

Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications

Erik M. Bollt, Jie Sun, and Jakob Runge

Citation: *Chaos* **28**, 075201 (2018); doi: 10.1063/1.5046848

View online: <https://doi.org/10.1063/1.5046848>

View Table of Contents: <http://aip.scitation.org/toc/cha/28/7>

Published by the [American Institute of Physics](#)



Chaos
An Interdisciplinary Journal of Nonlinear Science

Fast Track Your Research. *Submit Today!*

Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications

Erik M. Bollt,^{1,2,3,4} Jie Sun,^{1,2,3,5} and Jakob Runge^{6,7}

¹Clarkson Center for Complex Systems Science (C^3S^2), Clarkson University, Potsdam, New York 13699, USA

²Department of Mathematics, Clarkson University, Potsdam, New York 13699, USA

³Department of Physics, Clarkson University, Potsdam, New York 13699, USA

⁴Department of Electrical and Computer Engineering, Clarkson University, Potsdam, New York 13699, USA

⁵Department of Computer Science, Clarkson University, Potsdam, New York 13699, USA

⁶German Aerospace Center (DLR), Institute of Data Science, Maelzerstrasse 3, 07745 Jena, Germany

⁷Grantham Institute, Imperial College, London SW7 2AZ, United Kingdom

(Received 2 July 2018; accepted 4 July 2018; published online 31 July 2018)

Questions of causation are foundational across science and often relate further to problems of control, policy decisions, and forecasts. In nonlinear dynamics and complex systems science, causation inference and information flow are closely related concepts, whereby “information” or knowledge of certain states can be thought of as coupling influence onto the future states of other processes in a complex system. While causation inference and information flow are by now classical topics, incorporating methods from statistics and time series analysis, information theory, dynamical systems, and statistical mechanics, to name a few, there remain important advancements in continuing to strengthen the theory, and pushing the context of applications, especially with the ever-increasing abundance of data collected across many fields and systems. This Focus Issue considers different aspects of these questions, both in terms of founding theory and several topical applications.

Published by AIP Publishing. <https://doi.org/10.1063/1.5046848>

I. INTRODUCTION

A basic and fundamental pursuit in science is to infer causal interactions and relationships. In terms of dynamical systems science, one may ask which dynamic variables influence other variables, either directly or indirectly through intermediate variables, and which variables only appear to be related due to the influence of a common driver. While traditional approaches to scientific discovery of causation between variables are through close iterations of forming a hypothesis and conducting controlled real experiments, in the past few decades, novel data-oriented approaches have been proposed that attempt to detect causal relations from purely observational data, driven by a growing availability of large-scale datasets. Observational causal inference constitutes a challenging problem for complex dynamical systems, from theoretical foundations to practical computational issues. Papers in this Focus Issue cover both theory and applications to a broad range of problems in social, physical, and biological systems.

The gold standard of scientific discovery is validation through controlled experiments: For example, in a standard physics experiment, to test whether a variable X has an influence on Y , one physically intervenes in X , changes its state, and measures whether this has an effect on Y , while keeping all other conditions unchanged as much as possible. In many problems, however, such controlled experiments can be infeasible—for example, when Johannes Kepler made the discovery of the laws of planetary motion, it was based on observations of orbital dynamics. In general, the central question that arises is how one might perform causal inference from purely observational data. The statistics of causation by interventions and observations is thoroughly described in

Pearl’s extensive works, summarized in Ref. 36, which lays out a mathematical framework for causation and develops the conditions under which interventions can be predicted from observational data.

In counterpoint to controlled experiments, there is the notion of observing a “free-running” complex system without actively probing it, and from (passive) time dependent observations thereof, asking the comparable question of what variables and factors are causal of others. In this context, causality can be interpreted as a form of influence on predictability (or the lack of predictability), that is, if the knowledge of one time series is useful in forecasting another time series, then the former can be seen and be interpreted as potentially “causal” for the latter. This formulation was put forward specifically in the work of Granger¹⁶ that led to his 2003 Nobel memorial prize in Economic Sciences,¹⁹ which in fact was closely related to the work of Norbert Wiener that predated by more than a decade.⁶⁸ In analogy to the controlled experimental setting above, to test whether a variable X has an influence on Y , one builds a predictive model of Y from observational data (most commonly, a linear autoregressive model) based on Y ’s and other covariates’ past and measures whether the inclusion of X in the model improves the predictability of Y . Here, other variables are not actively kept under control, but they are statistically controlled for.

