Supplementary Material

Decomposing causality into its synergistic, unique, and redundant components

Álvaro Martínez-Sánchez[1], Gonzalo Arranz[1], and Adrián Lozano-Durán[1,2]

[1]*Department of Aeronautics and Astronautics, Massachusetts Institute of Technology Cambridge, MA 02139, USA*

[2]*Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA*

Contents

S1 Method formulation

S1.1 Fundamentals of information theory

Consider N quantities of interest at time t represented by the vector of observable variables $\boldsymbol{Q} = [Q_1(t), Q_2(t), \ldots, Q_N(t)]$. We treat $\boldsymbol{Q}$ as a random variable and consider a finite partition of the observable phase space $D = \{D_1, D_2, \ldots, D_{N_D}\}$, where N_D is the number of partitions, such that $D = \cup_{i=1}^{N_D} D_i$ and $D_i \cap D_j = \emptyset$ for all $i \neq j$ (i.e., non-overlapping partitions that cover all the space D). We use upper case Q to denote the random variable itself; and lower case q to denote a particular state contained in one D_i (also referred to as a value or an event) of Q. The probability of finding the system at state D_i at time t is $p(\boldsymbol{Q}(t) \in D_i)$, that in general depends on the partition D. For simplicity, we refer to the latter probability as $p(\boldsymbol{q})$.

The information contained in the variable $\boldsymbol{Q}$ is given by [44]:

$$H(\boldsymbol{Q}) = \sum_{\boldsymbol{q}} -p(\boldsymbol{q}) \log_2[p(\boldsymbol{q})] \geq 0, \tag{S1}$$

where the summation is over all the states of $\boldsymbol{Q}$. The quantity H is referred to as the Shannon information or entropy [44]. The units of H are set by the base chosen, in this case 'bits' for base 2. For example, consider a fair coin with $Q \in \{\text{heads}, \text{tails}\}$ such that $p(\text{heads}) = p(\text{tails}) = 0.5$. The information of the system 'tossing a fair coin n times' is $H = -\sum 0.5^n \log_2(0.5^n) = n$ bits, where the summation is carried out across all possible outcomes (namely, 2^n). If the coin is completely biased towards heads, $p(\text{heads}) = 1$, then $H = 0$ bits (taking $0 \log 0 = 0$), i.e., no information is gained as the outcome was already known before tossing the coin. The Shannon information can also be interpreted in terms of uncertainty: $H(\boldsymbol{Q})$ is the average number of bits required to unambiguously determine $\boldsymbol{Q}$. H is maximum when all the possible outcomes are equiprobable (indicating a high level of uncertainty in the state of the system) and zero when the process is completely deterministic (indicating no uncertainty in the outcome).

The Shannon information of $\boldsymbol{Q}$ conditioned on another variable $\boldsymbol{Q}'$ is defined as:

$$H(\boldsymbol{Q}|\boldsymbol{Q}') = \sum_{\boldsymbol{q},\boldsymbol{q}'} -p(\boldsymbol{q}, \boldsymbol{q}') \log_2[p(\boldsymbol{q}|\boldsymbol{q}')]. \tag{S2}$$

where $p(\boldsymbol{q}|\boldsymbol{q}') = p(\boldsymbol{q}, \boldsymbol{q}')/p(\boldsymbol{q}')$ with $p(\boldsymbol{q}') \neq 0$ is the conditional probability distribution, and $p(\boldsymbol{q}') = \sum_{\boldsymbol{q}} p(\boldsymbol{q}, \boldsymbol{q}')$ is the marginal probability distribution of $\boldsymbol{q}'$. It is useful to interpret $H(\boldsymbol{Q}|\boldsymbol{Q}')$ as the uncertainty in the variable $\boldsymbol{Q}$ after conducting the 'measurement' of $\boldsymbol{Q}'$. If $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ are independent random variables, then $H(\boldsymbol{Q}|\boldsymbol{Q}') = H(\boldsymbol{Q})$, i.e., knowing $\boldsymbol{Q}'$ does not reduce the uncertainty in $\boldsymbol{Q}$. Conversely, $H(\boldsymbol{Q}|\boldsymbol{Q}') = 0$ if knowing $\boldsymbol{Q}'$ implies that $\boldsymbol{Q}$ is completely determined. Finally, the mutual information between the random variables $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ is

$$I(\boldsymbol{Q}; \boldsymbol{Q}') = H(\boldsymbol{Q}) - H(\boldsymbol{Q}|\boldsymbol{Q}') = H(\boldsymbol{Q}') - H(\boldsymbol{Q}'|\boldsymbol{Q}), \tag{S3}$$

which is a symmetric measure $I(\boldsymbol{Q}; \boldsymbol{Q}') = I(\boldsymbol{Q}'; \boldsymbol{Q})$ representing the information shared among the variables $\boldsymbol{Q}$ and $\boldsymbol{Q}'$. Figure S1 depicts the relationship between the Shannon information, conditional Shannon information, and mutual information.

The definitions above can be extended to continuous random variables by replacing summation by integration and the probability mass functions by probability density functions:

$$H_c(\boldsymbol{Q}) = \int_{\boldsymbol{Q}} -\rho(\boldsymbol{q}) \log_2[\rho(\boldsymbol{q})] \, \mathrm{d}\boldsymbol{q}, \tag{S4a}$$

$$H_c(\boldsymbol{Q}|\boldsymbol{Q}') = \int_{\boldsymbol{Q},\boldsymbol{Q}'} -\rho(\boldsymbol{q}, \boldsymbol{q}') \log_2[\rho(\boldsymbol{q}|\boldsymbol{q}')] \, \mathrm{d}\boldsymbol{q} \, \mathrm{d}\boldsymbol{q}', \tag{S4b}$$

$$I_c(\boldsymbol{Q}; \boldsymbol{Q}') = H_c(\boldsymbol{Q}) - H_c(\boldsymbol{Q}|\boldsymbol{Q}') = H_c(\boldsymbol{Q}') - H_c(\boldsymbol{Q}'|\boldsymbol{Q}), \tag{S4c}$$

where H_c is referred to as the differential entropy, $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ are now continuous random variables, ρ denotes probability density function, and the integrals are performed over the support set of $\boldsymbol{Q}$ and $\boldsymbol{Q}'$. The differential entropy shares many of the properties of the discrete entropy. However, it can be infinitely large, positive or negative. SURD relies on the use of mutual information, which is non-negative in the continuous case. Additionally, it can be shown that if $\rho(\boldsymbol{q}, \boldsymbol{q}') \log_2[\rho(\boldsymbol{q}, \boldsymbol{q}')]$ is Riemann integrable, then $I(\boldsymbol{Q}^{\Delta}; \boldsymbol{Q}'^{\Delta}) \to I_c(\boldsymbol{Q}; \boldsymbol{Q}')$ for $\Delta \to 0$, where $\boldsymbol{Q}^{\Delta}$ and $\boldsymbol{Q}'^{\Delta}$ are the quantized versions of $\boldsymbol{Q}$ and $\boldsymbol{Q}'$, respectively, defined over a finite partition $D = \cup_{i=1}^{N_D} D_i$ with a characteristic partition size for D_i equal to Δ. In the following section, SURD is presented using discrete mutual information; nevertheless, a similar formulation is applicable to the continuous case.

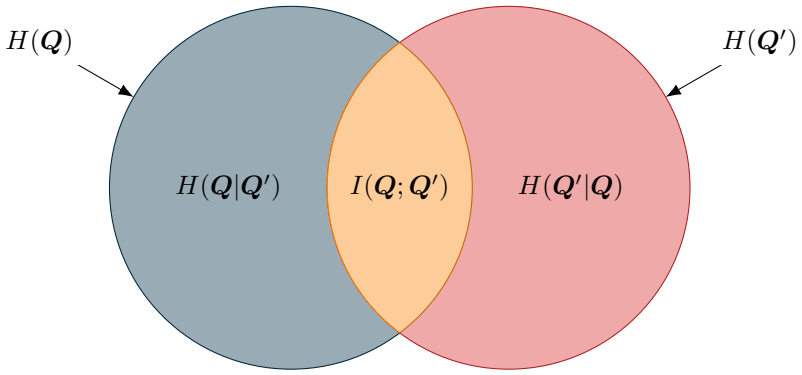


Figure S1: Venn diagram of the Shannon information, conditional Shannon information and mutual information between two random variables $\boldsymbol{Q}$ and $\boldsymbol{Q}'$.

S1.2 Synergistic-Unique-Redundant Decomposition of causality (SURD)

Our objective is to quantify the causality from the components of $\boldsymbol{Q}(t)$ to the future of the variable $Q_j^+ = Q_j(t+\Delta T)$, where Q_j could be one of the components of $\boldsymbol{Q}$ and $\Delta T > 0$ represents an arbitrary time lag. Moreover, for each component of $\boldsymbol{Q}$, the causality is decomposed into *redundant*, *unique*, and *synergistic* contributions to Q_j^+. The theoretical foundation of the method is rooted in the forward propagation of information in dynamical systems: information can only flow toward the future. Let us consider the information in the variable Q_j^+, given by $H(Q_j^+)$. Assuming that all the information in Q_j^+ is determined by the past states of the system, we can write the equation for the forward propagation of information [45]

$$H(Q_j^+) = \Delta I(Q_j^+; \boldsymbol{Q}) + \Delta I_{\text{leak}\to j}, \tag{S5}$$

where $\Delta I(Q_j^+; \boldsymbol{Q})$ is the information flow from $\boldsymbol{Q}$ to Q_j^+, and $\Delta I_{\text{leak}\to j}$ is the causality *leak*, representing the causality from unobserved variables that influence the dynamics of Q_j^+ but are not part of $\boldsymbol{Q}$. The causality leak can be expressed in closed form as a function of the observed variables:

$$\Delta I_{\text{leak}\to j} = H(Q_j^+|\boldsymbol{Q}), \tag{S6}$$

that is the uncertainty in Q_j^+ given the information in $\boldsymbol{Q}$. The amount of available information about Q_j^+ given $\boldsymbol{Q}$ is

$$H(Q_j^+) - \Delta I_{\text{leak}\to j} = \Delta I(Q_j^+; \boldsymbol{Q}) = H(Q_j^+) - H(Q_j^+|\boldsymbol{Q}) = I(Q_j^+; \boldsymbol{Q}), \tag{S7}$$

which is the mutual information between Q_j^+ and $\boldsymbol{Q}$,

$$I(Q_j^+; \boldsymbol{Q}) = \sum_{q_j^+, \boldsymbol{q}} p(q_j^+, \boldsymbol{q}) \log_2 \left(\frac{p(q_j^+|\boldsymbol{q})}{p(q_j^+)} \right) = \sum_{q_j^+, \boldsymbol{q}} p(q_j^+, \boldsymbol{q}) \log_2 \left(\frac{p(q_j^+, \boldsymbol{q})}{p(q_j^+)p(\boldsymbol{q})} \right), \tag{S8}$$

or in continuous form

$$I(Q_j^+; \boldsymbol{Q}) = \int_{Q_j^+, \boldsymbol{Q}} \rho(q_j^+, \boldsymbol{q}) \log_2 \left(\frac{\rho(q_j^+|\boldsymbol{q})}{\rho(q_j^+)} \right) \mathrm{d}q_j^+ \, \mathrm{d}\boldsymbol{q} = \int_{Q_j^+, \boldsymbol{Q}} \rho(q_j^+, \boldsymbol{q}) \log_2 \left(\frac{\rho(q_j^+, \boldsymbol{q})}{\rho(q_j^+)p(\boldsymbol{q})} \right) \mathrm{d}q_j^+ \, \mathrm{d}\boldsymbol{q}. \tag{S9}$$

Equation (S8), quantifies the average dissimilarity between $p(q_j^+)$ and $p(q_j^+|\boldsymbol{q})$. In terms of the Kullback-Leibler divergence [65], Equation (S8) measures the dissimilarity between $p(q_j^+, \boldsymbol{q})$ and the distribution obtained under the assumption of independence between Q_j^+ and $\boldsymbol{Q}$, viz. $p(q_j^+)p(\boldsymbol{q})$. Hence, SURD quantifies the causality from all the components of $\boldsymbol{Q}$ to Q_j^+ by examining how the probability of Q_j^+ changes when accounting for $\boldsymbol{Q}$. Figure S2 provides an interpretation of the quantification of causality based on Equation (S8).

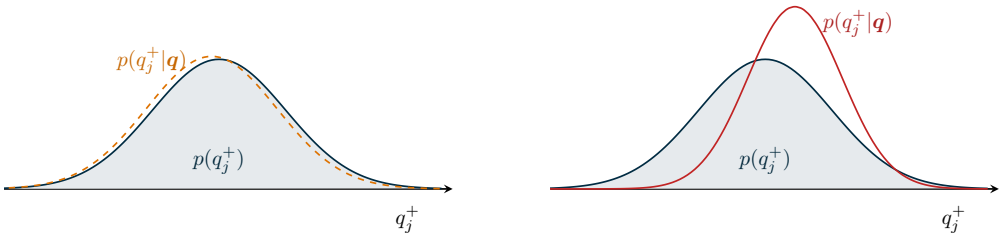


Figure S2: Dissimilarity between $p(q_j^+)$ and $p(q_j^+|\boldsymbol{q})$ contributing to $I(Q_j^+;\boldsymbol{Q})$. Examples of (a) $p(q_j^+|\boldsymbol{q})$ resembling $p(q_j^+)$, which barely contributes to $I(Q_j^+;\boldsymbol{Q})$; and (b) $p(q_j^+|\boldsymbol{q})$ different from $p(q_j^+)$, which increases the value of $I(Q_j^+;\boldsymbol{Q})$. Causality from $\boldsymbol{Q}$ to Q_j^+ is quantified by the expectation of $\log_2[p(q_j^+|\boldsymbol{q})/p(q_j^+)]$.

The next step involves decomposing $I(Q_j^+;\boldsymbol{Q})$ into its unique, redundant, and synergistic components as

$$I(Q_j^+;\boldsymbol{Q}) \equiv \sum_{i=1}^{N} \Delta I^U_{i\to j} + \sum_{\boldsymbol{i}\in\mathcal{C}} \Delta I^R_{\boldsymbol{i}\to j} + \sum_{\boldsymbol{i}\in\mathcal{C}} \Delta I^S_{\boldsymbol{i}\to j}, \tag{S10}$$

where $\Delta I^U_{i\to j}$ is the unique causality from Q_i to Q_j^+, $\Delta I^R_{\boldsymbol{i}\to j}$ is the redundant causality among the variables in $\boldsymbol{Q_i}$ with $\boldsymbol{i} = [i_1, i_2, \ldots]$ being a collection of indices, $\Delta I^S_{\boldsymbol{i}\to j}$ is the synergistic causality from the variables in $\boldsymbol{Q_i}$, and $\mathcal{C}$ is the set of all the combinations of numbers from 1 to N with more than one element and less than or equal to N elements. For example, Equation (S10) can be expanded for $N = 4$ as

$$\begin{aligned}
I(Q_j^+;\boldsymbol{Q}) \equiv\ & \Delta I^U_{1\to j} + \Delta I^U_{2\to j} + \Delta I^U_{3\to j} + \Delta I^U_{4\to j} && \text{(S11a)}\\
& + \Delta I^R_{12\to j} + \Delta I^R_{13\to j} + \Delta I^R_{14\to j} + \Delta I^R_{23\to j} + \Delta I^R_{24\to j} + \Delta I^R_{34\to j} + && \text{(S11b)}\\
& + \Delta I^S_{12\to j} + \Delta I^S_{13\to j} + \Delta I^S_{14\to j} + \Delta I^S_{23\to j} + \Delta I^S_{24\to j} + \Delta I^S_{34\to j} + && \text{(S11c)}\\
& + \Delta I^R_{123\to j} + \Delta I^R_{124\to j} + \Delta I^R_{134\to j} + \Delta I^R_{234\to j} + && \text{(S11d)}\\
& + \Delta I^S_{123\to j} + \Delta I^S_{124\to j} + \Delta I^S_{134\to j} + \Delta I^S_{234\to j} && \text{(S11e)}\\
& + \Delta I^R_{1234\to j} + \Delta I^S_{1234\to j}. && \text{(S11f)}
\end{aligned}$$

An important insight used to define SURD is the realization that the source of causality might change depending on the value of Q_j^+. For example, Q_1 can be only causal to positive values of Q_j^+, whereas Q_2 can be only causal to negative values of Q_j^+. For that reason, we define the specific mutual information [124] from $\boldsymbol{Q_i}$ to a particular event $Q_j^+ = q_j^+$ as

$$\tilde{\imath}(Q_j^+ = q_j^+;\boldsymbol{Q_i}) = \sum_{\boldsymbol{q_i}} p(\boldsymbol{q_i}|q_j^+) \log_2 \left(\frac{p(q_j^+|\boldsymbol{q_i})}{p(q_j^+)} \right) \geq 0. \tag{S12}$$

Note that the specific mutual information is a function of the random variable $\boldsymbol{Q_i}$ (which encompasses all its states) but only a function of one particular state of the target variable (namely, q_j^+). For the sake of simplicity, we will use the notation $\tilde{\imath}_{\boldsymbol{i}}(q_j^+) = \tilde{\imath}(Q_j^+ = q_j^+;\boldsymbol{Q_i})$. Similarly to Equation (S8), the specific mutual information quantifies the dissimilarity between $p(q_j^+)$ and $p(q_j^+|\boldsymbol{q})$ but in this case for the particular state $Q_j^+ = q_j^+$. The mutual information between $Q_j^+$and $\boldsymbol{Q_i}$ is recovered by $I(Q_j^+;\boldsymbol{Q_i}) = \sum_{q_j^+} p(q_j^+)\tilde{\imath}_{\boldsymbol{i}}(q_j^+)$.

We are now in the position of introducing the steps involved in the calculation of redundant, unique, and synergistic causalities (Figure S3). Our definitions are motivated by the following intuition:

- Redundant causality from $\boldsymbol{Q_i} = [Q_{i_1}, Q_{i_2}, \ldots]$ to Q_j^+ is the common causality shared among all the components of $\boldsymbol{Q_i}$, where $\boldsymbol{Q_i}$ is a subset of $\boldsymbol{Q}$ with two or more components.
- Unique causality from Q_i to Q_j^+ is the causality from Q_i that cannot be obtained from any other individual variable Q_k with $k \neq i$.
- Synergistic causality from $\boldsymbol{Q_i} = [Q_{i_1}, Q_{i_2}, \ldots]$ to Q_j^+ is the causality arising from the joint effect of the variables in $\boldsymbol{Q_i}$.

- Redundant and unique causalities must depend only on probability distributions based on Q_i and Q_j^+, that is, $p(q_i, q_j^+)$. On the other hand, synergistic causality must depend on the joint probability distribution of $\boldsymbol{Q_i}$ and Q_j^+, i.e., $p(\boldsymbol{q_i}, q_j^+)$.

For a given value $Q_j^+ = q_j^+$, the specific redundant, unique, and synergistic causalities are calculated as follows:

1. The specific mutual information are computed for all possible combinations of variables in $\boldsymbol{Q}$. This includes specific mutual information of order one $(\tilde{\imath}_1, \tilde{\imath}_2, \ldots)$, order two $(\tilde{\imath}_{12}, \tilde{\imath}_{13}, \ldots)$, order three $(\tilde{\imath}_{123}, \tilde{\imath}_{124}, \ldots)$, and so forth. One example is shown in Figure S3(a).

2. The tuples containing the specific mutual information of order M, denoted by $\tilde{\mathcal{G}}^M$, are constructed for $M = 1, \ldots, N$. The components of each $\tilde{\mathcal{G}}^M$ are organized in ascending order as shown in Figure S3(b).

3. The specific redundant causality is the increment in information gained about q_j^+ that is common to all the components of $\boldsymbol{Q_{j_k}}$ (blue contributions in Figure S3c):

$$\Delta\tilde{\imath}^R_{\boldsymbol{j}_k} = \begin{cases} \tilde{\imath}_{i_k} - \tilde{\imath}_{i_{k-1}}, & \text{for } \tilde{\imath}_{i_k}, \tilde{\imath}_{i_{k-1}} \in \tilde{\mathcal{G}}^1 \text{ and } k \neq n_1 \\ 0, & \text{otherwise,} \end{cases} \tag{S13}$$

   where we take $\tilde{\imath}_{i_0} = 0$, $\boldsymbol{j}_k = [j_{k1}, j_{k2}, \ldots]$ is the vector of indices satisfying $\tilde{\imath}_{j_{kl}} \geq \tilde{\imath}_{i_k}$ for $\tilde{\imath}_{j_{kl}}, \tilde{\imath}_{i_k} \in \tilde{\mathcal{G}}^1$, and n_1 is the number of elements in $\tilde{\mathcal{G}}^1$.

4. The specific unique causality is the increment in information gained by Q_{i_k} about q_j^+ that cannot be obtained by any other individual variable (red contribution in Figure S3c):

$$\Delta\tilde{\imath}^U_{i_k} = \begin{cases} \tilde{\imath}_{i_k} - \tilde{\imath}_{i_{k-1}}, & \text{for } i_k = n_1,\ \tilde{\imath}_{i_k}, \tilde{\imath}_{i_{k-1}} \in \tilde{\mathcal{G}}^1 \\ 0, & \text{otherwise.} \end{cases} \tag{S14}$$

5. The specific synergistic causality is the increment in information gained by the combined effect of all the variables in $\boldsymbol{Q}_{\boldsymbol{i}_k}$ that cannot be gained by other combination of variables $\boldsymbol{Q}_{\boldsymbol{j}_k}$ (yellow contributions in Figure S3c) such that $\tilde{\imath}_{\boldsymbol{j}_k} \leq \tilde{\imath}_{\boldsymbol{i}_k}$ for $\tilde{\imath}_{\boldsymbol{i}_k} \in \tilde{\mathcal{G}}^M$ and $\tilde{\imath}_{\boldsymbol{j}_k} \in \{\tilde{\mathcal{G}}^1, \ldots, \tilde{\mathcal{G}}^M\}$ with $M > 1$ (dotted line in Figure S3c):

$$\Delta\tilde{\imath}^S_{\boldsymbol{i}_k} = \begin{cases} \tilde{\imath}_{\boldsymbol{i}_k} - \tilde{\imath}_{\boldsymbol{i}_{k-1}}, & \text{for } \tilde{\imath}_{\boldsymbol{i}_{k-1}} \geq \max\{\tilde{\mathcal{G}}^{M-1}\}, \text{ and } \tilde{\imath}_{\boldsymbol{i}_k}, \tilde{\imath}_{\boldsymbol{i}_{k-1}} \in \tilde{\mathcal{G}}^M \\ \tilde{\imath}_{\boldsymbol{i}_k} - \max\{\tilde{\mathcal{G}}^{M-1}\}, & \text{for } \tilde{\imath}_{\boldsymbol{i}_k} > \max\{\tilde{\mathcal{G}}^{M-1}\} > \tilde{\imath}_{\boldsymbol{i}_{k-1}}, \text{ and } \tilde{\imath}_{\boldsymbol{i}_k}, \tilde{\imath}_{\boldsymbol{i}_{k-1}} \in \tilde{\mathcal{G}}^M \\ 0, & \text{otherwise.} \end{cases} \tag{S15}$$

6. The specific redundant, unique and synergistic causalities that do not appear in the steps above are set to zero.

7. The steps (1) to (6) are repeated for all the states of Q_j^+ (Figure S3d).

8. Redundant, unique, and synergistic causalities are obtained as the expectation of their corresponding specific values with respect to Q_j^+,

$$\Delta I^R_{\boldsymbol{i}\to j} = \sum_{q_j^+} p(q_j^+) \Delta\tilde{\imath}^R_{\boldsymbol{i}}(q_j^+), \tag{S16a}$$

$$\Delta I^U_{i\to j} = \sum_{q_j^+} p(q_j^+) \Delta\tilde{\imath}^U_{i}(q_j^+), \tag{S16b}$$

$$\Delta I^S_{\boldsymbol{i}\to j} = \sum_{q_j^+} p(q_j^+) \Delta\tilde{\imath}^S_{\boldsymbol{i}}(q_j^+). \tag{S16c}$$

9. Finally, we define the average order of the specific causalities with respect to Q_j^+ as

$$N^\alpha_{\boldsymbol{i}\to j} = \sum_{q_j^+} p(q_j^+) n^\alpha_{\boldsymbol{i}\to j}(q_j^+), \tag{S17}$$

where α denotes R, U, or S, $n^{\alpha}_{\boldsymbol{i}\to j}(q_j^+)$ is the order of appearance of $\Delta\tilde{\imath}^{\alpha}_{\boldsymbol{i}}(q_j^+)$ from left to right as in the example shown in Figure S3. The values of $N^{\alpha}_{\boldsymbol{i}\to j}$ are used to plot $\Delta I^{\alpha}_{\boldsymbol{i}\to j}$ following the expected order of appearance of $\Delta\tilde{\imath}^{\alpha}_{\boldsymbol{i}\to j}$. All the causalities from SURD presented in this work are plotted in order from left to right, following $N^{\alpha}_{\boldsymbol{i}\to j}$.

