# 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, kathpaliaadit@nias.res.in

## 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/ data-driven 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 largely 'model-free'/ 'data-driven' measures of causality. They make minimal assumptions about the underlying physical mechanisms and depend more on time series characteristics (Seth et al., 2015). Hence, they have a wider scope compared to specific model assumptions made by methods such as Dynamic Causal Modelling (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 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 scales of measurements. As a result, cause and effect may co-exist in a single measurement or overlap over blocks of measurements, making them inseparable. In such a scenario, it would be incorrect to estimate causality by means of correlations and/or joint probabilities which implicitly assumes the separability of 'cause' and 'effect'. Both GC and TE make this assumption of separability. Circularly, to characterize a time series sample as purely a 'cause' or an 'effect' is possible only if there is a clear linear/markovian separable relationship. When cause and effect are inseparable, 'associational' measures of causality such as GC and TE are insufficient and we need a method to climb up the ladder of causation.

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

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

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

Other methods for causality estimation based on compression have been proposed in literature (Budhathoki and Vreeken, 2016; Wieczorek and Roth, 2016), but the very philosophy behind our method and its implementation are very different from these existing methods.

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

## 2 DYNAMICAL COMPLEXITY (DC) AND DYNAMICAL COMPRESSION-COMPLEXITY (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 process, (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 -

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

where $\Delta X$ is a moving window of length $w$ samples and $X_{past}$ is a window consisting of immediate past $L$ samples of $\Delta X$. '+' refers to appending, for e.g., for time series $A = [1,2,3]$ and $B = [p,q]$, then $A + B = [1,2,3,p,q]$. $C(X)$ refers to complexity of time series $X$. *DC*, thus varies with the temporal index of $\Delta X$ 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 -

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

where $C(X_{past}, \Delta X)$ is the *joint* complexity of $X_{past}$ and $\Delta X$. 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 $\Delta X$ 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, 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). Recently, ETC has been used to propose a compression-complexity measure for networks (Virmani and Nagaraj, 2019).

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. Before that, we describe how individual and joint compression complexities are computed using ETC (Nagaraj et al., 2013) in the subsections below.

## 2.1 ETC measure for a time series: $ETC(X)$

Since ETC expects a symbolic sequence as its input (of length $> 1$), the given time series should be binned appropriately to generate such a sequence. Once such a symbolic sequence is available, ETC

proceeds by parsing the entire sequence (from left to right) to find that pair of symbols in the sequence which has the highest frequency of occurrence. This pair is replaced with a new symbol to create a new symbolic sequence (of shorter length). This procedure is repeated iteratively and terminates only when we end up with a constant sequence (whose entropy is zero since it consists of only one symbol). Since the length of the output sequence at every iteration decreases, the algorithm will surely halt. The number of iterations needed to convert the input sequence to a constant sequence is defined as the value of ETC complexity. For example, the input sequence '12121112' gets transformed as follows: $12121112 \mapsto 33113 \mapsto 4113 \mapsto 513 \mapsto 63 \mapsto 7$. Thus, $ETC(12121112) = 5$. ETC achieves its minimum value (0) for a constant sequence and maximum value $(m-1)$ for a $m$ length sequence with distinct symbols. Thus, we normalize the ETC complexity value by dividing by $m-1$. Thus normalized $ETC(12121112) = \frac{5}{7}$. Note that normalized ETC values are always between 0 and 1 with low values indicating low complexity and high values indicating high complexity.

## 2.2 Joint ETC measure for a pair of time series: $ETC(X,Y)$

We perform a straightforward extension of the above mentioned procedure $(ETC(X))$ for computing the joint ETC measure $ETC(X,Y)$ for a pair of input time series $X$ and $Y$ of the same length. At every iteration, the algorithm scans (from left to right) simultaneously $X$ and $Y$ sequences and replaces the most frequent jointly occurring pair with a new symbol for both the pairs. To illustrate it by an example, consider, $X = 121212$ and $Y = abacac$. The pair $(X,Y)$ gets transformed as follows: $(121212, abacac) \mapsto (1233, abdd) \mapsto (433, edd) \mapsto (53, fd) \mapsto (6,g)$. Thus, $ETC(X,Y) = 4$ and normalized value is $\frac{4}{5}$. It can be noted that $ETC(X,Y) \leq ETC(X) + ETC(Y)$.

## 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 $\Delta X$. These blocks are created by 'surgery' and do not exist in reality in the data that is already collected. Interventional Complexity Causality (ICC) is defined as the change in the dynamical complexity of time series $X$ when $\Delta X$ 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.

This formulation is actually in line with Wiener's idea, according to which, time series $Y$ causes $X$, if incorporating the past of $Y$ helps to improve the prediction of $X$ (Wiener, 1956). While GC is based on the notion of improved predictability and TE on reduction of uncertainty, ICC is based on the notion of change in 'dynamical complexity' when information from the past of $Y$ is brought in, in order to check its causal influence on $X$. The difference between existing approaches and the proposed measure is that the effect of $Y$ on $X$ is analyzed based on 'associational' means in case of the former and by 'interventional' means in case of the latter. With this formulation, ICC is designed to measure *effect*, like GC and TE and not the *mechanism*, as in Dynamic Causal Modelling (Seth et al., 2015; Barrett and Barnett, 2013). To elaborate on this aspect, ICC cannot explicitly quantify the interaction coefficients of the underlying generative model (physical mechanism), but will only estimate causal influence based on change in dynamical complexities. It is, however, expected that ICC will be closer to the underlying mechanism than existing methods, because, by its very formulation, it taps on causes and their effects based on dynamical evolution of processes.