Information-theoretic measures based on Shannon entropy and mutual information⁹ naturally allow for a very general characterization of dependencies in complex and dynamical systems from symbolic to continuous descriptions. Analogous to Wiener-Granger causality for linear systems, transfer entropy has become a highly popular way to consider

questions of pairwise information transfer between nonlinear dynamical systems.⁵³ In a basic sense, transfer from X to Y measures how much information the past of X contains about Y , beyond Y 's own past, which shows the close analogy to Granger causality. In the special case that the systems being analyzed are linear Gaussian stochastic processes, it has been shown that Granger causality is equivalent to transfer entropy.³ Hence, the entropic approach is more general, as it is applicable in the context of general distributions and nonlinear influences. Notice also that information-theoretic measures are fundamentally probabilistic in nature, as is Granger causality, since they are premised on comparing probability distribution functions of states. So, while it is naturally appropriate for a stochastic process, including random dynamical systems²⁹ or a stochastic differential equation,² it also can be well descriptive of a deterministic dynamical system. To this end, we simply recast the perspective by considering the evolution of ensembles of initial conditions,² which is essentially foundational to ergodic theory⁶³ and also transfer operators.^{5,6,29} Another aspect of causal inference that often is based on information-theoretic measures regards notions of causal coupling strength.^{24,38,42,44} Information-theoretic measures for continuous variables can be most efficiently estimated from data using nearest-neighbor estimators,²⁶ where permutation shuffle tests can be used for conditional-independence testing,⁴¹ that is, testing whether a (conditional) mutual information is zero.

The works of Arnhold, Hirata, Schiff, Sugihara, and others consider the causation problem through the perspective of dynamical attractors underlying nonlinear dynamical systems and the concept of generalized synchronization.^{1,17,20,21,39,51,56} Starting from the insight that higher-dimensional attractors can be reconstructed from univariate measured time series via time-delay embedding,⁶⁰ these works utilize what can be termed as the *closeness principle*. According to this principle, in causally related systems, states in the driver attractor that are temporally near to states in the response attractor should also be close to each other. As one implementation, *convergent cross-mapping*⁵⁶ tests the closeness principle via the convergence of nearest neighbors on the attractors of the two systems as the number of samples from the attractor increases.

More from the statistics and machine learning community, the problem of causal inference is treated within the framework of structural causal models, related to the work of Pearl.³⁶ The theory of structural causal models, summarized in Peters *et al.*,³⁷ lays out the conditions under which certain causal models, for example, linear models with non-Gaussian error terms, or nonlinear models, can be identified from observational data, not necessarily time series data, as in the Granger and nonlinear dynamics context, which is the major topic of this Focus Issue.

More broadly, a multi-faceted perspective is desired since while standard Granger causality and also transfer entropy are well suited for two-variable (or two-subsystem, two-component) settings of information flow, they are not designed to ask the question associated with three or more factors, including the issue of the curse of dimensionality.⁴⁵ For example, if we were to consider a system of three subsystems, X , Y , and Z , and ask: does X influence Z

directly or does X influence Z but only by an intermediary Y ? For these questions, one must cast a conditional variation of the above concepts. There exist conditional Granger causalities,^{7,12,34} conditional transfer entropies, and state conditioned transfer entropies,⁶⁷ and a special variation leading to causation entropy^{8,25,58,59}. Furthermore, if one wishes to uncover coupling structure, then we require an algorithm premised on these computations; for example, PC algorithm and momentary information-based causal discovery⁴⁶ which address inference of large-scale nonlinear causal networks in the presence of strong autocorrelation, or the optimal causation entropy (oCSE) approach^{58,59} which is designed to uncover the network of direct information flow influences using (CSE) as the underlying influence measure.

While most of the approaches are data-oriented in detail, and many are information-theoretic, there is notably the Liang-Kleeman formalism³¹ which takes a thermodynamic differential equations approach assuming known equations, in the spirit of information theory that is also shown in certain settings to coincide with transfer entropy. To counter point, a theoretical direct transfer operator description can be posed^{6,29} that utilizes Jensen-Shannon divergence instead of Kullback-Leibler divergence; these perspectives are also highlighted in this issue.

