal{P}_{X_m}^{min}| = r$. In order to facilitate reliable computation of CE (see STEP VI) to exclude non-causal neighbors, in all our applications we restrict ourselves here to $r = 3$ if $r > 4$, by choosing the three most dominant neighbors, based on their respective CE values.

Algorithm 4 Find minimal set of neighbors for conditioning

```

1: procedure FINDMINCONDITIONINGSET
2:   Input:  $\mathcal{P}_{X_n}$  at delays  $\{\tau_{p_1}, \dots, \tau_{p_k}\}$  and  $\mathcal{C}_{X_n}$  at delays  $\{\tau_{c_1}, \dots, \tau_{c_l}\}$  for each node  $X_n$  and  $r$ .
3:   Output:  $\mathcal{P}_{X_m}^{min}$ 
4:   for  $m = 1$  to  $M$  do
5:     for  $n = 1$  to  $N$  do
6:       Set  $\hat{\mathcal{C}}_{X_n} = \mathcal{C}_{X_n} \setminus X_m$ 
7:       if  $\hat{\mathcal{C}}_{X_n} = \emptyset$  then
8:          $\mathcal{C}_{X_n} = \{X_m\}$ 
9:          $\mathcal{P}_{X_m}^{min} = \{X_m\}$ 
10:        break
11:      end if
12:      Set  $\mathcal{P}_{X_m}^{min}$  according to Equation (11).
13:      if  $\mathcal{P}_{X_m}^{min} = \emptyset$  then
14:         $\mathcal{P}_{X_m}^{min} = \mathcal{P}_{X_m} \cap \mathcal{P}_{X_n}$ 
15:      end if
16:      if  $\mathcal{P}_{X_m}^{min} = \emptyset$  then
17:         $\mathcal{P}_{X_m}^{min} = \{X_m\}$ 
18:      end if
19:    end for
20:  end for
21:  if  $|\mathcal{P}_{X_m}^{min}| > r$  then
22:    Set  $|\mathcal{P}_{X_m}^{min}| = r$ 
23:  end if
24: end procedure

```

STEP VI: Removal of non-causal neighbors by proper conditioning

To check if X_n is a truly causal parent to X_m , we compute

$$\epsilon_{X_n} = H(X_m | \mathcal{P}_{X_m}^{min}) - H(X_m | \mathcal{P}_{X_m}^{min}, X_n). \quad (12)$$

If X_n is an indirect causal connection to X_m , then $\epsilon_{X_n} = 0$, since conditioning on X_n should not reduce $H(X_m | \mathcal{P}_{X_m}^{min})$ any further. However, since we are dealing with finite data, $\epsilon_{X_n} \approx 0$. In practice, we set another pre-defined threshold δ and if $\epsilon_{X_n} < \delta$, then X_n is considered as an indirect causal link to X_m . The CE $H(X_m | \mathcal{P}_{X_m}^{min})$, with $|\mathcal{P}_{X_m}^{min}| = r$, is given as

$$H(X_m | \mathcal{P}_{X_m}^{\min}) = - \sum_{i=1}^{M!} \sum_{j=1}^{M!} p(\pi_i^{p'_1, \tau'_1}, \dots, \pi_i^{p'_r, \tau'_r}, \pi_j^{x_m}) \log p(\pi_j^{x_m} | \pi_i^{p'_1, \tau'_1}, \dots, \pi_i^{p'_r, \tau'_r}). \tag{13}$$

The CE $H(X_m | \mathcal{P}_{X_m}^{\min}, X_n)$ is defined in an analogous way.

Algorithm 5 Remove non-causal neighbors based on CE differences

```

1: procedure REMOVE_NCN
2:   Input: Matrix H,  $\delta$ .
3:   Output: Matrix H.
4:   for  $m = 1$  to  $M$  do
5:     for  $n = 1$  to  $N$  do
6:       Compute  $\epsilon_{X_n}$  using Equations (12) and (13).
7:       if  $\epsilon_{X_n} < \delta$  then
8:         Set  $H(X_m | X_n) = H_{max}$ 
9:       end if
10:    end for
11:  end for
12: end procedure
    
```

3 Numerical examples

In the following, we present results from the application of the proposed methodology to simulations of linearly interacting stochastic processes as well as interacting nonlinear dynamical systems including Lorenz systems and a network of neural mass models. In the case of interacting stochastic processes and interacting Lorenz system, we varied δ from 0 to 0.5. We also added observational noise to the simulated data,

$$y(t) = x(t) + e(t) \tag{14}$$

where $e(t) \sim \beta \mathcal{N}(0, 1)$, where β is the noise level (NL), which is set at 0.1, 0.2 and 0.4 times the standard deviation of original noise free time series.

In the case of the network of neural mass models, we varied δ from 0 to 0.25 and added noise at the level of 0.5 and 1 times the standard deviation of original noise-free time series, which corresponds to realistic signal-to-noise ratios commonly found in EEG data (Amplitude signal-to-noise ratios of 2 and 1, respectively).

For evaluating the results of our methodology for the different simulations, we define the number of true positives (#TP) as the number of correctly identified links, and the number of false negatives (#FN) as the

number of missed links. The number of false positives (#FP) is defined as the number of incorrectly identified links, and the number of true negatives (#TN) corresponds to the number of correctly identified non-links. We then define the true positive rate (TPR) and false positive rate (FPR) as

$$\begin{aligned} \text{TPR} &= \frac{\#TP}{\#TP + \#FN}, \\ \text{FPR} &= \frac{\#FP}{\#FP + \#TN}. \end{aligned} \tag{15}$$

In addition, we use the F_1 -score to quantify the accuracy of the method, which is given as

$$F_1 = \frac{\#TP}{\#TP + 0.5(\#FP + \#FN)}. \tag{16}$$

3.1 Interacting stochastic processes

We first simulated the following multivariate autoregressive system such that it contains a directed chain as well as a fork, both of which can lead to spurious causal links:

$$\begin{aligned} x_{1,t} &= 3.4x_{1,t-1}(1 - x_{1,t-1}^2)e^{x_{1,t-1}^2} + c_{21}x_{2,t-4} \\ &\quad + c_{31}x_{3,t-2} + c_{41}x_{4,t-2} + 0.4u_{1,t} \\ x_{2,t} &= 3.4x_{2,t-1}(1 - x_{2,t-1}^2)e^{x_{2,t-1}^2} \\ &\quad + 0.4u_{2,t} \\ x_{3,t} &= 3.4x_{3,t-1}(1 - x_{3,t-1}^2)e^{x_{3,t-1}^2} \\ &\quad + c_{13}x_{1,t-1} + 0.4u_{3,t} \\ x_{4,t} &= 3.4x_{4,t-1}(1 - x_{4,t-1}^2)e^{x_{4,t-1}^2} \\ &\quad + c_{54}x_{5,t-3} + c_{64}x_{6,t-1} + 0.4u_{4,t} \\ x_{5,t} &= 3.4x_{5,t-1}(1 - x_{5,t-1}^2)e^{x_{5,t-1}^2} \\ &\quad + 0.4u_{5,t} \\ x_{6,t} &= 3.4x_{6,t-1}(1 - x_{6,t-1}^2)e^{x_{6,t-1}^2} + c_{76}x_{7,t-3} \\ &\quad + 0.4u_{6,t} \\ x_{7,t} &= 3.4x_{7,t-1}(1 - x_{7,t-1}^2)e^{x_{7,t-1}^2} + 0.4u_{7,t} \\ x_{8,t} &= 3.4x_{8,t-1}(1 - x_{8,t-1}^2)e^{x_{8,t-1}^2} + c_{78}x_{7,t-1} + 0.4u_{8,t} \\ x_{9,t} &= 3.4x_{9,t-1}(1 - x_{9,t-1}^2)e^{x_{9,t-1}^2} + c_{79}x_{7,t-1} + 0.4u_{9,t} \end{aligned} \tag{17}$$

with $c_{13} = 0.25$, $c_{21} = 2.5$, $c_{31} = 1.8$, $c_{41} = 1.5$, $c_{54} = 1.5$, $c_{64} = 1.2$, $c_{76} = 1.5$, $c_{79} = 1.8$, $c_{78} = 0.8$ and $u_{n,t}$ being zero mean Gaussian noise. The causal

structure of the system described above is shown in Fig. 4.

We vary the threshold δ from 0 to 0.5, and for each value of δ , we generate 50 realizations for the system described in Eq. (17). We varied the range of delays from 1 to 10 and the embedding dimension and delay were set to 3 and 100, respectively. We performed the simulations for generating data sets of a size of $T = 1000, 5000, 10000$ and 20000 samples.

Figure 5 shows the results from an exemplary simulation, where $\delta = 0.15, N = 10000$ samples and $NL = 0$. As we can see from the figure, the causal interactions among the stochastic processes are correctly identified along with their respective delays. In Fig. 5 the interaction $m \rightarrow n$ at a negative delay is again to be interpreted as $n \rightarrow m$. The causal interaction between two processes at a particular delay results in a drop in the CE value away from the H_{max} value, which in our case is given by $\log M! \approx 2.58$.

