Machine learning for dynamical models of human movement

Megan R. Ebers

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

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Katherine M. Steele, Chair

J. Nathan Kutz, Chair

Steven L. Brunton

Program Authorized to Offer Degree:

Department of Mechanical Engineering

University of Washington

Abstract

Machine learning for dynamical models of human movement

Megan R. Ebers

Co-Chairs of the Supervisory Committee:
Albert S. Kobayashi Endowed Professor Katherine M. Steele
Department of Mechanical Engineering

Robert Bolles and Yasuko Endo Professor J. Nathan Kutz
Departments of Applied Mathematics and Electrical and Computer Engineering

Data-driven dynamical modeling is an emerging and powerful tool for analyzing, predicting, and controlling complex systems in engineering and physical sciences. Traditionally, modeling of dynamical systems uses mathematical approaches like differential equations; modern approaches leverage advances in machine learning to discovery models directly from system measurements. In engineering and physical sciences, first-principles and physics-based models are ubiquitous, but only allow for the modeling of dynamics with a limited accuracy; purely data-driven models of dynamical systems are able to learn complex relationships, but can become unconstrained without leveraging known physics. As data-driven dynamical modeling continues to gain momentum, it is imperative that researchers utilize a hybrid modeling approach to combine domain knowledge and measurements to model complex systems.

Discrepancy modeling for dynamical systems is a hybrid modeling approach that aims to resolve the mismatch between model estimations and measurement data due to missing physics. In this dissertation, we focus on shifting the view of discrepancies as 'errors' or 'residuals' to highly valuable measures for model improvement and scientific insight. We discuss two nuanced, yet distinct, discrepancy modeling approaches for estimating

missing physics and demonstrate how different model discovery methods can be used interchangeably within this framework. Further, we emphasize discrepancy modeling considerations and trade-offs related to model interpretability and sensor constraints. The field of data-driven engineering for dynamical systems has an opportunity to improve system characterization and provide scientific insight by disambiguating deterministic and random effects within the model-measurement mismatch.

One such complex system is the human musculoskeletal system, and modeling the musculoskeletal system is of great interest to the biomechanics community. Great strides in clinical insight have emerged from modeling and simulation using physics-based and physiologically-detailed models. However, the complexity of the human body is challenging to represent as a musculoskeletal simulation, especially for clinical populations, limiting utility of such techniques. Machine learning has been employed to study human movement, but historically focuses on feature extraction and classification. In viewing the human body as a dynamical system, data-driven dynamical modeling techniques can be employed to tackle heterogeneous, nonlinear, and complex challenges in treating pathology, enhancing mobility, and personalizing rehabilitation.

In demonstrating data-driven dynamical modeling for human movement, we focus on two major challenges: (1) prescribing assistive devices and (2) collecting rich datasets of movement for health monitoring. In this dissertation, we apply discrepancy modeling to characterize individual responses to passive-elastic ankle exoskeletons during walking. A neural network-based discrepancy model successfully quantified complex changes in gait kinematics and electromyography during exoskeleton walking; yet, kinematics and electromyography alone were insufficient to fully capture muscle-level changes in responses to ankle exoskeletons. Understanding how an individual will respond to an assistive device and what data are needed to encode responses can help optimize the prescription process and uncover underlying mechanisms driving gait changes. The final study in this dissertation combined deep learning, time-delay embedding, and sparse sensing for full-state reconstruction of complex systems, including for personalized human movement tracking. For example, we can expand biomechanical datasets by mapping from as few as a single sensor to a full dataset of biomechanical states. This work will enable robust motion tracking in-the-wild for monitoring and supporting movement-based health outcomes; more generally, this mathematical architecture will enable state estimation of dynamical systems with mobile sensors without the challenge of sensor path planning.

Using data to discover dynamical models is transforming how we study complex systems. The work presented in this dissertation demonstrates how we can build models and extract insight from time-series data. While methodologically new to the field of biomechanics, modern advances in data-driven dynamical methods are a powerful tool for quantifying and understanding musculoskeletal biomechanics – and beyond. This dissertation provides the foundation for developing and deploying machine learning for dynamical models of human movement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

No amount of words on a page will ever encompass the gratitude I feel about the support I received during these past five years. Like most PhD students I'm sure, my journey was wrought with combinations of excitement, self-doubt, curiosity, frustration, exhaustion, and pride. Subsequently, it is truly through the efforts – both passive and active – of my mentors, friends, and family that I crossed the finish line. Therefore, it is only right that I acknowledge that my victory is their victory, too.

First and foremost, I must thank my advisors, Kat Steele and Nathan Kutz. It is with the utmost confidence I can say that I would not be the researcher I am today without their guidance, patience, and encouragement. Kat: thank you for setting a stellar example of what it is to be a badass in academia. Thank you for your willingness to confide that feelings of inadequacy never truly disappear, your dedication to fostering my ability to execute independent research (while also insisting it's okay to ask for help sometimes), and your trust in my wild ideas to bring together applied math and mechanical principles of musculoskeletal movement. Nathan: thank you for taking a chance on a naive PhD student whose mind was blown by linear algebra. Thank you for the countless cappuccinos, for generously sharing your time talking through paper outlines and research ideas and the spectrum of possible careers, and for always having my back when forward progress felt impossible. I would also like to recognize my collaborators, mentors, and committee members – Drs. Steven Brunton, Eric Rombokas, Brian Hafner, Michael Rosenberg, and Jan Williams – for providing invaluable discussion and helping shape my dissertation.

I want to recognize the Steele Lab – past and present – for being both a welcome distraction and an appreciated sounding board these past five years. To past members

– Ben Shuman, Michael Rosenberg, Momona Yamagami, Keshia Peters, Nick Biacoianu, Christina Papazian, Kim Ingraham, and Alyssa Spomer – thank you for accepting me with open arms, for shaping a positive lab culture, and for being people I could look up to. To current members – Nicole Zaino, Charlotte Caskey, Elijah Kuska, Mia Hoffman, Mackenzie Pitts, Yusuke Maruo, and Sasha Portnova – thank you for putting up with my being hell-bent on bringing applied math into biomechanics. I could not imagine a better group to have shared my PhD journey with.

I would like to recognize my funding sources, which enabled me to pursue uncommon and interdisciplinary research. My dissertation was supported by the University of Washington (UW) College of Engineering Fellowship, a National Science Foundation (NSF) Graduate Research Fellowship (grant no. DGE-1762114), and the NSF AI Institute for Dynamical Systems (grant no. 2112085).

I am grateful for all the random sources from which I was shaped and supported during my dissertation. Thanks to Louis Joslyn for providing management and mentorship during my PhD industry internship. Thanks to a few Colorado School of Mines folks that first fostered my love of mechanical engineering, biomechanics, and research: Drs. Ozkan Celik, Anne Silverman, Joel Bach, and Jenifer Blacklock. Thanks to the music that got me through tough and tired times: Mantaraybryn, Rainbow Kitten Surprise, and LoFi beats. Also, bless coffee.

Finally, I need to thank my support system. Above all, I want to acknowledge Joel, my husband and best friend: thank you endlessly for being the best cheerleader anyone could ask for. To my parents, thank you for encouraging me from a young age to problem solve, try hard things, ask good questions, and accomplish what people did not expect of me. To my sister, thank you for patiently listening to my woes and constantly shouting my praises. I am so lucky. I love you all fiercely.

Chapter 1
INTRODUCTION

Biomechanists collect a deluge of data to quantify human movement: from highly-controlled and precisely-measured laboratory datasets to portable devices like wearable sensors that estimate biomechanical variables in daily life. Data science and machine learning have been powerful tools in the quest to interpret and extract meaningful insights into human movement from this data. This dissertation seeks to develop and deploy machine learning techniques for time-series data with a focus on human movement. While a cornerstone of medical practice and discovery has utilized time series for comparing historic trends (for an individual or population), this time series can be used to develop data-driven dynamical models of human movement.

Dynamics provides a mathematical framework to describe the world around us and model the rich interactions between quantities that co-evolve in time. Traditional modeling of dynamical systems utilizes differential equations to describe the evolution of the state of a system. However, the complexity present in real-world dynamical systems (such as blood flow around the heart, mechanisms of flight, and dynamics of the brain) allows for the modeling of a system with limited accuracy. Improvements in sensing and computational technologies have enabled a hybrid approach for modeling complex systems by using first-principles physics and measurement data together. Thus, model and measurements work as complements to resolve the inevitable mismatch, or discrepancy, and improve complex system characterization.

While data-driven dynamical modeling is an emerging and powerful tool for analyzing, predicting, and controlling complex systems in engineering and physical sciences, such methods are in their infancy for understanding human movement. Data-driven modeling methods for biomechanical, neurological, and physiological applications (and beyond) are enticing, as they (i) make no explicit assumptions about an individual's physiology or motor control, (ii) they are appropriate for individual or population-based modeling, and (iii) have a rich pool of literature developed for other canonical dynamical systems. By adopting a data-driven dynamical modeling approach, not only challenging but untouched problems can be investigated, such as quantifying individual responses to an intervention or expanding a biomechanical dataset to unmeasured quantities.

1.1 Focus of the dissertation

This dissertation explores data-driven techniques to develop models of and extract insight from times series data of human movement. To overcome the limitations of first-principles modeling of complex dynamical systems, we investigated how data complements and improves mechanistic modeling of dynamical systems in physical and engineering applications (Chapter 3). We found that discrepancy modeling is a useful framework for resolving model-measurement mismatch due to missing physics in practical engineering applications. To apply these methods to the domain of human movement, we characterized heterogeneous responses to walking with an ankle exoskeleton. We leveraged a neural network-based discrepancy modeling framework to quantify complex changes in gait with ankle exoskeletons for nondisabled adults (Chapter 4). Discrepancy modeling successfully captured changes in the processes governing gait kinematics and electromyography recordings with ankle exoskeletons (*i.e.*, discrepancies) without requiring prior knowledge or assumptions about the physiological structure of the discrepancy. Finally, we combined deep learning, time-delay embedding, and sparse sensing to reconstruct full state-space estimates from sparse dynamical trajectories using a shallow recurrent decoder network. Such an architecture can enable the expansion of biomechanical datasets while using as few as one sensor, such as for personalized tracking of human motion in natural environments (Chapter 5). By leveraging the time histories of sensor measurements, we successfully reconstructed the full set of states from a sparse set of sensors, such as with joint kinematics or inertial measurement unit signals, during walking for nondisabled adults. We also demonstrated the success of this mathematical archiecture on other complex spatio-temporal systems like forced isotropic turbulence and global sea-surface temperature.

1.2 Significance

The research conducted in this dissertation contributes to the biomechanics, robotics, exoskeleton, wearable technology, sensing technology, data-driven engineering, and dynamical systems communities. The ability to learn models of complex behavior and quantify abstract concepts purely from data is paramount for personalized medicine and rehabilitation. These data-driven modeling methods are not solely intended to progress methodological innovation, but also to provide translational tools that improve movement-based health outcomes. The goal of this dissertation was to develop and deploy data-driven techniques for dynamical models to understand complex system behavior like human movement. The primary contributions of this dissertation are:

- **Resolving missing physics errors in time series of dynamical systems**

  We developed the first comprehensive and interchangeable framework for resolving missing physics in dynamical models of complex systems. Importantly, we distinguish between two distinct but nuanced approaches for estimating missing physics from the model-measurement mismatch. Previous strategies to model complex systems either used purely data-driven models (which misses the opportunity to leverage domain knowledge about the system) or used purely mathematical models like differential equations (which limits the ability to model complex behavior with prescribed accuracy). We demonstrate a hybrid approach that combines mechanistic and data-driven modeling techniques. This code has been made publicly available: *github.com/meganebers/Discrepancy-Modeling-Framework-code*.

- **Demonstrating that individual exoskeleton responses can be quantified**

  It is challenging to predict how an individual will respond to an exoskeleton. Instead of jumping to identify physiological mechanisms underlying exoskeleton responses – which have not been shown to generalize across individuals – we summarize the response dynamics as an all encompassing *discrepancy* to investigate the predictability of responses, as well as what measurements are needed to characterize it. We successfully demonstrated that individual exoskeleton responses were predictable, as quantified by neural network-based discrepancy modeling. However, joint kinematics and electromyography measurements alone were insufficient to capture the complex muscle-level changes in gait with ankle exoskeleton.

- **Leveraging time histories of sensor measurements to expand biomechanical datasets**

  Human motion tracking and analysis is essential for monitoring disease progression, guiding rehabilitation treatment, evaluating sports performance, and informing assistive device design. However, the technologies developed to track human movement are at odds: laboratory-based motion capture is the gold-standard for comprehensive data collection but is expensive and fixed, while devices like wearable sensors are cheaper and more portable but suffer from inaccuracies due to drift, placement, and signal interference. Therefore, we use a shallow recurrent decoder network that leverages time histories of sensor measurements; this algorithm learns a mapping from as few as one sensor to a full state estimate, enabling the monitoring of a laboratory-level comprehensive dataset with the portability of a single wearable sensor. While this methodology is exciting, it is important to extend this approach for meaningful clinical applications. Beyond human biomechanics, this mathematical framework can generalize to other complex dynamical systems with mobile sensing, such as forced isotropic turbulence or global sea-surface temperature; the use of dynamical trajectories to embed global information of the spatio-temporal dynamics of a sparsely measured system address the long-standing challenge of sensor path planning.

1.3 Dissertation Overview

This dissertation is focused around three research studies that are presented as self-contained journal articles. After this general introduction, Chapter 2 outlines the background material related to dynamical systems, data-driven modeling, and human movement that is useful for interpreting the following chapters. Chapter 3 introduces a framework for investigating model-measurement mismatch to resolve missing physics error (Ebers, et al, under review in *SIAM Journal on Applied Dynamical Systems*, 2022). Chapter 4 demonstrates that discrepancy modeling is a useful tool for characterizing individual responses to ankle exoskeletons (Ebers, et al, published in *Journal of Biomechanics*, 2023). Chapter 5 examines how time histories of sensor measurements can reconstruct full-state estimates of sparsely-measured systems, such as to expand biomechanical datasets to support collecting rich datasets representative of human movement 'in the wild' (Ebers, et al, *in preparation*). The final chapter of this dissertation summarizes the important findings of these studies and outlines areas for future research. The pronoun 'we' is used throughout this dissertation to recognize that research is a team process that requires input and collaboration from multi-disciplinary teams. The individuals who contributed to these studies are listed at the beginning of each chapter.

Chapter 2
BACKGROUND

2.1 Dynamical systems

The world around us is ever evolving. As researchers, our job is to observe how things change, how fast things change, their predictability, influences, and effects; all in the name of understanding. Curious people – whether mathematicians, physicists, chemists, or others – for hundreds of years have tried their hand at quantifying change. One could argue all scientific and technological innovation has come about through the observation and quantification of change. This dissertation adds to our theoretical framework for quantifying and interrogating change in our modeling methods, with an eye towards applications in human movement. The following sections outlines the over-arching theoretical frameworks underpinning this work for data-driven dynamical systems.

2.1.1 Mathematical modeling

Mathematicians developed a concept for quantifying change using dynamical systems. dynamical – exhibiting continual change – systems provide a framework to capture the rich interactions of quantities that co-evolve in time. In doing so, we can analyze, predict, and control how a system behaves. The French mathematician Henri Poincare's work in celestial mechanics is often credited as the qualitative founding of dynamical system theory [3, 4]. While dynamical systems theory focuses on methodologies to address phenomena such as nonlinear oscillations and chaos, the more general concept of dynamical systems has its origins in classical mechanics, such as Newton's laws of motion. Traditionally, dynamical systems are modeled using differential equations to describe the evolution of the state of a system. In physics and engineering systems, these traditional models rely on the development of governing equations that characterize underlying dynamical processes. Such governing equations are typically derived through asymptotic reductions, enforcing physical constraints or conservation laws, and/or positing empirical relations between variables. While these mathematical models of dynamical systems have supported immense technological advancements, the challenges we are facing in physical and engineering applications simply cannot be represented by first principles alone (due to dominant balance approximations).

2.1.2 *Data-driven dynamical modeling*

Modern approaches to model dynamical systems have leveraged advances in machine learning and data science to characterize, discover, and simulate these systems purely from data [5]. Data-driven dynamical models are mathematical models that are constructed using observational or measurement data, rather than theoretical principles or first-principles models and therefore don't require *a priori* knowledge of a system. Such models are advantageous when researchers and engineers need to capture the complexity of a system (*e.g.*, nonlinearity, multiscale, high-dimensionality) but may not have a complete or robust understanding of its underlying mechanisms. Data-driven dynamical systems have myriad applications, including predicting weather patterns, modeling infectious diseases, and analyzing brain activity. Algorithmic methods like dynamical mode decomposition [6], the sparse identification of nonlinear dynamics [7], and data-driven discovery of coordinates and governing equations [8] show immense promise for capturing the complexity of real-world systems such that we can monitor, predict, and control system behavior.

The emergence of data-driven approaches for modeling dynamical systems can be attributed to factors including advancements in computational capabilities, proliferation of data sources, and improvements in modeling techniques. With the availability of large-scale computing, we are able to process and analyze large amounts of information and in a relatively quick amount of time. This is especially important for machine learning methods like reduced order modeling [9] or compressive sensing [10] in which we can learn complex patterns and dynamics from large data. Regarding the data itself, we now have (i) more data than ever before, (ii) new types of data sources, and (iii) improved resolution of sensing technologies, all leading to an abundance of data from which to build dynamical models [5]. Improvements in modeling techniques, such as for real-time decision making or adaptability of models with updated information, have enabled data-driven models of dynamical systems to thrive where first-principles models could not.

Inevitably, there are drawbacks to data-driven modeling of dynamical systems. First, a model is only as good as its data ("garbage in, garbage out"). If a dataset is biased, incomplete, or noisy, the input-output relationship learned will inherently incorporate such effects. Overfitting is also a common problem in any data-driven modeling setting, and dynamical systems are no exception. Techniques like regularization and cross-validation are vital for ensuring generalizability of a learned model (*e.g.*, [7]). While using machine learning techniques like neural networks or support vector machines are powerful, interpretability into the physical mechanisms underlying the dynamics is limited. Innovations in physics-based or physically-interpretable model discovery methods have enabled more insights for explaining system behavior; the sparse identification of nonlinear dynamics [7] and dynamical mode decomposition [6] are two popular and interpretable data-driven methods. Finally, despite advancements in computational capabilities, training and evaluating machine learning models of dynamical systems can still be expensive, especially for large-scale and high-dimensional data. Careful consideration and validation of the model's performance and assumptions are critical to ensure that a model is providing useful and accurate insights.

2.1.3 Time-series data

Time-series data – a sequence of data points collected at regular intervals over time – are useful for modeling complex dynamics, as time series provides a rich basis of information about how a system or process evolves over time. Further, complex dynamics often emerge from quantities co-evolving in time; such interactions are encoded in time-series data. By analyzing models build from time series, we can extract patterns and trends that lend insight into the underlying mechanisms governing system behavior. Time-series data also allows feedback and control policies to be build into dynamical models, enabling mechanistic insight in the presence of external forcing or informing the design of control strategies for regulating system behavior. In machine learning and data science, coding programs such as python or MATLAB store time series as arrays or matrices. For example, if time-series data X has N states, or features, and these states are observed T times, the data is stored as a matrix of size $N \times T$, such that states are placed in rows and observations in time are placed in columns: $X \in \mathbb{R}^{N \times T}$.

In using time-series data for building models, there no standard consensus across fields – whether statistics, engineering, applied mathematics, or physics – how to define model applications. Terms like predict, forecast, reconstruct, and estimate are often used colloquially or interchangeably. In this dissertation, my working understanding of such terms are as follows:

1. **predict**: using a trained model to guess a new outcome

2. **forecast**: special case of predict that specifically attaches a temporal aspect for guessing a new outcome (*e.g.*, prediction of an outcome forward in time)

3. **reconstruct**: using a trained model to recreate the original data

4. **estimate**: inferring properties of the data, such as model parameters

Reconstruct and estimate are often used interchangeably; however, various fields use each term slightly different for model applications. Further, the appropriate choice of term may be affected by how training versus test data are partitioned (*e.g.*, train/test data randomly partitioned in a time domain versus train/test data partitioned based on time, such as the first 70% and last 30% of the time series).

2.2 Data-driven dynamical models of human movement

The human body can be thought of as a dynamical system; it exhibits the quintessential characteristics of one: complexity, nonlinearity, high-dimensionality, multi-scale. Further, the human body is well represented by Newtonian equations of motion, and how human movement evolves over time is well represented by differential equations [11, 12]. Machine learning for dynamical modeling is a powerful tool to complement current computational approaches for modeling human movement. Beyond complement, such approaches have the opportunity to fundamentally change how we investigate human movement.

2.2.1 Machine learning for modeling human movement

Machine learning is not a new approach to study human movement biomechanics; it has enabled researchers and clinicians to push the boundaries of sensing and analysis of human motion [13]. These algorithms have proven to be a powerful tool for extracting salient features and modeling complex relationships from the deluge of data collected from laboratory experiments, rehabilitation clinics, and wearable sensors. For example, principal component analysis is a staple in biomechanics for coordination and feature extraction [14–17]. Classification is the most common application of machine learning in biomechanics [13]; such techniques are most often used for identifying movement patterns, such as distinguishing between pathological and non-pathological kinematics [18–22]. More advanced data-driven methods have transformed human motion tracking and analysis for tasks like human activity recognition [21, 23–27], marker-less motion capture and pose estimation [28–31], fall detection [32–35], and sensor fusion [36–39], among others. Machine learning for regression models is a less common application than classification in human movement biomechanics [13]. As is explored more deeply in this dissertation, machine learning specifically for dynamical (regression) models is in its infancy for human movement biomechanics.

2.2.2 *Time-series data of human movement*

Human motion tracking and analysis is essential for monitoring disease progression, guiding rehabilitation treatment, evaluating sports performance, and informing assistive device design. Biomechanists traditionally characterize motion, such as gait, by measuring biomechanical variables like joint kinematics, kinetics, and spatio-temporal parameters [40]. Certain biomechanical variables have been established as biomarkers that correlate with meaningful outcomes, such as knee adduction angle for ACL injury or step width variability for aging/fall risk [41–45]. While motion tracking is used across clinical, research, and sports settings, the spectra of available technologies varies widely in practicality, accuracy, and expense. Optical motion capture and force plates are considered the gold standard to comprehensively capture kinematics (motions) and kinetics (forces) [46]. Motion capture cameras and sensors are highly specialized and require a fixed laboratory space, careful calibration, and expert professionals to conduct a gait study, thereby limiting the number of hospitals/clinics that have dedicated motion capture space. Conversely, more portable devices offer opportunities for out-of-lab motion tracking and across many repetitions, including electromyography (EMG), inertial measurement units (IMUs), depth cameras, red-green-blue (RGB) cameras, pedometers, and pressure insoles. Electromyography (EMG) is common technique in experimental research for measuring the electrical activity and timing produced during neural activation of muscles. While portable devices like video or wearable sensors can estimate biomechanical variables in any environment, such systems tend to be less accurate than optical motion capture. Other common time series included in data collection include physiological data – which are linked to movement – include heart rate, skin temperature, electrodermal activity, blood oxygen saturation, oxygen consumption, carbon dioxide production, breath frequency, and more.

Kinematics and EMG recordings are the most common time-series data used to quantify an individual's movement during a task. Kinematic data are generated by collecting marker motion record by an optical motion capture systems, such as Qualysis. Joint kinematics can then be estimated from marker data; a common workflow is estimating kinematics using the Inverse Kinematics algorithm in OpenSim 3.3 with a 19 degree-of-freedom skeletal model [47, 48]. Joint kinematics were low-pass filtered at 6Hz using a fourth-order Butterworth filter [1]. EMG data were high-pass filtered at 40Hz, rectified, and low-pass filtered at 10Hz using fourth-order Butterworth filters. All human subjects procedures were evaluated by the Institutional Review Board at the University of Washington. For nondisabled individuals, left and right kinematics variables exhibit relatively symmetric waveforms. However, some inter-step variability is expected, as the neuromusculoskeletal system is rhythmic yet noisy. EMG is recorded bilaterally from the major muscles responsible for the prescribed task motion; muscles typically include vastus medialis, soleus, gastrocnemius, gluteus medius, lateral hamstrings, in addition to others as needed/available. EMG data are high-pass filtered at 40Hz, rectified, and low-pass filtered at 10Hz using fourth-order Butterworth filters [1].

2.3 Data-driven model discovery for dynamical systems

This section is a brief mathematical overview of common model discovery methods for learning dynamical models from data. There are many variants and improvements that can be implemented for each method considered; this section focuses only on the general forms of each model discovery method. Hyper-parameter tuning of each technique is problem specific and again depends upon measurement characteristics (quantity, quality, resolution) and the intent of its use. Each of these methods are used to some degree in the subsequent chapters. For example, Chapter 3 introduces a mathematical framework for resolving model-measurement mismatch; this mismatch, or discrepancy, due to missing physics can be estimated using any one of the model discovery methods outlined in this section. In Chapter 4, neural networks are used to learn dynamical models of the biomechanical, neural, and sensory processes driving the time evolution of gait. In Chapter 5, shallow recurrent decoder networks are used to reconstruct full-state biomechanical variables from sparse sensors using time histories of sensor measurements.

2.3.1 Gaussian Process Regression (GPR)

Gaussian process regression (GPR) is a non-parametric supervised learning approach to regression that uses a Gaussian process prior for Bayesian inference [49]. Not only is GPR a powerful nonlinear interpolation tool, its inherent probabilistic architecture allows for uncertainty quantification of interpolated values. Impressively, this method is a universal approximator with a closed form solution. Consider the dataset $D\{(\mathbf{X}_i, \mathbf{Y}_i | i = 1, \ldots, n\}$ which is split into test and training subsets where $\mathbf{X}_i$ are individual observations and $\mathbf{Y}_i$ are observation labels. A Gaussian process $f(\mathbf{X}_i)$ is defined by a mean function $m(\mathbf{X}_i)$ and covariance function $\kappa(\mathbf{X}_i, \mathbf{X}_j)$ as

$$\begin{aligned} m(\mathbf{X}_i) &= \mathbb{E}[f(\mathbf{X}_i)] \\ \kappa(\mathbf{X}_i, \mathbf{X}_j) &= \mathbb{E}[(f(\mathbf{X}_i) - m(\mathbf{X}_i))(f(\mathbf{X}_j) - m(\mathbf{X}_j))] \end{aligned}$$

such that the Gaussian process is:

$$f(\mathbf{X}_i) \sim \mathcal{GP}(m(\mathbf{X}_i), \kappa(\mathbf{X}_i, \mathbf{X}_j)). \tag{2.1}$$

Kernel functions are used to calculate covariance and enforce assumptions about the data. Specifically, kernels are parameterized by hyperparameters controlling covariance characteristics (*e.g.* length-scales or periodicity) and are optimized by maximizing the log marginal likelihood during model selection. As the prior distribution is a Gaussian process $f(\mathbf{X}_i) \sim \mathcal{GP}$, the conditional distribution $f(\mathbf{X}_i)|D_n$ is the posterior, i.e. the *predictive distribution*. Computing on a finite-sized data set and partitioning it into training inputs ($\mathbf{X}$), training outputs ($\mathbf{f}$), test inputs ($\mathbf{X}_*$), and test outputs ($\mathbf{f}_*$), the prior joint distribution is assumed to take the form:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \kappa(\mathbf{X},\mathbf{X}) & \kappa(\mathbf{X},\mathbf{X}_*) \\ \kappa(\mathbf{X}_*,\mathbf{X}) & \kappa(\mathbf{X}_*,\mathbf{X}_*). \end{bmatrix}\right) \tag{2.2}$$

To find the posterior, or predictive, distribution, the joint prior distribution must be restricted, or conditioned, to contain functions that agree with the observations, $\mathbf{f}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{f} \sim \mathcal{N}(\bar{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*))$, which can be written in closed form [49]. An important note: GPR for prediction or parameter estimation is computationally expensive. Calculating the maximum likelihood requires finding the determinant and inverse of the covariance matrix, which has cubic computational complexity. Ultimately, GPR is about regressing to a Gaussian distribution and estimating the appropriate variances via (2.2). The mathematical details of GPR can be found here [49].

2.3.2 dynamical Mode Decomposition (DMD)

dynamical mode decomposition (DMD) is a modern system identification approach based on data-driven regression. DMD extracts spatio-temporal structures from time-series data and learns a low-dimensional linear model describing the evolution of features that encode salient system behavior. Consider a dynamical system measured at evenly-spaced time points $\mathbf{t} = [t_1, t_2, \cdots, t_m]$. From measurements $\mathbf{x}_k = \mathbf{x}(t_k)$, we construct a matrix of snapshots $\mathbf{X}(\mathbf{t}) = [\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_n] \in \mathbb{R}^{m \times n}$. A discrete time linear representation of a system is assumed to take the standard form, $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$ such that the linear operator $\mathbf{A}$ progresses the state vector $\mathbf{x}_t$ forward in time. Mirroring this standard form, two snapshot matrices are defined as:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix}, \tag{2.3}$$

$$\mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix} \tag{2.4}$$

where $\mathbf{X}'$ is the time-shifted matrix of snapshots $\mathbf{X}$, *i.e.* $\mathbf{X}' = \mathbf{A}\mathbf{X}$. The *exact* DMD is the best fit linear mapping $\mathbf{A}$ between snapshot pairs $\mathbf{X}$ and $\mathbf{X}'$:

$$\mathbf{A} = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X}' - \mathbf{A}\mathbf{X}\|_F = \mathbf{X}'\mathbf{X}^\dagger \tag{2.5}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\dagger$ is the Moore-Penrose pseudoinverse [6]. DMD exploits the singular value decomposition (SVD) to solve for:

$$\mathbf{A} = \mathbf{U}^*\mathbf{X}'\mathbf{V}\boldsymbol{\Sigma}^{-1}. \tag{2.6}$$

DMD exploits low-rank structure of high-dimensional systems, and thus projects $\mathbf{A}$ onto the first r modes of the principle components $\mathbf{U_r}$. This rank-r truncation of $\mathbf{X} \approx \mathbf{U_r}\boldsymbol{\Sigma}_\mathbf{r}\mathbf{V}_\mathbf{r}^*$ approximates the pseudo-inverse:

$$\tilde{\mathbf{A}} = \mathbf{U}_r^*\mathbf{A}\mathbf{U}_r = \mathbf{U}_r^*\mathbf{X}'\mathbf{V}_r\boldsymbol{\Sigma}_r^{-1}. \tag{2.7}$$

The eigendecomposition of $\tilde{\mathbf{A}}$ yields the eigenvalues and eigenvectors, $\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\boldsymbol{\Lambda}$, which provides insight into underlying system properties such as growth modes and resonance frequencies [6]. The eigenvectors of $\mathbf{A}$ are the *DMD modes* $\mathbf{\Phi}$:

$$\mathbf{\Phi} = \mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}\mathbf{W}. \tag{2.8}$$

Along with the mode amplitudes, $\mathbf{b} = \mathbf{\Phi}^{\dagger}\mathbf{x}_1$, the well known DMD solution takes the form

$$\mathbf{x}(t) = \sum_{i=1}^{r} \phi_i e^{\omega_i t} b_i = \mathbf{\Phi}\mathbf{exp}(\mathbf{\Omega t})\mathbf{b}. \tag{2.9}$$

However, exact DMD is prone to biased errors resulting from noisy measurements, affecting model fit and forecasting stability [50]. Therefore, Askham and Kutz [51] introduced *optimized* DMD, which uses variable projection to perform nonlinear optimization for de-biasing model fitting in the presence of observation noise. More specifically, the variable projection method optimally computes nonlinear least squares exponential fitting for DMD:

$$\underset{\omega,\Phi_b}{\text{argmin}} \|\mathbf{X} - \mathbf{\Phi}\mathbf{exp}(\mathbf{\Omega}t)\mathbf{b}\|_F. \tag{2.10}$$

Note: the rank of $\mathbf{A}$ cannot exceed the state dimension of $\mathbf{X}$, and DMD algorithms rely on the availability of full-state measurements, typically of high dimensions. When only partial observations or low-dimensional system measurements are available, it is helpful to build an augmented state vector that is 'lifted' into a higher dimension. This can be accomplished this via time-delay embedding, which also happens to result in an intrinsic coordinate system forming a Koopman-invariant subspace in which nonlinear dynamics appear linear [52].

