Open or closed? Information flow decided by transfer operators and forecastability quality metric

Erik M. Bollt

Open or closed? Information flow decided by transfer operators and forecastability quality metric

Erik M. Bollt[a)]

Department of Mathematics and Electrical and Computer Engineering and Clarkson Center for Complex Systems Science (C^3S^2), Clarkson University, Potsdam, New York 13699, USA

A basic systems question concerns the concept of closure, meaning autonomy (closed) in the sense of describing the (sub)system as fully consistent within itself. Alternatively, the system may be nonautonomous (open), meaning it receives influence from an outside subsystem. We assert here that the concept of information flow and the related concept of causation inference are summarized by this simple question of closure as we define herein. We take the forecasting perspective of Weiner-Granger causality that describes a causal relationship exists if a subsystem's forecast quality depends on considering states of another subsystem. Here, we develop a new direct analytic discussion, rather than a data oriented approach. That is, we refer to the underlying Frobenius-Perron (FP) transfer operator that moderates evolution of densities of ensembles of orbits, and two alternative forms of the restricted Frobenius-Perron operator, interpreted as if either closed (deterministic FP) or not closed (the unaccounted outside influence seems stochastic and we show correspondingly requires the stochastic FP operator). Thus follows contrasting the kernels of the variants of the operators, as if densities in their own rights. However, the corresponding differential entropy comparison by Kullback-Leibler divergence, as one would typically use when developing transfer entropy, becomes ill-defined. Instead, we build our Forecastability Quality Metric (FQM) upon the "symmetrized" variant known as Jensen-Shannon divergence, and we are also able to point out several useful resulting properties. We illustrate the FQM by a simple coupled chaotic system. Our analysis represents a new theoretical direction, but we do describe data oriented directions for the future. *Published by AIP Publishing.* https://doi.org/10.1063/1.5031109

Causation inference in the sense of G-causality (Granger causality) refers to the concept of reduction of variance. That is, to answer the basic question, does system X allow for sufficient information regarding forecasts of future states of system X or is there improved forecasts with observations from system Y. If the latter is the case, then we declare that X is not closed, as it is receiving influence, or information, from system Y. Such a reduction of uncertainty perspective of causal influence is not identical to the concept of allowing perturbations and experiments on two systems to decide what changes influence each other. Methods, such as Ganger causality, transfer entropy, causation entropy, and even cross correlation method, are premised on the concept of alternative formulations of the forecasting question, with and without considering the influence of an external state. Thus, the idea is to decide if a system is open or closed. Here we assert that the underlying transfer operator, called the Frobenius-Perron operator, that moderates not the evolution of single initial conditions but the evolution of ensembles of initial conditions allows for a direct and sensible analysis of information flow to decide the question of open or closed. Note that a restricted form of the transfer operator to a subsystem, queried either with or without the states of the other subsystem(s), allows for a new analytically tractable formulation of the question of information flow. In this philosophy, the exterior system becomes an "unknown" influence onto the interior system. Therefore, it becomes formulated as a stochastic influence with a corresponding stochastic transfer operator. In this philosophy, it becomes clear that even though the exterior system may be deterministic, it appears stochastic within focus on the interior system. As an explicit measurement for this concept, we build a Forecastability Quality Metric (FQM) based on Jensen-Shannon divergence, applied directly to alternative forms of the transfer operator, noting that a transfer entropy like application of Kullback-Leibler divergence would be impossible. However, this choice of metric like measurement allows for several especially elegant properties that we annunciate here. Application is described and future empirical directions are described.

I. INTRODUCTION

We assert that a basic question when defining the concept of information flow is to contrast versions of reality for a dynamical system. Either a subcomponent is closed or alternatively there is an outside influence due to another component. The details of how this question is posed and how it is decided gives rise to various versions of concepts of information flow, which are related to causation inference. This includes the celebrated Nobel work[1] behind Granger Causality[2–4] and closely related by Wiener.[5] The popular transfer entropy,[6,7] follows this logic, but the related causation entropy

a) Electronic mail: bolltem@clarkson.edu

furthermore uncovers the differences between direct and indirect influences.[8–11] Finally we mention direct forecast methods include Convergent Cross-Mapping method (CCM).[12] We shall generally interpret the problem causation inference, as estimated from observed data, to relate to the question of reduction of uncertainty associated with forecasts; that is, we ask if a subcomponent may be forecasted well on its own, or rather if a fuller model allowing for external variables provides for better forecasts. If the latter, then the subsystem is not closed since it must be receiving outside information. Information flow as a concept of reduction of uncertainty is often discussed as related to concepts of causation. Causation inference has overlapping philosophical roots,[13] and we have also allowed our own previous writings on these topics to overlap these distinct concepts, but here in this paper we will simply discuss information flow as a form of reduction of uncertainty. In fact, there is a beautiful connection between Granger causality and transfer entropy in the special case of Gaussian noise.[4] Furthermore, in Ref. 4 there is distinguished the concept of Weiner-Granger causality (G-causality)[2,5] that between two inter-dependent variables, X and Y, in a statistical sense "Y G-causes X," if measurements of Y can improve forecast of future values of measurements of X better than would be possible by measurements of X alone; this is what we mean by the reduction of uncertainty and this is the nonintervention philosophy that we will maintain here. This perspective is in contrast to the related but distinct concept of "physically instantiated causal relationship" in a sense that can only be truly uncovered by perturbations (also called interventions) to the system, as the statistics of causation by interventions and observations described in Pearl's extensive work enunciates.[14]

Most studies on information flow are in terms of data and the statistical inference concepts cited above,[2–4] sometimes by information theoretic methods.[6–11,15–20] Notably, however, see Liang-Kleeman,[21] as a more analytical approach that involves both the dynamical system and the concept of information flow and also leading to transfer operators. There is an important distinction in approach in that the Liang-Kleeman approach considers the intervention whereby one of the variables is held fixed, whereas we consider here the possible absence of the external variable; as such, our results are not identical but we do find both questions interesting. Note also our own prior work relates synchronization as a process of sharing information.[22] This current work then builds on Ref. 22 that we refer directly to transfer operators when describing the degree to which a system may or may not be open. That we work directly with the transfer operators is perhaps the most significant difference to previous approaches leading to transfer entropy, but also we will point out that there is a nuanced difference how this relates to the associated conditional probabilities, and then correspondingly a necessary difference in which kind of information divergence may be used.

In this paper, we describe a new approach formalism of analysis of the underlying concept of reduction of uncertainty in terms of evolution of densities. The question of how ensembles (densities) of initial conditions evolve under orbits of a dynamical system is handled by the Frobenius-Perron operator that is the dynamical system on the associated space of densities.[7,23] Within this framework of transfer operators, we may recast the question of information flow by more rigorously presenting the two versions of the basic question, which is to decide one of the two alternatives:

- Is the subsystem closed?
- Does the subsystem receive influence from another subcomponent?

