An Emergent Space for Distributed Data with Hidden Internal Order through Manifold Learning

FELIX P. KEMETH[1,2], SINDRE W. HAUGLAND[1,2], FELIX DIETRICH[4,5], TOM BERTALAN[4], KEVIN HÖHLEIN[1], QIANXIAO LI[6], ERIK M. BOLLT[7], RONEN TALMON[8], KATHARINA KRISCHER[1], AND IOANNIS G. KEVREKIDIS[2,4,5]

[1]Physik-Department, Nonequilibrium Chemical Physics, Technische Universität München, James-Franck-Str. 1, D-85748 Garching, Germany
[2]Institute for Advanced Study - Technische Universität München, Lichtenbergstr. 2a, D-85748 Garching, Germany
[3]Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany
[4]The Department of Chemical and Biological Engineering - Princeton University, Princeton, NJ 08544, USA
[5]Department of Chemical and Biomolecular Engineering, Department of Applied Mathematics and Statistics, Johns Hopkins University and JHMI
[6]Institute of High Performance Computing, 1 Fusionopolis Way, #16-16 Connexis North, Singapore 138632, Singapore
[7]Department of Mathematics, and Department of Electrical and Computer Engineering, Clarkson Center for Complex Systems Science, Clarkson University, Potsdam, NY 13699-5815, USA
[8]Department of Electrical Engineering - Technion - Israel Institute of Technology, Technion City, Haifa, Israel 32000

Corresponding author: Ioannis G. Kevrekidis (e-mail: yannis@princeton.edu).

ABSTRACT Manifold-learning techniques are routinely used in mining complex spatiotemporal data to extract useful, parsimonious data representations/parametrizations; these are, in turn, useful in nonlinear model identification tasks. We focus here on the case of time series data that can ultimately be modelled as a spatially distributed system (e.g. a partial differential equation, PDE), but where we do not know the space in which this PDE should be formulated. Hence, even the spatial coordinates for the distributed system themselves need to be identified - to "emerge from"- the data mining process. We will first validate this "emergent space" reconstruction for time series sampled without space labels in known PDEs; this brings up the issue of observability of physical space from temporal observation data, and the transition from spatially resolved to lumped (order-parameter-based) representations by tuning the scale of the data mining kernels. We will then present actual emergent space "discovery" illustrations. Our illustrative examples include chimera states (states of coexisting coherent and incoherent dynamics), and chaotic as well as quasiperiodic spatiotemporal dynamics, arising in partial differential equations and/or in heterogeneous networks. We also discuss how data-driven "spatial" coordinates can be extracted in ways invariant to the nature of the measuring instrument. Such gauge-invariant data mining can go beyond the fusion of heterogeneous observations of the same system, to the possible matching of apparently different systems.

INDEX TERMS Data Mining, Diffusion Maps, Dimensionality Reduction, Nonlinear Dynamical Systems

I. INTRODUCTION

IN 1979, when the American embassy was vacated in Tehran, sensitive documents were not incinerated; they were instead "strip-shredded" and considered destroyed. Yet local carpet-weavers painstakingly put them together again [1]. In 2011 DARPA issued a "shredder challenge": reconstructing five shredded pages (a handwritten page, a picture etc.); the $50,000 prize was claimed a few weeks later [2].

The subject of this paper is the extraction of useful in-formation *implicit* in apparently spatially disorganized data, such as the actual physical order of the paper shreds; this can be useful for winning the DARPA challenge, or, as we show below, for visualizing and hopefully better understanding spatiotemporal simulations. Extracting useful information, such as the dimension and geometry of the underlying physical space, is crucial in identifying a consistent distributed dynamic model. Establishing the correct global sequence of temporal observation windows, as well as the number and nature of state variables in this consistent dynamic model is

also crucial for the system identification task. In this paper we use nonlinear data mining tools (in particular, versions of diffusion maps [3]–[7]) to extract these types of information from spatially/temporally disorganized data by exploiting the intrinsic variabilities in the recorded data sequences. These extracted variabilities are the key to defining new spatial co-ordinates for the representation and modeling of the system.

In classical mechanics, generalized coordinates define a frame in which to represent the configuration of a system with respect to a reference state. Many possible choices of generalized coordinates exist, and there is an "anthropic" motivation in choosing them so as to make the formulation, or the solution, of the equations of motion in these coordinates easier. Data-driven coordinate discovery provides an even broader set of alternatives. Ideally such a set of coordinates would be intrinsic to the phenomenon observed and independent of the particular nature of the measurement entity.

The focus in this paper is threefold: We will start with the observability and reconstruction of physical space from disorganized (spatially unlabelled) observations of spatially distributed processes. The extent to which we can "redis-cover" the missing physical space validates the approach. In the case of networks, where no obvious physical space exists, we will demonstrate the identification of a "surrogate" embedding space, useful in reducing the network dynamics, demonstrate how it emerges from the network dynamic data, and discuss its possible physical interpretation in terms of network heterogeneity. Motivated by this heterogeneous net-work example, we choose the term "Emergent Space" for such spatial embeddings.

Physical dynamics can be observed at several different scales. Our second focus involves the adaptation of data min-ing to the degree of feature coarse-graining we are interested in: that is, to the scale of the observer. Investigating chimera states (hybrid states of coherent and incoherent dynamics) for a partial integro-differential equation we show how we can move from recovering space on a fine scale to the extraction of a more lumped description in terms of more coarse-grained order parameters.

Lastly, we focus on exploring the use of different ob-servation modalities. Our illustrative example involves low-dimensional spatiotemporal dynamics, namely modulated traveling waves (quasiperiodic solutions) of the Kuramoto-Sivashinsky equation, and we start by reconstructing the underlying toroidal attractor. We then show how one can re-cover the same emergent space reconstruction from different types of (disorganized) measurements. We draw connections between this and traditional dynamical system "embedol-ogy" [8]: the theorems by Whitney [9], Nash [10], and Takens [11]. If two systems are thus found to be observations of each other, we illustrate, in a simple example, how spectral techniques (gauge-invariant data mining, but also Koopman-operator techniques [12], [13]) can guide us in matching the two systems. Being able to consider (in principle) all possible diffeomorphisms of a given representation of a system, leads naturally to a discussion of which representation to choose

to work with and report on. For which variables should we try to identify a dynamic model? This, in turn, leads to some simple arguments about the interplay of data mining (and, more generally, machine learning) and systems modeling, opening up several research directions, some of which we outline.

Details on the individual models and algorithms are given in the Supplementary Information [14].

II. ON DIFFUSION MAPS

In recent years, many nonlinear manifold learning tech-niques have been proposed for embedding apparently high-dimensional nonlinear phenomena in lower-dimensional spaces. Examples include Isomap [15], locally linear em-bedding [16], Laplacian eigenmaps [17] and our method of choice here: Diffusion Maps [3], [4], [18], [19].

Assume a collection of N real-valued time series segments $\{a_i\}$, $i = 1, \ldots, N$. Each sampled time series a_i is a T-dimensional vector, containing the value of one recorded variable at T discrete points in time, with T depending on the sampling rate and the time window considered (longer vectors sampling more than one variable at each moment in time can be similarly constructed). Thus, each (time series) vector can be regarded as a point in a T-dimensional Eu-clidean space, with all time series together forming a point cloud embedded in T dimensions. We assume that this point cloud lies close to a smooth manifold $\mathcal{M}$, embedded in $\mathbb{R}^T$. Calculating all pairwise Euclidean distances between the points yields, through a diffusion kernel, a symmetric $N \times N$ matrix, whose spectral properties can reveal intrinsic structures in this cloud.

A diffusion kernel weighs the Euclidean distances between points that are close in T-dimensional space (i.e., between similar time series segments), much more strongly than those pairs of points at larger Euclidean distances. This effectively embodies a random walk on the data, where the probability of jumping from point a_i to $a_{i'}$ is large if their distance is small, and vice versa. The diffusion kernel on $\mathbb{R}^T$ is defined as

$$k\left(a_i, a_{i'}\right) = e^{-\frac{\|a_i - a_{i'}\|^2}{2\epsilon}} \, .$$

Here, the scale parameter ϵ is used to tune the rate of decay of the kernel compared to a characteristic scale present in the data. For relatively small ϵ, the kernel decays fast, and only pairs of points very close in the Euclidean distance are significantly weighted.

Using the similarity measure given through the kernel, we construct a graph between all data points. For a given scale parameter $\epsilon > 0$, the connectivity between two points $a_i, a_k \in \mathbb{R}^T$ is stored in the kernel matrix $K_{ik} = k(a_i, a_k)$. The Diffusion Maps concept is based on the convergence of the normalized graph Laplacian on this graph to the Laplace–Beltrami operator on the manifold $\mathcal{M}$, in the limit $n \to \infty, \epsilon \to 0$. If the data points are not sampled uniformly

on the manifold, the matrix K has to be normalized by an estimation of the density,

$$P_{ii} = \sum_{k=1}^{n} K_{ik}, \qquad (1)$$

$$\tilde{K} = P^{-\alpha} K P^{-\alpha}. \qquad (2)$$

Here, $\alpha = 0$ (no normalization [17]) can be used in the case of uniform sampling, and $\alpha = 1$ otherwise [3]. The kernel matrix $\tilde{K}$ then has to be normalized again, by the diagonal matrix $D_{ii} = \sum_{k=1}^{n} \tilde{K}_{ik}$, yielding the Markov matrix $A = D^{-1}\tilde{K}$. A non-linear parametrization (embedding) of the manifold is then given by a certain number l of eigenvectors of A, scaled by their respective eigenvalue. The new embedding dimension l can be much smaller than the previous ambient space dimension T, in which case the algorithm achieves dimensionality reduction. Selecting the eigenvectors is not as straight-forward as for linear dimensionality reduction methods (Principal Component Analysis), but can be achieved, for example, by sorting the eigenvectors based on the absolute value of their associated eigenvalue and removing eigenvectors that are functions of previous ones [20].

Changing the length scale ϵ can lead to very different representations of the same data, as described in more detail in the section "The scale of the observer" in the paper. For convenience, we write $\epsilon = \text{const} \cdot D_{\max}^2$, with $D_{\max}$ being the maximal distance in the data set.

While most of our results are obtained using a simple Euclidean-distance-based diffusion kernel, many extensions exist, such as vector diffusion maps and (non)orientable Diffusion Maps [21]; we will also use anisotropic diffusion kernels, based on the more refined, so-called Mahalanobis-like distance [6]. Recent Diffusion Maps research has progressed along a broad range of topics and applications, including intrinsic modeling [22], reduction in multiscale dynamical systems [7], multimodal data analysis [23], and data organization [24], just to name a few.

Diffusion maps help construct a nonlinear, data-based change of coordinates, embedding an intrinsically d-dimensional manifold in a low-dimensional Euclidean space. The extrinsic dimension of the manifold can be much larger than d, making this a suitable choice for dimensionality reduction. In this paper, we use Diffusion Maps to discover intrinsic order contained in a data set, focusing on the relation between this order and the spatiotemporal dynamic modeling of the data.

