Data based intervention approach for Complexity-Causality measure

Aditi Kathpalia [Corresp., 1] , Nithin Nagaraj [1]

[1] National Institute of Advanced Studies, Bengaluru, Karnataka, India

Corresponding Author: Aditi Kathpalia
Email address: kathpaliaaditi@gmail.com

Causality testing methods are being widely used in various disciplines of science. Model-free methods for causality estimation are very useful as the underlying model generating the data is often unknown. However, existing model-free measures assume separability of cause and effect at the level of individual samples of measurements and unlike model-based methods do not perform any intervention to learn causal relationships. These measures can thus only capture causality which is by the associational occurrence of 'cause' and 'effect' between well separated samples. In real-world processes, often 'cause' and 'effect' are inherently inseparable or become inseparable in the acquired measurements. We propose a novel measure that uses an adaptive interventional scheme to capture causality which is not merely associational. The scheme is based on characterizing complexities associated with the dynamical evolution of processes on short windows of measurements. The formulated measure, Compression- Complexity Causality is rigorously tested on simulated and real datasets and its performance is compared with that of existing measures such as Granger Causality and Transfer Entropy. The proposed measure is robust to presence of noise, long-term memory, filtering and decimation, low temporal resolution (including aliasing), non-uniform sampling, finite length signals and presence of common driving variables. Our measure outperforms existing state-of-the-art measures, establishing itself as an effective tool for causality testing in real world applications.

Data based Intervention Approach for Complexity-Causality Measure

Aditi Kathpalia[1] **and Nithin Nagaraj**[1]

[1]**National Institute of Advanced Studies, Bengaluru, India**

Corresponding author:
Aditi Kathpalia[1]

Email address: kathpaliaaditi@gmail.com

ABSTRACT

Causality testing methods are being widely used in various disciplines of science. Model-free methods for causality estimation are very useful as the underlying model generating the data is often unknown. However, existing model-free measures assume separability of cause and effect at the level of individual samples of measurements and unlike model-based methods do not perform any intervention to learn causal relationships. These measures can thus only capture causality which is by the associational occurrence of 'cause' and 'effect' between well separated samples. In real-world processes, often 'cause' and 'effect' are inherently inseparable or become inseparable in the acquired measurements. We propose a novel measure that uses an adaptive interventional scheme to capture causality which is not merely associational. The scheme is based on characterizing complexities associated with the dynamical evolution of processes on short windows of measurements. The formulated measure, Compression-Complexity Causality is rigorously tested on simulated and real datasets and its performance is compared with that of existing measures such as Granger Causality and Transfer Entropy. The proposed measure is robust to presence of noise, long-term memory, filtering and decimation, low temporal resolution (including aliasing), non-uniform sampling, finite length signals and presence of common driving variables. Our measure outperforms existing state-of-the-art measures, establishing itself as an effective tool for causality testing in real world applications.

1 INTRODUCTION

The 'Ladder of Causation' very rightly arranges hierarchically the abilities of a causal learner (Pearl and Mackenzie, 2018). The three levels proposed are 1.Association, 2. Intervention and 3. Counterfactuals, when arranged from the lower rung to the higher rung. Currently, causality learning and inferring algorithms using only data are still stuck at the lowermost rung of 'Association'.

Measures such as Granger Causality (GC) (Granger, 1969) and its various modifications (Dhamala et al., 2008; Marinazzo et al., 2008), as well as, Transfer Entropy (TE) (Schreiber, 2000) that are widely being used across various disciplines of science — neuroscience (Seth et al., 2015; Vicente et al., 2011), climatology (Stips et al., 2016; Mosedale et al., 2006), econometrics (Hiemstra and Jones, 1994; Chiou-Wei et al., 2008), engineering (Bauer et al., 2007) etc., are claimed to be 'model-free' measures of causality. This is because they have a wider scope compared to specific model assumptions made by methods such as Dynamical Causal Modeling (Friston et al., 2003) and Structural Equation Modeling (Pearl, 2009). However, the assumptions made by these methods are often ignored in practice resulting in erroneous causality estimates on real world datasets. These measures can accurately quantify the degree of coupling between the given time series only if assumptions (such as linearity, stationarity and presence of Gaussian noise in case of GC and stationarity, markovian in case of TE) are satisfied. Thus, these methods, when correctly applied, can infer the presence of causality when it is by 'association' alone and not due to higher levels on the Ladder of Causation. To explain this better, consider a case where the 'cause' and 'effect' are inseparable. This can happen even when the time series satisfies stationarity but is non-markovian or in several instances when it is non-stationary. In fact, the stated assumptions are quite unlikely to be met in practice considering that acquired data are typically samples of continuous/discrete evolution of real world processes. These processes might be evolving at spatio-temporal scales very different from the

47 scales of measurements. As a result, cause and effect may co-exist in a single measurement or overlap
48 over blocks of measurements thus making them inseparable. In such a scenario, it would be incorrect
49 to estimate causality by means of correlations and/or joint probabilities which implicitly assumes the
50 separability of 'cause' and 'effect'. Both GC and TE make this assumption of separability. Circularly,
51 to characterize a time series sample as purely a 'cause' or an 'effect' is possible only if there is a clear
52 linear/markovian separable relationship. When cause and effect are inseparable, 'associational' measures
53 of causality such as GC and TE are insufficient and we need a method to climb up the ladder of causation.

54 Intervention based approaches to causality rank higher than association. It involves not just observing
55 regularities in the data but actively changing what is there and observing its effect. In other words, we
56 are asking the question — what will happen if we 'do' something? Given only data and not the power to
57 intervene on the experimental set up, intervention can only be done by building strong, accurate models.
58 Model-based causality testing measures, alluded to before, will fall in this category. They invert the model
59 to obtain its various parameters, and then intervene to make predictions about situations for which data is
60 unavailable. However, these methods are very domain specific and the models require specific knowledge
61 about the data. With insufficient knowledge about the underlying model which generated the data, such
62 methods are inapplicable.

63 Given only data that has already been acquired without any knowledge of its generating model or
64 the power to intervene on the experimental/real-world setting, we can ask the question — what kind of
65 intervention is possible (if at all) to infer causality? The proposed 'interventional causality' approach will
66 not merely measure 'associational causality' because it does not make the assumption that the cause and
67 its effect are present sample by sample (separable) as is done by existing model-free, data based methods
68 of causality estimation.

69 Even in cases where cause and its effect are inseparable, which is probably true for most real-
70 world processes, the change in the dynamics of the processes would contain information about causal
71 influences between them. With this understanding, we propose the novel idea of data-based, model-
72 free **Interventional Complexity Causality (ICC)**. In this paper, we formalize the notion of ICC using
73 Compression-Complexity to define **Compression-Complexity Causality (CCC)**. CCC shows some
74 interesting properties. We test CCC on simulated and real datasets and compare its performance with
75 existing model-free causality methods. Our results demonstrate that CCC overcomes the limitations of
76 'associational' measures (GC and TE).

77 This paper is organized as follows. The idea of Dynamical Complexity and its specific realization
78 Dynamical Compression-Complexity are discussed in Section 2. Interventional Complexity Causality
79 and its specific case Compression-Complexity Causality (CCC) are discussed in Section 3. How positive
80 and negative CCC is a possibility and what is its implication on the kind of causal influence is detailed
81 in Section 4. Results and Discussion on the performance of CCC and its comparison with the existing
82 measures, GC and TE, are included in Section 5, followed by Conclusions and Future Work in Section 6.

83 2 DYNAMICAL COMPLEXITY (DC) AND DYNAMICAL COMPRESSION-COMPLEXITY
84 (CC)

There can be scenarios where cause and effect co-exist in a single temporal measurement or blocks
of measurements. For example, this can happen (a) inherently in the dynamics of the generated pro-
cess, (b) when cause and effect occur at different spatio-temporal scales, (c) when measurements are
acquired at a scale different from the spatio-temporal scale of the cause-effect dynamics (continuous
or discrete). In such a case, probabilities of joint occurrence is too simplistic an assumption to capture
causal influences. On the other hand, the very existence of causality here is actually resulting in a change
of joint probabilities/correlations which cannot be captured by an assumption of static probabilities.
To overcome this problem, we capture causality using the idea of *dynamical complexity*. Inseparable
causal influences within a time series (or between two time series) would be reflected in their dynamical
evolution. Dynamical Complexity (*DC*) of a single time series X is defined as below -

$$DC(\Delta X|X_{past}) = C(X_{past} + \Delta X) - C(X_{past}),
\qquad (1)$$

85 where ΔX is a moving window of length w samples and X_{past} is a window consisting of immediate
86 past L samples of ΔX. '+' refers to appending, for e.g., for time series $A = [1,2,3]$ and $B = [p,q]$, then
87 $A + B = [1,2,3,p,q]$. $C(X)$ refers to complexity of time series X. DC, thus varies with the temporal index
88 of ΔX and can be averaged over the entire time series to estimate its average DC.

It is important to note that dynamical complexity is very different from *complexity rate* (*CR*), which can be estimated as follows -

$$CR(\Delta X|X_{past}) = C(X_{past}, \Delta X) - C(X_{past}),\qquad(2)$$

where $C(X_{past}, \Delta X)$ is the *joint* complexity of X_{past} and ΔX. Complexity rate can be seen as a generalization of Shannon entropy rate (Cover and Thomas, 2012), the difference being that the former can be computed using any notion of complexity, not just entropy. As is evident from the equation, *CR* is estimated based on the joint occurrences of ΔX and X_{past}, while *DC* captures temporal change in complexities on the evolution of the process. In case of inseparability of cause and effect, it would be inappropriate to use *CR* to infer causal relationships.