Our own previous work considered the relationship of evolution of densities as moderated by the Frobenius-Perron operator, together with the information theoretic question of information flow by transfer entropy.[7,22] However, the details of our previous work were discussed in terms of estimating the associated probability density functions (pdf's) at steady state, and furthermore, through estimation of the transfer operator's action on the space of densities by the famous Ulam-Galerkin's methods of projection on to a linear subspace, Δ_N, as $P : L^2(\Omega) \rightarrow \Delta_N \subset L^2(\Omega)$ to describe finite matrix computations. There is a long history to the Ulam's method,[7,23–33] but this approach generally relies on covering the space with boxes and estimating probabilities in a histogram-like fashion at steady state so that the estimations can be statistically stationary. This current work takes a significant departure of the theme of steady state, and we do so directly within the scope of transfer operators by a new interpretation of external influences described analytically as a random variable like term.

A unique outcome of our study is that attempting to use the Kullback-Leibler divergence, D_{KL}, analogously to how it is done when developing transfer entropy, but here we wish to examine directly the kernel of the transfer operator, leads to an unbounded measure. Instead, we use the so-called symmetrized version of KL-divergence, called the Jensen-Shannon divergence, D_{JS}. Not only does this approach fix an otherwise unpleasant nonconvergence problem, but also it brings with it several beautiful new properties and interpretations that underlie the theory special to the JS divergence. With these new interpretations in mind, we call this variant of information flow, the Forecastability Quality Metric, written $FQM_{y \rightarrow x}$ in analogy to the notation one uses typically for transfer entropy, $T_{y \rightarrow x}$ between subsystem y and subsystem x.

The work presented herein could be considered theoretical in nature, marrying the theories of transfer operators, statistics, and information theory in a unique way to well define a concept of information flow in terms defining the difference between closed and not closed subsystems. Thus in standing up a new perspective of these questions within the formal language of these disparate fields, we hope to better sharpen the general understanding of these important questions. Nonetheless, we will point out at the end of the paper directions in which this perspective could be turned into a data oriented methodology.

II. BASIC PROBLEM SETUP

A most basic version of the discussion of a full system with subcomponents follows consideration of two coupled

oscillators:

$$x_{n+1} = f_1(x_n) + \epsilon_1 k(x_n, y_n),$$
$$y_{n+1} = f_2(y_n) + \epsilon_2 k(y_n, x_n). \tag{1}$$

We might ask if the "x-oscillator" is "talking to" the "y-oscillator," and vice versa. Defining the concept of "talking to" may be made in various forms. Avoiding philosophical notions, we take the perspective of predictability, by asking if x variables improve forecasts of future states of y-variables better than considering just y variables alone, in the sense of reduction of uncertainty, thus G-causality.

Now we recast the typical symmetrically coupled problem, Eq. (1), to a general form of partitioned dynamical systems on a skew product space $X \times Y$,

$$T : \Omega_X \times \Omega_Y \to \Omega_X \times \Omega_Y. \tag{2}$$

This emphasizes that the full system is a single dynamical system where the phase space is a skew product space, so examples such as Eq. (1) discuss information flow between the Ω_X and Ω_Y states. In this notation then, the two component coupled dynamical systems of the x and y component variables may be written

$$T(x, y) = [T_x(x, y), T_y(x, y)], \tag{3}$$

where

$$T_x : X \times Y \to X$$
$$x_n \mapsto x_{n+1} = T_x(x_n, y_n),$$
$$T_y : X \times Y \to Y$$
$$y_n \mapsto y_{n+1} = T_y(x_n, y_n). \tag{4}$$

In the case of Eq. (1), let

$$T_x(x_n, y_n) = f_1(x_n) + \epsilon_1 k(x_n, y_n) \quad \text{and}$$
$$T_y(x_n, y_n) = f_2(y_n) + \epsilon_2 k(y_n, xn). \tag{5}$$

The notation $x \in \Omega_X$ and $y \in \Omega_Y$ allows that each may be vector valued and generally even differing dimensionality. We write $\Omega = \Omega_X \times \Omega_Y$, but sometimes in the subsequent we will write Ω as the phase space of an unspecified transformation, and these phase spaces will also serve conveniently as outcome spaces when describing the dynamical systems as stochastic processes.

III. INFORMATION FLOW AS ALTERNATIVE VERSIONS OF FORECASTS IN PROBABILITY

Information flow is premised on a simple question of comparing alternative versions of forecasts, stated probabilistically. We ask the question as to if two different probability distributions are the same, or different, which can be stated[7]

$$P(x_{n+1}|x_n) \stackrel{?}{=} P(x_{n+1}|x_n, y_n), \tag{6}$$

and if they are different, the degree to which they are different. This describes a degree of deviation from a Markov-property. This statement as contrasted to Eq. (32) to come is a key difference in our measure, the FQM as derived directly from comparing contrasting version of transfer operator kernels, versus the transfer entropy (TE) information flow question highlighted here in Eq. (6).

A. Information flow as transfer entropy

Specifically, the transfer entropy[6] measures deviation from the Markov-property question, Eq. (6) using the Kullback-Leibler divergence

$$T_{y \to x} = D_{KL}[p(x_{n+1}|x_n)||p(x_{n+1}|x_n, y_n)], \tag{7}$$

in terms of the probability distributions associated with the probabilities of Eq. (6). A useful outcome in using this entropy-based measure to describe deviation from Markovness is that the answer is naturally describing information flow in units of bit/second. In subsequent sections, we will point out problems in the Kullback-Leibler divergence that are solved by answering the same question with the Shannon-Jensen divergence instead, with some lovely special properties to also follow. Generally, the transfer entropy was defined[6] in terms of k-previous states in each variable, but we take this simplification to single prior states to be associated with the related problem of true embedding in delay variables.[34–36]

Note that the probability density functions written in Eqs. (6) and (7) are not generally the same functions. Furthermore, they need not be assumed to be steady state probabilities; this is an important distinction in the course of this paper as departure from many previous works in information flow. Instead generally consider them as nonequilibrium functions representing the state of probabilities of ensembles of orbit states (x_n, y_n), at time-n, following a random ensemble of initial states (x_0, y_0) but observed at time n.

