

## Article

# A General Metric for the Similarity of Both Stochastic and Deterministic System Dynamics

Colin Shea-Blymyer <sup>1</sup>, Subhadeep Roy <sup>2</sup> and Benjamin Jantzen <sup>3,\*</sup>

<sup>1</sup> School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331-5501, USA; sheablyc@oregonstate.edu

<sup>2</sup> Department of Mechanical Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA; ROYS5@erau.edu

<sup>3</sup> Department of Philosophy, Virginia Tech, Blacksburg, VA 24060, USA

\* Correspondence: bjantzen@vt.edu

**Abstract:** Many problems in the study of dynamical systems—including identification of effective order, detection of nonlinearity or chaos, and change detection—can be reframed in terms of assessing the similarity between dynamical systems or between a given dynamical system and a reference. We introduce a general metric of dynamical similarity that is well posed for both stochastic and deterministic systems and is informative of the aforementioned dynamical features even when only partial information about the system is available. We describe methods for estimating this metric in a range of scenarios that differ in respect to control over the systems under study, the deterministic or stochastic nature of the underlying dynamics, and whether or not a fully informative set of variables is available. Through numerical simulation, we demonstrate the sensitivity of the proposed metric to a range of dynamical properties, its utility in mapping the dynamical properties of parameter space for a given model, and its power for detecting structural changes through time series data.

**Keywords:** nonlinearity; model selection; chaos detection; change detection; model behavior mapping; dynamical similarity; causal discovery



**Citation:** Shea-Blymyer, C.; Roy, S.; Jantzen, B. A General Metric for the Similarity of Both Stochastic and Deterministic System Dynamics. *Entropy* **2021**, *23*, 1191. <https://doi.org/10.3390/e23091191>

Academic Editor: Ernestina Menasalvas

Received: 1 August 2021  
Accepted: 7 September 2021  
Published: 9 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The term dynamical similarity refers to the degree to which the dynamics governing the evolution of a system over a period of time resembles the dynamics of another system, or the dynamics of the same system over a different period of time. Two fundamental problems of time series analysis are how best to measure dynamical similarity, and how to infer a given measure from noisy time series data. We address both problems, first by presenting a new metric of dynamical similarity that is indicative of underlying causal structure and that is well posed whether one is comparing two deterministic or two stochastic dynamical systems, and then by providing a range of tools for estimating this metric from data.

There is no uniquely best measure of dynamical similarity since the aptness of any given measure is relative to its intended use. However, there is broad interest in measures that are sensitive to the causal structure of a system in the sense of which variables directly determine the values (or stochastic distributions over the values) of other variables and with what functional form. For instance, the field of change detection is concerned with determining if and when the causal structure of a dynamical system has changed, as for example when an ecosystem (or climate) has been perturbed by external forcing, or when a mechanical component has begun to fail [1]. This amounts to determining whether one and the same system at a later time is dynamically similar to itself at an earlier time. Classically, this problem has been pursued under the assumption that the behavior of some stochastic system before and after a rapid change in parameter values is stationary [2], though progress has been made on the problem without the assumption

of stationarity [3]. From the perspective of deterministic complex systems, the problem of change detection manifests as either the problem of detecting incipient bifurcations [4] or transitions between regions of the system's attractor [5]. The more general problem encompassing both approaches is the detection of change in the structure of a dynamical system without assumptions of determinism or stationarity, such as detecting incipient bifurcation in an arbitrary stochastic dynamical system [6]. In general, we want to know whether a given system will continue to evolve and respond to perturbations in the same way as it did before.

It has been recognized for some time that it is often more important to know the degree of difference in dynamics after a change rather than the mere fact of change [7]. While change detection is often treated as a binary statistical decision problem [8], explicit measures of dynamical similarity have been used to assess whether a dynamical shift is practically (as opposed to statistically) significant [7]. Implicitly, such a notion of dynamical similarity plays a central role in system identification, where it is often important to know at the outset whether the system in question can be described with sufficient fidelity using a linear model [9], or in other words, whether the system is sufficiently dynamically similar to a linear one. Similarly, it is important to know the effective order of the dynamical process of interest in contexts where acquiring data points in a time series is expensive, such as community ecology, in order to know how long of a time series will be needed to fit a reliable model [10]. Relatedly, the degree of nonlinearity and chaos exhibited by a system is essential for managing error in system control or prediction [11–14]. Each of these problems—identifying effective order, nonlinearity, and chaos—can be seen in terms of dynamical similarity. Whether a system is strongly nonlinear is equivalent to asking whether and to what degree it is dynamically similar to a strongly nonlinear system and *mutatis mutandis* for effective order and chaos.

Whatever the appropriate measure of dynamical similarity, the problem of estimating its value from data is made more difficult with increasing sampling noise, and becomes significantly more challenging to address for stochastic systems. Model validation is notoriously difficult for stochastic systems [15,16]. Model validation amounts to assessing the similarity between one system—the model—whose dynamics is exactly characterized, and a target system with unknown dynamical properties; a valid model is one whose dynamics matches that of the target. The difficulty of this comparison is only compounded when one attempts to determine the similarity between two systems, both with unknown dynamics. Furthermore, existing techniques for detecting change or assessing, e.g., nonlinearity, are highly sensitive to sampling noise [2,17].

We here describe a general metric of dynamical similarity that is well posed for both stochastic and deterministic dynamical systems and which is sensitive to the effective order of dynamics, the degree of nonlinearity, and the presence of chaos. Importantly, this metric can be informative of these dynamical features even when only partial information about the dynamical state of a system is tracked, or a lossy function of the dynamical variables is observed, or in other words, if the system is only “partially observed” [18]. We introduce a variety of algorithms to show that this metric can be learned in a range of contexts, from situations in which one has full control of the dynamical system and complete dynamical information to situations in which only partial information is available for passively observed systems. We also demonstrate how this metric can be applied to the problem of change detection in this range of circumstances, and how it can be deployed to rapidly map out the varieties of dynamical behavior as a function of parameter values for a given dynamical model.

## 2. Related Work

### 2.1. Assessing Nonlinearity, Chaos, and Effective Order

The metric of dynamical similarity we describe is sensitive to the degree of nonlinearity in that the greater the magnitude of nonlinear terms in the governing differential equations of a system, the more it differs according to our metric from otherwise similar linear systems.

Though not typically framed in terms of dynamical similarity, a number of measures of dynamical properties that could be used as such have been introduced previously for the purpose of detecting nonlinearity in time series data. In one prominent approach, introduced by Theiler et al. [19], one or another of these dynamical measures is deployed as a discriminative statistic for binary hypothesis testing where the null hypothesis is a linearized dataset—the surrogate data—constructed with a model that preserves properties such as the mean and variance of the original data. While Theiler et al. [19] used a battery of statistics familiar from dynamical systems theory, including the correlation dimension, Lyapunov exponent, and forecasting error, much recent work draws upon information theoretic constructs. Paluš [20,21], for instance, pursues the method of surrogate data using redundancy (the multidimensional generalization of mutual information). Unfortunately, redundancy and similar information theoretic measures reflect not just the dynamical relations among variables but also the apparent coordination imposed by their shared history, forcing function, or boundary conditions. Other information theoretic entropies have been pursued that better capture the causal relations among variables. The transfer entropy [22,23], for example, considers the probabilities of state transitions rather than of the states themselves. It has been applied as a discriminative statistic in the surrogate data framework by Nichols et al. [24], and was shown to outperform time-delayed mutual information. All of these methods, however, still depend upon the assumption that the system being assessed is at most weakly nonstationary. Merging information theoretic and dynamical systems approaches, Bandt and Pompe [25] introduced the permutation entropy, which they describe explicitly as a measure of the complexity of system dynamics. For a given  $n$  (typically chosen to be on the order of 10), the permutation entropy is an information theoretic entropy based on the probability of each of the  $n!$  permutations of the ordinal ranks of the elements in each  $n$ -sample long subsequence of a time series. It closely tracks the Lyapunov exponent  $\lambda$ , and requires that a time series exhibit only a very modest sort of stationarity. Typically, it is employed in the surrogate framework as a discriminative statistic [26], as is the related quantity known as the “number of missing ordinal patterns” [27]. More to the point, the permutation entropy allows for the sort of comparison between dynamical systems that our metric does, though with one substantial restriction: it vanishes uniformly for monotonic functions, whether nonlinear or not.

There is particular interest in distinguishing chaotic nonlinear systems from non-chaotic systems. Some methods for doing so require comparison with an explicit contrast class of models such as the method of comparing the predictive power of linear and nonlinear models of Volterra-Wiener form [28], and the method of surrogate data using the correlation dimension as the discriminative statistic [29]. Most, however, focus on some endogenous property of a system that can be estimated directly from a time series. Measures of this sort include (but are not exhausted by) Lyapunov exponents [30,31], the correlation dimension itself [32], nonlinear forecasting [33] (which provides an estimate of the largest Lyapunov exponent and has been found more robust than the correlation coefficient [34]), the determinism test [35,36], Kolmogorov entropy [37], the noise titration test [38], and the 0–1 test [39]. More recently, Kulp and Zunino [40] used the encoding scheme of Bandt and Pompe [25] to construct a “symbolic spectrum” in the manner of Yang and Zhao [41]. By examining the spectrum for a time series and looking for missing ordinal patterns with zero standard deviation (a hallmark of determinism), deterministic dynamics can be distinguished from stochastic, and by noting variation for some patterns, chaos can be distinguished from periodicity. Because hyper-chaotic systems have fewer forbidden patterns with zero variance, this test has the potential to yield an integral degree of chaos. With the exception of the 0-1 test, all of the other endogenous measures indicate a continuous-valued degree of chaos. Such a quantitative degree in turn admits an obvious measure of dynamical similarity, at least with respect to chaos: the closer two systems are in their degree of chaos, the more dynamically similar they are.

The final feature of system dynamics with which we are concerned is the effective order or history dependence. Identifying the effective order of dynamics is an essential component of model selection in statistical forecasting and system identification. How such an identification can be made depends to a great deal on what is already known about the functional form of the connection between some finite number of past values of the system variables and the future values. Nonparametric system order identification techniques that can be applied to the more challenging classes of systems to which our algorithm can be applied, including nonlinear, stochastic systems, are diverse and well documented [42,43]. Broadly speaking, the problem is conceived as one of function estimation: the object is to find the form of a function  $f(\cdot)$  and values of the associated parameters  $\theta$  such that  $y(t) = f(\vec{x}(t), \theta) + \epsilon(t)$ , where  $y(t)$  is the target time series,  $\vec{x}(t)$  is the input vector, and  $\epsilon(t)$  is noise, typically presumed to be independent and identically distributed with constant variance [43]. As with function estimation in general, learning model structure—including model order—from a time series requires balancing the goodness of fit to the observed data and model complexity. There are a diversity of general approaches to achieving this balance, including cross-validation [44], stepwise selection (using measures such as AIC to bound complexity and regularize the search) [45–47], structural risk minimization [48], and LASSO [49]. A distinct approach is provided by methods for determining the embedding dimension in dynamical systems [50,51].

## 2.2. Dynamical Similarity and Change

One straightforward application of a measure of dynamical similarity is the problem of detecting a change in the behavior of a system over time. There are at least three specific versions of this general problem recognized in the literature. The first concerns detection of points or portions of a time series that are outliers with respect to an unknown but stationary (or slowly changing) distribution. This is a species of the generic anomaly or outlier detection problem [52,53]. The second problem is the detection of state change. From a statistical perspective, this means identifying points in a time series at which the parameters of a stationary or linearly changing distribution change suddenly [2,54]. From a dynamical perspective, this means detecting changes from one stable state to another as system parameters vary, or finding points at which an evolving system moves suddenly into a different region of its attractor [5].

The third problem of behavior change—and that which concerns us here—is the identification of a change in time of the dynamics of a system. Such a change may be due to a change in parameter values, a change in the functional form of relationships among variables, or even a change in the causal structure among the variables, and may occur fast or slow relative to the timescale of observation. Before, during, and after the change, the system need not exhibit a stationary distribution. In other words, the problem is to determine when a system at one time is dynamically different or distant from itself at an earlier time. A variety of approaches have been proposed for detecting changes of this sort. The most straightforward involve time series similarity measures. An overview and quantitative comparison of the similarity measures has been provided in [55] and more recently in [56]. These include distance measures such as the family of Minkowski distances (which include Manhattan and Euclidean distances) and the Mahalanobis distance [57] that treat each time series as a point in a high-dimensional space, as well as similarity indices constructed from correlation measures like Pearson's cross-correlation coefficient [58]. While these approaches work directly with time series (though often normalized), related approaches involve an initial transformation of the series, e.g., by replacing the data matrix with the first few principal components (linear functions of the original variables) as a function of time [59], or shifting to the frequency domain by FFT [60,61]. The point of these transformations is generally some combination of dimension reduction and the increase in sensitivity to significant features. More recently, efficient algorithms have been introduced for computing the matrix profile, which replaces a time series with a series of values of the minimum Euclidean distance between a moving

window and all similarly-sized sub-sequences [62]. In the resulting profile, anomalous segments of a time series are indicated by high profile values, while repeating motifs appear as minima. All of these metrics can be used as a measure of dynamical difference to detect changes in behavior. Unfortunately, none of these approaches to change detection can distinguish between dynamical changes and mere changes of state.