Now for this notion of "complexity", that has been referred to in this section several times, there is no single unique definition. As noted in Nagaraj and Balasubramanian (2017b), Shannon entropy (Cover and Thomas, 2012) is a very popular and intuitive measure of complexity. A low value of Shannon entropy indicates high redundancy and structure (low complexity) in the data and a high value indicates low redundancy and high randomness (high complexity). For ergodic sources, owing to Shannon's noiseless source coding theorem (Cover and Thomas, 2012), (lossless) compressibility of the data is directly related to Shannon entropy. However, robustly estimating compressibility using Shannon entropy for short and noisy time series is a challenge (Nagaraj and Balasubramanian, 2017a). Recently, the notion of compression-complexity has been introduced (Nagaraj and Balasubramanian, 2017a) to circumvent this problem. Compression-complexity defines the complexity of a time series by using optimal lossless data compression algorithms. It is well acknowledged that data compression algorithms are not only useful for compression of data for efficient transmission and storage, but also act as models for learning and statistical inference (Cilibrasi et al., 2007). Lempel-Ziv (LZ) Complexity (Lempel and Ziv, 1976) and Effort-To-Compress (ETC) (Nagaraj et al., 2013) are two measures which fall in this category.

As per the minimum description length principle (Rissanen, 1978) that formalizes the Occam's razor, the best hypothesis (model and its parameters) for a given set of data is the one that leads to its best compression. Extending this principle for causality, an estimation based on dynamical complexity (compressibility) of time series would be the best possible means to capture causally influenced dynamics.

Out of the complexity measures discussed before, ETC seemed to be most suitable for estimation of dynamical complexity. ETC is defined as the effort to compress the input sequence using the lossless compression algorithm known as Non-sequential Recursive Pair Substitution (NSRPS). It has been demonstrated that both LZ and ETC outperform Shannon entropy in accurately characterizing the dynamical complexity of both stochastic (Markov) and deterministic chaotic systems in the presence of noise (Nagaraj and Balasubramanian, 2017a,b). Further, ETC has shown to reliably capture complexity of very short time series where even LZ fails (Nagaraj and Balasubramanian, 2017a), and for analyzing short RR tachograms from healthy young and old subjects (Balasubramanian and Nagaraj, 2016).

In order to faithfully capture the process dynamics, *DC* is required to be estimated on overlapping short-length windows of time series data. Infotheoretic quantities (like shannon entropy) which are based on computation of probability densities are not the ideal choice here (owing to finite-length effects). Compression-Complexity measures are more appropriate choices. Because of the advantages of ETC over LZ mentioned above, we use ETC to formulate our measure of causality discussed in the next section.

3 INTERVENTIONAL COMPLEXITY CAUSALITY (ICC) AND COMPRESSION-COMPLEXITY CAUSALITY (CCC)

To measure how the dynamics of a process Y influence the dynamics of a process X, we intervene to create new hypothetical blocks of time series data, $Y_{past} + \Delta X$, where Y_{past} is a window of length L samples, taken from the immediate past of the window ΔX. These blocks are created by 'surgery' and do not exist in reality in the data collected. Interventional Complexity Causality (ICC) is defined as the change in the dynamical complexity of time series X when ΔX is seen to be generated jointly by the dynamical evolution of both Y_{past} and X_{past} as opposed to by the reality of the dynamical evolution of X_{past} alone. Mathematically,

$$ICC_{Y_{past} \to \Delta X} = DC(\Delta X|X_{past}) - DC(\Delta X|X_{past}, Y_{past}),\qquad(3)$$

where $DC(\Delta X|X_{past})$ is as defined in Eq. 1 and $DC(\Delta X|X_{past}, Y_{past})$ is as elaborated below:

$$DC(\Delta X|X_{past}, Y_{past}) = C(X_{past} + \Delta X, Y_{past} + \Delta X) - C(X_{past}, Y_{past}),\qquad(4)$$

3/19

128 where $C(\cdot,\cdot)$ refers to joint complexity. *ICC* varies with the moving temporal window ΔX and its
129 corresponding Y_{past}, X_{past}. To estimate the average causality from time series Y to X, $ICC_{Y_{past}\to\Delta X}$
130 obtained for all ΔXs are averaged.

The above is the generic description of ICC that can be estimated using any complexity measure. For the reasons discussed in Section 2, we would like to estimate ICC using the notion of Dynamical Compression-Complexity estimated by the measure ETC. The measure would then become Interventional Compression-Complexity Causality. For succinctness, we refer to it as Compression-Complexity Causality (CCC). To estimate CCC, time series blocks X_{past}, Y_{past}, $X_{past} + \Delta X$, and surgically created $Y_{past} + \Delta X$ are separately encoded (binned) — converted to a sequence of symbols using '*B*' uniformly sized bins for the application of ETC[1]. For the binned time series blocks, X_{past}, Y_{past}, $X_{past} + \Delta X$, $Y_{past} + \Delta X$, to determine whether Y_{past} caused ΔX or not, we first compute dynamical compression-complexities, denoted by *CC*,

$$CC(\Delta X | X_{past}) = ETC(X_{past} + \Delta X) - ETC(X_{past}), \tag{5}$$

$$CC(\Delta X | X_{past}, Y_{past}) = ETC(X_{past} + \Delta X, Y_{past} + \Delta X) \qquad - ETC(X_{past}, Y_{past}), \tag{6}$$

131 Eq. 5 gives the dynamical compression-complexity of ΔX as a dynamical evolution of X_{past} alone. Eq. 6
132 gives the dynamical compression-complexity for ΔX as a dynamical evolution of both X_{past} and Y_{past}.
133 $ETC(\cdot)$ and $ETC(\cdot,\cdot)$ refer to individual and joint *effort-to-compress* complexities (see Section 1 of
134 supplementary material). For estimating ETC from these small blocks of data, short-term stationarity of
135 X and Y is assumed.

We now define Compression-Complexity Causality $CCC_{Y_{past}\to\Delta X}$ as:

$$CCC_{Y_{past}\to\Delta X} = CC(\Delta X | X_{past}) - CC(\Delta X | X_{past}, Y_{past}). \tag{7}$$

Averaged CCC from Y to X over the entire length of the time series with the window ΔX being slided by a step-size of δ is estimated as —

$$CCC_{Y\to X} = \overline{CCC}_{Y_{past}\to\Delta X} = \overline{CC}(\Delta X | X_{past}) - \overline{CC}(\Delta X | X_{past}, Y_{past}), \tag{8}$$

136 If $\overline{CC}(\Delta X | X_{past}, Y_{past}) \approx \overline{CC}(\Delta X | X_{past})$, then $CCC_{Y\to X}$ is statistically zero, implying no causal influ-
137 ence from Y to X. If $CCC_{Y\to X}$ is statistically significantly different from zero, then we infer that Y causes
138 X. A higher magnitude of $CCC_{Y\to X}$ implies a higher degree of causation from Y to X. The length of
139 X_{past}, Y_{past}, L is chosen by determining the correct intervention point. This is the temporal scale at which
140 Y has a dynamical influence on X. Detailed criteria and rationale for estimating L and other parameters
141 used in *CCC* estimation — w (length of ΔX), δ and B for any given pair of time series are discussed in
142 Section 3 of the supplementary material.

For multivariate data, CCC can be estimated in a similar way by building dictionaries that encode information from all variables. Thus, to check conditional causality from Y to X amidst the presence of other variables (say Z and W), two time varying dictionaries are built — D that encodes information from all variables (X, Y, Z, W) and D' that encodes information from all variables except Y $(X, Z, W$ only$)$. Once synchronous time series blocks from each variable are binned, the dictionary at that time point is constructed by obtaining a new sequence of symbols, with each possible combination of symbols from all variables being replaced by a particular symbol. The mechanism for construction of these dictionaries are discussed in Section 2 of the supplementary material. Subsequently, dynamical compression-complexities are computed as:

$$CC(\Delta X | D'_{past}) = ETC(D'_{past} + \Delta X) - ETC(D'_{past}), \tag{9}$$

$$CC(\Delta X | D_{past}) = ETC(D_{past} + \Delta X) \qquad - ETC(D_{past}), \tag{10}$$

143 where $D'_{past} + \Delta X$ represents the lossless encoding of joint occurrences of binned time series blocks
144 $X_{past} + \Delta X$, $Z_{past} + \Delta X$, $W_{past} + \Delta X$ and D'_{past} refers to the lossless encoding of joint occurrences of
145 binned time series blocks X_{past}, Z_{past} and W_{past}. Similarly, $D_{past} + \Delta X$ represents the lossless encoding of
146 joint occurrences of binned time series blocks $X_{past} + \Delta X$, $Y_{past} + \Delta X$, $Z_{past} + \Delta X$, $W_{past} + \Delta X$ and D_{past}
147 refers to the the lossless encoding of joint occurrences of binned time series blocks X_{past}, Y_{past}, Z_{past} and
148 W_{past}.

---

[1]Henceforth, the same variables are used to denote the binned/encoded versions of the blocks.

Conditional Compression-Complexity Causality, $CCC_{Y_{past} \to \Delta X | Z_{past}, W_{past}}$, is then estimated as the difference of Eq. 9 and Eq. 10. Averaged Conditional Compression Complexity-Causality over the entire time series with the window ΔX being slid by a step-size of δ is given as below:

$$CCC_{Y \to X | Z, W} = \overline{CC}(\Delta X | D') - \overline{CC}(\Delta X | D). \tag{11}$$

4 POSITIVE AND NEGATIVE CCC

The dynamical compression-complexities estimated for the purpose of CCC estimation, $CC(\Delta X | X_{past})$ and $CC(\Delta X | X_{past}, Y_{past})$, can be either positive or negative. For instance, consider the case when $CC(\Delta X | X_{past})$ becomes negative. This happens when $ETC(X_{past} + \Delta X)$ is less than $ETC(X_{past})$, which means that with the appending of ΔX, the sequence X_{past} has become more structured resulting in reduction of its complexity. The value of $CC(\Delta X | X_{past})$ is positive when appending of ΔX makes X_{past} less structured (hence more complex). Similarly, $CC(\Delta X | X_{past}, Y_{past})$ can also become negative when ETC realizes $X_{past} + \Delta X, Y_{past} + \Delta X$ to be more structured than X_{past}, Y_{past}. When the opposite is true, $CC(\Delta X | X_{past}, Y_{past})$ is positive.

