Estimating the mechanisms underlying transient dynamics based on peri-event data

Anonymous Author(s)
Affiliation
Address
email

Abstract

Many important dynamical phenomena emerging in complex systems such as storms, stock market crashes, or reactivations of memory engrams in the mammalian brain are transient in nature. We consider the problem of learning accurate models of such phenomena based only on data gathered by detecting such transient events, and analyzing their peri-event dynamics. This approach is widely used to analyze spontaneous activity in brain recording, as it focuses on emerging events of particular significance to brain function. We show, however, that such an approach may misrepresent the properties of the system under study due to the event detection procedure that entails a selection bias. We develop the Debiased Snapshot (DeSnap) approach to de-bias the time-varying properties of the system estimated from such peri-event data and demonstrate its benefits in recovering state-dependent transient dynamics in toy examples and neural time series.

1 Introduction

Understanding the emergence and dynamics of transient phenomena in complex systems is a key challenge in many fields. While models are broadly used to investigate the underlying mechanisms, exploiting observational data to inform such modeling in a principled way remains largely elusive. One key challenge to devise such an approach is to formalize mathematically what is meant by emergence and what limitations it entails from the perspective of statistical data analysis. Emergence reflects the idea that the phenomenon is not triggered by an observable external input but instead results from the internal dynamics of the system. For example, in contrast to brain activity evoked by a visual stimulus, sharp-wave ripple phenomena are an internally generated transient brain oscillation associated with a previously experienced stimulus, which is observable in the hippocampus during offline states [Buzsaki et al., 1992].

Such events are ubiquitous in Neuroscience, and they are believed to be instrumental to brain function Friston [1995]. However, the analysis of their dynamics is based on an empirical detection followed by reporting "event-triggered" averages (see e.g. Logothetis et al. [2012], Sullivan et al. [2011], Lundqvist et al. [2018]). More advanced analyses, such as Granger causality, either focus on stimulus triggered activities, or simply apply the stimulus-triggered approaches to spontaneous activity without considering its specificity, namely, that the peri-event "trials" accumulated in this way are not from a randomized controlled trial, but instead are selected based on a specific signal detection procedure, and thus potentially subject to selection biases [Bareinboim et al., 2014].

In this work, we explicitly model the whole event-triggered analysis procedure to emphasize the specific issues to pay attention to when exploiting such data for fitting statistical models. After pointing out identifiability issues related to such an approach in a fully non-parametric setting, we investigate the linear autoregressive Gaussian case for which classical estimation procedures are

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

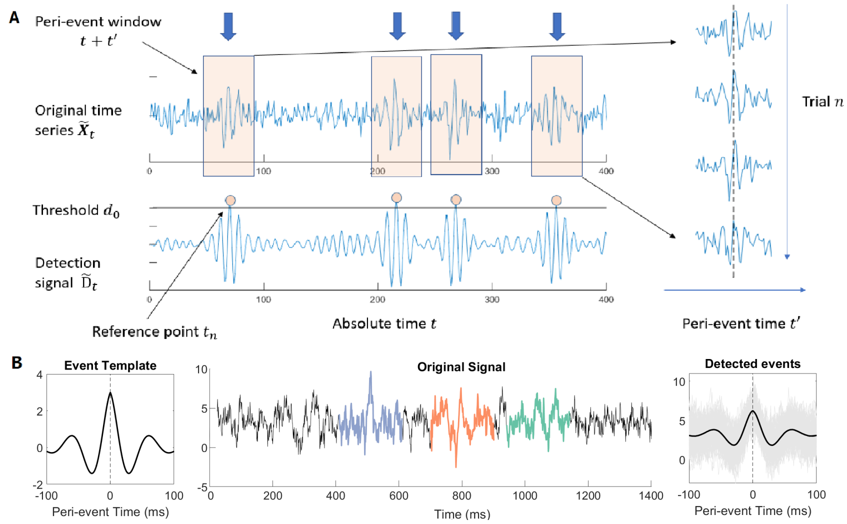

Figure 1: Illustration of the event detection procedure (A) and selection bias in a stationary time series (B).

shown to be biased. We then develop a bias correction procedure whose efficiency is illustrated on simulated data and further applied on neural recordings.

2 Methods

2.1 Modelling of peri-event snapshot detection procedure

Assume we want to analyze properties of a dynamical system for which we observe time series that repetitively exhibits a characteristic pattern of transient activity that we will call “event”. We may assume that this pattern is associated to the system visiting a particular region of its state space. Since the true state dynamics is usually not fully observed, we resort to the statistical analysis of the observed time series during transient events to learn properties of the system. This requires an event detection step for determining the location of putative transient events in the observed signals, which is typically performed by applying a filter to the original signal to get a *detection signal*.

We first provide a mathematical model of such detection procedure. As shown in Figure 1A, given a multivariate observation signal $\tilde{\boldsymbol{X}}_t$, the detection is typically based on a continuous-value *detection signal* $\tilde{D}_t$ ascribed to each (discrete) time point t. To ease notations, we will consider the detector is basing its decision on the last N_D samples, where $\tilde{\boldsymbol{X}}_{D,t} = \{\tilde{\boldsymbol{X}}_{t-1}, \cdots, \tilde{\boldsymbol{X}}_{t-N_D}\}$. Event occurrences are located based on a deterministic detector function that extracts information from N_D past samples

$$\tilde{D}_t = w(\tilde{\boldsymbol{X}}_{D,t})$$

such that only the time points satisfying $D_t \geq d_0$ are kept, used as *reference points*

$$\mathcal{T} = \{t_n\} = \{t | \tilde{D}_t \geq d_0\}\,,$$

where d_0 is referred to as a *threshold*. Samples from the long time series $\tilde{\boldsymbol{X}}_t$ covering a fixed peri-event time window around t_n, i.e., $\mathcal{I} = [-T/2,\, T/2]$, are extracted to build a two-way panel $\{\mathbf{X}_{t'}^{(n)}\}$ that contains $\#\mathcal{T}$ time-varying instances, of peri-event activity that we call *snapshots*:

$$\mathbf{X}_{t'}^{(n)} = \tilde{\mathbf{X}}_{t'+t_n}, t' \in \mathcal{I}, t_n \in \mathcal{T}\,. \tag{1}$$

Event-triggered analysis aims at exploiting the empirical distribution of this panel data to infer properties of the ground truth dynamical system.

2.1.1 A motivating example for selection bias in peri-event snapshots

To illustrate the potential problem caused by the detection procedure, let us consider an example of event detection with a Morlet wavelet-like discrete-time template w (exemplifying the detection of some oscillatory event) in a stationary time series:

$$w_t = \begin{cases} 3\exp(-|t|/4)\cos(t)\,, |t| \leq 10\,, \\ 0\,,\ \text{otherwise.} \end{cases}$$

Due to the symmetry of the template, we can implement the template matching procedure by computing a *detection signal* $\tilde{D}_t$ resulting from the convolution of this template with the observed time series $\tilde{D}_t = \left(w * \tilde{X}\right)_t$ and extract peri-event snapshots according to Section 2.1. We applied the above Morlet detector to a white noise signal made of *i.i.d.* normal samples (zero mean, unit variance) using a detection threshold of 3 SD (standard deviation). The Morlet template, original signal, example events, and resulting peri-event snapshots are provided in Figure 1B. Due to the choice of a large selection threshold, all snapshots are very similar to the template. While this is expected from a template matching approach, this also demonstrates that the selection of snapshots based on such procedure introduces a structure in $X_{t'+t_n}$ that is not related to the properties of the completely unstructured (*i.i.d.*) original time series $\tilde{X}_t$. Next, we provide a framework based on the theory of dynamical systems to shed light this form of *selection bias*.