It is worth noting that the problem of defining redundant, unique, and synergistic causalities can be generally framed as the task of decomposing the mutual information $I(Q_j^+;\boldsymbol{Q})$ into multiple components. The definitions proposed above are motivated by their consistency with the properties presented in the following sections along with the ease of interpretability. Alternative definitions are possible and other decompositions have been suggested in the literature [114, 115, 116, 117, 118, 45, 119]; however, these do not comply with the properties discussed next or result in an unmanageable number of terms. For instance, one of the most referenced decompositions by Williams & Beer [114] results in a number of terms that grows as the Dedekind numbers. In the case of 9 variables, this decomposition yields over 10^{23} terms, whereas SURD produces only 512 terms.

In scenarios with a large number of terms, synergistic and redundant causalities in SURD can be grouped by different orders to facilitate interpretation. For example, for three variables, we can define $\Delta I^S_{\text{2nd}} = \Delta I^S_{12\to1} + \Delta I^S_{23\to1} + \Delta I^S_{13\to1}$, which represents the second-order causal synergy to the target variable 1 (similarly for other orders and redundancies). This is possible in SURD due to the additivity of its causal components. Additionally, if synergistic causalities above a given order are not computed, they are accounted for by the causality leak. In § S3.8, we apply SURD to a system of eight interacting species and calculate synergistic causalities up to the fourth order. The remaining synergistic causalities are considered as causality leaks. This approach effectively manages the challenge of dimensionality by focusing on lower-order synergistic interactions, which are often sufficient for practical analyses.

S1.3 Application of SURD to multiple time lags

The vector of observables can also include variables at multiple time lags:

$$\boldsymbol{Q} = [Q_1(t), Q_1(t-\Delta T_1), \dots, Q_1(t-\Delta T_p), \dots, Q_2(t), Q_2(t-\Delta T_1), \dots, Q_2(t-\Delta T_p), \dots], \tag{S18}$$

where $\Delta T_i > 0$ for $i = 1, \dots, p$. For instance, for $N = 2$ and $p = 1$, the vector is

$$\boldsymbol{Q} = [Q_1(t), Q_1(t-\Delta T), Q_2(t), Q_2(t-\Delta T)]. \tag{S19}$$

For simplicity, we will use the notation $\boldsymbol{Q} = [Q_{1_1}, Q_{1_2}, Q_{2_1}, Q_{2_2}]$, where the first subindex denotes the variable number and the second subindex denotes the present time for $i = 1$ and past times for $i > 1$, e.g., $Q_{1_1} = Q_1(t)$, $Q_{1_2} = Q_1(t-\Delta T)$, $Q_{2_1} = Q_2(t)$, $Q_{2_2} = Q_2(t-\Delta T)$ and so on. The formulation of SURD presented above is equally applicable to the observable $\boldsymbol{Q}$ from Equation (S19). Following the example above for $N = 2$ and $p = 1$, the mutual information between the target variable Q_j^+ and $\boldsymbol{Q}$ is decomposed as:

$$\begin{aligned}
I(Q_j^+;\boldsymbol{Q}) = {} & \Delta I^U_{1_1\to j} + \Delta I^U_{2_1\to j} + \Delta I^U_{1_2\to j} + \Delta I^U_{2_2\to j} && \text{(S20a)}\\
& + \Delta I^R_{1_1 2_1\to j} + \Delta I^R_{1_1 1_2\to j} + \Delta I^R_{1_1 2_2\to j} + \Delta I^R_{2_1 1_2\to j} + \Delta I^R_{2_1 2_2\to j} + \Delta I^R_{1_2 2_2\to j} + && \text{(S20b)}\\
& + \Delta I^S_{1_1 2_1\to j} + \Delta I^S_{1_1 1_2\to j} + \Delta I^S_{1_1 2_2\to j} + \Delta I^S_{2_1 1_2\to j} + \Delta I^S_{2_1 2_2\to j} + \Delta I^S_{1_2 2_2\to j} + && \text{(S20c)}\\
& + \Delta I^R_{1_1 2_1 1_2\to j} + \Delta I^R_{1_1 2_1 2_2\to j} + \Delta I^R_{1_1 1_2 2_2\to j} + \Delta I^R_{2_1 1_2 2_2\to j} + && \text{(S20d)}\\
& + \Delta I^S_{1_1 2_1 1_2\to j} + \Delta I^S_{1_1 2_1 2_2\to j} + \Delta I^S_{1_1 1_2 2_2\to j} + \Delta I^S_{2_1 1_2 2_2\to j} && \text{(S20e)}\\
& + \Delta I^R_{1_1 2_1 1_2 2_2\to j} + \Delta I^S_{1_1 2_1 1_2 2_2\to j}. && \text{(S20f)}
\end{aligned}$$

S1.4 Properties of SURD

- *Non-negativity.* All the terms in Equation (S10) are non-negative by the definition of the redundant, unique and synergistic causalities, and the non-negativity of the specific mutual information [124].

$$\Delta I^U_{i\to j} \geq 0, \text{ for all } i = 1, \dots, N, \ \Delta I^R_{\boldsymbol{i}\to j} \geq 0, \ \Delta I^S_{\boldsymbol{i}\to j} \geq 0, \text{ for all } \boldsymbol{i} \in \mathcal{C}. \tag{S21}$$

- *Reconstruction of individual mutual information.* The mutual information between Q_i and Q_j^+ is equal to the

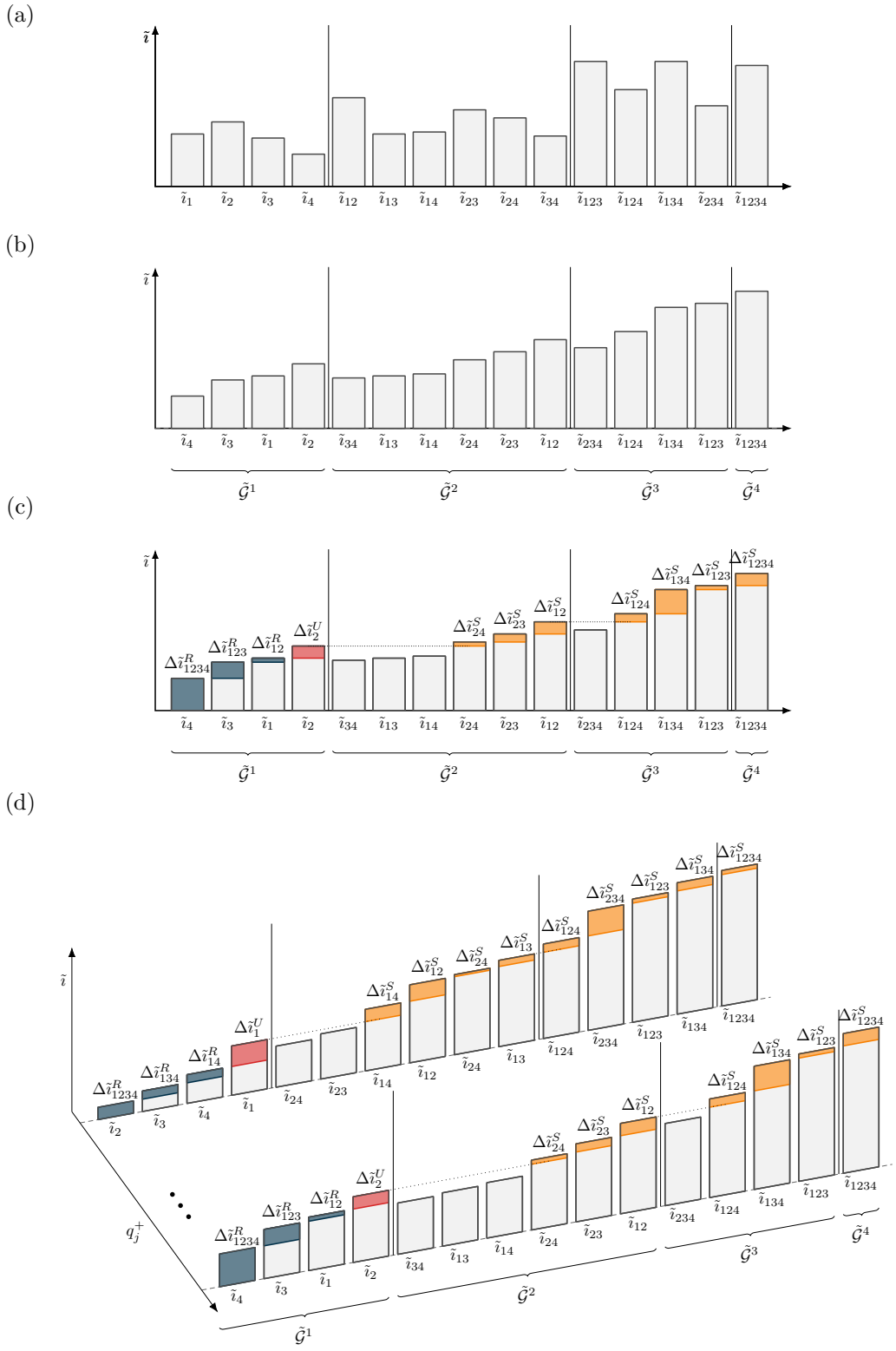


Figure S3: Schematic of the steps involved in the calculation of specific causalities. For a given state $Q_j^+ = q_j^+$, the panels illustrate: (a) all possible specific mutual information values for a collection of four variables; (b) tuples of specific mutual information with the components organized in ascending order; (c) the increments corresponding to specific redundant (blue), unique (red), and synergistic (yellow) causalities; and (d) examples of specific causalities for different states of Q_j^+.

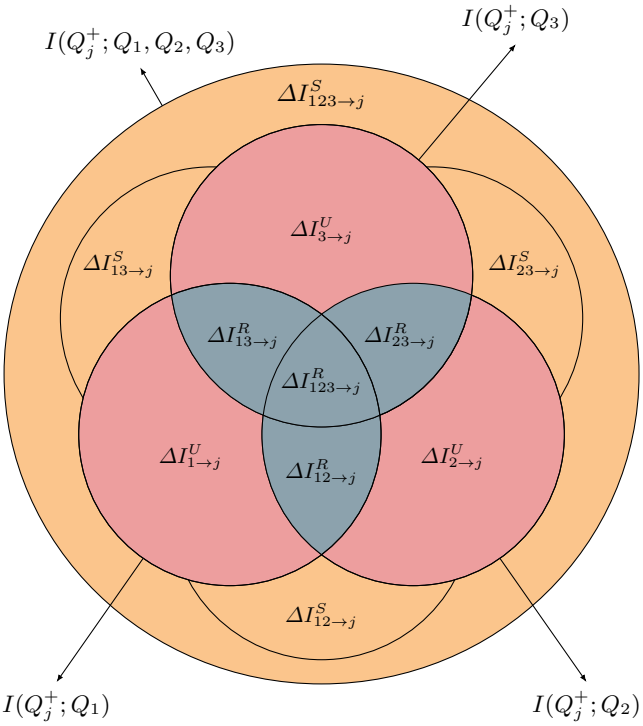


Figure S4: Diagram of the decomposition into redundant, unique, and synergistic causalities and contributions to total and individual mutual information for three observable variables $\boldsymbol{Q} = [Q_1, Q_2, Q_3]$ and the target Q_j^+.

unique and redundant causalities containing Q_i

$$I(Q_i; Q_j^+) = \Delta I^U_{i\to j} + \sum_{\boldsymbol{i}\in\mathcal{C}_i} \Delta I^R_{\boldsymbol{i}\to j}, \tag{S22}$$

where $\mathcal{C}_i$ is the set of the combinations in $\mathcal{C}$ containing the variable Q_i. This condition is consistent with the notion that the information shared between Q_i and Q_j^+ should comprise contributions from unique and redundant causalities only, whereas synergistic causalities should arise through the combined effects of variables. This property, along with the non-negativity and forward propagation of information, enables the construction of the causality diagrams as the example depicted in Figure S4 for three variables. The properties mentioned above are also responsible for avoiding the duplication of causalities within the system.

- *Zero-causality property.* If Q_j^+ is independent of Q_i, then $\Delta I^R_{\boldsymbol{i}\to j} = 0$ for $\boldsymbol{i} \in \mathcal{C}_i$ and $\Delta I^U_{i\to j} = 0$ as long as Q_i is observable.

- *Invariance under invertible transformations.* The redundant, unique, and synergistic causalities are invariant under invertible transformations of $\boldsymbol{Q}$. This property follows from the invariance of the mutual information.

S1.5 Example of SURD in logic gates

We illustrate the concept of redundant, unique, and synergistic causality in three simple examples. The examples represent a system with two inputs Q_1 and Q_2 and one output $Q_3^+ = f(Q_1, Q_2)$. Both input and output are binary variables distributed randomly and independently. The causal description of the system is characterized by the four components:

$$H(Q_3^+) = \Delta I^U_{1\to 3} + \Delta I^U_{2\to 3} + \Delta I^R_{12\to 3} + \Delta I^S_{12\to 3}, \tag{S23}$$

where $\Delta I_{\text{leak}\to 3} = 0$ as $H(Q_3^+|Q_1, Q_2) = 0$. The results for the three cases are summarized in Figure S5.

The first example represents a system in which $Q_2 \equiv Q_1$ (duplicated input) and the output is given by $Q_3^+ = Q_1$. In this case, both Q_1 and Q_2 provide the same information about the output and the only non-zero term in Equation (S23) is the redundant causality $\Delta I^R_{12\to 3} = 1$ bit. In the second example, the output is given by $Q_3^+ = Q_1$

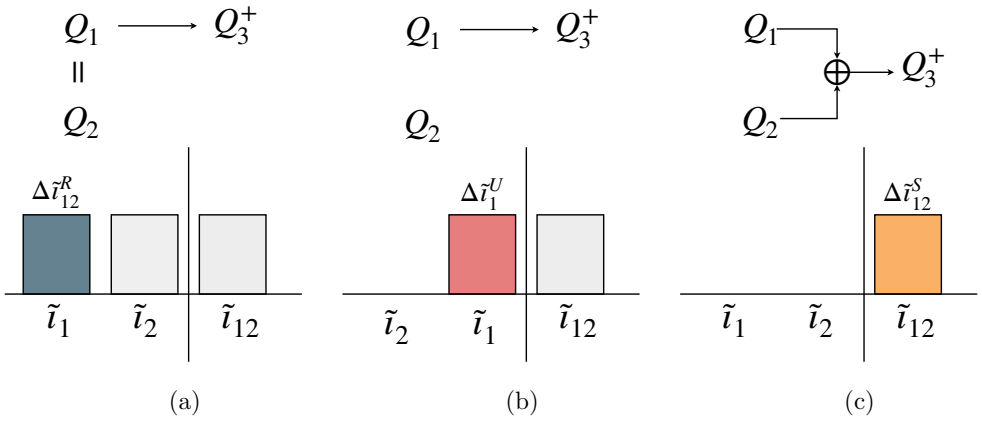


Figure S5: Schematic of logic gates (top panels) and associated specific mutual information (bottom panels) for (a) duplicated input (pure redundant causality), (b) output equal to first input (pure unique causality), and (c) exclusive-OR output (pure synergistic causality). The schematics of the specific mutual information apply to both states $Q_3^+ = 0$ and $Q_3^+ = 1$.

with no dependence on Q_2, which only results in the unique causality $\Delta I^U_{1\to 3} = 1$ bit. In the last example, the output is given by the exclusive-OR operator: $Q_3^+ = Q_1 \oplus Q_2$ such that $Q_3^+ = 1$ if $Q_1 \neq Q_2$ and $Q_3^+ = 0$ otherwise. In this case, the output behaves randomly when observing Q_1 or Q_2 independently. However, the outcome is completely determined when the joint variable $[Q_1, Q_2]$ is considered. Hence, $[Q_1, Q_2]$ contains more information than their individual components and all the causality is synergistic $\Delta I^S_{12\to 3} = 1$ bit.

S2 Other methods for causal inference: description and implementation

In this section, we discuss the methods for causal inference implemented in our study. For each method, we provide a description of the approach, details of the implementation, and its verification in documented cases.

S2.1 Conditional Granger causality (CGC)

Granger causality (GC) [25] is a statistical technique for assessing causality between two time series. This approach relies on the ability of past values of one time series, $Q_2(t)$, to enhance the predictability of another series, $Q_1(t)$. The time signal $Q_2(t)$ is considered to Granger-cause $Q_1(t)$ if the historical data of $Q_2(t)$ significantly improves the forecast of $Q_1(t)$. A foundational principle of Granger causality rests on the assumptions that the causative factor must provide unique information not available in the past values of the effect itself, and that this unique information can be detected via a forecasting model.

GC analysis is typically implemented by means of vector autoregressive (VAR) modeling of the time series. In the bivariate case, Granger causality between Q_1 and Q_2 is assessed by computing the change in the error between two VAR models:

$$Q_1(t) = \hat{a}_0 + \sum_{j=1}^{p} \hat{a}_j Q_1(t - \Delta T_j) + \hat{\varepsilon}(t), \tag{S24}$$

$$Q_1(t) = a_0 + \sum_{j=1}^{p} a_j Q_1(t - \Delta T_j) + \sum_{j=1}^{p} b_j Q_2(t - \Delta T_j) + \varepsilon(t), \tag{S25}$$

where $\hat{a}_j$, a_j and b_j are regression coefficients that represent the influences of the past values of the time series of $Q_1(t)$ and $Q_2(t)$, $\hat{\varepsilon}(t)$ and $\varepsilon(t)$ denote the errors of the models at time t, j is the lag into the past used to predict the values of $Q_1(t)$ and p represents the maximum time lag used. Granger causality from Q_2 to Q_1 is defined as a measure of the extent to which inclusion of Q_2 in the second model (S25) reduces the prediction error of the first model (S24). The standard measure of Granger causality is given by the natural logarithm of the ratio of the residual variance between both models:

$$\mathrm{GC}_{2\to 1} = \log_2 \left(\frac{\mathrm{var}\,(\hat{\varepsilon})}{\mathrm{var}\,(\varepsilon)} \right) \geq 0. \tag{S26}$$

Therefore, if the past values of $Q_2(t)$ improve the prediction of $Q_1(t)$, the residual variance of the second model will be smaller than that of the first model, i.e., $\mathrm{var}\,(\varepsilon) < \mathrm{var}\,(\hat{\varepsilon})$, and the Granger causality measure will be greater than zero, $\mathrm{GC}_{2\to 1} > 0$. Conversely, if the residual variance between both models is exactly the same after introducing Q_2 in the model, i.e., $\mathrm{var}\,(\varepsilon) = \mathrm{var}\,(\hat{\varepsilon})$, the past values of $Q_2(t)$ do not improve the prediction of $Q_1(t)$ and the Granger causality measure will be zero, $\mathrm{GC}_{2\to 1} = 0$.

In this work, we use the extension first proposed by Geweke [26], in which both models use an additional vector of variables, $\boldsymbol{Q}'(t) = [Q_3(t), Q_4(t), \ldots, Q_N(t)]$, which do not contain Q_1 and Q_2. The method is referred to as conditional Granger causality (CGC) and allows us to compute the Granger causality value from Q_2 to Q_1 conditioned on $\boldsymbol{Q}'$ by comparing the errors from:

$$Q_1(t) = \hat{a}_0 + \sum_{j=1}^{p} \hat{a}_j Q_1(t - \Delta T_j) + \sum_{j=1}^{p} \hat{\boldsymbol{c}}_j \boldsymbol{Q}'(t - \Delta T_j) + \hat{\varepsilon}(t), \tag{S27}$$

$$Q_1(t) = a_0 + \sum_{j=1}^{p} a_j Q_1(t - \Delta T_j) + \sum_{j=1}^{p} b_j Q_2(t - \Delta T_j) + \sum_{j=1}^{p} \boldsymbol{c}_j \boldsymbol{Q}'(t - \Delta T_j) + \varepsilon(t), \tag{S28}$$

where $\hat{\boldsymbol{c}}_j$ and $\boldsymbol{c}_j$ are the regression coefficient matrices associated with $\boldsymbol{Q}'(t - \Delta T_j)$. Causality is then computed as:

$$\mathrm{CGC}_{2\to 1} = \log_2 \left(\frac{\mathrm{var}\,(\hat{\varepsilon})}{\mathrm{var}\,(\varepsilon)} \right) \geq 0. \tag{S29}$$

The CGC formulation has been extensively investigated and further developed in a multivariate setting–the Multivariate Granger causality (MVGC) [29].

We have validated our implementation of CGC with the MVGC toolbox [28]. Here, we show as an example the

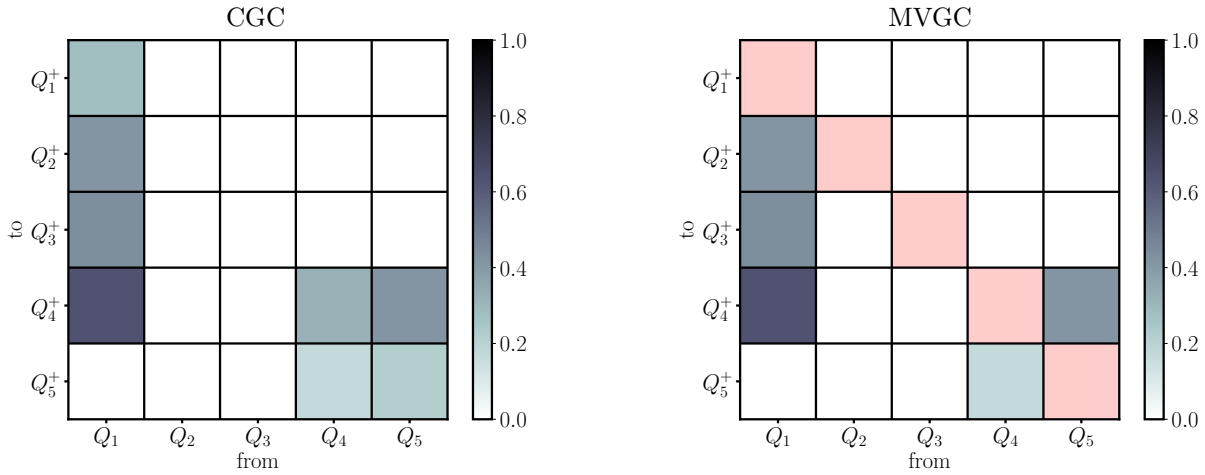


Figure S6: Comparison of current implementation of conditional Granger causality (CGC) and multivariate Granger causality toolbox (MVGC) [28] for a five-node oscillatory network. The MVGC does not provide self-causality values and the diagonal is colored in light pink. Note that in this case the results from CGC and MVGC have not been normalised, as compared with the results presented in the main text.

five-node oscillatory network given by the following equations [28]:

$$\begin{aligned}
Q_1(n+1) &= \left(0.95\sqrt{2} - 0.9025\right) Q_1(n) + \eta_1(n) \\
Q_2(n+1) &= 0.5Q_1(n) + \eta_2(n) \\
Q_3(n+1) &= -0.4Q_1(n) + \eta_3(n) \\
Q_4(n+1) &= -0.5Q_1(n) + 0.25\sqrt{2}Q_4(n) + 0.25\sqrt{2}Q_5(n) + \eta_4(n) \\
Q_5(n+1) &= -0.25\sqrt{2}Q_4(n) + 0.25\sqrt{2}Q_5(n) + \eta_5(n),
\end{aligned} \tag{S30}$$

where $\eta_i(t)$ for $i = 1, 2, 3, 4, 5$ represents Gaussian white noise processes with variances $\sigma_i^2 = [0.6, 0.5, 0.3, 0.3, 0.6]$. The results, shown in Figure S6, indicate that both CGC and the MVGC toolkit yield identical values that correctly identify the cross-induced causalities. Our implementation of CGC also computes the self-induced causalities, although MVGC does not provide this information.