III. RECOVERING SPACE FROM SPATIOTEMPORAL DATA

We start by investigating observations of spatiotemporal chaos in one spatial dimension. In particular, we consider intermittency in the complex Ginzburg-Landau equation, a nonlinear reaction-diffusion-type partial differential equation for a complex variable $W(x, t)$, arising in the modeling
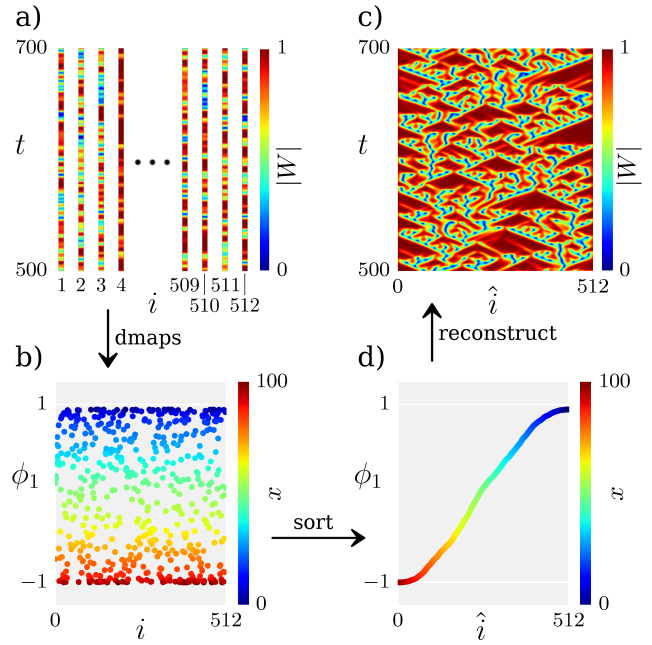

FIGURE 1: (a) Randomly shuffled time series segments (indexed by i) observed at different spatial locations in a simulation of the complex Ginzburg-Landau equation. The color corresponds to the modulus of the complex amplitude W. (b) First independent Diffusion Map coordinate obtained from the shuffled time series segments in (a). (d) Sorting the time series in (a) according to their respective first Diffusion Map coordinate yields a new indexing, $\hat{i}$. (c) The time series in (a) sorted according to this new, data-driven indexing $\hat{i}$; this identifies the correct topology of the original physical space, and reconstructs the original spatiotemporal simulation results.

of oscillatory systems [25]–[27]. The complex Ginzburg-Landau equation (CGLE) in a re-scaled form reads

$$\partial_t W = W + (1 + ic_1)\nabla^2 W - (1 + ic_2)|W|^2 W, \qquad (3)$$

with real parameters c_1 and c_2. The parameter values leading to our spatiotemporal intermittency [25] are $c_1 = 0$ and $c_2 = -3$ (see the supplemental information for details on the integration methods used).

The repeated appearance of temporally synchronous spatial patches at seemingly random locations in space and time is characteristic of this dynamical regime. Following their initial emergence, these patches shrink in size due to diffusion, giving rise to the triangular patterns in Fig. 1(c). The incoherent nature of the dynamics suggests that the (here 512) time series recorded at each individual discretization point, each consisting of 1000 steps, are mutually different (there is no spatial periodicity) Nevertheless, due to the smoothness induced by diffusion in the system, time series at points located close to each other in physical space tend to be similar: The Euclidean distance between observations at nearby points will be smaller than that between observa-
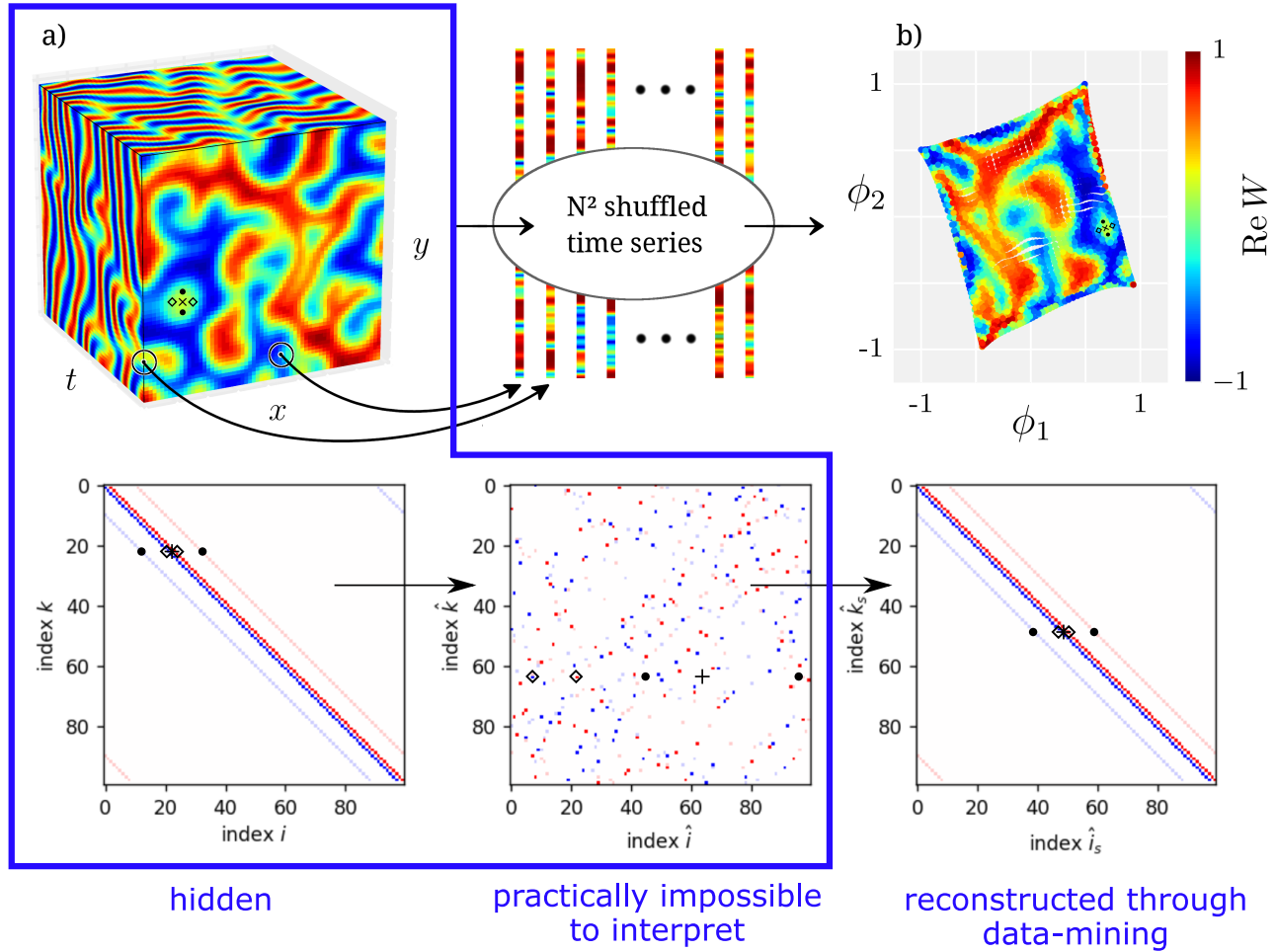
FIGURE 2: Top row: (a) Spatiotemporal data obtained from numerical simulations of the complex Ginzburg-Landau equation in two spatial dimensions. The data is colored by the real part of the complex amplitude W. To demonstrate that we can reconstruct (a version of) the original physical space from disorganized, spatially unlabelled time series data, we randomly shuffle the N^2 time series segments before using Diffusion Maps. (b) The first two independent Diffusion Map coordinates, ϕ_1 and ϕ_2, resulting from data mining and colored as in (a); the "rediscovered" space is visually homeomorphic to the hidden "true" one. Apart from a slight rotation, ϕ_1 corresponds roughly to $-x$ and ϕ_2 to y. Bottom row (schematic): in a finite difference discretization of the PDE, the "hidden" Jacobian would have a banded structure in the usual row-by-row numbering of the discretization points (a typical point five-point stencil is highlighted, left). This structure is lost upon random point labelling (middle), but recovered when the points are correctly reordered after data mining (right).

tions at points far apart in physical space. This fact will be exploited in the Diffusion Map process. In Fig. 1(d), it can be seen that the first Diffusion Map coordinate ϕ_1 associates (is one-to-one) with the spatial coordinate x: Each data point (each time series) has its own entry in the ϕ_1 vector; here large entries correspond to time series observed at small x, while small entries correspond to observations at large x (the original physical space was discovered "flipped"!). That ϕ_1 (which is one-to-one with x) turns out to be by far the most dominant mode, suggests that the data mainly vary along a single "distributed"direction; alternatively, that the system can be described using one spatial dimension. Note that for the kernel scale $\epsilon = 1.0 \approx 5.7 \cdot 10^{-4} D_{max}^2$ chosen,

with D_{max} being the maximal Euclidean distance between our time series, only a few nearest neighbors are effectively considered in the Diffusion Map computation. (The effects of the kernel scale are treated in more detail in section "The scale of the observer: Tuning the kernel parameter".)

Shuffling the individual time series, c.f. Fig. 1(a), i.e. shuffling the indices in the Diffusion Map matrix, does not affect its eigencomputations. The first and only significant Diffusion Map coordinate now appears nonmonotonic with the (shuffled) index i, as shown in Fig. 1(b). However, simply sorting the entries of the Diffusion Map coordinate in increasing order (and changing the indices of the time series accordingly to the new index $\hat{i}$) recovers the original

spatial arrangement, see Fig. 1(c). We thus argue that the dimensionality (one) and the topology (the correct ordering of the points) of the physical space x is contained in (is observable through, can be recovered from) the dynamic simulation data themselves, without spatial labels.

Shredding Fig. 1(a) vertically gave 512 time series; shredding it in the horizontal direction gives 1000 spatial snapshots. They are mutually different, yet the smoothness inherent in the time evolution implies that nearby time instances will yield similar snapshots. The above procedure applies again: We can now recover from temporally shuffled snapshot data the correct temporal sequence [28].

Clearly, this approach to recovering physical space is not restricted to line segments. It is also applicable to one-dimensional systems with periodic boundaries (rings, c.f. Fig. 4), as well as to systems with two or more spatial dimensions. Indeed, our procedure so far can be thought of as the putting-back together of effectively one-dimensional puzzles – the puzzle pieces are very long in the vertical (time series) or in the horizontal (snapshot) direction. It is easy to rationalize that by comparing edges of space-time patches (two-dimensional puzzle tiles) two-dimensional puzzles can also be reassembled. For spatiotemporal intermittency in the complex Ginzburg-Landau equation with one spatial dimension and periodic boundary conditions, one finds that the first two Diffusion Map coordinatesdefine a circle. Shuffling the data indices, applying Diffusion Maps and sorting the shuffled data along that circle, one recovers the original ring ordering modulo a rotation and possibly a reflection (not shown).

Fig. 2(a) illustrates the recovery of two-dimensional physical space from dynamical (time series segment) data. It depicts spatiotemporal intermittency on a square, with zero-flux boundary conditions, again for the complex Ginzburg-Landau equation (eq. 3). Recording individual time series at each point of a 64 by 64 discretization mesh, and applying Diffusion Maps to these 4096 data segments, one finds two independent Diffusion Map coordinates ϕ_1 and ϕ_2. Embedding the data points in (ϕ_1, ϕ_2) space, and coloring them by the real part of the complex amplitude W at $t = 0$ shows that the two Diffusion Map coordinates indeed span a space that is one-to-one with the original physical space – by visual inspection, apparently a homeomorphism, see Fig. 2(b).

It is smoothness (similarity of recorded time series from nearby points), that lets diffusion maps discover the right two-dimensional parametrization of the data and yield their actual relative positions. Neither the dimension of physical space, nor the physical location of each discretization (and thus, observation) point were used – everything is contained in (observable through, recoverable from) just the recorded dynamic data.

IV. EMERGENT SPACE RECONSTRUCTION IN NETWORK DYNAMICS

Nodes in a network will, in general, not correspond to points in a low-dimensional physical space (all-to-all coupled networks are an obvious example). Our approach can help discover an effective embedding space, based on the dimensionality and the intrinsic geometry of the node dynamics.

Consider a neuronal network that arises in modeling the pre-Bötzinger complex [29]–[31]. The states of these Hodgkin-Huxley-type neurons (a membrane potential V and a channel variable h) oscillate periodically. As in previous studies [31], the neurons are heterogeneous: Each is characterized by a different value of the intrinsic kinetic parameter I_{app}^i.

The neurons are also heterogeneous through their connectivity: They are not all-to-all coupled, but form a Chung-Lu-type network [14], [32]. The number of connections to a neuron i, its degree κ_i, varies strongly across the neurons, making it thus a second, structural, heterogeneity. The (synchronized) temporal evolution of our network of 1024 neurons is depicted in Fig. 3(a). The neurons oscillate clustered in the (V, h) plane, but with somewhat different phases and amplitudes each. Due to the heterogeneities, their instantaneous values differ. Nevertheless, the dynamics of the ensemble of neurons can be well approximated by a smooth function of the two heterogeneous parameters I_{app}^i and κ_i [31]. This is indicated by the color code in Figs. 3(b-c).