Mathematically,

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

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

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

where $C(\cdot, \cdot)$ refers to joint complexity. *ICC* varies with the moving temporal window $\Delta X$ and its corresponding $Y_{past}, X_{past}$. To estimate average causality from time series $Y$ to $X$, $ICC_{Y_{past} \to \Delta X}$ obtained for all $\Delta X$s 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 $\Delta X$ or not, we first compute dynamical compression-complexities, denoted by $CC$,

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

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

184 Eq. 5 gives the dynamical compression-complexity of $\Delta X$ as a dynamical evolution of $X_{past}$ alone. Eq. 6
185 gives the dynamical compression-complexity for $\Delta X$ as a dynamical evolution of both $X_{past}$ and $Y_{past}$.
186 $ETC(\cdot)$ and $ETC(\cdot, \cdot)$ refer to individual and joint *effort-to-compress* complexities. For estimating $ETC$
187 from these small blocks of data, short-term stationarity of $X$ 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 time series with the window $\Delta X$ being slided by a step-size of $\delta$ is estimated as —

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

188 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-
189 ence from $Y$ to $X$. If $CCC_{Y \to X}$ is statistically significantly different from zero, then we infer that $Y$ causes
190 $X$. A higher magnitude of $CCC_{Y \to X}$ implies a higher degree of causation from $Y$ to $X$. The length of
191 $X_{past}, Y_{past}, L$ is chosen by determining the correct intervention point. This is the temporal scale at which
192 $Y$ has a dynamical influence on $X$. Detailed criteria and rationale for estimating $L$ and other parameters
193 used in $CCC$ estimation: $w$ (length of $\Delta X$), $\delta$ and $B$ for any given pair of time series are discussed in
194 Section 3 of the supplementary material.

195 CCC is invariant to local/global scaling and addition of constant value to the time series. As CCC is
196 based on binning of small blocks of time series data, it is noise resistant. Furthermore, it is applicable to
197 non-linear and short term stationary time series. Being based on dynamical evolution of patterns in the
198 data, it is expected to be robust to sub-sampling and filtering.

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 1 of the supplementary material. Subsequently, dynamical compression-complexities are computed as:

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

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

199 where $D'_{past} + \Delta X$ represents the lossless encoding of joint occurrences of binned time series blocks
200 $X_{past} + \Delta X$, $Z_{past} + \Delta X$, $W_{past} + \Delta X$ and $D'_{past}$ refers to the lossless encoding of joint occurrences of
201 binned time series blocks $X_{past}$, $Z_{past}$ and $W_{past}$. Similarly, $D_{past} + \Delta X$ represents the lossless encoding of
202 joint occurrences of binned time series blocks $X_{past} + \Delta X$, $Y_{past} + \Delta X$, $Z_{past} + \Delta X$, $W_{past} + \Delta X$ and $D_{past}$
203 refers to the the lossless encoding of joint occurrences of binned time series blocks $X_{past}$, $Y_{past}$, $Z_{past}$ and
204 $W_{past}$.

---

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

Conditional Compression-Complexity Causality, $CCC_{Y_{past} \rightarrow \Delta X | Z_{past}, W_{past}}$, 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 $\Delta X$ being slided by a step-size of $\delta$ is given as below:

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

## 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 $\Delta X$, 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 $\Delta X$ 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, $CCC_{Y_{past} \rightarrow \Delta X}$ 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 S1 of the supplementary material (Section 2). We see that the sign of $CCC_{Y_{past} \rightarrow \Delta X}$ signifies the 'kind of dynamical influence' that $Y_{past}$ has on $\Delta X$, whether this dynamical influence is similar to or different from that of $X_{past}$ on $\Delta X$. When $CCC_{Y_{past} \rightarrow \Delta X}$ is $-ve$, it signifies that $Y_{past}$ has a different dynamical influence on $\Delta X$ than $X_{past}$. On the contrary, when $CCC_{Y_{past} \rightarrow \Delta X}$ is $+ve$, it signifies that $Y_{past}$ has a dynamical influence on $\Delta X$ that is similar to that of $X_{past}$. On estimating the averaged $CCC$ from time series $Y$ to $X$, expecting that $CCC_{Y_{past} \rightarrow \Delta X}$ 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 coupled tent maps were generated as per Eq. 17, 18. 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,

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

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

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 and linearly coupled tent maps tent maps are shown in Figures 1 and 2 respectively. Results for non-linear coupling of tent maps are similar to that for linear coupling and are included in the supplementary material (Figure S10, Section 4.1). The values displayed represent the mean over 50 trials. As the degree of

