Data-Driven Methods for Physics-Constrained Dynamical Systems

Daniel Dylewsky

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

J. Nathan Kutz, Chair

Gerald Seidler

Gordon Watts

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Data-Driven Methods for
Physics-Constrained Dynamical Systems

Daniel Dylewsky

Chair of the Supervisory Committee:
Professor J. Nathan Kutz
Applied Mathematics

As the availability of large data sets has risen and computation has become cheaper, the field of dynamical systems analysis has placed increased emphasis on data-driven numerical methods for diagnostics, forecasting, and control of complex systems. Results from machine learning and statistics offer a broad suite of techniques with which to approach these tasks, often with great efficacy. With respect to time series data gathered from sequential measurements on a physical system, however, these generic methods often fail to account for important dynamical properties which are obscured if the data is treated as a collection of unordered snapshots without attention to coherence phenomena or symmetries. This thesis presents three methodological results designed to address particular problems in systems analysis by taking a physics inspired, dynamics focused approach. Chapter 3 offers a method for decomposition of data from systems in which different physics phenomena unfold simultaneously on highly disparate time scales by regressing separate local dynamical models for each scale component. Chapter 4 presents a novel representation for complex multidimensional time series as superpositions of simple constituent trajectories. It is shown that working in this representation, a large class of nonlinear, spectrally continuous systems can be effectively reproduced by actuated linear models. Finally, Chapter 5 introduces a dynamical alternative to existing methods for stability analysis of networked power systems.

Instead of employing graph theory techniques directly on the topological structure of the power grid in question, a phenomenological graph representation learned directly from time series data is shown to offer greater practical insight into the structural basis for failure events. Taken together, these results contribute to a larger push toward effective data-driven analysis of physical systems which takes explicit account for geometry, scale, and coherence properties of observed dynamics.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I want to express my gratitude in particular to my advisor, Nathan Kutz, for his support and enthusiasm for my work over the past four years. He has been the source of a wealth of indispensable guidance, both academic and professional, for as long as we have worked together. In addition to Nathan I have been fortunate to work with a number of excellent collaborators, both at the UW and elsewhere. I want to specifically acknowledge Steve Brunton and Eurika Kaiser, who have contributed a great deal to my recent work, Alex Tartakovsky, Xiu Yang, and David Barajas-Solano of PNNL, Molei Tao, and Nina de Lacy. I am also grateful to all the current and former members of the Kutz and Brunton research groups who have offered a welcoming and stimulating work community, and to the staff of the Physics and Applied Mathematics departments who have made this experience possible. I thank my doctoral committee for all the effort they have dedicated on my behalf, and in particular my readers Gordon Watts and Jerry Seidler for their time and attention. Lastly, I want to express gratitude to all of my family and friends who have sustained me through my time as a graduate student, and in particular to my parents for their love and support.

Chapter 1

INTRODUCTION

The modern era of computation has seen a tremendous proliferation of methodological development in the fields of statistical learning and data science, due largely to the newfound availability of extensive data sets in almost every field of scientific inquiry. Among the many diverse applications for these tools, there is a growing interest in how they might be employed on sequential data obtained by repeated measurement of some dynamically evolving system. Whereas time series analysis has historically focused primarily on statistical profiling of observed data, advances in machine learning have lately shifted attention toward the harder problem of extracting underlying dynamical models directly from measurements. The successful discovery of governing laws is highly desirable, even when they only approximate the true dynamics (offering e.g. a coarse-grained representation or a time-local description), because it opens the door to a suite of diagnostic, forecasting, and control techniques rooted in dynamical systems theory.

This thesis presents three novel methods in the field of data-driven systems analysis. All of them leverage time series data for what might be termed a "dynamics-first" analytic approach: rather than repurposing more generic machine learning approaches, attention is focused on building and applying algorithms which take into account defining dynamical properties of the data. Time series measurements on physical systems commonly exhibit continuity, low-rank spatial structure, or sparse Fourier representation. Moreover, they can be framed as the output of some (usually unknown) generator function expressible by a differential equation. More general methods in statistical learning and data science have a great deal to offer in this domain, but they are unnecessarily hamstrung if they are not

further tailored to take advantage of these highly constraining properties.

1.1 Simplified Representations in Physics

Understanding of the various symmetries that characterize our universe and subsystems that it contains is central to the underpinning theories of physics. The invariance of physical properties under transformations of translation, rotation, charge inversion, etc. (and the conservation laws which arise from these symmetries) present crucial constraints to the mathematics that govern all observed behavior. One direct consequence of these constraints is order reduction in explanatory models: a system with n degrees of freedom subjected to energy conservation admits a reduced description of $n-1$ variables. A dimensionally-reduced representation can be exact or approximate, depending on the strictness of the generating symmetry. Thermodynamics, for example, seeks highly simplified models based on statistical properties which hold in the large-n limit. A system containing on the order of Avogadro's number of particles can be described quite accurately by the single-line ideal gas law $pV = Nk_BT$; the relevant conservation property here is not an ironclad physical law but rather a set of statistical assumptions which hold under the conditions of thermodynamic equilibrium. As such the simplified equation of state cannot be expected to give a complete account of the physics of the system, but for many applications it provides a useful and fully adequate description.

Physicists' preoccupation with identifying simplifying dynamical representations for complex phenomena also extends beyond dimensionality reduction. In many cases the right coordinate transformation can greatly simplify a model even without altering the number of free parameters it contains. The 2nd-century CE Ptolemaic model of the solar system, in which planets follow nested circular epicycles centered on the Earth, was remarkably successful as a predictive tool. It was eventually supplanted some 14 centuries later by the heliocentric Copernican theory, which is now heralded as a revolutionary advance in astronomy. But the Copernican model was not superior to its predecessor in computational accuracy; even Kepler acknowledged that the two were "for practical purposes equivalent within a hair's breadth"

[196]. The great triumph of the new theory was its naturalness. Where Ptolemy was forced to introduce unintuitive ad hoc modifications to his laws to fit the apparent retrograde motion of planets relative to the Earth, Copernicus offered clean elliptical trajectories simply by changing his observational reference frame. Both models were correct, in the sense that their predictions could later be reproduced by Newtonian mechanics carried out in their respective frames, but it is unlikely that Newton's laws could ever have been deduced from Ptolemy. The notion of naturalness that distinguishes these theories is difficult to rigourously define, but it offers a critical guiding principle in the path from empirical observation to explanatory theory.

1.2 Data-Driven Discovery of Simplified Representations

The lessons of this story have become more salient with the advent of modern computational methods. Data on evolving physical systems can be collected in quantities much too great for traditional methods of analysis. Increasingly, we turn to automated techniques to identify simplifying structure, symmetries, and governing laws with little to no human supervision. Implementing this programmatically requires attention to the two key aforementioned insights from physics: first, that the right coordinate transformation can bring to light the obscured underlying structure of complex data, and second that a given system can admit multiple representations which capture its properties from different scales or perspectives. Many-body quantum mechanics is in principle capable of describing global weather patterns, but in practice it is clearly not the right tool for the job: different dynamical representations are suited to different purposes. As successful as the field of physics has been in producing governing models for all manner of systems, modern computational methods have engendered new objectives for which these models are not well adapted. Complex fluid flows can be simulated accurately by numerical integration of the Navier-Stokes equations, but this is often extremely computationally demanding. Burgeoning interest in online data assimilation, forecasting, and control methods demands new reduced-order representations for these problems to allow for real-time simulation.

The interplay between practical use case and mathematical representation is of central importance to data-driven discovery methods. A low-dimensional approximation to observed dynamics can be thought of as a delineation between those degrees of freedom which are relevant to a given question and those which are not. The user's role is to programmatically distinguish between the two. In a fluids problem, where high-dimensional input data might consist of a velocity field sampled on densely spaced grid points in space and time, a useful reduced representation will not be achieved by a simple global affine coordinate transformation à la Copernicus. If there is a dynamical simplification to be found which preserves the important physics of the system, it will presumably lie in recurring spatiotemporally coherent structures in the flow field which can be shown to interact in some reproducible fashion. Furthermore a reduced coordinate system should naturally encode the symmetries of the original physics.

The field of data science, organized around the task of automated learning directly from data, dates back at least to 1962, when John Tukey called for a paradigm shift in statistics research to make use of the increasingly capable computer systems of the time [192, 53]. In the decades since, statistical and machine learning methods have become ubiquitous in physics. These approaches are leveraged for model discovery, signal processing, experiment design, sensor placement, and control applications, among many others. They have particular currency in problems involving high-dimensional, multiscale, and nonlinear systems. Learned classifiers are used to reconstruct jets in particle physics, neural networks are used to understand and simulate many-body quantum systems, material properties are programatically inferred from molecular chemistry, and turbulent fluid flows are controlled using learned actuators [111, 34]. Learning methods range from simple principal component decompositions and linear regression analysis to more sophisticated modern techniques such as genetic algorithms, support vector machines, regularized and sparse regression models, and artificial neural networks. Central to all of these approaches is the premise that the data in question exhibits underlying structure which, under proper coordinate transformation, can be leveraged to obtain a simpler or more useful representation.

1.3 Organization and Contributions

The work presented in this thesis seeks to bring these insights to bear on problems with properties which make them particularly resistant to certain analytic or numerical methods. In each case, dynamical structure is inferred directly from time series data and leveraged into an alternate representation designed with a practical objective in mind. The results are organized in sections as follows: Chapter 2 offers a brief methodological overview of Koopman theory and Dynamic Mode Decomposition, which form the basis of much of the work of this thesis. Chapter 3 presents a novel method for data-driven decompositions of systems in which dynamics unfold simultaneously on multiple highly disparate time scales. This is a property which seriously undermines the efficacy of most numerical methods for diagnostics and forecasting, so successful scale separation enables the use of many techniques which would otherwise fail on such data. The physical intuition that separate time scales exhibit distinct spatially coherent structure and that dynamical coupling between time scales is often very weak is essential to the success of the algorithm. Chapter 4 gives a highly generalizable approach to discovering linear dynamical representations for nonlinear systems using time-delay embedding. Building on previous work in the field of Koopman theory, this result presents an intuitive visual interpretation for the delay coordinate basis in which models are constructed and shows how these models can be extended to reproduce arbitrary nonlinear dynamics. Linear representation enables the use of many linear algebraic methods for spectral analysis, state prediction, and control which could not otherwise be implemented. Finally, Chapter 5 introduces a novel method for analyzing cascading instability events on networked power systems. Previous work on the subject has made use of results from graph theory to understand the complex and unpredictable behaviors that arise from large and asymmetric electrical grids, but this approach on its own has been shown to be largely insufficient in accounting for the systems' actual observed activity. The technique presented here seeks to integrate those network-topological results with a data-driven method which constructs an alternative graphical representation for the system based on phenomenological analysis of

the time series data in which cascading instability events play out.

Chapter 2

BACKGROUND: DYNAMIC MODE DECOMPOSITION AND KOOPMAN THEORY

Much of the work presented in this thesis makes use of the Dynamic Mode Decomposition (DMD) algorithm. DMD was first proposed as a method for decomposition of fluid flows [169], but it has since proliferated as a highly useful tool in data-driven analysis of many kinds of time series data [108]. Its practical value lies in its ability to represent complex spatiotemporal signals by a relatively low-dimensional superposition of coherent structures (or "dynamic modes"). Intuitively, it can be thought of as a hybrid of Singular Value Decomposition (SVD) and Fourier analysis. Performed on a data matrix of multivariate measurement snapshots of a system, SVD discovers dominant spatial modes by identifying the directions of maximal variance in the state space. Temporal information is discarded, however: reordering the snapshots would not affect the spatial singular vectors obtained. The Fourier transform accomplishes the opposite: it characterizes the time-frequency content of a signal, but if it is a multidimensional signal whose coordinates have a known spatial relationship to one another this information is ignored. DMD offers a marriage of these methods: each dynamic mode discovered pairs a static spatial vector $\boldsymbol{\phi}$ with a complex-valued temporal frequency ω. These quantities differ from the results one would get from SVD and Fourier decomposition, respectively, because each is regressed with explicit attention to coherency along the other space/time dimension. This chapter offers a brief introduction to DMD and its variants, as well as a discussion of its connection to Koopman theory.

2.1 The DMD Algorithm

Dynamic Mode Decomposition (DMD) is a model regression algorithm which produces a best-fit linear operator to reproduce the dynamics observed in some time series data set [170, 161, 190]. The resulting model offers the benefits of interpretability (via its eigenvalue spectrum and spatial eigenvectors) and future-state forecasting to arbitrary time. Given a set of sequential measurements $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_m] \in \mathbb{R}^{n \times m}$, DMD seeks to solve

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} \tag{2.1}$$

The simplest implementation of the algorithm, known as *exact DMD*, simply seeks a least-squares best fit which minimizes $\|\dot{\mathbf{X}} - \mathbf{A}\mathbf{X}\|_F$, solved by $\mathbf{A} = \dot{\mathbf{X}}\mathbf{X}^\dagger$ ($\dagger$ denotes the Moore-Penrose pseudo-inverse) [189]. Eq. 2.1 makes use of both the state $\mathbf{X}$ and its time derivative $\dot{\mathbf{X}}$. In most circumstances, the latter cannot be directly measured and must therefore by approximated by numerical differentiation. For this reason, exact DMD is often formulated in a discrete time representation instead. This can be accomplished by separating the uniformly-spaced data $\mathbf{X} \in \mathbb{R}^{n \times m}$ into two sequential matrices $\mathbf{X}_1 \in \mathbb{R}^{n \times (m-1)}$ and $\mathbf{X}_2 \in \mathbb{R}^{n \times (m-1)}$:

$$\begin{aligned} \mathbf{X}_1 &= \begin{bmatrix} | & | & & | \\ \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_{m-1} \\ | & | & & | \end{bmatrix} \\ \mathbf{X}_2 &= \begin{bmatrix} | & | & & | \\ \mathbf{X}_2 & \mathbf{X}_3 & \cdots & \mathbf{X}_m \\ | & | & & | \end{bmatrix} \end{aligned} \tag{2.2}$$

The operator $\mathbf{A}$ is then simply the matrix which minimizes the Frobenius norm $||\mathbf{X}_2 - \mathbf{A}\mathbf{X}_1||_F$, solved by $\mathbf{A} = \mathbf{X}_2\mathbf{X}_1^\dagger$. Note that this is equivalent to the continuous-time formulation of Eq. 2.1 if $\dot{\mathbf{X}}$ is approximated by the forward difference formula $\frac{1}{\Delta t}(\mathbf{X}_2 - \mathbf{X}_1)$ and $\mathbf{X}_1$ is substituted for $\mathbf{X}$. The discrete linear map $\mathbf{A}^d$ is then related to the continuous operator $\mathbf{A}^c$ by

$$\mathbf{A}^d = \mathbf{I} + \mathbf{A}^c \Delta t \tag{2.3}$$

The computation of the Moore-Penrose pseudo-inverse is straightforward, but can become prohibitively expensive for large state dimension n. It is therefore common to first project the data into a lower-dimensional space using SVD, and carry out DMD on a low-rank representation of the full data. For a more detailed overview of the method, see Tu et. al. (2014) [190]. A more thorough treatment of this discretization procedure and variations thereof is presented in [190]

2.2 Interpretations and Extensions of DMD

The operator $\mathbf{A}$ that DMD regresses defines a first-order linear differential matrix equation in the state space $\mathbb{R}^n$. Equations of this class generically admit closed-form solutions, which can be written using the eigendecomposition of $\mathbf{A}$:

$$\tilde{\mathbf{x}}(t) = \sum_j b_j \boldsymbol{\phi}_j e^{\omega_j t} \tag{2.4}$$

where $\boldsymbol{\phi}_j$ and ω_j are eigenvectors and eigenvalues of the DMD operator and b_j are scalar weighting coefficients. This illustrates the power of linear model regression: the obtained DMD model admits a closed-form analytic solution which can easily be broken down into simple exponential modes and can be evaluated over any time domain. Space and time dependencies are encoded separately for each mode (in $\boldsymbol{\phi}_j$ and $e^{i\omega_j t}$, respectively), but are regressed jointly. Spatiotemporal coherency is intrinsic to DMD, unlike SVD or Fourier decomposition.

DMD has become a heavily used algorithm in the analysis of spatiotemporal data analysis, with many recent innovations including a sparsity promoting variant [90], an optimized framework [9], extensions to control [157], a Bayesian formulation [182], a variational formulation [147, 146, 11], a higher-order formulation (HO-DMD) [120], a stochastic formulation [203] and a multi-resolution analysis [112, 56].

A well-known deficiency of the exact DMD approach is the adverse effect of the measurement errors (sensor noise) on its performance [9, 82, 51]. Previous studies indicate that presence of sensor noise can negatively influence the computation of eigenvalues and they would be biased, presenting a serious problem for studies that rely upon exact DMD to distinguish between stable and unstable modes. This is primarily because exact DMD treats data sequentially rather than as a whole, and thus favors the forward time direction. Recent studies have addressed this issue and proposed several techniques to mitigate this problem through employing various forms of ensemble averaging, cross-validation, windowing, and rank reduction. However, Hemati et al. [82] showed that the resulting analysis from the aforementioned techniques are subject to systematic bias errors when the measurements are inexact due to sensor noise or other effects.

Many of the results in this thesis make use of a variation on the standard DMD algorithm known as Optimized DMD, which seeks to address these shortcomings. Optimized DMD recasts the minimization problem outlined above as a task of exponential curve-fitting making use of the variable projection method. This comes at the price of convexity, but yields a decomposition which much more faithfully reconstructs the input data series. A full exposition of the Optimized DMD algorithm and its advantages is presented in [9].

2.3 Connections to Koopman Theory

Much of the current interest in the DMD algorithm stems from the observation that it functions as an efficient numerical approximator of the Koopman operator. Koopman theory, first introduced by B. O. Koopman in 1931, states that nonlinear dynamical systems can be reproduced by linear evolution in a space of observables on the state variables [101, 102]. In most cases, this dual representation is infinite dimensional. Consider the system defined by

$$\mathbf{x}_{j+1} = \mathbf{f}(\mathbf{x}_j) \tag{2.5}$$

for some nonlinear function $\mathbf{f} : \mathbb{C}^n \to \mathbb{C}^n$. The Koopman operator $\mathcal{K}$ acts on a space of observables $\mathbf{g}(\mathbf{x}_j) : \mathbb{C}^n \to \mathbb{C}^k$ where k is the (possibly infinite) dimension of the Koopman

invariant space. In this space, the evolution of Eq. 2.5 is expressed by the composition of $\mathbf{g}$ with $\mathbf{f}$:

$$\mathbf{g}(\mathbf{x}_{j+1}) = \mathcal{K}\mathbf{g}(\mathbf{x}_j) = \mathbf{g}\left(\mathbf{f}\left(\mathbf{x}_j\right)\right) \tag{2.6}$$

This confers all the advantages of linear algebraic and spectral analysis that come with linear dynamics, but does so at the price of finite representation. The existence of the Koopman operator which accomplishes this is in some sense trivial: any continuum state space can be discretized onto a grid of infinite resolution, and the nonlinear map which steps forward in time by $\Delta t \to 0$ can then be quite simply represented by an infinite-dimensional binary operator. Though theoretically simple, the practical implementation of this procedure is entirely infeasible. Consequently, Koopman's result was an obscure mathematical curiosity for decades until numerical methods caught up to it.

Beginning in 2004, the works of Mezić et. al. raised the possibility of finite approximations of the Koopman operator, which could reasonably accurately reproduce nonlinear dynamics while retaining all of the analytic benefits of linearity [133, 137]. Following the 2008 introduction of DMD by Schmid and Sesterhenn [169], Rowley, Mezić et. al. observed the Koopman-theoretic underpinning of this result [161]. DMD produces a linear operator which provides the least-squares optimal fit to any time series data, which is precisely the objective of a finite Koopman approximation. In its simplest form DMD acts directly on the measurement space, whereas the Koopman operator is free to act on any number of observable functions *on* the measurement space. DMD applied directly to observation data can be thought of as a regression for $\mathcal{K}$ in the case that $\mathbf{g}$ is the identity on the space of $\mathbf{f}$.

Much of the active methodological development that has taken place since the resurgence of interest in Koopman theory has centered on the discovery of alternate formulations for $\mathbf{g}$ which lift measurements into a higher- (but still finite-) dimensional space that is more amenable to linear representation, even for highly nonlinear dynamics. A variety of approaches have been taken, including constructing libraries of simple monomial functions evaluated on the data and regressed densely [200] or sparsely [30, 27], kernel methods such as

diffusion mapping [74], and deep learning of coordinate transformations [183, 127, 130, 198]. The work presented in Chapter 4 focuses on the use of time-delay embedding, taking inspiration from the seminal Takens' theorem [185].

With all of these promising Koopman approximation techniques being developed, it is also important to acknowledge the limitations of this approach. There are nonlinear systems for which finite linear representations necessarily fail to capture crucial properties of the dynamics. Chaos, for example, is a manifestly nonlinear phenomenon which cannot be reproduced linearly in any finite dimension. This can be understood more generally in the context of spectral theory. Any linear operator $\mathcal{K} \in \mathbb{C}^{k \times k}$ admits at most k distinct eigenvalues, and any non-transient dynamics it generates are guaranteed to be sparsely represented in Fourier space. Many nonlinear systems also generate discrete (point) Fourier spectra, but others produce continuous spectra which densely cover the frequency space. While a finite Koopman representation might plausibly reproduce their true dynamics over some length of time, it will inevitably fail in the long-time limit. Methods to circumvent this issue have sought to extend the DMD approach in a variety of ways. Brunton et. al., for example, showed how an intermittently-applied forcing signal could be used to actuate an otherwise-linear system and faithfully reconstruct the chaotic dynamics of the Lorenz attractor [31]. The algorithms introduced in Chapters 3 and 4 both offer such variations on DMD, representing nonlinear dynamics by piecewise combination of local models or by parametrically forced linear evolution.

Chapter 3

DYNAMIC MODE DECOMPOSITION FOR MULTISCALE NONLINEAR PHYSICS

Physical systems whose dynamics evolve on a broad range of scales simultaneously (spatial or temporal) have been the subject of much study in the development of diagnostic and modeling tools. These multiscale systems are ubiquitous in physics, so there is a great deal of practical interest in methods which are accommodating to scale disparities spanning orders of magnitude. Of particular note are those systems whose behavior can be decomposed into a finite number of discrete scales, as this lends an additional structural constraint which can be exploited in modeling. For instance, atmospheric climate data and/or simulations can be characterized by developing separate models for variations on the order of one day and one year, respectively, and then coupling them. Mathematical methods for exploiting these distinct and disparate scales can greatly simplify the problem of state estimation and forecasting. The method of dynamic mode decomposition (DMD) is extended in order to characterize multiscale physics and their coupling dynamics, showing that such a data-driven strategy provides a viable and adaptive strategy for diagnostics and dynamical modeling.

The task of identifying distinct multiscale temporal physics directly from data in a way that allows the signal to be decomposed into its constituent scale-separated components is a subject of ongoing investigation. Well-established methods use Fourier- and wavelet-based techniques to separate coarse-grain and fine-grain features in space or time (but generally not both at once) [50, 69, 199]. These approaches, though often useful, are purely diagnostic. They do not directly produce dynamical models from data. Moreover, their focus on exclusively temporal (or exclusively spatial) coherencies limits their utility as precursors to model discovery for state estimation and forecasting. Regardless, such techniques form the

mathematical basis of *multiresolution analysis* (MRA) [129, 111] which provides a rigorous foundation for multiscale decomposition.

To address the dynamical limitations of MRA, researchers have put forth a number of equation-free, data-driven modeling techniques tailored to multiscale spatio-temporal systems. Indeed, there is a significant body of research focused on modeling multiscale systems and linking scales: notably the heterogeneous multiscale modeling framework, equation-free methods, and structure preserving versions known as FLAVORs [97, 59, 58, 186]. Additional work has focused on testing for the presence of multiscale dynamics so that analyzing and simulating multiscale systems is more computationally efficient [66, 67]. Many of the same issues that make modeling multiscale systems difficult can also present challenges for model discovery and system identification [35]. This motivates the development of specialized methods for performing model discovery on problems with multiple time scales, taking into account the unique properties of multiscale systems. A purely data-driven approach was recently introduced by Kutz. et. al.[107] which recursively applies DMD to build closed-form linear models to approximate dynamics at all scales simultaneously. DMD was first proposed as a decompositional technique for complex fluid flows [171, 161, 168], but it has since been adopted more widely as a method for finite approximation of the Koopman operator in a large variety of data sets [190]. DMD produces a linearized model for a (generically nonlinear) data set. It can be thought of as a best-fit approximation of a signal generated by a linear combination of static spatial modes whose time-varying weights follow complex exponential trajectories of oscillation, growth, or decay. The technique proposed by Kutz et. al., dubbed *multi-resolution Dynamic Mode Decomposition* (mrDMD), builds on MRA wavelet techniques by recursively subdividing the data set to access different regimes of the time-frequency domain. The length of the window over which DMD is applied is repeatedly halved, and the most salient components of each iteration are interpreted as a simplified local model for the dynamics at that scale.

Decomposition of a multiscale signal can also be cast as a Blind Source Separation (BSS) problem by treating each time scale as an independent source contributing to the composite

signal being measured. Typical methods for BSS include Principal Component Analysis (PCA) and Independent Component Analysis (ICA). A comparison of these methods to DMD is presented in [106]. When the source signals occupy fairly narrow frequency bands, as is assumed to be the case here, DMD is shown to drastically outperform the other techniques. For this reason it is an obvious candidate for the decompositional method used in the sliding window framework presented here.

This work aims to extend and generalize the mrDMD algorithm. The essential insight of mrDMD is the sensitivity of results to duration of the input signal: given a time series containing dynamics on widely-varying time scales, the eigenfrequencies obtained by DMD could reflect any of these time scales depending on the duration and resolution of sampling. The window lengths tested in mrDMD are limited to some base time span and power-of-two subdivisions thereof. This can be problematic in systems whose multiscale frequency content does not follow that pattern—the ability of DMD to robustly identify a persistent component at a particular time scale turns out to be fairly sensitive to window size. The simple halving scheme could easily fail to resolve a component whose characteristic time scale falls between those given by powers of two. This problem is addressed by 1) implementing a protocol using sliding, overlapping windows on the data set to generate spectral bands of DMD eigenvalues, and 2) developing a diagnostic to use the narrowness of these bands to tune the window size for optimal resolution of a particular scale component. The method proposed here effectively identifies and isolates the constituent time scale components of two test systems. In addition to providing diagnostic information on the frequency content of a signal, it produces 1) faithful reconstructions of each constituent component with minimal cross-pollution between them, 2) closed-form expressions for these reconstructions which can be used for low-cost forecasting at any time scale, and 3) statistics on the parameters of windowed DMD models, whose distributions can be sampled for stochastic ensemble forecasting.

The method presented bears some similarity to the Frequency Map Analysis (FMA) technique often used in analysis of time series generated by nonlinear dynamics [118]. Both seek to identify dominant frequencies in a multivariate signal and fit the data to a linear

combination of sinusoids at these frequencies. However, FMA does this with a static basis of spatial modes (corresponding to the canonical coordinates of Hamiltonian mechanics) whereas the DMD approach allows spatial modes to vary over time using a sliding-window framework. This makes it more versatile in its ability to reconstruct a wide variety of input signals. Moreover, FMA typically restricts its analysis to real-valued frequencies which produce purely sinusoidal dynamics. The DMD method can be similarly constrained, but in general it admits complex-valued frequencies which also allow for exponential growth or decay in its local windowed reconstructions.

The rest of the chapter is outlined as follows: Sec. 3.1 outlines the protocol for the sliding-window scale separation technique and demonstrates it on a simple toy model. Sec. 3.2 briefly discusses the advantages of this method over traditional temporal filtering tools. A fully-fledged recursive, many-scale example using data from a three-body planetary system is presented in Sec. 3.3. The chapter is concluded in Sec. 3.5 with a discussion of theoretical context for the algorithm and possibilities for its future application.

The work presented in this chapter is based on a paper entitled "Dynamic mode decomposition for multiscale nonlinear physics" written in collaboration with Molei Tao of the Georgia Institute of Technology and J. Nathan Kutz of the University of Washington. The paper was published in Physical Review E in 2019 [56].

3.1 Methods: Time Scale Separation using DMD

The decomposition method introduced in this chapter consists of the following steps:

1. A sliding-window implementation of DMD to extract a large number of (complex) frequencies ω_j^k associated with spatially-coherent dynamics in the input time series

2. A clustering algorithm to identify the most highly-represented frequencies in the population of $\left\{|\omega_j^k|\right\}$. These clusters represent the multiple time scale regimes present in the input data

3. Retroactive labeling of modal components of each windowed DMD identified in Step 1 with labelling based on the cluster assignments of their associated frequencies

4. For each distinct scale regime identified in Step 2, a separate DMD reconstruction is produced by summing over the components assigned to that cluster. These reconstructions are produced separately for each iteration of the sliding-window DMD from Step 1

5. A single global reconstruction is produced for each time scale regime by combining weighted contributions from each windowed reconstruction

Note that Steps 1-3 are *offline* computations. Having carried them out on a representative data set for a given system, the clustering results can be used to label new data from the same system.

3.1.1 A Simple Toy Model

The decomposition method is introduced using a simple system with nonlinear dynamics on two distinct time scales. The model is given by the equations

$$\begin{aligned} \dot{v}_1 &= v_2 \\ \dot{v}_2 &= -w_1^2 v_1^3 \\ \dot{w}_1 &= w_2 \\ \dot{w}_2 &= -\epsilon^{-1} w_1 - \delta^{-1} w_1^3 \end{aligned} \tag{3.1}$$

The parameters which set the time scale separation were assigned the values $\delta = 0.25$ and $\epsilon = 0.01$, respectively. The system is initialized at $t = 0$ in the state $(v_1, v_2, w_1, w_2) = (0, 0.5, 0, 0.5)$. Taken alone, the w variables (i.e. the "fast scale") form an undamped Duffing oscillator in which the cubic nonlinearity term can be considered a small perturbation from simple harmonic motion. The v variables (representing the "slow scale") also take the form

of a cubic oscillator (sans linear term), but with a coefficient (w_1^2) which is dependent on the state of the w variables.