It is known that CGC is subject to some important limitations [21]. First, CGC cannot account for hidden confounding effects or capture non-linear causal relationships. This limitation stems from the assumed linear relationship between variables, which may not be appropriate for complex nonlinear systems. Several nonlinear extensions of Granger causality have been proposed in the literature [30, 31, 32, 34, 33, 33, 35], although their adoption is much more limited than that of the linear counterpart.

S2.2 Convergent cross-mapping (CCM)

Convergent cross-mapping (CCM) [36] is a statistical method for causal inference grounded in the theory of dynamical systems. The approach relies on Takens' embedding theorem [42], which states the conditions under which the dynamics of chaotic nonlinear systems can be captured by observing the trajectory of a single variable over time.

Consider two time series $Q_1(t)$ and $Q_2(t)$. We can define the vector of delay coordinates for Q_1 (similarly for Q_2) as $\boldsymbol{Q}_1 = [Q_1(t), Q_1(t - \Delta T), ..., Q_1(t - (E-1)\Delta T)]$, where ΔT is the time lag, and E is the embedding dimension that determines the number of time lagged observations. The lagged-coordinate embeddings $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ lie within the shadow manifolds M_1 and M_2, respectively. A consequence of Takens' embedding theorem is that if Q_1 and Q_2 belong to the same manifold, then M_1 and M_2 are topologically equivalent. CCM leverages this property and states that if Q_1 causally influences Q_2, then local neighborhoods on M_1 should correspond to local neighborhoods on M_2.

The reconstruction of $Q_2(t)$ from the manifold of Q_1, denoted as $\hat{Q}_2(t)|_{M_1}$, is calculated as a weighted average:

$$\hat{Q}_2(t)|_{M_1} = \sum_{i=1}^{E+1} w_i Q_2(t_i), \tag{S31}$$

where t_i represent the time indices of the nearest neighbors in M_1, and w_i are the weights. The process of estimating $Q_2(t)$ from $\hat{Q}_2(t)|_{M_1}$ is referred to as cross mapping. The weights are computed as

$$w_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j}, \quad u_i = \exp\left(-\frac{d(\boldsymbol{Q}_1(t), \boldsymbol{Q}_1(t_i))}{d(\boldsymbol{Q}_1(t), \boldsymbol{Q}_1(t_1))}\right), \tag{S32}$$

where $d(\boldsymbol{Q}_1(t), \boldsymbol{Q}_1(t_i))$ is the Euclidean distance between the lagged-coordinate vectors in M_1, and $d(\boldsymbol{Q}_1(t), \boldsymbol{Q}_1(t_1))$ is the distance to the nearest neighbor, serving as a normalization factor to scale the weights. This ensures that the prediction is most influenced by the states closest to the current state, whereas the effect of more distant states decays exponentially. The causality from Q_2 to Q_1 is evaluated by examining the correlation coefficient between the actual and estimated values of Q_2 across the timeline:

$$\mathrm{CCM}_{2\to1} = \mathrm{corr}(Q_2, \hat{Q}_2(t)|_{M_1}). \tag{S33}$$

where a causal link is detected for $\mathrm{CCM}_{2\to1} \to 1$ as the length of the time series increases. An analogous definition applies to $\mathrm{CCM}_{1\to2}$ using $Q_1(t)$ and $\hat{Q}_1(t)|_{M_2}$.

The assessment of causality in CCM is predicated upon the convergence of $\mathrm{CCM}_{1\to2}$ as the length of the time series N increases. As N grows, the shadow manifolds become more densely populated, leading to a reduction in the distances among the $E+1$ nearest neighbors. This increased density enables more accurate predictions as $\hat{Q}_2(t)|_{M_1}$ converges to $Q_2(t)$. Observing the convergence of the nearest neighbors and the resulting improvement in prediction accuracy is crucial for substantiating claims about the influence of one variable on another within the framework of CCM. Generally, the stronger the causal link between variables, the faster the convergence with N.

In this work, we employ the Python implementation of CCM by Erneszer [83]. The method was used with an embedding dimension equivalent to the number of variables of the system and executed with a library size that ensured convergence of the prediction capabilities for all cases. In this section, we illustrate the assessment of causality from CCM for the two test cases used by Sugihara *et al.* [36]. The system used consists of a nonlinear logistic difference system with constant coefficients that exhibits chaotic behaviour and represents the phenomenon of mirage correlation. The dynamical system is given by:

$$Q_1(n+1) = Q_1(n)\left[r_1 - r_1 Q_1(n) - \beta_{2\to1} Q_2(n)\right], \tag{S34a}$$
$$Q_2(n+1) = Q_2(n)\left[r_2 - r_2 Q_2(n) - \beta_{1\to2} Q_1(n)\right]. \tag{S34b}$$

For certain combinations of the coefficients, the variables of this system can be positively coupled for long periods of time and can spontaneously become anticorrelated or decoupled. This can become challenging when fitting models for the variables as in the Granger causality framework.

We analyze two cases proposed by Sugihara et al. [36]. The results, shown in Figure S7, aim to illustrate the role of convergence in CCM. For the first case, the variable Q_1 is decoupled from Q_2, while Q_2 is driven by Q_1, namely $\beta_{2\to1} = 0$ and $\beta_{1\to2} \neq 0$, respectively. The results for this case are displayed in Figure S7(left), where the aforementioned dependencies are clearly captured. For the second model, both variables are coupled, but $\beta_{1\to2} > \beta_{2\to1}$. In this scenario, observing the evolution of prediction accuracy can aid in understanding the influence of each variable on the other. Figure S7(right) demonstrates how the lower coupling coefficient $\beta_{2\to1}$ leads to a lesser prediction accuracy as the sample size increases: cross mapping Q_1 using M_2 converges faster than cross mapping Q_2 using M_1. Consequently, the causality from variable Q_2 to Q_1 is weaker compared to that from Q_1 to Q_2.

CCM is particularly suited for data measured from a deterministic nonlinear attractor. For time series of stochastic nature, CCM is known to underperform [123]. Additionally, the presence of noise in the signals complicates the reconstruction process of the attractor manifold and could reduce the reliability of the CCM, as observed in this work and previously reported in the literature [121, 122, 123]. Moreover, CCM fails to accurately predict the causality direction in cases where the coupling is strong enough to lead to the synchronization of variables.

Finally, this study focused on the original CCM algorithm introduced by Sugihara *et al.* [36]. However, more recent algorithms have been developed to examine the impact of noise, external signals, and synchronisation of variables such as pairwise asymmetric inference (PAI) [37], multispatial CCM (MCCM) [38], partial cross-mapping (PCM) [40], and latent CCM [41].

S2.3 Peter–Clark algorithm with momentary conditional independence test (PCMCI)

Conditional independence-based methods uncover the causal structure of interactions, often represented as directed acyclic graphs, by examining the conditional dependencies among variables [61]. The core idea is that if two variables are conditionally independent given a set of other variables, then there is no direct causal link between them. The

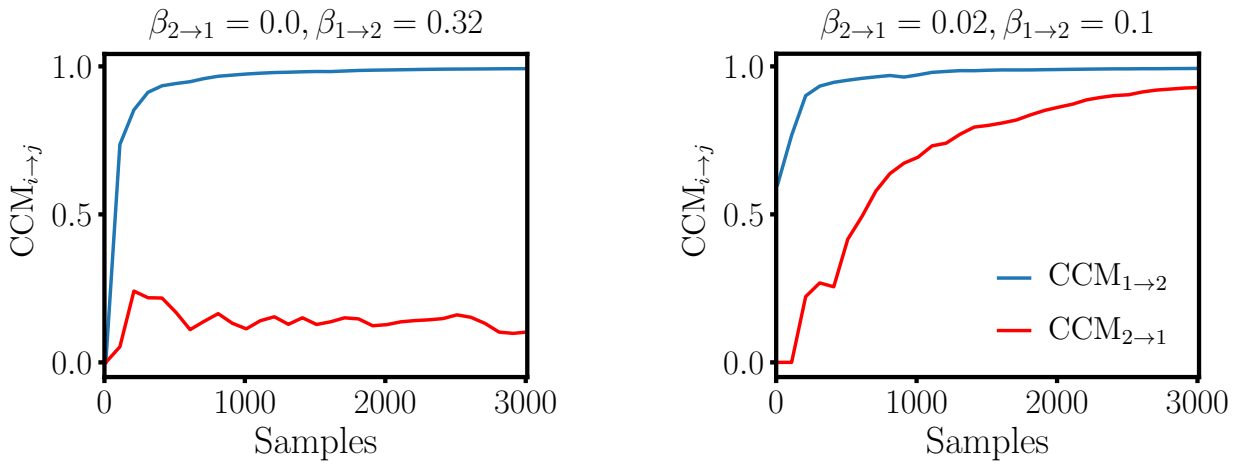


Figure S7: Convergent cross-mapping for two cases in the nonlinear logistic difference system from Equation (S34). $\mathrm{CCM}_{i\to j}$ is the prediction of cross-map estimates represented by the correlation coefficient as a function of the number of samples. (Left) $\mathrm{CCM}_{i\to j}$ for the system with $\beta_{2\to1} = 0$ and $\beta_{1\to2} = 0.32$ (i.e., Q_2 has no effect on Q_1) and (right) $\beta_{2\to1} = 0.02$ and $\beta_{1\to2} = 0.1$ (i.e., the effect of Q_1 on Q_2 is stronger than in the previous case).

approach was popularized by the Peter–Clark (PC) algorithm [59], with subsequent extensions incorporating tests for momentary conditional independence (PCMCI) [23]. PCMCI algorithms are formulated under two main assumptions: (i) the causal Markov condition, which assumes that a variable is independent of its non-descendants, given its parents in a causal graph, and (ii) the faithfulness assumption, which posits that if two variables are statistically independent, then they must be conditionally independent given some set of variables in the causal network. Here, the parents of a particular variable are all those variables in the causal graph that have a direct arrow pointing to it, whereas non-descendant refers to any variable in the graph that is not a direct or indirect outcome of the given variable.

Contrary to methods that condition on the entire past of all processes, PCMCI seeks to identify a reduced set of conditioning variables that includes, at a minimum, the parents of the target variable. These parents are associated with the particular time lag at which the causal relationship occurs in the system. The algorithm unfolds in two phases:

- The initial phase, based on the PC algorithm, is a selection stage aiming to infer a superset of the parents of each variable Q_j at time t, denoted as $\hat{\mathcal{P}}[Q_j(t)]$. This phase starts with a fully connected graph and tests the independence of $Q_i(t - \Delta T)$ and $Q_j(t)$, given conditioning sets of increasing size. The goal of this phase is the removal of irrelevant links and it is designed to have an initial estimate of the parents of each variable Q_j, namely $\hat{\mathcal{P}}[Q_j(t)]$.

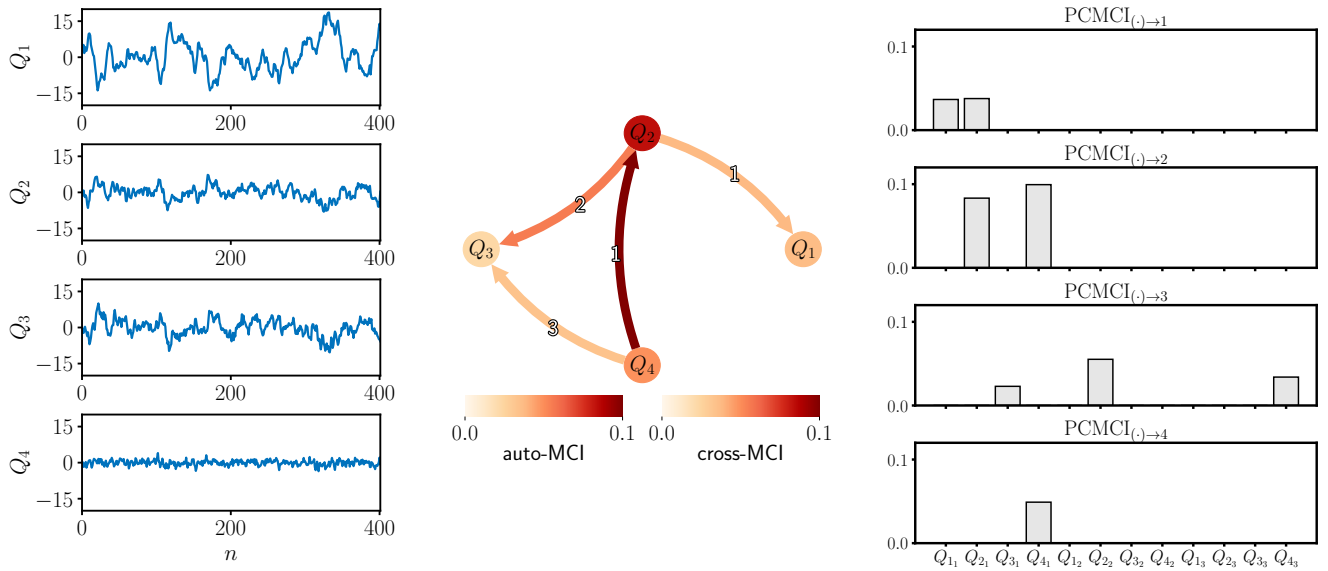
- The second phase of the algorithm conducts the momentary conditional independence (MCI) test for each pair of variables $Q_i(t - \Delta T)$ and $Q_j(t)$ using the estimated parents $\hat{\mathcal{P}}[Q_i(t - \Delta T)]$ and $\hat{\mathcal{P}}[Q_j(t)]$ as conditions. The test examines the null hypothesis at a significance threshold α_{MCI}:

$$Q_i(t - \Delta T) \perp\!\!\!\perp Q_j(t) | \hat{\mathcal{P}}\left[Q_j(t)\right] \setminus Q_i(t - \Delta T), \hat{\mathcal{P}}\left[Q_i(t - \Delta T)\right], \tag{S35}$$

  which denotes the conditional independence of $Q_i(t - \Delta T)$ and $Q_j(t)$, given the causes (or parents) of $Q_j(t)$ excluding $Q_i(t - \Delta T)$, and the causes of $Q_i(t - \Delta T)$. If this hypothesis is not rejected at a significance threshold α_{MCI}, the causal link between $Q_i(t - \Delta T)$ and $Q_j(t)$ is removed. This phase effectively eliminates autodependencies and controls false positives, while improving detection power compared to other adaptations of the PC algorithm.

In our study, we use the Python implementation of PCMCI provided by the package *Tigramite*. It is worth noting that PCMCI accommodates different independence tests, such as partial correlation, nonlinear two-step conditional independence test, and a fully non-parametric test based on conditional mutual information (CMI). For comparison purposes with SURD, we selected the CMI with the k-nearest neighbor (k-NN) estimator. The causal strength of the causal link is determined by the statistic value of the test in Equation S35:

$$\mathrm{PCMCI}_{i\to j} = I\left(Q_i(t - \Delta T); Q_j(t) | C\right), \tag{S36}$$



(a) (b) (c)

Figure S8: Validation for PCMCI. (a) Time series evolution for the signals of the system. (b) Graphical result of the causal connections identified by PCMCI. The indices on top of the arrows represent the time lag at which the causal links are identified. (c) Representation of PCMCI results in bar format. The first subindex represents the variable, and the second subindex represents the time lag. The height of the bar in (c) represents the color of the links in (b).

where $C = \hat{\mathcal{P}}\left[Q_j(t)\right] \setminus Q_i(t - \Delta T), \hat{\mathcal{P}}\left[Q_i(t - \Delta T)\right]$ and $I(\cdot, \cdot|\cdot)$ is the conditional mutual information. $\text{PCMCI}_{i\to j}$ is non-zero only if the null hypothesis is not rejected at a certain significance threshold. All cases examined in this study were evaluated at a significance level of 1%. Furthermore, PCMCI was estimated with $\alpha_{\text{PC}} = 0.05$ and CMI-kNN parameters $k_{\text{CMI}} = 0.1$, $k_{\text{sn}} = 5$, and $B = 200$ permutation surrogates. The parameter α_{PC} denotes the signficance level for the parent selection phase of the algorithm, k_{CMI} determines the size of hypercubes, i.e., the data-adaptive local length-scale used in the k-NN estimator, k_{sn} denotes the number of neighbours to which each sample is mapped randomly, and B the number of surrogates to approximate the null distribution. More details about these parameters and their effect on the PCMCI algorithm are provided in Ref. [123].

The graphical results for PCMCI are now reported for the test case provided by the package *Tigramite*. The system is a set of time series given by:

$$
\begin{aligned}
Q_1(n) &= 0.7Q_1(n-1) - 0.8Q_2(n-1) + \eta_1(n),\\
Q_2(n) &= 0.8Q_2(n-1) + 0.8Q_4(n-1) + \eta_2(n),\\
Q_3(n) &= 0.5Q_3(n-1) + 0.5Q_2(n-2) + 0.6Q_4(n-3) + \eta_3(n),\\
Q_4(n) &= 0.7Q_4(n-1) + \eta_4(n),
\end{aligned}
$$

where η are independent Gaussian variables. PCMCI results for this system are provided in Figure S8, where we also introduce the notation and organization of the results from PCMCI used in our study.

Finally, we discuss some of the limitations of the PCMCI algorithm. First, the method relies on the selection of parameters during the parent-selection and link-selection steps. The presence or absence of a link can be affected by the confidence level used during these steps. Certain parameters used in the k-NN estimator, such as permutation surrogates, might also affect the results: lower values alleviate computational cost; however, they could lead to reduced confidence in the results. The choice of independence test may also impact the results depending on the type of relationships in the data. The CMI is conceptually the most reliable approach but also computationally expensive and data demanding, particularly for a large number of neighbors and permutations. A summary of the results presented in this work for different independence tests is reported in Table S2. Lastly, the presence of redundant variables in the conditioning set can yield to unidentified links, even when the conditioning set encompasses the entire history of all other processes [23], as demonstrated in the examples from this work.

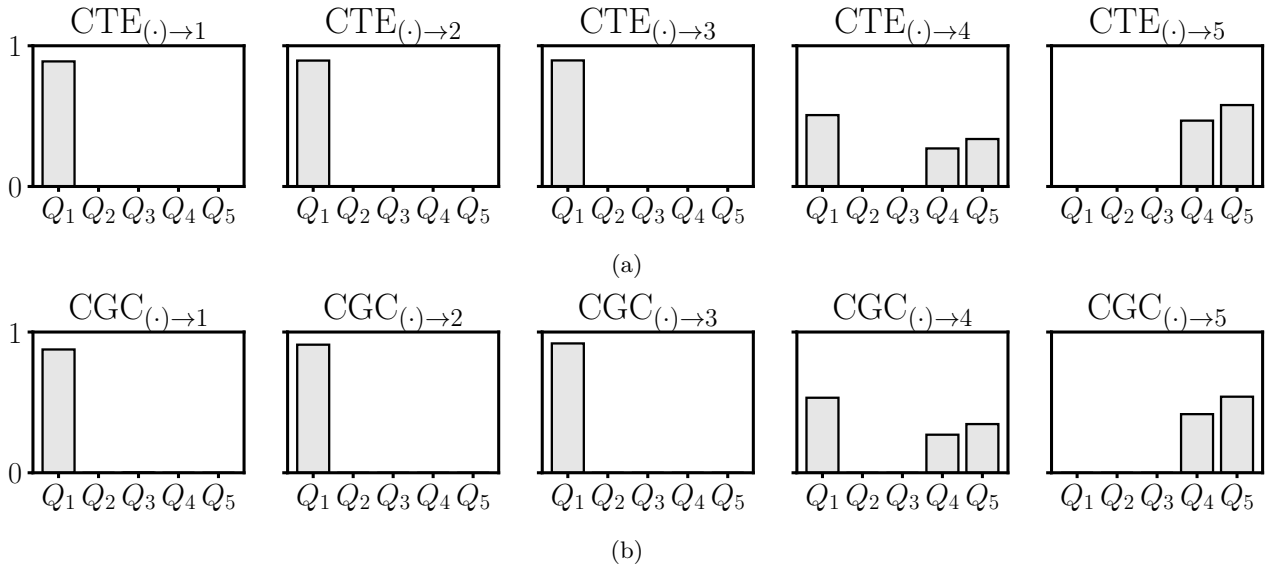


Figure S9: Comparison of current implementation of conditional transfer entropy (CTE) and conditional Granger causality (CGC) for a five-node oscillatory network [28]. $\mathrm{CGC}_{i\to j}$ is divided by 2 for consistency with $\mathrm{CTE}_{i\to j}$. Both quantities are then normalized with the mutual information between the target and the vector of observed variables. The system is integrated for $N = 10^8$ steps and the phase space was partitioned using 50 bins for each variable to calculate CTE.

S2.4 Conditional transfer entropy (CTE)

Transfer entropy (TE) is a method for assessing the directional information transfer between two time series in a non-parametric manner [49]. The TE between two time signals Q_1 and Q_2 is defined as follows:

$$\mathrm{TE}_{2\to 1} = H\left(Q_1(t)|\boldsymbol{Q}_1^{(k)}\right) - H\left(Q_1(t)|\boldsymbol{Q}_1^{(k)}, \boldsymbol{Q}_2^{(l)}\right), \tag{S37}$$

where $H(\cdot \mid \cdot)$ is the conditional Shannon entropy defined in Equation (S2), k and l are constants denoting the time lag of the variables, $\boldsymbol{Q}_1^{(k)} = Q_1(t-\Delta T), \ldots, Q_1(t-k\Delta T)$ and $\boldsymbol{Q}_2^{(l)} = Q_2(t-\Delta T), \ldots, Q_2(t-\Delta T)$. TE measures the amount of information about the future state of Q_2 that is exclusively provided by the current state of Q_1, beyond what is already known from the past of Q_2 itself. The method is particularly effective for analyzing complex systems where traditional linear methods may not be suitable, such as in the study of nonlinear dynamics.

The multivariate extension of TE is given by the conditional transfer entropy (CTE) [50, 51, 52, 45]. The approach is usually formulated for one single time lag as:

$$\mathrm{CTE}_{\boldsymbol{i}\to j}(\Delta T) = H(Q_j^+|\boldsymbol{Q}_{\not i}) - H(Q_j^+|\boldsymbol{Q}), \tag{S38}$$

where $\boldsymbol{Q} = [Q_1(t-\Delta T), Q_2(t-\Delta T), \ldots, Q_N(t-\Delta T)]$ represents the vector of observed variables and $\boldsymbol{Q}_{\not i}$ denotes the same vector excluding the components given by the indices in $\boldsymbol{i}$. In this work, we have restricted our analysis to applications of CTE with one single component in $\boldsymbol{i}$. The results of CTE using combinations of variables is discussed below in Figure S23.

Barnett *et al.* [27] showed that conditional Granger causality and CTE are equivalent up to a factor of 2 when the variables in $\boldsymbol{Q}$ follow a joint multivariate Gaussian distribution:

$$\mathrm{CTE}_{i\to j} = \frac{1}{2}\mathrm{CGC}_{i\to j}. \tag{S39}$$

Consequently, both measurements of causality share the same upper bound, given by the maximum value of the mutual information between the target and the vector of observed variables $\boldsymbol{Q}$. Throughout our study, $\mathrm{CGC}_{i\to j}$ is divided by 2 to allow for direct comparisons with CTE.