2.3.3 *Sparse identification of nonlinear dynamics (SINDy)*

Sparse identification of nonlinear dynamics (SINDy) recovers parsimonious representations of the dynamics from measurement data by sparse regression to a library of candidate models [7, 8, 53]. Consider a nonlinear dynamical system measured at time points $\mathbf{t} = [t_1, t_2, \cdots, t_m]$. From measurements, we construct the matrix $\mathbf{X}(\mathbf{t}) = [\mathbf{x}_1(\mathbf{t})\ \mathbf{x}_2(\mathbf{t})\ \ldots\ \mathbf{x}_n(\mathbf{t})]$

$\in \mathbb{R}^{m\times n}$. The method introduced in [7] seeks to identify $\mathbf{f}$ via sequential threshold least-squares, which is a proxy for the sparsifying zero-norm. The set of n state measurements are used to populate a library of candidate nonlinear terms $\boldsymbol{\Theta}(\mathbf{X}) = [\mathbf{1}^\top\ \mathbf{X}^\top\ (\mathbf{X}\otimes\mathbf{X})^\top\ \cdots\ \sin(\mathbf{X})^\top]$, where $\mathbf{x}\otimes\mathbf{y}$ defines the vector of all product combinations of the state components. Each candidate term should be unique, as a suitable library is crucial in the SINDy algorithm. A common strategy is to start with polynomials and increase the complexity of the library with other terms, such as trigonometric functions. Thus, a dynamical system can be re-written as:

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi}. \tag{2.11}$$

The time derivatives $\dot{\mathbf{X}}(\mathbf{t}) = [\dot{\mathbf{x}}_1(\mathbf{t})\ \dot{\mathbf{x}}_2(\mathbf{t})\ \ldots\ \dot{\mathbf{x}}_n(\mathbf{t})]$, if not measured directly, can be found via numerical differentiation and should be appropriately de-noised, if necessary [54–60]. The coefficients $\boldsymbol{\Xi}$ are the *sparse* weightings of the corresponding candidate library terms. Therefore, our regression relies on sparse regularization to enforce a parsimonious $\boldsymbol{\Xi}$ corresponding to the fewest nonlinear terms in our library that describe our dynamics well:

$$\boldsymbol{\Xi} = \arg\min_{\hat{\boldsymbol{\Xi}}} \|\boldsymbol{\Theta}(\mathbf{X})\hat{\boldsymbol{\Xi}} - \dot{\mathbf{X}}\|_2 + \lambda\|\hat{\boldsymbol{\Xi}}\|_0 \tag{2.12}$$

Regressing to the zero-norm is often achieved by relaxing the one-norm. However, modern optimization frameworks are allowing for computationally tractable proxies for the zero-norm that are superior to the one-norm relaxation [61].

2.3.4 Neural Networks (NN)

Artificial neural networks, or simply *neural networks* (NN), are mathematical models inspired by biological neural networks. While there is a wide array of literature on NN [62–69], we outline the basic concepts. NN learn a mapping between a set of input data and target outcomes. The middle, or hidden, layers form a compositional structure that optimizes the association between the training data set. The user designates the depth (number of hidden layers), the dimensionality (number of nodes) of each layer, and how each layer is connected.

Linearly, this means a NN optimizes over the compositional function to learn the neural network weights and biases matrices $\mathbf{A}_j$ between the k-hidden layers:

$$\underset{\mathbf{A}_j}{\operatorname{argmin}}(f_M(\mathbf{A}_M, \dots, f_2(\mathbf{A}_2, f_1(\mathbf{A}_1, \mathbf{x}))\dots) + \lambda g(\mathbf{A}_j)), \tag{2.13}$$

where $\lambda g(\mathbf{A}_j)$ is included to provide and appropriate regularization for the solution. For example, a simple, single hidden layer $(k = 1)$ NN is structured as:

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{A}_1 \mathbf{x} \\ \mathbf{y} &= \mathbf{A}_2 \mathbf{x}^{(1)} \end{aligned}. \tag{2.14}$$

Leveraging the compositional structure, a mapping is defined by:

$$\mathbf{y} = \mathbf{A}_2 \mathbf{A}_1 \mathbf{x}, \tag{2.15}$$

which generalizes to M layers

$$\mathbf{y} = \mathbf{A}_M \mathbf{A}_{M-1} \dots \mathbf{A}_2 \mathbf{A}_1 \mathbf{x}. \tag{2.16}$$

Nonlinear mappings are structured similarly. In this case, nonlinear activation functions connect hidden layers and are given by:

$$\begin{aligned} \mathbf{x}^{(1)} &= f_1(\mathbf{A}_1, \mathbf{x}) \\ \mathbf{y} &= f_2(\mathbf{A}_2, \mathbf{x}^{(1)}). \end{aligned} \tag{2.17}$$

Further, nonlinear activation functions, $f_j(\cdot)$, can differ between layers. Thus, nonlinear mapping between a given set of input and output data over M layers is structured as:

$$\mathbf{y} = f_M(\mathbf{A}_M, \dots, f_2(\mathbf{A}_2, f_1(\mathbf{A}_1, \mathbf{x}))\dots) = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) \tag{2.18}$$

where $\mathbf{f}_{\boldsymbol{\theta}}(\cdot)$ represents the overall network structure with weights and biases $\boldsymbol{\theta}$. While often used for classification, NNs can be structured to learn the evolution of dynamical systems. NN for dynamical systems provides a flexible and powerful architecture for high-dimensional supervised learning of system behavior for future state predictions [70].

2.3.5 *Shallow Recurrent Decoder Networks (SHRED)*

The shallow recurrent decoder (SHRED) model is a neural network mapping from a time history of sensor measurements to a high-dimensional, spatio-temporal state. The algorithm is expressed as:

$$\mathcal{H}(\{\mathbf{y}_i\}_{i=t-k}^{t}) = \mathcal{F}(\mathcal{G}(\{\mathbf{y}_i\}_{i=t-k}^{t}; \mathbf{W}_{RN}); \mathbf{W}_{SD}) \tag{2.19}$$

where $\mathbf{y}_t$ consists of measurements of the high-dimensional state $\mathbf{x}_t$, $\mathcal{F}$ is a fully-connected, feed-forward neural network parameterized by weights $\mathbf{W}_{SD}$ and $\mathcal{G}$ of an LSTM network parameterized by weights $\mathbf{W}_{RN}$. The SHRED architecture $\mathcal{H}$ minimizes reconstruction loss,

$$\mathcal{H} \in \underset{\bar{\mathcal{H}} \in \mathscr{H}}{argmin} \sum_{i=1}^{N} ||\mathbf{x}_i - \bar{\mathcal{H}}(\{\mathbf{y}_j\}_{i=t-k}^{t})||_2 \tag{2.20}$$

given a set of training states $\{\mathbf{x}_i\}_{i=1}^{N}$ and corresponding measurements $\{\mathbf{y}_i\}_{i=1}^{N}$. The network is trained to minimize reconstruction loss using the ADAM optimizer [71].

The reconstruction error is calculated as the averaged mean-squared error over each state in the partitioned test set:

$$\text{MSE Error} = \frac{1}{T} \sum_{i=1}^{T} \frac{||\mathcal{H}(\{\mathbf{y}_j\}_{i=t-k}^{t}) - \mathbf{x}_i||}{||\mathbf{x}_i||_2} \tag{2.21}$$

Due to SHRED's reliance on the time history of measurements for state estimation, each dataset is truncated such that only the final $N - k$ temporal snapshots are reconstructed, where N is the initial number of examples and k is the length of the utilized time history.

This dissertation adopts the perspective that dynamical models of human movement constructed from experimental measurements will provide substantial and unique insight into individual-specific responses to assistive devices, such as passive-elastic ankle exoskeletons, and enable rich reconstructions of both personalized and population-based datasets to infer movement-based health outcomes. The following chapters detail applications of machine learning of dynamical models for complex systems, particularly the human musculoskeletal system, but also the Van der Pol oscillator, the chaotic Lorenz attractor, Burgers' shock wave equation, forced isotropic turbulence, and global sea-surface temperature. This dissertation provides the foundation for building models and extracting insights from time-series data. The long term goal of this work is to advance how we investigate pathology, enhance mobility, and personalize rehabilitation by utilizing advancements in machine learning for dynamical systems.

Chapter 3

DISCREPANCY MODELING FRAMEWORK: LEARNING MISSING PHYSICS, MODELING SYSTEMATIC RESIDUALS, AND DISAMBIGUATING BETWEEN DETERMINISTIC AND RANDOM EFFECTS

Physics-based and first-principles models pervade the engineering and physical sciences, allowing for the ability to model the dynamics of complex systems with a prescribed accuracy. The approximations used in deriving governing equations often result in discrepancies between the model and sensor-based measurements of the system, revealing the approximate nature of the equations and/or the signal-to-noise ratio of the sensor itself. In modern dynamical systems, such discrepancies between model and measurement can lead to poor quantification, often undermining the ability to produce accurate and precise control algorithms. We introduce a discrepancy modeling framework to identify the missing physics and resolve the model-measurement mismatch with two distinct approaches: (i) by learning a model for the evolution of systematic state-space residual, and (ii) by discovering a model for the deterministic dynamical error. Regardless of approach, a common suite of data-driven model discovery methods can be used. Specifically, we use four fundamentally different methods to demonstrate the mathematical implementations of discrepancy modeling: (i) the *sparse identification of nonlinear dynamics* (SINDy), (ii) *dynamic mode decomposition* (DMD), (iii) *Gaussian process regression* (GPR), and (iv) *neural networks* (NN). The choice of method depends on one's intent (*e.g.*, mechanistic interpretability) for discrepancy modeling, sensor measurement characteristics (*e.g.*, quantity, quality, resolution), and constraints imposed by practical applications (*e.g.*, state- or dynamical-space operability). We demonstrate the utility and suitability for both discrepancy modeling approaches using the suite of data-driven modeling methods on three continuous dynamical systems under varying signal-to-noise ratios. Finally, we emphasize structural shortcomings of each discrepancy modeling approach depending on error type. In summary, if possible, one should try to discover missing physics as a top priority in building discrepancy models. If this is not possible, established data assimilation frameworks are the next preferred method.

3.1 Introduction

The traditional modeling of physics and engineering systems relies on the development of governing equations that characterize the underlying nonlinear, dynamical processes. Such governing equations are typically derived through asymptotic reductions, enforcing physical constraints or conservation laws, and/or positing empirical relations between variables [72]. Simulation of the governing equations then allows for prediction, control, and characterization of the complex system. However, it is well known that governing equations are often idealized and only approximate, either achieved through dominant balance physics arguments or neglecting higher-order effects [73–76]. In many emerging fields, the idealized models currently used are simply inadequate for modeling applications where precision is necessary, such as in robotics, biomechanics, precision manufacturing, and automated systems. The time evolution of complex nonlinear systems is often highly sensitive to small errors in system dynamics. This limits the utility of simulation, such as for control or inference. Generally, there are two kinds of errors that occur in modeling physical systems: missing physics and measurement error (which can be systematic or random). In practice, both errors exist and are difficult to disambiguate. Our framework provides approaches which help disambiguate between the dominant error forms to estimate the missing physics, either by learning the deterministic dynamical error or characterizing the state-space residual between model and measurement.

With the advancement of modern sensor technologies, there is opportunity to improve the characterization of system dynamics through modern data-driven methods to refine and augment the known, governing first-principles. Indeed, a better understanding of the underlying physical processes can be achieved by inspecting the error between first-principles theory and sensor measurements of dynamical systems. The error may contain deterministic effects, or *discrepancies*, and models of the discrepancy can be learned using data-driven model discovery. A number of machine learning techniques have been developed to learn model error or discrepancies using hybrid data assimilation techniques, in which additive correction models of the missing physics [57, 77–79] are learned for a diversity of applications. For example, Levine and Stuart [80] recently proposed leveraging machine learning methods, specifically recurrent neural networks, for modeling hybrid problems in non-Markovian settings. By augmenting known first-principles with a discrepancy model estimating the missing physics, an improved dynamical model can be learned. In certain practical applications, such as in controls engineering [81–85], the dynamical model may be unavailable or infeasible to interface with, thereby necessitating learning a discrepancy model of the state-space residual to correct system approximations. Although we have always improved our models through systematic approaches, we build on recent work [80, 86] and outline here two principled, data-driven approaches by which discrepancy modeling is automated. Moreover, we examine the relationship between deterministic and random effects within the model-measurement mismatch; experimental noise in sensor measurements dictate limitations in learning a discrepancy model.

Discrepancy modeling has deep historical context, especially since early models of any system are typically coarse approximations of the physics. Indeed, throughout the 1960s, before the advancement of scientific computing, asymptotic and perturbation methods [73, 74] that systematically introduced discrepancies were leading methods in the development of fluid dynamics. From the Prandtl number to the Reynolds number, various approximations led to different dominant balance physics approximations [75, 76, 87]. Thus by construction, such models allowed for improved understanding at the expense of detailed models. Computation has allowed us to move beyond such reductions, yet model-measurement mismatch continues to prevent accurate interpretation of system structure and/or prediction of time evolution. In modern applications, discrepancy modeling can play a foundational role in improved modeling across any data-driven application. As an example, consider emerging digital twin technologies which require a computational model of a real system. Such models are typically idealized from first principle physics, which fail to accurately match reality. This is especially problematic where precision is required in the digital twin. Thus almost any practical systems of interest can benefit from a discrepancy modeling improvement and evaluation. Kennedy and O'Hagan [88] provide a thorough review of uncertainty quantification techniques from a statistical perspective. The current methods advocated, which are focused on dynamical systems, give additional techniques which can be used for improved modeling purposes. Indeed, in almost every application area, mismatch exists between experiment and theory which is not due to noise. For instance, improvements in tracking planetary motion in the late 1800s and early 1900s allowed for the characterization of a discrepancy between Newton's gravitation laws and the observed physics, eventually leading to the development of general relativity by Einstein [89]. More recently, identifying missing deterministic effects (provided they are not obscured by noise) has allowed for the discovery of missing physics that are challenging to model with first principles, such as fluid drag forces of falling objects [90] or bearing chatter during double pendulum control [86]. Our goal is to leverage improved sensor observations; we automate the process of building better models by identifying a discrepancy model. Note that this is a significantly different task than characterizing the sensitivity of the system to initial condition which is a hallmark feature of chaotic systems.

The ideas in our discrepancy modeling framework are well established concepts in statistics, optimization, and controls [91–96]. In statistical regression analysis, measures of deviation are referred to as either *error* or *residual*. An *error* is the difference between the observed values and the true values. It is important to note that true values are unobservable and thus error values are inaccessible. A *residual* is the difference between the observed values and the estimated values. In practice, observed values are used as a proxy for the true values; therefore, residuals may contain both random and deterministic signals. Regression analysis seeks to evaluate how well a statistical model fits a data set; if the residual contains a bias, it suggests the model can be improved by capturing deterministic values within the residual. We posit that discrepancy modeling is to dynamical systems modeling what regression analysis is to statistical modeling. In controls engineering, traditional approaches to discrepancy modeling in a state-space representation include Kalman filtering for data assimilation; however, the dominant assumption is that the mismatch between model and measurement are given by normally-distributed variables, *i.e.*, random processes [96, 97]. Therefore, the state-space model is not updated through self-improvement to account for non-random processes. Whether modeling a dynamic system in a dynamical or state space, the error may reveal deterministic structure affecting the system's observed time evolution. As already noted, Levine and Stuart [80] recently have advanced the state-of-the-art by leveraging machine learning methods (neural networks) for learning dynamical error. More broadly, they provide a rigorous analysis for learning such models from data in a dynamical representation, *e.g.*, differential equations. In this work, we build on this theme by considering a broader class of models, including those constructed by *sparse identification of dynamical systems* (SINDy), *dynamic mode decomposition* (DMD), and *Gaussian process regression* (GPR). We also advocate for learning a discrepancy model of the systematic residuals while operating in the state space. Again, we demonstrate SINDy, DMD, and GPR along with neural networks (NN) to learn models of the systematic residual directly and correct our state space approximations. By construction, identifying missing physics in continuous dynamical systems is best treated by learning the discrepancy model in the dynamical space; however, circumstances may dictate learning a discrepancy model in the state space, such as for discrete dynamical systems or measurement error. Both approaches to discrepancy modeling are valid – despite one being more favorable by construction – towards resolving model-measurement mis-

Variable	Description
$\mathbf{x}(t) \in \mathbb{R}^n$	approximate state space
$\tilde{\mathbf{x}}(t) \in \mathbb{R}^n$	augmented state space
$\mathbf{x}_0(t) \in \mathbb{R}^n$	true state space (inaccessible)
$\mathbf{y}_k = \mathbf{y}(k\Delta t) \in \mathbb{R}^n$	measurements
$f(\cdot)$	approximate dynamics
$g(\cdot)$	missing physics
$F(\cdot)$	true dynamics
$\tilde{F}(\cdot)$	augmented dynamics
$\mathbf{x}_D(t_k) = \mathbf{y}_k - \mathbf{x}(t_k)$	state-space residual
$\dot{\mathbf{x}}_D(t_k) = \dot{\mathbf{y}}_k - f(\mathbf{y}_k)$	deterministic dynamical error
$\delta(\cdot)$	discrepancy model estimating the state-space residual
$\delta_f(\cdot)$	discrepancy model estimating the deterministic dynamical error

Table 3.1: Variable definitions. The goal of discrepancy modeling is to improve the Platonic or idealized model via augmentation of a learned discrepancy model, such that the model solution and measurement data converge. There are two approaches to build a discrepancy model and estimate missing physics in a dynamical system: learn the deterministic dynamical error or model the systematic state-space residual.

match in dynamical systems. This hybrid (mechanism+data) framework brings together domain knowledge from first principles and data-driven model discovery to provide a more comprehensive modeling space [98, 99].

From a modeling perspective (see definitions in Table 3.1), it is assumed that approximate dynamics of the system are known, *i.e.*, the Platonic model:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) \tag{3.1}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ and the approximate governing dynamics $f(\cdot)$ are known and derived from first principles, asymptotic reductions, enforcing physical constraints or conservation laws, or positing empirical relations between variables. However, in truth, the true dynamics $F(\cdot)$, *which we do not have access to*, are given by

$$\dot{\mathbf{x}}_0(t) = F(\mathbf{x}_0(t)) = f(\mathbf{x}_0(t)) + g(\mathbf{x}_0(t)) \tag{3.2}$$

where $\mathbf{x}_0(t) \in \mathbb{R}^n$ and $g(\cdot)$ is the missing (deterministic) physics that remains unmodeled due to some suitable approximation and/or lack of physics knowledge. Note that the missing physics comprising $g(\cdot)$ may be intentionally omitted (*e.g.*, through model reduction) or unintentionally omitted (*e.g.*, due to lack of first-principles knowledge of the system structure). While not discussed in this manuscript, there may exist a noise process that drives stochastic variability in the model. If the stochastic process also remains unmodeled, the residual becomes the combination of the missing deterministic and stochastic physics. Therefore, those two terms would need to be disambiguated in order to learn a discrepancy model of the deterministic effect.

A time series of the system's full state is observed at discrete time points so that

$$\mathbf{y}_k = \mathbf{x}_0(t_k) + \mathcal{N}(\mu, \sigma) \tag{3.3}$$

where $\mathbf{y}_k \in \mathbb{R}^n$ and $\mathcal{N}(\mu, \sigma)$ is a noise process describing observation (or sensor) noise. The processes driving the observed dynamical systems are often continuous, and measurements can be collected at regular time intervals Δt via the observation process in Eq. 3.3, such that $\mathbf{y}_k = \mathbf{y}(k\Delta t)$.

In this manuscript, fairly basic assumptions are made concerning the noise. We will consider Gaussian additive noise in the sensor that is parameterized by a mean and variance. For sufficiently large noise, or small signal-to-noise ratio, there are few methods capable of producing viable discrepancy models. Of course, smoothing of the data can be attempted, but for large noise such signal filtering can also compromise any ability to identify a reasonable discrepancy model. Thus we consider noise levels that are low or moderate in order to apply the model discovery methods advocated. For large noise, Gaussian process regression is perhaps the best technique available. This is an important consideration, as this formulation relies on the ability to compute a good approximation of derivatives from the observed measurements.

The goal of discrepancy modeling is to resolve model-measurement mismatch due to missing physics. This can be done in two distinct ways; a similar distinction has been discussed in [57]. First, one can generate an improved dynamical model:

$$\dot{\tilde{\mathbf{x}}}(t) = \tilde{F}(\tilde{\mathbf{x}}(t)) = f(\tilde{\mathbf{x}}(t)) + \delta_f(\tilde{\mathbf{x}}(t)) \tag{3.4}$$

by learning a discrepancy model of the deterministic dynamical error, $\delta_f(\cdot)$, and appending it to the known approximate dynamics, such that $\|\mathbf{y}_k - \tilde{\mathbf{x}}(t_k)\| < \|\mathbf{y}_k - \mathbf{x}(t_k)\|$, *i.e.*, the augmented model's system estimations are less erroneous than the approximate model's. Because our goal is to recover the missing physics to resolve mismatch, priority should be given to learning a discrepancy model in the dynamical space. In a second method, given the systematic state-space residual, one can decrease the model-measurement mismatch by correcting the approximate state values [88]:

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(t) + \delta(\mathbf{x}(t)). \tag{3.5}$$

Note that in both methods, $\tilde{\mathbf{x}}(t)$ is the augmented and improved state space. In the former approach, a discrepancy model of the missing physics is learned in the dynamical space, while in the latter approach, a discrepancy model for the systematic residual is constructed in the state space. Equations (3.4) and (3.5) are explicitly the focus of this manuscript. Table 3.1 summarizes all our variable definitions whereas Fig. 3.1 illustrates the discrepancy modeling framework.

By focusing explicitly on the discrepancy between measured and modeled dynamics, our framework shifts the view of discrepancies as 'errors' or 'residuals' to highly valuable measures for model improvement. The field of data-driven engineering for dynamical systems has an opportunity to improve system characterization by disambiguating deterministic and random effects within the model-measurement mismatch. In this paper, we formalize the mathematical infrastructure of discrepancy modeling for dynamical systems, highlighting the interplay and balance between deterministic and random effects. Specifically, we consider four data-driven modeling methods for identifying discrepancy models and provide two principled approaches for evaluating the utility and suitability of discrepancy modeling in practical engineering applications. We leverage recent mathematical advancements in data-driven model discovery and evaluate the interplay between missing physics, $g(\mathbf{x}(t))$, and random processes, $\mathcal{N}(\mu, \sigma)$, showing how their relative sizes determine the ability to disambiguate deterministic from random or noisy effects. To encourage exploration and expansion of discrepancy modeling, we employ base coding packages for each model discovery method implementation [6–8, 49–53, 61, 62, 100–103]. We found that the performance characteristics of our suite of model discovery methods matched their documented performance for dynamical systems modeling. All the methods proposed (SINDy, DMD, GPR & NN) can be used successfully for both modeling the missing physics or the systematic residual. Thus, the evaluation of an appropriate method involves the intent of the user (*e.g.*, interpretability of the discrepancy model), constraints of a practical engineering applications (*e.g.*, if the Platonic model is accessible), and the computational efficiency and robustness of the method for a given set of measurements (*e.g.*, data quantity, quality, resolution).

3.2 Methods for Data-Driven Modeling

See Chapter 2's background section on methods for data-driven modeling for an overview of each method's general form. Here, we provide discrepancy-modeling specific conjecture for each method.

3.2.1 Gaussian Process Regression

GPR is the most general method advocated for discrepancy modeling. It simply gives the best fit Gaussian distribution, with estimates for the mean and variance with minimal hyper-parameter tuning in the optimization. However, its interpretability lies in ones choice of kernel function to describe characteristics of the discrepancy, such as periodicity or smoothness. The algorithm is generalizable provided the data is drawn from a stationary distribution, which may not be the case in practice. Instead it simply quantifies the statistics of the error. It is especially useful when large noise, or small signal-to-noise ratios, are present in the system. It is recommended to use additional techniques when using large datasets to avoid costly computation due to covariance matrix inversion.

3.2.2 Dynamic Mode Decomposition

We specifically use optimized DMD, but generally refer to it as 'DMD'. The DMD discrepancy architecture provides a potentially interpretable model for characterizing the dynamics, allowing for the decomposition of the data into modes and frequencies. DMD is quite robust, especially in its newest versions such as the *bagging optimized DMD* (BOP-DMD) [104] which leverages statistical bagging to help stabilize the linear model while providing uncertainty quantification. It scales well and has been shown to have some degree of generalizability, with minimal hyper-parameter tuning. However, it is so efficient to compute that one can easily compute new DMD models on the fly. BOP-DMD can always be computed, making it an attractive method as an alternative to GPR, handling even larger noise fluctuations with statistical bagging.

3.2.3 Sparse Identification of Nonlinear Dynamics

The SINDy algorithm has the strongest potential in providing an interpretable model for discovering missing physics. However, this method is not expected to perform well in fitting a discrepancy model to the state-space residual, as residuals might not be amenable to a sparse representation. SINDy also has the best possibility for generalization since the model learned is minimally parameterized. The algorithm is also computationally efficient and scalable, with minimal hyper-parameter tuning. Like BOP-DMD [104], ensemble SINDy [105] has been developed to help make SINDy robust even with increasing noise. Even with such methods for improving performance, SINDy is not as robust as GPR and DMD in handling noise. Thus it is recommended for low and intermediate noise regimes.

3.2.4 Neural Networks

Of the methods advocated, neural networks are the least interpretable, providing a black-box algorithm that is trained from data. The quality of the models trained are also sensitive to noise, making it more delicate than some of the other algorithms. However, as with many NN applications, with high-quality data of sufficiently large volume, the NN can provide high-quality discrepancy models in practice. It is the most computationally expensive of all the algorithms, with computational costs greatly exceeding other algorithms and requires significant hyper-parameter tuning in general. Techniques such as stochastic gradient descent greatly improve network optimization. Neural networks are recommended when large quantities of data of low and intermediate noise are available.

3.3 Discrepancy Modeling Framework

3.3.1 The Two Approaches

Discrepancy modeling aims to improve system characterization from data by disambiguating deterministic and random effects within the model-measurement mismatch. There are two nuanced, yet distinct, means of building a discrepancy model of missing physics by learning a model of the (i) deterministic dynamical error, or (ii) systematic state-space residual. In the first approach, the discrepancy is learned in the dynamical space, *i.e.*, the discrepancy model learns the error between the derivative of the measurements, $\dot{\mathbf{y}}_k$, and the approximate dynamics given the state measurements, $f(\mathbf{y}_k)$. Thus, learning a discrepancy model is akin to learning missing physics and is to be appended to the approximate dynamics. In the second approach, the discrepancy is learned in the state space, *i.e.*, the discrepancy model learns the systematic residual between the approximate state space, $\mathbf{x}(t_k)$, and the measurements, $\mathbf{y}_k$. Thus, this discrepancy model acts as a correction to the approximate state-space solution.

3.3.2 *Learning the Deterministic Dynamical Error*

In this approach, the discrepancy model learns the deterministic dynamical error within the model-measurement mismatch to improve the known, approximate dynamical model. The discrepancy is learned in the dynamical space to resolve the error between measurement derivatives and approximate dynamics, $\dot{\mathbf{x}}_D(t_k) = \dot{\mathbf{y}}_k - f(\mathbf{y}_k)$. The formulation begins with Eqns. (3.1-3.3). The discrepancy model $\delta_f(\cdot)$ estimates the dynamical relationship between measurements and deterministic dynamical error:

$$\dot{\mathbf{x}}_{\mathbf{D}}(t_k) \approx \delta_f(\mathbf{y}_k)$$

Using the suite of data-driven methods proposed, we can build discrepancy models $\delta_f(\cdot)$:

$$\begin{aligned}
\text{GPR}: &\quad \dot{\mathbf{x}}_D(t_k) \sim \mathcal{GP}(m(\mathbf{y}_k), \kappa(\mathbf{y}_k, \mathbf{y}_{k+1})) \\
\text{DMD}: &\quad \dot{\mathbf{x}}_D(t_k) \approx \mathbf{\Phi}\ \text{diag}(\mathbf{b}) e^{\boldsymbol{\omega} t_k}, \quad \mathbf{b} = \mathbf{\Phi}^\dagger \mathbf{y}_1 \\
\text{SINDy}: &\quad \dot{\mathbf{x}}_D(t_k) = \Theta(\mathbf{y}_k)\Xi \\
\text{NN}: &\quad \dot{\mathbf{x}}_D(t_k) = f_{\boldsymbol{\theta}}(\mathbf{y}_k)
\end{aligned}$$

and appended it to the approximate dynamics:

$$\dot{\tilde{\mathbf{x}}}(t) = \tilde{F}(\tilde{\mathbf{x}}(t)) = f(\tilde{\mathbf{x}}(t)) + \delta_f(\tilde{\mathbf{x}}(t))$$

to minimize:

$$\dot{\mathbf{y}}_k - \tilde{F}(\mathbf{y}_k) \tag{3.6}$$

3.3.3 *Modeling the Systematic State-Space Residual*

In this approach, the discrepancy model learns the time evolution of systematic state-space residual and corrects the approximate state-space solution. The discrepancy is learned in the state space to resolve the residual between measurements and the approximate solution, $\mathbf{x}_D(t_k) = \mathbf{y}_k - \mathbf{x}(t_k)$. The discrepancy model framework begins with Eqs. (3.1-3.3). The discrepancy model $\delta(\cdot)$ estimates the state-space relationship between the approximate state space and the systematic residual:

$$\mathbf{x}_\mathbf{D}(t_k) \approx \delta(\mathbf{x}(t_k))$$

Using the suite of data-driven methods proposed, we can build discrepancy models $\delta(\cdot)$:

$$\begin{aligned}
\text{GPR}: &\quad \mathbf{x}_D(t_k) \sim \mathcal{GP}(m(\mathbf{x}(t_k)), \kappa(\mathbf{x}(t_k), \mathbf{x}(t_{k+1}))\\
\text{DMD}: &\quad \mathbf{x}_D(t_k) \approx \mathbf{\Phi}\ \text{diag}(\mathbf{b})e^{\boldsymbol{\omega} t_k}, \quad \mathbf{b} = \mathbf{\Phi}^\dagger \mathbf{x}(t_1)\\
\text{SINDy}: &\quad \mathbf{x}_D(t_k) = \Theta(\mathbf{x}(t_k))\Xi\\
\text{NN}: &\quad \mathbf{x}_D(t_k) = f_{\boldsymbol{\theta}}(\mathbf{x}(t_k))
\end{aligned}$$

and correct the approximate state-space solution (initialized using the first observation $\mathbf{y}_1$):

$$\tilde{\mathbf{x}}(t) = \mathbf{x}(t) + \delta(\mathbf{x}(t))$$

to minimize:

$$\mathbf{y}_k - \tilde{\mathbf{x}}(t_k).$$

3.4 Discrepancy Modeling Applications

Discrepancy modeling is both system- and situational-dependent; thus we evaluate the utility and suitability of each discrepancy modeling approach using a suite of four model discovery methods by comparing reconstruction and forecasting accuracy. We further probe discrepancy modeling performance on increasingly complex systems and for increasing levels of noise.

3.4.1 Van der Pol Oscillator: A Simple Example

We began with a simple model and no noise. The data used in this example was generated using the Van der Pol oscillator:

$$\frac{d^2x}{dt^2} - \mu(1-x^2)\frac{dx}{dt} + x = 0 \tag{3.7}$$

using $x_0(0) = x(0) = [0.1, 5]$, $t = [0, 50]$, and $\Delta t = 0.01$. We simulated Eqn. (3.7) to generate our Platonic or approximate dynamics. To this system, we added a small nonlinear term ϵx^3:

$$\frac{d^2x_0}{dt^2} - \mu(1-x_0^2)\frac{dx_0}{dt} + x_0 + \epsilon x_0^3 = 0 \tag{3.8}$$

Eqn. (3.8) was simulated to generate the true system behavior using $\epsilon = 0.01$. This ϵ-small nonlinearity represents the missing deterministic physics not captured in the approximate model. This ϵ cubic term added to the approximate dynamics perturbed the time evolution of Van der Pol, as seen in Fig. (3.2), while still maintaining salient characteristics associated with the oscillator.

Deterministic Dynamical Error

We first evaluated the ability of discrepancy modeling to recover missing physics for the Van der Pol oscillator. As seen in Fig (3.3), the suite of model discovery methods learned the deterministic dynamical error within the mismatch such that no error remained. Both the reconstruction and the forecasting errors (black dashed line) between the true and augmented models is zero for all model discovery methods. The discrepancy dynamics between the true and approximate models are denoted by the blue line. The Van der Pol oscillator has a parameter dictating the nonlinearity of its oscillations; we chose a mild level of nonlinearity to start. As the nonlinearity parameter increases, we would expect that *linear* model discovery methods (*e.g.*, DMD) would struggle to build an accurate discrepancy model and thus fail to fully recover missing physics.

Systematic State-Space Residual

We next evaluated the ability of discrepancy modeling to learn the evolution of the systematic state-space residual for the Van der Pol oscillator. As seen in Fig. (3.4), the suite of model discovery methods had various success in modeling the time evolution of the systematic residual, thus demonstrating promising utility of discrepancy modeling for correcting state-space solutions. SINDy failed to learn the systematic residual, thus resulting in reconstruction and forecasting errors similar to if no discrepancy model was used. The inability for SINDy to model systematic residual is unsurprising, as it is formulated to recover dynamical terms (See Eqn. 2.11). DMD, GPR, and NN fared well in modeling the systematic residual, resulting in reduced error as compared to the approximate model; however, these methods struggled to correct forecasts of the approximate model. DMD was able to capture the salient features of the residual, yet the augmented solution appeared to be slightly time-shifted from the true solution, as seen by the brief zero-error within the reconstruction regime. Both GPR and NN learn promising discrepancy models of the residual, as they both had zero error between the true and augmented solutions for the first ten seconds in the reconstruction region, but began to diverge, albeit minimally. However, both GPR and NN had large errors at the beginning of forecasting. We theorize this is due to a failure to appropriately initialize. Generally, both methods learn their training trajectories with high accuracy; often, large datasets with many trajectories, and thus many initial conditions, were used in model training. In our case, because only one training trajectory was used, *i.e.*, one set of initial conditions, GPR and NN were unable to extrapolate the expected systematic residual when initialized with a different initial condition for forecasting. We theorize this initialization error can be resolved to greatly reduce remaining error when using a state-space residual discrepancy model via: (i) increasing data *quantity* (number of trajectories, initial conditions, parameters, resolution) for training or (ii) sacrificing the first portion of the test data/new times series to update the trained model.