---

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

coupling is varied for AR processes, there is no clear pattern in KL and JSD values. 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 (both linear and non-linear coupling), 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 latter (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.
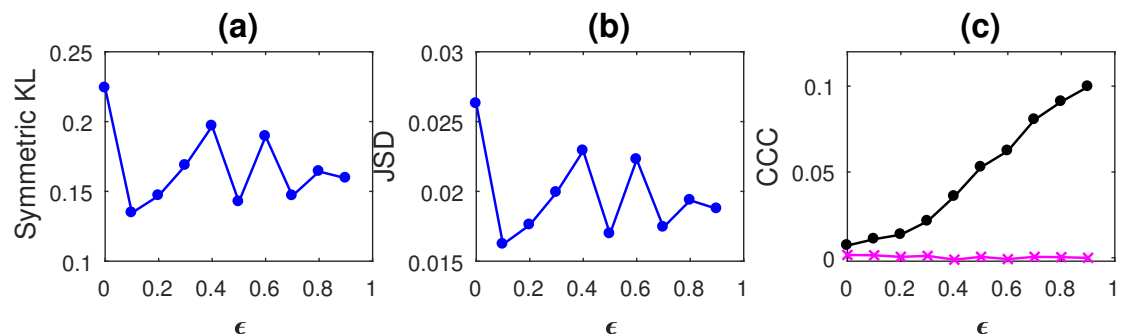
**Figure 1.** (color online). Mean values of divergence between distributions of coupled AR(1) processes using Symmetric Kullback-Leibler (KL) (subfigure (a)) and Jensen Shannon (JSD) divergences (in nats) (subfigure (b)), and the mean causality values estimated using CCC from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta), as the degree of coupling, $\varepsilon$ is varied (subfigure (c)). CCC values increase with increasing $\varepsilon$. There is no similarity in the trend of KL/JSD to CCC.
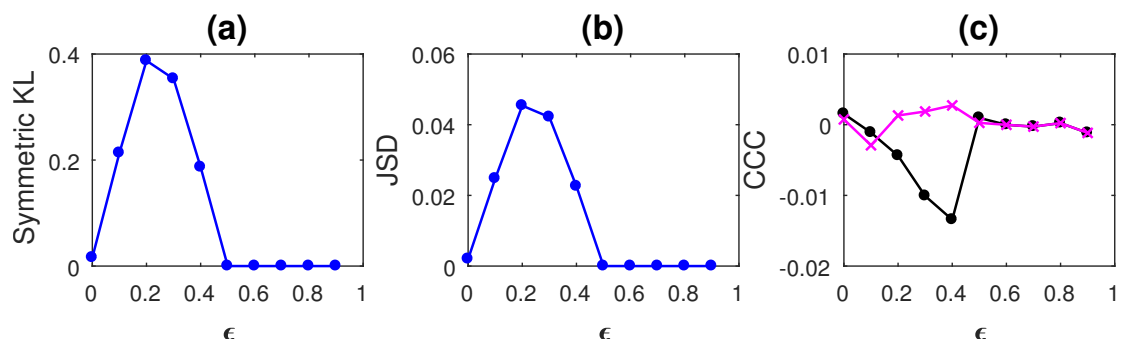
**Figure 2.** (color online). Mean values of divergence between distributions of linearly coupled tent maps using Symmetric Kullback Leibler (KL) (subfigure (a)) and Jensen Shannon (JSD) divergences (in nats) (subfigure (b)), and the mean causality values estimated using CCC from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) (subfigure (c)), as the degree of coupling, $\varepsilon$ is varied. For $\varepsilon < 0.5$, CCC and KL/JSD are highly negatively correlated.

Both positive and negative CCC imply significant causal influence (CCC≈0 implies either no causal

258   influence or identical processes), but the nature of the dynamical influence of the *cause* on the *effect* is very
259   different in these two cases. Causality turning 'negative' does not seem very intuitive at first, but all that it
260   signifies is that the past of the *cause* variable makes the dynamics of the *effect* variable less unpredictable
261   than its (*effect's*) own past. Such a unique feature could be very useful for real world applications in terms
262   of 'controlling' the dynamics of a variable being effected by several variables. If a particular *cause*, out of
263   several causes that makes the caused 'less predictable' and has 'intrinsically different' dynamics from that
264   of the *effect*, needs to be determined and eliminated, it can be readily identified by observing the sign of
265   CCC. Informed attempts to inhibit and enforce certain variables of the system can then be made.

266       As the existing model-free methods of causality can extract only 'associational causality' and ignore
267   the influence that the cause has on dynamics of the caused, it is impossible for them to comment on the
268   nature of this dynamical influence, something that CCC is uniquely able to accomplish. Obviously, model
269   based methods give full-fledged information about 'the kind of dynamical influence' owing to the model
270   equations assumed. However, if there are no equations assumed (or known), then the sign and magnitude
271   of CCC seems to be the best choice to capture the cause-effect relationship with additional information on
272   the similarity (or its lack of) between the two dynamics.