Without prior knowledge of these heterogeneity parameters, the two-dimensional nature of the collective dynamics can be recovered from temporal observations only, using Diffusion Maps. Using pairwise distances between node time series segments yields two Diffusion Map coordinates that parametrize a two-dimensional "variability manifold". The kernel width was chosen as $\epsilon = e^{10} \approx 0.39 D_{\max}^2$. Embedding the nodes in the resulting two-dimensional diffusion space, and coloring them by the heterogeneous parameter I_{app}^i, one observes an approximate one-to-one correspondence between this parameter and the first Diffusion Map coordinate (Fig. 3(d)). The second direction, transverse to the first, correlates with the degree κ_i, see Fig. 3(e).

The leading few eigenvectors following ϕ_2 are harmonics of the first two (not shown), indicating that only two major directions parametrize the variability of the node dynamics. Comparing these results with the outcome from the section "Recovering space from spatiotemporal data", an analogy arises between the two heterogeneity parameters in the neuronal network, and the spatial axes recovered in the 2-D reaction-diffusion system above. Data mining enables us, by extracting the dominant variabilities in the dynamics in both cases, to find our "Emergent Space" (an effective embedding space). In the reaction-diffusion example, this Emergent Space is one-to-one with actual physical space; in the network problem, it plays the role of a physical space (even though the latter does not really exist). Indeed, representing the behavior in terms of well-chosen basis function sets in these intrinsic variability dimensions makes the network description analogous to that of a discretized PDE in the to leading Diffusion Map coordinates (and time). This can be used to dramatically reduce the network model: a few
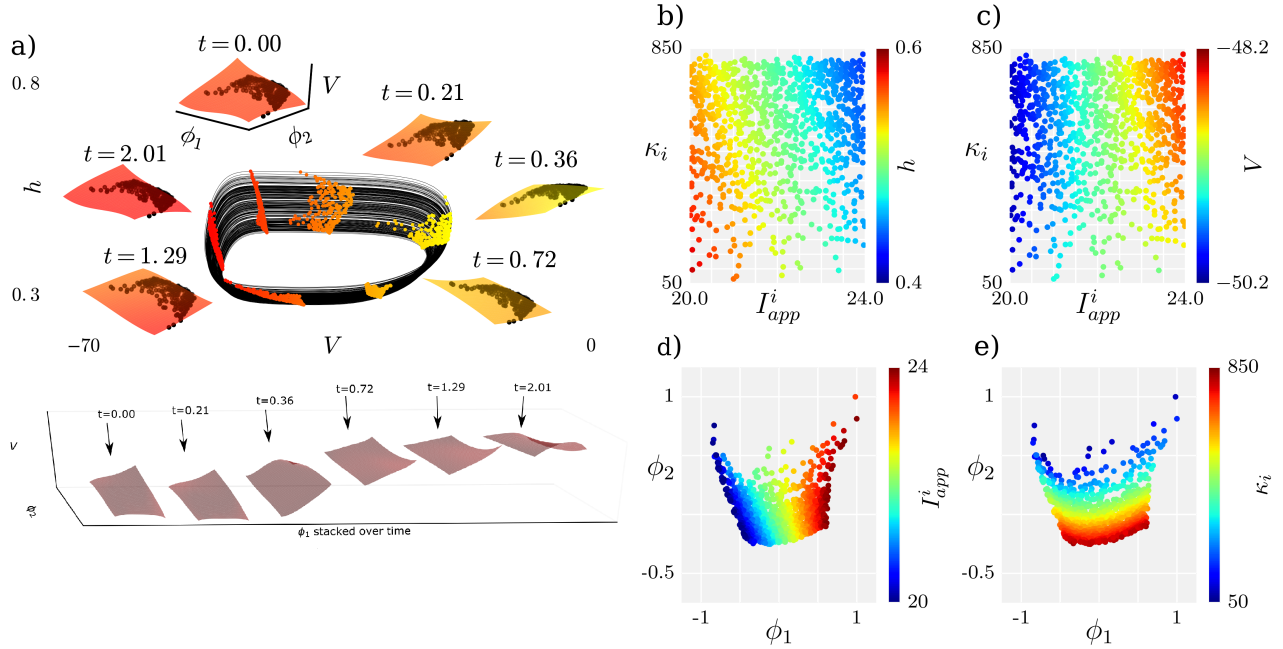
FIGURE 3: (a) Top: The closed curves in the center show temporal evolution of the two variables h and V of the 1024 individual oscillators in the Chung-Lu type network of pre-Bötzinger neurons. For six representative temporal snapshots along the trajectory (indicated by the color of the point clouds on the attractor) the potential V is plotted as a function of the first two Diffusion Map coordinates ϕ_1 and ϕ_2. The value of V for each neuron is marked by a dot. The apparent smooth dependence of V on the two Diffusion Map coordinates ([31]) strongly suggests that the network could be modelled as an irregular grid discretization of a PDE in ϕ_1 and ϕ_2 (whose time evolution is sketched around as well as below the attractor). (b,c) Each oscillator as a function of the two heterogeneous parameters κ_i and I_{app}^i, colored with the variable h and the voltages V at $t = 0$. (d,e) The two independent Diffusion Map coordinates ϕ_1 and ϕ_2, colored by applied heterogeneous current I_{app}^i and κ_i.

collocation polynomials in variability space, instead of tens of thousands of coupled ODEs accounting for each node separately [31].

V. THE SCALE OF THE OBSERVER: TUNING THE KERNEL PARAMETER

In the previous examples, a specific choice of the kernel scale ϵ was made. When recovering the Emergent Space for the PDE, we had to choose this ϵ to be very small, in the sense that only the nearest neighbors (only very similar time series) contribute to the computation. For the network, a coarser observation (a larger ϵ, taking into account more than just nearest neighbors) was required to extract the two dominant variabilities. In this section, we vary ϵ (the decay rate of the kernel, i.e. the scale of the observer) in order to explore how different features of our data are seen by observers with different perception sensitivity. Increasing ϵ decreases the ability to discriminate between nearby time series, thus "coarsening" the observation.

As a model example, we observe a chimera state, that is, a dynamical hybrid state of coexisting coherence and incoherence [33]–[36]. An example of such a state arises in a globally coupled version of the complex Ginzburg-Landau equation. A simulation in one spatial dimension is depicted in Fig. 4(a). Note that here we use periodic boundary conditions, so that the spatial axis is in fact a ring. This chimera state has an underlying two-cluster state: One of the two clusters develops incoherent dynamics while the other remains largely synchronized [37].

By choosing the kernel scale $\epsilon = e^{-2.5} \approx 3.5 \cdot 10^{-5} D_{\max}^2$ – very small, in the sense that only nearest neighbors, with very similar time series, contribute to the computation – we are able to reconstruct the full circular spatial arrangement, as depicted in Fig. 4(b). Due to the periodicity in x, two Diffusion Map coordinates are needed to embed the data. Note that this maps the coherent oscillations (by nature very similar) onto a dense cluster in Diffusion Map space. Nevertheless, by zooming in on this cluster, we find that the two Diffusion Map coordinates are still able to differentiate between the synchronous series, see the inset of Fig 4(b). This is possible since diffusion preserves a slight variation across the coherent cluster and therefore allows for distinguishing between the coherent time series.

We now vary ϵ; the embedding in the first two Diffusion Maps eigenvectors for $\epsilon = e^7 \approx 0.46 D_{\max}^2$ is shown in Fig 4(c). This figure clearly shows that the observability of physical space is now lost when observing the data in terms of its two most dominant variabilities. There are three "cusp like" regions in the data; a finer-scale plot would reveal
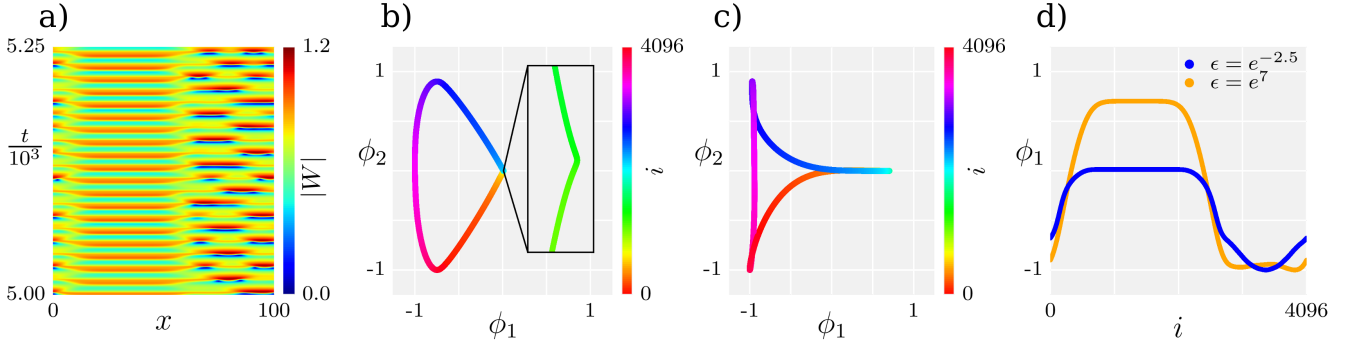
FIGURE 4: (a) Temporal evolution of a chimera state in a globally coupled version of the complex Ginzburg-Landau equation with one spatial dimension x and periodic boundary conditions. The pseudo-color corresponds to the modulus of the complex amplitude W. (b) The first two independent Diffusion Map coordinates ϕ_1 and ϕ_2 for $\epsilon = e^{-2.5} \approx 3.5 \cdot 10^{-5} D_{max}^2$, colored by the position i along the spatial coordinate x. (c) The first two independent Diffusion Map coordinates ϕ_1 and ϕ_2 for $\epsilon = e^7 \approx 0.46 D_{max}^2$, colored by the position i along the spatial coordinate x. (d) First independent Diffusion Map coordinate ϕ_1 for $\epsilon = e^{-2.5} \approx 3.5 \cdot 10^{-5} D_{max}^2$ and $\epsilon = e^7 \approx 0.46 D_{max}^2$.

intersections, which suggests that points at different physical locations (and known to have different dynamics) appear the same at this observation scale. Indeed, the entire coherent region effectively maps now to a point (the rightmost point on this plot); and while the incoherent region can be largely discriminated in terms of ϕ_2, practically all its constituents have the same ϕ_1 value. Observation in terms of only ϕ_1 now reveals two plateaus (see Fig. 4(d)); the structure within the incoherent plateau observed at small ϵ is lost at this coarser observation level (large ϵ). Thus the system naturally coarse-grains from a one-dimensional PDE to a "bistable" system, where two distinct scalar variables, two "oscillator densities" (the extent of the coherent and the incoherent regions) interact. A small percentage of phase space is taken up by the transition regions (the fronts) between the two plateaus. In a sufficiently large spatial simulation one might expect the extent (and the importance) of these transitions to be practically negligible. This perception of clustering is consistent with earlier studies, which have shown that the dynamics of the type-II chimera state can be well approximated by a modulated-amplitude two-cluster state [37], [38] - not a PDE any more, but two ODEs.

Note that, as in the example of the pre-Bötzinger neurons, the Euclidean distances between time series in the same ϕ_1 cluster can be quite large; yet observing these different time series through the right observable (their ϕ_1 component) helps cluster them meaningfully.

For very small kernel scales, the graph connecting data points is practically disconnected. Every data point is a "dimension" by itself (eigenvectors approximate indicator functions, with local support and eigenvalue one). As ϵ is increased, these many distinct dimensions start to interact and gradually merge into a coarser, one-dimensional "emergent space": the ring, described by the two Diffusion Map coordinates that we see dominating for $\epsilon = e^{-2.5}$.

This description is one-to-one with the physical space of the simulation, which thus emerges as a natural descriptor of the data for a range of ϵ values (Fig. 4(b)). Further coarsening destroys the observability of physical space from the data (as indicated by multiple loops in Fig. 4(c)), leading eventually to the "two-cluster" description through ϕ_1 only. Interestingly, this showcases the transition (in the eye of the beholder) between three distinct regimes, clearly characterized by different scales and different physical interpretations: the set of 4096 distinct coupled oscillators, the one-dimensional PDE with periodic boundary conditions, and the two-cluster amplitude description.

VI. OBSERVATIONS OF AN ATTRACTOR

In the previous examples, time series observations were used in order to discover parametrizations of variabilities intrinsic to the dynamics, yielding coordinates in which to embed the data. Yet, all our observations came from a single time window; we did not observe long-term variabilities along the time direction, important in the study of dynamics.