This construction separates the fast and slow dynamics for the sake of interpretability. Because no such separation is guaranteed in measurements made on a real multiscale physical system, a random linear mixing is applied to the above coordinates:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \mathbf{Q} \begin{bmatrix} v_1 \\ v_2 \\ w_1 \\ w_2 \end{bmatrix} \tag{3.2}$$

where $\mathbf{Q}$ is a randomly generated 4×4 orthogonal matrix. This system is numerically integrated (using Matlab's `ode45` solver) for a duration of 48 time units with a sampling interval of $\Delta t = 4 \times 10^{-4}$. The results are aggregated into a data matrix $\mathbf{X} \in \mathbb{C}^{N \times M}$ where $N = 4$ and $M = 120,000$.

3.1.2 Sliding Window DMD

The sliding window approach takes advantage of DMD's sensitivity to the duration and sampling rate of the time series input it receives. Consider an N-variable system measured over M time points. An application of DMD to the full data matrix $\mathbf{X}$ would identify a frequency spectrum that would likely look quite different from that of the same algorithm applied to a subset $\bar{\mathbf{X}} \in \mathbb{C}^{N \times W}$, ($W << M$). For the purposes of this investigation, the "correct" sample length is defined by the multiscale properties of the data: it must be long enough to capture variations on the slowest scale, but not so long as to fail to resolve the fastest scale. In the example model, the sample length T_W that most cleanly separated the two distinct time scales was $T_W \approx 2\ T_{\text{fast}}$ and $T_W \approx T_{\text{slow}}/20$ (The approximate periods of oscillation associated with the fast-scale and slow-scale dynamics, respectively). This is illustrated in Fig. 3.1 (though the width of the window drawn has been increased slightly for visual clarity). The step size for the foreward motion of the window is chosen to be much

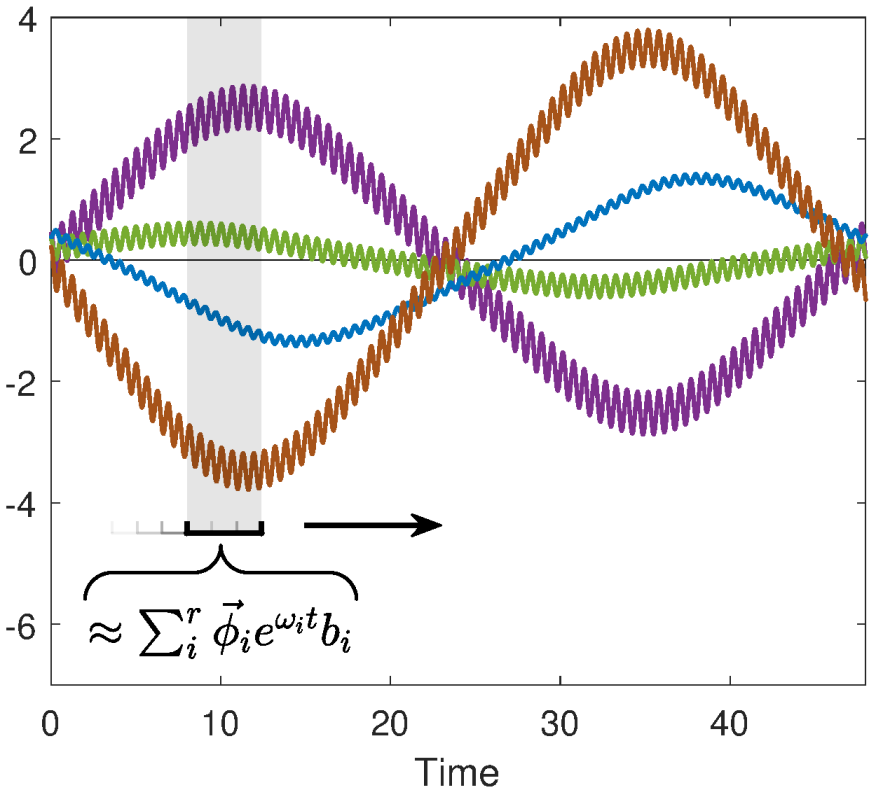

Figure 3.1: Sliding a window filter across a longer time-series data set. For each step of the window's movement, a new DMD result is obtained. Note that the width of the window is such that it contains multiple complete oscillations of the fast scale but only a fraction of a period of the slow scale.

smaller than the width of the window (about 4%) so that any given time point is contained by a large number of windows.

DMD approximates the dynamics contained in each window with a linear operator. The best-fit reconstruction of the windowed data can then be expressed in the standard form for solutions to a first-order linear system, using the eigendecomposition of that operator:

$$\bar{\mathbf{x}}^k(t) = \sum_j^r \boldsymbol{\phi}_j^k e^{i\omega_j^k t} b_j^k + \mathbf{c}_k \tag{3.3}$$

Here k is used to index the steps of the sliding window (i.e. $\bar{\mathbf{x}}^k(t)$ is the time series contained

by the kth window position). j indexes the eigenvalues (ω) and eigenvectors ($\boldsymbol{\phi}$) of the linear DMD operator (numbered from 1 to r for a rank-r decomposition). $\mathbf{c}_k$ denotes the (constant) mean of the input data, which must be reintroduced only if it was subtracted off before applying the DMD algorithm.

Note that while the form presented in Eq. (3.3) is manifestly complex, applying DMD to a real-valued input signal leads it to identify oscillatory modes in complex conjugate pairs whose imaginary components cancel each other out in the reconstruction process. In the discussion that follows, all $\bar{\mathbf{x}}^k(t)$ can be taken to be real.

3.1.3 Frequency Clustering

The spectral results of the sliding-window DMD procedure are then clustered to discover their dominant frequency content. Concatenating the set of all $|\omega_j^k|^2$ into a single vector, cluster centroids are obtained using the k-medians algorithm [138] (i.e. k-means using an L_1 distance metric to limit the influence of outliers). The choice to cluster in $|\omega_j^k|^2$ rather than $|\omega_j^k|$ has the effect of inflating the separation of higher frequencies and compressing that of lower frequencies. For this example, this has no practical effect. In Sec. 3.3, however, a recursion method is introduced which uses multiple clustering iterations working sequentially from the fastest time scales to the slowest. In this case, improved differentiation between higher frequencies is an asset, and the compression of the lower frequencies is inconsequential because they can simply be dealt with on the next iteration.

Plotted in Fig 3.2 are the spectra obtained by applying DMD to each windowed subset of the sample data. The multiscale structure is immediately obvious: there are two strong bands at $|\omega|^2 \approx 100$ and $|\omega|^2 \approx 1$. Although there are a number of outliers from these dominant bands (particularly in regions where the slow-scale dynamics are relatively flat), the full set of $\left\{|\omega_j^k|^2\right\}$ is unambiguously peaked about two centroids (depicted in Fig. 3.3). The clustering results are used to retroactively label each frequency (and, by association, each windowed DMD mode) based on its k-medians categorization.

A brief digression regarding clustering parameterization: k-medians requires that the

number of clusters be supplied *a priori*. In the above case the choice of $k = 2$ seemed quite obvious from the band structure of Fig. 3.2. But the window size was specifically tuned to be sensitive to the fast-scale dynamics at $|\omega|^2 \approx 100$ (as described in Sec. 3.1.2). If there had been a third, even slower time scale (e.g. at $|\omega|^2 \approx 0.01$), it would have gone entirely unnoticed—fully subsumed by Cluster #1—resulting in an incomplete separation of scales. This apparent failure is resolved by the recursion method which is outlined in Sec. 3.3. In the meantime, it bears noting that for any system with persistent dynamics on multiple, discrete time scales, each of these scale components can be resolved into a clean frequency band with an appropriate choice of DMD window length. Thus, for a given window size, choosing the number of clusters for k-medians can be easily accomplished by visual inspection or using the statistical cluster ennumeration method of your choice. Fully isolating all time scale components may require multiple clusterings at multiple window lengths (see Sec. 3.3.2), but the choice of k for each of these will be independent and greatly simplified by a well-delineated band structure.

3.1.4 Scale-Separated Reconstruction

Global Reconstruction from Windowed Results

Each windowed dynamic mode decomposition admits a linearized reconstruction of the signal, given in Eq. (3.3). The reconstruction $\bar{\mathbf{x}}^k(t)$ (green) is overlaid on the original data $x^k(t)$ (black) in Fig. 3.4. Results are only plotted for four non-overlapping windows (demarcated by vertical dotted lines) to avoid visual clutter.

$$\bar{\mathbf{x}}^{\text{global}}(t) = \frac{\sum_k e^{-(t-\mu_k)^2/\sigma^2} \bar{\mathbf{x}}^k(t)}{\sum_k e^{-(t-\mu_k)^2/\sigma^2}} \tag{3.4}$$

It is evident from the plot that the reconstructions tend to diverge from the true signal near the edges of each window. Converting an ensemble of windowed reconstructions into a single global reconstruction calls for a linear combination of results from all windows that

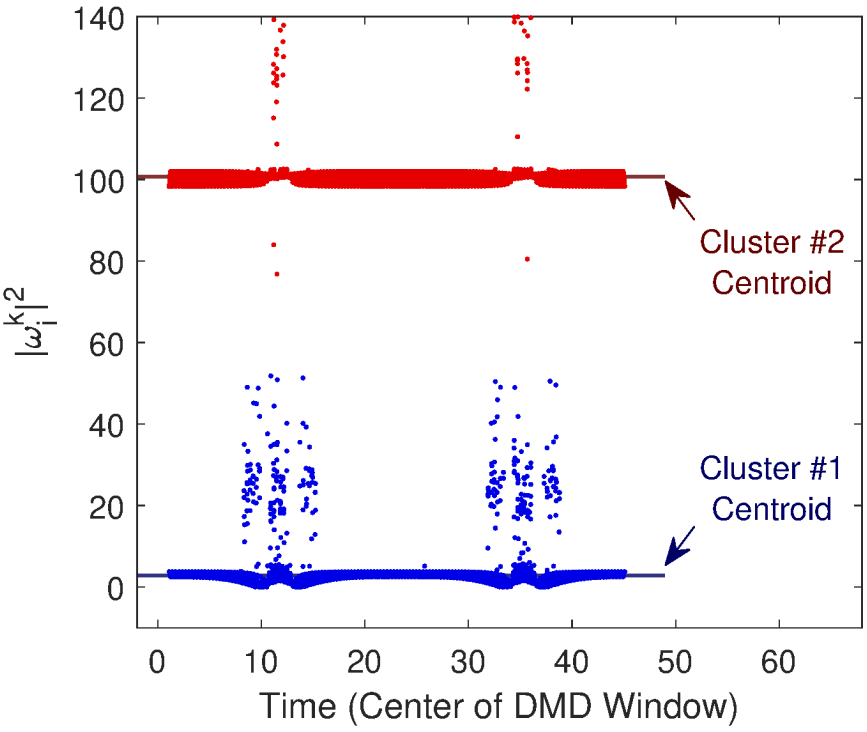

Figure 3.2: Spectra of the (modulus squared) frequencies obtained by the sliding-window DMD procedure. Frequencies are plotted at the midpoints of the windows from which they were computed. Colors denote the cluster labels assigned to each point retroactively.

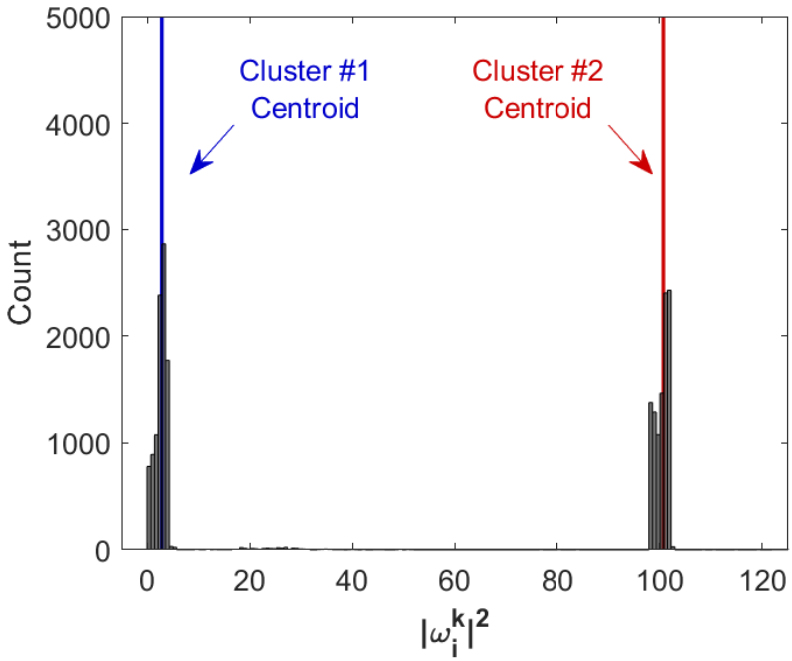

Figure 3.3: Histogram of all $|\omega_j^k|^2$ with the k-medians cluster centroids overlaid in color. Note that the outliers visible in Fig. 3.2 are vastly outnumbered by in-band data points

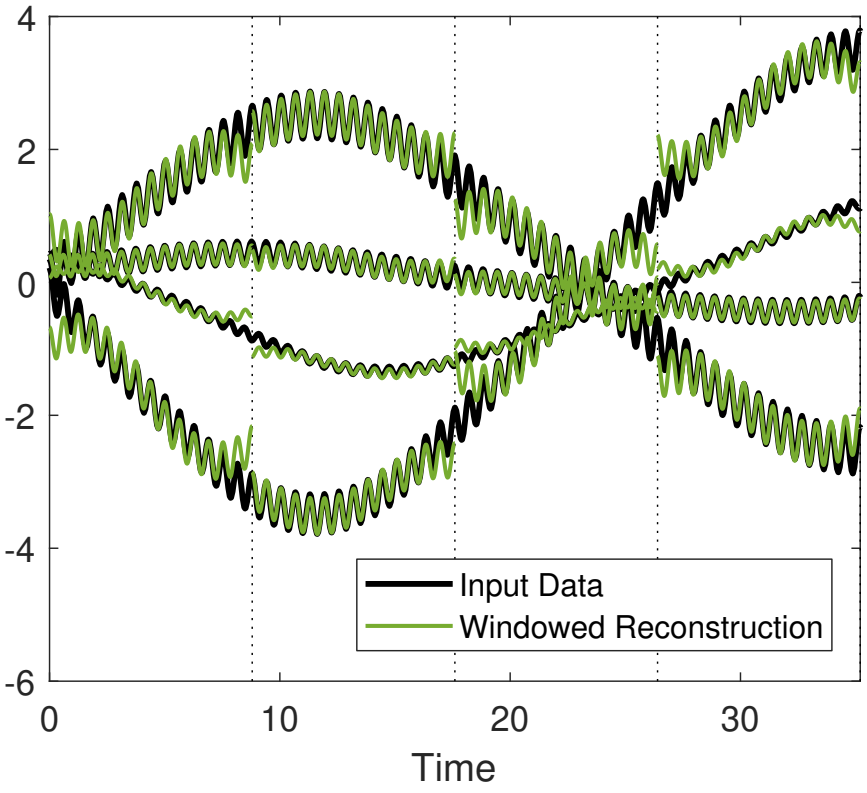

Figure 3.4: DMD reconstructions $\bar{\mathbf{x}}^k(t)$ plotted over input data $x(t)$ for four non-overlapping windows (delineated by dotted lines)

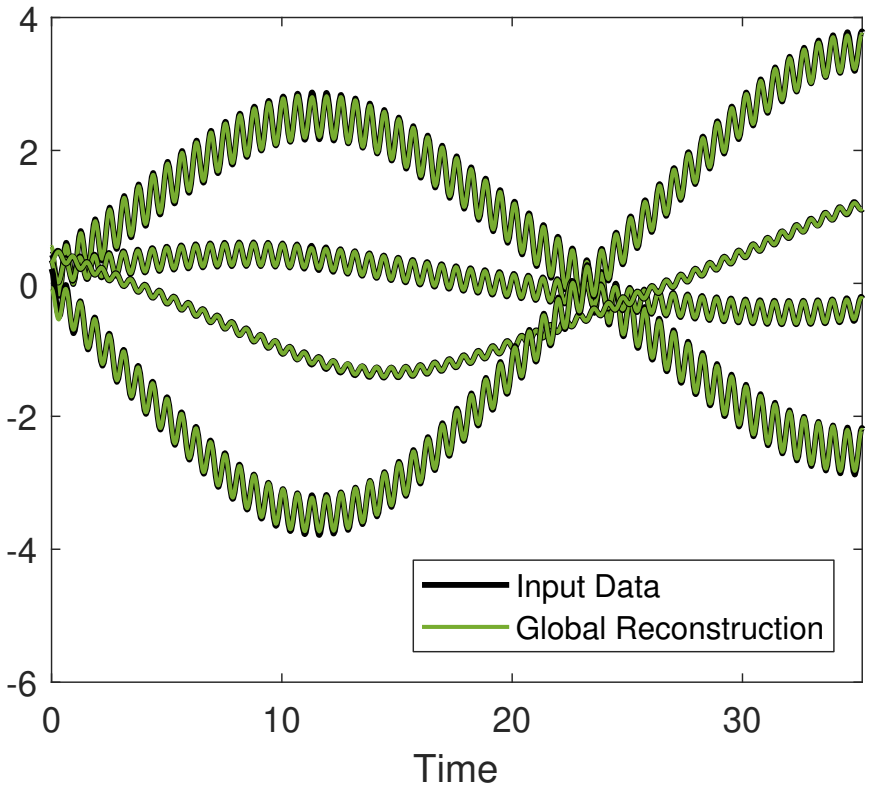

Figure 3.5: Global DMD reconstruction $\bar{\mathbf{x}}^{\text{global}}(t)$ plotted over the input data $x(t)$

contain a given point. But clearly they should not be weighted equally—to estimate the system state at $t = 12$, for example, the result from the window centered on $t = 12$ is more likely to be accurate than the result from a window whose boundary lies near $t = 12$. To address this issue, each windowed result is weighted using a Gaussian centered on the midpoint of the window μ_k and with standard deviation σ equal to one eighth of the window's width (see Eq. 3.4). The denominator simply acts as a normalization factor ensuring unit net contribution to every time point. The result of this method, plotted in Fig. 3.5, hews closely to the ground truth signal for the full duration of the simulation.

Separation of Time Scales

Having labeled the individual modes according to the clustering results, it is straightforward to separate this summation to obtain separate reconstructions for each identified time scale:

$$\begin{aligned} \bar{\mathbf{x}}^k_{\text{slow}}(t) &= \sum_{i \in \{\text{slow}\}}^{r} \boldsymbol{\phi}^k_j e^{i\omega^k_j t} b^k_j \\ \bar{\mathbf{x}}^k_{\text{fast}}(t) &= \sum_{i \in \{\text{fast}\}}^{r} \boldsymbol{\phi}^k_j e^{i\omega^k_j t} b^k_j \end{aligned} \tag{3.5}$$

In the same fashion, Eq. (3.4) can be separated to produce fast- and slow-scale global reconstructions (plotted in Fig. 3.6). This result has a number of desirable properties:

- Fidelity: The separated reconstructions sum to a very close approximation of the original time series (Fig. 3.5)

- Excellent time-scale separation: There is very little mixing of frequency content between the identified regimes. Plots of the signals' power spectra (Fig. 3.7) show that the separated reconstructions closely mirror the spectral content of the input signal near their respective peaks, and they contribute very little elsewhere.

- Spatial interpretability: Unlike other frequency filtering approaches, sliding-window DMD identifies spatial modes corresponding to dynamics of a given frequency. Concatenating the results from all windowed decompositions, one can construct time series of (complex) mode vectors which are already labeled by time scale category using the clustering results. Identifying patterns in the evolution of these modes presents a promising approach for model-building or forecasting

- Closed analytic form: The reconstructions $\bar{\mathbf{x}}^{\text{global}}_{\text{slow}}(t)$ and $\bar{\mathbf{x}}^{\text{global}}_{\text{fast}}(t)$ are simply weighted sums of exponentials. They therefore represent models for scale-separated variables whose values can be computed directly for arbitrary t (without need for any iterative integration scheme).

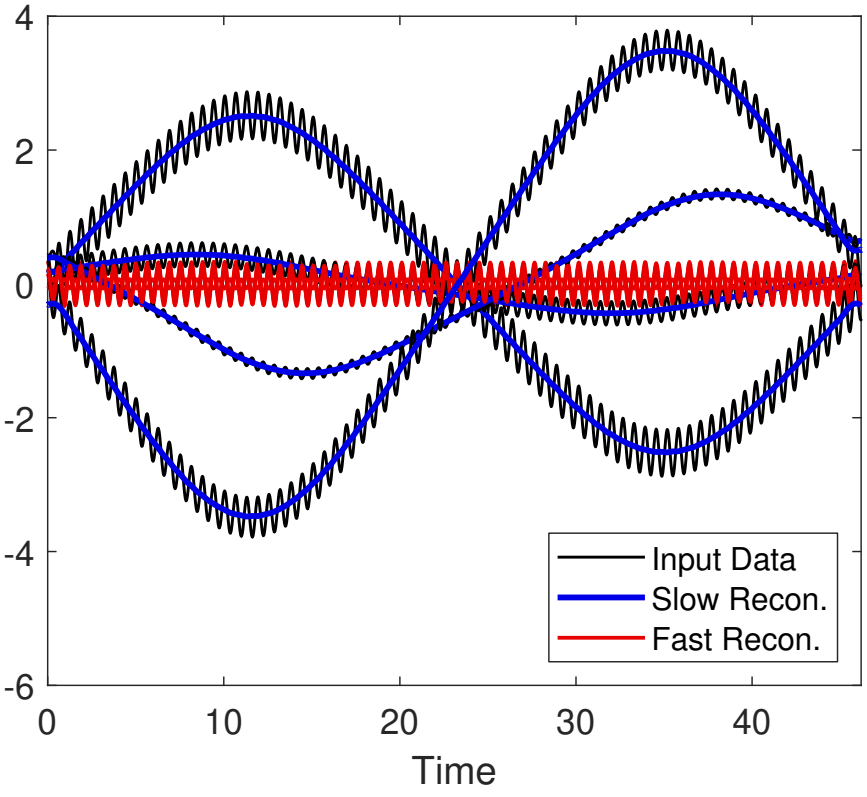

Figure 3.6: Scale separated reconstructions $\bar{\mathbf{x}}^{\text{global}}_{\text{slow}}(t)$ (blue) and $\bar{\mathbf{x}}^{\text{global}}_{\text{fast}}(t)$ (red) plotted over the input data $x(t)$ (black)

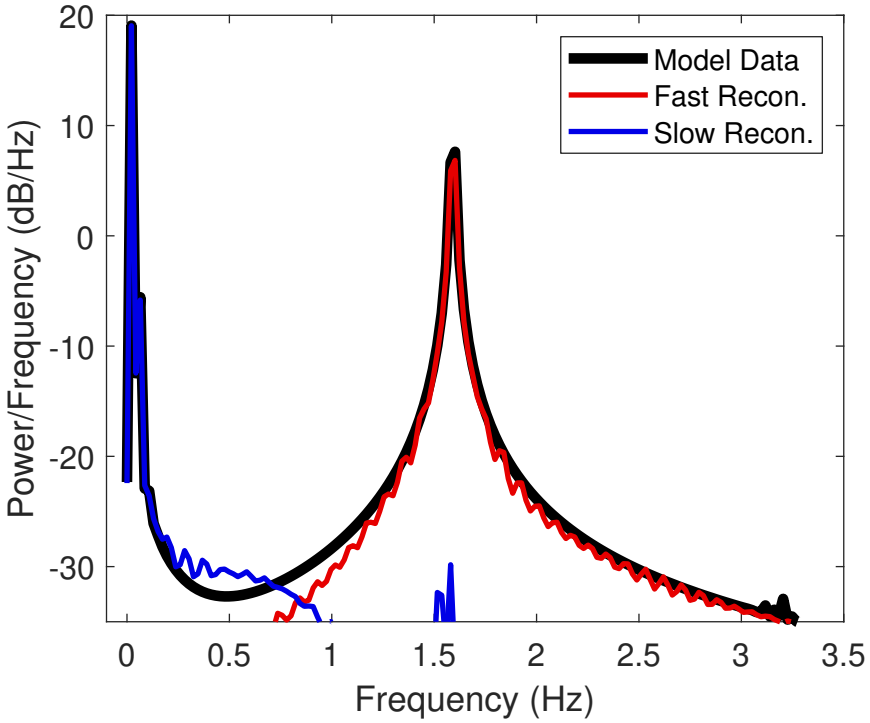

Figure 3.7: Power spectra of the input signal (black) and fast- and slow-scale reconstructions (red and blue, respectively). The 4 variables of each signal are summed to compute frequency content

These properties make this decomposition method a powerful tool for data-driven analyses of systems with multi-scale dynamics, with potential for application towards a variety of modeling and forecasting tasks.

3.2 Scale Separation Performance

Given the task of separating out time-scale regimes from a multiscale signal, one standard and well-known approach is Fourier filtering. Peaks in the power spectrum could be used to identify the constituent frequencies, and each component could then be isolated using an appropriately-designed bandpass filter. This method differs from the one presented here in that the former identifies only temporal coherencies in the signal, whereas the latter incorporates spatial coherencies as well. A brief example is presented here of a case in which the sliding-window DMD technique outperforms Fourier filtering.

Two separate signals with different characteristic time scales are generated using two simple models:

$$\begin{array}{ll} \textbf{FitzHugh-Nagumo} & \textbf{Unforced Duffing} \\ \dot{v} = v - \frac{1}{3}v^3 - w + 0.65 & \dot{p} = q \\ \dot{w} = \frac{1}{\tau_1}\left(v + 0.7 - 0.8w\right) & \dot{q} = -\frac{1}{\tau_2}\left(p + p^3\right) \end{array} \tag{3.6}$$

Characteristic time scales are set to $\tau_1 = 2$ and $\tau_2 = 0.2$, a factor of 10 apart. The FitzHugh-Nagumo model, used as a simple model for biological neuron dynamics, spikes sharply at intervals determined by its characteristic time scale. The Duffing model, on the other hand, is a simple nonlinear oscillator whose dynamics resemble a distorted sinusoid. Therefore, despite the disparity between τ_1 and τ_2, the "slow" component periodically acquires a rate of change comparable to that of the "fast" component. A combined signal x is generated from these by a randomized linear mapping into $\mathbb{R}^4$:

$$\mathbf{x} = \mathbf{A} \cdot \begin{bmatrix} v \\ p \end{bmatrix} \tag{3.7}$$

where $\mathbf{A}$ is a 4×2 Gaussian-random orthonormal matrix.

Signal separation is carried out using a simple Fourier filtering approach and the sliding-window DMD method. The Fourier processing uses Matlab's built-in low- and high-pass filter functions with passband frequencies of 0.04 Hz and 0.15 Hz, respectively. Results are plotted in Fig. 3.8.

Note that while the sliding-window DMD approach clearly performs better, neither method's reconstruction conforms perfectly to the ground truth (plotted underneath in black). Disambiguating truly overlapping scales without error is a highly nontrivial problem, beyond the scope of this work. This result is presented as evidence that sliding-window DMD is at least superior to purely-temporal methods in the case of a problem with nearly-overlapping scales, e.g. closely-spaced frequencies or nonlinear oscillations with spiking behavior.

It is also worth commenting that the data requirements for this method (i.e. duration and frequency of sampling) are at most only slightly greater than those for a Fourier-based decomposition. From an information theoretic perspective, DMD is subject to the same sampling rate restriction that applies to discrete-time Fourier analysis, i.e. the Nyquist criterion that sampling frequency must be greater than twice the highest frequency present in the signal. The lower bound on sampling duration is less strictly defined, but qualitatively it should of course be long enough to capture dynamical evolution of the lowest-frequency content of the signal. The DMD method introduces the additional requirement that there be enough distinct positions for the sliding window to obtain a sufficient set of frequency points for clustering. Windows can overlap with one another though, so this would (at most) perhaps double the requisite sampling duration relative to that of Fourier decomposition.