Adding Noise

We re-ran both discrepancy modeling approaches with the Van der Pol oscillator and increasing levels of Gaussian noise added to the 'true' Van der Pol observations to evaluate the effect of sensor measurement *quality*. We use noise level of $\sigma = [0.1\%,\ 1\%,\ 10\%]$. As seen in Fig. 3.5, discrepancy modeling can be successful in numerous ways. Therefore, implementation amounts to user intent for discrepancy modeling, as well as the quantity and quality of sensor measurements. For example, when using discrepancy modeling to learn missing physics of the Van der Pol oscillator, all methods reconstruct and forecast true dynamics successfully in the no, low, and medium noise regimes. Note in the high noise regime, accuracy will most likely increase for methods like SINDy (which has an accuracy increase but is non-sparse) and DMD with innovations from baseline code packages (*e.g.*, ensemble, bagging, culling) [101, 104, 105]. If the user's intent for discrepancy modeling is interpretability, for no/low/medium noise, SINDy and DMD perform well, with minimal data quantity requirements and small computational costs (See Fig. 3.11 for computational cost comparisons). Even if the user has no interest in interpretability, SINDy and DMD perform comparably to GPR and NN (methods that have increased data quantity requirements and greater computational costs). On the other hand, if the user's intent is to only reduce state-space error for example, correcting the state-space solution with a discrepancy model of the systematic residual is feasible with DMD, GPR, and NN. While forecasting errors appear abysmal, utility of these methods for forecasting may increase with better initialization, as discussed above.

3.4.2 Lorenz Attractor: Adding Chaos

We extended our discrepancy modeling analyses to a more complex canonical system. We simulated the Lorenz attractor, which has increasingly complex dynamical features like chaos:

$$\frac{dx_1}{dt} = \sigma(x_2 - x_1) \qquad \frac{dx_2}{dt} = x_1(\rho - x_3) - x_2 \qquad \frac{dx_3}{dt} = x_1 x_2 - \beta x_3 \tag{3.9}$$

with the oft-used parameters, $x_0(0) = x(0) = [-8; 8; 27]$,$\sigma = 10$, $\rho = 28$, and $\beta = 8/3$, as well as $t = [0, 50]$ and $\Delta t = 0.01$, to generate our Platonic or approximate dynamics. To this system, we again added an ϵ-small nonlinearity:

$$\frac{dx_{0_1}}{dt} = \sigma(x_{0_2} - x_{0_1}) + \epsilon x_{0_1}^3 \qquad \frac{dx_{0_2}}{dt} = x_{0_1}(\rho - x_{0_3}) - x_{0_2} \qquad \frac{dx_{0_3}}{dt} = x_{0_1} x_{0_2} - \beta x_{0_3} \tag{3.10}$$

and simulated to generate the true system behavior using $\epsilon = 0.01$. This ϵ cubic term added to the approximate dynamics perturbed the time evolution of Lorenz, as seen in Fig. 3.6, while still exhibiting salient characteristics associated with the attractor and maintaining dynamical stability.

Deterministic Dynamical Error

We evaluated the ability of discrepancy modeling to recover missing physics for the Lorenz attractor. As seen in Fig. 3.7, the suite of model discovery methods had various success in learning the deterministic dynamical error for Lorenz. SINDy, GPR, and NN did well at learning the dynamical error, especially considering the sensitivity of chaotic systems to small dynamic deviations. These methods were able to reconstruct the first five seconds of the true Lorenz solution, as denoted by the zero-error in the reconstruction region (black dashed line). Interestingly, the ability to reconstruct the true Lorenz behavior increased dramatically with increased measurement resolution (Δt). In the forecasting region, SINDy had zero-error briefly but quickly became erroneous; this was most likely due to a small parameter deviation from the true dynamics that allowed the augmented dynamics to diverge from the true dynamics. Both GPR and NN were able to forecast the first five seconds of the true dynamics. In both the reconstruction and forecasting regions, the jump in non-zero error (blue line) corresponds with the attractor jump of the Lorenz dynamics. Unsurprisingly, DMD — a linear model discovery method — fails to learn the deterministic error for the Lorenz attractor example. Innovations from DMD's base code package have addressed some of the dynamical challenges due to nonlinearity and chaos, and continued innovation may allow for recovery of missing physics [52].

Systematic State-Space Error

We next evaluated the ability of discrepancy modeling to learn the evolution of the systematic state-space residual for the Lorenz attractor. As seen in Fig. 3.8, the suite of model discovery methods struggled to model the time evolution of the residual. Similar to Van der Pol results above, SINDy was unable to learn a discrepancy model the state-space residual for the Lorenz attractor; the augmented state-space solution showed no change in time series solution as compared to the approximate state-space solution. Similarly, DMD showed minimal impact on the approximate state-space solution when corrected using the learned state-space residual discrepancy model. GPR and NN exhibited reduced error in the augmented state-space solutions as compared to the approximate; however, it does not appear that a discrepancy model of the residual alone can recover the true Lorenz attractor solution. Using discrepancy modeling to learn the state-space residual may benefit from improved data quantity, as well as data assimilation techniques to combat known challenges with chaotic deterministic systems [106–108].

Adding Noise

We re-ran both discrepancy modeling approaches with the Lorenz attractor and increasing levels of Gaussian noise added to the 'true' Lorenz dynamics to evaluate the effect of sensor measurement *quality*. We use noise level of $\sigma = [0.1\%,\ 1\%,\ 10\%]$. As seen in Fig. 3.10, discrepancy modeling shows promise for overcoming model-measurement mismatch, even for highly nonlinear, chaotic systems such as the Lorenz attractor. Firstly, despite seeing an increase in accuracy as compared to the approximate model in learning a discrepancy model of the systematic state-space residual, the model discovery method were unable to reliably predict the attractor 'jump', which is a known challenge for data-driven modeling of chaotic dynamical systems [52]. Secondly, we noticed a relationship between the accuracy increase and the forecasting window, as seen in Fig. 3.9. While discrepancy modeling appears to have little impact on reducing the model-measurement mismatch, the percent change in error in Fig. 3.10 was computed over the entire forecasting window. However, we see that percent change in root mean squared error (RMSE) is a function of forecasting. Accuracy increase with discrepancy model augmentation begins near 100% and decreases as the forecasting window is extended, eventually settling near zero percent change from the approximate model. Note the drop in RMSE around five seconds of forecasting for NN, GPR, and arguably the SINDy discrepancy models corresponds with the first attractor 'jump' of the Lorenz.

3.4.3 Burgers' Equation: Spatio-temporal Nonlinearity

Finally, we evaluate discrepancy modeling with a canonical system in partial differential equations. We applied the findings from our previous examples from the Van der Pol oscillator and the Lorenz attractor to inform our implementation of discrepancy modeling for the Burgers' equation. In this 'scenario', our intent for discrepancy modeling is for rapid evaluation with interpretability. Additionally, we have good data quality (no noise) and low data quantity (one set of spatio-temporal snapshots). Therefore, we chose to use

DMD to learn a discrepancy model of the missing physics. We simulated Burgers' equation, which contains increasingly complex dynamical features, such as spatio-temporal nonlinearity:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} \tag{3.11}$$

using $\nu = 0.1$ $t = [0, 50]$ and $\Delta t = 0.01$, to generate our Platonic or approximate dynamics. To this system, we again added an ϵ-small nonlinearity:

$$\frac{\partial u_0}{\partial t} + u_0\frac{\partial u_0}{\partial x} + \epsilon u_0^3 = \nu\frac{\partial^2 u_0}{\partial x^2} \tag{3.12}$$

and simulate to generate the 'true' system behavior using $\epsilon = 0.01$. To create the data matrix $x \in \mathcal{R}^{1xn}$, Eqs. 3.11 and 3.12 were evaluated on a gridspace comprising $n = 256$ equally spaced spatial points and initiated using $u_0(0) = u(0) = e^{(x+2)^2}$.

As seen in Fig. 3.12, DMD was successful in learning a discrepancy model recovering the missing physics in Burgers' equation. We note that trends from our non-chaotic ordinary differential equation example (Van der Pol) hold for this partial differential equation example (Burgers).

3.4.4 Mass-Spring-Damper System

One critique of this framework is that a systematic residual discrepancy model, as is formulated in Equation 3.5, is by construction a poor strategy for coping with imperfect dynamical models. As is demonstrated in the three previous continuous dynamical system examples, an imperfect model is best improved by recovering the missing physics instead of learning a corrective model of the systematic residual. However, learning a discrepancy model of the state-space error is by construction a suitable approach for resolving an imperfect discrete dynamical system or for addressing observation errors like measurement bias. For example, a discrete dynamical system model for controls engineering could benefit from a discrepancy model updating the state space:

$$x_{k+1} = f(x_k) + \delta(x_k) \tag{3.13}$$

In order to demonstrate where the systematic residual discrepancy modeling approach would succeed, we simulated the linear mass-spring-damper system:

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = F(t), \tag{3.14}$$

with an observation error in the form of measurement bias:

$$y_k = Cx(t_k) + \mathcal{N}(\mu, \sigma), \tag{3.15}$$

where the measurement bias, $\epsilon = 0.01$, is applied through the observation matrix C =[1+ϵ; 0]. We use the parameters $m = 2$, $b = 0.5$, $k = 2$, t =[0, 20], $\Delta t = 0.01$, $x_0 = x(0) =$ [0.5, 0], and $F(t) = x^3$.

As seen in Fig. 3.13, regardless which data-driven modeling method was used, the state-space measurement bias was accurately learned. Even with noisy and biased measurements, the state-space error was corrected. However, the numerical simulations shown for the linear mass-spring-damper – while simplistic – presents a neutral result: state-space error can be just as well addressed by a data assimilation framework, a standard technique that is well-founded and mature in its formulation. For example, the Kalman filter and its variants have long achieved state-of-the-art performance [109, 110] and perform well when a model of the underlying dynamics is readily available.

3.5 Conclusions and Guidelines for Discrepancy Modeling (Framework)

In conclusion, discrepancy modeling for learning missing physics, modeling systematic residuals, and disambiguating between deterministic and random effects emerged as an important framework for principled investigation of model-measurement mismatch. The methods advocated for provide a suite of algorithms that can be broadly used in almost any realistic system where data and models are jointly used, such as emerging digital twin technologies [111, 112]. By leveraging improved observations to automate the process of learning better models, we can improve the characterization of underlying system dynamics. The two discrepancy modeling approaches introduced are distinct, yet nuanced: (i) learning missing physics in the dynamical space to improve the underlying model and (ii) estimating the systematic residual in the state space to correct the model's state approximations. While an abundance of data-driven modeling methods exist, we adapted a suite of four model discovery methods to demonstrate the mathematical implications of discrepancy modeling for practical applications. Each combination of approach and method for discrepancy modeling demonstrates the need for clearly defined goals and an understanding of constraints, such as those imposed by data collection or model accessibility. Further, by taking a deeper look at the mathematical implementations of each data-driven modeling method, we can better anticipate the utility, suitability, and interpretability of learned discrepancy models. Indeed, certain model discovery methods are more appropriate than others in conjunction with the chosen discrepancy modeling approach based on the intended use of the learned model, as well as sensor capabilities (data quality, quantity, resolution). To summarize, in the presence of model-measurement mismatch in dynamical systems, one should try to discover missing physics as a top priority in building discrepancy models. If this is not possible, the results on modeling the systematic residual point towards using a data assimilation framework; systematic residual discrepancy modeling (except in trivial cases) can not compete in practice with the extensive and mature theoretical framework of Kalman filtering and data assimilation. Especially when a model of the underlying dynamics is known (which is assumed in this work's state-space error scenario), Kalman filtering achieves state-of-the-art performance in.

Because this manuscript is one of the first to comprehensively investigate how to handle the discrepancy of missing physics (*i.e.*, deterministic dynamical error) in practical applications, we made a couple assumptions to isolate only the effect of missing physics on discrepancy modeling. Firstly, we specifically studied a fully observed dynamical system. If only partial observations are available, additional techniques can be implemented. Autoencoders are a standard technique in machine/deep learning for learning a full state space from limited observations. For example, shallow neural networks have been used to reconstruct high-dimensional states from limited measurements [113–115]. Additionally, time delay embedding is a widely-used technique, originally established by Taken's theorem, to characterize a latent dimension from incomplete measurements [116, 117]. Second, in this manuscript we specifically address an erroneous model, not erroneous initial conditions. These are two important but distinct challenges in modeling dynamical systems. Even with the perfect initial condition, if the underlying model is incorrect (e.g., idealized or missing physics), the time series behavior will be wrong from observations. Concurrently, an imperfect initial condition with a perfect model will also alter the estimated system behavior. We specifically investigate discrepancy modeling of missing physics, to de-conflate effects of model error and initial conditions. It is important we isolate the effect of missing physics for discrepancy modeling because incorrect parameter estimation can affect reconstruction and forecasting accuracy of a dynamical system. Indeed, an incorrect parameter – even with the correct dynamical structure – can lead to inaccurate system behavior [86].

Demonstrated in this manuscript is discrepancy modeling for a number of canonical spatial and/or temporal systems. For each example system, we evaluate reconstruction and/or forecasting accuracy. In particular, we vary the signal-to-noise ratio to demonstrate the impact of random fluctuations on the ability to disambiguate between deterministic and random effects. An important implication of this work is that there is no 'silver bullet' to automatically resolve model-measurement mismatch. In fact, in certain cases when opposing priorities cannot be reconciled, not using discrepancy modeling may be most appropriate. Further, this work demonstrates the limitations of discrepancy modeling as a function of observation noise; deterministic effects may exist within the model-measurement mismatch, but if random effects dominate, learning a discrepancy model of missing physics will be impossible and require improved sensor technology.

We summarize the the effects of the various discrepancy modeling paradigms:

(i) **SINDy:** While SINDy requires more training data and a high signal-to-noise ratio, it results in the recovery of parsimonious dynamics and thus is agnostic to nonlinearity strength in the discrepancy dynamics. SINDy has minimal computational costs and provides immense opportunity to improve engineering design through its interpretable approach to learning the governing dynamics. Additionally, SINDy provides the architecture to evaluate discrepancy model capability and parameter dependence of the learned discrepancy dynamics.

(ii) **Dynamic Mode Decomposition:** DMD's strength is in its interpretable approach to rapid model evaluation. Further, DMD is a good discrepancy modeling method when the signal-to-noise is too low for SINDy and is not dominated by random fluctuations. As DMD is a linear data-driven modeling approach, a DMD discrepancy model may be sensitive to the nonlinearity strength of discrepancy dynamics.

(iii) **Gaussian process regression:** GPR is a powerful non-parametric Bayesian approach to discrepancy modeling in no/low/medium noise regime and lower data quantity requirements. It performed as well as — or, in certain cases, better than — a NN, along with lower associated computational costs. GPR is a universal approximator with a closed form solution. While SINDy and DMD provide convenient solutions, they are not universal function approximators; conversely, NNs are approximators, but do not have closed form solutions. The challenges of slowness and scalability with GPR can be addressed by using low rank or Random Feature approximations of Gaussian processes [118].

(iv) **Neural networks:** NNs — a parametric modeling approach — shine when data in capturing complicated functions and generalizing to non-local behavior. If one can afford it, data assimilation for correcting systematic state space residual, for example with a GPR or NN, will be highly effective. For example, in chaotic systems such as the Lorenz, data assimilation will provide a consistent state-space correction versus modeling systematic error without feedback [119].

Remark: With increased data resolution (*e.g.*, a high capacity sensor), a discrepancy model learning the deterministic dynamical error can almost perfectly reconstruct and forecast over short-time both the Van der Pol and Lorenz dynamical systems highlighted in this manuscript (*e.g.*, using a $\Delta t = 0.001$ versus the 0.01 used for the results presented in this manuscript). This is particularly important for learning a dynamical discrepancy model using the general form of SINDy whose coefficient estimations of a sparse set of dynamical terms are sensitive to data resolution, quality, and quantity. Additional techniques such as statistical bagging and ensembling methods [105] can greatly improve SINDy's robustness with lower resolution; such statistical techniques have successfully improved the general form of DMD as well [104]. Indeed, each architecture (i)-(iv) used can be enhanced and improved [6–8, 49–53, 61, 62, 100–103]; this comparative study does not claim to optimize the ability of each technique to perform at its absolute best. Indeed, hyper-parameter tuning of each technique is problem specific and again depends upon the characteristics of measurement data and the intent of its use. Rather, our goal is to demonstrate the various possibilities and their appropriate uses. It is evident from the study that each method considered has strengths and weaknesses that are appropriate to consider depending upon data and intent.

As data-driven modeling continues to gain momentum, it is imperative that researchers utilize domain knowledge (*e.g.*, first principles physics) to model complex systems. However, in all disciplines, model-measurement mismatch exists. The lack of investigation into this error leaves the missed opportunity to resolve model-measurement mismatch, disambiguate deterministic effects, and improve the underlying model. Importantly, there may exist a spectrum of other types of discrepancies, such as observation error, partial measurements, initial condition errors, stochastic processes, and non-Gaussian distributions — among others we may not be aware of yet. Discrepancy modeling for dynamical systems is in its infancy. While our comparative work is early and extensive, it is not exhaustive. Many open and exciting challenges exist for data-driven engineering of dynamical systems in practical engineering applications.

Acknowledgements

We are especially grateful to Kadierdan Kaheman and Steven Brunton for discussions related to discrepancy modeling. MRE acknowledges support from the NSF under award GRFP DGE-1762114. JNK acknowledges funding from the National Science Foundation AI Institute in Dynamic Systems grant number 2112085.

Code

Code can be found at *github.com/meganebers/Discrepancy-Modeling-Framework-code*.

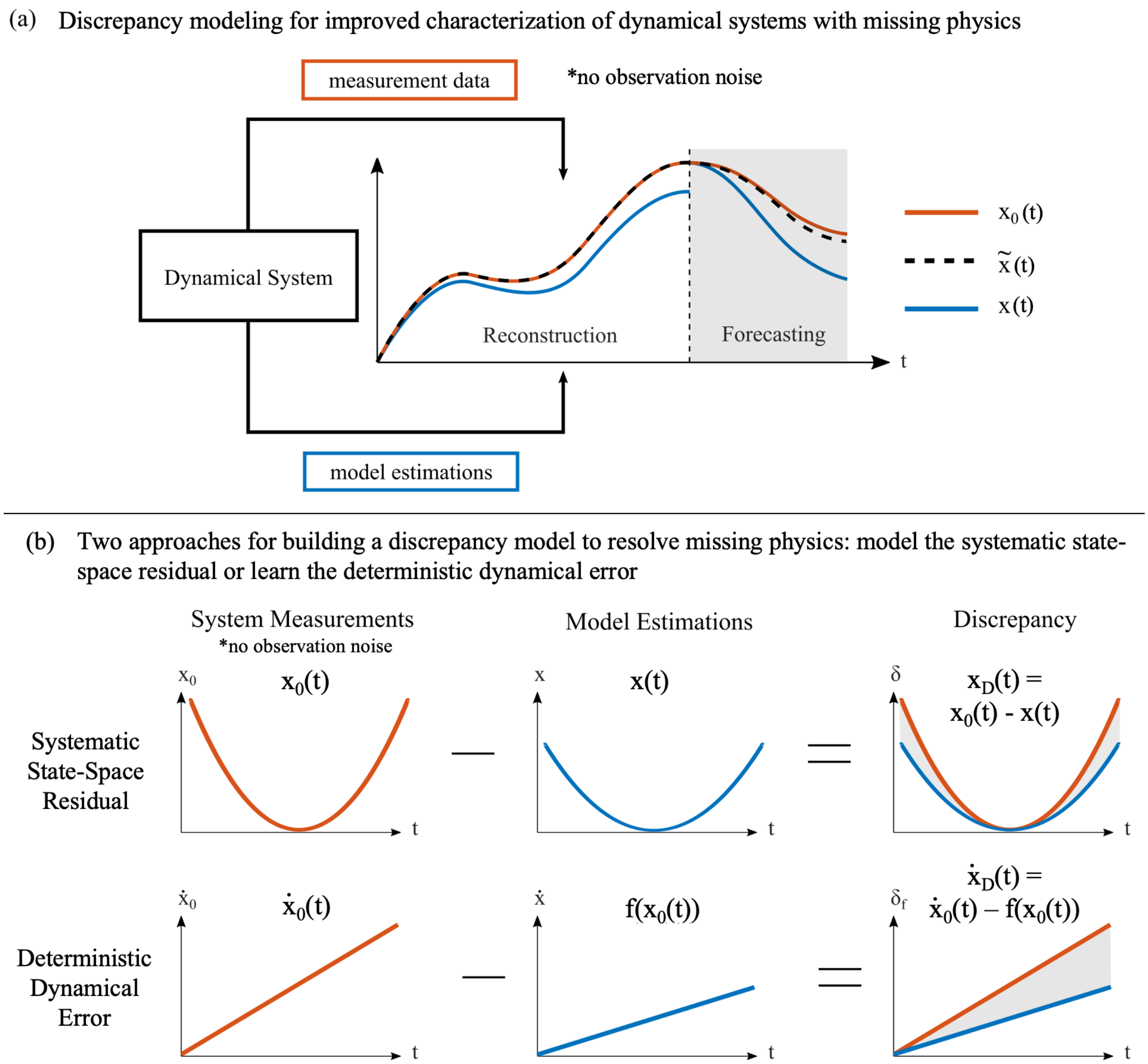

*Figure 3.1: **Top panel**: An approximate dynamical model $f(\cdot)$ provides estimates of system behavior used for both reconstruction and forecasting (shaded region), $\mathbf{x}(t)$. However, true behavior $\mathbf{x_0}(t)$ (without observation noise) deviates from these estimates. The goal of discrepancy modeling is to learn a discrepancy model that recovers the missing physics and augments the approximate dynamics to improve system characterization, $\tilde{\mathbf{x}}(t)$. **Bottom panel**: There are two approaches for building a discrepancy model to estimate missing physics: (i) modeling systematic state-space residual between the approximate state space, $\mathbf{x}(t)$, and true state space, $\mathbf{x}_0(t)$, and (ii) learning the deterministic dynamical error between the true dynamics, $\dot{\mathbf{x}}_0(t) = f(\mathbf{x}_0(t)) + g(\mathbf{x}_0(t))$, and the approximate dynamics, $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$. In real-world systems, the true system behavior is noisily observed, $\mathbf{y}_k = \mathbf{x}_0(t_k) + \mathcal{N}(\mu, \sigma)$, model-measurement mismatch contains both deterministic and random effects; measurements $\mathbf{y}_k = \mathbf{y}(k\Delta t)$ denote a continuous dynamical system's full state noisily observed at discrete time points.*

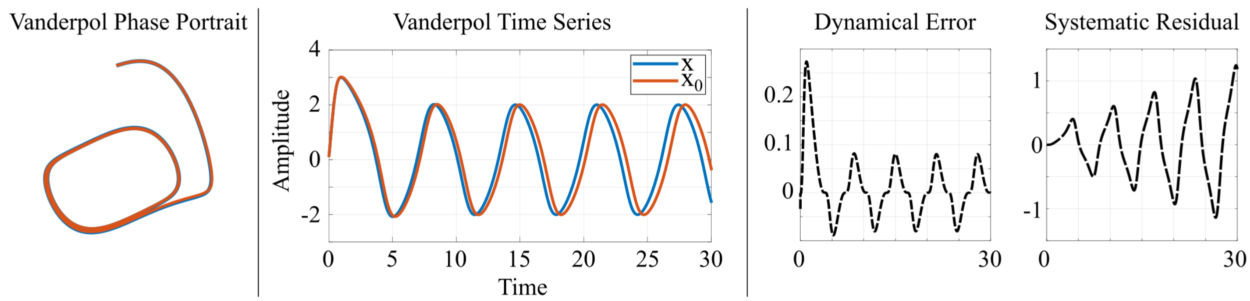

Figure 3.2: Van der Pol oscillator example (no noise) with and without a discrepancy. While the salient dynamical features are preserved as seen in the phase portrait (left panel), the time evolution (middle panel) diverges quickly with only an ϵ-small dynamical difference. The dynamical error and systematic residual (right panel) are plotted to demonstrate the two discrepancy types.

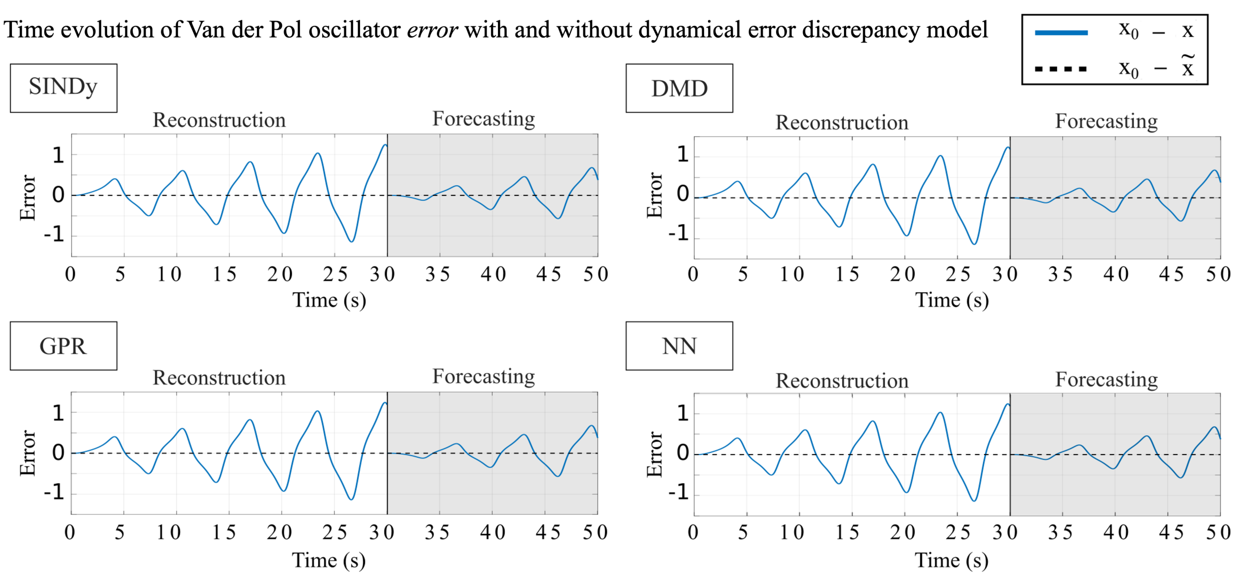

Figure 3.3: Remaining state-space error with and without a discrepancy model of the deterministic dynamical error appended to the approximate Van der Pol oscillator model (no noise). The blue line shows the error without a discrepancy model, and the black dashed line shows the error with a discrepancy model recovering the missing physics. The suite of model discovery methods learned the missing physics within the discrepancy such that no error remained.

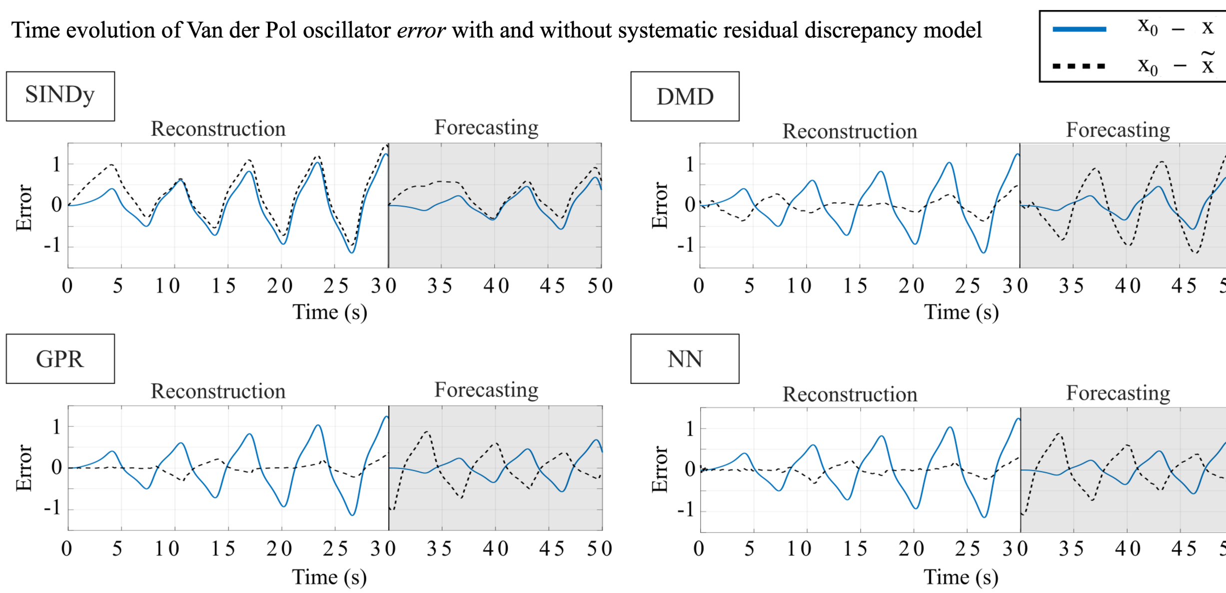

Figure 3.4: Remaining state-space error with and without a discrepancy model of the systematic state-space residual to correct the approximate Van der Pol oscillator state-space solution (no noise). The blue line shows the error without a discrepancy model, and the black dashed line shows the error with a discrepancy model modeling the systematic residual. The suite of model discovery methods had various success in modeling the systematic residual, resulting in promising utility of discrepancy modeling for correcting state-space solutions.

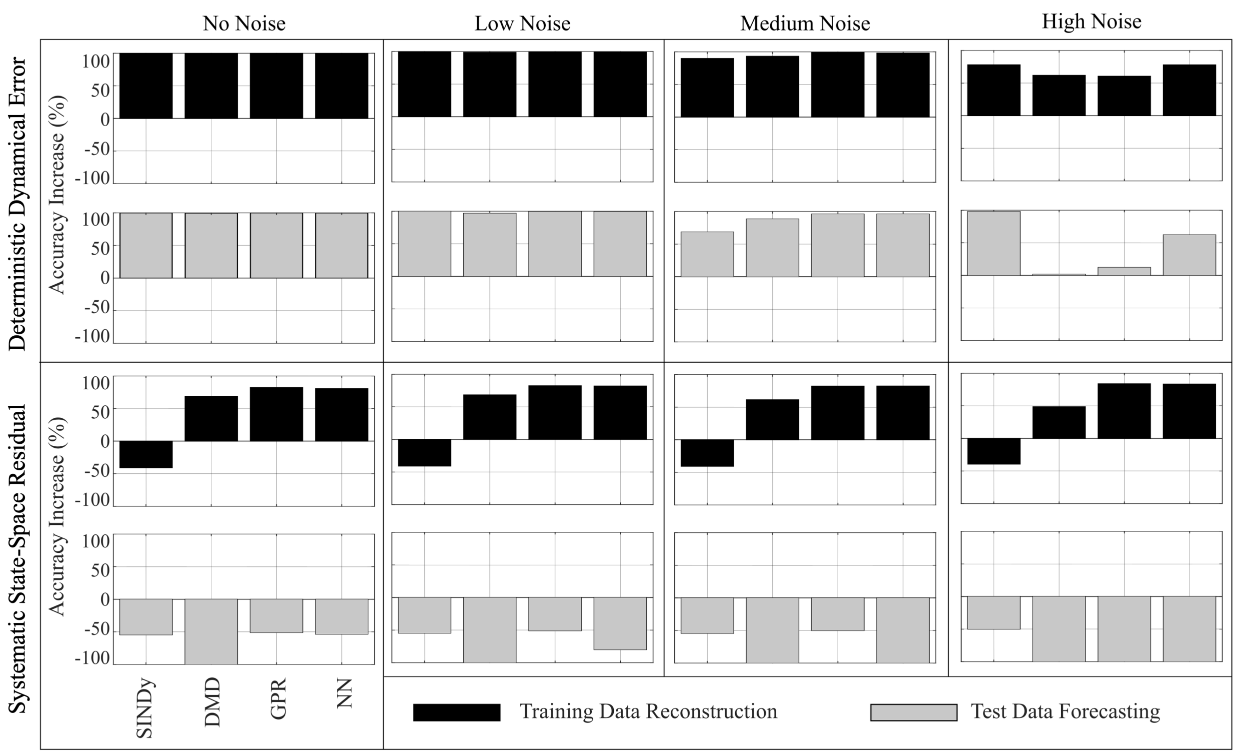

Figure 3.5: Increase in accuracy with discrepancy modeling augmentation in Van der Pol oscillator. Increase in % accuracy is calculated as the relative change between the root mean squared errors (RMSE) of the augmented and approximate solutions as compared to the true solution. Results are shown for no (0%), low (0.1%), medium (1%), and high (10%) noise levels. Discrepancy modeling can be successful in numerous ways and depends on user intent (e.g., interpretability), data quality (e.g., noise tolerance), and data quantity (e.g., rapid evaluation vs. computational cost).