Several works have shown limitations of the Shannon information-theoretic framework in measuring dependence and causation. One aspect being that entropy is not invariant to change of variables,⁹ which was a critique of differential entropy early on Ref. 65. Another more recent work notes that there are stochastic processes for which the past is entirely shielded from the future.²³ In terms of multivariate dependencies, James *et al.*²² elaborate on polyadic relationships that put in question, even fundamentally, the notion of a directed graph with pairwise links to represent what should be a hyper graph. Such polyadic, or synergistic, causal drivers can also be crucial in optimal time series prediction schemes.⁴³

In the recent decade in particular, there has been a significant resurgence in consideration of causation inference, and the closely related concepts of information flow. Perhaps there are several underlying reasons for this, but not least of which the growth in this field is associated with the even broader growth generally of data sciences, machine learning, and "big data" analytics, in particular in the context of complex systems. Reasons for this boom can be described in terms of the ever growing availability of large-scale and more affordable data collection platforms, ever more powerful and cost-effective computing facilities and equipment, and an abundance of massive datasets from an unimaginable number of sensors everywhere. Questions like asking what part of the earth's atmosphere may be influential in forecasting other parts of the earth's atmosphere^{4,10,33,54,64,69} would be too overwhelmingly complex to consider analytically, but in the vein of a data-oriented approach on massive computing platforms, such a question if possible is clearly important. Data-oriented answers to general scientific questions are broadly enjoying a golden age, and causation inference and information flow are no exception. Applications range from social,^{18,32,57,62} medical,^{35,40,49,61,66} earth sciences,^{13-15,27,28,47,48,52} engineering,^{50,55} as well as financial

and economic,¹¹ to name a small fraction of the massive growing literature on the exciting applications of these powerful scientific tools.

Many open problems remain, however, from spatial and temporal resolution effects to determinism, nonlinearity, and multi-element and synergistic interactions. In this Focus Issue, contributions include work on these topics in terms of theory as well as applications such as collective animal behavior, interdependent ecological dynamics, social communication and opinion diffusion, sensing and monitoring in mechanical and other engineering structures, and the global climate system, thus encompassing a broad range of problems in social, physical, and biological systems.

II. THIS FOCUS ISSUE

As the theory of causation inference evolves, the state of the art, algorithmic design, and applications are all advancing as reflected in the papers herein. Many of the general concepts discussed here are expanded upon, modernized, or reviewed in the articles of this Focus Issue. Below, we provide a brief summary of each of the papers that appear in this Focus Issue.