Because of the values that $CC(\Delta X | X_{past})$ and $CC(\Delta X | X_{past}, Y_{past})$ can take, $CCC_{Y_{past} \to \Delta X}$ can be both positive or negative. How different cases result with different signs of the two quantities along with their implication on CCC is shown in Table 1.

From the table we see that the sign of $CCC_{Y_{past} \to \Delta X}$ signifies the 'kind of dynamical influence' that Y_{past} has on ΔX, whether this dynamical influence is similar to or different from that of X_{past} on ΔX. When $CCC_{Y_{past} \to \Delta X}$ is $-ve$, it signifies that Y_{past} has a different dynamical influence on ΔX than X_{past}. On the contrary, when $CCC_{Y_{past} \to \Delta X}$ is $+ve$, it signifies that Y_{past} has a dynamical influence on ΔX that is similar to that of X_{past}. On estimating the averaged CCC from time series Y to X, expecting that $CCC_{Y_{past} \to \Delta X}$ values do not vary much with time, we can talk about the kind of dynamical influence that time series Y has on X. For weak sense stationary processes, it is intuitive that the influence of Y on X would be very different from that on X due to its own past when the distributions of coupled time series Y and X are very different.

We verify this intuition by measuring probability distribution distances[2] between coupled processes Y and X using symmetric Kullback-Leibler Divergence (KL) and Jensen-Shannon Divergence (JSD). The trend of values obtained by these divergence measures is compared with the trend of CCC for different cases such as when CCC is positive or negative.

Coupled autoregressive (AR) processes were generated as per Eq. 15. Also, linearly and non-linearly coupled tent maps were generated as per Eq. 17, 18 and Eq. 17, 19 respectively. Symmetric KL and JSD between distribution P and Q of coupled processes are estimated as per Eq. 12 and 14 respectively.

$$D_{\text{Symm KL}}(P, Q) = D_{\text{KL}}(P \| Q) + D_{\text{KL}}(Q \| P), \tag{12}$$

where,

$$
\begin{aligned}
D_{\text{KL}}(P \| Q) &= \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right), \\
D_{\text{KL}}(Q \| P) &= \sum_i Q(i) \log \left(\frac{Q(i)}{P(i)} \right).
\end{aligned}
\tag{13}
$$

$$\text{JSD}(P \| Q) = \frac{1}{2} D(P \| M) + \frac{1}{2} D(Q \| M), \tag{14}$$

where, $M = \frac{1}{2}(P + Q)$. KL and JSD values are in unit of nats.

Curves for KL, JSD and CCC estimated for increasing coupling between AR processes of order 1, linearly coupled tent maps and non-linearly coupled tent maps are shown in Figures 1, 2 and 3 respectively. The values displayed represent the mean over 50 trials. As the degree of coupling is varied for AR

---

[2]It should be mentioned that strictly speaking KL and JSD are not distance measures since they don't satisfy the triangle inequality.

Table 1. Sign of Dynamical Compression-Complexities, $CC(\Delta X | X_{past})$ and $CC(\Delta X | X_{past}, Y_{past})$, and their resulting implication on the sign of estimated Compression Complexity-Causality, $CCC_{Y_{past} \to \Delta X}$.

$CC(\Delta X \| X_{past}, Y_{past})$ / $CC(\Delta X \| X_{past})$	$-ve$	$+ve$
$-ve$	$X_{past} + \Delta X$ was more structured than X_{past}. Further, two cases arise. 1. When $\|CC(\Delta X\|X_{past})\| > \|CC(\Delta X\|X_{past}, Y_{past})\|$, $CCC_{Y_{past} \to \Delta X} < 0$. Here, intervention by Y_{past} in the joint case **degraded the structure** by bringing patterns different from X_{past}. Dynamical influence of Y_{past} on ΔX is very different from the dynamical influence of X_{past} on ΔX. e.g.: CCC from independent tent map to dependent tent map. 2. When $\|CC(\Delta X\|X_{past})\| < \|CC(\Delta X\|X_{past}, Y_{past})\|$, $CCC_{Y_{past} \to \Delta X} > 0$. Intervention by Y_{past} in the joint case **enhanced the structure** by bringing patterns similar to X_{past}. Dynamical influence of Y_{past} on ΔX is very similar to the dynamical influence of X_{past} on ΔX .e.g.: CCC from independent autoregressive (AR) process to dependent AR process.	$CCC_{Y_{past} \to \Delta X}$ is $-ve$ always. $X_{past} + \Delta X$ was more structured than X_{past}. Intervention by Y_{past} in the joint case **degraded the structure**. Dynamical influence of Y_{past} on ΔX is very different from the dynamical influence of X_{past} on ΔX. e.g.: CCC from independent tent map to dependent tent map.
$+ve$	$CCC_{Y_{past} \to \Delta X}$ is $+ve$ always. $X_{past} + \Delta X$ was less structured than X_{past}. Intervention by Y_{past} in the joint case **enhanced the structure** by bringing patterns similar to X_{past}. Dynamical influence of Y_{past} on ΔX is very similar to the dynamical influence of X_{past} on ΔX.	$X_{past} + \Delta X$ was less structured than X_{past}. Further, two cases arise. 1. When $\|CC(\Delta X\|X_{past})\| > \|CC(\Delta X\|X_{past}, Y_{past})\|$, $CCC_{Y_{past} \to \Delta X} > 0$. Here, intervention by Y_{past} in the joint case **enhanced the structure** by bringing patterns similar to X_{past}. Dynamical influence of Y_{past} on ΔX is very similar to the dynamical influence of X_{past} on ΔX. 2. When $\|CC(\Delta X\|X_{past})\| < \|CC(\Delta X\|X_{past}, Y_{past})\|$, $CCC_{Y_{past} \to \Delta X} < 0$. Intervention by Y_{past} in the joint case **degraded the structure** by bringing patterns different from X_{past}. Dynamical influence of Y_{past} on ΔX is very different from the dynamical influence of X_{past} on ΔX .

processes, there is no clear pattern in KL and JSD values. The CCC values increase in the positive
direction as expected for increasing coupling, signifying that the dynamical influence from Y to X is
similar to the influence on X from its own past. Also, when we took larger number of trials for AR, the
values obtained by KL and JSD become confined to a smaller range and seem to converge towards a
constant value indicating that the distributions of X and Y are quite similar. However, in case of coupled
tent maps, as coupling is increased, the divergence between the distributions of the two coupled processes
increases, indicating that their distributions are becoming very different. The values of CCC grow in the
negative direction showing that with increasing coupling the independent process Y has a very different
dynamical influence on X compared to X's own past. Subsequently, due to the synchronization of Y and
X, KL, JSD as well as CCC become zero. With these graphs, it may not be possible to find an universal
threshold for the absolute values of KL/JSD above which CCC will show negative sign. However, if the
distributions of the two coupled processes exhibit an increasing divergence (when the coupling parameter
is varied) then it does indicate that the independent process would have a very different dynamical
influence on the dependent one when compared with that of the dependent process' own past suggesting
that the value of CCC will grow in the negative direction. The fact that KL/JSD and CCC do not have a
one-to-one correspondence is because the former (KL and JSD) operate on first order distributions while
the later (CCC) is able to capture higher-order dynamical influences between the coupled processes. For
non-stationary processes, our measure would still be able to capture the kind of dynamical influence,
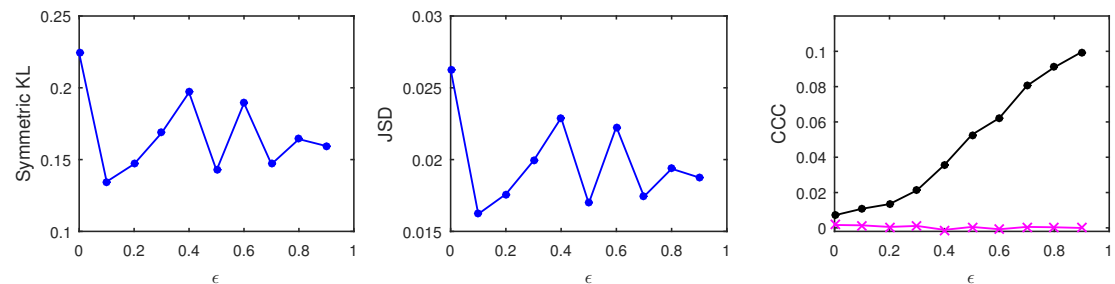though distributions are not static.

Figure 1. (color online). Divergence between distributions of coupled AR(1) processes using
Symmetric Kullback-Leibler (KL) and Jensen Shannon (JSD) divergences (in nats), and the causality
estimated using CCC from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta), as
the degree of coupling, ε is varied. CCC values increase with increasing ε. There is no similarity in the
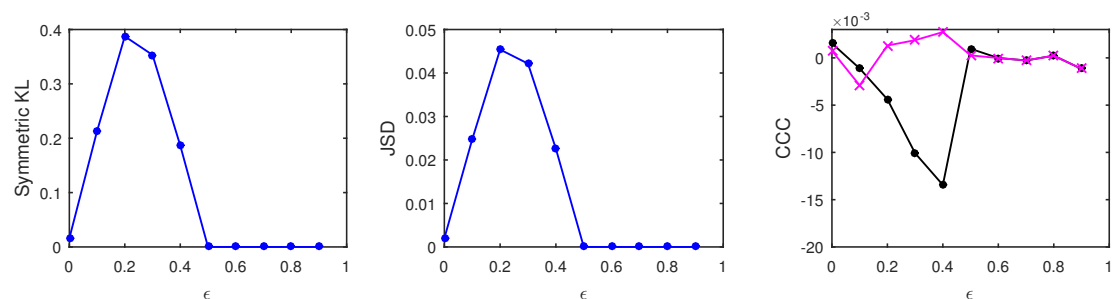trend of KL/JSD to CCC.