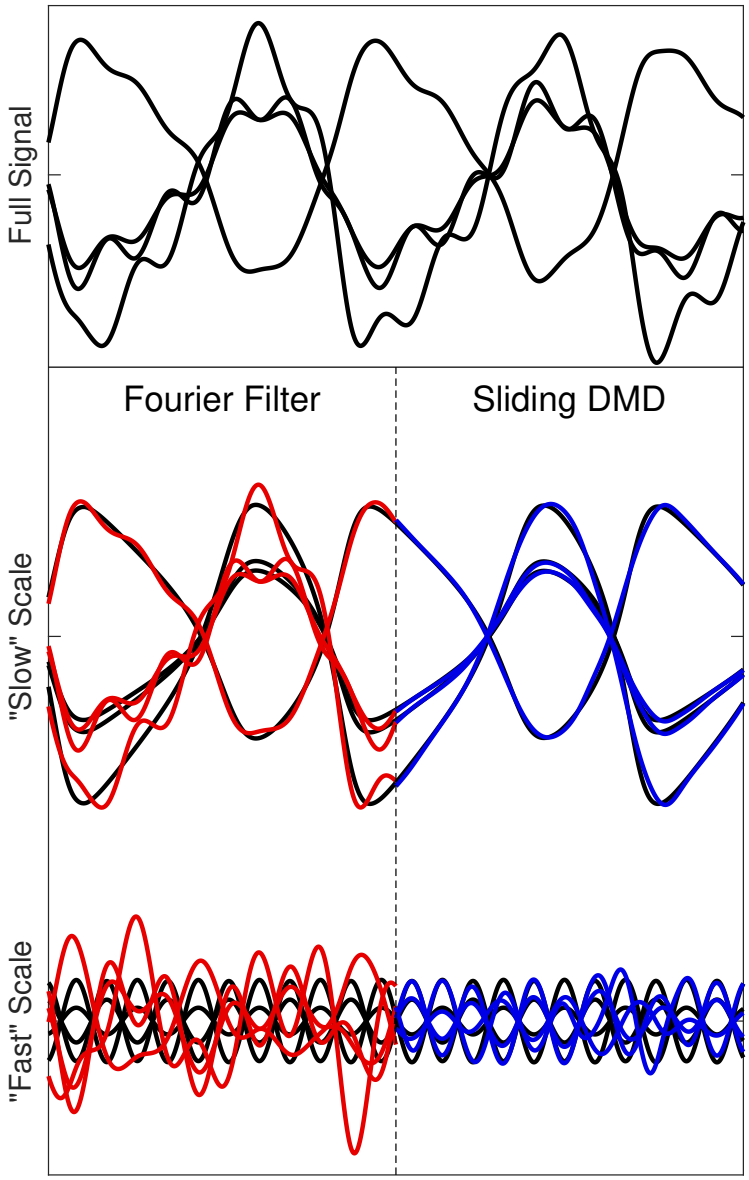

Figure 3.8: Comparison of two scale separation techniques. The measurement signal (top) is constructed from two components, generated by the FitzHugh-Nagumo (center) and Duffing (bottom) models. The ground-truth signal separation is plotted in black, with the results of the two data-driven methods overlaid in color. The sliding-window DMD approach is much more successful in recovering the true components

3.3 Application: A Three-Body Planetary System

3.3.1 Multiscale Properties of Nearly Keplerian Orbits

This section presents an application of the sliding-window decomposition technique to a real physical system with multiscale properties: three bodies interacting gravitationally in bounded orbits, with relative masses comparable to those of Jupiter, Saturn, and the Sun. Because the Sun is larger than the planets by several orders of magnitude, the system resembles two fairly stable elliptical orbits which interact weakly with one another. This suggests the presence of at least three well separated time scales in the dynamics: two "fast" frequencies corresponding to the planetary orbits, and one "slow" regime capturing the evolution of the orbits over much longer durations (which may itself have a multiscale makeup).

3.3.2 Recursive Application and Results

Data was generated for the three-body planetary system using a 4th-order symplectic integrator in Cartesian coordinates over a time span of 1,000,000 years. Applying the same sliding-window DMD procedure outlined in the previous section (window size ~ 600 years), the frequency content very cleanly separated into the three expected regimes (Fig. 3.9).

This highlights a key limitation of the sliding-window DMD approach as it has been presented thus far. The technique is sensitive to the chosen window duration, and data spanning 600 years simply does not contain sufficient information to characterize processes taking place over many millennia. Dynamics unfolding on a scale of 10,000 years would be indistinguishable from those unfolding over 100,000 years: both would just appear as a constant-valued background. While the window size used here does an excellent job of separating out the orbital frequencies of the two planets, it relegates everything taking place on time scales longer than those to a single "slow" regime. Zooming out on Fig. 3.10 to see the evolution of this component, it is evident that it itself constitutes a rich multiscale signal with nontrivially complex dynamics (Fig. 3.12).

To better characterize these dynamics, the sliding-window DMD approach is re-applied

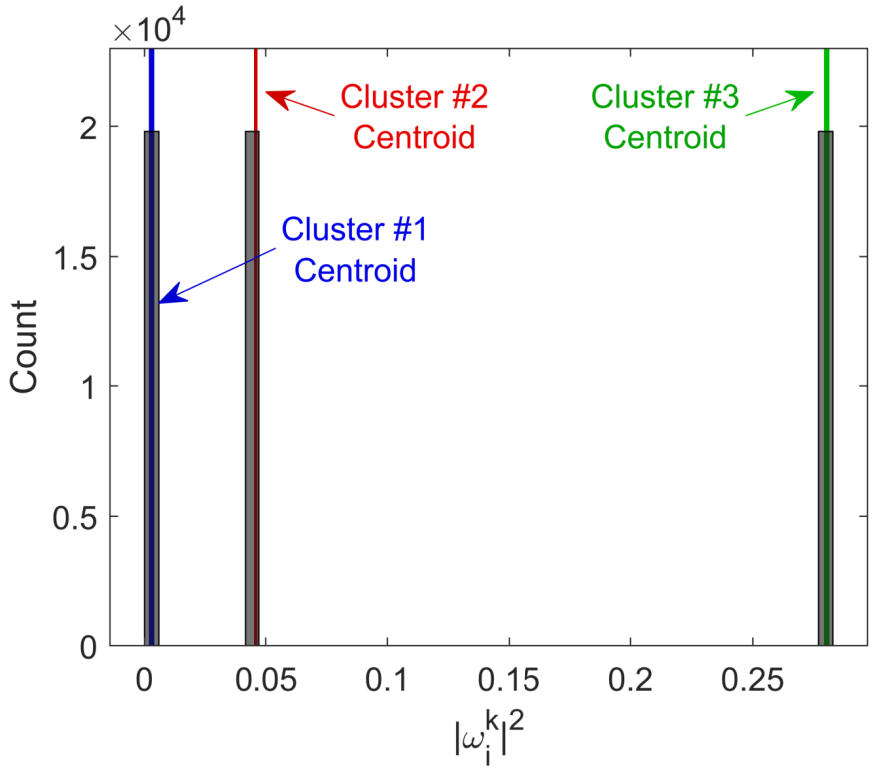

Figure 3.9: Histogram of frequency results of sliding window DMD on the three-body planetary system. k-medians cluster centroids ($k = 3$) are overlaid in color

recursively to identify and isolate signals present in this slow component at different time scales. The methodology is identical to that of the first iteration, but now uses a window of length $\sim 4,600$ years. Repeating this process with successively longer windows yields a decomposition of the original data into 5 distinct time scale components. These components are plotted separately in Fig. 3.13. The full reconstruction obtained by summing them is plotted against the input data in Fig. 3.14. It successfully captures the true dynamics across all time scales.

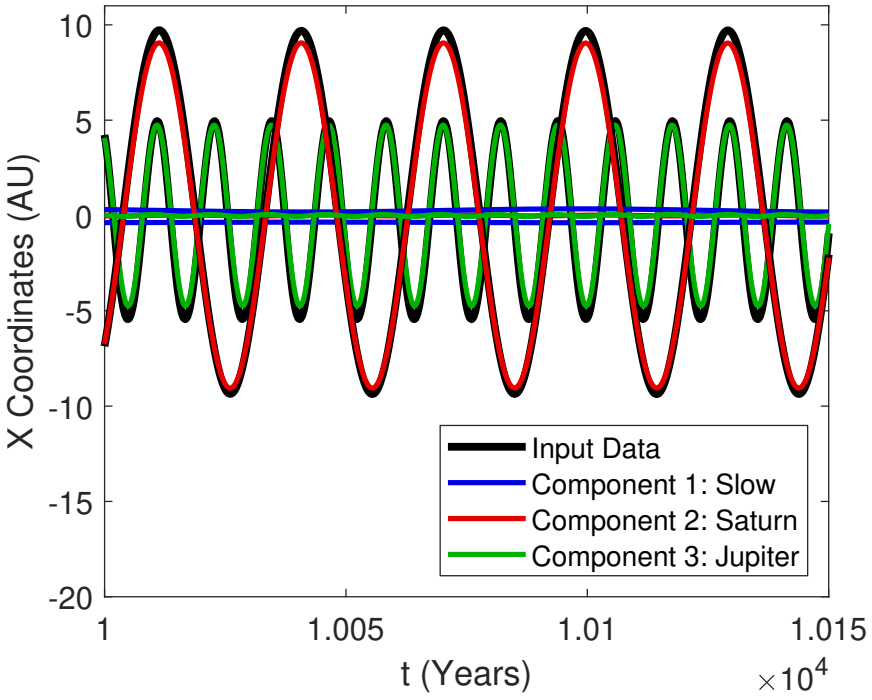

Figure 3.10: Scale-separated reconstructions of three components (color) overlaid on the Cartesian input data (black). Note that for the short domain plotted (150 years), the slow-scale component in blue looks like a constant

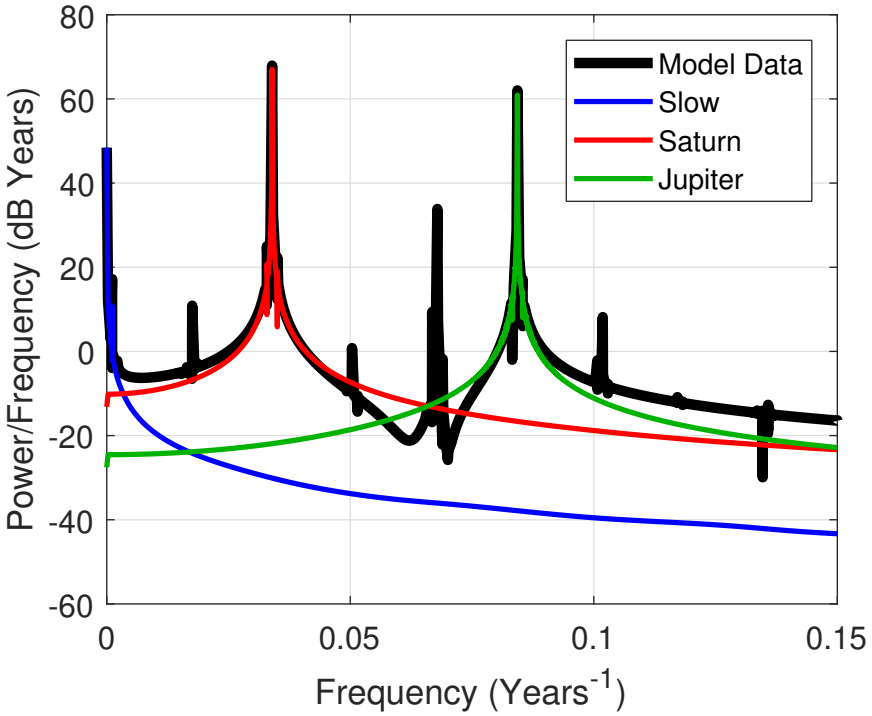

Figure 3.11: Power spectra of the scale-separated reconstructions (color) overlaid on the spectrum of the input data (black). The latter contains peaks not captured by the reconstructions (mostly corresponding to higher harmonics of the base frequency components), but these are mostly quite small compared to the three primary peaks (the vertical scale is logarithmic)

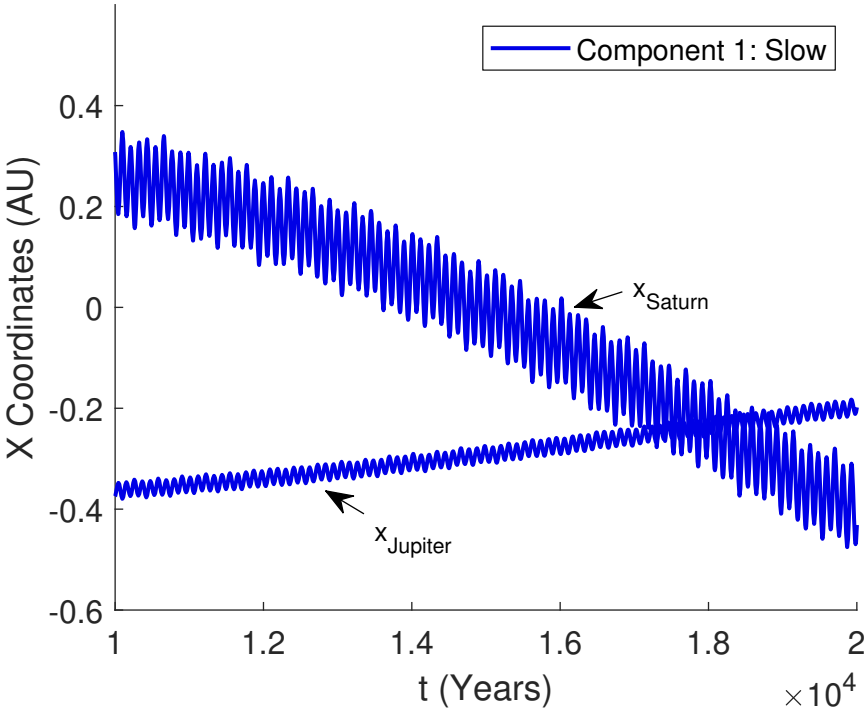

Figure 3.12: The isolated slow-scale component obtained by the first pass of sliding-window DMD on the three-body planetary system. Plotted in blue as Component 1 in Fig. 3.10, the full scope of its dynamics is revealed over this longer time domain. The multiscale behavior evident within this single component motivates a recursive approach to scale separation

3.3.3 Physical Interpretation of Results

This example was chosen in part because it is a well-studied system with a known set of scale-separated parameters: the six Keplerian Orbital Elements provide the minimum information necessary to unambiguously define a (two-body) orbit. For each planet, these quantities are computed over the duration of the simulation: eccentricity (e), semimajor axis (a), inclination (i), longitude of ascending node (Ω), argument of periapsis (ω), and true anomaly (θ). The multiscale properties of the planetary orbits can be observed by plotting these elements over different time scales. Fig. 3.15 suggests three distinct scale regimes: θ

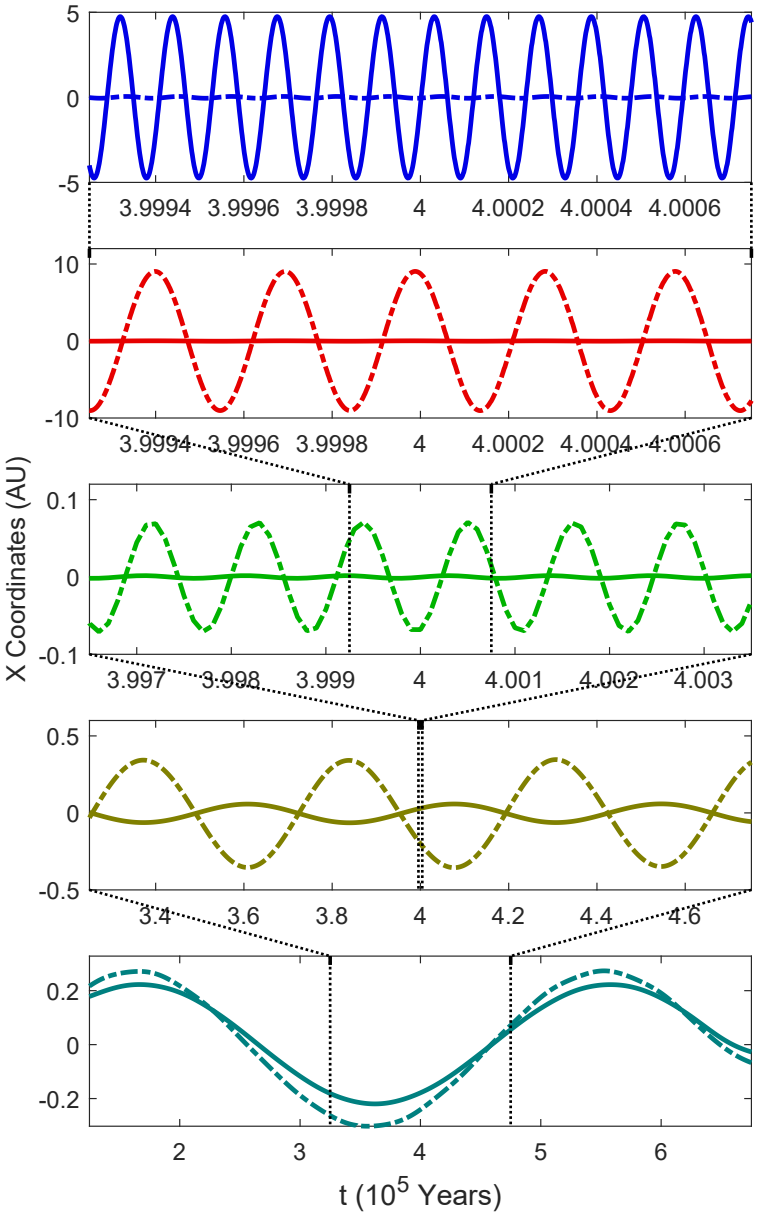

Figure 3.13: Scale-separated reconstructions obtained from 4 recursive applications of the sliding-window DMD procedure to the three-body planetary system. Dotted lines are used to indicate the relative time scales between each successive recursion. Only the x coordinates are plotted. Solid lines denote x_{Jupiter} and dot-dashed lines denote x_{Saturn}.

and a vary at a time scale corresponding to the planetary year, Ω, e and i oscillate at some precessional frequency with a period of about 53,000 years, and the outer envelope of $\sin(\omega)$ has a period of over 300,000 years.

While the DMD components identified (Fig. 3.13) do not all correspond precisely to the dominant frequencies of these elements, they mostly fall neatly into the same three regimes. Like θ, components 1 and 2 have periods corresponding to the revolutions of the two planets. Components 4 and 5 have periods of 47,000 and 383,000 years, respectively, which fall neatly into the two slowest regimes of the orbital elements. Physically, these oscillations seem to relate to the eccentricity cycles of the planets, whose periods have been estimated from numerical models to be 45,900 years (Saturn) and 305,000 years (Jupiter) [117]. The discrepancy between these reported values and those obtained from the simulation likely results from the tertiary effects of other planets and moons in the solar system (which were omitted from the numerical model).

The only DMD component that is not closely matched to any of the orbital elements is the third, with a period of 107 years. This may represent some minor resonance phenomenon of the planetary revolution (it is almost exactly 9 Jovian years), but its specific origin is not clear. In any case it is not a dominant effect; this component has the smallest amplitude of all those identified and it could be omitted entirely without dramatically affecting the full reconstruction.

3.3.4 Limitations

From the power spectra in Fig. 3.11, it is clear that the scale-separated reconstruction has captured the most dominant frequency content of the input data but missed lesser peaks at a few other discrete frequencies. This is further evident in comparison of the slow-scale reconstruction to the "ground truth" of the slow-varying Keplerian Orbital Elements. The singular value spectra of their respective time series (Fig. 3.16) match very closely for the first few modes, but the latter has a fatter tail resulting from the additional frequencies not present in the reconstruction.

This difference demonstrates a key limitation of the scale-separated decomposition method as it has been implemented here: for a system of N variables it returns at most $N/2$ oscillatory modes (in complex conjugate pairs). While it correctly identified the three most dominant peaks in the power spectrum, it entirely passed over the less-prominent harmonics present in the data. This will occur in any system whose dynamics occur at a number of discrete frequencies larger than half the number of measured variables. If a scale-separated reconstruction with more than $N/2$ components is required, it is advised that the input data set be augmented via time-delay coordinate embedding. This approach is discussed at length in Chapter 4.

3.4 Multiscale Forecasting

This section offers a brief example of how the sliding-window DMD approach might be put to practical use. While the decomposition process can be somewhat costly in computational overhead, the execution of the resulting dynamical model is quite efficient (it is simply a closed-form sum of complex exponentials which can be evaluated at arbitrary t). The cost-benefit assessment therefore favors applications in which many calls are made to the predictive model. It is thus quite a natural fit to explore the use of this technique as a precursor to Ensemble Kalman Filtering (EnKF). EnKF is a well-established data assimilation technique that integrates measurement data with an ensemble of modeled forecasts [63]. For many models, this ensemble must be built by stochastically perturbing the parameters of some governing differential equation and then numerically integrating out to the target time. With a DMD-based model, however, no integration is necessary. Perturbations can be applied directly to the model parameters b_j^k, $\boldsymbol{\phi}_j^k$, and ω_j^k according to distributions obtained from the spread of those values over all windowed iterations. A very large ensemble could therefore be built quite efficiently, which would be of particular use for online EnKF. A sample of such an ensemble is plotted component-wise in Fig. 3.17.

3.5 Conclusions

This chapter has introduced a data-driven method for separating complex, multiscale systems into their constituent time-scale components using a recursive implementation of *dynamic mode decomposition* (DMD). The method provides a robust mathematical architecture for regressing to a hierarchy of linear models approximating the nonlinear dynamics at different temporal scales. It even applies to multiscale dynamics produced by coupled, strongly nonlinear oscillators. For integrable systems, for instance, it can extract the constant frequencies of nonlinear oscillations. For nearly integrable systems, these frequencies may no longer be constant, but they change slowly and such variations can be captured by the sliding windows. If there are fast chaotic dynamics, however, then it has no theoretical guarantee to work. In addition to providing diagnostic information on the frequency content of a signal, the sliding-window DMD method produces 1) faithful reconstructions of each constituent component with minimal cross-pollution between them, 2) closed-form expressions for these reconstructions which can be used for low-cost forecasting at any time scale, and 3) statistics on the parameters of windowed DMD models, whose distributions can be sampled for stochastic ensemble forecasting.

3.5.1 Connection to Koopman Theory

The underlying DMD algorithm exploited has a well-documented relationship to the field of Koopman theory, so the scale-separation method presented here can be expressed in those terms. The Koopman operator is a linear operator in some measure space which is fully represents the nonlinear dynamics in the original state space of some system. It is typically infinite-dimensional, but can sometimes be well-approximated in finite dimensions. DMD is one of a number of methods which accomplish this: the matrix A which steps the data forward in time plays the same role in r dimensions that the full Koopman operator would in infinite dimensions.

The sliding-window approach presented in this work generates a new approximator to

the Koopman operator for each DMD iteration. As such, the scale discovery protocol can be viewed as an ensemble approach to building a statistical distribution for the Koopman eigenvalue spectrum and identifying peaks that correspond to discrete time scales present in the original data. While the eigenvectors of A vary from one window to the next, and so cannot be interpreted as global Koopman eigenfunctions, the spectral distribution of eigenvalues could serve as a valuable starting point for an algorithm seeking these functions.

3.5.2 Utility and Applications

The recent ascendance of machine learning techniques for analyzing complex systems, along with advances in hardware to support these techniques, has dramatically overhauled engineering approaches for diagnostics and control of such systems. These methods of course require high quality data sets, but also often rely heavily on an interstitial preprocessing step. For time series data with highly disparate time scale content, scale separation is an integral preprocessing procedure for many tasks. Modeling, forecasting, and control of discretely-multiscale systems are much more effective when the scale components can be treated separately. This is particularly true in the common case where the governing dynamics of these components are only weakly coupled to one another: modeling them independently can produce excellent approximations to the true dynamics at a fraction of the computational cost.

With this in mind, the method outlined in this chapter is presented as a possible precursor to any data-driven application seeking to exploit a system's multiscale properties. While its output is not entirely dissimilar from well-established multiresolution analysis methods, its differences from other approaches render it particularly well-suited to this role. Its synthesis of spatial and temporal coherencies in the data integrate well into dynamics-focused applications; it can more robustly separate components even when one briefly encroaches on the other's characteristic time scale. Clean separation on this basis is crucial for scale-separated model discovery. Furthermore, it generates a closed-form parametric model for time-local dynamics, which opens up possibilities for forecasting explored in Section 3.4. The example

presented there is fairly rudimentary, but a more nuanced approach might prove a useful forecasting tool in and of itself. One possible approach is a two-step algorithm which first predicts time evolution of DMD eigenvectors and then builds a full data prediction from those results.

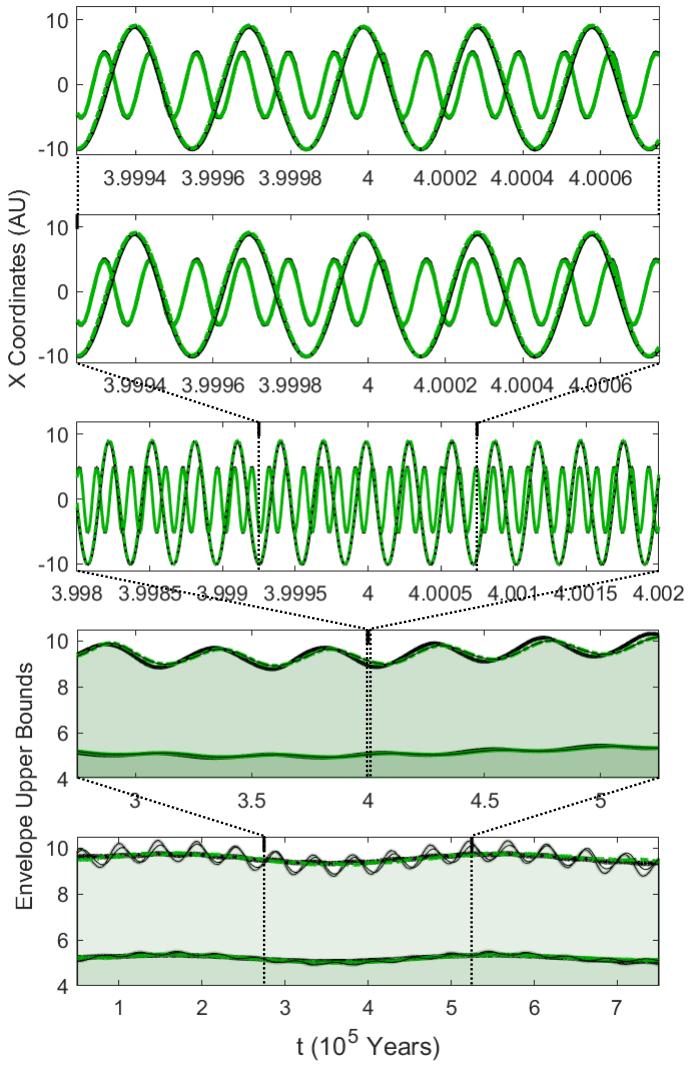

Figure 3.14: The full reconstruction to the three-body planetary system obtained by summing the scale-separated components (x_{Jupiter} in solid green, x_{Saturn} in dot-dashed green) plotted over the original simulation data (black). Each plot here contains the same result, visualized over progressively longer timespans. In the last two plots, only the upper-bound envelope has been plotted for ease of visualization. The final plot shows moving averages to show the conformity even over the longest time scales

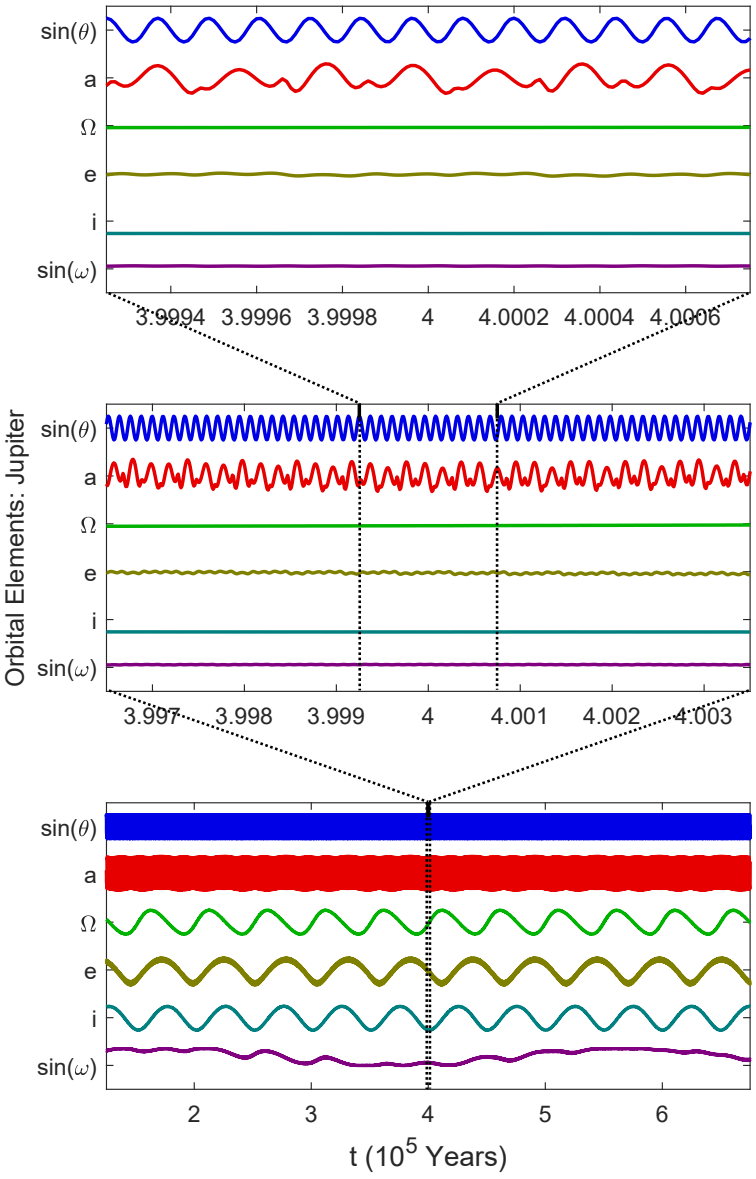

Figure 3.15: Keplerian Orbital Elements for the orbit of Jupiter, computed from the Cartesian simulation data of the three-body planetary system. The three plots show the same data over three distinct time scales