2.2 Snapshot analysis framework

2.2.1 Continuous time dynamics perspective

We first expose informally a continuous time dynamical system framework to justify the discrete time snapshot analysis presented above. We assume that a given type of neural event is associated to a single specific region of the state space favoring their emergence. The dynamics of hidden states in this region is inferred by collecting multiple "trials" that each comprises the sequence of measurements recorded from the system during one occurrence of certain events. We assume these trials correspond to portions of state trajectories passing through the specific region of the state space where events are prone to emerge.

Assume a deterministic continuous time dynamical system governed by the autonomous differential equation

$$\begin{cases} \frac{d\mathbf{z}}{dt}(t) = \mathbf{H}(\mathbf{z}(t))\,, \\ \mathbf{z}(t_0) = \mathbf{z}_0\,, \end{cases} \tag{2}$$

where $\mathbf{z}(t)$ represents the state of the system at time t. The flow of vector field $\mathbf{H}$ is then defined as

$$\varphi(\mathbf{z}_0, t) = \mathbf{z}(t)\,,\ \mathbf{z}_0 \in \mathcal{Z},\ t \in \mathbb{R}\,,$$

and is such that

$$\varphi(\varphi(\mathbf{z}_0, t_1), t_2) = \varphi(\mathbf{z}_0, t_1 + t_2),\ t_1, t_2 \in \mathbb{R}\,.$$

As illustrated in Figure 2A, we assume the vector of observations $\tilde{\mathbf{x}}(t)$ for a given event instance are deterministic functions of the current state ($\tilde{\mathbf{x}}(t) = \tilde{f}(\mathbf{z}(t))$) and an event is detected when the state trajectory crosses a set $\mathcal{E}_0$ included in an hyperplane of the state space. For each such event, the absolute $\mathcal{E}_0$-crossing time is called the reference time $t = t_0$ and is mapped to the peri-event time $t' = 0$, and we sample the time series at regular time intervals around this event. As a consequence, states corresponding to a given perievent time sample t' belong to a corresponding set $\mathcal{E}_{t'}$, $t' \in \{\ldots, -1, 0, +1, \ldots\}$. Given the observation $\tilde{\mathbf{x}}(t)$ the deterministic mapping between two successive sets implies that it is also a deterministic function of the past state $\mathbf{z}(\mathbf{t}-\mathbf{1})$. Following the principle of the *Takens theorem* [Takens, 1981], information about $\mathbf{z}(\mathbf{t}-\mathbf{1})$ can be gathered by collecting values of the observations at multiple lags k in the past $\{\tilde{\mathbf{x}}(t-k)\}_{k=1..p}$. However, this information may still remain incomplete especially if the number of lags is small and the dimension of $\mathcal{Z}$ is large, which is likely the case for complex physical, biological or social systems such as the brain. Under *ergodicity* and *mixing* assumptions for our dynamical system (see e.g. [Lasota and Mackey, 2013]), if the event occurs long enough after the initialization of the dynamics, $\mathbf{z}(\mathbf{t}-\mathbf{1})$ is approximately distributed according to the invariant measure $\boldsymbol{\mu}$ of the system. As a consequence, it can be modeled as a random vector $\mathbf{Z}_{t-1} \sim \boldsymbol{\mu}$, and the knowledge of the vector of past observations $\mathbf{x}_{p,t}$ up to lag p reduces the uncertainty on the state trough the conditional $\mathbf{Z}_{t-1}|\mathbf{x}_{p,t}$. The deterministic (and invertible) mapping between $\mathcal{E}_{t-1}$ and $\mathcal{E}_t$ through φ leads to a stochastic model for the state $\mathbf{Z}_t|\mathbf{x}_{p,t}$ as well as current observations $\mathbf{X}_t|\mathbf{x}_{p,t}$. We can thus parameterize each conditional distribution as

$$\mathbf{X}_t = \tilde{f}(\varphi(\mathbf{Z}_{t-1}|\mathbf{x}_{p,t}, 1)) = f_t(\mathbf{x}_{p,t},\ \boldsymbol{\eta}_t) \tag{3}$$

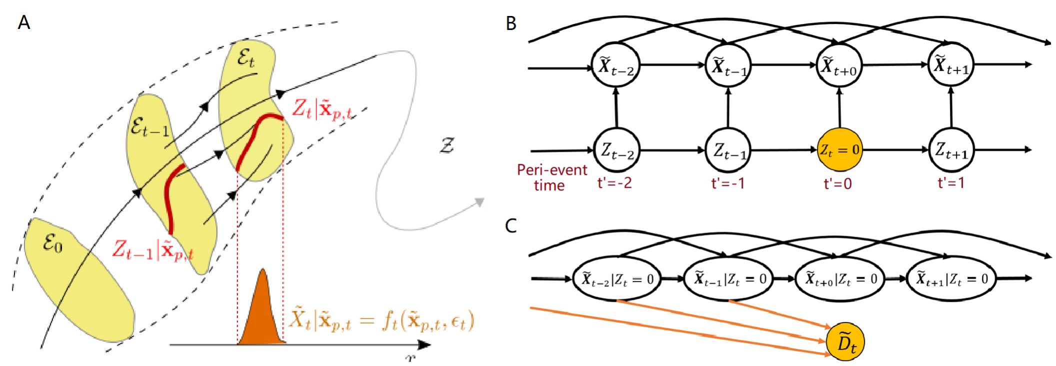

Figure 2: (A) Continuous time dynamical system perspective on transient events. (B) Causal graph for the VAR model of peri-event snapshots. (C) Detection introduces an additional node into the causal graph.

where f_t is a deterministic function and $\boldsymbol{\eta}_t$ models the randomness of the prediction of $\mathbf{X}_t$ given past observations $\mathbf{x}_{p,t}$. This randomness is itself due to the remaining uncertainty of the location in state space $\mathbf{Z}_{t-1}$ based on past observations. It is noteworthy that this uncertainty also entails that f_t and the distribution of $\boldsymbol{\eta}_t$ depend on t. [1]

2.2.2 Selection bias in a discrete-time snapshot model

From this previous section, we see the interest of modeling transient events as a state dependent time series, where the focus is put on a selected location of the state space. Assuming we sample the continuous dynamics with a sufficiently large sampling rate, we can make a linear approximation of equation 3, justifying the use of time-inhomogeneous linear VAR type models, for which coefficient estimation procedures are established. For a simplified representation of the state dependency of the overall dynamics of the system, we use Markov Switching Models (MSM) that combine a discrete state dependency with vector VAR dynamics [Hamilton, 1989]. More precisely, the MSM state Z_t is a discrete Markov chain with m-states and transition matrix M such that

$$p(Z_t = k | Z_{t-1} = j) = M_{k,j}$$

and this state controls the time varying parameters of the VAR model for $\tilde{\boldsymbol{X}}_t$

$$\tilde{\boldsymbol{X}}_t = A_{Z_t} \tilde{X}_{p,t} + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(b_{Z_t}, \Sigma_{Z_t})\,. \tag{4}$$

Applying the detection procedure under the snapshot analysis framework when targeting the discrete hidden state $Z_t = 0$, we make the following key assumption of "perfect detection":

Assumption 1 (Perfect detection) *Assume that $\tilde{D}_t$ being above a certain known threshold d_0 entails with probability one that the observed system is in target state $Z_t = 0$, i.e., $P(Z_t = 0 | D_t \geq d_0) = 1$*

With this assumption, for each *reference point* t_n, $P(Z_{t_n} = 0 | D_{t_n} \geq d_0) = 1$. Notably, this assumption provides only a sufficient condition to have $Z_t = 0$, but not a necessary one, i.e., we can have $P(D_t \geq d_0 | Z_t = 0) < 1$. This suggests that such thresholding detection only select a subset of all time points satisfying $Z_t = 0$ thus leading to selection bias. As a consequence, collected snapshots at peri-event time t' are distributed according to $\tilde{\boldsymbol{X}}_{t'+t_n} | \tilde{D}_{t_n} \geq d_0$ which typically differs from $\tilde{\boldsymbol{X}}_{t'+t} | Z_t = 0$. Finding a better approximation of this last distribution based on the snapshot panel data is the main goal of this paper.