Here, we will keep with the description that the outcome spaces may be continuous and state the differential entropy version of a Kullback-Leibler divergence definition for transfer entropy. A general definition that suits our purposes is as follows. Let outcome space Ω have a measure μ, so that probability measures P_1 and P_2 are absolutely continuous to μ, so that $p_1 = \frac{dP_1}{d\mu}$ and $p_2 = \frac{dP_2}{d\mu}$, then $D_{KL}(P_1||P_2)$ may be written

$$D_{KL}(P_1||P_2) = \int_\Omega p_1 \log \frac{p_1}{p_2} d\mu = -h(p_1) - \int_\Omega p_1 \log p_2 d\mu, \tag{8}$$

using the standard notation for differential entropy, $h(p_1) = \int_\Omega p_1 \log = p_1 = d\mu$. We will allow the abuse of notation to write the KL-divergence in terms of the pdf's as the arguments, $D_{KL}(p_1||p_2)$. Therefore, when there are continuous state spaces, let

$$T_{y \to x} = D_{KL}[p(x_{n+1}|x_n, y_n)||p(x_{n+1}|x_n)]$$
$$= \int_\Omega p(x_{n+1}|x_n, y_n)[\log p(x_{n+1}|x_n, y_n)$$
$$- \log p(x_{n+1}|x_n)]d\Omega, \tag{9}$$

and in this integral, $\Omega = X_n \times X_{n+1} \times Y_n$. The expression for $T_{x \to y}$ is similar,

$$T_{x \to y} = D_{KL}[p(y_{n+1}|x_n, y_n)||p(y_{n+1}|y_n)]. \tag{10}$$

There is, however, a significant technical difficulty with using the Kullback-Leibler divergence in this way as generally $D_{KL}(p_1||p_2)$ is only bounded if the support of p_1 is contained in the support of p_2.[37] This turns out to be untrue in

a natural interpretation that follows when directly approaching the description of the densities by the kernel of the transfer operators. This will motivate our fix to the problem by developing the $FQM_{y \to x}$. Also the usual interpretation to assign $0 \log \frac{0}{0} = 0$ is useful here to emphasize continuity.

At this step, it is important to point out that there is a significant technical difficulty with using the Kullback-Leibler divergence in this way as is a theorem that $D_{KL}(p_1 || p_2)$ is only bounded if the support of p_1 is contained in the support of p_2.[37] This is easy to see since, if $support(p_2) \subsetneq support(p_1)$ where support denotes the set where the function is nonzero, then there are values x such that $p_1(x) \log[p_1(x)/p_2(x)] = p_1(x) \log[p_1(x)] - p_1(x) \log[p_2(x)]$, but $\log[p_2(x)]$ is not defined when $p_2(x) = 0$. Normally this may not be a problem, for example, in standard application by transfer entropy, but this important detail turns out to arise in a natural interpretation of transfer entropy that follows when directly applying the description of the densities by the kernel of the transfer operators. This can be seen in the illustration of example variants of the kernel functions already in Fig. 2. On the other hand for standard transfer entropy or also the Liang-Kleeman formalism, the issue does not arise as by neither approach is the KL-divergence applied to the kernel directly as it is here. This issue will motivate our fix to the problem by developing the $FQM_{y \to x}$. Also, the usual interpretation, to assign $0 \log \frac{0}{0} = 0$, is useful here to emphasize continuity.

IV. EVOLUTION OF DENSITIES OF INITIAL CONDITIONS BY THE FROBENIUS-PERRON OPERATOR

The evolution of single initial conditions proceeds by the mapping $T : \Omega_X \times \Omega_Y \to \Omega_X \times \Omega_Y$. But the evolution of many initial conditions all at once follows evolution of ensemble densities of many states both before and after the mapping is applied. The Frobenius-Perron operator is defined to describe the associated dynamical system for evolution of densities. First, we review this theory, and then we will specialize the concepts to both the full problem and the marginalized problem, both considering with and without the coupling term. What is especially new here is that in the coupled case, the coupling influence of the other dynamical system enters in a way that may be interpreted as a stochastic perturbation, so associated to the stochastically perturbed Frobenius-Perron operator.

A. The deterministic Frobenius-Perron operators

Remarkably, even considering a nonlinear dynamical system

$$F : M \to M, \tag{11}$$

the one-step action of the map in the space of (ensembles of initial conditions) densities is that of a linear transfer operator,[7,23] for a phase space, $M \subset \mathbb{R}^n$. The Frobenius-Perron operator generates an associated linear dynamical system on the space of densities,

$$P_F : L^1(M) \to L^1(M), \tag{12}$$

defined by

$$P_F[\rho](x') = \int_M \delta[x - F(y)]\rho(y)dy = \sum_{\{x:F(x)=x'\}} \frac{\rho(x)}{|F'(x)|}, \tag{13}$$

where the sum is taken over all pre-images, s, when the map has a multiple branched "inverse." Note also that in the case of a multi-variate transformation F, $m > 1$, then the term $\sum_{\{x:F(x)=x'\}} \frac{\rho(x)}{|F'(x)|}$ is instead replaced by $\sum_{\{x:F(x)=x'\}} \rho(x)|DF^{-1}(x)|$, meaning that the determinant of the Jacobian derivative matrix of the inverse of the map must be used. While this infinite dimensional operator is typically not realizable in closed form, except for special cases,[7,23] there are matrix-methods in terms of approximating the action of the transformation as a stochastic matrix, and weak convergence to the true invariant density is called Ulam's method,[25,38–41] as a technique to project this operator to a finite dimensional linear subspace $\Delta_N \subset L^2(M)$ generated by the set characteristic functions supported over the partitioning grid.[25] The idea is that refining the grid yields weak approximants of invariant density. The projection is exact when the map is "Markov" using basis functions supported on the Markov partition.[42–44] Roughly speaking, the infinitesimal transfer operator[45]

$$\mathcal{L}(x', x) = \delta[x' - F(x)], \tag{14}$$

when integrated over a grid square B_i which are small enough so that $DF(x)$ is approximately constant, and then this action is approximated by a constant matrix entry $S_{i,j}$. Under special assumptions on F, statements concerning quality of the approximation can be made rigorous. Recently, many researchers have been using Ulam's method to describe global statistics of a dynamical system,[39–41,46,47] such as invariant measure, Lyapunov exponent, dimension, etc. A point of this paper is to get away from three major aspects of this kind of computation which are as follows:

1. The estimations based on the finite rank matrix computations.
2. The statistical approximations based on estimation of the matrix entries.
3. The inherently steady state stationarity concept assumptions for collecting the statistics of $S_{i,j}$; those assumptions were previously built into our own Ulam-Galerkin based approach to transfer entropy by Frobenius-Perron operator methods.[22]

Instead, our descriptions will be in terms of the full integral describing the transfer operator adapted to notions of information flow, with no underlying assumption of steady state.

B. The stochastic Frobenius-Perron operators

Now consider the stochastically perturbed dynamical system,

$$F_g : M \to M, \tag{15}$$

$$x \mapsto F(x) + y, \tag{16}$$

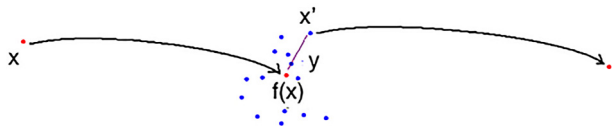where y is a random variable with pdf g, which is applied once

FIG. 1. Stochastic iteration, Eq. (15). At time n, $x \mapsto F(x)$ "deterministically" and then also by convention at the same time n, random value y is added to yield $x' = F(x) + y$ now at time, $n + 1$. A multiplicative scenario can be handled comparably, according to Eqs. (18) and (19).