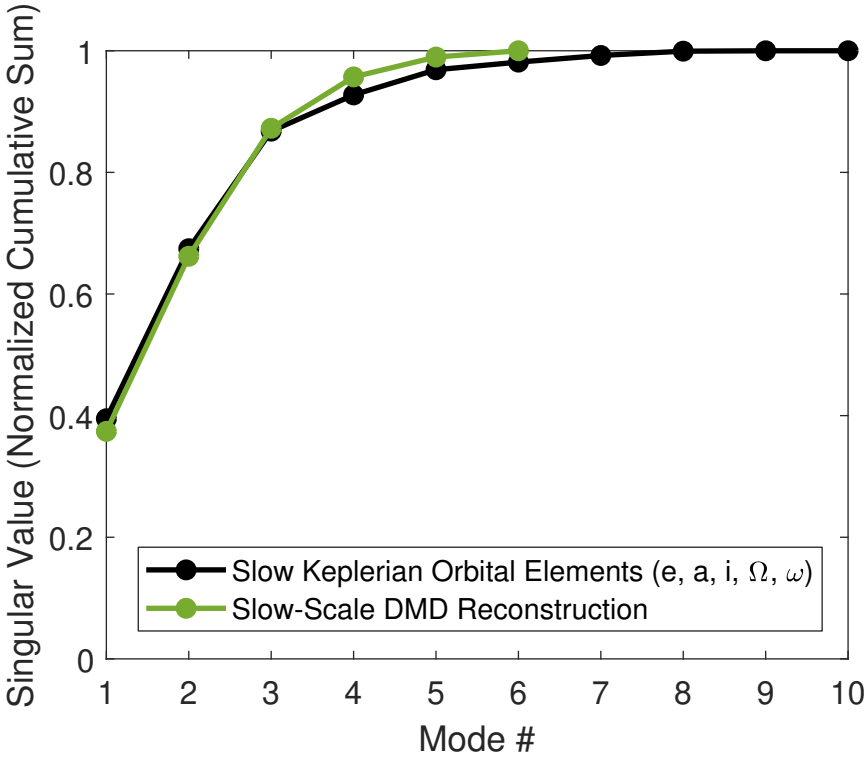

Figure 3.16: Comparison of the singular value spectra of the "true" slow-varying dynamics (obtained by analytically computing the Keplerian Orbital Elements) and the slow-scale reconstruction obtained with the sliding-window DMD method

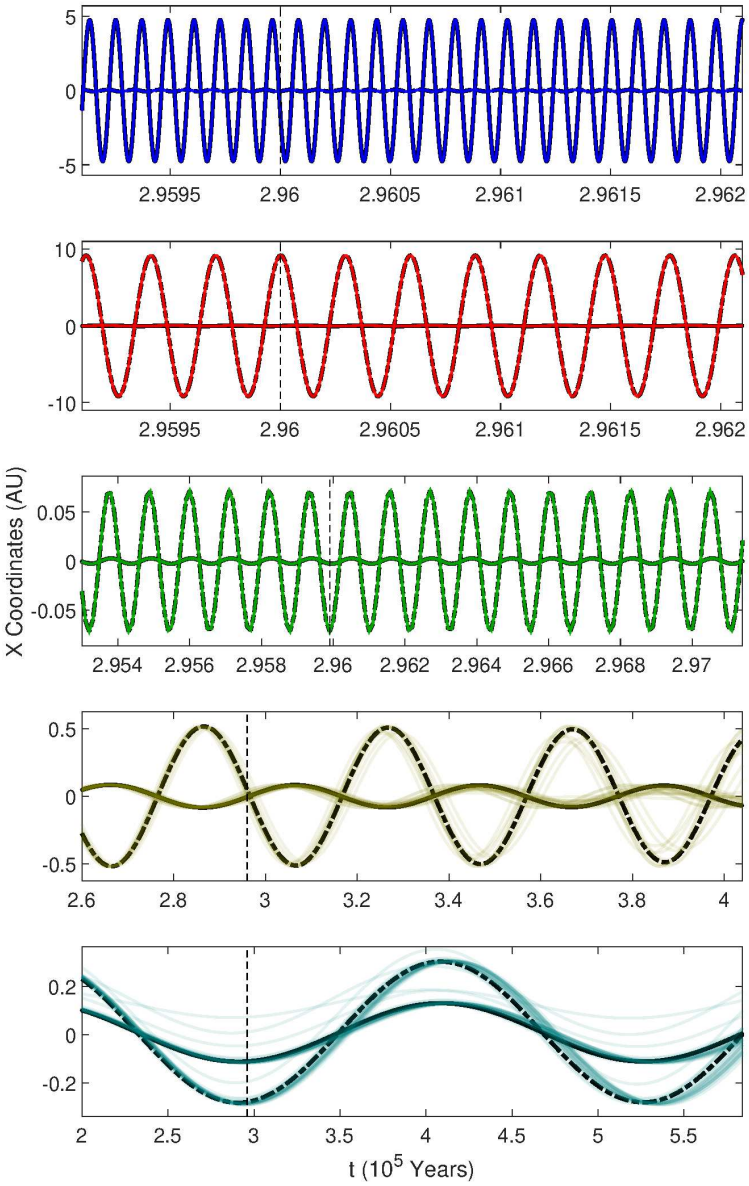

Figure 3.17: Ensemble component-wise forecasts from the sliding-window DMD model on the three-body planetary system. The dotted vertical lines represent the "current time;" everything to their right is the forecasted trajectory. At each scale, the unperturbed DMD forecast (black) is overlaid by an ensemble of predictions generated by sampling b_j^k, $\boldsymbol{\phi}_j^k$, and ω_j^k according to their statistics in the preceding windowed iterations (constraining all $\boldsymbol{\phi}_j^k$ to maintain unit length). Note that these trajectories are all nearly identical in the fast-scale components, where the variance of DMD parameters is minimal

Chapter 4

DELAY-COORDINATE LINEARIZATION MODELS

Modern time series analysis has been transformed by emerging and innovative mathematical methods from machine learning and data science. The ubiquitous availability of measurement data across the sciences, coupled with advances in storage and processing power, has led to a resurgence of interest in the question of how diagnostic and predictive models can be discovered directly from time series data. Using a dynamical systems perspective allows one to take advantage of a broad variety of existing systems analysis methods with applications to forecasting and control. Of particular interest are techniques which are able to map an observed measurement series into a space where it can be accurately reproduced by a set of linear governing equations. This is the central tenet of Koopman theory, as discussed in Chapter 2. This linear representation typically comes at the price of dimensionality: finite-dimensional nonlinear evolution generically requires an infinite-dimensional lifting to be accurately recreated by a linear operator. Linear models are desirable because they allow for the use of many powerful linear algebraic methods for estimation, prediction, and control, all of which can be extended to nonlinear systems if one is able to construct an accurate finite-dimensional coordinate system for the Koopman operator [28]. Time-delay embeddings are shown to provide such a coordinate system, providing a mathematical framework for the extraction of *principal component trajectories* (PCT) which allows for the construction of dynamics, actuated or not, via superposition.

Dynamic Mode Decomposition (DMD) is the leading approach to approximating Koopman operators through regression [170, 161, 108]. As outlined in Chapter 2, it offers a means of efficiently discovering a finite-rank linear approximator to nonlinear dynamics directly from observation data, even in very high dimensions. The simplest implementation of the

algorithm regresses an operator which acts directly on the measured state variables, but there is no reason, *a priori*, to expect this space to admit a faithful linear representation. Central to the successful implementation of DMD is the problem of identifying a space of Koopman observables into which the dynamics can be lifted to better suit the assumption of linearity [137, 133, 33, 134]. The approach of interest in this work is the use of time-delay embedding of measurement data, which has been used to great effect in the HAVOK [31] and the subsequent Hankel DMD [6] algorithms. DMD on time-delay coordinates offers advantageous properties with respect to Koopman approximation which are universal to any input data [94].

Time-delay embedding refers to a coordinate transformation in which a time-localized measurement $x(t)$ is augmented by time-shifted copies of itself $x(t-\tau)$. The use of this technique for data-driven modeling dates back to the seminal Takens embedding theorem, which showed that a chaotic attractor can be reconstructed, up to a diffeomorphism, from the time series of a single measured variable [185]. The method has since found purchase in a number of widely-used algorithms, including *singular spectrum analysis* (SSA) [24], *nonlinear Laplacian spectral analysis* (NLSA) [75], and the *eigensystem realization algorithm* (ERA) [91]. Particular attention has been devoted to the use of techniques such as singular value decomposition (SVD) to identify dominant modal content of data represented in delay coordinates [81, 94]. For a time-delayed scalar measurement series, the principal components obtained by SVD form a temporal basis, with functions similar to those in a Fourier or wavelet basis. Indeed, it has been demonstrated that for sufficiently long embedding windows the delay coordinate SVD converges to the discrete Fourier decomposition [193]. This property suggests a strong compatibility with the Koopman operator-theoretic approach to systems modeling, in which the eigenvalue spectrum of the desired operator has been shown to relate to the same harmonic averages used to compute the Fourier transform [133]. This connection has been borne out by recent work on time-delay DMD [189, 26], variations on which form analogous models on their SVD projections [31, 6] or on exact Fourier basis projections [153].

The fidelity of a DMD model depends heavily on the Fourier spectral properties of the

true dynamics. A finite-dimensional linear operator generates dynamics on a discrete set of frequencies determined by its eigenvalues; if the input data has a continuous spectrum it cannot be fully reproduced by DMD (time-delayed or otherwise). Moreover, the number of discrete spectral peaks that can be captured by a linear model depends on its dimension. DMD represents oscillatory modes by complex conjugate eigenvalue pairs, so a rank-r model will admit at most $r/2$ distinct frequency components. This offers one perspective on the motivation for building models in delay coordinates: a discrete eigenvalue spectrum of arbitrary finite length can be obtained simply by embedding additional shifted copies of the data to increase the dimension of the augmented space. It can be shown that for scalar time series, the minimum number of such embeddings is determined fully by the sparsity of the Fourier spectrum [124, 153]. The continuous spectrum case, however, presents a much greater challenge for Koopman modeling. While a sufficiently dense point spectrum can approximate a continuous one in the high-dimensional embedding limit [49], the practical utility of this for numerical methods is limited. It has long been suggested [133] that a Koopman approach might be taken for the "almost periodic" (i.e. spectrally discrete) portion of dynamics, and augmented somehow to account for the continuous remainder. Previous attempts at this have used intermittent parametric forcing derived from rank-reduced SVD time series [31] or direct translational manipulation of eigenvalues [127] to fill out the spectrum. In this work this problem is approached by incorporating the DMD with Control (DMDc) algorithm [157], which fits measurement data to a linear control model using a known control signal. A fully unsupervised method is presented to learn a control model and a concomitant forcing signal which together fully reproduce the observed dynamics, with discrete-spectrum behavior modeled by endogenous linear evolution and the continuous-spectrum remainder isolated in the exogenous control.

The remainder of this chapter is structured as follows: Sec. 4.1 summarizes the process by which delay coordinate representations are obtained from data. The approach is modeled on that of SSA, but focuses on the less common case of multivariate state delay embedding. A novel interpretation is presented for the coordinate bases that this generates and illustrate

the information-richness of the resultant modes relative to the highly generic output of the same approach on scalar data. Sec. 4.2 discusses the regression of linear DMD models in the learned delay-coordinate space, using data with discrete and mixed spectra. A procedure is introduced for the extraction of an external forcing signal to account for observed nonperiodic dynamics, with results validating its success on systems with known exogenous forcing. Sec. 4.4 joins this method with the existing DMDc algorithm, allowing these results to be integrated into a single unified linear control model. Finally, Sec. 4.5 offers an example of how this technique might be applied to a real-world data set as an informative diagnostic tool.

The work presented in this chapter is based on a paper entitled "Principal Component Trajectories (PCT): Nonlinear dynamics as a superposition of time-delayed periodic orbits" written in collaboration with Eurika Kaiser, Steven Brunton, and J. Nathan Kutz, all of the University of Washington. The paper is still under review, at the time of submission of this thesis, but a preprint is available on the ArXiV: `https://arxiv.org/abs/2005.14321`.

4.1 Full-State Embedding: Methods and Interpretations

This section outlines a procedure for time-delay analysis of a measurement time series. The steps taken are similar to those of SSA. But where SSA typically entails delay embedding on a scalar signal, focus is here extended to the case of a multivariate vector signal and briefly discuss the interpretation of the principal components that this produces.

4.1.1 Time Delay Embedding

Delay embedding is the process of lifting a time series signal into a higher dimensional space by stacking it with time-shifted copies of itself. For a scalar $y(t) \in \mathbb{R}$, the lifted Hankel matrix $\mathbf{H}$ is obtained as follows:

$$\mathbf{Y} = \begin{bmatrix} y(t_0) & y(t_1) & \cdots & y(t_m) \end{bmatrix} \tag{4.1}$$

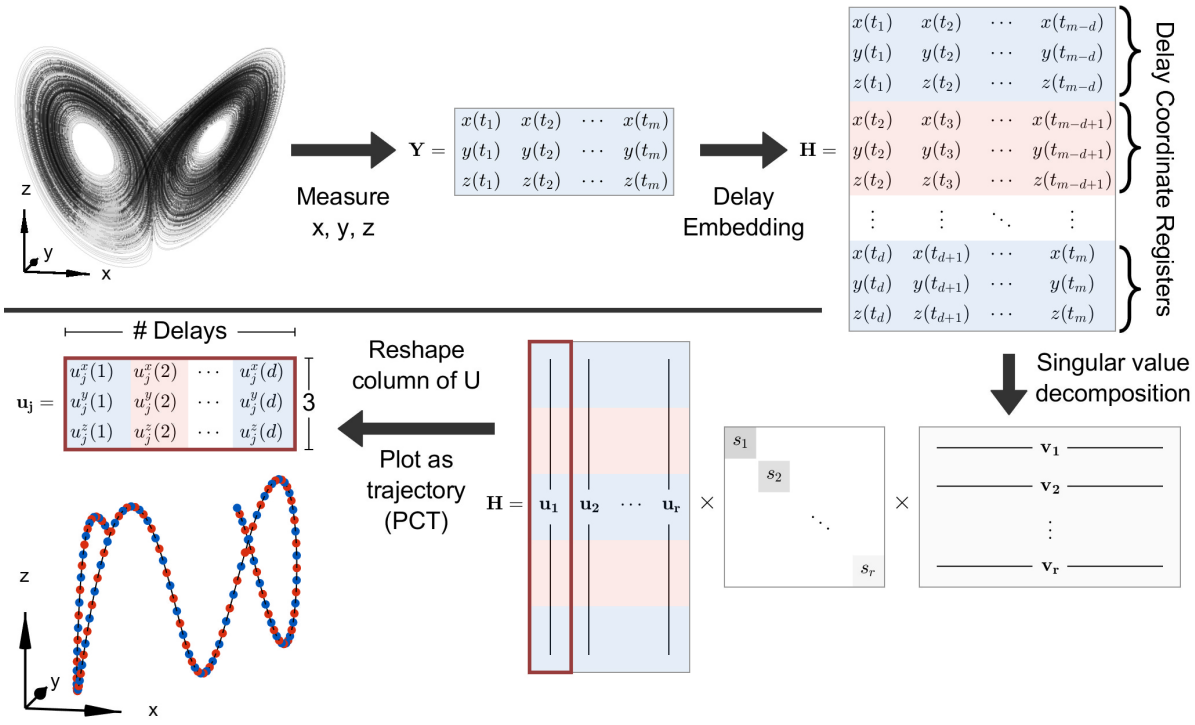

Figure 4.1: Protocol for extracting delay-embedded SVD trajectories, or *principal component trajectories* (PCT), from time series data.

becomes

$$\mathbf{H} = \begin{bmatrix} y(t_0) & y(t_1) & \cdots & y(t_{m-d}) \\ y(t_1) & y(t_2) & \cdots & y(t_{m-d+1}) \\ \vdots & \vdots & \ddots & \vdots \\ y(t_d) & y(t_{d+1}) & \cdots & y(t_m) \end{bmatrix}. \tag{4.2}$$

The matrix $\mathbf{H}$ contains d shifted copies of the original signal $\mathbf{Y}$, each offset from the last by one time step. If $\mathbf{Y}$ had dimension $1 \times m$, $\mathbf{H}$ has dimension $d \times (m - d + 1)$. Note that there is considerable redundancy in the elements of $\mathbf{H}$: all elements belonging to diagonal sets $\{H_{ij}\}$ such that $i + j = (\text{const})$ are equal.

The extension of this procedure to a vector signal $\mathbf{y} \in \mathbb{R}^n$ is straightforward. Each column of $\mathbf{H}$ now consists of d vectors in $\mathbb{R}^n$ stacked on top of each other with the same time shifting scheme as before:

$$\mathbf{Y} = \begin{bmatrix} | & | & & | \\ \mathbf{y}(t_0) & \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_m) \\ | & | & & | \end{bmatrix} \tag{4.3}$$

becomes

$$\mathbf{H} = \begin{bmatrix} | & | & & | \\ \mathbf{y}(t_0) & \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_{m-d}) \\ | & | & & | \\ | & | & & | \\ \mathbf{y}(t_1) & \mathbf{y}(t_2) & \cdots & \mathbf{y}(t_{m-d+1}) \\ | & | & & | \\ \vdots & \vdots & \ddots & \vdots \\ | & | & & | \\ \mathbf{y}(t_d) & \mathbf{y}(t_{d+1}) & \cdots & \mathbf{y}(t_m) \\ | & | & & | \end{bmatrix}. \tag{4.4}$$

The vectorized Hankel matrix is often used for linear system identification in control applications via ERA [91]. This procedure offers two tunable parameters to be independently chosen.

Firstly, the number of delay embeddings d performed on the input signal. This determines the dimension of the delay coordinate space into which the columns of $\mathbf{y}$ are lifted. Previous work on this topic has sought to quantify the lowest value d can take while still admitting a linear dynamical model which faithfully reproduces observed dynamics [25, 76, 153, 124]. For the purposes of this analysis, however, it is assumed that input data of arbitrary duration is available, so there is no need for parsimony in the selection of d. While a high value of d does lead to a high-dimensional embedding space, the subsequent model-building efforts will be carried out in a rank-reduced basis spanned by r principal component vectors, or the *principal componant trajectories* (PCT) of the time-series. Thus a large number of embeddings d does not carry with it any risk of causing the model regression to be underdetermined.

The second parameter to be selected is the embedding period T^d. In Eq. (4.4) the time lag introduced in each successive embedding is equal to the sampling resolution of $\mathbf{y}(t)$, so $T^d = t_d - t_0 = d\Delta t$, assuming the columns of $\mathbf{Y}$ are sampled at evenly spaced intervals of Δt. However, the time shift per embedding could just as easily be set to any integer multiple of Δt, leading to a modified embedding period $T^d = qd\Delta t$, where $q \in \mathbb{Z}^+$. The governing principle for making this choice should be to match the time scale of the dynamics of interest: in the low-T^d limit the stacked state vectors which comprise the columns of $\mathbf{H}$ will be highly redundant, encoding the passage of a time interval too short for any meaningful dynamical evolution of the system. In the high-T^d limit, the delay coordinate sampling resolution is too coarse to resolve the dynamics; unless $\mathbf{y}(t)$ is perfectly periodic the delay registers of $\mathbf{H}$ will become completely uncorrelated. If the system in question contains multiscale dynamics with periods separated by orders of magnitude, multiple sets of embedding parameters may be required. A more extensive discussion of this case with respect to the application of dynamic mode decomposition can be found in [56]. For the purposes of this work, analysis is restricted to monoscale dynamics and choose T^d such that it spans a few periods of the lowest-frequency peak of the Fourier spectrum of $\mathbf{y}(t)$. A more detailed discussion of this parameter selection is available in [76].

4.1.2 SVD and Full-State Embedding

Of central importance to the signal processing technique of SSA [25, 65, 193], and to more recent dynamical systems work with time delay coordinates [6, 31, 94], is the application of singular value decomposition (SVD) to the delay-embedded Hankel matrix $\mathbf{H}$ (as defined in Eq. (4.4)). This section presents a brief discussion of the motivation for this step, convergence properties of the results, and interpretations of the modes in delay space specifically for the case of full-state vector embedding.

The SVD of a matrix $\mathbf{Y}$ yields $\mathbf{Y} = \mathbf{USV}^T$. If $\mathbf{Y}^{n\times m}$ is a time series with state dimension n and time dimension m, then the columns (or *modes*) of $\mathbf{U}$ form an orthonormal basis for the state space, the columns of $\mathbf{V}$ are time series representing normalized projections of the state onto the $\mathbf{U}$ modes, and the elements of the diagonal matrix $\mathbf{S}$ carry the energetic (correlation) weight of each mode's contribution to the full signal $\mathbf{Y}$. Because modal amplitudes are encoded entirely by $\mathbf{S}$, modes can be hierarchically ranked by energy (variance). Truncating the decomposition to contain only the top r modes yields a rank-reduced representation which produces a least-squares optimal reconstruction of $\mathbf{Y}$.

The SVD can be applied to a time series lifted into delay coordinates, as in Eq. 4.4, to obtain a decomposition of the same form:

$$\mathbf{H} = \mathbf{USV}^T \tag{4.5}$$

Resultant modes have additional degree of interpretability owing to the temporal structure embedded in $\mathbf{H}$. Columns of $\mathbf{U}$ are vectors in the input state space, which in this case is the high-dimensional delay space. If $\mathbf{H}$ was constructed from a scalar time series $y(t)$, the SVD modes can be treated as d-element time series spanning a duration of T^d. If $\mathbf{H}$ were instead the full-state embedding of a vector times series $\mathbf{y}(t) \in \mathbb{R}^n$, its modes can be interpreted as trajectories in $\mathbb{R}^n$, also of duration T^d. An intuitive parallel for these decompositions can be found in time-frequency analysis methods such as windowed Fourier and wavelet transforms, both of which have been applied to the DMD method [112, 56]. Thus a vector projection of $\mathbf{H}$ onto the columns of $\mathbf{U}$ in delay coordinates is analogous to a time-localized projection

onto some template function, e.g. a wavelet confined to a length of T^d.

This procedure is shown in Fig. 4.1. Note that when the input signal is multivariate, vectors in the delay embedding space have elements spanning both spatial coordinates x, y, z and measurement times $t_j, t_{j+1}, ..., t_{j+d}$. To treat this as a normed vector space (as SVD does) is to make an implicit assumption about the relative weighting of spatial versus temporal coordinate separations. This assumption is carried in the choice of time spacing between delay embeddings, or equivalently in the choice of embedding period T^d.

4.1.3 Interpreting SVD Modes in Delay Coordinates

The result of the decomposition process outlined in Fig. 4.1 depends heavily on the structure of $\mathbf{H}$. By construction, elements of $\mathbf{H}$ are represented redundantly up to d times (specifically, all H_{ij} where $i + n \times j = (\text{const})$ are equal). This property holds regardless of the dynamical content of the signal in question, and leads to certain guarantees on the SVD modes of $\mathbf{H}$. In particular, it can be shown that for embeddings of a scalar time series ($n = 1$), modes converge to simple functions in the limiting cases of T^d. For small T^d, the columns of $\mathbf{U}$ resemble Legendre polynomials, while for large T^d they become sinusoidal [25, 193]. Indeed, in some circumstances exact instantiations of these functions have been used instead of the data-driven approximations obtained with the SVD [76, 153]. This work focuses on the large-T^d limit, in which delay-coordinate SVD can be thought of as reproducing the discrete Fourier transform on a sparsified frequency basis: the near-sinusoidal $\mathbf{U}$ modes resemble the projection basis used in the DFT, but instead of a large number of basis functions densely populating frequency space, only the (SVD rank) r most prominently represented frequencies are retained.

Shifting attention to the case of multivariate vector embeddings, similar behavior is observed in the long embedding time limit. Some sample modes obtained for the chaotic Lorenz system are plotted in Fig. 4.2. The trajectories in $\mathbb{R}^3$ traced out by the modes result from sinusoidal oscillation in x, y, and z. The frequencies of oscillation for different state variables within the same mode are typically equal or related by an integer ratio, but the relative

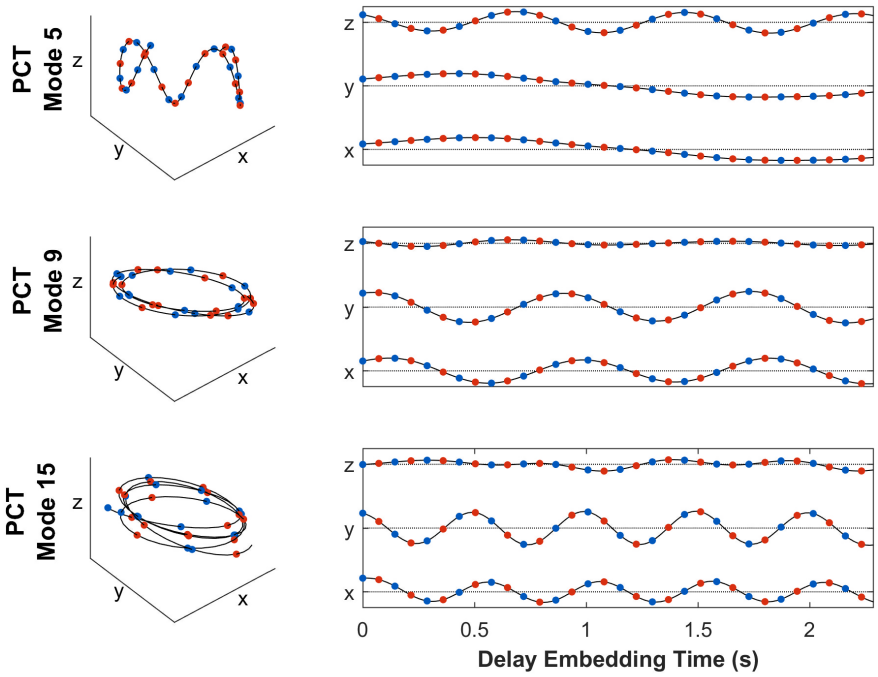

Figure 4.2: Time-delayed SVD modes for a chaotic Lorenz system. Full 3D representations of the PCTs (left) are composed of approximately sinusoidal oscillation in the x, y, z coordinates with (often) incommensurate frequencies and amplitudes.

phases and amplitudes are free to vary. Consequently, modes which in the scalar high-T^d case could be completely defined by a single frequency value take on much richer geometric structure in the vector high-T^d case. The figures which can be thus formed by parametric sinusoids belong to the class of generalized N-dimensional Lissajous figures [2], in the special case where frequencies differ by integer ratios. Note that the periods of oscillation are independent of the embedding period T^d, so these curves do not in general form closed orbits.

The SVD reconstruction (4.5) can be equivalently written as a sum over individual modes:

$$\mathbf{H}(t) = \sum_j \mathbf{u}_j s_j \mathbf{v}_j^*(t) \tag{4.6}$$

Each column of $\mathbf{H}$, which represents a windowed snapshot of the dynamics over a period of

T^d, is reproduced by a linear combination of the orbits $\{\mathbf{u}_j\}$ weighted by the singular values $\{s_j\}$ and the time series projections $\{\mathbf{v}_j(t)\}$. The delay-coordinate SVD therefore functions as an automated means of decomposing a complex trajectory into simple orbital components (Fig. 4.3), or the PCT coordinates of interest. The result is reminiscent of the Ptolemaic model of the solar system, in which the apparent retrograde motion of planets in the sky is explained using a hierarchy of superimposed epicycles in geocentric coordinates.

4.1.4 PCT of a Nonlinear Oscillator with Discrete Fourier Spectrum

A more concrete illustration of the analogy to multivariate Fourier decomposition can be found in the example of the Van der Pol system. This canonical nonlinear oscillator is defined by the coupled first-order equations:

$$\begin{aligned} \dot{y}_1 &= y_2 \\ \dot{y}_2 &= \mu \left(1 - y_1^2\right) y_2 - y_1 \end{aligned} \tag{4.7}$$

with μ, which controls the strength of the nonlinearity, taken to be 1 from here on. This system admits solutions which form a closed periodic orbit, so the Fourier spectrum of the trajectory it generates is discrete (only integer harmonics of the base periodic frequency are present). Applying the procedure from Fig. 4.1 with $T^d = 20$ produces the delay-coordinate modes plotted in Fig. 4.4. The state space of Eq. 4.7 is represented by the $x - y$ plane, and the time coordinate on the interval $[0, T^d]$ is extended up the z axis to better illustrate frequency distinctions. In the rightmost column, Fourier spectral content of each (combined) mode pair is overlaid on the spectrum of the full Van der Pol trajectory.

