Detectability of Granger causality for subsampled continuous-time neurophysiological processes

Lionel Barnett[*], Anil K. Seth

Sackler Centre for Consciousness Science and School of Engineering and Informatics,
University of Sussex, Brighton BN1 9QJ, UK

Abstract

Background

Granger causality is well established within the neurosciences for inference of directed functional connectivity from neurophysiological data. These data usually consist of time series which subsample a continuous-time biophysiological process. While it is well known that subsampling can lead to imputation of spurious causal connections where none exist, less is known about the effects of subsampling on the ability to reliably *detect* causal connections which *do* exist.

New Method

We present a theoretical analysis of the effects of subsampling on Granger-causal inference. Neurophysiological processes typically feature signal propagation delays on multiple time scales; accordingly, we base our analysis on a distributed-lag, continuous-time stochastic model, and consider Granger causality in continuous time at finite prediction horizons. Via exact analytical solutions, we identify relationships among sampling frequency, underlying causal time scales and detectability of causalities.

Results

We reveal complex interactions between the time scale(s) of neural signal propagation and sampling frequency. We demonstrate that detectability decays exponentially as the sample time interval increases beyond causal delay times, identify detectability "black spots" and "sweet spots", and show that downsampling may potentially improve detectability. We also demonstrate that the invariance of Granger causality under causal, invertible filtering fails at finite prediction horizons, with particular implications for inference of Granger causality from fMRI data.

Comparison with Existing Method(s)

Our analysis emphasises that sampling rates for causal analysis of neurophysiological time series should be informed by domain-specific time scales, and that state-space modelling should be preferred to purely autoregressive modelling.

Conclusions

On the basis of a very general model that captures the structure of neurophysiological processes, we are able to help identify confounds, and offer practical insights, for successful detection of causal connectivity from neurophysiological recordings.

Keywords: Granger causality, subsampling, continuous-time process, distributed lags

1. Introduction

Neurophysiological recordings are generally obtained by sampling, at regular discrete time intervals, a continuous-time analogue signal associated with some underlying biophysiological processes. Thus, for example, electroencephalography (EEG) records electrical activity arising from ionic current flows in the brain, magnetoencephalography (MEG) records the weak magnetic fields produced by neuronal currents, while functional magnetic resonance imaging (fMRI) measures changes in blood oxygenation level associated with neural activity (Logothetis et al., 2001). Even spike train recordings are typically derived from a continuous analogue measurement of cellular membrane potentials.

Wiener-Granger causality (Wiener, 1956; Granger, 1963, 1969, 1981; Geweke, 1982)—henceforth just Granger causality, or GC—is a popular technique for inferring directed functional connectivity of the underlying process in the neurosciences (Seth et al., 2015), from (discrete-time) subsampled[1]

---

[*]Corresponding author

Email addresses: `l.c.barnett@sussex.acf.uk` (Lionel Barnett), `a.k.seth@sussex.ac.uk` (Anil K. Seth)

---

[1]The term "subsample" refers throughout to sampling of a discrete- or continuous-time process at *regular intervals*. We reserve the term "downsample" for the further subsampling of an already-sampled discrete-time process.

process. Granger causality is premised on a notion of causality whereby cause (a) precedes effect, and (b) contains unique information about effect. This idea is commonly (but not exclusively) operationalised within a vector autoregressive (VAR) modelling framework. At this point, we recognise that the ascription of a "causal" interpretation to GC is clearly problematic to some. Our view is that Granger causality represents *a* rather than *the* notion of causality, an avowedly *statistical*, as opposed, e.g., to "interventionist" notions (Pearl, 2009). As such, its strengths and limitations have been widely discussed [see e.g. Valdes-Sosa et al. (2011) for a review of the issues involved with regard to biophysical modelling; also Chicharro and Panzeri (2014)], and we do not enter that debate here. We remark, however, that Granger causality also has a principled interpretation—through its intimate relationship (Barnett et al., 2009; Barnett and Bossomaier, 2013) with the information-theoretic *transfer entropy* (Schreiber, 2000; Paluš et al., 2001)—as a measure of *information transfer*, and we generally prefer this interpretation (Lizier and Prokopenko, 2010), particularly with regard to functional connectivity analysis.

Problems associated with Granger-causal inference from subsampled (or otherwise aggregated) time series have long been noted (Granger, 1969; Sims, 1971; Wei, 1981; Marcellino, 1999; Breitung and Swanson, 2002). Specifically, it has been observed that subsampling may distort GC values. This may be considered especially problematic in two distinct aspects:

i *Spurious causality*, where GC is absent at the finer time scale, but non-zero for the subsampled process (Comte and Renault, 1996; Renault et al., 1998; Breitung and Swanson, 2002; McCrorie and Chambers, 2006; Solo, 2007, 2016), and

ii *Undetectable causality*, where GC is present at the finer time scale, but zero (or too small to detect reliably) for the subsampled process (Barnett and Seth, 2011; Seth et al., 2013; Zhou et al., 2014).

Subsampling may, in addition, distort the *relative strengths* of causalities (Solo, 2016).

Solo (2007, 2016), drawing on previous work by Caines (1976), distinguishes between the conventional "weak" causality and "strong" causality (see Section 2.1), and concludes that only strong causality remains undistorted by subsampling. Seth et al. (2013) demonstrate that GC inference from fMRI data may be severely degraded by the sample rates, slow in comparison to underlying neural time scales, of fMRI recording technologies. More recently, Zhou et al. (2014) report oscillations in estimated causalities with varying sampling frequency, with causal estimates almost vanishing at some frequencies, as well as inference of spurious causalities.

Although Granger himself was clearly concerned about the detectability problem—in Granger (1969) he notes that "[...] a simple causal mechanism can appear to be a feedback mechanism[2] if the sampling period for the data is so long that details

of causality cannot be picked out"—subsequent studies have concentrated mostly on spurious causality. Here we investigate detectability: specifically, we examine how the relationship between the underlying time scale of causal mechanisms and the sampling time scale mediates the distortion of (non-zero) Granger causalities, and how this distortion impacts on statistical inference of Granger causality from empirical data. We discuss the implications of our results with regard to the successful inference of Granger causalities at the structural (neural) level, from neurophysiological recordings.

1.1. Contributions of this study

A significant feature of the neuronal systems underlying such measurements is the potential range of signal propagation delays due to variation in biophysical parameters such as axonal length, diameter, conduction velocity and myelination (Miller, 1994; Budd and Kisvárday, 2012; Caminiti et al., 2013). Here we model the underlying analogue signal as a stochastic linear autoregression in continuous time. Unlike prevailing continuous-time stochastic process models in the neurosciences, our model accommodates distributed lags on arbitrary time scales, and is thus able to reflect variability of signal propagation delays. This leads, via consideration of prediction at *finite* time horizons, to a novel and intuitive definition of Granger causality at multiple time scales for continuous-time processes. In contrast to previous work on continuous-time Granger causality, in which various statistical (non)causality test criteria have been proposed, our definition is *quantitative*, furnishing a Granger-Geweke measure with an information-theoretic interpretation.

Using discrete-time VAR modelling, we then analyse the properties of processes obtained by subsampling the temporally multiscale continuous-time process, and relate the spectral and causal properties of the subsampled process to those of the underlying continuous-time model. Having defined continuous-time, finite-horizon GC—which represents a target for statistical analysis—we investigate the extent to which it may be inferred, and in particular *detected*, by discrete-time VAR analysis of the subsampled processes.

We focus on the practical questions of the feasibility and reliability of causal inference on sampling frequency and the (dominant) time scale of causal feedback in the generative process. We investigate in detail the relationship between sampling frequency and the quality of causal inference via a fully analytic solution of a minimal, but non-trivial, bivariate model in continuous time, with finite causal delay.

On the basis of our theoretical and empirical analysis, we identify critical relationships between causal delay, sampling interval and detectability of Granger causality. These include exponential decay of subsampled Granger causalities with increasing sampling interval, resonance between sampling frequency and causal delay frequency, potential detectability

---

[2]Here, by "feedback mechanism", Granger refers to *contemporaneous* feed-

back between time series [Geweke (1982) terms this "instantaneous feedback"], as opposed to *time-delayed* feedback, which in his theory underpins "causal mechanism"; see Section 2.1 for details.

"black spots", and the existence of a non-zero optimal sampling interval (i.e., detectability may sometimes be improved by downsampling). We also discover a hitherto unremarked non-invariance of finite-horizon/multistep GC under causal, invertible filtering (in contrast with the known invariance of single-step discrete-time GC).

Finally, we discuss the implications of our findings for Granger-causal inference of neural functional relationships from neurophysiological recordings under various technologies - including fMRI, which continues to generate controversy.

1.2. Organisation

The paper is organised as follows: in Section 2 we review essential aspects of the theory of VAR processes and Granger causality in discrete time. In Section 3 we introduce CTVAR (continuous-time vector autoregressive) processes as continuous-time, distributed-lag analogues of discrete-time VAR processes, and derive a principled extension of Granger causality to such processes, based on finite-temporal horizon prediction. We analyse discrete-time processes derived by sub-sampling a CTVAR process, and demonstrate the consistency of GC in the limit as the subsampling interval shrinks to zero. In Section 4 we present a detailed analytic solution of the sub-sampling problem for a non-trivial minimal bivariate CTVAR process with finite causal delay, and address the issue of statistical inference (detectability) for GC. Lastly, in Section 5 we discuss the implications of our results presented in the setting of analysis of neurophysiological data. Technical details, where they would detract from the narrative flow, are presented in Appendices.

1.3. Notation and conventions

The principal objects of study in this paper are random vectors in a real Euclidean space $\mathbb{R}^n$ and vector stochastic processes; i.e. sequences of random vectors in discrete or continuous time. Time sequences are generally written as $x = \{x_k \mid k \in \mathbb{Z}\}$ ($\mathbb{Z}$ denotes the set of integers) and $x = \{x(t) \mid t \in \mathbb{R}\}$ in discrete and continuous time respectively, where the x_k or $x(t)$ could be real or complex, random or deterministic scalars, vectors, matrices, etc.; note that when we refer to a sequence as a whole, we shall frequently drop the time index/variable. Vectors in $\mathbb{R}^n$ are generally written in bold type and random variables in upper case; thus, e.g., a vector stochastic process in discrete time is generally represented as $\{X_k \mid k \in \mathbb{Z}\}$, and the entire process referred to simply as X. For avoidance of consideration of initial conditions, process time (discrete or continuous) is assumed to extend into the infinite past.

Time is assumed measured in a standard unit, which we take to be milliseconds (ms). For discrete-time sequences, we require that a sample interval (time step) Δ be specified [equivalently, a sampling frequency $f_s \equiv 1/\Delta$, measured in kilohertz (kHz)]. To emphasize the dependence of a quantity on sample interval, Δ is included as a function argument. In particular, this study is concerned with the *regular subsampling* of continuous-time processes. If $x = \{x(t) \mid t \in \mathbb{R}\}$ is a (random or deterministic, scalar, vector, etc.) continuous-time sequence,

we write $x(\Delta) \equiv \{x(k\Delta) \mid k \in \mathbb{Z}\}$ for the *discrete*-time sequence obtained by sampling x at regular intervals Δ, which we refer to as a Δ-*subsampling* of x.

Much of the analysis presented here takes place naturally in the spectral domain. Ordinary frequencies are generally written as $-\infty < \lambda < \infty$, measured in kHz. In discrete time with sampling interval Δ, spectral quantities are periodic in λ with period $f_s = 1/\Delta$, the sampling frequency; we shall sometime restrict such quantities to the interval $-1/(2\Delta) \leq \lambda < 1/(2\Delta)$, where $1/(2\Delta) = f_s/2$ is the *Nyqvist frequency*. For continuous-time sequences, spectral quantities are not generally periodic. For discrete-time spectral quantities, it is sometimes convenient to work instead with the *angular frequency* $\omega \equiv 2\pi\Delta\lambda$. We may then consider spectral quantities as defined on the unit circle $|z| = 1$ in the complex plane $\mathbb{C}$, with $z = e^{-i\omega}$, $-\pi \leq \omega < \pi$. In continuous time we occasionally use a normalised frequency $\omega \equiv 2\pi\lambda$, and spectral quantities may be considered defined on the imaginary line $\Re\mathfrak{e}(\zeta) = 0$ in the complex plane, with $\zeta = i\omega$, $-\infty < \omega < \infty$. In the time domain, z and ζ may be interpreted as *lag (backshift) operators* in discrete and continuous time respectively.

The link between time and frequency domains is the *Fourier transform*. Here, Fourier transforms are always defined in terms of ordinary frequency λ and indicated with a hat symbol over over the corresponding sequence specifier; e.g. $\hat{x}(\lambda)$, or just $\hat{x}$ when the entire transform is referenced. In light of the proliferation of conventions, our definitions for Fourier transforms are set out in Appendix A; discrete-time transforms are scaled by the sample time step Δ in order to ensure that dimensions are always the same in discrete and continuous time, and that, in particular, limiting values for Δ-subsampled continuous-time sequences tend to corresponding continuous-time values as $\Delta \to 0$; see Appendix A for details. Generally, we take care to scale by sampling interval so that (almost) all measurable quantities (Table 1) have the same dimensions in discrete and continuous time, and comparisons of magnitudes are thus meaningful, in particular in the limit $\Delta \to 0$. Although this convention may appear cumbersome, particularly in our analysis of discrete-time processes (Section 2), the payoff is a more harmonious and intuitive tie-in with the continuous-time case and subsampling analysis (Section 3).

Throughout, superscript "⊤" denotes matrix transpose, superscript "∗" matrix conjugate transpose and $|\cdot|$ the determinant of a square matrix. A dot over a symbol denotes differentiation with respect to a continuous time parameter.

2. Discrete VAR processes and Granger Causality

We briefly outline the VAR theory that we shall require; the reader is referred to standard texts (Hamilton, 1994; Lütkepohl, 2005) for details. Let $X \equiv \{X_k \mid k \in \mathbb{Z}\}$ be a discrete-time, purely nondeterministic, zero-mean, wide-sense stationary vector process. With a view to Granger-causal analysis, we assume the same conditions on the process X as in Geweke (1982). By Wold's Theorem (Wold, 1938; Lütkepohl, 2005) X has a *vector*

quantity	description	unit
$k, \ell, m, \dots$	discrete time index	1
$t, u, h, \dots$	continuous time parameter	t
Δ	sample interval	t
λ	ordinary frequency	t^{-1}
ω	angular frequency	rad.
$X_k, X(t)$	stochastic process	x
Σ	residual noise intensity	$x^2 t^{-1}$
$\mathcal{E}_m, \mathcal{E}(h)$	mean-square prediction error (MSPE)	x^2
$\Gamma_k, \Gamma(t)$	autocovariance sequence/funtion	x^2
$H(\lambda)$	transfer function	t
$S(\lambda)$	cross-power spectral density (CPSD)	$x^2 t$
$\mathcal{F}_{Y \to X}$	Granger causality (time domain)	b
$f_{Y \to X}(\lambda)$	Granger causality (frequency domain)	b

Table 1: Notation and dimensions: t denotes units of time (e.g., ms, so that t^{-1} is measured in kHz), x denotes the units of the neural signal under consideration (e.g., volts, tesla, etc.) and b denotes the unit of information (bits or nats, depending on whether base 2 or natural logarithms are used).

moving-average (VMA) representation

$$X_k = \sum_{\ell=0}^{\infty} B_\ell \boldsymbol{\varepsilon}_{k-\ell} \qquad (1)$$

with square-summable coefficient matrices B_ℓ ($B_0 = I$), and $\boldsymbol{\varepsilon}$ a white-noise process. Considering z as the *lag operator* $z \cdot X_k = X_{k-1}$, we can write (1) in compact form as

$$X_k = \Psi(z) \cdot \boldsymbol{\varepsilon}_k \qquad (2)$$

where the MA operator

$$\Psi(z) \equiv \sum_{\ell=0}^{\infty} B_\ell z^\ell \qquad (3)$$

has the minimum-phase property that $|\Psi(z)| \neq 0$ for complex z on the unit disc $|z| \leq 1$. We shall require that the VMA representation (1) may be inverted to yield a *vector autoregressive* (VAR) representation

$$X_k = \sum_{\ell=1}^{\infty} A_\ell X_{k-\ell} + \boldsymbol{\varepsilon}_k \qquad (4)$$

with square-summable coefficient matrices A_ℓ. Geweke (1982) supplies a condition on the spectrum of X (see below)—that it is bounded away from zero almost everywhere—which suffices for invertibility of the VMA representation (Rozanov, 1967). The condition (see Geweke, 1982, eq. 2.4), which we assume here, also guarantees that any vector *sub*process of X has a VAR representation. We may write the VAR (4) as

$$\Phi(z) \cdot X_k = \boldsymbol{\varepsilon}_k \qquad (5)$$

with

$$\Phi(z) \equiv I - \sum_{\ell=1}^{\infty} A_\ell z^\ell \qquad (6)$$

Stability of the process requires that $|\Phi(z)| \neq 0$ on $|z| \leq 1$, and we have $\Phi(z) = \Psi(z)^{-1}$ on $|z| \leq 1$. Henceforth, by "VAR process" we mean a process satisfying all of the above conditions.

In accordance with the conventions described in the Introduction, we assume a sample time step Δ is always given, and parametrise the magnitude of the residual noise $\boldsymbol{\varepsilon}$ by its *intensity* $\Sigma \equiv \Delta^{-1} \operatorname{cov}(\boldsymbol{\varepsilon}_k)$. This is consistent with the additivity of variance, and ensures that the dimensions of Σ are consistent with the corresponding quantity in the continuous-time processes we shall encounter later (Section 3).

The *autocovariance sequence* $\Gamma \equiv \{\Gamma_k \mid k \in \mathbb{Z}\}$ is given by

$$\Gamma_k \equiv \operatorname{cov}(X_{k+\ell}, X_\ell) \qquad (7)$$

(by stationarity, this does not depend on ℓ) and satisfies $\Gamma_{-k} = \Gamma_k{}^\mathsf{T}$, and the *Yule-Walker equations*

$$\Gamma_k = \sum_{\ell=1}^{\infty} A_\ell \Gamma_{k-\ell} + \delta_{k0} \Delta \Sigma \qquad k \geq 0 \qquad (8)$$

In terms of the VMA coefficients, it is straightforward to show that

$$\Gamma_k = \Delta \sum_{\ell=0}^{\infty} B_{k+\ell} \Sigma B_\ell{}^\mathsf{T} \qquad k \geq 0 \qquad (9)$$

The *cross-power spectral density* (CPSD) S for X is defined for $-\infty < \lambda < \infty$ by

$$S(\lambda) \equiv \lim_{K \to \infty} \frac{1}{2K\Delta} \operatorname{cov}(\widehat{X}_K(\lambda)) \qquad (10)$$

where $\widehat{X}_K(\lambda) \equiv \Delta \sum_{k=-K}^{K} X_k e^{-2\pi i \Delta \lambda k}$ is the truncated Fourier transform of the process X. The $S(\lambda)$ are Hermitian matrices and the *Wiener-Khintchine Theorem* (Wiener, 1930; Khintchine, 1934) states that:

$$S(\lambda) = \widehat{\Gamma}(\lambda) \qquad (11)$$

at all frequencies; i.e. the CPSD is the Fourier transform of the autocovariance sequence.

The *transfer function* for the VAR (4) is defined to be the Fourier transform of the MA coefficients:

$$H(\lambda) \equiv \widehat{B}(\lambda) = \Delta \Psi\left(e^{-2\pi i \Delta \lambda}\right) = \Delta \sum_{\ell=0}^{\infty} B_\ell e^{-2\pi i \Delta \lambda \ell} \qquad (12)$$

which may also be written as

$$H(\lambda) = \Delta \Phi\left(e^{-2\pi i \Delta \lambda \ell}\right)^{-1} = \Delta \left(I - \sum_{\ell=1}^{\infty} A_\ell e^{-2\pi i \Delta \lambda \ell}\right)^{-1} \qquad (13)$$

We then have the *spectral factorisation* formula (Masani, 1966)

$$S(\lambda) = H(\lambda) \Sigma H(\lambda)^* \qquad (14)$$

which holds for all λ. A classical result states that, given a CPSD $S(\lambda)$ satisfying certain regularity conditions (Masani, 1966; Wilson, 1972), there exists a unique $\Psi(z)$ holomorphic on the disc $|z| \leq 1$ with $\Psi(0) = I$, and a unique positive-definite

symmetric matrix Σ, such that setting $H(\lambda) = \Delta\Psi(e^{-2i\pi\Delta\lambda})$, (14) is satisfied. In other words, for a class of CPSDs the spectral factorisation (14) is *uniquely solvable* for $H(\lambda)$ and Σ, and hence parameters for a VAR model with the given CPSD may be obtained. Although this result is not constructive—there is no known algorithm for analytic factorisation of an arbitrary CPSD—in specific cases, in particular for *rational* spectral densities (Kučera, 1991)], it is frequently feasible; see Section 4.2 and Appendix J for a concrete, nontrivial example.

A VAR of the form (4) is equivalently specified by the VAR parameters (A, Σ), the autocovariance sequence Γ or the CPSD S. The Yule-Walker equations (8), Wiener-Khintchine Theorem (11) and spectral factorisation formula (14) establish reciprocal relationships between the respective representations. Barnett and Seth (2014) exploit these relationships to design efficient computational pathways for the numerical computation of Granger causalities (Section 2.1 below). Analytically, we are also free to choose the representation appropriate to the task at hand.

2.1. Granger Causality

Granger causality is most commonly framed in terms of *prediction*. Usually, only 1-step-ahead prediction is considered. Here, for reasons that will become clear (Section 3.3), we consider Granger causality at arbitrary prediction horizons (Lütkepohl, 1993; Dufour and Renault, 1998). The optimal (in the least-squares sense) m-step-ahead prediction ($m = 1, 2, \ldots$) of the stable VAR (4) based on all information contained in its own (infinite) past—i.e., the optimal prediction of X_{k+m} given the history $X_k^- \equiv \{\ldots, X_{k-2}, X_{k-1}, X_k\}$ of the process up to and including the kth step—is given by the orthogonal projection $\mathbf{E}\left[X_{k+m} \mid X_k^-\right]$ of X_{k+m} onto X_k^-. A standard result (Hamilton, 1994) states that

$$\mathbf{E}\left[X_{k+m} \mid X_k^-\right] = \sum_{\ell=m}^{\infty} B_\ell \boldsymbol{\varepsilon}_{k+m-\ell} \tag{15}$$

It follows that the *mean-square prediction error* (MSPE) at a prediction horizon of time $m\Delta$ into the future, is given by

$$\mathcal{E}_m \equiv \mathrm{cov}\left(\mathbf{E}\left[X_{k+m} \mid X_k^-\right] - X_{k+m}\right) = \Delta \sum_{\ell=0}^{m-1} B_\ell \Sigma B_\ell^{\mathsf{T}} \tag{16}$$

In particular, $\mathcal{E}_1$ is just the residual noise covariance $\Delta\Sigma$, while from (9), $\mathcal{E}_m \to \Gamma_0$ as $m \to \infty$; it also makes sense to define $\mathcal{E}_0 \equiv 0$, as prediction at zero horizon is exact.

Our exposition of Granger causality follows in spirit the standard formulation of Geweke (1982). Suppose that we have two jointly distributed discrete-time vector stochastic processes ("variables") X, Y so that the joint process $[X^{\mathsf{T}} Y^{\mathsf{T}}]^{\mathsf{T}}$ is a VAR of the form (4). By assumption, the subprocesses X and Y also have VAR representations. We may then, for a given prediction horizon $m\Delta$ ($m = 1, 2, \ldots$), compare the MSPE $\mathcal{E}_{m,xx}$ of the prediction $\mathbf{E}\left[X_{k+m} \mid X_k^-, Y_k^-\right]$ of X_{k+m} based on the *joint* history of X and Y (the "full regression"), with the MSPE $\mathcal{E}'_{m,xx}$ of the prediction $\mathbf{E}\left[X_{k+m} \mid X_k^-\right]$ of X_{k+m} based only on the *self*-history

of the subprocess X (the "reduced regression"[3]). If inclusion of the history Y_k^- improves the prediction of X_{k+m}, then we say that Y (the "source" variable) *Granger-causes* X (the "target" variable) at prediction horizon $m\Delta$. Geweke (1982) proposed that prediction be quantified by *generalised variance*[4] (Wilks, 1932)—that is, the determinant of the MSPE—leading to the definition

$$\mathcal{F}_{Y \to X, m} \equiv \log \frac{\left|\mathcal{E}'_{m,xx}\right|}{\left|\mathcal{E}_{m,xx}\right|} \tag{17}$$

$\mathcal{F}_{X \to Y, m}$ is defined symmetrically.

If $m = 1$ (1-step prediction), we drop the m subscript. Note that $\mathcal{E}_{1,xx} = \Delta\Sigma_{xx}$, where Σ_{xx} is the xx component of the noise intensity Σ of the joint process, while $\mathcal{E}'_{1,xx} = \Delta\Sigma'_{xx}$ where Σ'_{xx} is the noise intensity of the subprocess X, considered as a VAR. Thus we obtain the standard 1-step Geweke measure

$$\mathcal{F}_{Y \to X} \equiv \mathcal{F}_{Y \to X, 1} = \log \frac{\left|\Sigma'_{xx}\right|}{\left|\Sigma_{xx}\right|} \tag{18}$$

$\mathcal{F}_{Y \to X, m} \geq 0$ always, since inclusion of the history Y_k^- in the full regression can only decrease the prediction error. It may be shown, furthermore (Sims, 1972; Caines, 1976), that $\mathcal{F}_{Y \to X} = 0$ iff $\Psi_{xy}(z) \equiv 0$. But from (2) it follows that $\Psi_{xy}(z) \equiv 0 \implies \Psi'_{xx}(z) = \Psi_{xx}(z)$, and, since all the B_ℓ are lower block-triangular, from (16) we have $\mathcal{E}'_{m,xx} = \mathcal{E}_{m,xx}$ for any m. Thus we may state:

$$\mathcal{F}_{Y \to X} = 0 \iff \Psi_{xy}(z) \equiv 0 \iff \mathcal{F}_{Y \to X, m} = 0 \; \forall m > 0 \tag{19}$$

That is, *vanishing* 1-*step GC implies vanishing GC at any prediction horizon*. The converse does not hold, though: $\mathcal{F}_{Y \to X, m}$ may vanish for $m > 1$ even if $\mathcal{F}_{Y \to X} > 0$ (*cf.* Appendix C). In general, $\mathcal{F}_{Y \to X, m}$ will depend on the prediction horizon. Since both $\mathcal{E}_{m,xx}$ and $\mathcal{E}'_{m,xx} \to \Gamma_{0,xx}$ as $m \to \infty$, $\mathcal{F}_{Y \to X, m} \to 0$ as $m \to \infty$, so that $\mathcal{F}_{Y \to X, m}$ attains a maximum at some finite value(s) of m.

At this point we note, as alluded to in the Introduction (Section 1), that Granger causality has a clear information-theoretic interpretation: Barnett et al. (2009) show that for Gaussian processes, Granger causality is entirely equivalent to the non-parametric information-theoretic *transfer entropy* measure (Schreiber, 2000; Paluš et al., 2001), and for general Markovian processes (under a mild ergodicity assumption) the log-likelihood ratio statistic for the Markov model [*cf.* (D.3)] converges in the large-sample limit to the corresponding transfer entropy (Barnett and Bossomaier, 2013). TE, as a conditional mutual information—and by extension GC—is naturally measured in units of bits (or nats, if natural logarithms are used).

Regarding Granger's other requirement for causal effect, that the information that Y contains about (the future of) X be

---

[3]We generally indicate quantities associated with the reduced, as opposed to full regression, with a prime.

[4]For a discussion on the preferability of the generalised variance $|\Sigma|$ over the *total variance* trace $[\Sigma]$, see Barrett et al. (2010); see also the maximum likelihood interpretation outlined in Appendix D.

unique, here we note just that the effect on Granger causalities of other (accessible) variables jointly distributed with X and Y may be discounted by including them in both the full and reduced predictor sets[5]. This leads to the definition of *conditional* Granger causality (Geweke, 1984). While all Granger causalities discussed in this paper have conditional counterparts, here we restrict our attention to the unconditional case.

