BAYESIAN IDENTIFICATION OF
NONLINEAR STRUCTURAL SYSTEMS:
INNOVATIONS TO ADDRESS PRACTICAL UNCERTAINTY

by

Alana Lund

A Dissertation

Submitted to the Faculty of Purdue University
In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Lyles School of Civil Engineering

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Shirley J. Dyke, Co-Chair

Lyles School of Civil Engineering and School of Mechanical Engineering

Dr. Ilias Bilionis, Co-Chair

School of Mechanical Engineering

Dr. Julio Ramirez

Lyles School of Civil Engineering

Dr. Wei Song

Department of Civil, Construction, and Environmental Engineering

University of Alabama, Tuscaloosa, AL

Dr. Erik Johnson

Sonny Astani Department of Civil and Environmental Engineering

University of Southern California, Los Angeles, CA

Approved by:

Dr. Dulcy M. Abraham

"We have not succeeded in answering all our problems – indeed we sometimes feel we have not completely answered any of them. The answers we have found have only served to raise a whole new set of questions. In some ways, we feel we are as confused as ever, but we think we are confused on a higher level about more important things."   *- Earl C. Kelley*

To God, my light and my strength,
to Jarrod, who helps me see I'm making a difference,
to Vanessa and Eliza, who push me to be more,
and to my mother, who got me across the finish line.

ACKNOWLEDGMENTS

Finally, I would like to thank my family for their loving encouragement over the years. My husband, Jarrod Lund, is a blessing and a wonder in my life. He is an endless well of support, and is always ready with an interesting idea or new way of looking at things that helps me move my work forward. Our parents are amazing and so invested in and supportive of my progress. I love you all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

ABSTRACT

The ability to rapidly assess the condition of a structure in a manner which enables the accurate prediction of its remaining capacity has long been viewed as a crucial step in allowing communities to make safe and efficient use of their public infrastructure. This objective has become even more relevant in recent years as both the interdependency and state of deterioration in infrastructure systems throughout the world have increased. Current practice for structural condition assessment emphasizes visual inspection, in which trained professionals will routinely survey a structure to estimate its remaining capacity. Though these methods have the ability to monitor gross structural changes, their ability to rapidly and cost-effectively assess the detailed condition of the structure with respect to its future behavior is limited.

Vibration-based monitoring techniques offer a promising alternative to this approach. As opposed to visually observing the surface of the structure, these methods judge its condition and infer its future performance by generating and updating models calibrated to its dynamic behavior. Bayesian inference approaches are particularly well suited to this model updating problem as they are able to identify the structure using sparse observations while simultaneously assessing the uncertainty in the identified parameters. However, a lack of consensus on efficient methods for their implementation to full-scale structural systems has led to a diverse set of Bayesian approaches, from which no clear method can be selected for full-scale implementation. The objective of this work is therefore to assess and enhance those techniques currently used for structural identification and make strides toward developing unified strategies for robustly implementing them on full-scale structures. This is accomplished by addressing several key research questions regarding the ability of these methods to overcome issues in identifiability, sensitivity to uncertain experimental conditions, and scalability. These questions are investigated by applying novel adaptations of several prominent Bayesian identification strategies to small-scale experimental systems equipped with nonlinear devices. Through these illustrative examples I explore the robustness and practicality of these algorithms, while also considering their extensibility to higher-dimensional systems. Addressing these core concerns underlying full-scale structural identification will enable the practical application of Bayesian inference techniques and thereby enhance the ability of communities to detect and respond to the condition of their infrastructure.

1. INTRODUCTION

Our communities are inexorably enmeshed with the complex and evolving environment we've built to support them. Despite the foundational role these infrastructure systems play in society, they are often viewed as static monuments as opposed to the dynamic, and often deteriorating, structures we trust for health, safety, and productivity. Due to insufficient rehabilitation and expansion efforts, the United States infrastructure system is risking the safety of its citizens and impeding the nation's ability to compete in a global economy [1]. The 2017 ASCE infrastructure report card gives the nation a grade of D+ overall and shows major concerns across different sectors, such as the 9% of bridges with known structurally deficiencies [2]. They predict that a $4.5 trillion overhaul will be necessary to bring the system back into shape for a competitive and sustainable future. However, the current national investment in infrastructure, approximately 2.4% of the GDP [3], is not sufficient to overcome this gap and lags far behind both the performance of and investment in infrastructure internationally, with European nations on average spending 5% of their GDP on infrastructure [1]. With the limited resources currently available for improving the condition of the nation's infrastructure, it is critical to develop and implement information-based strategies for condition monitoring to prioritize rehabilitation planning that support the safety and productivity of our communities. This is the role of structural health monitoring.

Infrastructure monitoring is most commonly approached through visual inspection, in which structures are visited regularly by trained inspectors who judge the level of the deterioration in the structure by its outward appearance. Although inspection-based procedures and processes are already in place for a number of critical systems, such as the bridge [4] and dam [5] networks, this approach to infrastructure assessment has been shown to fall short in a number of cases. Take, for example, the collapse of the I-35W bridge in 2007. This structure had been inspected annually since 1993, and although classified as "structurally deficient" was thought to have sufficient capacity to remain in service until replacement in 2020-2025 [6]. Though the root of this particular collapse was found to be due to errors in construction rather than deterioration in the members, this and other tragedies have brought increased attention to the challenges of visual inspection for performance estimation, such as the reliability and consistency of infrastructure ratings, the accessibility of structures, the cost of inspections, and the safety of inspectors [7]–[9]. Some of

these concerns are being addressed by recent research in robotic inspection and image processing [10], but at its core this method of structural health monitoring still stretches to predict structural behavior from surface-level observations.

A promising alternative to this approach is to directly observe the dynamic behavior of the structure and make judgements concerning its condition through vibration-based monitoring techniques. These automated infrastructure assessment approaches can be broadly classified into two categories, signal processing methods and model updating methods. *Signal processing methods* assess the current condition of a structure in relation to pre-determined quantitative behavioral thresholds. *Model updating methods,* in contrast, use observations from the structure to iteratively identify a model which describes its current state and predicts its future behavior. Variations in the parameters of this model over time are indicative of localized areas of concern in the physical structure. Ideally, the updated model can also be used to forecast the performance of the structure with regard to future events. Although each of these methods are valuable for infrastructure inspection and monitoring, model updating methods are uniquely powerful due to the level of actionable information they provide for infrastructure planning and rehabilitation efforts.

Methods based on Bayesian inference are well-suited to the model updating problem due to their ability to produce models of varying fidelity with sparse observations while simultaneously assessing the uncertainty in the identified parameters. The application of these methods to full-scale structural identification has been studied from a variety of perspectives, such as their robustness to noise and system nonlinearity [11], [12], their ability to continuously track and update changes in the model parameters [13], [14], and their real-time implementation [15], [16]. However, to date there has been little agreement as to which particular method adequately addresses all of the concerns raised in structural health monitoring applications. For example, particle filters have been recognized for their increased accuracy and ability to handle nonlinear models with non-Gaussian distributions, but they lag far behind the extended and unscented Kalman filters in terms of computational speed, making them inefficient for real-time or even rapid implementation [12]. Effective implementation of Bayesian inference methods for structural health monitoring requires that the selected approach address these and other unresolved challenges, including: scalability [17], model selection [18], [19], sparse observations of structural behavior [20], [21], identifiability [22], and sensitivity to uncertain experimental conditions. To evolve

Bayesian structural identification methods toward general full-scale structural health monitoring implementations, the available approaches must first be assessed and enhanced to address these issues.

The objective of my dissertation is to develop a scalable approach to Bayesian structural identification which also addresses concerns specifically related to real-time implementation and experimental uncertainty. My research therefore centers around the following questions:

1. For a given nonlinear system, how can appropriate excitation signals be selected in advance of experimentation to ensure the successful identification of the system model from its dynamic response?

2. How can Bayesian identification approaches be implemented to reliably produce a representative model of the structure in the face of experimental uncertainties?

3. How can the flexibility of the Bayesian approach best be balanced with its scalability such that Bayesian identification methods can be applied to realistic engineering structures?

I pursue the answers to these questions through the experimental identification of a small-scale nonlinear energy sink device, which is described in detail in Chpt. 4. This device poses a particular challenge for identification due to the dual stability it exhibits in its response behavior, such that two separate modes of operation are evident when different levels of input energy are applied. The complexity of the device behavior, coupled with the inherent limits of modeling techniques to fully capture experimental variability, provide a rich experimental basis to evaluate novel approaches to Bayesian predictive modeling.

The theoretical framework for Bayesian inference on structural systems and some seminal literature regarding several methods common to structural health monitoring research is summarized in Chpt. 2 to provide a common foundation for the developments in this dissertation. Of these methods, the unscented Kalman filter is identified as the most likely to be effective in practical structural health monitoring scenarios, and is therefore the focus of the innovations in scalability and experimental reliability in Chpts. 5 and 6, respectively. Recent innovations in Bayesian approximations in the field of variational inference are discussed in Chpt. 3. This promising approach to Bayesian approximation is adapted for use in structural systems in Chpt. 7, and a novel approach to variational filtering for structural systems is developed in Chpt. 8. Conclusions from this project as well as extensions of the research for future study are discussed in Chpt. 9. It is the hope of the author that by addressing the fundamental challenges of

identifiability, experimental robustness, and scalability in a controlled experimental setting, the practical implementation of these methods for information-based infrastructure management will gain greater acceptance and utility in the field.

1.1   References

1.   Mcbride, J. The state of U.S. infrastructure. 1–13 (2018).

2.   ASCE. *2017 infrastructure report card: a comprehensive assessment of America's infrastructure*. (2017). doi:10.1007/978-3-319-13296-9_12

3.   Shirley, C. Spending on Infrastructure and Investment. (2017).

4.   FHWA. 23 CFR 650 — National bridge inspection standards. **69,** 650.401-650.517 (2004).

5.   FEMA. *Federal guidelines for dam safety*. (2004).

6.   Narayanan, S. I-35W Mississippi river bridge failure - is it a wake up call? *Indian Concr. J.* **19,** 29–38 (2008).

7.   Whitehead, J. M. Probability of detection study for visual inspection of steel. (Purdue University, 2015).

8.   Graybeal, B. A., Rolander, D. D., Phares, B. M., Moore, M. E. & Washer, G. A. Reliability and accuracy of in-depth inspection of highway bridges. *Transp. Res. Rec. J. Transp. Res. Board* **1749,** 93–99 (2007).

9.   Phares, B. M., Washer, G. A., Rolander, D. D., Graybeal, B. A. & Moore, M. Routine highway bridge inspection condition documentation accuracy and reliability. *J. Bridg. Eng.* **9,** 403–413 (2004).

10.   Lattanzi, D., Asce, M. & Miller, G. Review of robotic infrastructure inspection systems. *J. Infrastruct. Syst.* **23,** (2017).

11.   Wu, M. & Smyth, A. W. Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Struct. Control Heal. Monit.* **14,** 971–990 (2007).

12.   Chatzi, E. N. & Smyth, A. The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Struct. Control Heal. Monit.* **16,** 99–123 (2009).

13.   Yang, J. N., Lin, S., Huang, H. & Zhou, L. An adaptive extended Kalman filter for structural damage identification. *Struct. Control Heal. Monit.* **13,** 849–867 (2006).

14. Zhou, L., Wu, S. & Yang, J. N. Experimental study of an adaptive extended Kalman filter for structural damage identification. *J. Infrastruct. Syst.* **14,** 42–51 (2008).

15. Shao, X., Mueller, A. & Mohammed, B. A. Real-Time hybrid simulation with online model updating: methodology and implementation. *J. Eng. Mech.* **142,** 1–19 (2015).

16. Song, W. & Dyke, S. J. Real-time dynamic model updating of a hysteretic structural system. *J. Struct. Eng.* **140,** 1–14 (2014).

17. Azam, S. E. *Online damage detection in structural systems: applications of proper orthogonal decomposition, and Kalman and particle filters*. (Springer Science+Business Media, 2014). doi:10.1007/978-3-319-02559-9

18. Yuen, K.-V. & Mu, H.-Q. Real-time system identification: an algorithm for simultaneous model class selection and parametric identification. *Comput. Civ. Infrastruct. Eng.* **30,** 785–801 (2015).

19. Muto, M. & Beck, J. L. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *J. Vib. Control* **14,** 7–34 (2008).

20. Al-hussein, A. & Haldar, A. Novel unscented Kalman filter for health assessment of structural systems with unknown input. *J. Eng. Mech.* **141,** 1–13 (2015).

21. Azam, S. E., Chatzi, E. & Papadimitriou, C. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mech. Syst. Signal Process.* **60,** 866–886 (2015).

22. Olivier, A. & Smyth, A. W. On the performance of online parameter estimation algorithms in systems with various identifiability properties. *Front. Built Environ.* **3,** 1–18 (2017).

2. BAYESIAN FILTERS FOR THE PREDICTIVE MODELING OF STRUCTURAL SYSTEMS

Bayesian methods for system identification allow us to address the inherent epistemic uncertainties that arise in the formulation of predictive models for structural systems. These methods operate according to Bayes Theorem

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{2.1}$$

in which beliefs concerning the latent model, $\mathbf{x}$, and its calibration parameters, $\boldsymbol{\theta}$, are updated through observations, $\mathbf{y}$, of the true system. The epistemic uncertainties associated with predictive modeling, which are generally classified as parametric uncertainty, model uncertainty, observation uncertainty, and interpolation uncertainty [1]–[3], are incorporated directly into the Bayesian formulation by their implied association with the probability density functions (PDFs) of different variables.

Parameter uncertainty expresses the lack of knowledge concerning the precise values of particular calibration parameters in the model, such as stiffness or damping for structural systems. Though these parameters may not be known precisely, they can typically be specified with some certainty within a particular range. This belief as to the range of values of the parameters can be stated precisely by expressing it in the form of a PDF, $p(\boldsymbol{\theta})$. Likewise, the model uncertainty describes the limited capacity of a computational model to fully capture the range of behavior expressed by the physical system, and is encapsulated within $p(\mathbf{x}|\boldsymbol{\theta})$. Collectively, the term $p(\mathbf{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$ forms what is called the *prior*, referring to modeler's prior knowledge of the system which can be concisely defined in terms of a stochastic computational model.

Observation uncertainty is defined through the *likelihood function*, $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$, and describes the errors associated with data collection. The likelihood function itself then represents the probability that computations with the prior will yield accurate predictions of the true experimental observations.

Interpolation uncertainty refers to the inherent errors in our formulation of the predictive model which arise due to limited data which gives only partial information concerning the system behavior. This uncertainty is captured in the posterior, $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$, which gives an updated expression of our model based on all available information. As more data becomes available and

the model is continuously updated, the variance of the posterior naturally decreases. A single step of this process is shown in Fig. 2.1, in which the prior is adjusted by the likelihood to form a more confident system model in the posterior.
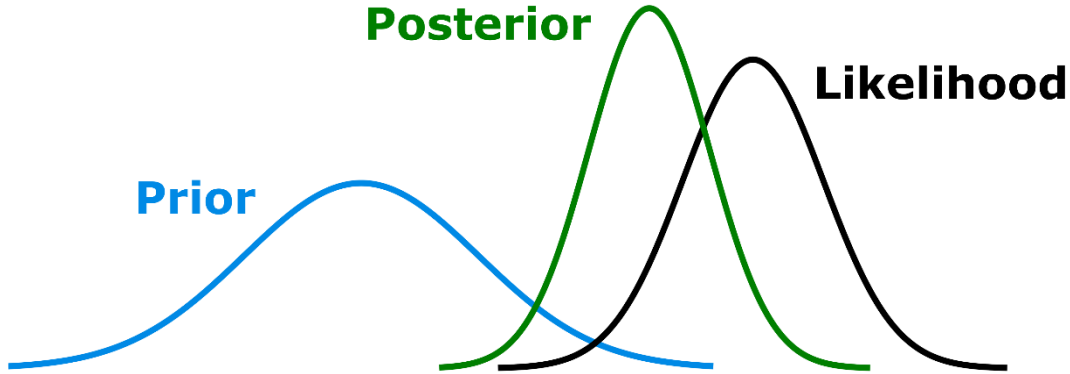
Fig. 2.1 – Graphical representation of Bayes' Theorem

The final term in the Bayesian formulation is the marginal likelihood, $p(\mathbf{y})$, which is also referred to as the evidence. Because all the data has been observed, this term evaluates to a constant which is used to normalize the computation such that the posterior meets the definition of a PDF (i.e. integrates to 1). For the majority of physical systems however, the evidence is difficult to compute. As will be discussed later in the chapter, this complexity has inspired a number of approximate approaches to Bayesian system identification.

To apply the Bayesian formulation to structural systems, it is first necessary to define the interactions of the basic system variables, which are shown by the graphical interaction model in Fig. 2.2. For structural systems, the latent model, $\mathbf{x}$, is described by the interaction of a variety of dynamics states which characterize the system behavior, such as displacement and velocity. As Fig. 2.2 shows, these states exhibit an inherent dependency structure in time, which is typically described using a Markovian transmission probability $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1})$, where $\mathbf{u}_{k-1}$ is the control input, at discrete time intervals $t_k = k\Delta t$, where $k = 1, 2, 3, \ldots, \mathrm{K}$. As evidenced in the transmission probability model, the dynamic behavior of these states is also influenced by the geometric and material properties of the system, such as stiffness and damping. These properties form the calibration parameters, $\boldsymbol{\theta}$, of the system, and are typically modeled as constant, global system variables. Changes in the value of these variables due to environmental variations or system

degradation can be included in the model, but do not form the focus of this study. Our understanding of the system behavior is also characterized by its emission probability $p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k)$, which forms the likelihood model. If they were to be interpreted deterministically, the transmission and emission probabilities would be equivalent to the transition and observation equations that form the state space model for a general dynamical system. The use of the transmission and emission probabilities in the description of the system model simply allows for the explicit incorporation of uncertainty into the known computational model.
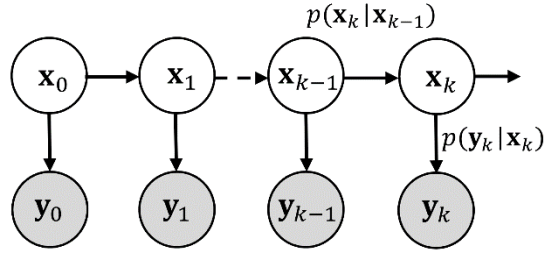
Fig. 2.2 – Probabilistic model of a general dynamical system.

With the stochastic system model fully defined, the posterior can be formulated according to Bayes theorem,

$$p(\mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{y}_{1:K}, \mathbf{u}_{1:K}) = \frac{\prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k) \prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}) \, p(\mathbf{x}_0, \boldsymbol{\theta}|\mathbf{u}_{1:K})}{p(\mathbf{y}_{1:K}, \mathbf{u}_{1:K})}, \quad (2.2)$$

where $p(\mathbf{x}_0, \boldsymbol{\theta})$ is the joint prior distribution on the initial conditions of the dynamic states and system parameters and $p(\mathbf{y}_{1:K}, \mathbf{u}_{1:K})$ is the model evidence given all observations of the control and the response in the time series. This interpretation of Bayes theorem infers the full joint posterior on the states and parameters, $p(\mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{y}_{1:K}, \mathbf{u}_{1:K})$, from all data points simultaneously, meaning that either all data must be available at the time of inference or that the posterior must be recalculated with increasingly large data sets as new observations become available. In light of this dependence on the full data set, methods which use this approach are typically referred to as *batch* methods.

Batch methods for predictive modeling of structural systems are advantageous in that they are designed to completely characterize the posterior of complex, nonlinear system models using all data and knowledge of the uncertainty simultaneously. Using approximation techniques tailored for batch Bayesian analysis, researchers have had great success in tackling practical issues such as

model class selection [4]–[9], model updating under ambient/environmental conditions [10], and damage detection [11],[12]. A comprehensive review of the state-of-the-art in these techniques is available from Huang et al. [13]. Despite their many benefits, these approaches are limited due to their computational complexity. Because of the longer times and memory capabilities needed to compute the posterior model approximation, real-time techniques, referred to as *Bayesian filtering* approaches, are often preferred for structural health monitoring applications.

Filtering methods operate recursively by determining $p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k})$, the marginal posterior of states and parameters at the current time step, as opposed to the full joint posterior, $p(\mathbf{x}_{0:K}, \boldsymbol{\theta} | \mathbf{y}_{1:K}, \mathbf{u}_{1:K})$. The marginal posterior is constructed by first determining the marginal prior through the Chapman-Kolmogorov equation,

$$p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1}, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1}) \, d\mathbf{x}_{k-1}, \qquad (2.3)$$

and then developing the marginal posterior with Bayes' Theorem,

$$p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k) p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1})}{p(\mathbf{y}_k, \mathbf{u}_k | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1})}. \qquad (2.4)$$

If necessary, the full posterior, $p(\mathbf{x}_{0:K}, \boldsymbol{\theta} | \mathbf{y}_{1:K}, \mathbf{u}_{1:K})$, can then be constructed by iterating backwards in time with a smoothing algorithm [3], though it should be noted smoothing algorithms have disadvantages similar to batch methods with regard to real-time applications. Even without the implementation of a smoothing algorithm, the marginal posterior on the parameters at time K, $p(\boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k})$, is equivalent between the batch and filtering interpretations due to the treatment of $\boldsymbol{\theta}$ as constant in Eq. (2.3).

As mentioned previously, the analytical tractably of this solution scheme is dependent on the computation of the model evidence $p(\mathbf{y}_k, \mathbf{u}_k | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1})$. For linear systems in which the transmission and emission probabilities can be adequately described by Gaussian probability densities, the evidence can be explicitly defined and requires fixed computational time at each step, k. The majority of systems for which structural health monitoring would be beneficial do not fit this mold. The inherent uncertainties in modeling these complex nonlinear systems often result in issues in constructing the model evidence, such that it either cannot be expressed in closed form or requires exponential time to compute [14]. Significant research has therefore been devoted to the development of strategies which circumvent these issues through various computational methods.

Over the past 60 years, a number of solutions have been developed to address these issues. For the purposes of structural health monitoring we are primarily interested in those algorithms which can be implemented sequentially, as in the construction given above, as they have the potential to provide real-time health monitoring and model updating information concerning the structures of interest. Such approaches can be broadly classified as analytical methods, sampling methods, and optimization methods. The analytical and sampling methods, which are more common to current research and implementation in structural health monitoring, will be described in the remainder of this chapter. Optimization methods for predictive modeling of structural systems are described in Chpt. 3.

2.1 Analytical Methods for Bayesian Filtering

Exact analytical solutions to the Bayesian filtering equations are possible for linear, Gaussian distributed, stochastic dynamical systems. Succinct algorithms were developed for this purpose in the 1950s and 60s, the most notable being the Kalman filter, developed by R.E. Kalman in 1960 [15] as a special case of the Weiner filter.

The Kalman filter treats systems with transmission and emission probabilities of the form

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\theta}_{k-1}) = N(\mathbf{x}_k|\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1}, \mathbf{Q}_{k-1}) \qquad (2.5)$$

$$p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}_k) = N(\mathbf{y}_k|\mathbf{C}_k\mathbf{x}_k + \mathbf{D}_k\mathbf{u}_k, \mathbf{R}_k) \qquad (2.6)$$

where $\mathbf{A}$ is the state transition matrix, $\mathbf{B}$ is the input matrix, $\mathbf{C}$ is the measurement matrix, $\mathbf{D}$ is the direct transmission matrix, and $\mathbf{u}$ is the system input vector. The covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ are representative of the process noise and the measurement noise, respectively. These terms are noted as time varying in the Eqs. (2.5) and (2.6), but can often be interpreted as time invariant. The prior distribution which initializes the algorithm is inferred using information from the known behavior of the system such that $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$. The full Kalman filter algorithm then develops the Gaussian posterior over the states through matrix manipulations on the computational model, as shown in Algorithm 2.1.

A critical implication of the Kalman filter structure is its lack of an inherent capacity for structural identification. Expressing the parameters as unknown would imply nonlinearity in the system and therefore violate fundamental assumptions of the analytical filtering solution. This shortcoming has most commonly been overcome through direct optimization of the parameters to

find the maximum of the likelihood (ML) or posterior (MAP) distributions using methods such as the expectation-maximization algorithm [14] or by sampling from the posterior distribution using Markov chain Monte Carlo (MCMC) methods [3]. However, these approaches to parameter identification negate the advantages in computational speed that are the hallmark of the Kalman filter, as they require several iterations over the data to determine the optimal parameters. Wan and Nelson [16],[17] proposed an alternative approach that allows for parameter estimation in real time, which they term the dual Kalman filter. Dual estimation splits the problem into two interconnected filtering implementations, in which one filter estimates the state and the other estimates the parameters. This approach has been shown to be an effective and stable form of predictive modeling [18],[19].

Initialize with:

$$\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0]$$
$$\mathbf{P}_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$$

State Prediction:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1}$$
$$\mathbf{P}_k^- = \mathbf{A}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}$$

Observation Prediction:

$$\hat{\mathbf{y}}_k = \mathbf{C}_k\hat{\mathbf{x}}_k^- + \mathbf{D}_k\mathbf{u}_k$$
$$\mathbf{P}_{\mathbf{y}_k\mathbf{y}_k} = \mathbf{C}_k\mathbf{P}_k^-\mathbf{C}_k^T + \mathbf{R}_k$$
$$\mathbf{P}_{\mathbf{x}_k\mathbf{y}_k} = \mathbf{P}_k^-\mathbf{C}_k^T$$

Correction:

$$\mathcal{K}_k = \mathbf{P}_{\mathbf{x}_k\mathbf{y}_k}\mathbf{P}_{\mathbf{y}_k\mathbf{y}_k}^{-1}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathcal{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k)$$
$$\mathbf{P}_k = \mathbf{P}_k^- - \mathcal{K}_k\mathbf{P}_{\mathbf{y}_k\mathbf{y}_k}\mathcal{K}_k^T$$

Algorithm 2.1 The Kalman Filter

Though the analytical nature of the Kalman filter promotes scalability to high-dimensional structural systems, the constraint to a linear dynamical model severely limits its applicability to the complex nonlinear systems typical of structural health monitoring applications. To alleviate this concern while still taking advantage of the efficient Gaussian framework of the Kalman filter, several approximations to the algorithm have been developed. These approaches can be divided into extended Kalman filters (EKF) and sigma point Kalman filters (SPKF).

Common among these approximate analytical approaches is their philosophy for parameter identification. The overwhelming majority of applications apply joint [20]–[23] parameter

identification, though dual [16],[24],[25] estimation is also fairly common. In joint estimation, the state vector, $\mathbf{x}_k$, is augmented with the unknown system parameters such that a new state vector, $\mathbf{z}_k = [\mathbf{x}_k^T, \boldsymbol{\theta}_k^T]^T$, is formed. The identification then progresses by replacing $\mathbf{x}_k$ with $\mathbf{z}_k$ in the selected filtering algorithm. Theoretically, each of these applications can develop the MAP solution over the parameters [26]. In practice, these approaches, and the joint filtering approach in particular, may experience convergence issues due to modeling the parameters as constant dynamical states, which can generate numerical errors in the algorithm such that the covariance matrix $\mathbf{P}$ becomes singular [3],[27].

2.1.1   Extended Kalman Filters

The EKF was developed in response to the problems in nonlinear state estimation that arose in configuring the orbital guidance system for the Apollo missions in the early 1960s [28]. A natural extension of the Kalman filter algorithm to the nonlinear filtering problem, the method operates by linearizing the transition and observation equations about the current state of the nonlinear system [29],[30]. A detailed outline of the approach is given in Algorithm 2.2.

Initialize with:	
	$\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0]$
	$\mathbf{P}_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$
State Prediction:	
	$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})$
	$\mathbf{F}_{\mathbf{x}_k} = [\nabla_{\mathbf{x}} \mathbf{f}^T(\mathbf{x}, \mathbf{u})]_{\mathbf{x}=\hat{\mathbf{x}}_{k-1}}^T$
	$\mathbf{P}_k^- = \mathbf{F}_{\mathbf{x}_k} \mathbf{P}_{k-1} \mathbf{F}_{\mathbf{x}_k}^T + \mathbf{Q}_{k-1}$
Observation Prediction:	
	$\hat{\mathbf{y}}_k = \mathbf{h}(\hat{\mathbf{x}}_k^-, \mathbf{u}_k)$
	$\mathbf{H}_{\mathbf{x}_k} = [\nabla_{\mathbf{x}} \mathbf{h}^T(\mathbf{x}, \mathbf{u})]_{\mathbf{x}=\hat{\mathbf{x}}_k^-}^T$
	$\mathbf{P}_{\mathbf{y}_k \mathbf{y}_k} = \mathbf{H}_{\mathbf{x}_k} \mathbf{P}_k^- \mathbf{H}_{\mathbf{x}_k}^T + \mathbf{R}_k$
	$\mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} = \mathbf{P}_k^- \mathbf{H}_{\mathbf{x}_k}^T$
Correction:	
	$\mathcal{K}_k = \mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} \mathbf{P}_{\mathbf{y}_k \mathbf{y}_k}^{-1}$
	$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathcal{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k)$
	$\mathbf{P}_k = \mathbf{P}_k^- - \mathcal{K}_k \mathbf{P}_{\mathbf{y}_k \mathbf{y}_k} \mathcal{K}_k^T$

Algorithm 2.2 The Extended Kalman Filter

Since Yun and Shinozuka demonstrated in 1980 that the method was valid for identifying parameters of linear structural models [31], a number adaptations have been proposed to better suit the algorithm to the purpose of civil structural identification. Hoshiya and Saito [32] implemented the EKF with a weighted global iteration procedure to enhance its stability and convergence on some common nonlinear structural models. Solonen et al. [33] and Yuen and Kuok [34] adjusted the algorithm for online estimation of the error covariance matrices, $\mathbf{Q}$ and $\mathbf{R}$, thereby eliminating some of the uncertainty in constructing the necessary prior information. Naets et al. [20] tackled the issue of output-only structural identification, which has since been explored in other works, such as [35],[36]. The development of these enhanced EKF algorithms has been supplemented by their practical implementation on a number of experimental structures, such as in [35],[37],[38]. Overall, EKF methods have been and continue to be one of the more widely used tools in vibration based structural identification [21]. However, some limitations still exist that prevent its broad application for structural health monitoring.