The results for this simple example are illustrative of three key features of vector-embedded PCT. First, the modes naturally adhere to the frequencies present in the power spectrum of the input signal. In other words, the PCT not only converge to a sinusoidal basis, but it discovers a sparse Fourier basis made up of those sinusoids most strongly represented in the data. Second, the ordering of the modes corresponds to the heights of their corresponding spectral peaks. This follows from the fact that in the limit as the columns of $\mathbf{U}$ converge to sinusoids, the singular values associated with them become directly analogous to Fourier power measurements in that both represent a magnitude of the energetic content of the full dynamics contained in a projection onto a given oscillatory mode. Third, the PCT modes naturally self-organize into pairs which share a common frequency and singular value, which is often observed in the fluid flow past bluff bodies [145, 181]. In each row in Fig. 4.4, the modes are identical up to a 90° rotation. This again suggests a parallel to Fourier analysis: a pair of out-of-phase sinusoids of equal frequency (e.g. $\sin(\omega t)$ and $\cos(\omega t)$) form

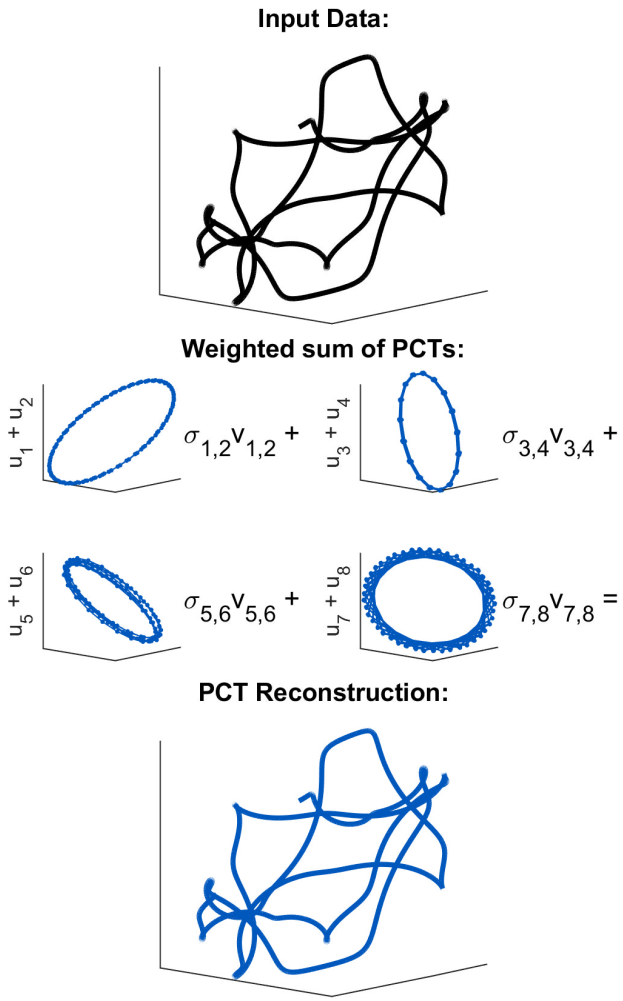

Figure 4.3: Pointwise linear combinations of simple modal orbits can reconstruct complex trajectories in $\mathbb{R}^n$.

a complete basis for dynamics of arbitrary phase at that frequency.

4.1.5 PCT and Continuous-Spectrum Dynamics

As the spectral plots in Fig. 4.4 suggest, PCT is a method suited to systems whose dynamics admit a discrete Fourier representation. In the long-embedding limit in which PCT modes converge to sinusoids, a decomposition of finite rank r can at most reproduce $r/2$ spectral peaks in its reconstruction. For a system with a continuous spectrum, this is clearly insufficient: no finite number of discrete frequencies can densely cover a continuous interval on the frequency space. Thus, while the delay method of Fig. 4.1 can be applied to any uniformly-sampled time series data in n dimensions, its utility as a means of sparsely representing the dynamical content of the given signal is limited to the discrete-spectrum case.

This is a highly restrictive condition, given that many systems of mathematical interest and almost all real-world systems do not exhibit fully discrete Fourier spectra. Even when dynamics are very nearly periodic, the introduction of any measurement noise or stochastic/chaotic forcing will populate the gaps in the spectrum and undermine any attempt to build a linear model for the system in delay coordinates. This is illustrated in Fig. 4.5, in which the Van der Pol oscillator from Eq. 4.7 is subjected to different types of weak parametric forcing on the y_2 variable. Though the magnitude of forcing is sufficiently small that the state-space trajectories (top) look nearly identical, the differences in the power spectra (bottom) are quite obvious. The consequences of this spectral contamination are evident when the plotted time series are subjected to linear model discovery using DMD, which is discussed in the next section. Even when the exogenous forcing is an order of magnitude weaker than the intrinsic dynamics, it yields a model whose reconstruction error is many times greater than that of the unforced system. Because the base periodic oscillation still dominates the observed dynamics, a linear delay-coordinate model should offer a useful approximation to the true behavior. The remainder of this chapter is devoted to understanding how such a model can be obtained by isolating the continuous-spectrum forcing from the underlying periodic motion and combining them in the framework of linear control.

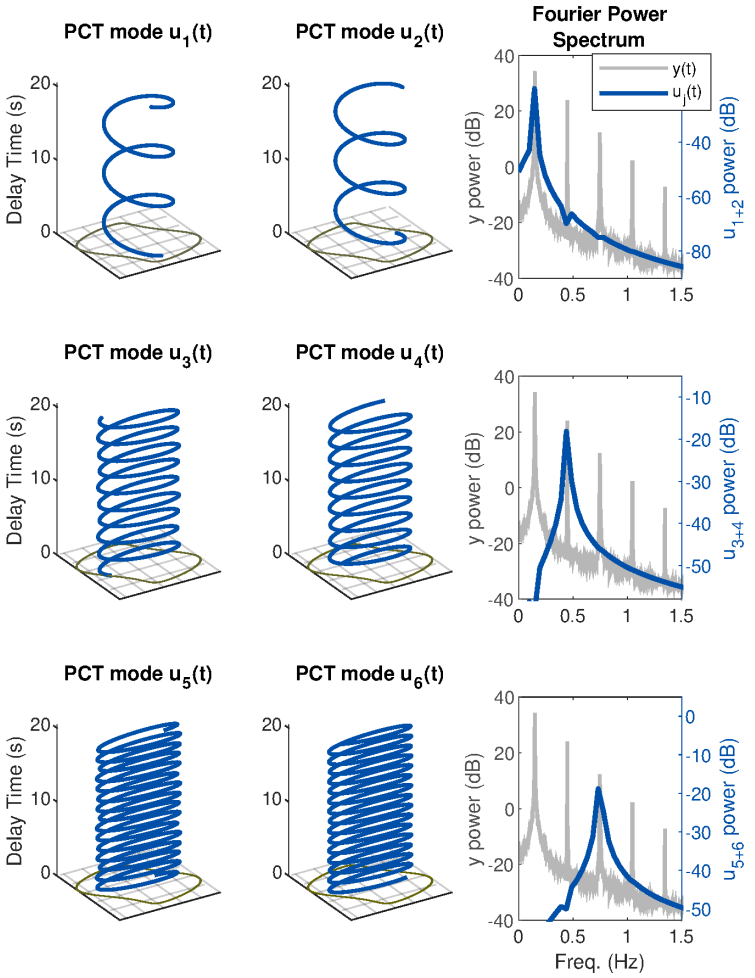

Figure 4.4: First 6 SVD modes for the delay-embedded Van der Pol oscillator. Each mode can be thought of as a trajectory in $\mathbb{R}^2$, with the z coordinate here representing the delay time. They organize into pairs 90° out of phase in order to form a complete orthogonal basis for a given frequency. Power spectra plotted on the right show that the dominant SVD modes naturally correspond to the frequencies most prominently represented in the signal.

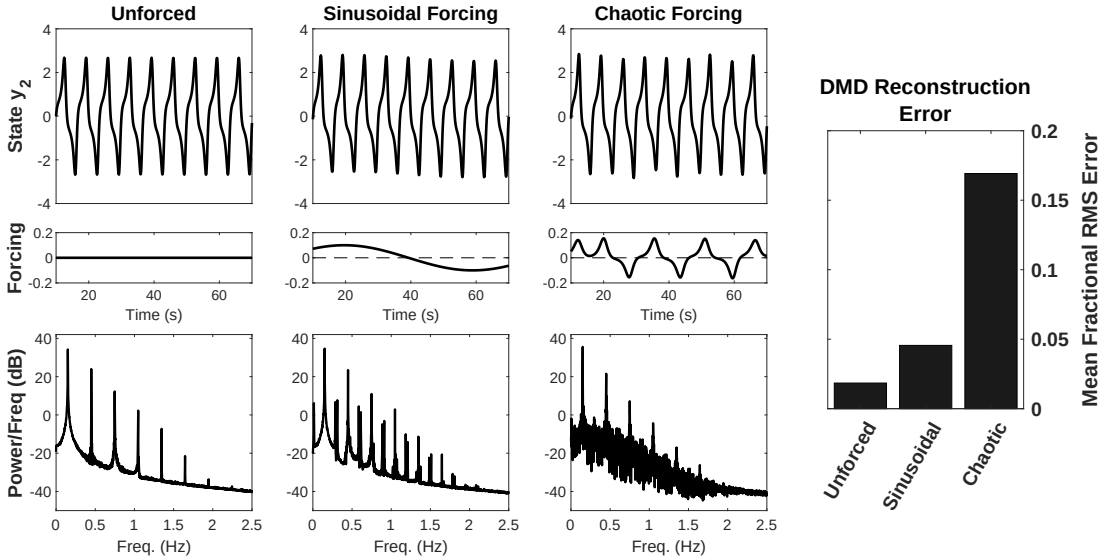

Figure 4.5: Power spectrum of a Van der Pol oscillator system subjected to different types of forcing.

4.2 Dynamic Mode Decomposition and Time-Delay Embedding

4.2.1 Time-Delay DMD

A resurgence of interest in Koopman theory and consequently in DMD in recent years has led to a variety of extensions of the algorithm focusing on two principal tasks: modifying the optimization objective function to redefine what constitutes a "best fit" for $\mathbf{A}$ and identifying lifting transformations which render the data more amenable to linear representation [200, 74, 127]. The results presented in this work remain agnostic on the former issue and focus on a particular solution to the latter: namely, the use of delay-embedding coordinates, or PCT. The lifting procedure described in Eq. (4.2) has been shown to quite generically improve DMD representations for continuous time-series data [133, 120]. Of particular note is the original Hankel alternative view of Koopman (HAVOK) formulation [31], and the subsequent and closely related Hankel DMD [6].

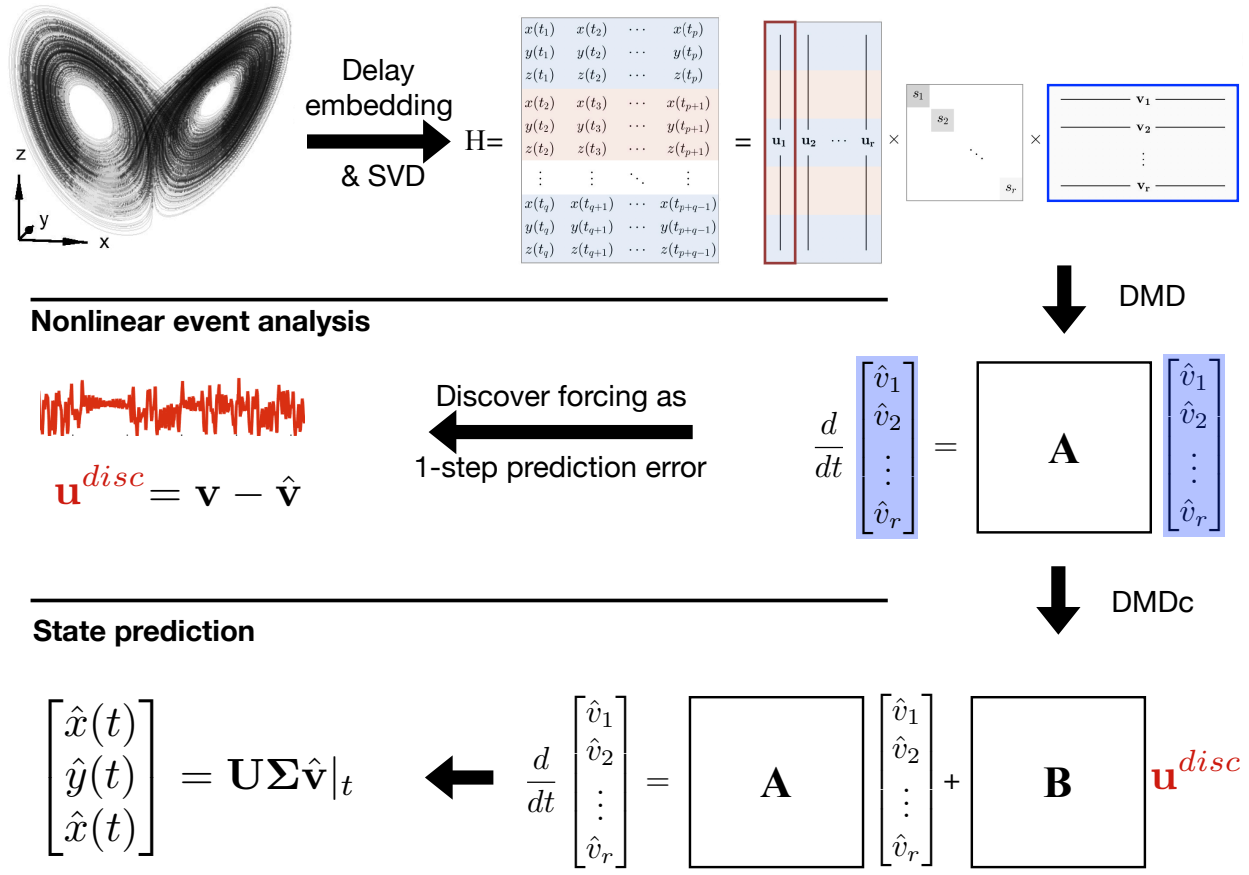

Figure 4.6: Procedure for building linear models from the delay-coordinate SVD illustrated in Fig. 4.1. DMD performed on the time series V matrix yields a linear model whose stepwise prediction error is interpreted as a learned forcing signal u^{disc}. The same V matrix can then be fed to the DMD with control (DMDc) algorithm along with u^{disc} in order to construct a linear control model of the form shown.

The procedure for constructing a delay-coordinate linear model is diagrammed in the first two rows of Fig. 4.6. Starting from the SVD approach illustrated in Fig. 4.1, a linear DMD operator $\mathbf{A}$ is regressed using data from the first r rows of the time-series projection matrix $\mathbf{V}$:

$$\dot{\mathbf{V}} = \mathbf{AV} \tag{4.8}$$

The resultant model can reproduce the dynamics in the space of the SVD modes by integrating this differential equation (or using the analytic solution from Eq. 2.4). It also enables state forecasting simply by integrating further out to some future time. Results can be transformed back into the original delay-coordinate space via matrix multiplication: $\hat{\mathbf{H}} = \mathbf{U}\mathbf{S}\hat{\mathbf{V}}^T$, where $\hat{\mathbf{V}}$ is the DMD reconstruction of the data matrix $\mathbf{V}$.

The Hankel matrix formed by the embedding procedure of Eq. (4.4) is guaranteed by construction to have redundant values along matrix diagonals $\{H_{ij}$ where $i+j = (\text{const})\}$. No such property is assured for the DMD reconstruction $\hat{\mathbf{H}}$, so the task of collapsing the resultant signal back down into the original state space is nontrivial. There is an approach used in SSA known as Hankelization, or diagonal averaging, in which a state-space reconstruction $\hat{\mathbf{x}}$ is obtained from $\hat{\mathbf{H}}$ by averaging over the elements which would in a true Hankel matrix be redundant [81]. However, in the case of model forecast residuals discussed in the following section, there is often coherent oscillatory behavior observed along these diagonals which can be obfuscated by averaging. In these cases, the first n rows of $\hat{\mathbf{H}}$ are instead sampled to compress results back into the state space. The delay registers largely differ from one another only by a global phase factor, so the choice to use the first register is not particularly consequential.

4.3 Data-Driven Decomposition of Nonlinear Systems into Forced Linear Models

DMD generates models with complex eigenvalues, but for real and stationary input data its modes converge to pure-imaginary conjugate pairs. In this case, a rank-r DMD model yields

dynamics at at most $r/2$ distinct frequencies. Thus, while Hankel DMD can produce a linear model which reproduces a discrete-spectrum nonlinear oscillator to arbitrary precision, it will unavoidably fall short when it comes to systems with continuous spectra. This section presents an extension of Hankel DMD to model nonlinear dynamics as a forced linear system, for which the model and the forcing signal are learned simultaneously. This approach retains the benefits of linearity for the portion of the model which captures the dominant, (quasi)periodic dynamics, while adding the versatility to simulate a much broader class of nonlinear systems.

A standard delay-coordinate DMD model is first generated as explained in the previous section. Working in the space of PCT modes, this means learning the linear operator $\mathbf{A}$ which offers a best-fit solution to $\dot{\mathbf{v}}(t) = \mathbf{A}\mathbf{v}(t)$ over the full duration of available data. The resulting model is then used for stepwise forecasting: given some $\mathbf{v}(t_j)$, the succeeding observation $\mathbf{v}(t_{j+1})$ can be approximated as follows:

$$\hat{\mathbf{v}}(t_{j+1}) = \mathbf{v}(t_j) + \Delta t \mathbf{A}\mathbf{v}(t_j) \tag{4.9}$$

(This simple Euler time step could be replaced by any forward numerical integration method). Iterating over the first $(m-1)$ columns of $\mathbf{V}$, this approach can be used to construct a full single-step prediction matrix $\hat{\mathbf{V}}$:

$$\hat{\mathbf{V}} = \begin{bmatrix} | & | & | & \\ \hat{\mathbf{v}}(t_1) & \hat{\mathbf{v}}(t_2) & \hat{\mathbf{v}}(t_3) & \cdots \\ | & | & | & \end{bmatrix} \tag{4.10}$$

The discrepancy between $\mathbf{V}$ and $\hat{\mathbf{V}}$ can be thought of as a quantitative account of the shortcoming of the DMD representation. The true dynamics can be asserted to take the form

$$\dot{\mathbf{v}} = \mathbf{A}\mathbf{v} + \mathbf{u}(t) \tag{4.11}$$

for some parametric exogenous forcing $\mathbf{u}(t)$. In this case a forward Euler step would yield

$$\mathbf{v}(t_{j+1}) = \mathbf{v}(t_j) + \Delta t \mathbf{A}\mathbf{v}(t_j) + \Delta t \mathbf{u}(t_j) \tag{4.12}$$

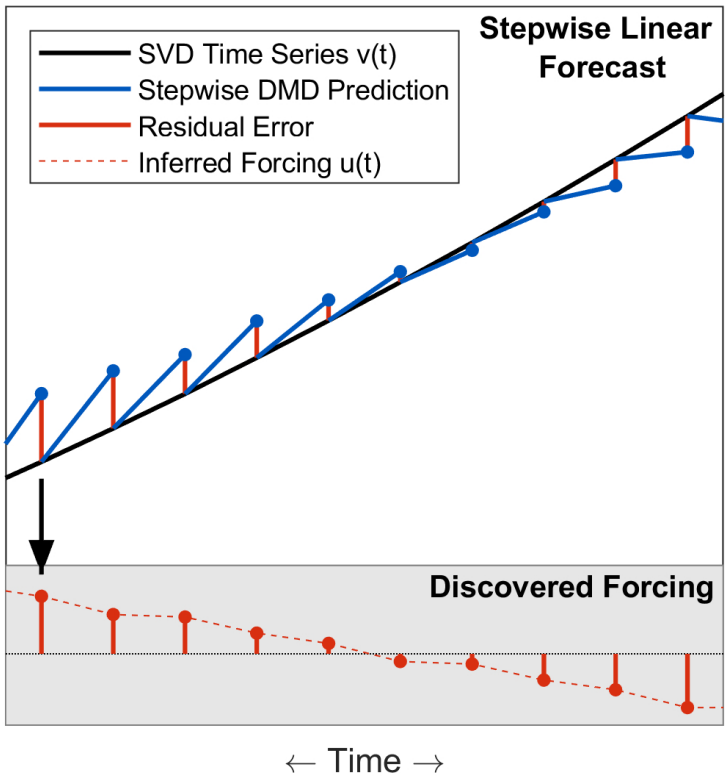

Figure 4.7: Extracting stepwise DMD forecast error to construct a forcing time series. Actuated by this signal, the linear DMD model will perfectly reproduce the true nonlinear dynamics over the observation interval. The model error shown in red has been scaled up for visual clarity.

and so

$$\mathbf{v}(t_{j+1}) - \hat{\mathbf{v}}(t_{j+1}) = \Delta t \mathbf{u}(t_j) \tag{4.13}$$

leading to

$$\mathbf{u}(t) = \frac{\mathbf{v}(t+\Delta t) - \hat{\mathbf{v}}(t+\Delta t)}{\Delta t}. \tag{4.14}$$

An inferred forcing signal can thus be extracted simply by computing the stepwise forecast $\hat{\mathbf{V}}$ using Eq. (4.9). This process is illustrated schematically in Fig. 4.7. In one sense

this process is tautological: the learned $\mathbf{u}(t)$ takes on whatever value is required to match prediction to reality at a given time. The control model posited in Eq. (4.11) is fully accurate by construction. However, the decomposition it achieves is nontrivial. For any system whose dynamics are dominated by near-periodic evolution (or linear combinations thereof), Hankel DMD can generate a model which captures a large fraction of observed behavior. Under these circumstances, and assuming the rank of $\mathbf{A}$ is sufficient to account for all prominent spectral peaks, the discovered forcing consists solely of the continuous-spectrum mixing dynamics which lie beyond the reach of DMD. In other words, to the extent that Hankel DMD can be considered an optimal means of discovering a set of r observables in whose space dynamics are most nearly linear (known as a Koopman invariant subspace), $\mathbf{u}(t)$ offers a window into the residual dynamics which unfold in the orthogonal complement to this subspace.

The following forced Van der Pol system offers an instructive example:

$$\begin{aligned} \dot{y}_1 &= y_2 \\ \dot{y}_2 &= \left(1 - y_1^2\right) y_2 - y_1 + u(t) \end{aligned} \tag{4.15}$$

As plotted in Fig. 4.5, cases are considered where the forcing $u(t)$ is zero (unforced), where it is sinusoidal, and where it is a chaotic oscillatory signal derived from the z coordinate of a trajectory on the canonical Lorenz attractor. In each case, a Hankel DMD model is generated from data generated over an interval of 1000 time units sampled at a resolution of 0.01 time units. The Hankel matrix is constructed using $d = 256$ delay embeddings spaced $q = 4$ time steps apart, for an embedding period of $T^d = qd\Delta t = 10.24$ (which is on the same order as the Van der Pol oscillatory period). The results of the forcing discovery procedure are plotted side by side with the true forcings used to simulate the data in Fig. 4.8. While the phases of the corresponding signals clearly do not match, their structures are quite similar and their autocorrelations plotted on the right match very closely.

The DMD models used to generate Fig. 4.8 were selected from models with ranks varying from 2 to 48. The results plotted (models of $r = 12$ and $r = 6$, respectively) were chosen as particularly clean representations of the desired results. Some of the models of other

ranks produced forcing signals which were somewhat obfuscated by additional high-frequency oscillation. This is because the ability of this method to separate periodic and aperiodic contributions to observed dynamics is limited by the capacity of DMD to fully capture the periodic behavior without any additional, spurious oscillatory modes. The strength of a DMD model in this regard can be evaluated by comparison between its eigenvalue spectrum and the Fourier spectrum of the input data: there should be eigenfrequencies corresponding to all prominent Fourier peaks no more than that. In Fig. 4.5 it is apparent that the sinusoidal forcing introduced additional peaks at half-integer harmonics of the base frequency, so it is not surprising that a higher-rank DMD model was required to isolate a clean forcing signal relative to the chaotically forced system.

It should also be noted that in the case of sinusoidal forcing, the separation that has occurred is not between periodic and aperiodic dynamics: the forcing signal itself is, of course, periodic. But $\mathbf{u}(t)$ oscillates at a frequency orders of magnitude lower than those which dominate the Fourier power spectrum, and DMD is well known to be ill-equipped to capture dynamics at such disparate time scales [56]. The obtained linear model therefore ignores that slow oscillatory behavior and fully relegates it to the exogenous forcing.

4.4 Integration with DMDc: Discovering Linear Control Systems Without Prior Knowledge of the Forcing Signal

DMD with control (DMDc) extends traditional DMD to fit a model of the form [157]:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}(t). \tag{4.16}$$

The algorithm offers a means of simultaneous regression of the intrinsic linear dynamics $\mathbf{A}$ and the control mapping $\mathbf{B}$, but requires a known control signal $\mathbf{u}(t)$ in addition to the state data $\mathbf{x}(t)$. In many circumstances this information is not available, either because a known forcing variable cannot be measured or because the exogenous influence on the system is carried by many environmental variables which cannot even be enumerated, much less measured. The control architecture can be extended to a Koopman theoretic framework

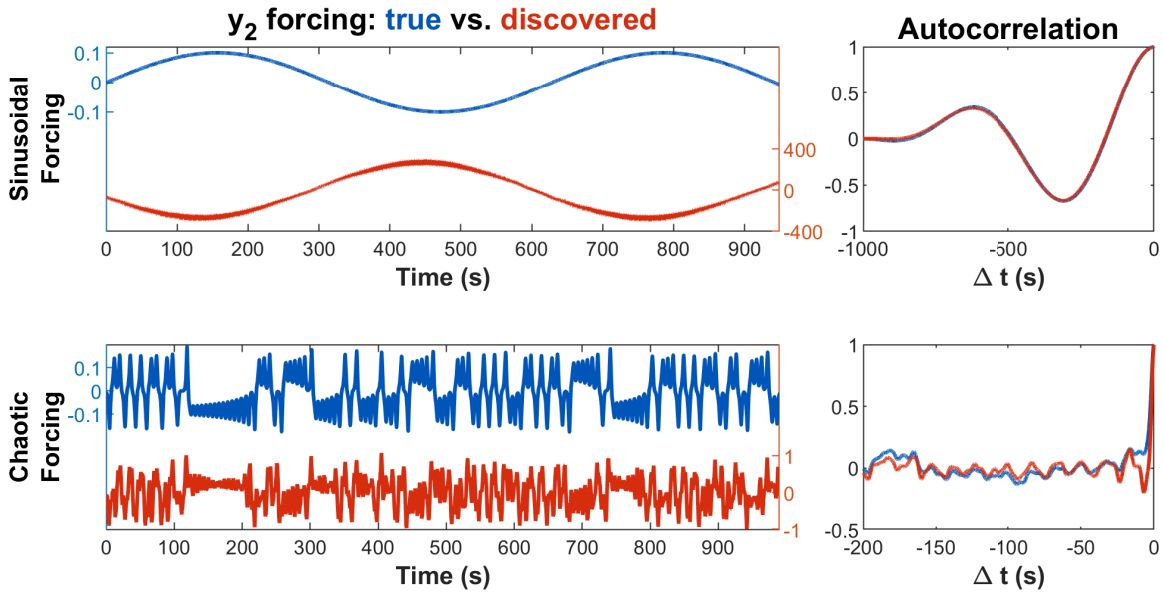

Figure 4.8: Discovered forcing signals (red) plotted alongside the true signals (blue) used in the simulations. Autocorrelation profiles (right) illustrate their similarity up to a global phase factor.

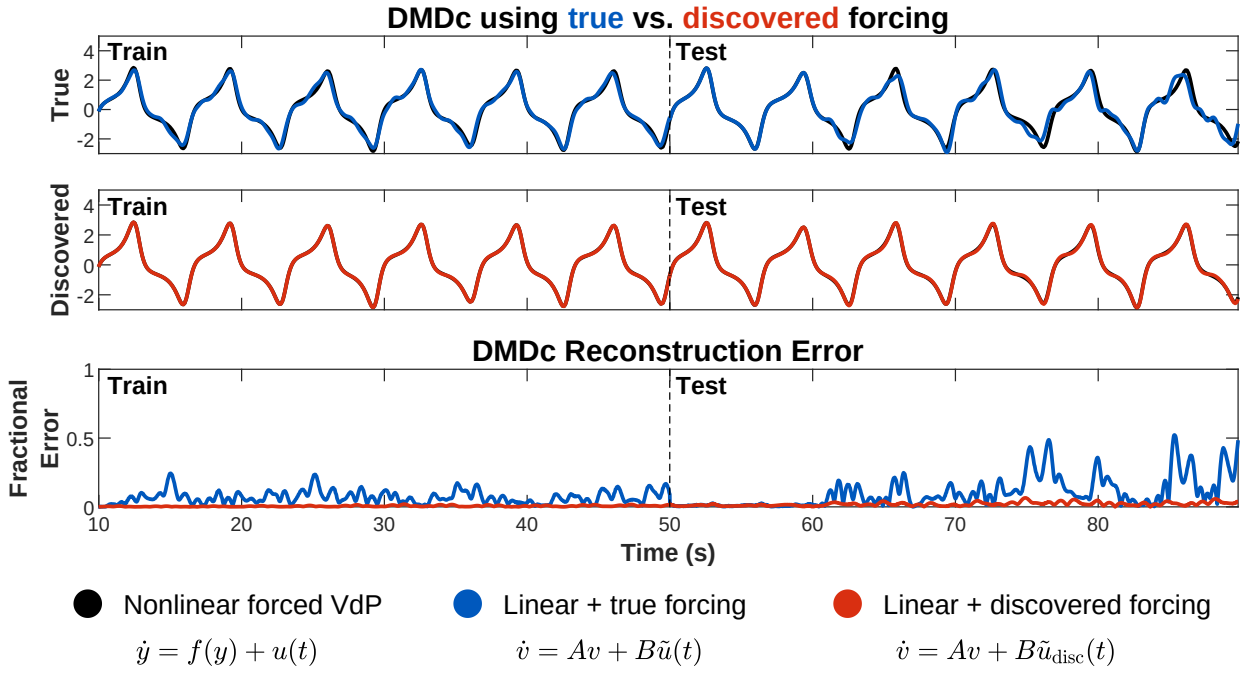

Figure 4.9: $u(t)$ is the control signal (in state space) used to generate the full nonlinear signal (black). $\tilde{u}$ is the same control signal lifted into delay embedding space and projected onto the top r SVD modes of the state Hankel matrix $\mathbf{H}$. $\tilde{u}_{\text{disc}}$ is the discovered forcing obtained by the procedure diagrammed in Fig. 4.7.

as well [104, 154, 92, 158].

