nature portfolio

Peer Review File

Decomposing causality into its synergistic, unique, and redundant components

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In the manuscript titled "Decomposing causality into its synergistic, unique, and redundant components", the authors develop a novel causal inference method named SURD: Synergistic-Unique-Redundant Decomposition. This method is based on information theory and decomposes causality into synergistic, unique, and redundant components, quantifying it as the information about future events obtained from observing variables. It effectively addresses many challenges present in current causal inference methods such as Granger causality, convergent cross mapping, transfer entropy, and others. The numerical results from the toy models and the example of energy cascade in turbulence demonstrate that SURD has great potential in revealing causality in various complex systems. The manuscript is very well-written, and the methods and results are generally interesting to the scientific community of many disciplines. The underlying idea is intuitive and effective. I enjoyed reading this paper and can recommend it to publish in Nature Communications after the following concerns being properly addressed:

1. SURD is developed based on information theory, which is an essential part of the work. However, the introduction section contains only a brief discussion on the current progress of the application of information theory in causal inference. This makes the proposal of the method in the last paragraph somewhat abrupt. The authors should consider integrating parts of the content from the supplementary material into the main body.
2. When the number of individuals in a system is large, using the SURD method to decompose causality requires breaking it down into a very large number of parts. How should the issue of the resulting dimensionality problem be addressed in such cases?
3. In Sugihara et al. Science 2012, they claimed that data from nonlinear dynamical systems suffers from the non-separability, restricting the use of methods such as GC and from information theory. How the developed SURD overcomes the non-separability problem?
4. In Table 1 of the manuscript, the authors mention that the SURD method can address the problem of synchronization of two variables in logistic maps. This claim seems strange because a common sense is that the causality can hardly be defined if two variables are completely synchronous or generalized synchronous. Therefore no method can deal with the synchronization problem.
5. The authors mentioned that causality is often defined by confirming the results of a variable when intervening the other, which can usually not be realized in practice. How causality is defined mathematically in this manuscript? Does it denote coupling relationship of variables in the systems' equations? See "Continuity scaling: A rigorous framework for detecting and quantifying causality accurately, Research, 2022". If so, how the information-related methods can be used to infer this kind of causality?
6. The application of SURD to the energy cascade in turbulence is an excellent example, but the data is still simulated using existing models. It would be better to include some experiments to test SURD on real-world time series data.
7. The code provided in the manuscript requires an MIT account, which makes it inaccessible. The authors should find an alternative way to make the code available.

Reviewer #1 (Remarks on code availability):

The code provided in the manuscript requires an MIT account, which makes it inaccessible.

Reviewer #2 (Remarks to the Author):

In this paper, the authors present an exhaustive analysis of causality detection in physical systems following the formalism of information transfer. The authors take this formalism substantially further by decomposing information transfer into three well-defined and physically-motivated components (SURD). The causal relations detected with this method are then exhaustively compared with other causality-detection methods, providing strong and well-supported evidence of the ability of SURD to successfully detect and decompose causality in wide range of physical systems. The theoretical results of the paper are justified on physical and mathematical grounds, and
the authors make a complex topic such as information transfer accessible to a wide audience.

In general, this paper presents important advancements in the field of causal analysis and opens avenues to novel causal discovery in complex system. That is why I am happy to recommend its publication in Nature Communications after the following comments have been taken into account:

1) The information transfer decomposition is very well explained in the supplemental material. This is the right place due to the heavy mathematical content. However, I miss a somewhat deeper explanation of the physical meaning of the decomposition in the main text. For instance, in the bullet point list in the second page the authors could extend briefly on the meaning of "common causality" and "the causality that cannot be obtained by any other individual variables" for instance in terms of uncertainty reduction or information gain.

2) Sometimes across the text (also in the Supplemental Material), it is a little bit unclear what is uncertainty and what is information. For instance, the authors say "The information in Q+j is measured by the Shannon entropy, denoted by H(Q+j), which represents the average number of bits required to unambiguously determine Q+j" . The way it is defined, H is an entropy and thus reflects lack of knowledge (typically, large entropy goes in the direction of large uncertainty). In fact, the reduction of H(Q+j) due to knowledge of the past, is then used in the SM to define a mutual information. Perhaps, the distinction is subjective in the sense that H is the uncertainty in the absence of measurements and also the information gained by reducing this totally reducing this uncertainty when a measurement is performed. Maybe these ideas are implicitly clear to the authors, but I think that perhaps the paper would gain in clarity for the general reader with more consistency.