The nature of the EKF approximation, in which the system is linearized through a first-order Taylor series approximation, inherently limits the accuracy of the resulting estimation when applied to highly nonlinear systems [39]. Other challenges arise in the implementation of the method, such as the nontrivial computation of the necessary Jacobian matrices and the high sampling frequencies needed to ensure stability [40]. Therefore, despite its real-time operation and potential to scale to high-dimensional systems, the EKF appears ill-suited to the general structural health monitoring problem.

2.1.2   Sigma Point Kalman Filters

SPKFs were first introduced through the unscented Kalman filter (UKF), which was developed by Julier et al. in a series of papers from 1995-1998 [40]–[42] in answer to concerns over the accuracy and analytical complexity of the EKF. While the EKF addresses the nonlinear system directly through linear first-order approximations, the UKF instead take the stance that Gaussian distributions are far easier to approximate than arbitrary nonlinear functions [40]. The resulting filter captures better estimates of the higher order moments caused by the nonlinear transformation than the EKF, achieving 3rd order accuracy in mean estimate and 1st order accuracy in the covariance estimate [43]. Attempts to generalize and improve this implementation have led to the development of the Gauss-Hermite Kalman filter and central difference filter by Ito and Xiong

[44], the divided difference filter by Norgaard et al. [45], and the cubature Kalman filter by Arasaratnam and Haykin [46], among others [39]. This broad group of SPKF approaches, also known as Gaussian filters, is characterized by their use of deterministic Gaussian integral approximations to estimate the mean and covariance of the nonlinear transformation on the state distributions [47]. These methods progress through the data by recursively forming and propagating a deterministically selected and weighted sample of 'sigma points' through the nonlinear system equations, from which the appropriate distributional parameters can be approximated to update the estimate of the state vector. A general algorithm for this class of Kalman filters is given in Algorithm 2.3. A more exhaustive definition of the different types of SPKFs can be found in [3] and [39].

SPKFs have been found in a number of comparative studies to be superior in accuracy and stability to the EKF [21],[24],[40],[44],[48],[49]. Further increases in accuracy have been obtained by varying the deterministic approach to sigma point selection to further enhance this higher order moment approximation [50]–[55]. However, these modifications often result in an increased number of sigma points. Because the number of sigma points increases in proportion to the dimensionality of the problem, such modifications present some difficulty in scaling the method to high-dimensional physical systems. As such, several approaches have been proposed to minimize the number of sigma points while maintaining the accuracy of the algorithm [56],[57]. Other improvements to the SPKF methods have primarily addressed the treatment of the noise covariances. Kontoroupi and Smyth [22], for example, developed an approach for the online estimation of the error covariance matrices with the UKF, while Wu et al. [58] determined that augmenting the error covariances in the implementation of the UKF typically generated more accurate results than using the method in its original, non-augmented form.

Of the available SPKF methods, applications of the UKF are by far the most common in structural health monitoring applications, reinforcing its potential to be applied to general structural systems. The performance of the UKF with respect to Bouc-Wen hysteretic structures has been explored in simulation [59],[60] and in experimental tests [61], with significant success in its use for real-time dynamic model updating [62],[63] and real-time hybrid simulation [23]. Mariani and Ghisi demonstrated that the UKF can also be used to effectively determine the model of softening dynamic systems [64]. Though the scalability of this method to high-dimensional systems is still a challenge, Ghorbani and Cha [65],[66] were able to implement a combined

cubature-unsigned Kalman filter on a 10 story linear structural system, while Olivier and Smyth [67] have had some success with high-fidelity finite element models through the use of marginalization techniques. Azam [68] takes a different approach by suggesting a reduction in the dimensionality of the structural model through orthogonal decomposition. Song et al. approach another practical identification issue by developing an adaptive unscented Kalman filter to jointly estimate the process and measurement noise during the monitoring process [69]. Chen and Feng seek to further the effectiveness of these strategies by applying them in a predictive modeling context rather than a monitoring context, and do so by evaluating a central difference filtering approach to predictive modeling on an experimental bridge specimen [70].

Initialize with:
$$\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0]$$
$$\mathbf{P}_0 = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$$
$$\text{Set } \omega_j^{\dagger}, \qquad j = 0, \dots, N_{sig}$$

State Prediction:
$$\mathcal{X}_{k-1} = \text{SIG}^{\ddagger}(\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1})$$
$$\hat{\mathbf{x}}_k^- = \sum_{j=0}^{N_{sig}} \omega_j \, \mathbf{f}\left(\mathcal{X}_{k-1}^{(j)}, \mathbf{u}_{k-1}\right)$$
$$\mathbf{P}_k^- = \sum_{j=0}^{N_{sig}} \omega_j \left[\mathbf{f}\left(\mathcal{X}_{k-1}^{(j)}, \mathbf{u}_{k-1}\right) - \hat{\mathbf{x}}_k^-\right]\left[\mathbf{f}\left(\mathcal{X}_{k-1}^{(j)}, \mathbf{u}_{k-1}\right) - \hat{\mathbf{x}}_k^-\right]^T + \mathbf{Q}_{k-1}$$

Observation Prediction:
$$\mathcal{X}_k^- = \text{SIG}(\hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)$$
$$\hat{\mathbf{y}}_k = \sum_{j=0}^{N_{sig}} \omega_j \, \mathbf{h}\left(\mathcal{X}_k^{-(j)}, \mathbf{u}_k\right)$$
$$\mathbf{P}_{\mathbf{y}_k \mathbf{y}_k} = \sum_{j=0}^{N_{sig}} \omega_j \left[\mathbf{h}\left(\mathcal{X}_k^{-(j)}, \mathbf{u}_k\right) - \hat{\mathbf{y}}_k\right]\left[\mathbf{h}\left(\mathcal{X}_k^{-(j)}, \mathbf{u}_k\right) - \hat{\mathbf{y}}_k\right]^T + \mathbf{R}_k$$
$$\mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} = \sum_{j=0}^{N_{sig}} \omega_j \left[\mathbf{f}\left(\mathcal{X}_k^{-(j)}, \mathbf{u}_{k-1}\right) - \hat{\mathbf{x}}_k^-\right]\left[\mathbf{h}\left(\mathcal{X}_k^{-(j)}, \mathbf{u}_k\right) - \hat{\mathbf{y}}_k\right]^T$$

Correction:
$$\mathcal{K}_k = \mathbf{P}_{\mathbf{x}_k \mathbf{y}_k} \mathbf{P}_{\mathbf{y}_k \mathbf{y}_k}^{-1}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathcal{K}_k (\mathbf{y}_k - \hat{\mathbf{y}}_k)$$
$$\mathbf{P}_k = \mathbf{P}_k^- - \mathcal{K}_k \mathbf{P}_{\mathbf{y}_k \mathbf{y}_k} \mathcal{K}_k^T$$

†Weights associated with sigma points   ‡Sigma point generation function (deterministic)

Algorithm 2.3 Generalized Sigma Point Kalman Filter

Overall, the class of SPKFs is very promising for application to the high-dimensional, highly nonlinear problem of structural health monitoring. Though some challenges still exist, such as the

sensitivity of the method to the selection of tuning parameters [53] and the inference of prior information [22], there is a general consensus that the method is highly informative for structural identification and model updating. Further examination of these issues, as well as detailed experimental examination of algorithm scalability, will be required prior to general application.

2.2 Sampling Methods for Bayesian Filtering

Sampling approaches to Bayesian filtering, which are more commonly referred to as particle filters (PF) or sequential Monte Carlo (MC) approaches, determine an approximate solution to the Bayesian inference problem by exploiting MC methods to generate numerical approximations of analytically intractable posterior densities. The accuracy of the approximation is therefore dependent on the number of samples; the approximation converges in expectation to the true posterior as the number of samples approaches infinity [14]. These methods are particularly attractive for generalized nonlinear systems as they place no constraints on the linearity of the model nor on the form of the associated probability densities.

Sampling methods have their foundation in sequential importance sampling (SIS), which was developed in the early 1970s [71]. The basic premise of SIS is that the expectation over the posterior density $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, which is difficult to draw samples from, can be decomposed with respect to an importance distribution $\pi(\mathbf{x}_k|\mathbf{y}_{1:k})$, which is simple to draw samples from, such that

$$\int \mathrm{h}(\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{y}_{1:k}) \mathrm{d}\mathbf{x} = \int \left[\frac{\mathrm{h}(\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k})}{\pi(\mathbf{x}_k|\mathbf{y}_{1:k})}\right]\pi(\mathbf{x}_k|\mathbf{y}_{1:k})\mathrm{d}\mathbf{x} \tag{1}$$

where $\mathrm{h}(\mathbf{x}_k)$ is a nonlinear function on the states [3]. By expressing the Bayesian update equations in terms of the importance distribution, a weighted sampling approximation to the true posterior can easily be developed. Issues with degeneracy in SIS, where a single particle becomes weighted significantly more heavily than the remaining particles, limited its use prior to 1993, when Gordon, Salmond, and Smith [72] proposed a resampling step in their bootstrap filter. Since then, a number of techniques have been developed which improve upon this resampling approach [71],[73]–[77], including an adaptive resampling approaches based on an effective number of particles [78], generally referred to as sequential importance resampling (SIR) [3], which is given in Algorithm 2.4.

Initialize with:
$$\hat{\mathbf{x}}_0^{(j)} \sim p(\mathbf{x}_0), \qquad j = 1, \dots, N^\dagger$$
$$\omega_0^{(j)} = \frac{1}{N}$$

Sequential Importance Sampling ($\forall$j):
$$\hat{\mathbf{x}}_k^{(j)} \sim \pi\left(\hat{\mathbf{x}}_k^{(j)} | \mathbf{x}_{k-1}^{(j)}, \mathbf{y}_k\right)$$
$$\hat{\omega}_k^{(j)} = \omega_{k-1}^{(j)} \frac{p\left(\mathbf{y}_k \middle| \hat{\mathbf{x}}_k^{(j)}\right) p\left(\hat{\mathbf{x}}_k^{(j)} \middle| \mathbf{x}_{k-1}^{(j)}\right)}{\pi\left(\hat{\mathbf{x}}_k^{(j)} | \mathbf{x}_{k-1}^{(j)}, \mathbf{y}_k\right)}$$
$$\omega_k^{(j)} = \frac{\hat{\omega}_k^{(j)}}{\sum_{j=1}^N \hat{\omega}_k^{(j)}}$$
$$N_{eff} = \frac{1}{\sum_{j=1}^N \left(\omega_k^{(j)}\right)^2}$$

If $N_{eff} \ll N$:
$$\text{Resample}^\ddagger \left(\hat{\mathbf{x}}_k^{(j)}, \omega_k^{(j)}\right)$$
$$\text{Regularize}^\ddagger \left(\hat{\mathbf{x}}_k^{(j)}, \omega_k^{(j)}\right)$$
$$\omega_k^{(j)} = \frac{1}{N}$$

Else:
$$\mathbf{x}_k^{(j)} = \hat{\mathbf{x}}_k^{(j)}$$

†Number of particles $\qquad\qquad$ ‡Several strategies available

Algorithm 2.4 Sequential Importance Resampling Particle Filter

The selection of the importance distribution has a nontrivial effect on the performance of the filter. Using the optimal importance distribution [73],[79]

$$\pi(\mathbf{x}_k | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{1:k}), \tag{2.7}$$

allows for a filtering strategy that is with minimal variance in the sampling distributions and generally quick convergence with a relatively small number of particles. When the optimal importance distribution cannot be directly used, other distributions have been used successfully in its place. Gordon et al. [72], for example, use the transition probability $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1})$ as the importance distribution. This model significantly simplifies the implementation of the particle filter algorithm at the cost of increasing the variance of the sampling distribution, and therefore requiring more particles for concise convergence. Alternatively, Pitt and Shephard [80] proposed

an auxiliary sequential importance resampling (SIR) filter which mimics the optimal importance distribution by resampling at step $k-1$ using the measurement available at time t_k. Other common approaches for the selection of the importance distribution emphasize local linearization through by combining the particle filter approach with various Kalman filtering methods. These combination approaches will be discussed further in Section 2.3.

Sample impoverishment represents the final major computational issue that impedes the implementation of the particle filter, and can be particularly detrimental for joint state and parameter estimation applications. This effect occurs when the noise variance on a particular state is set to be, or converges to, a very small value. When this occurs, many of the particles will have exactly the same value, yielding what is essentially a point distribution that is propagated forward in time and is not indicative of the true statistics of the state. Techniques such as 'roughening', where a slight artificial increase in variance is introduced to the state, were initially proposed and used to great effect by several researchers [72],[73],[81]. Chatzi and Smyth [82], for example, used this approach to great effect in determining a predictive model for a three story structure with a Bouc-Wen hysteretic component. However, the non-trivial issue of selecting a roughening variance which provides sufficient reduction of sample impoverishment without significantly increasing the sampling variance led researchers to generate a variety of other approaches to combat this issue, such as evolutionary mutation [83], resample-move algorithms, regularization, and MCMC steps [73],[79]. Most recently Rao-Blackwellization approaches, which allow for computations on some of the states to be performed analytically, have been shown to be particularly effective in approaching this issue [73],[74],[84]. Oliver and Smyth [85] and Storvik [86] have had particular success with this approach as it applies to the identification of static parameters in structural systems.

Research into the application of particle filtering techniques for the predictive modeling of structural systems has had some success, though application of Kalman filtering techniques are still far more common. Wang et al. for example, developed a method to extract the fundamental mode of a bridge from ambient vehicle excitations using a particle filter approach [87], Wan et al. develop a modified particle filter for joint input-state-parameter identification [88], and Xue et al. applied the auxiliary particle filter to the problem of inferring abrupt parameter changes [89]. Olivier and Smyth addressed some practical uncertainties in these methods through a study on their effectiveness with systems having various identifiability properties, which determined that

the particle filter is informative in developing distributional characteristics of parameters in unidentifiable and locally identifiable structural systems [90]. Several researchers, such as Erazo and Hernandez [49] and Azam [68], have performed comparison studies which demonstrate that though the particle filter can be useful in some cases, it nonlinear Kalman and combination approaches can often give better stability and computational efficiency.

Despite the success of the particle filter in relatively low-dimensional systems, degeneracy and impoverishment issues are known to be exacerbated in high-dimensional systems. The convergence of the algorithm to the expectation of the true posterior is guaranteed through the central limit theorem as the number of samples approaches infinity [3], regardless of dimension. However, the number of samples required to maintain the accuracy of the approximation have been shown to increase exponentially with the dimensionality of the system [85],[91]. Various solutions have been suggested to mitigate these issues, including marginalization [3],[71],[85] and dimension reduction [68] approaches. Though these solutions have shown promise in mitigating the issues associated with the dimensionality of the solution, they remain computationally intensive, limiting their effectiveness for structural health monitoring applications.

2.3   Combination Methods

In addition to traditional implementations of analytical and sampling filters, several adaptive Bayesian filtering methods have been generated which seek to take advantage of the strengths of these respective methods. This section seeks to provide a brief summary of these methods, with examples of their application to predictive modeling in structural health monitoring.

2.3.1   Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) was developed by Evensen in 1994 [92] as a Monte Carlo based approach to solve some of the issues that arise with the EKF. Specifically, Evensen seeks to use the EnKF to address the computational complexity of covariance matrix propagation in large-scale systems and the neglect of higher-order nonlinearities seen in the EKF. The EnKF achieves these goals by first generating an initial 'ensemble' of states by selecting a best-guess estimate of the initial states and then adding perturbations to form other ensemble members. This initial ensemble is then propagated forward through the filter, being used in the predictor and

corrector steps to form Gaussian error statistics to develop the Kalman gain on the stochastic dynamical system. A detailed look on the theoretical developments, major adaptations, and applications of this method are available from Evensen in [93].

The EnKF has shown improved performance over the EKF in several areas of research, and in particular with respect to structural health monitoring problems. Practical applications of the EnKF include the work by Akita et al. [94] and Slika and Saad [95], who respectively use the filter to perform joint state and noise identification on a deployable frame structure and employ the filter to develop a predictive model of chloride-induced corrosion in reinforced concrete structures. Ghanem and Ferro [96] directly compare the EKF and EnKF in their ability to develop non-parametric models of complex nonlinear systems with non-Gaussian process noise, finding that the EnKF to have a robust performance in damage identification and state estimation. Perhaps most enlightening is the comprehensive comparison study by Erazo et al. [49], in which he uses the case study of a simulated 10-story structure with nonlinear elements to compare the state estimation performance of the EKF, UKF, EnKF, PF. He found that the UKF and EnKF generally performed better than the other filters, though the UKF had the best performance overall.

2.3.2   Monte Carlo Kalman Filter

Similar to the EnKF, the Monte Carlo Kalman filter (MCKF) develops the Gaussian error statistics of the Kalman filter scheme through samples of the marginal prior. The key distinction between the methods, and indeed between the MCKF and the particle filter, is that new samples are drawn in each predictor and corrector phase instead of the samples being propagated through the filter from some initial draw [39]. Applications of the MCKF are uncommon in structural health monitoring applications.

2.3.3   Gaussian Particle Filter

As described in Section 2.2, proposal distributions for the particle filter govern the stability and convergence rate of the approximation. Ideal implementations of the particle filter take advantage of the optimal proposal distribution, given in Eq. (2.7). However, determining this proposal distribution in practical filtering scenarios can be extremely difficult, resulting in the use of simpler proposals, such as the transition probability used by Gordon et al. [72], which are prone to fail

when new measurements appear in the tail of the prior or if the likelihood function is has significantly lower variance than the prior.

In response to these issues, researchers have proposed several adaptations of the particle filter method which augment the proposal distribution with Kalman filter updates. Initial implementations worked with the EKF [97], but van der Merwe et al. found greater success in using the UKF to augment the proposal distribution [98]. The methods proceed by propagating each particle of the filter through the nonlinear Kalman filter to develop a posterior approximation. Particles are then samples, and are weighted from a proposal distribution characterized by that posterior, $\pi\left(x_t^{(i)}\middle|x_{0:t-1}^{(i)}, y_{1:t}\right) = N(\bar{x}_t^{(i)}, \bar{P}_t^{(i)})$, where $i = 1, \dots, N$ refers to each individual particle. The advantage of this approach over the transition probability is that it incorporates the current observation into the proposal distribution, essentially eliminating issues in convergence due to an inability to account for overly precise or unlikely measurements. Several variants of this approach have been introduced which use the wide variety of sigma point Kalman filters to generate a couple Kalman-particle filter approach [99]–[103].

A few examples the implementation of these methods exist structural health monitoring applications. For example, Azam and Mariani [18] improve upon the scalability and accuracy of the unscented particle filter on simulated structural systems with a dual EKF-PF approach. Additionally, Monsouri et al. [104] proposed an iterated square-root unscented particle filter, and subsequently improved upon it with an iterated square-root central difference Kalman particle filter [100], validating the performance of both of these methods against several other Kalman and particle filter variance on simulated structural systems. These researchers found, in general, great improvement in these methods over traditional sigma-point Kalman and particle filter approaches.

2.4   Other Topics in Practical Predictive Modeling

To be effective in a practical health monitoring setting, Bayesian filtering methods must also address a number of non-algorithmic considerations. Some issues of note that are currently being addressed by active research are summarized below. Though these issues characterize much of the current research narrative in the practical implementation of Bayesian predictive modeling tools for structural health monitoring, this dissertation addresses in detail only a small subset of these topics, particularly the selection of excitation signals for system identifiability, as will be discussed

in detail in Chpt. 5. I instead focus in more detail on issues of algorithm robustness which are not well studied in the current literature, such as the selection of the prior distributions on the states and parameters, sensitivity to inaccurate model error predictions, and the adaptability to varying levels of sensors noise. These issues are explored in great detail in Chpts. 5, 6, and 7.

2.4.1 Experimental Design for System Identifiability

Experimental design includes the selection of excitation signals and sensor locations to improve the identifiability of the experimental system. These topics have been addressed in a number of ways for structural systems, including through Bayesian analysis frameworks.

Traditionally, experimental identifiability from the standpoint of a given experimental system from either a structural, practical, or sensitivity-based standpoint. Structural methods directly analyze the system under the assumption of a perfect model with no measurement noise in the response and are typically mathematically intensive [105]. With the advent of a variety of generalized software tools [106]–[108] and specific methods for engineering applications [109]–[111], these methods have become more accessible to the general user. However, experimental systems exhibit behaviors that often deviate significantly from the computational models, leading to which are structurally identifiable, but practically unidentifiable [90],[105],[112]. Practical methods seek to overcome some of these limitations by assessing the identifiability of nonlinear system models in the presence of noise, though the absence of modeling errors is still assumed. Several methods have been proposed for offline analysis prior to experimentation [113],[114] and online assessment operating in parallel with Bayesian inference techniques [7],[90]. Sensitivity-based approaches strike a balance between the other methods by interpreting identifiable parameters as those which the model is most sensitive to, or which produce the most variation in its output. Though these approaches share the same assumptions as the structural methods regarding model form and measurement noise, they also require the user to identify nominal parameter values or ranges over which to evaluate the sensitivity, as with practical methods [105]. Applications of this class of methods for the purpose of establishing identifiability almost exclusively use local analysis methods [114],[115], which are not computationally intensive, but are less effective for models with severe nonlinearities and parameters of varying levels of uncertainty [116]. Global sensitivity analysis methods [117]–[122] can offer a greater depth of

information and are explored in Chpt. 5 as an alternative way to establish identifiability with respect to excitation signal selection.

Specific attention has also been given to alternative methods for sensor network optimization for structural dynamics, resulting in a number of new methods and adaptations of the above techniques. Sensor network optimization techniques seek to develop, either sequentially or as a batch, the optimal network of sensors to promote identifiability in a system with respect to a particular metric. A clear summary of the classes of metrics used for optimization are given by Yuen and Kuok [123] in their paper on multi-type sensor network optimization. They classify these methods as being

- modal parameter based, in which the network is optimized for the determination of modal parameters, [124]–[126];
- response reconstruction based, in which the network is optimized to minimize the mean square error in the reconstructed structural response, [127]–[130];
- energy based, in which the network is optimized to maximize the kinetic energy, and therefore the modal contributions, of the observed structure [131],[132]; or
- information based, in which the network is optimized through either deterministic or Bayesian means to maximize the information delivered by each sensor [124]–[126],[133]–[139].

As is clear by the numerous examples associated with these methods, the topic of sensor network optimization is still an open question in experimental design, though methods using Bayesian information theory for optimization do seem to be an effective and simple approach, particularly for sequential sensor selection and multi-type sensor systems.

2.4.2 Model Selection

The common objective of predictive modeling problems is to efficiently find the most useful model to capture the desired aspect of the system response behavior, given the knowledge that no computational model will be able to capture all aspects of the response perfectly. Model selection algorithms typically seek to achieve this by proposing a single flexible model or set of increasingly complex models which are tuned and eventually narrowed down to a single representative system model as observations of the system become available. Bayesian techniques for model selection are particularly efficient, as the structure of the Bayesian algorithm enforces the principle of

Occam's Razor, that the selected model should be the most simple version which adequately describes the system behavior, thereby reducing the possibility of over-fitting [140].

Current practice in predictive structural modeling tends to favor the Bayesian approach to model selection. Several authors have adapted Bayesian model selection techniques for structural systems, such as Kerschen et al., who used a simple Markov Chain Monte Carlo variant [6], Yuen, who combined Bayesian model selection techniques with artificial neural networks for structural damage detection [9], Worden and Hensmen, who apply the approach to Duffing and Bouc-Wen hysteretic systems [8], and De et al., who seek to increase the computational efficiency of the technique for high-dimensional systems [141]. Beck et al. in particular have proposed several incremental improvements to Bayesian model selection algorithms for structural systems over the past 20 years, moving from a basic Bayes-theoretic approach [4] to take advantage of new techniques such as transitional Markov Chain Monte Carlo [7], asymptotically independent Markov sampling [142], hierarchical sparse Bayesian learning and Gibbs sampling [11],[143], and second-order Langevin Monte Carlo [144]. Though these methods primarily emphasize batch techniques, several real-time model selection algorithms have also been proposed by Chatzi et al. [61], Yuen and Mu [145] and Kontoroupi and Smyth [146].

Regardless of the quality of the models which result from these model selection techniques, there will inevitably be discrepancies between the true experimental system and the selected computational model. Quantifying and mitigating the effect of these uncertainties form the other side of the model selection issue. Kennedy and O'Hagan first defined the model discrepancy in detail in 2001 [2], highlighting the issues involved in calibrating computational models in the face of uncertainty. Arendt et al. [1] and Brynjarsdottir and O'Hagan [147] further clarify the impact of misrepresenting this model uncertainty on the calibration of model parameters. When the model uncertainty is either extremely large or inadequately quantified, the remaining computational model may suffer from miscalibration, where the identified model parameters are not representative of the physical system. Current efforts for mitigating these issues include grey-box modeling [148], Bayesian history matching [149], and detailed uncertainty quantification techniques [150],[151].

2.4.3 Input Identification

The joint identification of the input signal is currently of vital interest to the structural health monitoring community, particularly in applications related to wind, wave, and vehicle vibrations. These applications where the exact location of the input force is often unknown, pose a particular challenge for predictive modeling. These techniques can be most easily categorized by their approach to joint identification, i.e. whether they perform only input identification, input-state identification, or input-state-parameter identification.

Input-only identification techniques tend to focus on the robustness and practical implementation of the general class of methods, in preparation for their application in conjunction with other identification techniques. For example, Maes et al. derive the specific conditions and sensor network configurations necessary for the identifiability of the system input, with particular consideration of the stability and uniqueness of the identification [152]. Nguyen et al. [153] develop two approaches, using the minimum variance unbiased filter and an augmented Kalman filter, with the intent of increasingly the computational efficiency and accuracy of the force identification. Song et al. [154] use a square-root cubature Kalman filter to estimate both the magnitude and location of input forces in experimental and simulated application. Though such studies are extremely helpful in determining ideal practical implementations for force identification, most of the work in this field emphasize joint identification techniques.

Joint input-state identification is well studied in its application to civil infrastructure systems, and has been found to be particularly useful for condition monitoring and damage detection. Many applications us a state augmentation approach to develop the joint inference on the states and input forces, such as Lourens et al. [155] and He et al. [35] who respectively use state augmentation with a standard and extended Kalman filter. Nayek et al. [156] also use an augmented EKF to identify the state and input forces, but vary the system characterization by modeling the input force as a Gaussian process. Azam et al. [19] improve the stability of the state augmentation approach using a dual UKF, where the states are estimated using a UKF and the inputs are inferred used a parallel Kalman filter. Gillijns and De Moor [157] develop a novel filter to approach the problem which separates the recursive process between the state and input estimation. Maes et al. improve upon and apply this approach to civil structural systems in [158] and further adapt it to consider time-delay in force estimation due to non-collocated sensing in [159].

Joint input-state-parameter identification adds another level of difficulty to the inference process, but is extremely helpful in generating predictive structural models in practical monitoring scenarios. Yuen and Kuok [34], Ebrahimian et al. [160], and Naets et al. [20] all demonstrated successful joint input-state-parameter identification through variations on the augmented EKF, while Lei et al. [161] found greater success in adapting an augmented UKF approach. Al-Hussein and Haldar [162] seek to improve the accuracy and stability of joint input-state-parameter identification through the combined use of the UKF and an iterative least squares technique and successfully illustrate the approach on a number of simulated nonlinear structural systems in [163]. They further improve the approach to identify damage states in high-dimensional systems through the incorporation of a weighted global iteration procedure in [164]. Dertimanis et al. [165] approach the same problem through the combination of a dual and augmented UKF, where a Kalman filter is used to estimate the input force in parallel with a UKF to identify the parameters and states. This method was further validated through a numerical study on fatigue damage prediction in steel structural systems [166]. Erazo and Nagarajaiah [167] also endeavor to avoid the instability caused by state augmentation, but achieve it instead through a re-parameterize of the state predictive distribution. They then implement an offline method which combines the UKF and MCMC sampling to generate a recursive algorithm which can handle arbitrary priors and posteriors. Rogers et al. [168] take work in a different direction through a Gaussian process latent force modeling approach, through which they infer the parameters and latent force using MCMC inference.

Despite the multitude of methods which have been developed for output-only inference, there is no clear consensus on a set of effective approaches for predictive structural modeling. The research community is very active on this topic and will continue to unravel the practical issues barring these methods from real-world implementation.

2.5   Conclusions

One of the key points to draw from this review is that there has been an immense amount of work done on the topic of Bayesian predictive modeling without a clear collection of the innovations in algorithm development and practical usage into a single set of techniques that can be generally applied. From this review it should be evident that no single technique will work well for the identification of an arbitrary structural system, let alone in a practical setting where issues such as

identifiability, model uncertainty, and unknown input forces are a reality. Though the definition of a suite of effective predictive modeling techniques for structural engineering is still under development, the remainder of this dissertation seeks to address some of the core issues in algorithmic robustness and scalability which could bring us closer to this goal.

2.6 References

1. Arendt, P. D., Apley, D. W. & Chen, W. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *J. Mech. Des. Trans. ASME* **134,** 1–12 (2012).

2. Kennedy, M. C. & O'Hagan, A. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **63,** 425–464 (2001).

3. Sarkka, S. *Bayesian filtering and smoothing*. *Cambridge University Press* (Cambridge University Press, 2013). doi:10.1017/CBO9781139344203

4. Beck, J. L. & Yuen, K. V. Model selection using response measurements: Bayesian probabilistic approach. *J. Eng. Mech.* **130,** 192–203 (2004).

5. Beck, J. L. Bayesian system identification based on probability logic. *Struct. Control Heal. Monit.* **17,** 825–847 (2010).

6. Kerschen, G., Golinval, J.-C. & Hemez, F. M. Bayesian model screening for the identification of nonlinear mechanical structures. *J. Vib. Acoust.* **125,** 389–397 (2003).

7. Muto, M. & Beck, J. L. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *J. Vib. Control* **14,** 7–34 (2008).

8. Worden, K. & Hensman, J. J. Parameter estimation and model selection for a class of hysteretic systems using Bayesian inference. *Mech. Syst. Signal Process.* **32,** 153–169 (2012).

9. Yuen, K. V. Recent developments of Bayesian model class selection and applications in civil engineering. *Struct. Saf.* **32,** 338–346 (2010).

10. Behmanesh, I., Moaveni, B., Lombaert, G. & Papadimitriou, C. Hierarchical Bayesian model updating for structural identification. *Mech. Syst. Signal Process.* **64–65,** 360–376 (2015).

11. Huang, Y., Beck, J. L. & Li, H. Hierarchical sparse Bayesian learning for structural damage detection: Theory, computation and application. *Struct. Saf.* **64,** 37–53 (2017).