per each iteration. See Fig. 1 where we illustrate this simple additive stochastic iteration, where we describe that x evolves deterministically to $F(x)$ and then a "random" value of y is added, which we describe by convention as if at the same time instant at time step, n. Then $x' = F(x) + y$ denotes the value at time $n + 1$. Multiplicative can also be handled, according to Eqs. (18) and (19). The usual assumption at this stage is that the realizations y_n of y added to subsequent iterations form an i.i.d. (identical independently distributed) sequence, but since we are allowing for just one application of the dynamic process, the assumption is not necessary, and g maybe simply be the distribution of y_n at time n. If x is relatively small to x', then the deterministic part F has primary influence, but this is not even a necessary assumption for this stochastic Frobenius-Perron operator formalism. Neither is a standard assumption for many stochastic analyses that require certain forms of the noise term, such as Gaussian distributed, as we do not require anything other than g is a measurable function, which likely is the weakest kind of assumption possible. The "stochastic Frobenius-Perron operator" has a similar form to the deterministic case[7,23]

$$P_{F_g}[\rho](x) = \int_M g[x' - F(x)]\rho(x)dx. \quad (17)$$

It is interesting to compare this integral kernel to the delta function in Eq. (13). Now a stochastic kernel describes the pdf of the noise perturbation. We denote the stochastic Frobenius-Perron operator to be P_{F_g}, vice P_F for no noise version in Eq. (13). In the case that the random map Eq. (15) arises

from the usual continuous Langevin process, the infinitesimal generator of the Frobenius-Perron operator for Gaussian g corresponds to a general solution of a standard Fokker-Planck equation.[23]

Within the same formalism, we can also study multiplicative noise,

$$z \to \eta\tau(x), \quad (18)$$

(modeling parametric noise). It can be proved[7,48] that the kernel-type integral transfer operator is

$$\mathcal{K}(z, s) = g[z/F(s)]/F(s). \quad (19)$$

More generally, the theory of random dynamical systems[49] classifies those random systems which give rise to explicit transfer operators with corresponding infinitesimal generators, and there are well defined connections between the theories of random dynamical systems and of stochastic differential equations.

V. INTERPRETING CLOSURE BY EVOLUTION OF DENSITY IN TERMS OF TRANSFER OPERATORS

Consider now the Frobenius-Perron operator Eq. (13), term by term, as associated with relevant conditional and joint probabilities. First, let $y = x' - F(x)$, which upon substitution into Eq. (13) yields the following simplifications. The notation relates to when the stochastic process interpretation of the variables take values, $X_{n+1} = x'$, $X_n = x$, and $Y_n = y$. The substitution yields

$$\overline{\rho}(x') = P_F[\rho](x') = \int_M \delta[x' - F(x)]\rho(x)dx$$

$$= \int_M \frac{\delta(y)}{F'[F^{-1}(x' - y)]}\rho[F^{-1}(x' - y)]dy. \quad (20)$$

By a similar computation, with the same substitution, the stochastic version of the Frobenius-Perron operator, Eq. (17)
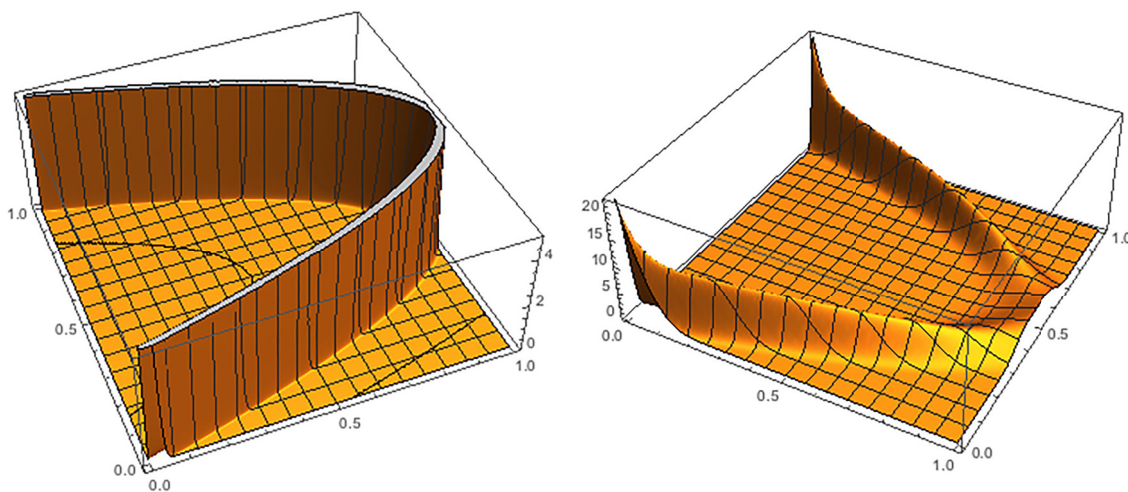

FIG. 2. Contrasting the kernel functions, (left) $\delta[x - F(s)]$ and (right) $g[x - F(s)]$, for contrasting versions of the transfer operator corresponding to a logistic map, Eqs. (43)–(45), for closed versus open versions of the concept of the primary question of the associated conditional probabilities question in Eqs. (32) and (33) as decided by the $FQM_{y \to x}$.

can be written as

$$\overline{\rho}(x') = P_{F_g}[\rho](x) = \int_M g[x' - F(x)]\rho(x)dx$$

$$= \int_M \frac{g(y)}{F'[F^{-1}(x' - y)]}\rho[F^{-1}(x' - y)]dy. \quad (21)$$

(Again if the transformation is multivariate, then the determinant of the Jacobian, or so-called Wronskian, must be used, $|DF^{-1}|$.) Now we have written the new distribution of points as, $\overline{\rho}(x)$, evaluated at a point $x \in M$. Notice that these Eqs. (20) and (21) are essentially the same in the special case that the distribution g is taken to be a delta function, as if the noise limits to a zero variance, in the sense of weak convergence.

Let us interpret these pdf's as describing probabilities as follows. It is useful at time n to associate

$$P[X_n \in (x, x + dx)] = \rho(x)dx, \quad (22)$$

and

$$P[X_{n+1} \in (x', x' + dx')] = \overline{\rho}(x')dx', \quad (23)$$

and $(x', x' + dx')$ may denote small measurable sets containing at x' in the general multivariate scenario.

Take ρ to be the probability distribution associated with samples of the ensemble of points along orbits, at time n and likewise $\overline{\rho}$, at time $n + 1$. Interpreted in this way as a stochastic system (where the randomness is associated with the initial selection from the ensemble) depends on which version of the dynamics (with or without randomness, upon iteration) whether version Eq. (11) or Eq. (15).