3) As I understand the paper, the information transfer method defines causality by how much a variable can be used to reduce the uncertainty of another variable in the future, i.e., how useful it is to predict. This is a strong point that connects the method with the fundamentals of scientific

discovery (scientific theories are meant to predict nature), but also to more practical problems. From the point of view of temporal forecasting, could the authors discuss briefly (perhaps in the SM) on the meaning of their decomposition? Could it be used to construct minimal predictive models, for instance, of turbulent flows? Or to discard and select the appropriate variables with which to construct these models?

4) The application of the method to the energy cascade is very interesting. I think the authors could connect their findings with the dissipative anomaly (dissipation does not vanish with vanishing viscosity) or Taylor's surrogate dissipation, which point to the idea that large-scale dynamics determine small-scale dynamics. These are classical empirical observations which lack a convincing explanation beyond the phenomenological theory of the cascade. Why the cascade happens the way it does is, in my opinion, an open question which could potentially benefit from the analysis presented in this paper. In this direction, an important problem in turbulence is to determine what parts of the flow are relevant to the cascade, which is connected to intermittency and LES modelling. Maybe, the authors could briefly comment on this in the energy cascade section.

5) Also in the energy cascade section, I think that the reason why CCM fails to detect causality in the forward energy cascade is because it is not well-suited for strongly synchronized variables such as the average interscale energy fluxes in turbulence (e.g. reference 106). This limitation of CCM was reported and corrected in a follow-up paper (Ye, H., Deyle, E., Gilarranz, L. et al.Sci Rep 5, 14750 (2015)) in which time-lags are explicitly introduced in the method. In my personal experience, this method works fine to detect the causality of the forward cascade. I think this also means that the SURD is adequate for strongly synchronized systems.

NCOMMS-24-21987 – Response to Reviewer 1

Decomposing causality into its synergistic, unique, and redundant components

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

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

We would like to thank the reviewer for taking the time to read our manuscript and for providing constructive feedback on the work. As explained in this point-by-point response, we have addressed all the concerns of the reviewer. Overall, we believe the modifications have improved the quality of the manuscript. The most significant changes are highlighted in blue in the revised version of the manuscript, which also includes corrections suggested by other referees.

R1: **SURD is developed based on information theory, which is an essential part of the work. However, the introduction section contains only a brief discussion on the current progress of the application of information theory in causal inference. This makes the proposal of the method in the last paragraph somewhat abrupt. The authors should consider integrating parts of the content from the supplementary material into the main body.**

A1: We agree with the referee. To address this concern, we have integrated more detailed content from the supplementary material into the main body of the introduction section. We have also extended the introduction to account for recent publications on information theory for causal inference. The new discussion is reproduced below:

Information theory, the science of message communication [44], has also served as a framework for model-free causality quantification. The success of information theory relies on the notion of information as a fundamental property of physical systems, closely tied to the restrictions and possibilities of the laws of physics [45, 46]. The grounds for causality as information are rooted in the intimate connection between information and the arrow of time. Time-asymmetries present in the system at a macroscopic level can be leveraged to measure the causality of events using information-theoretic metrics based on the Shannon entropy [44]. The initial applications of information theory for causality were formally established through the use of conditional entropies, employing what is known as directed information [47, 48]. Among the most recognized contributions is transfer entropy (TE) [49], which measures the reduction in entropy about the future state of a variable by knowing the past states of another. Various improvements have been proposed to address the inherent limitations of TE. Among them, we can cite conditional transfer entropy (CTE) [50, 51, 52, 53], which stands as the nonlinear, nonparametric extension of conditional GC [27]. Subsequent advancements of the method include multivariate formulations of CTE [45] and momentary information transfer [54], which extends TE by examining the transfer of information at each time step. Other information-theoretic methods, derived from dynamical system theory [55, 56, 57, 58], quantify causality as the amount of information that flows from one process to another as dictated by the governing equations.

R2: **When the number of individuals in a system is large, using the SURD method to decompose causality requires breaking it down into a very large number of parts. How should the issue of the resulting dimensionality problem be addressed in such cases?**

A2: We thank the reviewer for raising this important point. Defining redundant, unique, and synergistic causalities indeed involves decomposing mutual information into multiple components. The definition proposed in SURD is motivated by consistency with the properties presented in Section S1.2, as well as their interpretability. In the literature, alternative definitions and decompositions of mutual information have been suggested, but these often lack ease of interpretation and scale poorly with the number of variables involved. One of the most cited decompositions of mutual information is from Williams & Beer [114], where the number of terms grows as the Dedekind numbers. This yields an absolutely unreasonable number of terms. To give the reviewer an idea, in the case of 9 variables, previous decompositions of mutual information, such as Williams & Beer, would result in over 10^{23} terms! Some decompositions do not even guarantee the nonnegativity of the redundant and synergistic components, introducing difficulties in their interpretation.