Figures 6 and 7 show the TPR and FPR values, respectively, for the interacting stochastic system, as δ, T and NL are varied. We can see that at $T = 1000$, high TPR as well as FPR values are obtained, with no significant changes in their values as NL or δ is varied. In case of the TPR values (see Fig. 6), we find that for $T \geq 5000$, the δ value at which TPR starts to drop below 1 slightly decreases as NL increases, but in all cases, $TPR = 1$ for $\delta < 0.2$. For $\delta > 0.3$, the TPR

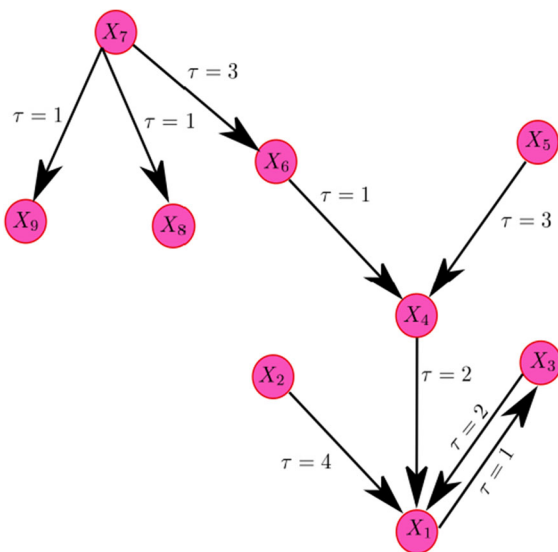


Fig. 4 Multivariate autoregressive system described in Eq. (17)

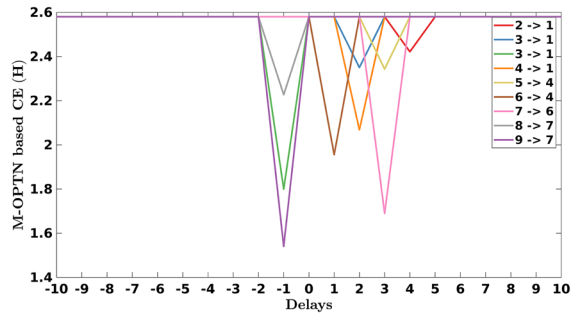


Fig. 5 Causality detection based on M-OPTNs for an exemplary simulation of the multivariate autoregressive system for $\delta = 0.15$ and $T = 10000$ samples

values tend to remain at ≈ 0.25 , except for $NL = 0.40$, where it continues to decrease for $\delta > 0.4$. When the $T \geq 5000$, we can see that the FPR values drop below 0.1 and for $\delta > 0.2, FPR = 0$ for $T = 5000$, and for $\delta > 0.05 FPR = 0$ for $T = 10000$ or $T = 20000$, irrespective of the value of NL (see Fig. 7).

Figure 8 shows the accuracy of the causal inference algorithm in terms of the F_1 -score for the interacting stochastic system. We observe that for $T = 1000, F_1 = 0$, irrespective of the choice of δ or the level of the noise. For $T = 5000$, we find that F_1 tends to increase as δ is increased and reaches ≈ 0.9 only when $\delta > 0.2$. F_1 tends to decrease for $\delta > 0.25$ for $NL < 0.40$. At $NL = 0.40$, for $T = 5000, F_1$ reaches a maximum value of 0.8 for $0.2 < \delta < 0.25$, after which it tends to decrease. For $T = 10000$ or 20000 , we again observe that F_1 tends to increase when δ is increased and it reaches ≈ 1 for the range $0.1 < \delta < 0.2$. The range of δ tends to get narrower only at $NL = 0.40$, when $F_1 \approx 1$ for $0.1 < \delta < 0.15$. F_1 starts to decrease as $\delta > 0.2$ for $NL < 0.4$ and $\delta > 0.15$ for $NL = 0.4$.

3.2 Interacting Lorenz systems

The next example of three interacting identical Lorenz systems with the structure $X_1 \rightarrow X_2 \rightarrow X_3$ is defined by the following set of ordinary differential equations,

$$\begin{aligned} \frac{dx_1(t)}{dt} &= 10(y_1(t) - x_1(t)) \\ \frac{dy_1(t)}{dt} &= 28x_1(t) - y_1(t) - x_1(t)z_1(t) \\ \frac{dz_1(t)}{dt} &= x_1(t)y_1(t) - 8/3z_1(t) \end{aligned}$$

Fig. 6 TPR computed for 50 independent realizations of the multivariate autoregressive system as δ , T , and NL are varied

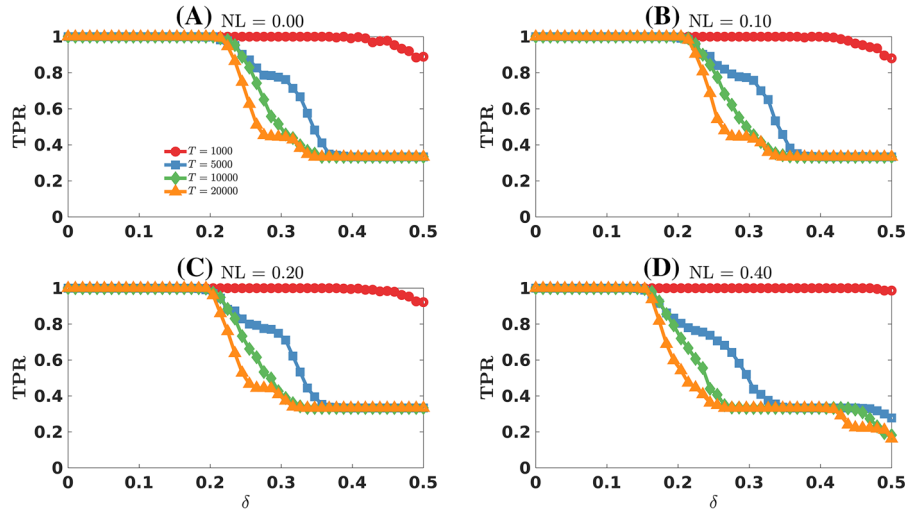


Fig. 7 FPR computed for 50 independent realizations of the multivariate autoregressive system as δ , T , and NL are varied

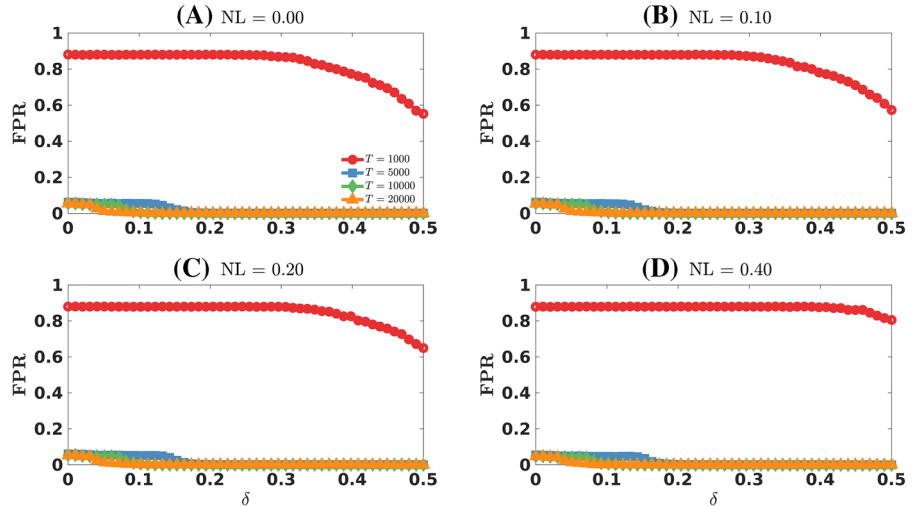
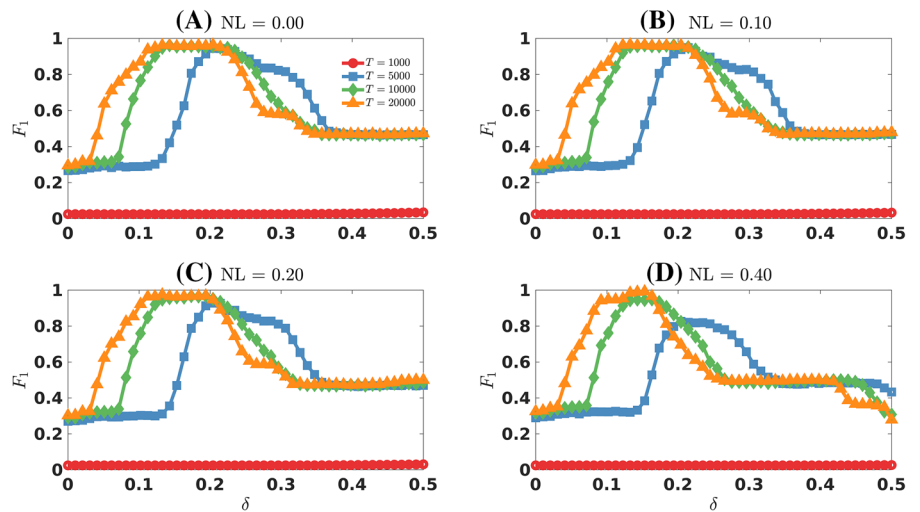


Fig. 8 F_1 computed for 50 independent realizations of the multivariate autoregressive system as δ , T , and NL are varied



$$\begin{aligned}
 \frac{dx_2(t)}{dt} &= 10(y_2(t) - x_2(t)) + c(x_1(t) - x_2(t)) \\
 \frac{dy_2(t)}{dt} &= 28x_2(t) - y_2(t) - x_2(t)z_2(t) \\
 \frac{dz_2(t)}{dt} &= x_2(t)y_2(t) - 8/3z_2(t) \\
 \frac{dx_3(t)}{dt} &= 10(y_3(t) - x_3(t)) + c(x_2(t) - x_3(t)) \\
 \frac{dy_3(t)}{dt} &= 28x_3(t) - y_3(t) - x_3(t)z_3(t) \\
 \frac{dz_3(t)}{dt} &= x_3(t)y_3(t) - 8/3z_3(t)
 \end{aligned} \tag{18}$$

We have set the coupling strength to $c = 0.6$ and used a Runge-Kutta integrator with a step size of $dt = 0.001$ to numerically solve the above set of ordinary differential equations. We then used the time series $\{x_k(t)\}_{t=1}^T$ with $k = 1, 2, 3$ as the observations from the interacting Lorenz systems, to which we added noise at $NL = 0.1, 0.2, 0.4$. We set $\lambda = 0.995$ and performed 50 simulations for every combination of the threshold δ , time series length T and noise level NL including the noise-free condition (i.e., $NL = 0$).

Figures 9 and 10 show the TPR and FPR values obtained for the interacting Lorenz systems as δ and T is varied for different values of NL . Here $NL = 0$ represents the noise-free condition. We can see that for $T = 1000$, the proposed causal inference algorithm gives high TPR (≈ 1) as well as FPR (≈ 0.60), which starts to drop for $\delta > 0.15$ and $\delta > 0.1$, respectively, for $NL = 0$. As NL is increased, the δ at which TPR and FPR start to drop, also increases. However, we observe that for all values of NL , for $T = 1000$, the accuracy of the algorithm as given by the F_1 score (see Fig. 11) is the highest at $\delta \approx 0.2$ for $NL = 0$ and for other values of NL , F_1 remains mostly at 0.5.