2.3 Correction of detection-dependent selection bias

2.3.1 Recoverability under Structural Causal Models

The recoverability under sample selection bias has been investigated within the framework of Structural Causal Models (SCMs) [Pearl, 2000], by using causal graphical models equipped with

[1] if the state would be fully observed, the mapping would be independent of time, due to the autonomous differential equation (2)

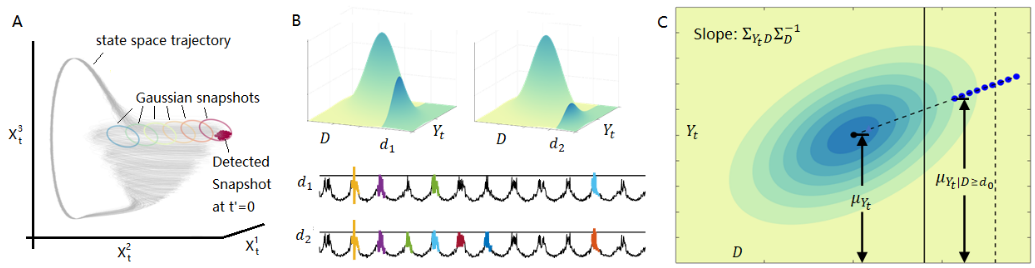

Figure 3: (A) Non-linear dynamical system [Montbrió et al., 2015] trajectories and typical regions where local Gaussian approximations are made. (B) Illustration of the support's regions sampled for different detection thresholds. (C) Principle of DeSnap for retrieving unconditional mean based on data for several detection threshold.

a special node representing the sampling process Bareinboim and Pearl [2012], Bareinboim et al. [2014]. Here we address the recoverability of detection-dependent selection bias by treating the VAR model of snapshots (Section 2.2.2) as an SCM and applying the recoverability theories.

For state-dependent peri-event data, the unbiased peri-event snapshots $\tilde{\boldsymbol{X}}_{t'+t}$ for the state $Z_t = 0$ can be obtained by gathering observation signals $\tilde{\boldsymbol{X}}_t$ for peri-event time $t' + t$ where $Z_t = 0$ and $t' = [-T/2, T/2]$ with the peri-event window T (see Section 2.1). As seen in Figure 2B, the SCM formalism for the unbiased peri-event snapshots $\tilde{\boldsymbol{X}}_{t+t'}|Z_t = 0$ can be seen as conditioned on the yellow hidden state node for $t' = 0$ such that $Z_t = 0$. With the assumption that all peri-event $\tilde{\boldsymbol{X}}_{t+t'}$ belong to the state $Z_t = 0$, all paths through the hidden states are blocked such that this SCM can be approximated by the 1-layer VAR model of $\tilde{\boldsymbol{X}}_{t+t'}|Z_t = 0$, as shown in Figure 2C, where $\tilde{\boldsymbol{X}}_{t'+t}|Z_t = 0$ can be assumed only dependent on $\tilde{\boldsymbol{X}}_{p,t'+t}|Z_t = 0$ for arbitrary perievent times t'.

The detection procedure is equivalent to adding a node for detection in the VAR model of $\tilde{\boldsymbol{X}}_{t'+t}|Z_t = 0$ (Figure 2C). For a Markov-switching VAR model of Eq. 4, we are interested in using $P(\tilde{\mathbf{X}}_{t'+t}|\tilde{\mathbf{X}}_{p,t'+t}, Z_t = 0, \tilde{D}_t \geq d_0)$ to recover the conditional probability characterizing the markovian dynamics $P(\tilde{\mathbf{X}}_{t'+t}|\tilde{\mathbf{X}}_{p,t'+t}, Z_t = 0)$ for t' in a peri-event time window.

Based on d-separation and the non-parametric recoverability theories in [Bareinboim and Pearl, 2012, Bareinboim et al., 2014], the conditional probability of two variables $P(Y|X)$ in an SCM can be recovered from samples selected using $S = 1$ if we have the d-separation $Y \perp\!\!\!\perp_d S|X$, such that $P(Y|X, S = 1) = P(Y|X)$ (for an elaboration see Section A.1.2). Thus, $P(\tilde{\mathbf{X}}_{t'+t}|\tilde{\mathbf{X}}_{p,t'+t}, Z_t = 0)$ is identifiable from the snapshot data $P(\tilde{\mathbf{X}}_{t'+t}|\ \tilde{\mathbf{X}}_{p,t'+t}, Z_t = 0, \tilde{D}_t \geq d_0)$ for points after the detection time point $t' \geq 0$, but not before (i.e., $t' < 0$). This theoretical result provides insights about challenges for identifying the peri-event dynamics in a non-parametric setting. However, we will show that by enforcing parametric assumptions on the model would lead to identifiability for a broader range of time points.

2.3.2 Bias in Gaussian parametric VAR models for peri-event snapshots

Following Section 2.2.2, a Gaussian VAR model of the peri-event dynamics takes the form:

$$\tilde{\boldsymbol{X}}_{t'+t} := A_{t'+t}\tilde{\boldsymbol{X}}_{p,t'+t} + \boldsymbol{\eta}_{t'+t}\,, \boldsymbol{\eta}_{t'+t} \sim \mathcal{N}(\boldsymbol{k}_{t'+t}, \Sigma_{t'+t})\,. \tag{5}$$

Notably, these quantities are all conditioned on $Z_t = 0$, while we omit this condition to ease notation. Modelling peri-event data with this model is essentially finding a local linear map between consecutive time point that are assume Gaussian distributed. As Figure 3A illustrates, transient trajectories of a stochastic non-linear dynamical systems may be clearly non-Gaussian when looking at the full state space. However, approximating local dynamics by Gaussian VAR process in a small region of the state space may be reasonable, for example when considering processes whose stochasticity stems from a Wiener process. Detection-dependent selection bias of peri-event data can then be modeled as sampling from a portion of the support of the joint Gaussian distributions, as illustrated in Figure 3B.