12. Behmanesh, I., Moaveni, B. & Papadimitriou, C. Probabilistic damage identification of a designed 9-story building using modal data in the presence of modeling errors. *Eng. Struct.* **131,** 542–552 (2017).

13. Huang, Y., Shao, C., Wu, B., Beck, J. L. & Li, H. State-of-the-art review on Bayesian inference in structural system identification and damage assessment. *Adv. Struct. Eng.* **22,** 1329–1351 (2019).

14. Bishop, C. M. *Pattern recognition and machine learning*. *Pattern Recognition* (Springer, 2006). doi:10.1117/1.2819119

15. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82,** 35–45 (1960).

16. Wan, E. A. & Nelson, A. T. Dual Kalman filtering methods for nonlinear prediction, smoothing, and estimation. in *Advances in Neural Information Processing Systems* 793–799 (1997).

17. Nelson, A. T. *Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods*. (2000).

18. Eftekhar Azam, S. & Mariani, S. Dual estimation of partially observed nonlinear structural systems: A particle filter approach. *Mech. Res. Commun.* **46,** 54–61 (2012).

19. Azam, S. E., Chatzi, E. & Papadimitriou, C. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mech. Syst. Signal Process.* **60,** 866–886 (2015).

20. Naets, F., Croes, J. & Desmet, W. An online coupled state/input/parameter estimation approach for structural dynamics. *Comput. Methods Appl. Mech. Eng.* **283,** 1167–1188 (2015).

21. Wu, M. & Smyth, A. W. Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Struct. Control Heal. Monit.* **14,** 971–990 (2007).

22. Kontoroupi, T. & Smyth, A. W. Online noise identification for joints state and parameter estimation of nonlinear systems. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2,** 1–12 (2016).

23. Shao, X., Mueller, A. & Mohammed, B. A. Real-Time hybrid simulation with online model updating: methodology and implementation. *J. Eng. Mech.* **142,** 1–19 (2015).

24. Wan, E., van der Merwe, R. & Nelson, A. T. Dual estimation and the unscented transformation. in *Advances in Neural Information Processing Systems* 666–672 (2000).

25. Wan, E. A. & Nelson, A. T. in *Kalman Filtering and Neural Networks* (ed. Haykin, S.) 123–173 (John Wiley & Sons, 2001).

26. Tremain Nelson, A. Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods. (Oregon Health and Science University, 2000).

27. Lund, A., Dyke, S. J., Song, W. & Bilionis, I. Identification of an experimental nonlinear energy sink device using the unscented Kalman filter. *Mech. Syst. Signal Process.* **136,** (2020).

28. Schmidt, S. F. The Kalman filter - Its recognition and development for aerospace applications. *J. Guid. Control* **4,** 4–7 (1981).

29. Sorenson, H. W. Kalman filtering techniques. *Adv. Control Syst.* **3,** 219–292 (1966).

30. Jazwinski, A. in *Stochastic Processes and Filtering Theory* (Academic Press, 1970). doi:10.1016/S0076-5392(09)60377-5

31. Yun, C.-B. & Shinozuka, M. Identification of nonlinear structural dynamic systems. *J. Struct. Mech.* **8,** 187–203 (1980).

32. Hoshiya, M. & Saito, E. Structural identification by extended Kalman filter. *J. Eng. Mech.* **110,** 1757–1770 (1984).

33. Solonen, A., Hakkarainen, J., Ilin, A., Abbas, M. & Bibov, A. Estimating model error covariance matrix parameters in extended Kalman filtering. *Nonlinear Process. Geophys.* **21,** 919–927 (2014).

34. Yuen, K. V. & Kuok, S. C. Online updating and uncertainty quantification using nonstationary output-only measurement. *Mech. Syst. Signal Process.* **66,** 62–77 (2016).

35. He, J., Xu, Y. L., Zhan, S. & Huang, Q. Structural control and health monitoring of building structures with unknown ground excitations: Experimental investigation. *J. Sound Vib.* **390,** 23–38 (2017).

36. Wang, D. & Haldar, A. System identification with limited observations and without input. *J. Eng. Mech.* **123,** 504–511 (1997).

37. Ghanem, R. G., Gavin, H. & Shinozuka, M. *Experimental verification of a number of structural system identification algorithms*. *Technical Report NCEER-91-0024* (1991).

38. Zhou, L., Wu, S. & Yang, J. N. Experimental study of an adaptive extended Kalman filter for structural damage identification. *J. Infrastruct. Syst.* **14,** 42–51 (2008).

39. Haug, A. J. *Bayesian Estimation and Tracking: A Practical Guide*. (John Wiley & Sons, 2012).

40. Julier, S. J., Uhlmann, J. K. & Durrant-Whyte, H. F. A new approach for filtering nonlinear systems. in *Proceedings of American Control Conference* **3,** 1628–1632 (1995).

41. Julier, S. J. & Uhlmann, J. K. A new extension of the Kalman filter to nonlinear systems. in *Signal processing, sensor fusion, and target recognition VI* **3068,** 182–194 (International Society for Optics and Photonics, 1997).

42. Julier, S. J. A skewed approach to filtering. in *Signal and Data Processing of Small Targets* 271–282 (International Society for Optics and Photonics, 1998).

43. Julier, S. J. & Uhlmann, J. K. Unscented filtering and nonlinear estimation. *Proc. IEEE* **92,** 401–422 (2004).

44. Ito, K. & Xiong, K. Gaussian filters for nonlinear filtering problems. *IEEE Trans. Automat. Contr.* **45,** 910–927 (2000).

45. Nørgaard, M., Poulsen, N. K. & Ravn, O. New developments in state estimation for nonlinear systems. *Automatica* **36,** 1627–1638 (2000).

46. Arasaratnam, I. & Haykin, S. Cubature kalman filters. *IEEE Trans. Automat. Contr.* **54,** 1254–1269 (2009).

47. Wu, Y., Hu, D., Wu, M. & Hu, X. A numerical-integration perspective on gaussian filters. *IEEE Trans. Signal Process.* **54,** 2910–2921 (2006).

48. van der Merwe, R., Wan, E. A. & Julier, S. I. Sigma-point Kalman filters for nonlinear estimation and sensor-fusion - applications to integrated navigation. in *AIAA Guidance, Navigation, and Control Conference and Exhibit* 5120 (2004). doi:10.2514/6.2004-5120

49. Erazo, K. & Hernandez, E. M. Uncertainty quantification of state estimation in nonlinear structural systems with application to seismic response in buildings. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2,** 1–17 (2016).

50. van Zandt, J. R. A more robust unscented transform. in *Signal and Data Processing of Small Targets* 371–381 (International Society for Optics and Photonics, 2001). doi:10.1117/12.492745

51.   Julier, S. J. The scaled unscented transformation. in *Proceedings of the American Control Conference* 4555–4559 (2002).

52.   Tenne, D. & Singh, T. The higher order unscented filter. in *American Control Conference* 2441–2446 (2003).

53.   Duník, J., Simandl, M. & Štraka, O. Unscented kalman filter: Aspects and adaptive setting of scaling parameter. *IEEE Trans. Automat. Contr.* **57,** 2411–2416 (2012).

54.   Liu, J., Wang, Y. & Zhang, J. A linear extension of unscented Kalman filter to higher-order moment-matching. in *Proceedings of the IEEE Conference on Decision and Control* 5021–5026 (IEEE, 2014).

55.   Wang, Y., Liu, J., Yang, W. & Zhang, J. A heuristic for sigma set selection of UKF. in *International Conference on Signal Processing* 72–77 (IEEE, 2014).

56.   Julier, S. J. & Uhlmann, J. K. Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. in *American Control Conference* 887–892 (IEEE, 2002).

57.   Jia, B., Xin, M. & Cheng, Y. High-degree cubature Kalman filter. *Automatica* **49,** 510–518 (2013).

58.   Wu, Y., Hu, D., Wu, M. & Hu, X. Unscented Kalman filtering for additive noise case: augmented versus nonaugmented. in *American Control Conference* 4051–4055 (IEEE, 2005). doi:10.1109/LSP.2005.845592

59.   Omrani, R., Hudson, R. E. & Taciroglu, E. Parametric identification of nondegrading hysteresis in a laterally and torsionally coupled building using an unscented Kalman filter. *J. Eng. Mech.* **139,** 452–468 (2013).

60.   Xie, Z. & Feng, J. Real-time nonlinear structural system identification via iterated unscented Kalman filter. *Mech. Syst. Signal Process.* **28,** 309–322 (2012).

61.   Chatzi, E. N., Smyth, A. W. & Masri, S. F. Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty. *Struct. Saf.* **32,** 326–337 (2010).

62.   Song, W. & Dyke, S. Development of a cyber-physical experimental platform for real-time dynamic model updating. *Mech. Syst. Signal Process.* **37,** 388–402 (2013).

63.   Song, W. & Dyke, S. J. Real-time dynamic model updating of a hysteretic structural system. *J. Struct. Eng.* **140,** 1–14 (2014).

64. Mariani, S. & Ghisi, A. Unscented Kalman filtering for nonlinear structural dynamics. *Nonlinear Dyn.* **49,** 131–150 (2007).

65. Ghorbani, E. & Cha, Y. J. An iterated cubature unscented Kalman filter for large-DoF systems identification with noisy data. *J. Sound Vib.* **420,** 21–34 (2018).

66. Cha, Y.-J. & Ghorbani, E. Identification of large-scale systems with noisy data using an iterated cubature unscented Kalman filter. 64 (2018). doi:10.1117/12.2295952

67. Olivier, A. & Smyth, A. W. A marginalized unscented Kalman filter for efficient parameter estimation with applications to finite element models. *Comput. Methods Appl. Mech. Eng.* **339,** 615–643 (2018).

68. Azam, S. E. *Online damage detection in structural systems: applications of proper orthogonal decomposition, and Kalman and particle filters*. (Springer Science+Business Media, 2014). doi:10.1007/978-3-319-02559-9

69. Song, M., Astroza, R., Ebrahimian, H., Moaveni, B. & Papadimitriou, C. Adaptive Kalman filters for nonlinear finite element model updating. *Mech. Syst. Signal Process.* **143,** 106837 (2020).

70. Chen, Y. & Feng, M. Q. Structural health monitoring by recursive bayesian filtering. *J. Eng. Mech.* **135,** 231–242 (2009).

71. Cappe, O., Godsill, S. J. & Moulines, E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proc. IEEE* **95,** 899–924 (2007).

72. Gordon, N. J., Salmond, D. J. & Smith, A. F. M. Novel approach to nonlinear / non-Gaussian Bayesian state estimation. *IEE Proc. (Radar Signal Process.* **140,** 107–113 (1993).

73. Doucet, A., de Freitas, N. & Gordon, N. *Sequential Monte Carlo methods in practice*. (Springer, 2001). doi:10.1198/tech.2003.s23

74. Doucet, A., Godsill, S. & Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10,** 197–208 (2000).

75. Doucet, A. & Johansen, A. M. in *Handbook of nonlinear filtering* 656–704 (2009).

76. Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50,** 174–188 (2002).

77. Liu, J. S. & Chen, R. Sequential monte carlo methods for dynamic systems. *J. Am. Stat. Assoc.* **93,** 1032–1044 (1998).

78. Liu, J. S. & Chen, R. Blind deconvolution via sequential imputations. *J. Am. Stat. Assoc.* **90,** 567–576 (1995).

79. Ristic, B., Arulampalam, S. & Gordon, N. *Beyond the Kalman filter: particle filters for tracking applications*. (Artech House, 2004).

80. Pitt, M. K. & Shephard, N. Filtering via simulation: auxiliary particle filters. *J. Am. Stat. Assoc.* **94,** 590–599 (1999).

81. Gustafsson, F. Particle filter theory and practice with positioning applications. *IEEE Aerosp. Electron. Syst. Mag.* **25,** 53–81 (2010).

82. Chatzi, E. N. & Smyth, A. The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Struct. Control Heal. Monit.* **16,** 99–123 (2009).

83. Chatzi, E. & Smyth, A. Particle filter scheme with mutation for the estimation of time-invariant parameters in structural health monitoring applications. *Struct. Control Heal. Monit.* **20,** 1081–1095 (2013).

84. Casella, G. & Robert, C. P. Rao-blackwellisation of sampling schemes. *Biometrika* **83,** 81–94 (1996).

85. Olivier, A. & Smyth, A. W. Particle filtering and marginalization for parameter identification in structural systems. *Struct. Control Heal. Monit.* **24,** 1–25 (2017).

86. Storvik, G. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Trans. Signal Process.* **50,** 281–289 (2002).

87. Wang, H., Nagayama, T., Nakasuka, J., Zhao, B. & Su, D. Extraction of bridge fundamental frequency from estimated vehicle excitation through a particle filter approach. *J. Sound Vib.* **428,** 44–58 (2018).

88. Wan, Z., Wang, T., Li, S. & Zhang, Z. A modified particle filter for parameter identification with unknown inputs. *Struct. Control Heal. Monit.* **25,** 1–14 (2018).

89. Xue, S., Tang, H. & Xie, Q. Structural damage detection using auxiliary particle filtering method. *Struct. Heal. Monit.* **8,** 101–112 (2009).

90. Olivier, A. & Smyth, A. W. On the performance of online parameter estimation algorithms in systems with various identifiability properties. *Front. Built Environ.* **3,** 1–18 (2017).

91. Snyder, C., Bengtsson, T., Bickel, P. & Anderson, J. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136,** 4629–4640 (2008).

92. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99,** (1994).

93. Evensen, G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53,** 343–367 (2003).

94. Akita, T., Takaki, R. & Kogiso, N. An adaptive estimation of nonlinear structural deformations by using the ensemble Kalman filter. *Trans. Japan Soc. Aeronaut. Sp. Sci. Aerosp. Technol. Japan* **14,** 43–49 (2016).

95. Slika, W. & Saad, G. An Ensemble Kalman Filter approach for service life prediction of reinforced concrete structures subject to chloride-induced corrosion. *Constr. Build. Mater.* **115,** 132–142 (2016).

96. Ghanem, R. & Ferro, G. Health monitoring for strongly non-linear systems using the Ensemble Kalman Filter. *Struct. Control Heal. Monit.* **13,** 245–259 (2006).

97. De Freitas, J. F. G., Niranjan, M., Gee, A. H. & Doucet, A. Sequential Monte Carlo methods to train neural network models. *Neural Comput.* **12,** 955–993 (2000).

98. Van Der Merwe, R., Ando De Freitas, N., Doucet, A. & Wan, E. The Unscented Particle Filter. *Adv. Neural Inf. Process. Syst.* 584–590 (2001).

99. Zandara, S. & Nicholson, A. Square root unscented particle filtering for grid mapping. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **5866 LNAI,** 121–130 (2009).

100. Chaabane, M., Mansouri, M., Nounou, H., Nounou, M. & Hamida, A. Ben. Enhanced particle filter for states and parameters estimation in structural health monitoring applications. *J. Civ. Struct. Heal. Monit.* **6,** 521–543 (2016).

101. Ze-jian, Y., Nan-ning, Z. & Xin-chun, J. The Gauss-Hermite Particle Filter. *Acta Electron. Sin.* **31,** 970–973 (2003).

102. Feng, S. & Li-jun, T. Cubature particle filter. *Syst. Eng. Electron.* **33,** 2554–2557 (2011).

103. Mu, J., Cai, Y. L. & Zhang, J. M. Square root cubature particle filter. *Adv. Mater. Res.* **219–220,** 727–731 (2011).

104. Mansouri, M., Avci, O., Nounou, H. & Nounou, M. A comparative assessment of nonlinear state estimation methods for structural health monitoring. in *Proceedings of the Society for Experimental Mechanics Series* **3,** 45–54 (2015).

105. Miao, H., Xia, X., Perelson, A. S. & Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **53,** 3–39 (2011).

106. Bellu, G., Saccomani, M. P., Audoly, S. & D'angì, L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **88,** 52–61 (2007).

107. Chis, O.-T., Banga, J. R. & Balsa-Canto, E. GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* **27,** 2610–2611 (2011).

108. Meshkat, N., Er-zhen Kuo, C. & DiStefano III, J. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: A novel web implementation. *PLoS One* **9,** 110261 (2014).

109. Chatzis, M. N., Chatzi, E. N. & Smyth, A. W. On the observability and identifiability of nonlinear structural systems. *Struct. Control Heal. Monit.* **22,** 574–593 (2015).

110. Grandjean, T. R. B., McGordon, A. & Jennings, P. A. Structural identifiability of equivalent circuit models for li-ion batteries. *Energies* **10,** 1–16 (2017).

111. Maes, K., Chatzis, M. N., Vandebril, R. & Lombaert, G. Observability of modally reduced order models with unknown parameters. *Mech. Syst. Signal Process.* **146,** 106993 (2020).

112. Margaria, G., Riccomagno, E. & White, L. J. Structural identifiability analysis of some highly structured families of statespace models using differential algebra. *J. Math. Biol.* **49,** 433–454 (2004).

113. Miao, H. *et al.* Modeling and estimation of kinetic parameters and replicative fitness of HIV-1 from flow-cytometry-based growth competition experiments. *Bull. Math. Biol.* **70,** 1749–1771 (2008).

114. Rodriguez-Fernandez, M., Mendes, P. & Banga, J. R. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems* **83,** 248–265 (2006).

115. Quaiser, T. & Mönnigmann, M. Systematic identifiability testing for unambiguous mechanistic modeling - Application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models. *BMC Syst. Biol.* **3,** 1–21 (2009).

116. Saltelli, A., Chan, K. & Scott, E. M. (Eds. . *Sensitivity Analysis*. (John Wiley & Sons, 2000).

117. Homma, T. & Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* **52,** 1–17 (1996).

118. Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55,** 271–280 (2001).

119. Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. & Schaibly, J. H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *J. Chem. Phys.* **59,** 3873–3878 (1973).

120. Cukier, R. I., Levine, H. B. & Shuler, K. E. Review nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.* **26,** 1–42 (1978).

121. Saltelli, A. *et al. Global Sensitivity Analysis. The Primer*. (John Wiley & Sons, 2008).

122. Saltelli, A., Tarantola, S. & Chan, K. P. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **41,** 39–56 (1999).

123. Yuen, K.-V. & Kuok, S.-C. Efficient Bayesian sensor placement algorithm for structural identification: a general approach for multi-type sensory systems. *Earthq. Eng. Struct. Dyn.* **44,** 757–774 (2015).

124. Leyder, C., Dertimanis, V., Frangi, A., Chatzi, E. & Lombaert, G. Optimal sensor placement methods and metrics–comparison and implementation on a timber frame structure. *Struct. Infrastruct. Eng.* **14,** 997–1010 (2018).

125. Li, B., Li, D., Zhao, X. & Ou, J. Optimal sensor placement in health monitoring of suspension bridge. *Sci. China Technol. Sci.* **55,** 2039–2047 (2012).

126. Worden, K. & Burrows, A. P. Optimal sensor placement for fault detection. *Eng. Struct.* **23,** 885–901 (2001).

127. Guo, H. Y., Zhang, L., Zhang, L. L. & Zhou, J. X. Optimal placement of sensors for structural health monitoring using improved genetic algorithms. *Smart Mater. Struct.* **13,** 528–534 (2004).

128. Limongelli, M. P. Optimal location of sensors for reconstruction of seismic responses through spline function interpolation. *Earthq. Eng. Struct. Dyn.* **32,** 1055–1074 (2003).

129. Papadimitriou, C., Haralampidis, Y. & Sobczyk, K. Optimal experimental design in stochastic structural dynamics. *Probabilistic Eng. Mech.* **20,** 67–78 (2005).

130. Zhu, S., Zhang, X. H., Xu, Y. L. & Zhan, S. Multi-type sensor placement for multi-scale response reconstruction. *Adv. Struct. Eng.* **16,** 1779–1797 (2013).

131. Yi, T.-H., Li, H.-N. & Gu, M. Optimal sensor placement for structural health monitoring based on multiple optimization strategies. *Struct. Des. Tall Spec. Build.* **20,** 881–900 (2011).

132. Li, D. S., Li, H. N. & Fritzen, C. P. The connection between effective independence and modal kinetic energy methods for sensor placement. *J. Sound Vib.* **305,** 945–955 (2007).

133. Nasr, D. E. & Saad, G. A. Optimal sensor placement using a combined genetic algorithm-ensemble Kalman filter framework. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **3,** 1–14 (2017).

134. Capellari, G., Chatzi, E. & Mariani, S. Parameter identifiability through information theory. *Proc. 2nd Int. Conf. Uncertain. Quantif. Comput. Sci. Eng. (UNCECOMP 2017)* 372–380 (2017). doi:10.7712/120217.5376.17179

135. Giordano, E. *et al.* Damage detection and optimal sensor placement in health monitoring of " Collegiata di Santa Maria " in Visso (Central Italy). in *Convegno Anidis* 44–53 (2019).

136. Papadimitriou, C. Optimal sensor placement methodology for parametric identification of structural systems. *J. Sound Vib.* **278,** 923–947 (2004).

137. Papadimitriou, C. & Lombaert, G. The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mech. Syst. Signal Process.* **28,** 105–127 (2012).

138. Capellari, G., Chatzi, E. & Mariani, S. An optimal sensor placement method for SHM based on Bayesian experimental design and Polynomial Chaos Expansion. *ECCOMAS Congr. 2016 - Proc. 7th Eur. Congr. Comput. Methods Appl. Sci. Eng.* **3,** 6272–6282 (2016).

139. Capellari, G., Chatzi, E. & Mariani, S. Structural Health Monitoring Sensor Network Optimization through Bayesian Experimental Design. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **4,** 04018016 (2018).

140. MacKay, D. J. C. Bayesian Interpolation. *Neural Comput.* **4,** 415–447 (1992).

141. De, S., Johnson, E. A., Wojtkiewicz, S. F. & Brewick, P. T. Computationally efficient Bayesian model selection for locally nonlinear structural dynamic systems. *J. Eng. Mech.* **144,** (2018).

142. Beck, J. L. & Zuev, K. M. Asymptotically Independent Markov Sampling: A New MCMC Scheme for Bayesian Inference. *Vulnerability, Uncertainty, Risk Quantif. Mitigation, Manag. - Proc. 2nd Int. Conf. Vulnerability Risk Anal. Manag. ICVRAM 2014 6th Int. Symp. Uncertain. Model. a* 2022–2031 (2014). doi:10.1061/9780784413609.203

143. Huang, Y., Beck, J. L. & Li, H. Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment. *Comput. Methods Appl. Mech. Eng.* **318,** 382–411 (2017).

144. Catanach, T. A. & Beck, J. L. Bayesian System Identification using auxiliary stochastic dynamical systems. *Int. J. Non. Linear. Mech.* **94,** 72–83 (2017).

145. Yuen, K.-V. & Mu, H.-Q. Real-time system identification: an algorithm for simultaneous model class selection and parametric identification. *Comput. Civ. Infrastruct. Eng.* **30,** 785–801 (2015).

146. Kontoroupi, T. & Smyth, A. W. Online Bayesian model assessment using nonlinear filters. *Struct. Control Heal. Monit.* **24,** 1–15 (2017).

147. Brynjarsdóttir, J. & Ohagan, A. Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30,** (2014).

148. Rogers, T. J., Holmes, G. R., Cross, E. J. & Worden, K. On a grey box modelling framework for nonlinear system identification. in *Special Topics in Structural Dynamics* **6,** 167–178 (2017).

149. Gardner, P., Rogers, T. J., Lord, C. & Barthorpe, R. J. Learning of model discrepancy for structural dynamics applications using Bayesian history matching. *J. Phys. Conf. Ser.* **1264,** 012052 (2019).

150. Ceravolo, R., Faraci, A. & Miraglia, G. Bayesian calibration of hysteretic parameters with consideration of the model discrepancy for use in structural health monitoring. *Appl. Sci.* **10,** 5813 (2020).

151. Marelli, S. & Sudret, B. *UQLab: A framework for uncertainty quantication in Matlab.* (2014). doi:10.13140/RG.2.1.3778.7366

152. Maes, K. *et al.* Design of sensor networks for instantaneous inversion of modally reduced order models in structural dynamics. *Mech. Syst. Signal Process.* **52–53,** 628–644 (2015).

153. Nguyen, H. Van, Ranasinghe, D. C., Skvortsov, A. & Arulampalam, S. Computationally efficient methods for estimating unknown input forces on structural systems. in *IEEE 23rd International Conference on Information Fusion* (2020).

154. Song, X., Zhang, Y. & Liang, D. Input forces estimation for nonlinear systems by applying a square-root cubature Kalman filter. *Materials (Basel).* **10,** (2017).

155. Lourens, E., Reynders, E., De Roeck, G., Degrande, G. & Lombaert, G. An augmented Kalman filter for force identification in structural dynamics. *Mech. Syst. Signal Process.* **27,** 446–460 (2012).

156. Nayek, R., Chakraborty, S. & Narasimhan, S. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mech. Syst. Signal Process.* **128,** 497–530 (2019).

157. Gillijns, S. & De Moor, B. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica* **43,** 934–937 (2007).

158. Maes, K., Smyth, A. W., De Roeck, G. & Lombaert, G. Joint input-state estimation in structural dynamics. *Mech. Syst. Signal Process.* **70–71,** (2016).

159. Maes, K., Gillijns, S. & Lombaert, G. A smoothing algorithm for joint input-state estimation in structural dynamics. *Mech. Syst. Signal Process.* **98,** 292–309 (2018).

160. Ebrahimian, H., Astroza, R., Conte, J. P. & Papadimitriou, C. Bayesian optimal estimation for output-only nonlinear system and damage identification of civil structures. *Struct. Control Heal. Monit.* **25,** 1–32 (2018).

161. Lei, Y., Xia, D., Erazo, K. & Nagarajaiah, S. A novel unscented Kalman filter for recursive state-input-system identification of nonlinear systems. *Mech. Syst. Signal Process.* **127,** 120–135 (2019).

162. Al-Hussein, A. & Haldar, A. A new extension of unscented Kalman filter for structural health assessment with unknown input. *Sensors Smart Struct. Technol. Civil, Mech. Aerosp. Syst. 2014* **9061,** 90612Y (2014).

163. Al-hussein, A. & Haldar, A. Novel unscented Kalman filter for health assessment of structural systems with unknown input. *J. Eng. Mech.* **141,** 1–13 (2015).

164. Al-Hussein, A. & Haldar, A. Unscented Kalman Filter with unknown input and weighted global iteration for health assessment of large structural systems. *Struct. Control Heal. Monit.* **23,** 156–175 (2016).

165. Dertimanis, V. K., Chatzi, E. N., Eftekhar Azam, S. & Papadimitriou, C. Input-state-parameter estimation of structural systems from limited output information. *Mech. Syst. Signal Process.* **126,** 711–746 (2019).

166. Dertimanis, V., Chatzi, E., Azam, S. E. & Papadimitriou, C. Output-only fatigue prediction of uncertain steel structures. *8th Eur. Work. Struct. Heal. Monit. EWSHM 2016* **2,** 869–879 (2016).

167. Erazo, K. & Nagarajaiah, S. An offline approach for output-only Bayesian identification of stochastic nonlinear systems using unscented Kalman filtering. *J. Sound Vib.* **397,** 222–240 (2017).

168. Rogers, T. J., Worden, K., Manson, G., Tygesen, U. T. & Cross, E. J. A Bayesian filtering approach to operational modal analysis with recovery of forcing signals. *Proc. ISMA 2018 - Int. Conf. Noise Vib. Eng. USD 2018 - Int. Conf. Uncertain. Struct. Dyn.* 5181–5194 (2018).

3. VARIATIONAL INFERENCE FOR THE PREDICTIVE MODELING OF STRUCTURAL SYSTEMS

Current methods common to predictive modeling for civil infrastructure systems exploit analytical or sampling approximations to Bayesian inference, as discussed in Chpt. 3. Variational inference represents a third and relatively new class of inference approximations, which develop the posterior through optimization. This approach can be seen as striking a balance between the speed of analytical techniques and the flexibility of sampling approaches. Variational inference tends to be more computationally efficient than sampling approximations when faced with complex models and large data sets [1]. Though its speed cannot compare to the efficiency of analytical methods, its overall flexibility in comparison with these techniques make it a promising alternative for predictive structural modeling.

Variational inference can be described by the four key steps shown in Fig. 3.1. The algorithm begins with the definition of the stochastic model of the system $p(\mathbf{z}, \mathbf{y}|\mathbf{u})$, which asserts any prior beliefs about the latent variables, $\mathbf{z}$, of the system and the relationship of these variables with observed system behavior, $\mathbf{y}$, given the control input, $\mathbf{u}$. A distributional form, G, is then proposed which serves as an approximation of the posterior density, $p(\mathbf{z}|\mathbf{y}, \mathbf{u})$. This distributional form, called the *variational family* or *guide*, represents a family of distributions whose members, $g(\mathbf{z}; \boldsymbol{\phi}_i)$, can be specified by tuning the distributional parameters, $\boldsymbol{\phi}_i$, to different values. For succinctness, I will refer to members of the variational family as $g(\mathbf{z})$. Optimization is then performed by finding the member of the variational family which is most similar to the true posterior as defined by the Kullback-Leibler (KL) divergence

$$
\begin{aligned}
\hat{g} &= \underset{g \in G}{\arg\min} \, \text{KL}\big(g(\mathbf{z}) \parallel p(\mathbf{z}|\mathbf{y}, \mathbf{u})\big) \\
&= \underset{g \in G}{\arg\min} \, \big(\mathbb{E}_{g(\mathbf{z})}[\log q(\mathbf{z})] - \mathbb{E}_{g(\mathbf{z})}[\log p(\mathbf{z}, \mathbf{y}|\mathbf{u})] + \log p(\mathbf{y}, \mathbf{u})\big),
\end{aligned} \qquad (3.1)
$$

a measure of the information lost by approximating $p(\mathbf{z}|\mathbf{y}, \mathbf{u})$ with $g(\mathbf{z})$ [2]. Although it is possible to propose a guide for any distributional form, using a simple and flexible guide allows for efficient computation of a posterior approximation which is well representative of the true posterior.

Define Problem Domain	→	Propose Variational Family	→	Specify Loss Function	→	Perform Optimization
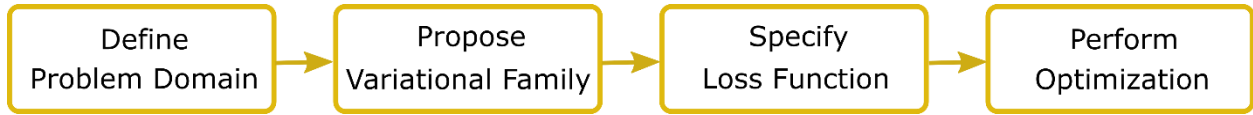
Fig. 3.1 – Key steps of the variational inference algorithm.