## 5 RESULTS AND DISCUSSION

274   A measure of causality, to be robust for real data, needs to perform well in the presence of noise, filtering,
275   low temporal and amplitude resolution, non-uniformly sampled signals, short length time series as well as
276   presence of other causal variables in the system. In this section, we rigorously simulate these cases and
277   evaluate the performance of CCC measure by comparing with existing measures — Granger Causality
278   (GC) and Transfer Entropy (TE). Owing to space constraints, some of these results are included in Section
279   4 of the supplementary material. In the last sub-section, we test CCC on real-world datasets. In all
280   cases, we take the averaged value of CCC over entire time series as computed by Eq. 8 (or Eq. 11 in
281   the conditional case) and the parameters for CCC estimation are chosen as per the selection criteria
282   and rationale discussed in Section 3 of the supplementary material. GC estimation is done using the
283   MVGC toolbox (Barnett and Seth, 2014) in its default settings and TE estimation is done using MuTE
284   toolbox (Montalto et al., 2014). Akaike Information Criteria is used for model order estimation with the
285   maximum model order set to 20 in the MVGC toolbox, except where specified. Maximum number of lags
286   to take for autocorrelation computation is done automatically by the toolbox. In the MuTE toolbox, the
287   approach of Non Uniform Embedding for representation of the history of the observed processes and of
288   Nearest Neighbor estimator for estimating the probability density functions is used for all results in this
289   paper. The number of lags to consider for observed processes was set to 5 and the maximum number of
290   nearest neighbors to consider was set to 10.

### 5.1 Varying unidirectional coupling

#### 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.

$$X(t) = aX(t-1) + \varepsilon Y(t-1) + \varepsilon_{X,t}$$
$$Y(t) = bY(t-1) + \varepsilon_{Y,t},$$
(15)

293   where $a = 0.9$, $b = 0.8$, $t = 1$ to 1000s, sampling period = 1s. $\varepsilon$ is varied from $0 - 0.9$ in steps of 0.1.
294   Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where $\nu$ = noise intensity = 0.03 and $\eta$ follows standard normal distribution.
295   Figure 3 shows the performance of CCC along with that of TE and GC as mean values over 50 trials,
296   (CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$). Standard deviation of CCC, TE and GC values are
297   shown in Figure 4.

298       With increasing coupling, the causality estimated by CCC, TE as well as GC increases.

#### 5.1.2 AR(100)

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

$$X(t) = aX(t-1) + \varepsilon Y(t-100) + \varepsilon_{X,t}$$
$$Y(t) = bY(t-1) + \varepsilon_{Y,t},$$
(16)
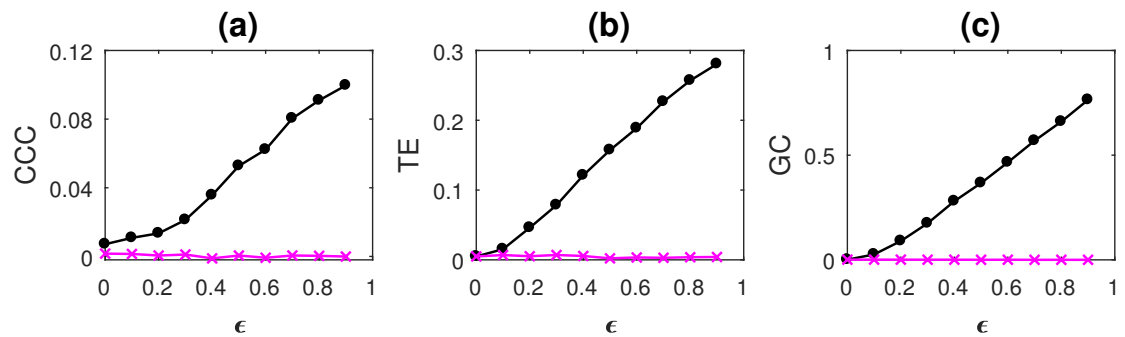
**Figure 3.** (color online). Mean causality values estimated using CCC (a), TE (b) and GC (c) for coupled AR(1) processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling, $\varepsilon$ is varied. CCC, TE as well as GC are able to correctly quantify causality.
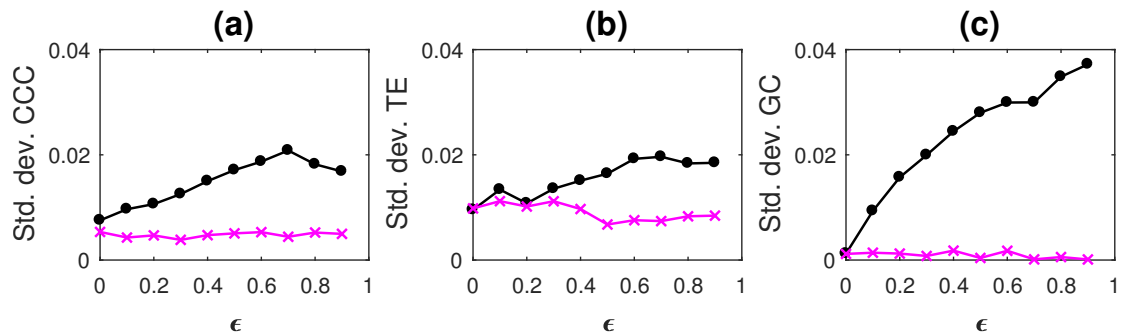
**Figure 4.** (color online). Standard deviation of causality values estimated using CCC (a), TE (b) and GC (c) for coupled AR(1) processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling, $\varepsilon$ is varied.