Recall that since [by general conditional probability formula, $P(A|B) \cdot P(B) = P(A, B)$], or a chain statement for compound events,

$$P(A, B, C) = P(A|B, C) \cdot P(B|C) \cdot P(C). \quad (24)$$

Then let events be defined

$$A = [X_{n+1} = x'],$$
$$B = [X_n = F^{-1}(x' - y)], \quad (25)$$
$$C = [Y_{n+1} = y].$$

Again we refer to Fig. 1 for the notation. For convenience, we will now drop the formal descriptions of small intervals as dx, dx', dy and the careful notation of probability events in intervals, as noted in Eqs. (22) and (23). So more loosely in notation now, we describe

$$P[X_{n+1} = x'|X_n = F^{-1}(x' - y), Y_n = y]$$
$$\cdot P[X_n = F^{-1}(x' - y)|Y_n = y] \cdot P(Y_n = y)$$
$$= \frac{1}{F'[F^{-1}(x' - y)]} \cdot \rho[F^{-1}(x' - y) \cdot g(y)], \quad (26)$$

with the interpretation,

$$P[X_{n+1} = x'|X_n = F^{-1}(x' - y), Y_n = y] = \frac{1}{F'[F^{-1}(x' - y)]}, \quad (27)$$

$$P[X_n = F^{-1}(x' - y)|Y_n = y] = \rho[F^{-1}(x' - y)], \quad (28)$$

$$P(Y_{n+1} = y) = g(y), \quad (29)$$

(but not necessarily normalized). The rigorous details behind this interpretation bring us into the functional analysis behind the Ulam's method,[26,27,42,50–52] for descriptions of regularity and estimation of the action of a Frobenius-Perron operator, which has an extensive literature of its own beyond the scope of this paper, with many remaining open problems especially for multivariate transformations.[53] For simplified interpretation and description, we may presume a fine grid of cells covering the domain and the functions described here are piecewise constant in those cells and the transformation is Markov. Beyond the rigorous analysis, these interpretation allow us to compute a conditional entropy of evolution both with and without full consideration of externals to the partitioned subsystem effects.

To explicitly interpret a transfer entropy described seamlessly together with the evolution of densities derived by the Frobenius-Perron transfer operator, we may be interested to understand the *propensity* of the mapping F to move densities and then in this context we may therefore assume that a specific simple form, ρ, is uniform. This is not a necessary but a simplifying assumption, since otherwise we would need to include ρ in the subsequent discussion. Therefore in this context, recombining Eqs. (27) and (29) suggests an interpretation,

$$P[X_{n+1} = x', Y_{n+1} = y|X_n = F^{-1}(x' - y)]$$
$$= P[X_{n+1} = x'|X_n = F^{-1}(x' - y), Y_n = y]P(Y_{n+1} = y)$$
$$= \frac{g(y)}{F'[F^{-1}(x' - y)]}. \quad (30)$$

Now we may work directly with this quantity in the subsequent, but instead we use this form simply for interpretation. Instead, we find it more convenient in the subsequent, to work directly with the original kernel, despite that it may be different in scale, and we will explicitly normalize. Also noting by Eq. (15) that $x'(x, y)$ is a function of the initial position x and the realization of the noise y, let

$$q(x, x', y) = \frac{g[x' - F(x)]}{\int g[x' - F(x)]dx}. \quad (31)$$

This is just the integral kernel and we have explicitly normalized as a probability distribution, for each x', to be used in Eq. (33).

While this is not the same as the original question leading to transfer entropy, Eq. (6), $P(x_{n+1}|x_n) \overset{?}{=} P(x_{n+1}|x_n, y_n)$, we find comparing the kernel's corresponding to a system that is closed unto itself, versus that of a system that is receiving the information at each step by the action of the associated transfer operator, to be extremely informative. Now as we see, this amounts to a slightly different but perhaps related question,

$$P(x_{n+1}|x_n) \overset{?}{=} P(x_{n+1}, y_{n+1}|x_n). \quad (32)$$

Here too, for the sake of simplifying computation, we use the related term as described above, $q(x, x', y)$, interpreted as a variables changed version of Eq. (30). These two alternative stories, closed, or open, of what may moderate the x-subsystem of the dynamical system,

$$q(x, x', y) \overset{?}{=} \delta[x' - F(x)], \quad (33)$$

distinguish the cases whether the x-subsystem is closed, or if it is open—receiving information from the y-subsystem.

Therefore in the subsequent we will describe how to compare these, within the language of information theory. See contrasting versions of Eq. (33) in Fig. 2, described in details as the example in Sec. VII.

VI. FORECASTABILITY QUALITY METRIC

To decide the forecasting question, by comparing alternative versions of the underlying transfer operator kernels for closure of the system, Eq. (33), the seemingly obvious way by a Kullback-Leibler divergence $D_{KL}[q(x, x', y)||\delta(x' - F(x))]$ is generally not well defined. The reason is in part because it is a theorem that[37] the KL-divergence is that not well defined when the support of the second argument is properly contained within the support of the first argument, which will generally be a problem when stating a δ-function as the second argument. Notice that this critical detail arises in our use of the conditionals directly by the kernels of the associated transfer operators, but the arguments do not lead here in other formulations of information flow such as either to transfer entropy or Liang-Kleeman formalism. So in the spirit of transfer entropy, considering $D_{KL}[q(x, x', y)||\delta(x' - F(x))]$ may seem relevant but it is not fruitful.

Instead, the Jensen-Shannon divergence gives an alternative that allows several natural associated interpretations. Let us define the Forecastability Quality Metric, from the y-subsystem to the x-subsystem,

$$FQM_{y \to x} = D_{JS}[(x, x', y)||\delta(x - T_x(s))]$$
$$= \lim_{\epsilon \to 0} D_{JS}[q(x, x', y)||\delta_\epsilon(x' - T_x(x))], \quad (34)$$

using the notation of Eqs. (1) and (3)–(5), and replacing the general F with the component function T_x. The influence of y is encoded in the distribution g that has been normalized to the form q from Eq. (31). More will be said on this below. The Jensen-Shannon divergence is defined as usual[54–56]

$$D_{JS}(p_1||p_2) = \frac{D_{KL}(p_1||m) + D_{KL}(p_2||m)}{2}, \quad (35)$$

where

$$m = \frac{p_1 + p_2}{2}, \quad (36)$$

the mean distribution. An important result is that the necessity of support containment is no longer an issue.

The statement of the limit of terms, δ_ϵ, may be taken as any one of the many variants of smooth functions that progressively (weakly) approximate the action of the delta function, such as

$$\delta_\epsilon(s) = \frac{e^{-\frac{s^2}{4\epsilon}}}{2\sqrt{\pi\epsilon}}, \quad (37)$$

but normalized as in Eq. (31) for each s related to $x' - T_x(x)$.

The Jensen-Shannon divergence has several useful properties and interpretations that are inherited therefore by the FQM. We summarize some of these here. $\sqrt{D_{JS}(p_1||p_2)}$ is a metric, stated in the usual sense. Recall that a function $d : M \times M \to \mathbb{R}^+$ is a metric if

1. Non-negative, $[d(x, y) \geq 0, \forall x, y \in M]$,
2. Identity and discernible, $[d(x, y) = 0$ if $x = y]$,
3. Symmetric, $d(x, y) = d(y, x), \forall x, y \in M$, and
4. Triangle inequality, $d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in M$.

The terminology *metric* is reserved for those functions d which satisfy 1-4, and *distance* while sometimes used interchangeably with metric is sometimes used to denote a function that satisfies perhaps just properties 1-3. The term *divergence* is used to denote a function that may only satisfy property one, but it is only "distance-like." So the Kullback-Leibler divergence D_{KL} is clearly not a distance, and only a divergence because it is not symmetric.