Determining such a model requires the estimation of time-varying model parameters (autoregressive coeffcients, covariance matrices, etc.) from peri-event data, whose estimation can be done following

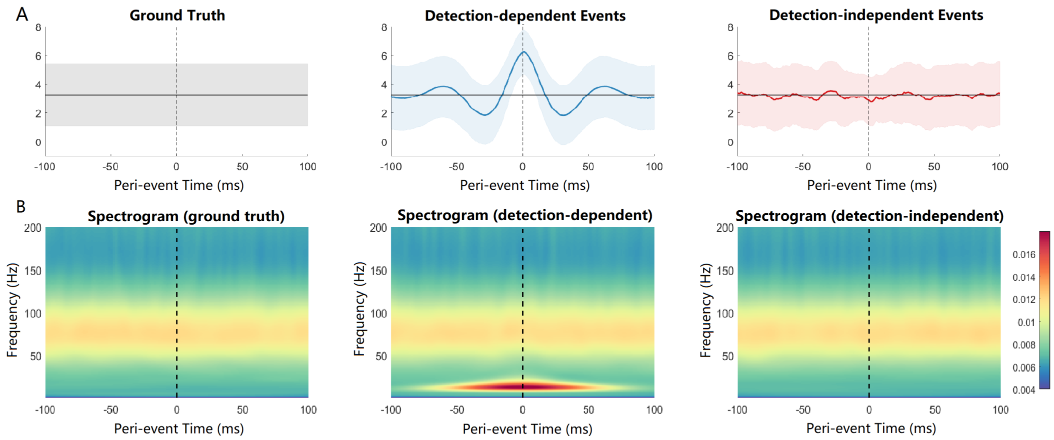

Figure 4: DeSnap recovers ground truth peri-event dynamics and spectrograms in Figure. 1, comparing the detection-dependent and detection-independent cases.

[Shao et al., 2022a]. Critically, model coefficients $A_{t'+t}$ are inferred from the time-resolved empirical statistics of the snapshots, making it sensitive to detection (see illustration Figure 3C for the mean):

$$\widehat{A}_{t'+t} = \widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{t'+t}\tilde{\boldsymbol{X}}_{p,t'+t}} (\widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{p,t'+t}})^{-1} \tag{6}$$

where $\widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{t'+t}\tilde{\boldsymbol{X}}_{p,t'+t}}$ and $\widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{p,t'+t}}$ denote empirical (cross-)covariance matrices. Specifically, as snapshots are detected in observed time series $\tilde{\boldsymbol{X}}_t$ using condition $\tilde{D}_{t_n} > d_0$, the covariance matrices we obtain directly from the detected peri-events snapshots are estimates of the detection-dependent covariance matrices $\widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{t'+t_n}\tilde{\boldsymbol{X}}_{p,t'+t_n}|Z_{t_n}=0,\tilde{D}_{t_n}>d_0}$ and $\Sigma_{\tilde{\boldsymbol{X}}_{p,t'+t_n}|Z_{t_n}=0,\tilde{D}_{t_n}>d_0}$, may differ from the detection-independent ones, i.e. $\widehat{\Sigma}_{\tilde{\boldsymbol{X}}_{t'+t}\tilde{\boldsymbol{X}}_{p,t'+t}|Z_t=0}$ and $\Sigma_{\tilde{\boldsymbol{X}}_{p,t'+t}|Z_t=0}$.

2.3.3 Debiasing based on threshold variations: the *DeSnap* algorithm

We propose here a novel method, named *DeSnap*, to correct for the detection-dependent bias in peri-event dynamic modelling by setting multiple thresholds during detection.

By stacking the current and lagged snapshots as $\mathbf{Y}_t = \left[\tilde{\boldsymbol{X}}^T_{t'+t}|Z_t = 0, \tilde{\boldsymbol{X}}^T_{p,t'+t}|Z_t = 0\right]^T$ and assuming joint Gaussianity, we have established the relationship between the detection-dependent and detection-independent time-varying state statistics (see Appendix A.2 for derivations):

$$\mu_{\mathbf{Y}_t|D\geq d_0} = \mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}\left(\overline{d} - \mu_D\right), \tag{7}$$

$$\Sigma_{\mathbf{Y}_t|D\geq d_0} = \Sigma_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}c(d_0)\Sigma_D^{-1}\Sigma^T_{\mathbf{Y}_t D}. \tag{8}$$

where $\overline{d}$ is the average of D_t over the threshold d_0 and $c(d_0)$ is a scalar statistic of D_t.

The left hand sides of Eq. 7 and Eq 8, as well as $\overline{d}$ can be estimated empirically, while $\Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}$, μ_D and $c(d_0)$ are unknown state-dependent variables. Then it is possible to get different samples for the empirically obtainable variables by setting multiple detection thresholds and detect snapshots multiple times. Then by performing three linear regressions over the samples with different thresholds and time points, detection-independent statistics $\mu_{\mathbf{Y}_t}$ and $\Sigma_{\mathbf{Y}_t}$ can be retrieved (for details see Appendix A.2). An illustration of the DeSnap method can be found in Figure 3.

3 Results

3.1 Univariate stationary process with Morlet-shaped detection

We first test the DeSnap algorithm on the correction of the motivation example in Section 2.1.1 and Figure 1, which is simplest uni-state system to test the performance of DeSnap. The dynamics of the

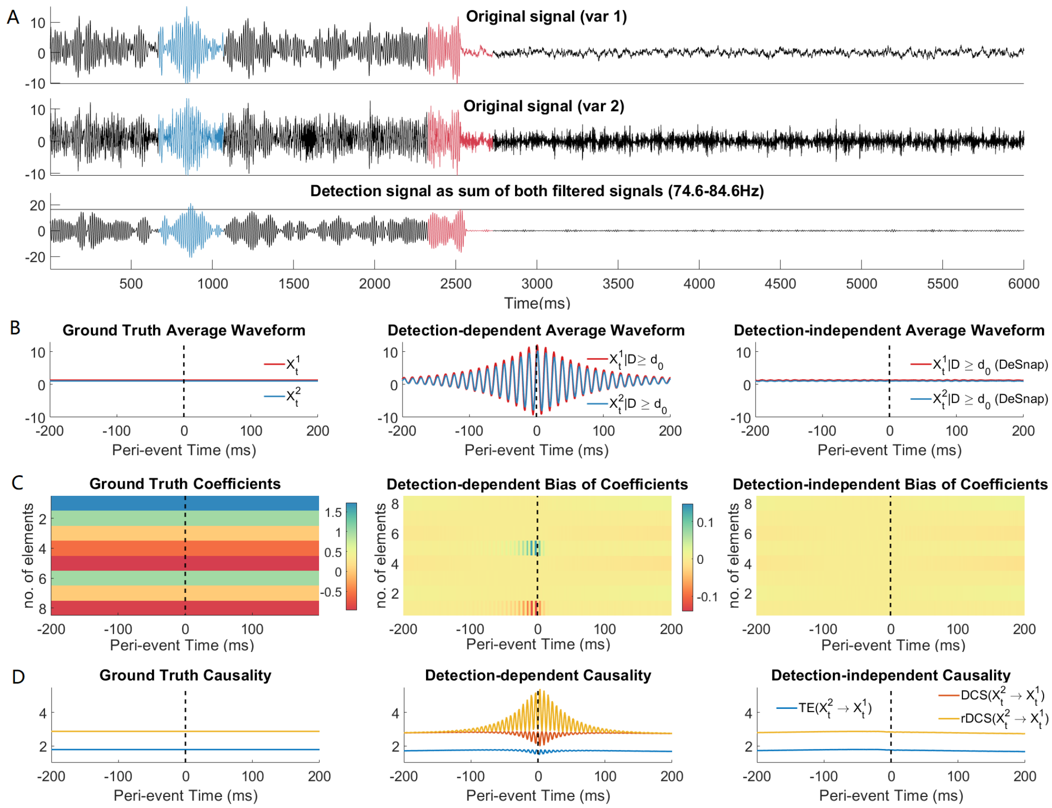

Figure 5: DeSnap recovers state-dependent dynamics and causality measures in an two-state Markov-switching VAR(2) model

system is controlled by a constant coefficient matrix A_t and a non-zero innovations mean η_t:

$$A_t = [.8 \quad -0.64 \quad 0.512 \quad -0.4096]\,, \eta_t = 0.2$$

The detection procedure is identical to what is described in Section 2.1.1.