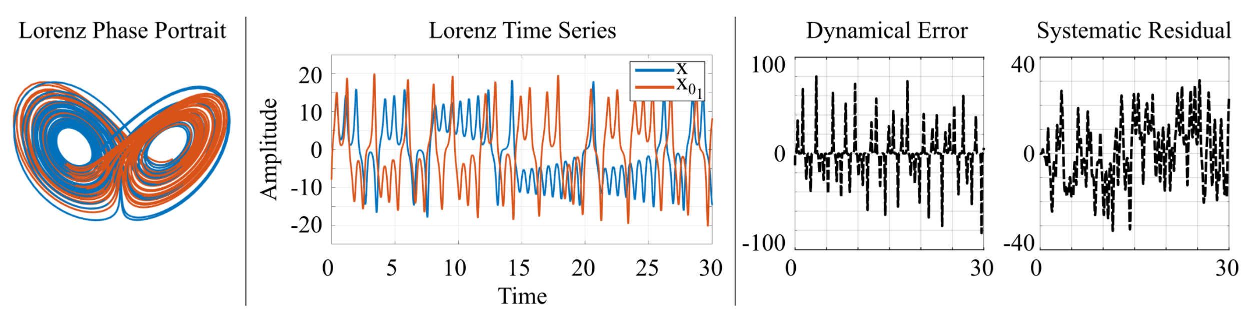

Figure 3.6: Lorenz attractor example (no noise) with and without a discrepancy. While the salient dynamical features are preserved as seen in the phase portrait (left panel), the time evolution (middle panel) bifurcates quickly with only an ϵ-small dynamical difference. Chaotic dynamical systems like the Lorenz attractor are particularly sensitive to small errors in system dynamics. The deterministic dynamical error and systematic state-space residual (right panel) are plotted to demonstrate the two types of discrepancies.

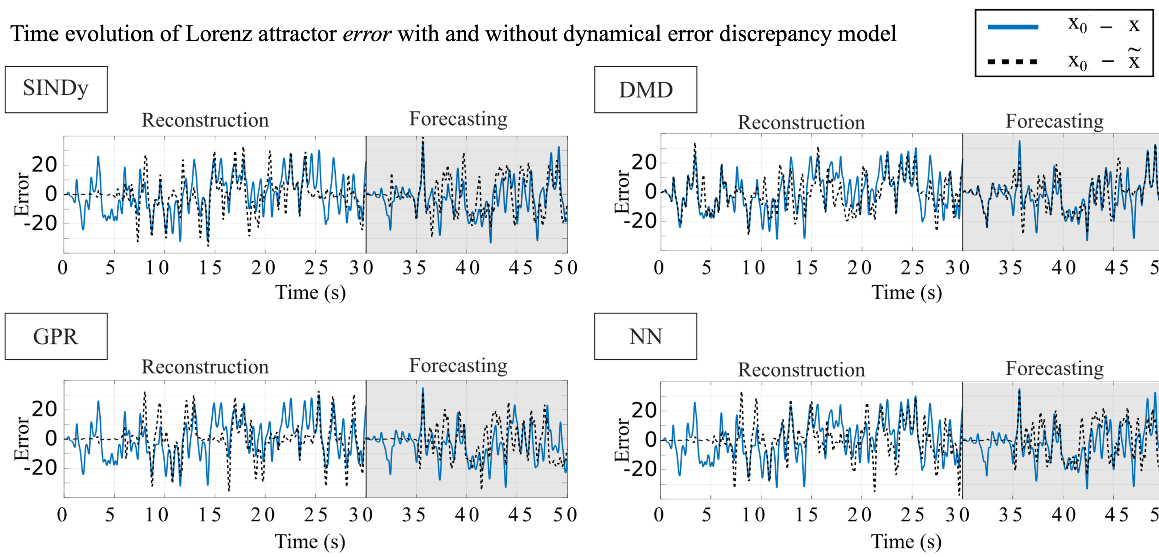

Figure 3.7: Remaining state-space error with and without a discrepancy model of the deterministic dynamical error appended to the approximate Lorenz attractor dynamics (no noise). The blue line shows the error without a discrepancy model, and the black dashed line shows the error with a discrepancy model recovering the missing physics.

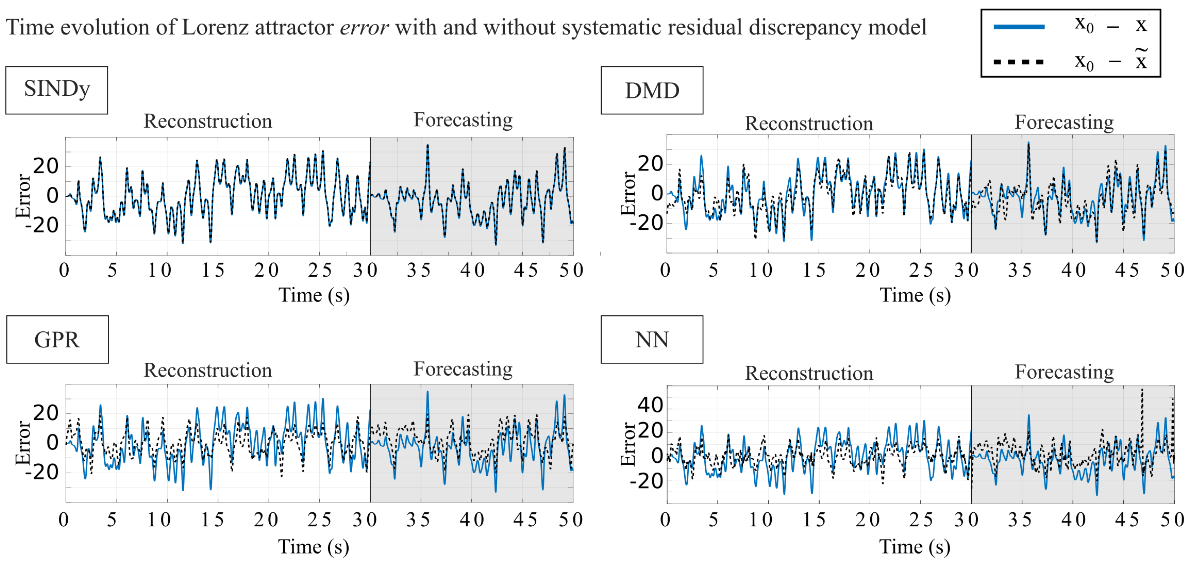

Figure 3.8: Remaining state-space error with and without a discrepancy model of the state-space systematic residual to correct the approximate Lorenz attractor state-space solution (no noise). The blue line shows the error without a discrepancy model, and the black dashed line shows the error with a discrepancy model of the systematic residual. All model discovery methods struggled to learn the time evolution of this residual.

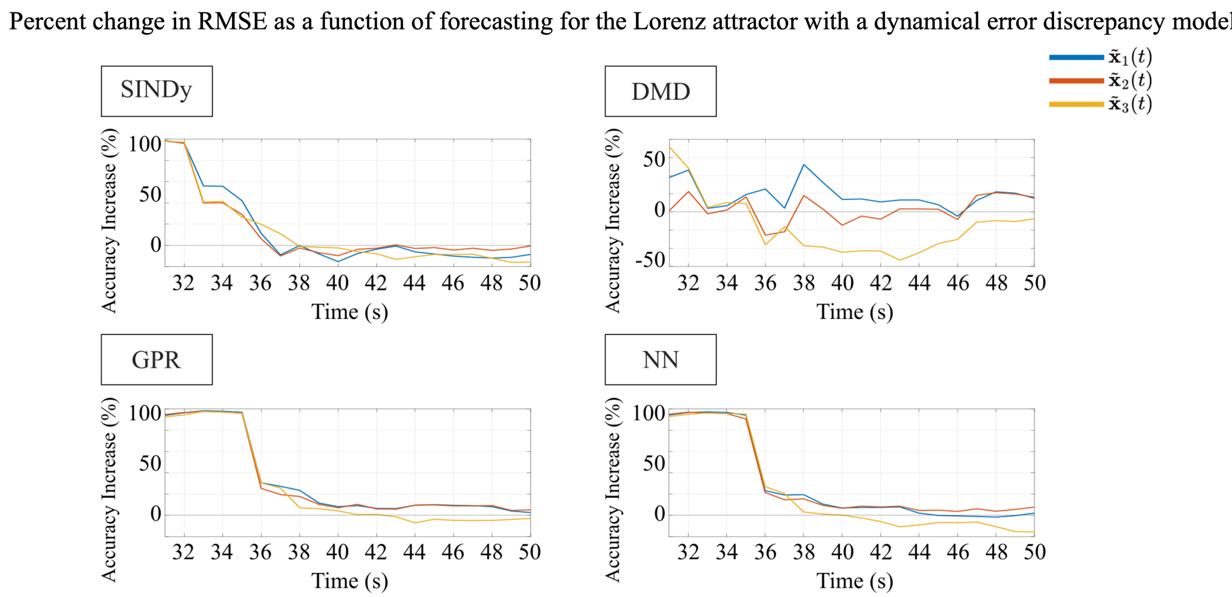

Figure 3.9: Percent change in RMSE as a function of forecasting. The discrepancy model for learning deterministic dynamical error in the Lorenz attractor example (no noise) demonstrates the relationship between the accuracy increase and the forecasting window; here, we see a decrease in accuracy with discrepancy model augmentation as the forecasting window is extended and settle near zero percent change from the approximate model. Note: the drop in forecasting RMSE for NN, GPR and arguably the SINDy discrepancy models corresponds with the first attractor 'jump' of the Lorenz. Predicting Lorenz bifurcations continues to be a challenging task in the field of dynamical systems.

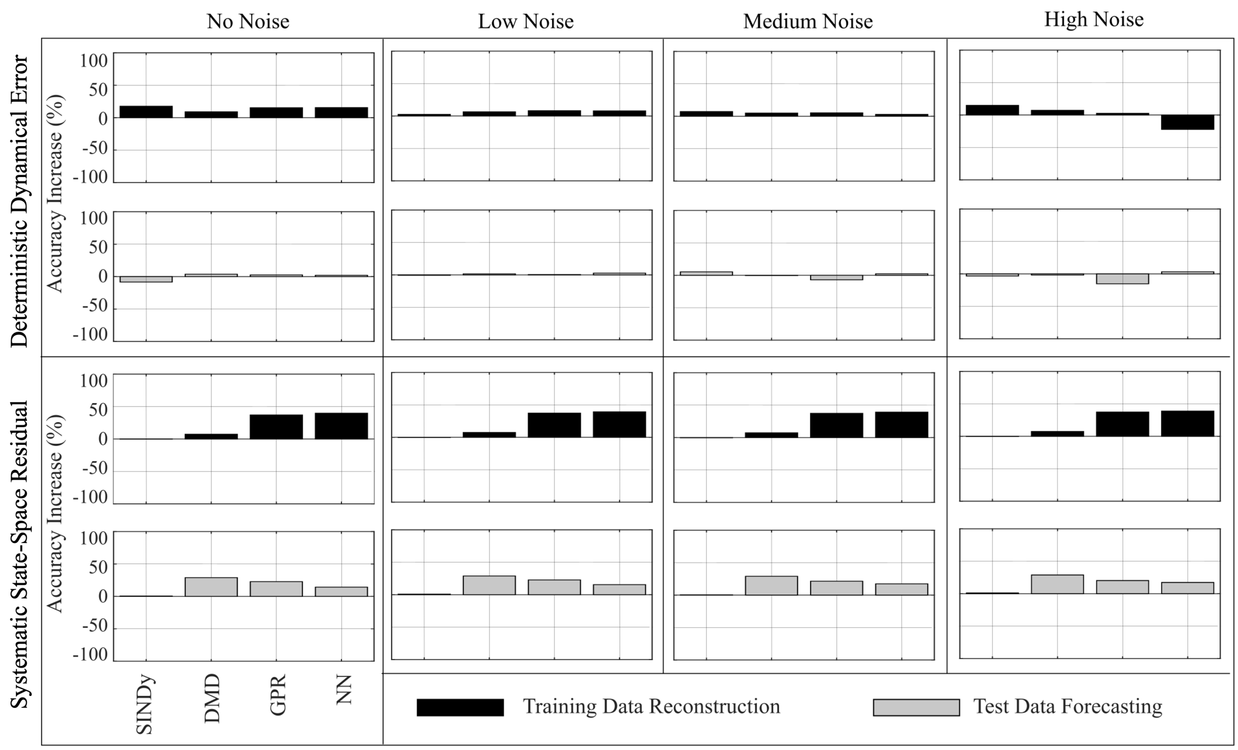

Figure 3.10: Increase in accuracy with discrepancy modeling augmentation in Lorenz attractor. Increase in % accuracy is calculated as the root mean squared error (RMSE) between true and augmented state space solutions for reconstruction of the training region and forecasting in the test region. Results are shown for no (0%), low (0.1%), medium (1%), and high (10%) levels of noise. Discrepancy modeling shows promise for overcoming model-measurement mismatch, even for nonlinear and chaotic systems such as the Lorenz attractor.

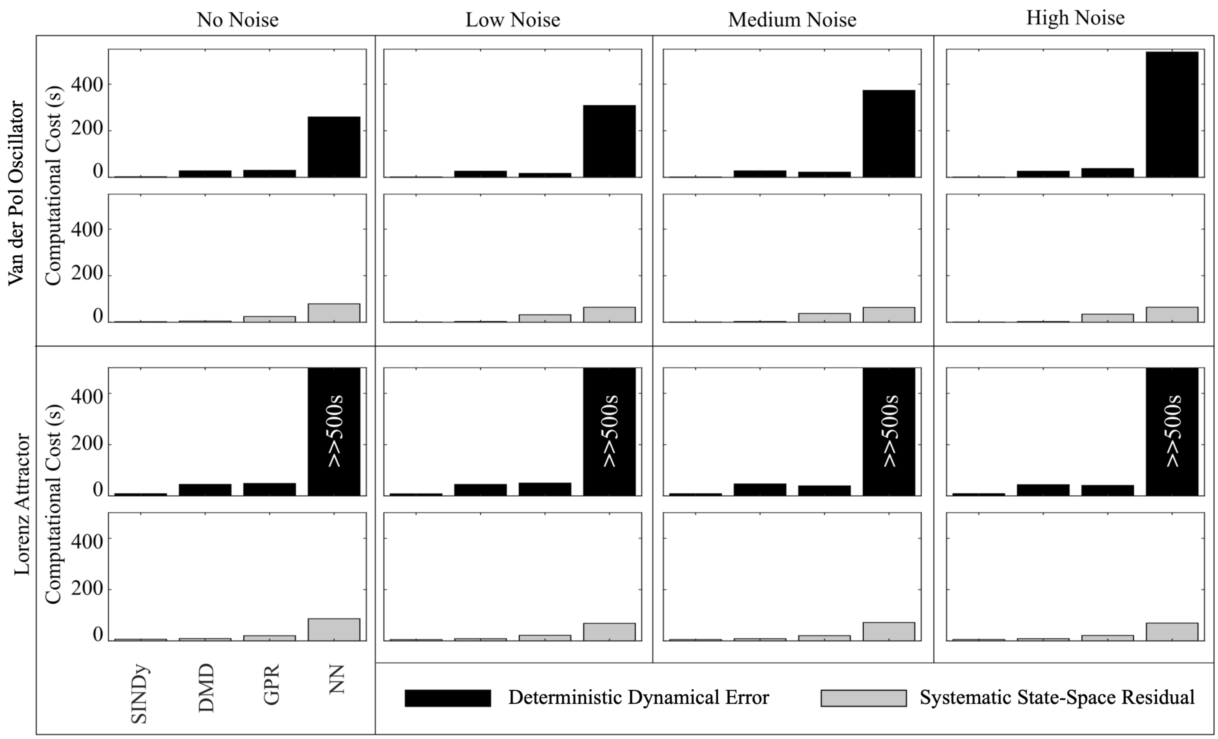

Figure 3.11: Computational costs (seconds) corresponding to executing the discrepancy modeling framework in MATLAB. Comparison across both approaches for each of the suite of model discovery methods. Costs include data generation for approximate and true models, computation of discrepancy dynamics, and reconstruction and forecasting of augmented model. The computational cost for the Van der Pol oscillator increases as noise increases for all model discover methods. SINDy has the lowest computational cost, followed by DMD, GPR, and NN, for both discrepancy modeling approaches. Learning missing physics with a NN has a noteably higher computational cost. This occurs because of how the discrepancy model is appended to the approximate dynamical model; the discrepancy dynamics are computed at each time step in MATLAB's ODE45 function, which greatly increases computational cost. These trends hold for the Lorenz attractor. Note: the time to execute the 'learning deterministic dynamical error' script using the NN was much higher than 500 seconds, and peaked around 2100 seconds.

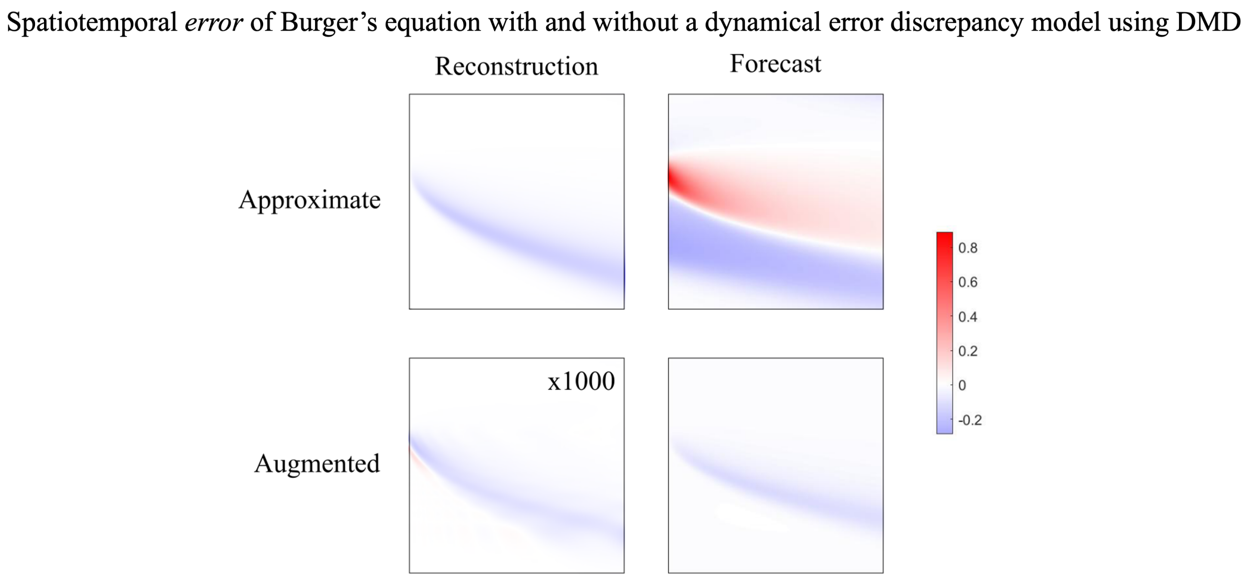

Figure 3.12: Remaining error with and without a discrepancy model of the deterministic dynamical error appended to the approximate Burgers' dynamics (no noise). No color (white) represents zero error as compared to the true system. Both red and blue denote non-zero error; the different colors only distinguish positive and negative error, respectively. Note that the augmented reconstruction error is multiplied by 1000. The dynamical error discrepancy model using DMD reconstructed true spatiotemporal dynamics with virtually zero remaining error and greatly diminished the error during forecasting. These results follow the same trends as seen in the Van der Pol example.

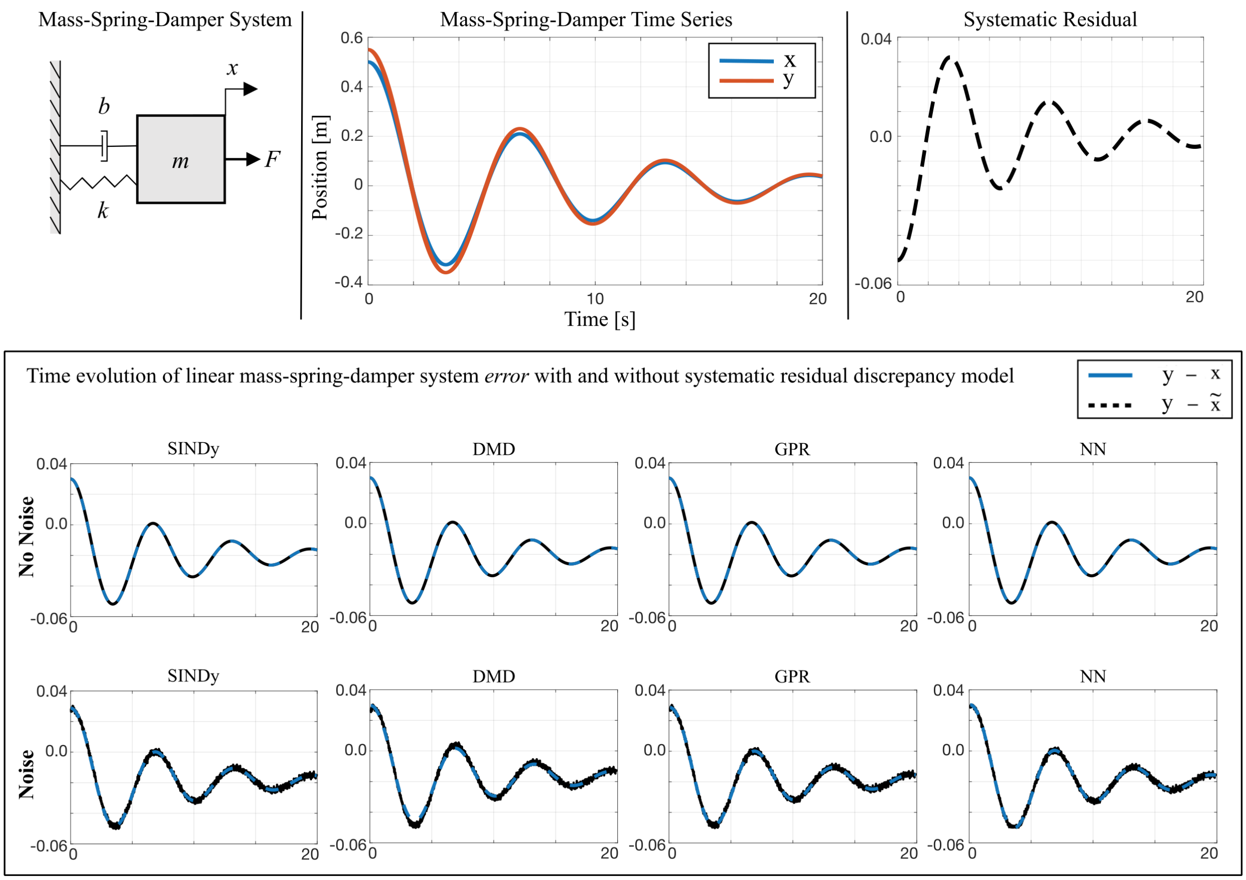

Figure 3.13: (Top) A linear mass-spring-damper system with measurement bias, leading to a systematic state-space error. (Bottom) State-space error like measurement bias, as in this mass-spring-damper system, is well resolved using systematic residual discrepancy modeling, regardless of data-driven modeling method or noise. Importantly, this presents a neutral result: data assimilation like Kalman filtering can just as well resolve state-space observation error when the underlying model is known.

Chapter 4

A MACHINE LEARNING APPROACH TO QUANTIFY INDIVIDUAL GAIT RESPONSES TO ANKLE EXOSKELETONS

We currently lack a theoretical framework capable of characterizing heterogeneous responses to exoskeleton interventions. Predicting an individual's response to an exoskeleton and understanding what data are needed to characterize responses has been a persistent challenge. In this study, we leverage a neural network-based discrepancy modeling framework to quantify complex changes in gait in response to passive ankle exoskeletons in nondisabled adults. Discrepancy modeling aims to resolve dynamical inconsistencies between model predictions and real-world measurements. Neural networks identified models of (i) *Nominal* gait, (ii) *Exoskeleton (Exo)* gait, and (iii) the *Discrepancy* (*i.e.*, response) between them. If an *Augmented* (Nominal+Discrepancy) model captured exoskeleton responses, its predictions should account for comparable amounts of variance in *Exo* gait data as the *Exo* model. Discrepancy modeling successfully quantified individuals' exoskeleton responses without requiring knowledge about physiological structure or motor control: a model of *Nominal* gait augmented with a *Discrepancy* model of response accounted for significantly more variance in *Exo* gait (median R^2 for kinematics $(0.928 - 0.963)$ and electromyography $(0.665 - 0.788)$, ($p < 0.042$)) than the *Nominal* model (median R^2 for kinematics $(0.863 - 0.939)$ and electromyography $(0.516 - 0.664)$). However, additional measurement modalities and/or improved resolution are needed to characterize *Exo* gait, as the discrepancy may not comprehensively capture response due to unexplained variance in *Exo* gait (median R^2 for kinematics $(0.954 - 0.977)$ and electromyography $(0.724 - 0.815)$). These techniques can be used to accelerate the discovery of individual-specific mechanisms driving exoskeleton responses, thus enabling personalized rehabilitation.

4.1 Introduction

Ankle exoskeletons and other assistive devices augment locomotion in nondisabled adults and improve kinematics, provide musculoskeletal alignment, and reduce the energetic demands of walking for individuals with neurologic injuries [120–128]. However, it is difficult to predict how exoskeleton forces and torques acting on the body alter an individual's gait biomechanics, motor control, and sensory feedback [129–131]. Specifically, an incomplete understanding of the factors driving heterogeneous responses to exoskeletons makes identifying exoskeleton design parameters that optimize an individual's gait challenging [123, 125, 132–134]. The most successful methods for personalizing exoskeletons are experimental techniques – such as human-in-the-loop optimization – that are resource intensive and iterative, placing additional burden on participants [122, 126, 135]. Conversely, computational approaches require shorter experimental sessions, but are currently unsuccessful due to the poor modeling of individual differences in neuromuscular physiology and control [136–140]. Despite diverse approaches to characterizing exoskeleton impacts on gait, predicting responses to exoskeletons (*i.e.*, complex changes in gait with changes in exoskeleton properties) is an open problem [1, 125–127, 130, 132, 133, 136, 141–153]. Emerging machine learning methods may help overcome experimental and theoretical barriers to quantifying individual-specific responses to exoskeletons.

While studies have investigated potential physiological mechanisms underlying exoskeleton responses, the complexity of the neuromusculoskeletal system limits our ability to uncover individual-specific physiological mechanisms governing exoskeleton gait [130, 146, 147, 154]. For example, muscle-tendon mechanics are known to explain unexpected effects of exoskeleton assistance on gait energetics [130, 147]. This mechanism alone, however, is unlikely to explain responses across individuals, which may be influenced by other factors, such as sensation or motor control [146, 155]. Other researchers have used physics-based or data-driven models to capture the physiological processes driving exoskeleton responses without explicit physiological mechanisms [1, 132, 156]. For example, a Random Forest algorithm with kinematic and clinical exam measurements from over 300 children with cerebral palsy was used to predict changes in kinematics with passive ankle exoskeletons [156]. The algorithm, however, only explained 19-28% of the variance in kinematic responses. Quantifying individual-specific processes – complex interactions between biomechanical, neural, and sensory mechanisms – that can explain heterogeneous exoskeleton response could inform personalized device designs or improve accuracy of gait simulations with exoskeletons.

An unexplored approach to accelerate discovery of individual-specific processes driving complex gait changes with exoskeletons is characterizing the response itself. Gait kinematics, kinetics, and muscle activity change with exoskeletons, and our inability to explain this response represents a *discrepancy* in our understanding of the neuromusculoskeletal system and how it responds to an exoskeleton. This discrepancy may reflect many individual factors, such as inter-individual differences in musculoskeletal physiology or motor control, an incomplete understanding of gait adaptation with exoskeletons, or an inability to model these complex processes. It is precisely because *we do not know* what processes (or combinations thereof) underlie responses that, in this paper, we choose to summarize an all-encompassing discrepancy to describe exoskeleton responses. Additionally, it is unknown what data are needed to capture discrepancies. In previous studies, some researchers used only a few strides to tune models [146, 152, 153], while others used large datasets [1]. In tuning models, a variety of data were used: kinematics [157], electromyography (EMG) [152, 158], and clinical exams [146, 156], among others. Quantifying this discrepancy is, therefore, a critical step in understanding individual responses to ankle exoskeletons.

Discrepancy modeling is a novel tool in machine learning developed to identify missing physics in complex systems described by differential equations [159], such as those describing human movement [160–164]. A variety of machine learning techniques can be used to quantify discrepancies [159]. Deep learning is a sub-field of machine learning, with neural networks comprising the backbone of deep learning algorithms. We employ a simple neural network underpinning more complicated deep learning architectures [165]. Neural networks have been used to model complex behavior from time-series gait data [1, 166, 167], making them an ideal candidate to model exoskeleton responses as discrepancies. Neural networks (i) can learn complex patterns in experimental data that coalesce from diverse sources, (ii) do not make explicit assumptions about physiology and motor control, and (iii) learn time-variant features of exploration. In biomechanics research, neural networks have been used to classify and predict human movement [28, 30, 168–170]. While neural networks applied to lower-limb exoskeletons and prostheses are most often used for trajectory prediction and device control [1, 171–175], they have not quantified exoskeleton responses.

The purpose of this study was to determine if a discrepancy modeling framework could quantify individual-specific gait responses to ankle exoskeletons (Fig. 4.1). We employed neural network-based discrepancy modeling to encode processes governing joint kinematic and EMG responses of nondisabled adults to bilateral passive ankle exoskeletons. Specifically, we modeled the discrepancy between the processes governing gait in (i) a *Nominal Condition* (*i.e.*, zero-stiffness exoskeletons) and (ii) an *Exo Condition* (*i.e.*, exoskeletons with stiff springs resisting ankle dorsiflexion). If a *Discrepancy* model can encode exoskeleton responses, then augmenting a model of *Nominal* gait with the *Discrepancy* model should predict *Exo* kinematics and EMG more accurately than the *Nominal* model alone. Therefore, we hypothesized that (i) the *Nominal* model would predict *Exo* kinematics and EMG less accurately than for the *Nominal* condition, and (ii) the *Augmented* (Nominal+Discrepancy) model would capture greater variance in *Exo* kinematics and EMG than the *Nominal* model. To assess whether standard gait measurements have sufficient resolution for encoding exoskeleton gait, we evaluated the extent to which the *Exo* model could predict *Exo* gait. Finally, to assess the viability of discrepancy modeling in gait analysis settings, we evaluated the effect of limiting training data quantity on predictions of *Exo* gait.

