Generalized Theorems for Nonlinear State Space Reconstruction

Ethan R. Deyle, George Sugihara*

Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, United States of America

Abstract

Takens' theorem (1981) shows how lagged variables of a single time series can be used as proxy variables to reconstruct an attractor for an underlying dynamic process. State space reconstruction (SSR) from single time series has been a powerful approach for the analysis of the complex, non-linear systems that appear ubiquitous in the natural and human world. The main shortcoming of these methods is the phenomenological nature of attractor reconstructions. Moreover, applied studies show that these single time series reconstructions can often be improved *ad hoc* by including multiple dynamically coupled time series in the reconstructions, to provide a more mechanistic model. Here we provide three analytical proofs that add to the growing literature to generalize Takens' work and that demonstrate how multiple time series can be used in attractor reconstructions. These expanded results (Takens' theorem is a special case) apply to a wide variety of natural systems having parallel time series observations for variables believed to be related to the same dynamic manifold. The potential information leverage provided by multiple embeddings created from different combinations of variables (and their lags) can pave the way for new applied techniques to exploit the time-limited, but parallel observations of natural systems, such as coupled ecological systems, geophysical systems, and financial systems. This paper aims to justify and help open this potential growth area for SSR applications in the natural sciences.

Introduction

A growing realization in many natural sciences is that simple idealized notions of linearly decomposable, fixed equilibrium systems often do not accord with reality. Rather, empirical measurements on ecosystems, metabolic systems, financial networks, and the like suggest a more complex, but potentially more information-rich paradigm at work [1–14]. Despite a long history of linear methods development in the engineering sciences, natural systems are generally not well described as sums of independent frequencies that can be sensibly decomposed, analyzed as non-interacting, and reassembled (e.g. Fourier or spectral analysis) in the style of traditional reductionism [15,16]. Rather, quantitative measurements show many systems to be fundamentally non-equilibrium and unstable, in a manner more consistent with nonlinear (state dependent) dynamics occurring on a strange attractor manifold M, where relationships between state variables cannot be studied independently of the overall system state [17–27]. This emergent comprehensive view may help explain why many natural systems, such a those mentioned above, appear so difficult to understand and predict. Mirage correlations are commonplace in nonlinear systems where the manifold may contain trajectories that can temporarily exhibit positive correlations between variables for surprisingly long time periods (and in some regions of the state space) and can subsequently and rapidly exhibit negative correlations or no relationship in other time periods (and other regions of M). This transient property of

apparent non-stationarity in correlations is one of the confounding phenomena faced by traditional linear models that require continual refitting and exhibit little or no predictive power.

In this paper, we present two general theorems that addresses the problem of characterizing the coupled dynamics of nonlinear systems using time series observations on a manifold M. A special case of this theorem, attributed originally to Takens [12], provided the first sketch of a mathematical proof for reconstructing a diffeomorphic shadow manifold M' using lags of a single time series as coordinate axes. The basic idea, that was earlier demonstrated by Packard, Crutchfield, Farmer, and Shaw [28] and Crutchfield [2], is that under generic conditions, a shadow manifold M' can be created using time-lagged observations of M based on a single observation function (Cartesian coordinate variable) that is a smooth and smoothly invertible 1 : 1 mapping with M. Subsequently, Sauer, Yorke, and Casdagli [29] provided a definitive proof and an explicit extension of Takens' theorem to fractal sets; their theorems are also more powerful than the original theorem, as they show embeddings are not just generic in the sense of being open and dense in the set of all mappings, but in fact almost every mapping in the sense of prevalence [30] is an embedding (see [30] for an in-depth explanation of the advantages of "prevalence" over "generic"). The theorem was also extended by Stark, Broomhead, Davies, and Huke [31,32] and Stark [33] to include certain classes of stochastic systems. Practical methods for reconstruction have also been explored, particularly to address the presence of noise in real data (e.g. [29,34]). Casdagli et al. [35]

give a thorough treatment of such techniques based on transformations of univariate maps, showing how optimal noise reduction can be achieved. These very important prior results all focused on reconstruction from a single time series; however, as proven below, they can be extended to the more practically significant case where multiple observation functions are used to generate M'.

Here we prove the more general case of multivariate embeddings (embeddings using multiple time series and lags thereof), and show how time series information can be leveraged if multiple time series and their lags are used to construct embeddings of M'. These theorems pave the way for more extensive use of state space reconstruction methods in practical applications where long time series may not be available, so that multiple diffeomorphic embeddings may be created in factorial fashion to more fully exploit the coupled non-redundant information that can be extracted from multiple time series (multiple observation functions of dynamics on a manifold) to create predictive shadow manifolds [36]. The use of multiple time series allows the possibility of noise reduction that exceeds the limitations of univariate reconstructions in the presence of noise [35].

The possibility of extending Takens' theorem to allow lags of multiple observation functions was mentioned in Remark 2.9 from [29], but was not explicitly proven. The remark was also restricted to mappings strictly formed from consecutive lags, which is not the only possibility that needs to be considered in the multivariate case. Given the potential importance of multivariate reconstructions, we believe a full proof is required—in particular, one that extends the generalization to non-consecutive lags. We show how Takens' theorem is a special case of our more general Theorem 2 (below) and by following the structure of Takens' original proof we clarify the logic and highlight the restrictions and special cases (non-generic cases) that can arise in its application to real world systems. We then give explicit proof of a stronger version of Remark 2.9 from Sauer et al. [29] that allows non-consecutive lags. This third theorem is stronger than the first two in the sense that it shows embeddings are prevalent and not just generic. For those less familiar, we begin with a brief overview of some basic terms and concepts used in our proofs.

Some Basic Concepts of Embedding Theory

Consider the classic Lorenz attractor [37] shown in Figure 1(a), consisting of trajectories in three-dimensional space that together define a butterfly shaped surface or *manifold*. For simplicity, a manifold can be thought of as a generalized, n-dimensional surface embedded in some higher dimensional space, where the dimension of the manifold may be fractal (as is the case for the Lorenz attractor). More generally, an *embedding* is a multivariate transformation of a manifold that resolves all trajectories on the original manifold without crossings. That is, an embedding is globally $1:1$ in that it resolves all *singularities* in trajectories that define the manifold (singularities are points on the manifold where trajectories cross so that future paths are not uniquely determined).

An *immersion* is a local embedding that may not preserve the global topology of a manifold. Rather an immersion preserves the topology of every local neighborhood of the original manifold, so that each point of the tangent space of the immersed manifold has the same dimensionality as the true manifold. Thus, an immersion is a mapping that is $1:1$ between any given "piece" of the true manifold and the immersed manifold. However, this condition does not guarantee that the global topology is preserved. This is illustrated in Figure 1(c), where two different pieces of the original manifold are mapped to the same piece of the immersed manifold,

producing an immersion that is not an embedding. Immersions are nonetheless a useful conceptual stepping stone for constructing proofs about embeddings, since all embeddings are necessarily immersions.

The Lorenz attractor, Figure 1(a), provides an excellent example to illustrate both of these concepts. Consider two different multivariate functions that transform the original manifold, $\Phi_y = (y(t), y(t-\tau), y(t-2\tau))$ and $\Phi_z = (z(t), z(t-\tau), z(t-2\tau))$ where τ is a small time lag as in Takens' theorem. Both of these functions map points on the true manifold to points on a shadow manifold, shown in Figures 1(b) and 1(c). Examining these shadow manifolds, it is evident that both are immersions of the Lorenz attractor, because zooming in on a particular piece of either will reveal that the tangent spaces have the same dimensionality as the original. However, only Figure 1(b) is an embedding that successfully reproduces the two lobes of the butterfly. The reconstruction in Figure 1(c), based on lags of the z-coordinate, fails to do so, because the two fixed points of the original attractor have the same z-coordinate; they are mapped to the same point on the shadow manifold, so the two lobes are stacked on top of each other. This singularity is a consequence of a special, non-generic symmetry in the Lorenz system that violates an assumption of Takens' theorem. Figure 1(d) shows an embedding based on lags of both y- and z-coordinates and is an example of the generalized mappings addressed in this paper.