3.1 Literature Review

Variational inference emerged in the late 1980s as an efficient learning algorithm for neural networks, resulting from the adaptation of mean-field theory from statistical physics by Peterson and Anderson [3] and in parallel by Hinton and van Camp [4]. These initial innovations, coupled with important insights from Neal and Hinton [5] on their relation to the expectation-maximization algorithm, soon led to the adaptation of the method to a wide variety of statistical models [6]–[14], with the first comprehensive statement of the method and its general application introduced by Jordan et al. in 1999 [15]. Initially, the explicit dependence of the optimization process on the analytical form of the approximate posterior restricted the implementation of the variational inference approach to small-scale data sets described by a limited selection of approximate posterior densities [1]. Adaptations of the method were primarily developed in the form of optimization frameworks to work with different types of stochastic models [16]–[21] and implementations for specific applications, such as linguistics [22],[23], image processing [24],[25], and computational biology [26]–[28]. Significant attention was also given to adapting variational inference to online learning [28]–[31], particularly by merging it with other well-known filtering methods such as the Kalman filter [32]–[35]. The development of filtering approaches for state space models will be discussed in more detail in Chpt. 8.

The true power of variational inference was revealed with the development of two key algorithmic improvements in the early 2010s. The first, referred to as *stochastic variational inference,* resulted from a push toward online learning on massive data, which was at the time infeasible due to the lack of scalability of the coordinate ascent method [1] commonly used to optimize the KL divergence. Researchers at Princeton replaced this approach with the stochastic gradient descent method introduced by Robbins and Monro [36] and leveraged the natural gradients of the loss function [37] to iteratively optimize the global parameters on local subsets of the data. The resulting method, called *stochastic variational inference,* was successfully applied to massive data sets in probabilistic topic modeling [38],[39] before being fully refined and

presented by Hoffman et al. [40] in 2013. The second advancement, referred to as *black box variational inference*, arose in response to the need to optimize complex variational objectives whose analytical gradients were difficult to define. This approach instead allows the user to write the gradient of the loss function as an expectation, compute MC estimates of it, and then use those estimates to update the global parameters of the system. Early implementations of this idea were developed independently by several groups from 2010-2013 [41]–[44], with full development of the method for conjugate and non-conjugate inference appearing independently from several authors in 2014 [45]–[48]. These advances, coupled with the power of probabilistic programming languages like PyMC3 [49], STAN [50], Edward [51], and PyTorch [52] have allowed for the development of fast variational inference algorithms that are almost completely model independent. Automatic Differentiation Variational Inference (ADVI) [53],[54] represents the current state of the art of these methods and combines stochastic variational inference, black box variational inference, probabilistic programming tools, and distributional transformations to build a generalized variational inference interface that is accessible for wide variety of stochastic models. These advancements are particularly beneficial for predictive structural modeling as they make variational inference accessible for non-specialist users and give it the potential to expand to the massive data associated with structural health monitoring.

Current research on variational inference tends to focus on increasing the flexibility and accuracy of the variational family with respect to the true posterior. The most common way this is done is through modifications to the loss function. Early work on the loss function by Minka [55] explored the effectiveness of a variety of divergence metrics such as inclusive KL divergence, $\mathrm{KL}\big(p(\mathbf{z}|\mathbf{y},\mathbf{u}) \parallel g(\mathbf{z})\big)$, and the α-divergence, concluding that the different metrics are suitable for different inference goals. Recent works have revisited this topic due to a dissatisfaction with the propensity of the exclusive KL divergence, $\mathrm{KL}\big(g(\mathbf{z}) \parallel p(\mathbf{z}|\mathbf{y},\mathbf{u})\big)$, to underestimate the variance of the true posterior [53],[56]. Naesseth et al. [57] developed a method for unbiased optimization toward $\mathrm{KL}\big(p(\mathbf{z}|\mathbf{y},\mathbf{u}) \parallel g(\mathbf{z})\big)$ called Markovian score climbing, which results in more mass covering approximate posteriors. Daudel et al. [58] and Bamler et al. [59] pursued similar capabilities by replacing the exclusive KL divergence with the α-divergence and a divergence approach based in variational perturbation theory, respectively. A recent study by Geffner and Domke [60] has expressed doubt as to the feasibility of practical α-divergence implementations however, due to the gross computational cost of these methods in scaling to high-dimensional

problems. Burda et al. [61] and Mnih and Rezende [62] achieved tighter lower bounds and more mass-covering posterior approximations by generating loss functions which derive from likelihood estimates averaged over independent samples, or *Monte Carlo Objectives*. Alternatively, Ji et al. [41],[63] and Bamler et al. [64] have pursued strategies to impose bounds on both sides the optimization process and in so doing have improved convergence speed, variance matching to the true posterior, and confidence in the resulting model approximation. The use of a trust-region update to reduce sensitivity to hyperparameters [65],[66] and the replacement of the lower bound optimization on the loss function with a more flexible operator-based optimization [67] have also increased the precision and versatility of the method.

The other means by which increased flexibility and accuracy is accomplished is through direct modifications to the variational family. Several researchers have expanded on stochastic and black box techniques by exploring structured variational parameters which move beyond the mean-field (mutually independent) model. To broaden the impact of stochastic variational inference, which typically depends on the fully factorized variational distributions, Hoffman and Blei [68] proposed an update to the algorithm which explicitly allows for dependence between the local and global parameters. Ambrogioni et al. [69] made similar updates to the black box approach to variational inference by developing an automated method to construct the variational distribution by combining the inherent structure of the prior with evidence from the observed data. With regard to dynamical systems, Archer et al. [70] and Bamler and Mandt [71] have both proposed black box techniques which use a tri-diagonal dependence structure to indicate the Markovian relationship among the latent states. Liu et al. have taken an alternate approach to introducing structure into the variational family, through the development of Stein variational gradient descent (SVGD) [72]–[74]. The key idea of this approach is to separate variational inference from the specific distributional forms which govern its accuracy and computational cost. Instead, the posterior is approximated by a set of particles, which are driven toward the true posterior using a form of functional gradient descent to minimize the KL divergence. The method is easily able to approach multi-modal posterior distributions and works well with complex stochastic models. SVGD has been shown to perform at the level of accuracy of exact MC sampling and is extremely promising for the development of robust practical applications. This approach represents a huge step forward in variational inference as it removes from the user the

need to specify a variational family to fit an obscure posterior distribution. As such, this method becomes accessible and robust to a practical level of user knowledge.

Variational inference is a promising area of research for predictive structural modeling as it offers increased model flexibility over analytical methods at a reduced computational cost as compared with sampling methods. Beyond these immediate benefits, methods like ADVI and SVGD offer relatively few barriers to non-expert users and suggest theoretical robustness to practical use cases. However, these theoretical suggestions are far different than a practical guarantee. Research is still required to adapt these methods to manage the practical uncertainties evident in structural monitoring settings, such as low signal-to-noise ratios in the data, poor specification of the priors, scalability of the parameter space, real-time operation, etc. In the remainder of this chapter I will explore the derivation of the batch variational inference approach for structural systems with particular attention to these practical uncertainties. Chpt. 7 will then analyze the performance of this algorithms in simulation through a comparison with the unscented Kalman filter, which is the current standard for predictive structural modeling.

3.2 Batch Automatic Differentiation Variational Inference for Structural Systems

The derivation of ADVI provided herein for dynamical system identification is adapted from the work of Kucukelbir et al. [53]. The interpretation of variational inference for this class of problems enforces the conditional independence of the data and the Markovian nature of the stochastic dynamical system. The base algorithm remains otherwise unchanged.

The goal of ADVI is to transform the problem domain such that unconstrained optimization on the variational parameters can be performed in Euclidean space. As shown in Fig. 3.1, the algorithm begins with defining the problem domain. The model identification problem for a stochastic dynamical system can be expressed in terms of the prior distributions on the initial state $p(\mathbf{x}_0)$ and the physical parameters $p(\boldsymbol{\theta})$, as well as through the transmission probabilities $p(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\theta},\mathbf{u}_{k-1})$ and the emission probabilities $p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta},\mathbf{u}_k)$, where $\mathbf{u}_k$ is the known input excitation. The joint probability of the hidden states, initial conditions, parameters, and observations can then be expressed as

$$p(\mathbf{y}_{1:K},\mathbf{x}_{0:K},\boldsymbol{\theta}|\mathbf{u}_{0:k}) =$$
$$\prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta},\mathbf{u}_k) \prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\theta},\mathbf{u}_{k-1}) \prod_{m=1}^{M} p(\mathbf{x}_{0,m}|\mathbf{u}_0) \prod_{n=1}^{N} p(\theta_n), \quad (3.2)$$

where K is the number of observations, M is the dimension of the state space, and N is the dimension of the parameter space. The factorization of the prior distributions in Eq. (3.2) implies mutual independence of the parameters and the initial conditions on the states, which is commonly a valid assumption for the inference of dynamical systems. Note that the parameters are assumed constant in this derivation.

To automate the solution approach, these distributions are transformed to have support on the Euclidean space $\mathbb{R}^{K+M+N}$. The joint density is therefore expressed in terms of transformed states $\hat{\mathbf{x}}_{0:K}$ and parameters $\hat{\boldsymbol{\theta}}$ given by $[\hat{\mathbf{x}}_{0:K}, \hat{\boldsymbol{\theta}}] = \Upsilon([\mathbf{x}_{0:K}, \boldsymbol{\theta}])$, such that

$$p(\mathbf{y}_{1:K}, \hat{\mathbf{x}}_{0:K}, \hat{\boldsymbol{\theta}} | \mathbf{u}_{0:k}) = p(\mathbf{y}_{1:K}, [\mathbf{x}_{0:K}, \boldsymbol{\theta}] = \Upsilon^{-1}([\hat{\mathbf{x}}_{0:K}, \hat{\boldsymbol{\theta}}]) | \mathbf{u}_{0:k}) |\det J_{\Upsilon^{-1}}([\hat{\mathbf{x}}_{0:K}, \hat{\boldsymbol{\theta}}])|, \quad (3.3)$$

where $\Upsilon(\cdot)$ is a one-to-one differentiable function which transforms $[\mathbf{z}_{0:K}, \boldsymbol{\theta}]$ to have full support in $\mathbb{R}^{K+M+N}$ and $J_{\Upsilon^{-1}}(\cdot)$ is the Jacobian of the inverse of Υ [53].

In the second step, proposing a variational family, the form of the approximate posterior model is specified such that it generates an adequate representation of the system. As the hidden states and parameters have been transformed to have support in $\mathbb{R}^{K+M+N}$, the Gaussian distribution is a simple and effective choice for the variational family. I represent the approximate posterior of the system state vectors (e.g. displacement and velocity) to be mutually independent and capture the Markovian transitions within the time history of each state through a tri-diagonal covariance matrix [70],[71]. Though some correlations exist between non-adjacent states, the tri-diagonal covariance matrix has been found to be sufficient in capturing the major correlations in the progression of the state histories such that an accurate posterior approximation can be formulated. The assumption of mutual independence between the state vectors is selected to reduce the dimensionality of the inference problem (i.e. reduce the number of covariance terms in the optimization) while asserting minimal influence on the accuracy of the result, and has been found to be effective for single degree of freedom structural systems. Such an assumption has the effect of generating some overconfidence in the posterior, while still retaining an accurate approximation of the mean given the data observed [53]. Relaxing this assumption has the potential to increase the effectiveness of the method for multi-degree of freedom problems and for the development of recursive identification approaches, and merits future study. However, these assumptions have been found to be sufficiently accurate for the cases described within this work, and yield the variational families

$$g(\hat{\mathbf{x}}; \boldsymbol{\phi}_{\hat{\mathbf{x}}}) = N(\hat{\mathbf{x}} | \boldsymbol{\mu}_{\hat{\mathbf{x}}}, \mathbf{L}_{\hat{\mathbf{x}}} \mathbf{L}_{\hat{\mathbf{x}}}^T) \quad (3.4)$$

which are parameterized by the variational parameters $\boldsymbol{\phi}_{\hat{\mathbf{x}}} = (\boldsymbol{\mu}_{\hat{\mathbf{x}}}, \mathbf{L}_{\hat{\mathbf{x}}})$, where $\boldsymbol{\mu}_{\hat{\mathbf{x}}}$ and $\mathbf{L}_{\hat{\mathbf{x}}}$ are respectively the mean and Cholesky factor of the covariance matrix, $\boldsymbol{\Sigma} = \mathbf{L}_{\hat{\mathbf{x}}} \mathbf{L}_{\hat{\mathbf{x}}}^{T}$, of the variational parameters related to $\hat{\boldsymbol{x}}$. These variational parameters are unconstrained in $\mathbb{R}^{M(3K-1)}$.

Likewise, we propose a Gaussian variational family on the hidden parameters which enforces their assumed mutual independence through a diagonal covariance matrix, yielding

$$g(\widehat{\boldsymbol{\theta}}; \boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}}) = N\left(\widehat{\boldsymbol{\theta}} | \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}, \operatorname{diag}(\exp(\boldsymbol{\rho}_{\widehat{\boldsymbol{\theta}}})^2)\right) \tag{3.5}$$

which is parameterized by $\boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}} = (\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}, \boldsymbol{\rho}_{\widehat{\boldsymbol{\theta}}})$, where $\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\rho}_{\widehat{\boldsymbol{\theta}}}$ are respectively the mean and log standard deviation of the variational parameters related to $\widehat{\boldsymbol{\theta}}$ and are unconstrained in $\mathbb{R}^{2N}$. The full model then requires optimization on $\mathbb{R}^{M(3K-1)+2N}$.

In the third step, specifying the loss function, we set up our optimization problem. To do this, we first define the evidence lower bound (ELBO) as

$$\begin{aligned} \operatorname{ELBO}(g) &= -\operatorname{KL}\big(g(\mathbf{z}) \,||\, p(\mathbf{z}|\mathbf{y}, \mathbf{u})\big) + \log p(\mathbf{y}, \mathbf{u}) \\ &= \mathbb{E}_{g(\mathbf{x}, \boldsymbol{\theta})}[\log p(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}, \boldsymbol{\theta}, \mathbf{u}_{0:k})] + \mathbb{E}_{g(\mathbf{x}, \boldsymbol{\theta})}[\log p(\mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{u}_{0:k})] \\ &\qquad\qquad - \mathbb{E}_{g(\mathbf{x}, \boldsymbol{\theta})}[\log g(\mathbf{x}_{0:K}, \boldsymbol{\theta})] \\ &= \mathbb{E}_{g(\mathbf{x}, \boldsymbol{\theta})}[\log p(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}, \boldsymbol{\theta}, \mathbf{u}_{0:k})] - \operatorname{KL}(g(\mathbf{x}_{0:K}, \boldsymbol{\theta}) \,||\, p(\mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{u}_{0:k})), \end{aligned} \tag{3.6}$$

and note that the ELBO is equivalent to the negative KL divergence save for the addition of the log of the model evidence, which is constant for this class of problems. Maximization of the ELBO is therefore equivalent to minimization of the KL divergence and eliminates the need to compute the model evidence, $p(\mathbf{y}, \mathbf{u})$, which can be difficult to determine. We therefore select the optimal member of the variational family $\boldsymbol{G}$ by maximizing the ELBO. The ELBO can be intuitively understood from its two components to strike a balance between encouraging densities which fit the observed data, $\mathbb{E}_{g(\mathbf{z}, \boldsymbol{\theta})}[\log p(\mathbf{y}_{1:K}|\mathbf{x}_{0:K}, \boldsymbol{\theta}, \mathbf{u}_{0:k})]$, and encouraging densities which stay close to the prior, $\operatorname{KL}(g(\mathbf{x}_{0:K}, \boldsymbol{\theta}) \,||\, p(\mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{u}_{0:k}))$ [1]. When this balance is achieved, the optimal density will express the behavior of the true system without overfitting to the limited data set used for optimization. The ELBO can be expressed in terms of the transformed states and parameters in Eq. (3.3) as

$$\begin{aligned} \operatorname{ELBO}(g) &= \mathbb{E}_{g(\hat{\mathbf{x}}, \widehat{\boldsymbol{\theta}}; \boldsymbol{\phi})}\big[\log p\big(\mathbf{y}_{1:K}, \Upsilon^{-1}([\hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}])\big)|\mathbf{u}_{0:k}\big) + \log\big|\det J_{\Upsilon^{-1}}([\hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}])\big|\big] \\ &\qquad\qquad - \mathbb{E}_{g(\hat{\mathbf{x}}, \widehat{\boldsymbol{\theta}}; \boldsymbol{\phi})}\big[\log g(\hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}; \boldsymbol{\phi})\big]. \end{aligned} \tag{3.7}$$

In the final step, performing stochastic optimization, our goal is to use a noisy estimate of the gradient of the ELBO to walk toward locally optimal values of the hidden states. It is difficult to take the gradient of the ELBO directly, however, as the ELBO involves the computation of an intractable expectation. To resolve this issue and allow for the use of automatic differentiation to evaluate the gradient, an additional transformation, referred to as *elliptical standardization*, is used. This transformation can be expressed as $\eta_{\hat{\mathbf{x}}} = S_{\phi_{\hat{\mathbf{x}}}}(\hat{\mathbf{x}}) = \mathbf{L}_{\hat{\mathbf{x}}}^{-1}(\hat{\mathbf{x}} - \boldsymbol{\mu}_{\hat{\mathbf{x}}})$ for the states and $\eta_{\hat{\boldsymbol{\theta}}} = S_{\phi_{\hat{\boldsymbol{\theta}}}}(\hat{\boldsymbol{\theta}}) = \mathrm{diag}(\exp(\boldsymbol{\rho}_{\hat{\boldsymbol{\theta}}}))^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}})$ for the parameters, yielding the modified ELBO

$$\mathrm{ELBO}(g) = \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)}\left[\log p\left(\mathbf{y}, \Upsilon^{-1}\left(\left[S_{\phi_{\hat{\mathbf{x}}}}^{-1}(\eta_{\hat{\mathbf{x}}}), S_{\phi_{\hat{\boldsymbol{\theta}}}}^{-1}(\eta_{\hat{\boldsymbol{\theta}}})\right]\right) | \mathbf{u}\right)\right]$$
$$+ \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)}\left[\log\left|\det J_{\Upsilon^{-1}}\left(\left[S_{\phi_{\hat{\mathbf{x}}}}^{-1}(\eta_{\hat{\mathbf{x}}}), S_{\phi_{\hat{\boldsymbol{\theta}}}}^{-1}(\eta_{\hat{\boldsymbol{\theta}}})\right]\right)\right|\right] - \mathbb{E}_{g(\hat{\mathbf{x}},\hat{\boldsymbol{\theta}};\boldsymbol{\phi})}[\log g(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})]. \quad (3.8)$$

Note that the expectation containing the joint probability on the observations, states, and parameters is now given in terms of a standard normal distribution, without any explicit dependence on the variables to be optimized. This modification allows for the use of Monte Carlo methods, typically with only 1 sample, to obtain a noisy approximation of the ELBO for the automatic evaluation of the gradient. Stochastic optimization can then be performed using a number of algorithms available with probabilistic programming libraries. In this study, I implement ADVI using the ADAM stochastic optimization algorithm [75] as part of the python library PyTorch [52].

The formulation given by Eqs. (3.2-3.8) expresses a simple black box implementation for using variational inference to identify the time-varying states and constant parameters of a stochastic dynamical systems, as structured by the ADVI methodology. A more succinct version of the complete framework is given in Algorithm 3.1 to aid the reader in visualizing the computational process.

A key feature of this process is the evaluation of the ELBO over the full data set during each optimization iteration. Due to the highly correlated nature of the time series data, the problem cannot readily be sub-sampled according to the standard stochastic variational inference paradigm. Adaptations to the algorithm which address this issue were recently proposed Johnson et al. [76] and Foti et al. [77]. Johnson et al. addressed the solution in a setting where each mini-batch could be considered an independent time series, and simply conducted the standard SVI approach using these independent batches of time series data [76]. This approach is obviously not applicable to the continuous vibrations observed in structural monitoring situations. In contrast, Foti et al.

present a method for subsampling the correlated time series in which cohesive sequences of data are selected from the full chain. The optimization is then broken down into an update on the local parameters, in which edge effects in the data subset are decreased by adding additional data points at the beginning and end of the sequence as a buffer, and the global update, in which the buffering points are eliminated and the optimized local parameters are used to take a gradient step on the global parameters [77]. This approach could be extremely helpful for structural identification in limited data scenarios, as it would allow for the use of data at a much faster rate. However, similar speed, as well as adaptivity to continuous monitoring, have the potential to be achieved with a variational filtering approach, which will be developed and examined in Chpt. 8.

Initialize with:
$$p(\mathbf{y}_{1:K}, \mathbf{x}_{0:K}, \boldsymbol{\theta}|\mathbf{u}_k)$$
$$= \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}) \prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\theta}) \prod_{m=1}^{M} p(\mathbf{x}_{0,m}) \prod_{n=1}^{N} p(\theta_n)$$
$$p(\mathbf{y}_{1:K}, \hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}|\mathbf{u}_k) = p(\mathbf{y}_{1:K}, [\mathbf{x}_{0:K}, \boldsymbol{\theta}] = \Upsilon^{-1}([\hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}])|\mathbf{u}_k)|\det J_{\Upsilon^{-1}}([\hat{\mathbf{x}}_{0:K}, \widehat{\boldsymbol{\theta}}])|$$
$$g(\hat{\mathbf{x}}; \boldsymbol{\phi}_{\hat{\mathbf{x}}}) = N(\hat{\mathbf{x}}|\boldsymbol{\mu}_{\hat{\mathbf{x}}}, \mathbf{L}_{\hat{\mathbf{x}}}\mathbf{L}_{\hat{\mathbf{x}}}^T)$$
$$g(\widehat{\boldsymbol{\theta}}; \boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}}) = N\left(\widehat{\boldsymbol{\theta}}|\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}, \mathrm{diag}(\exp(\boldsymbol{\rho}_{\widehat{\boldsymbol{\theta}}})^2)\right)$$

Optimize variational family G:
    While ELBO has not converged
        Sample from $g(\hat{\mathbf{x}}; \boldsymbol{\phi}_{\hat{\mathbf{x}}})$ and $g(\widehat{\boldsymbol{\theta}}; \boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}})$
        Compute $\mathrm{ELBO}(g)$ from samples
        Evaluate the gradient of $\mathrm{ELBO}(g)$ over the variational parameters $\boldsymbol{\phi}_{\hat{\mathbf{x}}}$ and $\boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}}$
        Update variational parameters $\boldsymbol{\phi}_{\hat{\mathbf{x}}}$ and $\boldsymbol{\phi}_{\widehat{\boldsymbol{\theta}}}$ with stochastic optimization

Algorithm 3.1 Batch Variational Inference with ADVI

Other innovations to this basic approach have been suggested to improve the scalability of these algorithms with the dimension of the problem space, i.e. the number of states and parameters considered. Gorbach et al. [78] address this problem by tackling high computational costs in numerical integration used for inference of dynamical systems. They defined variational inference over a Gaussian process gradient matching which takes advantage of local linearities in the nonlinear system dynamics. Loaiza-Maya et al. [79] suggest a different approach in which the variational family is adapted for sparsity and accuracy among the parameters and states of the

model. Specifically, they use a copula model adapted from Smith et al. [80] to generate a sparse approximation of the global parameters and define the exact conditional posterior model over the latent states. This variational family, coupled with a specialized approach for estimating the gradient of the ELBO, allows the authors to perform efficient batch variational inference on extremely high-dimensional systems.

Batch variational inference on structural systems, as described herein, is shown in Chpt. 7 to be effective for nonlinear structural identification in small-scale, small-data systems. For structures with higher dimensional problem spaces or continuous monitoring data, a filtering approach to inference may be preferable, as developed and explored in Chpt. 8.

3.3   Conclusions

In this chapter, I have explored the history of variational inference and its particular adaptation toward predictive modeling for structural systems, including a novel variational filtering technique. Variational inference stands distinct from other Bayesian inference methods due to its focus on the optimization of an approximate posterior. This approach allows for increased flexibility in the selection of the form of the approximate posterior, which can be both a blessing and a curse for the user. Sufficiently malleable posterior approximations can allow the user to reach a suitable understanding of the posterior at significantly less computational cost than sampling methods, and yet with more accuracy than analytical methods. However, incorrect assertions on the approximate posterior can easily lead to overconfident posteriors. Variational families that have been tailored to for robust performance in specific applications are key to delivering the method to general users.

In the case of predictive modeling, it is also vital that the method be compatible with near real-time implementation to manage the continuous data streams expected from structural health monitoring scenarios. Currently, this goal cannot be readily achieved through typical adaptations on batch inference methods. The exploration of filtering adaptations for this method provides an intuitive and practical step toward needs of real-time predictive modeling for structural systems, and will be explored in detail in Chpt. 8.

3.4 References

1. Blei, D. M., Kucukelbir, A. & Mcauliffe, J. D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **112,** 859–877 (2017).

2. Bishop, C. M. *Pattern recognition and machine learning*. *Pattern Recognition* (Springer, 2006). doi:10.1117/1.2819119

3. Peterson, C. & Anderson, J. R. A Mean Field Theory Learning Algorithm for Neural Networks. *Complex Syst.* **1,** 995–1019 (1987).

4. Hinton, G. E. & Van Camp, D. Keeping neural networks simple by minimizing the description length of the weights. in *6th Annual Conference on Computational Learning Theory* (1993).

5. Neal, R. M. & Hinton, G. E. in *Learning in Graphical Models* 355–368 (Springer, 1993).

6. Saul, L. K., Jaakkola, T. & Jordan, M. I. Mean field theory for sigmoid belief networks. *J. Artif. Intell. Res.* **4,** 61–76 (1996).

7. Jaakkola, T. S. & Jordan, M. I. Computing upper and lower bounds on likelihoods in intractable networks. *Uncertain. Artif. Intell.* 340–348 (1996).

8. Ghahramani, Z. & Jordan, M. I. Factorial hidden Markov models. in *Advances in Neural Information Processing Systems* 472–478 (1996).

9. Waterhouse, S., Mackay, D. & Robinson, T. Bayesian methods for mixtures of experts. in *Advances in Neural Information Processing Systems* 351–357 (1996).

10. MacKay, D. J. C. *Ensemble learning for hidden Markov models*. (1997).

11. Barber, D. & Bishop, C. M. Ensemble learning in Bayesian neural networks. *NATO ASI Ser. F Comput. Syst. Sci.* **168,** 215–238 (1998).

12. Saul, L. K. & Jordan, M. I. Exploiting tractable substructures in intractable networks. in *Advances in Neural Information Processing Systems* 486–492 (1996).

13. Jaakkola, T. S. & Jordan, M. I. A variational approach to Bayesian logistic regression models and their extensions. in *6th International Workshop on Artificial Intelligence and Statistics* 4–15 (1996).

14. Barber, D. & Wiegernick, W. Tractable variational structures for approximating graphical models. in *Advances in Neural Information Processing Systems* 183–199 (1999).

15. Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. An introduction to variational methods for graphical models. *Mach. Learn.* **37,** 183–233 (1999).

16. Wainwright, M. J. & Jordan, M. I. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1,** 1–305 (2008).

17. Honkela, A., Tornio, M., Raiko, T. & Karhunen, J. Natural conjugate gradient in variational inference. in *International Conference on Neural Information Processing* 305–314 (Springer, 2007).

18. Braun, M. & Mcauliffe, J. Variational inference for large-scale models of discrete choice. *J. Am. Stat. Assoc.* **105,** 324–335 (2010).

19. Wand, M. P., Ormerod, J. T., Padoan, S. A. & Frühwirth, R. Mean mield variational Bayes for elaborate distributions. *Bayesian Anal.* **6,** 847–900 (2011).

20. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3,** 993–1022 (2003).

21. Beal, M. J. Variational algorithms for approximate Bayesian inference. (Ph. D. Thesis, University College, London, 2003).

22. Blei, D. M. & Lafferty, J. D. A correlated topic model of science. *Ann. Appl. Stat.* **1,** 17–35 (2007).

23. Cohen, S. B. & Smith, N. A. Covariance in unsupervised learning of probabilistic grammars. *J. Mach. Learn. Res.* **11,** 3017–3051 (2010).

24. Likas, A. & Galatsanos, N. P. A variational method for Bayesian blind image deconvolution. in *International Conference on Image Processing* **52,** 2222–2233 (IEEE, 2004).

25. Jojic, N. & Frey, B. J. Learning flexible sprites in video layers. in *Computer Society Conference on Computer Vision and Pattern Recognition* **1,** I199–I206 (IEEE, 2001).

26. Logsdon, B. A., Hoffman, G. E. & Mezey, J. G. A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC Bioinformatics* **11,** 58–70 (2010).

27. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197,** 573–589 (2014).

28. Friston, K. J., Trujillo-Barreto, N. & Daunizeau, J. DEM: A variational treatment of dynamic systems. *Neuroimage* (2008). doi:10.1016/j.neuroimage.2008.02.054

29. Ghahramani, Z. & Hinton, G. E. Variational learning for switching state space models. *Neural Comput.* **12,** 831–864 (2000).

30. Sato, M.-A. Online model selection based on the variational Bayes. *Neural Comput.* **13,** 1649–1681 (2001).

31. Honkela, A. & Valpola, H. Online variational Bayesian learning. in *4th International Symposium on Independent Component Analysis and Blind Signal Separation* 803–808 (2003).

32. Ghahramani, Z. & Beal, M. J. Propagation algorithms for variational Bayesian learning. *Adv. Neural Inf. Process. Syst.* (2001).

33. Beal, M. J. & Ghahramani, Z. *The Variational Kalman Smoother*. *Gatsby University Technical Reports* (2001).

34. Sarkka, S. & Nummenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Trans. Automat. Contr.* **54,** 596–600 (2009).

35. Auvinen, H., Bardsley, J. M., Haario, H. & Kauranne, T. The variational Kalman filter and an efficient implementation using limited memory BFGS. *Int. J. Numer. Methods Fluids* **64,** 314–335 (2010).

36. Robbins, H. & Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **22,** 400–407 (1951).

37. Amari, S.-I. Natural gradient works efficiently in learning. *Neural Comput.* **10,** 251–276 (1998).

38. Hoffman, M. D., Blei, D. M. & Bach, F. Online learning for latent Dirichlet allocation. in *Advances in Neural Information Processing Systems* 856–864 (2010).