Geweke (1982) refers to the Granger causality $\mathcal{F}_{Y \to X}$ as the "linear feedback" from Y to X, and goes on to define the "instantaneous feedback" or *instantaneous causality*:

$$\mathcal{F}_{X \cdot Y} \equiv \log \frac{|\Sigma_{xx}| \, |\Sigma_{yy}|}{|\Sigma|} \qquad (20)$$

which vanishes iff the residuals $\varepsilon_{x,k}, \varepsilon_{y,k}$ are contemporaneously uncorrelated. Solo (2007) distinguishes "strong" Granger causality from the conventional ("weak") variety, noting that only (the existence of) strong causality is strictly preserved under subsampling. In a VAR framework, strong causality from $Y \to X$ replaces the full (1-step) predictor set X_k^-, Y_k^- with the predictor set X_k^-, Y_{k+1}^- [*cf.* Geweke (1982, eq. 2.9)]; that is, the contemporaneous source term Y_{k+1} is included in the full predictor set (the reduced predictor set remains unaltered). The residual errors of the strong least-squares prediction are $\varepsilon_{x,k} - \Sigma_{xy}\Sigma_{yy}^{-1}\varepsilon_{y,k}$ (Geweke, 1982), so that the MSPE is $\Delta \times$ the *partial* residual noise intensity matrix

$$\Sigma_{xx|y} \equiv \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \qquad (21)$$

This leads to the statistic

$$\mathcal{F}_{Y \to X}^{\text{strong}} \equiv \log \frac{\left| \Sigma_{xx}' \right|}{\left| \Sigma_{xx|y} \right|} = \mathcal{F}_{Y \to X} + \mathcal{F}_{X \cdot Y} \qquad (22)$$

(the last equality follows from block-decomposition of the determinant $|\Sigma|$). Strong GC, while invariant under subsampling is, however, unsatisfactory as a *directional* measure, since it is not generally possible to disentangle the directional and instantaneous contributions.

Although not the focus of this paper, a significant feature of Granger-causal analysis is that (time domain) GC may be decomposed in a natural way by frequency. The resulting frequency-domain, or *spectral* Granger causality integrates to the time-domain GC (18). For a full derivation and discussion we refer to Geweke (1982); here we just present the definition of the spectral GC from Y to X:

$$f_{Y \to X}(\lambda) \equiv \log \frac{|S_{xx}(\lambda)|}{\left| S_{xx}(\lambda) - H_{xy}(\lambda)\Sigma_{yy|x}H_{xy}(\lambda)^* \right|} \qquad (23)$$

where $\Sigma_{yy|x} \equiv \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$ [*cf.* (21)]. $f_{Y \to X}(\lambda)$ is always non-negative, and Geweke's fundamental spectral decomposition of

Granger causality applies[6]

$$\mathcal{F}_{Y \to X} = \Delta \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} f_{Y \to X}(\lambda) \, d\lambda \qquad (24)$$

The spectral GC (23) is, at any specific frequency λ, also a quantity of information measured in bits or nats, and (24) presents time-domain GC as an average over all frequencies of spectral GC. We remark that a spectral counterpart for $\mathcal{F}_{X \cdot Y}$ has been defined (Ding et al., 2006), but is somewhat unsatisfactory insofar as it may become negative at some frequencies and lacks a compelling physical interpretation.

It is known (Geweke, 1982; Barnett and Seth, 2011; Solo, 2016) that 1-step Granger causality, in both time and frequency domains, is invariant under (almost) arbitrary causal, invertible (stable, minimum-phase[7]) filtering; see Appendix B for more detail. However, as demonstrated in Appendix C, invariance does *not* extend to *m*-step GC for $m > 1$, unless $\Psi_{xy}(z) \equiv 0$ - equivalently, $\mathcal{F}_{Y \to X} = 0$. In that case filter-invariance does hold for $m \geq 1$; that is, causal, invertible filtering will not induce *spurious* Granger causalities at any prediction horizon.

VAR modelling is particularly suited to data-driven approaches to functional analysis (Appendix D) and its applicability quite general. By the Wold decomposition theorem (Hannan, 1970), any covariance-stationary stochastic process in discrete time has a moving-average (MA) representation. Further spectral conditions may be imposed so that the Wold MA representation may be inverted to yield a stable VAR representation (4) (Rozanov, 1967). We assume that these conditions apply for all discrete stochastic processes encountered in this study. We note that if there is nonlinear (delayed) feedback in the generative process, while this does not preclude VAR-based estimation of Granger causalities (provided the VAR representation criteria mentioned above hold), a linear model will not be parsimonious and transfer entropy (or a suitable nonlinear model-based version of Granger causality) may be preferable (Barnett and Bossomaier, 2013). In general, though, VAR-based Granger causality has the advantages of simplicity, ease of estimation, a known sampling distribution and a natural spectral decomposition.

More recently, a theory of Granger causality has been developed for *state-space* processes (Barnett and Seth, 2015; Solo, 2016). The state-space approach offers some significant advantages from modelling, estimation and computational perspectives. This, as well as estimation, statistical inference and detection of (discrete-time) Granger causality from empirical time series data, is discussed in Appendix D.

---

[5]Of course this is only possible for *accessible* variables - inaccessible (hidden, latent) influences are in general problematic for causal analysis in a broader sense (Valdes-Sosa et al., 2011).

[6]Strictly speaking, equality in (24) holds provided the condition $\left| A_{yy}(z) - \Sigma_{yx}\Sigma_{xx}^{-1}A_{xy}(z) \right| \neq 0$ is satisfied for all z on the unit disc $|z| \leq 1$; otherwise it should be replaced by $\leq$. In practice, according to Geweke (1982), the equality condition is "almost always" satisfied.

[7]Minimum phase requires that the inverse filter also be stable; note that in Barnett and Seth (2011) this requirement is erroneously overlooked [thanks to Victor Solo (personal communication) for bringing this to our attention].

3. Distributed-lag vector autoregressive processes in continuous time

Following the discussion in Section 1 regarding the essentially continuous-time nature of biophysical processes, in order to address the impact of subsampling we require appropriate continuous-time generative processes for which Granger causality may be defined. Accordingly, we start with an underlying analogue neurophysiological process $U(t)$ in continuous time[8] t and an *observation function* $\xi(\cdot)$. The observed (multivariate) signal $X(t) = \xi(U(t))$ is then sampled at regular discrete time intervals. $U(t)$ may be stochastic (endogenous noise), as may be the observation function (exogenous, measurement noise), so that $X(t)$ is considered a continuous-time stochastic process. Our approach is to assume that $X(t)$ admits a continuous-time linear autoregressive representation.

The standard multivariate linear autoregressive model in continuous time is the vector Ornstein-Uhlenbeck (VOU) process (Uhlenbeck and Ornstein, 1930; Doob, 1953) defined by a linear stochastic differential equation (SDE)

$$dX(t) = AX(t)\,dt + dW(t) \qquad (25)$$

where $W(t)$ is a vector Wiener process. The process (25) must, however, be considered implausible as a model for an observed neurophysiological processes, since it fails to model delayed feedback at *finite* time scales. To address this we generalise the VOU process to the CTVAR (continuous-time vector autoregressive) process described below.

Our construction closely mirrors that of the discrete-time VAR case (Section 2). Thus we assume that the continuous-time, wide-sense stationary, stable, minimum-phase (and zero-mean) vector process $X \equiv \{X(t) \,|\, t \in \mathbb{R}\}$ admits a moving-average representation (Caines and Chan, 1975; Comte and Renault, 1996)

$$X(t) = \int_{u=0}^{\infty} B(u)\,dW(t-u) \qquad (26)$$

where $W(t)$ is again a vector Wiener process[9], the MA kernel $B(u)$ [with $B(0) = I$] is square-integrable and the integral is to be interpreted as an *Itō integral* (Øksendal, 2003). In continuous time we define the lag operator ζ as follows: suppose that a complex-valued function $\mathcal{L}(\zeta)$ may be written (uniquely) as a Laplace transform $\mathcal{L}(\zeta) = \int_0^{\infty} L(u)\,e^{-\zeta u}\,du$. Then for a continuous-time process $U(t)$, we define $\mathcal{L}(\zeta) \cdot U(t)$ as the Itō integral $\int_{u=0}^{\infty} L(u)\,dU(t-u)$, and (26) may be written as

$$X(t) = \Psi(\zeta) \cdot W(t) \qquad (27)$$

where

$$\Psi(\zeta) \equiv \int_0^{\infty} B(u)\,e^{-\zeta u}\,du \qquad (28)$$

The minimum-phase property requires that $|\Psi(\zeta)| \neq 0$ on the right half-plane $\Re\mathfrak{e}(\zeta) \geq 0$.

As in the discrete-time case, we assume that the MA representation (26) may be inverted to yield a continuous-time vector autoregressive (CTVAR) representation as a stochastic linear integro-differential equation (Comte and Renault, 1996)

$$dX(t) = \int_0^{\infty} A(u)X(t-u)\,du\,dt + dW(t) \qquad (29)$$

or

$$\Phi(\zeta) \cdot X(t) = W(t) \qquad (30)$$

with

$$\Phi(\zeta) \equiv \zeta I - \int_0^{\infty} A(u)e^{-\zeta u}\,du \qquad (31)$$

To verify (30,31), note that $\Phi(\zeta)$ may be written as the Laplace transform of $\dot{\delta}(u)I - A(u)$, where $\dot{\delta}(u)$ denotes the derivative, in the generalised function sense, of the delta function $\delta(u)$; (30) then follows from the relation $\int_{-\infty}^{\infty} \dot{\delta}(u)\,\varphi(t-u)\,du = \dot{\varphi}(t)$ for any function $\varphi(u)$ (Friedlander and Joshi, 1998). Stability requires that $|\Phi(\zeta)| \neq 0$ on the right half-plane $\Re\mathfrak{e}(\zeta) \geq 0$, and in Appendix E we prove that $\Psi(\zeta) = \Phi(\zeta)^{-1}$ on the right half-plane $\Re\mathfrak{e}(\zeta) \geq 0$

The AR kernel $A(u)$ specifies causal, time-lagged coupling between nodes over a range of feedback delays u, while $dW(t)$ represents continuous-time white noise with (positive-definite) covariance matrix $\Sigma\,dt$, so that Σ again represents residual noise *intensity*. We assume $A(u)$ be be *square-integrable*[10] and allow it to be a *generalised function* (Friedlander and Joshi, 1998), so it might, for example, include delta functions. The integral over u in (29) is then taken to be a Lebesgue integral. Note that the VOU process (25) is a special case of (29) with $A(u) = A\delta(u)$ a delta function at "infinitesimal lag" $u = 0$. Analogous to the discrete-time case, we also assume that any vector *sub*-process of $X(t)$ may be represented as a CTVAR[11].

Stochastic integro-differential equations similar to (29) have been studied *in abstracto*, as models for various physical, engineering and biological phenomena, and (more along the present lines) in the econometrics literature (Sims, 1971; Geweke, 1978; McCrorie and Chambers, 2006). They have not however, as far as we are aware, been deployed previously in the neurosciences. Our approach most closely resembles the "CIMA" processes presented in Comte and Renault (1996); our emphasis, however, is more on the autoregressive and (as we shall see later) *predictive* aspects of the model.

In Appendix F we show that the MA kernel $B(u)$ satisfies

$$\dot{B}(u) = \int_0^u A(s)B(u-s)\,ds \qquad u \geq 0 \qquad (32a)$$

$$B(0) = I \qquad (32b)$$

where $\dot{B}(u)$ denotes differentiation from the right[12] with respect to t.

---

[8]The unit of time is taken to be the same as for discrete-time processes.

[9]This might be generalised to continuous-time white noise processes as defined for the continuous-time Wold decomposition theorem (Rozanov, 1967).

[10]This condition may be unnecessarily restrictive; we require at least that the CPSD of $X(t)$ (see below) exists (Lighthill, 1958).

[11]It seems plausible, although we have not established this rigorously, that this may follow from a similar boundedness condition on the CPSD to that described in Geweke (1982).

[12]We shall generally assume that appropriate derivatives exist wherever they appear in a formula.

The autocovariance function of a stationary continuous-time vector stochastic process is defined as

$$\Gamma(t) \equiv \mathrm{cov}(X(t+u), X(u)) \qquad (33)$$

which again, by stationarity, does not depend on u, and $\Gamma(-t) = \Gamma(t)^\mathsf{T}$. From the MA representation (27) an application of the *Itō isometry* (Øksendal, 2003) yields

$$\Gamma(t) = \int_0^\infty B(t+u)\Sigma B(u)^\mathsf{T}\, du \qquad t \geq 0 \qquad (34)$$

From (34) we may derive the continuous-time Yule-Walker equations for the process (29)

$$\dot\Gamma(t) = \int_0^\infty A(u)\Gamma(t-u)\, du \qquad t > 0 \qquad (35a)$$

$$\dot\Gamma(0) + \dot\Gamma(0)^\mathsf{T} = -\Sigma \qquad (35b)$$

where $\dot\Gamma(t)$ denotes differentiation from the right.

Analogous to the discrete-time case, the CPSD for the process is defined by

$$S(\lambda) \equiv \lim_{T\to\infty} \frac{1}{2T}\,\mathrm{cov}(\widehat{X}_T(\lambda)) \qquad (36)$$

on $-\infty < \lambda < \infty$, now with $\widehat{X}_T(\lambda) \equiv \int_{-T}^T X(t)e^{-2\pi i\lambda t}\, dt$ [*cf.* (10)], and the Wiener-Kintchine Theorem in continuous time again reads:

$$S(\lambda) = \widehat{\Gamma}(\lambda) \qquad (37)$$

The continuous-time transfer function is again defined as

$$H(\lambda) \equiv \widehat{B}(\lambda) = \Psi(2\pi i\lambda) = \int_0^\infty B(u)e^{-2\pi i\lambda u}\, du \qquad (38)$$

which may also be written as

$$H(\lambda) = \Phi(2\pi i\lambda)^{-1} = \left(2\pi i\lambda I - \int_0^\infty A(u)e^{-2\pi i\lambda u}\, du\right)^{-1} \qquad (39)$$

and from (34) it is not hard to establish the continuous-time spectral factorisation[13]

$$S(\lambda) = H(\lambda)\Sigma H(\lambda)^* \qquad (40)$$

We note that $H(\lambda)$ satisfies

$$\lim_{|\lambda|\to\infty} 2\pi i\lambda\, H(\lambda) = I \qquad (41)$$

so that from (40)

$$\lim_{|\lambda|\to\infty} 4\pi^2\lambda^2\, S(\lambda) = \Sigma \qquad (42)$$

i.e., $S(\lambda)$ decays as λ^{-2}, as $|\lambda| \to \infty$. We conjecture that, analogous to the discrete-time case, given a continuous-time CPSD $S(\lambda)$ satisfying suitable regularity conditions, there exists a unique $\Psi(\zeta)$ holomorphic on $\mathfrak{Re}(\zeta) \geq 0$ with $\lim_{|\omega|\to\infty} i\omega\Psi(i\omega) = I$ and positive-definite Σ, such that (40) is satisfied for $H(\lambda) = \Psi(2\pi i\lambda)$ on $-\infty < \lambda < \infty$.

3.1. Subsampling a CTVAR process

We next examine some properties of the discrete-time processes $X(\Delta)$ obtained by subsampling a CTVAR process X at fixed time intervals Δ; i.e., $X_k(\Delta) \equiv X(k\Delta)$. It is these Δ-subsampled processes which stand as models for discretely-sampled neurophysiological recordings of an underlying bio-physiological process (Section 1). A subtlety which we must address is that, while a Δ-subsampling of a (stable, minimum-phase) CTVAR is itself always stable, there is no guarantee that it will be minimum-phase for all Δ - see e.g., Åström et al. (1984). We thus assume that the minimum-phase condition for Δ-subsamplings holds as necessary [in worked examples (*cf.* Section 4.2) it must be tested explicitly], and that in particular (*cf.* Section 3.3 below) it holds in the limit of fine subsampling; that is, there exists a sampling interval Δ_0 such that for any Δ-subsampling with $0 < \Delta \leq \Delta_0$, $X(\Delta)$ is minimum phase.

With a view to calculation of (multistep) Granger causalities (Section 2.1), we require expressions for the transfer function, residual noise intensity, MA coefficients and CPSD of the Δ-subsampled processes. The crucial observation is that the autocovariance sequence $\Gamma(\Delta)$ of the subsampled process $X(\Delta)$ is just $\Gamma_k(\Delta) = \Gamma(k\Delta)$, where Γ is the autocovariance function of the original continuous-time process - this follows immediately from (33) and (7). Recall that for calculation of time-domain multistep Granger causalities for a discrete-time process, we require the residual noise intensity matrices and the first $m - 1$ MA coefficients (where m is the number of prediction steps) of the process itself and also of subprocesses (16,17). In the frequency domain we need the transfer function, evaluated at the requisite frequencies (12,14,23). Analytically, while in principle the Δ-subsampled VAR parameters might be derived from the autocovariance sequence via the discrete-time Yule-Walker equations (8) (*cf.* our remarks in Section 2 regarding the multiple representations for a VAR process), in practice this is generally intractable, and it is more convenient to calculate them from the discrete-time CPSD by spectral factorisation[14].

Given a CTVAR specified by an autoregressive coefficients kernel $A(u)$ and residuals covariance matrix Σ, a procedure for analytic calculation of multistep Granger causalities for the discrete-time Δ-subsampled process is described in Table 2. In Section 4.1 below we follow precisely this procedure for a non-trivial analytic example.

In Appendix G we establish firstly an asymptotic expansion for the CPSD of the subsampled process in the limit $\Delta \to 0$

$$S(\lambda; \Delta) = S(\lambda) + \tfrac{1}{12}\Delta^2\Sigma + \tfrac{1}{720}\Delta^4(\Omega + 12\pi^2\lambda^2\Sigma) + O(\Delta^5) \qquad (43)$$

where $\Omega \equiv \ddot\Gamma(0) + \ddot\Gamma(0)^\mathsf{T}$, and also the scaling relations

---

[13]This follows from the continuous-time versions of the Wiener-Kintchine and Convolution theorems, noting that (34) may be written as $\Gamma(t) = (\mathsf{B} * \mathsf{B}^\mathsf{T})(t)$ where $\mathsf{B}(u) \equiv B(u)L$, with L a matrix square root of Σ satisfying $LL^\mathsf{T} = \Sigma$ (by positive-definiteness, such an L exists).

[14]Solving for A_k, Σ from (8) involves a matrix deconvolution, which is generally difficult to perform analytically. In the frequency domain, the Convolution Theorem—which underlies the spectral factorisation formula (14)—renders the deconvolution more tractable. In continuous time, however, the integro-differential Yule-Walker equations (35) may well be more tractable (*cf.* Section 4.1).

1. Calculate the continuous-time MA kernel B by direct solution of (32).

2. Calculate the continuous-time autocovariance function Γ as follows: either

   (a) Calculate the continuous-time transfer function H (38).

   (b) Calculate the continuous-time CPSD S (40).

   (c) Calculate Γ by inverse Fourier transform (37).

   or

   (d) Calculate Γ by integration (34).

   or

   (e) Calculate Γ by direct solution of the continuous-time Yule-Walker equations (35).

3. Calculate the discrete-time subsampled process autocovariance sequence $\Gamma(\Delta)$ by $\Gamma_k(\Delta) = \Gamma(k\Delta)$.

4. Calculate the subsampled process CPSD $S(\Delta)$ by discrete-time Fourier transform of $\Gamma(\Delta)$ (11).

5. Calculate the subsampled process transfer function $H(\Delta)$ and residuals intensity $\Sigma(\Delta)$ by discrete-time spectral factorisation of $S(\Delta)$, for both full and (time-domain only) reduced models (14).

6. Time domain (1-step): calculate Granger causality from full and reduced subsampled residuals intensities (18).

7. Time domain (m-step):

   (a) Calculate subsampled MA coefficients $B_k(\Delta)$ up to $k = m - 1$ by inverse Fourier transform of the subsampled transfer function $H(\Delta)$, for both full and reduced models (12).

   (b) Calculate MSPEs $\mathcal{E}_m(\Delta)$ from $\Sigma(\Delta)$ and the $B_k(\Delta)$, for both full and reduced models (16).

   (c) Calculate subsampled m-step Granger causality from the full and reduced MSPEs (17).

8. Frequency domain: calculate frequency domain Granger causality from (full model) $\Sigma(\Delta)$, $S(\Delta)$ and $H(\Delta)$ (23).

Table 2: Procedure for analytical calculation of multistep time-domain and/or spectral Granger causalities for a Δ-subsampled CTVAR process from known CTVAR parameters $A(u), \Sigma$.

(Zhou et al., 2014)

$$H(\lambda; \Delta) = H(\lambda) + \boldsymbol{O}(\Delta) \qquad (44a)$$
$$\Sigma(\Delta) = \Sigma + \boldsymbol{O}(\Delta), \qquad (44b)$$

while from (9) and (34) we have:

$$B_k(\Delta) = B(k\Delta) + \boldsymbol{O}(\Delta) \qquad (45)$$

3.2. Subsampling a VOU process

The special case of subsampling a vector Ornstein-Uhlenbeck process, i.e., where $A(u) = A\delta(u)$, may be solved exactly. The Yule-Walker equation (35a) becomes the ordinary differential equation $\dot{\Gamma}(t) = A\Gamma(t)$, with solution

$$\Gamma(t) = e^{At}\Gamma(0) \qquad t \geq 0 \qquad (46)$$

and from the initial condition (35b), $\Gamma(0)$ satisfies the continuous-time Lyapunov equation

$$A\Gamma(0) + \Gamma(0)A^{\mathsf{T}} = -\Sigma \qquad (47)$$

The autocovariance sequence for the Δ-subsampled process is thus

$$\Gamma_k(\Delta) = e^{\Delta A k}\Gamma(0) \qquad (48)$$

Now it is easily calculated from the discrete-time Yule-Walker equations (8) that the discrete-time VAR(1) process $X_k = AX_{k-1} + \boldsymbol{\varepsilon}_k$ has autocovariance sequence $\Gamma_k = A^k\Gamma_0$, where Γ_0 satisfies the discrete-time Lyapunov equation $\Gamma_0 - A\Gamma_0 A^{\mathsf{T}} = \Delta\Sigma$. Since a VAR process is uniquely identified by its autocovariance sequence, we thus find that the subsampled process $X(\Delta)$

is VAR(1) (which underlines the unsuitability of VOU processes as models for neurophysiological data). The (1-lag) coefficient matrix and residual noise intensity are given respectively by

$$A(\Delta) = e^{\Delta A} \qquad (49a)$$
$$\Sigma(\Delta) = \Delta^{-1}\left[\Gamma(0) - e^{\Delta A}\Gamma(0)e^{\Delta A^{\mathsf{T}}}\right] \qquad (49b)$$

Note that in general a subprocess of a VOU process will *not* be VOU, nor will a subsampled subprocess be VAR(1).

3.3. Granger causality for CTVAR processes

It is not immediately clear how we should *define* Granger causality for continuous-time processes in general, and for CTVAR processes (considered as natural continuous-time analogues of VAR processes) in particular. As we shall see, if we attempt to calculate Granger causality at an "infinitesimal" prediction horizon, then prediction errors becomes negligible and, in particular, full and reduced prediction errors decay to zero *at the same rate* (Renault and Szafarz, 1991; Comte and Renault, 1996; Renault et al., 1998); thus Granger causality vanishes in the infinitesimal horizon limit. This suggests that we consider prediction at *finite* time horizons; that is, a Granger causality measure $\mathcal{F}_{Y\to X}(h)$ based on a prediction horizon a finite time h into the future (Comte and Renault, 1996; Florens and Fougère, 1996). We would also like continuous-time GC to be, in a precise sense, the limiting case of discrete-time GC under increasingly fine subsampling.

We are thus lead to consider optimal prediction of $X(t + h)$ given the history $X^-(t) \equiv \{X(s) \mid s \leq t\}$ of the process X up to and including time t. The orthogonal projection $\mathbf{E}[X(t + h) \mid X^-(t)]$ may be expressed as the limiting case, as

9

$\Delta \to 0$, of the expectation of $X(t + h)$ conditioned on a Δ-subsampling of the history $X^-(t)$:

$$\mathbf{E}[\,X(t + h)\,|\,X^-(t)\,] =$$
$$\lim_{\Delta \to 0} \mathbf{E}[\,X(t + h)\,|\,X(t), X(t - \Delta), X(t - 2\Delta), \ldots] \quad (50)$$

Then, setting $\Delta = h/m$, this is just the limit as $m \to \infty$ of the optimal m-step prediction of the h/m-subsampled process $X(h/m)$—recall that by assumption (Section 3.1) it has a stable, minimum-phase VAR representation, at least for large enough m—and from (15) and (45) we obtain in the limit

$$\mathbf{E}[\,X(t + h)\,|\,X^-(t)\,] = \int_{u=h}^{\infty} B(u)\,dW(t + h - u) \quad (51)$$

An application of the Itō isometry then yields the continuous-time MSPE

$$\mathcal{E}(h) \equiv \mathrm{cov}(\mathbf{E}[\,X(t + h)\,|\,X^-(t)\,] - X(t + h))$$
$$= \int_0^h B(u)\Sigma B(u)^\mathsf{T}\,du, \qquad h \geq 0 \quad (52)$$

Alternatively, we might have defined the continuous-time MSPE at horizon h as the limit of its subsampled counterpart:

$$\mathcal{E}(h) \equiv \lim_{m \to \infty} \mathcal{E}_m(h/m), \qquad h \geq 0 \quad (53)$$

where $\mathcal{E}_m(h/m)$ denotes the m-step MSPE of the h/m-subsampled process. By (16), (44b) and (45) the definitions coincide. Note that convergence in (53) is from above: for fixed $m \geq 1$ and any integer $r > 1$, $\mathcal{E}_{rm}(h/rm) \leq \mathcal{E}_m(h/m)$, since the corresponding orthogonal projections predict the process at the same horizon (i.e., time h into the future), but the rm-step prediction is based on a superset of the historic predictor set of the m-step prediction. Since $\mathcal{E}_m(h/m) \geq 0$ for any $m \geq 1$, the limit (53) thus exists. From (52) $\mathcal{E}(h)$ satisfies the ordinary differential equations

$$\dot{\mathcal{E}}(h) = B(h)\Sigma B(h)^\mathsf{T} \qquad h \geq 0 \quad (54a)$$
$$\mathcal{E}(0) = 0 \quad (54b)$$

For a joint CTVAR process $[X^\mathsf{T}\,Y^\mathsf{T}]^\mathsf{T}$, we now define Granger causality at horizon h in continuous time analagously to the discrete-time case (17) as

$$\mathcal{F}_{Y \to X}(h) \equiv \log \frac{|\mathcal{E}'_{xx}(h)|}{|\mathcal{E}_{xx}(h)|}, \qquad h \geq 0 \quad (55)$$

where $\mathcal{E}'_{xx}(h)$ denotes the continuous-time MSPE at horizon h for the subprocess X (recall that by assumption X has a CTVAR representation). From (17) and (53) we have

$$\mathcal{F}_{Y \to X}(h) = \lim_{m \to \infty} \mathcal{F}_{Y(h/m) \to X(h/m), m}, \qquad h \geq 0 \quad (56)$$

so that continuous-time GC may be *defined* as the limit of discrete-time GC under progressively finer subsampling, whilst holding the prediction horizon h fixed (FIG. 1).

Clearly $\mathcal{F}_{Y \to X}(h) \geq 0$ always. We now show that $\mathcal{F}_{Y \to X}(h) \to 0$ linearly as $h \to 0$ [*cf.* Zhou et al. (2014)]. From (54) we have $\mathcal{E}(0) = 0$ and $\dot{\mathcal{E}}(0) = \Sigma$, so that from (55)

$$\mathcal{F}_{Y \to X}(h) = \log \frac{\left|h\Sigma'_{xx} + \boldsymbol{O}(h^2)\right|}{\left|h\Sigma_{xx} + \boldsymbol{O}(h^2)\right|} \quad (57)$$

as $h \to 0$. Now the CPSD of X may be written in two ways as

$$S_{xx}(\lambda) = [H(\lambda)\Sigma H(\lambda)^*]_{xx} = H'_{xx}(\lambda)\Sigma'_{xx}H'_{xx}(\lambda)^* \quad (58)$$

where $H'_{xx}(\lambda), \Sigma'_{xx}$ denote respectively the transfer function and residual noise intensity associated with the reduced CTVAR. Multiplying through by λ^2 and letting $\lambda \to \infty$, from (41) we obtain $\Sigma'_{xx} = \Sigma_{xx}$ and by (57) we see that [as noted by Renault and Szafarz (1991); Florens and Fougère (1996) and Comte and Renault (1996)], $\mathcal{F}_{Y \to X}(h) \to 0$ as $h \to 0$.