Results

Two Theorems in the Style of Takens: The Generic Case

Let M be a compact manifold of dimension m. A dynamical system is a diffeophorism ϕ defining the trajectories or "flow" on M for discrete time or a vector field X on M for continuous time. Takens [12] proved generically that given ϕ and M, a smooth observation function $y : M \rightarrow \mathbb{R}$ can be used to construct an embedding of M in $2m+1$ dimensions under the transformation $\Phi_{(\phi,y)} : M \rightarrow \mathbb{R}^{2m+1}$ where $\Phi_{(\phi,y)}(\mathbf{x}) = \langle y(\mathbf{x}), y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots, y(\phi^{2m}(\mathbf{x})) \rangle$. Here the components $\langle y(\mathbf{x}), y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots, y(\phi^{2m}(\mathbf{x})) \rangle$ correspond to time-lagged observations of the dynamics on M defined by ϕ. Notice that such mappings involve a single distinct observation function (i.e. a single time series), and represent a small subset in the larger set $\mathcal{Y}^{2m+1}$ of all possible mappings $M \rightarrow \mathbb{R}^{2m+1}$ that could, for example, involve multiple time series and their lags.

Takens explicitly refers only to the unlagged y as an observation function, but in its most general sense an observation function is any $y : M \rightarrow \mathbb{R}$. Thus, the functions $y(\phi(\mathbf{x})), y(\phi^2(\mathbf{x})), \ldots$, corresponding to the lags of the time series are technically observation functions as well. This bears mention, because in the more general case of mappings $\Phi : M \rightarrow \mathbb{R}^{2m+1}$, the observation functions making up the components of Φ are not all derived from a single time series, but can be various lags of multiple time-series. To treat these cases, it is necessary to acknowledge that these are all observation functions, and we will refer to distinct time series as "unlagged" observation functions.

For a mapping Φ in the larger set $\mathcal{Y}^{2m+1}$ of all mappings $M \rightarrow \mathbb{R}^{2m+1}$, consider the case with $2m+1$ component functions $y_k : M \rightarrow \mathbb{R}$ which are multiple unlagged observation functions of M (i.e. multiple time series). Again, an observation function is any function $M \rightarrow \mathbb{R}$ that assigns a real number to each point on the manifold M. For a mapping $\Phi : M \rightarrow \mathbb{R}^{2m+1}$, we can think of Φ in terms of its $2m+1$ component functions, which correspond to the coordinates in $\mathbb{R}^{2m+1}$. These component functions may all be lags of a single distinct observation function tracking a dynamical system, as in the case of Takens, or they may be multiple observation functions,
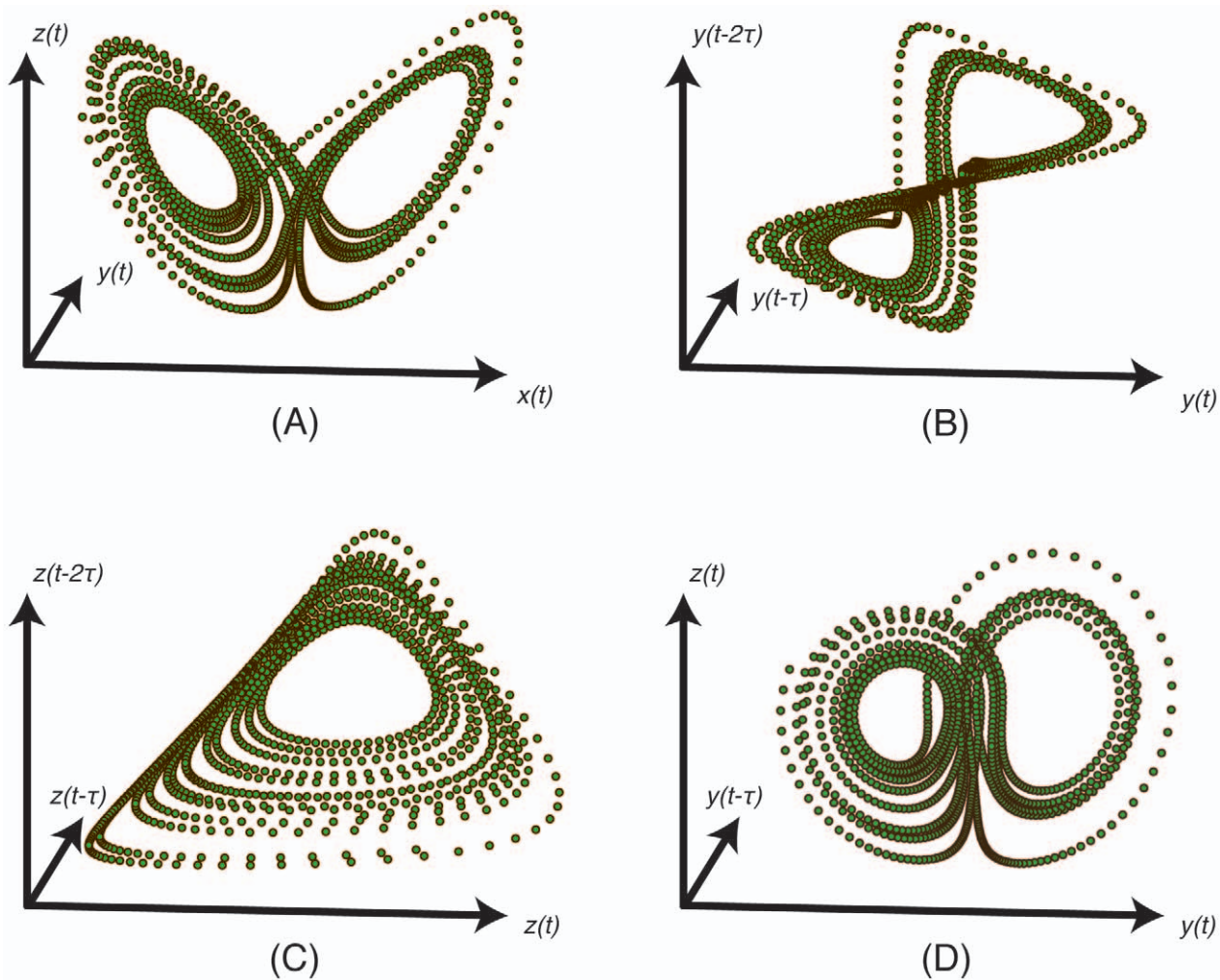
Figure 1. Lorenz attractor with three shadow manifolds. The Lorenz attractor [37] is shown with three shadow manifolds created from lag-coordinate transformations. The typical parameters were used: $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$, giving the three coupled equations as $\dot{x} = 10(y - x)$, $\dot{y} = 28x - xz - y$, and $\dot{z} = xy - (8/3)z$. The solution was computed using a fourth order Runge-Kutta method with a time step of $\delta t = 0.01$, and the time lag used to create the shadow manifolds was $\tau = 8\delta t = 0.08$. (A) The trajectory shown in the x, y, and z coordinates of the original system reveals a two-lobed manifold. (B) A univariate transformation using time lags of the y-coordinate, $\Phi = (y(t), y(t-\tau), y(t-2\tau))$, preserves this two-lobed structure (and other topological properties), verifying Takens' theorem. (C) A univariate transformation using time lags of the z-coordinate, $\Phi = (z(t), z(t-\tau), z(t-2\tau))$, does not preserve the two-lobed structure. Local neighborhoods of the original attractor are, however, preserved. Thus, though this mapping violates a genericity assumption of the original theorem and is not an embedding, it is an immersion of the original manifold. (D) A multivariate transformation using both the y- and z-coordinates, $\Phi = (y(t), y(t-\tau), z(t))$, fulfills the assumptions of Theorems 2 and 7. As predicted, it also preserves the two-lobed structure of the Lorenz and is a valid embedding.
doi:10.1371/journal.pone.0018295.g001

as in the case of Whitney, or they may be lags of multiple observation functions, as in Theorems 2 and 7 below.

The question arises whether general multivariate mappings $\Phi(\mathbf{x}) = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$ form legitimate embeddings. Here we present two theorems: one that demonstrates that maps created from $2m+1$ distinct observation functions are generically embeddings and another that shows that maps created from lags of multiple observation functions are also generically embeddings. Both of these theorems generalize Takens' theorem for which the component functions only involve a single observation function.