In this section we start using as data points partially overlapping time-series windows, obtained during long-term simulations, in order to explore such temporal variabilities. These window overlaps are chosen so that distances in the time- and the space-directions are (loosely) comparable. Our illustrative example comes from observations of a Modulated Traveling Wave (a quasiperiodic attractor) of the one-dimensional Kuramoto-Sivashinsky equation with periodic boundary conditions, shown in Fig. 5(a)

The Kuramoto-Sivashinsky equation is a fourth-order partial differential equation used to model spatiotemporal instabilities in a number of physical settings [39], [40]. In this case, the dynamics possess two frequencies: one determined by the speed of traveling along the periodic domain, and one coming from the temporal modulation. The nature of the latter becomes apparent when observing the dynamics in a frame that is co-traveling with the wave, as depicted in

Fig. 5(b). Here, every spatial point oscillates differently, but with the same constant (modulation) frequency. Altogether, the dynamics (the short time window observations) at any particular point x belongs to a 2-parameter family of possible behaviors, parametrized by the phases with respect to the traveling and to the modulation frequency. In other words, the dynamics of the system lives on a (two-)torus, as schematically depicted in Fig. 5(c). There, the toroidal angle ζ denotes the phase with respect to the modulation, and the poloidal angle θ the phase with respect to the traveling of the wave. A time-series window can be depicted as a short line-segment lying along this torus; all possible time segments constitute a two-parameter family, filling out the torus surface.

As a particular example, the dynamics shown in Fig. 5(a-b) are in the form of time series from $N = 100$ equidistributed spatial points, each of total length $T = 500$ (the periods of the two oscillations are ≈ 56 and ≈ 250). These time series are then subdivided into partially overlapping windows of length $l_{\text{string}} = 100$, at several degrees of overlap, as will be discussed below. For an overlap of $l_{\text{string}} - n$ time steps and evenly distributed segments in time, this will give a total of $(T - l_{\text{string}})/n + 1$ time windows.

How the torus is reconstructed by applying Diffusion Maps to these time windows, will, as in the previous sections, depend on the scale of the observer: the kernel parameter ϵ. For $\epsilon = e^{-0.5} \approx 1.7 \cdot 10^{-2} D_{\text{max}}^2$, the dynamics projected on the first three non-trivial Diffusion Map coordinates visually appear one-dimensional: a closed loop spanned by ϕ_1 and ϕ_2 (see Fig. 5(d)). None of the next at least 8 Diffusion Map coordinates encode new, independent directions.

A single visible ring indicates that, at this kernel scale, only one of the two torus frequencies is observable. While at this value of ϵ, the temporal modulation is simply missed by the Diffusion Maps to leading order; the traveling in the spatial domain corresponds to the rotation along this ring. The last 100 points plotted correspond to the 100 time windows obtained from the 100 different spatial locations during our last sampling interval in time. The smoothness of the coloring by the spatial coordinates at which these observations were made, shows that the Diffusion Maps can recover the physical space x.

Decreasing the scale parameter, the ring in Diffusion Map space partially unfolds into several rings, as the kernel allows us to begin detecting variability along the modulation direction. For $\epsilon = e^{-2.0} \approx 3.8 \cdot 10^{-3} D_{\text{max}}^2$, this is shown in Fig. 5(e). Note that each of the rings is still parametrized by physical space. A further decrease of ϵ leads to further unfolding of the rings, see Figs. 5(f-g). At $\epsilon = e^{-3.2} \approx 1.2 \cdot 10^{-3} D_{\text{max}}^2$ we get a full separation of the rings, making the entire torus finally visible in Fig. 5(h). The direction along each of the rings corresponds to the underlying traveling wave. We assert that the other direction, from one ring to the next, corresponds to the modulating oscillation: The time difference between two successive rings is one increment: $n = 8$ time units. Running through seven such overlaps yields a total time length of $7 \cdot 8 = 56$ time

steps, the modulation period. A finer discretization in time and larger overlaps would clearly "fill in" the torus surface. In the opposite direction, less overlap will "depopulate" the torus surface, see Fig. 5(i).

A strong relationship exists between our varying the scale of the observer, and topological data mining/persistent homology [41]. Successive plateaus in the identified surface genus in the latter correspond to successive plateaus in the dimension of our identified manifold: from a cloud of individual discretization points at very small ϵ, to the (desired) torus, then to a ring, and, ultimately, to a single point at very large ϵ. Our Figs. 5(d-h) provide an interesting study of the transition between two of the plateaus, where the dimensionality of the manifold appears to vary along it [42].

In what follows, we will assume that we have chosen the scale at which we want to observe the system, and revisit the effect of different types of observations. The dynamics live on a T^2 in function space. We want to be able to observe this T^2, that is, we want a mapping between each of our observations and a corresponding unique point on the surface of this torus. How many quantities (variables) do we need to observe to construct such a mapping? This has been a long-standing research issue in topology as well as in dynamical systems. Starting with Whitney's theorem in the 1930s [43], guaranteeing homeomorphisms between manifolds and their Euclidean embeddings, and through the work of Nash in the 1950s providing isometries [10], we have the work of Takens in 1980 [11] (and also of Farmer et al. [44]) guaranteeing homeomorphisms between the system state space and a space spanned by just a few delayed measurements of a single scalar observable (see [8]). This is exactly what we did, using time windows at various spatial locations in our discussion above. Since the dimension of the attractor is 2, our 99 delays were certainly sufficient to guarantee an embedding.

In fact, what we observe is the function $u(x, t)$ for $x \in [0, 2\pi]$ and t within the modulation period. This is a two-parameter family of points, and, at each one of them, we took short temporal measurements (a value and 99 delays). Clearly, there are other ways by which one can learn (observe) this surface. One can take horizontal measurement segments (short discretized solution profiles at fixed time). One can consider traveling observers: short measurement segments along an arbitrary angle in (x, t). Alternatively, as discussed above, we can consider small spatiotemporal patches centered at each point (or, to make the observation finite, several leading Fourier components of the function within each patch, or several leading PCA components of the histogram of the function values in the patches). Each of these different types of measurement data simply constitutes a different way to observe "the same" points on the attractor in function space (or, alternatively, points on our $u(x, t)$ surface). All of these different types of measurements of points on the attractor (all these different embeddings of the original torus) can therefore be diffeomorphically mapped to each other.

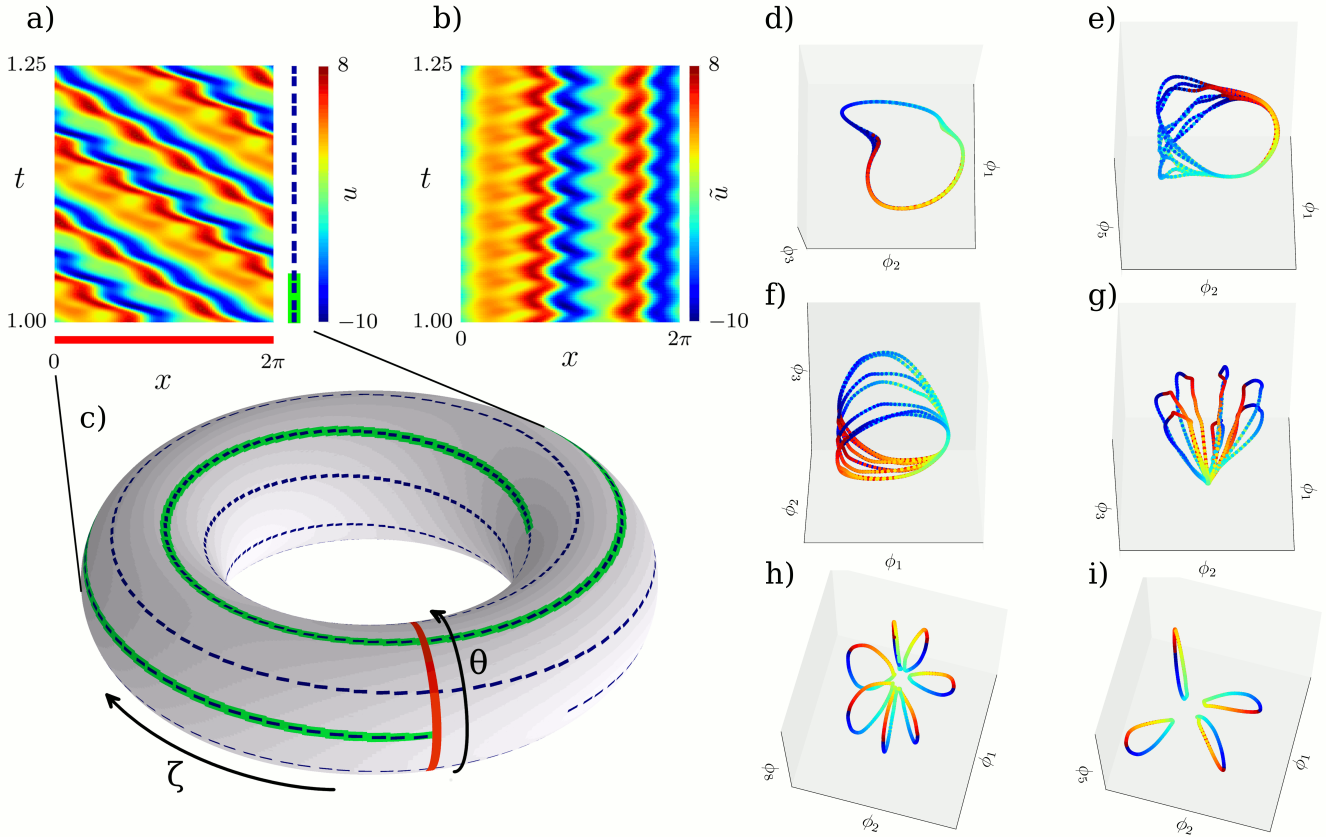Interestingly, if we can agree on a reference set of observ-

FIGURE 5: (a) Temporal evolution of u in the 1-D Kuramoto-Sivashinsky equation with periodic boundaries, showing modulated traveling waves. (b) The dynamics of (a) in a co-rotating frame. The color corresponds to u. (c) Torus spanned by the phases with respect to traveling wave, θ, and modulation, ζ, respectively. Going along space (red line) only changes the former, whereas moving in time (green line, dashed blue line) changes both phases. (d) First three independent Diffusion Map coordinates obtained from overlapping time segments created with a time shift of $n = 8$ and for $\epsilon = e^{-0.5} \approx 1.7 \cdot 10^{-2} D_{\max}^2$, colored by spatial position. (e-h) Equivalents to (d) for $\epsilon \approx 3.8 \cdot 10^{-3} D_{\max}^2$, $\epsilon \approx 2.1 \cdot 10^{-3} D_{\max}^2$, $\epsilon \approx 2.0 \cdot 10^{-3} D_{\max}^2$ and $\epsilon \approx 1.2 \cdot 10^{-3} D_{\max}^2$, respectively. (i) First three independent Diffusion Map coordinates obtained for $n = 14$ and $\epsilon = e^{-2.3}$, colored by spatial position. See discussion and interpretation of the images in the text.

ables and construct the conjugacy between this set and our various types of observation, we can then easily fuse information from the various types of observation (i.e. the different measurement instruments/modalities). It would make sense for this reference set of observables to be in some sense intrinsic to the manifold, and not to depend on the particular embedding; observables based on the manifold curvature tensor, like the ones resulting from the Codazzi-Mainardi equations, might be good candidates for this [45].

Manifold learning can therefore go beyond the recovery of physical space or the creation of useful embeddings for dynamical observations. It holds the promise of fusing heterogeneous observations of the same dynamical system and of realizing when different dynamical systems are indeed observations of each other (are conjugate). The property of obtaining a description of a system that does not depend on the measurement instrument is sometimes termed "gauge invariance", and indeed performing gauge-invariant data

mining constitutes a promising research direction. Diffusion maps based on a Mahalanobis-like distance (involving the local pseudo-inverse of the noise covariance) were proposed by Singer and Coifman in 2008 [6], and have been used to this effect in e.g. [7], [18], [22].