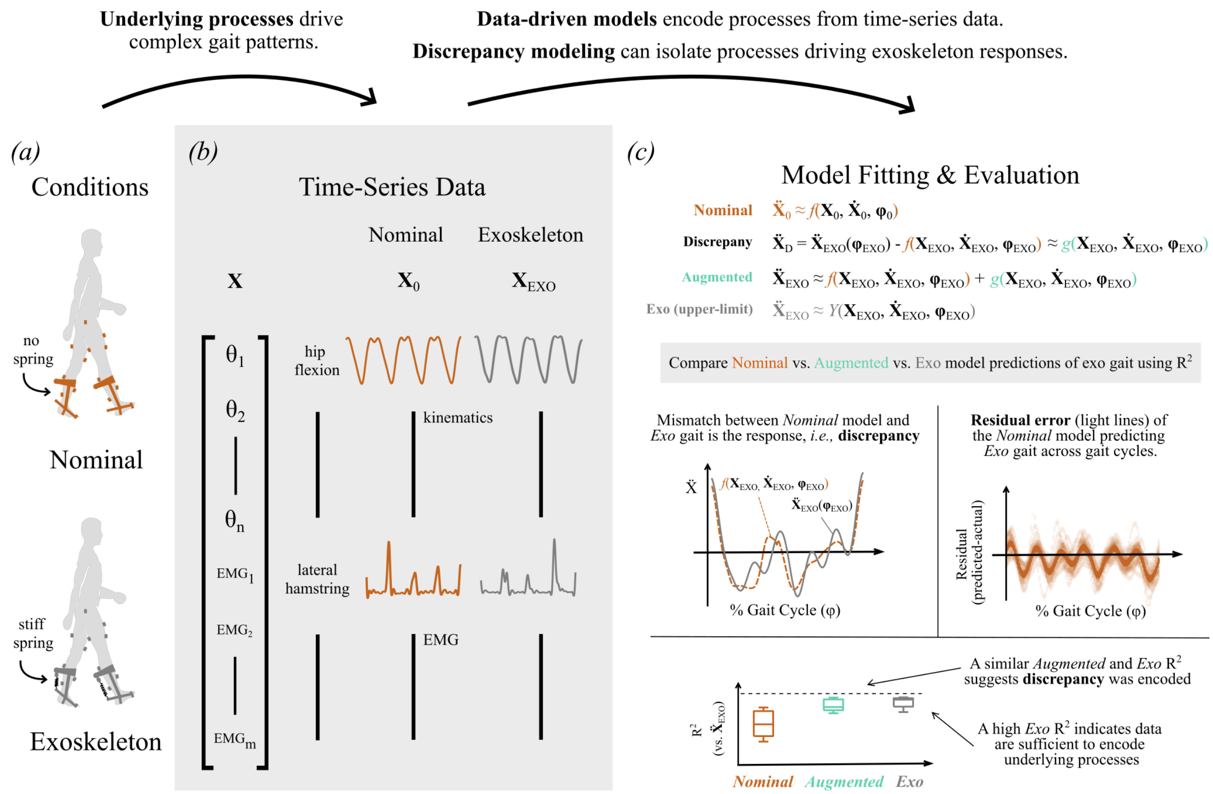

Figure 4.1: ***Framework outlining a machine learning approach using discrepancy modeling to quantify individual exoskeleton responses.*** *(a) Joint kinematics (θ) and electromyography (EMG) data were collected from nondisabled participants during treadmill walking in bilateral passive ankle exoskeletons. Two conditions were analyzed: Nominal (exoskeleton with no spring) and Exo (exoskeleton with a $5\ Nm\ deg^{-1}$ spring). (b) We encoded the processes governing the time evolution of gait by identifying a data-driven differential equation function transforming model inputs (time-series gait data, $\mathbf{X}$, their first time derivatives, $\dot{\mathbf{X}}$, and percent stride, ϕ) into their outputs (second time derivatives of gait, $\ddot{\mathbf{X}}$). (c) Feed-forward neural network models identified: (1) a* ***Nominal model*** *(orange) using Nominal gait data, $\mathbf{X}_0$. The discrepancy between Exo gait and the Nominal model predictions is the exoskeleton response. (2) A* ***Discrepancy model*** *was trained using the identified response for each individual. This Discrepancy model was added to the Nominal model to create an* ***Augmented model*** *(green). (3) Finally, an* ***Exo model*** *(gray) was trained using the Exo gait data, $\mathbf{X}_{EXO}$; this served as the maximum expected variance. We compared how much variance in Exo gait is accounted for (R^2) using the Nominal, Augmented (Nominal+Discrepancy), and Exo models and performed residual analysis on model predictions.*

4.2 Methods

4.2.1 Human subjects

We analyzed gait data for 12 nondisabled adults (6F/6M; age = 23.9 ± 1.8 yrs; height = 1.69 ± 0.10 m; mass = 66.5 ± 11.7 kg) during treadmill walking in bilateral passive ankle exoskeletons at a self-selected speed (speed = 1.36 ± 0.11 m/s; data available: https://simtk.org/projects/ankleexopred) [1]. Exoskeletons resisted ankle dorsiflexion using linear springs attached in parallel to the shank. Data were captured in two conditions: The *Nominal Condition* involved walking while wearing exoskeletons without springs resisting ankle flexion (*i.e.*, zero stiffness); The *Exo Condition* involved walking in exoskeletons resisting ankle dorsiflexion with a stiffness of 5 $Nm\ deg^{-1}$ (Fig. 4.1a).

The complete experimental protocol, including exoskeleton fitting and walking practice, is described in [1]. Briefly, after an exoskeleton fitting and practice session on a prior day, participants walked on a split-belt instrumented treadmill (Bertec Corp., Columbus, USA) for six minutes per condition: a two-minute familiarization period, followed by four minutes at their self-selected speed. Conditions were randomized. Marker motion was recorded using a 10-camera optical motion capture system (Qualisys AB, Gothenburg, SE) and EMG signals were recorded bilaterally from five muscles (vastus medialis, soleus, gastrocnemius, gluteus medius, lateral hamstrings) (Delsys Inc., Natick, USA). Joint kinematics were estimated from marker data using the Inverse Kinematics algorithm in OpenSim 3.3 with a 19 degree-of-freedom skeletal model [47, 48]. Joint kinematics were low-pass filtered at 6Hz using a fourth-order Butterworth filter [1]. EMG data were high-pass filtered at 40Hz, rectified, and low-pass filtered at 10Hz using fourth-order Butterworth filters. All human subject procedures were evaluated by the Institutional Review Board at the University of Washington.

4.2.2 Continuous-time neural network models of walking with ankle exoskeletons

To characterize discrepancies between processes underlying *Nominal* and *Exo* gait, we fit individual-specific neural networks to each participant's data (MATLAB, Deep Learning Toolbox). The neural networks learned continuous-time nonlinear transformations from joint kinematics and EMG to their time derivatives, thereby encoding the biomechanical, neural, and sensory processes driving the time evolution of gait and how those processes change with ankle exoskeletons [11, 12, 176]. We defined a state vector that included n = 10 lower-limb joint angles kinematics and m = 14 EMG recordings:

$$\mathbf{X}(\mathbf{t}) = [\boldsymbol{\theta}_1(\mathbf{t}),\ \boldsymbol{\theta}_2(\mathbf{t}),\ \cdots,\ \boldsymbol{\theta}_n(\mathbf{t}),\ \mathbf{EMG}_1(\mathbf{t}),\ \mathbf{EMG}_2(\mathbf{t}),\ \cdots,\ \mathbf{EMG}_m(\mathbf{t})], \tag{4.1}$$

where $\boldsymbol{\theta}(\mathbf{t}) \in \mathbb{R}^{\mathbf{n}\times\mathbf{t}}$ were n kinematic states and $\mathbf{EMG}(\mathbf{t}) \in \mathbb{R}^{\mathbf{m}\times\mathbf{t}}$ were m electromyography states uniformly sampled in time $\mathbf{t} = [t_1,\ t_2,\ \cdots\ t_k]$. These variables reflect common measurements in gait analysis that are relevant to quantifying and synthesizing human movement [1, 125, 134, 177–180]. We included include EMG signals as state variables because we think EMG encodes important information about individual-specific responses to exoskeletons that may not be captured by kinematics alone [130, 131].

The *Nominal*, *Exo*, and *Discrepancy* inputs included the state vector and its first derivative, as well as percent stride, ϕ (Fig. 4.1b). Output data for the *Nominal* and *Exo* models were a subset of the variables from the state vector: hip, knee, and ankle angular accelerations and the second time derivatives of the HAM, RF, VAS, GAS, and SOL EMG recordings from their respective exoskeleton condition. Output data for the *Discrepancy* model were the exoskeleton responses, as identified in Eq. 4.4, for the hip, knee, ankle, HAM, RF, VAS, GAS, and SOL. For each participant, model, and walking condition, we withheld the final 30 seconds (12.5%) of each trial from the training dataset. The held-out data were used for model evaluation.

We modeled discrepancies up to the state vector's second derivative, consistent with the Newtonian physics framework commonly used to define biomechanical dynamics

[181].

For each participant, we trained feed-forward neural network models (Fig. 4.1c) of walking for the Eq. 4.2: *Nominal Condition*, Eq. 4.3: *Exo Condition*, and Eq. 4.4: the *Discrepancy* between the *Nominal* and *Exo* conditions:

$$\textbf{Nominal} : \ddot{\mathbf{X}}_0 \approx f(\mathbf{X}_0, \dot{\mathbf{X}}_0, \boldsymbol{\phi}_0), \tag{4.2}$$

$$\textbf{Exo} : \ddot{\mathbf{X}}_{EXO} \approx Y(\mathbf{X}_{EXO}, \dot{\mathbf{X}}_{EXO}, \boldsymbol{\phi}_{EXO}), \tag{4.3}$$

$$\textbf{Discrepancy} : \ddot{\mathbf{X}}_D = \ddot{\mathbf{X}}_{EXO}(\boldsymbol{\phi}_{EXO}) - f(\mathbf{X}_{EXO}, \dot{\mathbf{X}}_{EXO}, \boldsymbol{\phi}_{EXO}) \approx g(\mathbf{X}_{EXO}, \dot{\mathbf{X}}_{EXO}, \boldsymbol{\phi}_{EXO}) \tag{4.4}$$

The *Discrepancy* model can then be used to *Augment* the *Nominal* model:

$$\textbf{Augmented} : \ddot{\mathbf{X}}_{EXO} \approx f(\mathbf{X}_{EXO}, \dot{\mathbf{X}}_{EXO}, \phi_{EXO}) + g(\mathbf{X}_{EXO}, \dot{\mathbf{X}}_{EXO}, \phi_{EXO}). \tag{4.5}$$

By adding the *Discrepancy* model to the *Nominal* model, we are testing whether there are organized and predictable features captured in the *Discrepancy* that can be used to quantify differences in individual exoskeleton responses. If the kinematic and EMG data are sufficient to capture responses, then the *Augmented* model (Eq. 4.5) should predict *Exo* gait as accurately as the *Exo* model (Eq. 4.3). The *Exo* model thus serves as an idealized 'upper-limit' case (*i.e.,* predictive potential when the response is known).

The feed-forward neural networks used three layers with activation functions: (1) *log-sigmoid*, to encode nonlinear relationships between features, (2) *radial basis*, to approximate functions and time-series relationships, and (3) *pure linear*, to transform activations to an appropriate scale for output. To avoid information loss due to dimensionality reduction, each layer had 64 nodes, lifting the 49 input features of gait into a higher-dimensional latent representation [182].

4.2.3 Statistical analysis

We quantified, using the coefficient of determination (R^2), the extent to which processes underlying *Nominal*, *Augmented* (Nominal+Discrepancy), and *Exo* models could predict the held-out joint angle accelerations and second time derivatives of EMG during *Exo* gait. For each variable and across individuals, we compared R^2 between (i) *Nominal* and *Augmented* models, (ii) *Nominal* and *Exo* models, and (iii) *Augmented* and *Exo* models using Wilcoxon Signed-Rank tests with Holm-Sidak Stepdown corrections for multiple comparisons ($\alpha = 0.05$) [1, 183]. One participant was excluded from analysis due to changes in gait – possibly due to adaptation or conscious exploration of different gait patterns – late in the *Exo* condition walking trial. Additionally, we compared variance accounted for by the *Exo* model against $R^2 = 1$. If these data alone could explain *Exo* gait, R^2 would approach 1, and $R^2 < 1$ would suggest additional measurement modalities or higher-resolution signals are needed to encode *Exo* gait.

Additionally, we performed residual analysis on model predictions [93, 184]. We plotted residuals across all gait cycles for the *Nominal*, *Augmented* (Nominal+Discrepancy), and *Exo* model predictions. Non-zero stride-averaged residuals indicate that a model does not capture processes driving gait. If the stride-averaged residual is zero, either all processes driving gait were captured or noise obscured the disambiguation of missing processes (a limit imposed by sensor technology).

4.2.4 Data quantity

To assess the data quantity required to encode *Exo* gait, we evaluated the effect of training data length on variance accounted for. The required data quantity will dictate the settings in which discrepancy modeling is practical, such as in gait analysis where datasets contain only a few gait cycles [134, 156]. Therefore, we iteratively changed training data length (n

= [1, 10, 20, 30, 45, 60, 90, 120, 180] seconds) for the *Nominal*, *Exo*, and *Augmented* models and calculated *Exo* variance accounted for. The full training set contained 210 seconds of data for each participant. For all training set lengths, we evaluated models using the full 30-second validation dataset. For all variables and across individuals, we compared (i) the *Augmented* versus *Exo* models at each training data length, and (ii) the *Exo* models with the smaller versus full training dataset at each length.

4.3 Results

Discrepancy modeling captured the complex changes in gait underlying individual responses to ankle exoskeletons. Across participants, the *Nominal* model accounted for $96.0 - 98.1\%$ and $69.7 - 76.8\%$ of the *Nominal* condition's kinematic and EMG median variance, respectively, while only accounting for $86.3 - 94.0\%$ and $51.6 - 66.4\%$ of the *Exo* condition's kinematic and EMG median variance, suggesting the processes underlying *Nominal* and *Exo* gait differ. While joint kinematics during *Exo* gait were well predicted using the *Nominal* model (orange in Fig. 4.2; median $R^2 = 0.863 - 0.939$), the *Augmented* model (green in Fig. 4.2) significantly increased variance accounted for ($p < 0.042$, median $R^2 = 0.928 - 0.963$). For EMG, the *Augmented* model (median $R^2 = 0.665 - 0.788$) accounted for significantly more variance than the *Nominal* model (median $R^2 = 0.516 - 0.664$). Indeed, the *Augmented* model accounted for $2.34 - 14.91\%$ more variance (median across participants) in *Exo* gait than the *Nominal* model, suggesting that the discrepancy model captured processes driving responses. The *Augmented* model also predicted kinematics and EMG with similar accuracy to the upper-limit *Exo* model (gray in Fig. 4.2). While the *Augmented* model explained significantly less variance in the knee, ankle, HAM, and VAS compared to the *Exo* model (Sidak $\alpha = 0.013$), differences in median R^2 were small (knee = 0.0144, ankle = 0.0267, HAM = 0.0469, VAS = 0.0936). The remaining variables'

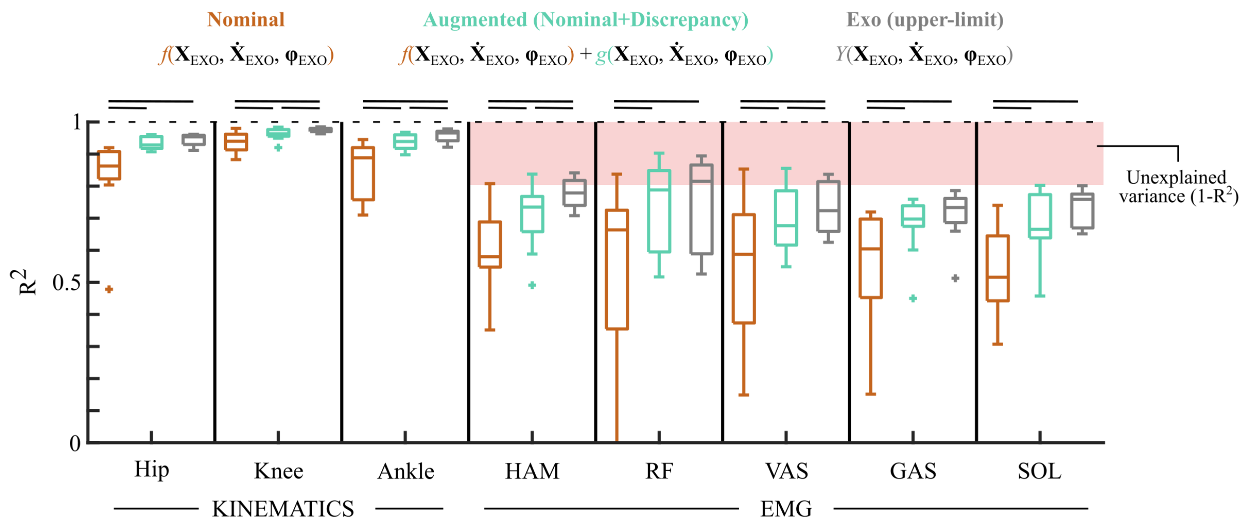

Figure 4.2: ***Discrepancy modeling captured individual responses to ankle exoskeletons and quantified sufficiency of kinematic and EMG measurements for encoding exoskeleton gait.*** *The Augmented (Nominal+Discrepancy) model (green) explained significantly more variance in joint kinematics and EMG measurements compared to the Nominal model (orange). Boxplots show variance accounted for (R^2) across participants (N = 11) on 30 seconds of held-out exoskeleton gait data. Horizontal bars denote significant differences between the models according to Wilcoxon Signed-Rank tests with Holm-Sidak Step-down corrections for multiple comparisons ($\alpha = 0.05$). The Exo model (gray) indicates expected maximum variance.*

variance was not significantly different between the *Augmented* and *Exo* models. Minimal kinematic variance was left unexplained by the *Exo* model (median $R^2 = 0.954 - 0.978$), but only accounted for $72.4\% - 81.5\%$ of the median variance in EMG during *Exo* gait across all individuals.

Residual analysis revealed non-zero trends in the stride-averaged residuals from the *Nominal* model across variables; Fig. 4.3 shows results from one representative subject. If the data were sufficient to fully encode the processes governing exoskeleton responses, the stride-averaged residuals would be at or near zero. Compared to the *Nominal* model (orange in Fig. 4.3), both the *Augmented* model (green in Fig. 4.3) and *Exo* (gray in Fig. 4.3) reduced the variance in the stride-averaged residuals, especially for the kinematic vari-

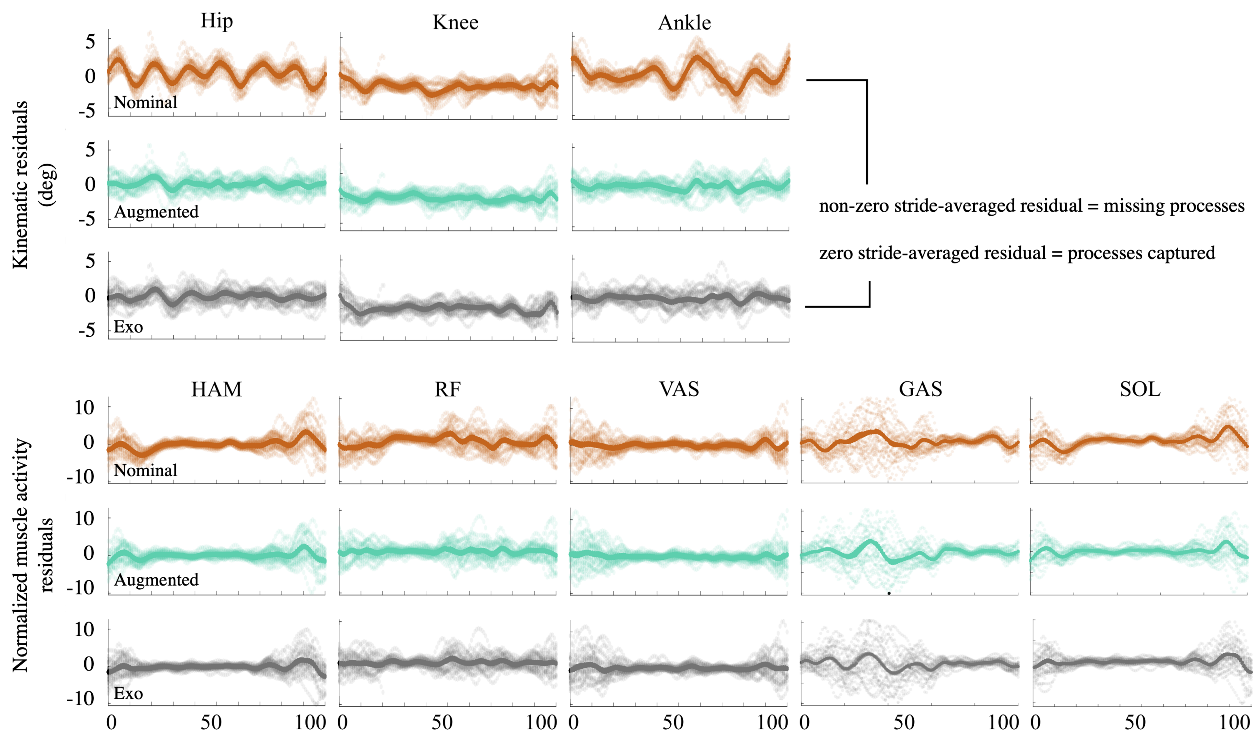

Figure 4.3: ***Residual analysis for one representative participant indicates discrepancy modeling identified processes underlying complex gait changes****. Less saturated colors represent residual between model outputs and measured exoskeleton gait; more saturated colors represent the stride-averaged residuals. Patterns in the Augmented model's residual diminished, as compared to Nominal, suggesting discrepancy modeling captured processes underlying gait responses to ankle exoskeletons.*

ables. For example, the mean of the *Nominal* model's hip residual has a clear non-zero oscillatory pattern, indicative of an incomplete model. The magnitude of this pattern was reduced by 52.8% when the *Augmented* model predicted *Exo* gait. When evaluating residuals in predicted EMG, we focus on the portions of the gait cycle in which muscles are active. We observed a decrease in magnitude of the stride-averaged residuals across all muscles with the *Augmented* model. Except for 1 second of training data – which predicted negative R^2 for all models (Fig. 4.4) – the *Augmented* model did not predict significantly different R^2 than the *Exo* model regardless of data quantity, for all variables and across individuals. This suggests discrepancy modeling is effective, even with small data quan-

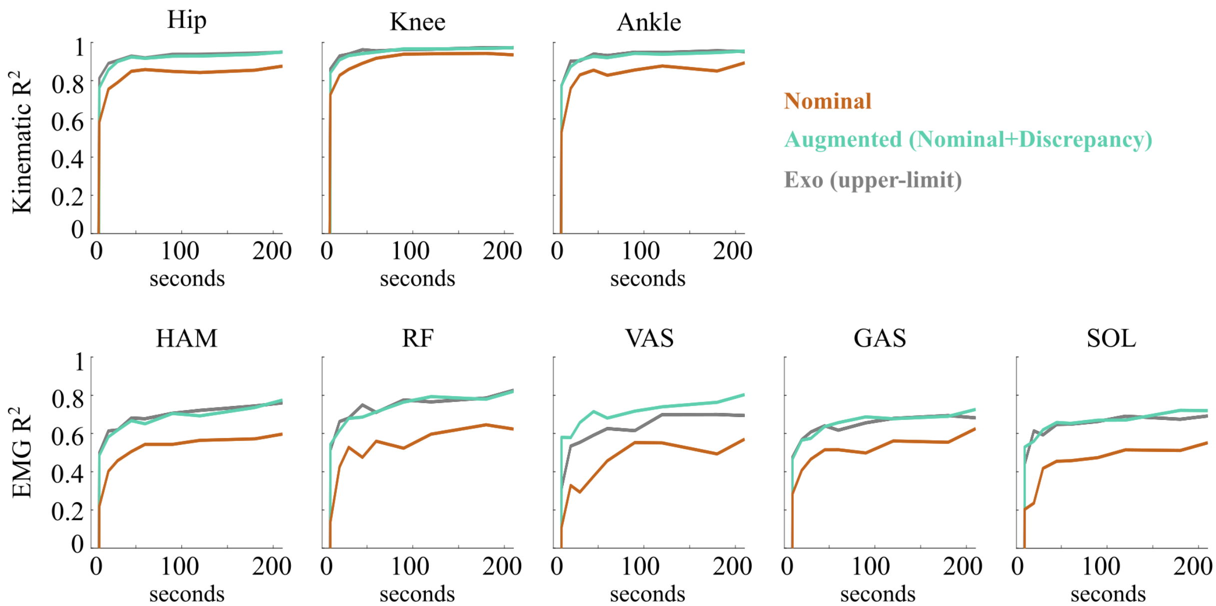

Figure 4.4: **Discrepancy modeling captures exoskeleton responses, even with small training datasets.** *Median R^2 for kinematics (top) and EMG (bottom) during exoskeleton walking for the Nominal (orange), Augmented (Nominal+Discrepancy) (green), and upper-limit Exo (gray) models over training set sizes ranging from 1 to 210 (full set) seconds.*

tities. However, the ability to encode the response with the discrepancy diminishes with less data. For all variables and across all individuals, 180 seconds of data are needed for the *Exo* model to predict statistically similar R^2 to the *Exo* model's full training dataset.

4.4 Discussion

A neural network-based discrepancy modeling framework successfully captured changes in the processes governing gait kinematics and EMG with ankle exoskeletons (*i.e.*, discrepancies) without requiring prior knowledge or assumptions about the physiological structure of the discrepancy. Compared to the *Nominal* model alone, improved predictions of gait using with the *Augmented* (Nominal+Discrepancy) model highlights this framework's ability to quantify exoskeleton responses as discrepancies. However, non-zero residuals, even for the upper-limit *Exo* model, indicate additional measurement modalities and/or improved measurement resolution are needed to capture the complex muscle-level changes in gait with ankle exoskeletons. Our framework's ability to isolate discrepancies in *Exo* walking provides a first step towards leveraging machine learning to uncover physiological mechanisms driving responses to exoskeletons or other assistive devices.

As suggested by our prior work developing discrepancy models in synthetic systems [159], we found that neural networks captured individual-specific processes governing changes in kinematics and EMG with exoskeletons from experimental data. As described below, the ability to quantify discrepancies is relevant to existing modeling frameworks because it can (i) capture differences in exoskeleton responses, (ii) encode physiological mechanisms underlying responses, and (iii) resolve dynamical inconsistencies between observations and predictions.

Our results support that discrepancy modeling can quantify individual-specific exoskeleton responses without assuming neuromusculoskeletal structure. Because few studies report kinematic and EMG accuracy, and because we predict accelerations and second time-derivatives, head-to-head comparisons are challenging. Regardless, we expect our prediction accuracy is comparable to predictions from physiological-detailed models [154]. The ability of our framework to encode discrepancies from data is relevant for modeling exoskeleton responses in patient populations: existing physiologically-detailed models are challenging to tune to an individual's physiology and motor control, especially for individuals with highly-heterogeneous physiology [146, 151–154, 185, 186]. Our discrepancy framework's ability to encode processes governing gait changes with exoskeletons could, therefore, enable precise individual-specific representations of exoskeleton walking, even in the absence of explicit understanding of an individual's physiology.

These structures may enable more accurate predictions of exoskeleton responses and suggest physiological features of gait explaining these responses [1, 130, 154].

Discrepancies may encode physiological mechanisms underlying exoskeleton responses. Increased residuals when predicting *Exo* gait with the *Nominal* model suggests missing processes. Reductions in these residuals with the *Augmented* model further supports that discrepancies captured exoskeleton responses. Yet, we do not know what physiological mechanisms constitute the discrepancy. Prior studies have found motor, muscle-tendon, and optimal control mechanisms that impact responses to exoskeletons and other assistive devices [130, 147, 154]. Therefore, it is likely that muscle-tendon properties, sensory processes, and motor control underlie discrepancies [147, 154, 187–190]. However, these studies do not reveal which feature(s) underlying responses. Discrepancy models, by definition, isolate the exoskeleton response, such that the features underlying responses could be grounded in physiology. This work shows the discrepancy modeling framework can capture individual-specific processes driving changes in exoskeleton kinematics and EMG. Uncovering which of the above physiological mechanisms – and possibly others [191–193] – explain discrepancies is an interesting avenue of future research.

This discrepancy modeling framework represents a complementary approach to established biomechanical models. Existing modeling approaches typically use iterative optimization to select model parameters(*e.g.*, cost functions, muscle-tendon parameters, joint moments, and activation dynamics) explaining changes in kinematics, kinetics, and muscle activity across experimental conditions [146, 152, 154, 194]. Conversely, the fitting process of the discrepancy modeling framework tunes model parameters based solely on the consistency of model predictions to kinematic accelerations and EMG second time-derivatives. As a result, discrepancy modeling does not require iterative optimization: the neural networks used here could be replaced with a least-squares regression model to more-rapidly (but possibly less-accurately) discover discrepancies. Further, discrepancy modeling aims to resolve dynamical inconsistences, not just reproduce gait kinematics. Resolving dynamical inconsistencies with exoskeletons in this study is analogous to OpenSim's residual reduction algorithm (RRA), which ensures dynamical (*i.e.*, forces and moments) consistency between a skeletal model and measured external forces [127, 157, 195]. Therefore, while this work supports the validity of discrepancy modeling for exoskeleton responses, the framework's utility may generalize to myriad biomechanical applications.

The models in this work, similar to prior work [1, 146, 153], struggled to capture the complex changes in EMG. If gait with exoskeletons is not fully described by the data, it is unlikely the discrepancy will completely encode exoskeleton responses. The limited prediction accuracy of the *Exo* model ($R^2 < 1$) indicates that kinematic and EMG data lack information needed to encode processes driving EMG. Therefore, discrepancies describing EMG responses between *Nominal* and *Exo* gait could not be comprehensively captured. Why did we not capture EMG responses? If the *Exo* model provided the expected maximum variance, and variance remained unexplained, we may need additional measurement modalities and/or improved resolution. For example, Sawicki and colleagues (2015) demonstrated that including muscle-tendon mechanics could improve predictions of individual responses to ankle exoskeletons during hopping [147]. Incorporating additional data, such as measures of muscle-tendon mechanics from ultrasound [131] or musculoskeletal simulation could improve response predictions. The training data quantity may also limit our ability to encode from data the processes driving gait and exoskeleton responses. Particularly for EMG, unexplained *Exo* variance remained. Prediction accuracies, however, plateaued for all variables, similar to a prior study of gait with exoskeletons [1], suggesting that predictions would not improve with more training data alone. Finally, experimental protocol may impact our ability to encode exoskeleton responses from data. For example, the exoskeletons may not have elicited large or diverse enough perturbations to distinguish processes underlying responses from noise during model fitting. Experimental paradigms eliciting larger responses may increase the ability to encode discrepancies from data.

One limitation of this study is our evaluation of a single passive exoskeleton condition, similar to current standard practice for clinical prescription of ankle foot orthoses. However, understanding whether discrepancy modeling can encode responses to changes in other exoskeleton parameters (*e.g.*, other mechanical properties or powered exoskeleton control policies) represents important areas of future work. Additionally, we applied discrepancy modeling to nondisabled participants; the framework's ability and/or the data needed to capture discrepancies may change for individuals with disabilities [196, 197]. Further, the participant omitted from analysis highlights an important limitation of discrepancy modeling: the discrepancy relies on training data emerging from the same processes that govern observed behavior (the test data, in this study). This participant may have fatigued or adapted during the trial, such that training the discrepancy on data early in the trial may fail to capture discrepancies later in the trial. Another potential limitation was in our selection of processing parameters (*e.g.* low-pass filtering cutoff frequency), as such parameters may remove non-noise information from the signals [198]. Finally, while the discrepancy modeling framework is model agnostic, it is important to consider limitations of the model's structure (*e.g.*, neural networks), which may impact the ability to capture and interpret exoskeleton responses. Therefore, we hyperparameter tuned layer quantity, nodes per layer, and activation functions to approximate gait without overfitting. Despite the network's simple architecture, it captured discrepancies; more sophisticated deep learning architectures could be employed in future studies.

4.5 Conclusion

This study used a neural network-based discrepancy modeling framework to quantify individual responses to ankle exoskeletons. We found that discrepancy modeling captured processes underlying complex changes in kinematic and EMG data with exoskeletons. Additionally, we demonstrated how discrepancy modeling can be used to identify what and how much data are needed to capture responses. Discrepancy modeling is a unique and innovative tool that complements current biomechanical modeling approaches and may accelerate the discovery of individual-specific mechanisms driving responses to exoskeletons, other assistive devices, and clinical interventions.

4.6 Declaration of Competing Interest

There are no conflicts of interest to report.

4.7 Acknowledgements

MRE acknowledges support from the National Science Foundation under award GRFP DGE-1762114. MCR acknowledges support from the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1762114. JNK acknowledges funding from the National Science Foundation AI Institute in dynamical Systems grant number 2112085.

Chapter 5

LEVERAGING ARBITRARY MOBILE SENSOR TRAJECTORIES WITH SHALLOW RECURRENT DECODER NETWORKS FOR FULL-STATE RECONSTRUCTION

Sensing is one of the most fundamental tasks for the monitoring, forecasting and control of complex, spatio-temporal systems. In many applications, a limited number of sensors are mobile and move with the dynamics, with examples including wearable technology, ocean monitoring buoys, and weather balloons. In these dynamic systems (without regions of statistical-independence), the measurement time history encodes a significant amount of information that can be extracted for critical tasks. Most model-free sensing paradigms aim to map current sparse sensor measurements to the high-dimensional state space, ignoring the time-history all together. Using modern deep learning architectures, we show that a sequence-to-vector model, such as an LSTM (long, short-term memory) network, with a decoder network, dynamic trajectory information can be mapped to full state-space estimates. Indeed, we demonstrate that by leveraging mobile sensor trajectories with shallow recurrent decoder networks, we can train the network (i) to accurately reconstruct the full state space using arbitrary dynamical trajectories of the sensors, (ii) the architecture reduces the variance of the mean-square error of the reconstruction error in comparison with immobile sensors, and (iii) the architecture also allows for rapid generalization (parameterization of dynamics) for data outside the training set. Moreover, the path of the sensor can be chosen arbitrarily, provided training data for the spatial trajectory of the sensor is available. The exceptional performance of the network architecture is demonstrated on three applications: turbulent flows, global sea-surface temperature data, and human movement biomechanics.