For $T = 5000$, the TPR remains at 1 for $NL = 0, 0.10$ and 0.20 and starts to drop for $\delta > 0.1$ (see Fig. 9). Only in the case of $NL = 0$, the FPR drops to ≈ 0.20 at $\delta \approx 0.08$ (see Fig. 9). For $NL = 0.10$ and 0.20 , FPR remains at 0.6 and only starts to drop when $\delta > 0.1$. For $NL = 0.40$, we observe low FPR (< 0.4) as well as TPR (< 0.5) values for all values of δ . Figure 11 shows that for $T = 5000$, at $NL = 0.0$ and $\delta \approx 0.15$, the F_1 -score is ≈ 0.78 and starts to drop to 0 and δ is increased. The F_1 -score remains mostly at about 0.5 and drops to 0 for other values NL as δ is increased.

From Fig. 9, for $T = 10000$ and 20000 , we observe that TPR remains at 1 and the δ value at which it starts to drop and eventually reaches zero decreases as NL increases. However, we find $TPR > 0.8$ even with the addition of noise at $NL = 0.1$ and 0.2 , for $\delta < 0.05$, for which we also observe that the FPR starts to drop (see Fig. 10). This is also reflected in the accuracy as given by the F_1 -score shown in Fig. 11, where $F_1 \approx 0.88$ for $NL = 0$ and $\delta = 0.1$, and $F_1 \approx 0.88$ for $NL = 0.1$ and $\delta \approx 0.05$. As NL is increased to 0.20 , F_1 -score drops to 0.6. At $NL = 0.40$, we observe that the accuracy is 0 for all values of δ , except for very small values (< 0.03), for which $F_1 \approx 0.4$.

3.3 Network of neural mass models

The simulations described in Sects. 3.1 and 3.2 have been restricted to linear stochastic systems and paradigmatic nonlinear dynamical systems, both of which may not fully characterize the typical nonlinear characteristics in neural time series. In order to also demonstrate the ability of the proposed method to capture interactions in nonlinear dynamical systems such as neuronal networks, we finally consider a network of neural mass models [41]. To this end, we created a network of eight neural mass models (the ordinary differential equations describing each neural mass model are provided in the Appendix, and the parameters are set as given in [41]), with the % of directed interactions (\mathcal{K}) between the eight regions varying as 5%, 10% and 25% of the overall possible connections (N^2 including N self-connections), at a delay of 40 milliseconds. We vary the threshold λ from 0.99 to 1.0 and the threshold δ from 0 to 0.2. To also investigate the effect of noise on the performance of the method – in addition to the noise free observations from neural mass models, Gaussian noise at a NL of 0.5 and 1.0 was added to the output of the neural mass models. For each of the combinations of these parameters (\mathcal{K} , λ , δ , and NL), we generated 25 simulations. The embedding parameters were the same as in Sect. 3.1 ($M = 3$ and $d = 1$). We computed CE based on M-OPTNs for delays ranging from 10 milliseconds to 100 milliseconds. Since the interaction in the simulated network happens at around 40 milliseconds, any interaction in the estimated networks at a delay other than 40 milliseconds is counted as a false positive.

Fig. 9 TPR of the proposed causal inference algorithm for the interacting Lorenz systems as δ , T , and NL are varied

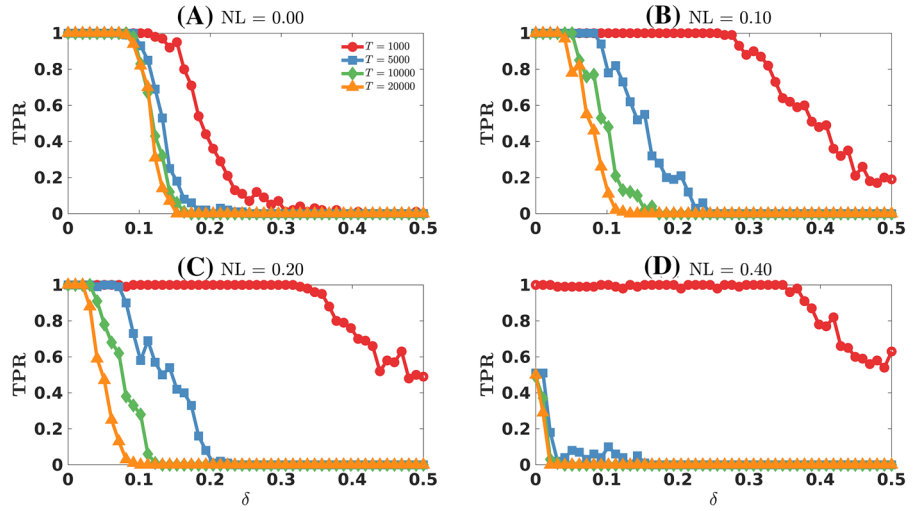


Fig. 10 FPR of the proposed causal inference algorithm for the interacting Lorenz systems as δ , T , and NL are varied

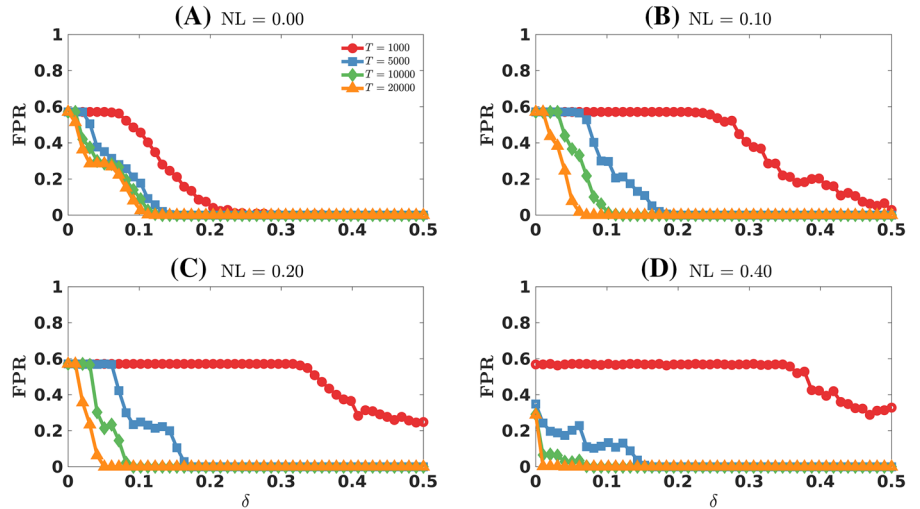


Fig. 11 F_1 -score of the proposed causal inference algorithm for the interacting Lorenz system as δ , T , and NL are varied

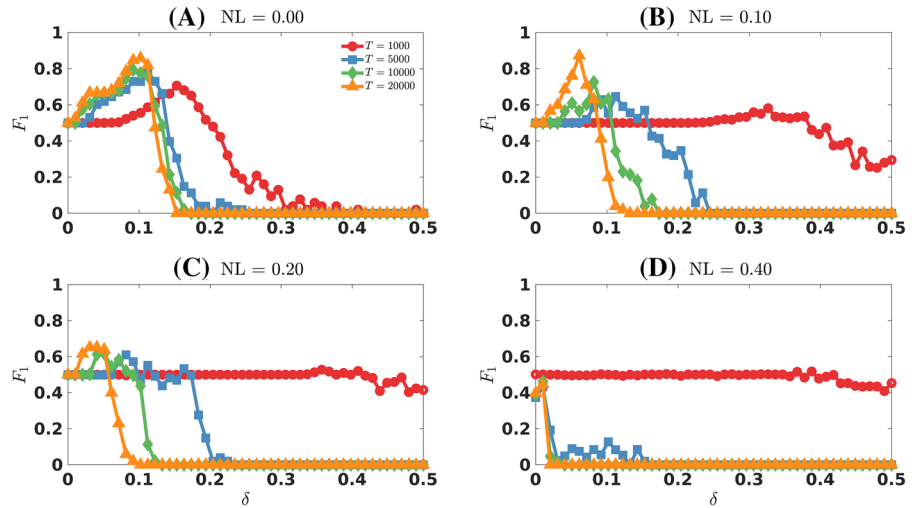


Figure 12 shows the TPR values for the network of neural mass models for varying NL , λ , δ and \mathcal{K} . We can see that for all values of \mathcal{K} and $NL = 0$ and 0.5 , the value of δ at which TPR remains at 1 decreases and TPR values are largely unaffected by the choice of λ . At $NL = 1.0$, we observe that as λ is increased, the range of δ for which $TPR = 1$ increases and at higher connection densities ($\mathcal{K} = 0.25$), we observe that high TPR values are only observed for a narrow range of $\lambda > 0.995$ and δ values.

Figure 13 shows the associated FPR values. We observe that for $\delta < 0.03$, FPR remains very high and close to 1 irrespective of the choice of δ , λ or \mathcal{K} when the $NL = 0$ or 0.5 . When $NL = 1.0$, high values of FPR are observed for $\lambda > 0.998$. When $\delta > 0.05$ and $NL = 0$ or 0.5 , the FPR values drop to zero irrespective of the choice of λ and for all \mathcal{K} . When NL is increased to 1.0 , we observe in general lower values of FPR as λ is decreased. However for $\delta > 0.05$, FPR drops to zero irrespective of the choice of λ or \mathcal{K} .

The accuracy of the proposed approach for the network of neural mass models is finally shown in terms of the F_1 -score in Fig. 14. As it is evident from the figure, as NL and \mathcal{K} increases, the choice of λ and δ for which we obtain high values for F_1 gets narrower. For example, at $NL = 0$ and $\mathcal{K} = 0.1$, for $0.1 < \delta < 0.2$, we can see that $F_1 \approx 1$ irrespective of the choice of λ .

But for the same \mathcal{K} , we observe that in order to get good accuracy, i.e., high F_1 -score, we need to choose lower values of λ and δ as NL is increased. When $\mathcal{K} = 0.25$, and $NL < 0.5$, we observe that $F_1 \approx 0.7$ for certain choices of λ and δ but when $NL = 1.0$, $F_1 < 0.5$.

4 Causality detection in MEA electrophysiology data

To validate our approach against real-world experimental data, we have applied the developed algorithm to electrophysiological recordings of epileptiform patterns generated by 4-aminopyridine (4AP)-treated rodent hippocampus-cortex (CTX) slices. In order to visualize the network activity, we transform the CE values obtained from two MEA signals i and j , $H(i|j)$ as

$$S(i, j) = 2.58 - H(i, j) \tag{19}$$

and normalize them such that the strongest pairwise interaction takes a value of 1, i.e., $\tilde{S} = S(i, j) / \max(S)$.