- J. M. Amigo and Y. Hirata (this focus issue, Chaos **28**, 075302 (2018)) re-examine the method of the joint distance distribution to identify directional nonlinear couplings (“Detecting directional couplings from multivariate flows by the joint distance distribution”). The method lies at the intersection of nonlinear dynamics and time series analysis and utilizes the *closeness principle* according to which the states in the driver attractor that are concurrent with close states in the response attractor are also close to each other. Also, convergent cross-mapping falls into this framework. The paper insightfully illustrates advantages and pitfalls of the method, for example, the issue of phase synchronization.
- H. Ashikaga and R. James (this focus issue, Chaos **28**, 075306 (2018)) explored asymmetry in the information flow across different spatial scales in a mathematical model of cardiac dynamics, aimed at determining the relationship between rotors and inter-scale information flow (paper title “Inter-scale information flow as a surrogate for downward causation that maintains spiral waves”). By comparing transfer entropy and intrinsic transfer entropy, the paper concludes that information flow from macro- to micro-scale is adequately captured by transfer entropy, and no synergistic effects are present. Using transfer entropy as a surrogate, the paper focuses on information flow from macro-scale behavior of the system to its corresponding micro-scale states and found that such “downward causation” correlates with the number of rotors, which are rotation centers of cardiac spiral waves. However, no significant association was found between a higher number of rotors and higher downward information flow. Such a finding has the potential to challenge the existing paradigm in cardiac research that rotors are the causal factors to maintain atrial fibrillation.
- J. Bagrow and L. Mitchell (this focus issue, Chaos **28**, 075304 (2018)) discuss a model of the social flow of written information (“The quoter model: A paradigmatic model of the social flow of written information”). The quoter model simulates the posting and sharing of short social media posts where information propagates over a graph via a quoting mechanism. The authors provide an in-depth information-theoretic analysis with analytical derivations of information flow validated by numerical experiments. The model can serve as a benchmark for how information flow measures applied to text deal with spurious interactions and confounds.
- E. M. Bollt (this focus issue, Chaos **28**, 075309 (2018)) presents a new perspective on information flow in terms of directly inspecting the associated transfer operators. Since entropy is fundamentally measured in terms of inspecting the underlying probability distributions, the argument here is that information flow should consider the evolution of such densities, and therefore two competing versions of the evolutions of densities are inspected, one without and one with considering the possibility of the outside influence of an extra factor. These two competing possibilities of closed versus open as it turns out are elegantly considered in terms of either a standard Frobenius-Perron operator for the autonomous (closed) deterministic system or a stochastic kernel for the corresponding stochastic Frobenius-Perron operator for the nonautonomous (open) dynamical system, whereby the ensemble of collective perturbations can be thought of as a random influence. Then, directly measuring the differences between these kernels leads to the need for a Jensen-Shannon divergence due to proper inclusion, and this is called the forecastability quality metric.
- The paper by J. Garland, A. M. Berdahl, J. Sun, and E. M. Bollt (this focus issue, Chaos **28**, 075308 (2018)), “Anatomy of leadership in collective behaviour,” studies causality and information flow in the form of leadership in collective behaviour of mobile animal groups, which is facilitated by the recent emergence of large datasets of trajectory time series of individuals in animal groups. Starting from the insight that heterogeneous individuals in such groups feature different types of influence over group behaviour, the authors develop an anatomy of leadership and provide a framework for evaluating and discussing leadership and models of animal group behaviour.
- In “Causality and information flow with respect to predictability,” by X. S.-Liang (this focus issue, Chaos **28**, 075311 (2018)), the author continues in advancing the Liang-Kleeman formalism,^{30,31} which is a rigorous formalism for information transfer assuming the system’s dynamics is analytically described. Here, however, important thermodynamic issues related to over a time epoch T are related in terms of Shannon information despite arbitrary dimensionality, and the property of so-called “nil causality” is considered, whereby classical methods may fail to verify.
- U. Ozturk, N. Marwan, O. Korup, H. Saito, A. Agarwal, M. J. Grossman, M. Zaiki, and J. Kurths (this focus issue, Chaos **28**, 075301 (2018)) use event synchronization to

construct complex networks of information flow to track extreme rainfall over Japan and surrounding seas (paper “Complex networks for tracking extreme rainfall during typhoons”). Directionality of information flow in the network is used to analyze regional sources and sinks of extreme weather patterns. In addition, the paper found, among several interesting results derived from a network and information flow perspective of the system, that for typhoon seasons, extreme rainfall tends to follow the southwest-northeast motion of typhoons and mean rainfall gradient of Japan.

- M. Paluš, A. Krakovská, J. Jakubík, and M. Chvosteková (this focus issue, Chaos **28**, 075307 (2018)) set out to investigate the interesting topic of how measured causality would change if one were to reverse the temporal order of observations (paper “Causality, dynamical systems and the arrow of time”). They found, using several distinct methods as ways for measuring causality, that Granger’s first principle of causality which states that “The cause occurs before the effect” can in fact be violated for chaotic systems under time reversal. Even though such a violation occurs only in hypothetical situations since chaotic processes in the real-world are not reversible, it does send a warning signal for practitioners who wish to detect causality in experimental data to not rely solely on a single measure of causality, but instead consider additional analysis, such as tests for nonlinearity, synchronization, as well as spectral and time-frequency analysis.
- S. Roy and B. Jantzen (this focus issue, Chaos **28**, 075305 (2018)) present a novel method for detecting the direction of influence between two nonlinearly coupled dynamical processes (“Detecting causality using symmetry transformations”). The method utilizes the property of dynamical symmetries, which can be considered as the set of transformations of the system trajectories that commute with its time evolution. Comparisons with transfer entropy and convergent cross mapping show that the method is especially robust in the presence of observational noise. Currently, the method is limited to first-order autonomous systems leaving extensions to the important multivariate case to future work.
- The paper by J. Runge, “Causal network reconstruction from time series: From theoretical assumptions to practical estimation” (this focus issue, Chaos **28**, 075310 (2018)), offers a comprehensive computational and theoretical perspective on what can be learned from observed experimental multivariate time series. The paper discusses the underlying assumptions and computational issues of the broad framework of conditional independence-based methods that encompasses Granger causality, transfer entropy, optimal causation entropy, and momentary information transfer, among others. Which effects occur if important assumptions are not satisfied, such as those due to unobserved variables, sampling issues, determinism, stationarity, nonlinearity, or measurement error? How are causal reconstructions affected by computational issues due to autocorrelation and high dimensionality? The article is

intended to briefly review and accessibly illustrate the foundations and practical problems of time series-based causal discovery and stimulate further methodological developments.