300  where $a = 0.9$, $b = 0.8$, $t = 1$ to 1000s, sampling period = 1s. $\varepsilon$ is varied from $0 - 0.9$ in steps of 0.1.
301  Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where $\nu$ = noise intensity = 0.03 and $\eta$ follows standard normal distribution.
302  Figure 5 shows the performance of CCC along with that of TE and GC, as mean values over 50 trials
303  (CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$). Maximum model order was set to 110 in the MVGC
304  toolbox.
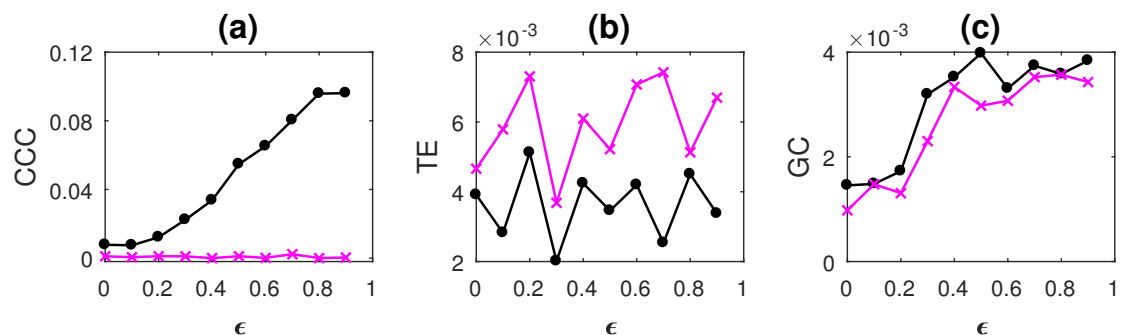
**Figure 5.** (color online). Mean causality values estimated using CCC (a), TE (b) and GC (c) for coupled AR(100) processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the degree of coupling, $\varepsilon$ is varied. Only CCC is able to reliably estimate the correct causal relationship for all values of $\varepsilon$ while TE and GC fail.

305  CCC values increase steadily with increasing coupling for the correct direction of causation. TE fails
306  as it shows higher causality from $X$ to $Y$ for all $\varepsilon$. GC also shows confounding of causality values in two
307  directions. Thus, causality in coupled AR processes with long-range memory can be reliably estimated

308 using CCC and not using TE or GC. Range of standard deviation of CCC values from $Y$ to $X$ is 0.0076 to
309 0.0221 for varying parameter $\varepsilon$ and that from $X$ to $Y$ is 0.0039 to 0.0053. These values are much smaller
310 than the mean CCC estimates and thus, causality estimated in the direction of causation and opposite to it
311 remain well separable. For TE, $Y$ to $X$, standard deviation range is 0.0061 to 0.0090 and $X$ to $Y$, standard
312 deviation range is 0.0082 to 0.0118. For GC, $Y$ to $X$, standard deviation range is 0.0012 to 0.0033 and $X$
313 to $Y$, standard deviation range is 0.0015 to 0.0034.

### 314 *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:

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

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

$$
\begin{aligned}
X(t) &= \varepsilon Y(t) + (1-\varepsilon)h(t), \\
h(t) &= 2X(t-1), & 0 \leq X(t-1) < 1/2, \\
h(t) &= 2 - 2X(t-1), & 1/2 \leq X(t-1) \leq 1,
\end{aligned}
\tag{18}
$$

315 where $\varepsilon$ is the degree of linear coupling.
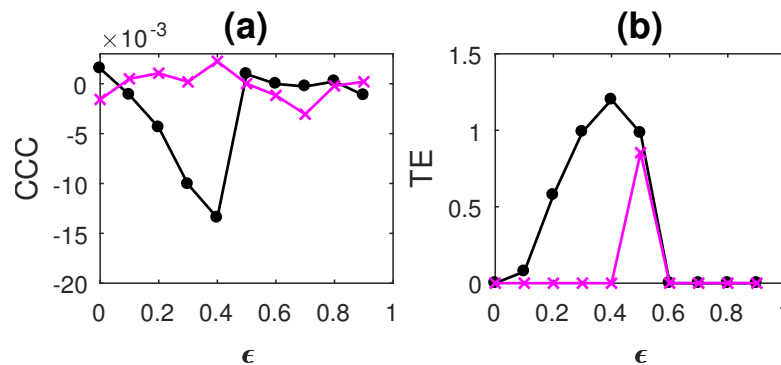
**Figure 6.** (color online). Mean of causality values estimated using CCC (a) and TE (b) for linearly
coupled tent maps, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the
degree of coupling is increased. With increasing coupling (until synchronization), magnitude of CCC and
TE values increases. CCC values are negative while TE are positive.

316 The length of the signals simulated in this case was 3000, i.e. $t = 1$ to 3000s, sampling period = 1s
317 and the first 2000 transients were removed to yield 1000 points for causality estimation. Figure 6 shows
318 the performance of CCC and TE for linearly coupled tent maps as $\varepsilon$ is varied (CCC settings: $L = 100$,
319 $w = 15$, $\delta = 80$, $B = 8$). CCC and TE comparison was also done for increasing coupling in the case of
320 non-linearly coupled tent maps. These results are included in the supplementary material, Section 4.1.
321 Results obtained are similar to the linear-coupling case. The assumption of a linear model for estimation
322 of GC was proved to be erroneous for most trials and hence GC values are not displayed. As $\varepsilon$ is increased
323 for both linear and non-linear coupling, $TE_{Y \rightarrow X}$ increases in the positive direction and then falls to zero
324 when the two series become completely synchronized at $\varepsilon = 0.5$. The trend of the magnitude of CCC
325 values is similar to TE, however, $CCC_{Y \rightarrow X}$ increment is in negative direction. This is because of the fact
326 that with increasing coupling the kind of dynamical influence from $Y$ to $X$ becomes increasingly different
327 than the dynamical influence from the past values of $X$ to itself.