We validate the current implementation of CTE in the five-node oscillatory network [28] from Equation (S30). In this case, CTE and CGC should yield similar results, as the variables are Gaussian distributed. The results for CTE

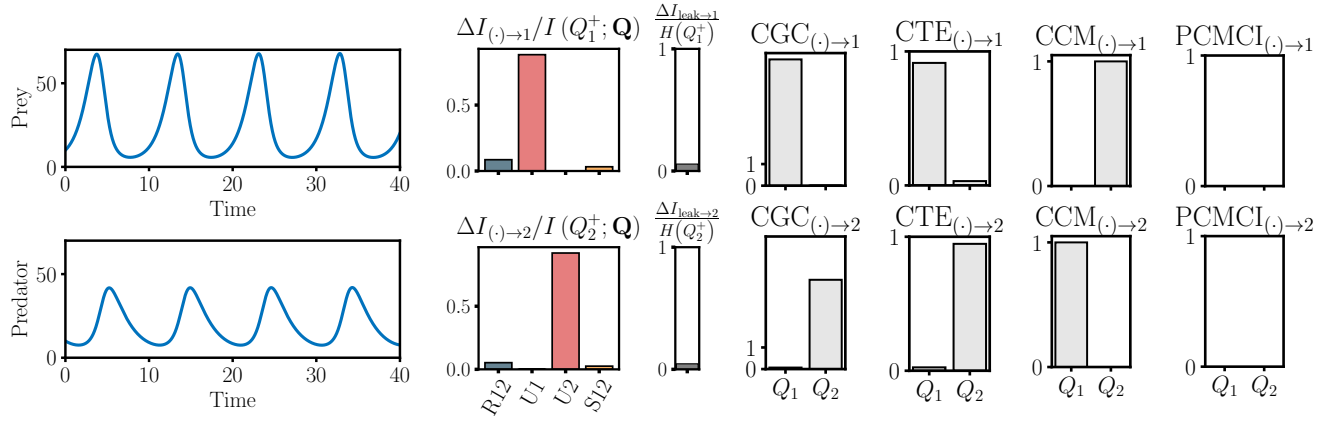


Figure S10: Lotka–Volterra prey-predator model [68, 69]. Redundant (R), unique (U), and synergistic (S) causalities. The gray bar is the causality leak. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD. The methods are applied at a time lag $\Delta t = 1.5$ corresponding to the delay between the maximum in the predator and prey population numbers.

and CGC are shown in Figure S9, where both methods provide same causal links with almost identical strengths.

While CTE offers a robust framework for causal inference in time series data, its practical application requires careful consideration of sample size and estimation methods. Various approaches, such as binning, nearest neighbors, and non-uniform embedding, are employed to effectively estimate the probabilities in entropy calculations. Regarding the effect of multiple variables, some attempts have been made in the literature to account for causality from combinations of variables, for example, through the calculation of CTE in its multivariate form [45]. However, these methods may yield negative values of causality, which can limit the interpretability of the results. Examples are discuss below in Figure S23.

S3 Additional validation cases

We discuss additional validation cases for SURD, CGC, CTE, CCM, and PCMCI. The metric for success is based on whether the results are consistent with the functional dependencies of the system, rather than on the concrete value of the causal strength provided by each method.

S3.1 Lotka–Volterra prey-predator model

The Lotka–Volterra predator-prey model [68, 69] was envisioned to describe the dynamics of biological systems in which two species interact. The model can be expressed as a pair of first-order nonlinear differential equations:

$$\frac{\mathrm{d}Q_1}{\mathrm{d}t} = \alpha Q_1 - \beta Q_1 Q_2, \tag{S40}$$

$$\frac{\mathrm{d}Q_2}{\mathrm{d}t} = \Gamma Q_1 Q_2 - \gamma Q_2, \tag{S41}$$

where Q_1 and Q_2 denote the prey and the predator population number, respectively, and $\alpha = 1$, $\beta = 0.05$, $\Gamma = 0.02$, and $\gamma = 0.5$ represent the prey reproduction rate, predation rate, predator reproduction rate and predator death rate, respectively. Figure S10 shows a visualization of the time signals of the model together with the results from SURD, CGC, CTE, CCM and PCMCI. SURD identifies synergistic causality as the most significant contribution for both variables, i.e., $\Delta I^U_{1\to 1}$ and $\Delta I^U_{2\to 2}$, together with some smaller redundant and synergistic causalities for both variables. This is in agreement with the low values of the nonlinear coupling between variables in Equation (S40). In relation to the other approaches, we observe relationships consistent with the functional dependencies in Equation (S40) for CCM, CGC and CTE; however, PCMCI does not detect any significant causal link.

S3.2 Moran effect model

The Moran effect [70] refers to a phenomenon in population ecology that describes how spatially separated populations can exhibit synchronous fluctuations in their sizes due to a common environmental factor affecting them all simultaneously. The system comprises two variables, N_1 and N_2, that do not exhibit causal relationships but are

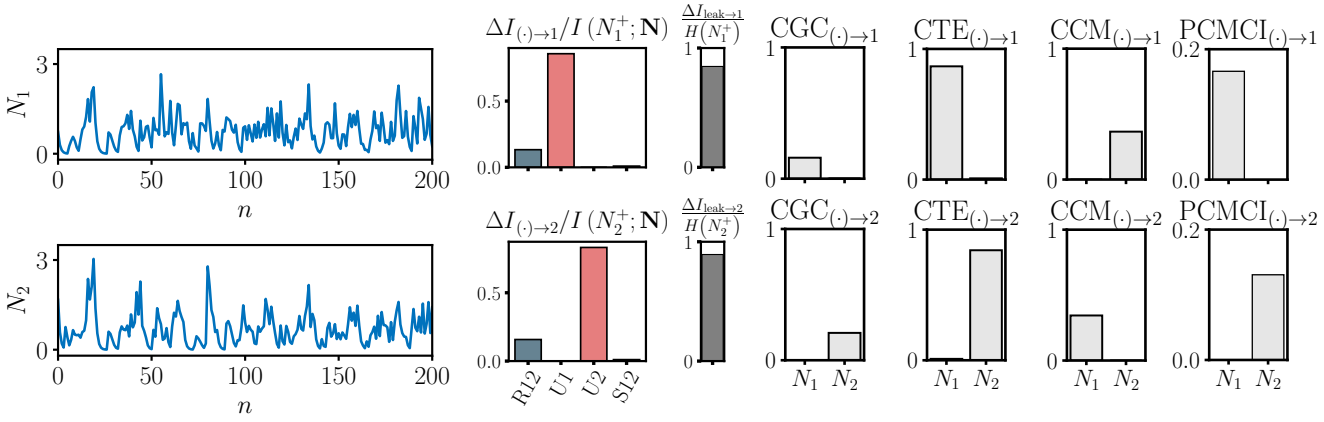


Figure S11: Moran effect model [70]. (Left panel) Time evolution of independent variables N_1 and N_2. (Right panel) Redundant (R), unique (U), and synergistic (S) causalities with the vector of observed variables $\boldsymbol{N} = [N_1, N_2]$. The gray bar is the causality leak. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD. The model was run for 10^8 time steps with the set of parameters: $r_1 = 3.4, r_2 = 2.9, \Psi_1 = 0.5, \Psi_2 = 0.6, s_1 = 0.4, s_2 = 0.35, D_1 = 3, D_2 = 3, R_1(0) = R_2(0) = 1, N_1(0) = N_2(0) = 0.5$.

significantly correlated in their time series due to shared external forcing, V. The latter follows a Gaussian distribution with a mean of zero and a standard deviation of one, acting on N_1 and N_2 through the mediator variables R_1 and R_2, respectively. The equations of the model are given by:

$$R_i(n+1) = r_i N_i(n)[1 - N_i(n)]e^{-\Psi_i V(n)}$$
$$N_i(n+1) = s_i N_i(n) + \max\left[R_i(n - D_i), 0\right].$$

The vector of observed variables is $\boldsymbol{N} = [N_1, N_2]$, and the aim is to assess whether different methods can discern the causal independence between N_1 and N_2 despite their significant correlation due to a shared confounder. The results are provided in Figure S11. The most significant causal interactions detected by SURD are self-unique causalities, while the remaining components are redundant causalities. These outcomes align with the functional dependencies between the variables in Figure S11, and SURD accurately captures the confounding effect. The other approaches also identify the causal independence among variables. CCM yields a non-zero value for cross-induced causal relationships. However, the inspection of this value for increasing number of samples revealed that it does not converge to 1, suggesting no causal link between the variables.

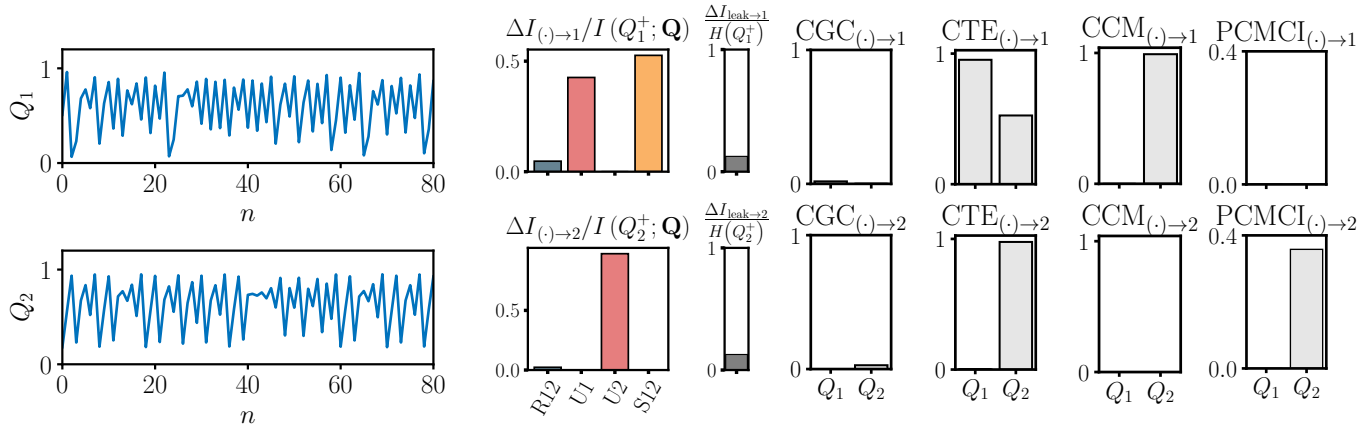
S3.3 Coupled logistic difference system

We consider the deterministic, nonlinear, logistic difference system proposed by Sugihara et al. [36]. The system, given in Equation (S34), illustrates the concept of mirage correlation, namely, a perceived but spurious correlation between two variables. To illustrate this phenomenon, we use four versions of the models for different values of the coupling parameters $\beta_{2\to 1}$ and $\beta_{1\to 2}$. A comparison of the results for all methods is shown in Figure S12 for one-way coupling between variables and in Figure S13 for the two-way coupling case.

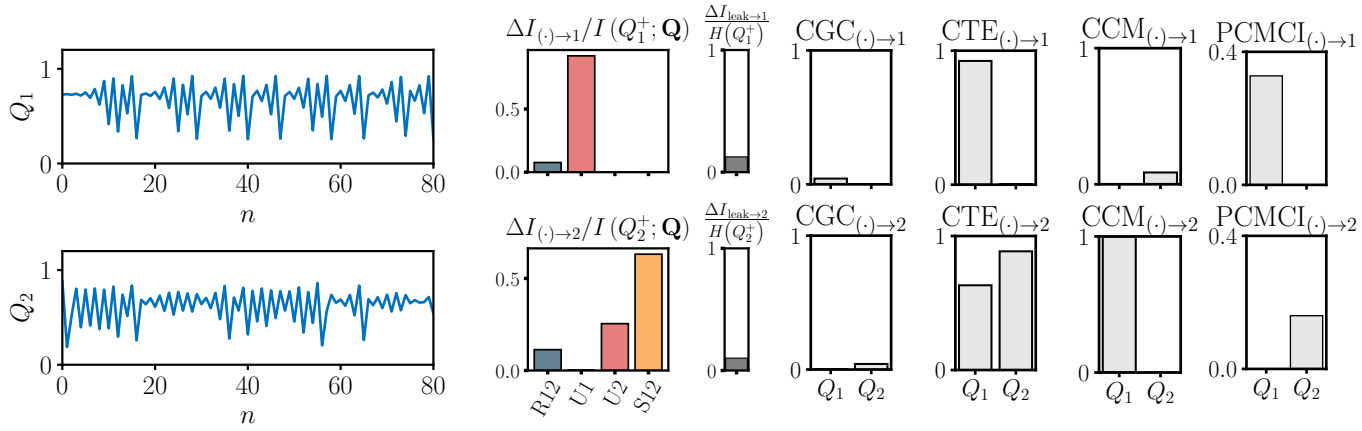
- One-way coupling $Q_2 \to Q_1$: SURD identifies the causalities $\Delta I^S_{12\to 1}$ and $\Delta I^U_{1\to 1}$ as the most important causalities driving Q_1, whereas only $\Delta I^U_{2\to 2}$ is identified for Q_2. This is consistent with the non-zero coupling constant $\beta_{2\to 1}$.

- One-way coupling $Q_1 \to Q_2$: similar conclusions can be drawn from this case. SURD identifies the causalities $\Delta I^S_{12\to 2}$ and $\Delta I^U_{2\to 2}$ as the most important variables driving Q_2, whereas only $\Delta I^U_{1\to 1}$ is identified for Q_1. This is consistent with the non-zero coupling constant $\beta_{1\to 2}$.

- Two-way coupling $\beta_{2\to 1} > \beta_{1\to 2}$: SURD identifies the causality $\Delta I^U_{2\to 2}$ for variable Q_2, which implies that Q_2 is mostly self-causal. This is consistent with the low value of the coupling constant $\beta_{1\to 2}$. There are also redundant and synergistic contributions given by $\Delta I^R_{12\to 2}$ and $\Delta I^S_{12\to 2}$, respectively, although these are smaller than the self-induced causality. In the case of Q_1, the synergistic contribution dominates over the unique causality $\Delta I^U_{1\to 1}$. This is agrees with the fact that the coupling parameter is significantly larger from $Q_2 \to Q_1$ (i.e., $\beta_{2\to 1}$) than from $Q_1 \to Q_2$ ($\beta_{1\to 2}$).



(a) $\beta_{2\to 1} = 0.1$, $\beta_{1\to 2} = 0$



(b) $\beta_{2\to 1} = 0$, $\beta_{1\to 2} = 0.32$

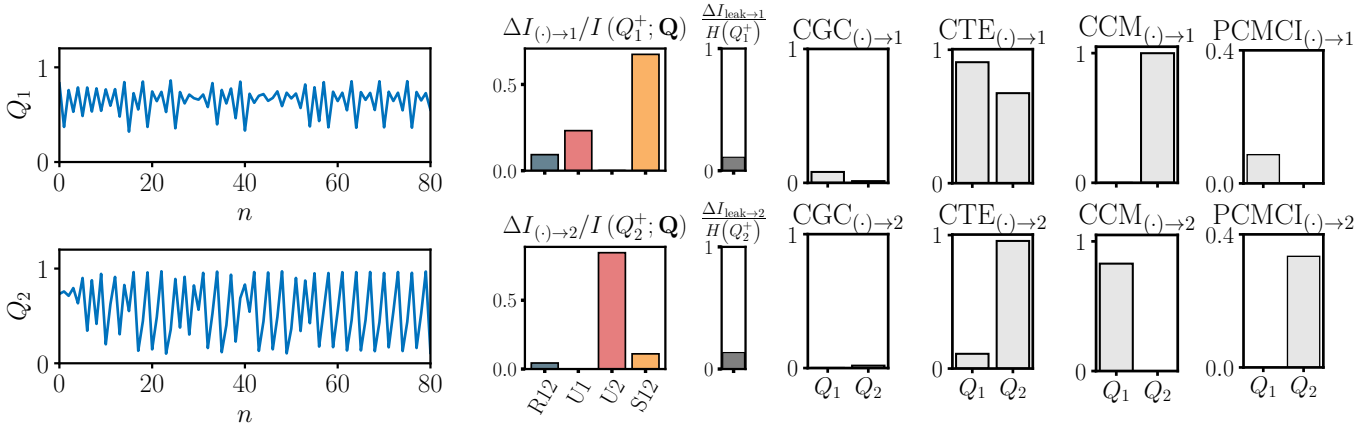
Figure S12: Nonlinear logistic difference system with one-way coupled components [36] for (a) $Q_2 \to Q_1$ and (b) $Q_1 \to Q_2$. Redundant (R), unique (U), and synergistic (S) causalities. The gray bar is the causality leak. The results of CGC, CTE, PCMCI and CCM are depicted on the right. CGC and CTE use the same normalization as SURD.

- Two-way coupling $\beta_{2\to 1} < \beta_{1\to 2}$: SURD identifies the causality $\Delta I^U_{1\to 1}$ for variable Q_1, which implies that Q_1 is mostly self-causal. This is consistent with the low value of the coupling constant $\beta_{1\to 2}$ from $Q_2 \to Q_1$. There are also redundant and synergistic contributions given by $\Delta I^R_{12\to 1}$ and $\Delta I^S_{12\to 1}$, respectively, although these are smaller than the self-induced causality. In the case of Q_2, the redundant and synergistic contributions dominate over the unique causality from Q_2 to Q_2. This is consistent with the fact that the coupling parameter is significantly larger from $Q_1 \to Q_2$ than from $Q_2 \to Q_1$ (i.e., $\beta_{1\to 2} > \beta_{2\to 1}$).

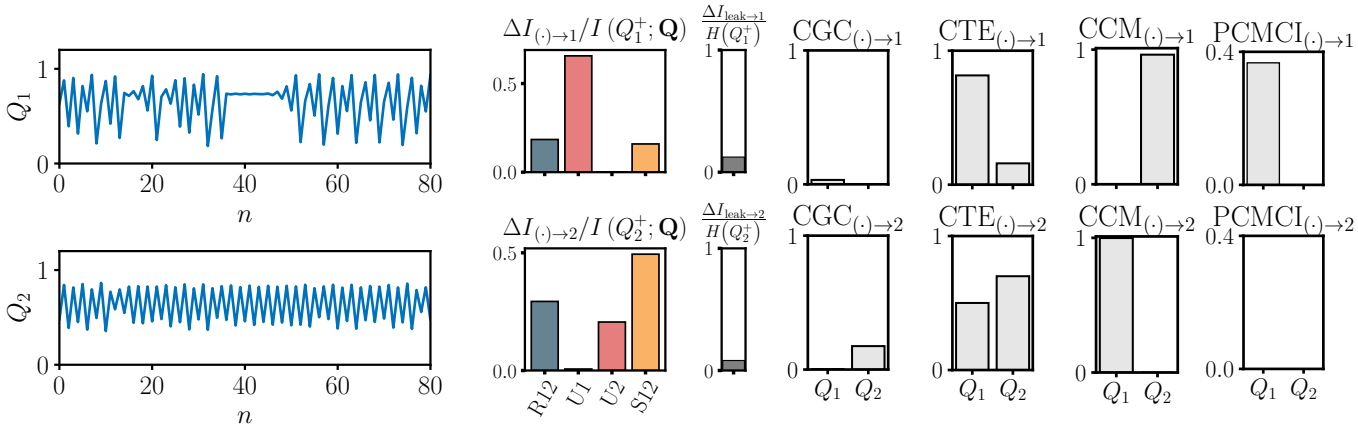
Among the other methods, only CTE and CCM can identify all the functional dependencies between variables in all cases. PCMCI correctly identifies self-causal links in all cases; however, it detects cross-induced causalities only in the two-way coupling case, which are very low in intensity. It is also important to mention that these methods fail to provide information on whether the coupling in the system is introduced by individual variables acting alone, or whether it is the combined effect of multiple variables that drives the coupling. In these scenarios, SURD offers a more comprehensive understanding of the system through the synergistic contribution $\Delta I^S_{12\to j}$.

S3.4 Stochastic system with time-lagged dependencies

We examine bivariate stochastic systems with linear and nonlinear time-lagged dependencies previously studied in [71] and [35], respectively. These cases demonstrate the performance of causal methods when the relationships are introduced with different time lags. The notation for the observable vector is $\boldsymbol{Q} = [Q_{1_1}, Q_{1_2}, Q_{2_1}, Q_{2_2}]$, where the second subindex denotes the time lag, e.g., $Q_{1_1} = Q_1(n-1)$ and $Q_{1_2} = Q_1(n-2)$.



(a) $\beta_{2\to 1} = 0.2$, $\beta_{1\to 2} = 0.01$



(b) $\beta_{2\to 1} = 0.02$, $\beta_{1\to 2} = 0.1$

Figure S13: Nonlinear logistic difference system with two-way coupled components [36] for (a) $\beta_{2\to 1} > \beta_{1\to 2}$ and (b) $\beta_{2\to 1} < \beta_{1\to 2}$. Redundant (R), unique (U), and synergistic (S) causalities. The gray bar is the causality leak. The results of CGC, CTE, PCMCI and CCM are depicted on the right. CGC and CTE use the same normalization as SURD.

For the first case, the relationships between variables are linear and they are described as:

$$Q_1(n+1) = 0.95\sqrt{2}Q_1(n) - 0.9025Q_1(n-1) + W_1(n)$$
$$Q_2(n+1) = 0.5 \cdot Q_1(n-1) + W_2(n)$$

where $W_i \sim \mathcal{N}(0,1)$ represents a stochastic forcing following a Gaussian distribution with a mean of zero and a variance of one. In this case, $Q_1(n)$ is self-caused at two different time lags, i.e., $Q_1(n-1)$ and $Q_1(n-2)$, while $Q_1(n-2)$ drives $Q_2(n)$. These relationships are well captured by the SURD, where the most significant causal relationships are $\Delta I^S_{1_1 1_2 \to 1}$ and $\Delta I^U_{1_2 \to 2}$, with the subscript indicating the time lag of the variable. Note that since the relationships between the variables are linear, the functional dependencies of Q_1 can also be expressed as a function of $Q_2(n-1)$. This is why SURD identifies synergistic causalities $\Delta I^S_{1_1 2_1 \to 1}$ and $\Delta I^S_{1_1 2_2 \to 1}$.

Among the other methods, we observe that only CGC and CTE can identify all causal relationships between variables across all time lags. PCMCI clearly detects the links $Q_1(n-1) \to Q_1(n)$ and $Q_1(n-2) \to Q_2(n)$, but depending on the threshold used the remaining identified connections exhibit lower intensity values compared to these. The identification of these latter links depends on the confidence threshold used. For a strict threshold of $\alpha = 0.01$, only $Q_1(n-1) \to Q_1(n)$ and $Q_1(n-2) \to Q_2(n)$ are identified. For CCM, causality needs to be assessed upon the convergence of the prediction skill to 1 as the length of the time series increases. In this system, the prediction skill for Q_2 using variable $Q_1(n-2)$ is slightly higher than that for $Q_1(n-1)$, implying that the manifold associated with $Q_1(n-2)$ enables a better reconstruction of the states of $Q_2(n)$ than the states of $Q_1(n-1)$. A

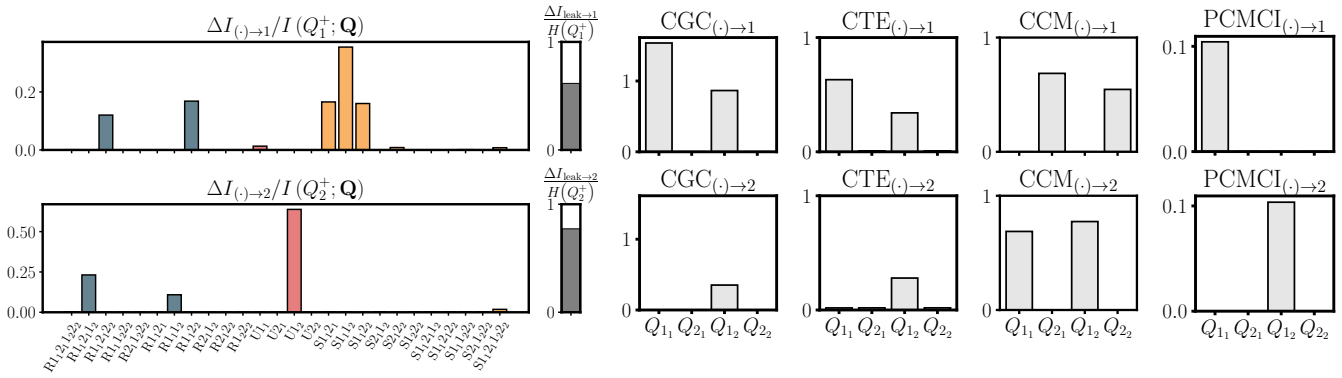


Figure S14: Stochastic systems with linear time-lagged dependencies [71]. Redundant (R), unique (U), and synergistic (S) causalities in blue, orange and yellow, respectively. The subindex of the labels represents the time delay associated with the variable. The gray bar is the causality leak. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

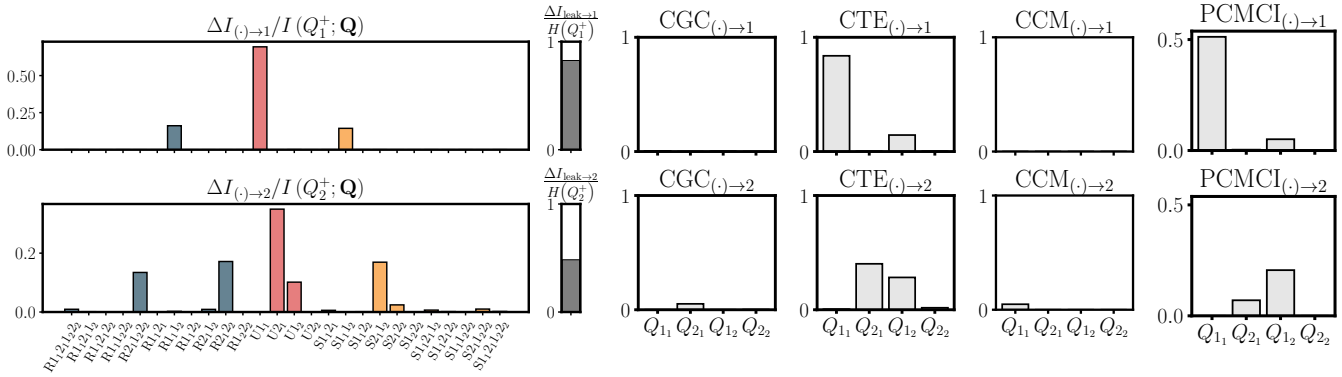


Figure S15: Stochastic systems with nonlinear time-lagged dependencies [35]. Redundant (R), unique (U), and synergistic (S) causalities in blue, orange and yellow, respectively. The subindex of the labels represents the time delay associated with the variable. The gray bar is the causality leak. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

similar conclusion can be drawn for Q_2, where the manifold associated with $Q_2(n-1)$ provides better reconstruction of the states of $Q_1(n)$.