There are, however, alternative approaches in the literature that are specifically sensitive to dynamical structure. One family of methods attempts to detect a variety of symptoms of an incipient bifurcation, such as an increase in variance, slower recovery from perturbations and a consequent increase in autocorrelation [63]. Methods of detection often focus on trends in fit coefficients of autoregressive models [4,64]. A different sort of approach focuses on properties of the system attractor that are invariant for fixed parameter values, such as the recurrence plot or density function of visitation over cells in a discretized phase space or the fractal dimension [7,65,66]. Hively et al. [7], for example, define two measures on the reconstructed phase space of a system that provide a dynamical distance between a reference case time series and a test case. They first convert each univariate time series,  $x_i$ , into a sequence of integers,  $s_i$  between 0 and  $S - 1$  via the function  $\text{INT}[S(x_i - x_{\min}) / (x_{\max} - x_{\min})]$ . If  $d$  is the embedding dimension of the reconstructed phase space, this effectively divides the phase space into  $S^d$  hypercubes or “bins.” They then compute the empirical distribution functions  $Q_i$  and  $R_i$  reflecting the population of the  $i$ th bin for the reference time series and base time series, respectively. The measures of dynamical distance they introduce are:  $\chi^2 = \sum_i (Q_i - R_i)^2 / (Q_i + R_i)$ , and  $L = \sum_i |Q_i - R_i|$ . When applied to a simulated Bondarenko neuron model [67], both of these measures increase monotonically as a key parameter is varied through a region of known chaotic behavior. In other words, the degree of dynamical dissimilarity was shown to track known structural dynamical changes. These measures (and related “connected” variants) were also applied to EEG data as a tool for detecting incipient seizures.

### 2.3. Dynamical Kinds

Effective order, nonlinearity, and chaos are all aspects of the causal structure of a dynamical system. The existing tests contrived to assess these aspects consider a target system in isolation; the details of a particular system’s behavior are used to determine, e.g., the effective order of its dynamics, and only after the fact is it recognized that two systems share such causal features in common. The theory of dynamical kinds offers an alternative approach: by determining that two systems are of the same or different dynamical kinds, we learn whether they share any of these dynamical properties. Information about any one can then be transferred to the class.

Dynamical kinds were first defined in [68], and have since been applied explicitly to problems of model validation [69] and causal discovery [70]. The dynamical kinds theory partitions the space of dynamical systems into equivalence classes—dynamical kinds—on the basis of dynamical symmetry. A dynamical symmetry is an intervention [71,72]—an externally induced change in the values of some of the dynamical variables of a system—that commutes with the time evolution of the system ([69], p. 162):

**Definition 1.** Let  $t$  be the variable representing time, let  $V$  be a set of dynamical variables, and let  $\Omega$  be the space of states that can be jointly realized by the variables in  $V$ . Let  $\sigma : \Omega \rightarrow \Omega$  be an intervention on the variables in  $\text{Int} \subseteq V$ , and  $\Lambda_{t_0, t_1}$  the time-evolution operator that advances the state of the system from  $t_0$  to  $t_1$ . The transformation  $\sigma$  is a dynamical symmetry with respect to time if and only if for all intervals  $\Delta t$  and initial states  $\omega_i \in \Omega$ , the final state of the system  $\tilde{\omega}_f \in \Omega$  is the same whether  $\sigma$  is applied at some time  $t_0$  and the system evolved until  $t_0 + \Delta t$ , or the system first allowed to evolve from  $t_0$  to  $t_0 + \Delta t$  and then  $\sigma$  is applied. This property is represented by the following commutative diagram:

$$\begin{array}{ccc}
 \tilde{\omega}_i & \xrightarrow{\Lambda_{t_0, t_0+\Delta}} & \tilde{\omega}_f \\
 \sigma \uparrow & & \sigma \uparrow \\
 \omega_i & \xrightarrow{\Lambda_{t_0, t_0+\Delta}} & \omega_f
 \end{array} \tag{1}$$

Biological growth offers a simple illustration of the concept. For an exponentially growing population of bacteria, for which the population  $x$  changes according to  $dx/dt = rx$ , any transformation that scales the population by a positive constant  $k$  is a dynamical symmetry of the system—scaling by  $k$  and then allowing the bacteria to grow for  $\Delta t$  versus growing for  $\Delta t$  and then scaling the resulting population by  $k$  results in the same final population.

The composition of any two dynamical symmetries (by successive intervention on the variables of a system) is itself a dynamical symmetry, and for any given system, its dynamical symmetries typically exhibit nontrivial algebraic structure under composition. It is the collection of symmetries and their structure under composition that characterizes a dynamical kind [68].

Jantzen [73] provides a method for determining whether or not two physical systems with continuous deterministic dynamics belong to the same dynamical kind directly from time series data without first constructing dynamical models of either system. The method exploits two facts: (i) that a necessary condition for two systems to belong to the same dynamical kind is that they share all of their dynamical symmetries, and (ii) that for every state-determined system in the same dynamical kind, there is exactly one symmetry that maps the unique trajectory passing through one point in phase space,  $\vec{x}$ , to the trajectory passing through another point  $\vec{x}'$ . That is, even if systems A and B exhibit different trajectories given initial conditions  $\vec{x}$  or  $\vec{x}'$ , the symmetry connecting these two trajectories for system A must be the same as that for system B if they belong to the same dynamical kind. Numerically estimating and then comparing these symmetries from time series data thus provides a sensitive test for the sameness of dynamical kind that is robust under significant sampling noise.

### 3. Theory and Algorithms

#### 3.1. Stochastic Dynamical Kinds

Dynamical kinds partition the space of possible dynamical systems based on their causal structure. However, as defined above, sameness of dynamical kind is binary; it does not indicate the degree to which two systems that are not of the same kind differ in their causal structure. It also fails to apply to systems with stochastic dynamics. This latter problem can be addressed with a more expansive definition of dynamical kind. Previously, it has been suggested that the definition of dynamical similarity should be generalized to accommodate stochastic dynamics [69]. When restricted to dynamical symmetries with respect to time (other types of symmetry are considered in [69]), that proposed definition reduces to the following:

**Definition 2.** Let  $V$  be a set of random variables,  $\Omega$  the set of states that can be jointly realized by the variables in  $V$ , and  $\Gamma$  the space of probability distributions over  $\Omega$ . Let  $\sigma : \Gamma \rightarrow \Gamma$  be an intervention on the variables in  $Int \subset V$ . The transformation  $\sigma$  is a dynamical symmetry with respect to time if and only if  $\sigma$  has the following property: for all initial joint distributions  $\gamma_i \in \Gamma$ , the final joint probability distribution over  $V$ ,  $\gamma_f \in \Gamma$ , is the same whether  $\sigma$  is applied at time  $t_0$  and then time evolution  $\Lambda_{t_0, t_0+\Delta} : \Gamma \rightarrow \Gamma$  evolves the joint distribution, or the system is first allowed to evolve over an interval  $\Delta$ , and then  $\sigma$  is applied. This property is represented in the following commutative diagram:

$$\begin{array}{ccc}
 \tilde{\gamma}_i & \xrightarrow{\Lambda_{t_0,t_0+\Delta}} & \tilde{\gamma}_f \\
 \sigma \uparrow & & \uparrow \sigma \\
 \gamma_i & \xrightarrow{\Lambda_{t_0,t_0+\Delta}} & \gamma_f
 \end{array} \tag{2}$$

We accept this definition, and use it to develop a natural metric over dynamical systems that solves the problem of degree. Specifically, we present a metric based on Definition 2 that provides a well-grounded degree of causal difference, and show that it is sensitive to variations in linearity, effective order, and the presence of chaos.

### 3.2. Constructing a Metric of Dynamical Similarity

Consider a dynamical system described by  $n$  variables, some of which may be derivatives or time-lagged versions of other variables. The states of such a system can be represented by  $n$ -dimensional vectors  $\vec{x} \in \Omega$ . Let  $\Gamma$  be the space of probability density functions over  $\Omega$ . We say that such a system is *stochastically state-determined* (SSD) if and only if the (possibly stochastic) dynamics of the system is such that the probability density  $\gamma_i \in \Gamma$  over possible states at time  $t_i > t_1$  is completely determined by the density  $\gamma_1 \in \Gamma$  over states at  $t_1$ . In other words, for an SSD system there exists a map,  $\Lambda_{t_1,t_i} : \Gamma \rightarrow \Gamma$  that, for any  $t_i > t_1$  advances the probability density over states of the system from  $\gamma_1(\vec{x})$  to  $\gamma_i(\vec{x})$  (“random dynamical systems” in the sense defined in [74] are thus SSD, though we do not insist that SSD systems be measure preserving). According to Definition 2, any function,  $\sigma : \Gamma \rightarrow \Gamma$  that commutes with this map is a dynamical symmetry of the system. More precisely,  $\sigma$  is a dynamical symmetry of a system if, for every  $t_i > t_1$

$$\sigma \circ \Lambda_{t_1,t_i} \circ \gamma_1(\vec{x}) = \Lambda_{t_1,t_i} \circ \sigma \circ \gamma_1(\vec{x}) \tag{3}$$

where  $\sigma \circ \Lambda_{t_1,t_i} \circ \gamma_1$  is the probability distribution that results from successive application of the maps  $\Lambda_{t_1,t_i}$  and then  $\sigma$  to the original distribution  $\gamma_1$  (and similarly for  $\Lambda_{t_1,t_i} \circ \sigma \circ \gamma_1$ ). This property is represented in by the following commutative diagram:

$$\begin{array}{ccc}
 \tilde{\gamma}_1(\vec{x}) = \sigma \circ \gamma_1(\vec{x}) & \xrightarrow{\Lambda_{t_1,t_i}} & \tilde{\gamma}_i(\vec{x}) = \sigma \circ \Lambda_{t_1,t_i} \circ \gamma_1(\vec{x}) = \Lambda_{t_1,t_i} \circ \sigma \circ \gamma_1(\vec{x}) \\
 \uparrow \sigma & & \uparrow \sigma \\
 \gamma_1(\vec{x}) & \xrightarrow{\Lambda_{t_1,t_i}} & \gamma_i(\vec{x}) = \Lambda_{t_1,t_i} \circ \gamma_1(\vec{x})
 \end{array} \tag{4}$$

Consider two probability densities,  $\gamma_1$  and  $\sigma \circ \gamma_1$ , connected by one such dynamical symmetry at time  $t_1$ . We call a time series  $\vec{x}(t_i)$  *untransformed* if, for every  $t_i$ , it is the value of a random variable distributed according to  $\gamma_i(\vec{x})$ . In other words, an untransformed time series is a time series for a system that evolves from an initial value selected according to the distribution  $\gamma_1$ . Similarly, we call a time series  $\tilde{\vec{x}}(t_i)$  *transformed* if, for every  $t_i$ , it is the value of a random variable distributed according to  $\sigma \circ \gamma_i(\vec{x})$ . We focus on a particular probability density function that relates the time evolution of untransformed and transformed time series. Specifically, if one of  $n$  distinct times,  $t_i$ , in the evolution of the system is selected at random from a uniform distribution over the  $n$  possibilities, we seek the probability density that any untransformed time series exhibits a system state  $\vec{x}(t_i)$  and any given transformed time series presents a system state  $\tilde{\vec{x}}(t_i)$  at the same time. This joint density is given by:

$$\begin{aligned}
 \gamma^*(\vec{x}, \tilde{\vec{x}}) &= \sum_i (\text{prob. that } t = t_i) (\text{prob. of } \vec{x} \text{ given } \gamma_i = \Lambda_{t_1,t_i} \circ \gamma_1) (\text{prob. of } \tilde{\vec{x}} \text{ given } \tilde{\gamma}_i = \sigma \circ \gamma_i) \\
 &= \sum_i \frac{1}{n} \gamma_i(\vec{x}) \sigma \circ \gamma_i(\tilde{\vec{x}}).
 \end{aligned} \tag{5}$$

Denote the cumulative distribution corresponding to  $\gamma^*$  by  $cdf^*$ . This distribution for any given dynamical system is shaped by its causal structure. In order to compare the degree to which dynamical structure differs between systems, our approach is to compare  $cdf^*$ . A suitable metric for doing so is the energy distance [75–77].

The energy distance for any two cumulative distributions,  $cdf^A$  and  $cdf^B$  over random variables  $X^A$  and  $X^B$  is defined as follows:

$$D_E^2(cdf^A, cdf^B) = 2\mathbb{E}[\|X^A - X^B\|] + \mathbb{E}[\|X^A - X^{A'}\|] - \mathbb{E}[\|X^B - X^{B'}\|], \tag{6}$$

where  $X^{A'}$  and  $X^{B'}$  are equivalent in distribution to  $X^A$  and  $X^B$ . The square root of the right hand side of Equation (6),  $D_E$ , is a proper metric, and is zero if and only if  $cdf^A = cdf^B$ .