- In, “Local causal states and discrete coherent structures,” Rupe and Crutchfield (this focus issue, Chaos **28**, 075312 (2018)) study computational mechanics and causal influences in spatiotemporal processes to extract so-called local causal states and uncover “local symmetries.” This causal approach is an original departure from many other set-oriented methods, more often used in the study of coherent structures, with applications here in terms of cellular automata for a rigorous discussion, but there is an outlook of real world application of this new approach.
- The paper by Smirnov, “Transient and equilibrium causal effects in coupled oscillators” (this focus issue, Chaos **28**, 075303 (2018)), notably describes nonequilibrium effects as contrasted to equilibria effects from the viewpoint of causal influences. Coupling parameter variations and also on-off switching of coupling are considered in the context of Wiener-Granger causality, and two kinds of influences are categorized for unidirectional couplings.

ACKNOWLEDGMENTS

This work was funded in part by the Simons Foundation (Grant No. 318812), the U.S. Army Research Office (Grant No. W911NF-16-1-0081), the U.S. Office of Naval Research (Grant No. N00014-15-1-2093), and also DARPA. J.R. was funded by a postdoctoral fellowship by the James S. McDonnell foundation.

¹J. Arnold, P. Grassberger, K. Lehnertz, and C. E. Elger, “A robust method for detecting interdependences: Application to intracranially recorded EEG,” *Physica D* **134**(4), 419–430 (1999).

²L. Arnold, *Stochastic Differential Equations* (Wiley-Interscience, New York, 1974).

³L. Barnett, A. B. Barrett, and A. K. Seth, “Granger causality and transfer entropy are equivalent for Gaussian variables,” *Phys. Rev. Lett.* **103**(23), 238701 (2009).

⁴A. Bhattacharya, “Carbon dioxide drove climate change during longest interglacial,” *Eos Trans. Am. Geophys. Union* **93**(37), 360–360 (2012).

⁵E. M. Boltt, “Synchronization as a process of sharing and transferring information,” *Int. J. Bifurcat. Chaos* **22**(11), 1250261 (2012).

⁶E. M. Boltt and N. Santitissadeekorn, *Applied and Computational Measurable Dynamics* (SIAM, 2013).

⁷S. L. Bressler and A. K. Seth, “Wiener–Granger causality: a well established methodology,” *Neuroimage* **58**(2), 323–329 (2011).

⁸C. Cafaro, W. M. Lord, J. Sun, and E. M. Boltt, “Causation entropy from symbolic representations of dynamical systems,” *Chaos: Interdiscip. J. Nonlinear Sci.* **25**(4), 043106 (2015).

⁹T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).

¹⁰P. De Michelis, G. Consolini, M. Materassi, and R. Tozzi, “An information theory approach to the storm-substorm relationship,” *J. Geophys. Res.* **116**(A8), A08225 doi:10.1029/2011JA016535 (2011).

¹¹T. Dimpfl and F. J. Peter, “Using transfer entropy to measure information flows between financial markets,” *Stud. Nonlinear Dyn. Econ.* **17**(1), 85–102 (2013).

¹²M. Ding, Y. Chen, and S. L. Bressler, “Granger causality: basic theory and application to neuroscience,” in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, edited by S. Schelter, M. Winterhalder, and J. Timmer (2006), pp. 437–460.

¹³I. Ebert-Uphoff and Y. Deng, “Causal discovery for climate research using graphical models,” *J. Clim.* **25**(17), 5648–5665 (2012).