Comparing the ground-truth temporal waveforms, the detection dependent waveforms reflect a bias consistent with the template, as already illustrated in Figure 1B. Figure 4B shows the power spectrograms computed from the estimated VAR model as a frequency-domain representation of the peri-event dynamics. It can be easily seen that in addition to the time-invariant prominent activities in the band [50-100]Hz, detection with the Morlet-shaped template introduces a strong transient pattern in the lower frequencies. After applying the DeSnap approach, both the ground-truth time-varying waveforms and spectrograms are recovered, supporting the ability of DeSnap to correct detection-dependent bias.

3.2 Recovery of state-dependent statistics and causal interactions

We further demonstrate with another toy model that DeSnap is able to recover the state-dependent dynamics when the system undergoes transition. A perfect example of such a condition is a bi-variate uni-directionally-coupled Markov-switching VAR model implementing alternations between a non-oscillatory regime and the oscillatory regime.

The dynamics of the two regimes are determined by two sets of VAR parameters, where the parameters remain time-invariant within each regime:

$$A = \begin{bmatrix} -0.5751 & 1 & -0.9408 & 1 \\ 0 & 1.7263 & 0 & -0.9737 \end{bmatrix}, \boldsymbol{k} = \begin{bmatrix} 0 \\ 0.65 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

and

$$A' = \begin{bmatrix} 0.5 & 1 & 0.3 & 1 \\ 0 & -1.5 & 0 & -0.7 \end{bmatrix}, \boldsymbol{k}' = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

We denote the oscillatory regime “state 0” and the other “state 1”. The parameters are designed such that “state 0” show strong oscillations in the 74.6-84.6-Hz band, which can be detected as events. The Markov switching model’s hidden state dynamics is determined by the transition probabilities: $P(\text{state } 0|\text{state } 1) = P(\text{state } 1|\text{state } 0) = 0.0001$. Example traces of the original signals and the detection signal are presented in Figure 5A.

As the events occur only during the stationary ’state 0’, peri-event dynamics should be time-invariant. Similar to Figure 4, detection introduced spurious oscillatory patterns in the average waveform of the events and the estimated VAR coefficients, as seen in Figure 5 B,C (middle column), reflecting the detection-dependent bias estimated directly from event ensembles. These bias are clearly removed after applying DeSnap to the detection-dependent model, as seen in Figure 5B,C (right column), supporting DeSnap as a promising tool to uncover state-dependent dynamics.

As this two-state model incorporates uni-directional coupling between the variables (also seen from the coefficients), we compared 3 time-varying causality measures, Transfer entropy(TE), Dynamical Causal Strength(DCS), and relative Dynamic Causal Strength (rDCS) (as proposed in Shao et al. [2022b]) in the ground truth direction, as seen in Figure 5D. All three measures oscillate around the peri-event time $t' = 0$ although the ground-truth causal connectivity remain constant within a regime. Such time-invariance properties of the causality measures are restored after applying DeSnap. These results suggest that combining DeSnap with causality measures provides knowledge of the state-dependent causal interaction underlying events.

3.3 Application to in-vivo hippocampal recordings

We now applied the methodology to real data. Ramirez-Villegas et al. [2021] has characterized transient hippocampal events prominent in 3 frequency bands (i.e. hTheta, high Gamma and Ripples) and hypothesized that the first type of events and the later two are associated to two distinct brain states. Here we use the *DeSnap* approach to identify the key transient dynamics underlying these three types of events.

We defined and detected events with 14 distinct frequency bands distributed into 4 classical groups: hTheta, Beta, Gamma and Ripple bands, in 16 pairs of local field potential signals recorded in the pyramidal layer of CA1 hippocampal subfield in an anesthetized macaque (sampling rate 667Hz). The detection is performed with band pass filters with the thresholds suggested by Ramirez-Villegas et al. [2021]. The resulting average power spectrograms are shown in Figure 6A,B(upper rows), reflecting the detection-dependent and threshold-dependent dynamics underlying each type of events.

We applied DeSnap to the 14 dataset of extracted event ensembles to estimate the detection-independent dynamics, and plotted the reconstructed spectrograms in Figure 6A,B(lower rows). Interestingly, we found that after DeSnap, the spectrograms forms two filter-band invariant patterns within the hTheta band group and within the high-Gamma/Ripple band group. The similarity between spectrograms within each band group and discrepencies between band groups are further characterized by concentrated patterns in Similarity Matrix (Figure 6C) and sample distributions in a dimension reduced spaces obtained by Multidimensional Scaling (Figure 6D), where clustering quality illustrates the formation of two clusters in the sample space (Figure 6E). By comparison, the detection-dependent spectrograms do not show obvious clusters. This result suggests that high-Gamma and ripple events are manifestations of the same underlying transient phenomenon, while hTheta events may be generated by another state-dependent mechanism. On the methodological level, this supports the ability of DeSnap to better recover ground truth nonlinear transient dynamics, independent of the detection procedure.

Discussion

In summary, in this paper we focused on the spontaneity of transient phenomenon observed in dynamical systems. We characterized the effect of detection procedure on reconstructing hidden transient dynamics from peri-event data in the form of selection bias and proposed a new method - *DeSnap* - to correct for the bias. Consistent results on applying DeSnap to toy models and electrophysiological siganls has confirmed its performance of identifying state-dependent dynamic properties of systems. Therefore, DeSnap has the capability to deepen the understanding of the transient mechanism underlying certain transient events. Koopman-Operator-based theories [Brunton et al., 2021] may yield further generalizations of DeSnap for non-Gaussian or deterministic systems.