Figure 2. (color online). Distance between distributions of linearly coupled tent maps using Symmetric
Kullback Leibler (KL) and Jensen Shannon (JSD) divergences (in nats), and the causality estimated using
CCC from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta), as the degree of
coupling, ε is varied. For $\varepsilon < 0.5$, CCC and KL/JSD are highly negatively correlated.

Both positive and negative CCC implies significant causal influence (CCC$\approx$0 implies either no causal
influence or identical processes), but the nature of the dynamical influence of the *cause* on the *effect* is very
different in these two cases. Causality turning 'negative' does not seem very intuitive at first, but all that it
signifies is that the past of the *cause* variable makes the dynamics of the *effect* variable less unpredictable
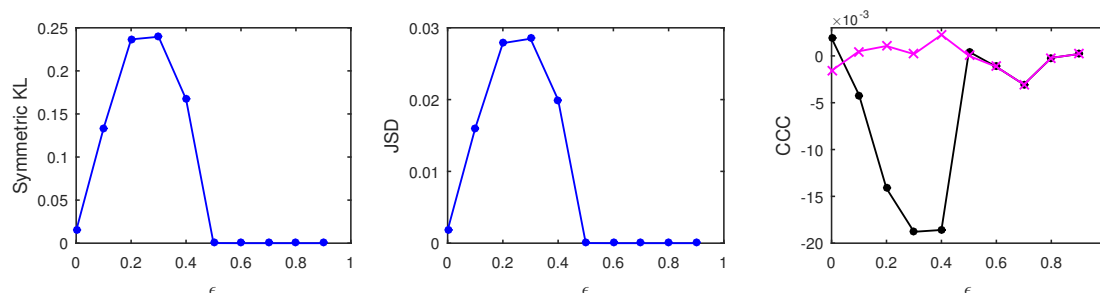
7/19

Figure 3. (color online). Distance between distributions of non-linearly coupled tent maps using Symmetric Kullback Leibler (KL) and Jensen Shannon (JSD) divergences (in nats), and the causality estimated using CCC from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta), as the degree of coupling, ε is varied. For $\varepsilon < 0.5$, CCC and KL/JSD are highly negatively correlated.

205 than its (*effect's*) own past. Such a unique feature could be very useful for real world applications in terms
206 of 'controlling' the dynamics of a variable being effected by several variables. If a particular *cause*, out of
207 several causes that makes the caused 'less predictable' and has 'intrinsically different' dynamics from that
208 of the *effect*, needs to be determined and eliminated, it can be readily identified by observing the sign of
209 CCC. Informed attempts to inhibit and enforce certain variables of the system can then be made.
210   As the existing model-free methods of causality can extract only 'associational causality' and ignore
211 the influence that the cause has on dynamics of the caused, it is impossible for them to comment on the
212 nature of this dynamical influence, something that CCC is uniquely able to accomplish. Obviously, model
213 based methods give full-fledged information about 'the kind of dynamical influence' owing to the model
214 equations assumed. However, if there are no equations assumed (or known), then the sign and magnitude
215 of CCC seems to be the best choice to capture the cause-effect relationship with additional information on
216 the similarity (or its lack of) between the two dynamics.

217 5 RESULTS AND DISCUSSION

218 A measure of causality to be robust for real data needs to perform well in the presence of noise, filtering,
219 low temporal and amplitude resolution, non-uniformly sampled signals, short length time series as well as
220 presence of other causal variables in the system. In this section, we rigorously simulate these cases and
221 evaluate the performance of CCC measure by comparing with existing measures — Granger Causality
222 (GC) and Transfer Entropy (TE). In the last sub-section, we also test CCC on real-world datasets. In
223 all cases, we take the averaged value of CCC over entire time series as computed by Eq. 11 and the
224 parameters for CCC estimation are chosen as per the selection criteria and rationale discussed in Section 3
225 of the supplementary material. GC estimation is done using the MVGC toolbox (Barnett and Seth, 2014)
226 in its default settings and TE estimation is done using MuTE toolbox (Montalto et al., 2014). Akaike
227 Information Criteria is used for model order estimation with the maximum model order set to 20 in the
228 MVGC toolbox, except where specified. In the MuTE toolbox, the approach of Non Uniform Embedding
229 (NUE) for representation of the history of the observed processes and of Nearest Neighbor (NN) estimator
230 for estimating the probability density functions is used for all results in this paper. The number of lags to
231 consider for observed processes was set to 5 and the maximum number of nearest neighbors to consider
232 was set to 10.

233 5.1 Varying unidirectional coupling

234 *5.1.1 AR(1)*

Autoregressive processes of order one (AR(1)) were simulated as follows. X and Y are the dependent and independent processes respectively.

$$
\begin{aligned}
X(t) &= aX(t-1) + \varepsilon Y(t-1) + \varepsilon_{X,t} \\
Y(t) &= bY(t-1) + \varepsilon_{Y,t},
\end{aligned}
\tag{15}
$$

235 where $a = 0.9$, $b = 0.8$, $t = 1$ to 1000s, sampling period = 1s. ε is varied from $0-0.9$ in steps of 0.1.
236 Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where ν = noise intensity = 0.03 and η follows standard normal distribution.

8/19

Figure 4 shows the performance of CCC along with that of TE and GC as mean values over 50 trials, (CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$).
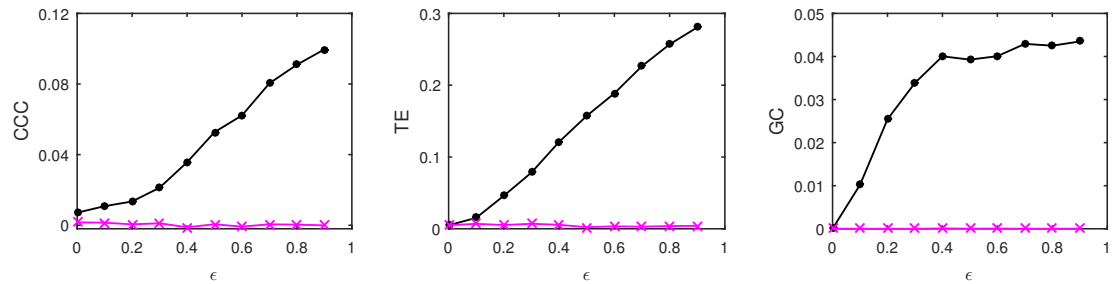

Figure 4. (color online). Causality estimated using CCC, TE and GC for coupled AR(1) processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling, ε is varied. CCC, TE as well as GC are able to correctly quantify causality.

With increasing coupling, the causality estimated by CCC and TE increases. For GC, however, causality increases with increasing coupling initially then settles to a constant value.

5.1.2 AR(100)

Autoregressive processes of order hundred (AR(100): X dependent, Y independent) were simulated as follows.

$$
\begin{aligned}
X(t) &= aX(t-1) + \varepsilon Y(t-100) + \varepsilon_{X,t} \\
Y(t) &= bY(t-1) + \varepsilon_{Y,t},
\end{aligned}
\tag{16}
$$

where $a = 0.9$, $b = 0.8$, $t = 1$ to 1000s, sampling period $= 1$s. ε is varied from $0 - 0.9$ in steps of 0.1. Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where ν = noise intensity = 0.03 and η follows standard normal distribution. Figure 5 shows the performance of CCC along with that of TE and GC, as mean values over 50 trials (CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$). Maximum model order was set to 110 in the MVGC toolbox.
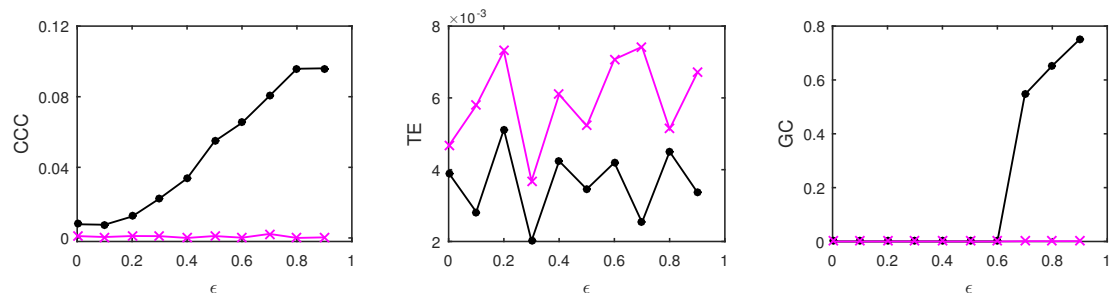

Figure 5. (color online). Causality estimated using CCC, TE and GC for coupled AR(100) processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling, ε is varied. Only CCC is able to reliably estimate the correct causal relationship for all values of ε.

CCC values increase steadily with increasing coupling for the correct direction of causation. TE fails as it shows higher causality from X to Y for all ε. GC performs poorly for low values of coupling. GC shows significant causality from Y to X only when $\varepsilon \geq 0.7$. Thus, causality in coupled AR processes with long-range memory can be reliably estimated using CCC and not using TE or GC.