In contrast, SURD is rooted in a new decomposition of mutual information that retains nonnegativity while keeping the number of terms as low as possible. For instance, with 9 variables, SURD yields 512 terms, which is a significantly lower number compared to other methods (recall 10^{23} terms). We understand that 512 might still be perceived as a large number, but this is the minimum number of terms needed to account for all redundancies and synergies. In scenarios with a large number of variables, synergistic and redundant causalities can be grouped by different orders or included in the causality leak. In summary, SURD offers a decomposition of causality into unique, synergistic, and redundant components with a minimum number of terms, which can additionally be grouped to facilitate interpretation when the number of variables is large. The following discussion has been extended in the supplementary material:

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 $\Delta I^S_{2\text{nd}} = \Delta I^S_{12\to 1} + \Delta I^S_{23\to 1} + \Delta I^S_{13\to 1}$, 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.

R3: **In Sugihara *et al.*, *Science* 2012, they claimed that data from nonlinear dynamical systems suffers from the non-separability, restricting the use of methods such as GC and from information theory. How the developed SURD overcomes the non-separability problem?**

A3: This is an interesting point. Similarly to CCM, SURD is also consistent with Takens' embedding theorem as it accounts for the flow of information among variables. We discuss this point in a new section included in the supplementary material. We also demonstrate the robustness of SURD using the same example suggested by Sugihara [36] to illustrate the problem of non-separability. The new section is reproduced below.

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 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 [36] to illustrate the robustness of SURD under the effect of multiple time lags with non-separable variables. The system is given by:

$$Q_1(n+1) = Q_1(n)\left[r_1 - r_1 Q_1(n) - \beta_{2\to 1} Q_2(n)\right], \tag{R1a}$$

$$Q_2(n+1) = Q_2(n)\left[r_2 - r_2 Q_2(n) - \beta_{1\to 2} Q_1(n)\right], \tag{R1b}$$

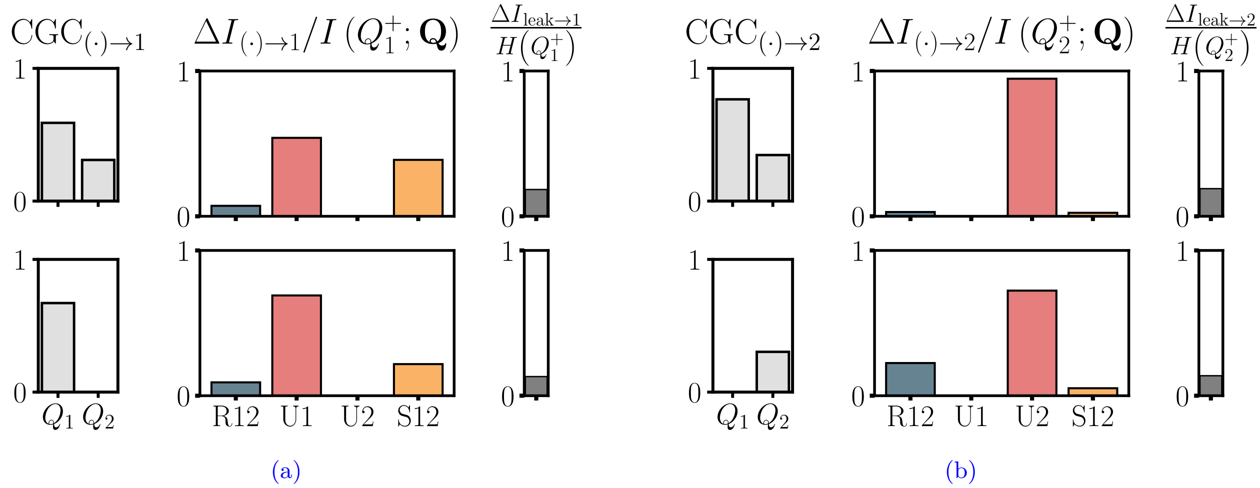


Figure R1: Effect of non-separability. Performance of non-linear CGC and SURD for the target variables (a) Q_1 and (b) Q_2. The results in the top row are for $\Delta n = 0$ ($\mathbf{Q} = [Q_1(n), Q_2(n)]$) and in the bottom row for $\Delta n = 1$ ($\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$ where $\mathbf{Q}_1 = [Q_1(n), Q_1(n-1)]$ and $\mathbf{Q}_2 = [Q_2(n), Q_2(n-1)]$). The system is simulated for the parameters $[r_1, r_2, \beta_{2\to 1}, \beta_{1\to 1}] = [3.8, 3.5, 0.2, 0.01]$.