328 In case of linear coupling, range of standard deviation of CCC values from $Y$ to $X$ is 0.0050 to 0.0087
329 for different values of $\varepsilon$ and that from $X$ to $Y$ is 0.0051 to 0.0100. For TE, $Y$ to $X$, standard deviation
330 range is 0 to 1.4851 and $X$ to $Y$, standard deviation range is 0 to 1.4225. For non-linear coupling, the
331 range of standard deviation values are included in Section 4.1 of the supplementary material.

332 For both CCC and TE, standard deviation values obtained indicate that there might be confounding in
333 the causality values in the direction of causation and the direction opposite to causation for low values of
334 $\varepsilon$.

### 5.2 Varying process noise

The performance of measures as process noise is varied is shown in Figure 7 for coupled AR processes simulated as in Eq. 15, where $a = 0.9$, $b = 0.8$, $\varepsilon = 0.8$, $t = 1$ to 1000s, sampling period = 1s, number of trials = 50. Noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where $\nu$ = noise intensity is varied from 0.01 to 0.1 and $\eta$ follows standard normal distribution. CCC settings: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$. The range of standard deviation of CCC values from $Y$ to $X$ is 0.0162 to 0.0223 for different values of $\varepsilon$ and that from $X$ to $Y$ is 0.0038 to 0.0058. For TE, $Y$ to $X$, standard deviation range is 0.0182 to 0.0267 and $X$ to $Y$, standard deviation range is 0.0063 to 0.0104. For GC, $Y$ to $X$, standard deviation range is 0.0314 to 0.0569 and $X$ to $Y$, standard deviation range is 0.0001 to 0.0002.
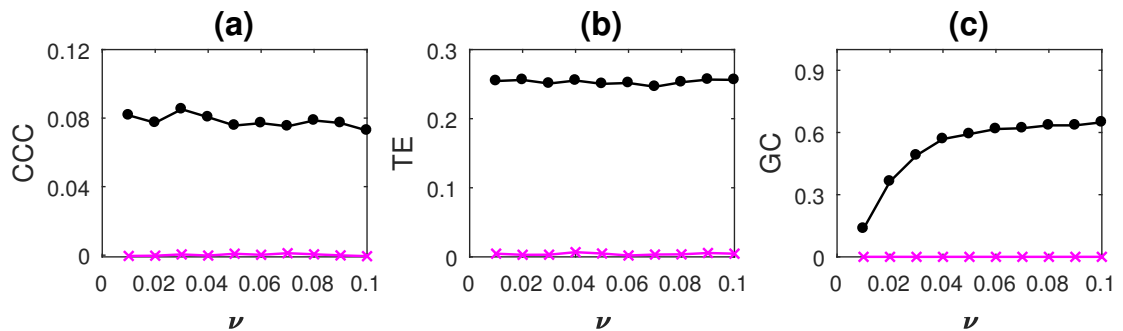
**Figure 7.** (color online). Mean causality values estimated using CCC (a), TE (b) and GC (c) for coupled AR processes, from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the intensity of noise, $\nu$ is varied. All the three measures perform well in this case.

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

### 5.3 Non uniform sampling

Results for cauality testing on uniformly downsampled signals are included in the Section 4.2 of the supplementary material. 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 Vỳrost, 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 ($\alpha$) 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 $\nu$ = noise intensity = 0.03 and $\eta$ follows standard normal distribution. Length of original time series, $N = 2000$, and is reduced upon increasing the percentage non-uniform sampling $\alpha$. 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 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 $\alpha$, while $\nu = 0.03$, are shown in Figure 8.

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 $\alpha$. 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 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 $\alpha$, while $\nu = 0.03$, are shown in Figure 9.