5.1.3 Tent Map

Linearly and non-linearly coupled tent maps were simulated as per the following equations. Independent process, Y, is generated as:

$$
\begin{aligned}
Y(t) &= 2Y(t-1), &\quad 0 \leq Y(t-1) < 1/2, \\
Y(t) &= 2 - 2Y(t-1), &\quad 1/2 \leq Y(t-1) \leq 1.
\end{aligned}
\tag{17}
$$

The linearly coupled dependent process, X, is as below:

$$X(t) = \varepsilon Y(t) + (1-\varepsilon)h(t),$$
$$h(t) = 2X(t-1), \qquad 0 \leq X(t-1) < 1/2,$$
$$h(t) = 2 - 2X(t-1), \qquad 1/2 \leq X(t-1) \leq 1,$$

(18)

254 where ε is the degree of linear coupling.

The non-linearly coupled dependent process, X, is as below:

$$X(t) = 2f(t), \qquad 0 \leq f(t) < 1/2,$$
$$X(t) = 2 - 2f(t), \qquad 1/2 \leq f(t) \leq 1,$$
$$f(t) = \varepsilon Y(t-1) + (1-\varepsilon)X(t-1),$$

(19)

255 where ε is the degree of non-linear coupling.
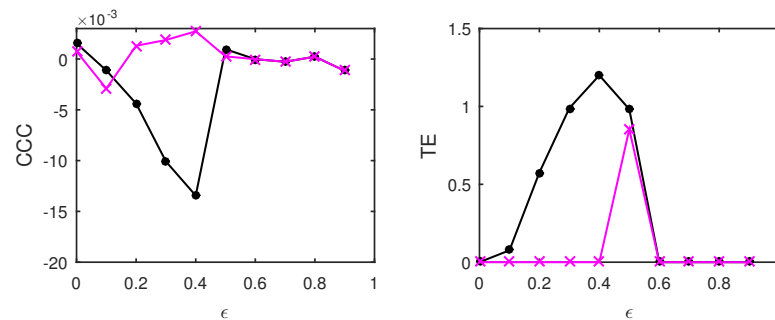

Figure 6. (color online). Causality estimated using CCC and TE for linearly coupled tent maps, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling is increased. With increasing coupling (until synchronization), magnitude of CCC and TE values increases. CCC values are negative while TE are positive.
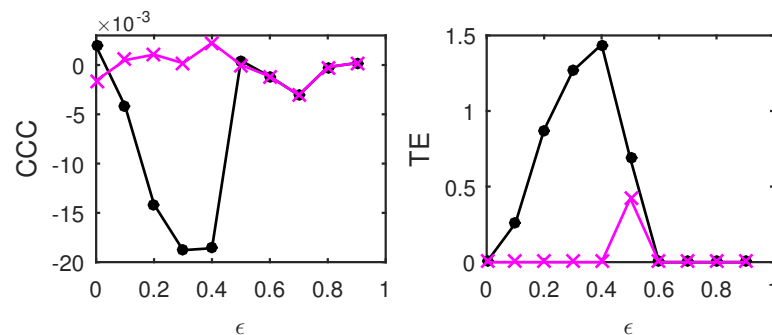

Figure 7. (color online). Causality estimated using CCC and TE for non-linearly coupled tent maps, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling is increased. With increasing coupling (until synchronization), magnitude of CCC and TE values increases. CCC values are negative while TE are positive.

256 The length of the signals simulated in both cases was 3000, i.e. $t = 1$ to 3000s, sampling period = 1s
257 and the first 2000 transients were removed to yield 1000 points for causality estimation. Figure 6 shows
258 the performance of CCC and TE for linearly coupled tent maps and Figure 7 shows the same for non
259 linearly coupled tent maps as ε is varied (CCC settings: $L = 100$, $w = 15$, $\delta = 80$, $B = 8$). The assumption
260 of a linear model for estimation of GC was proved to be erroneous for most trials and hence GC values
261 are not displayed. As ε is increased for both linear and non-linear coupling, $TE_{Y \to X}$ increases in the
262 positive direction and then falls to zero when the two series become completely synchronized at $\varepsilon = 0.6$.
263 The trend of the magnitude of CCC values is similar to TE, however, $CCC_{Y \to X}$ increment is in negative

²⁶⁴ direction. This is because of the fact that with increasing coupling the kind of dynamical influence from Y
²⁶⁵ to X becomes increasingly different than the dynamical influence from the past values of X to itself.

5.2 Varying process noise

²⁶⁷ The performance of measures as the process noise is varied is shown in Figure 8 for coupled AR processes
²⁶⁸ simulated as in Eq. 15, where $a = 0.9$, $b = 0.8$, $\varepsilon = 0.8$, $t = 1$ to 1000s, sampling period = 1s, number of
²⁶⁹ trials = 50. Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where ν = noise intensity is varied from 0.01 to 0.1 and η follows
²⁷⁰ standard normal distribution. CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$.
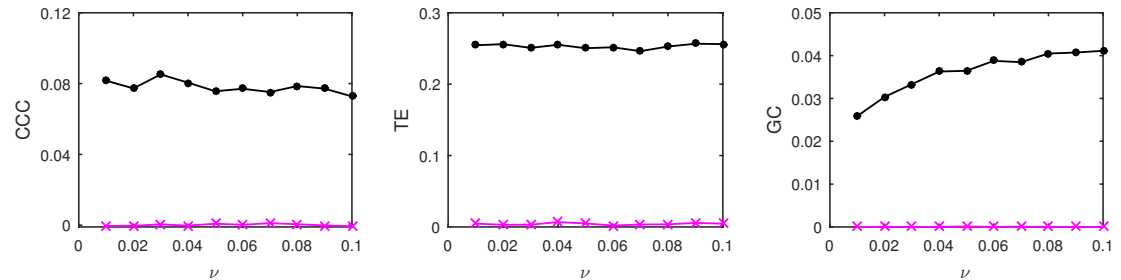

Figure 8. (color online). Causality estimated using CCC, TE and GC for coupled AR processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the intensity of noise, ν is varied. All the three measures perform well in this case.

²⁷¹   The performance of all three measures is fairly good in this case. Only GC values show a slightly
²⁷² increasing trend with increasing noise intensity.

5.3 Decimated coupled signals with uniform sampling

²⁷⁴ It is often the case that the rate of sampling of acquired measurements is not equal to the rate of generation
²⁷⁵ of the process. Causal inferences are regularly made from such data (Smirnov and Bezruchko, 2012), for
²⁷⁶ e.g., fMRI signals (Glover, 2011; Kim et al., 1997) as well as other neurophysiological recordings (de Abril
²⁷⁷ et al., 2018), climate data (Moberg et al., 2005). Two sets of coupled AR processes were first simulated
²⁷⁸ and subsequently decimated.

Set 1 of AR processes, of order 1, were simulated as below:

$$\begin{aligned} Y(t) &= 0.7Y(t-1) + \varepsilon_{Y,t}, \\ X(t) &= 0.9X(t-1) + 0.8Y(t-1) + \varepsilon_{X,t}. \end{aligned} \tag{20}$$

Set 2 of AR processes, of order 5, were simulated as below:

$$\begin{aligned} Y(t) &= 0.7Y(t-5) + \varepsilon_{Y,t}, \\ X(t) &= 0.9X(t-5) + 0.8Y(t-1) + \varepsilon_{X,t}, \end{aligned} \tag{21}$$

²⁷⁹   where, noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where ν = noise intensity = 0.03 and η follows standard normal
²⁸⁰ distribution. The original length of X and Y simulated in both the sets is 2000. Upon decimation, the
²⁸¹ length of the time series reduces. β represents the decimation factor that scales the sampling frequency.
²⁸² As β is varied from 1 to 0.5, sampling frequency is scaled from its original value to half its value.

²⁸³   *Set 1* of processes, being of order one, have low frequency components in the signal. As a result,
²⁸⁴ even when β is reduced to 0.5, it does not lead to frequency folding in the spectrum of process Y and X.
²⁸⁵ Frequency spectrum for a trial of X in this process is shown in Figure 9 for the original case and the case
²⁸⁶ where β is reduced to 0.5. In case of *Set 2*, decimation of the signals Y and X leads to aliasing. This is
²⁸⁷ because higher frequency components are present in the signals, leading to folding of these frequencies
²⁸⁸ even as β is reduced to 0.8. The frequency spectrum for a trial of X for its non-decimated version and for
²⁸⁹ decimation with β equal to 0.75 is shown in Figure 10.