It follows from Whitney [38] that generically $\Phi \in \mathcal{Y}^{2m+1}$ is an embedding. Note, however, that Whitney's work does not apply to the specific subsets of $\mathcal{Y}^{2m+1}$ involving fixed lagged relationships as discussed by Takens for reconstructing attractor manifolds M for dynamic systems. That is, Whitney's theorem is generic and does not address these specific subsets of $\mathcal{Y}^{2m+1}$ which have "measure

zero" (e.g. in the sense of "shy" defined in [30]). To tackle this problem, we look to the proof of Takens and see that it can be readily generalized to the other subsets of $\mathcal{Y}^{2m+1}$, including the case of generic $\Phi \in \mathcal{Y}^{2m+1}$.

Recall that, for a compact manifold, a mapping that is an immersion and injective is also necessarily an embedding. Thus, Takens' general approach was to first show that (i) immersions are dense in the set of mappings $\{\Phi_{(\phi,y)}\}$, then that (ii) there is a dense set of $1:1$ mappings within this set of immersions. Since the set of embeddings is open in the set of all possible mappings, Takens concludes that mappings in $\{\Phi_{(\phi,y)}\}$ are generically embeddings. The critical word here is "generically," meaning there can be exceptions (and as explained in [30], the set of such exceptions doesn't necessarily have zero measure).

To demonstrate both (i) and (ii), Takens argues that even when the property of interest (e.g. the $1:1$ property) does not hold for

some particular mapping, by making an arbitrarily small perturbation, it is possible to find a nearby mapping for which that property holds. The key to the theorem and also to adapting it to other sets of mappings is finding how to make these perturbations. The proof is most straightforward for the general case involving $2m+1$ distinct observation functions (each a distinct time series) because it is possible to perturb the component functions of $\Phi_{\langle y_k \rangle}$ independently. Thus we begin with this proof to add clarity to the more powerful main theorem 2 involving lags of multiple observation functions.

Theorem 1. *Consider a compact, m-dimensional manifold M and a set of $2m+1$ observation functions $\langle y_1, \ldots, y_{2m+1} \rangle$, where $y_k : M \to \mathbb{R}$ smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Then it is a generic property of all possible $\langle y_k \rangle$ that the mapping $\Phi_{\langle y_k \rangle} : M \to \mathbb{R}^{2m+1}$ defined as*

$$\Phi_{\langle y_k \rangle} = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. Consider an arbitrary set of $2m+1$ observation functions $\langle \bar{y}_k \rangle$ on M. We define a corresponding mapping $\Phi_{\langle \bar{y}_k \rangle} \in \mathcal{Y}^{2m+1}$ by letting each of these $2m+1$ observation functions be one of the component functions of $\Phi_{\langle \bar{y}_k \rangle}$. Now, recall that an immersion is a map with a derivative that is globally injective, i.e. $1:1$. We denote the total derivative of a function f as Df. If the derivative is evaluated at a particular point $\mathbf{x}$ in the domain of f, we will write $(Df)_\mathbf{x}$, and if Df is a matrix, then we denote the derivative at a particular point and along a particular tangent vector $\mathbf{v}$ as $(Df)_\mathbf{x}(\mathbf{v})$.

For any point $\mathbf{x} \in M$, we can perturb the co-vectors $(D\bar{y}_k)_\mathbf{x} \in \mathcal{T}^*(M)$ independently by perturbing individual $\bar{y}_k$. By making infinitesimal perturbations at points $\mathbf{x} \in M$ for which rank $(D\Phi_{\langle \bar{y}_k \rangle})_\mathbf{x} < m$, we can get a set of observables $\langle \bar{\bar{y}}_k \rangle$ arbitrarily close to $\langle \bar{y}_k \rangle$ such that rank $(D\Phi_{\langle \bar{\bar{y}}_k \rangle})_\mathbf{x} = m$ for all $\mathbf{x} \in M$—i.e., $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ is an immersion. Since the set of immersions is open in the set of all mappings, there is a neighborhood $\mathcal{U} \subset \mathcal{Y}^{2m+1}$ around this $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ such that every $\Phi_{\langle y_k \rangle} \in \mathcal{U}$ is an immersion.

Since immersions are local embeddings, we can find a $\delta > 0$ such that on the manifold, $0 < \rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies $\Phi_{\langle \bar{y}_k \rangle}(\mathbf{x}) \neq \Phi_{\langle \bar{y}_k \rangle}(\mathbf{x}')$. Here we depart from Takens' notation and let δ denote infinitesimal separations between two points on the manifold M to avoid confusion with the later defined ε which is used to perturb the observable; ρ is any fixed metric on M. In fact for this fixed δ, there is a subset $\mathcal{U}' \subset \mathcal{U}$ such that for any $\langle y_k \rangle$ in $\mathcal{U}'$, the associated map $\Phi_{\langle y_k \rangle}$ is an immersion, and $\rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies that $\Phi_{\langle y_k \rangle}(\mathbf{x}) \neq \Phi_{\langle y_k \rangle}(\mathbf{x}')$.

Next, we show that we can find a globally $1:1$ $\Phi_{\langle y_k \rangle} \in \mathcal{U}'$ arbitrarily close to $\Phi_{\langle \bar{y}_k \rangle}$. To do this, we construct a finite collection of subsets $\{U_i\}_{i=1}^N$ such that the U_i are open subsets of M, the collection covers M, and diameter $(U_i) < \delta$ for every i. Then, we take a partition of unity $\{\lambda_i\}$ corresponding to these U_i, so that we can vary the value of any $\bar{\bar{y}}_k$ by an infinitesimal amount $\bar{\bar{y}}_k \to \bar{\bar{y}}_k + \varepsilon_{ki} \lambda_i$ without altering the value of $\Phi_{\bar{\bar{y}}_k}(\mathbf{x})$ for $\mathbf{x} \notin U_i$.

We now consider the mapping $\Psi : M \times M \to \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ defined as $\Psi(\mathbf{x}, \mathbf{x}') = (\Phi_{\bar{\bar{y}}_k}(\mathbf{x}), \Phi_{\bar{\bar{y}}_k}(\mathbf{x}'))$. We define the set $W \subset M \times M$ as $W = \{(\mathbf{x}, \mathbf{x}') \in M \times M | \rho(\mathbf{x}, \mathbf{x}') \geq \delta\}$, so that (by our choice of δ), the mapping $\Phi_{\bar{\bar{y}}_k}$ is necessarily injective on the complement of W in $M \times M$. Furthermore, note that the intersection of $\Psi(W)$ with the diagonal of $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ gives the set of points $\{(\mathbf{x}, \mathbf{x}') \in M \times M | \Phi_{\bar{\bar{y}}_k}(\mathbf{x}) = \Phi_{\bar{\bar{y}}_k}(\mathbf{x}')\}$, and therefore $\Psi(W) \cap \Delta = \emptyset$ is equivalent to $\Phi_{\langle y_k \rangle}$ injective. Our task, then, is to perturb the manifold $\Psi(W)$ using the ε_{ki} and $\varepsilon_{ki'}$ so that it does not intersect the diagonal manifold Δ.

At each $\mathbf{p} \in \Psi(W) \cap \Delta$ we know that $\rho(\mathbf{x}, \mathbf{x}') > \delta$, so $\mathbf{x}$ and $\mathbf{x}'$ cannot belong to the same U_i. Consequently, varying an ε_{ki} or $\varepsilon_{ki'}$ only alters the value of $\Phi_{\langle \bar{\bar{y}}_k \rangle}$ at either $\mathbf{x}$ or $\mathbf{x}'$ (respectively). In the tangent space $\mathcal{T}_\mathbf{p}^*(\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1})$, then, the direction of the $(2m+1) + (2m+1)$ infinitesimal changes given by the ε_{ki} and $\varepsilon_{ki'}$ are all linearly independent (indeed orthogonal) and as such span $\mathcal{T}_\mathbf{p}^*(\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1})$. Since the tangent spaces of $\Psi(W)$ and Δ are at most $2m$ and $2m+1$ dimensional, respectively, we can construct a vector from a linear combination of $\left(\frac{\partial \Psi}{\partial \varepsilon_{ki}} \right)_\mathbf{p}$ and $\left(\frac{\partial \Psi}{\partial \varepsilon_{ki'}} \right)_\mathbf{p}$ that lies outside of both $\mathcal{T}_\mathbf{p}^*(\Psi(W))$ and $\mathcal{T}_\mathbf{p}^*(\Delta)$. Therefore, an infinitesimal perturbation corresponding to this linear combination will move the sub-manifolds $\Psi(W)$ and Δ away from each other at the point $\mathbf{p}$ without creating a new intersection at another point. By keeping the size of these perturbations sufficiently small, we ensure that we stay confined to $\mathcal{U}'$, so that $\Phi_{\langle y_k \rangle}$ is still an immersion. This is a more transparent statement of the transversality argument used in the Takens proof (1981).