Incorporating the PCT method for unsupervised forcing discovery into the DMDc approach allows for the construction of linear control models from time series measurements of the state alone. This is accomplished by first running the standard DMD algorithm and collecting the stepwise forecast errors as explained in the previous section, and then performing a second decomposition this time using the DMDc formalism to obtain a modified $\mathbf{A}$ approximating the intrinsic linear dynamics and a control mapping $\mathbf{B}$. The latter is not strictly necessary; the method by which the control signal is generated implicitly assumes equally-weighted forcing on all state variables, so working in the space of the truncated SVD on delay coordinates it can be assumed that $\mathbf{B}$ is the identity $\mathbf{I}_r$. There exists a variation on DMDc which assumes $\mathbf{B}$ is known, but even the general algorithm reliably discovers this to a good approximation, so there is little practical difference.

Figure 4.9 shows results from the application of this method to the chaotically forced Van der Pol system used in the previous section. Two DMDc models are trained: one using the ground truth forcing used in the initial simulation, and one using the discovered forcing obtained from the delay DMD method. In order to apply the original forcing to the model in rank-reduced SVD delay coordinates, it is transformed $\mathbf{u}(t) \in \mathbb{R}^2 \to \tilde{\mathbf{u}}(t) \in \mathbb{R}^r$ by time-delay embedding and projecting the result onto the top r PCT modes.

4.5 Application: Discovering forcing on real-world data with inherent periodicity

As an example of the practical diagnostic utility of this method, it is applied to power grid demand data published as part of the RE-Europe data set [89]. Electrical load is sampled hourly over a period from 2012 through 2014. Data is presented for many geolocated nodes of a reduced network representation of the European grid, which are averaged over to obtain a single time series for each country surveyed.

This subject was chosen as an example of a real-world data series which contains strong periodic signatures: daily, weekly, and yearly cycles are all quite obviously present, and in-

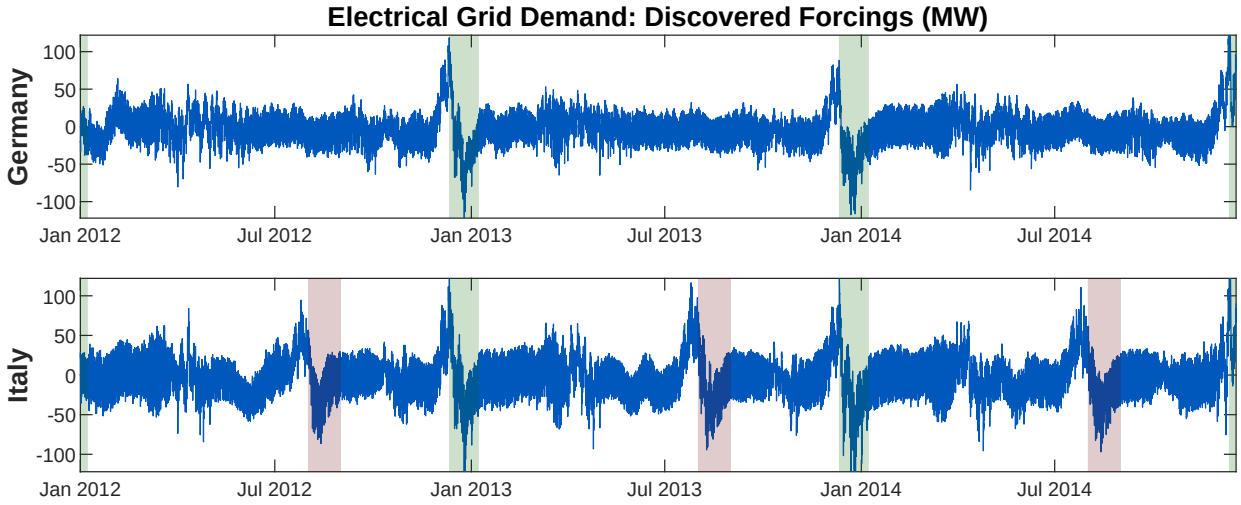

Figure 4.10: Discovered forcing signal for mean grid demand per network node in Germany (top) and Italy (bottom). The most obvious feature of the forcing signal is a large annual spike around Christmas in both countries, and an additional annual spike in late summer in Italy. Many businesses in Italy close in August. For reference, the weeks around Christmas (Dec. 11 - Jan. 8) are notated in green and the month of August (Aug. 1 - Sept. 1) in red.

deed dominate much of the observed behavior. But the dynamics are of course not entirely periodic; electricity demand is mediated by collective human behaviors, which are in turn influenced by weather patterns, economic trends, and uncountably many other variables. Casting these dynamics as a linear control system in time-delay coordinates effectively separates these factors, with predictable periodicities represented in eigenvalues of the DMD $\mathbf{A}$ matrix and everything else contained in the discovered control signal $\tilde{\mathbf{u}}(t)$. There is no ground truth against which to validate results, as the "true" forcing is not a measurable quantity, but some qualitative observations can be made to corroborate the idea that $\tilde{\mathbf{u}}(t)$ should encode anomalies in human behaviors as they pertain to electricity consumption.

Results are plotted in Fig. 4.10 for the load data from Germany and Italy. As discussed previously, DMD (like most numerical methods) cannot easily accommodate models with dynamics on highly disparate timescales, such as daily and yearly oscillations. To circumvent this issue decompositions are performed on a sliding window of approximately 2 weeks over the full sampling period. This is a sufficiently narrow domain to render seasonal variations negligible over the course of any given sub-series. Forcing signals are then inferred individually for each windowed iteration and then averaged to form a global $\tilde{\mathbf{u}}(t)$ for the full two-year span.

The most salient features in the results are large annual spikes in which forcing takes on an anomalously high value followed immediately by an anomalous low. Both nations exhibit this pattern once per year in December, and Italy additionally evinces a similar phenomenon (albeit slightly horizontally stretched) in the late summer of every year. This is highly consistent with known behavioral signatures in these countries: both celebrate Christmas with school holidays, business closures, etc., and August is widely observed as a national vacation month in Italy. That these breaks in human routine lead to by far the most prominent spikes in the learned forcing signals (while their influence is overshadowed by standard daily oscillations in the original data) suggests even in the absence of ground truth that this method has to some nontrivial extent successfully disambiguated the quasilinear oscillations which dominate the load data from anomalous activity resulting from more complex environmental

variables. This could have far-reaching implications for forecasting of time series such as these, in that it separates the more easily predicted, relatively deterministic dynamics from aperiodic factors which may be better modeled stochastically.

4.6 Conclusions

Since the seminal contributions of Takens, time-delay embeddings of dynamical systems have long been known to contain critical and representative information about the underlying dynamical system measured. In recent years, the advent of high-quality time-series measurements from spatio-temporal systems have afforded the community unprecedented opportunities for constructing data-driven models from such measurement data alone. Not only can time-delay embeddings extract meaningful information from unmeasured latent variables, but such embeddings can be used as a data-driven coordinate system approximating the Koopman operator. The work here advocates the use of these time-delay embeddings for constructing *principal component trajectories* (PCT) which provide a time-delay coordinate system which can be used to reconstruct dynamical trajectories via superposition. Indeed, the PCT provide an ideal representation of many dynamical systems and their time-series, especially in systems where an unmeasured latent space is critical to the dynamics. The method is shown to be intimately connected to SSA where various theoretical guarantees are available for infinite time-delay embeddings.

PCT also provide a coordinate system that can be used in conjunction with the DMD algorithm for producing good Koopman operator approximations. It can even be augmented to include external forcing terms in order to model the effects of a continuous spectrum. Such forcing terms are determined and constructed in a completely unsupervised fashion, allowing for a data-driven discovery process that can produce Koopman models which approximate both discrete and continuous spectral dynamics. A demonstration is offered of this method applied to real-world power grid load data to show its utility for diagnostics and interpretation on systems in which somewhat periodic behavior is strongly actuated by unknown and unmeasurable environmental variables. This example illustrates how a completely

data-driven method and PCT coordinates can be used in practice. It will be interesting to apply this approach to analyze other complex systems, such as for fluid flow control [29], which would benefit from the disambiguation of the linear dynamics and external forcing. Extending these approaches to more complex systems is an area of future research.

Chapter 5

ENGINEERING STRUCTURAL ROBUSTNESS IN POWER GRID NETWORKS SUSCEPTIBLE TO COMMUNITY DESYNCHRONIZATION

Disruptions of power grid systems can have a severe, negative impact on performance and lead to *Coherent Swing Instability* (CSI) [175, 180, 176], whereby a subset of machines in the grid lose synchrony with the rest of the network, thus shutting the entire network down and leading to unacceptable blackouts. CSI is in essence a manifestation of community structure in a networked system, a collective dynamical divergence of one subgroup of nodal oscillations from another. Instigated by the work of Girvan and Newman on modular structures in networks [77, 142, 141], a great deal of attention has been devoted to the development of methods for identification and characterization of modular components. *Community detection* has become a broadly-defined term which is used to refer to a variety of such approaches. A comprehensive review of its uses is presented by Schaub et. al. [167]. Much of the prominent work in the field has focused on topological approaches, i.e. methods which take as input an adjacency matrix describing a (weighted or unweighted, directed or undirected) graph. As Schaub notes, however, dynamics on a network are constrained by topology but cannot be fully described by it.

This distinction has led to work on community detection methods which place primary importance on dynamics. However, as many networks of interest lack a known set of laws governing their evolution (e.g. neurological, social, transportation, epidemiological, etc.), much of this research has modeled dynamics with Markovian diffusion processes describing random flows on the network [52, 160, 12]. Power grids have an advantage over these systems in that the physics of generator oscillations and current transmission are well-understood:

taken in isolation, each node of the network behaves predictably. Only when they are combined on a complex and non-symmetric graph structure do they begin to exhibit the group behaviors which resist simple characterization. The results presented in this work capitalize on this property by applying a dynamics-focused community detection perspective to simulation data generated by a realistic, machine-level power grid model.

Previous work on applying network-topological analysis to real-world power grids has met with mixed success. The extensive theoretical framework that has been developed in the field of Complex Networks has given rise to a variety of methods which assess functional properties of power grids directly from their network topology, using metrics such as node centrality, betweenness, degree distribution, community structure, and clustering coefficients [41, 1, 23, 174, 205, 201, 151, 152]. However, topological methods alone have consistently failed to fully account for observed network vulnerabilities [20, 22, 150, 85]. This has led to a variety of hybrid approaches which incorporate electrical properties of the system not captured by its graph structure, often by using them to assign edge weights or to compute modified versions of existing topological metrics [140, 85, 202, 22, 21]. The goal of these studies has generally been to develop a heuristic to estimate a power grid's vulnerability to failure, which they then validate using historical data or numerical simulations. This work presents an alternative method which evaluates the functional consequences of structural modifications directly from simulation data. While this approach has previously been applied to steady-state systems [13], its use for grid disturbances has generally been dismissed as infeasible due to the combinatorially large search space of possible faults and network structures [85]. The contribution of this work is a methodology for incorporating topological properties (specifically, community structure) and statistics on measurement data to significantly reduce this space. This technique identifies a set of candidate locations for single-line additions to an existing network; this is a small structural perturbation relative to the size of the full grid and therefore unlikely to significantly change its global topological characteristics, such as degree distribution. Robustification occurs not by directly tuning some topological (or hybrid topological/electrical) metric, but rather by using such a metric to inform a minor

structural modification. This lends itself to practical engineering application, as it suggests an inexpensive change that could be made to an existing grid rather than a design principle for the construction of a new grid from the ground up.

The data-driven approach introduced here is motivated by the proliferation of real-time monitoring strategies for power grids that have been deployed in recent years [132, 98, 15], with event location strategies gaining increasing attention in order to localize pernicious effects [123, 131, 18]. These strategies are directed to provide a system-wide awareness of events such as faults and other disturbances, taking advantage of the increasing coverage of *wide area measurement systems* (WAMS) technology which enables the implementation of wide area emergency and restorative control applications [123, 131, 18]. Even though these strategies have achieved positive results, additional technical challenges arise as the modern WAMS-generated data become high dimensional and more distributed thorough large areas of the system. This work proposes an additional technique for power grid network robustification. Specifically, it is shown that the robustness of the network can be diagnosed from ensemble fault simulations. Moreover, the power grid can be made significantly more robust to disturbances with proper engineering of the network design and attention to the community structure observed in instances of CSI. Such considerations are critical in considering future power systems deployments, or for upgrading current networks in order to circumvent susceptibility to CSI.

This chapter details a simulation model used for characterizing the power grid dynamics and disruptions. It further introduces a procedure for dynamics-based community detection based on the results of these simulations, showing through diagnostic tools that the network's sensitivity to CSI depends on the location and severity of a fault. This provides an engineering approach capable of characterizing network connectivity modifications capable of robustifying the network to decoherence.

The work presented in this chapter is based on a paper entitled "Engineering structural robustness in power grid networks susceptible to community desynchronization Daniel" written in collaboration with Xiu Yang and Alexandre Tartakovsky of the Pacific Northwest

National Laboratory, and J. Nathan Kutz of the University of Washington. The paper was published in Applied Network Science in 2019 [57].

5.1 Numerical Simulations of Power Grids

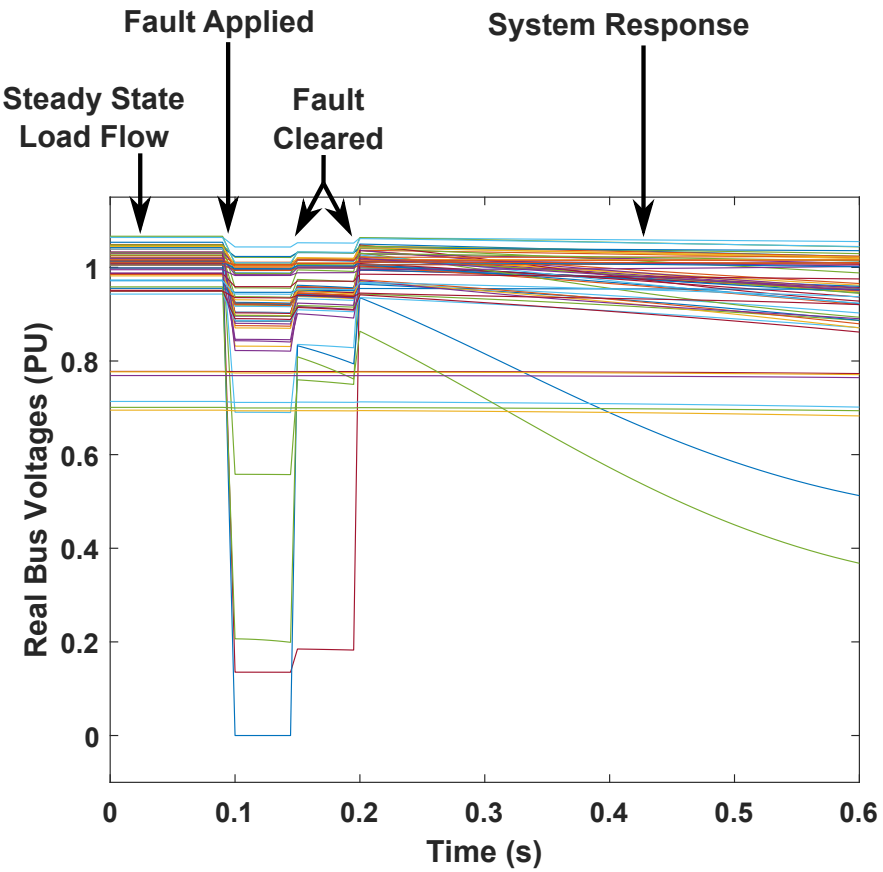

Figure 5.1: Stages of a grid simulation in PST: The system is in a steady state until a fault is applied at 0.1s. The fault is cleared in two stages, at 0.15 and 0.2s. After the fault has been cleared, the system has been perturbed from its steady state so it continues to evolve dynamically

This section highlights the numerical simulation architecture used to evaluate power grid systems and their connectivity structure. Importantly, a prescription of the disturbances

applied to the network to induce CSI is considered in order to evaluate the network robustness in a principled way.

5.1.1 The Power System Toolbox

Power grid simulations for this study were produced using Power System Toolbox (PST), a Matlab software package originally developed by Kwok W. Cheung and Joe Chow of Rensselaer Polytechnic Institute [36]. Supplied with both the topological structure of a power network and the specific electromechanical parameters of the grid's generators, nodes, and lines, PST performs dynamic simulations of both steady-state and nonequilibrium dynamics. This study employs system fault simulations structured as follows:

1. The system is initialized in a steady state for power flow through the network by solving the nonlinear algebraic network equations which specify the load flow problem.

2. The dynamic portion of the simulation is initiated by applying a transient three-phase fault to a single line in the network, as though the line were brought into momentary contact with a grounding object such as a tree.

3. When a fault occurs, power system protection equipment acts to isolate the disturbance. If the fault is transient, the line can be reconnected after a short time. PST treats this as a two-step process, clearing the fault first at the near end and then at the remote end of the line. These two time intervals, labeled here as τ_1 and τ_2, are supplied as input parameters in simulations. Figure 5.1 illustrates the two time scales, τ_1 and τ_2, for a simulation of the northeastern power grid system. The dynamics of interest unfold following the fault application and fault clearing time scales.

4. When the fault is cleared, the grid recovers its full original network structure. The fault has perturbed it from its initial steady-state configuration, so dynamic evolution continues. The simulation is carried out over a long time (relative to the fault duration)

and analyze the network's response to the disturbance. The CSI often is induced by a disturbance event (fault application) for which the power grid system does not recover to its original stable (steady-state) behavior.

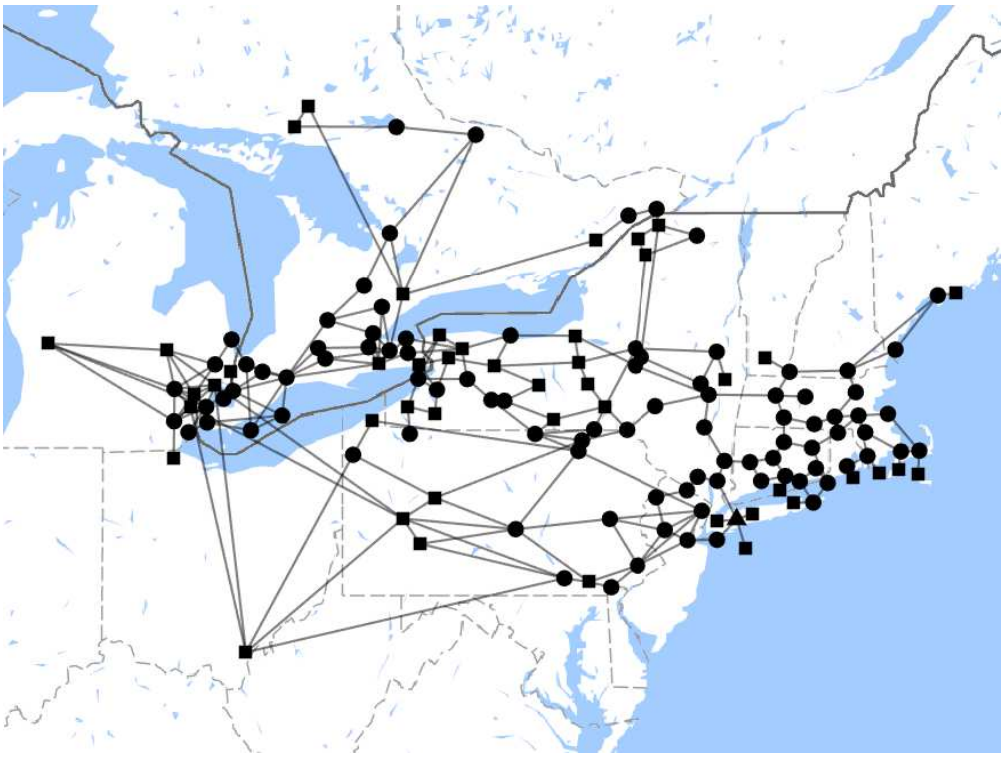

Figure 5.2: The NPCC 140-Bus system is a reduced model for the power grid of the north-eastern United States and Canada, plotted here overlaid on a map of the region. In this figure only, bus types are distinguished by the nodes' symbols (circles are load buses, squares are generator buses, and triangles are swing buses).

5.1.2 The NPCC 140-Bus System

Simulations in this study are carried out on the NPCC 140-bus test system, which is a reduced model based on the power grid of the North American northeast (Fig. 5.2). This

network was chosen because its machine parameters are representative of those in a major real-world grid and its graph structure is sufficiently large and complex that it gives rise to coherent dynamics at subnetwork level.

5.2 Discovering Community Structure in the Grid

Simulations of the northeastern power grid are sufficient to illustrate many of the key features of power grid networks and their induced CSI. By perturbing the various nodes of this specific network, one can characterize the instability structures, and their commonalities, induced in the power grid dynamics.

5.2.1 Identifying Coherent Swing Instabilities

The focus of this specific study is the phenonon of CSI, in which a subgroup of buses which are strongly coupled to one another, but only weakly coupled to other nodes, collectively lose synchronicity with the remainder of the network. Real-world power systems implement controllers to damp the oscillations of relative rotor angles which give rise to CSI, but no such safeguards are implemented in the PST simulation toolbox. The onset of CSI therefore is manifested as a clear qualitative transition in the dynamics of a subset of the buses. Specifically, a group of machines will begin to oscillate with linearly increasing frequency while the remainder of the network continues to evolve with dynamics on a slower and roughly constant time scale (Fig. 5.3). Although the PST model does not accurately model the behavior of the unstable network, since local controllers would activate to dampen growing oscillations, it does highlight the lack of robustness of the network to the intrinsic dynamics induced by the disturbances. By engineering a more robust system, the intrinsic dynamics itself acts to stabilize the system. This aspect of engineering a power grid network is considered below in the section entitled *Engineering Network Structure to Reduce CSI*.

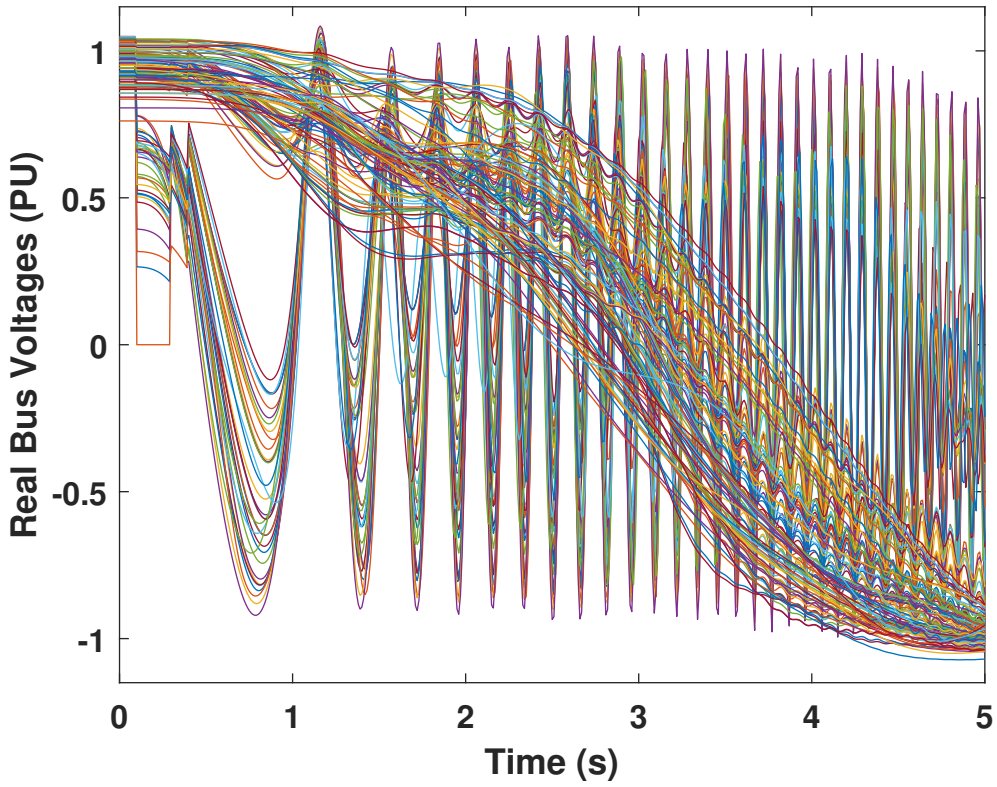

Figure 5.3: CSI in numerical simulations: Plotting bus voltages shows two qualitatively different types of dynamics. Each bus belongs to one of two coherent groups, one of which varies at a slow, consistent time scale and the other of which oscillates with linearly increasing frequency.

5.2.2 Community Detection

To investigate the incidence of CSI, faults are systematically applied to each line of the network in succession to track which buses (if any) exhibit unstable dynamics as a result. The results, plotted in Fig. 5.4, suggest that the majority of instabilities take place in a particular subgroup of buses. To formally characterize this structure it is approached as a network community detection problem: by treating the buses and fault locations as two disjoint populations of nodes with (unweighted, undirected) connections given by the nonzero

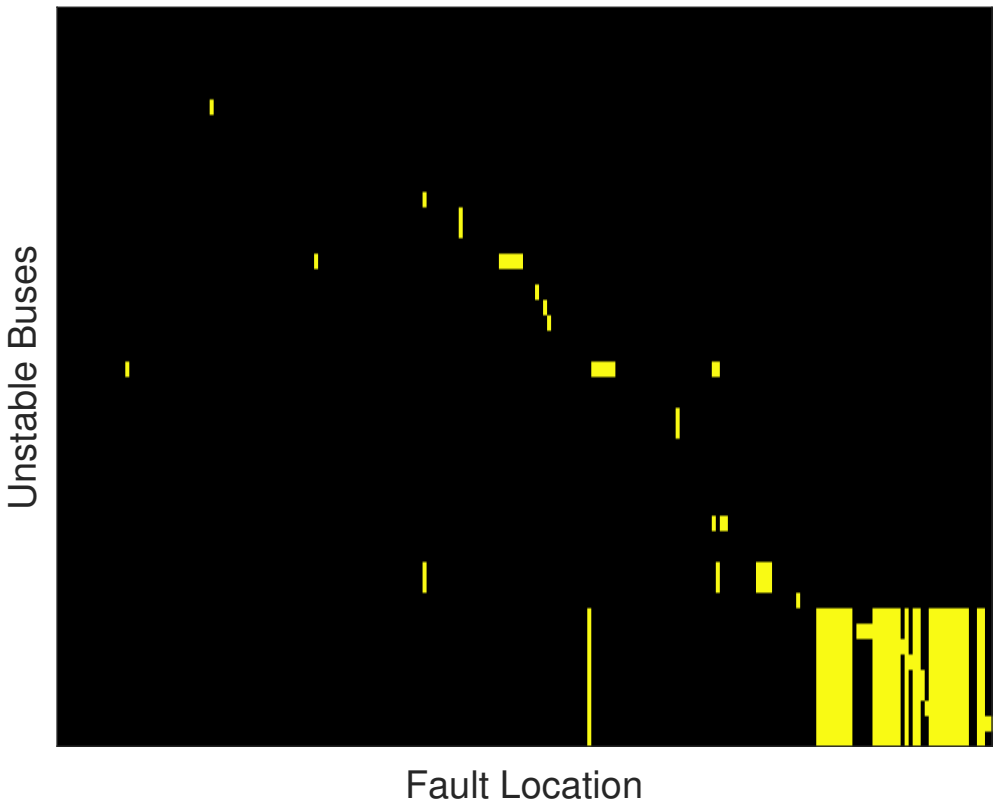

Figure 5.4: A systematic exploration of the network locations of unstable dynamics as a function of the line where the initial fault was applied. Each column represents a different simulation with a different fault location, with yellow pixels denoting which buses exhibited instability during that run.

elements in Fig. 5.4, one can build a square adjacency matrix which casts the results as a bipartite instability matrix (with $(N_{\text{bus}} + N_{\text{line}})$ rows and columns).

In this form, the results are amenable to any network-topological community detection scheme desired. Results presented in this chapter use the Adaptive BRIM algorithm, developed by Michael J. Barber for modularity-based community detection in bipartite networks [14]. Very similar results were obtained using other bipartite algorithms, including those introduced in Newman (2008) [141] and Liu and Murata (2009) [126].

The Adaptive BRIM algorithm detected 12 distinct communities from the bipartite instability matrix. However, many of these were quite small. In the interest of restricting focus to network-wide swing instabilities, communities containing fewer than 2 buses or lines were reassigned to a neighboring cluster based on a vote of graphical nearest neighbors. This led to a reduced population of 3 dominant communities, plotted in Fig. 5.5.

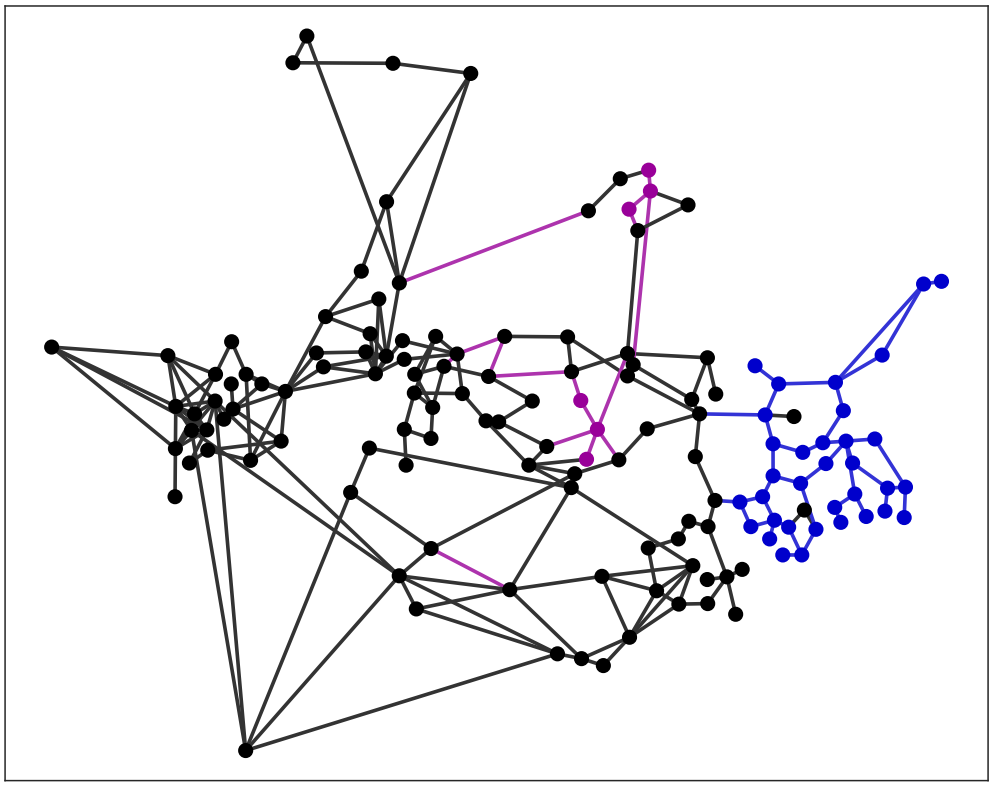

Figure 5.5: Results of Adaptive BRIM community detection on the adjacency matrix of the bipartite instability graph. Small communities (containing fewer than 2 lines or buses) have been subsumed into neighboring clusters.