Intuitively, this result may be thought of as follows: insofar as the transfer function $H(\lambda)$ represents the input $\to$ output response of the system, (41) indicates that on short timescales ($\lambda \to \infty$), the off-diagonal elements of $H(\lambda)$ (cross-response) decay to zero faster than the on-diagonal elements (self-response). Thus at short predictive time scales, conditional on their past joint history, the variables X, Y effectively decouple.

From (52) and (34), se see that both $\mathcal{E}_{xx}(h)$ and $\mathcal{E}'_{xx}(h) \to \Gamma_{xx}(0)$ as $h \to \infty$, so that $\mathcal{F}_{Y \to X}(h) \to 0$ as $h \to \infty$. Thus, unless identically zero, $\mathcal{F}_{Y \to X}(h)$ will attain a maximum at some finite horizon $0 < h < \infty$.

In contrast to $\mathcal{F}_{Y \to X}(0)$, the *zero-horizon Granger causality rate*[15]

$$\mathcal{R}_{Y \to X} \equiv \lim_{h \to 0} \frac{1}{h}\mathcal{F}_{Y \to X}(h) = \lim_{\Delta \to 0} \frac{1}{\Delta}\mathcal{F}_{Y(\Delta) \to X(\Delta)} \quad (59)$$

[the last equality follows from (57) and (44b)] will generally be non-zero. Setting

$$\mathsf{D} \equiv \tfrac{1}{2}\ddot{\mathcal{E}}(0) = \tfrac{1}{2}[\dot{B}(0)\Sigma + \Sigma\dot{B}(0)^\mathsf{T}] \quad (60)$$

[the last equality follows from differentiating (54a)], from (59) and (55) and noting that $\mathcal{R}_{Y \to X} = \dot{\mathcal{F}}_{Y \to X}(0)$, we have[16]

$$\mathcal{R}_{Y \to X} = \mathrm{trace}\left[\Sigma_{xx}^{-1}\left(\mathsf{D}'_{xx} - \mathsf{D}_{xx}\right)\right] \quad (61)$$

where D'_{xx} denotes the corresponding quantity for the reduced CTVAR. $\mathcal{R}_{Y \to X}$ may be considered an information transfer rate, measured in bits or nats per unit time.

While (as for the discrete-time, multistep case) we do not have a workable definition for spectral GC $\mathsf{f}_{Y \to X}(\lambda; h)$ at finite

---

[15]The Granger-causal concept underlying this quantity has been described in the econometrics literature as "local causality" or "instantaneous causality". Here we do not use the former term, since "local" is more commonly associated with *spatial* rather than *temporal* proximity, nor the latter, to avoid confusion with what Geweke (1982) terms "instantaneous feedback", an entirely distinct concept.

[16]This follows from the standard formula for the derivative of the log-determinant of a square matrix function: $\frac{d}{dt}\log|M(t)| = \mathrm{trace}\left[M(t)^{-1}\dot{M}(t)\right]$.
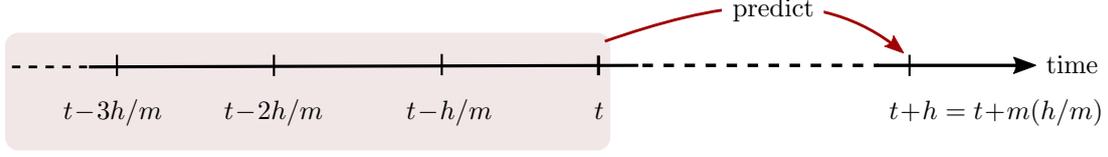
FIG. 1: Illustration of prediction underlying eq. (56): $\mathcal{F}_{Y \to X}(h)$ is the limit of the subsampled discrete-time m-step GC $\mathcal{F}_{Y(h/m) \to X(h/m),m}$ at fixed prediction horizon $h = m(h/m)$ under progressively finer subsampling ($m \to \infty$). Note that the historic predictor set $\{X(s), Y(s) \mid s = t, t - h/m, t - 2h/m, t - 3h/m, \ldots\}$ becomes progressively more detailed as m increases, approaching the continuous-time predictor set $\{X(s), Y(s) \mid s \leq t\}$ as $m \to \infty$.

prediction horizon h, we define at least the zero-horizon spectral GC in continuous time (again measured in bits or nats) as

$$f_{Y \to X}(\lambda; 0) \equiv \log \frac{|S_{xx}(\lambda)|}{\left| S_{xx}(\lambda) - H_{xy}(\lambda)\Sigma_{yy|x}H_{xy}(\lambda)^* \right|} \quad (62)$$

In contrast to the time-domain GC, spectral GC does not generally vanish at zero prediction horizon; for any λ, the pointwise limit as $\Delta \to 0$ of the Δ-subsampled spectral GC is equal to $f_{Y \to X}(\lambda; 0)$:

$$\lim_{\Delta \to 0} f_{Y(\Delta) \to X(\Delta)}(\lambda) = f_{Y \to X}(\lambda; 0) \quad (63)$$

This follows from the discrete-time spectral GC definition (23) via (43), (44a) and (44b). From (24) we then obtain a spectral decomposition for the continuous-time zero-horizon GC rate:

$$\mathcal{R}_{Y \to X} = \int_{-\infty}^{\infty} f_{Y \to X}(\lambda; 0) \, d\lambda \quad (64)$$

It is not quite obvious that $\Psi_{xy}(\zeta) \equiv 0 \implies \mathcal{F}_{Y \to X}(h) = 0$ for all $h > 0$. This may be seen as follows: $\Psi_{xy}(\zeta) \equiv 0$ implies that the MA kernel $B(u)$ and transfer function $H(\lambda)$ are lower block-triangular. From (40) it follows that the CPSD of X is given by $S_{xx}(\lambda) = [H(\lambda)\Sigma H(\lambda)^*]_{xx} = H_{xx}(\lambda)\Sigma_{xx}H_{xx}(\lambda)^*$, so that [cf. (58)] $\Sigma'_{xx} = \Sigma_{xx}$ and, since the MA kernel is the inverse Fourier transform of the transfer function, $B'_{xx}(u) = B_{xx}(u)$. From (52) and (55) it then follows that $\mathcal{F}_{Y \to X}(h) = 0$ for any h, and thence that $\mathcal{R}_{Y \to X} = 0$. By a result of Comte and Renault (1996, Prop. 17) the converse also holds; that is, $\mathcal{R}_{Y \to X} = 0 \implies \Psi_{xy}(\zeta) \equiv 0$, so that we may state[17]

$$\mathcal{R}_{Y \to X} = 0 \iff \Psi_{xy}(\zeta) \equiv 0 \iff \mathcal{F}_{Y \to X}(h) = 0 \; \forall h > 0 \quad (65)$$

This result may be considered a continuous-time analogue of (19). In contrast to the discrete-time case, where it is possible that $\mathcal{F}_{Y \to X} > 0$ but $\mathcal{F}_{Y \to X,m} = 0$ for some $m > 1$ (Section 2.1 and Appendix C), it is not clear whether we may have $\mathcal{R}_{Y \to X} >$

0 but $\mathcal{F}_{Y \to X}(h) = 0$ for some $h > 0$ (we have not found any examples of such behaviour, either analytically or numerically).

In general, there is no reason to suppose that $\mathcal{F}_{Y \to X}(h) \equiv 0$ will imply the vanishing of $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ for a Δ-subsampling; that is (Comte and Renault, 1996), *subsampling a CTVAR may induce spurious Granger causality*. We remark that it is non-trivial to verify this phenomenon analytically by example (*cf.* Section 4 below). Indeed, it is not hard to see that spurious causality cannot occur for a subsampled VOU process (Comte and Renault, 1996, Prop. 21). In this case (Section 3.2), we have $\Psi_{xy}(\zeta) \equiv 0 \iff A_{xy} = 0$, where the VOU AR kernel is $A(u) = A\delta(u)$, and (49a) implies immediately that $A_{xy}(\Delta) = 0$ where $A(\Delta)$ is the VAR(1) AR coefficient matrix for the Δ-subsampled process, so that $\mathcal{F}_{Y(\Delta) \to X(\Delta)} = 0$ for any Δ. Furthermore, the analysis of higher-order SDEs in Comte and Renault (1996, Sec. 3) would appear to imply that for a 2×1-dim (bivariate) CTVAR, spurious causality cannot arise (*cf.* Section 4.2). In general, it is possible that $\mathcal{F}_{Y(\Delta) \to X(\Delta)} > \mathcal{F}_{Y \to X}(\Delta)$ for some Δ values (*cf.* Section 4.2, FIG. 8).

Our discussion (Section 2.1) regarding filter-invariance of GC in discrete time suggests that, for $h > 0$, $\mathcal{F}_{Y \to X}(h)$ will not in general be invariant under a continuous-time causal invertible filter $\mathcal{G}(\zeta) = \int_0^\infty G(u) \, e^{-\zeta u} \, du$ with $\lim_{|\omega| \to \infty} \mathcal{G}(i\omega) = I$ and $\mathcal{G}_{xy}(\zeta) \equiv 0$; this is indeed the case - see Section 4.1 below for an example. As in the discrete-time case, filter-invariance does hold if $\Psi_{xy}(z) \equiv 0$, so that again causal, invertible filtering will not induce spurious causality at any prediction horizon. It may also be confirmed that filter invariance always holds at zero prediction horizon[18]; that is, $\mathcal{R}_{Y \to X}$ and $f_{Y \to X}(\lambda; 0)$ are invariant under causal, invertible filtering.

3.4. Estimation and inference of Granger causalities for sub-sampled continuous-time data

In an empirical setting, given neural data in the form of a discrete subsampling of an underlying continuous-time neurophysiological process, our standpoint is that the objective of GC-based functional analysis is to estimate as best we can (and perform statistical inference about) Granger causalities *for the underlying neurophysiological process*. That is, having access only to a Δ-subsampling of a joint continuous-time process $[X^\mathsf{T} \, Y^\mathsf{T}]^\mathsf{T}$, our aim is to estimate as well as possible

---

[17]$\mathcal{R}_{Y \to X} = 0$ is equivalent to what Comte and Renault (1996) describe as "local noncausality", while $\mathcal{F}_{Y \to X}(h) = 0 \; \forall h > 0$ corresponds to "global non-causality". The former is shown to be equivalent to $\Phi_{xy}(\zeta) \equiv 0$; in the *unconditional* GC case considered here, this is equivalent to $\Psi_{xy}(\zeta) \equiv 0$. We note also that Caines and Chan (1975), regarding some results which would seem to support this result (at least for rational transfer functions), remark that: "[...] the definitions and results in this paper are also applicable to continuous time processes", where by "continuous time processes" they refer explicitly to processes of the form (26).

[18]The argument of Appendix B goes through verbatim for $f_{Y \to X}(\lambda; 0)$; then (64) establishes invariance for $\mathcal{R}_{Y \to X}$. Alternatively, we may take the limiting case $\Delta \to 0$ in (59) under a suitable discretisation of $\mathcal{G}(\zeta)$.

the continuous-time Granger causalities $\mathcal{F}_{Y \to X}(h)$ at prediction horizons of interest, which we regard as reflecting "true" directed functional—as distinct from mechanistic (Barrett and Barnett, 2013)—relationships between the neurophysiological variables at various time scales. This suggests, on the basis of the preceeding analysis, two possible avenues

1. Estimate a CTVAR for the underlying process from the subsampled data and calculate continuous-time GC directly.

2. Calculate discrete-time GC based on a VAR estimate of the subsampled data.

The second is, of course, the standard route for GC-based functional connectivity analysis. As we have seen, in the limit that the sample increment $\Delta \to 0$, then *on a theoretical level*—i.e., given *exact* models—cases 1 and 2 converge to the same result. With limited data, however, several issues arise: regarding case 1, existing theory for the identification of continuous-time models from subsampled data (Åström, 1969; Larsson et al., 2006; Garnier and Wang, 2008) in general does not extend to distributed-lag multivariate stochastic integro-differential equation models[19]. In any case, numerical computation of GCs from (known) CTVAR parameters is likely to be nuch harder than for VARs, where standard algorithms are available (Appendix D).

In neither case is it clear how the relationship between sample rate and time scales underpinning the underlying neurophysiological process interact with GC inference - the principal theme of this study. In this paper we do not view CTVAR models as a *constructive* model with regard to Granger-causal inference; that is, we reject approach 1 above as a practical alternative (at least under the current theoretical background). Rather, we adopt approach 2 and regard the CTVAR model as an appropriate *analytical* tool for examining the effects of subsampling on Granger-causal inference. This approach is exemplified in the next section, where we solve, analytically, arguably the simplest non-trivial scenario, which demonstrates a basic mode of interaction between neural and sampling time scales.

A further question must also be addressed: given a continuous-time process accessible only via a Δ-subsampling, then in ascertaining e.g., whether we have obtained a spurious causality or failed to detect a non-zero causality what should we consider to be the "ground truth" Granger causality? Here we take the pragmatic view that, since our prediction horizon is constrained by the sampling interval Δ, the "true causality" is the underlying continuous-time GC at prediction horizon Δ; that is, an empirical estimate $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$ of discrete-time Δ-subsampled GC should be compared against the continuous-time GC $\mathcal{F}_{Y \to X}(\Delta)$. This makes sense since $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ and $\mathcal{F}_{Y \to X}(\Delta)$ are both based on prediction of $X(t+\Delta)$ by histories

of X, Y up to and including time t, the difference being that the former is based on sparse, discrete (i.e., Δ-subsampled) histories (*cf.* FIG. 1).

It might be argued that, for completeness, we should estimate and perform statistical inference for m-step Δ-subsampled GC; i.e., we should estimate $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta),m}$, to be compared against $\mathcal{F}_{Y \to X}(m\Delta)$, for $m = 1, 2, \ldots$. For simplicity we omit this analysis for our worked example (Section 4), although we do consider the situation where, for a fixed quantity of data, we have a *choice* of sampling interval Δ (i.e., we may downsample). Other possibilities worth consideration include comparison of the "empirical GC rate" $\frac{1}{\Delta} \widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$, at least for small Δ, with the zero-horizon GC rate $\mathcal{R}_{Y \to X}$, or comparison of an estimated "total GC" $\Delta \sum_{m=1}^{\infty} \widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta),m}$ against the corresponding continuous-time quantity $\int_0^\infty \mathcal{F}_{Y \to X}(h) \, dh$.

4. Subsampling analysis for a minimal CTVAR process with finite causal delay

Having established a consistent theoretical framework for the analysis of Granger causality for distributed-lag continuous-time processes and time series extracted by subsampling, we now work through a detailed Granger-causal subsampling analysis of a minimal (but certainly non-trivial) CTVAR process with finite-time feedback delay. This example is solvable analytically—in both discrete and continuous time—and serves to highlight some key modes of interaction between sampling frequency, causal delay and statistical power of Granger-causal inference. (Of course we should be cautious in assuming generalisation of results obtained here to more realistic neurophysiological models.) Analytical results are compared with detailed simulation. Although we concentrate on time-domain GCs (and, in the subsampled case, 1-step GC), for illustrational purposes we include some results on spectral GC.

We consider a CTVAR process (29) with AR kernel

$$A(u) \equiv -A \, \delta(u) + C \, \delta(t - \tau) \tag{66}$$

and residual noise intensity Σ, where

$$A \equiv \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \qquad C \equiv \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix}, \qquad \Sigma \equiv \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{67}$$

i.e.,

$$dX(t) = -a \, X(t) \, dt + c \, Y(t - \tau) \, dt + dW_x(t) \tag{68a}$$
$$dY(t) = -b \, Y(t) \, dt \qquad\qquad + dW_y(t) \tag{68b}$$

The process $Y(t)$ is a standard Ornstein-Uhlenbeck process, which drives the process $X(t)$ at a fixed delay of τ. $1/a, 1/b$ represent the (exponential decay) relaxation time of the x and y nodes in the absence of input, while c controls the strength of feedback from node y to node x at delay τ. For simplicity we assume $b \neq a$ (the special case $b = a$, which we may confirm behaves qualitatively similarly, may be solved along similar lines). The residual (instantaneous) correlation is $-1 < \rho < 1$. We henceforth refer to (68) as the "minimal CTVAR".

---

[19]Note that the parameter space for a general CTVAR (29) is infinite-dimensional. Existing theory appears to consider at best continuous-time VARMA models/higher-order SDEs, and does not accommodate distributed lags. A viable CTVAR approach might be to limit the model to a finite number of point lags (see e.g. McKetterick and Giuggioli, 2014, and references therein); but even in that restricted case we are not aware of any useful results on system identification from discretely-sampled data.

Processes like (68) (and non-linear generalisations) have been widely studied in the literature under the name *Stochastic Delay-Differential Equations* (SDDEs) (Longtin, 2010). Usually, though, only the univariate case is considered (but see e.g. McKetterick and Giuggioli, 2014) and the emphasis is generally on questions of stability, convergence, approximation, numerical simulation and perturbation theory, rather than Granger-causal analysis of stationary, stable systems.

Notation: In this section, for compactness we shall use the scaled frequency $\omega \equiv 2\pi\lambda$ in continuous time and the angular frequency $\omega \equiv 2\pi\Delta\lambda$ in discrete time for mathematical analyses, where Δ is as usual the sample interval. However, for displaying results we always convert back to ordinary frequency λ, so that spectral quantities scale appropriately with sample interval.

For the various plots and simulations in this section, we shall (unless otherwise stated) use the reference parameters:

$$1/a = 5, \quad 1/b = 6, \quad 1/c = 8, \quad \tau = 30 \tag{69}$$

(all times in milliseconds), while the residuals correlation coefficint ρ will take stated values.

4.1. Continuous-time analysis

From (66) using definition (31) we have

$$\Phi(\zeta) = \begin{bmatrix} a + \zeta & -ce^{-\tau\zeta} \\ 0 & b + \zeta \end{bmatrix} \tag{70}$$

which we invert to obtain

$$\Psi(\zeta) = \begin{bmatrix} \dfrac{1}{a + \zeta} & \dfrac{ce^{-\tau\zeta}}{(a + \zeta)(b + \zeta)} \\ 0 & \dfrac{1}{b + \zeta} \end{bmatrix} \tag{71}$$

from which we see that the minimum-phase condition is satisfied, and stability requires just that $a, b > 0$. From (32) it is straightforward to calculate the MA kernel as

$$B(u) = e^{-Au} + \mathsf{h}(u - \tau)\, q(u - \tau)\, C \tag{72}$$

where $\mathsf{h}(u)$ is the Heaviside step function equal to 0 for $u < 0$ and 1 for $u \geq 0$, and

$$q(u) \equiv \frac{1}{b - a}\left(e^{-au} - e^{-bu}\right) \tag{73}$$

From (71) we have

$$H(\omega) = \begin{bmatrix} \dfrac{1}{a + i\omega} & \dfrac{ce^{-i\tau\omega}}{(a + i\omega)(b + i\omega)} \\ 0 & \dfrac{1}{b + i\omega} \end{bmatrix} \tag{74}$$

and from (40) we obtain the continuous-time CPSD

$$S_{xx}(\omega) = \frac{b^2 + 2\rho c\, \upsilon(\omega; b, \tau) + c^2 + \omega^2}{(a^2 + \omega^2)(b^2 + \omega^2)} \tag{75a}$$

$$S_{xy}(\omega) = \frac{(a - i\omega)\left[ce^{-i\tau\omega} + \rho(b + i\omega)\right]}{(a^2 + \omega^2)(b^2 + \omega^2)} \tag{75b}$$

$$S_{yy}(\omega) = \frac{1}{b^2 + \omega^2} \tag{75c}$$

where

$$\upsilon(\omega; b, \tau) \equiv b \cos \tau\omega - \omega \sin \tau\omega \tag{76}$$

(75) is plotted in FIG. 2 for the reference parameters (69) and a few values of ρ. We see that power attenuates as ω^{-2} [*cf.* (42)], for $\rho \neq 0$ the xx-compenent oscillates with period τ, while for $\rho = 0$ the feedback delay τ appears only in the phase angle of the xy-component.

The autocovariance function is most easily calculated from $B(t)$ using (34); alternatively, we might solve the Yule-Walker equations (35), or invert the continuous-time Fourier transform (37). For convenience we define the dimensionless quantities

$$\theta \equiv \frac{c}{b - a}, \qquad \eta \equiv \frac{c}{a + b} \tag{77}$$

and we may calculate that for $t \geq 0$,

$$\Gamma(t) = \Gamma^{(0)}(t) + \mathsf{h}(\tau - t)\,\Gamma^{(1)}(\tau - t) + \mathsf{h}(t - \tau)\,\Gamma^{(2)}(t - \tau) \tag{78}$$

where

$$\Gamma_{xx}^{(0)}(t) = \frac{1 + \theta\eta}{2a}e^{-at} - \frac{\theta\eta}{2b}e^{-bt} + \rho\frac{\eta}{2a}e^{-a(t+\tau)} \tag{79a}$$

$$\Gamma_{xy}^{(0)}(t) = \rho\frac{1}{a + b}e^{-at} \tag{79b}$$

$$\Gamma_{yx}^{(0)}(t) = \rho\frac{1}{a + b}e^{-bt} + \frac{\eta}{2b}e^{-b(t+\tau)} \tag{79c}$$

$$\Gamma_{yy}^{(0)}(t) = \frac{1}{2b}e^{-bt} \tag{79d}$$

$$\Gamma_{xx}^{(1)}(t) = \rho\frac{\eta}{2a}e^{-at} \tag{79e}$$

$$\Gamma_{xy}^{(1)}(t) = \frac{\eta}{2b}e^{-bt} \tag{79f}$$

$$\Gamma_{xx}^{(2)}(t) = \rho\theta\left(\frac{1}{2a}e^{-at} - \frac{1}{a + b}e^{-bt}\right) \tag{79g}$$

$$\Gamma_{xy}^{(2)}(t) = \theta\left(\frac{1}{a + b}e^{-at} - \frac{1}{2b}e^{-bt}\right) \tag{79h}$$

and all other entries vanish. $\Gamma(t)$ is plotted for a few values of ρ in FIG. 3. We see that the xx- and yy-components have local peaks just after $t = \tau$ and autocovariance decays exponentially for large t.

To calculate the continuous-time Granger causality $\mathcal{F}_{Y \to X}(h)$ [clearly $\mathcal{F}_{X \to Y}(h)$ vanishes identically, since $H(\omega)$ is upper-triangular] we firstly require $\mathcal{E}_{xx}(h)$, which may be calculated from (72) by straightforward integration as per (52). We find:

$$\mathcal{E}_{xx}(h) = \frac{1}{2a}\left(1 - e^{-2ah}\right) + \mathsf{h}(h - \tau)r(h - \tau) \tag{80}$$

where

$$r(h) \equiv 2\rho\theta e^{-a\tau}\left\{\frac{1}{2a}\left(1 - e^{-2ah}\right) - \frac{1}{a + b}\left[1 - e^{-(a+b)h}\right]\right\}$$

$$+ \theta^2\left\{\frac{1}{2a}\left(1 - e^{-2ah}\right) - \frac{2}{a + b}\left[1 - e^{-(a+b)h}\right]\right.$$

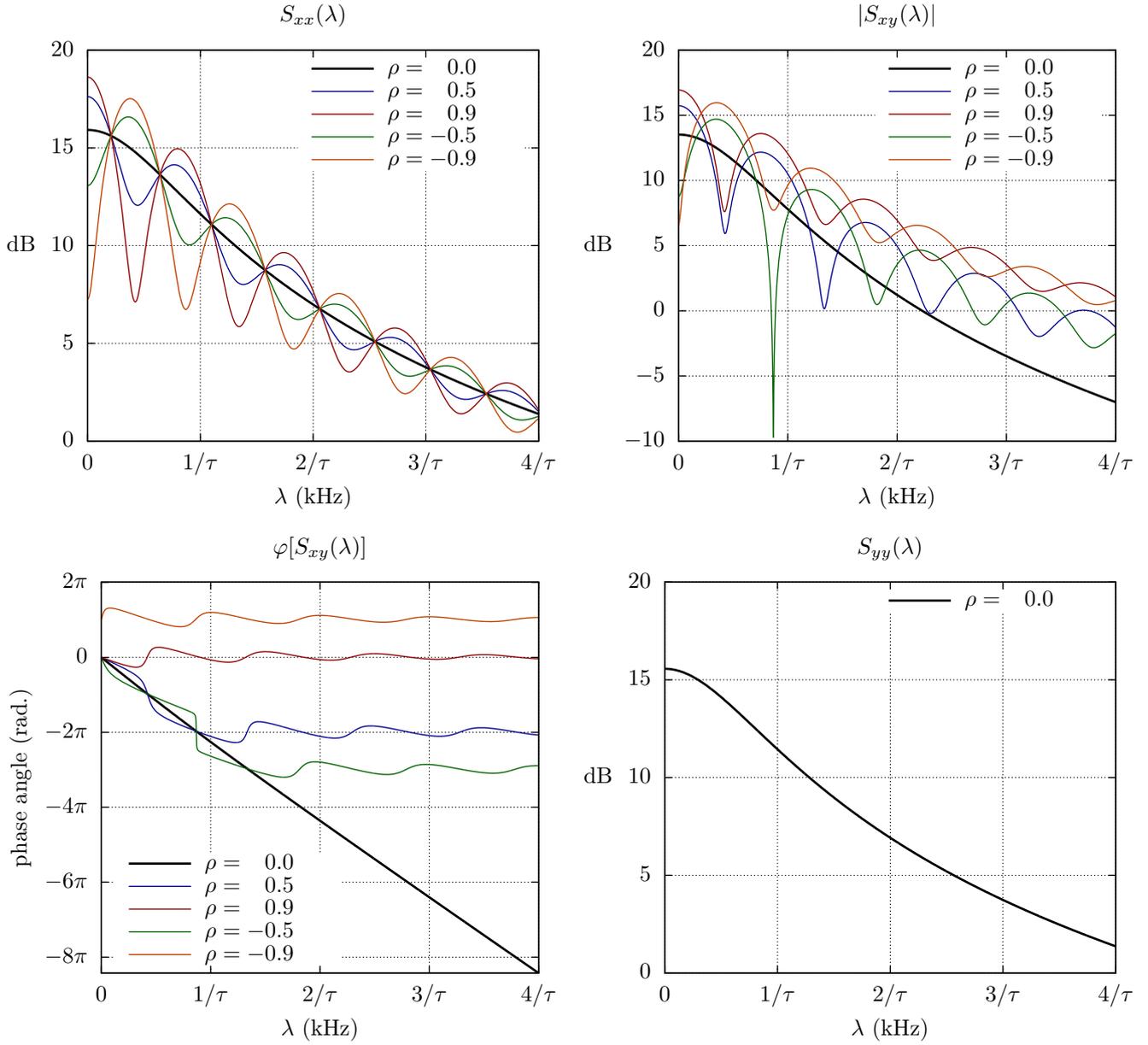$$\left. + \frac{1}{2b}\left(1 - e^{-2bh}\right)\right\} \tag{81}$$

FIG. 2: The CPSD (75) for the minimal CTVAR (68) with reference parameters (69) plotted (in dB) against ordinary frequency λ (kHz), for a few values of the residual noise correlation parameter ρ. The upper right figure plots the modulus, and the lower left figure the (unwrapped) phase angle, of the xy-cross-power term (75b). Note that the y-autospectrum (75c) (bottom right) does not depend on ρ.
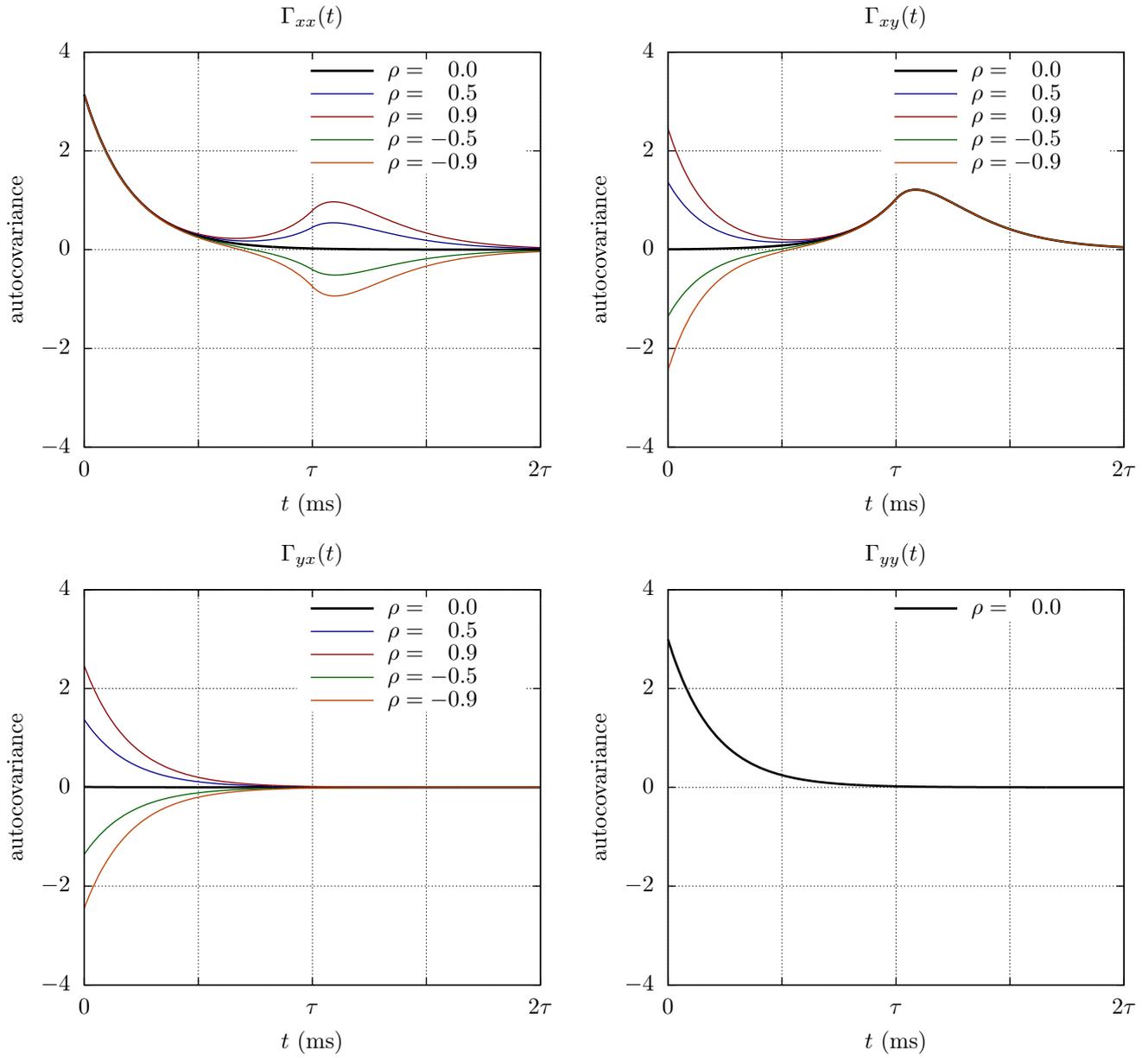
FIG. 3: The autocovariance function (78) for the minimal CTVAR (68) with reference parameters (69) plotted against lag time t (ms), for a few values of the residual noise correlation parameter ρ. Note that the y-autocovariance (79d) (bottom right) does not depend on ρ.