Thus, we have shown that for any arbitrary set of $2m+1$ observables $\langle \bar{y} \rangle$, we can find a set of observables $\langle y_k \rangle$ arbitrarily close to $\langle \bar{y}_k \rangle$ such that $\Phi_{\langle y_k \rangle}$ is an embedding—i.e., there is a dense set of observables $\{\langle y_k \rangle\} \subset \mathcal{Y}^{2m+1}$ such that $\Phi_{\langle y_k \rangle}$ is an embedding. The set of embeddings is open in the set of all mappings, so this set is dense and open, meaning that the embedding property is generic over all mappings.

When mappings are confined to fixed lag relationships, Takens showed it is valid to independently perturb each component of Φ at a given point of the domain by perturbing the unlagged observation function, y, in the other parts of the domain corresponding to neighborhoods of the lagged states $\phi^{-1}(\mathbf{x})$, $\phi^{-2}(\mathbf{x})$, etc. This ensures that the perturbations to Φ maintain the structure of the lag relationships and that we have not inadvertently left the subset of interest. As we now show, this allows the above result to be easily extended to families of maps having component functions that are the lags of multiple observation functions. This is the relevant case for many practical examples where lags of multiple time series (multiple variables or observation functions) are required to achieve a mechanistic reconstruction of M (e.g. [20]). It also allows information on M to be leveraged when the time series are short, as is the case in many physical and biological problems [22,36].

Before starting the proof, however, we must clarify exactly what the "subsets of interest" are. We define these sets as follows. First, we say y_q is a lag of the observable y if we can write $y_q = \phi^b(y)$ for positive b. We consider the lags in the positive time direction only to simplify notation in the proof, noting that the results apply equally to negative lags. Let $\mathbf{r} = \{r_1, r_2, \ldots\}$ be the subset of $k = 1, \ldots, 2m+1$ for which $y_r, r \in \mathbf{r}$ is an unlagged observable, i.e. y_r is not a lag of another $y \in \langle y_k \rangle$. We begin with the "unlagged" observation functions, y_r, or observation functions that are not a lag of another observable in $\langle y_k \rangle$. Now define a set C_r for each $r \in \mathbf{r}$ that contains y_r and any other observation function in $\langle y_k \rangle$ which is a lag of it. That is, C_r is the set of $y_q \in \langle y_k \rangle$ that are lags of y_r given as $y_q = \phi^{b_q}(y_r)$, where the lags b_q are distinct for fixed r. This choice of $\mathbf{C} = \{C_r : r \in \mathbf{r}\}$ and $\mathbf{b} = \{b_k : k = 1, \ldots, 2m+1\}$ determine a subset $\mathcal{Y}_{\mathbf{C},\mathbf{b}}^{2m+1} \subset \mathcal{Y}^{2m+1}$ containing all choices of $2m+1$ observables $\langle y_k \rangle$ which obey the correct lag relationships under a dynamical system ϕ. Note that each element of $\mathcal{Y}_{\mathbf{C},\mathbf{b}}^{2m+1}$ can be identified by the dynamical system and the y_r. We denote such an element, then, as $(\phi, \langle y_r \rangle)$.

Theorem 2. *Consider a diffeomorphism $\phi : M \to M$ on some compact manifold M of dimension m, along with $2m+1$ observation functions $y_k : M \to \mathbb{R}$, smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Restrict the y_k to have the lag relationships corresponding to a collection of sets $\mathbf{C}$ and lags $\mathbf{b}$*

under the dynamical system ϕ, and impose the following generic [12,39] properties on ϕ:

1. *The set A of periodic points with period $p < max(b_k)$ has finitely many points,*

2. *The eigenvalues of $(D\phi^b)_\mathbf{x}$ at each $\mathbf{x}$ in a compact neighborhood A are distinct and not equal to 1.*

Then, for generic $\langle y_k \rangle \in \tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$, the mapping described by

$$\Phi_{(\phi,\langle y_r \rangle)} = (y_1(\mathbf{x}), y_2(\mathbf{x}), \ldots, y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. The proof of this theorem closely follows the logic of the previous proof and the original argument of Takens [12]. As noted above, any perturbations to Φ via its component functions $\langle y_k \rangle$ must remain within $\tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$ (the set of observables having the desired lag relationships under ϕ prescribed by the C_r and the b_q). Here we must also deal with points of M that are fixed points or periodic under the dynamical system ϕ, i.e. the points for which there exists a b such that $\phi^b(\mathbf{x}) = \mathbf{x}$ (including the fixed point case, $b = 1$). The above proof shows that the mapping $\Phi_{\langle y_k \rangle}$ is generically an immersion because the co-vectors $(D\bar{y}_k)_\mathbf{x} \in \mathcal{T}^*(M)$ can be independently perturbed. This is also true for non-periodic points where there are fixed lag relationships between some observables, as we can perturb y_r in the neighborhood of $\phi^{-b_q}(\mathbf{x})$ and thus perturb $y_q = y_r(\phi^{b_q}(\mathbf{x}))$ without affecting y_r in the neighborhood of $\mathbf{x}$.

Note that periodic points $\mathbf{x}$ can exist such that the period b or some integer multiple of it, $n \cdot b$, is the fixed time lag between two observables $y_{q_1}, y_{q_2} \in \langle y_k \rangle$ belonging to the same C_r. Let $V \subset M$ be a compact neighborhood of all such points. For $\mathbf{x} \in V$, the vectors $(Dy_{q_1})_\mathbf{x}$ and $(D(y_{q_1} \circ \phi^{n \cdot b}))_\mathbf{x}$ cannot necessarily be perturbed independently. Nonetheless, while $y_{q_1}(\mathbf{x}) = y_{q_1}(\phi^{n \cdot b}(\mathbf{x}))$ for such a point, it is not generally true that $(Dy_{q_1})_\mathbf{x} = (D(y_{q_1} \circ \phi^{n \cdot b}))_\mathbf{x}$. By assumption, for each $\mathbf{x} \in V$, the eigenvalues of the $(D\phi^b)_\mathbf{x}$ are distinct and not equal to 1. Thus, by the chain rule, it is clear that $(Dy_{q_1})_\mathbf{x}$ and $(D(y_{q_1} \circ \phi^{n \cdot b}))_\mathbf{x}$ are linearly independent. As noted above, all the other $(Dy_k)_\mathbf{x}$ can be perturbed independently, so we can find a set of observables $\langle \bar{y}_k \rangle$ arbitrarily near $\langle y_k \rangle$ in $\tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$ for which $\Phi_{(\phi,\langle \bar{y}_r \rangle)}$ is an immersion on V. Note that because the set of immersions is open, there is an open neighborhood in $\tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$ around this $\langle \bar{y}_k \rangle$ for which every set of observables in that neighborhood gives an immersion.

We must also satisfy $\Phi_{(\phi,\langle y_r \rangle)}$ injective. The proof above relied on the ability to independently perturb the manifold $\Psi(W) \subset \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ at any point $(\mathbf{x}, \mathbf{x}')$ by an infinitesimal amount in any coordinate direction. For a periodic point on M with period b and two observables related as y_q and $y_q \circ \phi^{n \cdot b}$, it is impossible to independently perturb $\Psi(W)$ locally in the coordinate $y_q(\mathbf{x})$ or $y_q(\mathbf{x}')$, as you also perturb $y_q(\mathbf{x}) \circ \phi^{n \cdot b}$ or $y_q(\mathbf{x}') \circ \phi^{n \cdot b}$. By assumption, the set V has a finite number of elements. For such a generic ϕ and any set $\langle y_k \rangle \in \tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$, any neighborhood of the $\langle y_k \rangle$ will contain a set of observables $\langle \bar{y}_k \rangle$ for which the unlagged observation functions $\langle \bar{y}_r \rangle$ take distinct values at each point in V.