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 be employed in such scenarios.
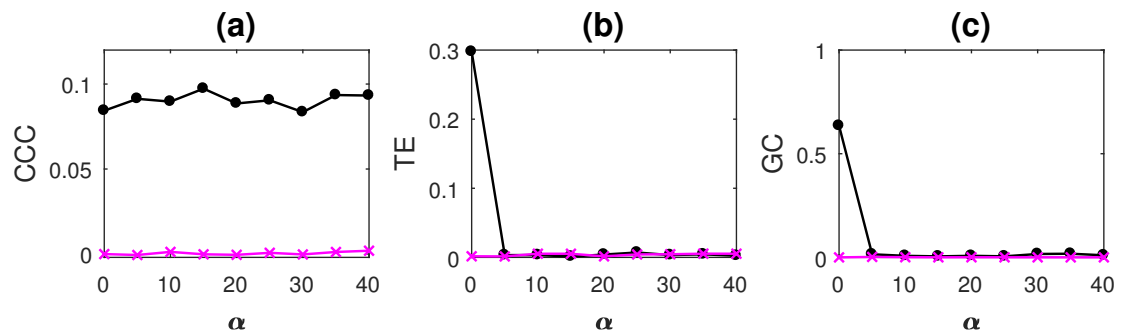
**Figure 8.** (color online). Mean causality values estimated using CCC (a), TE (b) and GC (c) for coupled AR processes from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the percentage of non-uniform sampling $\alpha$ is varied. CCC is the only measure that shows reliable, consistent and correct estimates of causality.
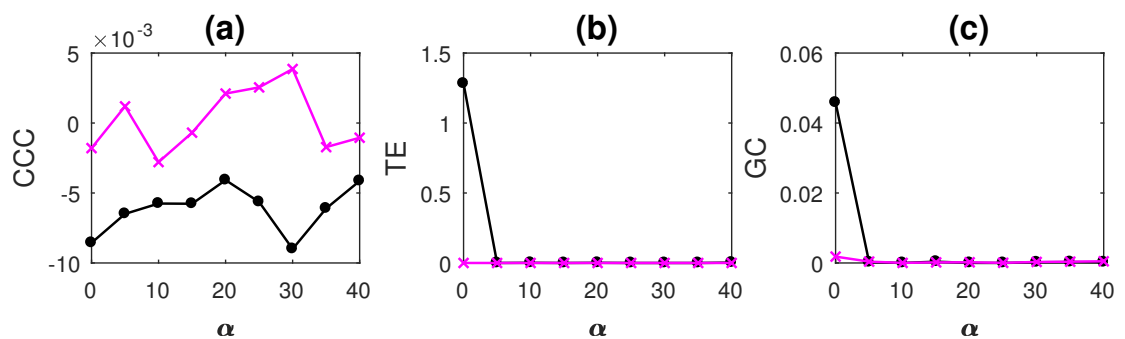
**Figure 9.** (color online). Mean causality values estimated using CCC (a), TE (b) and GC (c) for coupled tent maps from Y to X (solid line-circles, black) and X to Y (solid line-crosses, magenta) as the percentage of non-uniform sampling is varied. CCC is able to distinguish the causality direction but the separation between values is small. TE and GC fail completely fail.

## 5.4 Filtering of Coupled Signals

Acquired data preprocessing often involves low pass filtering to smooth out the signal (Teplan, 2002). At other times, high pass filtering is required to remove low frequency glitches from a high frequency signal. Also when the signals acquired are sampled at low frequencies, the effects due to decimation and filtering may add up and result in poorer estimates of causality. This is often the case in fMRI signals (Glover, 2011; Kim et al., 1997).

To test these scenarios, AR processes were simulated as below:

$$Y(t) = 0.7Y(t-5) + \varepsilon_{Y,t},$$
$$X(t) = 0.9X(t-5) + 0.8Y(t-1) + \varepsilon_{X,t},$$
(19)

where, noise terms, $\varepsilon_Y, \varepsilon_X = \nu\eta$, where $\nu$ = noise intensity = 0.03 and $\eta$ follows standard normal distribution.

Causality values were estimated using CCC, TE and GC when simulated signals are low pass filtered using a moving average window of length 3 with step size 1. The results are shown in Table 1 as mean values over 10 trials. CCC settings used: $L = 150$, $w = 15$, $\delta = 80$, $B = 2$. The performance of the measures when coupled signals are decimated to half the sampling rate and then low pass filtered are also included in the table. The length of the original signal simulated is 2000 and is reduced to 1998 upon filtering and to 998 upon filtering and decimation.

From the table, we see that CCC can distinguish the direction of causality in the original case as well as in the filtering and decimation plus filtering case. Erroneously, TE shows significant causality in the direction opposite to causation upon filtering as well as upon decimation and filtering and GC shows significant causality in the direction opposite to causation upon decimation and filtering. By this we can

**Table 1.** Mean 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 | |
|---|---|---|---|---|---|---|
| | $Y \to X$ | $X \to Y$ | $Y \to X$ | $X \to Y$ | $Y \to X$ | $X \to Y$ |
| 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 |

**(a)** Connectivity

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

**(b)** CCC

| To \ From | X | Y | Z |
|---|---|---|---|
| X | 0 | 0.0615 | 0.0023 |
| Y | 0.1122 | 0 | 0.0022 |
| Z | 0.0055 | 0.0205 | 0 |

**(c)** TE

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

**(d)** GC
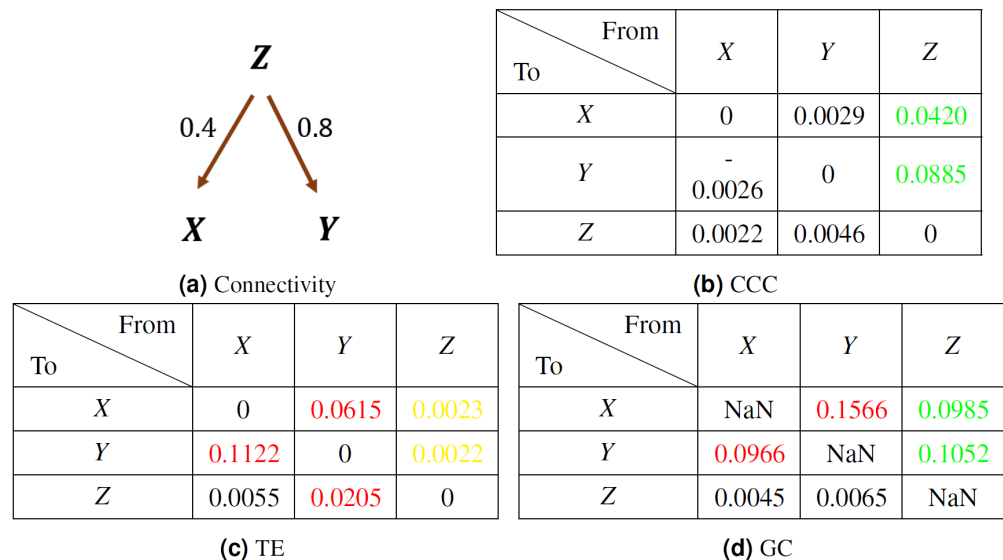
**Figure 10.** Mean causality values estimated using CCC (b), TE (c) and GC (d) for a system of three AR variables coupled as in (a). True positives are in green, true negatives in black, false positives in red and false negatives in yellow.