- ¹⁴I. Ebert-Uphoff and Y. Deng, "A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer," *Geophys. Res. Lett.* **39**(19), L19701 (2012).
- ¹⁵I. Ebert-Uphoff and Y. Deng, "Causal discovery from spatio-temporal data with applications to climate science," in *13th International Conference on Machine Learning and Applications (ICMLA), 2014* (IEEE, 2014), pp. 606–613.
- ¹⁶C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica* **37**(3), 424–438 (1969).
- ¹⁷D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik, "Topological causality in dynamical systems," *Phys. Rev. Lett.* **119**(9), 098301 (2017).
- ¹⁸S. He, X. Zheng, D. Zeng, K. Cui, Z. Zhang, and C. Luo, "Identifying peer influence in online social networks using transfer entropy," in *Pacific-Asia Workshop on Intelligence and Security Informatics* (Springer, 2013), pp. 47–61.
- ¹⁹D. F. Hendry, "The Nobel Memorial Prize for Clive W. J. Granger," *Scand. J. Econ.* **106**(2), 187–213 (2004).
- ²⁰Y. Hirata and K. Aihara, "Identifying hidden common causes from bivariate time series: A method using recurrence plots," *Phys. Rev. E* **81**(1), 016203 (2010).
- ²¹Y. Hirata, J. M. Amigó, Y. Matsuzaka, R. Yokota, H. Mushiaka, and K. Aihara, "Detecting causality by combined use of multiple methods: Climate and brain examples," *PLoS One* **11**(7), e0158572 (2016).
- ²²R. G. James, N. Barnett, and J. P. Crutchfield, "Information flows? A critique of transfer entropies," *Phys. Rev. Lett.* **116**(23), 238701 (2016).
- ²³R. G. James, C. J. Ellison, and J. P. Crutchfield, "Anatomy of a bit: Information in a time series observation," *Chaos* **21**(3), 037109 (2011).
- ²⁴D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, "Quantifying causal influences," *Ann. Stat.* **41**(5), 2324–2358 (2013).
- ²⁵P. Kim, J. Rogers, J. Sun, and E. Boltt, "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems," *J. Comput. Nonlinear Dyn.* **12**(1), 011008 (2017).
- ²⁶A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E* **69**(6), 066138 (2004).
- ²⁷M. Kretschmer, D. Coumou, J. F. Donges, and J. Runge, "Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation," *J. Clim.* **29**(11), 4069–4081 (2016).
- ²⁸M. Kretschmer, J. Runge, and D. Coumou, "Early prediction of weak stratospheric polar vortex states using causal precursors," *Geophys. Res. Lett.* **44**(16), 8592–8600 (2017).
- ²⁹A. Lasota and M. C. Mackay, *Chaos, Fractals and Noise*, Applied Mathematical Sciences Vol. 97 (Springer, 1994).
- ³⁰X. San Liang, "The Liang-Kleeman information flow: theory and applications," *Entropy* **15**(1), 327–360 (2013).
- ³¹X. San Liang and R. Kleeman, "Information transfer between dynamical system components," *Phys. Rev. Lett.* **95**(24), 244101–244101 (2005).
- ³²W. M. Lord, J. Sun, N. T. Ouellette, and E. M. Boltt, "Inference of causal information flow in collective animal behavior," *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**(1), 107–116 (2016).
- ³³H. Ma, L. Larsen, and W. Wagner, *Ecogeomorphic Feedbacks that Grow Deltas* (AGU Fall Meeting Abstracts, 2017).
- ³⁴S. Malekpour and W. A. Sethares, "Conditional Granger causality and partitioned Granger causality: differences and similarities," *Biol. Cybern.* **109**(6), 627–637 (2015).
- ³⁵S. Nemati, B. A. Edwards, J. Lee, B. Pittman-Polletta, J. P. Butler, and A. Malhotra, "Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy," *Respir. Physiol. Neurobiol.* **189**(1), 27–33 (2013).
- ³⁶J. Pearl, *Causality* (Cambridge University Press, 2009).
- ³⁷J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).
- ³⁸B. Pompe and J. Runge, "Momentary information transfer as a coupling measure of time series," *Phys. Rev. E* **83**(5), 1–12 (2011).
- ³⁹R. Quian Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Phys. Rev. E* **61**(5), 5142 (2000).
- ⁴⁰M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage* **52**(3), 1059–1069 (2010).
- ⁴¹J. Runge, "Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information," in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, 2018.
- ⁴²J. Runge, "Quantifying information transfer and mediation along causal pathways in complex systems," *Phys. Rev. E* **92**(6), 062829 (2015).
- ⁴³J. Runge, R. V. Donner, and J. Kurths, "Optimal model-free prediction from multivariate time series," *Phys. Rev. E* **91**(5), 052909 (2015).