We first perturb the y_r to find an open neighborhood of observables which give immersions when restricted to the set V. We then further perturb the observables to find within this neighborhood a set of observables $\langle \bar{y}_k \rangle$ for which $\Phi_{(\phi,\langle \bar{y}_r \rangle)}|V$ is also injective and therefore an embedding (on $V \subset M$). Since embeddings are dense in the space of all mappings, there is a neighborhood $\mathcal{U} \subset \tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$ such that for all $(\phi,\langle y_r \rangle) \in \mathcal{U}$, the map $\Phi_{(\phi,\langle y_r \rangle)}|V$ is an embedding.

We now show that we can find a $(\phi,\langle \bar{\bar{y}}_r \rangle) \in \mathcal{U}$ such that $\Phi_{(\phi,\langle \bar{\bar{y}}_r \rangle)}$ is an embedding on *all* of M. We first note that at points $\mathbf{x} \in M \backslash V$, the vectors $(D\bar{y}_k)_\mathbf{x} \in \mathcal{T}^*(M)$ can be perturbed independently, so we can find $(\phi,\langle \bar{\bar{y}} \rangle) \in \mathcal{U}$ for which $\Phi_{(\phi,\langle \bar{y}_k \rangle)}$ is an immersion. Because an immersion is a local embedding, there is a δ such that for $\mathbf{x}, \mathbf{x}' \in M$, $0 < \rho(\mathbf{x}, \mathbf{x}') < \delta$ implies that $\Phi_{(\phi,\langle \bar{y}_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle \bar{y}_r \rangle)}(\mathbf{x}')$. Since the set of immersions is open in the set of possible mappings, there is a neighborhood $\mathcal{U}' \subset \mathcal{U}$ such that for any $(\phi,\langle y_r \rangle) \in \mathcal{U}'$, the corresponding mapping $\Phi_{(\phi,\langle y_r \rangle)}$ is an immersion. Thus, for the same δ as above, $0 < \rho(\mathbf{x}, \mathbf{x}') \leq \delta$ implies $\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}')$.

Now we need to show that there is a $(\phi,\langle y_r \rangle) \in \mathcal{U}'$ such that $\Phi_{(\phi,\langle y_r \rangle)}$ is also injective on M. As noted in the first proof, this is equivalent to $\Psi(V) \cap \Delta = \emptyset$ for the mapping $\Psi : M \times M \to \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$ defined as $\Psi(\mathbf{x}, \mathbf{x}') = (\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}), \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}'))$. If $\mathbf{x}$ and $\mathbf{x}'$ are both in V or $\rho(\mathbf{x}, \mathbf{x}') \leq \delta$, we already know that $\Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}) \neq \Phi_{(\phi,\langle y_r \rangle)}(\mathbf{x}')$. Thus we restrict ourselves to the set $W = \{\mathbf{x}, \mathbf{x}' \in M \times M | \phi(\mathbf{x}, \mathbf{x}') > \delta$ and not both $\mathbf{x}, \mathbf{x}' \in \text{int}(V)\}$.

To perturb the manifold $\Psi(W)$ away from Δ at points of intersection, $\mathbf{p} \in \Psi(W) \cap \Delta$, we must be able to find variations for which the tangent vector $\left(\frac{\partial \Psi}{\partial \varepsilon}\right)_\mathbf{p}$ is linearly independent from the $2m$ tangent vectors $\left(\frac{\partial \Psi}{\partial x_i}\right)_\mathbf{p}$ and $\left(\frac{\partial \Psi}{\partial x'_i}\right)_\mathbf{p}$ and lies outside of $\mathcal{T}_\mathbf{p}^*(\Delta)$. In the first proof, it was obvious that each component of Ψ could be perturbed independently. Now we must be more careful. We do this by first creating a collection of N open subsets of M, $\{U_i\}$, with the following properties:

1. The $\{U_i\}$ cover the closure of $M \backslash V$.

2. For each b_q and $i = 1, \ldots, N$, the diameter of $\phi^{-b_q}(U_i)$ is less than δ.

3. For all choices of $i, j \in \{1, \ldots, N\}$, the set U_j intersects with $\phi^{-b_q}(U_i)$ for at most one b_q.

4. For $\mathbf{x}$ and $\mathbf{x}'$ such that $\phi^{-b}(\mathbf{x}) \in M \backslash \cup U_i$ for some $b \in \mathbf{b}$, $\mathbf{x}' \notin V$, and $\rho(\mathbf{x}, \mathbf{x}') > \delta$, no two of $\mathbf{x}, \{\phi^{b_q}(\mathbf{x})\}, \mathbf{x}', \{\phi^b(\mathbf{x}')\}$ belong to the same U_i.

Take a partition of unity $\{\lambda_i\}$ corresponding to this $\{U_i\}$. Because of the way we constructed the $\{U_i\}$, we can vary the value of each $\bar{y}_k$ at any point $\mathbf{x} \in M \backslash V$ by an infinitesimal amount without altering the value of the other $\bar{y}_k$ in the neighborhood of $\mathbf{x}$. We make this explicit as follows. To perturb the y_r, we take $\bar{\bar{y}}_r \to \bar{\bar{y}}_r + \varepsilon_{ri} \lambda_i$ for i corresponding to $\mathbf{x} \in U_i$. To perturb the other y_k ($y_k = y_r \circ \phi^{b_q}$ for some r), we perturb $\bar{\bar{y}}_r \to \bar{\bar{y}}_r + \varepsilon_{ri} \lambda_i$ for i corresponding to $\phi^{-b_q}(\mathbf{x}) \in U_i$. Consider the $2m+1$ perturbations, ε_{ri}, which are independent shifts at $\mathbf{x}$ in distinct y_k. In $\mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1}$, we note that each corresponding tangent vector $\left(\frac{\partial \Psi}{\partial \varepsilon_{ri}}\right)_\mathbf{p}$ lies outside of $\mathcal{T}_\mathbf{p}^*(\Delta)$. Note the $\left(\frac{\partial \Psi}{\partial \varepsilon_{ri}}\right)_\mathbf{p}$ together with any basis of $\mathcal{T}_\mathbf{p}^*(\Delta)$ form a linearly independent set of vectors. Since the dimension of $\text{span}\left(\left(\frac{\partial \Psi}{\partial x_i}\right)_\mathbf{p}, \left(\frac{\partial \Psi}{\partial x_{i'}}\right)_\mathbf{p}\right)$ is at most $2m$, there must be a linear combination of the $\left(\frac{\partial \Psi}{\partial \varepsilon_{ri}}\right)_\mathbf{p}$ that lies outside of both $\mathcal{T}_\mathbf{p}^*(\Psi(W))$ and $\mathcal{T}_\mathbf{p}^*(\Delta)$, which can be used to perturb $\Psi(W)$ away from Δ. By keeping variations in the ε_{ri} sufficiently small, we can find a set of $\langle y_k \rangle$ such that $(\phi,\langle y_r \rangle) \in \mathcal{U}'$ and $\Psi(\mathbf{x}, \mathbf{x}') \cap \Delta = \emptyset$ (where Ψ now corresponds to the $\Phi_{\phi,\langle y_k \rangle}$ map). This pair gives a mapping $\Phi_{(\phi,\langle y_k \rangle)}$ that is both an immersion and injective, and thus is an embedding. Because $\mathcal{U}'$ was an arbitrarily small neighborhood of any point in $\tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$, this means embeddings are

dense in $\tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$, and the set of embeddings is open in the set of mappings. Thus, the map $\Phi_{(\phi,\langle y_r\rangle)}$ given by $(\phi,\langle y_r\rangle) \in \mathcal{Y}_{\mathbf{C},\mathbf{b}}^{2m+1}$ is generically an embedding.