39. Wang, C., Paisely, J. & Blei, D. M. Online variational inference for the hierarchical Dirichlet process. in *14th International Conference on Artifical Intelligence and Statistics* **15,** 752–760 (Journal of Machine Learning Research, 2011).

40. Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. Stochastic variational inference. *J. Mach. Learn. Res.* **14,** 1303–1347 (2013).

41. Ji, C., Shen, H. & West, M. *Bounded approximations for marginal likelihoods*. (2010).

42. Nott, D. J., Li Tan, S., Villani, M. & Kohn, R. Regression density estimation with variational methods and stochastic approximation. *J. Comput. Graph. Stat.* **21,** 797–820 (2012).

43. Paisley, J., Blei, D. M. & Jordan, M. I. Variational Bayesian inference with stochastic search. *arXiv Prepr.* (2012). doi:arXiv1206.6430

44. Wingate, D. & Weber, T. Automated variational inference in probabilistic programming. *arXiv Prepr.* (2013).

45. Ranganath, R., Gerrish, S. & Blei, D. M. Black box variational inference. *Artif. Intell. Stat.* 814–822 (2014).

46. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *arXiv Prepr.* (2014).

47. Titsias, M. K. & Lázaro-Gredilla, M. Doubly stochastic variational Bayes for non-conjugate inference. in *International Conference on Machine Learning* 1971–1979 (2014).

48. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *31st Int. Conf. Mach. Learn. ICML 2014* **4,** 3057–3070 (2014).

49. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016,** 1–24 (2016).

50. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76,** (2017).

51. Tran, D. *et al.* Edward: A library for probabilistic modeling, inference, and criticism. (2016).

52. Paszke, A. *et al.* PyTorch: an imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems* 8024–8035 (2019).

53. Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. & Blei, D. M. Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18,** 430–474 (2017).

54. Kucukelbir, A., Ranganath, R., Gelman, A. & Blei, D. M. Automatic variational inference in Stan. in *Advances in Neural Information Processing Systems* (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 568–576 (2015).

55. Minka, T. P. Divergence measures and message passing. *Microsoft Res. Tech. Rep.* 17 (2005).

56. Jihan, N., Jayasinghe, M. & Perera, S. Streaming stochastic variational Bayes; An improved approach for Bayesian inference with data streams. *PeerJ Prepr.* (2019). doi:10.7287/peerj.preprints.27790v2

57. Naesseth, C. A., Lindsten, F. & Blei, D. Markovian score climbing: variational inference with KL(p‖q). *arXiv Prepr.* (2020).

58. Daudel, K., Douc, R. & Portier, F. Infinite-dimensional gradient-based descent for alpha-divergence minimisation. *arXiv Prepr.* (2020). doi:10.1214/15-AOS1380

59.   Bamler, R., Zhang, C., Opper, M. & Mandt, S. Perturbative black box variational inference. in *Advances in Neural Information Processing Systems* 5079–5088 (2017).

60.   Geffner, T. & Domke, J. On the difficulty of unbiased alpha divergence minimization. *arXiv Prepr.* (2020).

61.   Burda, Y., Grosse, R. & Salakhutdinov, R. Importance weighted autoencoders. *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.* 1–14 (2016).

62.   Mnih, A. & Rezende, D. J. Variational inference for Monte Carlo objectives. *arXiv Prepr.* (2016).

63.   Ji, C. & Shen, H. Stochastic variational inference via upper bound. in *4th Workshop on Bayesian Deep Learning* 1–8 (2019).

64.   Bamler, R., Zhang, C., Opper, M. & Mandt, S. Tightening bounds for variational inference by revisiting perturbation theory. *J. Stat. Mech. Theory Exp.* (2019). doi:https://doi.org/10.1088/1742-5468/ab43d3

65.   Theis, L., May, M. L., Hoffman, M. D. & Com, M. A. A trust-region method for stochastic variational inference with applications to streaming data. in *32nd International Conference on Machine Learning* (Journal of Machine Learning Research:W&CP Volume 37, 2015).

66.   Kim, G. & Yang, H. Trust region sequential variational inference. in *Asian Conference on Machine Learning* 1033–1048 (2019).

67.   Ranganath, R., Altosaar, J., Tran, D. & Blei, D. M. Operator variational inference. *Adv. Neural Inf. Process. Syst.* 496–504 (2016).

68.   Hoffman, M. D. & Blei, D. M. Structured stochastic variational inference. in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics* (eds. Lebanon, G. & Vishwanathan, S. V. N.) 361–369 (2015).

69.   Ambrogioni, L., Hinne, M. & van Gerven, M. Automatic structured variational inference. (2020).

70.   Archer, E., Park, I. M., Buesing, L., Cunningham, J. & Paninski, L. Black Box Variational Inference for State Space Models. in *6th International Conference on Learning Representations* 1–11 (2015).

71.   Bamler, R. & Mandt, S. Structured black box variational inference for latent time series models. in *International Conference on Machine Learning* 1–5 (2017).

72. Wang, D. & Liu, Q. Learning to draw samples: with application to amortized MLE for generative adversarial learning. in *International Conference on Learning Representations* 1–13 (2017).

73. Liu, Q. Stein variational gradient descent as gradient flow. in *Advances in Neural Information Processing Systems* 3116–3124 (2017).

74. Liu, Q. & Wang, D. Stein variational gradient descent: A general purpose Bayesian inference algorithm. in *Advances in Neural Information Processing Systems* 2378–2386 (2016).

75. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. in *International Conference on Learning Representations* 1–15 (2014).

76. Johnson, M. J. & Willsky, A. S. Stochastic variational inference for Bayesian time series models. in *Proceedings of the 31st International Conference on Machine Learning* (Journal of Machine Learning:W&CP Volume 32, 2014).

77. Foti, N. J., Xu, J., Laird, D. & Fox, E. B. Stochastic variational inference for hidden Markov models. in *Advances in Neural Information Processing Systems* 3599–3607 (2014).

78. Gorbach, N. S., Bauer, S. & Buhmann, J. M. Scalable variational inference for dynamical systems. in *Advances in Neural Information Processing Systems* 4807–4816 (2017).

79. Loaiza-Maya, R., Smith, M. S., Nott, D. J. & Danaher, P. J. Fast and accurate variational inference for models with many latent variables. *arXiv Prepr.* 1–57 (2020).

80. Smith, M. S., Loaiza-Maya, R. & Nott, D. J. High-dimensional copula variational approximation through transformation. *J. Comput. Graph. Stat.* **0,** 1–35 (2020).

4. EXPERIMENTAL CASE STUDY: NONLINEAR ENERGY SINK DEVICE

A small-scale structural system expressing extreme nonlinear behaviors is used to evaluate the experimental performance of the various Bayesian filtering algorithms used in this study. Details regarding this experimental system, the nonlinear energy sink (NES), is given in the chapter. The content of this chapter has been adapted from the published work of the author [1],[2].

4.1 Theoretical Development

Nonlinear energy sink (NES) devices have been of interest for applications in vibration reduction since they were first studied by Gendelmen et al. [3] and Vakakis et al. [4] in 2001. The ability of these devices to generate a one-way transfer of energy from the primary structure over a wide range of frequencies is particularly desirable for civil engineering applications [5],[6], where long life spans, exposure to hazards, and rehabilitation processes can change the dynamic characteristics of a structure. This effect, which is termed *targeted energy transfer* or *nonlinear energy pumping* [7], is realized through the essential stiffness nonlinearity of the non-conservative structural attachment [8]. Due to this nonlinearity, these devices have no preferential natural frequency and are therefore theoretically capable of extracting energy from any mode of the primary linear system [9],[10]. The attachment therefore extracts and dissipates the energy induced by transient vibrations experienced by the primary structure which occur above a certain input energy threshold [3]. In addition to their utility for damping vibrations in structural systems, these devices form an interesting and complex case study for predictive modeling due to the clear bifurcation in their behavior about the input energy threshold.

The NES device explored in this study belongs to the class of Duffing-type devices experimentally investigated by McFarland et al. [11] and is similar in construction to the small-scale Type 1 NES device developed by Wierschem [10] and later studied by Silva et al. [12] with respect to its ability to transfer and dissipate energy from the structural systems. The device consists of a mass that is constrained to unidirectional motion and a wire that is threaded through the mass perpendicular to its motion, as shown in Fig. 4.1.
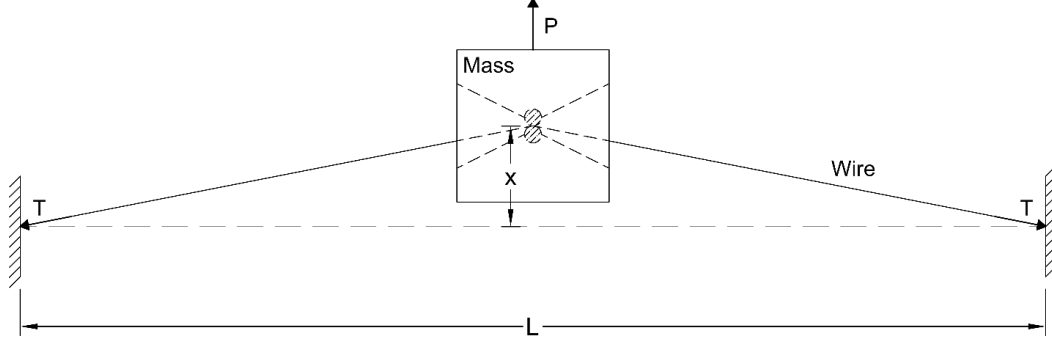
Fig. 4.1 – Schematic diagram of the NES Device

The essential nonlinearity in the system is generated by the geometric configuration of the fixed-end wire with respect to the moving mass. For any displacement of the mass eliciting a linear elastic response from the wire, the resulting force, f_s, can be expressed as

$$f_s = 2k_w x \left[1 - (L/2)((L/2)^2 + x^2)^{-1/2}\right] + Tx[(L/2)^2 + x^2]^{-1/2}, \qquad (4.1)$$

where k_w is the elastic stiffness of the wire, T is the initial tension in the wire, and x is the displacement of the mass relative to the fixed ends of the wire. This model is typically simplified by assuming small displacements and replacing $[(L/2)^2 + x^2]^{-1/2}$ with its Taylor expansion about $x = 0$. The force can then be expressed as

$$f_s = \frac{4T}{L} x + \frac{4(Lk_w - 2T)}{L^3} x^3 + O(x^5), \qquad (4.2)$$

where the higher order terms have been observed to not contribute significantly to the dynamics of the NES under normal operating conditions.

The Duffing-type NES device is typically assumed to dissipate energy through viscous damping, a modeling choice which, coupled with the stiffness forces given in Eq. (4.2), has been shown to effectively represent the behavior of this class of device [11],[13]–[15]. Experimental observations of the NES used in this study, shown in Fig. 4.2, also suggest the action of a Coulomb damping force. Fig. 4.2(a) shows the free response behavior of the NES after being excited by a step signal (signal 6, Table 4.1), in which the linear decay in response amplitude evident through the majority of the response history clearly indicates the presence of a non-negligible Coulomb friction interaction in the dynamics of the system. Friction induced behavior can also be seen in the low amplitude restoring force versus the displacement response, given in Fig. 4.2(b), where

73

the restoring force is computed using the experimental response of the device to a low amplitude band-limited white noise signal (signal 15, Table 4.1) as

$$R(\ddot{x}, \ddot{x}_g) = -m(\ddot{x} + \ddot{x}_g).$$

(4.3)

Because the contributions of higher order terms on the stiffness force, f_s, are negligible at small amplitudes of vibration, the damping forces can be more easily observed. The measured response data does not exhibit the elliptical behavior expected from a purely viscous damped system, but rather exhibits the nearly vertical unloading branches typical of Coulomb damping, plus a small linear stiffness. I therefore chose to incorporate Coulomb damping in the dynamic model of the NES device.
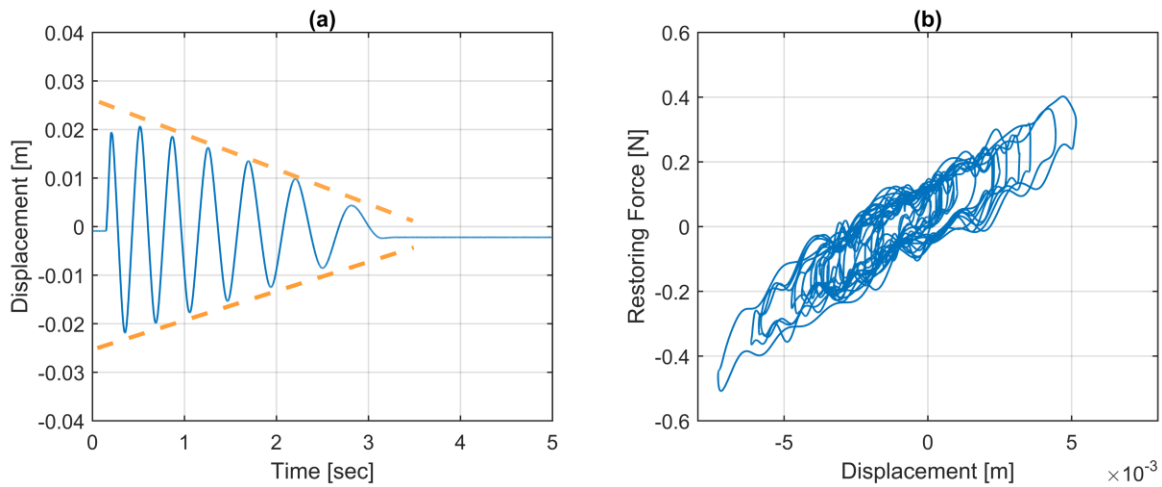
Fig. 4.2 - (a) Response history of signal 6 (Table 4.1) with imposed lines showing linear decay and (b) Restoring force versus displacement response to signal 15 (Table 4.1), where acceleration has been low-pass filtered to 50 Hz.

Using this interpretation of the device behavior, the equation of motion for this system becomes

$$m\ddot{x} + c_v \dot{x} + c_f \text{sign}(\dot{x}) + kx + zx^3 = P,$$

(4.4)

where m is the mass, c is the viscous damping coefficient, $k = 4T/L$ from Eq. (4.2), $z = 4(Lk_w - 2T)/L^3$ from Eq. (4.2), and P is a time-dependent external excitation force. To achieve the irreversible energy transfer noted in [3] and [4], the linear stiffness term (k) must be set to 0, which is achieved by eliminating any wire pretension (T). For more detailed information on the

vibration characteristics and energy dissipation properties of this class of devices, the reader is referred to [3],[4],[10],[11],[16],[17].

In preliminary unscented Kalman filter (UKF) identification trials using Eq. (4.4) it was determined that the replacement of the discontinuous $\text{sign}(\dot{x})$ term with the continuous approximation $\tanh(\upsilon\dot{x})$ results in a more accurate representation of the device behavior, where υ is a pre-determined parameter which governs the accuracy with which $\text{sign}(\dot{x})$ is approximated. This enhanced identification accuracy is not a reflection of the capacity of the UKF to work with discontinuous systems; due to the identifiability of the parameters in all smooth branches of the response for the model expressed in Eq. (4.4), the discontinuous UKF [18] would have no benefit over the UKF in this case. Rather, these results suggest that the continuous approximation is a more representative model of the dynamics of the system.

As υ increases the approximation becomes a better representation of the base function, but begins to exhibit discontinuous behavior at lower sampling frequencies. The identification process is sensitive to these constraints. At a sampling frequency of 4096 Hz, a υ of 200 was found to provide an appropriate balance between model accuracy and continuity, and therefore the greatest enhancement to the model. This adjustment yields the dynamic model

$$m\ddot{x} + c_v\dot{x} + c_f\tanh(200\dot{x}) + kx + zx^3 = P, \qquad (4.5)$$

from which the parameters c_v, c_f, k, and z will be identified using different approximate Bayesian inference techniques in later chapters.

4.1.1 Experimental Setup

A photograph of the test configuration is shown in Fig. 4.3(a). The NES device used in this study is constructed with a 27.5-gauge steel wire and a central mass of 0.65kg with dimensions 4.5cm by 4.5cm by 7cm. The mass is constrained to move along a rail of length 23cm by the action of a ball-bearing slider attachment, which also reduces friction at the interface. The ends of the wire are pinned at support towers on either side of the mass, leaving a section of length 65.5cm between the supports. To reduce wire pre-tension (T in Eq. (4.2)), the wire is hand tightened to eliminate slack only. After tightening, the potential for slippage at the support towers during large excitations is reduced by the installation of a bead clamp, as shown in Fig. 4.3(b).

Base motion is used to excite the NES device. The excitation force P in Eq. (4.5) is therefore replaced with $-m\ddot{x}_g$ for the purpose of representing these experiments, where $\ddot{x}_g$ is the acceleration of the fixed ends of the wire. To achieve this experimentally, the device is mounted on a rigid steel plate driven by a Shore Western 1.1 kip 91-Series double-ended hydraulic actuator. The plate is supported vertically by flexible steel columns. Displacements are measured using Keyence LK-G157 laser displacement sensors [19], which have a range of +/- 4cm. Accelerations are measured using PCB model 3711d1fa20g accelerometers [20], which have a working range of +/- 20g at 0-1.5kHz and a mass of 0.0137kg. Data is collected at a 4096 Hz sampling frequency to capture the complex dynamic behavior of the NES device with sufficient accuracy for identification. A m+p Vibpilot station is used for data acquisition, which is equipped with 24-bit A/D converters and anti-aliasing filters.
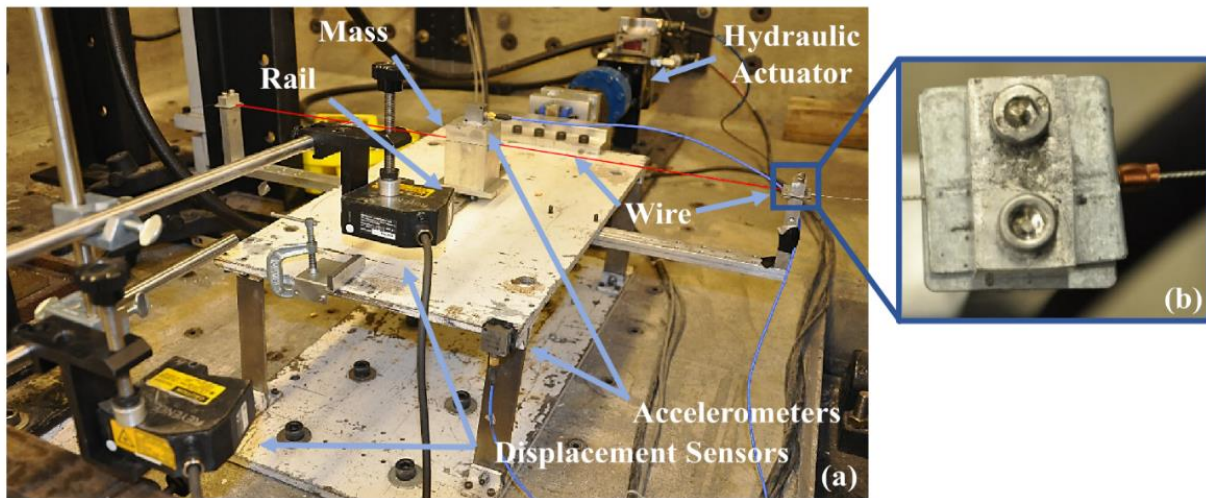
Fig. 4.3 – Experimental Setup (a) NES Device Configuration (b) Bead Clamp Configuration

Four common classes of excitations signals are used to generate motion in the NES device. These include a sine sweep (to generate a clear variation in NES response over a frequency range of interest), a sine wave (to generate a clear variation in the NES behavior with amplitude at a single frequency near the range of resonance), a step signal (to generate free response behavior in the system), and a band-limited white noise signal (to simulate the identification process under realistic operating conditions). The analytical forms of these excitation signals are shown in Table

4.1 and were given as displacement commands to the actuators during experimental testing. Where possible, the amplitude of these excitation signals was varied using multiple independent tests.

Table 4.1 – Base Excitation Signals for Identifiability Analysis

#	Input Type	Use	Mathematical Description of Input Displacement		A [mm]	F [Hz]	T [sec]
1	Sine Sweep	Chpt. 5 – Sobol' Analysis	$x_g = A\sin(\pi t^2 F/T)$	$t < T$	1.5	5	60
2					2.7	5	60
3			$x_g = \begin{cases} A\sin(\pi t^2 F/T) & t < T \\ A\sin(\pi t F[2T-t]/T) & T \le t < 2T \end{cases}$		1.5	5	40
4					2.7	5	40
5	Sine Wave		$x_g = \begin{cases} (At/T)\sin(2\pi Ft) & t < T \\ (A[2T-t]/T)\sin(2\pi Ft) & T \le t < 2T \end{cases}$		7.1	2.5	40
6	Step		$x_g = \begin{cases} (A/T)t & t < T \\ A & T \le t \end{cases}$		26.6	--	0.1
7	BLWN		$x_g = \text{lowPassFilter}(r,F)$	$r_{1xT} \sim N(0,A^2)$	0.8	6	60
8					2.6	6	60
9	Sine Sweep	Chpt. 6 – Identify	$x_g = \begin{cases} A\sin(\pi t^2 F/T) & t < T \\ A\sin(\pi t F[2T-t]/T) & T \le t < 2T \end{cases}$		2.7	5	40
10	Sine Wave		$x_g = \begin{cases} (At/T)\sin(2\pi Ft) & t < T \\ (A[2T-t]/T)\sin(2\pi Ft) & T \le t < 2T \end{cases}$		7.1	2.5	40
11	Sine Sweep	Chpt. 6 – Evaluate	$x_g = \begin{cases} A\sin(\pi t^2 F/T) & t < T \\ A\sin(\pi t F[2T-t]/T) & T \le t < 2T \end{cases}$		1.1	4	40
12					3.1	4	40
13	Sine Wave		$x_g = \begin{cases} (At/T)\sin(2\pi Ft) & t < T \\ (A[2T-t]/T)\sin(2\pi Ft) & T \le t < 2T \end{cases}$		5.8	2.0	40
14	Step		$x_g = \begin{cases} (A/T)t, & t < T \\ A, & T \le t \end{cases}$ $\begin{matrix} t < T \\ T \le t \end{matrix}$		26.6	--	0.1
15	BLWN		$x_g = \text{lowPassFilter}(r,F)$	$r_{1xT} \sim N(0,A^2)$	2.1	6	60
16					2.6	6	60

The Duffing-Type NES device considered in this study is characterized by a bifurcation of its stable response behavior about its range of resonance, where two stable steady state responses exist and are accessible depending on input energy. For input excitations exceeding a certain threshold, the device will exhibit a high amplitude response behavior that is desirable for energy pumping [3]. Below this threshold, low amplitude behavior is observed that is dominated by the minimal but nonzero amount of friction in the device. Varying the amplitudes of the input excitation signals ensures that the input energy dependence of the device will be captured in the identified model, thereby providing a more complete characterization of the device behavior. All

data from these tests, including measurements and recorded media, are published in Lund et al. [21].

The excitation signals in Table 4.1 are also partitioned according to their use in the subsequent identification and analysis. Excitation signals 1-8 are used in Chpt. 5 to assess the identifiability of the NES device with respect to various input excitations through an adapted global sensitivity analysis approach. Excitation signals 9-16 are used in Chpt. 6 to develop a robust method for UKF identification. Though there are similarities in the signals used for each analytical study, distinct amplitude and frequencies are typically used in the evaluation for Chpt. 6 to demonstrate the efficacy of the proposed predictive modeling technique.

4.1.2 Device Behavior

Signals 1 and 2 are used to observe the behavior of the NES device both below and above the input energy threshold, as shown in Fig. 4.4. The difference in input energy between the otherwise equivalent excitation signals allows for observation of the bifurcation in the response of the device.

The displacement response relative to the moving base, given as x in Fig. 4.1, is observed from three perspectives for each base excitation. The responses are displayed in terms of the relative displacement history, the relative displacement spectrogram, and the restoring force vs. relative displacement, where the restoring force is computed using the experimental measurements as in Eq. (4.3). The device response under signal 1 is provided in Fig. 4.4(a-c) and its response under signal 2 is provided in Fig. 4.4(d-f). These responses demonstrate the change in behavior in the device when it is excited about the input energy threshold and the impact of that bifurcation on the energy available at different frequencies of the response. For example, the relative displacement history in Fig. 4.4(a) shows a consistent low amplitude response after an initial transient displacement increase, whereas Fig. 4.4(d) shows a consistent increase in displacement amplitude. These responses are characteristic of the low and high amplitude behaviors of the device, respectively. Further information can be observed from the spectrograms for each signal, provided in Fig. 4.4(b,e). Fig. 4.4(b) shows that the device vibrates primarily in the 1st and 2nd harmonic modes of the forcing frequency, with the majority of the energy present in the 1st mode. In contrast, Fig. 4.4(e) shows that the device vibrates primarily in the 1st and 3rd harmonic modes of the forcing frequency, with significant energy in the 3rd mode. The initiation of vibration in the

3rd mode occurs at ~25sec, in conjunction with a transition in the displacement behavior in Fig. 4.4(a,d). These results confirm the expected bifurcation in the behavior of the device with respect to the amplitude of the excitation signal. Examination of the restoring force curves in Fig. 4.4(c,f) give some final insight into the impact of the equation of motion at different response amplitudes. Whereas the linear stiffness force and Coulomb damping forces govern at low amplitudes of displacement, as shown in Fig. 4.4(c), the cubic stiffness force clearly governs the response behavior when the device is excited to higher amplitudes of displacement.
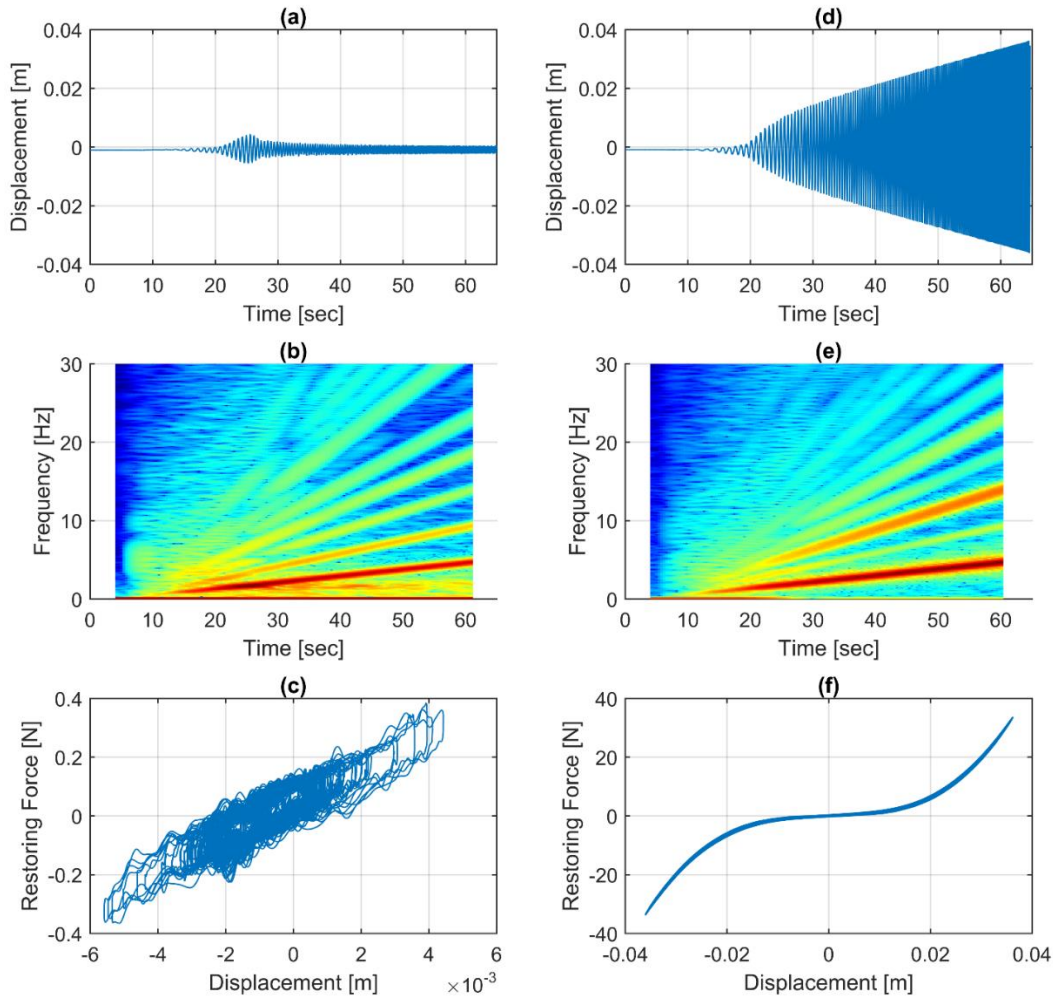
Fig. 4.4 – NES responses including displacement history, spectrogram, and hysteretic responses under: (a-c) Signal 1; (d-f) Signal 2. Acceleration responses have been low-pass filtered to 50 Hz for (c,f).

4.2   References

1.   Lund, A., Dyke, S. J., Song, W. & Bilionis, I. Global sensitivity analysis for the design of nonlinear identification experiments. *Nonlinear Dyn.* **98,** 375–394 (2019).

2.   Lund, A., Dyke, S. J., Song, W. & Bilionis, I. Identification of an experimental nonlinear energy sink device using the unscented Kalman filter. *Mech. Syst. Signal Process.* **136,** (2020).

3.   Gendelman, O., Manevitch, L. I., Vakakis, A. F. & M'Closkey, R. Energy pumping in nonlinear mechanical oscillators: Part I—Dynamics of the underlying hamiltonian systems. *J. Appl. Mech.* **68,** 34–41 (2001).

4.   Vakakis, A. F. & Gendelman, O. Energy pumping in nonlinear mechanical oscillators: Part II—Resonance capture. *J. Appl. Mech.* **68,** 42–48 (2001).

5.   Roberson, R. E. Synthesis of a nonlinear dynamic vibration absorber. *J. Franklin Inst.* **254,** 205–220 (1952).

6.   Nissen, J. C., Popp, K. & Schmalhorst, B. Optimization of a nonlinear dynamic vibration absorber. *J. Sound Vib.* **99,** 149–154 (1985).

7.   Kerschen, G., Kowtko, J. J., Mcfarland, D. M., Bergman, L. A. & Vakakis, A. F. Theoretical and experimental study of multimodal targeted energy transfer in a system of coupled oscillators. *Nonlinear Dyn.* **47,** 285–309 (2007).