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):

$$\beta_{2\to 1} Q_2(n) = 1 - Q_1(n) - \frac{Q_1(n+1)}{r_1 Q_1(n)}, \tag{R2a}$$

$$\beta_{1\to 2} Q_1(n) = 1 - Q_2(n) - \frac{Q_2(n+1)}{r_2 Q_2(n)}. \tag{R2b}$$

We can substitute Equation (R2a) into (R1b) and obtain an expression for $Q_2(n)$ as a function of $Q_1(n)$ and $Q_1(n-1)$:

$$Q_2(n) = \frac{r_2}{\beta_{2\to 1}} \left[(1 - \beta_{1\to 2} Q_1(n-1)) \left(1 - Q_1(n-1) - \frac{Q_1(n)}{r_1 Q_1(n-1)} \right) - \frac{1}{\beta_{2\to 1}} \left(1 - Q_1(n-1) - \frac{Q_1(n)}{r_1 Q_1(n-1)} \right)^2 \right]. \tag{R3}$$

Introducing Equation (R3) into (R1a), we obtain an expression for Q_1 that is exclusively a function of its own past, i.e. $Q_1(n)$ and $Q_1(n-1)$:

$$Q_1(n+1) = f(Q_1(n), Q_1(n-1)). \tag{R4}$$

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

$$Q_1(n+1) = \text{ANN}_1(\boldsymbol{Q}_1) + \hat{\varepsilon}(n+1), \tag{R5a}$$

$$Q_1(n+1) = \text{ANN}_{12}(\boldsymbol{Q}_1, \boldsymbol{Q}_2) + \varepsilon(n+1), \tag{R5b}$$

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 (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 R1 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.

R4: **In Table 1 of the manuscript, the authors mention that the SURD method can address the problem of synchronization of two variables in logistic maps. This claim seems strange because a common sense is that the causality can hardly be defined if two variables are completely synchronous or generalized synchronous. Therefore no method can deal with the synchronization problem.**

A4: We completely agree with the reviewer and apologize for the misunderstanding. Identifying the directionality of causality in the case of completely synchronous or generalized synchronous variables is not meaningful. Our intention was not to claim otherwise but to highlight the ability of SURD to identify synchronization between variables in such scenarios. More specifically, when two variables are fully synchronized, SURD detects this through a prevalence of the redundant causality component. Thus, while directionality of causality cannot be established, SURD effectively identifies the state of synchronization. This concept is discussed in § S3.5, where we analyze the performance of SURD in logistic systems with synchronized variables. We have further clarified this point. The discussion is reproduced below:

Strong coupling $Q_1 \to Q_2$ ($c_{1\to 2} = 1$ and $c_{12\to 3} = 0$). Taking the limit $c_{1\to 2} \to \infty$, it can be seen that $Q_2 \equiv Q_1$. It is also known that even for lower values of $c_{12\to 3} \sim 1$, 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 $\Delta I^R_{12\to 1} = \Delta I^R_{12\to 2} \neq 0$. 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 ($\Delta I^U_{3\to 3} \neq 0$).

R5: **The authors mentioned that causality is often defined by confirming the results of a variable when intervening the other, which can usually not be realized in practice. How causality is defined mathematically in this manuscript? Does it denote coupling relationship of variables in the systems' equations? See "Continuity scaling: A rigorous framework for detecting and quantifying causality accurately, Research, 2022". If so, how the information-related methods can be used to infer this kind of causality?**

A5: This is a complicated yet interesting question that may even verge on the philosophical. Causality with interventions is mentioned in the introduction as an intuitive definition of causality; however, we immediately enumerate the many limitations and caveats of this approach. Overall, we avoid providing any "absolute" definition of causality in mathematical terms, as this is not agreed upon within the causal inference community (from there the myriad methods for causal inference). Nonetheless, we require that causality (however defined by each method) must be at least consistent with the dependencies dictated by the governing equations of the system. We believe the mathematical definition of causality given by SURD, although not absolute, offers results consistent with the functional dependency of the variables in the multiple examples analyzed.

We also thank the reviewer for pointing out the new reference. In the revised version of the manuscript, we have also included a citation to the continuity scaling method in the introduction.