For $\mathcal{E}'_{xx}(h)$, we need to solve the reduced spectral factorisation problem. It may be verified with (75a) that

$$H'_{xx}(\omega) = $$
$$\frac{\sqrt{(1-\rho^2)c^2 + (b + \rho c \cos \tau\omega)^2} + i(\omega - \rho c \sin \tau\omega)}{(a + i\omega)(b + i\omega)} \quad (82)$$

satisfies $S_{xx}(\omega) = H'_{xx}(\omega)H'_{xx}(\omega)^*$ (we know that $\Sigma'_{xx} = \Sigma_{xx} = 1$). Then, at least in principal, we may calculate the reduced MA kernel as the inverse Fourier transform[20] $B'_{xx}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H'_{xx}(\omega) e^{it\omega} d\omega$. In the general case $\rho \neq 0, \tau > 0$ the inverse transform appears to be analytically intractable, although $B'_{xx}(u)$ may be calculated numerically. To this end we note that for large ω, $H'_{xx}(\omega)e^{i\omega u}$ is dominated by $(\sin \omega u)/\omega$, and we find that $B'_{xx}(u) \approx \frac{1}{\pi} \int_0^W H'_{xx}(\omega) e^{i\omega u} d\omega + \frac{1}{2} - \frac{1}{\pi} \mathrm{Si}(Wu)$ with $W \gg \max(a, b, |\rho c|)$, where $\mathrm{Si}(x) \equiv \int_0^x (\sin \xi)/\xi \, d\xi$ is the sine integral function and the integral may be approximated by numerical quadrature[21]. We may then approximate $\mathcal{E}'_{xx}(h)$ from (52), again by numerical quadrature, and $\mathcal{F}_{Y \to X}(h)$ is calculated as per (55).

The zero-horizon spectral GC (62) may be calculated from (74) and (75), noting that $\Sigma_{yy|x} = 1 - \rho^2$. We have

$$\mathrm{f}_{Y \to X}(\omega; 0) = \log\left(1 + \frac{(1-\rho^2)c^2}{b^2 + 2\rho c \, \upsilon(\omega; b, \tau) + \rho^2 c^2 + \omega^2}\right) \quad (83)$$

The zero-horizon GC rate $\mathcal{R}_{Y \to X}$ may then in principal be obtained from the spectral decomposition (64) by integrating (83). Again the integral appears analytically intractable, but may be calculated numerically. For large ω, the denominator of the fraction in (83) is dominated by the ω^2 term, from which we may calculate that $\mathcal{R}_{Y \to X} \approx \frac{1}{\pi} \int_0^W \mathrm{f}_{Y \to X}(\omega; 0) \, d\omega + \frac{1}{\pi}(1-\rho^2)c^2/W$ with $W \gg \max(b, |\rho c|)$ and the integral may again be approximated by numerical quadrature.

Full analytical calculation of $\mathcal{F}_{Y \to X}(h)$ and $\mathcal{R}_{Y \to X}$ is tractable in the special cases $\tau = 0$ and $\rho = 0$. Recall (Section 3) that we do not consider zero feedback delay to be plausible for neurophysiological processes. While it seems unlikely that residuals correlation will be completely absent, we present the $\rho = 0$ case analytically both in continuous time and under subsampling. In particular,

$$H'_{xx}(\omega) = \frac{k + i\omega}{(a + i\omega)(b + i\omega)} \quad (84)$$

with $k \equiv \sqrt{b^2 + c^2}$, and performing the inverse Fourier transform, we may calculate that

$$B'_{xx}(u) = \frac{(k-a)e^{-au} - (k-b)e^{-bu}}{b - a} \quad (85)$$
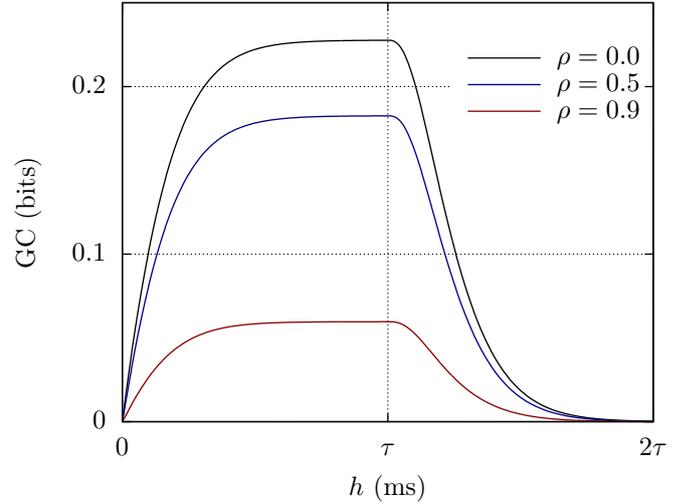

FIG. 4: Granger causality $\mathcal{F}_{Y \to X}(h)$ for the minimal CTVAR (68) with reference parameters (69) plotted against prediction horizon h (ms), for a few values of the residual noise correlation parameter ρ. Values for $\rho < 0$ (not displayed) are close to their positive counterparts.

We may then calculate $\mathcal{E}'_{xx}(h)$ by integration from (52):

$$\mathcal{E}'_{xx}(h) = \frac{1}{(b-a)^2} \left\{ \frac{(k-a)^2}{2a} \left(1 - e^{-2ah}\right) \right.$$
$$\left. - \frac{2(k-a)(k-b)}{a+b} \left[1 - e^{-(a+b)h}\right] + \frac{(k-b)^2}{2b} \left(1 - e^{-2bh}\right) \right\} \quad (86)$$

$\mathcal{F}_{Y \to X}(h)$ may then be calculated according to (55) with (80,86). From the resulting expression, we may confirm that for $h \gg \tau$, $\mathcal{F}_{Y \to X}(h)$ decays exponentially with h, with exponent $\min(a, b)$.

From (59) or (61) we may calculate

$$\mathcal{R}_{Y \to X} = \sqrt{b^2 + c^2} - b \quad (87)$$

and from (62) we have

$$\mathrm{f}_{Y \to X}(\omega; 0) = \log\left(1 + \frac{c^2}{b^2 + \omega^2}\right) \quad (88)$$

Note that, unlike in the general $\rho \neq 0$ case, the feedback delay time τ does not appear in these expressions. (We would, however, expect to see dependence on causal delay if more than a single delayed feedback were present.)

FIG. 4 plots $\mathcal{F}_{Y \to X}(h)$ against prediction horizon h for a few values of ρ. We see that for small h, $\mathcal{F}_{Y \to X}(h)$ rises approximately linearly with slope $\mathcal{R}_{Y \to X}$ [cf. (57,59)]. It then flattens out and attains a maximum just beyond $h = \tau$, before decaying exponentially as described above. Overall, $\mathcal{F}_{Y \to X}(h)$ is reduced for larger values of $|\rho|$ ($\rho < 0$ plots are not displayed; for our reference parameters they lie very close to the corresponding positive ρ plots).

We also make the following observation: suppose that we define a filter $\mathcal{G}(\zeta) = \begin{bmatrix} \dfrac{a + \zeta}{g + \zeta} & 0 \\ 0 & 1 \end{bmatrix}$, which is stable, minimum-phase and invertible provided $g > 0$. Now applying the filter to

---
[20]The factor of $1/\pi$ arises from the use of scaled frequency $\omega = 2\pi\lambda$; *cf.* (64).
[21]To avoid aliasing artefacts, for numerical quadrature we should ensure that for each u we have $d\omega \leq 2\pi/u$.

(71), we see that the filtered CTVAR is identical to the original CTVAR (68), but with a replaced by g (*cf.* Appendix C). Since $\mathcal{F}_{Y \to X}(h)$ depends on a [*cf.* (80,86)], we have demonstrated the non-invariance of finite-horizon GC in the continuous-time case, as claimed at the end of Section 3.3. Since the zero-horizon spectral GC (83) does not depend on a, we see that, as expected, zero-horizon GC, in both time and frequency domains, is invariant under $\mathcal{G}(\zeta)$.

4.2. Subsampling analysis

A procedure for calculating Δ-subsampled Granger causalities analytically from the CTVAR parameters is described in Section 3.1, Table 2. We have already (78) performed the first stage, calculation of the continuous-time autocovariance function Γ (Table 2, steps 1,2d), while step 3, subsampling Γ, is trivial. Step 4, calculation of the discrete-time CPSD, is straightforward, if laborious.

Step 5, spectral factorisation of the discrete-time CPSD, is particularly demanding in the general case. In lieu of an analytical factorisation, we employed the following numerical method: since the autocovariance sequence $\Gamma(\Delta)$ of the subsampled process is just the Δ-subsampling of the continuous-time autocovariance function (78), it is easily calculated. We then use Whittle's time-domain multivariate spectral factorisation algorithm (Whittle, 1963), which takes as input a discrete-time autocovariance sequence and yields the corresponding VAR parameters as output[22] (the autocovariance sequence was truncated at sufficient lags for all autocovariances to have decayed below machine precision). Here, these calculations were performed using the Multivariate Granger Causality (MVGC) Matlab® toolbox (Barnett and Seth, 2014). An interesting result of this experiment, was that, even with $\rho \neq 0$, subsampled GC in the non-causal direction, $\mathcal{F}_{X(\Delta) \to Y(\Delta)}$, was seen to be zero for any sampling interval Δ. That is, *for our minimal CTVAR, there is no spurious GC*. This is somewhat surprising as it is known that, as previously noted, subsampling is in general likely to induce spurious causalities in both discrete and continuous time[23].

We thus concentrate on the detectability of GC under subsampling in the causal $Y \to X$ direction. Since non-zero ρ was not found to have a profound qualitative effect on $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$, from here on we set $\rho = 0$; the subsampling problem may then be solved entirely analytically for $\tau \geq 0$. We start, again, by Δ-subsampling the continuous-time autocovariance function (78) to obtain $\Gamma(\Delta)$. We may then calculate[24] the CPSD $S(\Delta)$ of the subsampled process by discrete-time Fourier transform (11) of

---

[22]An alternative approach would be to calculate the CPSD analytically (Table 2, step 4) and apply Wilson's frequency-domain multivariate spectral factorisation algorithm (Wilson, 1972).

[23]But see also our remarks in Section 3.3 regarding non-occurrence of subsampling-induced spurious causality in 2-dim bivariate CTVAR processes.

[24]The only awkward case is the cross-power term $S_{xy}(\Delta)$, which we calculate in Appendix I.

$\Gamma(\Delta)$. For convenience, we define the following quantities:

$$\alpha \equiv e^{-a\Delta} \qquad\qquad q \equiv \lceil \tau/\Delta \rceil \qquad (89a)$$

$$\beta \equiv e^{-b\Delta} \qquad\qquad \kappa \equiv \lceil \tau/\Delta \rceil - \tau/\Delta \qquad (89b)$$

$$\gamma \equiv e^{b\Delta} = \beta^{-1} \qquad\qquad\qquad (89c)$$

where $\lceil x \rceil \equiv \min\{n \in \mathbb{N} \mid n \geq x\}$ denotes the ceiling function. Note that (i) $q \geq 1$, with $q = 1 \iff \Delta \geq \tau$ and (ii) $0 \leq \kappa < 1$, with $\kappa = 0 \iff \Delta$ divides τ exactly. We also define

$$u_1 \equiv \frac{1 + \theta\eta}{2a}\,(1 - \alpha^2) \qquad v_1 \equiv \frac{\theta + \eta}{2b}\,\alpha^\kappa = \frac{\theta\eta}{c}\,\alpha^\kappa \qquad (90a)$$

$$u_2 \equiv -\frac{\theta\eta}{2b}\,(1 - \beta^2) \qquad v_2 \equiv -\frac{\theta}{2b}\,\beta^\kappa \qquad (90b)$$

$$w \equiv \frac{1}{2b}\,(1 - \beta^2) \qquad v_3 \equiv -\frac{\eta}{2b}\,\gamma^\kappa \qquad (90c)$$

We then find (for compactness we express spectral quantities in terms of $z = e^{-i\omega}$ on the circle $|z| = 1$ in the complex plane):

$$S_{xx}(z; \Delta) = \frac{u_1}{|1 - \alpha z|^2} + \frac{u_2}{|1 - \beta z|^2} \qquad (91a)$$

$$S_{xy}(z; \Delta) = \left(\frac{v_1}{1 - \alpha z} + \frac{v_2}{1 - \beta z} + \frac{v_3}{1 - \gamma z} \right) z^q \qquad (91b)$$

$$S_{yy}(z; \Delta) = \frac{w}{|1 - \beta z|^2} \qquad (91c)$$

Note that $S(z; \Delta)$ depends on the causal delay τ only via the cross-power term, through dependency on q and κ. FIG. 5 plots the CPSD (91) against ordinary frequency λ and subsampling interval Δ, for the reference parameters (69). In all plots the $\Delta = 0$ spectral power corresponds to the continuous-time CPSD (75) (FIG. 2). Note in particular the cross-spectral power $|S_{xy}|$, which undergoes a series of increasingly rapid oscillations with decreasing Δ, peaking just past the values $\tau, \tau/2, \tau/3, \ldots$, reflecting the oscillations in κ (89b) as the sample frequency $1/\Delta$ resonates with the causal delay frequency $1/\tau$ (FIG. 5, bottom right).

Factorisation of $S(\lambda; \Delta)$ for both full and reduced regressions, is non-trivial and detailed in Appendix J. Results may be summarized as follows:

1. The joint subsampled process $X(\Delta), Y(\Delta)$ is VARMA$(2, q + 1)$, the subprocess $X(\Delta)$ is VARMA$(2, 1)$ and the subprocess $Y(\Delta)$ is VAR(1).

2. The residuals covariance matrix $\Sigma(\Delta) \equiv \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}$ of the full regression of the joint process is specified by

$$\sigma_{xy} = \begin{cases} -\beta P & q = 1 \\ 0 & q > 1 \end{cases} \qquad (92a)$$

$$\sigma_{yy} = w \qquad (92b)$$

$$\sigma_{xx} = \frac{D + \sigma_{xy}^2}{\sigma_{yy}} \qquad (92c)$$

where P is given by (J.9) and D by (J.23) of Appendix J.

17

$S_{xx}(\lambda; \Delta)$

$S_{yy}(\lambda; \Delta)$

$|S_{xy}(\lambda; \Delta)|$

$|S_{xy}(\lambda; \Delta)|$ (sections)

$\lambda = 0.0$
$\lambda = 0.1$
$\lambda = 0.2$
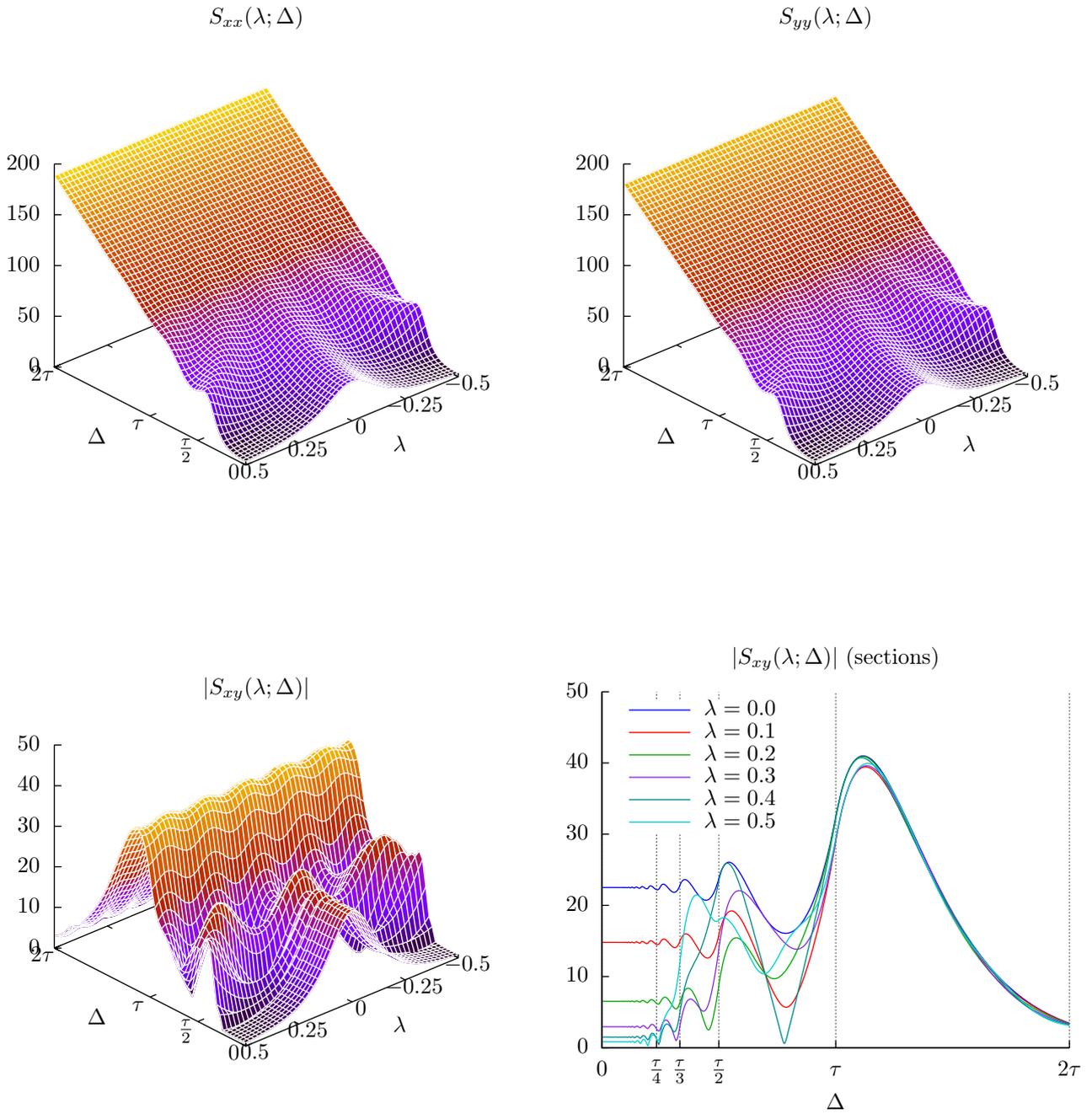$\lambda = 0.3$
$\lambda = 0.4$
$\lambda = 0.5$

FIG. 5: The CPSD $S(\lambda; \Delta)$ (91) for the subsampled minimal CTVAR (68) with reference parameters (69) and $\rho = 0$, plotted against ordinary frequency λ and sample interval Δ. In all plots the $\Delta = 0$ spectral power corresponds to the continuous-time CPSD (75) [*cf.* FIG. 2].

18

3. The residuals variances $\Sigma'_{xx}(\Delta) \equiv \sigma'_{xx}$ and $\Sigma'_{yy}(\Delta) \equiv \sigma'_{yy}$ of the reduced regressions of $X(\Delta), Y(\Delta)$ respectively are given by

$$\sigma'_{xx} = \varphi + \sqrt{\varphi^2 - \psi^2} \qquad (93a)$$

$$\sigma'_{yy} = \sigma_{yy} = w \qquad (93b)$$

where φ, ψ are given by (J.16) of Appendix J.

It was also verified, for all parameter values and subsampling intervals examined, that the joint Δ-subsampled process was minimum-phase[25] [see Appendix J, final paragraph; the sub-processes $X(\Delta)$ and $Y(\Delta)$ are always minimum-phase].

Since the subprocess Y is an Ornstein-Uhlenbeck process, the subsampled process $Y(\Delta)$ is VAR(1) (Section 3.2). Intuitively, the VARMA model orders for the full subsampled process $X(\Delta), Y(\Delta)$ and the subprocess $X(\Delta)$ may be explained as follows: firstly, inspection of the spectral factorisation process (Appendix J) reveals that the AR model order of 2 in both cases stems from the 2-dimensionality of the system. For the subprocess X (in both continuous time and subsampled), we have already noted that, with $\rho = 0$, the X auto-spectrum does not depend on the feedback delay time τ (75a,91a). To understand why this is so, we may consider the Y process as contributing lagged, additive (coloured) noise to the X process. With $\rho = 0$, this noise is uncorrelated with X (79), so that its contribution to the spectral power of X is independent of the lag length. The X process, considered in isolation, is thus structurally identical to the X process obtained by setting $\tau = 0$ in (68), which then becomes a 2-dim VOU process. Now from Section 3.2 we know that a subsampled VOU process is VAR(1), and it is straightforward to show (by eliminating the Y variable from the unlagged and singly-lagged structural equations) that a 1-dim subprocess of a 2-dim VAR(1) will in general be VARMA(2, 1). For the joint subsampled process, on the other hand, the cross-spectral term (91b) *does*, as we might expect, depend on the feedback delay, as does the transfer function cross-term (J.7), which is order $q + 1$; this reflects the delay of q time increments in sampling the contribution to X of an impulse from Y.

Our key result, analytical expressions for the directional and instantaneous Granger causalities of the subsampled process in the time-domain, follow immediately:

$$\mathcal{F}_{X(\Delta) \to Y(\Delta)} = \log\left(\frac{\sigma'_{yy}}{\sigma_{yy}}\right) = 0 \qquad (94a)$$

$$\mathcal{F}_{Y(\Delta) \to X(\Delta)} = \log\left(\frac{\sigma'_{xx}}{\sigma_{xx}}\right) \qquad (94b)$$

$$\mathcal{F}_{X(\Delta) \cdot Y(\Delta)} = -\log\left(1 - \tilde{\rho}^2\right) \qquad (94c)$$

where $\tilde{\rho} = \sigma_{xy}/\sqrt{\sigma_{xx}\sigma_{yy}}$ is the residuals correlation coefficient of the subsampled processes [*cf.* Section 2.1, eq. (20)]. Note

that, as might be expected with $\rho = 0$, $\mathcal{F}_{X(\Delta) \to Y(\Delta)}$ vanishes for any Δ, so that subsampling does not induce spurious causality in the non-causal $X \to Y$ direction. For completeness, we supply the formula for the frequency-domain GC

$$\mathrm{f}_{Y(\Delta) \to X(\Delta)}(z) =$$
$$- \log\left[1 - (1 - \rho^2)\frac{\left|\sigma_{xy}B_{xx}(z) + L(z)z^{q-1}\right|^2}{D|B_{xx}(z)|^2 + |L(z)|^2} \right] \qquad (95)$$

with $B_{xx}(z)$ given by (J.17) and $L(z)$ by (J.13) of Appendix J, while $\mathrm{f}_{X(\Delta) \to Y(\Delta)}(z)$ is again identically zero.

In FIG. 6 (top figure), the Δ-subsampled discrete-time GC $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ calculated according to (94b) is plotted, along with the continuous-time GC $\mathcal{F}_{Y \to X}(\Delta)$ as calculated from (80,86)—that is, for the same prediction horizon (*cf.* our remarks in Section 3.4)—against Δ for $\rho = 0$ and reference parameters (69). The pattern of distortion induced by subsampling is clear from the top figure: $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ sits under the envelope of the "true" GC $\mathcal{F}_{Y \to X}(\Delta)$, and closely follows both its linear rise from $\Delta = 0$ and also its exponential decay after peaking just beyond $\Delta = \tau$ [it is straightforward (if tedious) to show that in the limit of large sample interval, $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ decays with the same exponent as $\mathcal{F}_{Y \to X}(\Delta)$ as $\Delta \to \infty$].

As for the subsampled cross-spectral power $S_{xy}(\lambda; \Delta)$ (FIG. 5), we see increasingly rapid oscillations near the points $\Delta = \tau, \tau/2, \tau/3, \ldots$, which again arise as κ (89b) oscillates between 0 and 1, reflecting resonance between sampling and delayed feedback frequencies. Note that κ affects only the full prediction error $\sigma_{xx}(\Delta)$ (92c), while the reduced prediction errors $\sigma'_{xx}(\Delta)$ (93a) increases monotonically (FIG. 6, bottom figure): this reflects the fact that (for $\rho = 0$) the causal delay τ affects only the cross-power term of the CPSD, in both continuous time (75) and subsampled (91). In this case the positive-slope inflexion points of $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ lie almost exactly at $\Delta = \tau, \tau/2, \tau/3, \ldots$ (although this is not always the case; *cf.* FIG. 8 below.) Note that these oscillations are quite distinct from those reported in Zhou et al. (2014), which are of constant period, and are ascribed to periodicity in the time series. Since, as may be seen from the power spectra (FIG. 5), there is no strong periodic behaviour in the (subsampled) minimal CTVAR, we do not see the Zhou oscillations here (see also Section 5). A striking feature is that for $\tau/2 < \Delta < \tau$ there is strong distortion— a pronounced dip—in $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$.

In FIG. 7, $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ calculated according to (94b) is plotted, along with the continuous-time GC $\mathcal{F}_{Y \to X}(\Delta)$ for $\rho = 0$, against Δ for a range of node relaxation time parameters $1/a, 1/b$ around the reference values (69). We see that the dip at $\tau/2 < \Delta < \tau$ closely approaches zero for small values of $1/a$ (i.e., fast relaxation of the X variable). As $1/a$ increases (left to right), the $\mathcal{F}_{Y \to X}(\Delta)$ plateaux becomes more peaked towards $\Delta = \tau$. As the Y relaxation time $1/b$ increases (top to bottom) we see that the subsampled GC $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ "pulls away" from the continuous-time GC, indicating a higher degree of distortion.