We define the dynamical distance,  $D_D$ , according to the identity

$$D_D \equiv D_E(cdf^{*A}, cdf^{*B}) = \left( 2\mathbb{E}[\|X^A - X^B\|] + \mathbb{E}[\|X^A - X^{A'}\|] - \mathbb{E}[\|X^B - X^{B'}\|] \right)^{1/2}, \tag{7}$$

where the random variable  $X^A$  (equivalent in distribution to  $X^{A'}$ ) has values in the  $2n$ -dimensional space of joint states  $\langle \vec{x}^A, \tilde{\vec{x}}^A \rangle$ , and similarly for  $X^B$ . For two systems A and B, if at each time index  $i$ ,  $\gamma_i^A(\vec{x}) = \gamma_i^B(\vec{x})$ , then the difference between  $\gamma^{A*}(\vec{x}, \tilde{\vec{x}})$  and  $\gamma^{B*}(\vec{x}, \tilde{\vec{x}})$  and thus any difference between  $cdf^{*A}$  and  $cdf^{*B}$  must be due entirely to the action of the respective symmetries,  $\sigma^A$  and  $\sigma^B$ . In that case, measuring the dynamical distance  $D_D(cdf^{*A}, cdf^{*B})$  provides a quantitative comparison of the symmetries of systems A and B, and thus of the underlying casual structure that gives rise to them.

In general, an energy distance can be estimated by computing the sample mean for each of the expectation values on the right-hand side of Equation (6). For a sample of  $p$  points from system A and  $q$  from system B, there are  $p^2$  pairwise distances for estimating  $\mathbb{E}[\|X^A - X^{A'}\|]$ ,  $q^2$  for estimating  $\mathbb{E}[\|X^B - X^{B'}\|]$ , and  $pq$  for estimating the cross-term,  $\mathbb{E}[\|X^A - X^B\|]$ .

Accordingly,  $D_D$  can be estimated using a sample of  $2n$ -dimensional vectors,  $\langle \vec{x}^A(t_i), \tilde{\vec{x}}^A(t_i) \rangle$  from system A and  $\langle \vec{x}^B(t_i), \tilde{\vec{x}}^B(t_i) \rangle$  from system B, where  $\vec{x}^A(t_i)$  is the  $i$ -th point in a time series from system A beginning with an initial value  $\vec{x}^A(t_1)$  drawn according to  $\gamma_1^A(\vec{x}(t))$  and  $\tilde{\vec{x}}^A(t_i)$  is the contemporaneous  $i$ -th point in a time series from system A beginning with an initial value  $\tilde{\vec{x}}^A(t_1)$  drawn according to  $\tilde{\gamma}_1^A(\vec{x}(t))$ , and similarly for system B.

### 3.3. Temporal Scale Matching

The density  $\gamma^*$  depends on both the time-evolution operators  $\Lambda_{t_1, t_i}$  and the particular dynamical symmetry,  $\sigma$ , that maps  $\gamma_i$  to  $\sigma \circ \gamma_i$ . In general, two distinct systems will differ with respect to  $\Lambda_{t_1, t_i}$ , and so despite beginning with the same distribution it is impossible to find later times  $t, t'$  such that  $\gamma^A(\vec{x}(t)) = \gamma^B(\vec{x}(t'))$ . However, if  $\Lambda$  is sufficiently smooth and the means of the two distributions overlap over some nonempty time interval, then it is possible to find  $n$  times  $t_i$  and  $t'_i$  (such that  $t_{i+1} - t_i = c_1$  and  $t'_{i+1} - t'_i = c_2$  for some positive constants  $c_1, c_2$ ) for which  $\gamma^A(\vec{x}(t_i)) \approx \gamma^B(\vec{x}(t'_i))$ . In particular, one could sample each system over a sufficiently short interval of time relative to the natural scale of its dynamics in order to limit the evolution of each system to within acceptable deviation from the shared starting condition.

If the times  $t_i = t'_i$  are externally determined (as is often the case with data received for analysis by someone other than the experimenter or for data acquired by instrumentation with a fixed sampling period) but multiple time series are available from each system, then the same end can be achieved approximately by truncating or clipping one of the two sets of time series so as to effectively alter the time scale on which the corresponding system is sampled. We introduce an algorithm based again on an energy distance as defined in Equation (6). Suppose systems A and B are originally sampled at regular intervals at  $t_1, t_2, \dots, t_n$ , and for both systems, a number  $s$  of untransformed and a number  $s$  of transformed time series are provided. Then we seek the index  $m$  for either A or B such



that the following condition holds. If each time series in the untransformed set (from A or B) is truncated at  $m < n$  and for each index  $i$  of the truncated set of time series and each index,  $\text{floor}(i \frac{n}{m+1})$ , of the intact set of time series (from B or A) the energy distance between the sets of values at those indices is computed and the results averaged over  $i$ , there is no choice of  $m$  for either system A or system B that yields a lower average. The objective of clipping in this way is similar in spirit to the objective of dynamic time warping (to select a mapping of the indices of one time series to those of another that minimizes the sum of distances between the matched pairs) [78]. However, dynamic time warping requires that the first and last indices of the two time series compared coincide (leaving the overall time interval over which the system evolves intact), and unevenly alters the sampling interval throughout a given time series. Furthermore, dynamic time warping is not defined for ensembles of replicates from each system, while the energy distance is apt for comparing the similarity of  $\gamma^A(\vec{x}(t_i))$  and  $\gamma^B(\vec{x}(t_i))$  at each  $t_i$  given multiple time series samples from A and B. We therefore use the average energy distance over time to find an appropriate cut point for one or the other ensemble of time series.

If the set of transformed time series replicates for each system are clipped to match their corresponding untransformed curves, then the difference in  $D_D$  between the two systems is driven by the unknown dynamical symmetry,  $\sigma$ . For all numerical experiments reported here, we used the clipping algorithm expressed in Algorithm 1 when comparing time series from two systems. More specifically, we used Algorithm 1 to determine the optimal length at which to cut untransformed time series samples from A with the samples from B held fixed, and then repeated the process to determine the best length to cut samples from B with A held fixed. We ultimately clipped the data of whichever of the two systems resulted in the lowest value of  $D_D$  after clipping.

---

**Algorithm 1:** Clipping for temporal scale matching

---

**Input:** System states  $\vec{x}^A, \vec{x}^B$ , a minimum sequence length  $l_{min}$ , and a step size scalar  $s$

**Output:** An optimal clipping index  $m$

```

1  $\Delta \leftarrow (\text{len}(\vec{x}^A) - l_{min}) / s$ 
2  $distance_{min} \leftarrow inf$ 
3  $m \leftarrow l_{min}$ 
4 for  $i \leftarrow l_{min}$  to  $\text{len}(\vec{x}^A)$  step  $\Delta$  do
5    $\delta \leftarrow (\text{len}(\vec{x}^B) - l_{min}) / s$ 
6    $distances \leftarrow \emptyset$ 
7   for  $j \leftarrow 0$  to  $i$  do
8      $k \leftarrow \lfloor j * \delta \rfloor$ 
9      $distances_j \leftarrow D_E(\vec{x}_j^A, \vec{x}_k^B)$ 
10   $distance \leftarrow \text{mean}(distances)$ 
11  if  $distance < distance_{min}$  then
12     $distance_{min} \leftarrow distance$ 
13     $m \leftarrow i$ 
14 return  $m$ 

```

---

### 3.4. Partial Information and Degrees of Control

To make use of the dynamical similarity metric  $D_D$  to compare physical systems, one needs an appropriate collection of time series. Ideally, one would obtain multiple untransformed and transformed time series for each system beginning with initial values identically and independently distributed according to  $\gamma_i(\vec{x})$  and  $\tilde{\gamma}_i(\vec{x})$ , respectively, and where each system is sampled over a length of time such that for the pair of systems A and B to be compared,  $\gamma_i^A(\vec{x}(t_i)) \cong \gamma_i^B(\vec{x}(t_i))$  at each time index  $i$ . If one has control over the initial conditions (the ability to intervene on the system), this is manageable. However, in many cases of interest, this is impossible. Furthermore, it is often the case that the variables

in terms of which systems are described fail to satisfy the SSD condition because the set is incomplete or amounts to a noninvertible function of an SSD set. We describe methods for estimating our dynamical symmetry metric for every combination of these conditions.

### 3.4.1. SSD Variable Set and Full Control of the Initial Distribution

At one extreme, we are provided full control over the initial conditions from which each time series for an SSD set of variables begins. In this case, the simplest approach is to take multiple time series samples from each system, half of which begin with an untransformed initial value of  $\vec{x}_0$  and half of which begin at the transformed initial value,  $\vec{\tilde{x}}_0 \neq \vec{x}_0$ . The clipping algorithm described above can then be applied and the dynamical distance  $D_D$  estimated by using the resulting sets of time series to provide a sample of contemporaneous untransformed and transformed states,  $(\vec{x}^B(t_i), \vec{\tilde{x}}^B(t_i))$ , for each system.

### 3.4.2. SSD Variable Set and No Control of the Initial Distribution

Even if a set of variables is known or suspected to be SSD, it is often or typically the case that time series are provided for analysis without the ability to manipulate the system to select specific initial values. In the most difficult case, only a single extended time series is available. While one cannot set the initial conditions of such a passively observed system, one can—under the additional assumption that the dynamics of the system are autonomous—imagine that subsequences within the given time series are each an initialization of that system. By carefully selecting such subsequences as instances of a system's untransformed and transformed time series, it is still possible to estimate  $D_D$ .

In describing the algorithm, we assume that two systems (A and B) are to be compared, but the method generalizes straightforwardly to arbitrarily many systems for which one seeks to estimate pairwise commensurable dynamical distances. First, the time series for both systems A and B are broken into subsequences of uniform length  $l$ . To select a subset of these sequences that will be treated as the untransformed set of replicates and the subset that will be treated as the transformed set for each system, we pool the initial values of all subsequences and compute the normalized eigenvector  $\vec{v}$  with the largest eigenvalue  $\lambda$  for the covariance matrix  $\Sigma$  as well as the overall mean  $\vec{\mu}$  of the pool. We then compute two new means,  $\vec{\mu}_{\text{untrans}} = \vec{\mu} - \alpha\lambda\vec{v}$  and  $\vec{\mu}_{\text{trans}} = \vec{\mu} + \alpha\lambda\vec{v}$  where  $\alpha$  is a metaparameter controlling the degree of separation between the means of the two new distributions. Throughout, we have used a default value of  $\alpha = 1$  unless otherwise noted. We compute the singular value decomposition of the original  $n \times n$  covariance matrix  $\Sigma = USV^T$  where  $U$  and  $V^T$  are real,  $n \times n$  orthogonal matrices and  $S$  is an  $n \times n$  diagonal matrix with the singular values of  $\Sigma$  along the diagonal. We then construct a new covariance matrix,  $\Sigma' = U(\beta S)V^T$  where  $\beta$  is a second metaparameter determining the relative spread of the constructed distributions. We use  $\beta = 0.2$  unless otherwise indicated. Finally, for each system A and B, a set of untransformed replicates is chosen by selecting from the candidate subsequences those whose initial values have the highest density according to the  $n$ -dimensional normal distribution with mean  $\vec{\mu}_{\text{untrans}}$  and covariance matrix  $\Sigma'$ . Likewise, a set of transformed replicates is selected based on their density according to the  $n$ -dimensional normal distribution with mean  $\vec{\mu}_{\text{trans}}$  and covariance matrix  $\Sigma'$ . This selection procedure is detailed in Algorithm 2. Once sets of replicates have been chosen in this way, the dynamical distance can be estimated as usual. This procedure is shown schematically in Figure 1 and presented in detail in Algorithm 2.