The second case is a system with non-linear relationships between variables. This system was previously studied by Bueso *et al.* [35] for a single-lag dependency. In this work, we employ an extension of the system for multiple time lags to demonstrate how SURD can also be applied to identify nonlinear relationships at multiple lags. The equations describing the system are given by:

$$Q_1(n+1) = 3.4Q_1(n)(1-Q_1(n)^2)\exp\left(-Q_1(n-1)^2\right) + W_1(n)$$
$$Q_2(n+1) = 3.4Q_2(n)(1-Q_2(n)^2)\exp\left(-Q_2(n)^2\right) + \frac{Q_1(n-1)Q_2(n)}{2} + W_2(n)$$

where the time-varying stochastic forcing W_i that affects each variable follows a Gaussian distribution with a mean of zero and a variance of 0.4. For this model, $Q_1(n-1)$ and $Q_1(n-2)$ are common drivers of $Q_1(n)$, and $Q_2(n-1)$ and $Q_1(n-2)$ for $Q_2(n)$.

The results are shown in Figure S15. The causal connections identified by SURD $\Delta I^U_{1_1\to 1}$, $\Delta I^S_{1_1 1_2\to 1}$, $\Delta I^U_{2_1\to 2}$, and $\Delta I^U_{1_2\to 2}$ correctly capture the functional dependencies of the system. Since the relationships between variables are highly nonlinear, CGC and CCM are no longer able to identify the causal connections between variables. PCMCI and CTE are still capable of discerning the relationships $[Q_1(n-1), Q_1(n-2)] \to Q_1(n)$ and $[Q_2(n-1), Q_1(n-2)] \to Q_2(n)$ in a consistent manner with the functional dependencies between variables.

S3.5 Synchronization in logistic maps

The one-dimensional logistic map is a recurrence given by the relationship,

$$Q_1(n+1) = \alpha_1 Q_1(n)[1 - Q_1(n)], \tag{S42}$$

where n is the time step and α_1 is a constant. Equation (S42) exhibits a chaotic behavior for $\alpha_1 \approx 3.57 - 4$ [72]. We consider the three logistic maps:

$$Q_1(n+1) = \alpha_1 Q_1(n)[1 - Q_1(n)], \tag{S43a}$$
$$Q_2(n+1) = \alpha_2 f_{12}[1 - f_{12}], \tag{S43b}$$
$$Q_3(n+1) = \alpha_3 f_{123}[1 - f_{123}], \tag{S43c}$$

which are coupled by

$$f_{12} = \frac{Q_2(n) + c_{1\to2} Q_1(n)}{1 + c_{1\to2}}, \tag{S44a}$$
$$f_{123} = \frac{Q_3(n) + c_{12\to3} Q_1(n) + c_{12\to3} Q_2(n)}{1 + 2c_{12\to3}}, \tag{S44b}$$

where $\alpha_1 = 3.68$, $\alpha_2 = 3.67$, and $\alpha_3 = 3.78$ are constants, $c_{1\to2}$ is the parameter coupling Q_2 with Q_1, and $c_{12\to3}$ is the parameter coupling Q_3 with Q_2 and Q_1. The clear directionality of the variables in this system for different values of $c_{1\to2}$ and $c_{12\to3}$ offers a simple testbed to illustrate the behavior of the causality. The causal analysis for SURD is performed for one time-step lag after integrating the system for 10^8 steps. The phase-space was partitioned using 100 bins for each variables.

First, we consider three cases with different degrees of coupling between Q_1 and Q_2 while maintaining Q_3 uncoupled. The results are shown in Figure S16.

- Uncoupled systems ($c_{1\to2} = c_{12\to3} = 0$). In this case, Q_1, Q_2, and Q_3 are completely uncoupled and the only non-zero causalities are the self-induced unique components $\mathit{\Delta}I^U_{1\to1}$, $\mathit{\Delta}I^U_{2\to2}$, and $\mathit{\Delta}I^U_{3\to3}$, as shown by the left panels in Figure S16(a).

- Intermediate coupling $Q_1 \to Q_2$ ($c_{1\to2} = 0.1$ and $c_{12\to3} = 0$). In this case, the dynamics of Q_2 are affected by Q_1. This is shown in Figure S16(b) by the non-zero terms $\mathit{\Delta}I^R_{12\to2} \neq 0$, $\mathit{\Delta}I^U_{1\to1} \neq 0$ and $\mathit{\Delta}I^S_{12\to2} \neq 0$. The latter is the synergistic causality due to the combined effect of Q_1 and Q_2, which is a manifestation of the coupling term $f_{1\to2}$. We can also observe that $\mathit{\Delta}I^R_{12\to1} \neq 0$. The latter redundant causality appears due to the emerging synchronization between Q_1 and Q_2. However, note that there is no other contribution (unique or synergistic) from Q_2 to Q_1. Hence, the redundant causality $\mathit{\Delta}I^R_{12\to1}$ does not necessarily imply that conducting an intervention on Q_2 (e.g., a perturbation) will alter the value of Q_1. Instead, it should be interpreted as Q_2 being able to inform about the future of Q_1, which is expected since Q_1 is contained in the right-hand side of the equation for Q_2. The only non-zero causality for Q_3 is again $\mathit{\Delta}I^U_{3\to3}$, as it is uncoupled from Q_1 and Q_2. This result also demonstrates that the detection power of SURD is not affected by the inclusion of new independent variables in the analysis.

- Strong coupling $Q_1 \to Q_2$ ($c_{1\to2} = 1$ and $c_{12\to3} = 0$). Taking the limit $c_{1\to2} \to \infty$, it can be seen that $Q_2 \equiv Q_1$. It is also known that even for lower values of $c_{12\to3} \sim 1$, Q_1 and Q_2 synchronize and both variables exhibit identical dynamics. This is revealed in Figure S16(c), where the only non-zero causalities are $\mathit{\Delta}I^R_{12\to1} = \mathit{\Delta}I^R_{12\to2} \neq 0$. The identical redundant causalities along with the absence of any unique or synergistic causality between Q_1 and Q_2, imply that both variables are fully synchronized. In this situation, the directionality of the causality cannot be established, as Q_1 and Q_2 behave as a single variable, but SURD still effectively identifies the state of synchronization. Similar to the two previous cases, Q_3 remains unaffected ($\mathit{\Delta}I^U_{3\to3} \neq 0$).

Compared to other methods, SURD provides consistent results across the three coupling scenarios. While all methods correctly identified links in the uncoupled case, introducing intermediate and strong coupling between Q_1 and Q_2 hindered the identification of causal relationships. In such cases, the synchronization of Q_1 and Q_2 can be observed in SURD through redundant causality, a capability that other methods lack.

Next, we consider two additional cases in which Q_3 is coupled with Q_1 and Q_2. The results are shown in Figure S17.

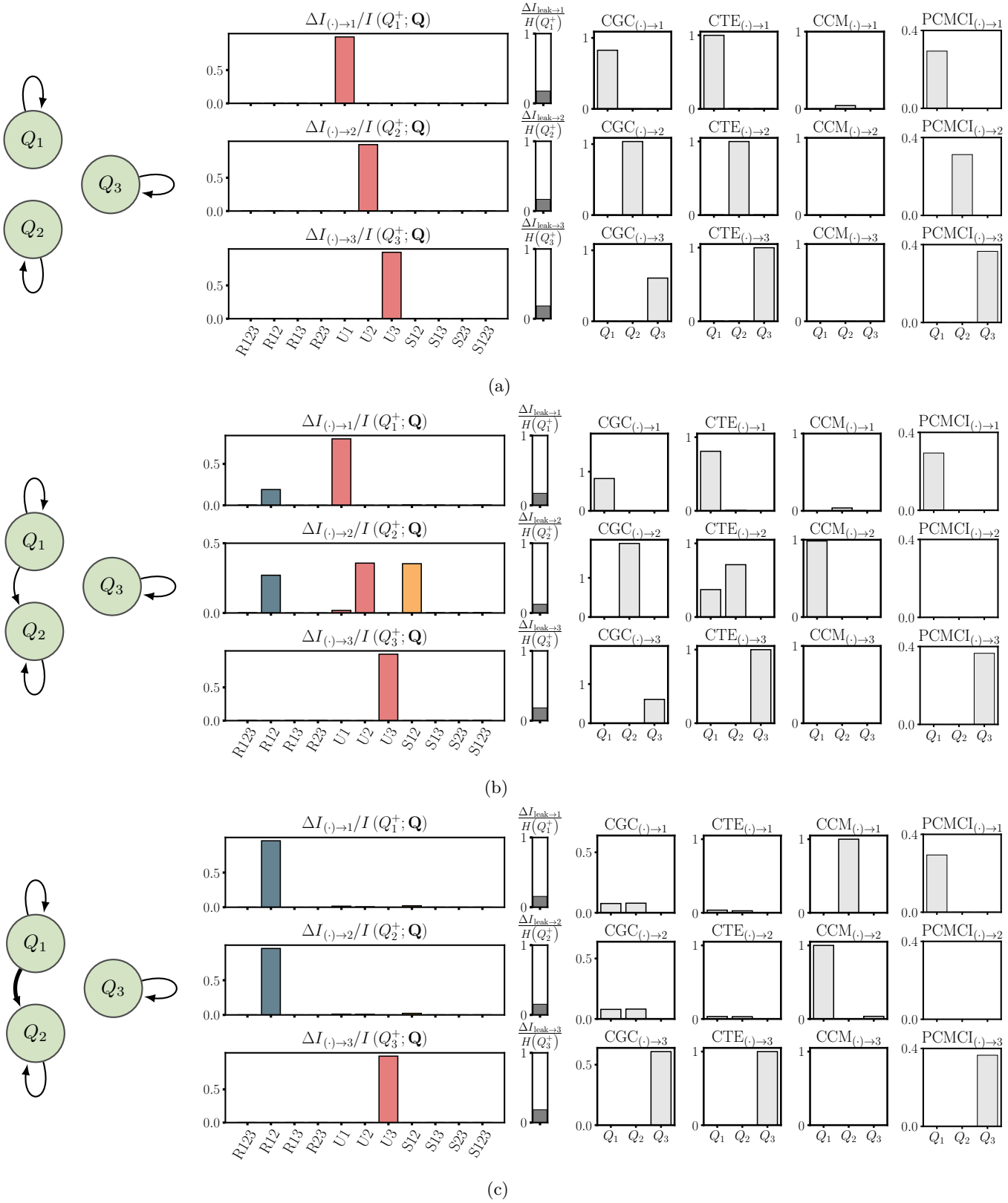
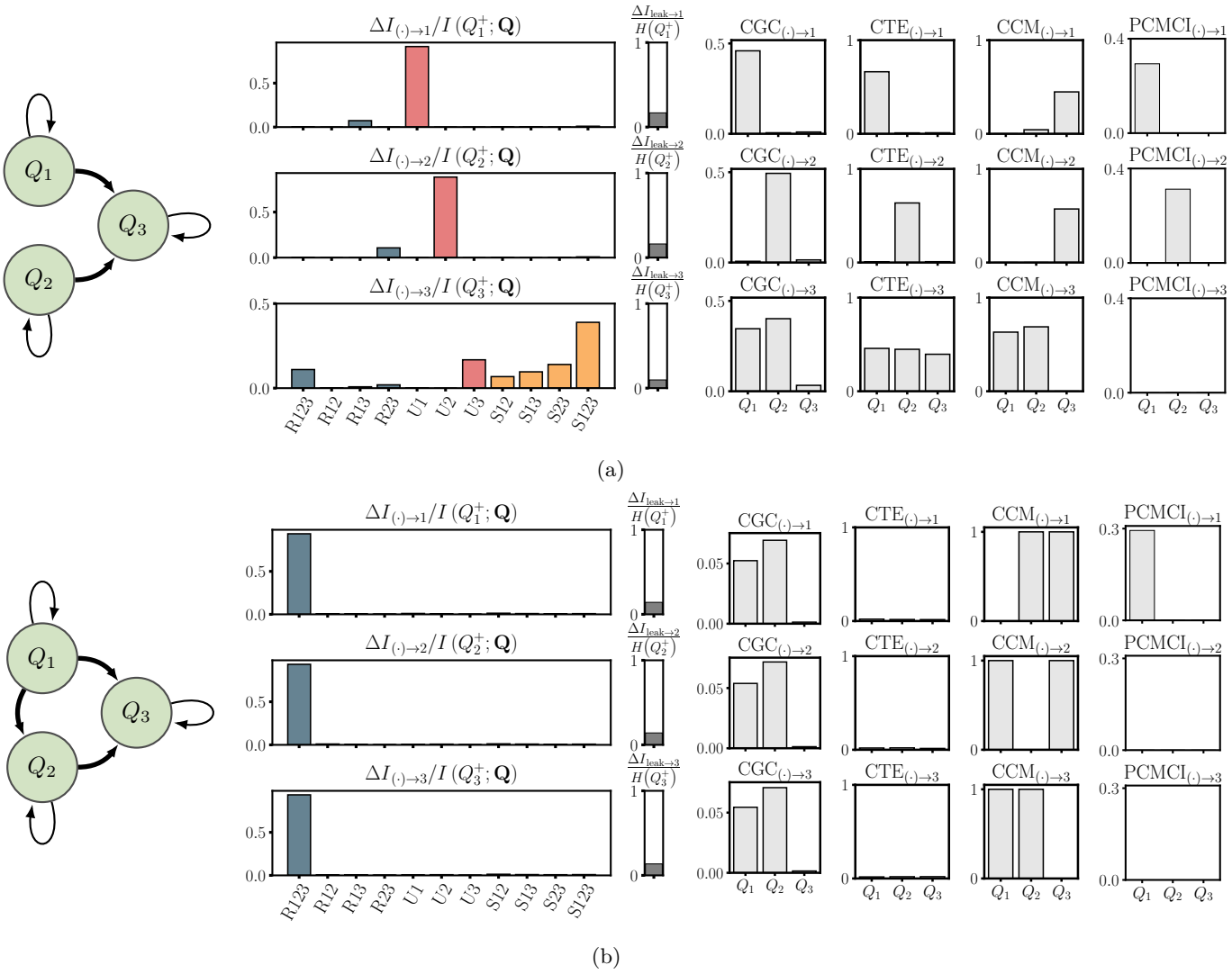


Figure S16: Logistic maps with none or one variables coupled. Redundant (R), unique (U) and synergistic (S) causalities for Q_1, Q_2, Q_3 in coupled logistic maps for (a) uncoupled variables, $c_{1\to 2} = 0$ and $c_{12\to 3} = 0$, (b) intermediate coupling $Q_1 \to Q_2$, $c_{1\to 2} = 0.1$ and $c_{12\to 3} = 0$ and (c) strong coupling $Q_1 \to Q_2$ $c_{1\to 2} = 1$ and $c_{12\to 3} = 0$. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.



(a)

(b)

Figure S17: Logistic maps with two or three variables coupled. Redundant (R), unique (U) and synergistic (S) causalities for Q_1, Q_2, Q_3 for (a) uncoupled variables, $c_{1\to 2} = 0$ and $c_{12\to 3} = 1$, (b) strong coupling $Q_1 \to Q_2$ and $[Q_1, Q_2] \to Q_3$, $c_{12\to 3} = 1$ and $c_{12\to 3} = 1$. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

- Strong coupling $[Q_2, Q_1] \to Q_3$ and no coupling $Q_1 \to Q_2$ ($c_{1\to 2} = 0$ and $c_{12\to 3} = 1$). The results, included in Figure S17(a), show that most of the causality to Q_1 and Q_2 is self-induced and unique ($\Delta I^U_{1\to 1} \neq 0$ and $\Delta I^U_{2\to 2} \neq 0$, respectively). There is also a strong causality from Q_1 and Q_2 to Q_3 in the form of synergistic causality, being $\Delta I^S_{123\to 3}$ the dominant component consistent with the coupling term f_{123}.
- Strong coupling $[Q_2, Q_1] \to Q_3$ and $Q_1 \to Q_2$ ($c_{1\to 2} = 1$ and $c_{12\to 3} = 1$). In this case, the three variables synchronize such that $\Delta I^R_{123\to 1} = \Delta I^R_{123\to 2} = \Delta I^R_{123\to 3} \neq 0$ (i.e., they can be interpreted as exact copies of each other). The results are shown in Figure S17(b).

A summary of the results provided by other methods is provided in Table S1. We note that CCM is the only method that can offer insights into the dynamics of the system, since the logistic maps analyzed in this section are given by a deterministic dynamical system where the coupled variables are part of the same manifold. The only case in which this method failed is the one of strong coupling between variables $[Q_1, Q_2] \to Q_3$, where the CCM did not converge to a value of one for a relatively high number of samples. This is the case where synergistic effects are important according to SURD and this might be playing a role in the identification of causalities from CCM.

CTE also provides consistent results for those cases in which the coupling between variables is inexistent or intermediate. However, when the coupling is strong the method completely cases. For these cases, SURD identifies

Case	CGC	CTE	CCM	PCMCI	SURD
No coupling	✓	✓	✓	✓	✓
One-way intermediate coupling	✗	✓	✓	✗	✓
One-way strong coupling	✗	✗	✓	✗	✓
Two-way strong coupling	✓	✓	✗†	✗	✓
Three-way strong coupling	✗	✗	✓	✗	✓

Table S1: Summary of the performance of the different methods for logistic maps with none to three coupled variables. †The value of the prediction skill from CCM does not converge to a value of one.

significant redundant causalities, which implies those variables are fully synchronized and act as the same variables. Therefore, introducing those variables into the conditioning set of CTE for a given variable hinders the identification of causalities. A similar conclusion can be drawn for CGC and PCMCI, which only identified consistent causal relationships for the case in which there is no coupling between variables and for $[Q_1, Q_2] \to Q_3$ in the CGC case.

S3.6 Coupled Rössler–Lorenz system

We study a coupled version of the Lorenz system [73] and the Rössler system [74]. The former was developed by Lorenz as a simplified model of a viscous fluid flow. Rössler proposed a simpler version of the Lorenz's equations in order to facilitate the study of its chaotic properties. The governing equations are

$$\frac{\mathrm{d}Q_1}{\mathrm{d}t} = -6[Q_2 + Q_3], \tag{S45a}$$
$$\frac{\mathrm{d}Q_2}{\mathrm{d}t} = 6[Q_1 + 0.2Q_2], \tag{S45b}$$
$$\frac{\mathrm{d}Q_3}{\mathrm{d}t} = 6\left[0.2 + Q_3[Q_1 - 5.7]\right], \tag{S45c}$$
$$\frac{\mathrm{d}Q_4}{\mathrm{d}t} = 10[Q_5 - Q_4], \tag{S45d}$$
$$\frac{\mathrm{d}Q_5}{\mathrm{d}t} = Q_4[28 - Q_6] - Q_5 + cQ_2^2, \tag{S45e}$$
$$\frac{\mathrm{d}Q_6}{\mathrm{d}t} = Q_4Q_5 - \frac{8}{3}Q_6, \tag{S45f}$$

where $[Q_1, Q_2, Q_3]$ correspond to the Rössler system and $[Q_4, Q_5, Q_6]$ to the Lorenz system. The coupling is unidirectional from the Rössler system to the Lorenz system from $Q_2 \to Q_5$ via the parameter c. This coupled system has previously been studied by [81] and [82]. We use this case to study the behavior of SURD in a continuous dynamical system when some of the variables are hidden. The observable variables are $\boldsymbol{Q} = [Q_1, Q_2, Q_5, Q_6]$. The system was integrated for $10^6 t_{\mathrm{ref}}$ where t_{ref} is the time for which $I(Q_1^+; Q_1)/I(Q_1; Q_1) = 0.5$. The time-lag selected for causal inference is $\Delta T \approx t_{\mathrm{ref}}$ and 50 bins per variable were used to partition the observed phase space.

The results for uncoupled systems ($c = 0$) are shown in Figure S18. The upper panel portrays typical trajectories of the systems. Unsurprisingly, SURD shows that both systems are uncoupled. Moreover, the unique, redundant and synergistic causal structure identified in the Rössler and Lorenz systems are consistent with structure of Equation (S45a). The causality leak is roughly 25% due to the unobserved variables.

The results for the coupled system ($c = 2$) are shown in Figure S19. The left panel shows how the trajectory of the Lorenz system is severely impacted by the coupling. The causalities in the Rössler system remain comparable to the uncoupled case besides some small redundancies and synergies due to the effect of unobserved variables. On the contrary, the causalities in the Lorenz system undergo deeper changes. This is evidenced by the emergence of multiple synergistic causalities involving Q_1 and Q_2. This effect is consistent with the coupling of both systems.

Among the other methods, only the CTE and CCM provide insight into the dynamics of the system, given that the system is given by a deterministic dynamical system where the group of variables $[Q_1, Q_2]$ and $[Q_5, Q_6]$ are part of the same manifold, respectively, in the uncoupled Rössler–Lorenz system and $[Q_1, Q_2, Q_5, Q_6]$ in the coupled Rössler–Lorenz case. However, CCM cannot clearly identify that the coupling is from Q_2 to Q_5, since the prediction skill of variables Q_1 and Q_2 using Q_5 and Q_6 is also very high, although lower than in the reverse case. For CTE, the direction of this coupling is properly identified. Finally, the methods CGC and PCMCI completely fail in both cases.