The Jensen-Shannon divergence is not only a divergence but "essentially" a metric. More specifically its square root, $\sqrt{D_{JS}(p_1||p_2)}$, is a metric on a space of distributions, as proved in Refs. 57 and 58. However, nonetheless through Pinsker's inequality there are metric-like interpretations of the Kullback-Leibler divergence, which bounds from above, $\sqrt{\frac{D_{KL}(p_1||p_2)}{2}} \geq ||p_1 - p_2||_{TV}$, by the total variation distance, and for a finite probability space this even relates to the L^1 norm.[59,60] However, a most exciting insight into the meaning of $1/D_{JS}$ follows the interpretation that relates the number of samples one would have to draw from two probability distributions with confidence that they were selected from p_1 or p_2 is inversely proportional to the Jensen-Shannon divergence.[61] Thus the Jensen-Shannon divergence is well known as a multi-purpose measure of dissimilarity between probability distributions, and we find it to be particularly useful to build our information flow concept of "forecasting" as defined, $FQM_{y \to x}$ by Eq. (34) following comparing the operator kernels of Eq. (33) as interpreted as conditional probabilities. $FQM_{x \to y}$ is likewise defined. Finally, we remark that the property,

$$0 \leq FQM_{y \to x} \leq 1, \quad (38)$$

is inherited from the similar bound for the underlying Jensen-Shannon divergence. Therefore, the $FQM_{y \to x}$ makes a particularly useful *score* for information flow.

VII. EXAMPLE—ONE WAY COUPLING AND THE FQM

Now we specialize the general two oscillator problem Eq. (5) to specify just one way coupling as an explicit computation of $FQM_{y \to x}$. Let $\epsilon_2 = 0$,

$$x_{n+1} = T_x(x_n, y_n) = f_1(x_n) + \epsilon_1 k(x_n, y_n),$$
$$y_{n+1} = T_y(x_n, y_n) = f_2(y_n). \quad (39)$$

For simplicity of presentation assume diffusive coupling,

$$k(x, y) = (y - x), \quad (40)$$

so that

$$x_{n+1} = f_1(x_n) + \epsilon_1(y_n - x_n) = \tilde{f}_1(x_n) + \epsilon_1 y_n, \quad (41)$$

and that

$$\tilde{f}_1(x) = f_1(x) - \epsilon_1 x. \quad (42)$$

Thus we have a special case of a coupled map lattice.[62,63]

Further for developing an explicit example,

$$f_1(s) = f_2(s) = 4s(1-s), \qquad (43)$$

the logistic map. We take $f_i : \mathbb{R} \to \mathbb{R}$, but in the uncoupled cases we know that $[0, 1]$ is an invariant set for each component. Since the y-subsystem is uncoupled, and we know its absolutely continuous invariant density in $[0, 1]$ is[7,23]

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}. \qquad (44)$$

We may take this as the distribution of $y_n \in \Omega_y = [0, 1]$ if the y-subsystem is taken to be at steady state. However, we emphasize a *steady state distribution need not be assumed* if we assume simply that a distribution of initial conditions may be chosen from the outside forcing y-subsystem. Since considering the form of the stochastic Frobenius-Perron operator, Eq. (21), the outside influence onto the x-subsystem looks like the noise coupling term $\epsilon_1 y_n$ in Eq. (41). Notice that the distribution of "noise" g is in fact

$$g(s) = \frac{\rho\left(\frac{s}{\epsilon_1}\right)}{\epsilon_1}, \qquad (45)$$

which may seem as noise to the x-subsystem not knowing the details of a y-subsystem, even if the evolution of the full system may even be deterministic. In fact, *this may be taken as a story explaining noise generally as the (unknown) accumulated outside influences on a given subsystem.* So therefore the appearance of "noise" of y-subsystem influence onto x is simply the lack of knowledge of the outside influence onto the not-closed subsystem x. It is a common scenario in chaotic dynamical system that lack of knowledge of states has entropy, and this is the foundation concept of ergodic theory to treat even a deterministic system as a stochastic dynamical system in this sense, as we expanded upon in Ref. 7.

We see in Fig. 2 the contrasting versions of Eq. (7), $P(x_{n+1}|x_n) \rightleftarrows P(x_{n+1}, y_n|x_n)$ associated with contrasting $q[x - F(s)]$ to $q_\epsilon[x - F(s)]$ corresponding to alternative truths, that the x-subsystem is closed, or open depending on y now considered as a stochastic influence. The point is within the transfer operator formalism, the outside influence may be as if stochastic, but nonetheless, the q is a well defined function, and the question of $FQM_{y \to x}$ is well defined by contrasting the two kernels of the associated transfer operators as if pdf's by the D_{JS} in Eq. (34).

In Fig. 3, we show a sequence of estimators illustrating $FQM_{y \to x}$ for Eq. (34). The system shown is relative to the one-way coupled logistic map systems, Eqs. (1)–(42). Note that nothing in the current computation requires a steady state hypothesis since considering an ensemble of y values then the resulting integration is well defined by whatever may be the transient distribution. However, as $\epsilon \to 0$ in the definition, then even though the $FQM_{y \to x}$ is described by a limit of closed form integrals, they become exceedingly stiff to capture reliable values for both ϵ and ϵ_1 small. In another note, notice that since our discussion in no way requires steady state, the two way coupled problem is just as straightforward as the one way coupled problem, which we highlighted purely for simplicity and pedagogy reasons. Finally, we restate that since $1/D_{JS}$ is descriptive of the number of samples required to distinguish
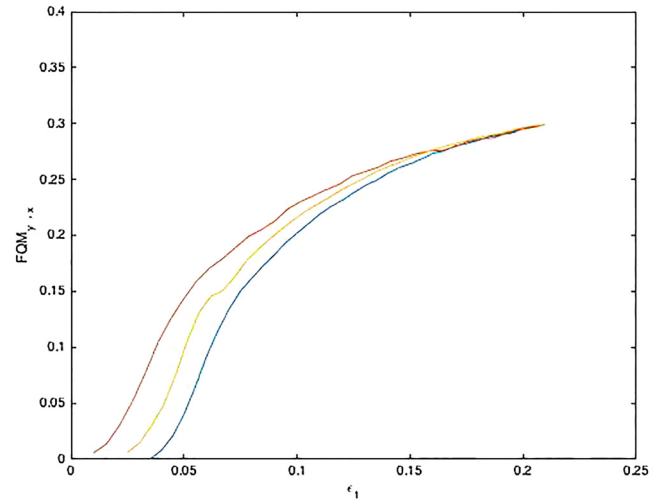

FIG. 3. Computed $FQM_{y \to x}$ for coupled logistic maps, in units of bits per time unit of iteration, with varying ϵ_1 increasing as shown on the horizontal axis. According to the definition of $FQM_{y \to x}$, Eq. (34), Jensen-Shannon divergences are computed for successive approximating values of ϵ decreasing, $\epsilon = 0.035, 0.025, 0.01$ shown from bottom to top listed in the same order. By definition as a Jensen-Shannon divergence, note that $0 \leq FQM_{y \to x} \leq 1$, and 0 is achieved if the distributions in Fig. 2 match, which is closely true when $\epsilon = \epsilon_1$ in terms of the coupling. However, as $\epsilon \to 0$ for a fixed positive but exceedingly small coupling $0 < \epsilon_1 \ll 1$ then the limit is numerically difficult to estimate since the integration becomes singular, and we perform these estimators of the integrals by the Monte-Carlo method; the estimation becomes much more reliable for larger coupling $\epsilon_1 > 0$ where the direct numerical integration becomes more stable.

the underlying two distributions, this sheds lights as interpretation onto the $FQM_{y \to x}$ curves in Fig. 3 which therefore may be interpreted that as coupling ϵ_1 decreases, the decreasing entropy indicates that significantly more observations, either more time, or more states from many initial conditions, are correspondingly required to decide if there is a second coupling system (open), or the system observed is autonomous (closed).