**Algorithm 2:** Choose untransformed and transformed representatives from data

---

**Input:** Time series  $data$  of a system, sequence length  $l$ , and hyperparameters  $\alpha, \beta$   
**Output:** Untransformed replicates  $p$ , transformed replicates  $\tilde{p}$

```

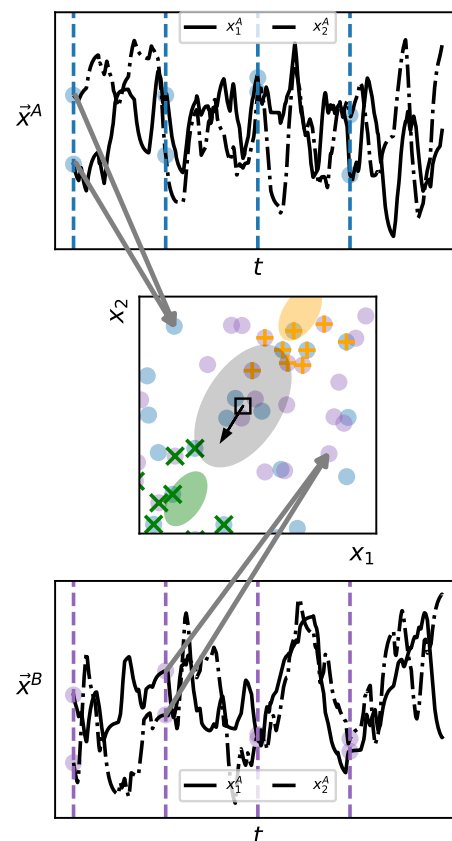
1  $\bar{\mu} \leftarrow \text{mean}(x_0 \in \vec{x} \in P \in data)$  // average of initial conditions
2  $\Sigma \leftarrow \text{covariance}(x_0 \in \vec{x} \in P \in data)$  // covariance matrix of initial
   conditions
3  $\vec{\lambda}, \mathbf{v} \leftarrow \text{eigendecomposition}(\Sigma)$  // compute eigenvalues and right
   eigenvectors
4  $\lambda \leftarrow \text{sqrt}(\max(\vec{\lambda}))$  // select largest eigenvalue
5  $\vec{v} \leftarrow \mathbf{v}[\text{index\_of}(\lambda \in \vec{\lambda})]$  // select eigenvector paired with  $\lambda$ 
6  $\bar{\mu}_{\text{untrans}} \leftarrow \bar{\mu} - \alpha * \lambda * \vec{v}$ 
7  $\bar{\mu}_{\text{trans}} \leftarrow \bar{\mu} + \alpha * \lambda * \vec{v}$ 
8  $U, S, V^T \leftarrow \text{svd}(\Sigma)$  // singular value decomposition of  $\Sigma$ 
9  $\Sigma' = U(\beta S)V^T$ 
10  $p \leftarrow \emptyset$ 
11  $\tilde{p} \leftarrow \emptyset$ 
12 foreach  $P \in data$  do
13    $q \leftarrow \emptyset$ 
14   foreach  $\vec{x} \in P$  do
15      $\lfloor \text{append\_a\_to\_b}(\mathcal{N}(x_0 | \bar{\mu}_{\text{untrans}}, \Sigma'), q)$  // value of  $\mathcal{N}$  at  $x_0$ 
16    $idx = \text{index\_of}(\text{sort}(q))[-r :]$  // select indices of  $r$  points with
     densest value in  $\mathcal{N}$ 
17    $p \leftarrow [data_i | i \in idx]$ 
18    $\text{delete}(data_i | i \in idx)$  // remove already selected sequences from
     candidates
19 foreach  $\tilde{P} \in data$  do
20    $\tilde{q} \leftarrow \emptyset$ 
21   foreach  $\vec{x} \in \tilde{P}$  do
22      $\lfloor \text{append\_a\_to\_b}(\mathcal{N}(x_0 | \bar{\mu}_{\text{trans}}, \Sigma'), \tilde{q})$ 
23    $i\tilde{d}x = \text{index\_of}(\text{sort}(\tilde{q}))[-r :]$ 
24    $\tilde{p} \leftarrow [data_i | i \in i\tilde{d}x]$ 
25 return  $p, \tilde{p}$ 

```

---

## 3.4.3. Non-SSD Variable Set and Full Control of the Initial Distribution

Not every set of variables captures enough about the dynamically or causally relevant aspects of a physical system to determine future states (or distributions over future states) from the state (or distribution over states) at a given time. When dynamical variables go unobserved or when an observed set of variables is a lossy function of a complete set then the set of variables will generally fail to meet the SSD condition. We generically refer to such a set of variables as “partial” and the systems they describe as “partially observed.” While the methods described above for estimating  $D_D$  do not work directly for partial variable sets, it is still possible in many circumstances to use the dynamical distance to discriminate among partially observed systems. For a system that is not SSD, the distribution over states  $\gamma(\vec{x}(t))$  at a time  $t$  does not uniquely determine the distribution at a later time  $t' > t$ . However, when the failure to meet the SSD condition is because the system is partially observed, there exists an unknown set of variables the states of which would uniquely determine  $\gamma(\vec{x}(t))$  when supplemented by the observed variables. If transformed and untransformed time series can be obtained for each of two systems where the marginal distribution over the unobserved variables at  $t_1$  is the same for both sets of untransformed series and the same for both sets of transformed series, then differences in the apparent values of  $D_D$  will still be driven by the dynamical symmetries of the systems.



**Figure 1.** By fragmenting long time series, sets of appropriately selected fragments can serve as untransformed and transformed sets for estimating  $D_D$ . For two systems A and B, time series are divided into fragments of equal length (**top** and **bottom** plots), and the initial values of each segment from A and B are pooled in a common  $n$ -dimensional space (where  $n$  is the number of variables for each system) (gray arrows to **center** figure). For the pooled initial values, an overall mean is computed (black box, **center** figure) and the unit eigenvector  $\vec{v}$  corresponding to the largest eigenvalue of the covariance matrix  $\Sigma$  is determined (black arrow, **center** figure). The major and semi-major axes of the gray ellipse at the center indicates the magnitude of the first and second singular values of  $\Sigma$ . Two new means are determined by moving in opposite directions along  $\vec{v}$  (indicated by the centers of the green and orange ellipses), and two new covariance matrices constructed by scaling the original (indicated with the major and semi-major axes of the green and orange ellipses). Finally, a predetermined number of fragments are selected for inclusion in the untransformed and transformed sets (orange '+'s and green 'x's respectively) for both systems by identifying initial values (points in the plane, center figure) with the highest density according to two  $n$ -dimensional normal distributions with the newly determined means and covariance matrices.

Given control of the initial state of each system to be compared, it is often the case that the procedure for setting the initial state in terms of the partial set of variables does fix a distribution over the unobserved variables. While this cannot be guaranteed, it can be tested given sufficient numbers of untransformed and transformed time series by verifying a fixed distribution over later times for a given initial (partial) state. The procedure for systems for which it is possible to intervene to set initial conditions but which are suspected of failing the SSD condition is thus to acquire multiple time series samples for each system for each initial condition, and then to estimate  $D_D$  as above.

#### 3.4.4. Non-SSD Variable Set and No Control of the Initial Distribution

The most difficult scenario for accurately assessing dynamical similarity, regardless of the method used, is the case in which data are passively acquired such that there is no opportunity to set initial conditions and partially observed such that the given system of

variables is therefore non-SSD. Even in this context, it is still possible to estimate  $D_D$  at the price of some additional assumptions about the systems under consideration, namely that the all dynamical variables are bounded, that the dynamics is such that given sufficient time the system will pass nearby any system state previously observed, and that for any observed state there is a stationary distribution over states of the unobserved variables. Such a bounded system observed over a sufficiently long time will often exhibit an approximately stationary “distribution of distributions” that makes it appear SSD. For example, observing only the angular position of a pendulum does not determine its future position (angular position alone is not a state-determined set of variables). However, watching the pendulum swing for a while, one builds up a time series in which, from nonextremal positions, it moves left half the time and right half the time. So the system, though not state-determined, appears SSD. A similar situation obtains for stochastic dynamics.

Given that the available time series is long enough to meet these assumptions, then the dynamical distance  $D_D$  can be estimated using the same algorithm as described in Section 3.4.2 for the case of a passively observed, SSD set of variables.

### 3.5. Change Detection

The methods for computing  $D_D$  described above for conditions in which there is no control over the initial distribution can be used to detect changes in dynamics manifest in an extended time series. To identify such change points, a rolling window method can be used. For a window width of  $w$  and a lag (the space between the leading and trailing window) of  $l$  and a discrete time series,  $\vec{x}(t_i), i \in \{1, \dots, n\}$ , we treat the subsequences  $\langle \vec{x}_i, \dots, \vec{x}_{i+w} \rangle$  and  $\langle \vec{x}_{i+w+l}, \dots, \vec{x}_{i+2w+l} \rangle$  as time series from separate systems, and compute the value of  $D_D(t_{i+w+l/2})$  between them using the method of either Sections 3.4.2 or 3.4.4, depending on whether the system is thought to be SSD. The appropriate window width depends upon how many replicates are required, and the minimum length of each. The best choice for the lag  $l$  will depend on the time over which the change in dynamical structure occurs. If  $l$  is shorter than the transition period, then both the leading and trailing windows will contain part of the transitional dynamics as they pass over the change, which tends to suppress the elevation of  $D_D$ , making it easier to miss the transition. If  $l$  is overly large, then it limits one’s ability to pinpoint the time of transition.

## 4. Numerical Experiments and Results

### 4.1. Difference in Dynamical Kind

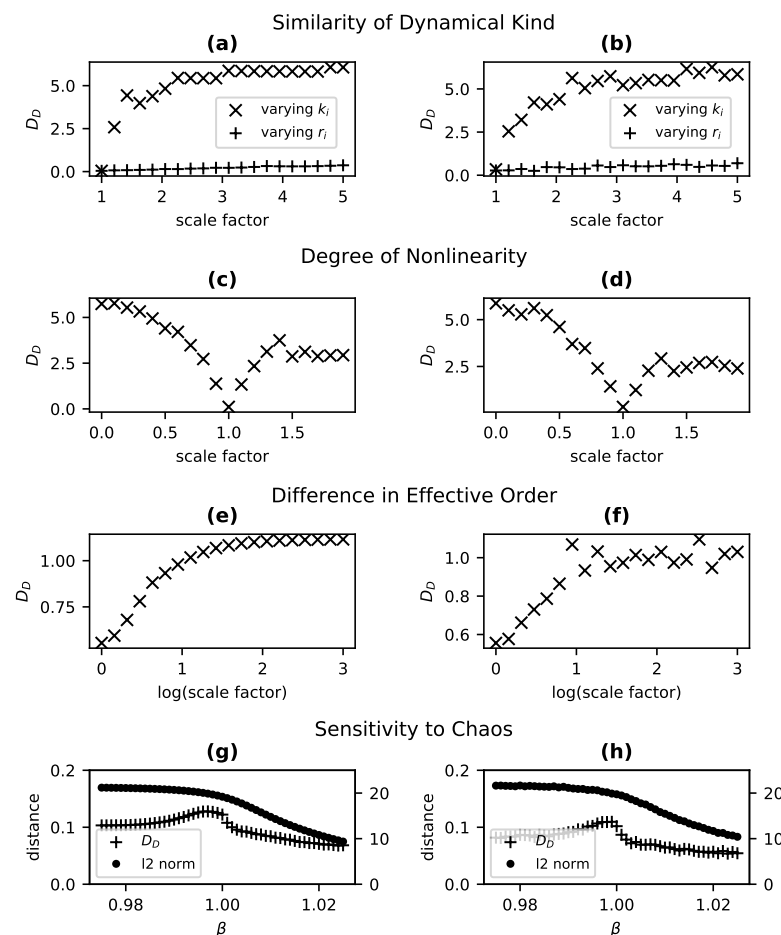
We conducted numerical experiments with Lotka–Volterra ODE models and various derivatives designed to isolate one or another dynamical aspect in order to assess the extent to which  $D_D$  is sensitive to linearity, effective order, chaos, and similarity of dynamical kind. For assessing sensitivity to dynamical kind as defined in Section 2.3, we numerically integrated a two-species instance of the general  $n$ -species Lotka–Volterra predator–prey system [79]:

$$\frac{dx_i}{dt} = r_i x_i \left( 1 - \frac{\sum_{j=0}^n a_{ij} x_j}{k_i} \right), i = 0, 1, \dots, n, \quad (8)$$

where  $x_i$  denotes the population size of the  $i^{\text{th}}$  species,  $k_i$  denotes its carrying capacity,  $r_i$  denotes the intrinsic growth rate of the species, and  $a_{ij}$  is the interaction coefficient of species  $j$  on species  $i$ . Any two systems related by scaling all  $r_i$  by a positive constant share the same dynamical symmetries and thus belong to the same dynamical kind. Conversely, scaling the carrying capacities  $k_i$  results in a different dynamical kind [73].

In order to demonstrate that the metric  $D_D$  provides a degree of difference in dynamical kind (and so generalizes the binary decision process of [73]), we use a reference system A with  $\vec{r}^A \equiv [r_1^A, r_2^A] = [1, 2]$ ,  $\vec{k}^A \equiv [k_1^A, k_2^A] = [100, 100]$  and a comparison system B with  $\vec{r}^B = s_r * \vec{r}^A$  and  $\vec{k}^B = s_k * \vec{k}^A$ . For each test, we generate two time series for each system: an untransformed series with an initial population of  $\vec{x} = [5, 5]$ , and a transformed series with an initial state of  $[8, 8]$ . When  $s_r$  is fixed at a value of 1 and  $s_k$  is varied for

system B,  $D_D$  increases with increasing  $s_k$  (Figure 2a) and thus indicates divergence in dynamical kind, as expected. The same relationship is apparent in Figure 2b, for which the experiment was repeated using time series to which normally distributed noise with a mean of 0 and standard deviation of 5 (equivalent to 5% of the dynamical range of the system) has been added. Note that  $D_D$  approaches a maximum value. For any dynamical system for which states are bounded (such that there is some  $c > 0$  for which  $|\vec{x}| < c$ ),  $D_D$  is bounded from above.