Just as Takens extends the original result for discrete time to dynamical systems in continuous time, we can extend our result as follows:

Corollary 3. *Consider a smooth vector field X on some compact manifold M along with $2m+1$ observables $y_k : M \to \mathbb{R}$, smoothly; by "smooth" we mean at least $\mathbb{C}^2$. Define ϕ_t as the flow on X. Suppose we restrict the y_k to have the lag relationships corresponding to a collection of sets C_r and lags b_q under the discrete dynamical system ϕ_τ, where τ is a constant. We impose the following generic properties on X:*

1. *For points $\mathbf{x}$ such that $X(\mathbf{x})=0$, the eigenvalues of $(D\phi_\tau)_{\mathbf{x}}$ are distinct and not equal to 1.*
2. *No periodic integral curve of X has integer period $\leq 2m+1$.*

Then, for generic $\langle y_r\rangle \in \tilde{\mathcal{Y}}_{\mathbf{C},\mathbf{b}}^{2m+1}$, the mapping described by

$$\Phi_{(\phi_\tau,\langle y_r\rangle)} = (y_1(\mathbf{x}),y_2(\mathbf{x}),\ldots,y_{2m+1}(\mathbf{x}))$$

is an embedding.

Proof. In this case, ϕ_τ is a discrete time dynamical system on M satisfying the conditions imposed in the theorem above, and this corollary follows directly.

A Theorem in the Style of Sauer et al.: The Prevalent Case

We now give an explicit proof of Remark 2.9 from [29] using the framework constructed in their original paper, but we extend the language to cover reconstructions using non-consecutive lags (from multiple time series). The proof uses Lemma 4.1, 4.6, and 4.11 from [29] to show that $1:1$ mappings and immersions are prevalent in the space $\mathcal{Y}_{\mathbf{C},\mathbf{b}}^{2m+1}$, just as Sauer et al. use Lemma 4.6 to prove Theorem 3.3, and Lemmas 4.1 and 4.11 to prove Theorem 3.5. These lemmas are now stated (for the proofs, see their original paper).

Lemma 4. *(Originally part 2 of 4.1) Let n and k be positive integers, $\mathbf{x}_1,\ldots,\mathbf{x}_n$ distinct points in $\mathbb{R}^k$, $u_1,\ldots,u_n$ in $\mathbb{R}$, and $\mathbf{v}_1,\ldots,\mathbf{v}_n$ in $\mathbb{R}^k$. Then there exists a polynomial h in k variables of degree at most n such that for $i=1,\ldots,n$, $\nabla h(\mathbf{x}_i)=\mathbf{v}_i$.*

Lemma 5. *(Originally 4.6) Let A be a compact subset of $\mathbb{R}^k$. Let $\Phi_0,\Phi_1,\ldots,\Phi_t : A \to \mathbb{R}^n$ be Lipschitz maps. For each integer $r \geq 0$, let S_r be the set of pairs $\mathbf{x}_1 \neq \mathbf{x}_2$ in A for which the $n \times t$ matrix*

$$M_{x_1 x_2} = (\Phi_1(\mathbf{x}_1)-\Phi_1(\mathbf{x}_2),\ldots,\Phi_t(\mathbf{x}_1)-\Phi_t(\mathbf{x}_2))$$

has rank r, and let $d_r = $ lower boxdim $(\bar{S}_r)$. Define $\Phi_\alpha = \Phi_0 + \sum_{i=1}^t \alpha_i \Phi_i : A \to \mathbb{R}^n$. If $d_r < r$ for all integers $r \geq 0$, then for $\alpha = (\alpha_1,\ldots,\alpha_t)$ outside a measure zero subset of $\mathbb{R}^t$, the map Φ_α is $1:1$.

Lemma 6. *(Originally 4.11) Let A be a compact subset of a smooth manifold embedding in $\mathbb{R}^k$. Let $\Phi_0,\Phi_1,\ldots,\Phi_t$ be a set of smooth maps from an open neighborhood U of A to $\mathbb{R}^n$. For each positive integer r, let S_r be the subset of the unit tangent bundle $S(A)$ such that the $n \times t$ matrix*

$$((D\Phi_1)_{\mathbf{x}}(\mathbf{v}),\ldots,(D\Phi_t)_{\mathbf{x}}(\mathbf{v}))$$

has rank r, and let $d_r = $ lower boxdim $(\bar{S}_r)$. Define $\Phi_\alpha = \Phi_0 + \sum_{i=1}^t \alpha_i \Phi_i : A \to \mathbb{R}^n$. If $d_r < r$ for all integers $r \geq 0$, then for almost every $\alpha \in \mathbb{R}^t$, the map Φ_α is an immersion on A.

To apply these lemmas, it is necessary to restrict the dimension of the sets of periodic orbits—that is, the sets $A_p = \{\mathbf{x} \in A : \phi^p(\mathbf{x})=\mathbf{x}\}$ for $p < \max(\{b \in \mathbf{b}\})$. For the case of consecutive lags, Sauer et al. state sufficient conditions to be boxdim

$(A_p) < p/2$. A sufficient condition for non-consecutive lags is a bit more complicated. Define the constants $B_{pr} = $ number of $y_q \in C_r$ such that $b_q = m \cdot p + b_{q'}$ for at least one $b_{q'}$ and $m \in \mathbb{N}$. Also, define $B_p = \sum_r B_{pr}$. A sufficient condition on the A_p is $2 \cdot \text{boxdim}(A_p) < n - B_p$.

Theorem 7. *Let ϕ be a diffeomorphism on an open subset U of $\mathbb{R}^m$, and let A be a compact subset of U, $boxdim(A)=d$. Let $\mathbf{C}$ be a collection of sets and $\mathbf{b}$ a set of lag relationships as above, such that $n = \sum_r n_r > 2d$. Assume that for every positive integer $p \leq max(\{b \in \mathbf{b}\})$, the set A_p of periodic points of period p satisfies $2 \cdot boxdim(A_p) < n - B_p$, and that for each point of A_p, the Jacobian $D\phi^p$ has distinct eigenvalues. Then, for almost every set of n observation functions $\{y_k\}$ satisfying the given lag relationships, the map*

$$\Phi_{(\phi,\langle yr\rangle)}(\mathbf{x}) = (y_1(\mathbf{x}),\ldots,y_n(\mathbf{x}))$$

is an embedding on A.

Proof. Without loss of generality, assume we have ordered the components of $\Phi_{(\phi,\langle y_r\rangle)}$ with y_{r_1} and all its lags first, then y_{r_2} and its lags, etc. That is,

$$\Phi_{(\phi,\langle yr\rangle)}(\mathbf{x}) = (y_{r_1}(\mathbf{x}),y_{r_1}^{b_2}(\mathbf{x}),\ldots,y_{r_2}(\mathbf{x}),\ldots).$$

To show prevalence, we find a suitable probe space (see [29]). The infinite dimensional space for the univariate theorem is the observation functions $y : U \to \mathbb{R}$, smoothly. For maps constructed from multiple lags, this becomes the sets of $s_r = \text{size}(\mathbf{r})$ unlagged observation functions. Sauer et al. take the probe space for the univariate theorem to be any set H of polynomials in m variables which include all such polynomials up to degree $2n$. It is now necessary to have a set of polynomials for each of the y_r. Thus, we take the probe space for this theorem to be the Cartesian product of s_r copies of H.

Let $\langle h_1,\ldots,h_t\rangle$ be a basis for H. We want to show that for almost all choices of $s_r \times t$ coefficients $\alpha_{r,t}$, the map $\Phi_{(\phi,\langle \bar{y}_r\rangle)}$ defined by the observation functions $\bar{y}_r = y_r + \sum_{i=1}^t \alpha_{r,i} h_i$ is an embedding. We first demonstrate that almost every $\Phi_{(\phi,\langle \bar{y}_r\rangle)}$ is $1:1$, proceeding as in the proof of Theorem 4.3 in [29].