4.1 Brain slice preparation and maintenance

Combined hippocampus-cortex (CTX) brain slices ($n = 4$), $400 \mu\text{m}$ thick, were prepared from four male

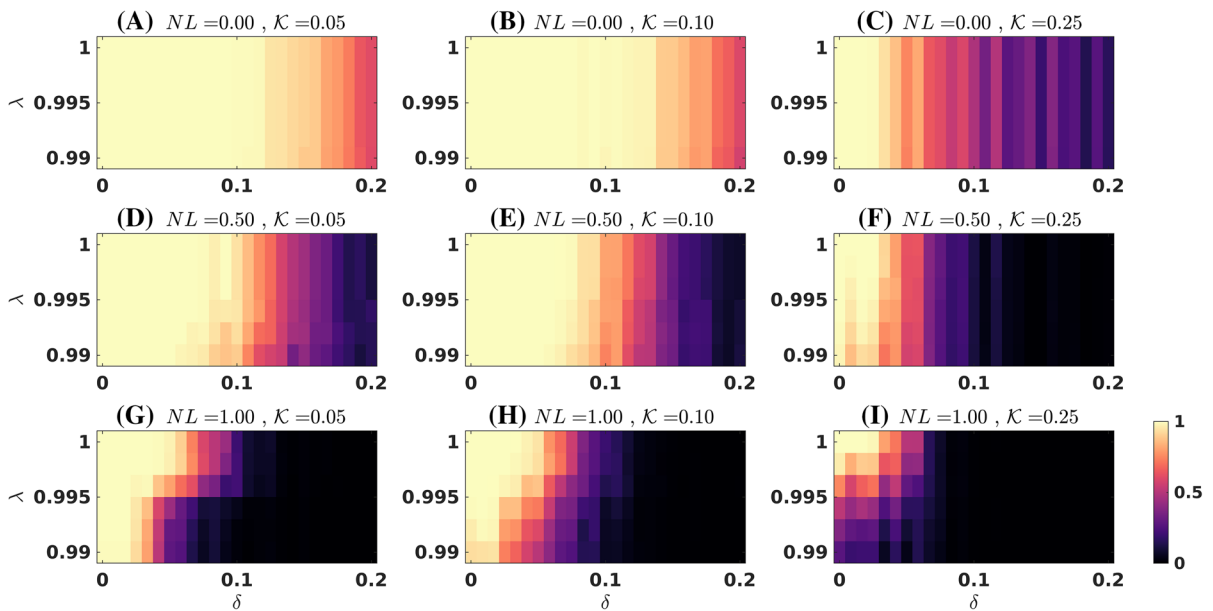


Fig. 12 TPR computed for 50 independent realizations of networks of neural mass models for every combination of \mathcal{K} , λ , δ , and NL

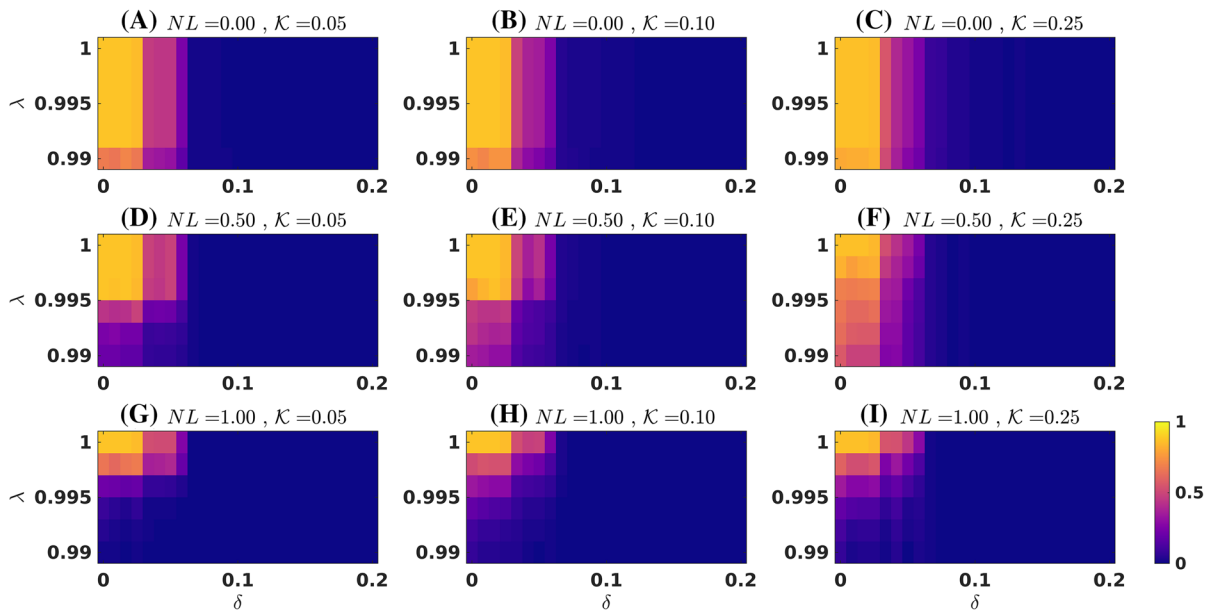


Fig. 13 FPR computed for 50 independent realizations of networks of neural mass models for every combination of κ , λ , δ , and NL

CD1 mice (4–8 weeks old), as previously described [42]. Briefly, animals were euthanized under deep isoflurane anesthesia, their brain was quickly removed and placed into ice-cold ($\approx 2^\circ\text{C}$) sucrose-based artificial cerebrospinal fluid (sucrose-ACSF) composed of (mM): Sucrose 208, KCl_2 , KH_2PO_4 1.25, MgCl_2 5, MgSO_4 , CaCl_2 0.5, D-glucose 10, NaHCO_3 26, L-Ascorbic Acid 1, Pyruvic Acid 3. The brain was let to chill for 2 min before slicing in ice-cold sucrose-ACSF using a vibratome (Leica VT1000S, Leica, Germany). Brain slices were immediately transferred to a submerged holding chamber containing room-temperature holding ACSF composed of (mM): NaCl 115, KCl_2 , KH_2PO_4 , 1.25, MgSO_4 1.3, CaCl_2 2, D-glucose 25, NaHCO_3 26, L-Ascorbic Acid 1. After at least 60 minutes recovery, individual slices were transferred to a submerged incubating chamber containing warm ($\approx 32^\circ\text{C}$) holding ACSF for 20–30 minutes and subsequently incubated in warm ACSF containing the K^+ channel blocker 4-aminopyridine (4AP, $250\ \mu\text{M}$), in which MgSO_4 concentration was lowered to 1 mM (4AP-ACSF, [42]). Brain slice treatment with 4AP is known to enhance both excitatory and inhibitory neurotransmission and induces the acute generation of epileptiform discharges [43]. All brain slices were incubated in 4AP-ACSF for at least 1 hour before beginning any recording session. All solutions were con-

stantly equilibrated at $\text{pH} \approx 7.35$ with 95% O_2 / 5% CO_2 gas mixture (carbogen) and had an osmolality of 300–305 mOsm/kg. Chemicals were acquired from Sigma-Aldrich. All procedures have been approved by the Institutional Animal Welfare Body and by the Italian Ministry of Health (authorization 176AA.NTN9), in accordance with the National Legislation (D.Lgs. 26/2014) and the European Directive 2010/63/EU. All efforts were made to minimize the number of animals used and their suffering.

4.1.1 MEA recording and signal pre-processing

Individual brain slices were placed on a 6×10 planar MEA (TiN electrodes, diameter $30\ \mu\text{m}$, inter-electrode distance $500\ \mu\text{m}$, impedance $< 100\ \text{k}\Omega$), held in place by a custom-made anchor, and continuously perfused at $\approx 1\ \text{ml/min}$ with 4AP-ACSF at ($\approx 32^\circ\text{C}$), equilibrated with carbogen gas mixture. To allow for laminar flow and a high exchange rate of the 4AP-ACSF, a custom-made low-volume ($\approx 500\ \mu\text{l}$) recording chamber (Crisel Instruments, Italy) replaced the default MEA ring [42].

Extracellular field potentials were acquired at 5 kHz (pre-sampling low-pass filter at 2 kHz) using the MEA2100-mini-HS60 system through the Multichannel Experimenter software (all from Multichannel Sys-

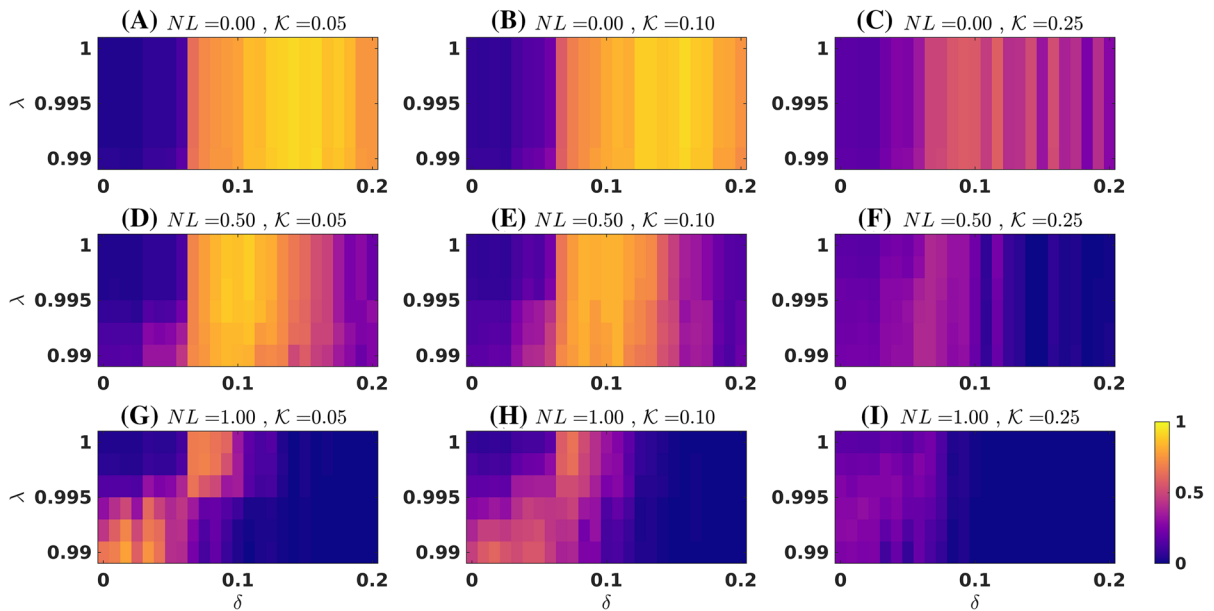


Fig. 14 F_1 -score computed for 50 independent realizations of networks of neural mass models for every combination of κ , λ , δ , and NL

tems – MCS, Reutlingen, Germany) and stored on the hard drive for off-line analysis.