5.1 Introduction

Sensing is a ubiquitous task, being critically important to every scientific and engineering discipline [199]. Sensor technologies accelerate scientific advancement by providing improved observational data that is capable of advancing new scientific theories and improving technological development. From the James Webb telescope [200] to single neuron recordings [201], sensors provide an interface with the physical world that allows us to interrogate diverse and complex systems across scientific domains. In many application fields, sensors are mobile by design, moving with dynamical trajectories as they record quantities of interest. Application areas such as motion tracking of human biomechanics or robotic systems, buoys that measure ocean dynamics, and weather balloons are all examples of technologies where sensors are not (or cannot be) immobile. From these mobile sensors, some or all of the canonical tasks of sensing are enacted, including reconstruction, forecasting, model discovery, control and uncertainty quantification [199]. Emerging sensor technologies are capable of producing exceptional quantities of data which can now be leveraged by machine learning algorithms. Specifically, we utilize a recurrent neural network [202–204] to learn the temporal sequences in a latent space that maps to full state-space estimates via a shallow decoder network [113, 205]. This *shallow recurrent decoder network* (SHRED) architecture has been successful for immobile sensors [206], but this work shows how it can be used in a mobile sensing framework. Indeed, mobile sensing affords distinct advantages over immobile sensors, including the reduction of the variance of the mean-square reconstruction error in comparison with

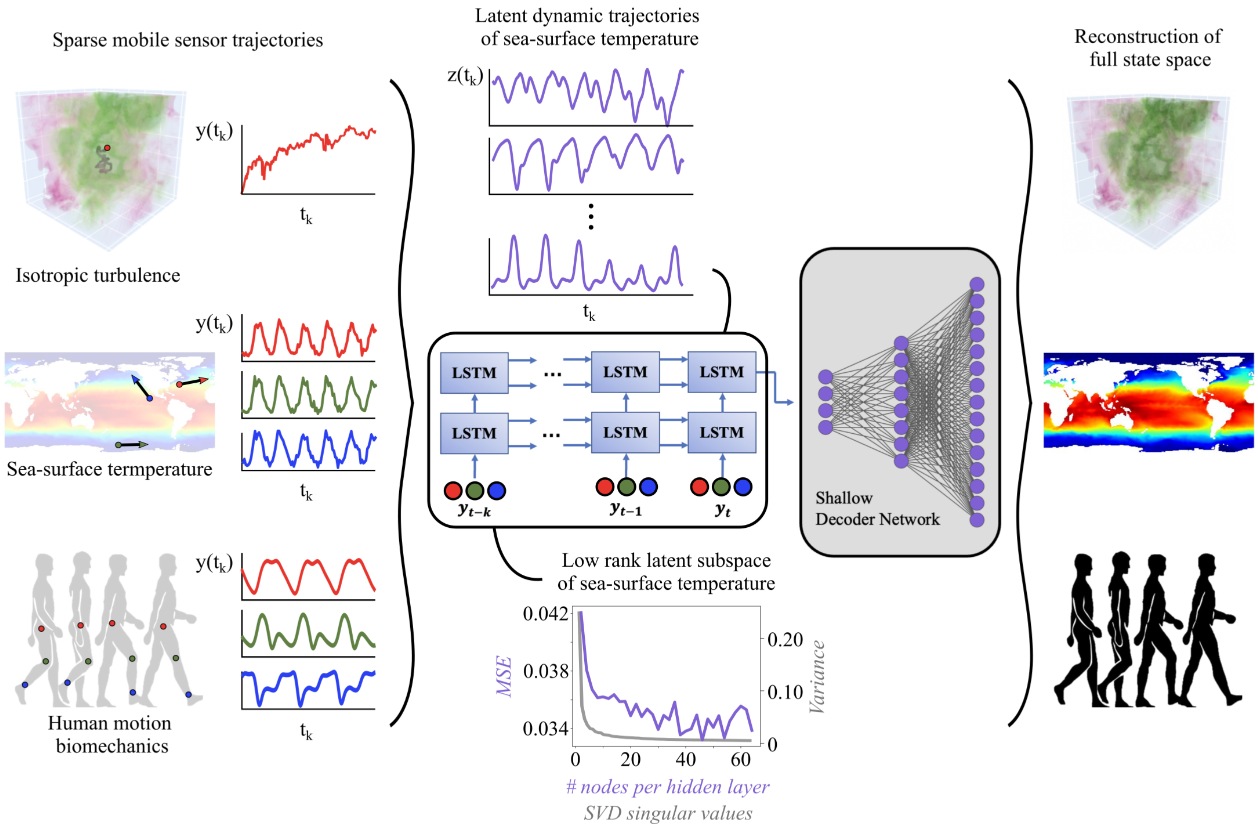

Figure 5.1: Summary figure of a shallow recurrent decoder network (SHRED) leveraging mobile sensors to reconstruct full state-space estimates from sparse dynamical trajectories. (Left) Sensor trajectory history encodes global information of the spatio-temporal dynamics of the sparsely measured system. In this work, we evaluate three challenging datasets, including forced isotropic turbulence, global sea-surface temperature, and human biomechanics. (Middle) The mobile SHRED architecture can (i) embed the multiscale physics of a system into a compact and low-dimensional latent space, and (ii) provide a mapping from the sparse mobile sensors to a full state estimate. (Right) The high-dimensional and complex system states can be reconstructed, provided training data for the dynamical trajectory of the sensor(s) is available.

immobile sensors and improved generalization (parameterization of dynamics) for data outside the training set.

Traditional system identification techniques such as the Kalman filter and its variants have long achieved state-of-the-art performance [109], but are limited to applications where a model of the underlying dynamics is readily available. With the recent explosion in quantity and quality of data from high-dimensional systems, we are increasingly required to perform state estimates of systems with unknown or computationally infeasible dynamics. State estimation in these cases is typically achieved by exploiting low-rank structures in the data learned by a singular value decomposition (SVD), otherwise known as proper orthogonal decomposition (POD) [207, 208]. Gappy POD was one of the first such techniques to reconstruct high-dimensional fields from limited, sparse sensor measurements [209]. Since then, much work has been aimed at optimizing sensor placement to well-condition the linear inverse problem upon which state estimates rely [199, 210–212]. There exist numerous techniques for determining optimal sensor locations in small search spaces [213–219], but in high-dimensional systems a greedy approximation, such as QR decomposition with column pivoting, is needed. These techniques for state-estimation in the absence of a governing model also differ from Kalman filtering in that they only utilize static measurements, rather than condition upon a history of sensor data. While this lends the methods flexibility to reconstruct data from a dataset of faces or a dynamical system, it ignores the rich information that can be gleaned from incorporating past sensor measurements in a method for reconstruction.

Leveraging the temporal histories of sensor measurements offers an opportunity to capitalize on both the rich information from measurement trajectories and enable greater inclusion of measurement modalities – beyond immobile sensors. To date, the inclusion and design of dynamic trajectories are addressed with sensor path planning. Path planning is a long-standing challenge in the field of engineering and robotics related to navigation and estimation in dynamical environments [220–224]. The majority of path planning has focused on the modeling and control of the mobile sensor positions [225–232]. For example, recent work proposed the use of mobile sensors with Kalman filtering to estimate spatio-temporal data [233]. The authors' approach to sensor path planning for dynamic estimation includes two main steps: (i) dynamic mode decomposition finds a low-rank representation of the data for Kalman filtering, and (ii) greedy path finding optimizes the observability matrix along the path and improves Kalman filter estimation. However, path planning for mobile sensing and state estimation may be unnecessary. The SHRED architecture has demonstrated the successful reconstruction of complex spatio-temporal data with arbitrarily placed – albeit immobile – sensors [206]. Therefore, this work aims to employ mobile sensors with SHRED to understand whether the use of arbitrary dynamic trajectories can feasibly and reliably reconstruct complex spatio-temporal data, thereby removing the challenge of path planning altogether.

As a mathematical architecture, SHRED leverages emerging deep learning paradigms [166] and their universal approximation capabilities [234] to map from sparse and minimal sensor measurements to high-dimensional spatio-temporal data. In contrast with common sensing modalities where principled sensor placement is critical [199], the LSTM embedded in the SHRED architecture extracts a latent representation of the low-dimensional dynamics using sensor trajectory (time history) information which is agnostic to the sensor location. Thus the sensor trajectory history encodes *global* information of the spatio-temporal dynamics of the measured system. And while previous research has considered mobile sensor features like the timescale of the spatio-temporal dynamics, velocity of the sensors, and rate of sampling as part of the path planning optimization [233], in this work, such features are treated hyperparameters to be tuned in the SHRED architecture. We demonstrate the performance of SHRED on three challenge data sets related to turbulence, sea-surface temperature, and human biomechanics. In all three cases, the mobile SHRED architecture provides a high-quality and accurate algorithm for estimating the original complex, multiscale, and high-dimensional system. In all the examples, it is demonstrated that by leveraging mobile sensor trajectories with SHRED, we can train the network to accurately reconstruct the full state space while also reducing the variance of the mean-square error of the reconstruction error in comparison with immobile sensors. Moreover, the path of the sensor can be chosen arbitrarily, provided training data for the spatial trajectory of the sensor is available. SHRED works with high-probability given that the data do not have regions of statistically independent data. In addition, when full state data is not available, proxy computational data with empirically similar data can be used for training.

5.2 Mathematical Formulation

The shallow recurrent decoder network (SHRED) can be understood as the amalgamation of an LSTM for processing a time-series of sensor measurements followed by a feedforward neural network, or decoder, for reconstructing a high-dimensional state from the learned latent representation of the LSTM [206]. Let the high-dimensional state to be reconstructed be denoted as $x_T \in \mathbb{R}^n$ and assume access to a set of sensor measurements $y_t = Cx_t \in \mathbb{R}^m$ for $t \in \{T-K, T-K+1, \ldots, T-1, T\}$. K can be determined by empirical analyses of the system at hand. We assume the measurements are sparse point measurements, that is $m << n$ and C consists of rows of the $n \times n$ identity matrix, although there is some evidence to suggest that neural network based reconstructions can be performed with nonlinear measurements [113]. The set of sensor measurements serve as inputs to an LSTM [235] with recursive update equations

$$h_t = \sigma\left(W_o\left[h_{t-1}, y_t\right] + b_o\right) \odot \tanh(c_t) \tag{5.1}$$

$$c_t = \sigma\left(W_f\left[h_{t-1}, y_t\right] + b_f\right) \odot c_{t-1} \tag{5.2}$$

$$+\sigma\left(W_i\left[h_{t-1}, y_t\right] + b_f\right) \odot \tanh\left(W_g\left[h_{t-1}, y_t\right] + b_g\right)$$

where $W_{RN} = \{W_o, W_f, W_i, W_g, b_o, b_f, b_i, b_g\}$ are the trainable weights and biases of the LSTM. We denote

$$h_T = \mathcal{G}(\{y_t\}_{T-K}^{T}; W_{RN}). \tag{5.3}$$

The latent state h_T learned by the LSTM has a variety of interesting properties that will be discussed in a later section.

The feedforward component of the SHRED architecture is a shallow decoder with b layers denoted by

$$\mathcal{F}(h; W_{SD}) := R(W^b R(W^{b-1} \cdots R(W^1 h))), \tag{5.4}$$

parameterized by trainable weights $W_{SD} = \{W^1, \ldots, W^b\}$ and with nonlinear scalar activation function R (chosen to be ReLU). In total, the SHRED network is given by

$$\mathcal{H}(\{y_t\}_{T-K}^{T}) = \mathcal{F}(\mathcal{G}(\{y_t\}_{T-K}^{T}; W_{RN}); W_{SD}). \tag{5.5}$$

The network is trained to minimize reconstruction loss over a set of training states $\{x_t\}_1^N$,

$$\mathcal{H} \in \underset{\widetilde{\mathcal{H}} \in \mathscr{H}}{\operatorname{argmin}} \sum_{t=1}^{N} ||x_i - \widetilde{\mathcal{H}}\left(\{y_i\}_{i=t-K}^{t}\right)||_2, \tag{5.6}$$

using the ADAM optimizer [236]. The assumption of access to a set of high-dimensional states for training in this manner is a strong one; simultaneous measurement of an entire high-dimensional system is sometimes simply impossible. In such cases, a high-fidelity simulation can be used to train the network, provided the simulation accurately approximates the statistics of the real system. Alternatively, if full-state measurements are possible, but prohibitively expensive in the long-term, the generation of training data can be viewed as a one-time up front cost.

Previous work has demonstrated that such networks outperform traditional, POD based techniques for state estimation while requiring fewer available sensors [206]. In this work, we consider the case of a time-dependent measurement matrix C. That is,

$$y_t = C_t x_t \tag{5.7}$$

where C_t varies in time. This corresponds to a mobile sensor, in contrast to existing work which considers only immobile sensors. We emphasize that while the set of measurement matrices C_t can be chosen arbitrarily, corresponding training data is necessary in order to train the network; the architecture as it stands cannot extrapolate to unseen sensor location trajectories. In the cases of periodic or quasi-periodic phenomena, this requirement amounts to the evolution of C_t being periodic over some characteristic time-scale in order to extrapolate beyond a temporally partitioned training dataset.

Despite the reputation of neural networks as lacking interpretability, we believe that there are a variety of potential theoretical connections between the latent state learned by the LSTM of a SHRED model and other domains of mathematics. In particular, there are parallels between the function of the LSTM and Takens' embedding, which states that with a sufficiently long time-history, the dynamics of a time-delayed state variable are diffeomorphic to the dynamics of the high-dimensional state space. While the LSTM does not explicitly construct a time-delayed state, we conjecture that, once trained, it does so implicitly in its processing of sensor measurements. The feedforward network that follows the LSTM layer might then be thought of as having learned a diffeomorphism between the analog of a time-delayed vector and the high-dimensional state space.

We also believe that theoretical exploration of the the LSTM's latent dimension size is warranted. The plot in Fig. 5.1 show the MSE of reconstructions of sea-surface temperature vs the number of nodes per hidden layer and the SVD singular value spectrum.

Both plots exhibit an clear elbow at ~ 5 nodes per hidden layer and the ~ 5 singular value. In traditional dimensionality reduction terms, the data exhibits a low-rank structure and can be well approximated using the ~ 5 dominant modes learned by the SVD. We conjecture that a similar relation might exist between the minimum required LSTM state dimension and the underlying structure of the data. In that case, the values of each entry in the latent state might be understood in analogy to the entries in the matrix V of the SVD $X = U\Sigma V^T$. The plot in Fig. 5.1 illustrates that these entries do appear to evolve as some quasi-periodic phenomena.

5.3 Mobile sensing applications

To demonstrate the mobile SHRED architecture, three example problems are selected: isotropic turbulence, sea-surface temperature and human biomechanics. The physics of all three systems produce high-dimensional, multiscale dynamics which is challenging to model in practice. However, we will show in the figures that follow that SHRED provides exceptional performance by mapping between arbitrary trajectories of sensor measurements and the full state space.

5.3.1 Forced isotropic turbulent flow

The first dataset upon which we evaluate the performance of SHRED with mobile sensors is a forced isotropic turbulent flow from the Johns Hopkins turbulence database [237]. The data was generated via the pseudo-spectral method with 1024^3 nodes and utilized time steps of 0.0002 seconds. From the simulation, we select a $50 \times 50 \times 50$ node volume with stride length of 4 in all dimensions. We select 1,257 snapshots of this volume with temporal spacing of 0.024 seconds. Of these 1,257 snapshots, the first 100 are set aside to account for our use of a time-history of 100 sensor measurements, 900 of the remaining

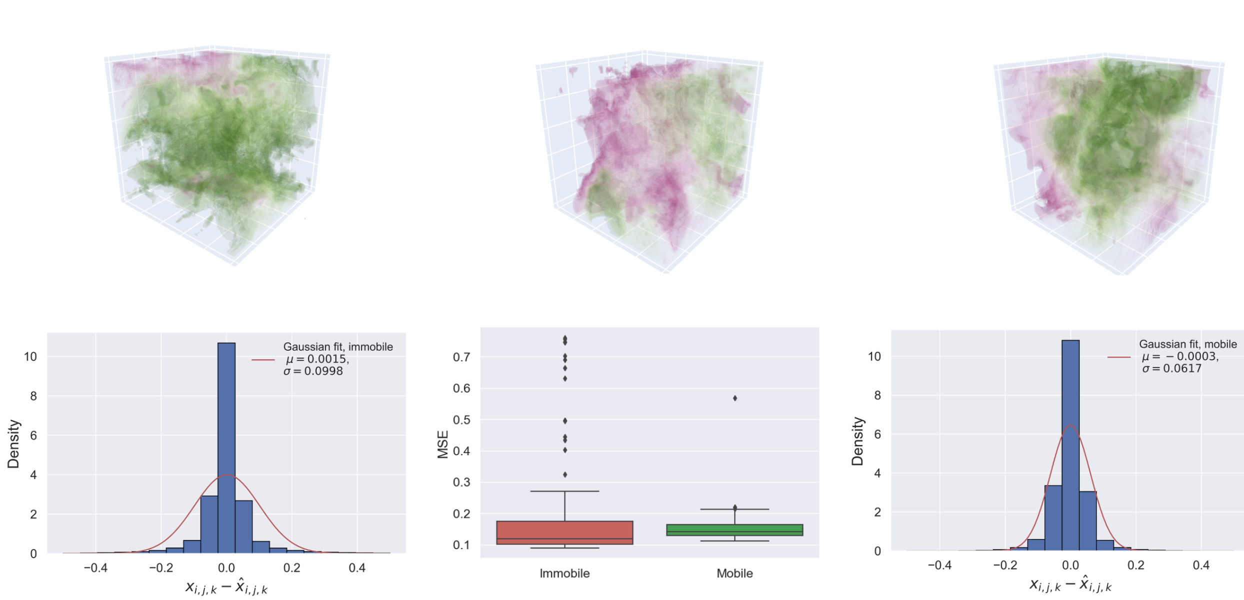

Figure 5.2: (Top) Example flow fields to be reconstructed from the test set. (Bottom) Histograms of the difference between ground truth and reconstruction across all nodes and samples in the test set for immobile sensors (left) and mobile sensors (right). Both distributions are approximately zero mean, but the variance for mobile sensors is lower. (Center) Box plot of the MSE evaluated across all samples in the test set for the 100 SHRED models with mobile and immobile sensors. While the median performance is similar, there are notably more outliers with poor performance in the case of immobile sensors.

snapshots are then randomly selected for training while the remainder are divided evenly between validation and test sets. In this example, we do not consider a strict, temporal partition of training and test data.

We train 100 SHRED models with one immobile, randomly placed sensors and 100 SHRED models with one mobile sensor following a modified random walk. Each random walk is initialized from the center of the considered volume and every three time-steps takes a one-node step in one of the three spatial dimensions, each with equal probability. With high probability, SHRED models with either mobile or immobile sensors accurately reconstruct the high-dimensional state as measured by MSE across the test set. However, the box plot in Fig. 5.2 shows that a greater proportion of models using mobile sensors

Sea surface temperature reconstruction error		
Route	MSE (random parition)	MSE (temporal partition)
Immobile sensors (3)	0.0279	0.0460
Immobile sensor (1)	0.0317	0.0481
Atlantic Ocean	0.0358	0.0442
Antarctica	0.0321	0.0516
USA West Coast	0.0321	0.0499
Atlantic Ocean + Antarctica	0.0395	0.1549
Atlantic Ocean + USA West Coast	0.0440	0.1349
Antarctica + USA West Coast	0.0328	0.1278
Atlantic Ocean + Antarctica + USA West Coast	0.0342	0.1345

Table 5.1: Mean-squared error for SHRED reconstructing sea-surface temperature for both randomly and temporaly partitioned training/test/validation data. Dynamic trajectories from mobile sensors in the Atlantic Ocean, Antarctic Ocean, and along the USA's West Coast were compared to one and three immobile sensors for reconstruction the complex spatio-temporal sea-surface temperature data [2].

achieve good performance in comparison to the models using immobile sensors. The histograms in Fig. 5.2 plot the deviation of the reconstruction from the ground truth at each node and across all samples in the test set. The estimates for both mobile and immobile sensors appear to be unbiased ($\mu = 0.0015$ immobile, $\mu = -0.0003$ mobile), but the variance of the mobile sensing reconstructions is less ($\sigma^2 = 0.0998^2$ immobile, $\sigma^2 = 0.0617^2$ mobile). Neither distribution is well approximated by a Gaussian, although the fit for mobile sensors is better.

5.3.2 Sea-surface temperature

For our second dataset, we evaluate sea-surface temperature (SST) as reported by NOAA [2]. This dataset includes the weekly mean sea-surface temperature from the years 1992 to 2019. Notably, NOAA produced the SST data using optimum interpolation with in situ and satellite data. The data comprises 1400 snapshots of a 180 by 360 grid; the exclu-

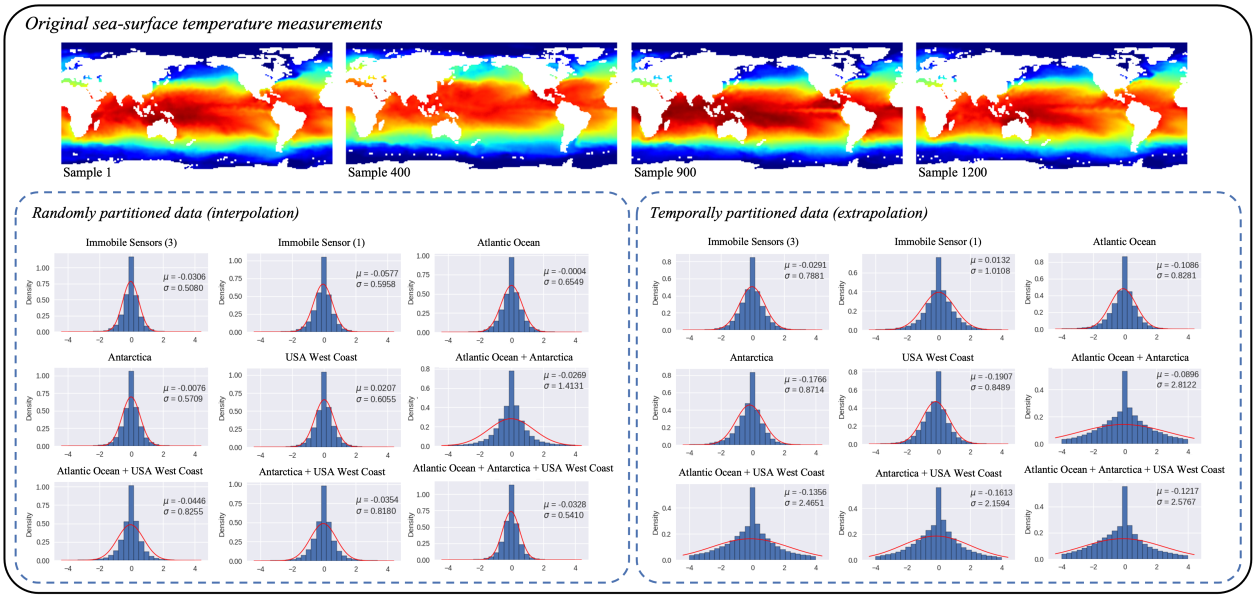

Figure 5.3: (Top) Example snapshots of global sea-surface temperature to be reconstructed from either the (i) randomly or (ii) temporally partitioned test set. (Bottom left) Histograms of the difference between ground truth and reconstructed states across all grid spaces and snapshots in the randomly-partitioned test set for immobile (1 and 3) and mobile sensors; dynamical trajectories of mobile sensors are one or more combinations of a year-long circuit in (i) the Atlantic Ocean, Antarctica Southern Ocean, and along the USA West Coast. (Bottom right) Histograms of the differences between ground truth and reconstructed states across all grid spaces and snapshots in the temporally-partitioned test set, again for immobile and mobile sensors; sensor positions and trajectories were the same as the randomly-partitioned dataset.

sion of landmass leaves 44,219 spatial locations corresponding to the sea surface. As in [206], we denote SHRED's input trajectory length to include 52 temporal measurements, resulting in one year of measurement time histories. The remaining 1348 snapshots are partitioned into training, test, and validation sets, comprised of 1000, 174, and 174 snapshots respectively. Regarding the dynamic trajectories of the mobile sensors, we consider three distinct moving sensors that complete a circuit over a year span: (i) an *Atlantic Ocean* route that travels from the east coast of the USA towards Europe and back, (ii) an out-and-back route near *Antarctica* Southern Ocean, and (iii) a route that follows the *West Coast of the USA*.

In Table 5.1, we compare how well each route – as well as their combinations – reconstruct SST, relative to one and three immobile sensors, for both random and temporal data partitions. For the random partition of training/test/validation data – an interpolation problem – we observe comparable mean-squared reconstruction errors across all mobile sensor routes relative to the immobile sensor reconstruction errors. For the temporal partition – an extrapolation problem – we see comparable errors for the single dynamic trajectory routes relative to the immobile sensors. The 2+ route combinations demonstrate much higher reconstruction errors; such results could be a potential indicator of overfitting, or even demonstrative of the challenges of extrapolating complex spatio-temporal data.

We additionally evaluated the mean and variance in the mean-squared reconstruction errors across routes and data-partitions, as seen in Fig. 5.3. For the randomly-partitioned data of global sea-surface temperature, immobile sensors are more biased than mobile sensors ($\mu = -0.0306$ 3 immobile sensors, $\mu = -0.0577$ 1 immobile sensor; across all mobile sensor routes, minimum $\mu = -0.0004$, maximum $\mu = 0.0446$, absolute mean $\mu = 0.0241$). However, the variance of the mobile versus immobile sensors are more variable ($\sigma^2 = 0.5080^2$ 3 immobile sensors, $\sigma^2 = 0.5958^2$ 1 immobile sensor; across all mobile sensor routes, minimum $\sigma^2 = 0.5410$, maximum $\sigma^2 = 1.4131$, mean $\sigma^2 = 0.7756$). For the temporally-partitioned data, immobile sensors are less biased than mobile sensors ($\mu = -0.0291$ 3 immobile sensors, $\mu = -0.0132$ 1 immobile sensor; across all mobile sensor routes, minimum $\mu = -0.0896$, maximum $\mu = 0.1907$, absolute mean $\mu = 0.1406$). We see comparable variance for the single dynamic trajectory routes relative to the immobile sensors; the 2+ route combinations have much higher error variance; ($\sigma^2 = 0.7881^2$ 3 immobile sensors, $\sigma^2 = 1.0108^2$ 1 immobile sensor; across all mobile sensor routes, minimum $\sigma^2 = 0.8281$, maximum $\sigma^2 = 2.8122$, mean $\sigma^2 = 1.7959$).

Marker-based Dataset results reconstructing kinematic states				
Type	Input Sensors	MSE (± s.d.) of Rotational Kinematics [°]		
		SHRED	SDN	Linear
Individual	3 random mobile sensors (transverse-plane pelvis rotation angle, medio-lateral pelvis position, right hip adduction angle)	0.056 ± 0.018	0.220 ± 0.097	0.419 ± 0.243
	3 non-random mobile sensors (right hip flexion angle, right knee flexion angle, right ankle dorsiflexion angle)	0.064 ± 0.027	0.153 ± 0.097	0.514 ± 0.328
	1 random mobile sensor (medio-lateral pelvis position)	0.080 ± 0.031	1.033 ± 0.440	1.096 ± 0.470
	1 non-random mobile sensor (right ankle dorsiflexion angle)	0.087 ± 0.042	0.607 ± 0.244	1.083 ± 0.705
Population	3 non-random mobile sensors (right hip flexion angle, right knee flexion angle, right ankle dorsiflexion angle)	0.098 ± 0.062	0.311 ± 0.168	0.944 ± 0.847
	1 non-random mobile sensor (right ankle dorsiflexion angle)	0.121 ± 0.073	1.208 ± 0.822	1.843 ± 2.110

Table 5.2: Results from the human biomechanics dataset [1]. SHRED with random and non-random measurement trajectory inputs is compared to a SDN and linear model for reconstructing kinematic states for individual-specific models and population-based models. Overall, SHRED successfully reconstructed kinematic states (rotational variables showed here) with low mean-squared error, far outperforming the other architectures. Translational kinematic variables (x, y, and z-direction pelvis position) were excluded from these MSE calculations, as rotational and translational variables have different units ([°] and [m], respectively; see Figs. 5.4 and 5.5 for all kinematic results.

5.3.3 Human biomechanics

The final dataset for which we leverage mobile sensor trajectories with SHRED is for capturing human biomechanics. Human motion tracking and analysis is essential for monitoring disease progression, guiding rehabilitation treatment, evaluating sports performance, and informing assistive device design. We use an open-source dataset that captures nondisabled human biomechanics during steady, rhythmic walking; such an approach can be generalized to other motion tracking such as robotic manipulation or computer animation. We use kinematics from 12 nondisabled adults (six female/six male; age 23.9 ± 1.8 years; height = 1.69 ± 0.10m; mass = 66.5 ± 11.7 kg) [1]. Participants walked at a self-selected speed (speed = 1.36 ± 0.11 m/s) on a split-belt instrumented treadmill

(Bertec Corp., Columbus, USA) for six minutes. While kinematics can be captured with multiple modalities, gold-standard marker data was recorded using a 10-camera optical motion capture system (Qualisys AB, Gothenburg, SE). Kinematics were estimated from marker data using the Inverse Kinematics algorithm in OpenSim 3.3 with a dynamically-constrained 19 degree-of-freedom skeletal model scaled to each participant [48, 157]. We evaluated the 18 kinematic states previously calculated in the open-source dataset which included three translational degrees of freedom at the pelvis, pelvis tilt/list/rotation, lumbar extension/bending, bilateral hip flexion/adduction/rotation, bilateral knee flexion, and bilateral ankle dorsiflexion. Kinematics were low-pass filtered at 6Hz using a fourth-order Butterworth filter [1]. An important distinction with this human biomechanics data, as opposed to the isotropic turbulence or sea-surface temperature, is that marker data recorded during motion capture is in a global coordinate frame, but through the standard procedures for marker data processing are transformed into kinematics in a relative frame of reference to the human subject.

Because this kinematic dataset cannot be compared to SHRED with immobile sensors, as the markers must move with the human, we compare the performance of mobile sensor trajectories and SHRED with that of a shallow decoder network (SDN) (*i.e.*, SHRED without leveraging time histories) and a linear model. Therefore, the aim of this example was to evaluate each modeling architecture's ability to learn a mapping between a sparse set of mobile sensors to reconstruct the full set of 18 kinematic states. We evaluate two conditions: (1) individual-specific models, in which a model is developed for each participant to reconstruct their respective set of full kinematics, and (2) population-based models, which uses data from all participants to reconstruct the full set of kinematics for an unseen individual, thus enabling rapid generalization to data outside the training set.

Overall, SHRED successfully maps between a sparse set of measurements to reconstruct the full biomechanical states for both the individual and population models. Additionally, SHRED far outperforms the linear and SDN architectures. While more input sensors generally increases reconstruction accuracy, performance with as little as 0.068 and 0.048 degrees (rotational variables) mean-squared error can be achieved with one and three sensors with SHRED, well below the accuracy and repeatability recommended even for clinical gait analyses ($2 \deg$ in sagittal-plane and $5 \deg$ in frontal-plane) [48, 238]. Table 5.2 provides a summary of the mean-squared errors for the rotational kinematic variables across modeling architectures and conditions. We take a closer look at each condition's results in the following sections:

Individual-specific biomechanics models

For the individual-specific biomechanics models, a mapping is trained for each of the 12 participants using a linear, SDN and SHRED architecture. Four combinations of dynamic trajectory inputs are tested: (1) three randomly chosen kinematic states (transverse-plane pelvis rotation angle, medio-lateral pelvis position, right hip adduction angle), (2) three non-randomly chosen kinematic states (right hip flexion angle, right knee flexion angle, right ankle dorsiflexion angle), (3) one randomly chosen kinematic state (medio-lateral pelvis position), and (4) one non-randomly chosen kinematic state (right ankle dorsiflexion angle). The non-randomly chosen states represent the kinematics most commonly used for biomechanical assessments, and which generally have the greatest accuracy and inter-session reliability (i.e, sagittal-plane kinematics) [238].