8.   Vakakis, A. F., Manevitch, L. I., Gendelman, O. & Bergman, L. Dynamics of linear discrete systems connected to local, essentially non-linear attachments. *J. Sound Vib.* **264,** 559–577 (2003).

9.   Vakakis, A. F. *et al. Nonlinear Targeted Energy Transfer in Mechanical and Structural Systems I.* (Springer Science+Business Media, 2008).

10.   Wierschem, N. E. Targeted energy transfer using nonlinear energy sinks for the attenuation of transient loads on building structures. (University of Illinois at Urbana-Champaign, 2014).

11.   McFarland, D. M., Bergman, L. A. & Vakakis, A. F. Experimental study of non-linear energy pumping occurring at a single fast frequency. *Int. J. Non. Linear. Mech.* **40,** 891–899 (2005).

12.   Silva, C. E., Maghareh, A., Tao, H., Dyke, S. J. & Gibert, J. Evaluation of energy and power flow in a nonlinear energy sink attached to a linear primary oscillator. *J. Vib. Acoust.* (2019). doi:10.1115/1.4044450

13. Gourdon, E., Alexander, N. A., Taylor, C. A., Lamarque, C. H. & Pernot, S. Nonlinear energy pumping under transient forcing with strongly nonlinear coupling: Theoretical and experimental results. *J. Sound Vib.* **300,** 522–551 (2007).

14. Wierschem, N. E. *et al.* Passive damping enhancement of a two-degree-of-freedom system through a strongly nonlinear two-degree-of-freedom attachment. *J. Sound Vib.* **331,** 5393–5407 (2012).

15. Mcfarland, D. M. *et al.* Experimental investigation of targeted energy transfers in strongly and nonlinearly coupled oscillators. *J. Acoust. Soc. Am.* **118,** 791–799 (2005).

16. Nucera, F., Mcfarland, D. M., Bergman, L. A. & Vakakis, A. F. Application of broadband nonlinear targeted energy transfers for seismic mitigation of a shear frame: Computational results. *J. Sound Vib.* **329,** 2973–2994 (2010).

17. Green, P. L., Worden, K., Atallah, K. & Sims, N. D. The effect of Duffing-type non-linearities and Coulomb damping on the response of an energy harvester to random excitations. *J. Intell. Mater. Syst. Struct.* **23,** 2039–2054 (2012).

18. Chatzis, M. N. & Chatzi, E. N. A discontinuous unscented kalman filter for non-smooth dynamic problems. *Front. Built Environ.* **3,** 1–15 (2017).

19. Keyence. Model LK-G157 Technical Specifications. (2019).

20. PCB Piezotronics. Model 3711D1FA20G DC Accelerometer Installation and Operating Manual. (2013).

21. Lund, A., Silva, C., Dyke, S. J., Song, W. & Bilionis, I. Response of a prototype nonlinear energy sink device to various base excitations. *DesignSafe-CI* (2019). doi:10.17603/ds2-nh34-e513

5. GLOBAL SENSITIVITY ANALYSIS FOR THE ASSESSMENT OF EXPERIMENTAL IDENTIFIABILITY FOR NONLINEAR SYSTEMS

A significant component of the challenges of nonlinear identification is due to the variation in the identifiability of nonlinear systems with respect to different input excitations. Though many formal definitions of identifiability exist to suit different analysis methods [1], there is a general consensus that a system is *globally identifiable* when the evaluation of its inverse problem yields a unique solution in the parameter space, and is *locally identifiable* when it yields a finite number of solutions [2]. The *a priori* determination of system identifiability is vital to inform the design of a successful identification experiment [3],[4]. Generalized techniques exist for defining the identifiability of linear systems, but the extension of such techniques to nonlinear systems is only feasible for a few simple nonlinear systems [1]. As such, a number of methods have been developed for assessing the identifiability of nonlinear systems. A summary of some of the more common methods is presented in Fig. 5.1.
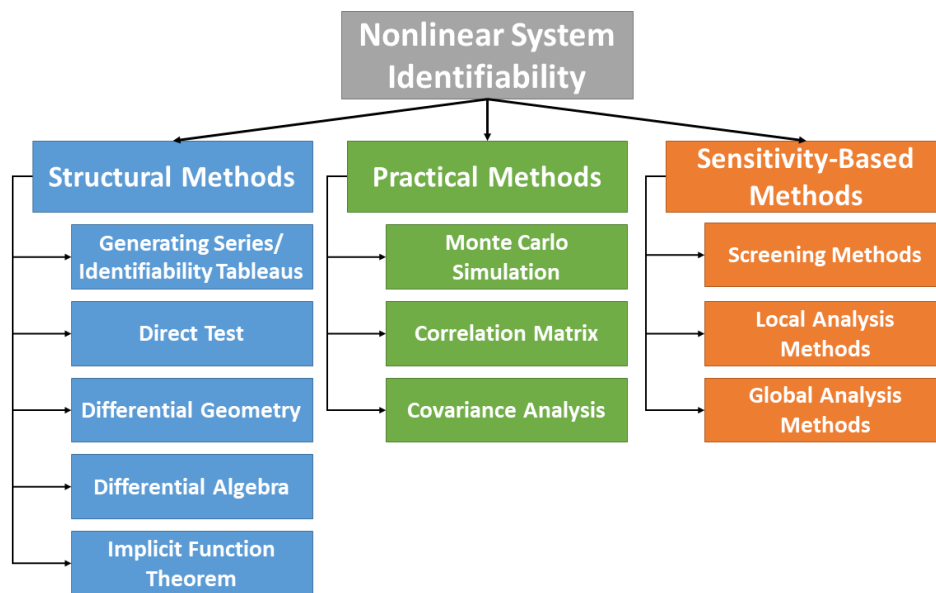
Fig. 5.1 – Summary of identifiability methods for nonlinear systems

The methods available for evaluating the identifiability of nonlinear systems can be grouped into three primary categories: structural methods, practical methods, and sensitivity-based

methods. Structural methods directly analyze the system under the assumption of a perfect model with no measurement noise in the response. These methods are typically analytically intensive, requiring advanced mathematical skill from the user [1]. In some cases, such as for the differential algebra [5],[6] and generating series approaches [7]–[9], the methods can be generalized to a broad range of models. As such, software tools have been developed to facilitate the application of these methods by general users [10]–[12]. The generating series and differential algebra methods, in particular, are used frequently in a variety of disciplines, such as biological sciences [13]–[15] and engineering [2],[3],[16].

Experimental systems exhibit behaviors that often deviate significantly from the theoretical models due to noise and modeling errors. Such cases can lead to systems which are structurally identifiable, but practically unidentifiable [1],[2],[14]. Practical methods seek to overcome some of these limitations by assessing the identifiability of nonlinear system models in the presence of noise, though the absence of modeling errors is still assumed. Offline methods, such as Monte Carlo (MC) simulations [17] or correlation matrix analysis [18], can be used to determine identifiability in advance of any experimentation for the purpose of experimental design. However, these methods can be challenging to implement due to the computational cost of the MC method and the reliability with which the correlation matrix must be estimated [1]. An online method for analyzing the identifiability which works in parallel with identification techniques using Bayesian inference has recently been introduced by Muto and Beck [19] and Olivier and Smyth [2]. This method, referred to in Fig. 5.1 as covariance analysis, compares the *a priori* and *a posteriori* covariance on the parameters as an indicator of identifiability, allowing the user to determine the reliability of the identified parameters. Though this method is extremely useful for a post-identification check on the identifiability of the parameters, it is less helpful in the design of experiments.

Sensitivity-based approaches offer a pragmatic perspective on the identifiability problem as they strike a balance between the methods previously discussed. Instead of directly determining identifiability from the model form, as structural methods do, sensitivity-based approaches interpret identifiable parameters as those which the model is most sensitive to, or which produce the most variation in its output. Though these approaches share the same assumptions as the structural methods regarding model form and measurement noise, they also require the user to identify nominal parameter values or ranges over which to evaluate the sensitivity, as with practical

methods [1]. Sensitivity analysis is in itself a rich field, and a range of methods exist that can be applied to models of varying complexity and computational demand [20]. Applications of this class of methods for the purpose of establishing identifiability almost exclusively use local analysis methods [18],[21], which are not computationally intensive, but are less effective for models with severe nonlinearities and parameters of varying levels of uncertainty [20]. In these cases, global sensitivity analysis methods present a promising alternative, despite their increased computational cost, as they present a greater depth of information that can be used to establish identifiability [22]–[27].

In this chapter the utility of one global sensitivity analysis method, the Sobol' approach [23],[28], is investigated for the design of nonlinear system identification experiments. This approach is examined for nonlinear system identifiability within the context of identifying the prototype nonlinear energy sink (NES) device described in Chpt. 4, subject to the base excitation signals 1-8 detailed in Table 4.1. It is important to note that the complexity introduced by the Coulomb damping term in the NES model prevents analysis by many of the more common structural identifiability approaches, such as the differential algebra software DAISY [10]. For the purpose of experimental design, the available identifiability analysis approaches for this system are essentially reduced to practical and sensitivity-based methods. Sobol' sensitivity analysis is therefore applied as a novel approach to indicate the success of the parameter identification with respect to a given excitation, prior to experimentation. To demonstrate the approach, the measured responses are used with the unscented Kalman filter (UKF) to identify the parameters of a representative device model. Through this analysis it is determined that Sobol' sensitivity analysis is particularly beneficial for experimental design as it establishes a simple, heuristic measure of the practical identifiability of a general nonlinear system prior to experimentation.

The chapter is organized as follows: Section 5.1 introduces the Sobol' methodologies; Section 5.2 discusses the implementation of the UKF for this case study; and Section 5.3 provides the results of both the sensitivity study and the parameter identification. The main conclusions are summarized in Section 5.4. The content of this chapter has been adapted from the published work of the author [29].

5.1 Sobol' Sensitivity Analysis

Variance-based sensitivity analysis methods are widely used to provide a summary measure of the response uncertainty associated with various input parameters. These global methods have the benefits of being model-independent, estimating the interaction effects among parameters through higher order sensitivity indices, and capturing the full range of variation of each input parameter [26].

Of the available variance-based methods, Saltelli's extended version [28] of the Sobol' approach [22],[23],[30] provides the most complete and accurate estimation of model sensitivity. The method generates MC sampling average estimates of the conditional variance of the model output, requiring $P(K + 1)$ simulation runs to determine the first-order sensitivity, where P is the number of MC samples and K is the number of input parameters. Typically, $500 - 1000$ MC samples are drawn to ensure sufficient accuracy in the estimation, where necessary increases in model evaluations due to the increased dimensionality of the system are handled through K. The computational cost accrued through these model evaluations is the key limitation in the practical application of this method, making it most applicable to low-dimensional models whose computation times are on the order of seconds [26].

The method is developed around the generalized system model

$$Y = f(\Theta_1, \Theta_2, \dots, \Theta_K), \tag{5.1}$$

where Θ_j $(j = 1, \cdots, K)$ are the K independent random variables describing the parameters of the model. The sensitivity of the model output Y to each of the model inputs Θ_j can then be expressed through the variance, denoted $\mathbb{V}[\cdot]$, of the conditional expectation, denoted $\mathbb{E}[\cdot]$, on the model output as,

$$\mathbb{V}_{\Theta_j}\left[\mathbb{E}_{\mathbf{\Theta}_{\sim j}}[Y|\Theta_j]\right] = \mathbb{V}[Y] - \mathbb{E}_{\Theta_j}\left[\mathbb{V}_{\mathbf{\Theta}_{\sim j}}[Y|\Theta_j]\right]$$

$$= \int \mathbb{E}^2[Y|\Theta_j = \theta_j]\, p_{\Theta_j}(\theta_j)d\theta_j - \mathbb{E}^2[Y], \tag{5.2}$$

where $\mathbf{\Theta}_{\sim j}$ refers to the set of all parameters except Θ_j. This measure of sensitivity examines the expected response of the model when all parameters are allowed to vary except for the parameter of interest (Θ_j). Tracking the variation of this expectation over all possible realizations of Θ_j captures the portion of the variance of the model which can be directly attributed to that particular

parameter. Division of this term by the unconditional variance yields the first-order sensitivity index

$$S_j = \mathbb{V}_{\Theta_j}\Big[\mathbb{E}_{\Theta_{\sim j}}[Y|\Theta_j]\Big]\Big/\mathbb{V}[Y],\tag{5.3}$$

which has the convenient property that $S_j \in [0,1]$. Sensitivity indices closer to 1 indicate that a larger portion of the total variance can be attributed to that parameter. The parameter is therefore more likely to be identifiable.

The MC computation of the sensitivity at first appears impractical due to the P^2 simulation runs implied by the computation of the integral in Eq. (5.2). However, Saltelli's extension of the Sobol method produces a convenient algorithm to reduce this computational cost [26]. First, two matrices,

$$\mathbf{A} = \begin{bmatrix} \theta_1^{(1)} & \theta_2^{(1)} & \dots & \theta_j^{(1)} & \dots & \theta_K^{(1)} \\ \theta_1^{(2)} & \theta_2^{(2)} & \dots & \theta_j^{(2)} & \dots & \theta_K^{(2)} \\ \vdots & \vdots & & & \ddots & \\ \theta_1^{(P)} & \theta_2^{(P)} & \dots & \theta_j^{(P)} & \dots & \theta_K^{(P)} \end{bmatrix},\tag{5.4}$$

and

$$\mathbf{B} = \begin{bmatrix} \hat{\theta}_1^{(1)} & \hat{\theta}_2^{(1)} & \dots & \hat{\theta}_j^{(1)} & \dots & \hat{\theta}_K^{(1)} \\ \hat{\theta}_1^{(2)} & \hat{\theta}_2^{(2)} & \dots & \hat{\theta}_j^{(2)} & \dots & \hat{\theta}_K^{(2)} \\ \vdots & \vdots & & & \ddots & \\ \hat{\theta}_1^{(P)} & \hat{\theta}_2^{(P)} & \dots & \hat{\theta}_j^{(P)} & \dots & \hat{\theta}_K^{(P)} \end{bmatrix},\tag{5.5}$$

are generated, each containing P quasi-random samples of the input parameters $\Theta_{1:K}$ from known parameter ranges according to the LP_τ method [31]. The notation $\theta_j^{(i)}$ and $\hat{\theta}_j^{(i)}$ is used to emphasize the fact that the samples contained in each matrix will take different values. $\mathbf{A}$ and $\mathbf{B}$ are then used to construct

$$\mathbf{C}_j = \begin{bmatrix} \hat{\theta}_1^{(1)} & \hat{\theta}_2^{(1)} & \dots & \theta_j^{(1)} & \dots & \hat{\theta}_K^{(1)} \\ \hat{\theta}_1^{(2)} & \hat{\theta}_2^{(2)} & \dots & \theta_j^{(2)} & \dots & \hat{\theta}_K^{(2)} \\ \vdots & \vdots & & & \ddots & \\ \hat{\theta}_1^{(P)} & \hat{\theta}_2^{(P)} & \dots & \theta_j^{(P)} & \dots & \hat{\theta}_K^{(P)} \end{bmatrix},\tag{5.6}$$

in which the sample values for the parameter of interest Θ_j are taken from **A** and the remainder of the sample values are taken from **B**. Model responses are then generated using the parameters in each row of **A** and $\mathbf{C}_{1:K}$, yielding the response vectors (Px1)

$$y_A = f(A), \tag{5.7}$$

$$y_{C_j} = f(C_j), \tag{5.8}$$

These response vectors provide the MC samples necessary to compute the unconditional variance (y_A) and the variance conditioned on a particular parameter (y_{C_j}). The first-order sensitivity index for the parameter of interest Θ_j can therefore be estimated by comparing the variance of y_{C_j} with y_A, such that

$$S_j = \frac{\mathbb{V}_{\Theta_j}\left[\mathbb{E}_{\Theta_{\sim j}}[Y|\Theta_j]\right]}{\mathbb{V}[Y]} \approx \frac{(1/P)\sum_{i=1}^{P} y_A^{(i)} y_{C_j}^{(i)} - f_0^2}{(1/P)\sum_{i=1}^{P}(y_A^{(i)})^2 - f_0^2}, \tag{5.9}$$

where

$$f_0 = (1/P)\sum_{i=1}^{P} y_A^{(i)}. \tag{5.10}$$

The cost of this approach amounts to N model evaluations to generate y_A and PK model evaluations to generate $y_{C_{1:K}}$, or a total of $P(K + 1)$ model evaluations to estimate the first-order sensitivity index with respect to all system parameters. Additional details concerning the derivation of the method can be found in [28].

The sensitivity analysis in this paper is performed using the open-source computational tool SaLib [32]. The parameters from the physical model given in Eq. (4.5) are uniformly sampled (P=1000) over the ranges $c_v \in [0,1]$ Ns/m, $c_f \in [0,1]$ N, $k \in [0,100]$ N/m, and $z \in [5 \times 10^5, 9 \times 10^5]$ N/m³, which describe the range of uncertainty of the device parameters. The sensitivity of the displacement and acceleration responses are analyzed with respect to simulated input signals corresponding to those given in Table 4.1.

5.2   Unscented Kalman Filter Implementation

The UKF is used in this study to assess the results of the Sobol' analysis and their implications on the practical identifiability of the NES system. The UKF algorithm is largely set forth in the description of sigma point Kalman filters given in Algorithm 2.3 and is visualized in Fig. 5.2. For

this particular sigma point filter adaptation, the unscented transform [33] is used to estimate the necessary moments of the PDFs through a set of $N_{sig} = 2n + 1$ sigma points, where n is the number of states. These sigma points are deterministically selected according to

$$\mathcal{X}_{k-1} = \left[\tilde{\mathbf{x}}_{k-1}, \tilde{\mathbf{x}}_{k-1} \pm \sqrt{n' + \lambda'} \, \text{chol}\big(\widetilde{\mathbf{P}}_{k-1}\big)\right] \tag{5.11}$$

$$\mathcal{X}_k^- = \left[\tilde{\mathbf{x}}_k^-, \tilde{\mathbf{x}}_k^- \pm \sqrt{n' + \lambda'} \cdot \text{chol}\big(\widetilde{\mathbf{P}}_k^-\big)\right] \tag{5.12}$$

and have the associated weights

$$\mathbf{W}^{(m)} = \left[\frac{\lambda}{N_{sig} + \lambda}, \left\{\frac{1}{2\big(N_{sig} + \lambda\big)}\right\}_i\right], \quad i = 1, \dots, 2n \tag{5.13}$$

$$\mathbf{W}^{(c)} = \left[\frac{\lambda}{N_{sig} + \lambda} + (1 - \alpha^2 + \beta), \left\{\frac{1}{2\big(N_{sig} + \lambda\big)}\right\}_i\right], \quad i = 1, \dots, 2n, \tag{5.14}$$

where the tuning parameters $\alpha = 1$, $\beta = 3$, and $\lambda = \alpha^2 \cdot \big(N_{sig} + \kappa\big) - N_{sig}$ with $\kappa = 0$ were found to perform well with this data set. Though a standard set of tuning parameters is recommended [34], varying these parameters can accommodate experimental deviations from the Gaussian assumption. Note that the sigma point set used herein is constructed in relation to the state of the system which has been augmented by the noise parameters, as $\tilde{\mathbf{x}}_{k-1} = [\mathbf{x}_{k-1} \quad \mathbf{w}_{k-1}]^T$ and $\tilde{\mathbf{x}}_k^- = [\mathbf{x}_k^- \quad \mathbf{v}_k]^T$, such that $\mathcal{X}_{k-1} = [\boldsymbol{\chi}^{(\mathbf{x}_{k-1})} \quad \boldsymbol{\chi}^{(\mathbf{w}_{k-1})}]^T$ and $\mathcal{X}_k^- = [\boldsymbol{\chi}^{(\mathbf{x}_k^-)} \quad \boldsymbol{\chi}^{(\mathbf{v}_k)}]^T$ [35].
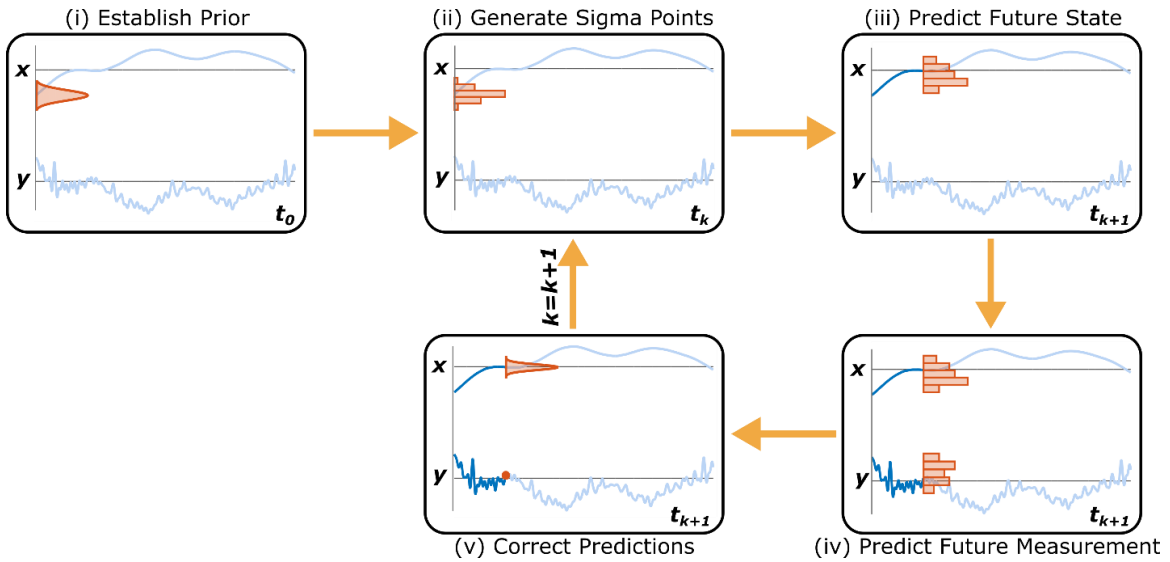
Fig. 5.2 - Overview of the UKF algorithm. The UKF approximates the filtering distributions of the Kalman filter through the selection and propagation of a weighted sigma point set.

Joint state and parameter identification is typically accomplished through state augmentation, which has been shown to work effectively with the UKF for a variety of nonlinear systems [2],[33]. The performance of the algorithm can be enhanced by exploiting prior information about the parameters to transform the augmented states [36]. For this system, both the relative magnitude of the parameters and their existence in the positive domain are known *a priori*. This information can be used to transform the problem into an unconstrained optimization of parameters which display similar magnitudes of variability and is therefore more suited to the construction of the UKF algorithm. The parameters are first scaled to similar orders of magnitude using the dimensionless scale factors shown in Eq. (5.15) and Table 5.1. The scaled parameters are then log-transformed to map their domain from the positive reals to the Euclidean space. These transformations are given in greater detail in Table 5.1.

Table 5.1 – Scaled and Transformed UKF States

NES Parameter	UKF State
$x_3' = c_v$ Ns/m	$x_3 = \log(c_v/0.1)$
$x_4' = c_f$ N	$x_4 = \log(c_f/0.1)$
$x_5' = k$ N/m	$x_5 = \log(k/10)$
$x_6' = z$ N/m^3	$x_6 = \log(z/10^5)$

Using these transformed states, an augmented state vector is defined for the NES device model given in Eq. (4.5) such that

$$\mathbf{x} = \left[x, \dot{x}, \log\left(\frac{c_v}{0.1}\right), \log\left(\frac{c_f}{0.1}\right), \log\left(\frac{k}{10}\right), \log\left(\frac{z}{10^5}\right) \right]^T$$
$$= [x_1, x_2, x_3, x_4, x_5, x_6]^T, \tag{5.15}$$

with the experimental responses and model parameters recorded in SI units. The state transition function, f(·), can be expressed as

$$f(\mathbf{x}, \ddot{x}_g) = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \end{bmatrix} = \begin{bmatrix} x_2 \\ -\ddot{x}_g - (x_3' x_2 + x_4' \tanh(v x_2) + x_5' x_1 + x_6' x_1{}^3)/m \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (5.16)$$

and the observation function, $h(\cdot)$, as

$$h(\mathbf{x}_k, \ddot{x}_g) = \begin{bmatrix} x_1 \\ -(x_3' x_2 + x_4' \tanh(v x_2) + x_5' x_1 + x_6' x_1{}^3)/m \end{bmatrix}. \qquad (5.17)$$

Note that the measurement function is constructed to use simultaneous observations of the relative displacement (x in Fig. 4.1) and absolute acceleration of the mass of the NES device.

The initial Gaussian distribution on the final augmented state vector can now be described by the mean and covariance

$$\mathbf{\mu}_0 = \begin{bmatrix} \mathbf{\mu}_0^{(x)} & \mathbf{\mu}_0^{(\theta)} \end{bmatrix}^T \qquad (5.18)$$

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{P}_0^{(x)} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0^{(\theta)} \end{bmatrix}, \qquad (5.19)$$

where $\mathbf{\mu}_0^{(x)}$ refers to the portion of the mean vector associated with the dynamic states and $\mathbf{\mu}_0^{(\theta)}$ refers to that portion associated with the unknown parameters. $\mathbf{P}_0$ is likewise partitioned. Note that the uncertainty in the states x_n ($n = 1, 2, \dots, 6$) is assumed independent, which implies that $\mathbf{P}_0$ is a diagonal matrix with entries $(\mathbf{P}_0)_{nn} = \sigma_n^2$.

The selection of the initial Gaussian distribution for the partitioned state vector is then determined by the judgement of the user and the known information on the system. For this system the initial distribution on the dynamic states is selected such that $\mathbf{\mu}_0^{(x)} = \mathbf{0}$, $\sigma_1 = 0.01$m, and $\sigma_2 = 0.01$ m/s, as the observation of the system is initiated prior to device excitation. Uncertainty as to the true range of variability in the parameters complicates the selection of a prior distribution on their respective states, as variation in the prior over a likely parameter range can yield wide swings in the identification results. In this study, a sampling approach on the prior distributions of the parameters is used to analyze this variability as a tool to study the relation between the Sobol' indices and the UKF identifiability. Details on this analysis process are given in Section 5.3.

For this experimental application the process and measurement noise covariance matrices are assumed constant and are estimated prior to UKF implementation through the analysis of

sensor outputs in a baseline experiment. To estimate the baseline noise in the system, measurements are taken while a constant zero command signal is sent to the hydraulic actuator, i.e. the hydraulic pump and actuator are powered on, but the actuator command is a zero input. The process noise covariance matrix ($\mathbf{Q}$) represents error in the assumed model form, and is estimated here as the variance of the base acceleration signal, which is $\sigma_b^2 = 1.03 \times 10^{-2}(\text{m/s}^2)^2$. The measurement noise covariance matrix ($\mathbf{R}$) represents the noise in the device response due to the sensors. The variance of the displacement signal in the baseline experiment is found to be $\sigma_d^2 = 1.44 \times 10^{-12}\text{m}^2$ and the variance of the NES acceleration signal is $\sigma_a^2 = 5.32 \times 10^{-3}(\text{m/s}^2)^2$. The noise in the various measurement signals are assumed to be independent of each other, and the accelerometer signals are left unfiltered for UKF implementation.

The accuracy of each identified model is assessed in terms of a normalized mean square error [37], given as

$$\text{MSE} = \frac{100}{M} \sum_{i=1}^{M} \left(\frac{(x_i - \hat{x}_i)^2}{\sigma_d^2} + \frac{(\ddot{x}_i - \hat{\ddot{x}}_i)^2}{\sigma_a^2} \right), \tag{5.20}$$

where M is the total number of observations of the system, x_i is the ith observed displacement value, and $\hat{x}_i$ is the displacement simulated for the experimental excitation using the identified model. It is important to note that the term inside the summation is equivalent to the log-likelihood of the measurements given information about the states and the parameters ($\log(\text{p}(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta}))$). The minimization of the log-likelihood of the measurements implies the minimization of the mean square error on the hidden states, which is what the Kalman filter is designed to accomplish [38].