For the implementation of the M-OPTN, signals were pre-processed by low-pass filtering (1 kHz) and resampling at 3 kHz. The signals were then further divided into overlapping (50%), 4 s windows, to obtain time-varying measures of causality.

4.1.2 Epileptiform activity generated by 4AP-treated hippocampus-CTX slices

The brain slice preparation used in this work includes the fundamental circuits involved in the generation of limbic seizures seen in temporal lobe epilepsy and enables analyzing the network interactions leading to seizure-like discharge generation. As shown in Fig. 15, the key regions of interest (ROIs) in this brain slice preparation are the dentate gyrus (DG), the hippocampal subfields Cornu Ammonis 3 and 1 (CA3 and CA1, respectively), the subiculum (SUB) and the parahippocampal cortex (CTX-1 and CTX-2). These regions communicate through the so-called hippocampal loop [44] (see Fig. 15 (A)). When challenged with convulsant drugs, such as 4AP, hippocampus-CTX slices generate a typical epileptiform pattern made of three types of activity [45]: (1) slow interictal events, recurring at 5–20 s interval, generated by and spreading to any ROI

with no specific site of origin, (2) fast interictal events, recurring at 0.5–2 s interval, generated specifically by the CA3, propagating to the CA1 via the Schaffer Collaterals and subsequently reaching the CTX through the SUB (output gate), (3) ictal (seizure-like) discharges, recurring at 3–5 min interval, originating primarily in the CTX and spreading to the hippocampus proper via the DG (input filter). It has been previously demonstrated that when the hippocampal loop circuitry is intact (connected brain slice), the fast CA3-driven interictal activity controls ictal discharge generation by the CTX, for which ictal discharges disappear within 1–2 hours of 4AP application, while only the interictal patterns remain. At variance, the disruption of the hippocampal loop upon Schaffer Collaterals damage, as seen in hippocampal sclerosis typical of temporal lobe epilepsy (disconnected brain slice) releases the CTX from the CA3 control permitting ictal activity generation and propagation [46].

Here, we have analyzed MEA recordings of epileptiform activity generated by disconnected hippocampus-CTX brain slices, in which the Schaffer Collaterals were mechanically severed.

The circuit diagram of a disconnected hippocampus-CTX slice is depicted in Fig. 15 (a) and 15 (b) shows a disconnected hippocampus-CTX slice placed on a

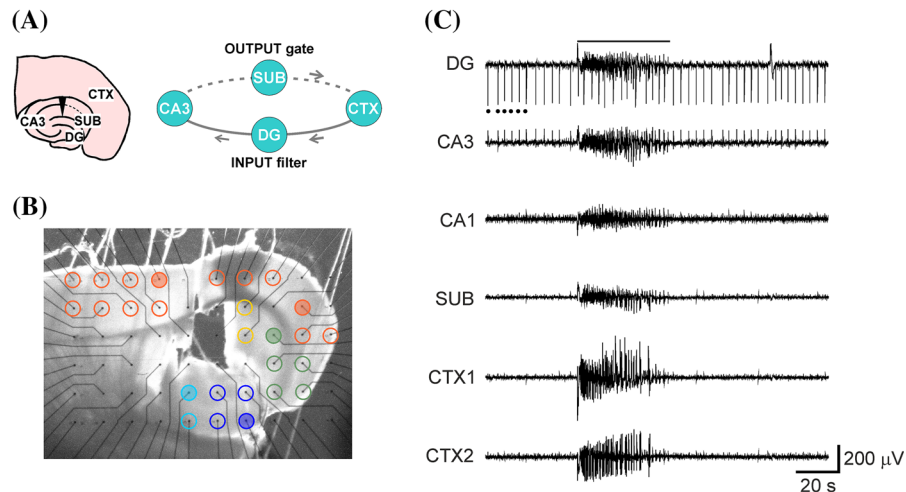


Fig. 15 **a** Schematic rendition of a disconnected hippocampus-CTX slice and its corresponding circuit. The solid line indicates the preserved pathway, whereas the dashed line indicates the disrupted pathway. Arrows indicate the signal propagation along the loop. **b** Hippocampus-CTX slice laying on a 6×10 planar MEA. MEA electrodes are placed in DG (blue), SUB (green),

CA1 (yellow), CA3 (cyan) and CTX (orange). **c** MEA recording of the epileptiform pattern generated by the brain slice in B. The pattern consists of short interictal events (dots) and prolonged ictal discharges (solid line). Note that the small-amplitude events in CA1, SUB and CTX are far fields originating in CA3

6×10 planar MEA, while Fig. 15 (c) shows the typical epileptiform pattern induced by 4AP in this brain slice preparation, consisting of brief interictal events and prolonged ictal discharges.

For the purpose of this study, we have selected the signals from six electrodes in each brain slice, to include each of the four hippocampal regions and two CTX locations, one proximal and one distal to the hippocampus with regards to the signal propagation pathway (see Fig. 15 (b)). For each MEA recording, we have selected a portion of the signal to include ictal activity preceded and followed by 30 – 100 s of interictal activity.

The CE based on M-OPTN was computed for each window and the resulting CE value was assigned to the mid-point of each window, to obtain a time-varying measure.

4.2 Results

The results from the application of our method are shown in Fig. 16, where the plots on the main diagonal show the MEA signals acquired from the six selected ROIs. The off-diagonal plots represent the time-varying interaction between two ROIs across varying time

delays (20 ms to 120 ms). From Fig. 16, it can be seen that significant network activity starts around the time of the ictal onset, with the strongest connections following the propagation paths CTX \rightarrow DG, CTX \rightarrow CA1, CTX \rightarrow SUB, SUB \rightarrow CTX and DG \rightarrow CA3. The strongest interactions are observed at 38 and 70 ms.

The results from the entire data sets are qualitatively similar (see Figs. 17, 18 and 19 in Sect. 7 for the results obtained from the other three brain slices). The propagation from SUB \rightarrow DG was also observed in three out of four slices. The interaction SUB \rightarrow CTX was observed in all slices although the strength of interaction from SUB \rightarrow CTX was found to be generally weaker compared to the interaction CTX \rightarrow SUB. In general, the results show that outward connections from CTX and DG are generally the strongest during the ictal event and that these interactions appear to be strongest at a delay of ≈ 38 ms.

5 Discussion

In this work, we have proposed a new method to detect causality from multivariate observational data by computing information theoretic measures such as CE upon

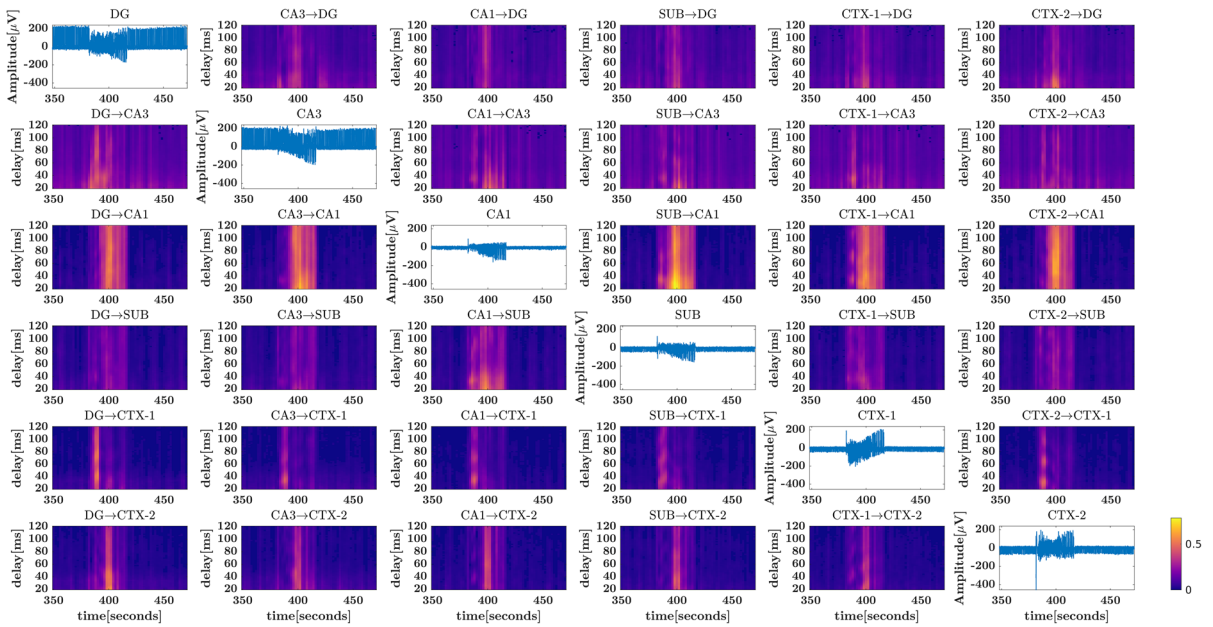


Fig. 16 Causality detection based on M-OPTNs for MEA recording of a representative brain slice. The subplots on the diagonal show the MEA signal (ictal activity preceded and followed by interictal events) recorded by the electrodes placed in each

of the six ROIs. Each off-diagonal subplot represents the time-varying interaction as given by CE based on M-OPTNs between two ROIs for varying time delays. The color bar indicates the connection strength