Fig. 5.4 shows the performance of SHRED relative to the SDN and linear model for reconstructing human motion across individuals. SHRED reconstructs the full set of 18 biomechanical variables more accurately that the shallow decoder and linear models, regardless of the number or choice of input sensors. Using three mobile sensors, SHRED

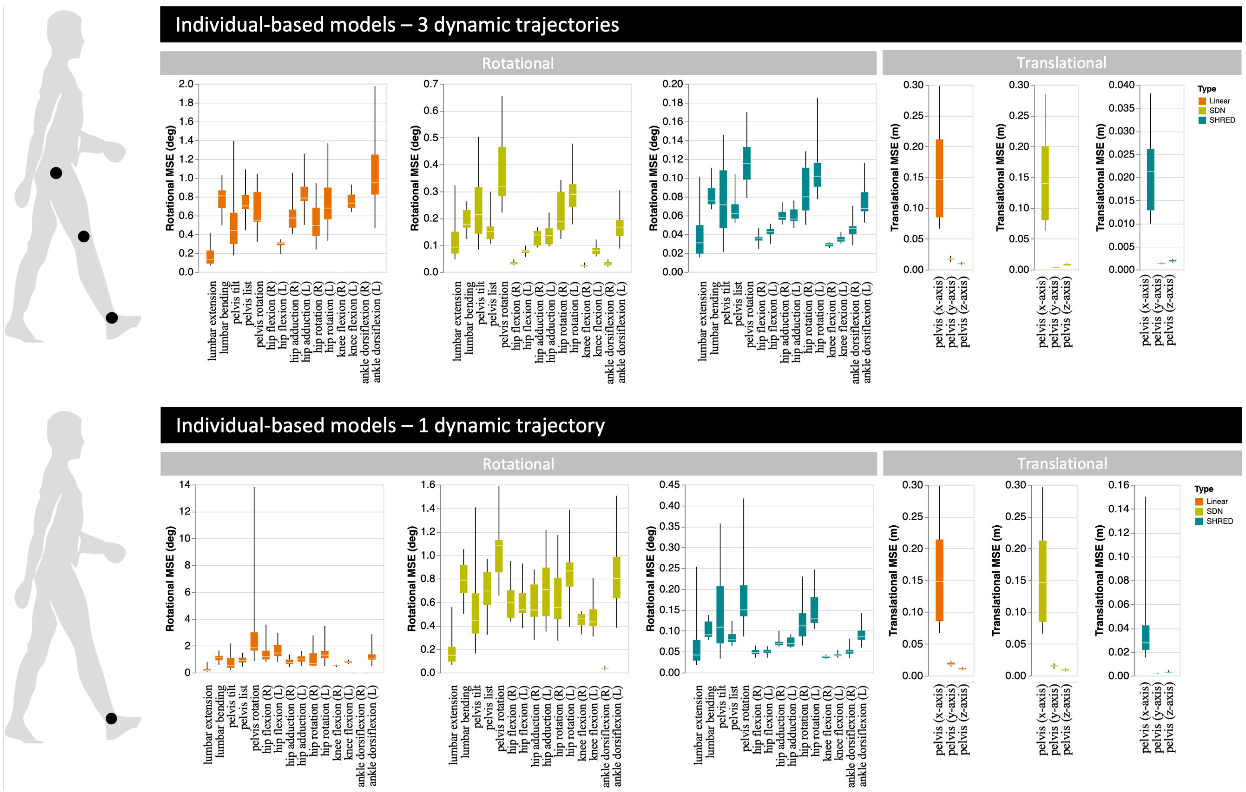

Figure 5.4: Visualization of individual-specific results with non-random sensor inputs for reconstructing human biomechanics [1]; see Table 5.2 for additional individual-specific results using random sensor inputs. A unique mapping was trained for each individual to transform their sparse set of sensor measurements to their full measurement data; each individual's mapping was evaluated by reconstructing their held-out test data. Performance was calculated using mean-square error (MSE); box plots show the aggregate MSE for all individuals and across kinematic states. Rotational and translational kinematic states are displayed separately, as rotational states use units of degrees while translational states use units of meters. Three modeling paradigms were trained to learn a mapping: linear regression (orange), a shallow decoder network (SDN) (green), and a shallow recurrent demcoder network (SHRED) (blue). Note the independent y-axes and respective scales for each graph. (Top) Three dynamic trajectories were purposefully chosen as sensor inputs to the reconstruction mapping: right hip flexion angle (degrees), right knee flexion angle (degrees), and right ankle dorsiflexion angle (degrees). (Bottom) One dynamic trajectory was purposefully chosen as sensor inputs to the reconstruction mapping: right ankle dorsiflexion angle (degrees).

has errors less than 0.064 ± 0.027 degrees (rotational variables), as much as 4.0x more accurate on average than the SDN and 8.0x more accurate than the linear model. Using a single mobile sensor, SHRED has errors less than 0.087 ± 0.042 degrees, as much as 12.8x

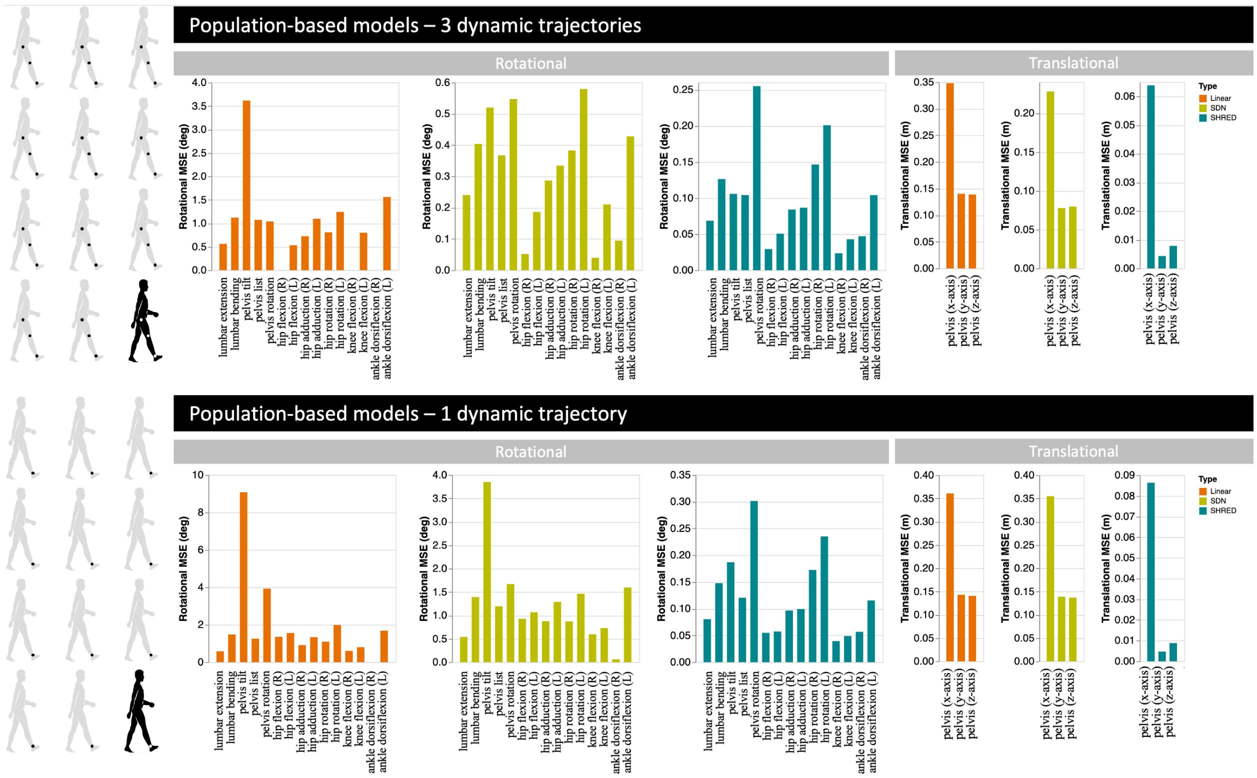

Figure 5.5: Visualization of population-based results for reconstructing human biomechanics [1]. The population models trained using 11 of the 12 individuals' data and were evaluated by reconstructing the held-out subject's full measurement data from their sparse input sensor(s); performance was calculated by using mean-square error (MSE). Rotational and translational kinematic states are displayed separately, as rotational states use units of degrees while translational states use units of meters. Three modeling paradigms were trained to learn the mapping: linear regression (orange), a shallow decoder neural network (green), and a shallow recurrent neural network (blue). Note the independent y-axes and respective scales for each graph. (Top) Three dynamic trajectories were purposefully chosen as sensor inputs to the reconstruction mapping: right hip flexion angle (degrees), right knee flexion angle (degrees), and right ankle dorsiflexion angle (degrees). (Bottom) One dynamic trajectory was purposefully chosen as sensor inputs to the reconstruction mapping: right ankle dorsiflexion angle (degrees).

more accurate on average than the SDN and 13.6x more accurate than the linear model.

Population-based biomechanics models

For the population-based biomechanics models, a reconstruction mapping was trained using 11 of the 12 participants' kinematic data, for a one-subject hold-out test using each architecture. Two combinations of dynamic trajectory inputs were tested: (1) three non-randomly chosen kinematic states (right hip flexion angle, right knee flexion angle, right ankle dorsiflexion angle), and (2) one non-randomly chosen kinematic state (right ankle dorsiflexion angle).

Fig. 5.5 shows the performance of SHRED relative to the SDN and linear model for reconstructing human motion within a population. SHRED far outperforms the SDN and linear model regardless of number or choice of input sensors. Using three mobile sensors, SHRED has errors less than 0.098 ± 0.062 degrees, which was on average 3.3x more accurate than the SDN and 9.6x more accurate than the linear model. Using a single mobile sensor, SHRED has errors less than 0.121 ± 0.073 degrees, which was on average 9.8x more accurate than the SDN and 14.8x more accurate than the linear model. These results demonstrate that SHRED can enable rapid generalization (parameterization of dynamics) for data outside the training set, which may be especially useful in human biomechanics where population-level models can be used to estimate parameters for a new individual (*e.g.*, a new patient visits a clinic for the first time).

5.4 Discussion & Conclusion

Sensing continues to be one of the most important tasks for science and engineering across disciplines. Machine learning methods are offering exceptional new paradigms for maximally exploiting sensor information for the diversity of tasks required of sensors, including reconstruction, forecasting, model discovery, control, and uncertainty quantification. Although there is a well established theory for sensing and sensor placement using immobile sensors, mobile sensing has remained exceptionally challenging with only recent and limited advancements made for leveraging the dynamical trajectories of the sensor. The mobile SHRED architecture advocated here provides a new paradigm for sensing whereby the time-history of the sensor is used to encode global information of the measured high-dimensional state space. SHRED not only allows for the embedding of the multiscale physics into a compact and low-dimensional latent space, it also provides through its decoder network a mapping from the minimal mobile sensors to full state estimates. Thus the SHRED architecture provides a nonlinear generalization of classical low-rank and linear embeddings such as the singular value decomposition.

The performance of SHRED is demonstrated on three challenging data sets which are all characterized by complex, nonlinear, and multiscale interactions. Thus the data considered represents some of the most challenging environments for sensing. Additionally, such applications often require the mobile sensing. Given the limited work on mobile sensing to date, along with the performance demonstrated by the method, the mobile SHRED architecture represents a significant innovation in mobile sensing technologies. Not only does the architecture reduce the variance in state estimation in comparison to immobile sensors, the decoder requires less data to train than typical networks due to its shallow structure. Empirically it is also observed that the mobile SHRED architecture is robust to hyper-parameter tuning, requiring little effort to train high performing models. The advantageous of such an architecture provide an application agnostic scheme which has significant potential for broad usage in science and engineering.

The mobile SHRED architecture works because the time history of the sensors embed global information of the system measured. For spatio-temporal data, for instance, the dynamics typically are governed by partial differential equations whose spatial derivative estimates rely on neighboring data points. These neighbors in turn rely on their neighbors and so forth. Thus a local spatial position is coupled to neighboring positions and ultimately globally by spatial derivatives. The sensor trajectories learn a representation of the spatio-temporal field via this global coupling. Of course, such an architecture would break down if there are regions in the dynamics that are statistically independent from each other, *i.e.* there are isolated dynamics that are not coupled in any way. There can also be "dead spots" in the measurement space, but the moving sensors can then typically move to new locations where information is captured. Thus with high-probability, a random sensor trajectory can be used to encode global information and estimate the full state space.

Finally, for the examples shown, there is clearly a requirement in the training data to have both the sensor trajectory data as well as the corresponding high-dimensional state space data. In many applications, there is no access to the high-dimensional state space. Thus proxy data, in the form of simulation data, can be used as a surrogate for the real data. As long as the proxy data is representative of the true data, it can be used to train the SHRED architecture. This gives a hybrid approach to training the neural network whereby simulation data can be used in partnership with real data to build a sensing model for producing otherwise inaccessible full state estimates.

Acknowledgements

JNK acknowledges support from the Air Force Office of Scientific Research (FA9550-19-1-0386). The authors also acknowledge support in part by the US National Science Foundation (NSF) AI Institute for Dynamical Systems (dynamicsai.org), grant 2112085.

Chapter 6
CONCLUSION

6.1 Summary

This dissertation presents foundational work developing methods for data-driven dynamical modeling to understand complex system behavior like human movement. This work was the first to establish discrepancy modeling as principled approach for resolving missing physics in dynamical systems, demonstrate that discrepancy modeling is a useful framework for characterizing individual exoskeleton responses, and leverage dynamical sensor trajectories with shallow recurrent decoder networks to reconstruct full-state estimates, such as for expanding biomechanical datasets to unmeasured/unavailable quantities.

The first objective of this dissertation was to resolve a missing physics discrepancy between the complex system measurements and a dynamical model's approximations for improved system characterization. Physics-based and first-principles models pervade the engineering and physical sciences, allowing for the ability to model the dynamics of complex systems with limited accuracy. Chapter 3 introduces a comparative discrepancy modeling framework to estimate the missing physics using data-driven model discovery methods by (i) learning a model for the evolution of systematic state-space residual, or (ii) discovering a model for the deterministic dynamical error. The results of this study demonstrated that, by construction, resolving a missing physics mismatch is best treated by learning the discrepancy model in the dynamical space. As data-driven modeling continues to gain momentum, it is imperative that researchers utilize domain knowledge (*e.g.*, first principles physics) to model complex systems; by focusing explicitly on the discrepancy between measured and modeled dynamics, this framework shifts the view of discrepancies as 'errors' or 'residuals' to highly valuable measures for model improvement and scientific insight. Code has been made freely-available for researchers (github.com/meganebers/Discrepancy-Modeling-Framework-code).

The second objective of this dissertation demonstrated that discrepancy modeling is a useful framework for quantifying how unique individuals respond to passive-elastic ankle exoskeletons (Chapter 4). Physiological responses to mechanical assistance from wearable devices (exoskeletons, robots, etc) are highly variable, especially for clinical populations. Neural network-based discrepancy models were able capture the complex gait changes with exoskeleton assistance for nondisabled adults without prior knowledge of physiology or motor control. Because individual response were shown to be predictable, the response dynamics can be used to optimize the prescription process by tuning device properties (*e.g.*, mechanical torque, control strategy) and uncover underlying mechanisms driving gait changes with wearable devices. However, additional biomechanical modalities (*e.g.*, muscle-tendon behavior) and/or improved resolution (*e.g.*, ultrasound) will be needed to capture muscle-level changes during walking with ankle exoskeleton assistance. Discrepancy modeling is a unique and innovative tool that complements current biomechanical modeling approaches and may accelerate the discovery of individual-specific mechanisms driving responses to exoskeletons, other assistive devices, and clinical interventions.

The final objective of this dissertation was to combine deep learning, time-delay embedding, and sparse sensing to reconstruct full-state estimates, with exciting applications for personalized human movement tracking. Human motion tracking and analysis is essential for monitoring disease progression, guiding rehabilitation treatment, evaluating sports performance, and informing assistive device design. In this dissertation, we leveraged the time histories (trajectories) of sensor measurements to expand biomechanical datasets (for both individuals and a population) to unmeasured or unavailable quantities. The results from this study demonstrated that a shallow decoder network can reconstruct the full set of states from a sparse set of sensors for three challenging datasets: forced isotropic turbulence, global sea-surface temperature, and human biomechanics. One of the most important outcomes of this work is demonstrating that dynamical deep learning can facilitate personalized tracking of individual and population-based human motion while ensuring a comprehensive set of measurements is collected with sparse sensing. The pipeline will enable the collection of a rich breadth of biomechanical data in a natural environment without imposing a burden to the user.

6.2 Future Work

The work presented in this dissertation demonstrates that data-driven dynamical models are a powerful tool for understanding human movement. The results of these studies suggest important and interesting future research directions in both human movement applications and data-driven modeling methods. The following sections outline avenues of future work that may expand our ability to analyze, predict, and control complex dynamical systems, identify physiological mechanisms that underlie individual responses to device prescription, and allow for rich biomechanical datasets to be collected for individuals in their natural environment:

6.2.1 Data-driven dynamical models from time-series data

- **Discrepancy modeling for complex dynamical systems**

  Discrepancy modeling for dynamical systems is in its infancy; the work in this dissertation is one of the first to comprehensively investigate how one should handle missing physics in practical applications, *i.e.*, a deterministic dynamical error. Further, it provides a first-step towards a principled approach guiding researchers/engineers to consider common challenges when modeling dynamical systems and errors in practical engineering applications. With this work being early and extensive, it is not exhaustive; there are many open questions yet unresolved within the framework of discrepancy modeling for dynamical systems, such as observation error, partially observed states, initial condition errors, stochastic processes, non-Gaussian distributions, and among others we may not be aware of yet. For example, the systematic state-space approach could be used to address observation error, a phenomena that occurs when measuring the system changes the state. Identifying other types of discrepancies responsible for model-measurement mismatch, beyond missing physics, will provide a principled and comprehensive hybrid modeling (mechanism+data) framework for improved dynamical system characterization, control, and simulation.

6.2.2 *Identifying physiological mechanisms underlying individual exoskeleton responses*

- **What additional modalities are needed to explain muscle-level changes in response to exoskeleton prescription?**

  The work in this dissertation highlights the challenges in understanding individual differences in motor control and exoskeleton responses from traditional biomechanical measurements alone. Understanding how we can capture these important differences in sensorimotor integration and control represents an exciting and challenging area for future work. Further, utilizing the discrepancy modeling framework, other outcomes could be predicted, such as joint moments, to directly translate to device design. Discrepancy modeling can offer a framework to evaluate whether and how new data streams, outcomes, or methods can explain and provide insight into these processes.

- **Are there interpretable mechanisms underlying individual exoskeleton responses?**

  The results from this dissertation demonstrated that exoskeleton response dynamics can be isolated. Neural networks as function approximators (*i.e.*, curve fitters) were trained to learn the response dynamics for each participant. However, neural networks do not offer insight into mechanisms underlying response dynamics. Advanced techniques in machine learning for discovering physically or physiologically interpretable and/or physics-constrained data-driven models can uncover meaningful (and potentially physiologically-ground) mechanisms underlying exoskeleton responses. Such methods to learn more interpretable models of response include the sparse identification of nonlinear dynamics [7] and dynamical mode decomposition [6, 104].

- **Are exoskeleton responses predictable and organized for more complicated scenarios?**

  Individual responses to exoskeletons are highly heterogeneous, therefore making it challenging to predict how exoskeleton forces and torques acting on the body alter an individual's biomechanics, motor control, and sensory feedback. In this dissertation, we used discrepancy modeling to quantify the simple, stereotypical responses (*i.e.*, evaluating gait changes in young nondisabled adults with bilateral passive-elastic ankle exoskeletons). An important area of future research is evaluating whether discrepancy modeling reveals that exoskeleton responses are predictable and organized for other types of more complicated assistive devices (*e.g.*, active exoskeletons that can apply assistance/resistance) or for populations that exhibit highly heterogeneous exoskeleton responses (*e.g.*, aging, CP, and stroke). However, additional population- or individual-specific factors that may influence the ability to encode discrepancies with assistive devices include noisy measurements due to soft tissue artifact or impedance between the muscle and EMG sensor, poor selective control of muscle excitation, or reduced automaticity of gait [239] with age or impairment.

- **Are there sub-groups of responders?**

  An important next step in evaluating individual responses to exoskeletons is evaluating if there are sub-groups of responders. This is akin to the Rozumalski 2009 paper that identified 5 main crouch gait patterns as defined using k-means cluster analysis (and how each gait pattern related to underlying clinical pathology). By using a similar classification methodology, future work could seek to identify whether groups of similar responders emerged and identify physically- or physiologically-interpretable mechanisms that underlie each response group. In this dissertation, we had a relatively limited number of participants, all who had similar responses. Extending these methods to a broader set of participants, exoskeletons that provide a larger perturbation (*e.g.*, powered exoskeletons) or a clinical population would provide the foundational dataset for investigating this question.

6.2.3 *Expansion of biomechanical datasets*

- **Can a shallow recurrent decoder network (SHRED) be used to reconstruct biomechanical datasets for cyclic/rhythmic activities other than walking, such as running, stair climbing, and biking?**

A dynamical systems framing can capture the nonlinear, high-dimensional, multi-scale, and complex characteristics of the human body. Further, dynamical systems are well-suited for representing periodicity, as observed in gait. For example, rhythmic gait is easily observed during walking on a fixed-speed treadmill, such that the current state of the system can be assumed to inform the temporal evolution of states. Research in motor control of human locomotion hypothesizes that locomotion is accomplished by groups of muscles co-activated by the central nervous systems. Therefore, we hypothesize that SHRED for biomechanical dataset expansion can be extended to other rhythmic tasks besides walking, such as running, biking, and stair climbing [1, 142, 167, 240–243].

- **Can we learn reconstruction mappings for nondisabled adults navigating unstructured environments or non-steady state actives?**

  We previously postulated that a dynamical systems framework is well suited for capturing the periodicity of rhythmic locomotion like walking or running. However, humans do not generally walk at constant speed; walking often involves starting, stopping and changing speeds [244]. Monitoring human motion 'in the wild' is vital for observing individuals' natural functionality and lifestyle, which includes non-steady state walking, a challenge not yet overcome in free-living environments [245]. Therefore, investigating whether SHRED can reconstruct a full kinematic dataset from a sparse set of sensors during non-periodic motion.

- **Can we learn unique and population-based mappings for clinical populations, such as Parkinson's or stroke?**

  The capacity for mobility is reflective of the physiological processes that support an individual's health. Therefore, tracking movement to quantify, monitor, and support movement-based health outcomes across the lifespan is paramount. We have demonstrated we can reconstruct a full biomechanical dataset from a sparse set of sensors for nondisabled adults. Yet, it is unknown if this method can be deployed on a clinical population, for example with pathological or aging gait, and show correlation with a meaningful clinical measure [246–249].

- **Can we extend this methodology beyond movement-to-movement mapping?**

  Extending this methodological approach beyond movement-to-movement mapping has potential to provide meaningful insight into movement-based health outcomes, as mobility is reflective of the physiological processes supporting an individual's health, such as cardio-respiratory fitness, neurologic plasticity and maintenance, and socio-emotional engagement. Learning a reconstruction mapping to estimate physiological data, such as metabolic cost or oxygen consumption, directly from movement-based measurements is an interesting and important future application of SHRED.

- **Does a shallow recurrent decoder network (SHRED) enable forecasting for biomechanical applications?**

  Leveraging the time histories of sensor measurements appears most compelling for the *reconstruction* of biomechanical datasets for expansion to unmeasured/unmodeled quantities. However, this method could be used for *forecasting* the states of the full datasets. Forecasting in biomechanical and robotic applications is important for control, such as with prosthesis, robot, or exoskeleton devices. Utilizing an expanded dataset for informing control strategies could improve the robustness of future state and trajectory prediction accuracy.

The long term goal of the work presented in this dissertation and the proposed studies outlined above is to develop models of and extract insight from time-series data of complex systems, such as the human musculoskeletal system. The utilization of machine learning and data science in human biomechanics has only just begun; data-driven methods for dynamical models of complex systems has the opportunity fundamentally change how we investigate pathology, enhance mobility, and personalize rehabilitation.

BIBLIOGRAPHY

[1] Michael C Rosenberg, Bora S Banjanin, Samuel A Burden, and Katherine M Steele. Predicting walking response to ankle exoskeletons using data-driven models. *Journal of the Royal Society Interface*, 17(171):20200487, 2020.

[2] Richard W Reynolds, Nick A Rayner, Thomas M Smith, Diane C Stokes, and Wanqiu Wang. An improved in situ and satellite sst analysis for climate. *Journal of climate*, 15(13):1609–1625, 2002.

[3] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, 13(1):A3–A270, 1890.

[4] Henri Poincaré. *Les méthodes nouvelles de la mécanique céleste*, volume 3. Gauthier-Villars et fils, imprimeurs-libraires, 1899.

[5] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

[6] Jonathan H Tu, Clarence W Rowley, Dirk M Luchtenburg, Steven L Brunton, and J Nathan Kutz. On dynamic mode decomposition: Theory and applications. *arXiv preprint arXiv:1312.0041*, 2013.

[7] Steven L Brunton, Joshua L Proctor, and J Nathan 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.

[8] Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

[9] Ariana Mendible, Steven L Brunton, Aleksandr Y Aravkin, Wes Lowrie, and J Nathan Kutz. Dimensionality reduction and reduced-order modeling for traveling wave physics. *Theoretical and Computational Fluid Dynamics*, 34:385–400, 2020.

[10] Steven L Brunton, Joshua L Proctor, Jonathan H Tu, and J Nathan Kutz. Compressed sensing and dynamic mode decomposition. *Journal of computational dynamics*, 2(2):165–191, 2016.

[11] Philip Hartman. Ordinary differential equations, classics in applied mathematics, vol. 38, society for industrial and applied mathematics (siam), philadelphia, pa, 2002, corrected reprint of the second (1982) edition. *Corrected reprint of the second*, 1982.

[12] Philip Hartman. *Ordinary differential equations*. SIAM, 2002.

[13] Eni Halilaj, Apoorva Rajagopal, Madalina Fiterau, Jennifer L Hicks, Trevor J Hastie, and Scott L Delp. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. *Journal of biomechanics*, 81:1–11, 2018.

[14] Andreas Daffertshofer, Claudine JC Lamoth, Onno G Meijer, and Peter J Beek. Pca in studying coordination and variability: a tutorial. *Clinical biomechanics*, 19(4):415–428, 2004.

[15] KJ Deluzio and JL Astephen. Biomechanical features of gait waveform data associated with knee osteoarthritis: an application of principal component analysis. *Gait & posture*, 25(1):86–93, 2007.

[16] Orna A Donoghue, Andrew J Harrison, Norma Coffey, and Kevin Hayes. Functional data analysis of running kinematics in chronic achilles tendon injury. *Medicine and science in sports and exercise*, 40(7):1323–1335, 2008.

[17] Willie Ryan, Andrew Harrison, and Kevin Hayes. Functional data analysis of knee joint kinematics in the vertical jump. *Sports Biomechanics*, 5(1):121–138, 2006.

[18] Adam Rozumalski and Michael H Schwartz. Crouch gait patterns defined using k-means cluster analysis are related to underlying clinical pathology. *Gait & posture*, 30(2):155–160, 2009.

[19] Joarder Kamruzzaman and Rezaul K Begg. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Transactions on Biomedical Engineering*, 53(12):2479–2490, 2006.

[20] Katarzyna Kaczmarczyk, Andrzej Wit, Maciej Krawczyk, and Jacek Zaborski. Gait classification in post-stroke patients using artificial neural networks. *Gait & posture*, 30(2):207–210, 2009.

[21] Dwaipayan Biswas, Andy Cranny, Nayaab Gupta, Koushik Maharatna, Josy Achner, Jasmin Klemke, Michael Jöbges, and Steffen Ortmann. Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification. *Human movement science*, 40:59–76, 2015.

[22] NA Capela, ED Lemaire, N Baddour, M Rudolf, N Goljar, and H Burger. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. *Journal of neuroengineering and rehabilitation*, 13(1):1–10, 2016.

[23] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and

Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4):1476, 2022.

[24] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, 15(12):31314–31338, 2015.

[25] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1254, 2018.

[26] Gary M Weiss, Jessica L Timko, Catherine M Gallagher, Kenichi Yoneda, and Andrew J Schreiber. Smartwatch-based activity recognition: A machine learning approach. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 426–429. IEEE, 2016.

[27] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 105:233–261, 2018.

[28] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[29] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[30] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

[31] Pedro Manuel Santos Ribeiro, Ana Clara Matos, Pedro Henrique Santos, and Jaime S Cardoso. Machine learning improvements to human motion tracking with imus. *Sensors*, 20(21):6383, 2020.

[32] Mitja Luštrek and Boštjan Kaluža. Fall detection and activity recognition with machine learning. *Informatica*, 33(2), 2009.

[33] Guto Leoni Santos, Patricia Takako Endo, Kayo Henrique de Carvalho Monteiro, Elisson da Silva Rocha, Ivanovitch Silva, and Theo Lynn. Accelerometer-based human fall detection using convolutional neural networks. *Sensors*, 19(7):1644, 2019.

[34] Faisal Hussain, Fawad Hussain, Muhammad Ehatisham-ul Haq, and Muhammad Awais Azam. Activity-aware fall detection and recognition based on wearable sensors. *IEEE Sensors Journal*, 19(12):4528–4536, 2019.

[35] Ta-Sen Wei, Peng-Ta Liu, Liang-Wey Chang, and Sen-Yung Liu. Gait asymmetry, ankle spasticity, and depression as independent predictors of falls in ambulatory stroke patients. *PloS one*, 12(5):e0177136, 2017.

[36] Eric Rapp, Soyong Shin, Wolf Thomsen, Reed Ferber, and Eni Halilaj. Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. *Journal of Biomechanics*, 116:110229, 2021.

[37] Panagiotis Tsinganos and Athanassios Skodras. On the comparison of wearable

sensor data fusion to a single sensor machine learning technique in fall detection. *Sensors*, 18(2):592, 2018.

[38] Henry Friday Nweke, Ying Wah Teh, Uzoma Rita Alo, and Ghulam Mujtaba. Analysis of multi-sensor fusion for mobile and wearable sensor based human activity recognition. In *Proceedings of the international conference on data processing and applications*, pages 22–26, 2018.

[39] Rachel C King, Emma Villeneuve, Ruth J White, R Simon Sherratt, William Holderbaum, and William S Harwin. Application of data fusion techniques and technologies for wearable health monitoring. *Medical engineering & physics*, 42:1–12, 2017.

[40] Verne Thompson Inman, Henry James Ralston, and Frank Todd. *Human walking*. Williams & Wilkins, 1981.

[41] Bart van Veen, Erica Montefiori, Luca Modenese, Claudia Mazzà, and Marco Viceconti. Muscle recruitment strategies can reduce joint loading during level walking. *Journal of biomechanics*, 97:109368, 2019.

[42] Casey A Myers, Peter J Laz, Kevin B Shelburne, Dana L Judd, Joshua D Winters, Jennifer E Stevens-Lapsley, and Bradley S Davidson. Simulated hip abductor strengthening reduces peak joint contact forces in patients with total hip arthroplasty. *Journal of biomechanics*, 93:18–27, 2019.

[43] Michael J Decker, Michael R Torry, Thomas J Noonan, William I Sterett, and J Richard Steadman. Gait retraining after anterior cruciate ligament reconstruction. *Archives of physical medicine and rehabilitation*, 85(5):848–856, 2004.

[44] John R Rebula, Lauro V Ojeda, Peter G Adamczyk, and Arthur D Kuo. Measure-

ment of foot placement and its variability with inertial sensors. *Gait & posture*, 38(4):974–980, 2013.

[45] Timothy E Hewett, Gregory D Myer, Kevin R Ford, Robert S Heidt Jr, Angelo J Colosimo, Scott G McLean, Antonie J Van den Bogert, Mark V Paterno, and Paul Succop. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *The American journal of sports medicine*, 33(4):492–501, 2005.

[46] Elena Ceseracciu, Zimi Sawacha, and Claudio Cobelli. Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: proof of concept. *PloS one*, 9(3):e87640, 2014.

[47] Apoorva Rajagopal, Łukasz Kidziński, Alec S McGlaughlin, Jennifer L Hicks, Scott L Delp, and Michael H Schwartz. Estimating the effect size of surgery to improve walking in children with cerebral palsy from retrospective observational clinical data. *Scientific reports*, 8(1):1–11, 2018.

[48] Jennifer L Hicks, Thomas K Uchida, Ajay Seth, Apoorva Rajagopal, and Scott L Delp. Is my model good enough? best practices for verification and validation of musculoskeletal models and simulations of movement. *Journal of biomechanical engineering*, 137(2), 2015.

[49] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.

[50] Shervin Bagheri. Effects of weak noise on oscillating flows: Linking quality factor, floquet modes, and koopman spectrum. *Physics of Fluids*, 26(9):094104, 2014.

[51] 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.

[52] 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(1):1–9, 2017.

[53] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

[54] Floris Van Van Breugel, J Nathan Kutz, and Bingni W Brunton. Numerical differentiation of noisy data: A unifying multi-objective optimization framework. *IEEE Access*, 8:196865–196877, 2020.

[55] Georg A Gottwald and Sebastian Reich. Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *Physica D: Nonlinear Phenomena*, 423:132911, 2021.

[56] Yuming Chen, Daniel Sanz-Alonso, and Rebecca Willett. Autodifferentiable ensemble kalman filters. *SIAM Journal on Mathematics of Data Science*, 4(2):801–833, 2022.

[57] Alban Farchi, Patrick Laloyaux, Massimo Bonavita, and Marc Bocquet. Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, 147(739):3067–3084, 2021.

[58] Rick Chartrand. Numerical differentiation of noisy, nonsmooth data. *International Scholarly Research Notices*, 2011, 2011.

[59] Jane Cullum. Numerical differentiation and regularization. *SIAM Journal on numerical analysis*, 8(2):254–265, 1971.

[60] Alexander Ramm and Alexandra Smirnova. On stable numerical differentiation. *Mathematics of computation*, 70(235):1131–1153, 2001.

[61] Kathleen Champion, Peng Zheng, Aleksandr Y Aravkin, Steven L Brunton, and J Nathan Kutz. A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access*, 8:169259–169271, 2020.

[62] Kumpati S Narendra and Kannan Parthasarathy. Neural networks and dynamical systems. *International Journal of Approximate Reasoning*, 6(2):109–131, 1992.

[63] Raul González-García, Ramiro Rico-Martìnez, and Ioannis G Kevrekidis. Identification of distributed parameter systems: A neural net based approach. *Computers & chemical engineering*, 22:S965–S968, 1998.