5.3   Results and Discussion

Identification is performed with the NES device model in Eq. (4.5) using the UKF with the experimental system responses to the inputs listed in Table 4.1. To address the uncertainty in the selection of an initial distribution of the parameters for the UKF algorithm, the joint identification process is conducted multiple times for each signal, starting from a variety of initial distributions over the unknown parameters. Latin hypercube sampling is used to generate a set of 500 initial means over the parameter ranges used for sensitivity analysis. For each set of initial means, 10 covariance matrices are generated such that $\sigma_n = \alpha(\mu_0)_n$ $(n = 3,4,5,6)$, where $\alpha \in \{0.1, 1, 10, 25, 50, 75, 100, 200, 300, 600\}\%$ is the coefficient of variation. This results in 5,000

identification trials for each experimental response signal. The device model identified from each experimental signal is selected as the one which minimizes the MSE.
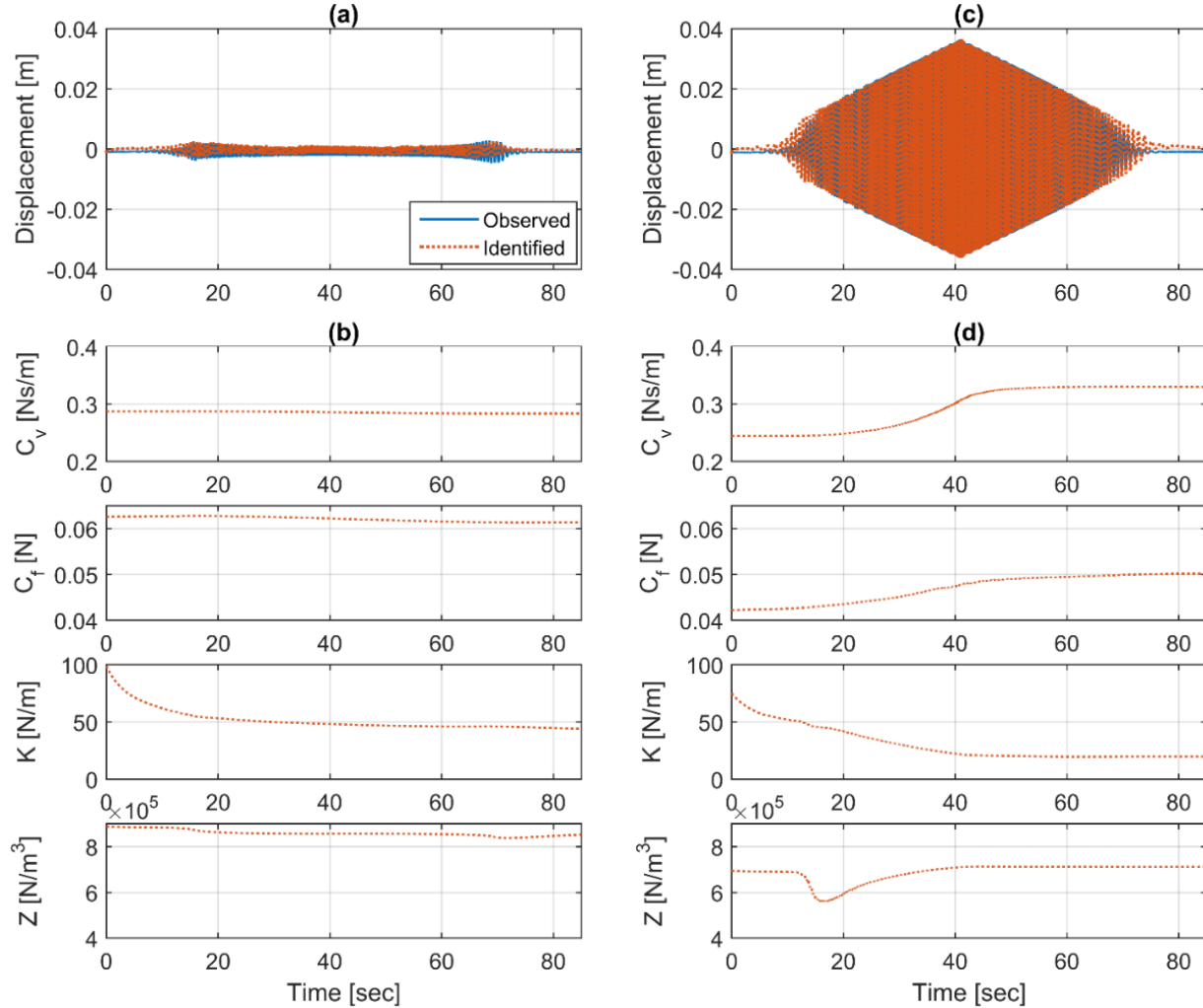
Fig. 5.3 – Sample UKF Results for Sine Sweep Signals: (a-b) Signal 3, (c-d) Signal 4

The parameters identified with signals 3 and 4 (Table 4.1) using the approach described herein provide an illustrative case study for an initial examination of system identifiability. These signals develop, over an extended time period, a clear variation in the device responses about the input energy threshold with otherwise equivalent signals. Simulated responses corresponding to these signals are generated from the identified device models and shown in Fig. 5.3. These simulated responses are compared with the observed experimental data, with low amplitude results given in Fig. 5.3(a) and high amplitude results given in Fig. 5.3(c). The parameter identification

histories for the respective signals are given in Fig. 5.3(b,d). The estimated responses demonstrate that a good fit to the experimental responses can be obtained through the UKF identification strategy. However, the identification process also yields multiple sets of parameters, which is more clearly evidenced by the numerical results in Table 5.2.

Table 5.2 – Parameters obtained with the UKF Method using Sine Sweep Responses

Parameter		Identified Value		Variance Ratio [%]	
		Signal 3 (Low)	Signal 4 (High)	Signal 3 (Low)	Signal 4 (High)
c_v	[Ns/m]	0.283	0.330	99.6	61.7
c_f	[N]	0.061	0.050	94.4	92.3
k	[N/m]	44.0	19.8	36.7	15.3
z	[N/m^3]	8.52×10^5	7.12×10^5	97.7	0.9

The occurrence of multiple parameter sets may generate some confusion over which model to select. Olivier and Smyth [2] suggest that the differences between the models may be interpreted as an identifiability issue. By their analysis approach, similarity between *a priori,* $\mathbf{P}_0$*,* and *a posteriori,* $\mathbf{P}_M$, covariance levels indicates that little information concerning the parameter is available in the signal and that the parameter is therefore unidentifiable. This similarity measure is stated in Table 5.2 as the variance ratio, which is defined as $(\sigma_M)_n/(\sigma_0)_n$. Values of the variance ratio that are close to 100% indicate high similarity between the *a priori* and *a posteriori* covariance and therefore suggest the parameter is not likely to be identifiable. Applying this measure to the identification histories for signals 3 and 4 suggests that the linear stiffness parameter is likely to be identifiable with either signal, and that the Coulomb damping coefficient will likely be identified with neither signal. However, this result does not appear to be in agreement with the identified values of the parameters. Comparing the UKF results, the Coulomb damping coefficient is identified with the greatest precision, whereas the linear stiffness term identified with signal 3 is more than twice the value identified with signal 4. Though these differences may be able to be understood by further analyzing the prior parameter distributions with respect to the noise present in the experimental system, the initial results can be misleading and are somewhat time consuming to generate. A Sobol' sensitivity analysis offers a different perspective on identifiability, *in*

advance of experimental implementation, that is helpful both in clarifying the identification results and in designing the experiment.

A Sobol' sensitivity analysis is conducted on the NES device model with respect to simulated low and high amplitude input signals modeled after experimental signals 3 and 4 (Table 4.1). The amplitudes of these simulated excitation signals are selected such that all $P(K + 2)$ sample responses fall either below the input energy threshold, in the case of the low amplitude excitation, or above it, in the case of the high amplitude excitation. The results of this analysis for the two signals are given in Fig. 5.4 and Fig. 5.5, respectively, and show a sample model response (a) and the time-varying first-order sensitivity of the displacement response to each parameter (b-d). The plots include the 95% confidence interval of that sensitivity. First-order sensitivities are used for examination because they quantify the information available in the response that is related to the direct impact of a parameter, instead of its potential interactions with other parameters. The sensitivity of the acceleration responses with respect to the parameters yields results similar to those developed for the displacement responses, and are therefore not shown.
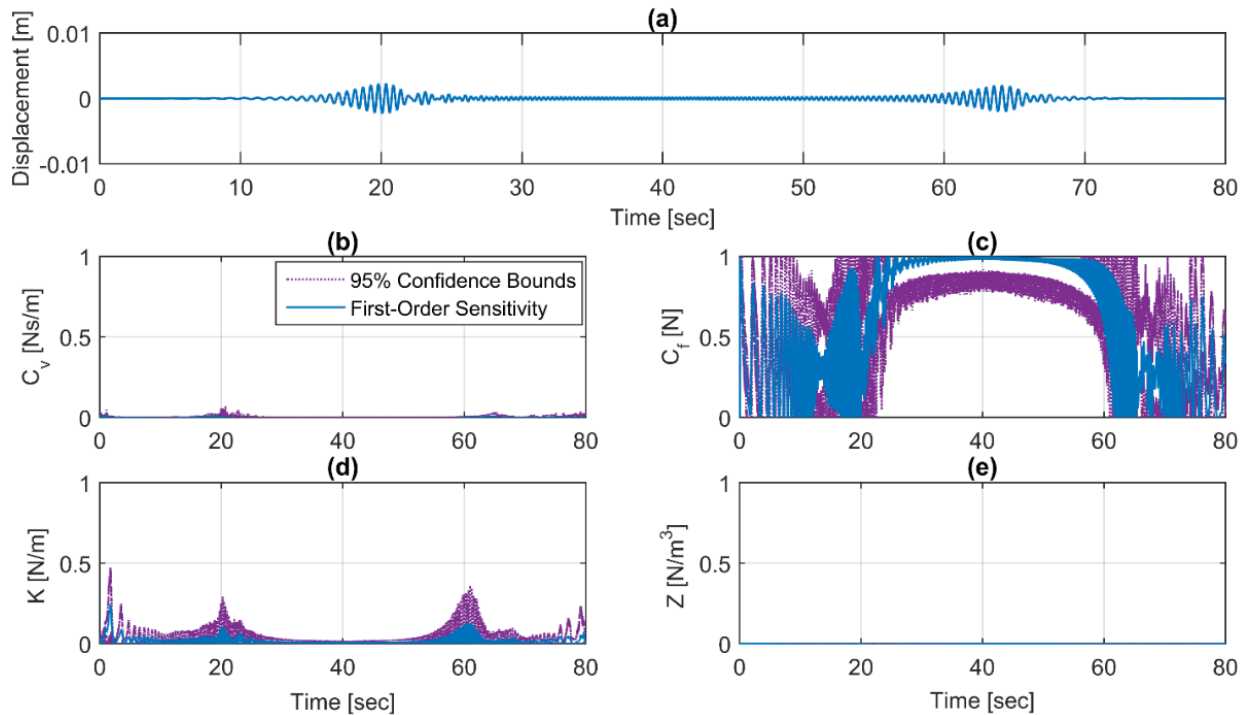
Fig. 5.4 – Sensitivity Analysis Results for a Low Amplitude Sine Sweep (Signal 3)

The sensitivity results given in Fig. 5.4 indicate that for a low amplitude sine sweep excitation, the NES model is not sensitive to the nonlinear stiffness parameter, z, and is minimally sensitive to the linear stiffness, k, and viscous damping, c_v, parameters. These results suggest that there is very little information available in this signal that can be used to reliably identify the nonlinear stiffness parameter. The only parameter for which a reliable estimate may be obtained appears to be the Coulomb damping coefficient, c_f. In contrast, Fig. 5.5 shows that for a high amplitude sine sweep excitation the NES model is quite sensitive to all parameters over the course of its response. The results of this sensitivity analysis agree with the observations from the identification results in Table 5.2. The responses of this nonlinear model to both signals are highly sensitive to the Coulomb damping coefficient, which appear to be identified with the great precision despite the low identifiability predicted by the variance ratio analysis. The response to the high amplitude signal is especially sensitive to the nonlinear stiffness, which does agree with the high likelihood of identifiability suggested by the variance ratio analysis.
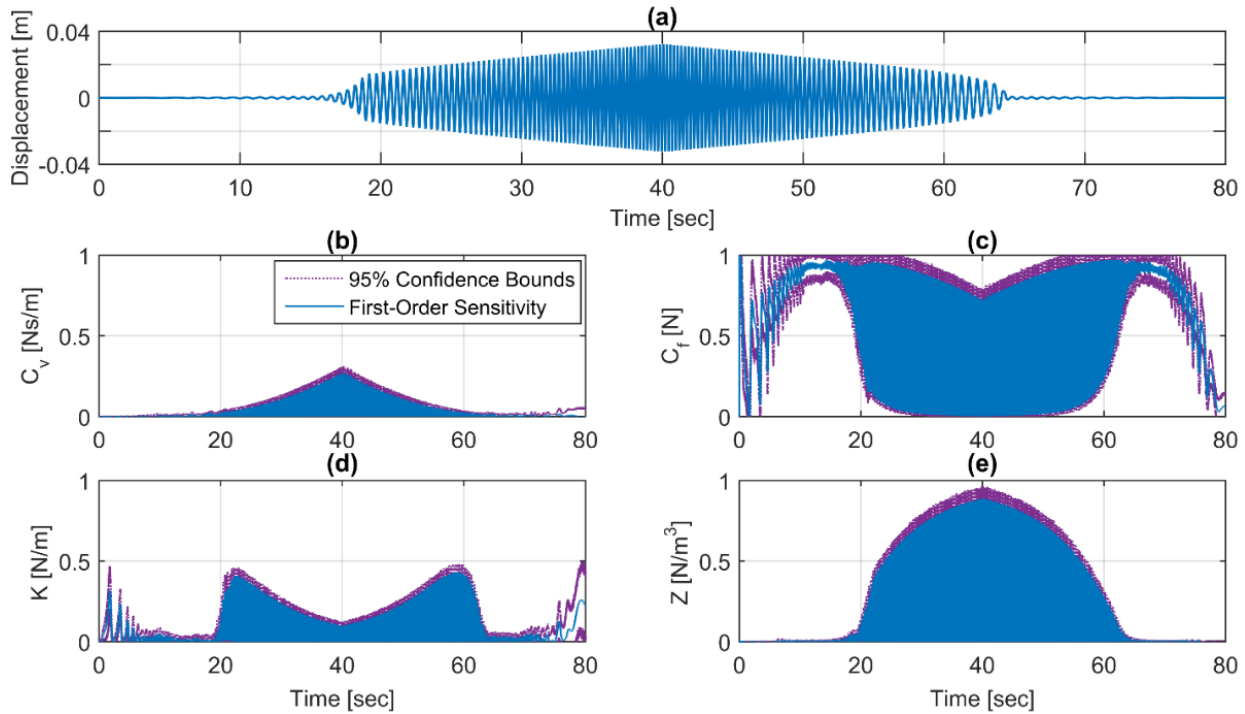
Fig. 5.5 – Sensitivity Analysis for a High Amplitude Sine Sweep (Signal 4)

The sensitivity results also provide an indication of how each parameter affects the various response behaviors in the NES device. For example, Fig. 5.5(c,d) shows that the sensitivity of the

linear stiffness parameter and the Coulomb damping coefficient both peak at ~20 sec and ~60 sec. Comparing these points in time with the displacement response history in Fig. 5.5(a) leads to the conclusion that these parameters play a primary role in governing the sudden shift in NES device behavior, which for this sample response occurs at ~18 sec and ~65 sec. In this transition region, the device switches from its characteristic low amplitude behavior in which the frictional damping force controls, to its characteristic high amplitude behavior in which the nonlinear stiffness force governs. The temporal location of the transition region is governed by the linear stiffness and Coulomb damping parameters, leading to the increased sensitivity of the parameters within that transition region.
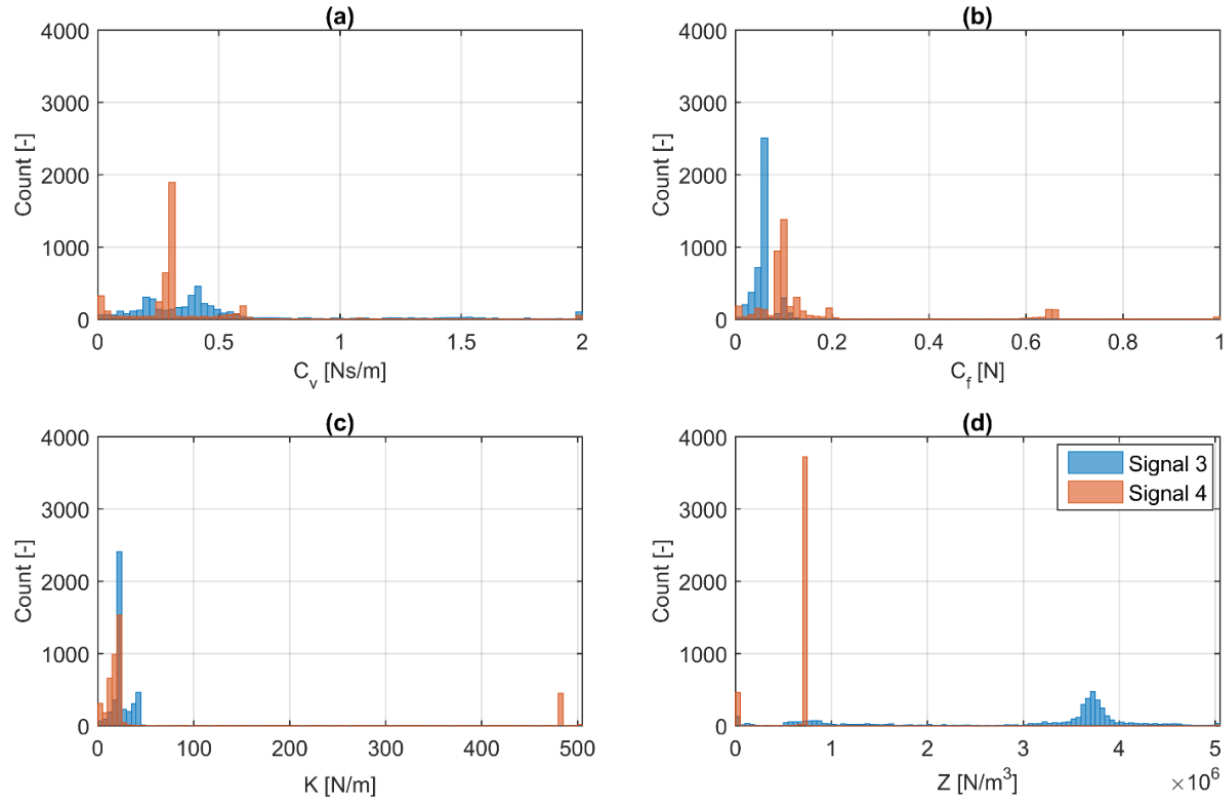
Fig. 5.6 – Distribution of Identified Parameters using Signals 3 and 4

The agreement between the first-order sensitivity results and the identifiability of the NES device is further supported through examination of the distribution of the models generated with experimental signals 3 and 4 in the 5,000 UKF identification trials, given in Fig. 5.6. The Sobol' sensitivity results suggest that there is little information in the low amplitude frequency sweep

response which can be used to accurately identify the viscous damping and nonlinear stiffness parameters. This lack of information is likely to translate into a large spread in the UKF identification results, as the variation of those parameters is less influential in approximating the system response. The identification results in Fig. 5.6 show that, of the four parameters, the distributions for the viscous damping (Fig. 5.6(a)) and nonlinear stiffness (Fig. 5.6(d)) exhibit the greatest differences in parameter distributions. The distribution identified for both parameters with the low amplitude sine sweep response conveys that a wide range of parameters can be used to generate an acceptable model of that response, whereas the narrow distribution of identified parameters for the high amplitude sine sweep response suggests that a precise value of the respective parameters is required to construct an accurate model. In contrast, the distributions for the Coulomb damping coefficient and the linear stiffness show similar spread between the two signals, indicating that they are similarly identifiable with both signals. This result is interesting as it suggests that relatively low levels of sensitivity, such as those shown in Fig. 5.4(d), may be sufficient to generate a good quality identified model.

Next the comparison between the sensitivity results and the distributions of identified parameters is extended to the other experimental response signals, as shown in Fig.5.7. Results are given with respect to the nonlinear stiffness parameter (z) only, as this parameter shows the widest variation in identifiability. Each horizontal pair of subplots provides a comparison between the sensitivity results and identified parameter distribution for a particular signal type. Note that the amplitude scale of the histogram in Fig.5.7(b) is extended to accommodate the concentration of the distribution.

For each of the signals, a significant portion of the identified nonlinear stiffness parameters appear at an unrealistically low number. This result occurs due to the method used to produce the 5,000 identification trials. Poorly tuned values of $\mathbf{P}_0^{(\theta)}$ (see (5.19)) have the potential to produce unrealistic identification results. Because the sampling method requires trials to be run for an especially wide range of $\mathbf{P}_0^{(\theta)}$, this result is expected and this set of identified parameters is disregarded.

All responses exhibit behaviors that are similar to those previously studied for signals 3 and 4. Cases with uniformly low sensitivity, as shown in Fig.5.7(a,e), exhibit a wide distribution of identification results. Those exhibiting periods of high sensitivity correspond to tight distributions. It is interesting to note that some gradation exists in these results. The sensitivity

response for signal 2, shown in Fig.5.7(b), exhibits the highest value of sensitivity over the longest signal duration, and also corresponds to the tightest distribution of the identified nonlinear stiffness parameter. Signal 5, in Fig.5.7(c), exhibits the second most significant sensitivity response, and corresponds to the second tightest parameter distribution. The sensitivity responses to signals 6 and 8 both exhibit a short duration of high sensitivity, with the majority of the response being minimally sensitive to the nonlinear stiffness parameter. The corresponding parameter distributions are more spread than those previously mentioned, indicating that the duration of the period of high sensitivity is also important in determining identifiability.
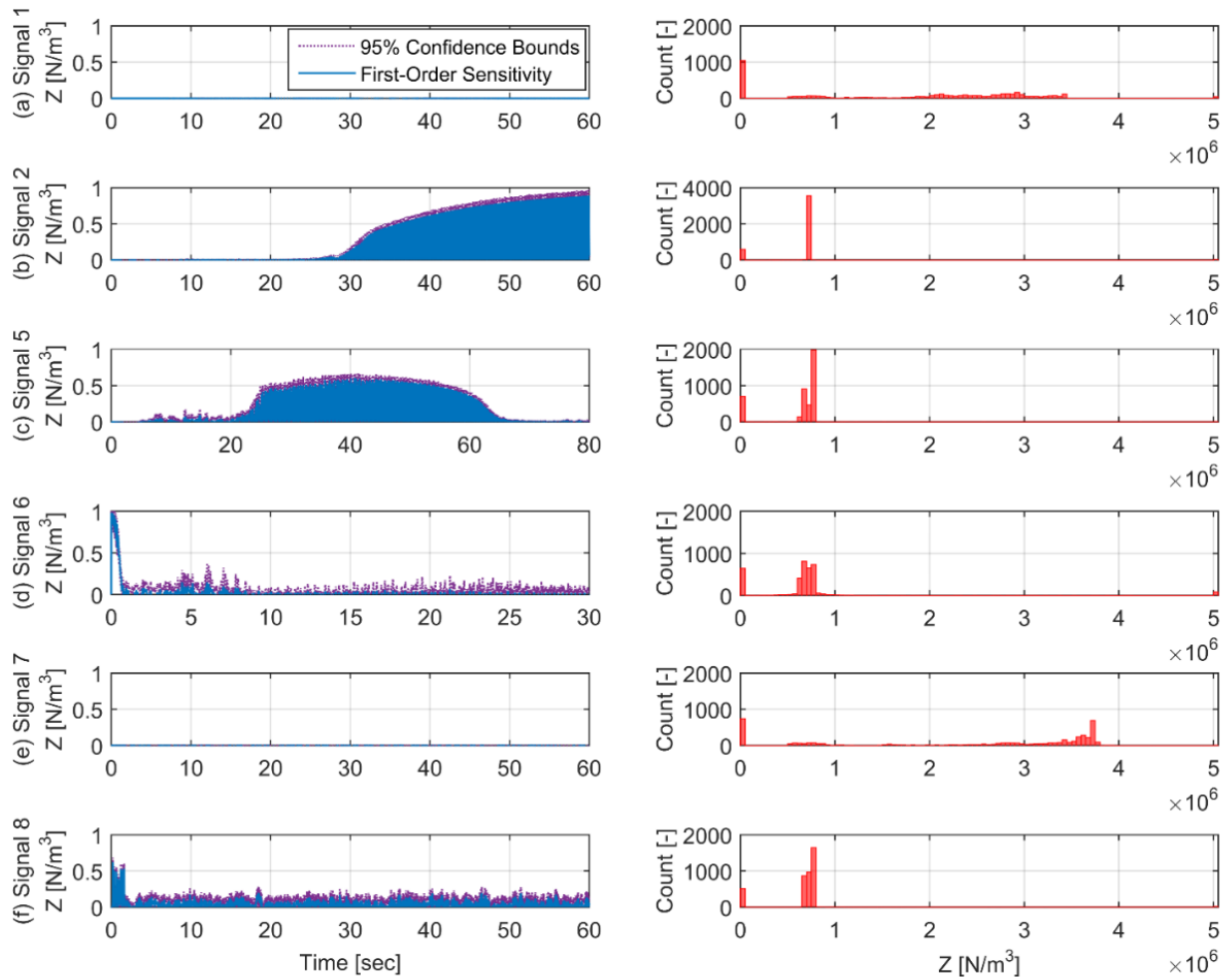
Fig.5.7 – Comparison of sensitivity and identified parameter distribution for the nonlinear stiffness parameter, z

A summary of the sensitivity and identification results for all parameters and experimental signals is given in Table 5.3. The sensitivity results in this table, in combination with those in Fig. 5.4-5.7, suggest that the selection of signals 2, 4, and 5 will lead to the most robust estimates of the NES device parameters. Among this set, signals 2 and 4 present similar information, though signal 4 includes the includes additional data concerning the decreasing amplitude behavior of the device over a longer time period. As such, the set of signals selected for the identification experiment is reduced to signals 4 and 5. A comparison of the displacement of the identified model and the experimental response signal used is shown for these signals in Fig.5.8 and Fig.5.9. It is interesting to note that even for those signals which have been determined to be identifiable, multiple high-quality models may be generated. This outcome does not mean that these signals are not identifiable, but instead indicates that the system response used for identification does not fully characterize the behavior of the device which the model was generated to capture. In essence, the identification with respect to a particular signal optimizes the model to replicate the behavior expressed in that signal, without regard to behavior that might be expressed in other signals. Thus, a model can be successfully identified, but that particular model may be insufficient to capture the behaviors of the experimental device that are exhibited in other untested signals. A method to complement this sensitivity analysis will be needed to distinguish between different models.

Table 5.3 – Results of the Sensitivity Analysis and the Experimental Identification

Signal			Maximum Sensitivity				Identified Value			
#	Type	Max Amp. [m/s^2]	C_v	C_f	K	Z	C_v [Ns/m]	C_f [N]	K [N/m]	Z [N/m^3]
1	Sweep	0.1	0.01	1.00	0.09	0.00	0.115	0.055	44.4	0.00×10^5
2	Sweep	9.4	0.27	1.00	0.42	0.89	0.330	0.074	53.6	7.12×10^5
3	Sweep	0.1	0.02	1.00	0.24	0.00	0.283	0.061	44.0	8.52×10^5
4	Sweep	9.4	0.27	1.00	0.43	0.89	0.330	0.050	19.8	7.12×10^5
5	Sine	4.9	0.09	1.00	0.36	0.61	0.351	0.064	37.8	6.39×10^5
6	Step	$x_0 = 5$ cm	0.10	0.11	0.12	0.98	0.452	0.049	26.0	6.67×10^5
7	BLWN	0.2	0.02	1.00	0.73	0.00	0.351	0.108	36.3	8.35×10^5
8	BLWN	18.2	0.22	1.00	0.34	0.62	0.313	0.067	49.5	6.20×10^5
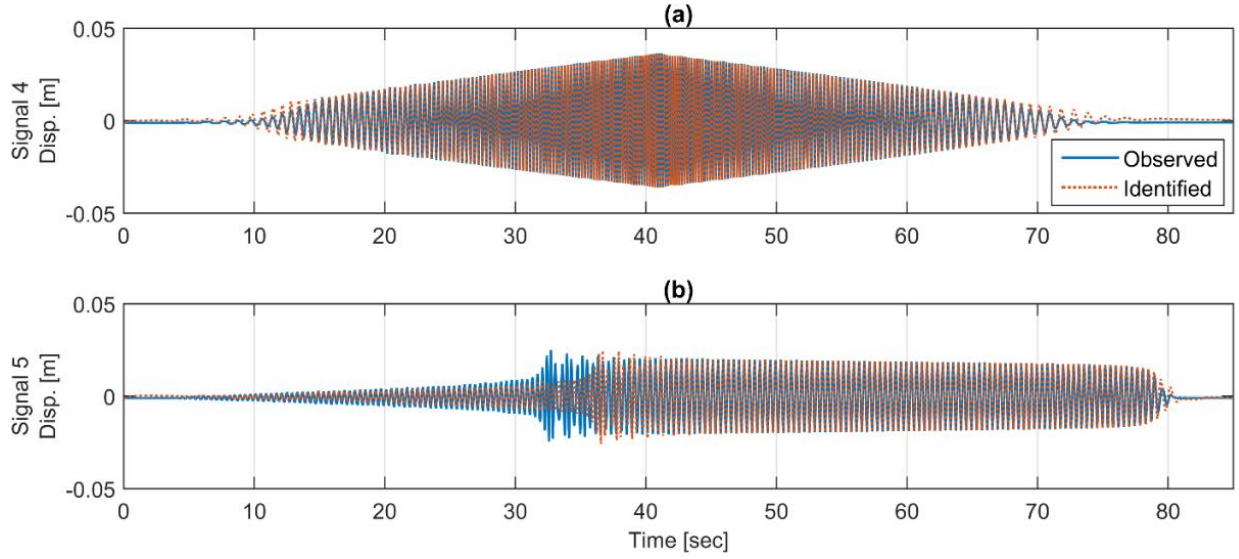
Fig.5.8 – Response comparison for model identified from Signal 4. The experimental response (observed) is plotted against the identified model response (identified) to the experimental excitation.
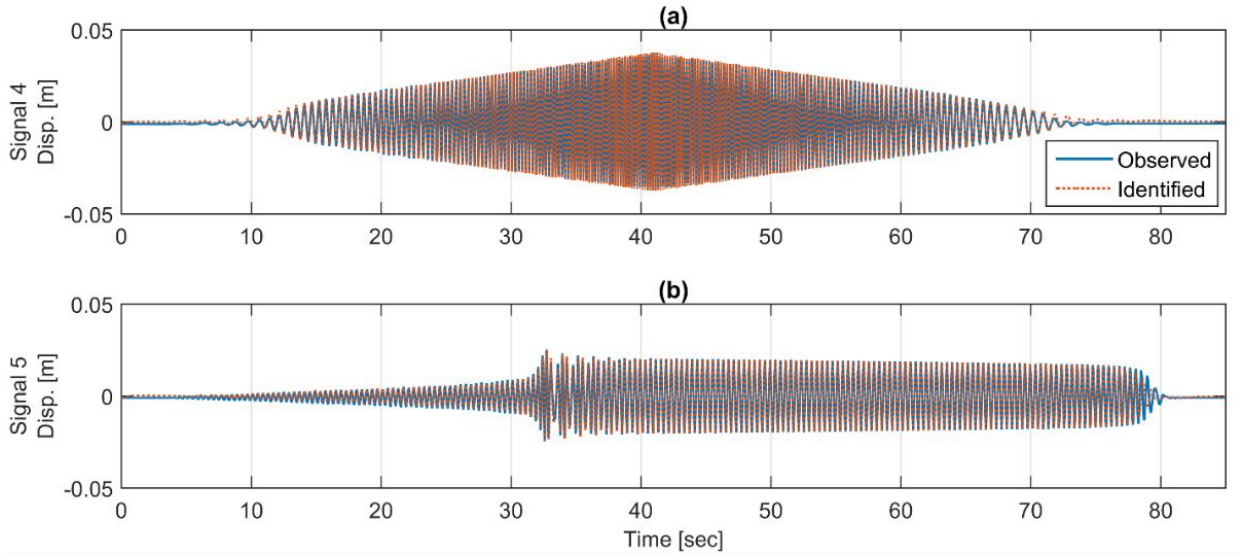
Fig.5.9 – Response comparison for model identified from Signal 5. The experimental response (observed) is plotted against the identified model response (identified) to the experimental excitation.