To sensibly apply Lemma 5, we adopt the following convention: think of $\Phi_{(\phi,\langle \bar{y}_r\rangle)}$ as a perturbation of $\Phi_{(\phi,\langle y_r\rangle)}$, which is the summed effect of perturbations on each y_r separately. For each pair (r,i), $r \in \mathbf{r}$ and $i \in \{1,\ldots,t\}$, there is a map $\Phi_{r,i} : U \to \mathbb{R}^n$ which is $\Phi_{(\phi,\langle \bar{y}_r\rangle)}$ for $\tilde{y}_{r'} = h_i$ if $r = r'$ and 0 otherwise. The components of $\Phi_{r,i}(\mathbf{x})$ are either 0 or of the form $h_i(\phi^{b_q}(\mathbf{x}))$. Consequently, $\Phi_{\phi,\langle \bar{y}_r\rangle} = \Phi_{\phi,\langle y_r\rangle} + \sum_r \sum_{i=1}^t \alpha_{r,i}\Phi_{r,i}(\mathbf{x})$, which matches the structure Lemma 5.

We now check that the rank of the matrix $M_{x_1 x_2}$ satisfies the conditions of Lemma 5 for each pair of distinct $\mathbf{x}_1,\mathbf{x}_2 \in A$. Note that to avoid confusion with the previous section of this paper and Takens' original work, we continue to use row vectors to describe the transformations Φ. However, Sauer et al. [29] prefer column vectors, so it is necessary to use of transposes in several instances. Thus, we have

$$M_{x_1 x_2}^T = \begin{pmatrix} \Phi_{r_1,1}(\mathbf{x}_1)-\Phi_{r_1,1}(\mathbf{x}_2) \\ \vdots \\ \Phi_{r_1,t}(\mathbf{x}_1)-\Phi_{r_1,t}(\mathbf{x}_2) \\ \Phi_{r_2,1}(\mathbf{x}_1)-\Phi_{r_2,1}(\mathbf{x}_2) \\ \vdots \end{pmatrix}.$$

Note that $M_{x_1 x_2}$ is a block diagonal matrix, and so it has rank

equal to the sum of the rank of the blocks. Each of the s_r blocks can be rewritten as the product of two matrices, J_r and H_r, where the entries of H_r are values of a single polynomial h and the entries of J_r are each one of $\{1, 0, -1\}$. Note, there are multiple possible choices for H_r and J_r that give the same $M_{x_1 x_2}$.

Case 1: First consider $\mathbf{x}_1$ and $\mathbf{x}_2$ that do not both lie in a periodic orbit of integer period less than $\max(\mathbf{b})$. We specify H_r so that the first n_r rows, where n_r is the size of the set C_r, correspond to the $h_{r,i}(\mathbf{x}_1), h_{r,i}(\phi^{b_{r+1}}(\mathbf{x}_1)), \ldots, h_{r,i}(\phi^{b_{r+n_r}}(\mathbf{x}_1))$, and the next n_r correspond to the $h_{r,i}(\phi^{b_{qr}}(\mathbf{x}_2))$. H_r is onto, so the rank of $M_{x_1 x_2}$ is just the sum of the ranks of the J_r. For this case, J_r contains a copy of I_{n_r}, and thus will have rank n_r. The entire matrix $M_{x_1 x_2}$ will thus have rank $n = \sum_r n_r$, which satisfies the conditions of Lemma 5.

Case 2: Now consider $\mathbf{x}_1$ and $\mathbf{x}_2$ in separate periodic orbits with periods p_1 and p_2 such that $1 \leq B_{p_1} \leq B_{p_2}$ and $p_1, p_2 < \max(\mathbf{b})$. H_r will have $B_{p_1 r}$ fewer rows corresponding to the $b_{q_1} = m \cdot p_1 + b_{q_2}$ for some $m \in \mathbb{N}$ (there will also be a reduction in the number of rows associated with B_{p_2}). In this case, J_r will still contain the column space of $I_{(n_r - B_{p_1 r})}$ and thus $\mathrm{rank}(J) \geq \sum_r n_r - B_{p_1 r} = n - B_{p_1}$. Again the H_r are onto, and so the rank of $M_{x_1 x_2}$ is the rank of J.

The dimension of the set S of all pairs $\mathbf{x}_1$ and $\mathbf{x}_2$ is $\dim(S) = \dim(A_{p_1}) + \max(\dim(A_{p_2}))$. By the conditions placed on the size of the A_p, we can conclude that $\dim(S) < n - B_{p_1} \leq \dim(M_{x_1 x_2})$, and thus that Lemma 5 applies to this case as well.

Case 3: Finally we consider $\mathbf{x}_1$ and $\mathbf{x}_2$ in the same p-periodic orbit, $p < \max(\mathbf{b})$. Now the matrix H_r becomes more complicated, since some of the $h(z)$ pertaining to x_2 may be equal to $h(z)$ pertaining to x_2. Consequently, the J_r are no longer guaranteed to contain the column space of the identity. Each J_r does contain the column space of an $n_r - B_{pr}$ dimensional matrix with 1 along the upper diagonal and a single -1 off the diagonal in each column. Using elementary operations, it is possible to make the first m columns of J_r upper diagonal for some integer $m \geq (n_r - B_{pr})/2$. Thus, the rank of each J_r is at least $(n_r - B_{pr})/2$ and the entire matrix has $\mathrm{rank}(J) \geq (n - B_p)/2$.

The dimension of the set S of all such $\mathbf{x}_1$ and $\mathbf{x}_2$ is just A_p. By the imposed conditions, $\dim(S) < (n - B_p)/2 \leq \dim(M_{x_1 x_2})$, and Lemma 5 applies.

Now we want show that almost every $\Phi_{(\phi, \langle \bar{y}_r \rangle)}$ is an immersion. We check that the matrix

$$\left((D\Phi_{r_1,1})_{\mathbf{x}}(\mathbf{v})^T, \ldots, (D\Phi_{r_1,t})_{\mathbf{x}}(\mathbf{v})^T, (D\Phi_{r_2,1})_{\mathbf{x}}(\mathbf{v})^T, \ldots \right)$$

has full rank and thus satisfies the conditions of Lemma 6 for each $(\mathbf{x}, \mathbf{v})$ in the tangent bundle $S(A)$. Note that this is a block diagonal matrix with s_r blocks, so it is sufficient to show that the columns of the ith block span the subspace $\mathbb{R}^{n_{r_i}}$ for $i = 1, \ldots, s_r$. We consider two cases.

Case 1: Consider first the subset S' of $\mathbf{x}$ that are not periodic with period $p < \max(\mathbf{b})$. The entries of each block are of the form $\nabla h(\phi^b(\mathbf{x}))^T (D\phi^b)_{\mathbf{x}}(\mathbf{v})$. Since ϕ is a diffeomorphism and $\mathbf{v} \neq 0$, we know that $(D\phi^b)_{\mathbf{x}}(\mathbf{v}) \neq 0$. Furthermore, the $\phi^b(\mathbf{x})$ are distinct points. Examining Lemma 4, it is clear that the columns span $\mathbb{R}^{n_r}$. The dimension of S' is at most $2d - 1$, so we may apply Lemma 6.

Case 2: Now consider the subset S' of $\mathbf{x}$ that are periodic with period $p < \max(\mathbf{b})$. By the conditions of the theorem, $(D\phi^{b_1})_{\mathbf{x}}$ has distinct eigenvalues from $(D\phi^{b_2})_{\mathbf{x}}$. Therefore, $\nabla h(\phi^{b_1}(\mathbf{x}))^T (D\phi^{b_1})_{\mathbf{x}}(\mathbf{v}) \neq \nabla h(\phi^{b_2}(\mathbf{x}))^T (D\phi^{b_2})_{\mathbf{x}}(\mathbf{v})$. Furthermore, the relationship depends on h, and again referencing Lemma 4, it is clear that

the columns span $\mathbb{R}^{n_r}$. The dimension of S' is certainly less than $2d - 1$, so we can safely apply Lemma 6.

Theorem 7 can be extended to continuous dynamical systems (smooth vector fields on a manifold) by letting the flow ϕ_t of X be ϕ in the statement of the theorem.