To further explore the parameter space of the model, in FIG. 8, $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ and $\mathcal{F}_{Y \to X}(\Delta)$ are again plotted against Δ

---

[25]We conjecture that the joint process is in fact minimum-phase for all parameters and subsampling intervals—indeed, we suspect that some results of Åström et al. (1984) for univariate rational transfer functions may generalise to cover our case—but have not succeeded in proving this conclusively due to the extreme algebraic complexity of the relevant condition.
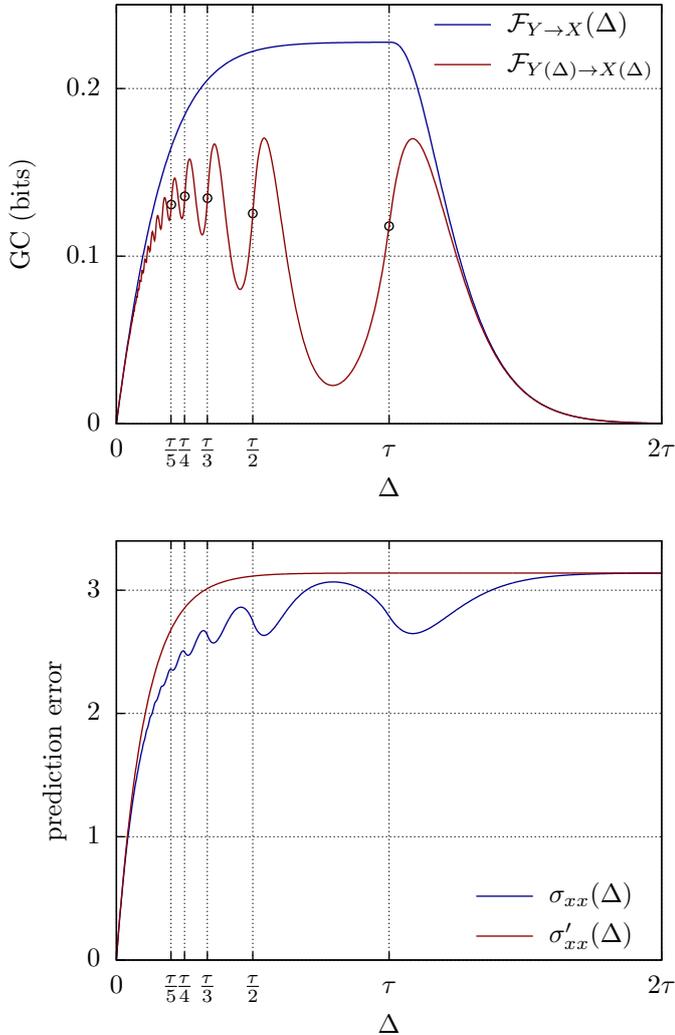
19

FIG. 6: *Top figure:* $\mathcal{F}_{Y \to X}(\Delta)$ (blue, eqs. 80,86) and $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (red, eq. 94a) plotted against sample interval Δ for the subsampled CTVAR process (68) with $\rho = 0$ and reference parameters (69). Black circles mark inflexion points of $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$. *Bottom figure:* subsampled process full prediction error $\sigma_{xx}(\Delta)$ (blue, eq. 92c) and reduced prediction errors $\sigma'_{xx}(\Delta)$ (red, eq. 93a) plotted against Δ.

for $\rho = 0$, for a selection of more extreme (if arguably biophysically implausible) parameters. In the top figure, the causal feedback delay τ is set to a similar time scale as the node decay times $1/a, 1/b$. Now subsampling has a comparatively mild distortional effect. We see that oscillations are weak, the plateau below $\Delta = \tau$ has virtually disappeared, GCs peak between τ and 2τ and decay of GCs for $\Delta > \tau$ is more gradual. The inflexional points for $\Delta \leq \tau$ no longer lie near $\tau, \tau/2, \tau/3, \ldots$. In the middle figure, the decay time $1/a$ for X is small (fast decay), while the decay time $1/b$ for Y is large (slow decay). Compared to the reference parameter settings, distortion is comparatively mild (although oscillations are sharply peaked) and we see that for small Δ the subsampled GC no longer lies completely below the continuous-time GC (*cf.* the discussion at the end of Section 3.3). In the bottom figure the situation is reversed: X decay is slow, while Y decay is fast. Distortion is now strong; subsampled GC is markedly smaller than continuous-time GC (implying poor detectability under subsampling) and oscillations are pronounced. We note that the c parameter (causal feedback strength) has a comparatively small qualitative effect on Granger causalities vs. subsample increment.

4.3. Statistical inference: detection of causalities

Next, we analyse the effects of subsampling frequency on the ability to detect non-zero continuous-time GC. We quantify detectability—i.e., statistical power—via the *Type II error (false negative) rate*: that is, the probability of failure to reject the null hypothesis that $\mathcal{F}_{Y \to X} = 0$, at a given significance level α. This is given by (Appendix D.2)

$$P_{II}(x; \alpha) = F_x \left(F_0^{-1}(1 - \alpha) \right) \tag{96}$$

where F_x denotes the cumulative distribution function (CDF) for the estimator $\widehat{\mathcal{F}}_{Y \to X}$, given that the actual causality $\mathcal{F}_{Y \to X}$ is equal to x [26]. Note that the distribution of $\widehat{\mathcal{F}}_{Y \to X}$ depends on the estimation method. Notwithstanding the potential advantages of the state-space approach (Section 2.1), all Granger causality estimates in this and the following Section are based on VAR modelling[27], for reasons outlined in Appendix D.1 - principally, the current lack of an analytical expression for the (asymptotic) distribution of $\widehat{\mathcal{F}}_{Y \to X}$ in the state-space case. We did in fact repeat, as far as possible, all experiments in this and the following Section via state-space modelling, using surrogate data techniques where appropriate to estimate distributions for $\mathcal{F}_{Y \to X}$. Results (not shown) did not differ significantly from the VAR case for our bivariate models with uni-directional causality; see Section 5.4 for further discussion.

For $\widehat{\mathcal{F}}_{Y \to X}$ based on a maximum-likelihood VAR model estimate of order p, for large sample size (i.e., number of observations) m, we have, asymptotically, $mF_x \sim \chi^2(d; mx)$ with $d = p n_x n_y$ degrees of freedom, where n_x, n_y are the dimensions

___

[26]Here $\mathcal{F}_{Y \to X}$, $\widehat{\mathcal{F}}_{Y \to X}$ and the distribution F_x refer to *discrete-time, subsampled* GC, since this is what is estimated in an empirical setting.

[27]VAR models were, however, converted to state-space form as a computational device, as explained in Appendix D.1; this does *not* affect $\widehat{\mathcal{F}}_{Y \to X}$, which is still distributed as for VAR-based causality estimation.

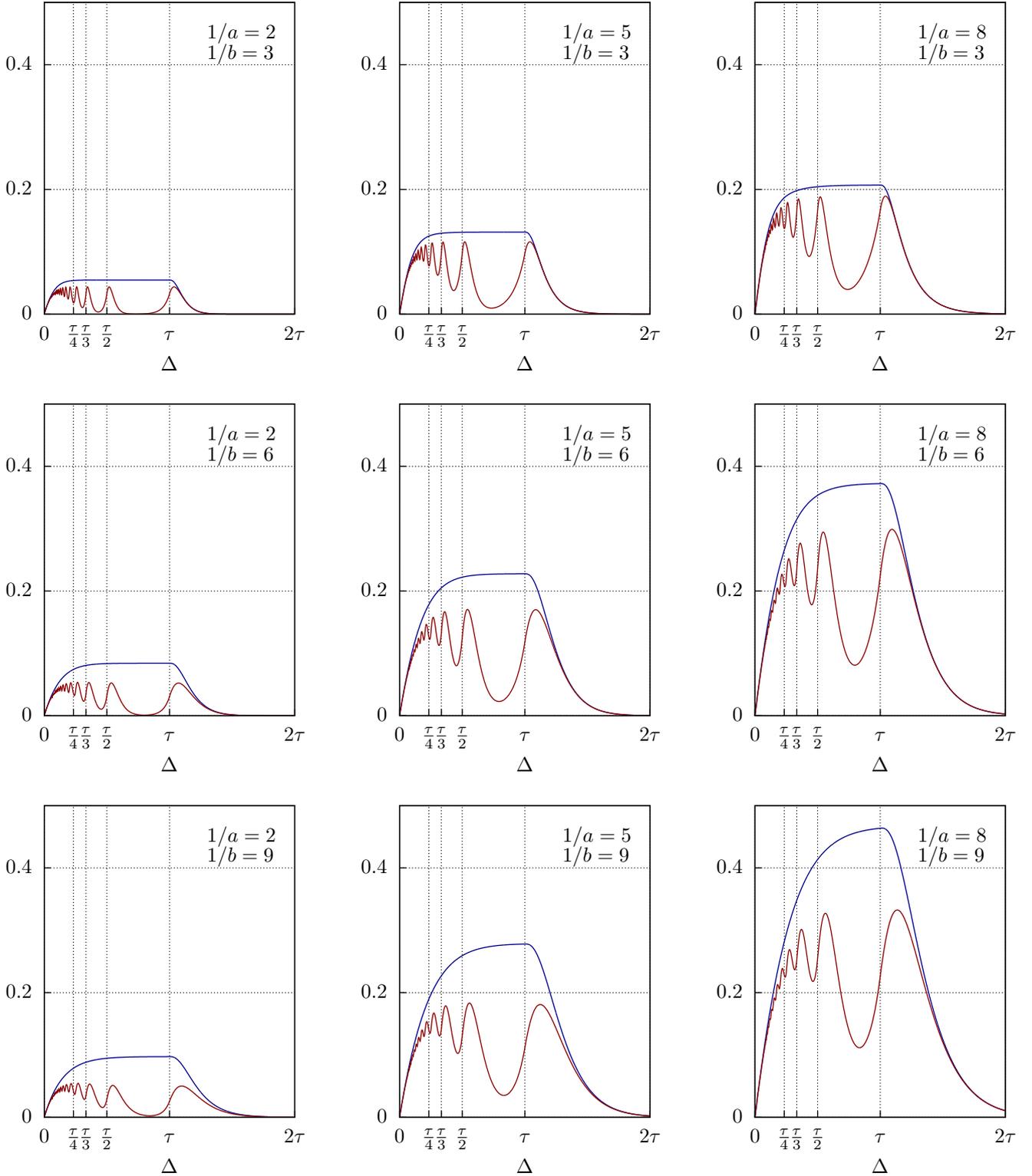FIG. 7: $\mathcal{F}_{Y \to X}(\Delta)$ (blue) and $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (red) plotted against sample interval Δ for the subsampled CTVAR process (68) with $\rho = 0$. Parameters are as in (69) except for a and b: $1/a$ takes the values $2, 5, 8$ from left to right, while $1/b$ takes the values $3, 6, 9$ from top to bottom (the centre figure thus corresponds to the exact reference parameters).
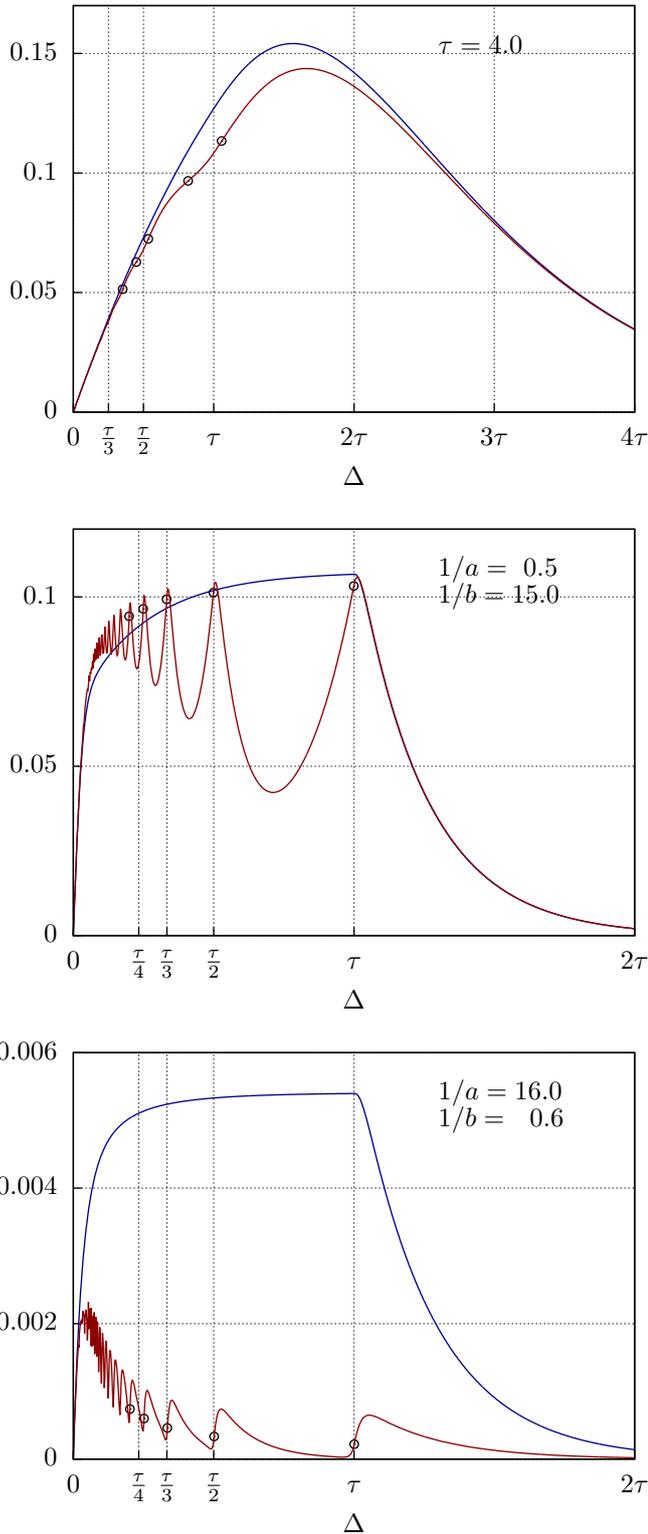
FIG. 9: Model orders (sample median) as estimated by the AIC from 1000×8000 ms realisations of the minimal CTVAR (68) with reference parameters (69) and $\rho = 0$, plotted alongside the approximation $p^*(\tau, \Delta)$ of (97). Error bars indicate ± 1 median absolute deviation. (Note the log scale.)

of the variables X, Y respectively (Appendix D). Since our sub-sampled minimal CTVAR is VARMA, rather than (finite-order) VAR (Section 4.2), in order to calculate $P_{II}(x; \alpha)$ (96) we need to select an empirical VAR model order p for the subsampled process model appropriate to the sample size m. Here we use the standard Akaike Information Criterion (AIC) (McQuarrie and Tsai, 1998) for model order selection (Appendix D). Note that, unlike our results so far, which have been purely analytical, this requires simulation. For a wide range of parameter values we simulated the CTVAR process (see Appendix H for details of our simulation method) and estimated the optimal model order according to the AIC. Results revealed that for a wide range of parameter values the optimal model order at sample interval Δ is well approximated by

$$p^*(\tau, \Delta) \equiv \max\left(\left[\tau/\Delta\right], 1\right) \tag{97}$$

(where $[\cdot]$ denotes rounding to the nearest integer) independently of the number m of observations (FIG. 9). This is intuitively reasonable[28]: although the joint process is VARMA$(2, q + 1)$, and thus in theory may only be represented as an infinite-order VAR, we might expect that with a sample

FIG. 8: $\mathcal{F}_{Y \to X}(\Delta)$ (blue) and $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (red) plotted against sample interval Δ for the subsampled CTVAR process (68) with $\rho = 0$, for a selection of parameter settings (note differing time and GC scales). Parameters are set to the reference parameters (69) unless otherwise labelled in the plots (see main text for discussion).
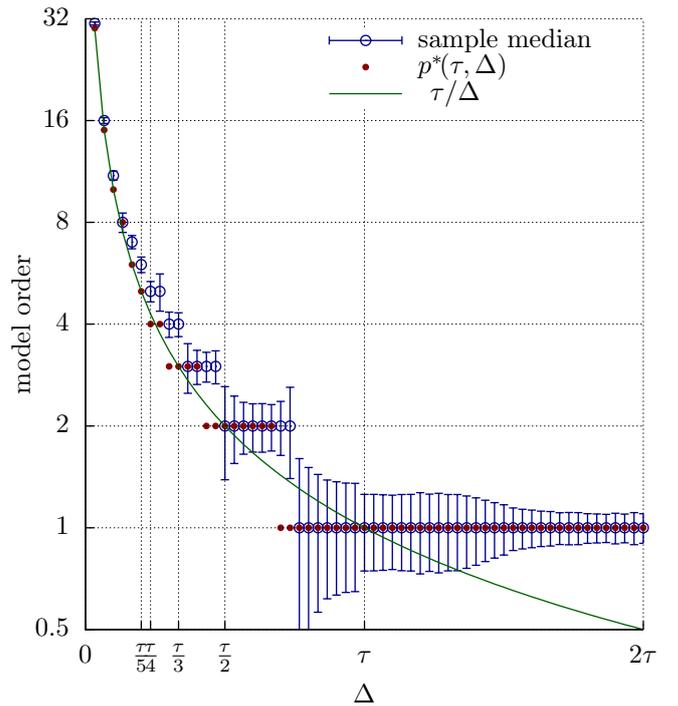
---

[28]For models with more causal delays and more complex interactions between variables, we might expect a more intricate dependence of empirical VAR model order on model parameter values. Our intuition is, however, that our broad conclusions on detectability which follow from the simple form of (97) would not change drastically.

22

interval of Δ, at least τ/Δ autoregression lags will be required to capture feedback at the causal delay τ.

Next we must consider the relationship between sample size m and sample interval Δ. We consider a scenario where we have a CTVAR time series of fixed duration T (measured in ms). The number of subsampled observations is then $m = \lfloor T/\Delta \rfloor$. This scenario corresponds to a realistic use case, where the experimenter has available an electrophysiological recording (at some base sampling frequency) of a given duration, and then has the option of downsampling the (discrete-time) recorded data.

We calculated $P_{II}(x; \alpha)$ from (96) at significance level[29] $\alpha = 0.05$. For a range of sample intervals Δ, we used the χ^2 asymptotic distribution of $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$ given the known actual causality $x = \mathcal{F}_{Y(\Delta) \to X(\Delta)}$ as calculated from (94b). Model order was specified by the approximation $p^*(\tau, \Delta)$ of (97); note that while (97) was empirically derived, the calculation of $P_{II}(x; \alpha)$ is purely analytic.

Results for the reference parameters and a range of data lengths from $T = 500\text{ms} - 8000\text{ms}$ are illustrated in FIG. 10. There are three competing effects at play here: firstly, the dependence of (Δ-subsampled) causal magnitude $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ on Δ (FIG. 6), secondly the decrease in sample size m as Δ increases and thirdly the decrease in model order (97) as Δ increases. Note that the last two factors, which affect the dispersion of the sample statistic, pull in opposite directions[30]. In the large Δ limit, we see that, as a result of exponential decay of the actual causality $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (FIG. 7), the ability to detect causality in the $Y \to X$ direction degrades abruptly and rapidly at a point beyond the causal delay τ. At the other extreme, if the subsampling frequency is too high ($\Delta \to 0$) detectability also becomes impossible as $\mathcal{F}_{Y(\Delta) \to X(\Delta)} \to 0$ (FIG. 6), exacerbated by the associated increase in model order and concommitant high variance of the sample statistic.

In between these extremes, we see a sequence of detectability "sweet spots" and "black" spots - values of Δ which locally minimise (resp. maximise) the Type II error rate. There is an optimal sweet spot (i.e., a value of Δ which *globally* minimises the Type II error rate) for detectability at a small sampling interval well below the causal delay τ (FIG. 11; for large data lengths it is hard to see these in FIG. 10, since they become indistinguishable from zero)[31]. This is followed by a series of oscillations in detectability as Δ increases. We see black spots—particularly just below the causal delay—where

detectability becomes difficult or unfeasible, then recovers, before finally degrading beyond τ. We also see that if the data length is too small, detectability becomes unreliable (or virtually impossible) at *any* subsample frequency.

We then tested our analytical predictions for subsampled GC $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (94b) and Type II error rate $P_{II}(x; \alpha)$ (96) against large-sample simulations, under a more realistic methodology where, rather than using the approximate model order $p^*(\tau, \Delta)$ of (97), model orders were estimated using the AIC per sample realisation of the CTVAR process[32]. For each sampling interval in the range $\Delta = 1\,\text{ms}$ to $\Delta = 2\tau = 60\,\text{ms}$, 10,000 stationary realisations of length 8000 ms were generated. Causalities were calculated in sample using VAR model estimation and a state-space computational method (Appendix D.1), as described previously. FIG. 12 (top figure) plots the mean $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$, de-biased according to (D.4) (note that due to sample fluctuations values can become slightly negative), for data sequences of length 8000 ms (points, with error bars at 95% empirical confidence intervals[33]), against the theoretical value $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ calculated according to (94b) (red line). The continuous-time GC $\mathcal{F}_{Y \to X}(\Delta)$ at horizon Δ is also displayed (black line). We see excellent agreement of the sample estimates with theory. In the bottom figure, Type II error rates at significance level $\alpha = 0.05$ (calculated as the fraction of sample causality values for which the corresponding p-value is $> \alpha$) are plotted against values calculated as before according to the theoretical χ^2 distributions of $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$. Note that in this figure[34] the error bars around the mean (again at 95% confidence intervals) apply to the *theoretical* values of P_{II}, since (for fixed Δ) these vary with the varying model orders as estimated per-sample by the AIC. Again, agreement with theory is excellent (since error rates are calculated on the basis of *asymptotic* statistics, we expect to see some deviation from theoretical predictions for larger values of Δ, where the number of observations is smaller), and we see clear evidence of a detectability black spot at the peak between $\Delta = \tau/2$ and $\Delta = \tau$.

For completeness, we repeated the above experiment, this time in the non-significant $Y \to X$ direction, along with estimation of the *Type I (false positive) error rate*. Results are displayed in FIG. 13. The top figure plots the mean de-biased $\widehat{\mathcal{F}}_{X(\Delta) \to Y(\Delta)}$ for data sequences of length 8000 ms, with error bars at 95% confidence intervals. Some structure in the variation of the sample statistic with Δ is apparent, with an expected increase in variance with increasing Δ. In the lower figure, Type I error rates at a significance level of $\alpha = 0.05$ are plotted. There is again some apparent structure in the variation of the error rate around the theoretical value of α, in particular a peak just before $\Delta = \tau$. For larger values of Δ the Type I error rate is somewhat

---

[29] As noted in Appendix D, a multiple hypothesis correction ought to be made for joint significance testing of causalities in both the $Y \to X$ and $X \to Y$ directions. For simplicity we don't apply any correction here, but note that e.g., for a Bonferroni correction (Hochberg and Tamhane, 1987), this would be equivalent to halving the significance level α.

[30] For parametric statistical estimation, decreasing the model order will result in fewer parameters to estimate and consequently better parameter estimation and lower dispersion of the sample statistic, while decreasing the sample size will result in a noisier model fit, resulting in increased dispersion.

[31] We remark that if, in contrast to the considered usage scenario, the number of *observations m* is held fixed and the duration of the signal T allowed to vary, then numerical computation (results not shown) reveals that the optimal sample interval does not change with the number of observation, and lies slightly *above* the causal delay at τ. We have not been able to establish this analytically. This scenario might, however, be considered less typical.

[32] If the AIC yielded a model order of zero (i.e., it "sees" the process as pure white noise), the model order was set to 1.

[33] Empirical confidence intervals were constructed at level α so that a fraction $1 - \alpha/2$ of the data points lie below the upper bound and the same fraction lie above the lower bound.

[34] For the value $T = 8000$ ms in the lower figure, the fine oscillatory structure in the value of $P_{II}(x; \alpha)$ at small values of Δ, visible for smaller T in FIG. 10, is not apparent; however, smaller T results in noisier estimation, compromising visual clarity.
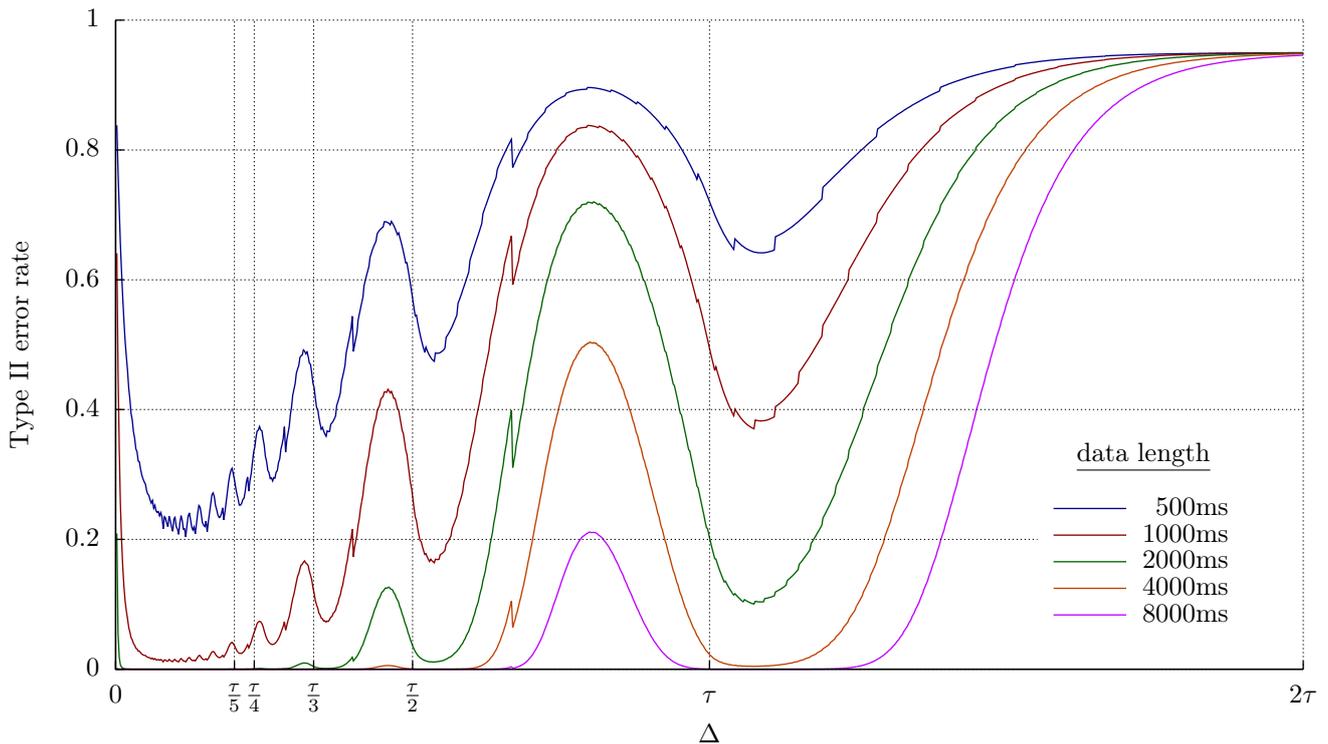
FIG. 10: Theoretical Type II error rate $P_{II}(x; \alpha)$ (D.5) at significance level $\alpha = 0.05$ for the CTVAR (68) with reference parameters (69) and $\rho = 0$, plotted against sample interval for a range of data sequence lengths. Calculations are based on model order $p^*(\tau, \Delta)$ of (97) (*cf.* FIG. 9).
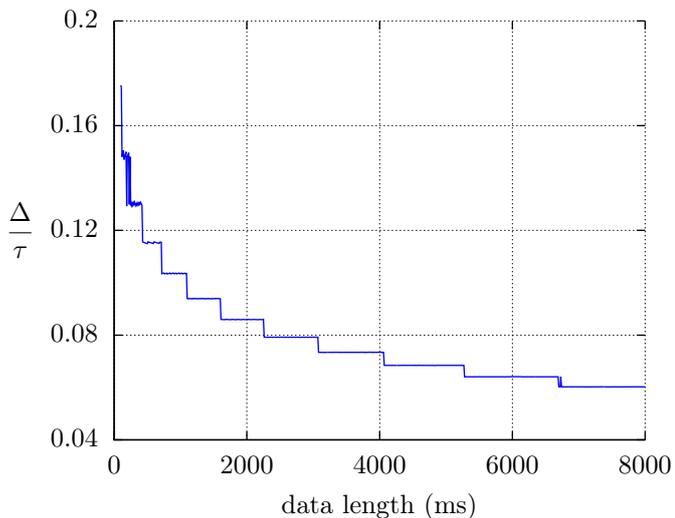

FIG. 11: Optimal detectability sweet spot Δ as a fraction of causal delay τ, plotted against data length.

higher than the theoretical value of $\alpha = 0.05$; this may be explained by deviation of the sampling distribution of $\widehat{\mathcal{F}}_{X(\Delta) \to Y(\Delta)}$ from the theoretical (null) distribution—which again is asymptotic rather than exact—and also the failure of finite-order VAR modelling to capture the VARMA characteristics of the subsampled process.