As a final remark, note that this discussion has been entirely for two oscillators, just as the original presentation of transfer entropy was for two oscillators. However by appropriately conditioning out intermediaries, to distinguish direct versus indirect effects, we generalized transfer entropy to become causation entropy,[8–10] and a comparable strategy might allow conditional FQM, by marginalizing and conditioning restricted versions of the transfer operators before measuring the differences using the Shannon-Jenson convergence. This will also be a consideration in our future works.

VIII. POSTSCRIPT AND CONCLUSIONS

We have described how noise and coupling of an outside influence onto a subsystem from another subsystem can be formally described as alternative views of the same phenomenon. Using these alternative descriptions of this concept, by using the kernels from deterministic versus stochastic Frobenius-Perron transfer operators to contrast the outside influence of a coupling system as if it were noise, we can explicitly enumerate the degree of information transferred

from one subsystem to another. This is the first time this formalism has been brought to consider information transfer. We show furthermore that motivated by transfer entropy, using the KL-divergence for the transfer operator concept based in this context produces problems regarding boundedness. The Jensen-Shannon divergence provides a useful alternative that furthermore comes with several pleasant extra interpretations.

Outside influences may be summarized by the following diagram and asking if it is possibly commuting, pointwise,

$$
\begin{array}{ccc}
\rho \in L^1(\Omega) & \xrightarrow{P_T} & \rho' \in L^1(\Omega) \\
R_y \downarrow & & \downarrow R_y \\
\rho_1 \in L^1(\Omega_y) & \xrightarrow{P_{T_y}} & \rho_1' \in L^1(\Omega_y)
\end{array}
\qquad (46)
$$

where we reiterate that $\Omega = \Omega_x \times \Omega_y$ states the proposed subsystems, and

$$
\begin{aligned}
r_y : \Omega &\to \Omega_y, \\
(x, y) &\mapsto y,
\end{aligned}
\qquad (47)
$$

denotes a projection function, from the full phase space Ω to the phase space of the y-subsystem, and likewise for the projection r_x. In this formulation, the main question of closure, if there is information flow or not, which we have already stated in Eq. (33) as $q[x - F(y)] \stackrel{?}{=} \delta[x - F(y)]$, also amounts to asking if advancing the density of states of the full system and then projecting by the operator corresponding to marginalizing [integrate density onto just y variables, $R_y[\rho(x, y)] = \int_\Omega \rho(x, y\,dx)]$ is the same as marginalizing first and then advancing by the transfer operator of the subsystem:

$$
R_y \circ P_T \stackrel{?}{=} P_{T_y} \circ R_y.
\qquad (48)
$$

In postscript, we already noted that the inverse of the Jensen-Shannon divergence is proportional to the expected number of samples required to distinguish the two distributions. Therefore, the $FQM_{y \to x}$ is inversely proportional to the number of samples required to distinguish the degree of coupling influence of the y-variables onto the x-variables subsystems. In this sense, in our follow-up work, we are planning a practical numerical scheme to associate data observations. Specifically, the Ulam's method allows for a cell-mapping method to cover the phase space with boxes (or say triangles), and then to collect statistics of transitions, and besides the usual discussion toward invariant density through the eigenvectors of the resulting stochastic matrix, known as Ulam's method, we have already pointed out[22] that there is information in this numerical estimate of the transfer operator that can be exploited to compute transfer operator. However, we realize that the operator itself bears a great deal of information regarding information flow, and so this points to the idea that FQM might be estimated from data, by using the data to build a stochastic matrix in the spirit of Ulam's method. Such a Markov chain model of the process can help distinguish open or closed, but building the transition matrix directly from data, and then applying the FQM, a D_{JS} computation in alternative formulations of the hypothesis. Therefore, we are working toward this for future research, and considering error analysis of the collected statistics has Markov-inequalities (including Chebyshev inequality) underlying. Therefore while this more practical data oriented approach is still in the works, what we have offered in this paper is a new view on information flow, which can be understood directly in terms of the underlying transfer operators, and computations of entropies directly from there.

[1]D. F. Hendry, "The nobel memorial prize for clive W.J. Granger," Scand. J. Econ. **106**, 187–213 (2004).

[2]C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," Econometrica **37**, 424–438 (1969).

[3]C. W. Granger, "Some recent development in a concept of causality," J. Econom. **39**, 199–211 (1988).

[4]L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," Phys. Rev. Lett. **103**, 238701 (2009).

[5]N. Wiener, *Modern Mathematics for Engineers* (McGraw-Hill, New York, 1956).

[6]T. Schreiber, "Measuring information transfer," Phys. Rev. Lett. **85**, 461 (2000).

[7]E. M. Bollt and N. Santitissadeekorn, *Applied and Computational Measurable Dynamics* (SIAM, 2013).

[8]J. Sun and E. M. Bollt, "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings," Physica D **267**, 49–57 (2014).

[9]J. Sun, D. Taylor, and E. M. Bollt, "Causal network inference by optimal causation entropy," SIAM J. Appl. Dyn. Syst. **14**, 73–106 (2015).

[10]J. Sun, C. Cafaro, and E. M. Bollt, "Identifying the coupling structure in complex systems through the optimal causation entropy principle," Entropy **16**, 3416–3433 (2014).

[11]C. Cafaro, W. M. Lord, J. Sun, and E. M. Bollt, "Causation entropy from symbolic representations of dynamical systems," Chaos **25**, 043106 (2015).

[12]G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," Science **338**, 1227079 (2012).

[13]B. Russell, in *Proceedings of the Aristotelian Society* (JSTOR, 1912), Vol. **13**, pp. 1–26.

[14]J. Pearl, *Causality* (Cambridge University Press, 2009).

[15]D. A. Smirnov, "Spurious causalities with transfer entropy," Phys. Rev. E **87**, 042917 (2013).

[16]D. Kugiumtzis, "Partial transfer entropy on rank vectors," Eur. Phys. J. **222**, 401–420 (2013).

[17]M. Prokopenko, J. T. Lizier, and D. C. Price, "On thermodynamic interpretation of transfer entropy," Entropy **15**, 524–543 (2013).

[18]M. Staniek and K. Lehnertz, "Symbolic transfer entropy," Phys. Rev. Lett. **100**, 158101 (2008).