5.3.1 Equal decimation of independent and dependent signal

²⁹¹ When both signals Y and X from the two sets are decimated by scaling their sampling rate by an equal
²⁹² decimation factor, β, ranging from 1 to 0.5 at intervals of 0.05, the results obtained using the three
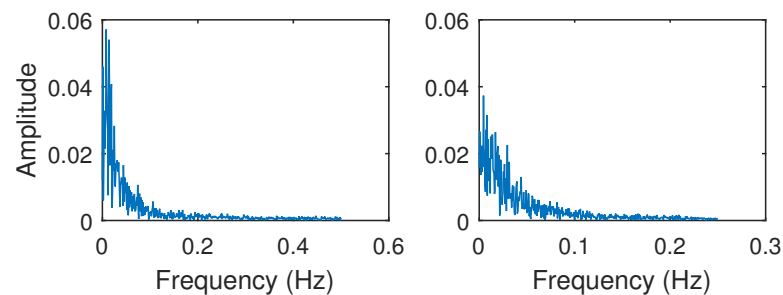
11/19

Figure 9. Frequency Spectrum of dependent AR(1) process from *Set 1* without decimation (left) and when decimation factor equals 0.5 (right). The process does not undergo aliasing on decimation.
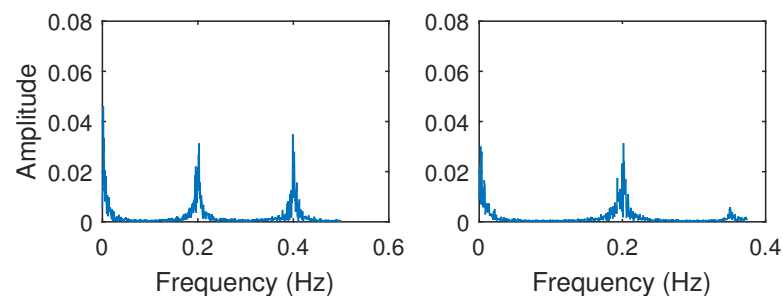

Figure 10. Frequency Spectrum of of dependent AR(5) process from *Set 2* without decimation (left) and when decimation factor equals 0.75 (right). The process undergoes aliasing on decimation.

methods, CCC, TE and GC are as shown in Figures 11 and 12. Figure 11 shows the results for Set 1 while Figures 12 shows results for Set 2 as mean causality values estimated over 10 trials. CCC settings for both sets: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$.
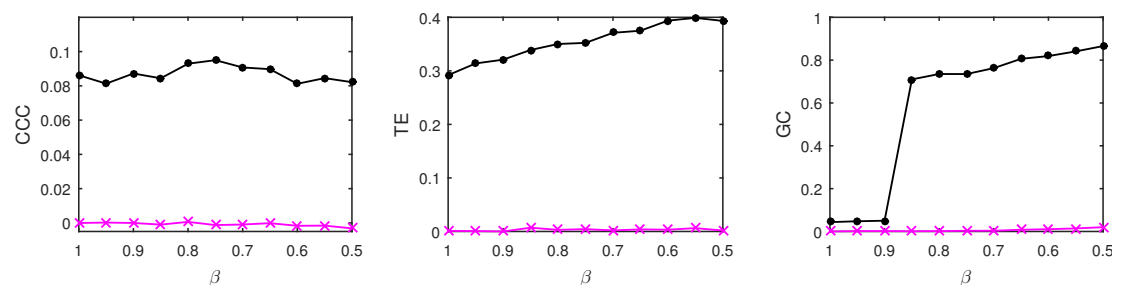

Figure 11. (color online). Causality estimated using CCC, TE and GC for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the decimation factor β is varied for both independent and dependent signal. The coupled AR processes simulated do not undergo frequency aliasing. CCC and TE values are relatively stable compared to GC.

While the values of CCC are relatively consistent even upon decimation (with or without aliasing) those of TE are stable only in case of non-aliased decimation. For GC (in both cases) and TE (in the aliased case), even though there is no confounding of the causality direction, the magnitude of causality estimated is not consistent and reliable.

5.3.2 Decimation of dependent signal

When only signal X is decimated by scaling its sampling rate by a decimation factor β, ranging from 1 to 0.5 at intervals of 0.05, the results obtained using the three methods CCC, TE and GC are as shown in Figures 13 and 14. Figure 13 shows the results for *Set 1* while Figure 14 shows results for *Set 2* as mean causality values estimated over 10 trials. For the length of the two signals to match, the independent
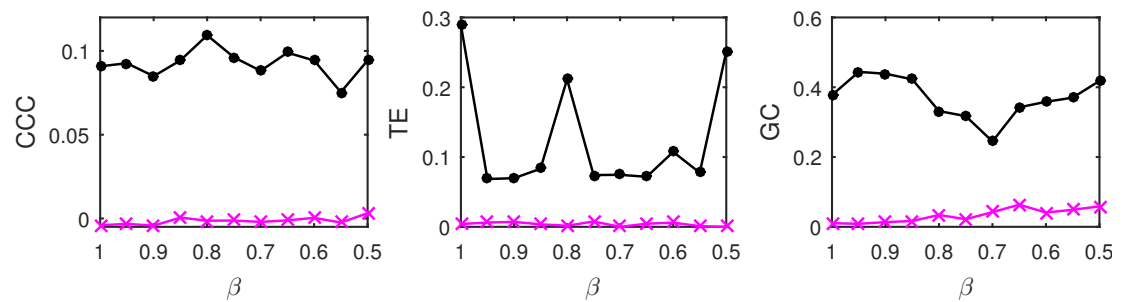
Figure 12. (color online). Causality estimated using CCC, TE and GC for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the decimation factor β is varied for both independent and dependent signal. The coupled AR processes simulated become frequency aliased. CCC values are stable compared to TE and GC.

306   signal considered is truncated at the length of the dependent signal. CCC settings for both sets: $L = 150$,
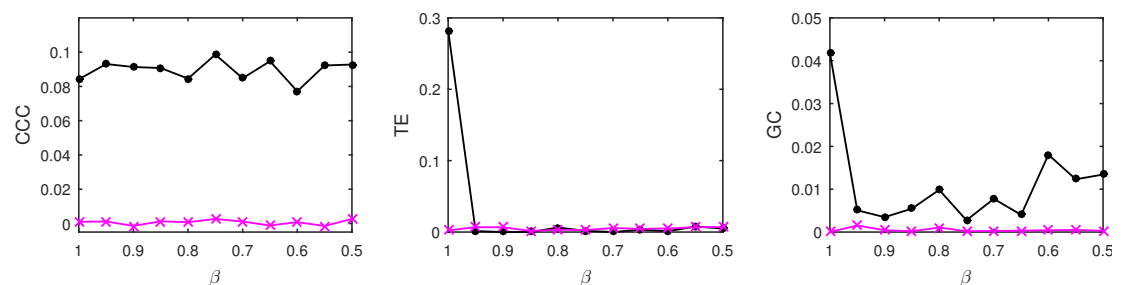307 $w = 15, \delta = 80, B = 2$.

Figure 13. (color online). Causality estimated using CCC, TE and GC for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the decimation factor β is varied for the dependent signal. The dependent AR process simulated does not undergo frequency-aliasing. Only CCC can capture the correct direction and strength of coupling when β is decreased.
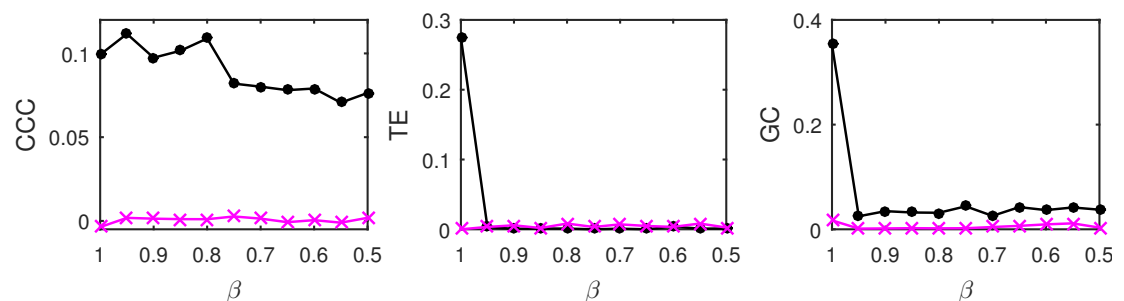

Figure 14. (color online). Causality estimated using CCC, TE and GC for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the decimation factor is varied for the dependent signal. The dependent AR process simulated becomes frequency aliased. Only CCC can capture the correct direction and strength of coupling when β is decreased.

308   In this scenario, for both non-aliased and aliased decimation, CCC estimates are much more stable
309   and consistent (across β) when compared to those of TE and GC, where confounding in the direction of
310   causality results even upon slightest decimation. It is clear from these results that CCC is the most robust,
311   reliable and consistent among the three causality measures.

5.4 Non uniform sampling

Non-uniformly sampled/non-synchronous measurements are common in real-world physiological data acquisition due to jitters/motion-artifacts as well as due to the inherent nature of signals such as heart rate signals (Laguna et al., 1998). Also, in economics, the case of missing data is common (Baumöhl and Vÿrost, 2010). To realistically simulate such a scenario, non-uniform sampling was introduced by eliminating data from random locations of the dependent time series and then presenting the resulting series as a set with no knowledge of the time-stamps of the missing data. The percentage of non-uniform sampling/non-synchronous measurements (α) is the percentage of these missing data points.

AR processes with non-uniformly sampled signals were simulated as per Eq. 15 with $b = 0.7$, $a = 0.9$, $\varepsilon = 0.8$. Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where ν = noise intensity = 0.03 and η follows standard normal distribution. Length of original time series, $N = 2000$, and is reduced upon increasing the percentage non-uniform sampling α. In order to match the lengths of the two time series, Y, the independent time series, is appropriately truncated to match the length of the dependent signal, X (this results in a non-synchronous pair of measurements). CCC settings used: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$. Mean causality estimated for 10 trials using the three measures with increasing increasing α, while $\nu = 0.03$, are shown in Figure 15.