- ⁴⁴J. Runge, J. Heitzig, N. Marwan, and J. Kurths, "Quantifying causal coupling strength: a lag-specific measure for multivariate time series related to transfer entropy," *Phys. Rev. E* **86**(6), 061121 (2012).
- ⁴⁵J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, "Escaping the curse of dimensionality in estimating multivariate transfer entropy," *Phys. Rev. Lett.* **108**(25), 258701 (2012).
- ⁴⁶J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting causal associations in large nonlinear time series datasets," 2018, preprint <http://arxiv.org/abs/1702.07007v2>.
- ⁴⁷J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, "Identifying causal gateways and mediators in complex spatio-temporal systems," *Nat. Commun.* **6**, 8502 (2015).
- ⁴⁸J. Runge, V. Petoukhov, and J. Kurths, "Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models," *J. Clim.* **27**(2), 720–739 (2014).
- ⁴⁹J. Runge, M. Riedl, A. Müller, H. Stepan, N. Wessel, and J. Kurths, "Quantifying the causal strength of multivariate cardiovascular couplings with momentary information transfer," *Physiol. Meas.* **36**(4), 813–825 (2015).
- ⁵⁰L. Sandoval, "Structure of a global network of financial companies based on transfer entropy," *Entropy* **16**(8), 4443–4482 (2014).
- ⁵¹S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," *Phys. Rev. E* **54**(6), 6708 (1996).
- ⁵²C.-F. Schleussner, J. Runge, J. Lehmann, and A. Levermann, "The role of the North Atlantic overturning and deep ocean for multi-decadal global-mean-temperature variability," *Earth Syst. Dyn.* **5**(1), 103–115 (2014).
- ⁵³T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.* **85**(2), 461 (2000).
- ⁵⁴C. Sturtevant, B. L. Ruddell, S. H. Knox, J. Verfaillie, J. H. Matthes, P. Y. Oikawa, and D. Baldocchi, "Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange," *J. Geophys. Res.: Biogeosci.* **121**(1), 188–204 (2016).
- ⁵⁵A. S. Ambedgedara, J. Sun, K. Janoyan, and E. Boltt, "Information-theoretical noninvasive damage detection in bridge structures," *Chaos* **26**(11), 116312 (2016).
- ⁵⁶G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *Science* **338**(6106), 496–500 (2012).
- ⁵⁷H. Sumioka, Y. Yoshikawa, and M. Asada, "Causality detected by transfer entropy leads acquisition of joint attention," in *IEEE 6th International Conference on Development and Learning, 2007 (ICDL 2007)* (IEEE, 2007), pp. 264–269.
- ⁵⁸J. Sun and E. M. Boltt, "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings," *Physica D* **267**, 49–57 (2014).
- ⁵⁹J. Sun, D. Taylor, and E. M. Boltt, "Causal network inference by optimal causation entropy," *SIAM J. Appl. Dyn. Syst.* **14**(1), 73–106 (2015).
- ⁶⁰F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. A. Rand and L.-S. Young (Springer, 1981), Vol. 898, pp. 366–381.
- ⁶¹T. Q. Tung, T. Ryu, K. H. Lee, and D. Lee, "Inferring gene regulatory networks from microarray time series data using transfer entropy," in *Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007 (CBMS'07)* (IEEE, 2007), pp. 383–388.
- ⁶²G. Ver Steeg and A. Galstyan, "Information transfer in social media," in *Proceedings of the 21st International Conference on World Wide Web (ACM, 2012)*, pp. 509–518.
- ⁶³P. Walters, *An Introduction to Ergodic Theory* (Springer Science & Business Media, 2000), Vol. 79.
- ⁶⁴S. Wan, J. Zhu, Z. Sun, L. Zhang, and Y. Dai, "Regional characteristics of subtropical high impact on precipitation and temperature in China," *Int. J. Clim. Change* **3**(3), 113–118 (2012).
- ⁶⁵H. White, "The entropy of a continuous distribution," *Bull. Math. Biophys.* **27**(1), 135–143 (1965).
- ⁶⁶M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser, "Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks," *Prog. Biophys. Mol. Biol.* **105**(1–2), 80–97 (2011).

⁶⁷M. Wibral, R. Vicente, and J. T. Lizier, *Directed Information Measures in Neuroscience* (Springer, 2014).

⁶⁸N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, edited by E. F. Beckenbach (McGraw Hill, New York 1956), pp. 165–190.

⁶⁹J. Y. Zhu, Y. Zheng, X. Yi, and V. O. K. Li, "A Gaussian Bayesian model to identify spatio-temporal causalities for air pollution based on urban big data," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2016* (IEEE, 2016), pp. 3–8.