We note that for the minimal CTVAR with reference parameters (69) and $\rho = 0$, we have $\mathcal{F}_{X(\Delta) \cdot Y(\Delta)} \ll \mathcal{F}_{Y(\Delta) \to X(\Delta)}$, so that Solo's strong causality measure $\mathcal{F}_{Y(\Delta) \to X(\Delta)}^{\text{strong}}$ of (22) (Solo, 2007) is virtually indistinguishable from $\mathcal{F}_{Y(\Delta) \to X(\Delta)}$ (this was also verified in simulation).

5. Discussion

A key question when applying Granger causality to empirically sampled data, is how the relationship between the sampling rate and the time scale(s) of the underlying neurophysiological process affects Granger causal inference. To address this, we introduce CTVAR processes—a generalisation of the standard vector Ornstein-Uhlenbeck process featuring finite-time, distributed lags—which capture the analogue, continuous-time nature of the underlying signal sampled by neurophysiological recording technologies. We develop a comprehensive theoretical basis, in both time and frequency domain, for CTVAR processes. We then analyse the process derived from a CTVAR process by subsampling at fixed time increments Δ. We show how key quantities such as the autocovariance, CPSD, transfer function and residual noise covariance
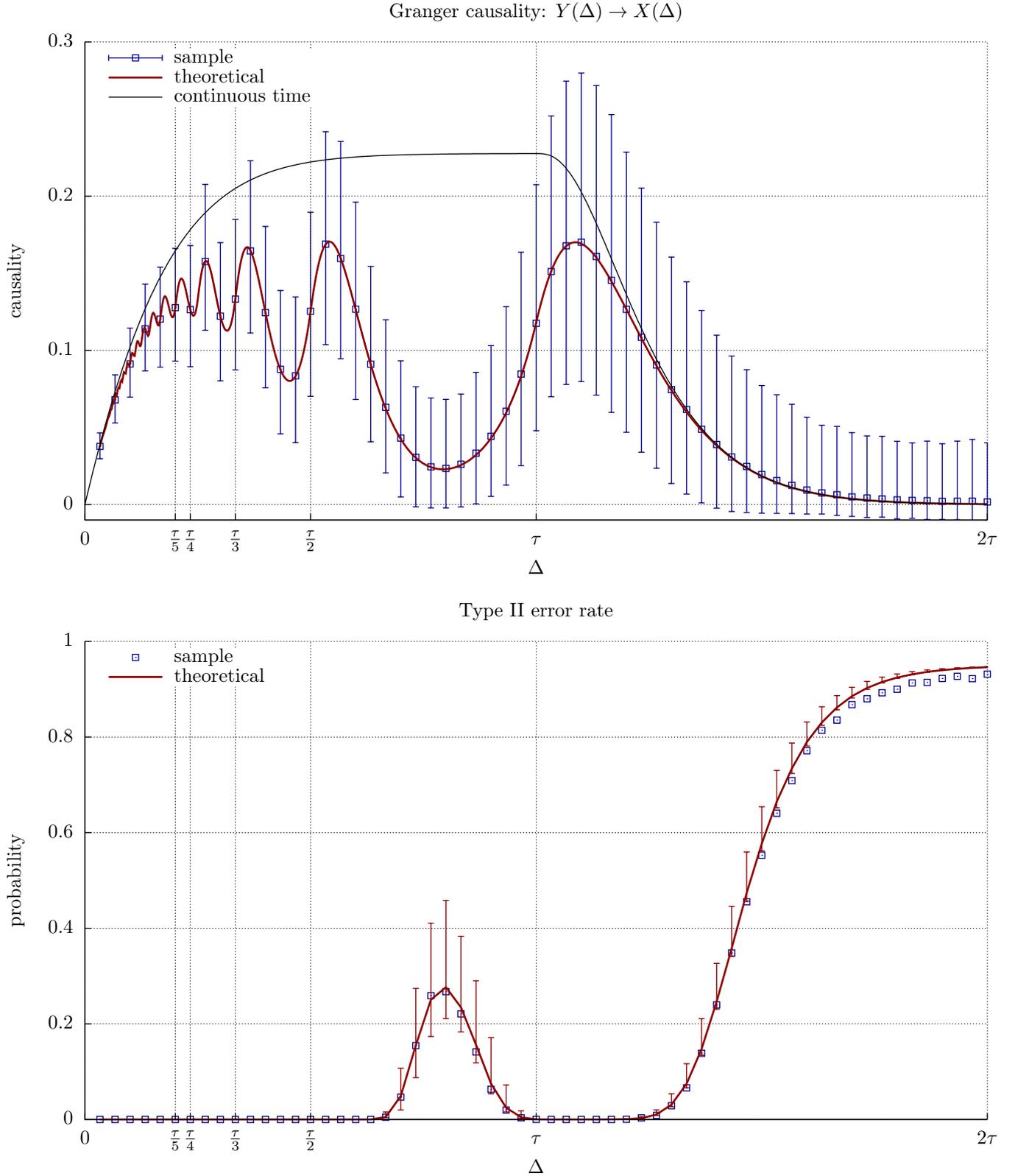
FIG. 12: Top figure: distribution of the de-biased (significant) sample Granger causality $\widehat{\mathcal{F}}_{Y(\Delta) \to X(\Delta)}$ for the minimal CTVAR (68) with reference parameters (69) and $\rho = 0$, based on $10,000$ realisations of length $8000\,\text{ms}$, plotted against sample interval Δ. Points denote the mean, while error bars depict 95% confidence intervals. The solid red line plots theoretical values (94b), while the solid black line plots the continuous-time GC $\mathcal{F}_{Y \to X}(\Delta)$ at horizon Δ. Bottom figure: Type II error rates at significance level $\alpha = 0.05$ for the same simulations (points). The solid red line plots the mean theoretical values (D.5) based on χ^2 distributions, with error bars at 95% confidence intervals (note that the dispersion in theoretical values is due to variance of the AIC-estimated model orders - see text for details).
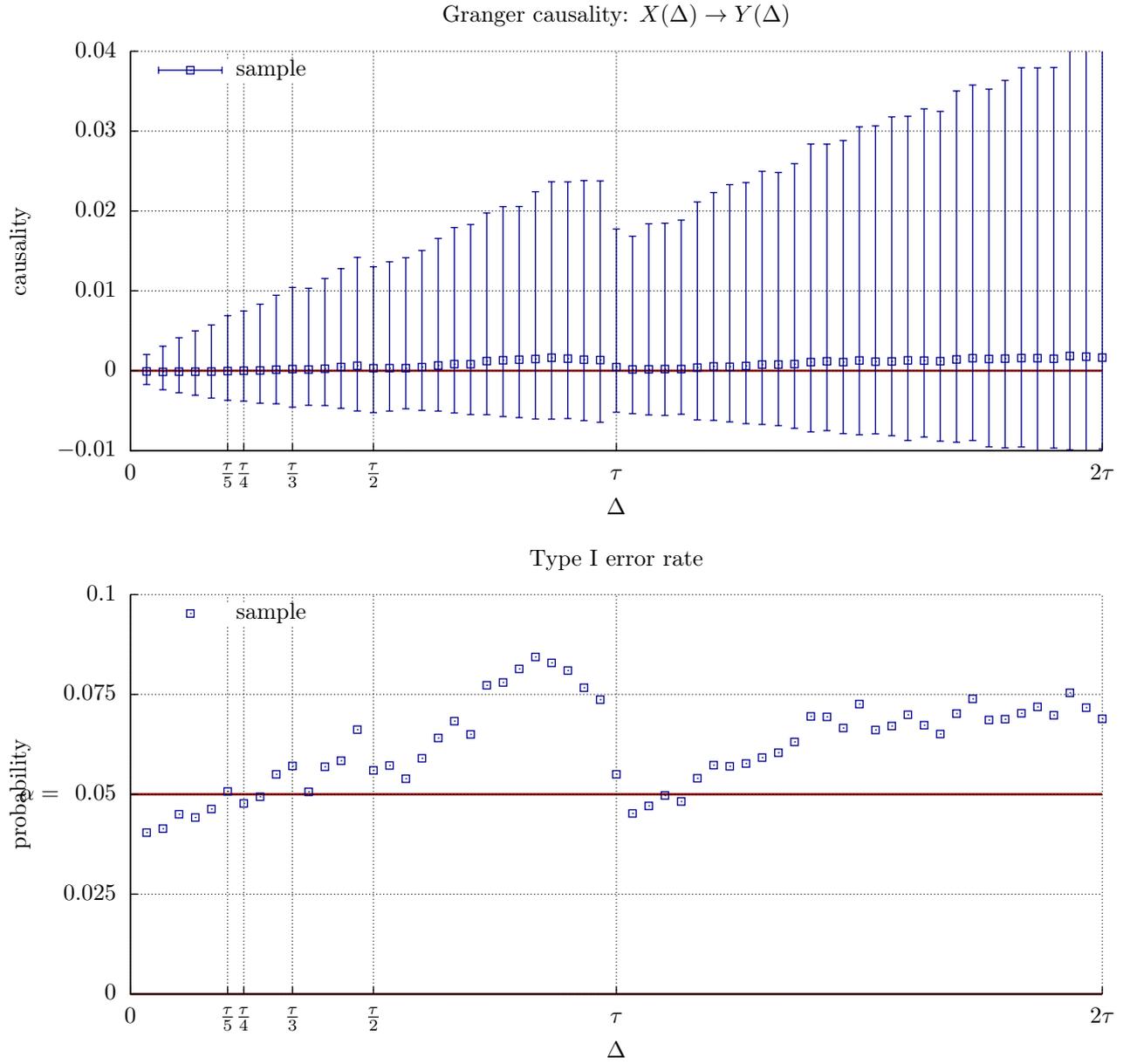
25

FIG. 13: Top figure: distribution of the de-biased (non-significant) sample Granger causality $\widehat{\mathcal{F}}_{X(\Delta) \to Y(\Delta)}$ for the CTVAR (68); parameters as in FIG. 12. Bottom figure: Type I error rates at significance level $\alpha = 0.05$ for the same simulations (points). Note that here the theoretical causality is zero, while the theoretical Type I error rate is just α (solid line).

associated with the VAR model of the subsampled process may be derived from the parameters of a CTVAR model by spectral factorisation, and demonstrate that in the limit $\Delta \to 0$ these quantities approach their continuous-time counterparts.

Then, under the premise that Granger causality in continuous time is best considered at *finite* prediction horizons, we develop a principled theoretical foundation for Granger causality in continuous time, based on the CTVAR formalism. For subsampled CTVAR processes, continuous-time GC appears as a natural limit of multistep discrete-time GC in the limit of a small subsampling increment. The existence of a zero-horizon Granger causality *rate* is demonstrated, and shown to satisfy a continuous-time Geweke frequency decomposition. The properties of subsampled CTVAR processes are discussed, including the possibility of spurious causalities, and the failure of finite-horizon GC invariance under causal, invertible filtering.

Armed with this theoretical background, we consider the effects of subsampling on the *detectability* of Granger causality. We proceed to an exact analytical solution of the subsampling problem for a minimal bivariate CTVAR process with finite causal delay, which we believe to be the first non-trivial full analytic solution in the literature for Granger causality in continuous time and subsampled continuous time. Analytic expressions are calculated for the autocovariance sequence, transfer function and CPSD of the subsampled process. We are able to factorise both the continuous-time and (in the case of uncorrelated residual noise) subsampled CPSD explicitly, leading to analytic expressions for Granger causalities for both the continuous-time and subsampled processes. This facilitates a detailed analysis of the effects of subsampling on detectability. Theoretical predictions are confirmed by large-sample simulations.

5.1. Relationship to previous work

Distributed-lag continuous-time processes have previously been considered in the econometrics literature, going back to Sims (1971, 1972); Geweke (1978), who also considered subsampling. Distortional effects of subsampling on Granger causality, in particular spurious causality, have been variously addressed, in both continuous time (Florens and Fougère, 1996; Comte and Renault, 1996; Renault et al., 1998; McCrorie and Chambers, 2006) and discrete time (Wei, 1981; Marcellino, 1999; Breitung and Swanson, 2002; Solo, 2007, 2016; Zhou et al., 2014).

Closest in spirit to our work is the "CIMA" (continuous-time invertible moving-average) model introduced by Comte and Renault (1996). There, however, only statistical test criteria for (non)causality are considered; the authors stop short of introducing, as we do here, statistical Granger-Geweke measures which quantify *magnitude*, rather than just (non)existence, of Granger-causal effects. Our measures, furthermore, are defined in terms of limits of the corresponding discrete-time subsampled quantities under progressively finer subsampling.

Solo (2007) distinguishes between "strong" and the conventional "weak" Granger causality in discrete time (Section 2.1), noting that only the former is strictly preserved under subsampling. We argue, however, that strong GC is unsatisfactory as

a *directional* causality measure, as it inextricably combines the effects of time-directed and contemporaneous feedback. Using a state-space approach, Solo (2016) expands on the distortion induced by subsampling, including the possibilty of spurious causality [we note (Section 4.2) that for our minimal CTVAR spurious GC does not in fact arise].

Zhou et al. (2014) consider Granger causality for discrete-time VAR models and a continuous-time integrate-and-fire neural simulation. They report oscillations at a near-constant frequency in estimated causalities plotted against sample interval, with causalities almost vanishing in the troughs. As we have noted, such oscillations are lacking in our minimal CTVAR. This is consistent with the explanation put forward by Zhou *et al.* (demonstrated also for discrete-time VAR models) that these oscillations are associated with periodic behaviour in the pre-subsampled signal, which manifest as peaks in the power spectra. For our minimal CTVAR we note that there are no peaks away from $\lambda = 0$ at $\Delta = 0$ (FIG. 5) - underlining that the simplicity of the model helps isolate the effects of subsampling on GC inference.

5.2. Original contributions

The original contributions of this study may be briefly summarised as:

1. Theoretical analysis of CTVAR continuous-time, distributed-lag vector autoregressive processes as an appropriate model for Granger-causal analysis of neural systems.

2. Natural and intuitive definitions of quantitative finite- and zero-horizon continuous-time Granger-Geweke measures, as limits of corresponding discrete-time quantities; these measures are proposed as "ground truth" targets for functional inference of neural systems based on discretely sampled neurophysiological recordings.

3. Analysis of the relationship between continuous-time GC for a CTVAR process and discrete-time GC for a subsampling of the process.

4. Demonstration of the non-invariance of finite-horizon GC under causal, invertible filtering.

5. Complete analytical solution of a minimal, but non-trivial bivariate finite-lag CTVAR, both in continuous time and subsampled, revealing:

   (a) Exponential decay of subsampled GC for sampling intervals beyond causal feedback time scales.

   (b) Resonance between causal and sampling frequencies, resulting in oscillations in subsampled GC for sampling intervals below causal feedback time scales.

6. Analysis of the *detectability* of Granger causalities under subsampling. This analysis reveals, in particular, (a) exponential decay of detectability beyond causal feedback time scales, and (b) the existence of detectability black spots and sweet spots in the sample rate.

5.3. Implications for causal inference from neurophysiological recordings

Analysis of the minimal CTVAR indicates that Granger causality in continuous time decays exponentially with increasing prediction horizon beyond causal feedback time scales, leading to exponential decay of subsampled GC with increasing sampling interval and a concomitant sharp drop in detectability. We conjecture that a similar effect will hold under more general conditions. Our analysis also reveals that, insofar as a choice of sample rates is available, faster is not necessarily better: results for the minimal CTVAR indicate that, for a data segment of fixed length, detectability approaches zero as the sample interval $\Delta \to 0$. This may be viewed in the context of a more general phenomenon: that for some class of systems there is a finite optimal sampling rate for *system identification* of a continuous-time model from subsampled data. Åström (1969), for example, proves that there is a finite optimal sampling rate for identification of the parameters of a (univariate) Ornstein-Uhlenbeck process, above which the variance of parameter estimates increases without bounds.

For our minimal CTVAR, then, there is an optimal sweet spot for sampling frequency that maximises detectability. Furthermore, as Δ decreases below the causal delay, Granger causality oscillates as the sampling frequency $1/\Delta$ resonates with the causal delay frequency $1/\tau$. These oscillations are distinct from the fixed-period oscillations noted by Zhou et al. (2014) and have not, as far as we are aware, been reported previously. They may manifest in detectability black spots below the causal delay. Again, we conjecture that these phenomena generalise, although in the case of distributed causal delays sub-causal oscillations are likely to be "smeared". An important implication for Granger-causal inference from electrophysiological recordings such as EEG, MEG, LFP, etc., where the sample rate may be high ($> 1000\,\mathrm{Hz}$) compared with neural time scales, is that downsampling may well be advantageous (anecdotally, it is common in these contexts to downsample to $100 - 500\,\mathrm{Hz}$ or slower). But this result also suggests, further, that to maximise detectability a range of sampling frequencies should be tested, so as to locate sweet spots and avoid black spots. In addition, we remark (*cf.* Section 3.1) that too high a sampling rate could potentially lead to a failure of the minimum-phase condition for the subsampled signal (Åström et al., 1984).

Exponential decay of detectability has particularly relevance for Granger-causal inference from fMRI recordings (Seth et al., 2013), since typical fMRI sample rates (currently $0.5 - 3$ secs) are substantially slower than typical synaptic delays ($10 - 50\,\mathrm{ms}$) in neural systems. fMRI/GC is already controversial for other reasons. Seth et al. (2013) argue that confounds due to regional variation of the hemodynamic response function (HRF), which mediates the generation of the BOLD (Blood-Oxygen-Level Dependent) signal from the underlying neural activity, are mitigated by the filter-invariance property of discrete-time (1-step) GC (Barnett and Seth, 2011). Solo (2016) claims that, while arguably causal, the HRF is unlikely to be minimum phase, so that filter invariance fails. Recent work[35], however,

---

[35]Personal communication.

casts some doubt on this contention, on the grounds that the HRF model analysed by Solo (2016) may be overly simplistic and thereby misleading. Notwithstanding, our finding that (non-zero) *finite-horizon* GC in continuous time is *not* in general invariant under filtering, implies that the HRF—*even if causal and invertible*—may still distort magnitudes and impact detectability of non-zero causalities at the neural level; that is, even if we could capture the BOLD signal in continuous time, GC analysis might still fail to reflect "ground truth" GC at the neural level. Importantly, filter-induced distortion will not give rise to spurious causalities. Note that this potential confound is distinct from subsampling-induced distortions identified for the various fMRI/GC scenarios discussed in Solo (2016), since only 1-step discrete-time GC is considered there. It is, however, far from clear what form finite-horizon GC filter-induced distortion might take for HRF-type filtering; more research is required.

5.4. General remarks, limitations and future research directions

Our research into continuous-time models and subsampling has thrown up a number of conjectures, caveats, technical issues and potential extensions, for which further research is required. These include:

1. Further investigation is required into the precise conditions under which subsampling a CTVAR process may induce spurious causalities.

2. While our analysis of the CTVAR model and continuous-time GC extends for the most part to the important case of *conditional* Granger causality (Geweke, 1984), the situation with subsampling is inevitably more complex. In particular, for conditional GC, eq. 19 in discrete time (Lütkepohl, 1993) and eq. 65 in continuous time (Dufour and Renault, 1998) no longer hold; that is, noncausality at the immediate prediction horizon will not generally imply noncausality at larger prediction horizons, since causal effects may propagate through an auxiliary variable. Since neurophysiological time series are in general highly multivariate, further research is necessary to extend our work beyond the bivariate scenario addressed in this study. Given the complexity already apparent in our minimal CTVAR model, this is likely to be highly challenging analytically, and empirical studies will probably be required.

3. Our minimal CTVAR restricts detailed analysis to the case of unidirectional causality and a single causal feedback delay. In the case of multiple (possibly bidirectional) feedback at varying delays, the situation will be far more complex. We would expect that multiple resonances bewteen sample frequency and causal feedback frequencies will result in more complex oscillations that, we suspect, may manifest in distortion and compromised inference of *relative* GC magnitudes by subsampling. Further research, both analytic and empirical, is required.

4. We have also, in this study, omitted analysis of the effects of measurement noise on Granger-causal inference (Solo, 2007), and its interaction with subsampling and filtering.

5. Our definition of CTVAR processes is not mathematically complete in terms of the (spectral) conditions under which our requirements that (i) a subprocess of a CTVAR be a CTVAR and (ii) a subsampled CTVAR be a VAR will be satisfied. In addition, we have not identified precise conditions on a continuous-time CPSD which will guarantee a unique spectral factorisation.

6. It seems likely that a principled definition for continuous-time finite-horizon spectral Granger causality exists. Such a measure should satisfy the Geweke frequency decomposition $\mathcal{F}_{Y \to X}(h) = \int_{-\infty}^{\infty} \mathrm{f}_{Y \to X}(\lambda; h) \, d\lambda$.

7. It would be of interest to investigate the relationship between subsampled GC resonance oscillations (and/or post-delay maxima) as observed in the minimal CTVAR and the general problem of inference of feedback time scales in linear systems (Björklund, 2003; Wibral et al., 2013).

8. As mentioned in Section 4, there is a substantial existing literature on Stochastic Delay-Differential Equations (SDDEs) (Longtin, 2010), which, although in general not focused on Granger-causal/information-theoretic analysis, may be useful in that direction. A related promising research direction, particularly for non-stationary and/or nonlinear systems in continuous time, is the inference of causal/driving mechanisms via Fokker-Planck equations (Prusseit and Lehnertz, 2008; Wahl et al., 2016).

In a more general vein, There is a growing consensus that state-space modelling should be the preferred method—at least in discrete time—for performing Granger-causal inference (Solo, 2007; Valdes-Sosa et al., 2011; Seth et al., 2013; Friston et al., 2014; Seth et al., 2015; Barnett and Seth, 2015; Solo, 2016). This consensus is based on observations that state-space processes—unlike VAR processes—are closed under subsampling, the addition of additive noise, linear digital filtering, and subprocess extraction). Consequently, powerful and efficient new methods for Granger-causal analysis of state-space systems have now been developed (Barnett and Seth, 2015; Solo, 2016). It is not clear, however, how we might extend the state-space/GC paradigm to continuous time *with distributed lags*, since a naïve approach would seem to require an infinite-dimensional state space. We have also noted a current sticking point for state-space Granger-causal inference: that the sampling distribution for Granger causality statistics based on (maximum-likelihood) state-space model estimation remains unclear, necessitating computationally costly surrogate methods for statistical inference; more research, both theoretical and empirical, is required in this area. A further promising avenue of research is the deployment of state-space methods to reconstruct models at finer time scales than the sampling frequency

5.5. Summary

Granger causal analysis of subsampled time-series data will inevitably be susceptible to distortion, and detectability in particular will degrade as the sampling interval increases beyond the natural time scale(s) of the underlying process. This is to be expected, since subsampling results in the loss of the predictive information which underpins Granger causality. In this study we have characterised how subsampling affects detectability through an exact analytic solution of the subsampling problem for a continuous-time processes with finite causal delay. Our analysis reveals a rapid decay of detectability for large subsampling intervals, but also the existence of detectability 'black' and "sweet" spots as the sampling frequency interacts with the underlying generative time scale(s). The theoretical basis for these findings provides a very general framework for further investigations of statistical inference on sampled continuous-time processes with causal interactions over multiple time scales. This encompasses a very wide range of possible scenarios. Overall, our results indicate that Granger causality analysis will be most successful when data are sampled fast enough to capture the relevant causal time scales, but not so fast as to impair detectability. Thus Granger causality analysis will be most effective when informed by sensible priors about domain-specific time scales. Further research using state-space approaches may shed light on these issues.

Appendix A. Fourier transforms in discrete and continuous time

For a discrete-time sequence $x = \{x_k \,|\, k \in \mathbb{Z}\}$ (the x_k may be random or deterministic, real or complex, scalar, vector, matrix, etc.) with time step Δ, we define the (two-sided) Fourier transform $\hat{x} \equiv \{\hat{x}(\lambda) \,|\, -\infty < \lambda < \infty\}$ as

$$\hat{x}(\lambda) \equiv \Delta \sum_{k=-\infty}^{\infty} x_k e^{-2\pi i \Delta \lambda k} \qquad (A.1)$$

which is periodic in λ with period $f_s = 1/\Delta$, the sampling frequency. We shall often restrict $\hat{x}(\lambda)$ to the interval $-1/(2\Delta) \leq \lambda < 1/(2\Delta)$, where $1/(2\Delta) = f_s/2$ is the Nyqvist frequency. Note that we scale $\hat{x}(\lambda)$ by the sample interval Δ; this ensures that the transform has the same dimensions for discrete and continuous-time transforms (see below). The original sequence may be recovered via the inverse transform

$$x_k = \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} \hat{x}(\lambda) e^{2\pi i \Delta \lambda k} \, d\lambda \qquad (A.2)$$

For a continuous-time sequence $x = \{x(t) \,|\, t \in \mathbb{R}\}$, we define the Fourier transform $\hat{x} \equiv \{\hat{x}(\lambda) \,|\, -\infty < \lambda < \infty\}$ as

$$\hat{x}(\lambda) \equiv \int_{-\infty}^{\infty} x(t) e^{-2\pi i \lambda t} \, dt \qquad (A.3)$$

and the original sequence may be recovered via the inverse transform

$$x(t) = \int_{-\infty}^{\infty} \hat{x}(\lambda) e^{2\pi i \lambda t} \, d\lambda \qquad \text{(A.4)}$$

We note, in particular, that for a discrete-time sequence $x(\Delta) \equiv \{x(k\Delta) \,|\, k \in \mathbb{Z}\}$ obtained by subsampling the continuous-time sequence $x = \{x(t) \,|\, t \in \mathbb{R}\}$ at regular intervals Δ, we have

$$\lim_{\Delta \to 0} \widehat{x(\Delta)} = \hat{x} \qquad \text{(A.5)}$$

pointwise; i.e., for fixed λ, $\widehat{x(\Delta)}(\lambda) \to \hat{x}(\lambda)$ as $\Delta \to 0$, at least insofar as the sum in (A.1) converges to the (Rieman) integral in (A.3).

Appendix B. Invariance of discrete-time 1-step Granger causality under causal, invertible filtering

Let $\mathcal{G}(z) = \sum_{k=0}^{\infty} G_k z^k$ be a causal invertible filter (without loss of generality we assume $G_0 = I$). Stability and minimum-phase require that both $|\mathcal{G}(z)|$ and $|\mathcal{G}(z)|^{-1}$ are non-zero on the unit disc $|z| \le 0$. If $X_k = \Psi(z) \cdot \varepsilon_k$ is the MA representation of a VAR process, then clearly (we denote quantities relating to the filtered process with a tilde) $\tilde{X}_k \equiv \mathcal{G}(z) \cdot X_k = \tilde{\Psi}(z) \cdot \varepsilon_k$, where $\tilde{\Psi}(z) = \mathcal{G}(z)\Psi(z)$, will also be a (stable, minimum-phase) VAR with residual noise intensity $\tilde{\Sigma} = \Sigma$.