**Figure 2.** Sensitivity of  $D_D$  to similarity of dynamical kind is demonstrated for two-species competitive Lotka–Volterra systems measured without sampling noise (a) or with normally distributed sampling noise with  $\mu = 0$  and  $\sigma = 5$  (b). Relative to a system with growth rates of  $\vec{r} = [1, 2]$  and carrying capacities,  $\vec{k} = [100, 100]$ ,  $D_D$  increases as the carrying capacities  $\vec{k}$  of the comparison system are multiplied by an increasing scaling constant such that the systems belong to diverging dynamical kinds ( $\times$ ), but remains approximately 0 as the growth rates  $\vec{r}$  are scaled ( $+$ ), which is a dynamical symmetry connecting systems of the same dynamical kind. When sensitivity to nonlinearity is assessed by multiplying the interaction matrix by a scale factor relative to the reference system with  $\alpha = [[1, 0.5], [0.7, 1]]$ , giving a linear system at a scale factor of 0,  $D_D$  increases as the nonlinearity of the comparison system is increased (c), even in the presence of sampling noise (d).  $D_D$  also increases between systems as their effective order diverges. For a modified Lotka–Volterra system that is second order with a scale factor of 1 and approaches first order as that scale factor tends to infinity,  $D_D$  increases rapidly from 0 relative to a reference system with a scale factor of 1, with sampling noise (e) and without (f).  $D_D$  also responds specifically to chaos, rising relative to a nonchaotic reference system of similar nonlinearity and order as a four-species Lotka–Volterra system passes through a chaotic transition in parameter space along the  $\beta$  direction (defined in Equation (11)) ( $+$ ), both with (g) and without (h) sampling noise. The  $l^2$  norm ( $\cdot$ ), on the other hand, changes monotonically across the chaotic transitions.

#### 4.2. Nonlinearity

Of broader interest is the sensitivity of the proposed metric  $D_D$  to the degree of linearity of one system relative to a benchmark system. To assess this aspect of performance, we contrived a modification of the Lotka–Volterra equations (Equation (8)) that allows us to modulate the system’s degree of nonlinearity:

$$\frac{dx_i}{dt} = r_i x_i \left( 1 - \frac{\sum_{j=0}^n a_{ij} l x_j}{k_i} \right), i = 0 \dots n \quad (9)$$

where the factor  $l$  scales the degree of nonlinearity of the system. When  $l$  is 0, the system is perfectly linear, while the nonlinear term dominates for large values of the scale factor  $l$ . We numerically integrated two parameterizations of a two-species version of this model, system A and system B. Both simulated systems are provided with identical growth rates and carrying capacities ( $\vec{r} = [1, 2], \vec{k} = [100, 100]$ ). System A’s value of  $l$  was fixed at 1, while the value of the scale factor  $l$  for system B was varied from 0 to 1.8. As before, the differential equations were numerically integrated from  $t = 0$  to 15. Each system’s untransformed initial population is 5 members, and each transformed initial population is 8. The calculation of  $D_D$  was repeated for an experiment using identical systems but for which normally distributed observation noise ( $\sigma = 5$ ) was added.

As seen in Figure 2c,d,  $D_D$  goes to 0 where the nonlinearity factor is 1 and both systems are identical. It reaches a maximum when the nonlinearity factor is 0, and thus one system is fully linear while the other involves substantial nonlinear interaction. As the value of  $l$  increases above 1,  $D_D$  appears to increase asymptotically as the nonlinear term comes to dominate. The dynamical distance is thus sensitive to relative decreases and increases in effective linearity of dynamics for this sort of system.

#### 4.3. Effective Order

The effective order of a system determines its dependence on prior history and the complexity of boundary conditions necessary to forecast the system. Isolating order required constructing a second order version of the Lotka–Volterra equations (Equation (8)) outfitted with a scale factor ( $\omega$ ):

$$\frac{1}{\omega} \frac{d^2 x_i}{dt^2} + \frac{dx_i}{dt} - r_i x_i \left( 1 - \frac{\sum_{j=0}^n a_{ij} x_j}{k_i} \right) = 0 \quad (10)$$

As  $\omega$  approaches infinity, the system approaches first order dynamics. In this test, the parameters of systems A and B are the same as the systems in Section 4.2, but for the absence of  $l$  and the presence of  $\omega$ . System A, whose  $\omega$  value set to 1, was compared to system B as its  $\omega$  value grew exponentially. These systems were also run from  $t = 0$  to 15 for our tests.

We expect to see that systems with diverging effective order are farther apart according to the metric  $D_D$ . This expected pattern is evident without sampling noise (Figure 2e) and with sampling noise (Figure 2f). The distance between the two systems grows with the order factor, leveling out as the large  $\omega$  value causes system B to approximate first order dynamics.

#### 4.4. Chaos

Testing the sensitivity of our distance metric to the presence of chaos requires us to compare a known chaotic system to similarly parameterized systems that, so far as possible, hold constant the effective order and degree of nonlinearity. Ideally, one would identify a single parameter that can be varied to move through the chaotic region of parameter space.

To meet these constraints, we use a four-species Lotka–Volterra system described by Equation (8) ( $n = 4$ ). This system is known to exhibit chaos for a range of parameterizations, and a portion of this parameter space was explored in [80]. The latter identify three chaotic points in parameter space, labeled  $(R_1, A_1)$ ,  $(R_2, A_2)$ , and  $(R_3, A_3)$ , where each  $R_i$  is a

particular vector of intrinsic growth rates ( $\vec{r}$  in our notation), and the  $A_i$  are interaction matrices ( $a$  in our notation). A 2-D plane in parameter space is then defined by linear combination of these points determined by a pair of coefficients  $\alpha$  and  $\beta$  as follows:

$$(R, A) = (R_1, A_1) + \alpha(R_2 - R_1, A_2 - A_1) + \beta(R_3 - R_1, A_3 - A_1) \quad (11)$$

In our experiments, we chose a nonchaotic reference system at  $\alpha = 0.2, \beta = 1.0$  and computed  $D_D$  relative to systems along a slice through this parameter space from  $\alpha = 0, \beta = 0.975$  to  $\alpha = 0, \beta = 1.025$ . For these experiments, carrying capacities are fixed at 1 for all species (equivalent to expressing each population as a fraction of its carrying capacity). The untransformed time series begin with all species at a population of 0.1, and 0.2 for the transformed time series. The result in the absence of noise (Figure 2g, + marks) is a clear rise in  $D_D$  in the vicinity of  $\beta = 1$  (where the behavior is known to be chaotic) with relatively constant distance over variations in  $\beta$  on either side of the chaotic (and near-chaotic) region. When the experiments are repeated with normally distributed observation noise,  $\mathcal{N}(0, 0.05)$ , added to each time series, a nearly identical pattern is observed, albeit shifted to a lower mean value of  $D_D$  (Figure 2h, + marks).

As a benchmark, we also computed the  $l^2$ -norm between the reference and comparison time series (specifically, the untransformed time series) for each value of  $\beta$  (Figure 2g, • marks). This provides a metric sensitive to differences in the gross shapes of trajectories. The  $l^2$  norm declined monotonically over the chaotic region, providing no indication that such a transition has taken place, regardless of whether or not sampling noise was included.

#### 4.5. Model Mapping: Lotka–Volterra

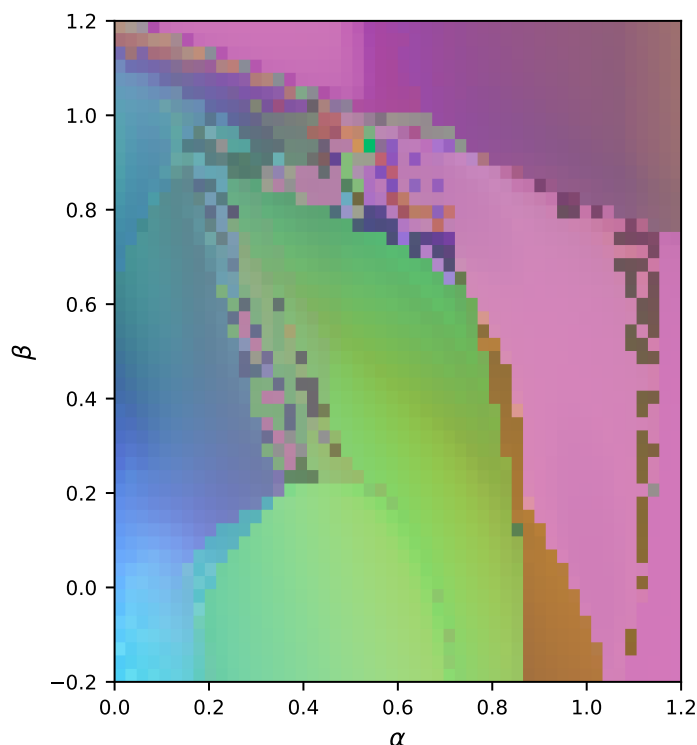
A generic metric of dynamical similarity provides a common yardstick for comparing arbitrarily many systems in a model-free way. Consequently, it offers the possibility of mapping out a portion of the parameter space for a system of interest, summarizing the overall similarity in dynamical behavior between any two such points. To demonstrate the efficacy and utility of such a unique dynamical map, we computed values of  $D_D$  for all pairs of parameter values in a discretization of the parameter space of the four-species Lotka–Volterra system described in Equation (8) and explored by [80].

Untransformed and transformed time series samples are collected from regularly arranged points in the rectangular region of parameter space spanned by values of  $\alpha$  between 0 and 1.2, and  $\beta$  from  $-0.2$  to 1.2, with 50 equally spaced divisions along each axis. For all systems we sample, we set the carrying capacity of each species to 100 and the initial populations for all species to 5 for the untransformed systems, and to 8 for the transformed systems. Each system is simulated for 4000 time steps, with a  $\Delta t$  of 0.5.

The result of these computations is a  $2500 \times 2500$  distance matrix. To visualize the similarity relations implied by this matrix, we deploy multidimensional scaling (MDS) to embed each point of our discrete parameter space in a continuous three-dimensional geometric space. Positions in this space are then normalized (such that positions in all three directions are within the interval  $[0, 1]$ ), and interpreted as colors in a standard color space (RGB color space). The map is thus constructed by coloring each cell of the  $50 \times 50$  grid in the  $\alpha - \beta$  parameter plane according to the color assigned by the MDS embedding. The result is a map of parameter space for which the hue of any one pixel is meaningless, but for which similarity of color corresponds to dynamical similarity, i.e., systems that are dynamically indistinguishable according to  $D_D$  are shown with the same color. In other words, the closer two pixels are in hue, the closer they are in the dynamical space defined by  $D_D$ .

The resulting map (Figure 3) can be directly compared with that in [80]. The latter colored the plane in parameter space based on the number of species remaining at steady-state. The dynamical similarity map captures much of the same broad structure, but reflects significantly more nuance since the mere number of species at equilibrium is a poor indicator of, e.g., effective order or degree of nonlinearity. Note that the chaotic points at  $(0, 0)$ ,  $(0, 1)$  and  $(1, 1)$  do not stand out as distinct from surrounding points. This is likely due to the coarse grain of our map and the extreme narrowness of the chaotic regions.





**Figure 3.** Exploration of the parameter space of a four-species Lotka–Volterra model in the same  $\alpha$ - $\beta$  plane explored in [80] and defined in Equation (11). The absolute hue of a pixel is meaningless. However, the closer two pixels are in color, the smaller the dynamical distance  $D_D$  between them. This color mapping is achieved by using multidimensional scaling and the distance matrix of all pairwise  $D_D$  values to embed every point of the depicted parameter space in a three-dimensional space that is then interpreted as RGB color space.

Using a low-dimensional embedding, constructed without normalization, we have discovered well-structured regions of similar dynamical structure, demonstrating the application of our algorithms as a novel method for exploring a system’s behavior.