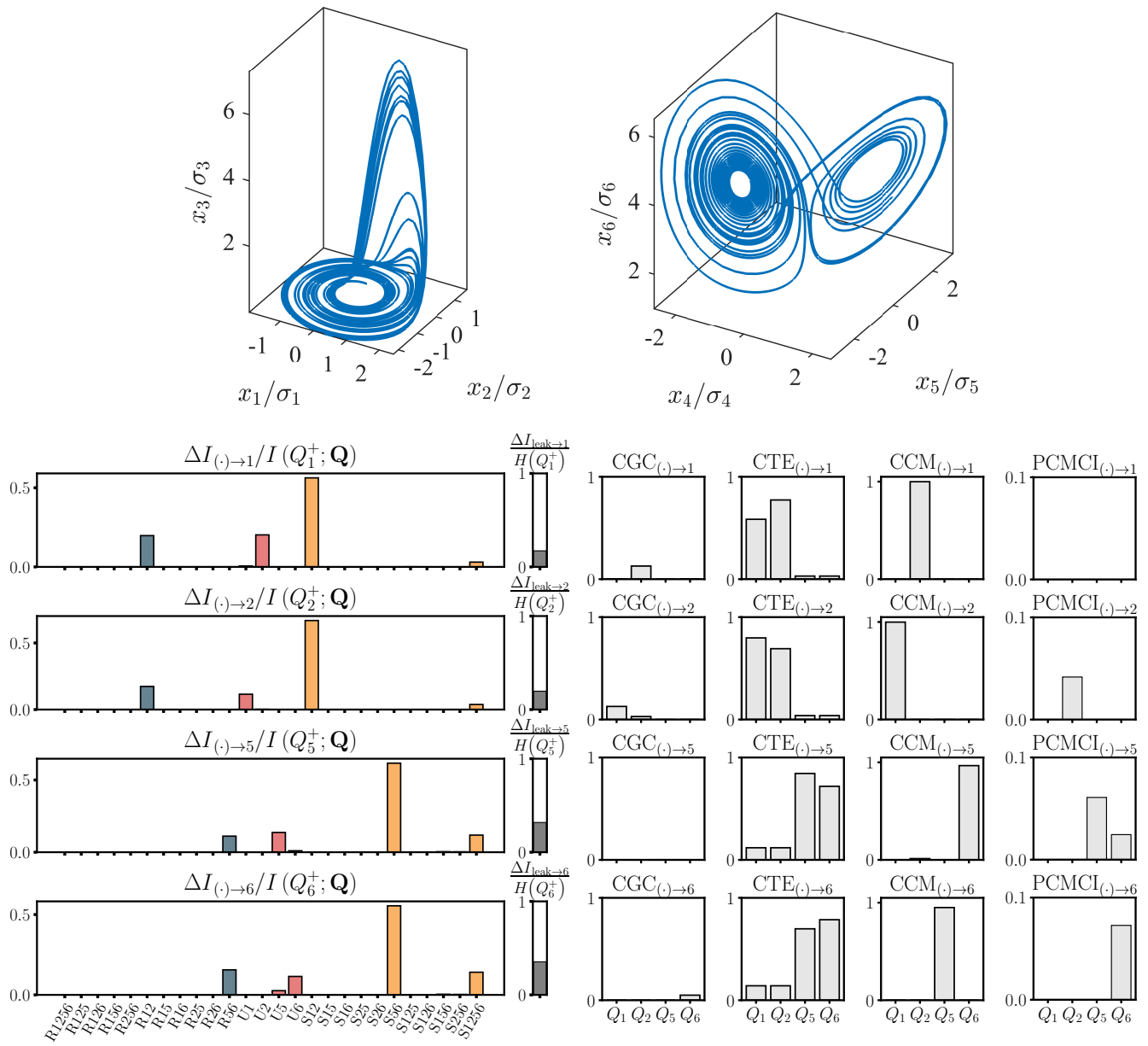


Figure S18: Uncoupled Rössler-Lorenz system ($c = 0$). The top panels show excerpts of the trajectories pertaining to Rössler systems $[Q_1, Q_2, Q_3]$ (left) and Lorenz system $[Q_4, Q_5, Q_6]$ (right). The bottom panels show the redundant (R), unique (U), and synergistic (S) causalities among $[Q_1, Q_2, Q_5, Q_6]$. The causality leak for each variable is also shown in the right-hand side bar. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

S3.7 Three interacting species

We analyze causality in a system of three interacting species. The case, proposed by Leng *et al.* [40], validates the Partial Cross Mapping (PCM) method –a variation of CCM that eliminates indirect causal influences. This benchmark served as an example where other methods, including CGC, CTE, and CCM, failed to detect causal links. The equations of the system are given by:

$$Q_1(n+1) = Q_1(n)\left[\alpha_1 - \alpha_1 Q_1(n) - \beta_{2\to 1} Q_2(n)\right] + \eta_1(n) \tag{S46}$$

$$Q_2(n+1) = Q_2(n)\left[\alpha_2 - \alpha_2 Q_2(n) - \beta_{1\to 2} Q_1(n) - \beta_{3\to 2} Q_3(n)\right] + \eta_2(n) \tag{S47}$$

$$Q_3(n+1) = Q_3(n)\left[\alpha_3 - \alpha_3 Q_3(n) - \beta_{1\to 3} Q_3(n)\right] + \eta_3(n) \tag{S48}$$

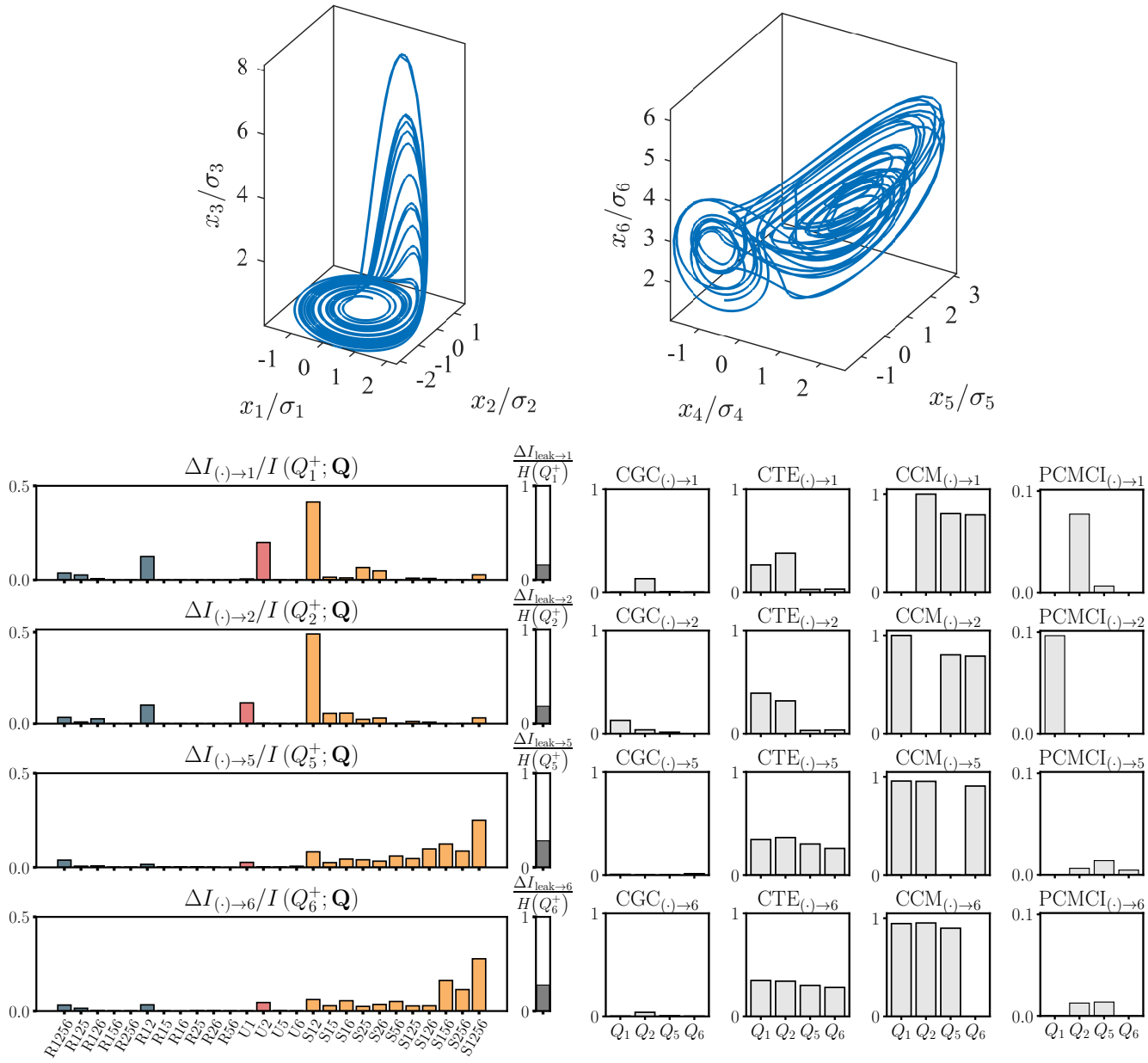
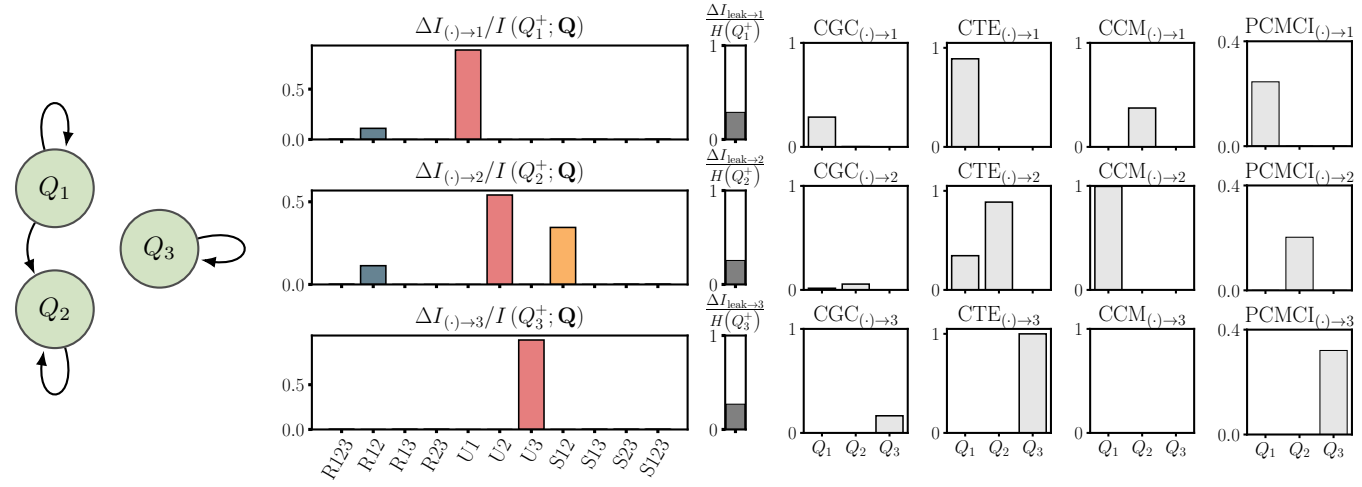


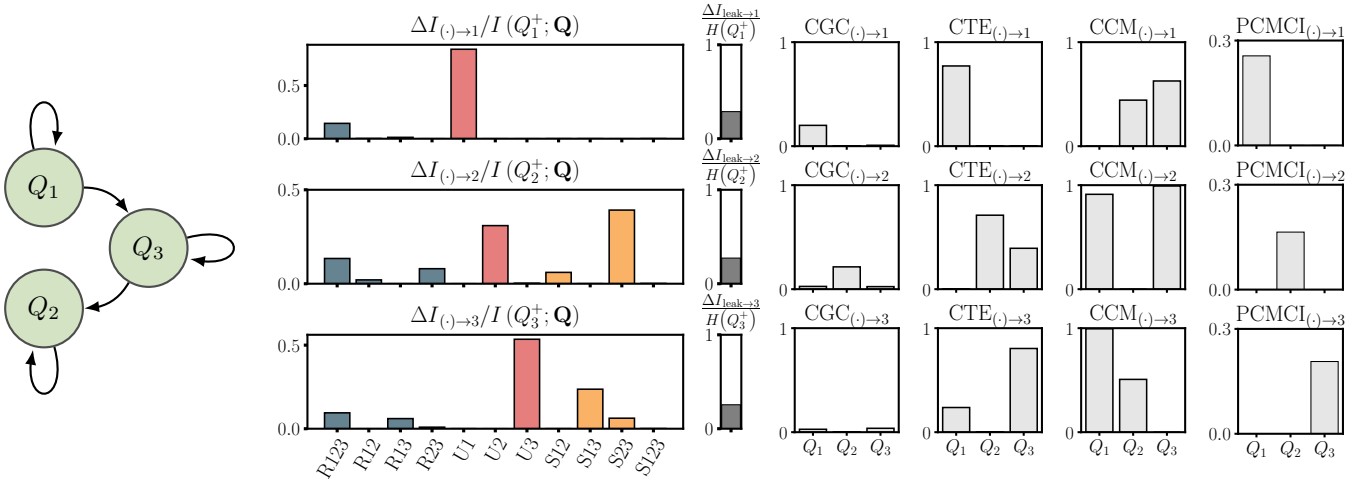
Figure S19: Coupled Rössler-Lorenz system ($c = 2$). The top panels show excerpts of the trajectories pertaining to Rössler systems $[Q_1, Q_2, Q_3]$ (left) and Lorenz system $[Q_4, Q_5, Q_6]$ (right). The bottom panels show the redundant (R), unique (U), and synergistic (S) causalities among $[Q_1, Q_2, Q_5, Q_6]$. The causality leak for each variable is also shown in the right-hand side bar. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

where $\alpha_1 = 3.6$, $\alpha_2 = 3.72$, $\alpha_3 = 3.68$, $\beta_{i\to j}$ denotes the coupling constants between variables, and η_i represents white noise with zero mean and a standard deviation of 0.005. Different choices of coupling parameters $\beta_{i\to j}$ can lead to various modes of interaction. We analyze the same combinations reported by Leng *et al.* [40], which represent three possible interaction structures among three species: fan-in ($Q_1 \to Q_2$), fan-out ($Q_1 \to Q_3 \to Q_2$), and cascading structures ($Q_1 \to Q_3 \to Q_2 \to Q_1$).

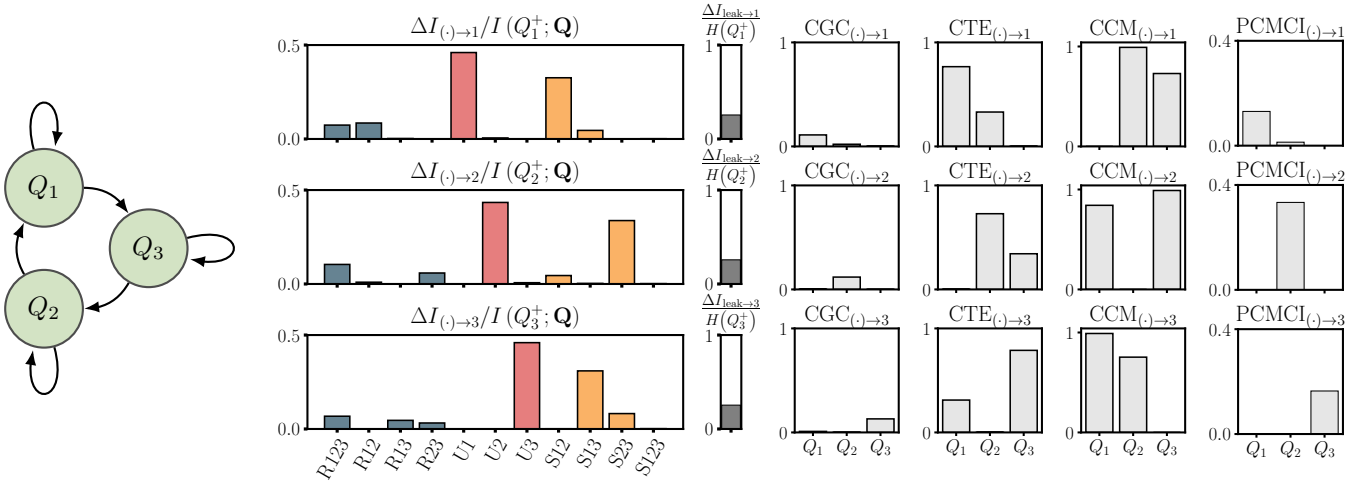
The results, shown in Figure S20, demonstrate that SURD identifies links consistent with the governing equations across all scenarios. CCM fails to identify correct causal links consistently as already reported by [40]. PCM (not shown) operates successfully; however, it does not provide information about the strength of self-causal links, a feature that SURD offers. CGC underperforms, while CTE can offer a good insight into the causal relationships between variables. Nonetheless, CTE cannot distinguish between unique and synergistic causalities. This distinction is useful



(a)



(b)



(c)

Figure S20: Three interacting species [40]. Redundant (R), unique (U) and synergistic (S) causalities for Q_1, Q_2, Q_3 in coupled logistic maps for (a) fan-in with $\beta_{1\to 2} = 0.35$, (b) fan-out with $\beta_{1\to 3} = \beta_{3\to 2} = 0.35$, and (c) cascading structures with $\beta_{1\to 3} = \beta_{3\to 2} = \beta_{2\to 1} = 0.35$. The results from CGC, CTE, CCM, and PCMCI are depicted on the right. CGC and CTE use the same normalization as SURD.

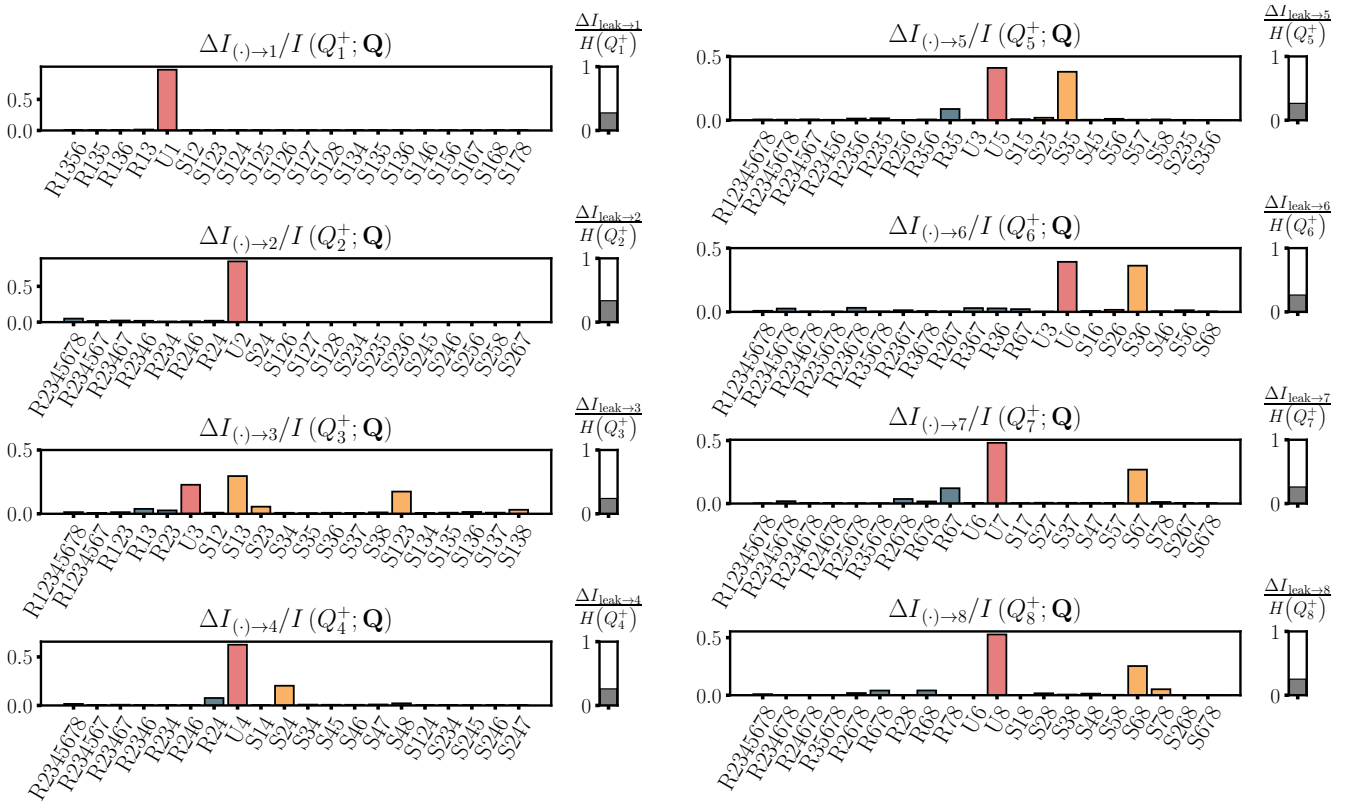


Figure S21: Eigth interacting species [40]. Redundant (R), unique (U) and synergistic (S) causalities for Q_i with $i = 1, \ldots, 8$. Only the top $N = 20$ contributions, satisfying the condition $\Delta I_{(\cdot)\to j}/I(Q_j^+;\boldsymbol{Q}) \geq 10^{-3}$, are represented for each variable. The gray bar is the causality leak.

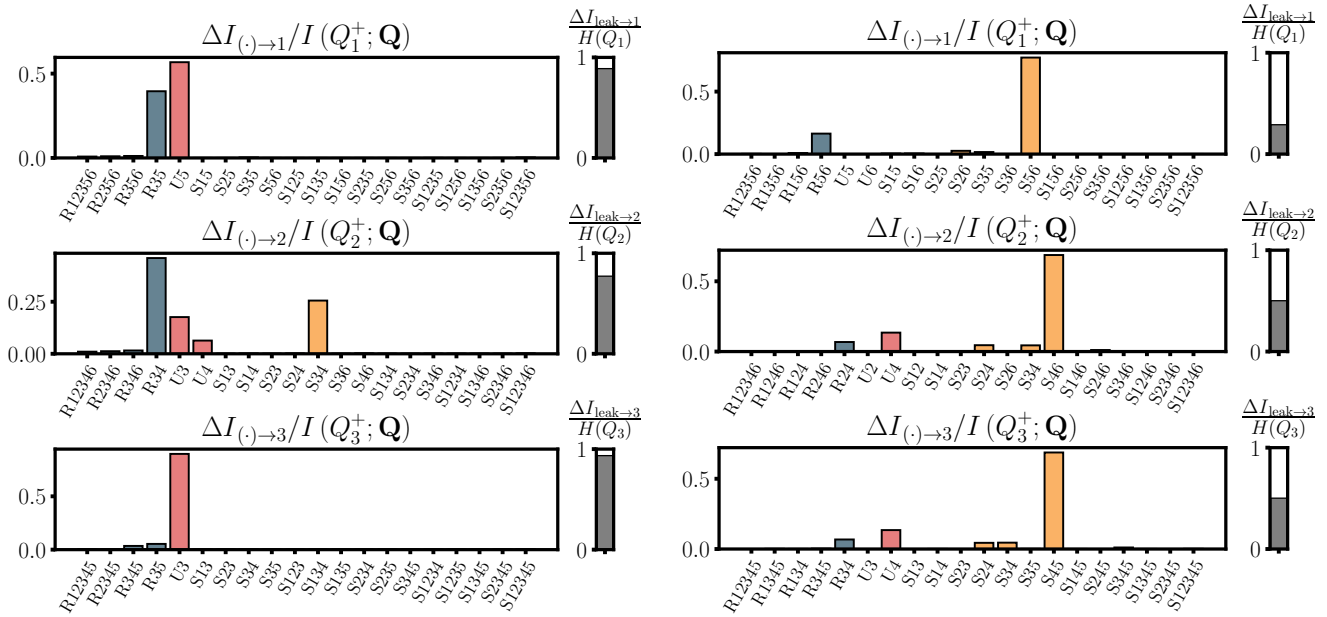
in these scenarios, as variables may be coupled through both cross-induced causal relationships and self-causal effects.

S3.8 Eight interacting species

We test the effectiveness of our method in a network model containing eight interacting species. This system was proposed by Leng *et al.* [40] to demonstrate the power of the partial cross-mapping (PCM) method in accurately reconstructing the underlying causal networks from multivariate time series of high dimensionality. The equations of the system are given by:

$$Q_1(n+1) = Q_1(n)\,[3.9 - 3.9Q_1(n)] + \eta_1(n), \tag{S49a}$$
$$Q_2(n+1) = Q_2(n)\,[3.5 - 3.5Q_2(n)] + \eta_2(n), \tag{S49b}$$
$$Q_3(n+1) = Q_3(n)\,[3.62 - 3.62Q_3(n) - 0.35Q_1(n)] + \eta_3(n), \tag{S49c}$$
$$Q_4(n+1) = Q_4(n)\,[3.75 - 3.75Q_4(n) - 0.35Q_2(n)] + \eta_4(n), \tag{S49d}$$
$$Q_5(n+1) = Q_5(n)\,[3.65 - 3.65Q_5(n) - 0.35Q_3(n)] + \eta_5(n), \tag{S49e}$$
$$Q_6(n+1) = Q_6(n)\,[3.72 - 3.72Q_6(n) - 0.35Q_3(n)] + \eta_6(n), \tag{S49f}$$
$$Q_7(n+1) = Q_7(n)\,[3.57 - 3.57Q_7(n) - 0.35Q_6(n)] + \eta_7(n), \tag{S49g}$$
$$Q_8(n+1) = Q_8(n)\,[3.68 - 3.68Q_8(n) - 0.35Q_6(n)] + \eta_8(n), \tag{S49h}$$

where $\eta_i(n)$ terms for $i = 1, \ldots, 8$, are white noise of zero mean and standard deviation of 0.005. The results provided in Figure S21 show the robustness of SURD to reconstruct all the causal links of the system according to their governing equations. Additionally, SURD is able to identify those variables that individually and synergistically cause the future of the target variables through unique and synergistic causalities, respectively. This validation case demonstrates that SURD is able to identify causal relationships in a setting with a larger number of variables.