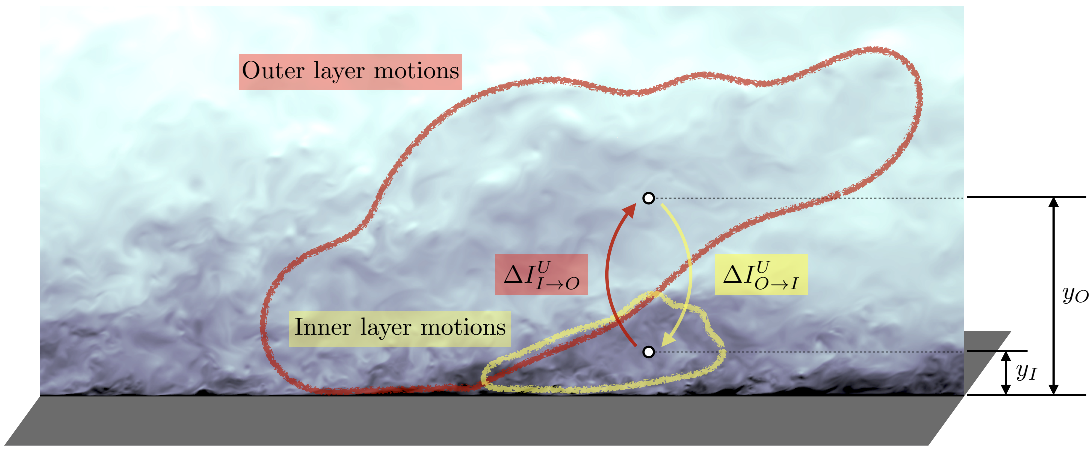
An alternative approach, known as continuity scaling [43], directly assesses causal relationships by examining the scaling laws governing the continuity of the system

We thank the reviewer again for bringing this new line of research to our attention. Interestingly, the concept of SURD appears to be analogous to continuity scaling, where instead of examining volumes of variables in phase space, we analyze volumes of conditional entropy in information space. As mentioned in one of the previous responses, our group have established some connections between Takens' embedding theorem and information theory [46]. We are currently working on providing rigorous connections between phase space and information space and would be very interested in exploring "information-theoretic" continuity scaling in the future.

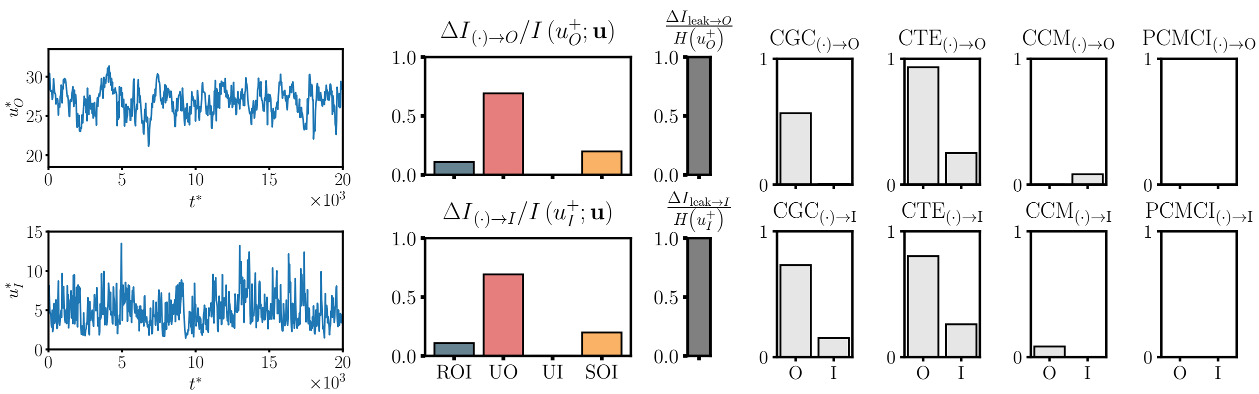
R6: **The application of SURD to the energy cascade in turbulence is an excellent example, but the data is still simulated using existing models. It would be better to include some experiments to test SURD on real-world time series data.**

A6: We appreciate the suggestion. Applying SURD to real-world experimental time series data would strengthen the study. To that end, we have applied SURD to experimental data from a turbulent boundary layer in the high Reynolds number wind tunnel at the University of Melbourne. The data is publicly available at `https://fluids.eng.unimelb.edu.au/`. The new section, included in the main text, is reproduced below.

Application to experimental data from a turbulent boundary layer



(a)



(b)

Figure R2: **Causality between streamwise velocity motions in a turbulent boundary layer**. (a) Schematic of outer-layer and inner-layer streamwise velocity motions in a turbulent boundary and their interactions via unique causality. The velocity signals $u_I(t)$ and $u_O(t)$ are experimentally measured at the wall-normal locations y_I and y_O, respectively, and are shown in the panel below. The superscript $*$ denotes the inner scaling with friction velocity, u_τ, and kinematic viscosity, ν. (b) Redundant (R), unique (U), and synergistic (S) causalities among velocity signals in the inner (I) and outer (O) layer of a turbulent boundary layer. The gray bar is the causality leak. The results of CGC, CTE, CCM, and PCMCI are shown on the right. Details about data are provided in Methods.