Linearly coupled tent maps with non-uniformly sampled signals were simulated as per Eq. 17 and 18 with $\varepsilon = 0.3$. Length of original time series, $N = 2000$, and is reduced upon increasing the percentage non-uniform sampling α. In order to match the lengths of the two time series, Y, the independent time series, is appropriately truncated to match the length of the dependent signal, X (this results in a non-synchronous pair of measurements). CCC settings used: $L = 100$, $w = 15$, $\delta = 80$, $B = 8$. Mean causality estimated for 10 trials using the three measures with increasing increasing α, while $\nu = 0.03$, are shown in Figure 16.
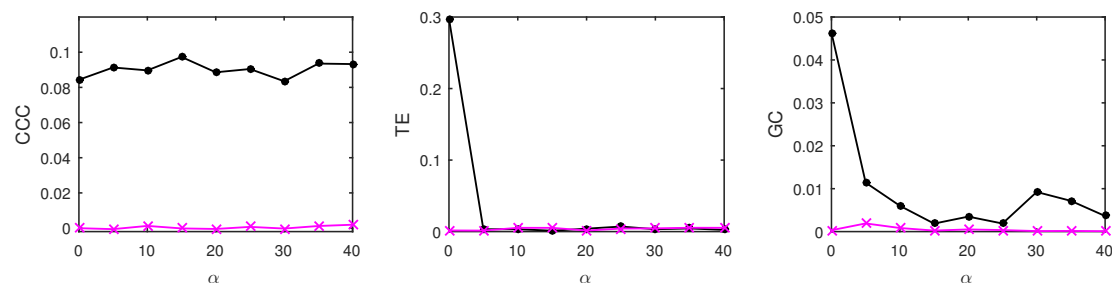

Figure 15. (color online). Causality estimated using CCC, TE and GC for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the percentage of non-uniform sampling α is varied. CCC is the only measure that shows reliable, consistent and correct estimates of causality.
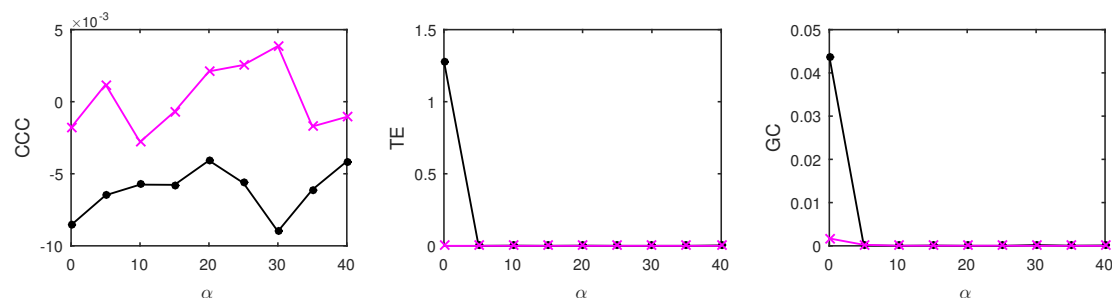

Figure 16. (color online). Causality estimated using CCC, TE and GC for coupled tent maps from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the percentage of non-uniform sampling is varied. CCC is able to estimate the correct causality direction while TE and GC fail.

As the results clearly indicate, both TE and GC fail when applied to non-uniformly sampled coupled AR and tent map processes. CCC values are relatively invariant to non-uniform sampling and thus could

337 be employed in such scenarios.

5.5 Filtering of Coupled Signals

339 Acquired data preprocessing often involves low pass filtering to smooth out the signal (Teplan et al., 2002).
340 At other times, high pass filtering is required to remove low frequency glitches from a high frequency
341 signal. Also when the signals acquired are sampled at low frequencies, the effects due to decimation
342 and filtering may add up and result in poorer estimates of causality. This is often the case in fMRI
343 signals (Glover, 2011; Kim et al., 1997).
344 To test these scenarios, causalities were estimated using the three measures when coupled AR processes
345 simulated as per Eq. 21, are low pass filtered using a moving average window of length 3 with step size
346 1. The results are shown in Table 2 as mean values over 10 trials. CCC settings used: $L = 150$, $w = 15$,
347 $\delta = 80$, $B = 2$. The performance of the measures when the coupled signals are decimated to a decimation
348 factor of 0.5 and then low pass filtered are also included in the table. The length of the original signal
349 simulated is 2000 and is reduced to 1998 upon filtering and to 998 upon filtering and decimation.

Table 2. CCC, TE and GC estimates for coupled AR processes Y (independent) and X (dependent) as it
is, upon filtering and upon decimation and filtering

System	CCC		TE		GC	
	$Y \rightarrow X$	$X \rightarrow Y$	$Y \rightarrow X$	$X \rightarrow Y$	$Y \rightarrow X$	$X \rightarrow Y$
Original	0.0908	-0.0041	0.2890	0.0040	0.3776	0.0104
Filtered	0.0988	0.0018	0.2398	0.0170	0.4787	0.0056
Decimated and Filtered	0.0753	0.0059	0.1270	0.0114	0.4321	0.0596

350 From the table, we see that CCC can distinguish the direction of causality in the original case as well
351 as in the filtering and decimation plus filtering case. Erroneously, TE shows significant causality in the
352 direction opposite to causation upon filtering as well as upon decimation and filtering and GC shows
353 significant causality in the direction opposite to causation upon decimation and filtering. By this we can
354 infer that CCC is highly suitable for practical applications which involve pre-processing such as filtering
355 and decimation of measurements.

5.6 Conditional CCC on short length MVAR system

357 A system of three variables was simulated as per the following equations —

$$
\begin{aligned}
Z(t) &= 0.8Z(t-1) + \varepsilon_{Z,t}, \\
X(t) &= 0.9X(t-1) + 0.4Z(t-100) + \varepsilon_{X,t}, \\
Y(t) &= 0.9Y(t-1) + 0.8Z(t-100) + \varepsilon_{Y,t},
\end{aligned}
\tag{22}
$$

358 where the noise terms, $\varepsilon_Z, \varepsilon_X, \varepsilon_Y = v\eta$, v = noise intensity = 0.03 and η follows standard normal
359 distribution. Length of time series simulated was 300 and first 50 transients were removed to yield short
360 length signals of 250 time points.
361 The coupling direction and strength between variables X, Y, Z are shown in Figure 17(a). The mean
362 values of causality estimated over 10 trials using CCC, TE and GC are shown in Figure 17 tables, (b), (c)
363 and (d) respectively. CCC settings used: $L = 150$, $w = 15$, $\delta = 20$, $B = 2$. In the tables, true positives
364 are in green, true negatives in black, false positives in red and false negatives in yellow. CCC detects
365 correctly the true positives and negatives. GC, detects the true positives but also shows some false positive
366 couplings. TE, performs very poorly, falsely detecting negatives where coupling is present and also
367 showing false positives where there is no coupling.

5.7 Real Data

369 CCC was applied to estimate causality on measurements from two real-world systems and compared
370 with TE. System (a) comprised of short time series for dynamics of a complex ecosystem, with 71 point

From / To	X	Y	Z
X	0	0.0029	0.0420
Y	-0.0026	0	0.0885
Z	0.0022	0.0046	0

(a) Connectivity   **(b)** CCC

From / To	X	Y	Z
X	0	0.0615	0.0023
Y	0.1122	0	0.0022
Z	0.0055	0.0205	0

(c) TE

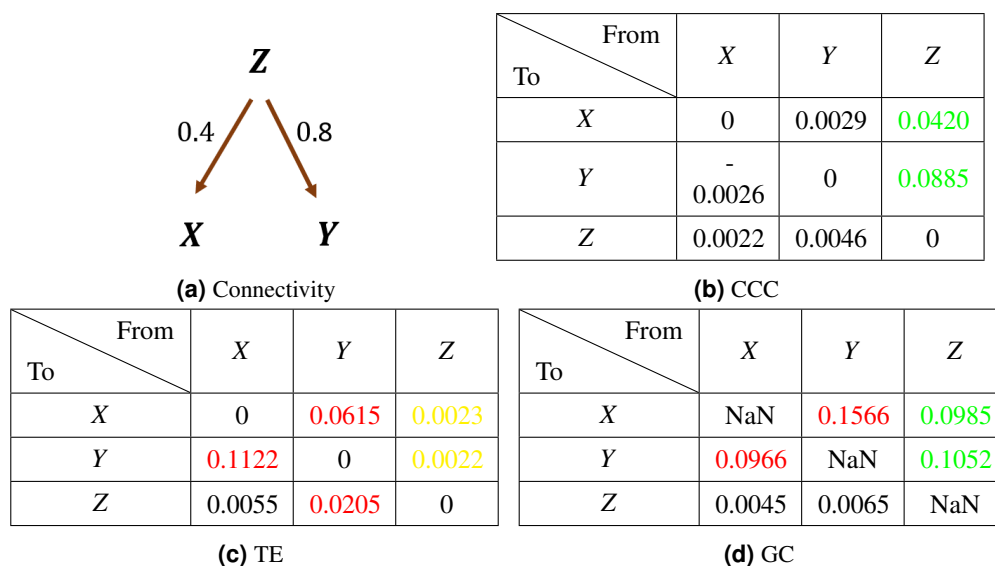| From / To	X	Y	Z
X	NaN	0.1566	0.0985
Y	0.0966	NaN	0.1052
Z	0.0045	0.0065	NaN

(d) GC

Figure 17. Causalities estimated using CCC, TE and GC for a system of three AR variables coupled as in (a). True positives are in green, true negatives in black, false positives in red and false negatives in yellow.
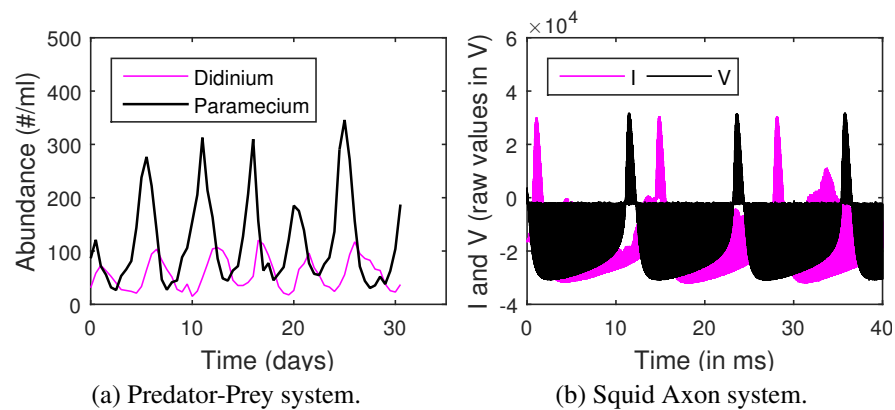
recording of predator (Didinium) and prey (Paramecium) populations, reported in Veilleux (1976) and originally acquired for Jost and Ellner (2000), with first 9 points from each series removed to eliminate transients (Figure 18(a)). Length of signal on which causality is computed, $N = 62$, CCC settings used: $L = 40$, $w = 15$, $\delta = 4$, $B = 8$. CCC is seen to aptly capture the higher (and direct) causal influence from predator to prey population and lower influence in the opposite direction (see Figure 18). The latter is expected, owing to the indirect effect of the change in prey population on predator. CCC results are in line with that obtained using Convergent Cross Mapping (Sugihara et al., 2012). TE, on the other hand, fails to capture the correct causality direction.