As a last example, Fig. 6 demonstrates this data-driven matching of a simple stochastic nonlinear oscillator (system (a)) with a nonlinear observation of it (system (b)). We parametrize the states $x = (x_1, x_2)$ and $\hat{x} = (\hat{x}_1, \hat{x}_2) = T(r(x_1, x_2), \theta(x_1, x_2))$ of both systems separately with Diffusion Maps. Instead of the Euclidean distance, we use the Mahalanobis distance between data points x^i, x^k (and $\hat{x}^i, \hat{x}^k$, respectively), defined through [6]

$$d(x^i, x^k)^2 = \frac{1}{2}(x^i - x^k)^T (C(x^i) + C(x^k))^\dagger (x^i - x^k). \quad (4)$$

Here, $C(x^i)$ and $C(x^i)$ are covariance matrices obtained by short bursts of trajectories started at the given points. In the original system, the short bursts result in isotropic
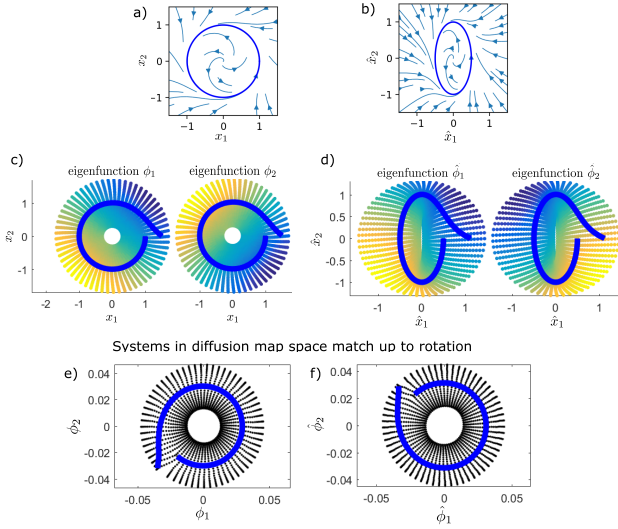
FIGURE 6: Panel (a) shows a limit cycle system with $\dot{\theta} = 1, \dot{r} = r - r^3$, where we add small amplitude white noise at each time step (isotropic in Cartesian space, not visible in the plots). Panel (b) shows the same limit cycle, but observed through $T(r, \theta) = [(r^2)/2\cos(\theta), \sqrt{r}\sin(\theta)]$. The second row shows the Diffusion Map eigenvectors obtained from the original system, $\phi_{1,2}$, and the transformed system, $\hat{\phi}_{1,2}$, as color on the phase space around the limit cycle. An example trajectory is depicted in blue. The third row shows the two systems in Diffusion Map space, where they match up to a rotation because we rescaled the distances based on the local covariances of the noise.

diffusion (so that the covariance is the identity), while the observation function distorts this to anisotropic diffusion in system (b). The pseudo-inverse $(\cdot)^\dagger$ is used in general, because the matrices are usually rank-deficient. The local noise covariance provides an estimate of the local Jacobian of the transformation that is used to re-scale the local distances, such that the metric on both spaces agree. Parametrization with Diffusion Maps then allows us to match the two dynamical systems modulo a rotation. An alternative approach to creating data-driven conjugacies between dynamical systems using Koopman operator eigenfunctions [12] is discussed in the Supplemental Information.

VII. DISCUSSION

We started by demonstrating that manifold learning techniques and, in particular, Diffusion Maps, can be used to reconstruct the topology of the physical space in which an observed process takes place, only from disorganized (spatially unlabelled) collections of temporal measurements (time series segments). This becomes especially useful in problems where no obvious physical space is involved, such as network dynamics. Here, by discovering intrinsic variability directions in the dynamics, the network can be conveniently reduced: Instead of thousands of coupled oscillators, the

response surface can be described in terms of a few collective Diffusion Map coordinates in our Emergent Space.

We then focused on chimera states arising in the dynamics of integrally coupled partial differential equations. Being able to tune systematically the kernel parameter ϵ gives us the ability to vary the scale of the observer and, ultimately, to construct representations of the system at different levels of coarse-graining [46]–[48]. In our examples, we always the same value of ϵ for all data points of a given data set. Note that for data where the density of points varies greatly over the data set, or where the density cannot be bounded from below, variable bandwidth kernels can be constructed [49].

We also discussed the observation of the same physical space topology through several different types of measurements: short time series windows, short spatial segments, small space-time patches. The ability to fuse heterogeneous observations of a dynamical system naturally brought forth the possibility of realizing that different dynamical systems are really observations of each other.

The question then arises: Given two dynamical systems, are their state spaces/their representations related by a diffeomorphism, such that the dynamics in the first system correspond to the dynamics in the second? This question is vital in the theory of dynamical systems, where such diffeomorphisms are typically constructed analytically. A famous closed-form example of such a transformation is the Cole-Hopf transform between the diffusion equation and the viscous Burgers equation [50]. The related idea of a normal form, shared by two systems, also provides an avenue towards matching them. We illustrated a data-driven path towards such a matching without access to closed-form models, but rather via processing observations of both systems' states with the help of a Mahalanobis-like distance in the Diffusion Map computations. Any model, even a quite inaccurate one, that lies in the same universality class as the true process we observe (that shares the same normal form with the true process), is but one diffeomorphism away from the truth. Modern machine-learning tools (like deep nets) can help find this diffeomorphism in a data-driven way. This "calibration curve" is a matching between the inaccurate model and the truth. Such calibration curves arise in the recent multifidelity modeling literature, e.g. [51].

If machine learning (gauge-invariant data mining) allows us to work practically indiscriminately with any possible diffeomorphic realization of a system, which one should we choose? Which one is "the best"? Until now, the optimality criteria were guided by human understanding of the process. Choosing physically interpretable variables for a model makes it easier to understand. It was this anthropocentric sense of simplicity that decided the observables and the form of the model equations. However, the temporal evolution in these variables might be rather complicated. Lax pairs provide an alternative point of view in choosing one's sense of "simplicity": here, the variables are quite complicated objects, resulting from eigenproblem solutions at every temporal step, yet the dynamics are almost trivial [52].

Machine learning can help us broaden our scope of possible simplicities, in that we can now choose from a much wider range of different system representations, and select the kind of simplicity we prefer: for example, linear dynamics in a minimal number of variables, if possible. This is emphatically not a new idea. Principal Component Analysis (PCA) is a time-honored reduction method. "Nonlinear" principal components (the auto-encoder bottleneck neural networks of the 1980s [53] that are now back as deep nets) can also provide parsimonious descriptions of processes in terms of a few observables (the states of the bottleneck neurons, that are, however, not easily physically interpretable). What has made these techniques widely applicable is the computational savings in producing accurate predictions, despite the "uninterpretability" of their state variables. It is reassuring that "easy" back-and-forth mappings between these and the interpretable, physical variables can be constructed.

We close this paper with a small gallery of examples, where we chose simplicity to mean a sense of beauty (admittedly, our own subjective criterion!). The first system is a chimera state, containing, in interpretable variables, spatiotemporal chaos. In Equal Space, it appears to admit periodic, or, at worst, quasi-periodic, visually remarkably coherent motion [Video 1]]. The second system is also a chimera state; in Equal Space it brings to the minds of each of us certain distinctive swarming behavior of the starlings of Rome [Video 2]. The last example system consists of 10000 globally coupled Stuart-Landau oscillators. In Equal Space, their probability density function evolves on what visibly suggests a homoclinic tangle involving the stable and unstable manifolds of a coarse periodic solution [Video 3]. In the words of the Little Prince, "c'est véritablement utile puisque c'est joli" [54].

The work we have discussed has an increasingly medieval flavor: We aspire to predictions, starting from observations, processing the data and never trying to grasp the physical nature of the observables or the interactions between them (obtaining closed-form expressions for physical laws).

It would appear that because of the complexity and interconnectedness of the systems we study today, the emphasis gradually shifts from understanding the system and its laws as a whole, to understanding the algorithms that create representations of the system, based on a possibly new, different sense of simplicity. Even though the new, data-driven variables may be difficult to rationalize, the mathematics involved in using them to make predictions remain the same.

REFERENCES

[1] Adrian Chen. Unshreddable. How to reassemble a shredded document. Slate, jul 2009.

[2] DARPA Shredder Challenge. 2014.

[3] David Donoho, Charles Chui, Ronald R. Coifman, and Stéphane Lafon. Special issue: Diffusion maps and wavelets diffusion maps. Applied and Computational Harmonic Analysis, 21(1):5–30, 2006.

[4] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National

Academy of Sciences of the United States of America, 102(21):7426–7431, 2005.

[5] Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Applied and Computational Harmonic Analysis, 21(1):113–127, 7 2006.

[6] Amit Singer and Ronald R. Coifman. Non-linear independent component analysis with diffusion maps. Applied and Computational Harmonic Analysis, 25(2):226–239, 2008.

[7] Carmeline J. Dsilva, Ronen Talmon, C. William Gear, Ronald R. Coifman, and Ioannis G. Kevrekidis. Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems. SIAM Journal on Applied Dynamical Systems, 15(3):1327–1351, 2016.

[8] Tim Sauer, James A Yorke, and Martin Casdagli. Embedology. Journal of statistical Physics, 65(3):579–616, 1991.

[9] J. Eells, D. Toledo, and Hassler Whitney. Collected Papers of Hassler Whitney (Contemporary Mathematicians). Birkhäuser, 1992.

[10] John Nash. C^1 isometric imbeddings. Annals of Mathematics, 60(3):383–396, 1954.

[11] Floris Takens. Detecting strange attractors in turbulence, pages 366–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 1981.

[12] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied koopmanism. Chaos, 22:047510, 2016.

[13] Erik M. Bollt, Qianxiao Li, Felix Dietrich, and Ioannis Kevrekidis. On matching, and even rectifying, dynamical systems through Koopman operator eigenfunctions. SIAM Journal on Applied Dynamical Systems, 17(2):1925–1960, jan 2018.

[14] See Supplemental Material at [URL will be inserted by publisher] for details on the individual models and algorithms., 2017.

[15] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319–2323, 2000.

[16] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323–2326, 2000.

[17] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15(6):1373–1396, 2003.

[18] Amit Singer, Radek Erban, Ioannis G. Kevrekidis, and Ronald R. Coifman. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. Proceedings of the National Academy of Sciences, 106(38):16090–16095, 2009.

[19] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. Multiscale Modeling & Simulation, 7(2):842–864, 2008.

[20] Carmeline J. Dsilva, Ronen Talmon, Ronald R. Coifman, and Ioannis G. Kevrekidis. Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study. Applied and Computational Harmonic Analysis, 44(3):759–773, may 2018.

[21] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. Communications on Pure and Applied Mathematics, 65(8):1067–1144, 2012.

[22] R. Talmon and R. R. Coifman. Empirical intrinsic geometry for nonlinear modeling and time series filtering. Proceedings of the National Academy of Sciences, 110(31):12535–12540, 2013.

[23] Roy R. Lederman and Ronen Talmon. Learning the geometry of common latent variables using alternating-diffusion. Applied and Computational Harmonic Analysis, in press, 2015.

[24] Ronald R. Coifman and Matan Gavish. Harmonic Analysis of Digital Data Bases, pages 161–197. Wavelets and Multiscale Analysis. Birkhäuser, 2011.

[25] B.I. Shraiman, A. Pumir, W. van Saarloos, P.C. Hohenberg, H. Chaté, and M. Holen. Spatiotemporal chaos in the one-dimensional complex Ginzburg-Landau equation. Physica D: Nonlinear Phenomena, 57(3):241–248, 1992.

[26] Hugues Chaté and Paul Manneville. Dynamics of complex systems phase diagram of the two-dimensional complex Ginzburg-Landau equation. Physica A: Statistical Mechanics and its Applications, 224(1):348–368, 1996.

[27] Vladimir García-Morales and Katharina Krischer. The complex Ginzburg-Landau equation: An introduction. Contemporary Physics, 53(2):79–95, 2012.

[28] C. J. Dsilva, B. Lim, H. Lu, A. Singer, I. G. Kevrekidis, and S. Y. Shvartsman. Temporal ordering and registration of images in studies of developmental dynamics. Development, 142(9):1717–1724, apr 2015.

[29] Jonathan Rubin and David Terman. Synchronized activity and loss of synchrony among heterogeneous conditional oscillators. SIAM Journal on Applied Dynamical Systems, 1(1):146–174, 2002.

[30] Carlo R. Laing, Yu Zou, Ben Smith, and Ioannis G. Kevrekidis. Managing heterogeneity in the study of neural oscillator dynamics. The Journal of Mathematical Neuroscience, 2(1):1–22, 2012.