(a) (b)

Figure S22: System with instantaneous causal dependencies for (a) mediator and (b) synergistic collider variables. Results from SURD with redundant (R), unique (U) and synergistic (S) causalities in blue, red and yellow, respectively. The notation employed is such that the present variables $Q_4 = Q_1^+$, $Q_5 = Q_2^+$ and $Q_6 = Q_3^+$. The gray bar is the causality leak.

S3.9 System with contemporaneous causal dependencies

We evaluate the performance of SURD in scenarios with contemporaneous causal links. In these cases, the vector of observed variables can also include variables at time $t + \Delta T$, excluding the target variable. For a system with N variables and target Q_i^+, the vector of observables is

$$\boldsymbol{Q} = [Q_1, Q_2, \ldots, Q_N, Q_1^+, \ldots, Q_{i-1}^+, Q_{i+1}^+, \ldots, Q_N^+]. \tag{S50}$$

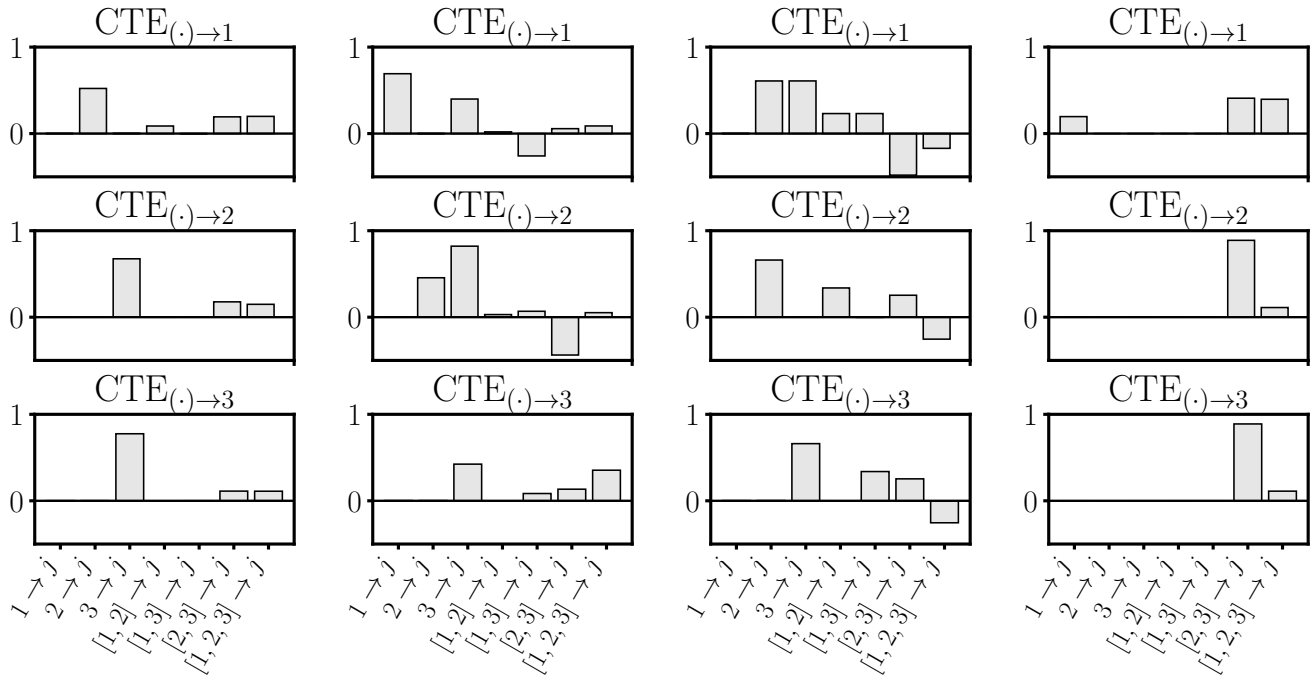
Note that Q_i^+ cannot be included in the vector of observed variables, since doing so would already reveal all the information about the target variable. For example, to calculate the causalities to Q_1^+ in a system with three variables, the vector of observed variables is $\boldsymbol{Q} = [Q_1, Q_2, Q_3, Q_2^+, Q_3^+]$. For simplicity, we use the notation $Q_4 = Q_1^+$, $Q_5 = Q_2^+$ and $Q_6 = Q_3^+$.

We use as test beds the systems with mediator and synergistic variables depicted in Figures 2 and 4. For the system with mediator variables, we introduce an contemporaneous dependence of Q_2 on Q_1, i.e., $Q_2^+ \to Q_1^+$. The equations of the system with mediator variables can be defined as follows:

$$Q_1(n+1) = \sin\left[Q_2(n+1)\right] + 0.01W_1(n) \tag{S51a}$$
$$Q_2(n+1) = \cos\left[Q_3(n)\right] + 0.01W_2(n) \tag{S51b}$$
$$Q_3(n+1) = 0.5Q_3(n) + 0.1W_3(n). \tag{S51c}$$

Figure S22(a) shows how SURD can identify the contemporaneous causal dependency of Q_2^+ on Q_1^+ through the unique causality $\Delta I^U_{5\to 1}$, where the index 5 refers to variable Q_2^+. For variables Q_2^+ and Q_3^+, the most relevant causalities are consistent with their dependencies on Q_3. However, a new synergistic causality, $\Delta I^S_{34\to 1}$, emerges in Q_2^+ due to its contemporaneous dependence on Q_1^+.

We also test the system with synergistic collider modified to include contemporaneous links. In this case, the



(a) $Q_3 \to Q_2 \to Q_1$ (b) $Q_1 \leftarrow Q_3 \to Q_2$ (c) $(Q_2, Q_3) \to Q_1$ (d) $Q_2 \equiv Q_3 \to Q_1$

Figure S23: Results from the multivariate version of the conditional transfer entropy (CTE) [45] for the systems with (a) mediator, (b) confounder, (c) synergistic collider, and (d) redundant collider variables. Further details about these systems are provided in the main text.

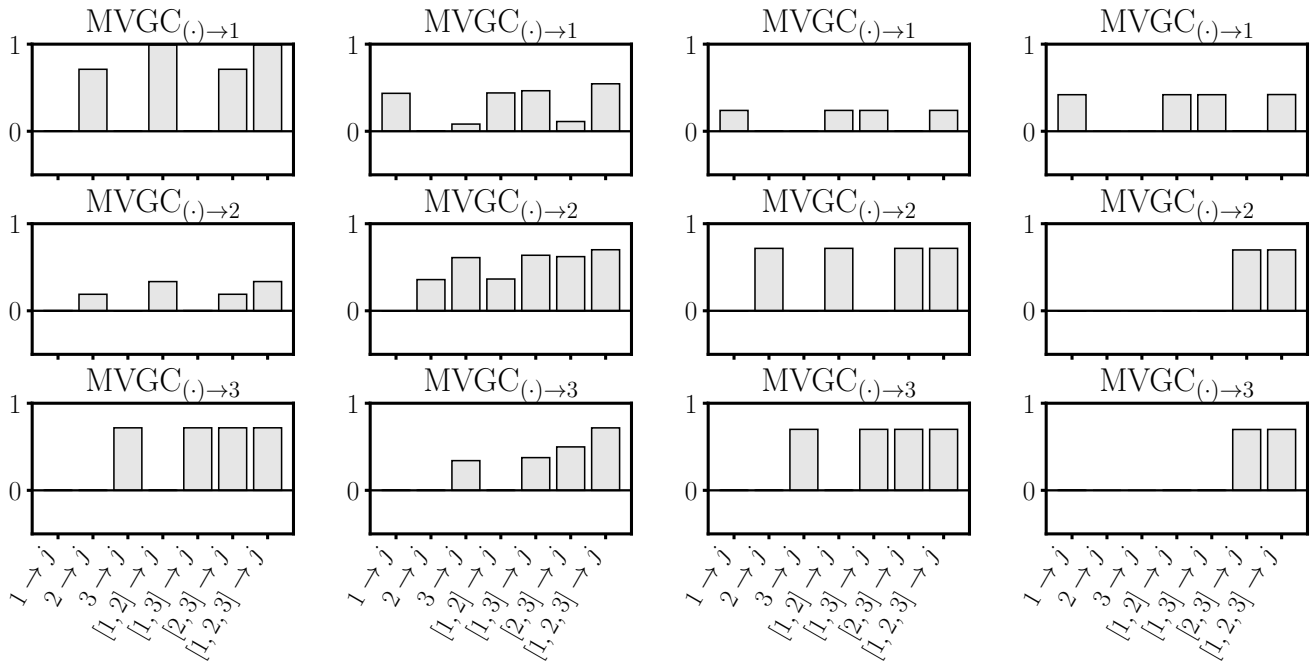
equations are given by:

$$Q_1(n+1) = \sin\left[Q_2(n+1)Q_3(n+1)\right] + 0.001W_1(n) \tag{S52a}$$
$$Q_2(n+1) = 0.5Q_2(n) + 0.1W_2(n) \tag{S52b}$$
$$Q_3(n+1) = 0.5Q_3(n) + 0.1W_3(n). \tag{S52c}$$

Figure S22(a) shows that the most relevant causality for Q_1^+ is the component $\Delta I^S_{56\to 1}$, where indices 5 and 6 represent variables Q_2^+ and Q_3^+, respectively. Furthermore, as the dependency between them is instantaneous and these variables are part of the observed vector, it is possible to write $Q_2^+ = f(Q_1^+, Q_3^+)$ and $Q_3^+ = f(Q_1^+, Q_2^+)$, indicating that the components $\Delta I^S_{45\to 2}$ and $\Delta I^S_{45\to 3}$ also play a significant role in determining the variables Q_2^+ and Q_3^+, respectively.

S3.10 Causality from combination of variables in mediator, confounder, and collider cases

In this section, we compare the redundant and synergistic causalities from SURD with the multivariate versions of CTE and CGC, referred to as MCTE and MVCG, respectively. The results are calculated for systems with mediator, confounder, synergistic collider, and redundant collider variables. Although MCTE and MVCG can quantify causality from combinations of variables (e.g., $[Q_1, Q_2, Q_3, \dots] \to Q_j$), we show here that the results are not easily interpretable or may be directly erroneous. Figure S23 shows how MCTE yields negative values of causality for cases in which synergistic effects play a major role, such as systems with confounder and synergistic collider variables. For example, in the synergistic collider $(Q_2, Q_3) \to Q_1$, $\text{MCTE}_{23\to 1}$ is strongly negative. The results from MVGC (Figure S24) are constrained to be nonnegative; however, MVGC does not provide information about the redundant or synergistic nature of the interactions. This makes it very challenging to interpret whether the identified links due to combined variables are consistent with the relationships between variables. Returning to the case of the synergistic collider $(Q_2, Q_3) \to Q_1$, SURD correctly detects $\Delta I^S_{23\to 1} > 0$ as the main causality to Q_1, whereas MVCG identifies $\text{MVGC}_{123\to 1}$, $\text{MVGC}_{12\to 1}$, and $\text{MVGC}_{13\to 1}$ as all equally important, while $\text{MVGC}_{23\to 1}$ is zero.



(a) $Q_3 \to Q_2 \to Q_1$ (b) $Q_1 \leftarrow Q_3 \to Q_2$ (c) $(Q_2, Q_3) \to Q_1$ (d) $Q_2 \equiv Q_3 \to Q_1$

Figure S24: Results from the multivariate version of the multivariate Granger causality (MVGC) [28] for the systems with (a) mediator, (b) confounder, (c) synergistic collider, and (d) redundant collider variables. Further details about these systems are provided in the main text.

S3.11 Causal graphs for PCMCI in multivariate systems

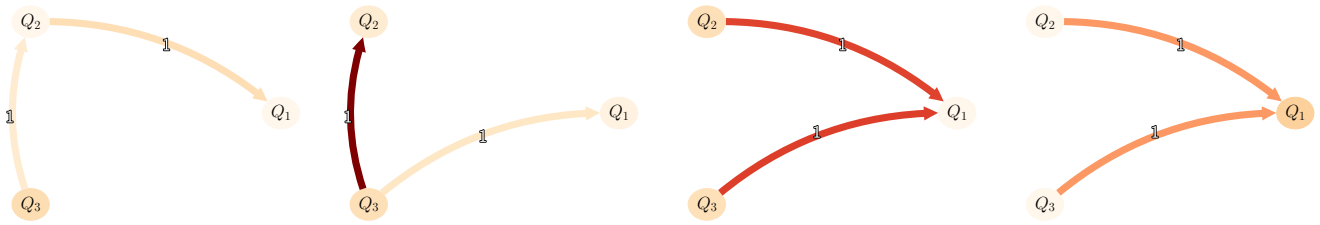
The primary outcome of the PCMCI algorithm for the mediator, confounder, and collider is illustrated through the causal graphs in Figure S25. In this visual representation, only the links that are statistically significant at a specified threshold level are shown. Additionally, a measure of the causal strength is included for both cross-causal and self-causal connections. This visualization facilitates a clear evaluation of the consistency of the causality analysis with respect to the functional relationships of the systems. Among all cases, the redundant collider stands out as the case where the summary graph lacks coherence. In this particular scenario, the approach identifies a fully interconnected graph, failing to correctly pinpoint the effects of the duplicated variables Q_2 and Q_3 on Q_1. This issue arises from incorporating duplicated variables into the conditional set of variables. Additionally, in the case of the confounder variable, although the causal connections across variables align with the interdependencies within the system, the strength of the self-causal connections for variables Q_1 and Q_2 is weak.

Figure S26(a) shows the summary graph of the causality analysis from PCMCI for the turbulent energy cascade (see Figure 6 in the main text). The most relevant causal links in the graph are $\langle \Sigma_1 \rangle \to \langle \Sigma_2 \rangle^+$ and $\langle \Sigma_1 \rangle \to \langle \Sigma_1 \rangle^+$. The other detected relationships, except for $\langle \Sigma_2 \rangle \to \langle \Sigma_1 \rangle^+$, are consistent with the hypothesis of forward propagation of energy in turbulence, which have very weak causal strength. The weak causalities can be attributed to the high importance of redundant causalities for variables other than $\langle \Sigma_1 \rangle$, as shown in the results from SURD in the main text.

Additionally, the results for the Rössler-Lorenz system discussed in Figures S18 and S19 are provided in graph format in Figures S26(b) and S26(c). For both cases, PCMCI fails to identify causal links that are consistent with the equations of the systems. Specifically, in the uncoupled Rössler-Lorenz system (see Figure S18), the method cannot identify a coupling between variables Q_1 and Q_2. In the coupled system (see Figure S19), the links provided by PCMCI are more representative of the real connections between variables. However, the link $Q_5 \to Q_1^+$ is inconsistent with the coupling in the Rössler-Lorenz system, which occurs through $Q_2 \to Q_5^+$.

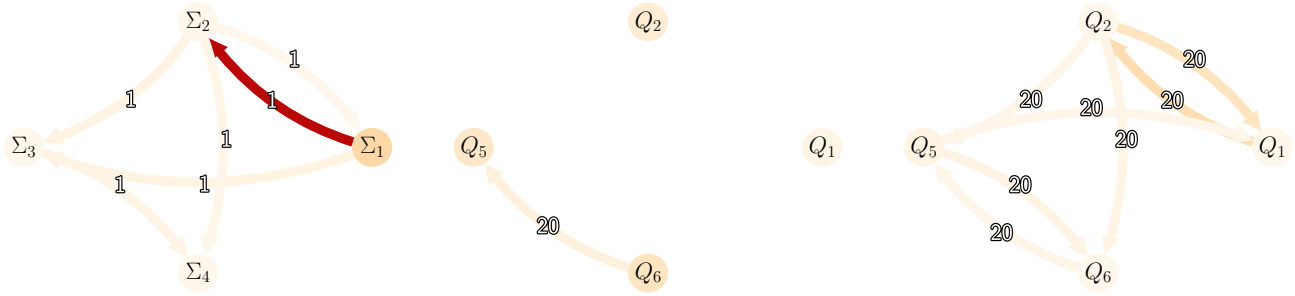
S3.12 PCMCI for different independence tests

We analyze the variability in the results of PCMCI for different independence tests. The optimal confidence interval α_{PC} for PCMCI is selected during the initial condition selection phase (PC phase) based on the Akaike Information



(a) Mediator variable (b) Confounder variable (c) Synergistic collider (d) Redundant collider

Figure S25: Summary graph of the causality analysis performed using the PCMCI method for the systems with (a) mediator, (b) confounder, (c) synergistic collider, and (d) redundant collider variables. Cross-causality is quantified with the color intensity of the links, whereas self-causality is quantified with the color intensity of the variables nodes. Both quantities range from 0 (light orange) to 0.5 (dark orange). The indices in the arrows represent the time lag at which the causal links are identified.



(a) Energy cascade (b) Uncoupled Rössler-Lorenz system (c) Coupled Rössler-Lorenz system

Figure S26: Summary graph of the causality analysis performed using the PCMCI method for the (a) turbulent energy cascade, (b) uncoupled Rössler-Lorenz system, and (c) coupled Rössler-Lorenz system. Cross-causality is quantified with the color intensity of the links, whereas self-causality is quantified with the color intensity of the variables nodes. Both quantities range from 0 (light orange) to 0.5 (dark orange). The indices in the arrows represent the time lag at which the causal links are identified for panels (b) and (c). In panel (a), an index of one is shown for all links, although the time lag for identification of causality used in the analysis is different for each variable. These values are consistent with the ones provided in the main text.

criterion [126] from a default list of values, i.e., $\alpha_{\text{PC}} = [0.05, 0.1, 0.2, 0.3, 0.4, 0.5]$. Next, the significance level for the momentary conditional independence phase (MCI phase) is set to $\alpha_{\text{MCI}} = 0.01$ to obtain the causal graph and strengths. The independence tests analyzed are partial correlation (ParCorr), robust partial correlation (rParCorr), Gaussian process regression and a distance correlation (GPDC), and conditional mutual information with a k-nearest neighbor estimator (CMI). For a detailed discussion of these independence tests and their assumptions, the reader is referred to Ref. [23].

The summary of the results for the other tests is compiled in Table S2. The results for the CMI test are the same as those reported in the main text. The table shows that there is strong variability in the conclusions depending on the independence test and the parameters used. The CMI test with the k-NN estimator might not be the best choice in systems with fully deterministic dynamics, such as the Lotka-Volterra prey-predator model and nonlinear logistic difference systems. In these cases, ParCorr and rParCorr offer a more consistent estimation of the causal graph. Moreover, the GPDC test is the only test that identifies the causal relationships in the system with confounder variables.

S4 Sensitivity of SURD

S4.1 Sensitivity of SURD to number of samples

SURD relies on the estimation of probability distributions, which becomes computationally intractable as the number of dimensions increases. To address this limitation, we use the concept of transport maps[120] to estimate high-

Case	ParCorr	rParCorr	GPDC	CMI
Mediator variable	✗	✗	✓	✓
Confounder variable	✓	✓	✓	✓
Synergistic collider variable	✗	✗	✓	✓
Redundant collider variable	✗	✗	✗	✗
Lotka-Volterra prey-predator model [68, 69]	✓	✓	✓	✗
Three-interacting species system [40]	✗	✓	✗	✗
Moran effect model [70]	✓	✓	✗	✓
One-way coupling nonlinear logistic difference system [36]	✓	✗	✗	✗
Two-way coupling nonlinear logistic difference system [36]	✗	✗	✓	✗
Stochastic system with linear dependencies [71]	✓	✓	✓	✓
Stochastic system with non-linear dependencies [35]	✓	✗	✗	✓
Synchronization of two variables in logistic maps [72]	✓	✓	✗	✓
Synchronization of three variables in logistic maps [72]	✗	✗	✗	✗
Uncoupled Rössler-Lorenz system [73, 74]	✗	✗	✗	✓
One-way coupled Rössler-Lorenz system [73, 74]	✗	✗	✗	✓

Table S2: Summary of the performance of different independence tests for PCMCI. The markers ✓ and ✗ denote consistent and inconsistent identification of causal links, respectively, according to the functional dependency of variables within the system. The independence tests considered are partial correlation (ParCorr), robust partial correlation (rParCorr), Gaussian process regression and a distance correlation (GPDC) and conditional mutual information with a k-nearest neighbor estimator (CMI).

dimensional probability distributions. The method relies on the estimation of a parsimonious and interpretable nonlinear transformation from a complex distribution $\pi(\boldsymbol{x})$ defined by the set of samples of the vector of observed variables to a simpler reference distribution $\eta(\boldsymbol{z})$, e.g. a Gaussian distribution $\mathcal{N}(0, I)$, as shown in Figure S27. Although there are infinitely many transformations that link these distributions, if π is absolutely continuous with respect to η, there exists a unique lower triangular and monotone function $S : \mathbb{R}^d \to \mathbb{R}^d$ that pushes forward π to η. This type of transformation is highly attractive since it provides a map that is differentiable along with a differentiable inverse[120].

This transformation is a mere approximation of the real distribution, which depends on the class of functions chosen. In this study, we used sixth-order polynomials to estimate the transport map. Figure S27 shows the results for the evolution of the relative error with the number of samples for redundant, unique, and synergistic causalities from the confounder example in Figure 3. We set the results estimated using the transport map with $N = 10^6$ samples as a reference. Using this method, we can obtain results with an error significantly lower than 5% by using only a number of samples in the order of a few hundred. Furthermore, if we compare the reference results with those obtained using the binning method with $N = 10^8$ samples, we obtain differences lower than 3% for all causalities. Therefore, the approximation from the transport map method allows us to estimate SURD causalities with relatively high accuracy when the number of samples is low. A more extensive analysis of the impact of sample size, partition refinement, and order of the polynomials on the calculation of SURD is provided in the Supplementary Materials for the binning and the transport map methods.

S4.2 Sensitivity of SURD to partition refinement

We investigate the sensitivity of SURD to the number of samples (N_{samples}) and number of bins (N_{bins}) used to partition the variables in situations where the binning method is employed to estimate the probability distributions. The Lorenz system is used as a testbed:

$$\frac{\mathrm{d}Q_1}{\mathrm{d}t} = 10[Q_2 - Q_1], \tag{S53a}$$

$$\frac{\mathrm{d}Q_2}{\mathrm{d}t} = Q_1[28 - Q_3] - Q_2, \tag{S53b}$$

$$\frac{\mathrm{d}Q_3}{\mathrm{d}t} = Q_1 Q_2 - \frac{8}{3} Q_3. \tag{S53c}$$

The system was integrated over time to collect $N_{\text{samples}} = 5 \times 10^3, 5 \times 10^4, 5 \times 10^5$, and 5×10^8 events after transients. Probability distributions were calculated using uniform bins with $N_{\text{bins}} = 10, 50, 100$, and 200 per variable. Our

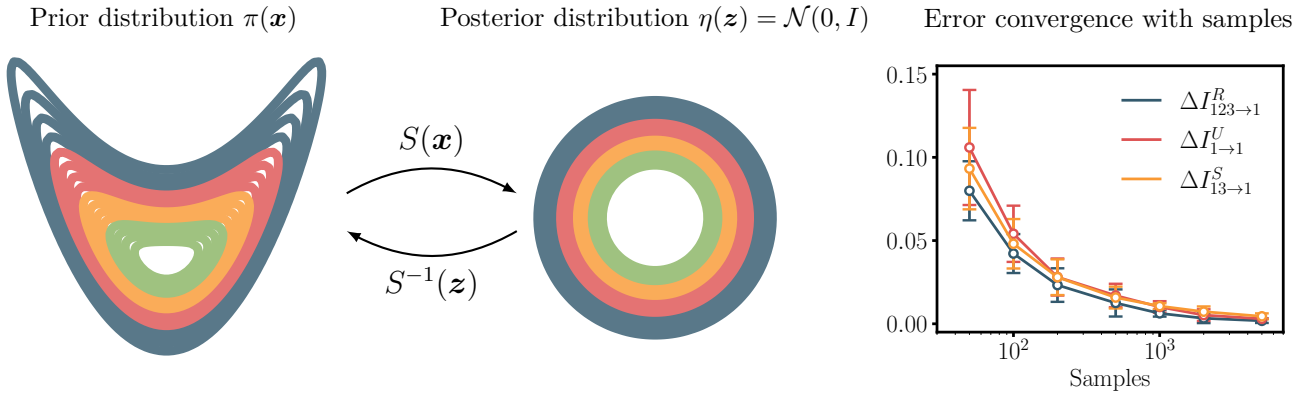


Figure S27: (Left) Schematic representation of the construction process of the transport map S that transforms an arbitrary distribution $\pi(\boldsymbol{x})$ to a Gaussian distribution $\eta(\boldsymbol{z}) = \mathcal{N}(0, I)$, where $\boldsymbol{x}$ and $\boldsymbol{z}$ denote samples from each of the distributions. (Right) Evolution of the relative error with the number of samples for three different causalities from the confounder validation case. The results estimated using the transport map with $N = 10^6$ samples and sixth-order polynomials are set as the ground truth. The error bars represent the variance of the relative error using different sets of random samples, which in total are equivalent to the number of samples used in the reference case.

primary focus is on causalities to Q_1, but the conclusions drawn also apply to Q_2 and Q_3.