System (b) comprised of raw single-unit neuronal membrane potential recordings (V, in 10V) of squid giant axon in response to stimulus current (I, in V, 1V=5 $\mu A/cm^2$), recorded in Paydarfar et al. (2006) and made available by Goldberger et al. (2000). We test for the causation from I to V for three axons (1 trial each) labeled 'a3t01', 'a5t01' and 'a7t01', extracting 5000 points from each recording. Length of signal on which causality is computed, $N = 5000$, CCC settings used: $L = 75$, $w = 15$, $\delta = 50$, $B = 2$. We find that $CCC_{I \to V}$ is less than or approximately equal to $CCC_{V \to I}$ and both values are less than zero for the three axons (Figure. 18), indicating negative causality in both directions. This implies bidirectional dependence between I and V. Each brings a different dynamical influence on the other when compared to its own past. TE fails to give consistent results for the three axons.

6 CONCLUSIONS

In this work, we have proposed a novel data-based, model-free intervention approach to estimate causality for given time series. The *Interventional Complexity Causality* measure (or ICC) based on capturing causal influences from the dynamical complexities of data is formalized as *Compression-Complexity Causality* (CCC) and is shown to have the following strengths —

- CCC operates on *windows* of the input time series (or measurements) instead of individual samples. It does not make any assumption of the separability of *cause* and *effect* samples.

- CCC doesn't make any assumptions of stochasticity, determinism, gaussianity, stationarity, linearity or markovian property. Thus, CCC is applicable even on non-stationary/ non-linear/ non-gaussian/ non-markovian, short-term and long-term memory processes, as well as chaotic processes. CCC characterizes causal relationship based on dynamical complexity computed from windows of the input data.

Figure 18. CCC, TE on real-world time series. Top (color online): (a) Time series showing population of *Didinium nasutum* (D_n) and *Paramecium aurelia* (P_n) as reported in Veilleux (1976), (b) Stimulus current (I) and voltage measurements (V) as recorded from a Squid Giant Axon ('a3t01') in Paydarfar et al. (2006). Bottom: Table showing CCC and TE values as estimated for systems (a) and (b).

System	Details	CCC		TE	
		$Y \rightarrow X$	$X \rightarrow Y$	$Y \rightarrow X$	$X \rightarrow Y$
Predator-Prey	$Y = D_n$ $X = P_n$	0.116	-0.021	0.249	0.252
Squid Axon	$Y = I_{a3t01}$ $X = V_{a3t01}$	-0.142	-0.129	0	0.049
	$Y = I_{a5t01}$ $X = V_{a5t01}$	-0.142	-0.135	0.055	0.108
	$Y = I_{a7t01}$ $X = V_{a7t01}$	-0.153	-0.154	0.196	0.127

- CCC is uniquely and distinctly novel in its approach since it does not estimate '*associational*' causality (first rung on Ladder of Causation) but performs '*intervention*' (second rung on the Ladder of Causation) to capture causal influences from the dynamics of the data.

- The point of 'intervention' (length L for creating the hypothetical data: $Y_{past} + \Delta X$) is dependent on the temporal scale at which causality exists within and between processes. It is determined adaptively based on the given data. This makes CCC a highly data-driven/data-adaptive method and thus suitable for a wide range of applications.

- Infotheoretic-based causality measures such as TE and others need to estimate joint probability densities which are very difficult to reliably estimate with short and noisy time series. On the other hand, CCC uses Effort-To-Compress (ETC) complexity measure over short windows to capture time-varying causality and it is well established in literature that ETC outperforms infotheoretic measures for short and noisy data (Nagaraj and Balasubramanian, 2017a; Balasubramanian and Nagaraj, 2016).

- CCC can be either positive or negative (unlike TE and GC). By this unique property, CCC gives information about the *kind* of causal influence that is brought by one time series on another, whether this influence is *similar* ($CCC > 0$) to or *different* ($CCC < 0$) from the influence that the series brings to its own present.

- Negative CCC could be used for 'control' of processes by intervening selectively on those variables which are dissimilar ($CCC < 0$)/similar ($CCC > 0$) in terms of their dynamics.

- CCC is highly robust and reliable, and overcomes the limitations of existing measures (GC and TE) in case of signals with long-term memory, low temporal resolution, noise, filtering, non-uniform

sampling (non-synchronous measurements), finite length signals, presence of common driving variables as well as on real datasets.

We have rigorously demonstrated the superior performance of CCC in this work. Given the above listed novel properties of CCC and its unique model-free, data-driven, data-adaptive intervention-based approach to causal reasoning, it has the potential to be applied in a wide variety of real-world applications. Future work would involve testing the measure on simulated networks with complex interactions as well as more real world datasets. We would like to further explore the idea of negative CCC and check its relation to Lyaupnov exponent (for chaotic systems) which can characterize the degree of chaos in a system. It is also worthwhile to explore the performance of other complexity measures such as Lempel-Ziv complexity for the proposed *Interventional Complexity Causality*.

We provide free open access to the CCC MATLAB toolbox developed as a part of this work. See Section 4 of supplementary material for details.

SUPPLEMENTARY MATERIALS

Supplementary text and code are made available.

ACKNOWLEDGMENTS

REFERENCES

Balasubramanian, K. and Nagaraj, N. (2016). Aging and cardiovascular complexity: effect of the length of RR tachograms. *PeerJ*, 4:e2755.

Barnett, L. and Seth, A. K. (2014). The mvgc multivariate granger causality toolbox: a new approach to granger-causal inference. *Journal of neuroscience methods*, 223:50–68.

Bauer, M., Cox, J. W., Caveness, M. H., Downs, J. J., and Thornhill, N. F. (2007). Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE transactions on control systems technology*, 15(1):12–21.

Baumöhl, E. and Vỳrost, T. (2010). Stock market integration: Granger causality testing with respect to nonsynchronous trading effects. *Finance a Uver*, 60(5):414.

Chiou-Wei, S. Z., Chen, C.-F., and Zhu, Z. (2008). Economic growth and energy consumption revisited: evidence from linear and nonlinear granger causality. *Energy Economics*, 30(6):3063–3076.

Cilibrasi, R. L. et al. (2007). Statistical inference through data compression.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

de Abril, I. M., Yoshimoto, J., and Doya, K. (2018). Connectivity inference from neural recording data: Challenges, mathematical bases and research directions. *Neural Networks*.

Dhamala, M., Rangarajan, G., and Ding, M. (2008). Estimating granger causality from fourier and wavelet transforms of time series data. *Physical review letters*, 100(1):018701.

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302.

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22(2):133–139.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424—-438.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664.

Jost, C. and Ellner, S. P. (2000). Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1453):1611–1620.

Kim, S.-G., Richter, W., and Uğurbil, K. (1997). Limitations of temporal resolution in functional mri. *Magnetic resonance in medicine*, 37(4):631–636.

Laguna, P., Moody, G. B., and Mark, R. G. (1998). Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715.

Lempel, A. and Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on information theory*, 22(1):75–81.

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2008). Kernel method for nonlinear granger causality. *Physical Review Letters*, 100(14):144103.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W. (2005). Highly variable northern hemisphere temperatures reconstructed from low-and high-resolution proxy data. *Nature*, 433(7026):613.

Montalto, A., Faes, L., and Marinazzo, D. (2014). Mute: a matlab toolbox to compare established and novel estimators of the multivariate transfer entropy. *PloS one*, 9(10):e109462.

Mosedale, T. J., Stephenson, D. B., Collins, M., and Mills, T. C. (2006). Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation. *Journal of climate*, 19(7):1182–1194.

Nagaraj, N. and Balasubramanian, K. (2017a). Dynamical complexity of short and noisy time series. *The European Physical Journal Special Topics*, pages 1–14.

Nagaraj, N. and Balasubramanian, K. (2017b). Three perspectives on complexity: entropy, compression, subsymmetry. *Eur. Phys. Journal Spec. Topics*, 226(15):3251–3272.

Nagaraj, N., Balasubramanian, K., and Dey, S. (2013). A new complexity measure for time series analysis and classification. *The European Physical Journal Special Topics*, 222(3-4):847–860.

Paydarfar, D., Forger, D. B., and Clay, J. R. (2006). Noisy inputs and the induction of on–off switching behavior in a neuronal pacemaker. *Journal of Neurophysiology*, 96(6):3338–3348.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85(2):461–464.

Seth, A. K., Barrett, A. B., and Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297.

Smirnov, D. and Bezruchko, B. (2012). Spurious causalities due to low temporal resolution: Towards detection of bidirectional coupling from time series. *EPL (Europhysics Letters)*, 100(1):10005.

Stips, A., Macias, D., Coughlan, C., Garcia-Gorriz, E., and San Liang, X. (2016). On the causal structure between co2 and global temperature. *Scientific reports*, 6.

Sugihara, G., May, R., Ye, H., Hsieh, C., and Deyle, E. (2012). Detecting causality in complex ecosystems. *Science*, 338(3):496–500.

Teplan, M. et al. (2002). Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11.

Veilleux, B. G. (1976). The analysis of a predatory interaction between didinium and paramecium. *Master's thesis. University of Alberta, Edmondton*.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy: a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1):45–67.