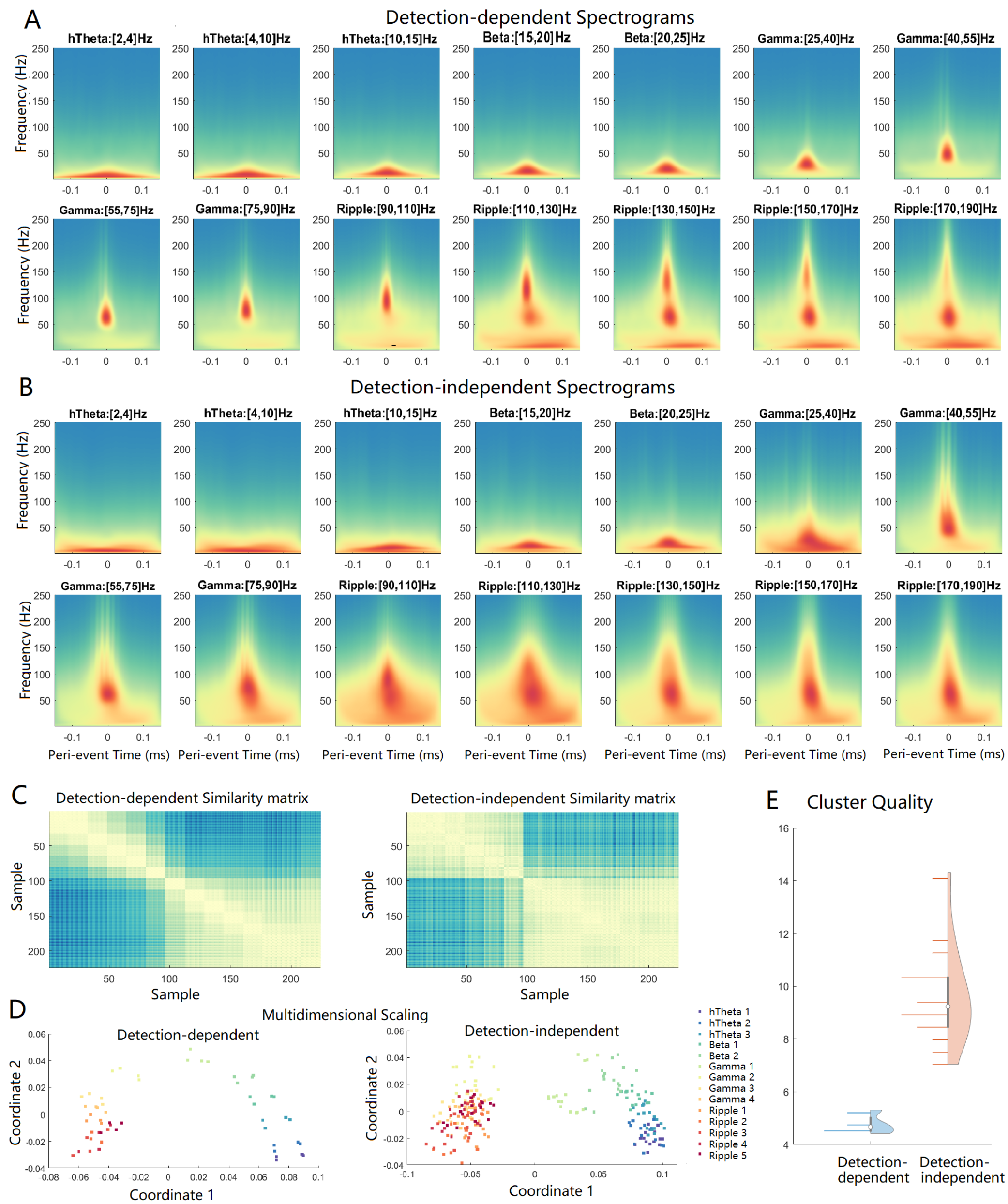

Figure 6: DeSnapped spectrograms identify two groups of filter-band-invariant events reflecting distinct state-dependent transient mechanisms.

References

G. Buzsaki, Z. Horvath, R. Urioste, J. Hetke, and K. Wise. High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025–7, 1992. ISSN 0036-8075 (Print) 0036-8075 (Linking).

Karl John Friston. Neuronal transients. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1362):401–405, 1995.

N K Logothetis, O Eschenko, Y Murayama, M Augath, T Steudel, H C Evrard, M Besserve, and A Oeltermann. Hippocampal-cortical interaction during periods of subcortical silence. *Nature*, 491 (7425):547–553, nov 2012. doi: 10.1038/nature11618. URL `http://dx.doi.org/10.1038/nature11618`.

David Sullivan, Jozsef Csicsvari, Kenji Mizuseki, Sean Montgomery, Kamran Diba, and György Buzsáki. Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. *Journal of Neuroscience*, 31(23):8605–8616, 2011.

Mikael Lundqvist, Pawel Herman, Melissa R Warden, Scott L Brincat, and Earl K Miller. Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nature communications*, 9(1):1–12, 2018.

Elias Bareinboim, Jin Tian, and Judea Pearl. Recovering from selection bias in causal and statistical inference. In *AAAI*, pages 2410–2416, 2014.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

Andrzej Lasota and Michael C Mackey. *Chaos, fractals, and noise: stochastic aspects of dynamics*, volume 97. Springer Science & Business Media, 2013.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384, 1989.

J. Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.

Elias Bareinboim and Judea Pearl. Controlling selection bias in causal inference. In *Artificial Intelligence and Statistics*, pages 100–108, 2012.

Ernest Montbrió, Diego Pazó, and Alex Roxin. Macroscopic description for networks of spiking neurons. *Physical Review X*, 5(2):021028, 2015.

Kaidi Shao, Nikos K. Logothetis, and Michel Besserve. Bayesian information criterion for event-based multi-trial ensemble data, 2022a. URL `https://arxiv.org/abs/2204.14096`.

Kaidi Shao, Nikos K Logothetis, and Michel Besserve. Information theoretic measures of causal influences during transient neural events. *arXiv preprint arXiv:2209.07508*, 2022b.

Juan F Ramirez-Villegas, Michel Besserve, Yusuke Murayama, Henry C Evrard, Axel Oeltermann, and Nikos K Logothetis. Coupling of hippocampal theta and ripples with pontogeniculooccipital waves. *Nature*, 589(7840):96–102, 2021.

Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*, 2021.

J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference – Foundations and Learning Algorithms*. MIT Press, 2017.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.

Checklist

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

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

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

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [No]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

A.1 Structural Causal Models and recoverability for selection bias

A.1.1 Basics of Structural Causal Models

SCMs are generalizations of Bayesian networks that combine Structural Equation Models (SEM) to incorporate directional information for causal analysis [Pearl, 2000]. A structural equations takes the form

$$Y := f(X_1, \cdots, X_k, \epsilon)$$

where the right hand side determines the assignment of values on the left-hand side. In the most usual case, Y and $\{X_j\}_{j\in\{1,\cdots,k\}}$ represent observed variables and ϵ a variable accounting for (unobserved) exogenous effects.

Based on this, a SCM is defined for a set of random variables $\{V_j\}$ associated to vertices in a graph as the follows.

Definition 1 (Structural Causal Model (SCM) (see e.g. Peters et al. [2017])) *A d-dimensional structural causal model is a triplet $(\mathbb{S}, P_N, \mathcal{G})$ consisting of:*

- *a directed acyclic graph $\mathcal{G}$ with d vertices*
- *a set $\mathbb{S}$ of structural equations*

  $$V_j := f_j(\boldsymbol{PA}_j, N_j), j = 1, \dots, d,$$

  where $\boldsymbol{PA}_j$ are the variables indexed by the set of parents of vertex j in $\mathcal{G}$
- *a joint distribution P_N over the exogenous variables N_j, which are assumed jointly independent.*

One attractive feature of this formalism is that the SCM's graph entails key properties of the join distribution of the nodes $\{V_j\}$, like the Markov properties and conditional independences (see e.g. Bishop [2006]).

Proposition 1 (Markov properties) *For a given SCM $(\mathbb{S}, P_N, \mathcal{G})$, the joint distribution P_V is Markovian with respect to $\mathcal{G}$, i.e. it satisfies the following properties:*

1. *(local Markov property) each variable V_j is independent of its non-descendants given its parents $\boldsymbol{PA}_j$,*
2. *(Markov factorization property) assume the joint distribution P_V has a density, then*

$$p(\boldsymbol{v}) = p(v_1, \dots, v_d) = \prod_{j=1}^{d} p(v_j|\boldsymbol{pa}_j)$$

With the conditional independence indicated in the local Markov property, the Bayesian network greatly simplifies the calculation of joint probabilities. In addition, the concept of d-separation allows assessing systematically the conditional independences between subsets of nodes in $\mathcal{G}$ based on graphical criteria of d-separation (see e.g. Pearl [2000]).