The blue community, which corresponds geographically to New England, is a contiguous and highly interconnected nodal group with few connections to the rest of the network. A simple topological community detection scheme applied directly to the NPCC-140 graph

would almost certainly have identified it as a highly modular cluster. The purple group, however, is thoroughly embedded within the larger black nodal population. This is a testament to the importance of a dynamics-based approach to community identification: modular graphical structure does not necessarily equate to dynamical coherency.

5.3 The Sensitivity of Network Connections

Given the diversity of dynamics observed for disruptions of the network, this work's analysis aims to understand the sensitivity of each node in the power grid to fault tolerances. By varying the fault severity, the nodes can be ranked by their susceptibility to CSI.

5.3.1 Varying Fault Severity

To identify the lines where a system fault is most likely to generate CSI, responses are measured for faults of varying intensity. Faults in PST simulations are parameterized by two time durations: τ_1 from the application of the fault to the clearing of the near end, and τ_2 from the clearing of the near end to that of the remote end. Generally speaking, a longer fault duration drives the system farther from its initial steady-state configuration, increasing the likelihood of instability. Indeed, the parametrization of faulty intensity through the (τ_1, τ_2) parameter space allows for characterization of the robustness of each node.

5.3.2 Ranking Lines by Sensitivity

Working in (τ_1, τ_2) parameter space, a domain is identified which captures the onset of instability for most fault locations. This gives a range of values broad enough so that the lowest values of τ_1, τ_2 yield fully stable dynamics, while the highest values of τ_1, τ_2 lead to instability at many fault locations. The precise choice of bounds is somewhat arbitrary as long as they meet these criteria; the performance over the specified (τ_1,τ_2) region will only ever be used as a relative metric for comparing lines. For each fault location, simulations are repeated over a grid in this domain of parameter space. The performance of each run is quantified by determining the number of buses which go unstable.

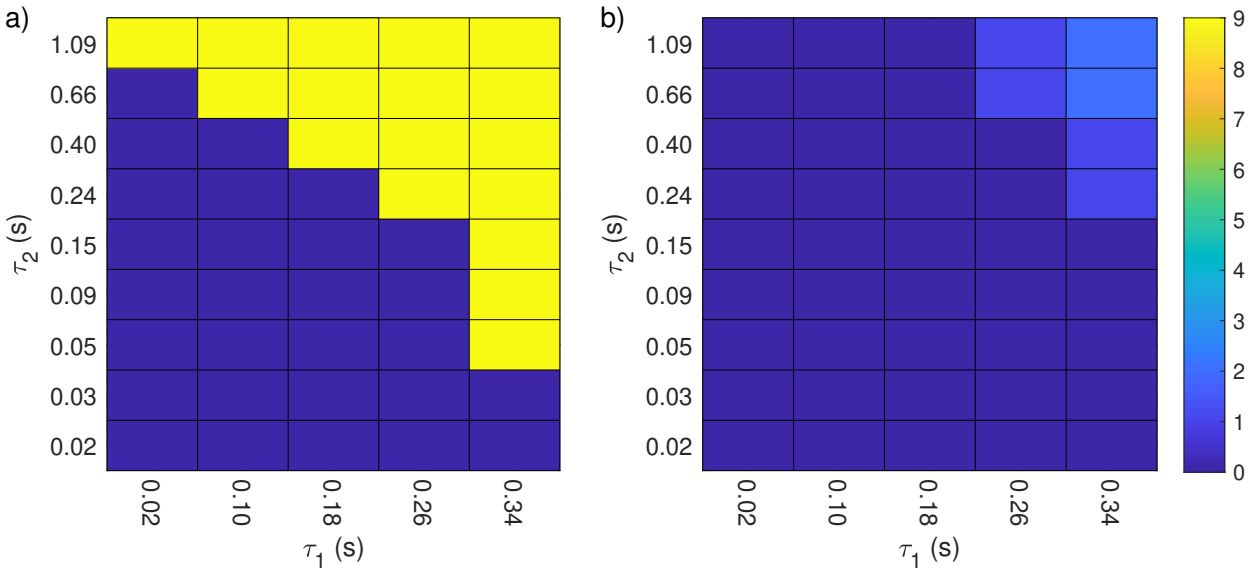

Figure 5.6: Sample results of simulations over a region of fault-time parameter space. Each box represents a simulation with fault times (τ_1, τ_2), with its color denoting the number of generator buses which exhibited unstable dynamics. A set of simulations like this was carried out for each line in the network. Examples are presented for a) a fault location likely to induce CSI, and b) a fault location unlikely to induce CSI. Note: the color scale is produced by ennumerating unstable *generator* buses (just those marked by squares in Fig. 5.2) so a count of 9 corresponds to the whole New England subnetwork.

The result of this analysis gives an "instability frontier" in the space of (τ_1, τ_2). As visualized in Fig. 5.6, the more sensitive lines of the network (e.g. Fig. 5.6a) have a frontier which extends farther down toward the bottom-left corner, whereas more robust lines (e.g. Fig. 5.6b) are fully stable until comparatively high values of (τ_1, τ_2). In the former case, it is observed that a small perturbation to the fault parameters can lead to a big jump in the number of unstable generators. This foreshadows the crucial role of community structure in understanding this behavior: instability often occurs collectively in coherent subnetworks.

Having performed exhaustive simulations for all possible fault locations, lines are ranked

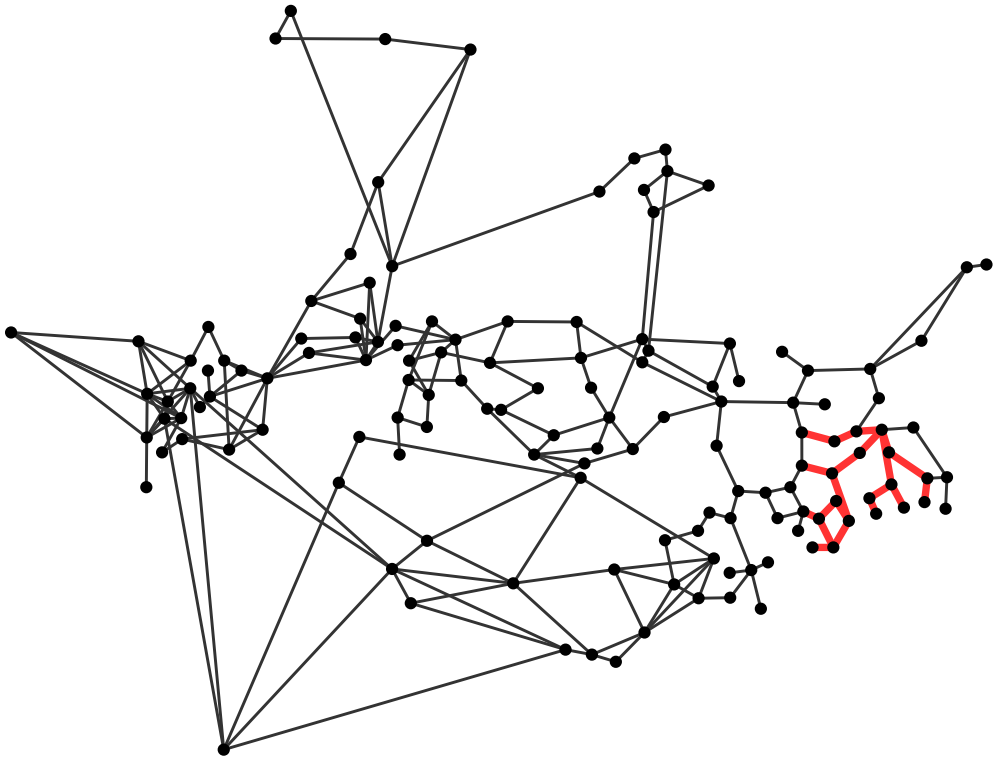

Figure 5.7: Highlighted in red: the twenty most sensitive lines of the network, i.e. those where a fault is most likely to induce CSI

according to sensitivity by averaging over the unstable nodes obtained at all (τ_1, τ_2) parameter combinations. The numerical values obtained are of course sensitive to one's choice of the fault-time domain, but they serve as a functional metric of comparison between different lines tested on this domain. Figure 5.7 illustrates the results of this process, with the 20 most sensitive lines highlighted in red.

These are the lines whose performance the remainder of this study attempts to improve. The network modifications considered in the next section belong to a combinatorially large space, so it is necessary to restrict the scope of analysis wherever possible to avoid having to do a prohibitively large number of simulations. Restricting simulations to these "worst offender" fault locations allows the parameter space to be constrained while still treating the

cases which are of greatest practical concern with respect to grid stability.

5.4 Engineering Network Structure to Reduce CSI

Although the specific criteria which lead a subgroup of buses to coherently desynchronize in a given network are not generally well understood, it is clear that the observed phenomenon of community structure is intimately related to the connectivity configuration of the network. Grids are prone to CSI when they contain a subnetwork which is relatively weakly coupled to its surrounding nodes. Thus a naive approach to engineering network stability would be to simply add connections between nodes inside and outside this instability prone community. The results presented in this section not only support this intuition, but also show that not all inter-community line additions yield significant improvements to stability. As such, the full simulation-based approach implemented here is necessary to determine *which* inter-community connections contribute the most to the grid's structural robustness.

5.4.1 Network Modification Protocol

The approach for assessing how the addition of a transmission line affects the incidence of CSI is as follows:

1. A line connecting the two chosen buses (with resistance and reactance specifications taken to be the median of those of the existing lines) is inserted into the PST network specifications.

2. For each of the most sensitive lines in the unmodified network (i.e. those highlighted in Fig. 5.7), simulations are carried out for all (τ_1, τ_2) fault-time combinations in the domain identified previously to obtain an instability frontier (denoted in Fig. 5.8 by a black dotted line).

3. Overall performance for each fault location is again obtained by averaging over the sum of unstable generators. These values are then averaged over all tested fault locations

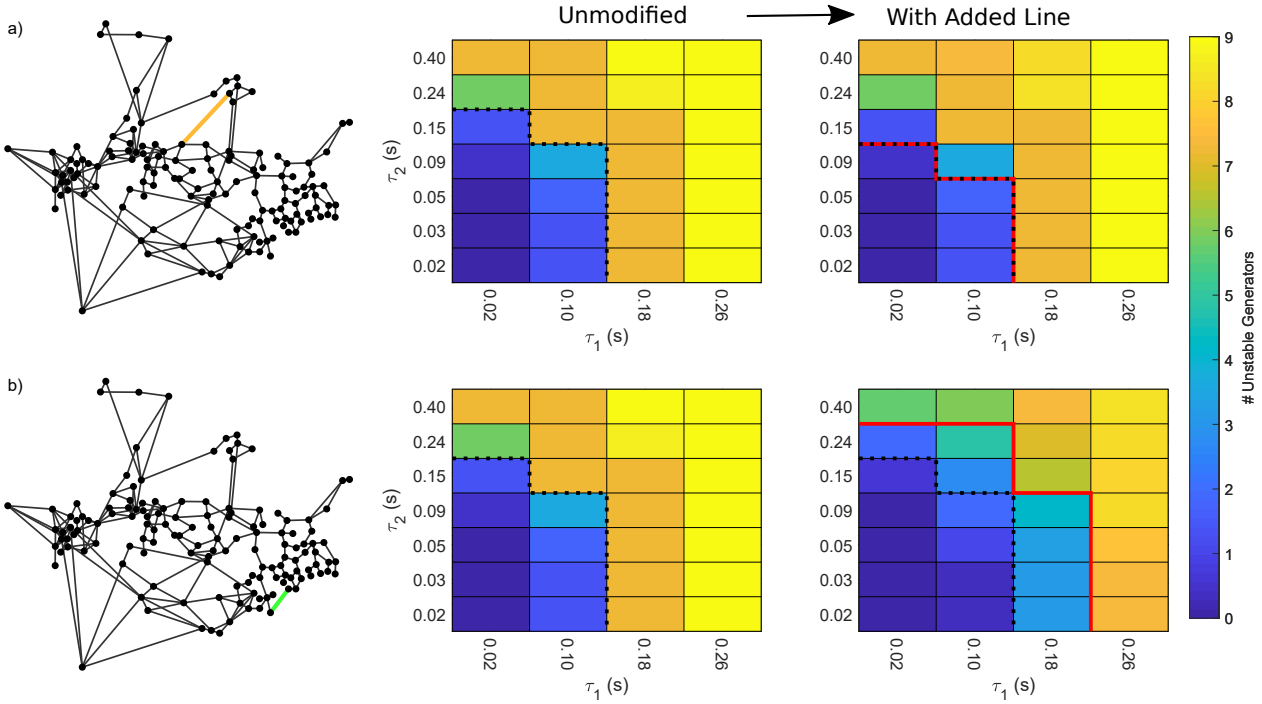

Figure 5.8: For each line addition considered (connections highlighted in color on the left), the mean instability frontier of the unmodified network (center, dotted black line) was compared to that of the network with the new connection (right, solid red line). Some candidates afforded little to no improvement ((a), with added line in orange), while others robustified the network so that it was resilient to considerably more severe faults ((b), with added line in green).

to obtain a single plot of the modified network's susceptibility to CSI (shown in the right column of Fig. 5.8).

4. The result of this process is compared to that of the unmodified network to obtain a ratio measuring the stability improvement afforded by the added connection.

The set of possible single-line additions to the network is combinatorially large ($N_{\text{additions}} = N_{\text{bus}}\,(N_{\text{bus}} - 1) - N_{\text{line}} = 19367$, in the case of the NPCC-140 system). The number of simu-

lations necessary to carry out the above steps for a single network modification is such that it is not computationally feasible to test all possible cases.

Results are presented for a semi-randomly selected subset of these possibilities ("semi-random" because they were chosen with a penalty on geographic distance between the buses to be connected, so as to better conform to practical engineering considerations). These candidates are separated into two groups: those which run between two distinct communities, and those which are internal to a single community.

5.4.2 Network Modification Results

The candidate lines tested using the above protocol are pictured in Fig. 5.9, colored according to their performance relative to the original network configuration. Additionally, the distributions of the stability change parameter for each of the two populations of line candidates are presented as histograms and box plots in Fig. 5.10.

It is observed that every line addition that significantly impacted the network's susceptibility to CSI is one that connects two distinct communities. This suggests that understanding the network's community structure is crucial to identifying candidates for new lines to robustify the grid. The search space of inter-community lines is substantially smaller ($N_{\text{additions}} = \frac{1}{2}\sum_{i}^{N_{\text{community}}}\left[N_{\text{bus}}^{i}\left(N_{\text{bus}}^{\text{tot}} - N_{\text{bus}}^{i}\right)\right] - N_{\text{line}}^{\text{inter}} = 4191$, in the case of the NPCC-140 system, which represents a 79% reduction of the search space). Coupled with practical distance considerations (and any other logistical criteria such as local topography or infrastructure), the number of line candidates for which this extensive simulation process must be carried out becomes much more tractable.

Within the inter-community population of line candidates, the criteria for significant robustification are not obvious. Examining the relationship between performance and common topological metrics such as betweenness and degree centrality reveals little to no correlation, which supports the literature consensus [20, 22, 150, 85] that these measures are insufficient as predictors of behavior in real power systems. Nonetheless, the dynamical community detection method successfully identified a number of lines which do offer a marked improve-

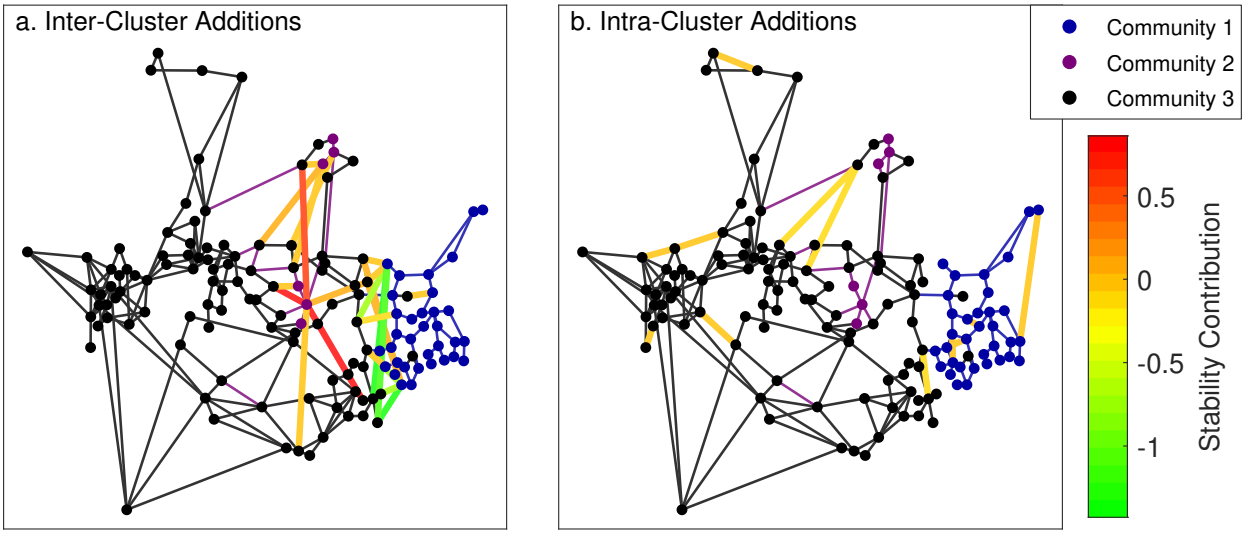

Figure 5.9: The NPCC 140-bus system with colored lines representing candidates for additions to the network. Their coloring denotes the extent to which they improved network stability relative to the original network. Green lines afforded the greatest improvement, while red lines left performance largely unchanged. The subnetworks obtained by community detection are colored in black, blue, and purple for reference. Line additions have been separated into a) connections between distinct communities, and b) connections within a single community

ment even by testing a very small subset of all possible inter-community connections. This shows that the approach to evaluating robustness of the network presented in this study can effectively identify potential connections capable of robustifying the power grid network.

5.5 Conclusions

This work has introduced a computational framework for the analysis of the network level dynamics and stability of power system disturbances. This analysis is critical for understanding how the network architecture itself can lead to subgraphs (communities) that are highly susceptible to CSI. By systematically parametrizing disturbances according to the temporal

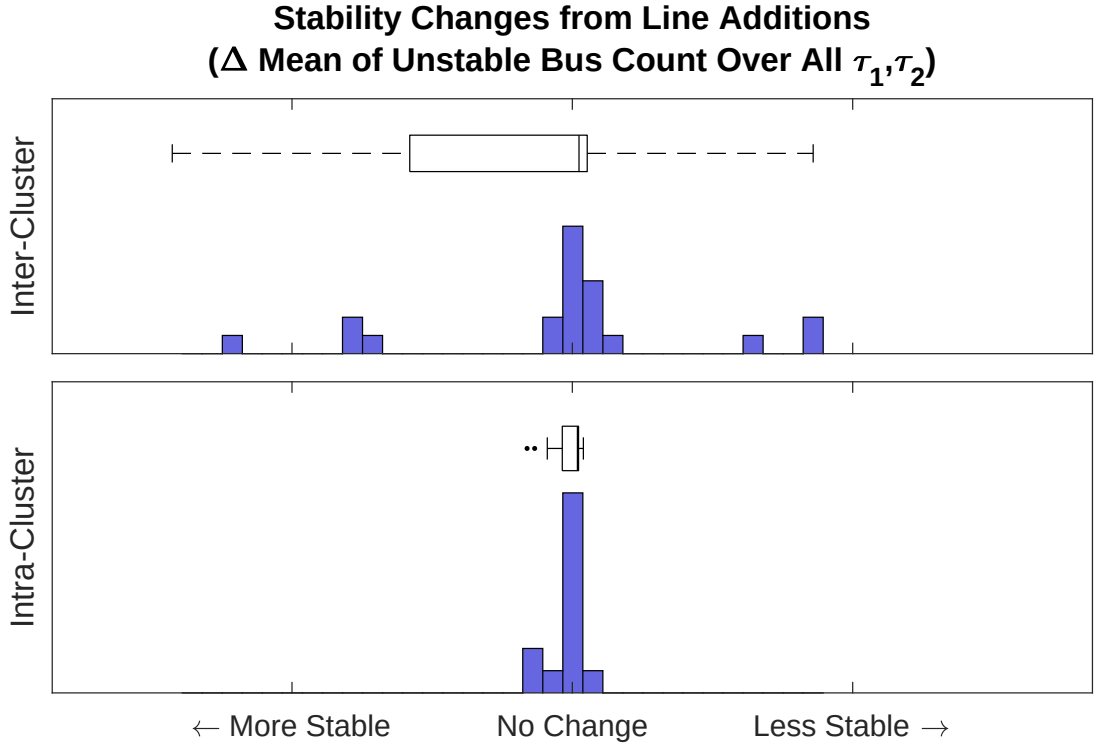

Figure 5.10: Distributions of stability impact parameter for a) inter-community connections, and b) intra-community connections. From the histogram outliers and the spread denoted by the box plot width, it is clear that the line additions between distinct communities are much more likely to effect significant changes to network stability (positive or negative).

parameters τ_1 (the time of the fault until it is cleared) and τ_2 (from the clearing time of the near end to that of the remote end), one can assess the effect of each node on the overall stability of the power grid network.

Evaluation metrics have been defined to (i) identify dynamics-based community sub-structures in the power grid network, (ii) determine weak points in the network that are particularly sensitive to faults, and (iii) produce an engineering approach for the addition of transmission lines to maximally reduce the incidence of CSI. For the example of the Northeast power grid, a strong dependence has been identified of line sensitivity on the New England

subnetwork. It has been shown how modifying the network's connectivity structure can robustify the network to CSI. The space of possible connectivity changes is combinatorially large, so modifications are constrained to a tractably small subset of single-line insertions to the original network. This work has shown that community detection can be used to substantially reduce this search space, as line additions which connect separate communities tend to much more significantly impact global stability. The addition of a line can markedly improve the grid's resilience to CSI, but the success is highly dependent on location within the network.

This analysis provides a versatile diagnostic for the efficacy of adding a particular line to a power grid which is known to be prone to CSI. This is a particularly relevant problem in large-scale power systems, where improving stability by increasing overall network connectivity is not feasible due to financial and infrastructural constraints. The approach presented in this study focuses principally on network topology, so its results are fairly robust to small variations in the model parameters used in simulation. This makes it a strong candidate for use in analyzing real-world power systems, as connectivity structure is a characteristic which can be perfectly reproduced in the translation from physical system to simulation model.

Chapter 6

FUTURE DIRECTIONS: GENERALIZATION AND FORECASTING

As with most data-driven methods, the results presented in this thesis leave open questions about their generalizability. The success of an approach on a few tested data sets does not guarantee its efficacy on any new test case. This is a core challenge in all machine learning applications, and presents a particularly formidable obstacle for extrapolation tasks. An algorithm trained on some collection of measurements is much more likely to succeed on a new, yet-unseen data point if that point lies within a domain that is already well populated in the training set; the new output can be estimated by interpolating between the nearby observations that have already been made. It is when a new input lies outside the training region that the learned function is forced to extrapolate, and performance in this regime is much less reliable. In a sense this problem is unavoidable: a purely data-driven method learns rules only from its observations, and an input from a region in which no measurements have been made forces it to generalize those rules without evidentiary guidance. The manner in which this generalization is carried out is necessarily non-universal; the information content of a function sampled on some compact input subspace is insufficient to reconstruct its behavior outside that subspace with complete certainty. Whereas for interpolation a conservative assertion of continuity and smoothness properties can typically assure a degree of accuracy (assuming enough training points have been provided), extrapolation demands more aggressive intervention by the user. Generic methods for distant extrapolation are therefore mostly unreliable, but nearby extrapolation is a more tractable problem and remains a subject of great interest in the field of machine learning.

For the time series analysis problems on which this thesis has primarily focused, the

desired extrapolative result is usually prediction of the future state of a system. This is naturally a task of great practical concern, and the nature of temporal causality ensures that it can never be brought into the interpolative regime (the future cannot be measured from the present). In many cases time series forecasting presents a better-posed problem than fully generic extrapolation because training data often exhibits constraining properties such as stationarity, boundedness, smoothness, etc. For any given learned reduced dynamical representation (such as those presented in Chapters 3 and 4), the task of prediction amounts to understanding how such constraints are encoded and how the forward-time mapping (e.g. DMD) introduces error over time. The inherent difficulty of long-term forecasting requires that any proposed method offer some evaluation of its horizon of accuracy, typically by gathering statistics on its deviation from the truth over many trials or by presenting an ensemble of predicted trajectories to form a future-state probability distribution which diffuses over time. Implementation of such a scheme offers a fertile research area for extension of the DMD methods introduced in this thesis.

In the case of the Principal Component Trajectory decomposition outlined in Chapter 4, the forecastability of a system is in some sense measured by its discovered representation. The PCT DMD method separates out the (quasi)periodic, spectrally discrete portion of a signal, whose intrinsic dynamics can be accurately extrapolated to arbitrary future time, from an external forcing on these dynamics which by construction contains all content of the signal that frustrates linear extrapolation. Represented thus, the challenge of forecasting lies entirely in the task of predicting the latter component. Moreover, all forecasting error lives in the forcing (again, by construction). This suggests that this decomposition may offer a powerful tool for future state prediction in systems where the forcing is a weak-to-moderate perturbation to the underlying linear model. In research which is ongoing at the time of submission of this thesis, forecasting is carried out by modeling the forcing signal as a stochastic process and using that to actuate the deterministic evolution of the periodic dynamics. With repeated stochastic realizations, it is expected that an ensemble of future trajectories can be constructed which shadows the unforced DMD extrapolation but allows

for amplitude and phase aberrations to occur in statistical accordance with their historically observed incidence. Recent results have shown the promise of Gaussian Process Regression (GPR) as a technique for estimating the forcing from the current delay-coordinate state of the full system.

The multiscale decomposition approach discussed in Chapter 3 also shows potential for utility in forecasting applications in systems where dynamics play out on multiple disparate time scales simultaneously. This was explored briefly in Sec. 3.4, wherein an ensemble Kalman filtering approach was used to demonstrate how data assimilation might be carried out separately on each discovered scale component. While this produced promising results for the three body planetary system, it could be rendered much more interpretable by integration with the linear control approach introduced in Chapter 4. The Kalman ensemble was constructed by gathering statistics on DMD eigenvalues, eigenvectors, and weight coefficients over all previous windowed iterations and generating a collection of autonomous linear systems which produced a distribution of possible future trajectories. This produces a highly constrained ensemble, in which members are required to take the form of complex exponentials (or linear combinations thereof). Incorporating the control system approach used with PCT decompositions would offer a much more diverse function space from which to sample these trajectories. A similar Kalman filtering technique could be applied, but using a model built from statistics on the forcing rather than those on the windowed DMD parameters. This would preserve the scale-separated property that makes this method so powerful, but offer a more flexible and robust framework for individual forecasting of time scale components.

More broadly, future extensions of these results should build on their capacity for deconstructing time series data based on its forecastability properties. As discussed in Secs. 2.3 and 4.3, Koopman theoretic modeling approaches are limited by the inability of finite-dimensional linear systems to universally reproduce nonlinear dynamics. From the perspective of future state prediction, this discrepancy seems to map onto the question of long-term forecasting. Those elements of a signal which recur on a periodic or quasiperiodic basis and are amenable

to finite-dimensional Koopman representation are also the only elements which can be reliably extrapolated into the far future. As such, Koopman methods offer a promising angle for frameworks of forecasting developed with explicit differentiation between components of the state which are predicted with differing levels of confidence. The work presented in this thesis (along with a number of referenced DMD and Koopman papers which preceded it) constitutes a foundation for this prediction methodology, but there remain a multitude of unanswered questions which will need to be addressed if a fully fledged forecasting framework is to emerge from it.

BIBLIOGRAPHY

[1] Réka Albert, István Albert, and Gary L. Nakarado. Structural vulnerability of the North American power grid. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69(2 2):1–10, 2004.

[2] Armando Albertazzi, Analucia V. Fantin, Allison F. Maia, Daniel P. Willemann, Mauro E. Benedet, and Matias Viotti. Use of generalized n-dimensional lissajous figures for phase retrieval from sequences of interferometric images with unknown phase shifts. In Wolfgang Osten, editor, *Fringe 2013*, pages 191–196, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[3] I. Antoniou. The time operator of the cusp map. *Chaos, Solitons and Fractals*, 12:1619–1627, 2001.