Suppose now that $[X^\intercal \ Y^\intercal]^\intercal$ is a joint VAR process, and that $\mathcal{G}_{xy}(z) \equiv 0$; i.e., $\mathcal{G}(z)$ is lower block-triangular. Let $X_k = \Psi'_{xx}(z) \cdot \varepsilon'_{x,k}$ be the reduced MA representation of X_k alone. Then $\tilde{X}_k = \mathcal{G}_{xx}(z) \cdot X_k = \tilde{\Psi}'_{xx}(z) \cdot \varepsilon'_{x,k}$ where $\tilde{\Psi}'_{xx}(z) = \mathcal{G}_{xx}(z)\Psi'_{xx}(z)$ so that the reduced residuals intensity $\tilde{\Sigma}'_{xx}$ of the filtered process $\tilde{X}$ alone is $\tilde{\Sigma}'_{xx} = \Sigma'_{xx}$. But $\tilde{\Sigma}_{xx} = \Sigma_{xx}$, so from (18) we have $\mathcal{F}_{\tilde{Y} \to \tilde{X}} = \mathcal{F}_{Y \to X}$. It follows straightforwardly from (12), (14) and (23) that $f_{\tilde{Y} \to \tilde{X}}(\lambda) = f_{Y \to X}(\lambda)$ for all λ.

Appendix C. Non-invariance of discrete-time multistep Granger causality under causal, invertible filtering

With reference to the argument in Appendix B, consider now the case $m > 1$: (16) gives $\tilde{\mathcal{E}}_m = \Delta \sum_{\ell=0}^{m-1} \tilde{B}_\ell \Sigma \tilde{B}_\ell^\intercal$, where the filtered MA coefficients are $\tilde{B}_\ell = \sum_{k=0}^{\ell} G_k B_{\ell-k}$. A moment's consideration reveals that, even with $\mathcal{G}_{xy}(z) \equiv 0$, $\tilde{\mathcal{E}}_m$ will not in general block-decompose conveniently, unless $\Psi_{xy}(z) \equiv 0$.

This is perhaps best illustrated by example. Consider the VAR(1)

$$X_k = aX_{k-1} + cY_{k-1} + \varepsilon_{x,k} \qquad \text{(C.1a)}$$
$$Y_k = bY_{k-1} \qquad\qquad + \varepsilon_{y,k} \qquad \text{(C.1b)}$$

with $|a|, |b| < 1$ (for stability) and $\Sigma = I$. We have $\Phi(z) = I - Az$ with $A = \begin{bmatrix} a & c \\ 0 & b \end{bmatrix}$, so that

$$\Psi(z) = \begin{bmatrix} \dfrac{1}{1-az} & \dfrac{cz}{(1-az)(1-bz)} \\ 0 & \dfrac{1}{1-bz} \end{bmatrix} \qquad \text{(C.2)}$$

We have $B_1 = A$, so that from (16)

$$\mathcal{E}_{2,xx} = [\Sigma + B_1 \Sigma B_1^\intercal]_{xx} = [I + AA^\intercal]_{xx} = 1 + a^2 + c^2 \qquad \text{(C.3)}$$

Spectral factorisation yields

$$S_{xx}(z) = [\Psi(z)\Psi(z)^*]_{xx} = \frac{|1-bz|^2 + c^2}{|1-az|^2 |1-bz|^2}, \quad |z| = 1 \qquad \text{(C.4)}$$

We seek a reduced spectral factorization of the form $S_{xx}(z) = \Psi''_{xx}(z)\Sigma'_{xx}\Psi''_{xx}(z)^*$ on $|z| = 1$ with

$$\Psi'_{xx}(z) = \frac{1-hz}{(1-az)(1-bz)}, \qquad \Sigma'_{xx} = \upsilon \qquad \text{(C.5)}$$

Comparing with (C.4), we obtain

$$\upsilon h = b \qquad \text{(C.6a)}$$
$$\upsilon\left(1 + h^2\right) = 1 + b^2 + c^2 \qquad \text{(C.6b)}$$

so that υ satisfies the quadratic equation $\upsilon^2 - 2D\upsilon + b^2$, $D \equiv \frac{1}{2}\left(1 + b^2 + c^2\right)$, and (C.6) has the solution[36]

$$\upsilon = D + \sqrt{D^2 - b^2}, \qquad h = \frac{1}{b}\left(D - \sqrt{D^2 - b^2}\right) \qquad \text{(C.7)}$$

From (18), the 1-step GC $\mathcal{F}_{Y \to X}$ is then just $\log \upsilon$ (Barnett and Seth, 2011). From (C.5), collecting the z^1 terms, we have $B'_{1,xx} = a + b - h$, so that from (16) we find

$$\mathcal{E}'_{2,xx} = \upsilon\left[1 + (a + b - h)^2\right] \qquad \text{(C.8)}$$

and from (17) with (C.3) we have

$$\mathcal{F}_{Y \to X, 2} = \log \frac{\upsilon\left[1 + (a + b - h)^2\right]}{1 + a^2 + c^2} \qquad \text{(C.9)}$$

Note that $\mathcal{F}_{Y \to X} = 0$ iff $c = 0$, in which case we may check that $\mathcal{F}_{Y \to X, 2} = 0$ as expected. Also, if $a + b = 0$ but $c \ne 0$, then from (C.6b) we see that $\mathcal{F}_{Y \to X, 2} = 0$ while $\mathcal{F}_{Y \to X} > 0$, confirming our observation in Section 2.1 that, for $m > 1$, vanishing $\mathcal{F}_{Y \to X, m}$ does not imply vanishing $\mathcal{F}_{Y \to X}$.

Now let us define the causal filter

$$\mathcal{G}(z) \equiv \begin{bmatrix} \dfrac{1-gz}{1-az} & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{(C.10)}$$

$\mathcal{G}_{xy}(z) \equiv 0$ and invertibility requires $|g| < 1$. Applying $\mathcal{G}_{xy}(z)$ to the MA operator (C.2) we see immediately that the filtered process is identical to the original process (C.1), but with a replaced by g. Then, since a appears explicitly in the expression (C.9) for $\mathcal{F}_{Y \to X, 2}$, we see that filter-invariance fails for 2-step GC (note that $\mathcal{F}_{Y \to X}$ does not depend on a, so that the 1-step GC is invariant under $\mathcal{G}$).

---

[36]It may be confirmed that the positive square root should be taken in the expression for υ; see e.g., Barnett and Seth (2011).

Appendix D. Estimation and statistical inference for discrete-time Granger causality

Further motivation for the definition (18) of (time domain) Granger causality stems from a *maximum likelihood* (ML) perspective. For simplicity we consider only 1-step GC, although much of our exposition translates to *m*-step prediction via the relations (16) and (17). Given a finite set of observations of the joint process $[X^\mathsf{T}\, Y^\mathsf{T}]^\mathsf{T}$, the full and reduced 1-step predictions correspond to the respective linear regression models

$$X_k = \sum_{\ell=1}^{\infty} A_{xx,\ell} X_{k-\ell} + \sum_{\ell=1}^{\infty} A_{xy,\ell} Y_{k-\ell} + \varepsilon_{x,k} \tag{D.1a}$$

$$X_k = \sum_{\ell=1}^{\infty} A'_{xx,\ell} X_{k-\ell} \qquad\qquad + \varepsilon'_{x,k} \tag{D.1b}$$

We may then ask whether the full model (D.1a) furnishes a more likely model (in the ML sense) for the data than the (nested) reduced model (D.1b). Truncating the regressions to an appropriate finite model order p, which in an empirical setting may be estimated by standard model selection techniques such as the Akaike or Bayesian error criterion, cross-validation, etc. (McQuarrie and Tsai, 1998), the Neyman-Pearson lemma (Neyman and Pearson, 1928, 1933) tells us that the uniformly most powerful (UMP) test for the null hypothesis

$$H_0 : A_{xy,1} = A_{xy,2} = \ldots = A_{xy,p} = 0 \tag{D.2}$$

against the alternative hypothesis that at least one of the $A_{xy,k}$ is non-zero, is the *log-likelihood ratio* statistic, which by standard theory is just

$$\widehat{\mathcal{F}}_{Y\to X} \equiv \log \frac{\left|\widehat{\Sigma}'_{xx}\right|}{\left|\widehat{\Sigma}_{xx}\right|} \tag{D.3}$$

where $\widehat{\Sigma}_{xx}, \widehat{\Sigma}'_{xx}$ are ML estimators for the residuals covariance matrices of the respective models. Thus, in principle, time-domain Granger causality (18) [and instantaneous causality (20)] may be estimated in sample by replacing the respective residuals covariance matrices by ML estimates (but see below). For Gaussian processes we note that the standard ordinary least squares (OLS: Hamilton, 1994) or Levinson-Wiggins-Robinson (LWR: Levinson, 1947; Whittle, 1963; Wiggins and Robinson, 1965; Morf et al., 1978) estimators for a residuals covariance matrix are ML estimators, while in the non-Gaussian case they are asymptotically equivalent (Lütkepohl, 2005). Standard large-sample theory (Wilks, 1938; Wald, 1943) then yields an asymptotic sampling distribution for the Granger causality estimator (D.3): specifically, if the statistic is based on m observations and the number of degrees of freedom is $d = pn_x n_y$, where n_x, n_y are the dimensionalities of X, Y respectively, then $m\widehat{\mathcal{F}}_{Y\to X} \sim \chi^2(d, m\mathcal{F}_{Y\to X})$ (non-central χ^2 distribution if $\mathcal{F}_{Y\to X} > 0$ or central if $\mathcal{F}_{Y\to X} = 0$)[37]. Note that

since Granger causality statistics are all non-negative, their estimators will be positively biased - but the expected bias may easily be calculated (Zhou et al., 2014) as

$$\mathbf{E}\left[\widehat{\mathcal{F}}_{Y\to X} - \mathcal{F}_{Y\to X}\right] = \frac{d}{m} \tag{D.4}$$

regardless of the actual causality $\mathcal{F}_{Y\to X}$; (D.4) may be used to obtain asymptotically unbiased estimates of Granger causalities.

To estimate $\mathcal{F}_{Y\to X}$ for discrete time-series data, a naïve implementation where full and reduced models (D.1a), (D.1b) are estimated *separately* in sample will not suffice - this leads to inaccurate (even potentially negative) causality estimates. Rather—after selecting a suitable model order—the full model must be estimated, and the reduced model estimate calculated from the full model parameters. This step essentially involves spectral factorisation (14) of the reduced model CPSD, which may be effected computationally e.g., via Wilson's frequency-domain algorithm (Wilson, 1972; Dhamala et al., 2008a,b), Whittle's time-domain algorithm (Whittle, 1963; Barnett and Seth, 2014), or via *state-space* methods (see below). Alternatively, the full model CPSD may be estimated directly from the data by standard methods (so-called "nonparametric" estimation), and factored separately for the full and reduced models (Dhamala et al., 2008a,b). To estimate $f_{Y\to X}(\lambda)$, from (23) we see that spectral factorisation is unnecessary; only the full model parameters are required[38]. As regards statistical inference, in contrast to the time-domain case no sampling distribution for $\widehat{f}_{Y\to X}(\lambda)$ (asymptotic or exact) is known [see Geweke (1984) for a fuller discussion on this issue] and nonparametric subsampling or surrogate data techniques are best deployed for significance testing and derivation of confidence intervals.

Appendix D.1. State-space methods

Recently, efficient and practical (linear, discrete-time) state-space methods have been proposed for Granger-causal estimation (Barnett and Seth, 2015; Solo, 2016). Unlike VAR models, state-space models accommodate a moving-average component in the data parsimoniously. It is known that sub-model extraction, invertible filtering, additive noise and—crucially—subsampling all induce a moving-average component (Nsiri and Roy, 1993; Solo, 2007; Barnett and Seth, 2011). The significant advantage of state-space over VAR-based estimation is that the class of state-space models—equivalently VARMA models—is closed under all these operations. This, together with the availability of efficient (non-iterative) *state-space subspace* system identification algorithms (van Overschee and de Moor, 1996), makes state-space methods an attractive approach to Granger causality estimation.

Pertinently to this study, we remark that a critical difference between state-space and VAR estimation is that the (minimal)

---

[37]In the case of a univariate causal target (i.e. $n_x = 1$) an alternative asymptotic sampling distribution is available for the R^2-like statistic $\exp(\widehat{\mathcal{F}}_{Y\to X}) - 1$, scaled by sample size, as a central F-distribution under the null and a non-central F-distribution under the alternative hypothesis. According to Hamilton (1994), for small samples in particular, the F-distribution may be preferable (it has a fatter tail than the corresponding χ^2 distribution).

[38]This is not true for the important case of *conditional* Granger causality (Geweke, 1984), where spectral factorisation is still required (Barnett and Seth, 2014). Chen et al. (2006), while recognising the issue, propose an invalid computational method which attempts to avoid spectral factorisation - see Solo (2016) for further commentary.

state-space model order *is not increased by subsampling* (Solo, 2016). An issue with state-space Granger causality estimation, however, is that the distribution of the sample statistic remains unclear. While eq. (17) in Barnett and Seth (2015) suggests that the number of degrees of freedom for a Granger-causal sample statistic based on a maximum-likelihood state-space estimate of order p should again be $d = pn_x n_y$, this has not been established rigorously (or indeed empirically)[39].

VAR-based Granger causality estimation may also be enhanced by state-space methods: an (estimated) VAR model may easily be converted to a state-space model in "innovations form" (Hannan and Deistler, 2012). The method of Barnett and Seth (2015) and Solo (2016) may then be applied to yield Granger causality estimates. This procedure exploits an additional, *computational*, advantage of the state-space approach, that calculation of a reduced innovations-form state-space model from the full model (the spectral factorisation step) is achieved by solution of a single discrete algebraic Riccati equation (DARE), for which stable and efficient algorithms exist (Arnold and Laub, 1984; Lancaster and Rodman, 1995). Note that in this case, since it is a VAR model which is actually estimated, causality estimates follow a χ^2 sampling distribution with $d = pn_x n_y$ degrees of freedom, where p is the VAR model order. This technique was used for computation of all empirical VAR-based Granger causality estimates in this study.

Appendix D.2. Detecting Granger causality

Suppose that the Granger causality sample statistic $\widehat{\mathcal{F}}$ (for some given number of variables, model order and data sample size) has a cumulative distribution function $F_x(u)$, given that the *actual* Granger causality of the underlying stochastic process is x. We know that in the large-sample limit, $F_x(u)$ approaches asymptotically a non-central χ^2 (scaled by sample size) with non-centrality parameter x, or a central χ^2 if $x = 0$. The number of degrees of freedom is $d = n_1 n_2 p$ where n_1 (resp. n_2) is the number of target (resp. source) variables and p the VAR model order. To test for significance at level α of a sample Granger causality value of u, the null hypothesis of zero Granger causality is $x = 0$, with null distribution $F_0(u)$. We thus *accept* the null hypothesis—i.e. we take u as *non*significant—if $u \leq F_0^{-1}(1 - \alpha)$ [equivalently, the p-value $1 - F_0(u)$ is $\geq \alpha$]. However, if x (the true Granger causality) is in fact > 0 then this is a Type II error; i.e. a false negative. Thus, given an actual causality $x > 0$, the Type II error rate (i.e. probability of a Type II error) is

$$P_{II}(x; \alpha) \equiv \mathbf{P}\left(\widehat{\mathcal{F}} \leq F_0^{-1}(1 - \alpha)\right) = F_x\left(F_0^{-1}(1 - \alpha)\right) \quad \text{(D.5)}$$

Note that (for fixed sample size) as $x \to \infty$, $P_{II}(x; \alpha) \to 0$ and as $x \to 0$, $P_{II}(x; \alpha) \to 1$. We consider the Type II error rate $P_{II}(x; \alpha)$ as a measure of *detectability*[40] in finite sample of a Granger causality known to be equal to some $x > 0$: if $P_{II}(x; \alpha)$ is too large then we will infer false negatives—that is, we

fail to detect a significant causality—unacceptably frequently. We might, then, describe a G-causal value x as "undetectable at significance level α" if $P_{II}(x; \alpha) > \alpha'$, where α' represents an "acceptable" incidence of Type II errors (we could take $\alpha' = \alpha$).

Note that the Type I error rate at significance level α (that is, the probability of falsely rejecting the null hypothesis of zero causality when the true causality actually *is* zero) is, trivially, just α. Note too, that for joint significance testing of *multiple* Granger causality statistics, a multiple hypothesis, family-wise error rate or false discovery rate correction should be applied (Hochberg and Tamhane, 1987).

Appendix E. Inversion of the MA operator for a CTVAR process

Integrating (28) by parts we find

$$
\begin{aligned}
\zeta \Psi(\zeta) &= \zeta \int_{u=0}^{\infty} B(u)\, e^{-\zeta u}\, du \\
&= -\int_{u=0}^{\infty} B(u)\, d\left(e^{-\zeta u}\right) \\
&= \int_{u=0}^{\infty} \dot{B}(u)\, e^{-\zeta u}\, du - \left[B(u)\, e^{-\zeta u}\right]_{u=0}^{\infty} \\
&= \int_{u=0}^{\infty} \dot{B}(u)\, e^{-\zeta u}\, du + I \quad \text{by (32b) and } \Re\mathfrak{e}(\zeta) \geq 0 \\
&= \int_{u=0}^{\infty} \int_{s=0}^{u} A(s) B(u-s)\, e^{-\zeta u}\, ds\, du + I \quad \text{by (32a)} \\
&= \int_{s=0}^{\infty} \int_{u=s}^{\infty} A(s) B(u-s)\, e^{-\zeta u}\, du\, ds + I \\
&= \int_{s=0}^{\infty} \int_{v=0}^{\infty} A(s) B(v)\, e^{-\zeta(s+v)}\, dv\, ds + I \quad v = u - s \\
&= \int_{s=0}^{\infty} A(s)\, e^{-\zeta s}\, ds \int_{v=0}^{\infty} B(v)\, e^{-\zeta v}\, dv + I \\
&= [\zeta I - \Phi(\zeta)]\, \Psi(\zeta) + I
\end{aligned}
$$

so that $\Psi(\zeta) = \Phi(\zeta)^{-1}$ on $\Re\mathfrak{e}(\zeta) \geq 0$ as required.

Appendix F. ODE for the MA kernel of a CTVAR process

Writing (26) as $X(s) = \int_{u=-\infty}^{s} B(s-u)\, dW(u)$ we have, working to first order in ds:

$$
\begin{aligned}
dX(s) &\\
&= \int_{u=-\infty}^{s+ds} B(s+ds-u)\, dW(u) - \int_{u=-\infty}^{s} B(s-u)\, dW(u) \\
&= \int_{u=-\infty}^{s} B(s+ds-u)\, dW(u) - \int_{u=-\infty}^{s} B(s-u)\, dW(u) \\
&\quad + \int_{u=s}^{s+ds} B(s+ds-u)\, dW(u) \\
&= \int_{u=-\infty}^{s} \dot{B}(s-u)\, dW(u)\, ds + \int_{u=s}^{s+ds} B(s+ds-u)\, dW(u) \\
&= \int_{u=-\infty}^{s} \dot{B}(s-u)\, dW(u)\, ds + B(0)\, dW(s)
\end{aligned}
$$

---

[39] Solo (2016) states without proof that the degrees of freedom is $d = 2pn_x n_y$; again, we have been unable to verify this.

[40] We prefer to talk about detectability rather than statistical *power*; the power of the statistical test is, of course, just $1 - P_{II}(x; \alpha)$.

so setting $t \equiv s - u$

$$dX(s) = \int_{t=0}^{\infty} \dot{B}(t)\, dW(s-t)\, ds + B(0)\, dW(s) \qquad \text{(F.1)}$$

Now

$$
\begin{aligned}
&\int_{u=0}^{\infty} A(u) X(s-u)\, du \\
&= \int_{u=0}^{\infty} A(u) \int_{v=0}^{\infty} B(v)\, dW(s-u-v)\, du \\
&= \int_{u=0}^{\infty} A(u) \int_{t=u}^{\infty} B(t-u)\, dW(s-t)\, du \quad t \equiv u + v \\
&= \int_{t=0}^{\infty} \left[\int_{u=0}^{t} A(u) B(t-u)\, du \right] dW(s-t)
\end{aligned}
$$

So from (29)

$$
\begin{aligned}
dX(s) = \int_{t=0}^{\infty} \left[\int_{u=0}^{t} A(u) B(t-u)\, du \right] dW(s-t)\, ds \\
+ dW(s) \quad \text{(F.2)}
\end{aligned}
$$

Equating the right-hand sides of (F.1) and (F.2), which must hold for all s, we obtain eqs. (32).

Appendix G. Limiting approximations for a subsampled CTVAR process

Firstly we derive an asymptotic expansion for the scaled CPSD $\Delta S(\Delta\lambda; \Delta)$ of a subsampled CTVAR process X. We assume that the continuous-time autocovariance function Γ has an inverse Laplace transform[41]; i.e. $\Gamma(t)$ may be represented as a Laplace transform

$$\Gamma(t) = \int_0^{\infty} \Lambda(s) e^{-ts}\, ds \qquad \text{for } t \geq 0 \qquad \text{(G.1)}$$

The Wiener-Kintshine theorem (37) applied to (G.1) then yields

$$S(\lambda) = \int_0^{\infty} \left[\frac{\Lambda(s)}{s + 2\pi i \lambda} + \frac{\Lambda(s)^{\mathsf{T}}}{s - 2\pi i \lambda} \right] ds \qquad \text{(G.2)}$$

We now calculate the CPSD $S(\lambda; \Delta)$ of the subsampled process as follows (we don't assume that Δ is small):

$$
\begin{aligned}
&S(\lambda; \Delta) \\
&= \Delta \sum_{k=-\infty}^{\infty} \Gamma(k\Delta) e^{-2\pi i \Delta \lambda k} \qquad \text{from eq. (11)} \\
&= \Delta \sum_{k=0}^{\infty} \Gamma(k\Delta) e^{-2\pi i \Delta \lambda k} + [*] - \Gamma(0) \\
&= \Delta \sum_{k=0}^{\infty} \left[\int_0^{\infty} \Lambda(s) e^{-\Delta s k}\, ds \right] e^{-2\pi i \Delta \lambda k} + [*] - \Gamma(0) \quad \text{from (G.1)} \\
&= \Delta \int_0^{\infty} \Lambda(s) \left[\sum_{k=0}^{\infty} e^{-\Delta(s+2\pi i\lambda)k} \right] ds + [*] - \Gamma(0) \quad \text{rearranging} \\
&= \Delta \int_0^{\infty} \Lambda(s) \left[1 - e^{-\Delta(s+2\pi i\lambda)} \right]^{-1} ds + [*] - \Gamma(0)
\end{aligned}
$$

[41]This will be the case if $\Gamma(t)$ is bounded continuous or in $L^{\infty}(0, \infty)$.

where $[*]$ indicates the Hermitian transpose of the preceding term, so that

$$S(\lambda; \Delta) = \Delta \int_0^{\infty} \left[\frac{\Lambda(s)}{1 - e^{-\Delta(s+2\pi i\lambda)}} + \frac{\Lambda(s)^{\mathsf{T}}}{1 - e^{-\Delta(s-2\pi i\lambda)}} \right] ds - \Gamma(0) \tag{G.3}$$

Now let us define the coefficients C_n by

$$\frac{x}{1 - e^{-x}} = \sum_{n=0}^{\infty} C_n x^n \qquad \text{(G.4)}$$

so $C_0 = 1, C_1 = \frac{1}{2}, C_2 = \frac{1}{12}, C_3 = 0, C_4 = -\frac{1}{720}$, etc. We then have (again $[*]$ indicates the Hermitian transpose of the preceding term)

$$
\begin{aligned}
&S(\lambda; \Delta) \\
&= \Delta \int_0^{\infty} \sum_{n=0}^{\infty} C_n [\Delta(s + 2\pi i\lambda)]^{n-1} \Lambda(s)\, ds + [*] - \Delta\Gamma(0) \\
&\hspace{6cm} \text{from (G.3) and (G.4)} \\
&= \sum_{n=0}^{\infty} \Delta^n C_n \int_0^{\infty} (s + 2\pi i\lambda)^{n-1} \Lambda(s)\, ds + [*] - \Delta\Gamma(0)
\end{aligned}
$$

From (G.2) we see that the $n = 0$ terms are just $S(\lambda)$ and from (G.1) $\Gamma(0) = \int_0^{\infty} \Lambda(s)\, ds$, so that the $n = 1$ terms cancel with the trailing $-\Delta\Gamma(0)$. We thus have have

$$S(\lambda; \Delta) = S(\lambda) + \sum_{n=2}^{\infty} \Delta^n C_n \int_0^{\infty} (s + 2\pi i\lambda)^{n-1} \Lambda(s)\, ds + [*] \tag{G.5}$$

From (G.1) we have

$$\overset{(k)}{\Gamma}(0) = (-1)^k \int_0^{\infty} s^k \Lambda(s)\, ds \qquad \text{(G.6)}$$

From the CTVAR Yule-Walker equations (35), $\overset{(k)}{\Gamma}(0)$ is symmetric for even k and $\dot{\Gamma}(0) + \dot{\Gamma}(0)^{\mathsf{T}} = -\Sigma$. Expanding (G.5) in powers of Δ and using (G.6), we find to $O(\Delta^4)$:

$$S(\lambda; \Delta) = S(\lambda) + \tfrac{1}{12}\Delta^2 \Sigma + \tfrac{1}{720}\Delta^4 (\Omega + 12\pi^2\lambda^2\Sigma) + O(\Delta^5) \tag{G.7}$$

where we have set $\Omega \equiv \dddot{\Gamma}(0) + \dddot{\Gamma}(0)^{\mathsf{T}}$. We note that the lowest-order approximation $S(\lambda; \Delta) = S(\lambda) + O(\Delta)$ may be derived more simply by replacing the sum in the expression $S(\lambda; \Delta) = \sum_{k=-\infty}^{\infty} \Delta\Gamma(k\Delta) e^{-2\pi i\lambda k\Delta}$ [*cf.* eq. (11)] by an integral over $t = k\Delta$ in the limit $\Delta \to 0$ (Zhou et al., 2014).

Next, we note with Zhou et al. (2014) that, since the analytic extension $\Psi(z; \Delta)$ of $\frac{1}{\Delta} H\left(\frac{\omega}{2\pi\Delta}; \Delta\right)$ is holomorphic on the interior of the unit disc $|z| \leq 1$ (Section 2), by the Mean-Value Property for holomorphic functions (Gamelin, 1998), we have

$$\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} H(\lambda; \Delta)\, d\lambda = I \qquad \text{(G.8)}$$

from which we may conclude that $H(\lambda; \Delta) = O(1)$ in Δ. Now, since

$$
\begin{aligned}
H(\lambda; \Delta) \Sigma(\Delta) H(\lambda; \Delta)^* &= S(\lambda; \Delta) \\
&\to S(\lambda) = H(\lambda) \Sigma H(\lambda)^* \tag{G.9}
\end{aligned}
$$

as $\Delta \to 0$, it follows that $H(\lambda; \Delta) = H(\lambda) + O(\Delta)$ and $\Sigma(\Delta) = \Sigma + O(\Delta)$ as required.

Appendix H. Simulating CTVAR processes

We simulate CTVAR processes via a straightforward generalisation of Newton's method to stochastic integro-differential equations. To this end, we develop a VAR approximation for the subsampled process[42]. Firstly, we define the coefficients[43]

$$A_k[\Delta] = \int_{(k-1)\Delta}^{k\Delta} A(u)\,du \qquad \text{for } k = 1, 2, \dots \tag{H.1}$$

A useful result here, is that if $\varphi(t)$ is *Lipschitz continuous*[44] on $[0, \infty)$—in particular if $\varphi(t)$ is everywhere differentiable and has *bounded derivatives* on $[0, \infty)$—then it is straightforward to show that

$$\sum_{k=1}^{\infty} A_k[\Delta]\varphi(k\Delta) = \int_0^{\infty} A(u)\varphi(u)\,du + O(\Delta) \tag{H.2}$$

Now consider the autoregression

$$X(k\Delta) = X((k-1)\Delta) + \Delta \sum_{l=1}^{\infty} A_\ell[\Delta] \cdot X((k-\ell)\Delta) + \eta_k \tag{H.3}$$

Assuming that both $\Gamma(t)$ and $\dot{\Gamma}(t)$ satisfy the condition for (H.2) to hold, then from (35) we may verify after some algebra that

$$\text{cov}(\eta_\ell, \eta_{\ell-k}) = \delta_{k0}\Delta\Sigma + O(\Delta^3) \tag{H.4}$$

so that the residuals are "almost" white and in this sense (H.3) "almost" specifies a VAR[45].