Discussion

Theorem 1 and the more general result presented in Theorem 2 (and its corollary) were given proofs intended to follow those presented by Takens. The original "transversality" argument, however, has been replaced with what we reckon is a simpler and more direct argument. These clarify how perturbations to the observation functions can be constructed and highlight why $2m + 1$ dimensions are necessary to have mappings that are generically embeddings. Theorem 7 is similar to Theorem 2, but takes advantage of the more powerful framework, built around the notion of prevalence, established by Sauer et al. [29]. It also provides more specific conditions on the periodic orbits than Theorem 2 and thus can be applied to certain non-generic situations that Takens' original framework would exclude. Namely, the set of periodic points need not be finite (as required in Takens' original theorem and Theorem 2), so long as the dimensionality does not exceed the bounds stated in Theorem 7. Theorem 7 is an extension of Remark 2.9 in [29], which we explicitly proved by determining a sufficient restriction for the periodic orbits when the lags composing Φ aren't necessarily consecutive.

This work also develops a language to describe a wider family of cases for reconstructing state space manifolds from multiple observational time series to encourage wider applicability of SSR in the natural sciences. For example, these results can be extended to another special case of interest for reconstructions using time derivatives [40], when multiple observation functions are available. The argument for this case is analogous to that used by Takens [12] for the case when all the derivatives are from a single observation function. Furthermore, these theorems validate heuristic work using spatial lag reconstructions and mixed spatial and temporal lag reconstructions to study spatially coupled dynamics [41].

More importantly, in terms of future applications, Theorems 2 and 7 set the stage for practical reconstruction of state space manifolds from multiple observation functions. This is significant in answering objections to single variable state space reconstruction (SSR) concerning the excessive phenomenology of lagged-coordinate embeddings [26]. These two theorems provide proof of principle for modeling attempts of nonlinear dynamics in the natural sciences involving multiple time series (e.g. [20]), and lays bare the rather non-restrictive assumptions required in such applications for building mechanistic models from multiple time series variables. Moreover, it gives support to the notion of using multiple embeddings as a potentially efficient way of extracting information from time series data of limited length, but where there are potentially many simultaneous observations of dynamics on the same attractor manifold. By reducing correlations in noise between the reconstructed coordinates, these techniques should allow reconstructions to exceed the limitations placed on univariate methods [35], as heuristic examples have already suggested [20]. The potential information leverage provided by multiple embeddings possible from novel combinations of variables (and their lags) can pave the way for a plethora of new applied techniques to exploit the time-limited, but parallel observations of nature [36]. This paper is intended to complement the existing literature on SSR and help promote this potential growth area in the natural sciences.

Acknowledgments

Author Contributions

References

1. Casdagli M (1989) Nonlinear prediction of chaotic time series. Physica D 35: 335–356.
2. Crutchfield JP (1979) Prediction and stability in classical mechanics. Senior thesis in physics and mathematics, University of California, Santa Cruz.
3. Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Physical Review Letters 59: 845–848.
4. Jenouvrier S, Weimerskirch H, Barbraud C, Park YH, Cazelles B (2005) Evidence of a shift in the cyclicity of antarctic seabird dynamics linked to climate. Proceedings of the Royal Society of London, Series B 272: 887–895.
5. Lo TT, Hsu HH (2010) Change in the dominant decadal patterns and the late 1980s abrupt warming in the extratropical northern hemisphere. Atmospheric Science Letters 11: 210–215.
6. Planque B, Batten SD (2000) Calanus finmarchicus in the north atlantic: the year of Calanus in the context of interdecadal change. ICES Journal of Marine Science 57: 1528–1535.
7. Ramanathan A, Wang C, Schreiber SL (2002) Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements. Proceedings of the National Academy of Sciences of the United States of America 102: 5992–5997.
8. Ruelle D (1989) Chaotic evolution and strange attractors: the statistical analysis of time series for deterministic nonlinear systems. Cambridge University Press, Cambridge.
9. Scheffer M, Carpenter SR (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in Ecology & Evolution 18: 648–656.
10. Seth AK, Izhikevich E, Reeke GN, Edelman GM (2006) Theories and measures of consciousness: an extended framework. Proceedings of the National Academy of Sciences of the United States of America 103: 10799–10804.
11. Soramaki K, Bech ML, Arnold J, Glass RJ, Beyeler WE (2007) The topology of interbank payment flows. Physica A 379: 317–333.
12. Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS, eds. Symposium on Dynamical Systems and Turbulence. Berlin: Springer-Verlag, volume 898 of Lecture Notes in Mathematics. pp 366–381.
13. Wagner BK, Kitami T, Gilbert TJ, Peck D, Ramanathan A, et al. (2008) Large-scale chemical dissection of mitochondrial function. Nature Biotechnology 26: 343–351.
14. May RM, Levin SA, Sugihara G (2008) Complex systems: Ecology for bankers. Nature 451: 893–895.
15. Sugihara G (2010) On early warning signs. Seed Magazine Global Reset 2010.
16. Sugihara G (2010) Nature is nonlinear. Kyoto Journal 75: 56–57.
17. Brock WA, Sayers CL (1988) Is the business cycle characterized by deterministic chaos? Journal of Monetary Economics 22: 71–90.
18. Brock WA, Malliaris AG (1989) Differential Equations, Stability and Chaos in Dynamic Economics, volume 27 of Advanced Textbooks in Economics. North-Holland.
19. Brock WA, Hsieh DA, LeBaron BD (1991) Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. MIT Press, Cambridge, Massachusettes.
20. Dixon PA, Milicich MJ, Sugihara G (1999) Episodic fluctuations in larval supply. Science 283: 1528–1530.
21. Hsieh CH, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435: 336–340.
22. Hsieh CH, Anderson C, Sugihara G (2008) Extending nonlinear analysis to short ecological time series. American Naturalist 171: 71–80.
23. Rodo X, Pascual M, Fuchs G, Faruque ASG (2002) ENSO and cholera: a nonstationary link related to climate change? Proceedings of the National Academy of Sciences of the United States of America 99: 12901–12906.
24. Schaffer WM (1984) Stretching and folding in lynx fur returns: evidence for a strange attractor in nature? American Naturalist 124: 789–820.
25. Sugihara G, Grenfell B, May RM (1990) Distinguishing error from chaos in ecological time series. Philosophical Transactions of the Royal Society of London, Series B: Biological Science 330: 235–250.
26. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in a data series. Nature 344: 734–741.
27. Sugihara G (1994) Nonlinear forecasting for the classification of natural time series. Philosophical Transactions of the Royal Society of London, Series A 348: 477–495.
28. Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geomtery from a time series. Physical Review Letters 45: 712–716.
29. Sauer T, Yorke JA, Casdagli M (1991) Embedology. Journal of Statistical Physics 65: 579–616.
30. Hunt BR, Sauer T, Yorke JA (1992) Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces. Bulletin of the American Mathematical Society 27: 217–238.
31. Stark J, Broomhead DS, Davies ME, Huke JP (1997) Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis: Theory, Methods, and Applications 30: 5303–5314.
32. Stark J, Broomhead DS, Davies ME, Huke JP (2003) Delay embeddings for forced systems: II. Stochastic forcing. Journal of Nonlinear Science 13: 519–577.
33. Stark J (1999) Delay embeddings for forced systems: I. Deterministic forcing. Journal of Nonlinear Science 9: 255–332.
34. Gibson J, Farmer JD, Casdagli M, Eubank S (1992) An analytic approach to practical state space reconstruction. Physica D 57: 1–30.
35. Casdagli M, Eubank S, Farmer JD, Gibson J (1991) State space reconstruction in the presence of noise. Physica D 41: 52–98.
36. Ye H, Sugihara G, Deyle ER (2011) Leverging information from multiple time series with ensemble state space reconstructions. In preparation.
37. Lorenz EN (1963) Deterministic nonperiodic flow. Journal of Atmospheric Science 20: 130–141.
38. Whitney H (1936) Differentiable manifolds. Annals of Mathematics 37: 645–680.
39. Huke JP (1993) Embedding nonlinear dynamical systems, a guide to Takens' theorem. Internal report, DRA, Malvern, UK.
40. Crutchfield JP, McNamara BS (1987) Equations of motion from a data series. Complex Systems 1: 417.
41. Ørstavik S, Stark J (1998) Reconstruction and cross-prediction in coupled map lattices using spatio-temporal embedding techniques. Physics Letters A 247: 145–160.