infer that CCC is highly suitable for practical applications which involve pre-processing such as filtering and decimation of measurements.

### 5.5 Conditional CCC on short length MVAR system

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

$$Z(t) = 0.8Z(t-1) + \varepsilon_{Z,t},$$
$$X(t) = 0.9X(t-1) + 0.4Z(t-100) + \varepsilon_{X,t},$$
$$Y(t) = 0.9Y(t-1) + 0.8Z(t-100) + \varepsilon_{Y,t},$$

(20)

where the noise terms, $\varepsilon_Z, \varepsilon_X, \varepsilon_Y = \nu\eta$, $\nu$ = noise intensity = 0.03 and $\eta$ follows standard normal distribution. Length of time series simulated was 300 and first 50 transients were removed to yield short length signals of 250 time points.

    The coupling direction and strength between variables $X, Y, Z$ are shown in Figure 10(a). The mean values of causality estimated over 10 trials using CCC, TE and GC are shown in Figure 10 tables, (b), (c) and (d) respectively. CCC settings used: $L = 150$, $w = 15$, $\delta = 20$, $B = 2$. In the tables, true positives are in green, true negatives in black, false positives in red and false negatives in yellow. CCC detects correctly the true positives and negatives. GC, detects the true positives but also shows some false positive couplings. TE, performs very poorly, falsely detecting negatives where coupling is present and also showing false positives where there is no coupling.

**Figure 11.** CCC, TE on real-world time series. (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). (c): Table showing CCC and TE values as estimated for systems (a) and (b).

### 5.6 Real Data

CCC was applied to estimate causality on measurements from two real-world systems and compared with TE. System (a) comprised of short time series for dynamics of a complex ecosystem, with 71 point 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 11(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 11). 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. 11), 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.

- 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 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 5 of supplementary material for details.

## LIST OF ABBREVIATIONS

| Abbreviation | Full form | Abbreviation | Full form |
|---|---|---|---|
| AR | Autoregressive | MVAR | Multivariate Autoregressive |
| $C(\cdot)$ | Complexity | $C(\cdot,\cdot)$ | Joint Complexity |
| $CC$ | Dynamical Compression-Complexity | $\overline{CC}$ | Averaged Dynamical Compression-Complexity |
| $CCC$ | Compression-Complexity Causality | $\overline{CCC}$ | Averaged Compression-Complexity Causality |
| $CR$ | Complexity Rate | $DC$ | Dynamical Compression-Complexity |
| $ETC(\cdot)$ | Effort-to-Compress | $ETC(\cdot,\cdot)$ | Joint Effort-to-Compress |
| GC | Granger Causality | $ICC$ | Interventional Complexity Causality |
| JSD | Jensen-Shannon Divergence | KL | Kullback-Leibler Divergence |
| LZ | Lempel-Ziv Complexity | TE | Transfer Entropy |

## ACKNOWLEDGMENTS

Aditi Kathpalia is thankful to Manipal Academy of Higher Education for permitting this research as part of the PhD programme.

## 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.

Barrett, A. B. and Barnett, L. (2013). Granger causality is designed to measure effect, not mechanism. *Frontiers in neuroinformatics*, 7:6.

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.

Budhathoki, K. and Vreeken, J. (2016). Causal inference by compression. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 41–50. IEEE.

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. (2007). Statistical inference through data compression.

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

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.

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

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

505   Jost, C. and Ellner, S. P. (2000). Testing for predator dependence in predator-prey dynamics: a
506     non-parametric approach. *Proceedings of the Royal Society of London B: Biological Sciences*,
507     267(1453):1611–1620.

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

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

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

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

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

519   Mosedale, T. J., Stephenson, D. B., Collins, M., and Mills, T. C. (2006). Granger causality of coupled
520     climate processes: Ocean feedback on the north atlantic oscillation. *Journal of climate*, 19(7):1182–
521     1194.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

546   Virmani, M. and Nagaraj, N. (2019). A novel perturbation based compression complexity measure for
547     networks. *Heliyon*, 5(2):e01181.

548   Wieczorek, A. and Roth, V. (2016). Causal compression. *arXiv preprint arXiv:1611.00261*.

549   Wiener, N. (1956). The theory of prediction. *Modern mathematics for engineers*, 1:125–139.