To generate an approximate realisation of the subsampled process, then, we generate realisations $x_k \approx X(k\delta t) : k = 1, 2, \dots$ for very small time increments $\delta t \ll 1$, by the recursion

$$x_k = x_{k-1} + \delta t \sum_{l=1}^{L} A_\ell[\Delta] \cdot x_{k-\ell} + \eta_k \tag{H.5}$$

where L is large enough that $A_\ell[\Delta] \ll 1$ for $\ell > L$, with iid normal residuals $\eta_k \sim \mathcal{N}(0; \delta t\Sigma)$. An integration time increment $\delta t = 0.01$ ms was used for all simulations in this study.

---

[42] For our minimal CTVAR, since we can explicitly solve the spectral factorisation problem for the subsampled process with a rational transfer function (see Appendix J below), we could in principle generate realisations of the subsampled process *exactly* as a VARMA at any desired time resolution. However, the factorisation is unwieldy, so we prefer to use an approximation which, furthermore, is also applicable to problems for which a spectral factorisation cannot easily be obtained analytically.

[43] We use the notation $A_k[\Delta]$ to distinguish these quantities from the $A_k(\Delta)$ which, under our notational convention, denote the VAR coefficients of the subsampled process; they will not generally coincide exactly.

[44] This condition might be relaxed, depending on the form of the kernel $A(u)$.

[45] See Sims (1971); Geweke (1978) for detailed analysis of the *exact* VAR satisfied by the subsampled CTVAR. Roughly, the approximation (H.3) can be expected to be reasonable provided $X(t)$ does not fluctuate too wildly at the time scale of the sample interval Δ. In particular, our minimal CTVAR satisfies this condition for $\Delta \ll \min(1/a, 1/b)$. See also Bergstrom (1966); Sargan (1974) for different approaches to discrete approximation of SDEs.

Appendix I. Calculation of the cross-power spectral term for the subsampled minimal CTVAR process with finite causal delay

We have

$$\begin{aligned}
S_{xy}(z; \Delta) &= \sum_{k=1}^{\infty} \Gamma_{xy}(k\Delta)z^k + \sum_{k=0}^{\infty} \Gamma_{yx}(k\Delta)z^{-k} \\
&= \sum_{k=1}^{q-1} \Gamma_{xy}(k\Delta)z^k + \sum_{k=q}^{\infty} \Gamma_{xy}(k\Delta)z^k + \sum_{k=0}^{\infty} \Gamma_{yx}(k\Delta)z^{-k} \\
&= \sum_{k=q}^{\infty} \theta\left(\frac{1}{a+b}e^{a\tau}\alpha^k - \frac{1}{2b}e^{b\tau}\beta^k\right)z^k \\
&\quad + \sum_{k=1}^{q-1} \frac{1}{2b}\eta e^{-b\tau}\beta^{-k}z^k + \sum_{k=0}^{\infty} \frac{1}{2b}\eta e^{-b\tau}\beta^k z^{-k} \\
&= \sum_{k=q}^{\infty} \theta\left(\frac{1}{a+b}e^{a\tau}\alpha^k - \frac{1}{2b}e^{b\tau}\beta^k\right)z^k \\
&\quad + \frac{1}{2b}\eta e^{-b\tau}\left[\sum_{k=1}^{q-1}\beta^{-k}z^k + \sum_{k=0}^{\infty}\beta^k z^{-k}\right] \\
&= \frac{1}{c}\theta\eta e^{a\tau}\frac{(\alpha z)^q}{1-\alpha z} - \frac{1}{2b}\theta e^{b\tau}\frac{(\beta z)^q}{1-\beta z} - \frac{1}{2b}\eta e^{-b\tau}\frac{(\gamma z)^q}{1-\gamma z}
\end{aligned}$$

Appendix J. Spectral factorisation of the subsampled minimal CTVAR: zero residuals correlation

Here we perform the spectral factorisation of the subsampled CPSD (91)—as required for discrete-time GC calculations in the time (18) and frequency (23) domains—for the full and reduced regressions. Since spectral factorisation are *unique* (Wilson, 1972), we achieve this via an *ad hoc* approach suggested by the structure of the process and the form of the CPSDs. Throughout this section $z = e^{-i\omega}$, where $\omega = 2\pi\Delta\lambda$ is the angular frequency. For compactness, we generally drop the Δ in $S(z; \Delta)$, etc.

For the full regression, we attempt a rational VARMA factorisation of the CPSD (91) of the form

$$A(z)S(z)A^*(z) = B(z)\Sigma B^*(z) \tag{J.1}$$

so that $A(z)$ represents the VAR factor, $B(z)$ the VMA factor and $\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}$ the residuals covariance matrix. $A(z), B(z)$ are matrix polynomials in z with $A(0) = B(0) \equiv I$ and Σ is symmetric positive-definite. Since Y [and hence $Y(\Delta)$] is autonomous, there is clearly no causality in the $X(\Delta) \to Y(\Delta)$ direction. We also note that any *instantaneous* GC must arise from the moving-average component of the joint subsampled process (see below), so we attempt a factorisation of the form

$$\begin{bmatrix} A_{xx}(z) & 0 \\ 0 & A_{yy}(z) \end{bmatrix}\begin{bmatrix} S_{xx}(z) & S_{xy}(z) \\ S_{xy}(\bar{z}) & S_{yy}(z) \end{bmatrix}\begin{bmatrix} A_{xx}(\bar{z}) & 0 \\ 0 & A_{yy}(\bar{z}) \end{bmatrix} =$$
$$\begin{bmatrix} B_{xx}(z) & B_{xy}(z) \\ 0 & B_{yy}(z) \end{bmatrix}\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}\begin{bmatrix} B_{xx}(\bar{z}) & 0 \\ B_{xy}(\bar{z}) & B_{yy}(\bar{z}) \end{bmatrix} \tag{J.2}$$

The component equations are then

$$xx: \quad |A_{xx}|^2 S_{xx} = \sigma_{xx}|B_{xx}|^2 + 2\sigma_{xy}\Re\left\{B_{xx}B_{xy}^*\right\} + \sigma_{yy}|B_{xy}|^2$$
$$\text{(J.3a)}$$

$$xy: \quad A_{yy}^* A_{xx} S_{xy} = B_{yy}^*\left[\sigma_{xy}B_{xx} + \sigma_{yy}B_{xy}\right] \quad \text{(J.3b)}$$

$$yy: \quad |A_{yy}|^2 S_{yy} = \sigma_{yy}|B_{yy}|^2 \quad \text{(J.3c)}$$

Examination of the CPSD suggests we try

$$A_{xx}(z) = (1 - \alpha z)(1 - \beta z) \quad \text{(J.4a)}$$
$$A_{yy}(z) = 1 - \beta z \quad \text{(J.4b)}$$

The yy equation yields $B_{yy}(z) \equiv 1$, so that

$$\sigma_{yy} = w \quad \text{(J.5)}$$

Defining

$$|M(z)|^2 \equiv w\left(u_1|1 - \beta z|^2 + u_2|1 - \alpha z|^2\right) \quad \text{(J.6a)}$$
$$L(z) \equiv \beta\left[v_1(1 - \beta z)(1 - \gamma z) + v_2(1 - \gamma z)(1 - \alpha z)\right.$$
$$\left. + v_3(1 - \alpha z)(1 - \beta z)\right] \quad \text{(J.6b)}$$

(note that $|M(z)|^2$ may be factorised), the xy equation then yields

$$\sigma_{xy}B_{xx}(z) + \sigma_{yy}B_{xy}(z) = -L(z)z^{q-1} \quad \text{(J.7)}$$

on $|z| = 1$. But since $B_{xx}(z), B_{xy}(z)$ are polynomials (by assumption), (J.7) holds for all z in the complex plane, so that in particular setting $z = 0$ we obtain

$$\sigma_{xy} = \begin{cases} -\beta P & q = 1 \\ 0 & q > 1 \end{cases} \quad \text{(J.8)}$$

with

$$P \equiv v_1 + v_2 + v_3 \quad \text{(J.9)}$$

Next, the xx equation yields (after some algebra)

$$D|B_{xx}(z)|^2 = |M(z)|^2 - |L(z)|^2 \quad \text{(J.10)}$$

on $|z| = 1$, where

$$D \equiv |\Sigma| = \sigma_{xx}\sigma_{yy} - \sigma_{xy}^2 \quad \text{(J.11)}$$

Thus we need to solve the factorisation problem (J.10) for $D, B_{xx}(z)$ (see below) and we then have the residuals variance σ_{xx} for the full regression.

Note that $|M(z)|^2$ for $z = e^{-i\omega}$ is linear in $\cos\omega$ while $|L(z)|^2$ is linear in $\cos\omega, \cos 2\omega$. Therefore (J.10) may be factored for $B_{xx}(z)$ a 2nd order polynomial and it follows from (J.7) that $B_{xy}(z)$ is of order $q + 1$. In summary, we have

$$A_{xx}(z) \text{ is of order } 2 \quad \text{(J.12a)}$$
$$A_{yy}(z) \text{ is of order } 1 \quad \text{(J.12b)}$$
$$B_{xx}(z) \text{ is of order } 2 \quad \text{(J.12c)}$$
$$B_{xy}(z) \text{ is of order } q + 1 \quad \text{(J.12d)}$$
$$B_{yy}(z) \text{ is of order } 0 \quad \text{(J.12e)}$$

while the remaining coefficients vanish. We see that the joint subsampled process $(X(\Delta), Y(\Delta))$ is thus VARMA$(2, q + 1)$. Note that $Y(\Delta)$ is just VAR(1).

For the full regression, we write

$$L(z) = \beta(P - Qz + Rz^2) \quad \text{(J.13)}$$

where

$$Q = v_1(\beta + \gamma) + v_2(\gamma + \alpha) + v_3(\alpha + \beta) \quad \text{(J.14a)}$$
$$R = v_1\beta\gamma + v_2\gamma\alpha + v_3\alpha\beta \quad \text{(J.14b)}$$

We have

$$|M(z)|^2 = 2w(\varphi - \psi\cos\omega) \quad \text{(J.15)}$$

where

$$\varphi \equiv \tfrac{1}{2}\left[u_2(1 + \alpha^2) + u_1(1 + \beta^2)\right] \quad \text{(J.16a)}$$
$$\psi \equiv u_2\alpha + u_1\beta \quad \text{(J.16b)}$$

and setting

$$B_{xx}(z) = 1 - Uz + Vz^2 \quad \text{(J.17)}$$

the factorisation problem (J.10) becomes

$$D\left[1 + U^2 + V^2 - 2U(1 + V)\cos\omega + 2V\cos 2\omega\right] =$$
$$2w(\varphi - \psi\cos\omega) - \beta^2\left[P^2 + Q^2 + R^2\right.$$
$$\left. -2Q(P + R)\cos\omega + 2PR\cos 2\omega\right] \quad \text{(J.18)}$$

Since this must hold for all ω, we have

$$D(1 + U^2 + V^2) = 2\mathfrak{a} \quad \text{(J.19a)}$$
$$DU(1 + V) = \mathfrak{b} \quad \text{(J.19b)}$$
$$DV = \mathfrak{c} \quad \text{(J.19c)}$$

where we have set

$$\mathfrak{a} \equiv w\varphi - \tfrac{1}{2}\beta^2(P^2 + Q^2 + R^2) \quad \text{(J.20a)}$$
$$\mathfrak{b} \equiv w\psi - \beta^2 Q(P + R) \quad \text{(J.20b)}$$
$$\mathfrak{c} \equiv -\beta^2 PR \quad \text{(J.20c)}$$

so that

$$U = \frac{\mathfrak{b}}{D + \mathfrak{c}} \quad \text{(J.21a)}$$
$$V = \frac{\mathfrak{c}}{D} \quad \text{(J.21b)}$$
$$\text{(J.21c)}$$

with D the largest root of the 4th degree polynomial equation

$$(D + \mathfrak{c})^2\left[(D - \mathfrak{a})^2 - \left(\mathfrak{a}^2 - \mathfrak{c}^2\right)\right] + \mathfrak{b}^2 D^2 = 0 \quad \text{(J.22)}$$

Setting

$$D = \mathfrak{d} + \sqrt{\mathfrak{d}^2 - \mathfrak{c}^2} \quad \text{(J.23)}$$

we find that (J.22) is satisfied if $4(\mathfrak{d} - \mathfrak{a})(\mathfrak{d} + \mathfrak{c}) + \mathfrak{b}^2 = 0$, so that (J.23) is the required solution with

$$\mathfrak{d} \equiv \tfrac{1}{2}\left(\mathfrak{a} - \mathfrak{c} + \sqrt{(\mathfrak{a} + \mathfrak{c})^2 - \mathfrak{b}^2}\right) \quad \text{(J.24)}$$

Note that if Δ divides τ exactly (for $\Delta \leq \tau$), then $P \equiv 0$, so $\mathfrak{c} \equiv 0$ and largest root is

$$D = \mathfrak{a} + \sqrt{\mathfrak{a}^2 - \mathfrak{b}^2} \qquad (J.25)$$

For the reduced regression we need to factorise $S_{xx}(z)$, the power spectrum of $X(\Delta)$ alone (91a). The VARMA factorisation takes the form

$$|a(z)|^2 S_{xx}(z) = \sigma'_{xx}|h(z)|^2 \qquad (J.26)$$

where σ'_{xx} is the residuals variance of the reduced regression, which is easily seen to be solvable by $a(z) = (1 - \alpha z)(1 - \beta z)$, $h(z) = 1 - pz$, so that $X(\Delta)$ is VARMA(2, 1). We find:

$$\sigma'_{xx}(1 + p^2) = 2\varphi \qquad (J.27a)$$
$$\sigma'_{xx}p = \psi \qquad (J.27b)$$

so that, eliminating p, the solution for σ'_{xx} is (again we need the largest root)

$$\sigma'_{xx} = \varphi + \sqrt{\varphi^2 - \psi^2} \qquad (J.28)$$

Finally, to calculate $f_{Y \to X}(z)$, (23) yields

$$f_{Y \to X}(z) = -\log\left[1 - S_{xx}(z)^{-1}\Sigma_{yy|x}|H_{xy}(z)|^2\right] \qquad (J.29)$$

We thus require $S_{xx}(z)$, $\Sigma_{yy|x}$ and $H_{xy}(z)$. Again, we already have $S_{xx}(z)$ from (91), while $\Sigma_{yy|x} = \sigma_{yy} - \sigma_{xx}^{-1}\sigma_{xy}^2 = \sigma_{xx}^{-1}D$. From (J.2), we have $H_{xy}(z) = A_{xx}(z)^{-1}B_{xy}(z)$. We already know $A_{xx}(z)$ from (J.4a) and from (J.7) we have $B_{xy}(z) = -\sigma_{yy}^{-1}\left[\sigma_{xy}B_{xx}(z) + L(z)z^{q-1}\right]$, with $B_{xx}(z)$ given by (J.17) and $L(z)$ by (J.13). The expression for $f_{Y \to X}(z)$ may be simplified somewhat by noting that $S_{xx}(z)|A_{xx}(z)|^2 = \sigma_{yy}^{-1}|M(z)|^2$, where, from (J.10), $|M(z)|^2 = D|B_{xx}(z)|^2 + |L(z)|^2$.

From (J.2) and $B_{yy}(z) \equiv 1$, we see that for the Δ-subsampled joint process to be minimum-phase it is necessary and sufficient that all roots of the equation $B_{xx}(z) = 1 - Uz + Vz^2 = 0$ (J.17) lie strictly outside the unit disc $|z| \leq 1$ in the complex plane. It is easily checked that the process $X(\Delta)$ is always minimum phase: from (J.27) we have $p = \left(\varphi - \sqrt{\varphi^2 - \psi^2}\right)/\psi$, so that $|p| < 1$ is always satisfied and the (single) root of $h(z) = 0$ thus always lies outside the unit disc. The process $Y(\Delta)$ is VAR(1) and thus trivially minimum-phase.

References

Arnold, III, W. F., Laub, A., 1984. Generalised eigenproblem algorithms and software for algebraic Riccati equations. Proc. IEEE 72 (12), 1746–1754.

Barnett, L., Barrett, A. B., Seth, A. K., 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett. 103 (23), 0238701.

Barnett, L., Bossomaier, T., 2013. Transfer entropy as a log-likelihood ratio. Phys. Rev. Lett. 109 (13), 0138105.

Barnett, L., Seth, A. K., 2011. Behaviour of Granger causality under filtering: Theoretical invariance and practical application. J. Neurosci. Methods 201 (2), 404–419.

Barnett, L., Seth, A. K., 2014. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. J. Neurosci. Methods 223, 50–68.

Barnett, L., Seth, A. K., 2015. Granger causality for state-space models. Phys. Rev. E (Rapid Communications) 91 (4), 040101(R).

Barrett, A. B., Barnett, L., 2013. Granger causality is designed to measure effect, not mechanism. Front. Neuroinform. 7, 6.

Barrett, A. B., Barnett, L., Seth, A. K., 2010. Multivariate Granger causality and generalized variance. Phys. Rev. E 81 (4), 041907.

Bergstrom, A. R., 1966. Nonrecursive models as discrete approximations to systems of stochastic differential equations. Econometrica 34 (1), 173–182.

Björklund, S., 2003. A survey and comparison of time-delay estimation methods in linear systems. Thesis no. 1061, Division of Automatic Control, Department of Electrical Engineering, Linköpings Universitet.

Breitung, J., Swanson, N. R., 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. J. Time Ser. Anal. 23 (6), 651–665.

Budd, J. M. L., Kisvárday, Z. F., 2012. Communication and wiring in the cortical connectome. Front. Neuroanat. 6 (42).

Caines, P. E., 1976. Weak and strong feedback free processes. IEEE. Trans. Autom. Contr. 21 (5), 737–739.

Caines, P. E., Chan, C. W., 1975. Feedback between stationary stochastic processes. IEEE. Trans. Autom. Contr. 20 (4), 498–508.

Caminiti, R., Carducci, F., Piervincenzi, C., Battaglia-Mayer, A., Confalone, G., Visco-Comandini, F., Pantano, P., Innocenti, G. M., 2013. Diameter, length, speed, and conduction delay of callosal axons in Macaque monkeys and humans: Comparing data from histology and magnetic resonance imaging diffusion tractography. J. Neurosci. 33 (36), 14501–14511.

Chen, Y., Bressler, S. L., Ding, M., 2006. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. J. Neuro. Methods 150, 228–237.

Chicharro, D., Panzeri, S., 2014. Algorithms of causal inference for the analysis of effective connectivity among brain regions. Front. Neuroinform. 8, 64.

Comte, F., Renault, E., 1996. Noncausality in continuous time models. Econ. Theory 12 (2), 215–256.

Dhamala, M., Rangarajan, G., Ding, M., 2008a. Analyzing information flow in brain networks with nonparametric Granger causality. Neuroimage 41, 354–362.

Dhamala, M., Rangarajan, G., Ding, M., 2008b. Estimating Granger causality from Fourier and wavelet transforms of time series data. Phys. Rev. Lett. 100, 018701.

Ding, M., Chen, Y., Bressler, S. L., 2006. Granger causality: Basic theory and application to neuroscience. In: Schelter, B., Winterhalder, M., Timmer, J. (Eds.), Handbook of Time Series Analysis. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 437–460.

Doob, J., 1953. Stochastic Processes. John Wiley, New York.

Dufour, J.-M., Renault, E., 09 1998. Short run and long run causality in time series: Theory. Econometrica 66 (5), 1099–1125.

Florens, J.-P., Fougère, D., 09 1996. Noncausality in continuous time. Econometrica 64 (5), 1195.

Friedlander, F. G., Joshi, M. S., 1998. Introduction to the Theory of Distributions, 2nd Edition. Cambridge University Press, Cambridge, UK.

Friston, K. J., Bastos, A. M., Oswal, A., van Wijk, B., Richter, C., Litvak, V., 2014. Granger causality revisited. NeuroImage 101, 796 – 808.

Gamelin, T. W., 1998. Complex Analysis. Springer Science+Business Media, Inc., New York, NY, USA.

Garnier, H., Wang, L. (Eds.), 2008. Identification of Continuous-time Models from Sampled Data. Advances in Industrial Control. Springer-Verlag, London.

Geweke, J., 1978. Temporal aggregation in the multiple regression model. Econometrica 46 (3), 643–661.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc. 77 (378), 304–313.

Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. J. Am. Stat. Assoc. 79 (388), 907–915.

Granger, C. W. J., 1963. Economic processes involving feedback. Inform. Control 6 (1), 28–48.

Granger, C. W. J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Granger, C. W. J., 1981. Some properties of time series data and their use in econometric model specification. J. Econometrics 16 (1), 121–130.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Hannan, E. J., 1970. Multiple Time Series. John Wiley, New York.

Hannan, E. J., Deistler, M., 2012. The Statistical Theory of Linear Systems. SIAM, Philadelphia, PA, USA.

Hochberg, Y., Tamhane, A. C., 1987. Multiple Comparison Procedures. John Wiley, New York.

Khintchine, A., 1934. Korrelationstheorie der stationären stochastischen prozesse. Math. Ann. 109 (1), 604–615.

Kučera, V., 1991. Factorization of rational spectral matrices: a survey of methods. In: Control 1991. International Conference on Control '91. Vol. 2. pp. 1074–1078.

Lancaster, P., Rodman, L., 1995. Algebraic Riccati Equations. Oxford University Press, Oxford, UK.

Larsson, K. E., Mossberg, M., Söderström, T., 2006. An overview of important practical aspects of continuous-time ARMA system identification. Circuits, Systems and Signal Processing 25 (1), 17–46.

Levinson, N., 1947. The Wiener RMS (root-mean-square) error criterion in filter design and prediction. J. Math. Phys. 25, 261–278.

Lighthill, M. J., 1958. An Introduction to Fourier Analysis and Generalised Functions. Cambridge University Press, Cambridge, UK.

Lizier, J. T., Prokopenko, M., 2010. Differentiating information transfer and causal effect. M. Eur. Phys. J. B 73 (4), 605–615.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.

Longtin, A., 2010. Stochastic delay-differential equations. In: Atay, F. M. (Ed.), Complex Time-Delay Systems: Theory and Applications. Springer-Verlag, Berlin - Heidelberg, Ch. 6, pp. 177–195.

Lütkepohl, H., 1993. Testing for causation between two variables in higher dimensional VAR models. In: Schneeweiß, H., Zimmerman, K. (Eds.), Studies in Applied Econometrics. Physica-Verlag HD, Heidelberg, pp. 75–91.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin.

Marcellino, M., 1999. Some consequences of temporal aggregation in empirical analysis. J. Bus. Econ. Stat. 17 (1), 129–136.

Masani, P., 1966. Recent trends in multivariate prediction theory. In: Krishnaiah, P. R. (Ed.), Multivariate Analysis. Academic Press, New York, pp. 351–382.

McCrorie, J. R., Chambers, M. J., 2006. Granger causality and the sampling of economic processes. J. Econometrics 132, 311–336.

McKetterick, T. J., Giuggioli, L., 2014. Exact dynamics of stochastic linear delayed systems: Application to spatiotemporal coordination of comoving agents. Phys. Rev. E 90 (4), 042135.

McQuarrie, A. D. R., Tsai, C.-L., 1998. Regression and Time Series Model Selection. World Scientific Publishing, Singapore.

Miller, R., 1994. What is the contribution of axonal conduction delay to temporal structure in brain dynamics? In: Pantev, C., Elbert, T., Lütkenhöner, B. (Eds.), Oscillatory Event-Related Brain Dynamics. Springer Science+Business Media, New York, pp. 53–57.

Morf, M., Viera, A., Lee, D. T. L., Kailath, T., 1978. Recursive multichannel maximum entropy spectral estimation. IEEE Trans. Geosci. Elec. 16 (2), 85 –94.

Neyman, J., Pearson, E. S., 1928. On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A, 175–240.

Neyman, J., Pearson, E. S., 1933. On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. R. Soc. A 231, 289–337.

Nsiri, S., Roy, R., 1993. On the invertibility of multivariate linear processes. J. Time Ser. Anal. 14 (3), 305–316.

Øksendal, B., 2003. Stochastic Differential Equations: An Introduction with Applications. Springer-Verlag, Berlin.

Paluš, M., Komárek, V., Hrnčíř, Z., Štěrbová, K., 2001. Synchronization as adjustment of information rates: Detection from bivariate time series. Phys. Rev. E 63 (4), 046211.

Pearl, J., 2009. Causality: Models, Reasoning and Inference, 2nd Edition. Cambridge University Press, Cambridge, UK.

Prusseit, J., Lehnertz, K., 2008. Measuring interdependences in dissipative dynamical systems with estimated Fokker-Planck coefficients. Phys. Rev. E 77 (4), 041914.

Åström, K. J., 1969. On the choice of sampling rates in parametric identification of time series. Inform. Sciences 1 (3), 273–278.

Åström, K. J., Hagander, P., Sternby, J., 1984. Zeros of sampled systems. Automatica 28 (1), 31–38.

Renault, E., Sekkat, K., Szafarz, A., 1998. Testing for spurious causality in exchange rates. J. Empiri. Financ. 5 (1), 47–66.

Renault, E., Szafarz, A., 1991. True versus spurious instantaneous causality.

Working papers, Universite Libre de Bruxelles - C.E.M.E.

Rozanov, Y. A., 1967. Stationary Random Processes. Holden-Day, San Francisco.

Sargan, J. D., 1974. Some discrete approximations to continuous time stochastic models. J. Roy. Stat. Soc. B Met. 36 (1), 74–90.

Schreiber, T., 2000. Measuring information transfer. Phys. Rev. Lett. 85 (2), 461–4.

Seth, A. K., Barrett, A. B., Barnett, L., 2015. Granger causality analysis in neuroscience and neuroimaging. J. Neurosci. 35 (8), 3293–3297.

Seth, A. K., Chorley, P., Barnett, L., 2013. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. Neuroimage 65, 540–555.

Sims, C. A., 1971. Discrete approximations to continuous time distributed lags in econometrics. Econometrica 39 (3), 545–563.

Sims, C. A., 1972. Money, income and causality. Am. Econ. Rev. 62 (4), 540–552.

Solo, V., Dec. 2007. On causality I: Sampling and noise. In: Proceedings of the 46th IEEE Conference on Decision and Control. IEEE, New Orleans, LA, USA, pp. 3634–3639.

Solo, V., May 2016. State-space analysis of Granger-Geweke causality measures with application to fMRI. Neural Comput. 28 (5), 914–949.

Uhlenbeck, G. E., Ornstein, L. S., 1930. On the theory of Brownian motion. Phys. Rev. 36, 823–841.

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: Influence, causality and biophysical modeling. Neuroimage 58 (2), 339–361.

van Overschee, P., de Moor, B. L. R., 1996. Subspace Identification for Linear Systems: Theory, Implementation, Applications. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wahl, B., Feudel, U., Hlinka, J., Wächter, M., Peinke, J., Freund, J. A., 2016. Granger-causality maps of diffusion processes. Phys. Rev. E 93 (2), 022213.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. T. Am. Math. Soc. 54 (3), 426–482.

Wei, W. W. S., 1981. Effect of systematic sampling on ARIMA models. Communications in Statistics - Theory and Methods 10 (23), 1389–2398.

Whittle, P., 1963. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 50 (1,2), 129–134.

Wibral, M., Pampu, N., Priesemann, V., Siebenhühner, F., Seiwert, H., Lindner, M., Lizier, J. T., Vicente, R., 2013. Measuring information-transfer delays. PloS One 8 (2).

Wiener, N., 1930. Generalized harmonic analysis. Acta Math. 55 (1), 117–258.

Wiener, N., 1956. The theory of prediction. In: Beckenbach, E. F. (Ed.), Modern Mathematics for Engineers. McGraw Hill, New York, pp. 165–190.

Wiggins, R. A., Robinson, E. A., 1965. Recursive solution of the multichannel filtering problem. J. Geophys. Res. 70, 1885–1891.

Wilks, S. S., 1932. Certain generalizations in the analysis of variance. Biometrika 24, 471–494.

Wilks, S. S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 6 (1), 60–62.

Wilson, G. T., 1972. The factorization of matricial spectral densities. SIAM J. Appl. Math. 23 (4), 420–426.

Wold, H., 1938. A Study in the Analysis of Stationary Time-Series. Almqvist & Wiksell, Uppsala.

Zhou, D., Zhang, Y., Xiao, Y., Cai, D., 2014. Analysis of sampling artifacts on the Granger causality analysis for topology extraction of neuronal dynamics. Front. Comput. Neurosci. 8 (75).