#### 4.6. Stochastic Dynamics and Partial Systems

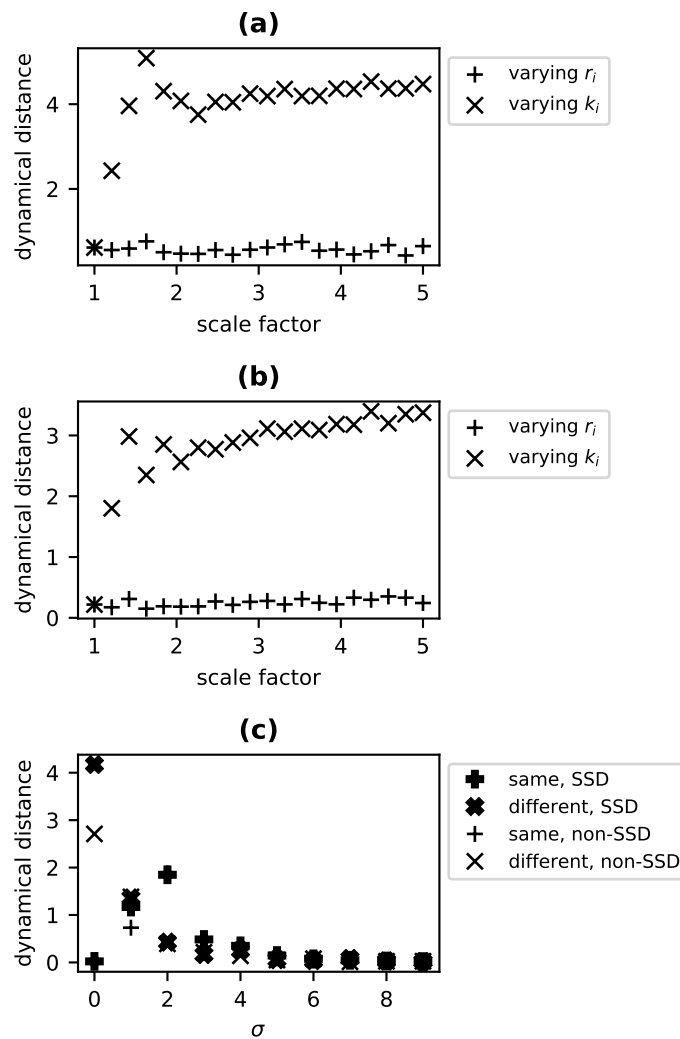
The simulation experiments described above all involve state-determined systems, with or without measurement noise. However,  $D_D$  applies to and can be estimated for systems that are only SSD. Furthermore, as described in Section 3.4.3,  $D_D$  can be estimated for systems that are SSD but only partially observed. To assess the effectiveness of these estimation procedures, and the sensitivity of  $D_D$  for SSD systems that are not state-determined, we constructed a stochastic version of the Lotka–Volterra equations based on [81]:

$$\Delta x_i(t) = x_i(t)r_i(t)\left(1 - \sum_{j=0}^n a_{ij}x_j(t)\right)\Delta t + \sigma_i x_i(t)\zeta_i(t)\sqrt{\Delta t} + \frac{\sigma_i^2}{2} x_i(t)\left(\left(\zeta_i(t)\right)^2 - 1\right)\Delta t, \quad (12)$$

where  $\sigma_i$  is the intensity of the noise for species  $i$ , and  $\zeta_i$  is a Gaussian random variable of the form  $\mathcal{N}(0, 1)$ .

We then numerically integrated the two-species version of this system in order to reproduce one of the basic tests described in Section 4.1, namely the test of sensitivity to the sameness of dynamical kind, in the context of stochastic systems in SSD and partially observed circumstances. As in the experiments of Section 4.1, a reference system with  $\vec{r} = [1, 2], \vec{k} = [100, 100]$  was compared with either a system for which the growth rates had been multiplied by a scale factor (which does not impact dynamical kind for this stochastic system), or a system for which the carrying capacities were multiplied by the scale factor (which results in increasingly distinct dynamical kinds). For these stochastic

experiments, ten replicates were included in each of the untransformed and transformed sets of time series (rather than the single replicates used for state-determined systems). The dynamical distance  $D_D$  as a function of the value of the scale factor for the two conditions (same dynamical kinds vs. different), depicted in Figure 4a, exhibits the expected pattern, identical with that observed for the state-determined Lotka–Volterra system: varying  $\vec{r}$  results in a constant  $D_D \approx 0$  while varying  $\vec{k}$  shows a rapid, apparently asymptotic increase in  $D_D$ . This test was repeated except only the average over species populations was provided at each time step, amounting to a one-dimensional, non-SSD set of variables. Nonetheless, when processed with the procedure indicated in Section 3.4.3, the same trend was apparent (Figure 4b).



**Figure 4.** (a) Dynamical distance  $D_D$  between a stochastic, two-species Lotka–Volterra system with  $\vec{r}$ ,  $\vec{k}$ , and  $\alpha$  as in Figure 2 and one for which  $\vec{r}$  is multiplied by the indicated scale factor (+), keeping both systems in the same dynamical kind, or for which  $\vec{k}$  is varied (x), moving the systems into increasingly distinct dynamical kinds. (b) The same comparison as in (a) except that only the average of number of species in each system (a partial set of variables that is not SSD) is provided for computing  $D_D$ . (c) Dynamical distance between stochastic Lotka–Volterra systems of the same dynamical kind ( $\vec{r}$ ,  $\vec{k}$ , and  $\alpha$  as in (a) for one system, and  $\vec{r}$  doubled for the other) described with an SSD (+) or non-SSD (x) set of variables, and between systems in different dynamical kinds ( $\vec{r}$ ,  $\vec{k}$ , and  $\alpha$  as in (a) for one system, and  $\vec{k}$  doubled for the other) as a function of the parameter  $\sigma$  which scales the Brownian term in the governing stochastic differential equation (higher  $\sigma$  corresponds to greater stochasticity, approaching a pure Brownian process as  $\sigma \rightarrow \infty$ ).

Finally, we examined the behavior of  $D_D$  as the degree of stochasticity is increased. Specifically, for 10 values of  $\sigma$  between 0 and 9, we generated 50 untransformed and transformed time series for three two-species systems: a reference system A for which  $\vec{r} = [1, 2], \vec{k} = [100, 100]$ , a system B of the same dynamical kind for which  $\vec{r} = [2, 4], \vec{k} = [100, 100]$ , and a system C of a distinct dynamical kind for which  $\vec{r} = [1, 2], \vec{k} = [200, 200]$ . We then computed  $D_D$  for system A and system B (SSD systems of the same dynamical kind;  $\blackspadesuit$  marks in Figure 4c) and for Systems A and C (of different dynamical kind;  $\blackcross$  marks in Figure 4c). When  $\sigma = 0$ , the system is state-determined and we merely replicate the findings of Section 4.1: for systems of the same kind,  $D_D \approx 0$  and systems of differing kinds,  $D_D \approx 4$ . As  $\sigma$  increases, however, these differences diminish. One would expect that as  $\sigma \rightarrow \infty$  the systems would all behave as Brownian process and thus converge on a distance of  $D_D = 0$ . This is clearly the case, with all values of  $D_D$  near 0 for  $\sigma \geq 6$ .

A similar pattern is obtained when non-SSD systems are considered. Specifically, we ran the same experiment for the same values of  $\sigma$  but provided only the mean value of species populations for computing  $D_D$ . Though the distances are somewhat suppressed for low values of  $\sigma$ , the metric  $D_D$  clearly distinguishes between systems of different dynamical kinds ( $+$  marks in Figure 4c) while systems of the same dynamical kind ( $\times$  marks in Figure 4c) exhibit distances near 0, and  $D_D$  approaches 0 for all pairwise comparisons as increasing  $\sigma$  results in a pure Brownian process.

#### 4.7. Change Detection

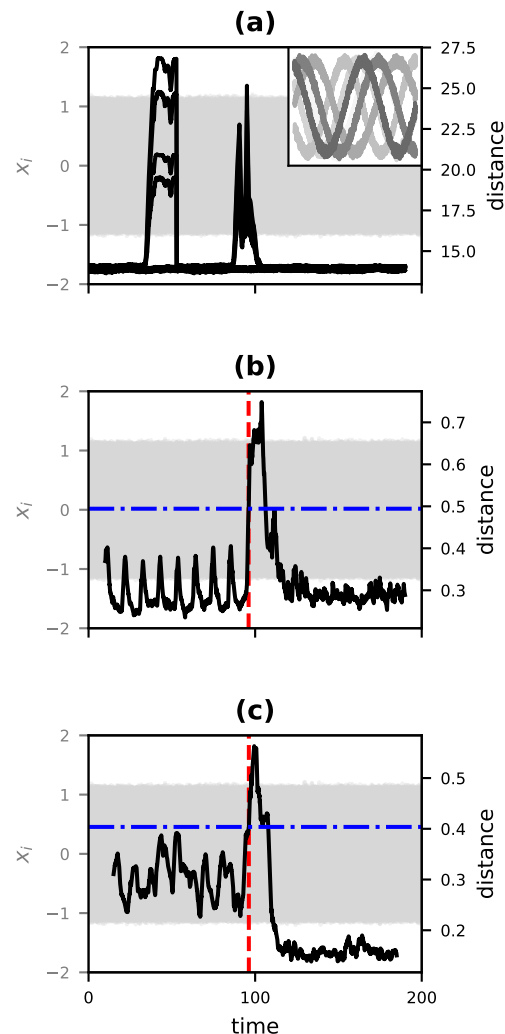
One of the principal applications of a general metric of dynamical similarity is the detection of shifts in the causal structure governing the behavior of a system. By computing the dynamical distance between time-lagged fragments of a time series as described in detail in Section 3.5, we can expect to find peaks of dissimilarity when the fragments differ in their underlying generative dynamics. This indicates that a structural change has occurred between the time of the first and second fragment. To test the specific approach described in Section 3.5, we simulated a three-member Kuramoto phase-oscillator system [82], generalized as:

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N} \sum_{j=0}^N \sin(\theta_j - \theta_i), i = 0 \dots N \quad (13)$$

where  $\theta_i$  is the phase of the  $i$ th oscillator,  $\omega_i$  is that oscillator's natural frequency,  $N$  is number of oscillators in the system, and  $K$ , the coupling coefficient, determines how much the difference in angle between oscillators affects an oscillator's future state. For this experiment, the three-member Kuramoto system was numerically integrated for 200 time units (with a step size of  $5 \times 10^{-4}$ ), and began with a coupling value of  $K = 1$ . From 95 time units to 105, the value of  $K$  is decreased linearly until at  $t = 105$  the oscillators are completely uncoupled ( $K = 0$ ) and so cannot influence one another. The data are then converted into rectangular coordinates, recording the sine and cosine of each phase, resulting in a 6-dimensional time series. A short segment of this series is shown in the inset of Figure 5a, and the full time series for all six variables is shown in light gray in all three panels of Figure 5.

As a benchmark for comparison, we computed the multidimensional matrix profile for the time series using the Stumpy Python package [83]. The result is shown as the six black curves in Figure 5a, and clearly demonstrates a false positive at a point (centered around  $t \approx 40$ ) well before the actual start of the transition in causal structure at  $t = 95$ . For using the rolling-window method of Section 3.5, we used a window size of  $2 \times 10^4$  time steps (10 time units) with the leading window and trailing window separated by 100 time steps (0.05 time units). At each step for which  $D_D$  was computed, the time series of each window was divided into 200 fragments and then processed as described in Section 3.4.2. The resulting time series of  $D_D$  values was smoothed as a rolling average with a window width of 50 time steps (0.025 time units). The result is shown as the black curve in Figure 5b.

The initial  $6 \times 10^4$  time steps (30 units of simulated time) was assumed to derive from a constant dynamical structure and was used as a reference series to set a threshold for anomalously high values of  $D_D$  that would indicate a change in dynamics. The dashed blue horizontal line shows this threshold, set at 3 standard deviations of the values of  $D_D$  for the reference series. The red dashed vertical lines indicate the location at which  $D_D$  crosses this anomaly threshold. This corresponds very closely with the onset of the transition at  $t = 95$ .



**Figure 5.** Comparison of change detection methods for a six-variable (3 phase-oscillator) Kuramoto system that transitions from uncoupled to weakly coupled over a short interval centered at 100 time units. (a) The matrix profile for each variable as a function of time (black lines); time series for all six variables are shown in the background in light gray, and over a short span of time in the inset. (b) For the same system and data as (a), the dynamical distance between two moving windows symmetrically arranged around each time is shown in black. The three standard deviation threshold for change detection is depicted as a dashed blue line while the vertical red dashed line indicates a detected change in dynamics. (c) The moving-window dynamical distance is shown (black line) for data from the same Kuramoto system but for which only one rectangular coordinate is provided from each phase oscillator (shown in light gray). The detection threshold is depicted as a dashed horizontal blue line, and the detected change event shown with a vertical dashed red line.

To assess the efficacy of using  $D_D$  for change detection in the more realistic scenario of a partially observed system, we repeated this experiment but discarded three of the six rectangular coordinates describing the system (preserving only the cosines of the phases). In order to compensate for the lost information, we increased the window width to  $3 \times 10^4$

time steps (15 time units). As shown in Figure 5c, the only change detected again occurs at the actual transition point at  $t = 95$ .

## 5. Discussion

The concept of a degree of dynamical similarity implicitly underwrites methods for solving a variety of problems, including detection of structural change, model selection and validation, and the discrimination of nonlinearity, effective order, and chaos in a system of interest. While there exist a range of bespoke metrics, measures, and methods for probing or comparing each of these aspects for one or more systems of interest, we propose the first generic metric of dynamical similarity sensitive to all of these aspects.

The dynamical distance,  $D_D$  is constructed of two parts. In the first place, we define a special cumulative probability density,  $cdf^*$ , for each dynamical system we wish to compare, such that the density in question differs if and only if the two dynamical systems differ with respect to their underlying dynamical symmetries. The latter are physical transformations of a system that commute with its time evolution, and are diagnostic of the underlying causal structure that generates the dynamics. Thus, the density  $cdf^*$  is an indicator of causal structure. The second component of  $D_D$  is the choice of a suitable metric for comparing  $cdf^*$  between two systems. For this, we deploy the energy distance defined by [75].