The interaction of turbulent motions of different size within the thin fluid layers immediately adjacent to solid boundaries poses a significant challenge for both physical understanding and prediction. These layers are responsible for nearly 50% of the aerodynamic drag on modern airliners and play a crucial role in the first hundred meters of the atmosphere, influencing broader meteorological phenomena[94]. Here, we leverage SURD to investigate the interaction between flow velocity motions in the outer layer (far from the wall) and inner layer (close to the wall) of a turbulent boundary layer. Figure R2(a) illustrates the configuration used to examine the causal interactions between velocity motions. More specifically, the hypotheses under consideration are either i) a dominant influence of motions far from the wall on those closer, indicating top-down causality (a.k.a. Townsend's outer-layer similarity hypothesis [105]), or ii) the opposite scenario, where influences move from areas closer to the wall outward, suggesting bottom-up causality.

We use experimental data from a zero-pressure gradient turbulent boundary layer from the high Reynolds number wind tunnel at the University of Melbourne[78, 79, 80]. The friction Reynolds number is $Re_\tau = u_\tau \delta/\nu = 14,750$, based on the thickness of the boundary layer δ, the kinematic viscosity ν, and the average friction velocity at the wall u_τ. The time signals consists of the streamwise velocity at two wall-normal locations within the inner (I) and outer (O) layers, denoted by $u_I(t)$ and $u_O(t)$, respectively.

Figure R2(b) shows the redundant, unique, and synergistic causalities from SURD between the inner and outer layers. We use the subindices I and O to refer to causalities from/to $u_I(t)$ or $u_O(t)$, respectively. The primary observation is that the inner layer motions are predominantly influenced by the unique causality from the outer layer, $\Delta I^U_{O\to I}$. The redundant and synergistic causalities are lower, but they remain significant. Curiously, the unique causality $\Delta I^U_{I\to I}$ is zero, implying that, at the time scale considered, the inner layer motions are independent of their past history. For the outer-layer motions, most of the causality is self-induced $\Delta I^U_{O\to O}$ with no apparent influence from the inner layer. The results distinctly support the prevalence of top-down interactions: causality flows predominantly from the outer-layer large-scale motions to the inner-layer small-scale motions. The outcome is consistent with the modulation of near-wall scales by large-scale motions reported in previous investigations [106, 107]. The lack of bottom-up causality from the inner to the outer layer also aligns with Townsend's outer-layer similarity hypothesis [105] and previous observations in the literature [108, 109, 110, 111, 112, 113].

The causality leak, also shown in Fig. R2(b), is 99% for both u_I and u_O. Such a high value implies that most of the causality determining the future of u_I and u_O is contained in other variables not considered in the analysis. This high value is unsurprising since most of the millions of degrees of freedom in the turbulent flow field have been neglected, and only two pointwise signals, u_I and u_O, are retained to evaluate the causality.

Finally, the results from SURD are contrasted with other methods. In this case, CCM and PCMCI do not support the hypothesis of top-down interactions between velocity motions. The reason behind the failure of these methods is unclear, but it might be related to the high causality leak. CGC and CTE are consistent with the flow of causality from the outer-layer large-scale motions to the inner-layer small-scale motions. However, as already highlighted in previous cases, none of these methods offer a detailed decomposition into redundant, unique, and synergistic causality, nor they account for the effect of unobserved variables as quantified by the causality leak in SURD.

R7: **The code provided in the manuscript requires an MIT account, which makes it inaccessible. The authors should find an alternative way to make the code available.**

A7: We have made our code publicly available using a public GitHub repository, which can be now accessed at: `https://github.com/Computational-Turbulence-Group/SURD`.

NCOMMS-24-21987 – Response to Reviewer 2

Decomposing causality into its synergistic, unique, and redundant components

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

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

We would like to thank the reviewer for taking the time to read our manuscript and for providing constructive feedback on the work. As explained in this point-by-point response, we have addressed all the concerns of the reviewer. Overall, we believe the modifications have improved the quality of the manuscript. The most significant changes are highlighted in blue in the revised version of the manuscript, which also includes corrections suggested by other referees.

R1: **The information transfer decomposition is very well explained in the supplemental material. This is the right place due to the heavy mathematical content. However, I miss a somewhat deeper explanation of the physical meaning of the decomposition in the main text. For instance, in the bullet point list in the second page the authors could extend briefly on the meaning of "common causality" and "the causality that cannot be obtained by any other individual variables" for instance in terms of uncertainty reduction or information gain.**

A1: We have expanded the explanations of redundant and unique causality to better convey their physical meanings in terms of information. The additions are short, as we would like to keep the discussion as compact as possible. The new descriptions are as follows:

- Redundant causality from $\boldsymbol{Q_i} = [Q_{i_1}, Q_{i_2}, \ldots]$ to Q_j^+ (denoted by $\Delta I^R_{\boldsymbol{i}\to j}$) is the common causality shared among all the components of $\boldsymbol{Q_i}$, where $\boldsymbol{Q_i}$ is a subset of $\boldsymbol{Q}$. Redundant causality occurs when all the variables in $\boldsymbol{Q_i}$ contain the same amount of information about Q_j^+. Therefore, any component of $\boldsymbol{Q_i}$ offers identical insight into the outcome of Q_j^+.
- Unique causality from Q_i to Q_j^+ (denoted by $\Delta I^U_{i\to j}$) is the causality from Q_i that cannot be obtained from any other individual variable $Q_k \neq Q_i$. This causality occurs when observing Q_i yields more information about some outcomes of Q_j^+ than observing any other isolated variable.
- Synergistic causality from $\boldsymbol{Q_i} = [Q_{i_1}, Q_{i_2}, \ldots]$ to Q_j^+ (denoted by $\Delta I^S_{\boldsymbol{i}\to j}$) is the causality arising from the joint effect of the variables in $\boldsymbol{Q_i}$. This causality occurs when more information about Q_j^+ is gained by observing a collection of variables simultaneously than by observing each variable individually.
- Causality leak represents the effect from unobserved variables that influence Q_j^+ but are not contained in $\boldsymbol{Q}$. This is the amount of information missing that would be required to unambiguously determine the future of Q_j after considering all observable variables collectively.

R2: **Sometimes across the text (also in the Supplemental Material), it is a little bit unclear what is uncertainty and what is information. For instance, the authors say "The information in Q_j^+ is measured by the Shannon entropy, denoted by $H(Q_j^+)$, which represents the average number of bits required to unambiguously determine Q_j^+" . The way it is defined, H is an entropy and thus reflects lack of knowledge (typically, large entropy goes in the direction of large uncertainty). In fact, the reduction of $H(Q_j^+)$ due to knowledge of the past, is then used in the SM to define a mutual information. Perhaps, the distinction is subjective in the sense that H is the uncertainty in the absence of measurements and also the information gained by reducing this totally reducing this uncertainty when a measurement is performed. Maybe these ideas are implicitly clear to the authors, but I think that perhaps the paper would gain in clarity for the general reader with more consistency.**

A2: We agree that a comment about the relationship between uncertainty and information would improve the clarity of the manuscript. We have added the following clarification in the main text after introducing the Shannon entropy:

It is also useful to interpret Shannon entropy as a measure of uncertainty. Processes that are highly uncertain (high entropy) are also the ones from which we gain the most information when their states are determined. Conversely, uncertainty is zero when the process is completely deterministic, indicating no information is gained when the outcome is revealed.

R3: **As I understand the paper, the information transfer method defines causality by how much a variable can be used to reduce the uncertainty of another variable in the future, i.e., how useful it is to predict. This is a strong point that connects the method with the fundamentals of scientific discovery (scientific theories are meant to predict nature), but also to more practical problems. From the point of view of temporal forecasting, could the authors discuss briefly (perhaps in the SM) on the meaning of their decomposition? Could it be used to construct minimal predictive models, for instance, of turbulent flows? Or to discard and select the appropriate variables with which to construct these models?**

A3: This is a valuable observation raised by the reviewer. Indeed, SURD can be used in the context of predictive modeling, including temporal forecasting. In response, we have incorporated a new section in the revised version of our Supplementary Materials, where we illustrate the application of SURD to select the most effective input variables for temporal forecasting in systems with synergistic and redundant causalities. This topic is very rich and warrants a more in-depth analysis, which will be presented in a follow-up work. The new section is included below:

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.

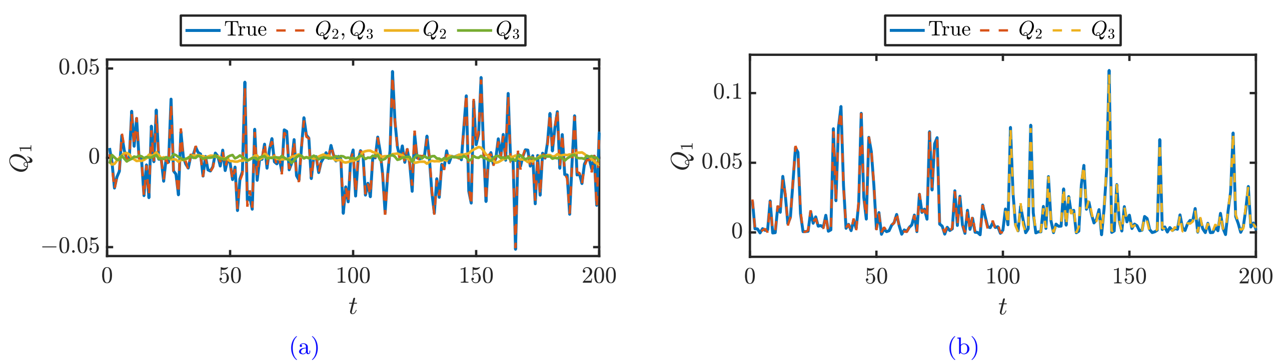


Figure S1: Comparative performance of LSTM models for forecasting the future of Q_1 using different input variables for (a) system with a synergistic collider (where Q_2 and Q_3 collectively influence the future of Q_1) and (b) system with redundant collider (where Q_2 and Q_3 contain the same information 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 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 S1(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 S1(b). 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].