[31] M. Choi, T. Bertalan, C.R. Laing, and I.G. Kevrekidis. Dimension reduction in heterogeneous neural networks: Generalized Polynomial Chaos (gPC) and ANalysis-Of-VAriance (ANOVA). The European Physical Journal Special Topics, 225(6):1165–1180, Sep 2016.

[32] Carlo R. Laing, Karthikeyan Rajendran, and Ioannis G. Kevrekidis. Chimeras in random non-complete networks of phase oscillators. Chaos, 22(1):013132, 2012.

[33] Yoshiki Kuramoto and Dorjsuren Battogtokh. Coexistence of coherence and incoherence in nonlocally coupled phase oscillators. Nonlinear Phenom. Complex Syst., 5:380–385, 2002.

[34] Daniel M. Abrams and Steven H. Strogatz. Chimera states for coupled oscillators. Phys. Rev. Lett., 93:174102, 2004.

[35] Mark J Panaggio and Daniel M Abrams. Chimera states: coexistence of coherence and incoherence in networks of coupled oscillators. Nonlinearity, 28(3):R67–R87, 2015.

[36] Felix P. Kemeth, Sindre W. Haugland, Lennart Schmidt, Ioannis G. Kevrekidis, and Katharina Krischer. A classification scheme for chimera states. Chaos, 26(9):094815, 2016.

[37] Lennart Schmidt and Katharina Krischer. Clustering as a prerequisite for chimera states in globally coupled systems. Phys. Rev. Lett., 114:034101, Jan 2015.

[38] Lennart Schmidt and Katharina Krischer. Chimeras in globally coupled oscillatory systems: From ensembles of oscillators to spatially continuous media. Chaos, 25(6):064401, 2015.

[39] James M. Hyman and Basil Nicolaenko. The Kuramoto-Sivashinsky equation: A bridge between PDE's and dynamical systems. Physica D: Nonlinear Phenomena, 18(1):113–126, 1986.

[40] Ioannis G. Kevrekidis, Basil Nicolaenko, and James C. Scovel. Back in the saddle again: A computer assisted study of the kuramoto–sivashinsky equation. SIAM Journal on Applied Mathematics, 50(3):760–790, jun 1990.

[41] T. Kaczynski, K. Mischaikow, and M. Mrozek. Computational Homology. Applied Mathematical Sciences. Springer New York, 2004.

[42] William K. Allarda, Guangliang Chena, and Mauro Maggioni. Multi-scale geometric methods for data sets ii: Geometric multi-resolution analysis. Applied and Computational Harmonic Analysis, 32:435–462, 2012.

[43] Hassler Whitney. Differentiable manifolds. Annals of Mathematics, 37(3):645–680, 1936.

[44] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. Phys. Rev. Lett., 45:712–716, Sep 1980.

[45] Wilhelm Blaschke and Kurt Leichtweiß. Elementare Differentialgeometrie. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg, 1973.

[46] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In NIPS, 2004.

[47] Matthias Hein and Jean-Yves Audibert. Intrinsic dimensionality estimation of submanifolds in rd. In Proceedings of the 22nd international conference on Machine learning - ICML '05, pages 289–296, 2005.

[48] Ronald R. Coifman, Yoel Shkolnisky, Fred J. Sigworth, and Amit Singer. Graph Laplacian tomography from unknown random projections. IEEE Trans. Image Process, 17(10):1891–1899, 2008.

[49] Tyrus Berry and John Harlim. Variable bandwdth diffusion kernels. Applied and Computational Harmonic Analysis, 40(1):68–96, 2016.

[50] Eberhard Hopf. The partial differential equation $u_t + uu_x = \mu u_{xx}$. Communications on Pure and Applied Mathematics, 3(3):201–230, 1950.

[51] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 473(2198), 2017.

[52] Peter D. Lax. Integrals of nonlinear equations of evolution and solitary waves. Communications on Pure and Applied Mathematics, 21(5):467–490, 1968.

[53] M.A. Kramer. Autoassociative neural networks. Computers & Chemical Engineering, 16(4):313–328, 1992. Neutral network applications in chemical engineering.

[54] A. de Saint-Exupéry. Le petit prince. Bonniers, 2001.

• • •

Supplemental Material

Felix P. Kemeth,[1,2] Sindre W. Haugland,[1,2] Felix Dietrich,[3,4] Tom Bertalan,[3] Qianxiao Li,[5]
Erik M. Bollt,[6] Ronen Talmon,[7] Katharina Krischer,[1] and Ioannis G. Kevrekidis[2,3,4,*]

[1]*Physik-Department, Nonequilibrium Chemical Physics,*
Technische Universität München, James-Franck-Str. 1, D-85748 Garching, Germany
[2]*Institute for Advanced Study - Technische Universität München,*
Lichtenbergstr. 2a, D-85748 Garching, Germany
[3]*The Department of Chemical and Biological Engineering - Princeton University, Princeton, NJ 08544, USA*
[4]*Department of Chemical and Biomolecular Engineering,*
Department of Applied Mathematics and Statistics, Johns Hopkins University and JHMI
[5]*Institute of High Performance Computing, 1 Fusionopolis Way,*
#16-16 Connexis North, Singapore 138632, Singapore
[6]*Department of Mathematics, and Department of Electrical and Computer Engineering,*
Clarkson Center for Complex Systems Science, Clarkson University, Potsdam, NY 13699-5815, USA
[7]*Department of Electrical Engineering – Technion – Israel Institute of Technology, Technion City, Haifa, Israel 32000*

FIGURES 1 AND 2 IN THE PAPER: SPATIOTEMPORAL INTERMITTENCY IN THE COMPLEX GINZBURG-LANDAU EQUATION

The complex Ginzburg-Landau equation (CGLE) in a re-scaled form reads

$$\partial_t W = W + (1 + ic_1) \nabla^2 W - (1 + ic_2) |W|^2 W,$$

with real parameters c_1 and c_2. The parameter values leading to our spatiotemporal intermittency [3] are $c_1 = 0$ and $c_2 = -3$. The equation is integrated using a pseudo-spectral integration method with exponential time-stepping [4] and fixed time step $dt = 0.05$. The system sizes considered are $L = 100$ with $N = 512$ grid points for one-dimensional simulations (Fig. 2) and $L_x = L_y = 40$ with $N_x = N_y = 64$ grid points for two-dimensional simulations (Fig. 3). The solution in both cases is sampled as $T = 1000$ snapshots between $t_0 = 500$ and $t_1 = 700$ time units. For the two-dimensional system, a kernel scale of $\epsilon = e^2 \approx 4.0 \cdot 10^{-3} D_{max}^2$ is chosen.

Fig. 1 in the paper discussed the spatial reconstruction of the solution from temporal observations. In that case, our data points consisted of 512 time series of length 1000 each. Fig. 1 below supports the claim made in the paper, that the same data-driven approach is also able to accomplish a temporal reconstruction of the solution from spatial observations. Here, the data points consisted of 1000 spatial snapshots of length 512. Clearly, the first diffusion maps component is now one-to-one with physical time, and the panels of the figure echo the ones of Fig. 2 in the paper.

---

* yannis@princeton.edu

FIGURE 3 IN THE PAPER: HETEROGENEOUS CHUNG-LU TYPE NETWORK OF PRE-BÖTZINGER NEURONS

As a toy example for complex dynamics without an obvious physical space, we consider an ensemble of 1024 neurons, a caricature of the pre-Bötzinger complex [5–7],

$$C\frac{dV_i}{dt} = -g_{Na}m(V_i)h_i(V_i - V_{Na}) - g_l(V_i - V_l)$$
$$+ I_{syn}^i + I_{app}^i$$
$$\frac{dh_i}{dt} = \frac{h_\infty(V_i) - h_i}{\tau(V_i)}.$$

with the coupling

$$I_{syn}^i = \frac{g_{syn}(V_{syn} - V_i)}{N}\sum_{j=1}^{N} A_{ij}s(V_j).$$

Here

$$m(V) = (1 + \exp(-(V + 37)/6))^{-1},$$
$$h_\infty(V) = (1 + \exp((V + 44)/6))^{-1},$$
$$\tau(V) = (\epsilon\cosh((V + 40)/5))^{-1},$$
$$s(V) = (1 + \exp(-(V + 40)/5))^{-1}$$

are nonlinear functions, whereas $C = 0.21$, $g_{Na} = 2.8$, $g_l = 2.4$, $g_{syn} = 0.3$, $V_{syn} = 0$, $V_{Na} = 50$, $V_l = -65$ and $\epsilon = 0.1$ are constants. As in previous studies [7], we take the intrinsic kinetic parameter I_{app}^i to vary across the ensemble ($I_{app}^i = 22 + 2\omega_i$, where ω_i is drawn from a uniform distribution on $[-1, 1]$), making the network kinetically heterogeneous.

We connect the neurons in the form of a Chung-Lu type network [8], where the network topology is given by a symmetric adjacency matrix $\mathbf{A}$. The entries A_{ij} are 1 if there is a coupling between oscillators i and j, and 0 otherwise. The algorithm for the creation of the A_{ij} uses
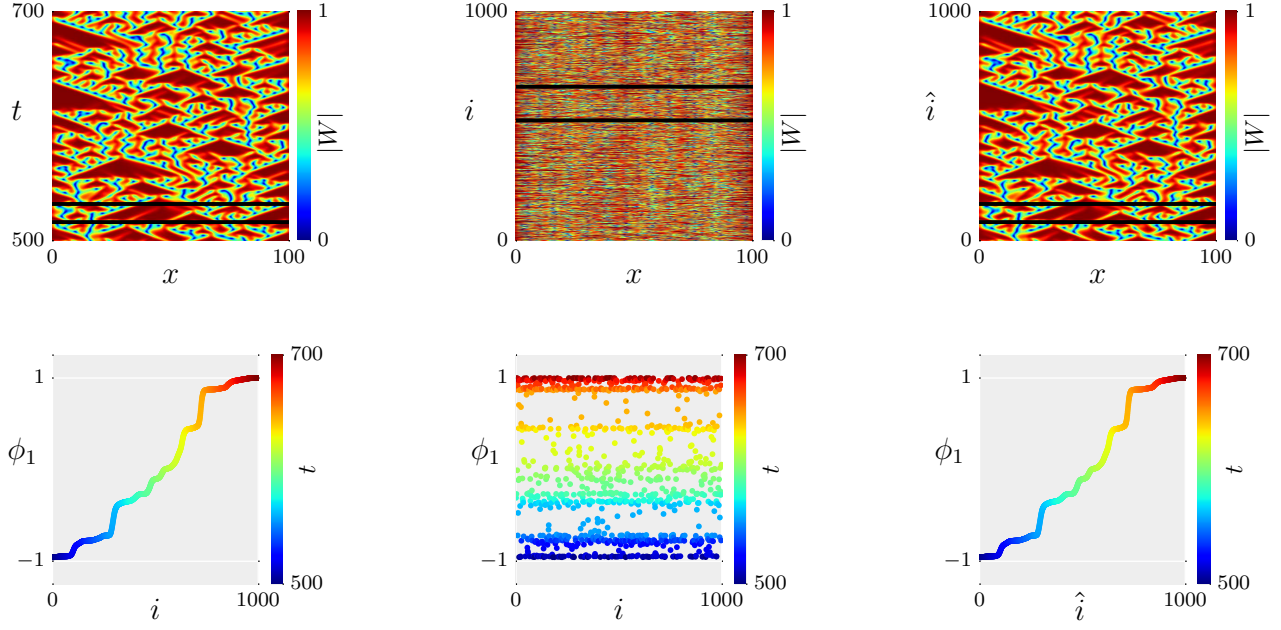
FIG. 1. Reconstruction of the temporal arrangement of the one-dimensional spatiotemporal intermittency in the CGLE following the spatial reconstruction, Fig. 1 in the paper.

a sequence of weights w_i for each oscillator i, defined as

$$w_i = pN(i/N)^r \quad , \quad i = 1, 2, ..., N$$

with parameters $p = 0.80$ and $r = 0.40$. From these weights, the entries P_{ij} of a matrix $\mathbf{P}$ of connection probabilities are defined as

$$P_{ij} = P_{ji} = \min\left(\frac{w_i w_j}{\sum_k w_k}, 1\right) .$$

The matrix is then mirrored along the diagonal, and all diagonal entries are set to zero to avoid self-loops. As initial conditions, $V = -60.0$ and $h = 0.0$ are taken for all oscillators. We simulated our realization of a Chung-Lu network of $N = 1024$ oscillators, using lsode [9, 10]. The time series observations from each neuron were taken between $t_0 = 20$ and $t_1 = 40$ in the form of $T = 2001$ time steps.

with $\langle \ldots \rangle$ denoting spatial averages. It has the property that for the spatially uniform mode,

$$\partial_t \langle W \rangle = -i\nu \langle W \rangle \Rightarrow \langle W \rangle = \eta e^{-i\nu t}$$

holds. Thus, the mean $\langle W \rangle$ is confined to harmonic motion with frequency ν and amplitude η [11]. This system is known to exhibit type-II chimeras for $c_1 = 0.2$, $c_2 = -0.63$, $\nu = 0.1$ and $\eta = 0.65$, starting from random initial conditions. Here, a system size of $L = 100$ with periodic boundaries and $N = 4096$ grid points is considered. For integration, a pseudo-spectral method with exponential time-stepping [4] and fixed time step of $dt = 0.01$ was used. The data was sampled between $t_0 = 5000$ and $t_1 = 6000$ at $T = 4000$ snapshots. Data mining for different values of the kernel scale parameter ϵ was performed on 4096 time series of length 4000 each.