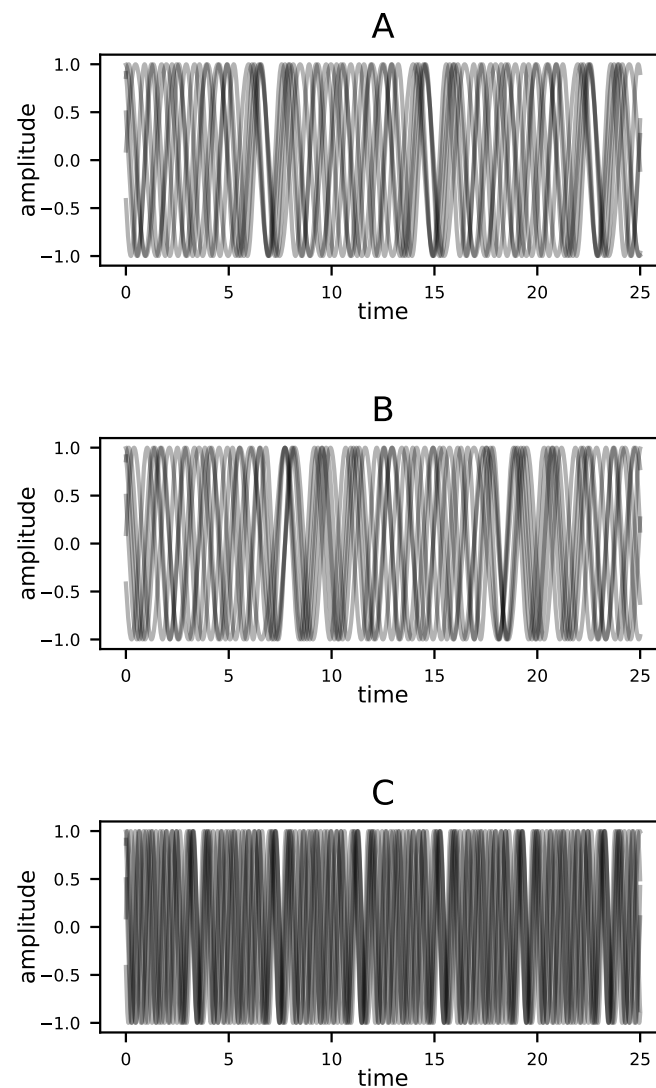
We have demonstrated that  $D_D$  is an indicator of the degree of similarity of the so-called dynamical kinds to which two systems belong. Specifically,  $D_D = 0$  if and only if two systems belong to the same dynamical kind and thus share a portion of their causal structure. More importantly, the use of  $D_D$  to assess dynamical kind improves upon the binary test of [73] by providing a degree to which two systems differ in dynamical kind and thus underlying causal structure. We have also demonstrated that  $D_D$  is sensitive to each of the traditional dynamical features mentioned above: nonlinearity, chaos, and effective order. Specifically,  $D_D$  increases as two systems diverge in character along one of these dimensions, e.g., with respect to the degree of nonlinearity or the presence of chaos.

The proposed metric  $D_D$  offers a number of unique advantages when it comes to detecting differences in these dynamical features from time series data. The metric is model free and domain general. It works in the presence of significant sampling noise, and the very same metric applies to both state-determined and stochastic systems. A disadvantage, however, lies in its comparative nature and the fact that a mere difference in  $D_D$  cannot be attributed to one or another aspect of dynamics (or the causal structure generating the dynamics) without additional information. However, with suitable reference systems—whether physical or modeled—or with assumptions about the plausible class of models, this can be overcome. For instance, divergence in  $D_D$  from a linear reference model for a system known to be nonchaotic and first order suggests nonlinearity. This approach requires no stronger assumptions than the specialist methods for detecting, e.g., linearity, and comes with the advantage of robustness to noise and of having a single tool equally applicable to state-determined and stochastic dynamics.

The greatest advantage of  $D_D$  is that it reflects structural features of the dynamics, whether deterministic or stochastic, and is otherwise indifferent to the shape or statistical properties of time series. To make this point clear, consider time series from three different Kuramoto oscillator systems, as shown in Figure 6. Each system, A, B, and C, comprises three phase oscillators governed by the equation,

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{3} \sum_{j=1}^3 \sin(\theta_j - \theta_i), \quad (14)$$

where  $i$  ranges from 1 to 3 [84]. The figure shows each phase,  $\theta_i$ , in rectangular coordinates as a pair of curves corresponding to  $\sin(\theta_i)$  and  $\cos(\theta_i)$  as functions of time.



**Figure 6.** Time series from three different Kuramoto phase oscillator systems (see Equation (14)) for which each oscillator phase  $\theta_i$  is represented by a pair of amplitudes  $\sin(\theta_i)$  and  $\cos(\theta_i)$ . Despite the superficial resemblance between the time series for systems A and B, the oscillators in System B are uncoupled. Systems A and C are structurally identical (every oscillator influences every other), and differ only in the set of natural frequencies,  $\omega$ .

For Systems A and C,  $K = 1$  and the oscillators are coupled. For System C, the natural frequencies  $\omega_i$  are twice the corresponding values for System A. However, the causal structure is the same; each oscillator influences every other oscillator. For System B, on the other hand,  $K = 0$  and the three oscillators are independent; none influences any other. Yet in terms of the statistics and gross appearance of trajectories, System C stands out as the exception. The dynamical distance proposed here is sensitive to the causal structure and not the appearance of the trajectories. Table 1 shows the values of  $D_D$  for each pairwise comparison of the three systems, computed using the single time series given for each system and the methods described in Section 3.4.2. The distances between A and B and between B and C are an order of magnitude greater than that between A and C. In other words, the dynamical distance  $D_D$  sees A and C, the systems with coupled oscillators, as similar and System B, which has the aberrant causal structure, as the standout. The second column of the table shows that this is not a matter of simply comparing time series statistics; the difference in mean values of the systems would suggest that A and B are most similar. More compellingly, the third column shows the minimum value of the globalized distance (the summed  $l_2$ -norm normalized by path length) between each pair of time series

after dynamical time warping using the dtw-python package [85]. This method, which is effective at detecting the relative similarity of time series shapes, identifies Systems A and B as the most similar. This is not wrong. Rather it is the correct answer to a question different from that which  $D_D$  addresses. The dynamical distance is concerned with similarity in the underlying dynamics, not with the similarity of particular time series those dynamics happen to produce.

**Table 1.** Distances between the systems represented in Figure 6.  $D_D$  is the dynamical distance computed between the indicated systems in each row using the methods described in Section 3.4.2 for singly sampled observational time series.  $\Delta\mu$  is the difference in mean values between the indicated time series, while  $DTW$  is the minimum global distance (normalized by path length) obtained after dynamic time warping to align the indicated time series. Minimum values in each column are bold-faced.

Systems Compared	$D_D$	$\Delta\mu$	$DTW$
A, B	3.38	<b>0.014</b>	<b>294</b>
A, C	<b>0.575</b>	0.032	425
B, C	3.36	0.028	426

There remain a number of open questions not addressed by the results reported here. It is unclear how best to leverage simultaneously the strengths of a general metric like  $D_D$  and the power of specialized tests. It should be possible to systematically bootstrap model identification by using  $D_D$  to rapidly and robustly identify systems or models with similar dynamics and then deploying more powerful but limited specialized methods for determining, e.g., the effective order needed. Additionally, the methods described in Section 3.4.4 for working with passively obtained data (in circumstances where intervening on the system or controlling boundary conditions is impossible) rely on metaparameters that must be tuned. How to do this efficiently, both in terms of computational complexity and the amount of data, remains to be determined. There are also a variety of questions concerning stochastic dynamical systems. It is known that the type of noise (additive or multiplicative) in a stochastic differential equation can dramatically shape gross features such as the onset of chaos and the types of bifurcations as well as the detailed evolution of probability densities over time. The dynamical symmetries that underwrite the distance  $D_D$  are shaped by these time evolutions. However, the relative sensitivity of  $D_D$  to the two types of noise is unknown, as is the ability to distinguish such dynamical noise, i.e., stochastic elements in the generating process of a time series, from noise in the measurement of a system. A likely candidate for investigating these questions in simulation involves numerically extracting the time evolution of probability density functions for a given system of stochastic differential equations using the Fokker–Planck equation [6]. However, it remains uncertain whether these methods would suggest a means of using  $D_D$  to distinguish the underlying dynamical differences between systems with different types of dynamical noise from empirically given time series.

Despite the aforementioned unknowns, the proposed metric is immediately useful for a range of applications. As demonstrated in Figure 3, the dynamical distance  $D_D$  can be used to systematically identify regions of the parameter space of a family of models with a common dynamical character, sometimes with surprising discontinuities. For instance, there are a scattering of points in the region  $0.2 < \alpha < 0.4, 0.2 < \beta < 0.6$  that are similar in dynamical behavior to versions of the four-species Lotka–Volterra model with radically different parameterizations ( $\alpha \approx 1$ ). Similarly,  $D_D$  can be used with empirical time series data to cluster physical systems with unknown underlying causal structure into classes likely to be described by similar dynamical models. This facilitates development of such models. Additionally,  $D_D$  can be used to validate models—especially stochastic models—when the ground truth is unknown and only time series from a target system are available, similar to the related approach of [69]. Insofar as time series produced with the model diverge in  $D_D$  from time series measured from the target system, one can determine

whether a model correctly captures the causal structure regardless of how well the model is able to mimic any particular trajectory. Finally, perhaps the most promising application of the dynamical distance metric is detecting structural change. We demonstrated using fully and partially observed Kuramoto phase oscillator systems (section 4.7) that  $D_D$  can identify points in a time series at which the underlying generative dynamics changes. The dynamical distance is uniquely apt for detecting a genuine change in the underlying causal structure rather than a shift to a previously unobserved portion of the system's phase space. This is precisely what is needed to identify, e.g., a change in ecosystem dynamics portending collapse or recovery, or a shift in the behavior of a volcano system suggesting worrisome internal structural changes. This list is only suggestive; there are myriad uses for a general metric of dynamical similarity that is tied to underlying causal structure, can be inferred from passively acquired time series observations, is robust under measurement noise, and is applicable to state-determined and stochastic dynamics alike.

**Author Contributions:** B.J. conceived this study. All authors contributed to algorithm development. B.J. and C.S.-B. analyzed the data. B.J. wrote the main manuscript. All authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Science Foundation grant numbers SES-1454190 and CMMI-1708622.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** A Python package implementing all of the methods reported here is available under an MIT license from <https://github.com/jantzen/eugene> (accessed on 6 September 2021). Scripts for conducting all tests reported in the text and generating the figures presented can be obtained from [https://github.com/jantzen/distance\\_metric](https://github.com/jantzen/distance_metric) (accessed on 6 September 2021). Because the numerical experiments reported involve a stochastic element, repeating the experiments will not give exactly the results reported. To recreate the exact analyses reported here, the specific data we used can be obtained from <https://doi.org/10.7294/16587179> (accessed on 6 September 2021).

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Basseville, M.; Nikiforov, I.V. *Detection of Abrupt Changes: Theory and Application*; Prentice Hall: Hoboken, NJ, USA, 1993; Volume 104.
2. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367, doi:10.1007/s10115-016-0987-z.
3. Yamanishi, K.; Takeuchi, J. A Unifying Framework for Detecting Outliers and Change Points from Non-stationary Time Series Data. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02, Edmonton, AB, Canada, 23–26 July 2002; ACM: New York, NY, USA, 2002; pp. 676–681, doi:10.1145/775047.775148.
4. Livina, V.N.; Lenton, T.M. A modified method for detecting incipient bifurcations in a dynamical system. *Geophys. Res. Lett.* **2007**, *34*, L03712, doi:10.1029/2006GL028672.
5. Da, C.; Li, F.; Shen, B.; Yan, P.; Song, J.; Ma, D. Detection of a sudden change of the field time series based on the Lorenz system. *PLoS ONE* **2017**, *12*, e0170720, doi:10.1371/journal.pone.0170720.
6. Mirzakhali, E.; Epureanu, B.I. Probabilistic Analysis of Bifurcations in Stochastic Nonlinear Dynamical Systems. *J. Comput. Nonlinear Dyn.* **2019**, *14*, doi:10.1115/1.4043669.
7. Hively, L.M.; Protopopescu, V.A.; Gailey, P.C. Timely detection of dynamical change in scalp EEG signals. *Chaos Interdiscip. J. Nonlinear Sci.* **2000**, *10*, 864–875, doi:10.1063/1.1312369.
8. Basseville, M. Detecting changes in signals and systems—A survey. *Automatica* **1988**, *24*, 309–326, doi:10.1016/0005-1098(88)90073-8.
9. Norton, J.P. *An Introduction to Identification*; Academic Press: London, UK; New York, NY, USA, 1986.
10. Berryman, A.A. On Choosing Models for Describing and Analyzing Ecological Time Series. *Ecology* **1992**, *73*, 694, doi:10.2307/1940776.
11. McSharry, P.E.; Smith, L.A. Better Nonlinear Models from Noisy Data: Attractors with Maximum Likelihood. *Phys. Rev. Lett.* **1999**, *83*, 4285–4288, doi:10.1103/PhysRevLett.83.4285.