R4: **The application of the method to the energy cascade is very interesting. I think the authors could connect their findings with the dissipative anomaly (dissipation does not vanish with vanishing viscosity) or Taylor's surrogate dissipation, which point to the idea that large-scale dynamics determine small-scale dynamics. These are classical empirical observations which lack a convincing explanation beyond the phenomenological theory of the cascade. Why the cascade happens the way it does is, in my opinion, an open question which could potentially benefit from the analysis presented in this paper. In this direction, an important problem in turbulence is to determine what parts of the flow are relevant to the cascade, which is connected to intermittency and LES modelling. Maybe, the authors could briefly comment on this in the energy cascade section.**

A4: We appreciate the suggestion by the reviewer. We have noted these observations in the revised version of the manuscript. The new additions are reproduced below:

Curiously, no unique causality is observed from smaller to larger scales, and any causality from the backward cascade arises solely through redundant relationships. In the context of SURD, this implies that no new information is conveyed from the smaller scales to the larger scales, which is consistent with recent views of the backward energy cascade in the literature [96, 97]. From the modeling perspective, this justifies the success of subgrid-scale modeling in large-eddy simulation, as the information contained in the smaller scales is redundant and does not constitute a key ingredient in solving the closure model problem. The results obtained from SURD also provide support for classic hypotheses about the energy cascade from a new causal-effect perspective. Among them, we can cite Taylor's dissipation surrogate assumption [98] and the dissipation anomaly [99]. The former posits that the dissipation rate can be determined by large-scale dynamics, even if dissipation is formally a small-scale feature of the flow. SURD clearly supports this assumption due to the lack of unique and synergistic causality from small to large scales. The results from SURD are also consistent with the dissipation anomaly (i.e., the constant rate of energy dissipation despite decreasing viscosity), which is enabled by the forward directionality of the energy cascade process.

Finally, in response to the last comment of the reviewer about which parts of the flow are relevant to the cascade, we are developing an extended version of SURD specifically designed for analyzing three-dimensional (3-D) fields, rather than being limited to time series data. The current version of SURD lays the groundwork for the forthcoming 3-D extention. We anticipate that this extension will improve our understanding of the 3-D causal structure of the energy cascade. This, in turn, could support the validation of existing theories on intermittency, among others, and contribute to the development of closure models for LES. While this aspect will be the focus of future research, we recognize its importance and are excited about its potential to provide deeper insights into the mechanisms of the energy cascade.

R5: **Also in the energy cascade section, I think that the reason why CCM fails to detect causality in the forward energy cascade is because it is not well-suited for strongly synchronized variables such as the average interscale energy fluxes in turbulence (e.g. reference 106). This limitation of CCM was reported and corrected in a follow-up paper (Ye, H., Deyle, E., Gilarranz, L. et al. *Sci Rep* 5, 14750 (2015)) in which time-lags are explicitly introduced in the method. In my personal experience, this method works fine to detect the causality of the forward cascade. I think this also means that the SURD is adequate for strongly synchronized systems.**

A5: We thank the reviewer for bringing to our attention the improved version of CCM. We have expanded the discussion about the energy cascade to clarify that the new iterations of CCM can perform satisfactorily in detecting causality within the energy cascade. The new addition is as follows:

The formulation of CCM used in this study adheres to the original work by Sugihara et al. [36]. However, more recent iterations of CCM, such as Extended CCM [39], which explicitly account for time delays, have demonstrated efficacy in accurately detecting causality in systems with strongly synchronized variables. Hence, these and other improved versions of CCM might be more suitable for analyzing the turbulent energy cascade, where smaller scales are enslaved to the larger ones.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

Thank the authors very much for considering and addressing my previous concerns. I find all my concerns have been addressed properly in the revised manuscript and I am now happy to recommend publication of this manuscript in NC. Moreover, as the authors mentioned that intervention-based causality inference method is important both in theory and practice, I suggest a recent work "Detecting dynamical causality via intervened reservoir computing", which demonstrates a representative method of this type.

Reviewer #1 (Remarks on code availability):

I have checked the provided link and code, and find it usable and useful for the readers.

Reviewer #2 (Remarks to the Author):

The authors have made a great work answering all my comments. I am happy to recommend the paper for publication.