Definition 2 (d-separation) *A path p in graph $\mathcal{G}$ is said to be blocked by a set of nodes Z if either: (1) p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node m is in Z, or (2) p contains a collider $i \to m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z.*

Z is said to d-separate X from Y in $\mathcal{G}$ if and only if Z blocks every path from a node in X to a node in Y. This property is denoted $X \perp\!\!\!\perp_{\mathcal{G}} Y|Z$.

Indeed, d-separation allows stating the *global Markov property* (see e.g. Peters et al. [2017]).

Proposition 2 (Global Markov property) *For a given SCM $(\mathbb{S}, P_N, \mathcal{G})$ and subsets of nodes X, Y, Z in $\mathcal{G}$, then*

$$X \perp\!\!\!\perp_{\mathcal{G}} Y|Z \Rightarrow X \perp\!\!\!\perp_{P_V} Y|Z\,.$$

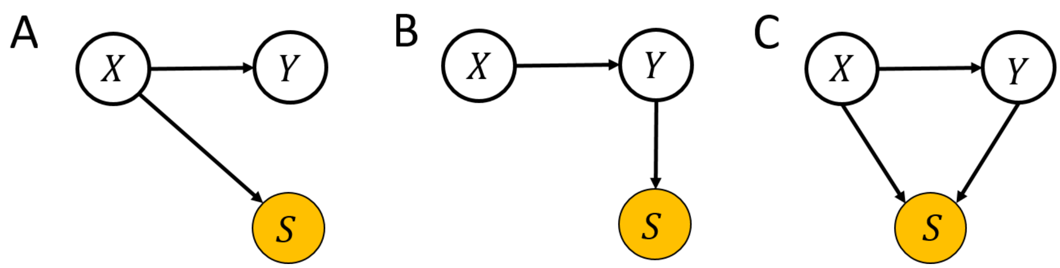

Figure 7: SCM, selection bias and recoverability (adapted from Bareinboim et al. [2014]). (A) SCM describing sample selection based on X, leading to identifiability of $P(Y|X)$ based on selected data. (B) SCM describing sample selection based on Y, leading to non-identifiability of $P(Y|X)$ based on selected data. (C) SCM describing sample selection based on both X and Y, leading to non-identifiability of $P(Y|X)$ based on selected data.

This proposition indicates that the conditional independences in the graph as defined by d-separation rules also hold for the corresponding random variables of the associated SCM.

The next sections will show how these basic concepts and properties of a SCM would fascilitate the understanding of sampling bias.

A.1.2 Recoverability with Sampling Selection Bias

In the simplest two-node SCM, the identifiability or recoverability of the effect based on different sampling methods has been investigated in [Bareinboim et al., 2014].

Figure 7A, B, C show three sampling conditions in a two-node SCM consisting of variables X and Y (with X causing Y). Sampling is represented by binary variable S in an additional node designed as descendant for either X or Y. S takes the value 1 when a data point is selected and zero otherwise. In Figure 7A, sample selection is a function of X only, while Figure 7B describes a sample selection based on Y only. Figure 7C presents the condition where sample selection depends on both variables.

In this model, we are interested in estimating the conditional propability of $P(Y|X)$ from sampled data. What is critical is whether $P(Y|X)$ can be recovered from the joint distribution of the selected samples $(X,Y)|S = 1$ given different sampling scenarios. Bareinboim et al. [2014] show that, under standard assumptions, a necessary and sufficient condition for recoverability is conditional independence between target variable Y and selection variable S, given conditioning variable X ($Y \perp\!\!\!\perp S|X)$) such that $P(Y|X, S = 1) = P(Y|X)$. For the scenarios of Figure 7, this implies that $P(Y|X)$ can be recovered $(X,Y)|S = 1$ in the case of Figure 7A, but not in Figure 7B and Figure 7C.

The rationale is simple according to the d-separation rules (see Section A.1.1 for details). In the condition of Figure 7A, conditioning on X corresponds to the "fork" case in the d-separation rules, indicating that the conditional independence $Y \perp\!\!\!\perp S|X$ is satisfied. On the contrary, Figure 7B shows the condition where such that $P(Y|X)$ is not recoverable from sample selected data because the above conditional independence requirement (Y independent of S given X) is not satisfied. For Figure 7B, detailed proof has been provided in Bareinboim et al. [2014]. The case in Figure 7C corresponds to the "collider" case of d-separation where a common observed descendant induces extra dependency between the ancestors.

However, it is important to point out that this negative theoretical result corresponds to a non-parametric case. In particular, putting further assumptions on the model that generated X and Y may help identify $P(Y|X)$.

A.2 *DeSnap*: Correction of coefficient estimation bias caused by selection

The calculation of many time-varying causality measures depends on the accurate estimation of autoregressive coefficient matrices. However, estimation with snapshots detected via thresholding tend to introduce selection bias into the estimated statistics, thus leading to erroneous estimation of causality measures.

If we assume that the peri-event snapshots under study can be modelled as a multi-variate autoregressive process $\mathbf{X}_t$, where the current state is a linear combination of the previous states:

$$\mathbf{X}_t = A_t \mathbf{X}_{p,t} + \boldsymbol{\eta}_t \,, \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{k}_t, \Sigma_t) \,.$$

Notably, the notations for peri-event snapshots $\mathbf{X}_t$ is different from notations for the general time series $\boldsymbol{X}_t$.

This is a k^{th}-order m-variate vector autoregressive process, with the current state defined as

$$\mathbf{X}_t = \left[X_t^1, X_t^2, \cdots, X_t^m\right]^T$$

and the past state as $\mathbf{X}_{p,t} = [\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \cdots, \mathbf{X}_{t-p}]^T$. Innovations $\boldsymbol{\eta}_t$ are time-inhomogeneous Gaussian random variables, where $\mathbb{E}[\eta] = \boldsymbol{k}_t, \mathrm{Cov}[\eta] = \Sigma_t$.

The estimation of two covariance matrices $\Sigma_{\mathbf{X}_t \mathbf{X}_p}$ and $\Sigma_{\mathbf{X}_p}$ determines the estimation of VAR coefficient matrix as $\widehat{A_t} = \widehat{\Sigma}_{\mathbf{X}_t \mathbf{X}_p} \left(\widehat{\Sigma}_{\mathbf{X}_p}\right)^{-1}$. The innovations mean and variances depends on the estimation of coefficient (see Shao et al. [2022a]).

As snapshots are detected in time series $\tilde{\boldsymbol{X}}_t$ using condition $D_{t_0} > d_0$, the covariances we obtain directly from the panel data estimation procedure are estimates of the conditional covariance matrices $\Sigma_{\mathbf{X}_t \mathbf{X}_p | D_{t_0} > d_0}$ and $\Sigma_{\mathbf{X}_p | D_{t_0} > d_0}$, may differ from the real (unconditional) ones.