[64] K Krischer, R Rico-Martínez, IG Kevrekidis, HH Rotermund, G Ertl, and JL Hudson. Model identification of a spatiotemporally varying catalytic reaction. *AIChE Journal*, 39(1):89–98, 1993.

[65] R Rico-Martinez, JS Anderson, and IG Kevrekidis. Continuous-time nonlinear signal processing: a neural network based approach for gray box identification. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, pages 596–605. IEEE, 1994.

[66] Isaac E Lagaris, Aristidis Likas, and Dimitrios I Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.

[67] Arka Daw, Anuj Karpatne, William Watkins, Jordan Read, and Vipin Kumar. Physics-guided neural networks (pgnn): An application in lake temperature modeling. *arXiv preprint arXiv:1710.11431*, 2017.

[68] Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

[69] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[70] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.

[71] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[72] Chia-Ch'iao Lin and Lee A Segel. *Mathematics applied to deterministic problems in the natural sciences*. SIAM, 1988.

[73] 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.

[74] J Nathan Kutz. Advanced differential equations: Asymptotics & perturbations. *arXiv preprint arXiv:2012.14591*, 2020.

[75] Jared L Callaham, James V Koch, Bingni W Brunton, J Nathan Kutz, and Steven L Brunton. Learning dominant physical processes with data-driven balance models. *Nature communications*, 12(1):1–10, 2021.

[76] Joseph Bakarji, Jared Callaham, Steven L Brunton, and J Nathan Kutz. Dimensionally consistent learning with buckingham pi. *Nature Computational Science*, pages 1–11, 2022.

[77] Matteo Saveriano, Yuchao Yin, Pietro Falco, and Dongheui Lee. Data-efficient control policy search using residual dynamics learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4709–4715. IEEE, 2017.

[78] John Harlim, Shixiao W Jiang, Senwei Liang, and Haizhao Yang. Machine learning for prediction with missing dynamics. *Journal of Computational Physics*, 428:109922, 2021.

[79] Guanya Shi, Xichen Shi, Michael O'Connell, Rose Yu, Kamyar Azizzadenesheli, Animashree Anandkumar, Yisong Yue, and Soon-Jo Chung. Neural lander: Stable drone landing control using learned dynamics. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9784–9790. IEEE, 2019.

[80] Matthew E Levine and Andrew M Stuart. A framework for machine learning of model error in dynamical systems. *arXiv preprint arXiv:2107.06658*, 2021.

[81] Karl Johan Åström and Richard M Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton university press, 2021.

[82] Norman S Nise. *Control systems engineering*. John Wiley & Sons, 2020.

[83] Farid Golnaraghi and Benjamin C Kuo. *Automatic control systems*. McGraw-Hill Education, 2017.

[84] Katsuhiko Ogata. System dynamics. *Englewood Cliffs*, 1978.

[85] Katsuhiko Ogata et al. *Modern control engineering*, volume 5. Prentice hall Upper Saddle River, NJ, 2010.

[86] Kadierdan Kaheman, Eurika Kaiser, Benjamin Strom, J Nathan Kutz, and Steven L Brunton. Learning discrepancy models from experimental data. *arXiv preprint arXiv:1909.08574*, 2019.

[87] Cx K Batchelor and GK Batchelor. *An introduction to fluid dynamics*. Cambridge university press, 2000.

[88] Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[89] Robert M Wald. *General relativity*. University of Chicago press, 2010.

[90] Brian M de Silva, David M Higdon, Steven L Brunton, and J Nathan Kutz. Discovery of physics from data: universal laws and discrepancies. *Frontiers in artificial intelligence*, 3:25, 2020.

[91] Philip R Bevington and D Keith Robinson. Data reduction and error analysis. *McGraw-Hill, New York*, 2003.

[92] John Taylor. *Introduction to error analysis, the study of uncertainties in physical measurements*. 1997.

[93] R Dennis Cook and Sanford Weisberg. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.

[94] David R Cox and E Joyce Snell. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265, 1968.

[95] George EP Box and David A Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.

[96] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.

[97] Kody Law, Andrew Stuart, and Kostas Zygalakis. Data assimilation. *Cham, Switzerland: Springer*, 214, 2015.

[98] Andrew C Miller, Nicholas J Foti, and Emily B Fox. Breiman's two cultures: You don't have to choose sides. *Observational Studies*, 7(1):161–169, 2021.

[99] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. Siam, 2013.

[100] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[101] Brian de Silva, Kathleen Champion, Markus Quade, Jean-Christophe Loiseau, J Kutz, and Steven Brunton. Pysindy: A python package for the sparse identification of nonlinear dynamical systems from data. 2020.

[102] Laura P Swiler, Mamikon Gulian, Ari L Frankel, Cosmin Safta, and John D Jakeman. A survey of constrained gaussian process regression: Approaches and implementation challenges. *Journal of Machine Learning for Modeling and Computing*, 1(2), 2020.

[103] Kadierdan Kaheman, Steven Brunton, and J Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *Machine Learning: Science and Technology*, 2022.

[104] Diya Sashidhar and J Nathan Kutz. Bagging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification. *Philosophical Transactions of the Royal Society A*, 380(2229):20210199, 2022.

[105] Urban Fasel, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *arXiv preprint arXiv:2111.10992*, 2021.

[106] Geir Evensen. Advanced data assimilation for strongly nonlinear dynamics. *Monthly weather review*, 125(6):1342–1354, 1997.

[107] Pierre Gauthier. Chaos and quadri-dimensional data assimilation: A study based on the lorenz model. *Tellus A: Dynamic Meteorology and Oceanography*, 44(1):2–17, 1992.

[108] Robert N Miller, Michael Ghil, and Francois Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of Atmospheric Sciences*, 51(8):1037–1056, 1994.

[109] Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer Cham, 1st edition edition, 2015.

[110] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.

[111] Fei Tao, He Zhang, Ang Liu, and Andrew YC Nee. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 15(4):2405–2415, 2018.

[112] David Jones, Chris Snider, Aydin Nassehi, Jason Yon, and Ben Hicks. Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29:36–52, 2020.

[113] N Benjamin Erichson, Lionel Mathelin, Zhewei Yao, Steven L Brunton, Michael W Mahoney, and J Nathan Kutz. Shallow neural networks for fluid flow reconstruction with limited sensors. *Proceedings of the Royal Society A*, 476(2238):20200097, 2020.

[114] Douglas W Carter, Francis De Voogt, Renan Soares, and Bharathram Ganapathisubramani. Data-driven sparse reconstruction of flow over a stalled aerofoil using experimental data. *Data-Centric Engineering*, 2:e5, 2021.

[115] Shervin Sahba, Christopher C Wilcox, Austin McDaniel, Benjamin D Shaffer, Steven L Brunton, and J Nathan Kutz. Wavefront sensor fusion via shallow decoder neural networks for aero-optical predictive control. In *Interferometry XXI*, volume 12223, pages 11–17. SPIE, 2022.

[116] Seth M Hirsh, Sara M Ichinaga, Steven L Brunton, J Nathan Kutz, and Bingni W Brunton. Structured time-delay models for dynamical systems with connections to frenet–serret frame. *Proceedings of the Royal Society A*, 477(2254):20210097, 2021.

[117] Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *arXiv preprint arXiv:2201.05136*, 2022.

[118] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

[119] Julien Brajard, Alberto Carrassi, Marc Bocquet, and Laurent Bertino. Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A*, 379(2194):20200086, 2021.

[120] Merel-Anne Brehm, Jaap Harlaar, and Michael Schwartz. Effect of ankle-foot orthoses on walking efficiency and gait in children with cerebral palsy. *Journal of rehabilitation medicine*, 40(7):529–534, 2008.

[121] DÉSIRÉE Maltais, ODED Bar-Or, VICTORIA Galea, and MICHAEL Pierrynowski. Use of orthoses lowers the o (2) cost of walking in children with spastic cerebral palsy. *Medicine and science in sports and exercise*, 33(2):320–325, 2001.

[122] Juanjuan Zhang, Pieter Fiers, Kirby A Witte, Rachel W Jackson, Katherine L Poggensee, Christopher G Atkeson, and Steven H Collins. Human-in-the-loop optimization of exoskeleton assistance during walking. *Science*, 356(6344):1280–1284, 2017.

[123] Emily M McCain, Taylor JM Dick, Tracy N Giest, Richard W Nuckols, Michael D Lewek, Katherine R Saul, and Gregory S Sawicki. Mechanics and energetics of post-stroke walking aided by a powered ankle exoskeleton with speed-adaptive myoelectric control. *Journal of neuroengineering and rehabilitation*, 16(1):1–12, 2019.

[124] Benjamin C Conner, Jason Luque, and Zachary F Lerner. Adaptive ankle resistance from a wearable robotic device to improve muscle recruitment in cerebral palsy. *Annals of biomedical engineering*, 48(4):1309–1321, 2020.

[125] Steven H Collins, M Bruce Wiggin, and Gregory S Sawicki. Reducing the energy cost of human walking using an unpowered exoskeleton. *Nature*, 522(7555):212–215, 2015.

[126] Ye Ding, Myunghee Kim, Scott Kuindersma, and Conor J Walsh. Human-in-the-loop optimization of hip assistance with a soft exosuit during walking. *Science robotics*, 3(15):eaar5438, 2018.

[127] Jeffrey R Koller, Daniel A Jacobs, Daniel P Ferris, and C David Remy. Learning to walk with an adaptive gain proportional myoelectric controller for a robotic ankle exoskeleton. *Journal of neuroengineering and rehabilitation*, 12(1):1–14, 2015.

[128] Philippe Malcolm, Wim Derave, Samuel Galle, and Dirk De Clercq. A simple exoskeleton that assists plantarflexion can reduce the metabolic cost of human walking. *PloS one*, 8(2):e56137, 2013.

[129] Katherine M Steele, Adam Rozumalski, and Michael H Schwartz. Muscle synergies and complexity of neuromuscular control during gait in cerebral palsy. *Developmental Medicine & Child Neurology*, 57(12):1176–1182, 2015.

[130] Rachel W Jackson, Christopher L Dembia, Scott L Delp, and Steven H Collins. Muscle–tendon mechanics explain unexpected effects of exoskeleton assistance on metabolic rate during walking. *Journal of Experimental Biology*, 220(11):2082–2095, 2017.

[131] Richard W Nuckols, Taylor JM Dick, Owen N Beck, and Gregory S Sawicki. Ultrasound imaging links soleus muscle neuromechanics and energetics during human walking with elastic ankle exoskeletons. *Scientific reports*, 10(1):1–15, 2020.

[132] Andrew J Ries, Tom F Novacheck, and Michael H Schwartz. The efficacy of ankle-foot orthoses on improving the gait of children with diplegic cerebral palsy: a multiple outcome analysis. *PM&R*, 7(9):922–929, 2015.

[133] Rachel W Jackson and Steven H Collins. An experimental comparison of the relative benefits of work and torque assistance in ankle exoskeletons. *Journal of applied physiology*, 119(5):541–557, 2015.

[134] Yvette L Kerkum, Annemieke I Buizer, Josien C Van Den Noort, Jules G Becher, Jaap Harlaar, and Merel-Anne Brehm. The effects of varying ankle foot orthosis stiffness on gait in children with spastic cerebral palsy who walk with excessive knee flexion. *PloS one*, 10(11):e0142878, 2015.

[135] KA Ingraham, CD Remy, and EJ Rouse. The role of user preference in the customized control of robotic exoskeletons. *Science robotics*, 7(64):eabj3487, 2022.

[136] Michael Rosenberg and Katherine M Steele. Simulated impacts of ankle foot orthoses on muscle demand and recruitment in typically-developing children and children with cerebral palsy and crouch gait. *PloS one*, 12(7):e0180219, 2017.

[137] Zachary F Lerner, Diane L Damiano, and Thomas C Bulea. Computational modeling of neuromuscular response to swing-phase robotic knee extension assistance in cerebral palsy. *Journal of biomechanics*, 87:142–149, 2019.

[138] Maarten Afschrift, Friedl De Groote, Joris De Schutter, and Ilse Jonkers. The effect of muscle weakness on the capability gap during gross motor function: a simulation study supporting design criteria for exoskeletons of the lower limb. *Biomedical engineering online*, 13(1):1–15, 2014.

[139] Charles A Crabtree and Jill S Higginson. Modeling neuromuscular effects of ankle foot orthoses (afos) in computer simulations of gait. *Gait & Posture*, 29(1):65–70, 2009.

[140] Elisa S Arch, Steven J Stanhope, and Jill S Higginson. Passive-dynamic ankle–foot orthosis replicates soleus but not gastrocnemius muscle function during stance in gait: insights for orthosis prescription. *Prosthetics and orthotics international*, 40(5):606–616, 2016.

[141] Arthur D Kuo. Energetics of actively powered locomotion using the simplest walking model. *J. Biomech. Eng.*, 124(1):113–120, 2002.

[142] Steve Collins, Andy Ruina, Russ Tedrake, and Martijn Wisse. Efficient bipedal robots based on passive-dynamic walkers. *Science*, 307(5712):1082–1085, 2005.

[143] Arthur D Kuo, J Maxwell Donelan, and Andy Ruina. Energetic consequences of walking like an inverted pendulum: step-to-step transitions. *Exercise and sport sciences reviews*, 33(2):88–97, 2005.

[144] Thomas K Uchida, Ajay Seth, Soha Pouya, Christopher L Dembia, Jennifer L Hicks, and Scott L Delp. Simulating ideal assistive devices to reduce the metabolic cost of running. *PloS one*, 11(9):e0163417, 2016.

[145] DJJ Bregman, MM Van der Krogt, V De Groot, J Harlaar, M Wisse, and SH Collins. The effect of ankle foot orthosis stiffness on the energy cost of walking: a simulation study. *Clinical Biomechanics*, 26(9):955–961, 2011.

[146] Lorenzo Pitto, Hans Kainz, Antoine Falisse, Mariska Wesseling, Sam Van Rossom, Hoa Hoang, Eirini Papageorgiou, Ann Hallemans, Kaat Desloovere, Guy Molenaers, et al. Simcp: A simulation platform to predict gait performance following orthopedic intervention in children with cerebral palsy. *Frontiers in neurorobotics*, 13:54, 2019.

[147] Gregory S Sawicki and Nabil S Khan. A simple model to estimate plantarflexor muscle–tendon mechanics and energetics during walking with elastic ankle exoskeletons. *IEEE Transactions on Biomedical Engineering*, 63(5):914–923, 2015.

[148] Daniel A Jacobs, Jeffrey R Koller, Katherine M Steele, and Daniel P Ferris. Motor modules during adaptation to walking in a powered ankle exoskeleton. *Journal of neuroengineering and rehabilitation*, 15(1):1–15, 2018.

[149] Zachary F Lerner, Taryn A Harvey, and Jennifer L Lawson. A battery-powered ankle exoskeleton improves gait mechanics in a feasibility study of individuals with cerebral palsy. *Annals of biomedical engineering*, 47(6):1345–1356, 2019.

[150] Patrick W Franks, Nicholas A Bianco, Gwendolyn M Bryan, Jennifer L Hicks, Scott L Delp, and Steven H Collins. Testing simulated assistance strategies on a hip-knee-ankle exoskeleton: a case study. In *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pages 700–707. IEEE, 2020.

[151] Kirsten Veerkamp, Wouter Schallig, Jaap Harlaar, Claudio Pizzolato, Christopher P Carty, David G Lloyd, and Marjolein M van der Krogt. The effects of electromyography-assisted modelling in estimating musculotendon forces during gait in children with cerebral palsy. *Journal of biomechanics*, 92:45–53, 2019.

[152] Andrew J Meyer, Carolynn Patten, and Benjamin J Fregly. Lower extremity emg-driven modeling of walking with automated adjustment of musculoskeletal geometry. *PloS one*, 12(7):e0179698, 2017.

[153] Nathan R Sauder, Andrew J Meyer, Jessica L Allen, Lena H Ting, Trisha M Kesar, and Benjamin J Fregly. Computational design of fastfes treatment to improve

propulsive force symmetry during post-stroke gait: a feasibility study. *Frontiers in Neurorobotics*, page 80, 2019.

[154] Antoine Falisse, Gil Serrancolí, Christopher L Dembia, Joris Gillis, Ilse Jonkers, and Friedl De Groote. Rapid predictive simulations with complex musculoskeletal models suggest that diverse healthy and pathological human gaits can emerge from similar control strategies. *Journal of The Royal Society Interface*, 16(157):20190402, 2019.

[155] Matthew B Yandell, Brendan T Quinlivan, Dmitry Popov, Conor Walsh, and Karl E Zelik. Physical interface dynamics alter how robotic exosuits augment human movement: implications for optimizing wearable assistive devices. *Journal of neuroengineering and rehabilitation*, 14(1):1–11, 2017.

[156] Andrew J Ries, Tom F Novacheck, and Michael H Schwartz. A data driven model for optimal orthosis selection in children with cerebral palsy. *Gait & posture*, 40(4):539–544, 2014.

[157] Scott L Delp, Frank C Anderson, Allison S Arnold, Peter Loan, Ayman Habib, Chand T John, Eran Guendelman, and Darryl G Thelen. Opensim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering*, 54(11):1940–1950, 2007.

[158] Gil Serrancolí, Allison L Kinney, Benjamin J Fregly, and Josep M Font-Llagunes. Neuromusculoskeletal model calibration significantly affects predicted knee contact forces for walking. *Journal of biomechanical engineering*, 138(8), 2016.

[159] Megan R Ebers, Katherine M Steele, and J Nathan Kutz. Discrepancy modeling framework: Learning missing physics, modeling systematic residuals,

and disambiguating between deterministic and random effects. *arXiv preprint arXiv:2203.05164*, 2022.

[160] Herbert Hatze. A complete set of control equations for the human musculo-skeletal system. *Journal of biomechanics*, 10(11-12):799–805, 1977.

[161] Herbert Hatze. The complete optimization of a human motion. *Mathematical Biosciences*, 28(1-2):99–135, 1976.

[162] Horst-Moritz Maus, Shai Revzen, John Guckenheimer, Christian Ludwig, Johann Reger, and Andre Seyfarth. Constructing predictive models of human running. *Journal of The Royal Society Interface*, 12(103):20140899, 2015.

[163] Ankaral M Mert, Shahin Sefati, Manu S Madhav, Andrew Long, Amy J Bastian, and Noah J Cowan. Walking dynamics are symmetric (enough). *Journal of the Royal Society Interface*, 12(108):20150209, 2015.

[164] Wen-Xu Wang, Ying-Cheng Lai, and Celso Grebogi. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644:1–76, 2016.

[165] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[166] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[167] Taniel S Winner, Michael C Rosenberg, Trisha M Kesar, Lena H Ting, and Gordon J Berman. Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. *bioRxiv*, 2022.

[168] Emilio Sansano, Raúl Montoliu, and Oscar Belmonte Fernandez. A study of deep neural networks for human activity recognition. *Computational Intelligence*, 36(3):1113–1139, 2020.

[169] Vincent Hernandez, Davood Dadkhah, Vahid Babakeshizadeh, and Dana Kulić. Lower body kinematics estimation from wearable sensors for walking and running: A deep learning approach. *Gait & Posture*, 83:185–193, 2021.

[170] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Sequential deep learning for human action recognition. In *International workshop on human behavior understanding*, pages 29–39. Springer, 2011.

[171] Du-Xin Liu, Xinyu Wu, Can Wang, and Chunjie Chen. Gait trajectory prediction for lower-limb exoskeleton based on deep spatial-temporal model (dstm). In *2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM)*, pages 564–569. IEEE, 2017.

[172] Abdelrahman Zaroug, Daniel TH Lai, Kurt Mudie, and Rezaul Begg. Lower limb kinematics trajectory prediction using long short-term memory neural networks. *Frontiers in Bioengineering and Biotechnology*, 8:362, 2020.

[173] Binbin Su and Elena M Gutierrez-Farewik. Gait trajectory and gait phase prediction based on an lstm network. *Sensors*, 20(24):7127, 2020.

[174] Rania Kolaghassi, Mohamad Kenan Al-Hares, Gianluca Marcelli, and Konstantinos Sirlantzis. Performance of deep learning models in forecasting gait trajectories of children with neurological disorders. *Sensors*, 22(8):2969, 2022.

[175] Huong Thi Thu Vu, Dianbiao Dong, Hoang-Long Cao, Tom Verstraten, Dirk

Lefeber, Bram Vanderborght, and Joost Geeroms. A review of gait phase detection algorithms for lower limb prostheses. *Sensors*, 20(14):3972, 2020.

[176] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42. Springer Science & Business Media, 2013.

[177] Felix E Zajac, Richard R Neptune, and Steven A Kautz. Biomechanics and muscle coordination of human walking: Part i: Introduction to concepts, power transfer, dynamics and simulations. *Gait & posture*, 16(3):215–232, 2002.

[178] Hartmut Geyer and Hugh Herr. A muscle-reflex model that encodes principles of legged mechanics produces human walking dynamics and muscle activities. *IEEE Transactions on neural systems and rehabilitation engineering*, 18(3):263–273, 2010.

[179] David J Clark, Lena H Ting, Felix E Zajac, Richard R Neptune, and Steven A Kautz. Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke. *Journal of neurophysiology*, 103(2):844–857, 2010.

[180] Katherine M Steele, Rachel W Jackson, Benjamin R Shuman, and Steven H Collins. Muscle recruitment and coordination with an ankle exoskeleton. *Journal of biomechanics*, 59:50–58, 2017.

[181] Felix E Zajac, Richard R Neptune, and Steven A Kautz. Biomechanics and muscle coordination of human walking: part ii: lessons from dynamical simulations and clinical implications. *Gait & posture*, 17(1):1–17, 2003.

[182] Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*, 2021.

[183] Stanton A Glantz. Primer of biostatistics. Technical report, 2002.

[184] JRJ Belloto and TD Sokolovski. Residual analysis in regression. *American Journal of Pharmaceutical Education*, 49(3):295–303, 1985.

[185] Ali Nikoo and Thomas K Uchida. Be careful what you wish for: Cost function sensitivity in predictive simulations for assistive device design. *Symmetry*, 14(12):2534, 2022.

[186] Geoffrey G Handsfield, Craig H Meyer, Mark F Abel, and Silvia S Blemker. Heterogeneity of muscle sizes in the lower limbs of children with cerebral palsy. *Muscle & nerve*, 53(6):933–945, 2016.

[187] Felix E Zajac. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17(4):359–411, 1989.

[188] Hartmut Geyer, Andre Seyfarth, and Reinhard Blickhan. Compliant leg behaviour explains basic dynamics of walking and running. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603):2861–2867, 2006.

[189] Saulo Martelli, Daniela Calvetti, Erkki Somersalo, and Marco Viceconti. Stochastic modelling of muscle recruitment during activity. *Interface focus*, 5(2):20140094, 2015.

[190] Jason R Franz. A sound approach to improving exoskeletons and exosuits. *Science Robotics*, 6(60):eabm6369, 2021.

[191] Lena H Ting and Hillel J Chiel. Muscle, biomechanics, and implications for neural control. *Neurobiology of Motor Control: Fundamental Concepts and New Directions*, pages 365–416, 2017.

[192] Andreas Sebastian Schroeder, Rüdiger Von Kries, Christina Riedel, Maria Homburg, Helene Auffermann, Astrid Blaschek, Klaus Jahn, Florian Heinen, Ingo Borggraefe, and Steffen Berweck. Patient-specific determinants of responsiveness to robot-enhanced treadmill therapy in children and adolescents with cerebral palsy. *Developmental Medicine & Child Neurology*, 56(12):1172–1179, 2014.

[193] Jonathon S Schofield, Courtney E Shell, Dylan T Beckler, Zachary C Thumser, and Paul D Marasco. Long-term home-use of sensory-motor-integrated bidirectional bionic prosthetic arms promotes functional, perceptual, and cognitive changes. *Frontiers in neuroscience*, 14:120, 2020.

[194] Michael C Rosenberg, Joshua L Proctor, and Katherine M Steele. Quantifying template signatures of center-of-mass motion during walking with ankle exoskeletons. *bioRxiv*, 2022.

[195] Jordan T Sturdy, Anne K Silverman, and Nathan T Pickle. Automated optimization of residual reduction algorithm parameters in opensim. *Journal of Biomechanics*, 137:111087, 2022.

[196] Tishya AL Wren, George E Gorton III, Sylvia Ounpuu, and Carole A Tucker. Efficacy of clinical gait analysis: A systematic review. *Gait & posture*, 34(2):149–153, 2011.

[197] Tishya AL Wren, Carole A Tucker, Susan A Rethlefsen, George E Gorton III, and Sylvia Õunpuu. Clinical efficacy of instrumented gait analysis: Systematic review 2020 update. *Gait & posture*, 80:274–279, 2020.

[198] Lahiru N Wimalasena, Jonas F Braun, Mohammad Reza Keshtkaran, David Hofmann, Juan Álvaro Gallego, Cristiano Alessandro, Matthew C Tresch, Lee E

Miller, and Chethan Pandarinath. Estimating muscle activation from emg using deep learning-based dynamical systems models. *Journal of Neural Engineering*, 19(3):036013, 2022.

[199] Krithika Manohar, Bingni W Brunton, J Nathan Kutz, and Steven L Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 38(3):63–86, 2018.

[200] Jonathan P Gardner, John C Mather, Mark Clampin, Rene Doyon, Matthew A Greenhouse, Heidi B Hammel, John B Hutchings, Peter Jakobsen, Simon J Lilly, Knox S Long, et al. The james webb space telescope. *Space Science Reviews*, 123:485–606, 2006.

[201] Ian H Stevenson and Konrad P Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–142, 2011.

[202] Larry R Medsker and LC Jain. Recurrent neural networks. *Design and Applications*, 5(64-67):2, 2001.

[203] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.

[204] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[205] Jan Williams, Olivia Zahn, and J Nathan Kutz. Data-driven sensor placement with shallow decoder networks. *arXiv preprint arXiv:2202.05330*, 2022.

[206] Jan P Williams, Olivia Zahn, and J Nathan Kutz. Sensing with shallow recurrent decoder networks. *arXiv preprint arXiv:2301.12011*, 2023.

[207] J. Nathan Kutz. *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*. Oxford University Press, Oxford, 1st edition edition, September 2013.

[208] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, Cambridge, 1st edition edition, April 2019.

[209] R. Everson and L. Sirovich. Karhunen–Loève procedure for gappy data. *JOSA A*, 12(8):1657–1664, August 1995. Publisher: Optica Publishing Group.

[210] Zlatko Drmač and Serkan Gugercin. A New Selection Operator for the Discrete Empirical Interpolation Method—Improved A Priori Error Bound and Extensions. *SIAM Journal on Scientific Computing*, 38(2):A631–A648, January 2016. Publisher: Society for Industrial and Applied Mathematics.

[211] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, November 2004.

[212] Saifon Chaturantabut and Danny C. Sorensen. Nonlinear Model Reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, January 2010. Publisher: Society for Industrial and Applied Mathematics.

[213] Stephen Boyd and Lieven Vandenberghe. Convex Optimization, March 2004. ISBN: 9780511804441 Publisher: Cambridge University Press.

[214] Siddharth Joshi and Stephen Boyd. Sensor Selection via Convex Optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, February 2009. Conference Name: IEEE Transactions on Signal Processing.

[215] W. F. Caselton and J. V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4):223–227, August 1984.

[216] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *The Journal of Machine Learning Research*, 9:235–284, June 2008.

[217] D. V. Lindley. On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956. Publisher: Institute of Mathematical Statistics.

[218] Paola Sebastiani and Henry P. Wynn. Maximum Entropy Sampling and Optimal Bayesian Experimental Design. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(1):145–157, 2000. Publisher: [Royal Statistical Society, Wiley].

[219] Liam Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17(7):1480–1507, July 2005.

[220] Peter Gunnarson, Ioannis Mandralis, Guido Novati, Petros Koumoutsakos, and John O Dabiri. Learning efficient navigation in vortical flow fields. *Nature communications*, 12(1):7143, 2021.

[221] Kartik Krishna, Zhuoyuan Song, and Steven L Brunton. Finite-horizon, energy-efficient trajectories in unsteady flows. *Proceedings of the Royal Society A*, 478(2258):20210255, 2022.

[222] Michele Buzzicotti, Luca Biferale, Fabio Bonaccorso, Patricio Clark di Leoni, and Kristian Gustavsson. Optimal control of point-to-point navigation in turbulent time dependent flows using reinforcement learning. In *International Conference of the Italian Association for Artificial Intelligence*, pages 223–234. Springer, 2020.

[223] Luca Biferale, Fabio Bonaccorso, Michele Buzzicotti, Patricio Clark Di Leoni, and Kristian Gustavsson. Zermelo's problem: optimal point-to-point navigation in 2d turbulent flows using reinforcement learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10), 2019.

[224] Angel Madridano, Abdulla Al-Kaff, David Martín, and Arturo De La Escalera. Trajectory planning for multi-robot systems: Methods and applications. *Expert Systems with Applications*, 173:114660, 2021.

[225] Sachin Shriwastav, Gregory Snyder, and Zhuoyuan Song. Dynamic compressed sensing of unsteady flows with a mobile robot. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11910–11915. IEEE, 2022.

[226] Kevin M Lynch, Ira B Schwartz, Peng Yang, and Randy A Freeman. Decentralized environmental modeling by mobile sensor networks. *IEEE transactions on robotics*, 24(3):710–724, 2008.

[227] Naomi Ehrich Leonard, Derek A Paley, Francois Lekien, Rodolphe Sepulchre, David M Fratantoni, and Russ E Davis. Collective motion, sensor networks, and ocean sampling. *Proceedings of the IEEE*, 95(1):48–74, 2007.

[228] Levi DeVries, Sharanya J Majumdar, and Derek A Paley. Observability-based optimization of coordinated sampling trajectories for recursive estimation of a strong,

spatially varying flowfield. *Journal of intelligent & robotic systems*, 70(1-4):527–544, 2013.

[229] Petter Ogren, Edward Fiorelli, and Naomi Ehrich Leonard. Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment. *IEEE Transactions on Automatic control*, 49(8):1292–1302, 2004.

[230] Fumin Zhang and Naomi Ehrich Leonard. Cooperative filters and control for cooperative exploration. *IEEE Transactions on Automatic Control*, 55(3):650–663, 2010.

[231] Derek A Paley and Artur Wolek. Mobile sensor networks and control: Adaptive sampling of spatiotemporal processes. *Annual Review of Control, Robotics, and Autonomous Systems*, 3:91–114, 2020.

[232] Liqian Peng, Doug Lipinski, and Kamran Mohseni. Dynamic data driven application system for plume estimation using uavs. *Journal of Intelligent & Robotic Systems*, 74(1-2):421–436, 2014.

[233] Jiazhong Mei, Steven L Brunton, and J Nathan Kutz. Mobile sensor path planning for kalman filter spatiotemporal estimation. *arXiv preprint arXiv:2212.08280*, 2022.

[234] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[235] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[236] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].

[237] Yi Li, Eric Perlman, Minping Wan, Yunke Yang, Charles Meneveau, Randal Burns, Shiyi Chen, Alexander Szalay, and Gregory Eyink. A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, 9:N31, January 2008. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14685240802376389.

[238] Anne Schmitz, Mao Ye, Robert Shapiro, Ruigang Yang, and Brian Noehren. Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. *Journal of biomechanics*, 47(2):587–591, 2014.

[239] David J Clark. Automaticity of walking: functional significance, mechanisms, measurement and rehabilitation strategies. *Frontiers in human neuroscience*, 9:246, 2015.

[240] Yang Wang and Manoj Srinivasan. System identification and stability analyses of steady human locomotion. *Foot*, 300(600):600, 2012.

[241] Yang Wang and Manoj Srinivasan. Stepping in the direction of the fall: the next foot placement can be predicted from current upper body state in steady-state walking. *Biology letters*, 10(9):20140405, 2014.

[242] Robert J Full and Daniel E Koditschek. Templates and anchors: neuromechanical hypotheses of legged locomotion on land. *Journal of experimental biology*, 202(23):3325–3332, 1999.

[243] Dan Zhao, Brian Bittner, Glenna Clifton, Nick Gravish, and Shai Revzen. Walking is like slithering: A unifying, data-driven view of locomotion. *Proceedings of the National Academy of Sciences*, 119(37):e2113222119, 2022.

[244] Michael S Orendurff, Jason A Schoen, Greta C Bernatz, Ava D Segal, and Glenn K

Klute. How humans walk: bout duration, steps per bout, and rest duration. *Journal of Rehabilitation Research & Development*, 45(7), 2008.

[245] Pietro Picerno. 25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches. *Gait & posture*, 51:239–246, 2017.

[246] David A Winter. Biomechanics of normal and pathological gait: implications for understanding human locomotor control. *Journal of motor behavior*, 21(4):337–355, 1989.

[247] David Grabli, Carine Karachi, Marie-Laure Welter, Brian Lau, Etienne C Hirsch, Marie Vidailhet, and Chantal François. Normal and pathological gait: what we learn from parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 83(10):979–985, 2012.

[248] Robert C Wagenaar and Richard EA van Emmerik. Dynamics of pathological gait. *Human Movement Science*, 13(3-4):441–471, 1994.

[249] Verne T Inman, Howard D Eberhart, et al. The major determinants in normal and pathological gait. *JBJS*, 35(3):543–558, 1953.