FIGURE 4 IN THE PAPER: TYPE-II CHIMERA IN THE MODIFIED COMPLEX GINZBURG-LANDAU EQUATION

The modified complex Ginzburg-Landau equation (MCGLE), that is, the CGLE with nonlinear global coupling, is described by

$$\begin{aligned} \partial_t W = \ & -i\nu W + (1 + ic_1)\nabla^2 W \\ & - (1 + i\nu)(\langle W \rangle - W) \\ & + (1 + ic_2)\left(\langle |W|^2 W \rangle - |W|^2 W\right) \end{aligned}$$

FIGURE 5 IN THE PAPER: MODULATED TRAVELING WAVES IN THE 1-D KURAMOTO-SIVASHINSKY EQUATION

The Kuramoto-Sivashinsky equation in one spatial dimension reads

$$\partial_t u + \alpha \left(u\partial_x u + \partial_{xx} u \right) + 4\partial_{xxxx} u = 0$$

with the real variable $u = u(x, t)$ and a single real parameter α. For system size $L = 2\pi$, with periodic boundary conditions, and $\alpha = 53.3$, () is known to exhibit stable
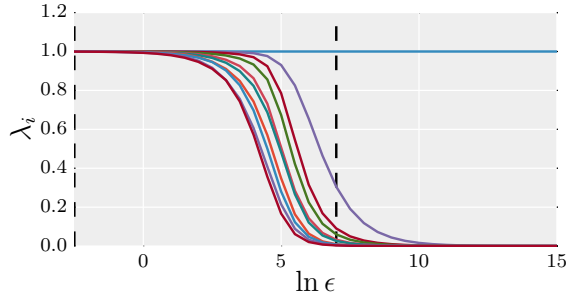
3

FIG. 2. The dominant ten eigenvalues of the eigenvalue spectrum obtained from the diffusion matrix for different values of the kernel parameter ϵ.
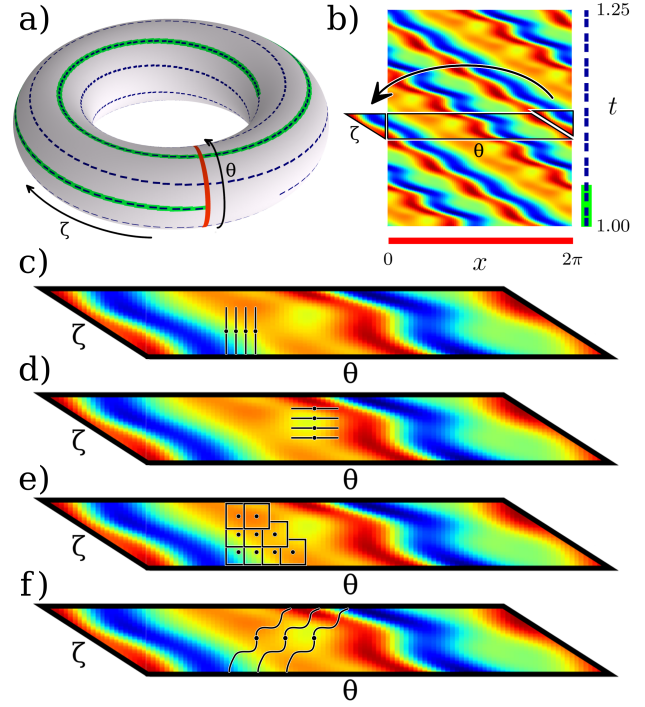

FIG. 3. (a) In a system exhibiting the dynamics of a modulated traveling wave, the state of any point in space and time can be described by its phase with respect to the modulation (ζ) and traveling wave (θ), respectively. Together, these two phases span the surface of a torus. (b) The data set from the Kuramoto-Sivashinsky equation treated in the paper section on observing an attractor is an example of such dynamics, containing the "minimal parallelogram" of the unfolded torus several times over.. (c-f) The points in this minimal parallelogram can be observed in several different ways, either as time strings (c), space strings (d), space-time patches (e) or through the trajectories of active observers (f) that travel with their own, known speeds along the torus surface.

modulated traveling waves [12]. Using $N = 100$ spatial grid points in a finite-difference discretization, we numerically integrate the Kuramoto-Sivashinsky equation using lsode [9, 10] with time step $dt = 0.001$ and sample the data between $t_0 = 1$ and $t_1 = 1.25$ as $T = 500$ snapshots. We are confident that our simulation has reasonably converged on the attractor by $t_0 = 1$, so that our 100 time series of length 400 are representative of it. For completeness, we note here that one period of the modulation contains ≈ 56 discrete snapshots and the "period" of traveling in the spatially periodic domain is known to consist of ≈ 250 snapshots.

FIGURE 6 IN THE PAPER (GAUGE-INVARIANT DIFFUSION MAPS), AND THE KOOPMAN OPERATOR

We illustrate the capability of performing gauge-invariant diffusion map computations on a simple two-dimensional nonlinear dynamical system possessing a stable limit cycle in its deterministic form, yet slightly perturbed by low-amplitude white noise. We start by recording the evolution of its state. Separately, we record observations of this evolution through a known "observation function" that transforms this state nonlinearly (and invertibly). In the original system, for short time intervals, the white noise causes states initialized at a single point (r, θ) to spread out to a normal distribution. In the observed system (here, observed through the observation function $T(r, \theta) = [(r^2)\cos(\theta)/2, \sqrt{r}\sin(\theta)])$, the normal distribution is transformed into an ellipsoid shape, with axes corresponding to the eigenvectors of the noise covariance matrix $C = J(r, \theta)J(r, \theta)^T$–where J is the Jacobian of the transformation function T. Thus, observing several brief bursts of (stochastic) simulation initialized at a particular state-space point through T allows us to estimate the covariance matrix C numerically. The pseudo-inverse of this matrix is used to compute the Mahalanobis-like distance between two data points x, y

(see [13, 14]):

$$d(x, y)^2 = 1/2(x - y)^T(C(x) + C(y))^{-1}(x - y).$$

The "gauge invariant" diffusion maps, based on using this Mahalanobis-like distance for each of the two observations, visibly yields (and can be shown to successfully approximate) *the same* embedding (Fig. 6 in the paper) modulo an orthogonal transformation. Clearly, the (invertible) mapping of any trajectory of the original system to the corresponding trajectory of the transformed (diffeomorphic) one provides a systematic way of matching the two systems, mapping one to the other, and even fusing observations from both of them (see e.g. [15]). This transformation is encoded in the eigenvectors of the gauge invariant diffusion map matrix (that approximate the eigenfunctions of the Laplace-Beltrami operator on the data manifold).

We now discuss an alternative approach to the con-

struction of data-driven conjugacies: one using the Koopman operator framework [16], which also employs a spectral representation. For a discrete dynamical system $x_{k+1} = f(x_k)$, $x \in \mathbb{R}^n$, the Koopman operator $\mathcal{K}$ acts linearly on a space of complex-valued observables $h : \mathbb{R}^n \to \mathbb{C}$. Specifically, for any given observable h, $(\mathcal{K}h)(x) = (h \circ f)(x)$. This property also gives it the name *composition operator*. Indeed, exploring the numerically obtained spectra of the Koopman operator (see [17]) for each of the two deterministic systems, we can also find matching eigenvalues and eigenfunctions, analogous to the gauge-invariant diffusion map case. If ψ is a Koopman eigenfunction of the original system associated to the eigenvalue λ, $\psi = \hat{\psi} \circ T$ defines an eigenfunction $\hat{\psi}$ of the observed system associated to the same eigenvalue. This relation can be used to approximate the function T if enough pairs of eigenfunctions $(\psi, \hat{\psi})$ are given [18]. There are already numerous algorithms available to approximate Koopman eigenfunctions, such as EDMD [19] in the general case and Fourier- or Laplace averaging for limit cycle systems [20]. Here we use the Fourier average defined as $f_\omega(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} (h \circ f^k)(x) \exp(-i\omega k)$, where ω is the period of the limit cycle and f is a generic observable of the system state. The resulting function $f_\omega : \mathbb{R}^n \to \mathbb{C}$ is an eigenfunction of the Koopman operator [20], such that $\mathcal{K}f_\omega = \exp(i\omega)f_\omega$. For continuous systems $\dot{x} = \frac{d}{dt}S^t|_{t=0}(x)$, the Fourier average is defined as $f_\omega(x) = \lim_{t \to \infty} \frac{1}{t} \int_0^t (h \circ S^{t'})(x) \exp(-i\omega t') dt'$. We compute the Fourier averages with the observable $h(x) = x_1 + x_2$ similar to [20], where x_1, x_2 are the Cartesian coordinates of the system. The (complex-valued) eigenfunction f_ω then defines the geometry in the Koopman eigenfunction space (see Fig. 4). The geometry in this space allows us to learn the (approximate) transformation function through a standard optimization formulation, in our case, a two-hidden-layer neural network. In particular, we solve $\min_{T:\mathbb{R}^2 \to \mathbb{R}^2} \|\psi - \phi \circ T\|_{L^2(\mathbb{R}^2)}^2$ by parameterizing T as a two-layer feed-forward neural network with 10 hidden units, i.e. $T(\mathbf{x}) = V\sigma(W\mathbf{x} + \mathbf{b}) + \mathbf{c}$, where $W \in \mathbb{R}^{10 \times 2}$, $\mathbf{b} \in \mathbb{R}^{10}$, $V \in \mathbb{R}^{2 \times 10}$, $\mathbf{c} \in \mathbb{R}^2$, and σ is a sigmoid point-wise nonlinearity $\sigma(\mathbf{x})_i = 1/(1 + e^{-x_i})$. We solve the minimization by standard stochastic gradient descent, with samples uniformly drawn in the set $|\mathbf{x}|^2 < 1.4$. The approximation by a neural network is scalable to larger problems because it circumvents the "curse of dimensionality" associated with the usual function approximation via linear combination of basis functions. Now, the transformation function is embodied in the neural network.
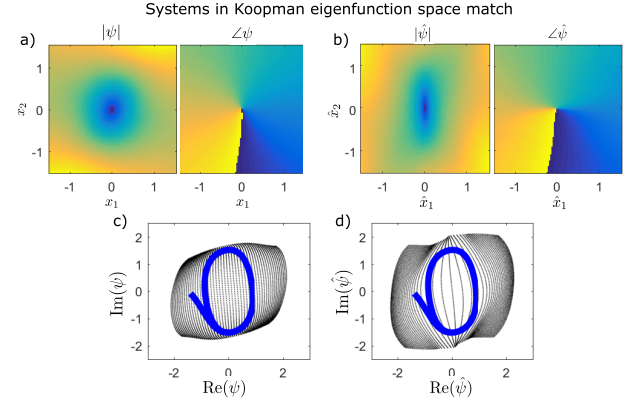

FIG. 4. Eigenfunctions of the Koopman operator are computed by Fourier average, and interpolated with a neural network, for the original system (a), and the transformed system (b). Drawn in the real and complex parts of the eigenfunction, the trajectories of both systems match (plots c,d). The black dots in plots (c,d) visualize the transformation of the plane $[-1,1]^2$ into the Koopman eigenfunction space.