Therefore our *DeSnap* procedure introduces a new approach to reduce the selection bias covariance matrices as follows. If we represent the snapshot values at peri-event time point t as a lagged state $\mathbf{Y}_t$, by concatenating $\mathbf{X}_t$ and $\mathbf{X}_p$, where $t \in [-T/2, T/2]$:

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{X}_{p,t} \end{bmatrix}$$

then second-order statistics of panel data approximate the conditional mean of the snapshots and can be written as :

$$\mu_{\mathbf{Y}_t | D_{t_0} \geq d_0} = \begin{bmatrix} \mu_{\mathbf{X}_t | D_{t_0} \geq d_0} \\ \mu_{\mathbf{X}p | D_{t_0} \geq d_0} \end{bmatrix}, \Sigma_{\mathbf{Y}_t | D_{t_0} \geq d_0} = \begin{bmatrix} \sum_{X_t | D_{t_0} \geq d_0} & \Sigma_{\mathbf{X}_t \mathbf{X}_p | D_{t_0} \geq d_0} \\ \sum_{\mathbf{X}_p \mathbf{X}_t | D_{t_0} \geq d_0} & \Sigma_{\mathbf{X}_p | D_{t_0} \geq d_0} \end{bmatrix} \tag{9}$$

For simplicity, we omit the time indices of D_{t_0} in the notations and refer to the *detection signal*, denoted by D. We now show how to exploit information in the snapshots to estimate the unconditional covariance under a joint Gaussian assumption of $\mathbf{Y}_t$ and D. For each values of $d \in D$ where $d \geq d_0$, the conditional distribution of $\mathbf{Y}_t | D = d$ is also Gaussian with mean $\mu_{\mathbf{Y}_t | D = d}$ and variance $\Sigma_{\mathbf{Y}_t | D = d}$, such that:

$$\mu_{\mathbf{Y}_t | D = d} = \mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D} \Sigma_D^{-1} (d - \mu_D) \tag{10}$$

$$\Sigma_{\mathbf{Y}_t | D = d} = \Sigma_{\mathbf{Y}_t} - \Sigma_{\mathbf{Y}_t D} \Sigma_D^{-1} \Sigma_{\mathbf{Y}_t D}^T \tag{11}$$

The conditional distribution of $\mathbf{Y}_t | D \geq d_0$ can then be computed as:

$$P\left(\mathbf{Y}_{\mathrm{t}} \mid \mathrm{D} \geq \mathrm{d}_0\right) = \int_{d_0}^{+\infty} \frac{P(D = d)}{P\left(D \geq d_0\right)} P\left(\mathbf{Y}_t \mid D = d\right) dd$$

The mean and covariance of this Gaussian mixture is a function of the mean and covariance of each element. For the mean we get

$$\begin{aligned}
&\mu_{\mathbf{Y}_t | D \geq d_0} \\
&= \textstyle\int_{d_0}^{+\infty} \frac{P(D=d)}{P(D \geq d_0)} \mu_{\mathbf{Y} | D = d} dd \,, \\
&= \textstyle\int_{d_0}^{+\infty} \frac{P(D=d)}{P(D \geq d_0)} \left(\mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D} \Sigma_D^{-1} (d - \mu_D)\right) dd \,, \\
&= \mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D} \Sigma_D^{-1} \textstyle\int_{d_0}^{+\infty} \frac{P(D=d)}{P(D \geq d_0)} (d - \mu_D) \, dd \,, \\
&= \mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D} \Sigma_D^{-1} \left(\overline{d} - \mu_D\right) \,,
\end{aligned}$$

where $\overline{d}$ is the average of $D = d \geq d_0$

For the covariance, we use the law of total covariance (for two random variables X and Y)

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[\mathrm{Cov}(X, Y, D)\right] + \mathrm{Cov}\left(\mathbb{E}\left[X|D\right], \mathbb{E}\left[Y|D\right]\right)$$

to obtain

$$\begin{aligned}
&\Sigma_{\mathbf{Y}_t|D \geq d}\,, \\
&= \int_{d_0}^{+\infty} \tfrac{P(D=d)}{P(D \geq d_0)} \left(\Sigma_{\mathbf{Y}_t} - \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}\Sigma_{\mathbf{Y}_t D}^T\right) dd\,, \\
&+ \int_{d_0}^{+\infty} \tfrac{P(D=d)}{P(D \geq d_0)} \left(\mu_{\mathbf{Y}|D=d} - \mu_{\mathbf{Y}|D \geq d_0}\right) \left(\mu_{\mathbf{Y}|D=d} - \mu_{\mathbf{Y}|D \geq d_0}\right)^T dd\,, \\
&= \Sigma_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1} c \Sigma_D^{-1}\Sigma_{\mathbf{Y}_t D}^T\,,
\end{aligned}$$

where $c = \int_{d_0}^{+\infty} \frac{P(D=d)}{P(D \geq d_0)} \left(d - \mu_D\right)^2 dd - \left(\bar{d} - \mu_D\right)^2 - \Sigma_D$.

As a result, we have

$$\mu_{\mathbf{Y}_t|D \geq d_0} = \mu_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}\left(\bar{d} - \mu_D\right)\,, \tag{12}$$

$$\Sigma_{\mathbf{Y}_t|D \geq d} = \Sigma_{\mathbf{Y}_t} + \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1} c \Sigma_D^{-1}\Sigma_{\mathbf{Y}_t D}^T\,. \tag{13}$$

What can be estimated from peri-event panels in Eq. 10, 12 and 13 are the conditional statistics $\mu_{\mathbf{Y}_t|D \geq d_0}$, $\Sigma_{\mathbf{Y}_t|D \geq d}$ (which we can estimate from Eq. 9, and the binned conditions d (which we can specify on our need). What we are interested in recovering, are the unconditional mean $\mu_{\mathbf{Y}_t}$ and covariance matrix $\Sigma_{\mathbf{Y}_t}$. Some intermediate unknown variables that help us estimated the unconditional statistics are $\Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}$, μ_D and c. For a uni-state signals, μ_D and c can be easily obtained by exploiting the distribution of D; however, if the signal is a mixture of multiple states, these statistics are largely unobserved. Actually, these intermediate variables and the unconditional statistics can all be retrieved by performing three linear regressions. First, with the snapshot and a given set of binned d (which must satisfy $d \geq d_0$ but should not be too large to limit the sample size of $P(\mathbf{Y}_t|D = d)$), we can regress d over $\mu_{\mathbf{Y}_t|D=d}$ in Eq. 12 to get the coefficient a_t and the intercept b_t corresponding to:

$$p_t = \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}\,, \tag{14}$$

$$q_t = \mu_{\mathbf{Y}_t} - \Sigma_{\mathbf{Y}_t D}\Sigma_D^{-1}\mu_D\,. \tag{15}$$

Secondly, b_t is a linear function of a_t as $q_t = \mu_{\mathbf{Y}_t} - p_t\mu_D$. Thus we can regress p_t over q_t to estimate the mean of D (μ_D) as the coefficient and $\mu_{\mathbf{Y}_t}$ as the intercept.

Finally, Eq. 13 can be reorganized as:

$$\Sigma_{\mathbf{Y}_t|D \geq d} = \Sigma_{\mathbf{Y}_t} + c p_t p_t^T\,, \tag{16}$$

For a given threshold d_0, $c(d_0)$ is a constant for all elements of the covariance matrix at all time points of the snapshots. Regressing $p_t p_t^T$ over $\Sigma_{\mathbf{Y}_t|D \geq d}$ for any single element across time, we can estimate $c(d_0)$, by which we are able to retrieve $\Sigma_{\mathbf{Y}_t}$ from Eq. 16. Sometimes, as event extraction induces temporal correlations, we can also apply a first order difference in the panel such that $\Delta_t(\Sigma_{\mathbf{Y}_t|D \geq d}) = \Delta_t(\Sigma_{\mathbf{Y}_t}) + c\Delta_t(p_t a_t^T) = c\Delta_t(p_t p_t^T)$. Then, regressing $\Delta_t(p_t p_t^T)$ over $\Delta_t(\Sigma_{\mathbf{Y}_t|D \geq d})$, we can similarly calculate c and retrieve $\Sigma_{\mathbf{Y}_t}$ from Eq. 16.