The sensitivity to N_{samples} is displayed in Figure S28(a), where N_{samples} varies while maintaining $N_{\text{bins}} = 50$ constant. For $N_{\text{samples}} > 5 \times 10^3$, the changes in causality remain within a few percentage points of difference. The sensitivity to the size of the partition is assessed in Figure S28(b), where N_{bins} varies while N_{samples} is held constant at $N_{\text{samples}} = 5 \times 10^5$. The causalities exhibit quantitative resemblance for all partitions, with the exception of $N_{\text{bins}} = 10$, which may be too coarse to capture the continuous dynamics of the variables.

S4.3 Sensitivity of SURD to polynomia order

We investigate the sensitivity of SURD to the number of samples (N_{samples}) and the order of polynomials (N_{order}) used to estimate the probability distributions using transport maps. The case considered for testing the results is the system with a confounder variable, as defined in the main text, but with bimodal stochastic forcing. The latter follows a bimodally distributed random variable $W = \alpha X + (1 - \alpha)Y$, where $X = \mathcal{N}(-2, 1)$ and $Y = \mathcal{N}(2, 1)$ are unimodal random variables following a Gaussian distribution and $\alpha = 0.3$ represents the mixture coefficient. The system was integrated over time to collect $N_{\text{samples}} = 5 \times 10^1, 5 \times 10^2, 5 \times 10^3$, and 5×10^4 events after transients. The probability distributions were estimated using the transport map method with $N_{\text{order}} = 4, 6, 8$, and 10. Our primary focus is on causalities to Q_1, but the conclusions drawn also apply to Q_2 and Q_3.

The impact of the number of samples on the causalities estimated with the transport map method is depicted in Figure S29(a), where N_{samples} is varied while keeping $N_{\text{order}} = 6$ constant. For $N_{\text{samples}} > 5 \times 10^2$, the changes in causality only differ within a few percent and are limited to synergistic contributions, which are linked to higher-order probability distributions. Figure S29(b) assesses the sensitivity to polynomial order, with N_{order} varying while N_{samples} is fixed at $N_{\text{samples}} = 200$. The causalities show similar quantitative values for all polynomial orders. Once again, the slight percentage differences noticed when increasing the polynomial order are associated with synergistic causalities. In conclusion, transport mapping theory provides an accurate estimation of the probability distribution when the number of samples is small or the number of variables is high. However, these estimates are only approximations of the true distribution, which might become more challenging in more complex distributions.

S5 Effect of non-separability of the variables

One of the prevailing weaknesses in some of the previous methods for causal inference arises from the non-separability of the variables [36]. The issue is a consequence of Takens' embedding theorem, which states that under the right conditions, the dynamics of a system can be captured by embedding a sequence of past observations into a higher-dimensional space. In such cases, the future of a variable can be fully forecasted using only its own past, without the need for any other variables. Consequently, methods for causal inference based on predictability, such as Granger causality, might miss causal connections when including past observations into the model. For example, consider

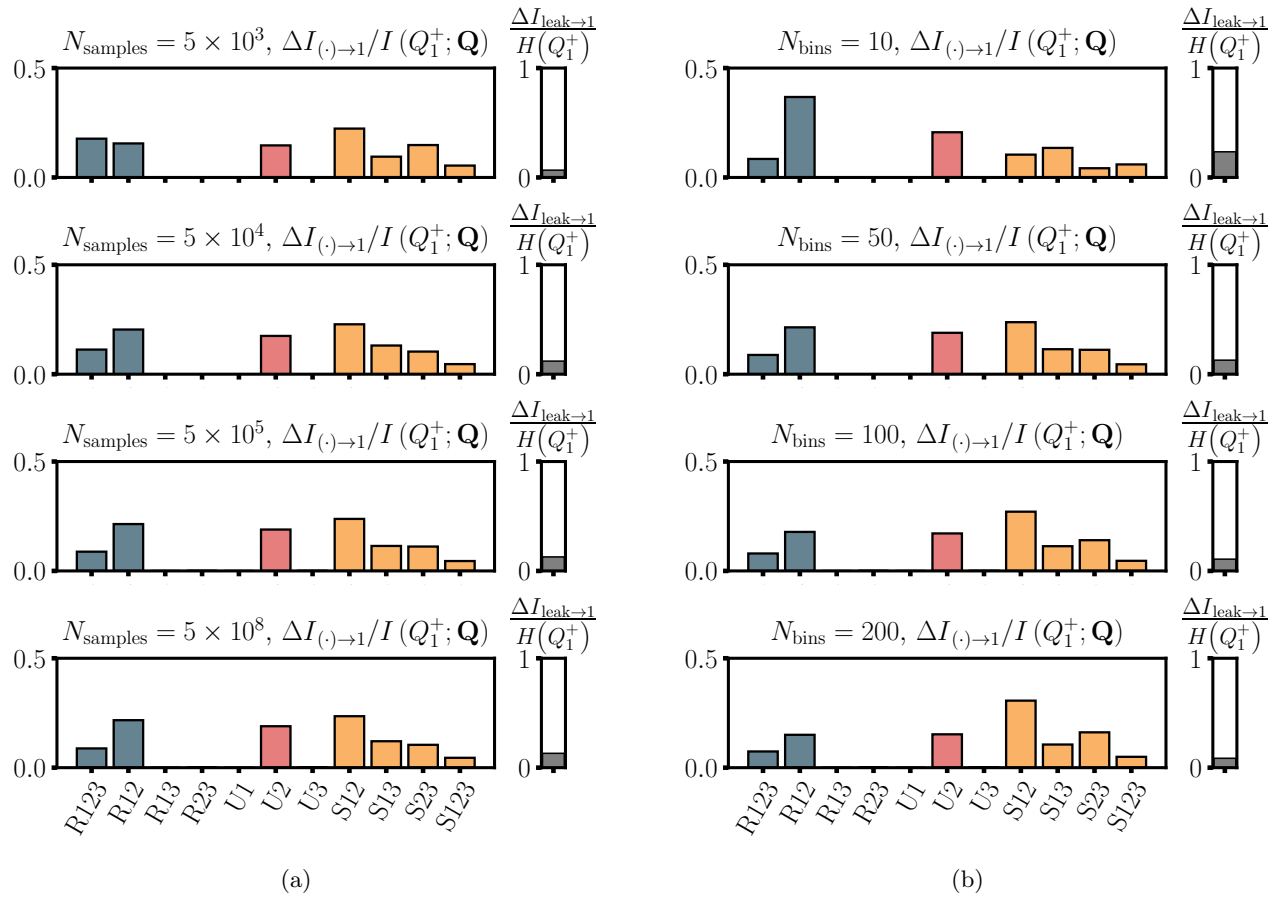


Figure S28: Sensitivity of SURD in the Lorenz system for (a) number of samples N_{samples} used to estimate the probability distributions using the binning method for $N_{\text{bins}} = 50$ held constant and (b) the number of bins N_{bins} used to partition values of the variables for $N_{\text{samples}} = 10^5$.

a system where $Q_1 \to Q_2$. By the Takens' embedding theorem, Q_2 could be forecasted using only its own past regardless of Q_1, leading to erroneous conclusions in Granger causality.

Takens' embedding theorem can also be interpreted within the framework of information theory [46]. If the variable Q_1 is causal to Q_2, then part of the past information from Q_1 is encoded into Q_2. Thus, from an information-theoretic viewpoint, non-separability arises from the flow of information among interacting variables. SURD is less susceptible to the issue of non-separability as it monitors all transfers of information among variables within the system, even if redundant.

Here, we employ the example introduced by Sugihara *et al.* [36] to illustrate the robustness of SURD under the effect of multiple time lags with non-separable variables. The system is given by:

$$Q_1(n+1) = Q_1(n)\left[r_1 - r_1 Q_1(n) - \beta_{2\to 1} Q_2(n)\right], \tag{S54a}$$
$$Q_2(n+1) = Q_2(n)\left[r_2 - r_2 Q_2(n) - \beta_{1\to 2} Q_1(n)\right], \tag{S54b}$$

where the coupling from Q_2 to Q_1 is controlled through $\beta_{2\to 1}$ and the coupling from Q_1 to Q_2, through $\beta_{1\to 2}$. In this simple system, we can recover algebraically the influence of Q_1 on Q_2 using $Q_2(n+1)$ and $Q_2(n)$ (and vice versa):

$$\beta_{2\to 1} Q_2(n) = 1 - Q_1(n) - \frac{Q_1(n+1)}{r_1 Q_1(n)}, \tag{S55a}$$
$$\beta_{1\to 2} Q_1(n) = 1 - Q_2(n) - \frac{Q_2(n+1)}{r_2 Q_2(n)}. \tag{S55b}$$

We can substitute Equation (S55a) into (S54b) and obtain an expression for $Q_2(n)$ as a function of $Q_1(n)$ and

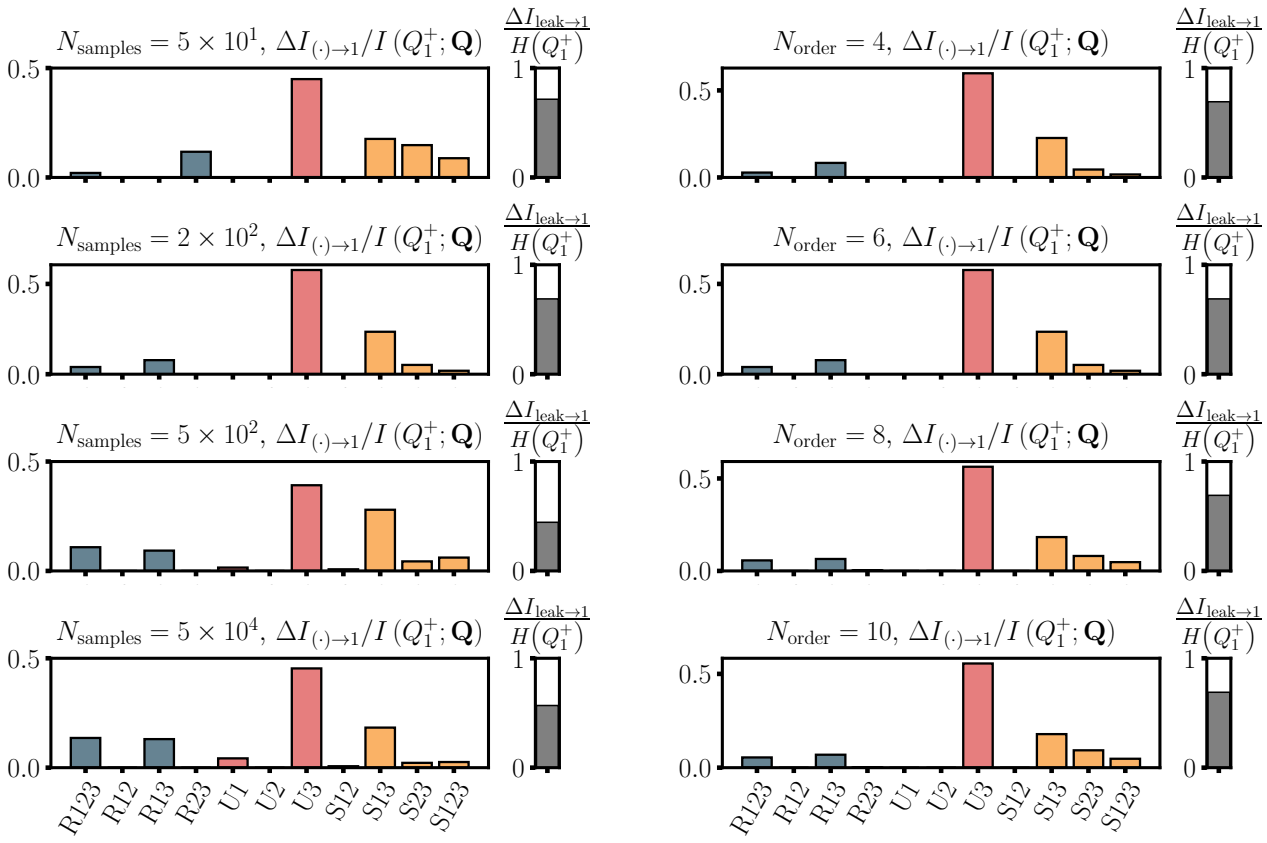


Figure S29: Sensitivity of causality in the system with confounder variables and bimodal stochastic forcing for (a) number of samples N_{samples} used to estimate the approximated probability distributions with the transport map method for $N_{\text{order}} = 6$ held constant and (b) the order of polynomials N_{order} used to partition values of the variables for $N_{\text{samples}} = 200$. The causalities are ordered from left to right according to $N^{\alpha}_{\boldsymbol{i}\to 1}$.

$Q_1(n-1)$:

$$Q_2(n) = \frac{r_2}{\beta_{2\to 1}}\left[(1-\beta_{1\to 2}Q_1(n-1))\left(1-Q_1(n-1)-\frac{Q_1(n)}{r_1 Q_1(n-1)}\right)-\frac{1}{\beta_{2\to 1}}\left(1-Q_1(n-1)-\frac{Q_1(n)}{r_1 Q_1(n-1)}\right)^2\right]. \tag{S56}$$

Introducing Equation (S56) into (S54a), we obtain an expression for Q_1 that is exclusively a function of its own past, i.e. $Q_1(n)$ and $Q_1(n-1)$:

$$Q_1(n+1) = f(Q_1(n), Q_1(n-1)). \tag{S57}$$

Methods for causal inference based on the predictability of Q_1 might incorrectly conclude that Q_2 does not cause Q_1 if the values $Q_1(n)$ and $Q_1(n-1)$ are included in the predictive model. To address this, we assess the causal connections to Q_1 and Q_2 using SURD and a non-linear version of CGC. We use a non-linear implementation of CGC because its linear counterpart failed in all considered scenarios, which does not allow us to demonstrate the problem of non-separability. The non-linear CGC consists of an artificial neural network (ANN) trained to predict the target variables $Q_1(n+1)$ and $Q_2(n+1)$, given different sets of past instances of Q_1 and Q_2. The model for Q_1 (similarly for Q_2) is

$$Q_1(n+1) = \text{ANN}_1(\boldsymbol{Q}_1) + \hat{\varepsilon}(n+1), \tag{S58a}$$
$$Q_1(n+1) = \text{ANN}_{12}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) + \varepsilon(n+1), \tag{S58b}$$

where vector of observables is defined as $\boldsymbol{Q} = [\boldsymbol{Q}_1, \boldsymbol{Q}_2]$ with $\boldsymbol{Q}_1 = \left[Q_1^n, Q_1^{n-1}, \cdots, Q_1^{n-\Delta n}\right]$ (similarly for $\boldsymbol{Q}_2$) and Δn is the maximum lag considered. Note that, from the point of view of SURD, $\boldsymbol{Q}$ only contains two variables

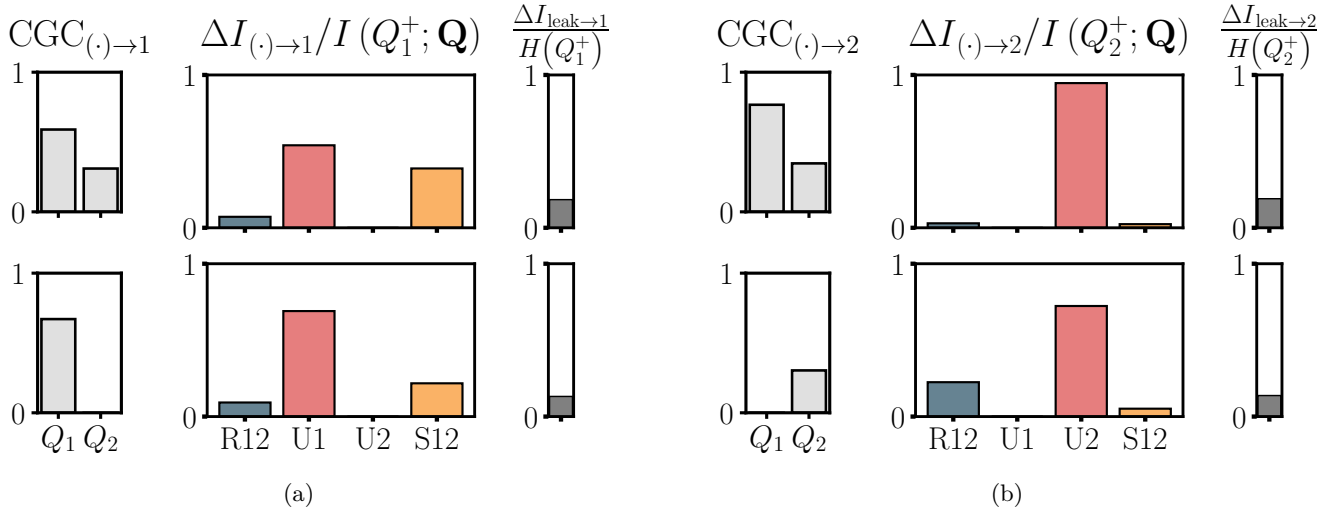


Figure S30: Effect of non-separability. Performance of non-linear CGC and SURD for the target variables (a) Q_1 and (b) Q_2. The results in the top row are for $\Delta n = 0$ ($\mathbf{Q} = [Q_1(n), Q_2(n)]$) and in the bottom row for $\Delta n = 1$ ($\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$ where $\mathbf{Q}_1 = [Q_1(n), Q_1(n-1)]$ and $\mathbf{Q}_2 = [Q_2(n), Q_2(n-1)]$). The system is simulated for the parameters $[r_1, r_2, \beta_{2\to 1}, \beta_{1\to 1}] = [3.8, 3.5, 0.2, 0.01]$.

(i.e., $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$), although these are vectors. This differs from the discussion in §S1.3, where different time lags are considered as different variables. The network architecture includes three hidden layers with 1024, 512 and 256 neurons, respectively, and it is trained using an Adam optimizer with a maximum of 200 epochs and an initial learning rate of 0.01, which is reduced by a factor of 0.3 with a period of 125 iterations.

Figure S30 displays the results from non-linear CGC and SURD using $\Delta n = 0$ ($\mathbf{Q} = [Q_1(n), Q_2(n)]$) and $\Delta n = 1$ ($\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$ where $\mathbf{Q}_1 = [Q_1(n), Q_1(n-1)]$ and $\mathbf{Q}_2 = [Q_2(n), Q_2(n-1)]$). For $\Delta n = 0$, both non-linear CGC and SURD identify the coupling between Q_1 and Q_2. However, with an additional time lag for both variables, non-linear CGC incorrectly determines that Q_1 does not influence Q_2 and vice versa, as these can be completely determined by their own past. In contrast, SURD continues to show the causal dependency between Q_1 and Q_2. The improved robustness of SURD is attributable to the fact that, under a statistical steady state, the flow of information between variables remains unchanged. The main difference observed in SURD is an increase in redundant causality due to duplicated information from the inclusion of additional time lags.

S6 Application of SURD to predictive modeling

SURD can also inform the development of predictive and/or reduced-order models of dynamical systems. By leveraging knowledge of the causal structure of the system, SURD enables the construction of minimal models by selecting the most effective input variables while disregarding those with irrelevant or duplicated information. This section illustrates an application of SURD to temporal forecasting of variables in the synergistic and redundant collider systems, as shown in Figures 4 and 5. The approach employs long-short-term memory (LSTM) artificial neural networks trained to predict $Q_1(n+1)$, using the exact values of $Q_1(n)$, $Q_2(n)$, and $Q_3(n)$. Several models are trained using different sets of input variables. The network architecture includes a sequence input layer with the corresponding number of input features, an LSTM layer with 200 hidden units to capture temporal dependencies between the signals, and a fully connected layer to map the previous layer to the output variable. The network is trained using an Adam optimizer with a maximum of 200 epochs and an initial learning rate of 0.01, which is reduced by a factor of 0.3 with a period of 125 iterations.

In the first case (Figure 4), Q_2 and Q_3 synergistically influence Q_1, as previously indicated by $\Delta I^S_{23\to 1}$. Therefore, it is crucial for models to incorporate both variables as inputs to ensure accurate predictions. This is illustrated in Figure S31(a), where the forecasting performance of the models using $[Q_2, Q_3]$ significantly surpasses those that include either variable alone. This outcome is consistent with the synergistic causality detected by SURD, where Q_2 and Q_3 collectively drive the future of Q_1. Generally, accurate forecasting of variables affected by synergistic causalities is achievable only when all synergistically interacting variables are incorporated into the model.

In the second case (Figure 5), Q_2 and Q_3 exhibit redundant causality to Q_1, as revealed by $\Delta I^R_{23\to(\cdot)}$. Hence, predictive models can use either Q_2 or Q_3 without compromising their predictive accuracy as shown in Figure S31(b).

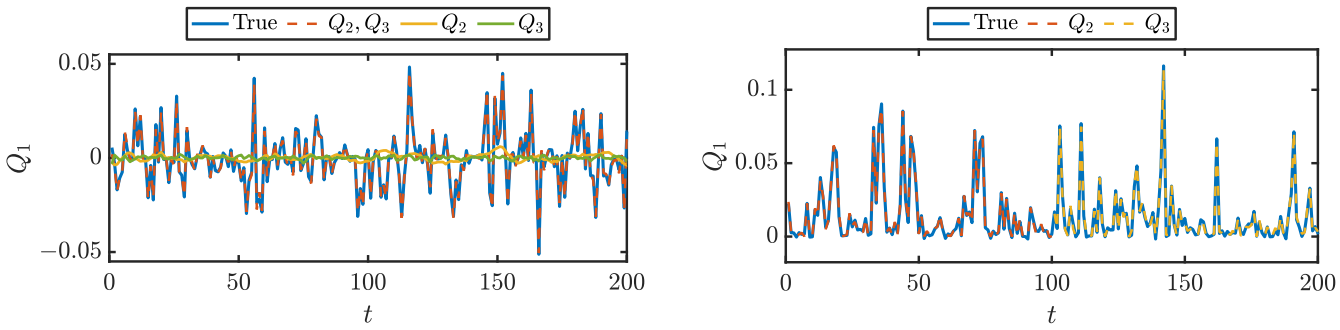


Figure S31: Comparative performance of LSTM models for forecasting the future of Q_1 using different input variables for (a) system with a synergistic collider (where Q_2 and Q_3 collectively influence the future of Q_1) and (b) system with redundant collider (where Q_2 and Q_3 contain the same information about the future of Q_1). The legend indicates the variables used as input to the LSTM model. In panel (b), the prediction is performed using Q_2 for the first half of the temporal sequence, while Q_3 is used for the second half.

In scenarios of high redundancy, minimal predictive models can be optimized by selecting the most convenient variable from the redundant set. This interchangeability provides a strategic advantage in model construction, allowing for the selection of variables based on practical considerations, such as measurement ease or data availability. For a more detailed discussion on information-theoretic causality for reduced-order modeling of chaotic dynamical systems, the reader is referred to Ref. [45].