[19]L. Barnett and T. Bossomaier, "Transfer entropy as a log-likelihood ratio," Phys. Rev. Lett. **109**, 138105 (2012).

[20]J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, "Escaping the curse of dimensionality in estimating multivariate transfer entropy," Phys. Rev. Lett. **108**, 258701 (2012).

[21]X. S. Liang and R. Kleeman, "Information transfer between dynamical system components," Phys. Rev. Lett. **95**, 244101–244101 (2005).

[22]E. M. Bollt, "Synchronization as a process of sharing and transferring information," Int. J. Bifurcat. Chaos **22**, 1250261 (2012).

[23]A. Lasota and M. C. Mackey, *Probabilistic Properties of Deterministic Systems* (Cambridge University Press, 1985).

[24]S. Ulam, *Problems of Modern Mathematics* (New York, 1964), science editions, originally published as: *A Collection of Mathematical Problems* (1960).

[25]T.-Y. Li, "Finite approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture," J. Approx. Theory **17**, 177–186 (1976).

[26]G. Froyland, "Ulam's method for random interval maps," Nonlinearity **12**, 1029 (1999).

[27]F. Y. Hunt, "Unique ergodicity and the approximation of attractors and their invariant measures using Ulam's method," Nonlinearity **11**, 307 (1998).

[28]R. Murray, "Ulam's method for some non-uniformly expanding maps," Discrete Cont. Dyn. Syst. **26**, 1007–1018 (2010).

[29]J. Matousek, *Using the Borsuk-Ulamtheorem: Lectures on Topological Methods in Combinatorics and Geometry* (Springer Science & Business Media, 2008).

[30]L. Billings and E. Bollt, "Probability density functions of some skew tent maps," Chaos Solitons Fractals **12**, 365–376 (2001).

[31]E. M. Bollt, A. Luttman, S. Kramer, and R. Basnayake, "Measurable dynamics analysis of transport in the gulf of Mexico during the oil spill," Int. J. Bifurcat. Chaos **22**, 1230012 (2012).

[32]E. M. Bollt, "Controlling chaos and the inverse Frobenius-Perron problem: global stabilization of arbitrary invariant measures," Int. J. Bifurcat. Chaos **10**, 1033–1050 (2000).

[33]E. M. Bollt, L. Billings, and I. B. Schwartz, "A manifold independent approach to understanding transport in stochastic dynamical systems," Physica D **173**, 153–177 (2002).

[34]J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," in *The Theory of Chaotic Attractors* (Springer, 1985), pp. 273–312.

[35]T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," J. Stat. Phys. **65**, 579–616 (1991).

[36]E. M. Bollt, "Model selection, confidence and scaling in predicting chaotic time-series," Int. J. Bifurcat. Chaos **10**, 1407–1422 (2000).

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

[38]S. M. Ulam, *Problems in Modern Mathematics*, 2nd ed. (John Wiley & Sons, Inc., New York, 1964), pp. xvii+150.

[39]G. Froyland, "Finite approximation of Sinai-Bowen-Ruelle measures for Anosov systems in two dimensions," Random Comput. Dynam. **3**, 251–263 (1995), available at http://web.maths.unsw.edu.au/~froyland/a1362773.pdf

[40]G. Froyland, "Approximating physical invariant measures of mixing dynamical systems in higher dimensions," Nonlinear Anal. **32**, 831–860 (1998).

[41]A. Boyarsky and Y.-S. Lou, "Approximating measures invariant under higher-dimensional chaotic transformations," J. Approx. Theory **65**, 231–244 (1991).

[42]A. Boyarsky and P. Gora, *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension* (Springer Science & Business Media, 2012).

[43]E. M. Bollt and J. D. Skufca, *Encyclopedia of Nonlinear Science* (Routledge, New York, 2005).

[44]E. Bollt, P. Góra, A. Ostruszka, and K. Życzkowski, "Basis Markov partitions and transition matrices for stochastic systems," SIAM J. Appl. Dyn. Syst. **7**, 341–360 (2008).

[45]P. Cvitanović, "Periodic orbits as the skeleton of classical and quantum chaos," Physica D **51**, 138–151 (1991).

[46]M. Dellnitz, A. Hohmann, O. Junge, and M. Rumpf, "Exploring invariant sets and invariant measures," Chaos **7**, 221–228 (1997).

[47]F. Y. Hunt and W. M. Miller, "On the approximation of invariant measures," J. Stat. Phys. **66**, 535–548 (1992).

[48]N. Santitissadeekorn and E. Bollt, "The infinitesimal operator for the semigroup of the Frobenius-Perron operator from image sequence data: Vector fields and transport barriers from movies," Chaos **17**, 023126 (2007).

[49]L. Arnold, *Random Dynamical Systems* (Springer Science & Business Media, 2013).

[50]C. Bose and R. Murray, The exact rate of approximation in Ulam's method, Discrete Cont. Dyn. Syst. A **7**, 219–235 (2001).

[51]J. Ding, T. Y. Li, and A. Zhou, "Finite approximations of Markov operators," J. Comput. Appl. Math. **147**, 137–152 (2002).

[52]J. Ding and A. Zhou, "A finite element method for the Frobenius-Perron operator equation," Appl. Math. Comput. **102**, 155–164 (1999).

[53]J. Ding and A. Zhou, "Finite approximations of Frobenius-Perron operators. A solution of Ulam's conjecture to multi-dimensional transformations," Physica D **92**, 61–68 (1996).

[54]M. Menéndez, J. Pardo, L. Pardo, and M. Pardo, "The Jensen-Shannon divergence," J. Franklin Inst. **334**, 307–318 (1997).

[55]J. Lin, "Divergence measures based on the Shannon entropy," IEEE Trans. Infor. Theory **37**, 145–151 (1991).

[56]B. Fuglede and F. Topsoe, "Jensen-Shannon divergence and Hilbert space embedding," in *International Symposium on Information Theory, 2004, ISIT 2004* (IEEE, 2004), p. 31.

[57]D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," IEEE Trans. Infor. Theory **49**, 1858–1860 (2003).

[58]I. Vajda *et al.*, "Anew class of metric divergences on probability spaces and and its statistical applications," Ann. Inst. Statist. Math. **55**, 639–653 (2003).

[59]M. S. Pinsker, *Information and Information Stability of Random Variables and Processes* (1960).

[60]E. Ordentlich and M. J. Weinberger, "A distribution dependent refinement of Pinsker's inequality," IEEE Trans. Inf. Theory **51**, 1836–1840 (2005).

[61]G. Tkacik, C. G. Callan, Jr., and W. Bialek, "Information capacity of genetic regulatory elements," Phys. Rev. E **78**, 011910 (2008).

[62]K. Kaneko, "Coupled map lattice," in *Chaos, Order, and Patterns* (Springer, 1991), pp. 237–247.

[63]S. D. Pethel, N. J. Corron, and E. Bollt, "Symbolic dynamics of coupled map lattices," Phys. Rev. Lett. **96**, 034105 (2006).