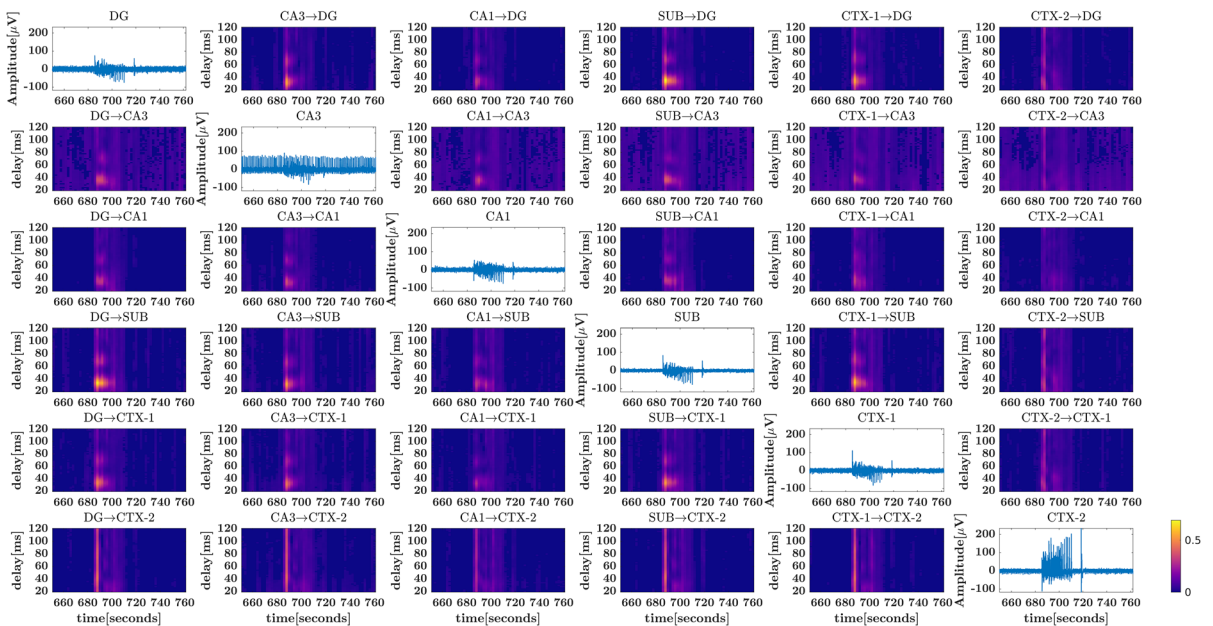


Fig. 17 Causality detection based on M-OPTNs for MEA data from slice no. 2. Each off-diagonal subplot represents the time-varying interaction as given by CE based on M-OPTNs between two ROIs for varying time delays. The subplots on the diagonal

show the MEA signal (including the ictal activity preceded and followed by inter-ictal activity) in the electrodes placed in each of the six ROIs. The color bar indicates the connection strength

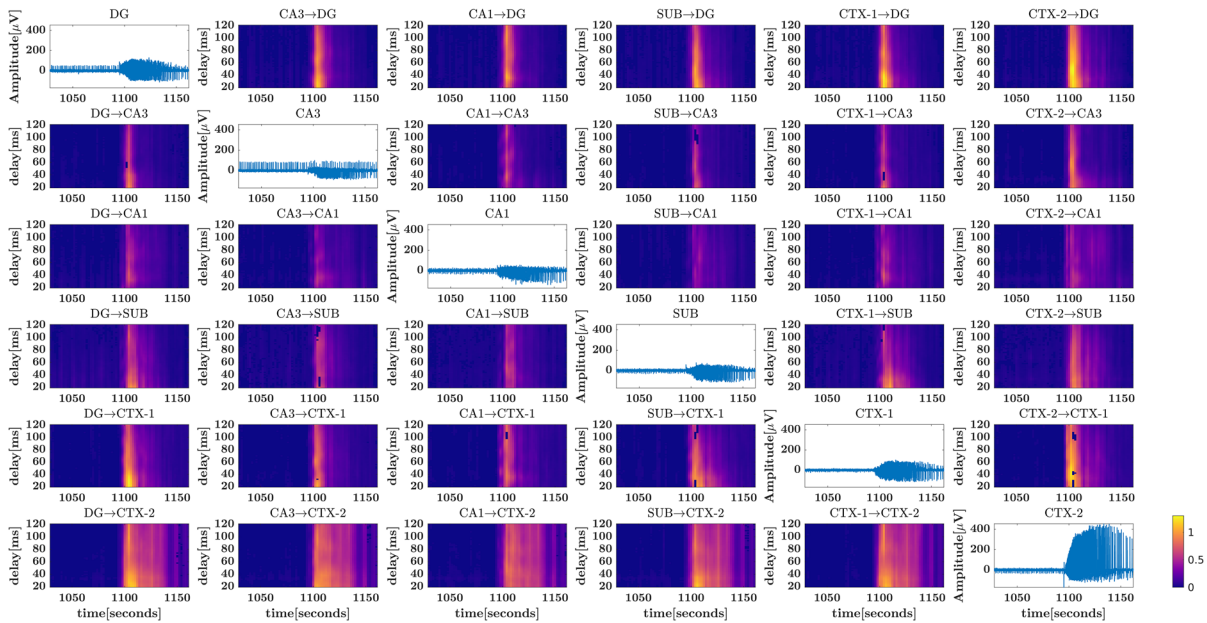


Fig. 18 Causality detection based on M-OPTNs for MEA data from slice no. 3. Each off-diagonal subplot represents the time-varying interaction as given by CE based on M-OPTNs between two ROIs for varying time delays. The subplots on the diagonal

show the MEA signal (including the ictal activity preceded and followed by inter-ictal activity) in the electrodes placed in each of the six ROIs. The color bar indicates the connection strength

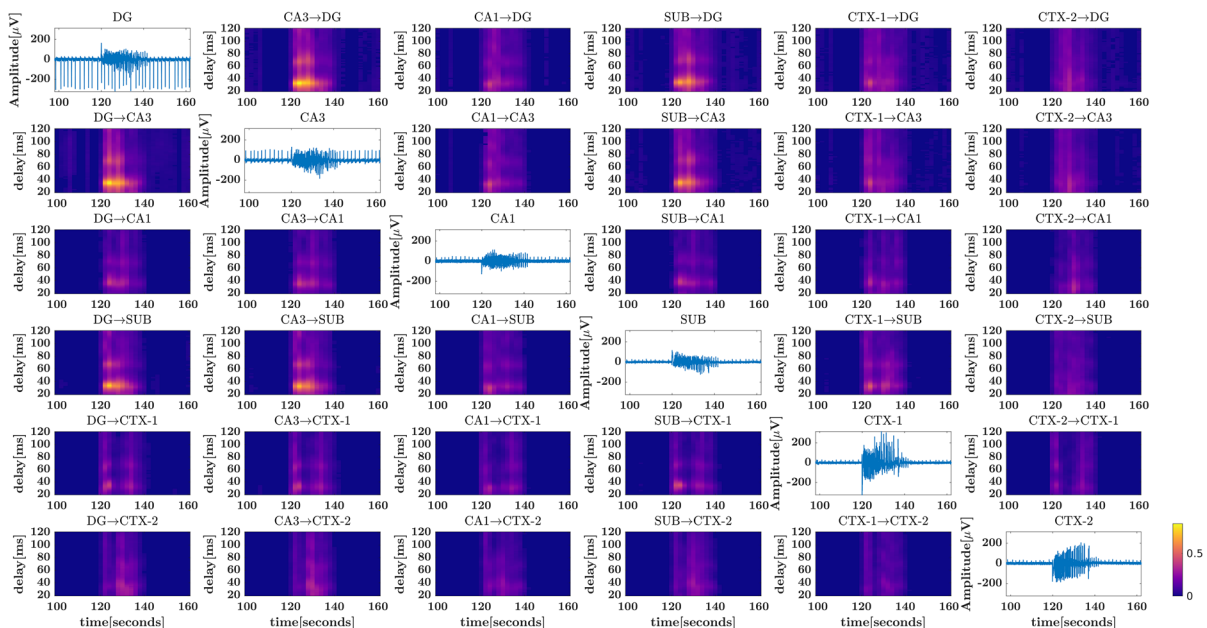


Fig. 19 Causality detection based on M-OPTNs for MEA data from slice no. 4. Each off-diagonal subplot represents the time-varying interaction as given by CE based on M-OPTNs between two ROIs for varying time delays. The subplots on the diagonal

show the MEA signal (including the ictal activity preceded and followed by inter-ictal activity) in the electrodes placed in each of the six ROIs. The color bar indicates the connection strength

the resulting M-OPTNs. For reliable computation of CE and removal of non-causal neighbors, we have also proposed a pragmatic methodology to define a minimal set for conditioning variables. Our numerical experiments show that our approach can be used to reliably infer both the directionality and the delay of the interactions between the signals even at considerably high NL . Causal network inference from real-world data of MEA recordings demonstrates that the application of the proposed method can infer network interactions during ictal activity and their associated delays.

5.1 Minimal set of neighbors for conditioning

In order to test if a node (signal) m has a causal influence on node (signal) n , the standard and most common approach is to use the Peter and Clarke algorithm [47], which tests the conditional independence between two nodes given all other variables. Non-existence of a causal relationship between m and n is established once the algorithm finds that m and n are conditionally independent given other variables. When inferring causality on multivariate time series, the conditioning set could have many variables resulting in unreliable estimates of information theoretic measures, particularly when the sample size is small. We alleviate this problem by defining a selection of variables to condition on, based on the common information shared between them. Furthermore, we restrict the number of variables to condition on, $r \leq 3$, according to the finite data size ($T \approx 10000$ samples), since conditioning on a higher number of variables resulted in unreliable estimates of CE. In addition to the Peter and Clarke algorithm, there are other iterative, constraint-based approaches that have been proposed for conditioning, including the modified Peter and Clarke algorithm [47] and fast causal inference algorithm [48] as well as score-based algorithms such as greedy equivalence search [49] for defining conditioning sets. However, most of these algorithms suffer from undesirable computational complexity and are not straightforward to implement. A systematic comparison of various conditioning approaches is beyond the scope of this work, where the motivation is to propose and demonstrate M-OPTNs as an extension of OPTNs to reliably infer causality among time series.

5.2 Effect of various parameters

The numerical results obtained from simulations of coupled stochastic processes and interacting Lorenz systems, have shown that the proposed approach can successfully capture coupling directions and the associated delays. When applied to more realistic simulations using a network of neural mass models, our approach could reliably recover the underlying causal coupling structure. However, as discussed in detail in the following sections, the performance of the proposed method depends on the choice of several parameters, which can overall be categorized into (1) number of time samples and (2) threshold for identifying significant connections.