[4] I. Antoniou, S. A. Shkarin, and E. Yarevsky. The resonance spectrum of the cusp map in the space of analytic functions. *Journal of Mathematical Physics*, 43(7):3746–3758, Jul 2002.

[5] Ioannis E. Antoniou and Karl E. Gustafson. From probabilistic descriptions to deterministic dynamics. *Physica A*, 197:153–166, 1993.

[6] Hassan Arbabi and Igor Mezić. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, 2017.

[7] Vladimir Igorevich Arnol'd and André Avez. Ergodic problems of classical mechanics. 1968.

[8] Roberto Artuso, Erik Aurell, and Predrag Cvitanovic. Recycling of strange sets: I. cycle expansions. *Nonlinearity*, 3:325–359, 1990.

[9] Travis Askham and J. Nathan Kutz. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 17(1):380–416, 2018.

[10] Ditza Auerbach, Predrag Cvitanovic, Jean-Pierre Eckmann, Gemunu Gunaratne, and Itamar Procaccia. Exploring chaotic motion through periodic orbits. *Physical Review Letters*, 58(23):2387–2389, 1987.

[11] Omri Azencot, Wotao Yin, and Andrea Bertozzi. Consistent dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 18(3):1565–1585, 2019.

[12] Karol A. Bacik, Michael T. Schaub, Mariano Beguerisse-Díaz, Yazan N. Billeh, and Mauricio Barahona. Flow-based network analysis of the caenorhabditis elegans connectome. *PLOS Computational Biology*, 12(8):1–27, 08 2016.

[13] R. Baldick, B. Chowdhury, I. Dobson, D. Hawkins, D. Kirschen, L. Mili, S. Miller, M. Nakayama, M. Papic, R. Podmore, J. Rossmaier, K. Schneider, and D. Wang. Vulnerability assessment for cascading failures in electric power systems. In *2009 IEEE/PES Power Systems Conference and Exposition*, pages 1–9, March 2009.

[14] Michael J. Barber. Modularity and community detection in bipartite networks. *Phys. Rev. E*, 76:066102, Dec 2007.

[15] E Barocio, Bikash C Pal, Davide Fabozzi, and Nina F Thornhill. Detection and visualization of power system disturbances using principal component analysis. In *Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*, pages 1–10. IEEE, 2013.

[16] Roberto Barrio, Andrey Shilnikov, and Leonid Shilnikov. Kneadings, symbolic dynamics and painting lorenz chaos. a tutorial. *arXiv preprint arXiv:1204.3278*, 2012.

[17] Carl M Bender and Steven A Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

[18] Pratyasa Bhui and Nilanjan Senroy. Application of recurrence quantification analysis to power system dynamic studies. *IEEE Transactions on Power Systems*, 31(1):581–591, 2016.

[19] Bruce M Boghosian, Aaron Brown, Jonas Lätt, Hui Tang, Luis M Fazendeiro, and Peter V Coveney. Unstable periodic orbits in the lorenz attractor. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 369(1944):2345–2353, 2011.

[20] Ettore Bompard, Roberto Napoli, and Fei Xue. Analysis of structural vulnerabilities in power transmission grids. *International Journal of Critical Infrastructure Protection*, 2(1-2):5–12, 2009.

[21] Ettore Bompard, Enrico Pons, and Di Wu. Extended topological metrics for the analysis of power grid vulnerability. *IEEE Systems Journal*, 6(3):481–487, 2012.

[22] Ettore Bompard, Di Wu, and Fei Xue. The concept of betweenness in the analysis of power grid vulnerability. *COMPENG 2010 - Complexity in Engineering*, pages 52–54, 2010.

[23] Ettore F. Bompard, Antonio Scala, Sandro Bologna, Gregorio D', Martí Rosas Casals, Giuliano Andrea Pagani, Trivik Verma, Wendy Ellens, and N.A. Agostino. Knowing power grids and understanding complexity science. *International Journal of Critical Infrastructures*, 11(1):4, 2015.

[24] D. S. Broomhead, Roger Jones, Martin Fleischmann, D. J. Tildesley, and R. C. Ball. Time-series analysis. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 423(1864):103–121, 1989.

[25] D.S. Broomhead and Gregory P. King. Extracting qualitative dynamics from experimental data. *Physica D*, 20:217–236, 1986.

[26] Bingni W. Brunton, Lise A. Johnson, Jeffrey G. Ojemann, and J. Nathan Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016.

[27] S. L. Brunton, B. W. Brunton, J. L. Proctor, and J. N Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE*, 11(2):e0150171, 2016.

[28] S. L. Brunton and J. N. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[29] S. L. Brunton and B. R. Noack. Closed-loop turbulence control: Progress and challenges. *Applied Mechanics Reviews*, 67:050801–1–050801–48, 2015.

[30] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[31] Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, Eurika Kaiser, and J. Nathan Kutz. Chaos as an intermittently forced linear system. *Nature Communications*, 8, 2017.

[32] Nazmi Burak Budanur, Daniel Borrero-Echeverry, and Predrag Cvitanovic. Periodic orbit analysis of a system with continuous symmetry—a tutorial. *Chaos*, 25(073112):1–17, 2015.

[33] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied Koopmanism a). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.

[34] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.

[35] Kathleen Champion, Steven L Brunton, and J Nathan Kutz. Discovery of nonlinear multiscale systems: Sampling strategies and embeddings. *arXiv preprint arXiv:1805.07411*, 2018.

[36] Joe H Chow and Kwok W Cheung. A toolbox for power system dynamics and control engineering education and research. *IEEE transactions on Power Systems*, 7(4):1559–1564, 1992.

[37] F. Christiansen, Predrag Cvitanovic, and V. Putkaradze. Hopf's last hope: spatiotemporal chaos in terms of unstable recurrent patterns. *arXiv preprint arXiv:9606016*, 2008.

[38] Freddy Christiansen, Giovanni Paladin, and Hans Henrik Rugh. Determination of correlation spectra in chaotic systems. *Physical review letters*, 65(17):2087, 1990.

[39] IP Cornfeld and SV Fomin. Sinaı ya. g. ergodic theory. translated from the russian, 1982.

[40] Daniel L. Crane, Ruslan L. Davidchack, and Alexander N. Gorban. Minimal cover of high-dimensional chaotic attractors by embedded coherent structures. *arXiv preprint arXiv:1607.02180*, 2016.

[41] Lucas Cuadra, Sancho Salcedo-Sanz, Javier Del Ser, Silvia Jiménez-Fernández, and Zong Woo Geem. A critical review of robustness in power grids using complex networks concepts. *Energies*, 8(9):9211–9265, 2015.

[42] Predrag Cvitanovic. Periodic orbits as the skeleton of classical and quantum chaos. *Physica D*, 51:138–151, 1991.

[43] Predrag Cvitanovic, Roberto Artuso, Ronnie Mainieri, Gregor Tanner, and Gabor Vattay. *Chaos: Classical and Quantum, Volume I: Deterministic Chaos*. Gone with the wind press, 2011.

[44] Predrag Cvitanovic and Bruno Eckhardt. Periodic-orbit quantization of chaotic systems. *Physical Review Letters*, 63(8):823–826, 1989.

[45] Predrag Cvitanovic and Bruno Eckhardt. Periodic orbit expansions for classical smooth flows. *J. Phys. A: Math. Gen*, 24:237–241, 1991.

[46] Predrag Cvitanovic and Bruno Eckhardt. Symmetry decomposition of chaotic dynamics. *arXiv preprint arXiv:9303016*, 1993.

[47] Predrag Cvitanovic, Mogens H. Jensen, Leo P. Kadanoff, and Itamar Procaccia. Renormalixation, unstable manifolds, and the fractal structure of mode locking. *Physical Review Letters*, 55(4):343–346, 1985.

[48] Predrag Cvitanovic and Domenico Lippolis. Knowing when to stop: How noise frees us from determinism. *arXiv preprint arXiv:1206.550*, 2012.

[49] Suddhasattwa Das and Dimitrios Giannakis. Delay-coordinate maps and the spectra of koopman operators. *Journal of Statistical Physics*, 175:1107–1145, 2019.

[50] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[51] Scott TM Dawson, Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):1–19, 2016.

[52] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, 107(29):12755–12760, 2010.

[53] David Donoho. 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4):745–766, 2017.

[54] M Dörfle. Spectrum and eigenfunctions of the frobenius-perron operator of the tent map. *Journal of statistical physics*, 40(1):93–132, 1985.

[55] Alexander Düring and Ulrich Behn. Selective stabilisation of very long unstable periodic orbits in chaotic systems. *Physics Letters A*, 230:177–182, 1997.

[56] Daniel Dylewsky, Molei Tao, and J. Nathan Kutz. Dynamic mode decomposition for multiscale nonlinear physics. *Phys. Rev. E*, 99:063311, Jun 2019.

[57] Daniel Dylewsky, Xiu Yang, Alexandre Tartakovsky, and J. Nathan Kutz. Engineering structural robustness in power grid networks susceptible to community desynchronization. *Applied Network Science*, 4(24), 2019.

[58] Weinan E. *Principles of multiscale modeling*. Cambridge University Press, 2011.

[59] Weinan E, Bjorn Engquist, and others. The heterogeneous multiscale methods. *Communications in Mathematical Sciences*, 1(1):87–132, 2003.

[60] Bruno Eckhardt. Periodic orbit theory. Lecture Notes, arXiv preprint: 9303015, December 18 2013.

[61] Jean-Pierre Eckmann. Roads to turbulence in dissipative dynamical systems. *Reviews of Modern Physics*, 53(4):643–654, 1981.

[62] Robert Eriksson and Lennart Soder. Wide-area measurement system-based subspace identification for obtaining linear models to centrally coordinate controllable devices. *IEEE Transactions on Power Delivery*, 26(2):988–997, 2011.

[63] Geir Evensen. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, Nov 2003.

[64] A. C. Fowler and M. J. McGuinness. A description of the lorenz attractor at high prandtl number. *Physica D*, pages 149–182, 1982.

[65] K. Fraedrich. Estimating dimensions of weather and climate attractors. *J. Atmos. Sci.*, 43:419–432, 1986.

[66] Gary Froyland, Georg A Gottwald, and Andy Hammerlindl. A computational method to extract macroscopic variables and their dynamics in multiscale systems. *SIAM Journal on Applied Dynamical Systems*, 13(4):1816–1846, 2014.

[67] Gary Froyland, Georg A Gottwald, and Andy Hammerlindl. A trajectory-free framework for analysing multiscale systems. *Physica D: Nonlinear Phenomena*, 328:34–43, 2016.

[68] Ioannis G. Kevrekidis, C Gear, and Gerhard Hummer. Equation-free: The computer-aided analysis of complex multiscale systems. *Aiche Journal - AICHE J*, 50, 07 2004.

[69] Rajesh Ganesan, Tapas K. Das, and Vivekanand Venkataraman. Wavelet-based multiscale statistical process monitoring: A literature review. *IIE Transactions*, 36(9):787–806, 2004.

[70] Pierre Gaspard. *Chaos, scattering and statistical mechanics*, volume 9. Cambridge University Press, 2005.

[71] Matan Gavish and David L Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.

[72] Alexey I. Rakhmanov George G. Malinetskii, Alexey B. Potapov. Limitations of delay reconstruction for chaotic dynamical systems. *Physical Review E*, 48(2):904–912, 1993.

[73] Dimitrios Giannakis. Dynamics-adapted cone kernels. *SIAM Journal on Applied Dynamical Systems*, 14(2):556–608, 2015.

[74] Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 47(2):338 – 396, 2019.

[75] Dimitrios Giannakis and Andrew J. Majda. Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, 109(7):2222–2227, 2012.

[76] John F. Gibson, J. Doyne Farmer, Martin Casdagli, and Stephen Eubank. An analytic approach to practical state space reconstruction. *Physica D*, 57:1–30, 1992.

[77] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*, 99(12):7821–7826, Jun 2002.

[78] Georg A Gottwald and Ian Melbourne. On the implementation of the 0–1 test for chaos. *SIAM Journal on Applied Dynamical Systems*, 8(1):129–145, 2009.

[79] Martin C. Gutzwiller. The quantization of a classically ergodic system. *Physica D*, pages 183–207, 1982.

[80] Hiroshi H. Hasegawa and William C. Saphir. Decaying eigenstates for simple chaotic systems. *Physics Letters A*, 161:471–476, 1992.

[81] Hossein Hassani. Singular spectrum analysis: Methodology and comparison. *University Library of Munich, Germany, MPRA Paper*, 5, 01 2007.

[82] Maziar S Hemati, Clarence W Rowley, Eric A. Deem, and Louis N. Cattafesta. Debiasing the dynamic mode decomposition for applied koopman spectral analysis. *Theoretical and Computational Fluid Dynamics*, 31(4):349–368, 2017.

[83] M. Henon. On the numerical computation of poincare maps. *Physica D*, pages 412–414, 1982.

[84] Gustav Eje Henter and W. Bastiaan Kleijn. Picking up the pieces: Causal states in noisy data, and how to recover them. *Pattern Recognition Letters*, 34:587–594, 2013.

[85] P. Hines, S. Blumsack, E. Cotilla Sanchez, and C. Barrows. The topological and electrical structure of power grids. *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 1–10, 2010.

[86] Leonhard Horstmeyer and Fatihcan M. Atayb. Characterization of exact lumpability for vector fields on smooth manifolds. *Differential Geometry and its Applications*, 48:46–60, 2016.

[87] E. Yarevsky I. Antoniou, S. A. Shkarin. Resonances of the cusp family. *J. Phys. A: Math. Gen*, 35:2833–2844, 2002.

[88] Z. Suchanecki I. Antoniou. Harmonic analysis of unstable systems. In Stuart A. Rice Ioannis Antoniou, I. Prigogine, editor, *Dynamical Systems and Irreversibility: A Special Volume of Advances in Chemical Physics*, volume 122. John Wiley & Sons, Inc, 2002.

[89] Tue V. Jensen, Hugo de Sevin, Martin Greiner, and Pierre Pinson. The re-europe data set, December 2015.

[90] Mihailo R Jovanović, Peter J Schmid, and Joseph W Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[91] Jer-Nan Juang and Richard S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985.

[92] E. Kaiser, J. N. Kutz, and S. L. Brunton. Data-driven discovery of Koopman eigenfunctions for control. *arXiv preprint arXiv:1707.01146*, 2017.

[93] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society of London A*, 474(2219), 2018.

[94] Mason Kamb, Eurika Kaiser, Steven L. Brunton, and J. Nathan Kutz. Time-delay observables for koopman: Theory and applications. *SIAM Journal on Applied Dynamical Systems*, 19(2):886–917, 2020.

[95] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.

[96] Jirayr Kevorkian and Julian D Cole. *Perturbation methods in applied mathematics*, volume 34. Springer Science & Business Media, 2013.

[97] Ioannis G. Kevrekidis, C. William Gear, James M. Hyman, Panagiotis G Kevrekidis, Olof Runborg, and Constantinos Theodoropoulos. Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 12 2003.

[98] Mladen Kezunovic, Le Xie, and Santiago Grijalva. The role of big data in improving power system operation and protection. In *Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*, pages 1–9. IEEE, 2013.

[99] Haris M Khalid and Jimmy C-H Peng. Improved recursive electromechanical oscillations monitoring scheme: A novel distributed approach. *IEEE Transactions on Power Systems*, 30(2):680–688, 2015.

[100] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 2018.

[101] B.O. Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315–318, 1931.

[102] B.O. Koopman and J. von Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences of the United States of America*, 18(3):255–263, 1932.

[103] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *Journal of Nonlinear Science*, 28:687–710, 2017.

[104] Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

[105] Boris Kramer, Piyush Grover, Petros Boufounos, Mouhacine Benosman, and Saleh Nabi. Sparse sensing and dmd based identification of flow regimes and bifurcations in complex flows. *arXiv preprint arXiv:1510.02831*, 2015.

[106] J. Kutz, S. Brunton, B. Brunton, and J. Proctor. *Dynamic Mode Decomposition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[107] J. Kutz, X. Fu, and S. Brunton. Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.

[108] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.

[109] J. N. Kutz, X. Fu, and S. L. Brunton. Multi-resolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016. Preprint. Available: arXiv:1506.00564.

[110] J. N. Kutz, J. L. Proctor, and S. L. Brunton. Koopman theory for partial differential equations. *Invited for Journal of Nonlinear Science*, 2016.

[111] J Nathan Kutz. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, 2013.

[112] J Nathan Kutz, Xing Fu, and Steven L Brunton. Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.

[113] J. Nathan Kutz, Joshua L. Proctor, and Steven L. Brunton. Koopman theory for partial differential equations. *Complexity (to appear)*.

[114] Ying-Cheng Lai and Nong Ye. Recent developments in chaotic time series analysis. *International Journal of Bifurcation and Chaos*, 13(6):1383–1422, 2003.

[115] Yueheng Lan and Igor Mezić. Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D: Nonlinear Phenomena*, 242(1):42–53, 2013.

[116] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

[117] J Laskar. Secular evolution of the solar system over 10 million years. *Astronomy and Astrophysics*, 198:341–362, 05 1988.

[118] Jacques Laskar. Frequency map analysis and quasiperiodic decompositions. 13 March 2003, May 2003.

[119] Andrzej Lasota and Michael C. Mackey. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, volume 97 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2 edition, 1994.

[120] Soledad Le Clainche and José M. Vega. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, 2017.

[121] Jure Leskovec, Kevin J Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pages 631–640. ACM, 2010.

[122] Qianxiao Li, Felix Dietrich, Erik M. Bollt, and Ioannis G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: a data-driven adaptive spectral decomposition of the Koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27, 2017.

[123] Wei Li, Ju Tang, Jing Ma, and Yilu Liu. Online detection of start time and location for hypocenter in north america power grid. *IEEE Transactions on Smart Grid*, 1(3):253–260, 2010.

[124] W. Liebert and H.G. Schuster. Proper choice of the time delay for the analysis of chaotic time series. *Physics Letters A.*, 142:107–111, 1988.

[125] Hesen Liu, Lin Zhu, Zhuohong Pan, Feifei Bai, Yong Liu, Yilu Liu, Mahendra Patel, Evangelos Farantatos, and Navin Bhatt. Armax-based transfer function model identification using wide-area measurement for adaptive and coordinated damping control. 2015.

[126] X. Liu and T. Murata. Community detection in large-scale bipartite networks. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 50–57, Sept 2009.

[127] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9:4950, 2018.

[128] Huanfei Ma, Wei Lin, and Ying-Cheng Lai. Detecting unstable periodic orbits in high-dimensional chaotic systems from time series: Reconstruction meeting with adaptation. *Physical Review E*, 87(5), May 2013.

[129] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11:674–693, 07 1989.

[130] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets: Deep learning of molecular kinetics. *Nature Communications*, 9(5), 2018.

[131] Kejun Mei, Steven M Rovnyak, and Chee-Mun Ong. Clustering-based dynamic event location using wide-area phasor measurements. *IEEE Transactions on Power Systems*, 23(2):673–679, 2008.

[132] Arturo Roman Messina and Arturo Rom Messina. *Wide-area monitoring of interconnected power systems*. The Institution of Engineering and Technology, 2015.

[133] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[134] Igor Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45:357–378, 2013.

[135] Igor Mezić. Koopman operator spectrum and data analysis. *arXiv preprint arXiv:1702.07597*, 2017.

[136] Igor Mezić. *Spectral operator methods in dynamical systems: Theory and applications*. Springer, 2018.

[137] Igor Mezić and Andrzej Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1):101 – 133, 2004.

[138] Kevin P Murphy. Machine learning: A probabilistic perspective. adaptive computation and machine learning, 2012.

[139] Seyedbehzad Nabavi, Jianhua Zhang, and Aranya Chakrabortty. Distributed optimization algorithms for wide-area oscillation monitoring in power systems using interregional pmu-pdc architectures. *IEEE Transactions on Smart Grid*, 6(5):2529–2538, 2015.

[140] A. B.M. Nasiruzzaman, H. R. Pota, and M. A. Mahmud. Application of centrality measures of complex network framework in power grid. *IECON Proceedings (Industrial Electronics Conference)*, pages 4660–4665, 2011.

[141] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[142] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.

[143] Jiawei Ning, Xueping Pan, and Vaithianathan Venkatasubramanian. Oscillation modal analysis from ambient synchrophasor data using distributed frequency domain optimization. *IEEE Transactions on Power Systems*, 28(2):1960–1968, 2013.

[144] Louis van Blargian Timothy Matchen Emma Tegling Nithin Govindarajan, Hassan Arbabi and Igor Mezić. An operator-theoretic viewpoint to non-smooth dynamical systems: Koopman analysis of a hybrid pendulum. *arXiv preprint arXiv:1608.08734*, 2016.

[145] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003.

[146] Frank Noé and Feliks Nuske. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Modeling & Simulation*, 11(2):635–655, 2013.

[147] Feliks Nüske, Bettina G Keller, Guillermo Pérez-Hernández, Antonia SJS Mey, and Frank Noé. Variational approach to molecular kinetics. *Journal of chemical theory and computation*, 10(4):1739–1752, 2014.

[148] Edward Ott. Strange attractors and chaotic motions of dynamical systems. *Reviews of Modern Physics*, 53(4):655–671, 1981.

[149] Samuel E Otto and Clarence W Rowley. Linearly-recurrent autoencoder networks for learning dynamics. *arXiv preprint arXiv:1712.01378*, 2017.

[150] Giuliano Andrea Pagani and Marco Aiello. The Power Grid as a complex network: A survey. *Physica A: Statistical Mechanics and its Applications*, 392(11):2688–2700, 2013.

[151] Giuliano Andrea Pagani and Marco Aiello. Power grid complex network evolutions for the smart grid. *Physica A: Statistical Mechanics and its Applications*, 396:248–266, 2014.

[152] S. Pahwa, M. Youssef, P. Schumm, C. Scoglio, and N. Schulz. Optimal intentional islanding to enhance the robustness of power grid networks. *Physica A: Statistical Mechanics and its Applications*, 392(17):3741–3754, 2013.

[153] Shaowu Pan and Karthik Duraisamy. On the structure of time-delay embedding in linear models of non-linear dynamical systems. 02 2019.

[154] Sebastian Peitz and Stefan Klus. Koopman operator-based model reduction for switched-system control of pdes. *arXiv preprint arXiv:1710.06759*, 2017.

[155] Mark Pollicott. Meromorphic extensions of generalised zeta functions. *Inventiones mathematicae*, 85(1):147–164, 1986.

[156] Ales Prochazka, Nicholas Kingsbury, PJW Payner, and J Uhlir. *Signal analysis and prediction*. Springer Science & Business Media, 2013.

[157] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[158] Joshua L Proctor, Steven L Brunton, and J Nathan Kutz. Generalizing koopman theory to allow for inputs and control. *SIAM Journal on Applied Dynamical Systems*, 17(1):909–930, 2018.

[159] Fredrik Raak, Yoshihiko Susuki, Takashi Hikihara, Harold R Chamorro, and Mehrdad Ghandhari. Partitioning power grids via nonlinear koopman mode analysis. In *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, pages 1–5. IEEE, 2014.

[160] Martin Rosvall, Alcides V. Esquivel, Andrea Lancichinetti, Jevin D. West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5:4630 EP –, Aug 2014. Article.

[161] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009.

[162] David Ruelle. Zeta-functions for expanding maps and anosov flows. *Inventiones mathematicae*, 34(3):231–242, 1976.

[163] David Ruelle. Locating resonances for axioma dynamical systems. *Journal of statistical physics*, 44(3):281–292, 1986.

[164] David Ruelle. One dimensional gibbs states and axiom a diffeomorphisms. *J. Diff. Geom.*, 25:117–137, 1987.

[165] David Ruelle. Dynamical zeta functions and transfer operators. *Notices of the AMS*, pages 887–895, 2002.

[166] Yoshitaka Saiki and Michio Yamada. Unstable periodic orbits embedded in a continuous time dynamical system: time averaged properties (nonlinear dynamics in macroeconomics). *???*, 1713:111–123, 2010.

[167] Michael T. Schaub, Jean-Charles Delvenne, Martin Rosvall, and Renaud Lambiotte. The many facets of community detection in complex networks. 4(2), 2017.

[168] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust. Applications of the dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, 25(1):249–259, Jun 2011.

[169] P. J. Schmid and J. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In *61st Annual Meeting of the APS Division of Fluid Dynamics*. American Physical Society, November 2008.

[170] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[171] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[172] M. E. W. Scholz. Locating unstable periodic orbits in recurrence plots. DCT rapporten 2003.122, Technische Universiteit Eindhoven, Eindhoven, 2003.

[173] Mohammad Shahraeini, Mohammad Hossein Javidi, and Mohammad Sadegh Ghazizadeh. Comparison between communication infrastructures of centralized and decentralized wide area measurement systems. *IEEE Transactions on Smart Grid*, 2(1):206–211, 2011.

[174] Ricard V. Solé, Martí Rosas-Casals, Bernat Corominas-Murtra, and Sergi Valverde. Robustness of the European power grids under intentional attack. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 77(2):1–7, 2008.

[175] Yoshihiko Susuki and Igor Mezić. Nonlinear koopman modes and coherency identification of coupled swing dynamics. *IEEE Transactions on Power Systems*, 26(4):1894–1904, 2011.

[176] Yoshihiko Susuki and Igor Mezić. Nonlinear koopman modes and a precursor to power system swing instabilities. *IEEE Transactions on Power Systems*, 27(3):1182–1191, 2012.

[177] Yoshihiko Susuki and Igor Mezić. Nonlinear koopman modes and power system stability assessment without models. *IEEE Transactions on Power Systems*, 29(2):899–907, 2014.

[178] Yoshihiko Susuki and Igor Mezić. A prony approximation of Koopman mode decomposition. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pages 7022–7027. IEEE, 2015.

[179] Yoshihiko Susuki, Igor Mezić, and Takashi Hikihara. Global swing instability in the new england power grid model. In *2009 American Control Conference*, pages 3446–3451. IEEE, 2009.

[180] Yoshihiko Susuki, Igor Mezić, and Takashi Hikihara. Coherent swing instability of power grids. *Journal of nonlinear science*, 21(3):403–439, 2011.

[181] Kunihiko Taira, Steven L Brunton, Scott Dawson, Clarence W Rowley, Tim Colonius, Beverley J McKeon, Oliver T Schmidt, Stanislav Gordeyev, Vassilios Theofilis, and Lawrence S Ukeiley. Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041, 2017.

[182] Naoya Takeishi, Yoshinobu Kawahara, Yasuo Tabei, and Takehisa Yairi. Bayesian dynamic mode decomposition. *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.

[183] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning Koopman invariant subspaces for dynamic mode decomposition. In *Advances in Neural Information Processing Systems*, pages 1130–1140, 2017.

[184] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Subspace dynamic mode decomposition for stochastic Koopman analysis. *Physical Review E*, 96:033310, 2017.

[185] Floris Takens. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, 1981.

[186] M. Tao, H. Owhadi, and J. Marsden. Nonintrusive and structure preserving multiscale integration of stiff odes, sdes, and hamiltonian systems with hidden slow dynamics via flow averaging. *Multiscale Modeling & Simulation*, 8(4):1269–1324, 2010.

[187] A. Townsend and L.N. Trefethen. Continuous analogues of matrix factorizations. *Proceedings of the Royal Society A*, 471, 2015.

[188] Lloyd N Trefethen. Pseudospectra of matrices. *Numerical analysis*, 91:234–266, 1991.

[189] Jonathan H. Tu, Clarence W. Rowle, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2158-2491 2014 2 391):391, 2014.

[190] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2158-2491):391, 2014.

[191] Warwick Tucker. The lorenz attractor exists. *C. R. Acad. Sci. Paris*, 328(1):1197–1202, 1999.

[192] John W. Tukey. The future of data analysis. *Ann. Math. Statist.*, 33(1):1–67, 03 1962.

[193] R. Vautard and M. Ghil. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, 35:395–424, 1989.

[194] Roberto Venegeroles. Leading pollicott-ruelle resonances for chaotic area-preserving maps. *Physical Review E*, 77(2):027201, 2008.

[195] Divakar Viswanath. Symbolic dynamics and periodic orbits of the lorenz attractor. *Nonlinearity*, 16:1035–1056, 2003.

[196] James R. Voelkel. *The composition of Kepler's Astronomia nova*. Princeton University Press, 2001.

[197] Yannan Wang, Pradeep Yemula, and Anjan Bose. Decentralized communication and control systems for power system operation. *IEEE Transactions on Smart Grid*, 6(2):885–893, 2015.

[198] Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(241703):1–9, 2018.

[199] Michael Weniger, Florian Kapp, and Petra Friederichs. Spatial verification using wavelet transforms: a review. *Quarterly Journal of the Royal Meteorological Society*, 143(702):120–136, 2017.

[200] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25:1307–1346, 2015.

[201] Yan Xu, Aleks Jacob Gurfinkel, and Per Arne Rikvold. Architecture of the Florida power grid as a complex network. *Physica A: Statistical Mechanics and its Applications*, 401:130–140, 2014.

[202] Jun Yan, Yihai Zhu, Haibo He, and Yan Sun. Multi-contingency cascading analysis of smart grid based on self-organizing map. *IEEE Transactions on Information Forensics and Security*, 8(4):646–656, 2013.

[203] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. *arXiv preprint arXiv:1708.06850*, 2017.

[204] Yongxiang Zhang and Guanwei Luo. Detecting unstable periodic orbits and unstable quasiperiodic orbits in vibro-impact systems. *International Journal of Non-Linear Mechanics*, 96:12–21, 2017.

[205] Jian Zhou, Ning Huang, Xiaolei Sun, Kunlong Wang, and Hongqi Yang. A new model of network cascading failures with dependent nodes. *Proceedings - Annual Reliability and Maintainability Symposium*, 2015-May:1–6, 2015.