12. Sitz, A.; Schwarz, U.; Kurths, J.; Voss, H.U. Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Phys. Rev. E* **2002**, *66*, 016210, doi:10.1103/PhysRevE.66.016210.
13. Aguirre, L.A.; Billings, S.A. Identification of models for chaotic systems from noisy data: Implications for performance and nonlinear filtering. *Phys. D Nonlinear Phenom.* **1995**, *85*, 239–258, doi:10.1016/0167-2789(95)00116-L.
14. Chen, G.; Dong, X. From chaos to order—Perspectives and methodologies in controlling chaotic nonlinear dynamical systems. *Int. J. Bifurc. Chaos* **1993**, *03*, 1363–1409, doi:10.1142/S0218127493001112.
15. Sokal, R.R.; Rohlf, F.J. *Biometry: The Principles and Practices of Statistics in Biological Research*, 3rd ed.; W. H. Freeman: New York, NY, USA, 1994.
16. McCarthy, M.A.; Broome, L.S. A method for validating stochastic models of population viability: A case study of the mountain pygmy-possum (*Burrhamys parvus*). *J. Anim. Ecol.* **2000**, *69*, 599–607, doi:10.1046/j.1365-2656.2000.00415.x.
17. Bradley, E.; Kantz, H. Nonlinear time-series analysis revisited. *Chaos Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 097610, doi:10.1063/1.4917289.
18. Vaswani, N. Change detection in partially observed nonlinear dynamic systems with unknown change parameters. In Proceedings of the 2004 American Control Conference, Boston, MA, USA, 30 June–2 July 2004; Volume 6, pp. 5387–5393, doi:10.23919/ACC.2004.1384709.
19. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Doynne Farmer, J. Testing for nonlinearity in time series: The method of surrogate data. *Phys. D Nonlinear Phenom.* **1992**, *58*, 77–94, doi:10.1016/0167-2789(92)90102-S.
20. Paluš, M. Testing for nonlinearity using redundancies: Quantitative and qualitative aspects. *Phys. D Nonlinear Phenom.* **1995**, *80*, 186–205.
21. Paluš, M. Detecting nonlinearity in multivariate time series. *Phys. Lett. A* **1996**, *213*, 138–147.
22. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461.
23. Butail, S.; Mwaffo, V.; Porfiri, M. Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Phys. Rev. E* **2016**, *93*, 042411.
24. Nichols, J.; Seaver, M.; Trickey, S.; Todd, M.; Olson, C.; Overbey, L. Detecting nonlinearity in structural systems using the transfer entropy. *Phys. Rev. E* **2005**, *72*, 046217.
25. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102.
26. Zunino, L.; Kulp, C.W. Detecting nonlinearity in short and noisy time series using the permutation entropy. *Phys. Lett. A* **2017**, *381*, 3627–3635.
27. Kulp, C.W.; Zunino, L.; Osborne, T.; Zawadzki, B. Using missing ordinal patterns to detect nonlinearity in time series data. *Phys. Rev. E* **2017**, *96*, 022218.
28. Barahona, M.; Poon, C.S. Detection of nonlinear dynamics in short, noisy time series. *Nature* **1996**, *381*, 215–217, doi:10.1038/381215a0.
29. Small, M.; Tse, C. Detecting determinism in time series: The method of surrogate data. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2003**, *50*, 663–672, doi:10.1109/TCSI.2003.811020.
30. Wolf, A.; Swift, J.B.; Swinney, H.L.; Vastano, J.A. Determining Lyapunov exponents from a time series. *Phys. D Nonlinear Phenom.* **1985**, *16*, 285–317, doi:10.1016/0167-2789(85)90011-9.
31. Eckmann, J.P.; Kamphorst, S.O.; Ruelle, D.; Ciliberto, S. Liapunov exponents from time series. *Phys. Rev. A* **1986**, *34*, 4971–4979, doi:10.1103/PhysRevA.34.4971.
32. Grassberger, P.; Procaccia, I. Measuring the strangeness of strange attractors. *Phys. D Nonlinear Phenom.* **1983**, *9*, 189–208, doi:10.1016/0167-2789(83)90298-1.
33. Sugihara, G.; May, R.M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **1990**, *344*, 734–741, doi:10.1038/344734a0.
34. Yunfan, G.; Jianxue, X.; Wei, R.; Sanjue, H.; Fuzhou, W. Determining the degree of chaos from analysis of ISI time series in the nervous system: A comparison between correlation dimension and nonlinear forecasting methods. *Biol. Cybern.* **1998**, *78*, 159–165, doi:10.1007/s004220050422.
35. Kaplan, D.T.; Glass, L. Direct test for determinism in a time series. *Phys. Rev. Lett.* **1992**, *68*, 427–430, doi:10.1103/PhysRevLett.68.427.
36. Kodba, S.; Perc, M.; Marhl, M. Detecting chaos from a time series. *Eur. J. Phys.* **2004**, *26*, 205–215, doi:10.1088/0143-0807/26/1/021.
37. Grassberger, P.; Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593, doi:10.1103/PhysRevA.28.2591.
38. Poon, C.S.; Barahona, M. Titration of chaos with added noise. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7107–7112, doi:10.1073/pnas.131173198.
39. Gottwald, G.A.; Melbourne, I. A new test for chaos in deterministic systems. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2004**, *460*, 603–611, doi:10.1098/rspa.2003.1183.
40. Kulp, C.W.; Zunino, L. Discriminating chaotic and stochastic dynamics through the permutation spectrum test. *Chaos Interdiscip. J. Nonlinear Sci.* **2014**, *24*, 033116, doi:10.1063/1.4891179.
41. Yang, Z.; Zhao, G. Application of symbolic techniques in detecting determinism in time series [and EMG signal]. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Volume 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286), Hong Kong, China, 1 November 1998; Volume 5, pp. 2670–2673, doi:10.1109/IEMBS.1998.745176.

42. Nelles, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*; Springer-Verlag: New York, NY, USA, 2001.
43. Hong, X.; Mitchell, R.J.; Chen, S.; Harris, C.J.; Li, K.; Irwin, G.W. Model selection approaches for non-linear system identification: A review. *Int. J. Syst. Sci.* **2008**, *39*, 925–946.
44. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 111–133, doi:10.1111/j.2517-6161.1974.tb00994.x.
45. Orr, M.J.L. Regularization in the Selection of Radial Basis Function Centers. *Neural Comput.* **1995**, *7*, 606–623, doi:10.1162/neco.1995.7.3.606.
46. Hong, X.; Harris, C.J. Neurofuzzy design and model construction of nonlinear dynamical processes from data. *IEE Proc.-Control Theory Appl.* **2001**, *148*, 530–538.
47. Hong, X.; Harris, C.J. Experimental design and model construction algorithms for radial basis function networks. *Int. J. Syst. Sci.* **2003**, *34*, 733–745, doi:10.1080/00207720310001640223.
48. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000; Volume 2.
49. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288.
50. Kennel, M.B.; Brown, R.; Abarbanel, H.D.I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **1992**, *45*, 3403–3411, doi:10.1103/PhysRevA.45.3403.
51. Rhodes, C.; Morari, M. Determining the model order of nonlinear input/output systems directly from data. In Proceedings of the 1995 American Control Conference—ACC'95, Seattle, WA, USA, 21–23 June 1995; Volume 3, pp. 2190–2194, doi:10.1109/ACC.1995.531288.
52. Hodge, V.; Austin, J. A Survey of Outlier Detection Methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126, doi:10.1023/B:AIRE.0000045502.10941.a9.
53. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–58.
54. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *arXiv* **2019**, arXiv:1801.00718.
55. Lhermitte, S.; Verbesselt, J.; Verstraeten, W.W.; Coppin, P. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sens. Environ.* **2011**, *115*, 3129–3152.
56. Wang, X.; Mueen, A.; Ding, H.; Trajcevski, G.; Scheuermann, P.; Keogh, E. Experimental comparison of representation methods and distance measures for time series data. *Data Min. Knowl. Discov.* **2013**, *26*, 275–309, doi:10.1007/s10618-012-0250-5.
57. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv.* **1999**, *31*, 264–323, doi:10.1145/331499.331504.
58. Liao, T.W. Clustering of time series data—A survey. *Pattern Recognit.* **2005**, *38*, 1857–1874, doi:10.1016/j.patcog.2005.01.025.
59. Shaw, C.T.; King, G.P. Using cluster analysis to classify time series. *Phys. D Nonlinear Phenom.* **1992**, *58*, 288–298, doi:10.1016/0167-2789(92)90117-6.
60. Faloutsos, C.; Ranganathan, M.; Manolopoulos, Y. Fast subsequence matching in time-series databases. *ACM SIGMOD Rec.* **1994**, *23*, 419–429, doi:10.1145/191843.191925.
61. Evans, J.; Geerken, R. Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure. *Remote Sens. Environ.* **2006**, *105*, 1–8.
62. Madrid, F.; Imani, S.; Mercer, R.; Zimmerman, Z.; Shakibay, N.; Keogh, E. Matrix Profile XX: Finding and Visualizing Time Series Motifs of All Lengths using the Matrix Profile. In Proceedings of the 2019 IEEE International Conference on Big Knowledge (ICBK), Beijing, China, 10–11 November 2019; pp. 175–182, doi:10.1109/ICBK.2019.00031.
63. Scheffer, M.; Bascompte, J.; Brock, W.A.; Brovkin, V.; Carpenter, S.R.; Dakos, V.; Held, H.; Nes, E.H.v.; Rietkerk, M.; Sugihara, G. Early-warning signals for critical transitions. *Nature* **2009**, *461*, 53–59, doi:10.1038/nature08227.
64. Ives, A.R.; Dakos, V. Detecting dynamical changes in nonlinear time series using locally linear state-space models. *Ecosphere* **2012**, *3*, art58, doi:10.1890/ES11-00347.1.
65. Tykierko, M. Using invariants to change detection in dynamical system with chaos. *Phys. D Nonlinear Phenom.* **2008**, *237*, 6–13, doi:10.1016/j.physd.2007.07.004.
66. Rao, C.; Ray, A.; Sarkar, S.; Yasar, M. Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal Image Video Process.* **2009**, *3*, 101–114, doi:10.1007/s11760-008-0061-8.
67. Bondarenko, V.E. Self-Organization Processes in Chaotic Neural Networks Under External Periodic Force. *Int. J. Bifurc. Chaos* **1997**, *07*, 1887–1895, doi:10.1142/S0218127497001461.
68. Jantzen, B.C. Projection, symmetry, and natural kinds. *Synthese* **2014**, *192*, 3617–3646, doi:10.1007/s11229-014-0637-5.
69. Jantzen, B.C. Dynamical Symmetries and Model Validation. In *Algorithms and Complexity in Mathematics, Epistemology, and Science*; Fields Institute Communications; Springer: New York, NY, USA, 2019; pp. 153–176.
70. Roy, S.; Jantzen, B. Detecting causality using symmetry transformations. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 075305, doi:10.1063/1.5018101.
71. Pearl, J. *Causality*; Cambridge University Press: New York, NY, USA, 2009.
72. Eberhardt, F.; Scheines, R. Interventions and Causal Inference. *Philos. Sci.* **2007**, *74*, 981–995, doi:10.1086/525638.
73. Jantzen, B.C. Dynamical Kinds and their Discovery. *arXiv* **2017**, arXiv:1612.04933.
74. Arnold, L. Random dynamical systems. In *Dynamical Systems*; Springer: Berlin, Germany, 1995; pp. 1–43.
75. Székely, G.J. E-Statistics: The energy of statistical samples. *Bowling Green State University, Department of Mathematics and Statistics Technical Report.* **2003**, *3*, 1–18.

76. Székely, G.J.; Rizzo, M.L. Energy statistics: A class of statistics based on distances. *J. Stat. Plan. Inference* **2013**, *143*, 1249–1272.
77. Rizzo, M.L.; Székely, G.J. Energy distance. *Wiley Interdiscip. Rev. Comput. Stat.* **2016**, *8*, 27–38, doi:10.1002/wics.1375.
78. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 43–49, doi:10.1109/TASSP.1978.1163055.
79. Volterra, V. Fluctuations in the Abundance of a Species considered Mathematically 1. *Nature* **1926**, *118*, 558–560, doi:10.1038/118558a0.
80. Roques, L.; Chekroun, M.D. Probing chaos and biodiversity in a simple competition model. *Ecol. Complex.* **2011**, *8*, 98–104.
81. Liu, M.; Fan, M. Permanence of stochastic Lotka–Volterra systems. *J. Nonlinear Sci.* **2017**, *27*, 425–452.
82. Kuramoto, Y. International symposium on mathematical problems in theoretical physics. *Lect. Notes Phys.* **1975**, *30*, 420.
83. Law, S.M. STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining. *J. Open Source Softw.* **2019**, *4*, 1504, doi:10.21105/joss.01504.
84. Strogatz, S.H. From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Phys. D Nonlinear Phenom.* **2000**, *143*, 1–20.
85. Giorgino, T. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *J. Stat. Softw. Artic.* **2009**, *31*, 1–24, doi:10.18637/jss.v031.i07.