5.2.1 Effect of varying the number of time samples

It is well known that the amount of data required for reliable reconstruction of the attractor depends on the embedding dimension M . Since the probability distributions required for the computation of CE are estimated from the ordinal patterns obtained after embedding, inadequate data length might result in unreliable estimates of CE. In our simulations we have set $M = 3$ and found that the results are reliable if $T \gtrsim 10^3$. For a given M , there are $(M!)^2$ possible pairs of ordinal patterns for which we have to estimate the co-occurrence frequencies. Having $T \leq 10^M$ samples, results in many spurious interactions being classified as causal links, which is reflected as an increase in the FPR and consequently low F_1 -score as shown in our simulations. For the stochastic model system, interacting Lorenz systems and the network of neural mass models, we used $M = 3$. Embedding in a higher dimension, for example $M = 5$ would require $T > 100000$ samples for reliable embedding and computation of entropy values. The typical sampling frequency of real-world data such as MEA recordings is of the order of 5000–10000 Hz, and to estimate dynamic changes in a causal effect network based on 100000 samples would mean using a window size of 10 to 20 seconds, which may be far too long compared to dynamical changes that occur in neural networks. Also, the use of 100000 or more samples, increases the computation time drastically. Thus, for electrophysiological recordings from neural data, a window size of 2 to 4 seconds seems more realistic, which amounts to having 10000 to 20000 samples per window, depending on the sampling frequency. This

in turn means that M should not be greater than 3 or at most 4. In contrast, we observed that varying the embedding delay d did not affect the results qualitatively (not shown here) and we used $d = 100$ for all our simulations and experimental data.

5.2.2 Effect of varying threshold parameters to define significant connections

The parameter λ determines the connections (direct and indirect) to be classified as significant. Lower values of λ prune away most of the connections, whereas higher values retain most of the connections. Based on our simulations of network of neural mass models, we observed that when the network connectivity is less than 25% and only a moderate amount of observation noise ($NL = 0.50$) is present, the choice of $\lambda \geq 0.99$ seems to be the optimal setting that results in high TPR and low FPR and, consequently, high F_1 -score, provided that δ is chosen in the range $0.08 \leq \delta \leq 0.12$. However when the NL increases ($NL > 0.5$), a setting of $\lambda > 0.995$ and $0.08 < \delta < 0.1$ leads to an optimal performance of our algorithm.

The aforesaid implies that the choice of δ , which determines the threshold to distinguish a causal neighbor from a non-causal neighbor, depends on the amount of noise present in the data. In the presence of low or moderate observational noise, a truly causal neighbor would result in a high δ as conditioning on this neighbor should reduce the entropy significantly. In contrast, a non-causal neighbor would result in a very small δ . Our results show that the setting $\delta \leq 0.1$ in such a scenario is a reasonable choice along with $\lambda \approx 0.995$. If the data is very noisy ($NL \geq 1$), then the necessary δ for identifying a truly causal neighbor would be very small, thereby making it hard to distinguish from a non-causal neighbour, for which δ should also be small. Thus setting δ too high will prune away all the true connections along with the spurious ones, while setting δ too low might retain some spurious connections.

Another factor that impacts the choice of λ and δ in addition to NL is the number of connections in the network. When the network is densely (in case of our simulations, more than 25%) connected, finding an optimal δ and λ that gives high TPR and low FPR , and consequently a high F_1 -score is more challenging as the estimated network has many spurious connections at multiple delays in addition to the interactions at the correct delays. Any choice of δ to prune away these

spurious connections will also yield the removal of true connections as the NL increases.

In summary, our results indicate that setting $\lambda \approx 0.995$ and varying δ between 0.05 and 0.1 should result in reliable network inference, assuming the underlying networks are sparse (see Figs. 12 – 14).

5.3 Causal network inference from MEA data

As a real-world example, we have applied the proposed method to MEA electrophysiology recordings obtained from an in vitro model of acute limbic seizures. Specifically, we have used 4AP-treated rodent hippocampus-CTX slices as a simplified model of the primary neural circuits involved in temporal lobe epilepsy. This model has been extensively characterized [46] and provides a solid ground-truth to validate our approach.

In a previous study [46], the reported mean time delay for the ictal discharge propagation in the directions $CTX \rightarrow DG$, $CTX \rightarrow CA3$ and $CTX \rightarrow CA1$ was 37.5 ± 9.6 ms, 71.7 ± 27.5 and 31 ± 6.3 ms. In keeping with this, our method has detected connectivity in the direction of $CTX \rightarrow DG$, $CTX \rightarrow CA3$ and $CTX \rightarrow CA1$ at delays of 30 – 38 ms during the ictal discharge. Note that the short delay in the $CTX \rightarrow CA1$ direction is due to the signal propagation along the direct temporoammonic pathway [46], which is known to short-circuit the hippocampal loop. Moreover, the connection between SUB and CTX is consistent with the previously reported role of SUB-CTX interactions through the temporoammonic pathway in reinforcing ictal synchronization in animal models of temporal lobe epilepsy [50].

We also found connections in the direction $DG \rightarrow CTX$, $CA3 \rightarrow CTX$, $CA3 \rightarrow SUB$ and $DG \rightarrow SUB$. However, as these connections are disrupted by the Schaffer Collaterals cut, they represent false positives due to far-field contamination [51] of the signals recorded from CA1, SUB and CTX, wherein far fields originating in CA3 can be seen in Fig. 15. In keeping with this, we found that the strength of the false positive connection is generally lower than the expected true connections. The observation of such false positive interactions could also stem from the common driver issue (see Fig. 1 (b)) wherein one of the CTX ROI is driving both DG and CTX

(in the other ROI) causing a spurious $DG \rightarrow CTX$ connection.

Overall, these results support the reliability and usefulness of our approach for analyzing interactions and their delays in real-world observations, such as electrophysiological time series.

5.4 Future work and perspectives

Although the proposed algorithm can reliably perform causal inference, there are certain issues that warrant our attention. First, we have not compared our method to other existing techniques that are commonly used to infer causality from electrophysiological recordings. In fact, the main motivation behind this work was to introduce and provide a proof-of-concept that complex network based time series analysis methods such as OPTNs can be generalized and improved to detect causality from multivariate observations. A systematic comparison of our approach with other commonly used techniques to infer causality is beyond the scope of this paper and will remain a subject of future studies.

Second, we applied our method to *in vitro* MEA electrophysiology data, which is not as widely used to map connections among brain regions as *in vivo* recordings. In the case of EEG data we typically do not have the ground-truth to validate our method against. In the case of the MEA data used in this study, previous studies have described the anatomical and functional circuits associated within this brain slice preparation, which served as the reference for our results. Furthermore, estimating causality directly from EEG recordings is not trivial due to the issue of volume conduction. In the case of EEG recordings, to mitigate the volume conduction effects, connectivity is estimated from source time-series obtained after solving the EEG inverse problem [52]. Thus, the proposed causal inference method has to be applied on the inverse solution, rather than directly to the EEG data, for reliable causal inference. It is not yet clear how such a transformation would alter the structural properties of the time series, and which impact it could have on the estimation of the ordinal patterns remains as a subject of future studies.

Third, we did not perform any surrogate data testing but rather relied on the theoretical maximum $H_{max} =$

$\log_2 M!$ and used λH_{max} as threshold for identifying significant bivariate connections, where $0.99 \leq \lambda < 1$, as due to the finite sample-size, two independent processes will not have a CE value exactly equal to H_{max} . We found this approach to be much faster than generating bivariate surrogates that gave essentially similar results (not shown). In STEP VI of the proposed algorithm, we use the threshold δ to distinguish between causal and non-causal neighbors. This step can be considerably improved by performing significance testing for the conditional independence test as proposed in [1], which preserves, for example, the association between X and Y in the coupling scheme $X \leftarrow Z \rightarrow Y$, that would otherwise be destroyed in a strictly bivariate permutation scheme.

6 Conclusions

In this paper, we have developed a new method based on OPTNs to infer causality from multivariate observational data. The proposed method allows to infer causality at different delays and can be adapted to provide a time-varying measure of causality. We have also proposed an iterative scheme to find a minimal set of neighbors for conditioning to yield a reliable estimation of the co-occurrence entropy as the employed coupling indicator. We have demonstrated the validity of our approach using different types of simulated signals as well as real-world electrophysiological time series.

In conclusion, the proposed approach provides a complementary tool for detecting causality from multivariate time series data and can be particularly useful in the area of neuroscience, where the estimation of (time-varying) causal networks from electrophysiology recordings has remained a fundamental problem so far.

Acknowledgements This project has received funding from the European Union's Horizon 2020 research and innovation programme FETPROACT-01-2018 (RIA) awarded to the project Hybrid Enhanced Regenerative Medicine Systems (HERMES) under grant agreement No 824164

Data availability The implementation of the causal inference algorithm and the datasets generated and an example MEA dataset analysed during the current study are available at https://github.com/narayanps/causal_inference_with_OPTNs. All the

MEA datasets are available on Zenodo (DOI : 10.5281/zenodo.4892673).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

7 Appendix

In the neural mass model studied in Sect. 3.3, the neuronal activity in one region is represented by the following set of ordinary differential equations:

Pyramidal neurons

$$\frac{dy_p(t)}{dt} = x_p(t), \quad (20)$$

$$\frac{x_p(t)}{dt} = G_e h_e z_p(t) - 2h_e x_p(t) - h_e^2 y_p(t) \quad (21)$$

$$z_p(t) = \frac{2e_0}{1 + e^{-rv_p}} - e_0 \quad (22)$$

$$v_p(t) = C_{pe} y_e(t) - C_{ps} y_s(t) - C_{pf} y_f(t) \quad (23)$$

Excitatory interneurons

$$\frac{dy_e(t)}{dt} = x_e(t), \quad (24)$$

$$\frac{x_e(t)}{dt} = G_e h_e \left(z_e(t) + \frac{u_p(t)}{C_{pe}} \right) - 2h_e x_e(t) - h_e^2 y_e(t) \quad (25)$$

$$z_e(t) = \frac{2e_0}{1 + e^{-rv_e}} - e_0 \quad (26)$$

$$v_e(t) = C_{ep} y_p(t) \quad (27)$$

Slow inhibitory interneurons

$$\frac{dy_s(t)}{dt} = x_s(t), \quad (28)$$

$$\frac{x_s(t)}{dt} = G_s h_s z_s(t) - 2h_s x_s(t) - h_s^2 y_s(t) \quad (29)$$

$$z_s(t) = \frac{2e_0}{1 + e^{-rv_s}} - e_0 \quad (30)$$

$$v_s(t) = C_{sp} y_p(t) \quad (31)$$

Fast inhibitory interneurons

$$\frac{dy_f(t)}{dt} = x_f(t), \quad (32)$$

$$\frac{x_f(t)}{dt} = G_f h_f z_f(t) - 2h_f x_f(t) - h_f^2 y_f(t), \quad (33)$$

$$\frac{dy_f(t)}{dt} = x_l(t), \quad (34)$$

$$\frac{x_l(t)}{dt} = G_e h_e u_f(t) - 2h_e x_l(t) - h_e^2 y_l(t), \quad (35)$$

$$z_f(t) = \frac{2e_0}{1 + e^{-rv_f}} - e_0 \quad (36)$$

$$v_f(t) = C_{fp} y_p(t) - C_{fs} y_s(t) - C_{ff} y_l(t) \quad (37)$$

A network of neural mass models can be constructed by connecting several such regions using a weight matrix W that describes the strength of connections. For example, if i and j represent two regions of neuronal population, then we can define

$$u_p^i(t) = n_p^i(t) + W_p^{ij} z_p^j(t - d), \quad (38)$$

where $u_p(t)$ and $z_p(t)$ correspond to the input and pulse density of the pyramidal neurons, respectively. Analogous definitions apply to the fast inhibitory interneurons. The term $n_p(t)$ represents Gaussian noise with mean $m = 0$ and variance $\sigma^2 = 5$ and d represents the connection delay. For further description on the model, the reader is kindly referred to [41].

References

1. Runge, J.: Causal network reconstruction from time series: from theoretical assumptions to practical estimation. *Chaos* **28**, 075310 (2018)
2. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., Bhatlacharya, J.: Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **441**, 1–46 (2007)

3. Sun, J., Bolt, E.M.: Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D* **267**, 49–57 (2014)
4. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969)
5. Baccala, L.A., Sameshima, K.: Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* **84**, 463–474 (2001)
6. Montalto, A., Faes, L., Marinazzo, D.: MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLoS One* **9**, e109462 (2014)
7. Schreiber, T.: Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000)
8. Vastano, J.A., Swinney, H.L.: Information transport in spatiotemporal systems. *Phys. Rev. Lett.* **60**, 1773 (1988)
9. Li, S., Xiao, Y., Zhou, D., Cai, D.: Causal inference in nonlinear systems: granger causality versus time-delayed mutual information. *Phys. Rev. E* **97**, 052216 (2018)
10. Barnett, L., Barrett, A.B., Seth, A.K.: Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009)
11. Friston, K.J.: Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapp.* **2**, 56–78 (1994)
12. Lachaux, J.P., Rodriguez, E., Martinerie, J., Francisco, F.J.: Measuring phase synchrony in brain signals. *Human Brain Mapp.* **8**, 194–208 (1999)
13. Mormann, F., Lehnertz, K., David, P., Elger, C.E.: Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D* **144**, 358–359 (2000)
14. Srinivasan, R., Nunez, P.L., Silberstein, R.B.: Spatial filtering and neocortical dynamics: estimates of EEG coherence. *IEEE Trans. Biomed. Engin.* **45**, 814–826 (1998)
15. Jeong, J., Gore, J.C., Peterson, B.S.: Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin. Neurophysiol.* **112**, 827–835 (2001)
16. Feldt, S., Bonifazi, P., Cossart, R.: Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends Neurosci.* **34**(5), 225–236 (2011)
17. de Abril, I.M., Yoshimoto, J., Doya, K.: Connectivity inference from neural recording data: challenges, mathematical bases and research directions. *Neural Netw.* **102**, 120–137 (2018)
18. Vicente, R., Wibral, M., Linder, M., Pipa, G.: Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**, 45–67 (2011)
19. Subramaniyam, N.P., Tronarp, F., Särkkä, S., Parkkonen, L.: Expectation-maximization algorithm with a nonlinear Kalman smoother for MEG/EEG connectivity estimation. *EMBE and NBC, IFMBE Proceedings*, 65 (2017)
20. Tronarp, F., Subramaniyam, N.P., Särkkä, S., Parkkonen, L., Subramaniyam, N.P.: Tracking of dynamic functional connectivity from MEG data with Kalman filtering. 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2018)
21. Subramaniyam, N.P., Tronarp, F., Särkkä, S., Parkkonen, L.: Joint estimation of neural sources and their functional connections from MEG data. *bioRxiv preprint* ; <https://www.biorxiv.org/content/early/2020/10/05/2020.10.04.325563> (2020)
22. Friston, K.J., Harrison, L., Penny, W.: Dynamic causal modelling. *Neuroimage* **19**, 1273–1302 (2003)
23. Ding, M., Chen, Y., Bressler, S.L.: Granger causality: basic theory and application to neuroscience. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications* **437–460**, (2006)
24. Bastod, A.M., Schoffelen, J.M.: A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* **9**, 175 (2016)
25. Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J.: Complex network approaches to nonlinear time series analysis. *Phys. Rep.* **787**, 1–97 (2019)
26. Donner, R.V., Zou, Y., Donges, J.F., Marwan, N., Kurths, J.: Recurrence networks—a novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**, 033025 (2010)
27. Subramaniyam, N.P., Hyttinen, J.: Characterization of dynamical systems under noise using recurrence networks: application to simulated and EEG data. *Phys. Lett. A* **378**, 3464–3474 (2014)
28. Subramaniyam, N.P., Donges, J.F., Hyttinen, J.: Signatures of chaotic and stochastic dynamics uncovered with ϵ -recurrence networks. *Proc. R. Soc. A* **471**(2183), 20150349 (2015)
29. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J.C.: From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA* **105**(13), 4972 (2008)
30. McCullough, M., Small, M., Stemler, T., Iu, H.H.C.: Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos* **25**, 053101 (2015)
31. Ruan, Y., Donner, R.V., Guan, S., Zou, Y.: Ordinal partition transition network based complexity measures for inferring coupling direction and delay from time series. *Chaos* **29**, 043111 (2019)
32. McCullough, M., Sakellariou, M.K., Stemler, T., Small, M.: Counting forbidden patterns in irregularly sampled time series I. The effects of under-sampling, random depletion, and timing jitter. *Chaos* **26**, 123103 (2016)
33. Bandt, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **88**, 174102 (2002)
34. Zhang, J.Y., Zhou, J., Tang, M., Guo, H., Small, M., Zou, Y.: Constructing ordinal partition transition networks from multivariate time series. *Sci. Rep.* **7**, 7795 (2017)
35. Amigó, J.M., Zambrano, S., Sanjuán, M.A.F.: Combinatorial detection of determinism in noisy time series. *Europhys. Lett.* **83**, 60005 (2008)
36. Amigó, J.M., Zambrano, S., Sanjuán, M.A.F.: Detecting determinism in time series with ordinal patterns: a comparative study. *Int. J. Bifurcation Chaos* **20**, 2915 (2010)
37. Kulp, C.W., Chobot, J.M., Niskala, B.J., Needhammer, C.J.: Using forbidden ordinal patterns to detect determinism in irregular sampled time series. *Chaos* **26**, 023107 (2016)
38. Amigo, J.M., Keller, K., Unakafova, V.A.: Ordinal symbolic analysis and its application to biomedical recordings. *Phil. Trans. R. Soc. A* **373**, 20140091 (2015)
39. Kulp, C.W., Chobot, J.M., Freitas, H.R., Sprechini, G.D.: Using ordinal partition transition networks to analyze ECG data. *Chaos* **26**, 073114 (2016)
40. Keller, K., Unakafov, A.M., Unakafova, V.A.: Ordinal patterns, entropy, and EEG. *Entropy* **16**, 6212–6239 (2014)

41. Zavaglia, M., Cona, F., Ursino, M.: A Neural Mass Model to Simulate Different Rhythms in a Cortical Region. *Comput. Intell. Neurosci.* **2010**, 456140 (2010)
42. Panuccio, G., Colombi, I., Chiappalone, M.: Recording and modulation of epileptiform activity in rodent brain slices coupled to micro electrode arrays. *J. Vis. Exp.* **135**, e57548 (2018)
43. Rutecki, P.A., Lebeda, F.J., Johnston, D.: 4-Aminopyridine produces epileptiform activity in hippocampus and enhances synaptic excitation and inhibition. *J. Neurophysiol.* **57**, 1911–1924 (1987)
44. Amaral, D.G., Witter, M.P.: The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* **31**, 571–591 (1989)
45. Avoli, M., D’Antuono, M., Louvel, J., Köhling, R., Biagini, G., Pumain, R., D’Arcangelo, G., Tancredi, V.: Network and pharmacological mechanisms leading to epileptiform synchronization in the limbic system in vitro. *Prog. Neurobiol.* **68**, 167–207 (2002)
46. Barbarosie, M., Louvel, J., Kurcewicz, I., Avoli, M.: CA3-released entorhinal seizures disclose dentate gyrus epileptogenicity and unmask a temporoammonic pathway. *J. Neurophysiol.* **83**, 1115–1124 (2000)
47. Spirtes, P., Glymour, C.N., Scheines, R., Heckerman, D.: *Causation, prediction, and search*. MIT Press, Boston (2000)
48. Ramsey, J., Zhang, J., Sprites, P.L. : Adjacency-faithfulness and conservative causal inference. arXiv preprint [arXiv:1206.6843](https://arxiv.org/abs/1206.6843) (2012)
49. Chickering, D.M.: Learning equivalence classes of Bayesian-network structures. *J. Mach. Learn. Res.* **2**, 445–498 (2002)
50. Panuccio, G., D’Antuono, M., Guzman, D.E.P., Lannoy, D.E.L., Biagini, G., Avoli, M.: In vitro ictogenesis and parahippocampal networks in a rodent model of temporal lobe epilepsy. *Neurobiol. Disease* **39**, 372–380 (2010)
51. Inaba, Y., Avoli, M.: Volume-conducted epileptiform events between adjacent neocortical slices in an interface tissue chamber. *J. Neurosci. Methods* **151**, 287–290 (2006)
52. Brunner, C., Billinger, M., Seeber, M., Mullen, T.R., Makeig, S.: Volume conduction influences scalp-based connectivity estimates. *Front. Comput. Neurosci.* **10**, 121 (2016)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.