VIDEO 1: TYPE-I CHIMERA STATE IN THE 2-D GLOBALLY COUPLED COMPLEX GINZBURG-LANDAU EQUATION

We integrate the complex Ginzburg-Landau equation with nonlinear global coupling (as described above) for parameters $c_1 = 0.2$, $c_2 = 0.61$, $\nu = 1.5$ and $\eta = 1.0$ on a two-dimensional spatial domain of size 750×750 and fixed time step $dt = 0.01$. The data is sampled from $t_0 = 1900$ to $t_1 = 2000$ as $T = 500$ snapshots. For these parameters, the system exhibits a so-called type-I chimera state, for which the absolute value $|W|$ of the complex system variable is used as z-axis in the right part of video 1.

Following earlier works on chimera states [21], we apply a scaled version of the discrete Laplacian to the data,

$$\begin{aligned} \mathbf{D}f(x,y,t) = \ & f(x + \Delta x, y, t) + f(x - \Delta x, y, t) \\ & + f(x, y + \Delta y, t) + f(x, y - \Delta y, t) \\ & - 4f(x,t) \ . \end{aligned}$$

Creating a histogram of these transformed data may reveal insightful features of different chimera states. In the left part of video 1, a two-dimensional histogram over the angle $\angle D$ of the complex-valued Laplacian and the absolute value $|W|$ of the complex system variable is shown.

A circular one-dimensional color scale is used to encode the individual bins of histogram. Every individual point in physical space (right part of the video) is then colored according to bin (and thus, the color) to which it corresponds. Special care is taken for the coherent regimes of the chimera state, as the value of D vanishes there and the angle $\angle D$ becomes thus not well-defined. Accordingly, all points with amplitude $|W|$ larger 1 are given the same blueish-purple color.

VIDEO 2: 3-D HISTOGRAM OF THE TYPE-I CHIMERA STATE

For the same data obtained in the previous section, a different representation is chosen in video 2. Here, a three-dimensional histogram is created over the absolute value $|W|$ of the complex system variable, and the real and imaginary parts of the discrete Laplacian D. The "height" of the bins in this histograms are encoded in the size and the color of the scattered points in the left part of the video. In addition, projections onto two-dimensional histograms are depicted in the three planes parallel to $(\mathrm{Re}D, \mathrm{Im}D)$, $(\mathrm{Re}D, |W|)$ and $(\mathrm{Im}D, |W|)$.

VIDEO 3: TYPE-I CHIMERA STATE IN A STUART-LANDAU ENSEMBLE

Video 3 shows the data of a chimera state arising in an ensemble of $N = 10000$ Stuart-Landau oscillators, globally coupled as

$$\partial_t W = - i\nu W - (1 + i\nu)\left(\langle W \rangle_\Sigma - W\right)$$
$$+ (1 + ic_2)\left(\langle |W|^2 W \rangle_\Sigma - |W|^2 W\right),$$

with $\langle W \rangle_\Sigma = 1/N \sum_i W_i$ and $\langle |W|^2 W \rangle_\Sigma = 1/N \sum_i |W_i|^2 W_i$ denoting ensemble averages. It has the property that for the spatially uniform mode, the relation

$$\partial_t \langle W \rangle = -i\nu \langle W \rangle \Rightarrow \langle W \rangle = \eta e^{-i\nu t}$$

holds. Thus, the mean $\langle W \rangle$ is confined to harmonic motion with frequency ν and amplitude η [11].

The system is integrated using lsode [9, 10] with fixed time step $dt = 0.01$ at parameter values $c_2 = 0.58$, $\nu = 1.49$ and $\eta = 1.02$, and the data is sampled from $t_0 = 500$ to $t_1 = 525$ as $T = 500$ snapshots.

The real parts of the individual oscillators are depicted on the right, and a two-dimensional histogram over the real and imaginary parts of the complex system variable W in a co-rotating frame, rotating with the frequency of the ensemble mean, ν, on the left. Moreover, the individual bins of histogram are encoded with a circular one-dimensional color scale, corresponding to the angle $\angle W$. Every individual oscillator in the right part of the video is then colored according to the bin (and thus, the color) to which it corresponds.

VIDEO 4: TYPE-I CHIMERA STATE IN THE 1-D GLOBALLY COUPLED COMPLEX GINZBURG-LANDAU EQUATION

This video illustrates the evolution of characteristics on an attractor: The temporal evolution of observables for fixed physical space (on the right) compared to how

physical space "visits" different observable values as time passes (on the left).

In order to generate the data for video 4, we integrate the complex Ginzburg-Landau equation with non-linear global coupling as described above, but on a one-dimensional spatial domain of length $L = 1000$. The simulation is carried out for parameters $c_1 = 0.2$, $c_2 = 0.61$, $\nu = 1.5$ and $\eta = 1.0$ with a fixed time step $dt = 0.01$. The data is sampled from $t_0 = 2500$ to $t_1 = 2530$ as $T = 300$ snapshots. For these parameters, this system exhibits a so-called type-I chimera state, for which the real value of the complex system W variable is shown as a function of space in the right part of video 4.

In order to create the two other parts of the video, we apply a scaled version of the 1-D discrete gradient, defined as

$$\mathbf{d}_\mathrm{x} f(x,t) = f\left(x + \Delta x, t\right) - f\left(x - \Delta x, t\right).$$

In the left part of video 4, a two-dimensional histogram over the real part $\mathrm{Re}d_\mathrm{x}$ of the complex-valued gradient and the absolute value $|W|$ of the complex system variable is shown. A projection of this histogram, showing all non-empty bins, is shown in the form of several black lines on the plane parallel to $(\mathrm{Re}d_\mathrm{x}, |W|)$ (below the histogram itself). The resulting long, "snake-like" self-intersecting black curve is parametrized by physical space - its evolution shows which observable values are realized (are "visited" by physical space) as time evolves.

In the middle part of the video, the last 15 such projections are plotted above each other, with the newest projection shown at the top, and every projection being moved one step downward every frame increment of the video.

Finally, a circular one-dimensional color scale is used to encode the individual bins of the histogram in the left part of the video. Every individual points in physical space (right part of the video) is then colored according to the bin (and thus, the color) to which it corresponds [cite vid_4.avi].

All figures are generated using matplotlib [22].

---

[1] Perron O (1907) Zur Theorie der Matrices. *Mathematische Annalen* 64(2):248–263.

[2] (2017) Reproduced from The Astronomical Almanac Online and produced by the U.S. Naval Observatory and H.M. Nautical Almanac Office.

[3] Shraiman B et al. (1992) Spatiotemporal chaos in the one-dimensional complex Ginzburg-Landau equation. *Physica D: Nonlinear Phenomena* 57(3):241–248.

[4] Cox S, Matthews P (2002) Exponential Time Differencing for Stiff Systems. *Journal of Computational Physics* 176(2):430–455.

[5] Rubin J, Terman D (2002) Synchronized activity and loss of synchrony among heterogeneous conditional os-

6

cillators. *SIAM Journal on Applied Dynamical Systems* 1(1):146–174.

[6] Laing CR, Zou Y, Smith B, Kevrekidis IG (2012) Managing heterogeneity in the study of neural oscillator dynamics. *The Journal of Mathematical Neuroscience* 2(1):1–22.

[7] Choi M, Bertalan T, Laing C, Kevrekidis I (2016) Dimension reduction in heterogeneous neural networks: Generalized Polynomial Chaos (gPC) and ANalysis-Of-VAriance (ANOVA). *The European Physical Journal Special Topics* 225(6):1165–1180.

[8] Laing CR, Rajendran K, Kevrekidis IG (2012) Chimeras in random non-complete networks of phase oscillators. *Chaos* 22(1):013132.

[9] Hindmarsh A (1986) *LSODE. Ordinary Differential Equation System Solver.*

[10] Radhakrishnan K, Hindmarsh AC (1993) *Description and use of LSODE, the Livemore Solver for Ordinary Differential Equations.*

[11] Schmidt L, Schönleber K, Krischer K, García-Morales V (2014) Coexistence of synchrony and incoherence in oscillatory media under nonlinear global coupling. *Chaos* 24(1).

[12] Brown HS, Kevrekidis IG (1996) Modulated traveling waves for the Kuramoto-Sivashinsky equation in *Pattern Formation: Symmetry Methods and Applications (Fields Institute Communication).* (American Mathematical Society, Providence, RI), Vol. 5.

[13] Nadler B, Lafon S, Coifman RR, Kevrekidis IG (2006) Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis* 21(1):113–127.

[14] Singer A, Coifman RR (2008) Non-linear independent component analysis with diffusion maps. *Applied and Computational Harmonic Analysis* 25(2):226–239.

[15] Dsilva CJ, Talmon R, Rabin N, Coifman RR, Kevrekidis IG (2013) Nonlinear intrinsic variables and state reconstruction in multiscale simulations. *The Journal of Chemical Physics* 139(18):184109.

[16] Koopman BO (1931) Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences of the USA* 17(5):315–318.

[17] Budišić M, Mohr R, Mezić I (2016) Applied koopmanism. *Chaos* 22:047510.

[18] Bollt EM, Li Q, Dietrich F, Kevrekidis I (2017) On matching, and even rectifying, dynamical systems through koopman operator eigenfunctions. *arXiv* 1712.07144v1.

[19] Williams MO, Kevrekidis IG, Rowley CW (2015) A data-driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science* 25(6):1307–1346.

[20] Mauroy A, Mezić I, Moehlis J (2013) Isostables, isochrons, and koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena* 261:19–30.

[21] Kemeth FP, Haugland SW, Schmidt L, Kevrekidis IG, Krischer K (2016) A classification scheme for chimera states. *Chaos* 26(9):094815.

[22] Hunter JD (2007) Matplotlib: A 2d graphics environment. *Computing In Science & Engineering* 9(3):90–95.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Felix P. Kemeth

Physik-Department, Nonequilibrium Chemical Physics, Technische Universit\"{a}t M\"{u}nchen, James-Franck-Str. 1, D-85748 Garching, Germany, and Institute for Advanced Study - Technische Universit\"{a}t München, Lichtenbergstr. 2a, D-85748 Garching, Germany.

Sindre W. Haugland

Physik-Department, Nonequilibrium Chemical Physics, Technische Universit\"{a}t M\"{u}nchen, James-Franck-Str. 1, D-85748 Garching, Germany, and Institute for Advanced Study - Technische Universit\"{a}t München, Lichtenbergstr. 2a, D-85748 Garching, Germany.

Felix Dietrich

The Department of Chemical and Biological Engineering - Princeton University, Princeton, NJ 08544, USA, and Department of Chemical and Biomolecular Engineering / Department of Applied Mathematics and Statistics, Johns Hopkins University/JHMI.

Tom Bertalan

The Department of Chemical and Biological Engineering - Princeton University, Princeton, NJ 08544, USA.

Kevin Höhlein

Physik-Department, Nonequilibrium Chemical Physics, Technische Universit\"{a}t M\"{u}nchen, James-Franck-Str. 1, D-85748 Garching, Germany.

Qianxiao Li

Institute of High Performance Computing, 1 Fusionopolis Way, #16-16 Connexis North, Singapore 138632, Singapore.

Erik M. Bollt

Department of Mathematics, and Department of Electrical and Computer Engineering, Clarkson Center for Complex Systems Science, Clarkson University, Potsdam, NY 13699-5815, USA.

Ronen Talmon

Department of Electrical Engineering - Technion – Israel Institute of Technology, Technion City, Haifa, Israel 32000.

Katharina Krischer

Physik-Department, Nonequilibrium Chemical Physics, Technische Universit\"{a}t M\"{u}nchen, James-Franck-Str. 1, D-85748 Garching, Germany.

Ioannis G. Kevrekidis

Institute for Advanced Study - Technische Universit\"{a}t München, Lichtenbergstr. 2a, D-85748 Garching, Germany, and The Department of Chemical and Biological Engineering - Princeton University, Princeton, NJ 08544, USA, and Department of Chemical and Biomolecular Engineering / Department of Applied Mathematics and Statistics, Johns Hopkins University/JHMI.