5.4 Conclusions

Sobol' sensitivity is examined as a practical indicator of identifiability and its effectiveness is demonstrated through the identification of an experimental NES device. The Sobol' sensitivity

results are shown to correlate well with the identification of the NES device, both over a single identification trial and across a suite of trials for a single signal. The results also indicate that both the magnitude of the sensitivity and the duration of high sensitivity regions influence the degree of identifiability. The approach is shown to give a unique and informative perspective on identifiability analysis in comparison with existing methods. Overall, Sobol' sensitivity analysis seems well-suited to informing the design of identification experiments for use with Bayesian inference techniques as it establishes a simple, heuristic measure of the practical identifiability of a general nonlinear system *in advance of the experiment*.

It is important to recognize the advantages of an analysis method while also respecting its limits. Sobol' sensitivity has the benefits of being model-independent, simple to execute, and useful for exploring the identifiability over the full range of variation of each model parameter. However, the computational requirements and error-free model assumption do suggest that practical implementation may be limited to systems for which the model form is reasonably predictable and the response of the model can be evaluated with moderate computational demands.

Although experimental design in terms of identifiability is a significant step in the design of system identification experiments, consideration must also be given to whether the signal selected for identification induces in the device the full range of dynamic behavior that needs to be captured for the purpose of modeling the response of the device to as yet unknown events. Further analysis of this topic is the focus of the following chapter.

5.5 References

1. Miao, H., Xia, X., Perelson, A. S. & Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **53,** 3–39 (2011).
2. Olivier, A. & Smyth, A. W. On the performance of online parameter estimation algorithms in systems with various identifiability properties. *Front. Built Environ.* **3,** 1–18 (2017).
3. Chatzis, M. N., Chatzi, E. N. & Smyth, A. W. On the observability and identifiability of nonlinear structural systems. *Struct. Control Heal. Monit.* **22,** 574–593 (2015).
4. Bellman, R. & Astrom, K. J. On structural identifiability. *Math. Biosci.* **7,** 329–339 (1970).
5. Ljung, L. & Glad, T. On global identifiability for arbitrary model parametrizations. *Automatica* **30,** 265–276 (1994).

6. Diop, S. & Fliess, M. Nonlinear observability, identifiability, and persistent trajectories. in *Proceedings of the 30th Conference on Decision and Control* 714–719 (1991).

7. Chis, O.-T., Banga, J. R. & Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS One* **6,** 27755 (2011).

8. Walter, E. & Lecourtier, Y. Global approaches to identifiability testing for linear and nonlinear state space models. *Math. Comput. Simul.* **24,** 472–482 (1982).

9. Pohjanpalo, H. System identifiability based on the power series expansion of the solution. *Math. Biosci.* **41,** 21–33 (1978).

10. Bellu, G., Saccomani, M. P., Audoly, S. & D'angì, L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **88,** 52–61 (2007).

11. Chis, O.-T., Banga, J. R. & Balsa-Canto, E. GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics* **27,** 2610–2611 (2011).

12. Meshkat, N., Er-zhen Kuo, C. & DiStefano III, J. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: A novel web implementation. *PLoS One* **9,** 110261 (2014).

13. Chis, O.-T., Villaverde, A. F., Banga, J. R. & Balsa-Canto, E. On the relationship between sloppiness and identifiability. *Math. Biosci.* **282,** 147–161 (2016).

14. Margaria, G., Riccomagno, E. & White, L. J. Structural identifiability analysis of some highly structured families of statespace models using differential algebra. *J. Math. Biol.* **49,** 433–454 (2004).

15. Kabanikhin, S. I., Voronov, D. A., Grodz, A. A. & Krivorotko, O. I. Identifiability of mathematical models in medical biology. *Russ. J. Genet. Appl. Res.* **6,** 838–844 (2016).

16. Grandjean, T. R. B., McGordon, A. & Jennings, P. A. Structural identifiability of equivalent circuit models for li-ion batteries. *Energies* **10,** 1–16 (2017).

17. Miao, H. *et al.* Modeling and estimation of kinetic parameters and replicative fitness of HIV-1 from flow-cytometry-based growth competition experiments. *Bull. Math. Biol.* **70,** 1749–1771 (2008).

18. Rodriguez-Fernandez, M., Mendes, P. & Banga, J. R. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems* **83,** 248–265 (2006).

19. Muto, M. & Beck, J. L. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *J. Vib. Control* **14,** 7–34 (2008).

20. Saltelli, A., Chan, K. & Scott, E. M. (Eds. . *Sensitivity Analysis*. (John Wiley & Sons, 2000).

21. Quaiser, T. & Mönnigmann, M. Systematic identifiability testing for unambiguous mechanistic modeling - Application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models. *BMC Syst. Biol.* **3,** 1–21 (2009).

22. Homma, T. & Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* **52,** 1–17 (1996).

23. Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55,** 271–280 (2001).

24. Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. & Schaibly, J. H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *J. Chem. Phys.* **59,** 3873–3878 (1973).

25. Cukier, R. I., Levine, H. B. & Shuler, K. E. Review nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.* **26,** 1–42 (1978).

26. Saltelli, A. *et al. Global Sensitivity Analysis. The Primer*. (John Wiley & Sons, 2008).

27. Saltelli, A., Tarantola, S. & Chan, K. P. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **41,** 39–56 (1999).

28. Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **145,** 280–297 (2002).

29. Lund, A., Dyke, S. J., Song, W. & Bilionis, I. Global sensitivity analysis for the design of nonlinear identification experiments. *Nonlinear Dyn.* **98,** 375–394 (2019).

30. Sobol', I. M. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* **1,** 407–414 (1993).

31. Sobol', I. M. *A Primer for the Monte Carlo Method*. (CRC Press, 1994).

32. Herman, J. & Usher, W. SaLib: An open-source Python library for sensitivity analysis. *J. Open Source Softw.* **2,** (2017).

33. Wu, M. & Smyth, A. W. Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Struct. Control Heal. Monit.* **14,** 971–990 (2007).

34. Wan, E. A. & Van Der Merwe, R. The unscented Kalman filter for nonlinear estimation. in *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium* 153–158 (IEEE, 2000).

35. Yuanxin Wu, Dewen Hu, Meiping Wu & Xiaoping Hu. Unscented Kalman filtering for additive noise case: augmented vs. non-augmented. *Proc. 2005, Am. Control Conf. 2005.* 4051–4055 (2005). doi:10.1109/ACC.2005.1470611

36. Sivia, D. S. & Skilling, J. *Data analysis: a Bayesian tutorial*. (Oxford University Press, 2006).

37. Worden, K. Data processing and experiment design for the restoring force surface methods, Part II : Choice of excitation signal. *Mech. Syst. Signal Process.* **4,** 321–344 (1990).

38. Sarkka, S. *Bayesian filtering and smoothing*. *Cambridge University Press* (Cambridge University Press, 2013). doi:10.1017/CBO9781139344203

6. DEVELOPMENT OF A ROBUST UNSCENTED KALMAN FILTER APROACH TO EXPERIMENTAL NONLINEAR IDENTIFICATION

As shown in the previous chapter, it is possible to determine in advance of experimentation the practical identifiability of a nonlinear dynamical model with respect to the behavior used for identification. However, for the identified model to be broadly applicable to predictive modeling tasks, it must be both robust, meaning that the same model is generated given reasonable variations on the prior, and extensible, meaning that the identified model should accurately represent the behavior of the system beyond the response it was trained against. The accuracy and robustness of the UKF to Bouc-Wen type nonlinearities has been shown in several experimental studies [1]–[5]. However, the application of the unscented Kalman filter (UKF) to experimental systems with other types of nonlinearities, especially those such as Duffing systems which exhibit bifurcations in their behavior, requires experimental validation prior to practical application.

In this study I develop and demonstrate a novel procedure for the experimental application of the UKF to a nonlinear energy sink (NES) device exhibiting a Duffing-type nonlinearity. Three identification approaches are proposed and analyzed with a focus on two typical concerns in experimental applications: the uncertainty in the selection of a prior distribution on the parameters, which is typically ill-defined in an experimental setting, and the assertion of a model form which is a useful, but imperfect, representation of the true device behavior. The robustness of the method to prior distribution selection is addressed in each case by evaluating the approach over a set of samples of the prior distribution. The issue of tuning an imperfect model to adequately represent the full range of device behavior is addressed by varying the manner by which each approach uses the experimental response signals during the identification process. By addressing these issues through the approaches developed herein, this study demonstrates a practical method for using limited response information to produce a device model which is well-suited to the responses of interest.

The paper is organized as follows: Section 6.1 examines the experimental behavior of the NES response signals used for identification; Section 6.2 introduces the approaches explored for the UKF identification; and Section 6.3 discusses the results of the three model identification approaches. The main conclusions are summarized in Section 0. The content of this chapter has been adapted from the published work of the author [6].

6.1 NES Behavior in Response to Identification Signals

Signals 9 and 10 from Table 4.1 are selected for model identification with the UKF as they provide a sufficient duration of excitation in the higher energy branch of the NES response such that all parameters of the NES model can be assumed to be fully engaged in the response. This selection is further supported by the Sobol' sensitivity analysis [7]–[9] on the device model with respect to these excitation signals given in the previous chapter. Of the experimental signals given in Table 4.1, signals 4 and 5, which are equivalent to signals 9 and 10, were found to be the most sensitive to variations in the NES model parameters and are therefore more likely to be identifiable. The remainder of the signals (Table 4.1, 11-16) are used to evaluate the accuracy of the identified model. The evaluation signals are selected to express the range of response likely to be seen by the device in both branches of its stable behavior. Where some of the evaluation signals are of the same type as those used for identification, their amplitude and frequency ranges are nominally modified to provide a more comprehensive basis for evaluation. All data from these tests, including measurements and recorded media, have been published by Lund et al. in [10].

The response of the NES device to the excitation signals used for identification is given in Fig. 6.1, with the response behavior to signal 9 given in Fig. 6.1(a-c) and the response behavior to signal 10 given in Fig. 6.1(d-f). The behavior of the device is shown from three perspectives: Fig. 6.1(a,d) show the displacement response histories; Fig. 6.1(b,e) show the displacement spectrograms; and Fig. 6.1(c,f) show the restoring force vs. displacement curves.

The device exhibits significant nonlinear behavior in each case, which is evident through different features in each perspective on the response. A shift in the stable response of the NES to a higher energy state is clearly evident in Fig. 6.1(c) at ~32 sec through the sudden variability in oscillation frequency and amplitude separating the two distinct regions of response. While not as obvious, a similar change in stable behavior is apparent in Fig. 6.1(a) at ~15 second by a shift in the slope of the response envelope.

These shifts in the response of the device are more clearly understood through their respective spectrograms. In each case, a shift in the time-domain behavior is accompanied by a spreading of energy from the 1st to the 3rd harmonic of the forcing frequency, where prior to the shift only the 1st harmonic had been significantly excited. It is particularly interesting to note that despite the symmetric variation in either the excitation frequency (signal 9) or amplitude (signal 10) about the ~40 sec mark, the device responses to these excitations are asymmetric. Once the

NES has achieved a stable high-energy response it is able to maintain that response despite a reduction in the input energy. This phenomenon is most clearly evident with respect to signal 9, but can be seen in signal 10 by comparing the responses at ~15 and ~70 sec.
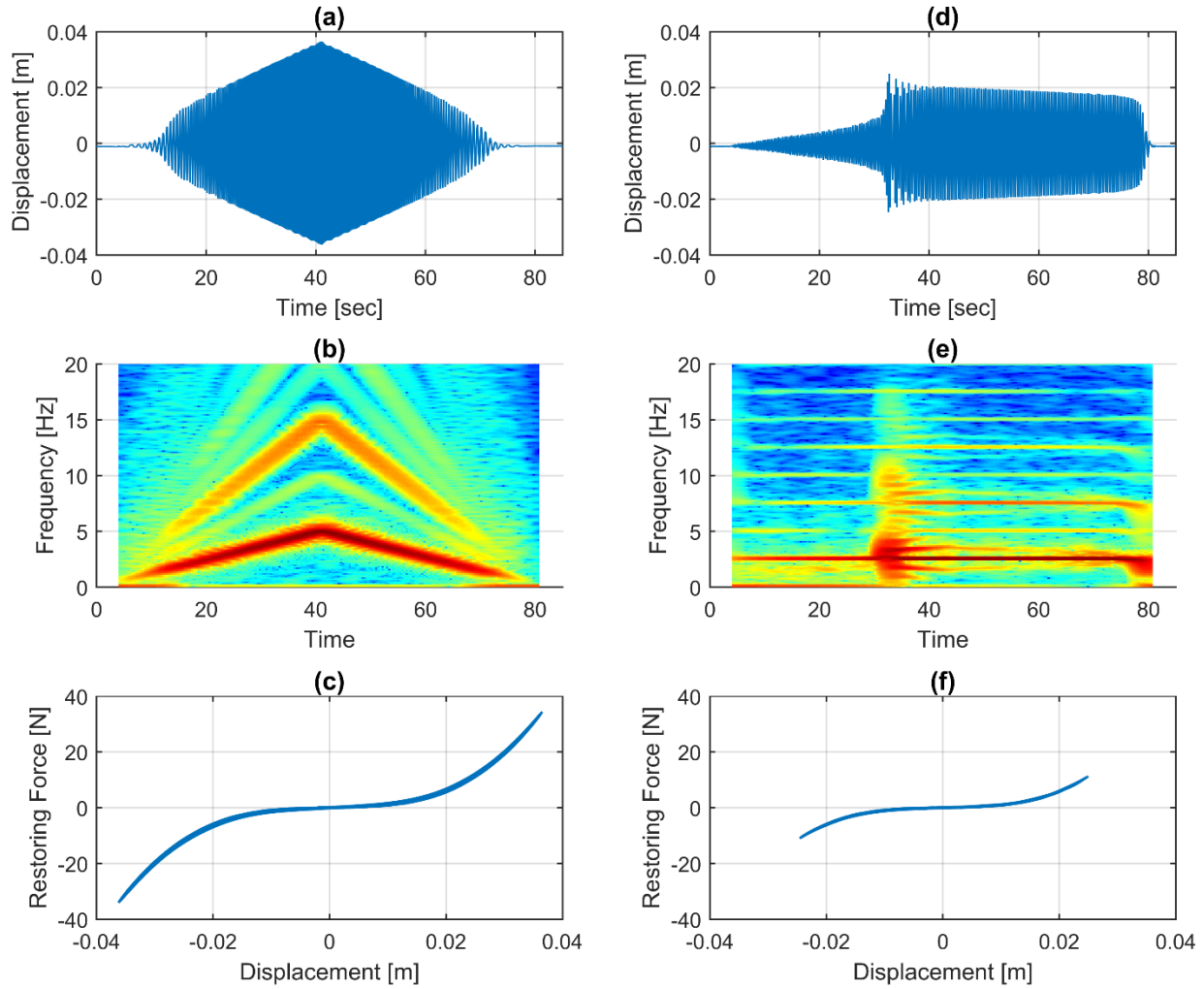
Fig. 6.1 – NES responses including displacement history, spectrogram, and restoring force vs. displacement respectively for signal 9 (a-c) and signal 10 (d-f). Acceleration has been low-pass filtered to 50 Hz for (c,f).

The results in Fig. 6.1(c,f) show that in each case the device is excited sufficiently to exhibit nonlinear restoring force behavior. The results confirm the expected dual stability of the NES response and the form of the nonlinearity governing its behavior. As a result, they also provide evidence of the engagement of all parameters in Eq. (4.5) for the purpose of identification.

The practical identifiability of the NES model with respect to signals 9 and 10 has further been established using Sobol' sensitivity analysis in the previous chapter, which demonstrated that the parameters of the model given by Eq. (4.5) exhibit high sensitivity to the respective identification signals. In particular, the parameters c_v, c_f, k, and z exhibited sensitivities of 0.27, 1.00, 0.43, and 0.89 respectively in response to excitation by signal 9 and 0.09, 1.00, 0.36, and 0.61 respectively in response to excitation by signal 10, where sensitivities closer to 1.00 indicate a higher likelihood of practical identifiability. The sensitivity results for these excitations were found to support the use of either signal for identification, though the parameters are more sensitive to, and therefore more likely to be consistently identified by, signal 9.

6.2 Analysis Approach

As in the previous chapter, the identification of the NES device is accomplished using the UKF. The approach to implement the UKF using the noise-augmented sigma points is equivalent to Chpt. 5, including the tuning parameters used. For clarity, the specific stochastic model implementation is redefined in this chapter.

6.2.1 UKF Implementation

Identification of the NES device using the UKF is accomplished through a state augmentation approach, which is also commonly referred to as joint state and parameter estimation [11]. As was done in the previous chapter, prior information concerning the system characteristics can be leveraged to enhance the state model for UKF implementation [12]. In this case, the relative scale of the parameters and their existence in the positive domain is known *a priori*. This prior information is used to define the augmented state vector

$$\mathbf{x} = \left[x, \dot{x}, \log\left(\frac{c_v}{0.1}\right), \log\left(\frac{c_f}{0.1}\right), \log\left(\frac{k}{10}\right), \log\left(\frac{z}{10^5}\right) \right]^T,$$

$$= [x_1, x_2, x_3, x_4, x_5, x_6]^T, \tag{6.1}$$

with the experimental responses and model parameters recorded in SI units. The augmented state is designed for joint state and parameter estimation with the UKF such that the states are at similar orders of magnitude and their associated parameters are constrained to identification in the positive domain. Note that the scale factors are dimensionless values selected to bring the parameter values

to a similar order of magnitude. The introduction of such a transformation provides a means to perform an unconstrained optimization on the parameters in a way that aligns with the base assumptions of the UKF, and therefore introduces no additional instabilities into the identification process. Though the transformation may appear to mask the form of the prior distribution on the true parameters (c_v Ns/m, c_f N, k N/m, and z N/m³), the assignment of these priors can easily be made in terms of a log-normal distribution, with the shape parameters then corresponding to the mean and variance of the log-transformed states.

The state transition function, $f(\cdot)$, can therefore be defined as

$$f\big(\mathbf{x}_{k-1}, \ddot{x}_{g,k-1}\big) = \mathbf{x}_{k-1} + \int_{t_{k-1}}^{t_k} F\big(\mathbf{x}(t), \ddot{x}_g(t)\big)\, dt + \mathbf{w}_k, \tag{6.2}$$

where

$$F\big(\mathbf{x}(t), \ddot{x}_g(t)\big) = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \end{bmatrix}$$

$$= \begin{bmatrix} x_2 \\ -\ddot{x}_g - (0.1e^{x_3}x_2 + 0.1e^{x_4}\tanh(vx_2) + 10e^{x_5}x_1 + 10^5 e^{x_6}x_1{}^3)/m \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6.3}$$

Similarly, the observation equation, $h(\cdot)$, is defined as

$$h\big(\mathbf{x}_k, \ddot{x}_g\big) = \left[-\frac{x_1}{(0.1e^{x_3}x_2 + 0.1e^{x_4}\tanh(vx_2) + 10e^{x_5}x_1 + 10^5 e^{x_6}x_1{}^3)/m} \right] + \mathbf{v}_k, \tag{6.4}$$

which implies the simultaneous observation of the relative displacement, x, and the absolute acceleration, $(\ddot{x} + \ddot{x}_g)$, of the experimental responses of the NES device. Base acceleration, $\ddot{x}_g$, is also observed from the experimental responses of the system.

The corresponding process noise ($\mathbf{Q}$) and measurement noise ($\mathbf{R}$) covariance matrices, which are taken as constant for this system, are obtained experimentally through a 60 sec observation of the NES system when a zero command voltage is sent to the hydraulic actuator. The process noise for this system is assumed to occur primarily due to noise in the base acceleration signal. Therefore, $\mathbf{Q}$ is a 6×6 sparse matrix in which the only non-zero entry is

$Q_{22} = \sigma_b^2 = 1.03 \times 10^{-2} (\text{m/s}^2)^2$, corresponding to the variance of the acceleration sensor over the observation period previously described. Likewise, **R** is a 2×2 diagonal matrix described by the sensor noise in the displacement and acceleration responses of the NES device, which have variances $\sigma_d^2 = 1.44 \times 10^{-12} \text{m}^2$ and $\sigma_a^2 = 5.32 \times 10^{-3} (\text{m/s}^2)^2$, respectively. The measurement noises in the various sensor measurements are assumed to be independent.

6.2.2   Identification Methods

Though the UKF is theoretically tractable for nonlinear systems, experimental deviations from the selected model form and the Gaussian noise assumption can make implementation of the algorithm quite challenging, leading to issues in the robust identification of a high-quality model. These issues commonly manifest in the identified model becoming overly tuned to the system response used for identification, without regard for other potential system behavior. This phenomenon is referred to as over-fitting.

Many of the issues associated with over-fitting can be addressed with an appropriate model and a sufficiently exciting input. Such a model should be selected such that the key physics of the system are captured and any extraneous terms are eliminated, which can be done prior to identification through methods such as sensitivity analysis [7], or during identification through Bayesian model selection techniques [13],[14] or automatic relevance determination methods [15]. In order for the parameters to be accurately identified from the resulting model form, the response signal used for identification must be sufficiently exciting such that all terms are active in the response. It has been shown in Section 6.1 that the NES model proposed in Eq. (4.5) meets these qualifications.

Even when these key concerns have been addressed, models that appear overly tuned to the responses used for identification can still occur in the identification of experimental systems. Over-fitting occurs in this case due to the combination of a potentially simplified system model and inadequate or overconfident modeling of the noise present in the experimental response. Even a useful, high-quality model of the system will not perfectly represent its experimental behavior. These imperfections, coupled with any inaccuracies in the estimation of the form and magnitude of the noise present in the system, are evidenced in the tuning of the identified model slightly toward the response used for identification instead of the general system behavior. This problem is exacerbated in systems with a clear bifurcation in their response, such as the NES device, as

slight shifts in the identified model likewise shift the regions of dual stability in the model away from those evidenced in the experimental device behavior. A reliable method for the experimental identification of the NES device must therefore address this issue.

I approach this second over-fitting problem from the standpoint of algorithm deployment, reasoning that if the UKF is implemented strategically to take advantage of the information available in the various device responses, a model which observes the true range of device behavior can be reliably identified. As such, three approaches to identify the NES device with the UKF algorithm are proposed and analyzed, as shown in Fig. 6.2. Each of these approaches consists of two phases: training and validation.
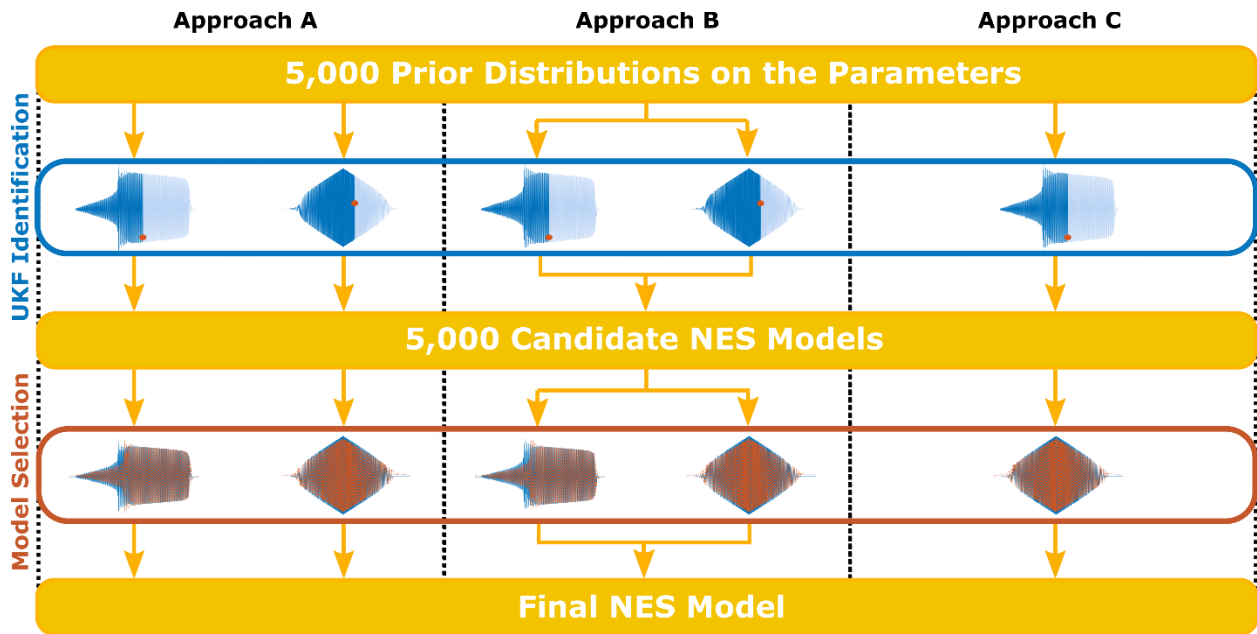
Fig. 6.2 – Training approaches for NES identification

The training phase, labeled *UKF Identification* in Fig. 6.2, is initiated by the selection of a set of prior distributions on the parameters. Models generated using the UKF algorithm are sensitive to the selection of the prior distribution, which can be difficult to determine in an experimental setting for states related to the parameters. To compensate for this uncertainty and to analyze the impact of the prior distribution on the UKF, multiple identification trials are performed with each of the three approaches. The trials are structured around the prior distribution on the state vector, which is characterized by

$$\boldsymbol{\mu}_0 = \left[\boldsymbol{\mu}_0^{(x)} \ \boldsymbol{\mu}_0^{(\theta)}\right]^T, \tag{6.5}$$

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{P}_0^{(x)} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0^{(\theta)} \end{bmatrix}, \tag{6.6}$$

where $\boldsymbol{\mu}_0$ is partitioned such that $\boldsymbol{\mu}_0^{(x)}$ is representative of the initial mean values belonging to the dynamic states of the system and $\boldsymbol{\mu}_0^{(\theta)}$ is representative of those belonging to the parameters. $\mathbf{P}_0$ is likewise partitioned. All states are assumed to be independent, which implies that $\mathbf{P}_0$ is a diagonal matrix with entries $(\mathbf{P}_0)_{nn} = \sigma_n^2$ for $n = 1,2,\dots,6$. Because all observations of the system begin with the NES at rest, it is assumed that $\boldsymbol{\mu}_0^{(x)} = \mathbf{0}$, $\sigma_1 = 0.01$m, and $\sigma_2 = 0.01$ m/s. Variability in the prior distribution on the parameters is achieved through a Latin hypercube sample of 500 vectors of initial means, $\boldsymbol{\mu}_0^{(\theta)}$, from the parameter space $c_v \in [0,1]$ Ns/m, $c_f \in [0,1]$ N, $k \in [0,100]$ N/m, and $z \in [5 \times 10^5, 9 \times 10^5]$ N/m³. A set of 10 initial covariance samples, $\mathbf{P}_0^{(\theta)}$, is associated with each sample of $\boldsymbol{\mu}_0^{(\theta)}$. These covariance samples are defined such that $\sigma_n = \alpha(\mu_0)_n$ ($n = 3,4,5,6$), where $\alpha \in \{0.1, 1, 10, 25, 50, 75, 100, 200, 300, 600\}\%$. This yields a total of 5,000 identification trials to examine each approach. The same trials are used for all three approaches.

The validation phase, labeled *Model Selection* in Fig. 6.2, consists of selecting a representative model from the 5,000 candidate models generated through UKF identification. The relative quality of the candidate models is assessed using the normalized mean square error indicator [16] given by

$$\text{MSE} = \frac{100}{N} \sum_{i=1}^{N} \left(\frac{(x_i - \hat{x}_i)^2}{\sigma_d^2} + \frac{(\ddot{x}_i - \hat{\ddot{x}}_i)^2}{\sigma_a^2} \right), \tag{6.7}$$

where N is the total number of experimental observations of the system, x_i is the ith observed displacement in time for a particular experimental excitation, and $\hat{x}_i$ is the displacement value obtained by simulating the candidate model's response to that same excitation. The final NES model for each approach is selected as the one which minimizes the MSE with respect to the experimental response of the NES.

The identification approaches vary in how identification signals 9 and 10 (Table 4.1) are used in each of these phases, as shown in Fig. 6.2. In approach A, independent identification trials

are run with the UKF using the device responses to signals 9 and 10, respectively. The candidate models are then validated against the response that was used for training. In approach B, signals 9 and 10 are used simultaneously with the UKF to train a model of the NES device. To accomplish this, the state vector for the UKF identification is modified such that

$$\mathbf{x} = \left[x^{(1)}, x^{(2)}, \dot{x}^{(1)}, \dot{x}^{(2)}, \log\left(\frac{c_v}{0.1}\right), \log\left(\frac{c_f}{0.1}\right), \log\left(\frac{k}{10}\right), \log\left(\frac{z}{10^5}\right) \right]^T, \qquad (6.8)$$

where $x^{(1)}$ is the displacement of the NES in response to signal 9, and $x^{(2)}$ is the displacement of the NES in response to signal 10. The base function for the state transition, $F(\cdot)$, is then redefined as

$$F\left(\mathbf{x}(t), \ddot{x}_g(t)\right) = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{x}_7 \\ \dot{x}_8 \end{bmatrix}$$

$$= \begin{bmatrix} x_2^{(1)} \\ x_2^{(2)} \\ -\ddot{x}_g^{(1)} - \left(0.1 e^{x_3} x_2^{(1)} + 0.1 e^{x_4} \tanh\left(v x_2^{(1)}\right) + 10 e^{x_5} x_1^{(1)} + 10^5 e^{x_6} x_1^{(1)^3}\right)/m \\ -\ddot{x}_g^{(2)} - \left(0.1 e^{x_3} x_2^{(2)} + 0.1 e^{x_4} \tanh\left(v x_2^{(2)}\right) + 10 e^{x_5} x_1^{(2)} + 10^5 e^{x_6} x_1^{(2)^3}\right)/m \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (6.9)$$

and the observation equation, $h(\cdot)$, is defined as

$$h\left(\mathbf{x}_k, \ddot{x}_g\right) =$$

$$\begin{bmatrix} x_1^{(1)} \\ x_1^{(2)} \\ -\left(0.1 e^{x_3} x_2^{(1)} + 0.1 e^{x_4} \tanh\left(v x_2^{(1)}\right) + 10 e^{x_5} x_1^{(1)} + 10^5 e^{x_6} x_1^{(1)^3}\right)/m \\ -\left(0.1 e^{x_3} x_2^{(2)} + 0.1 e^{x_4} \tanh\left(v x_2^{(2)}\right) + 10 e^{x_5} x_1^{(2)} + 10^5 e^{x_6} x_1^{(2)^3}\right)/m \end{bmatrix} + \mathbf{v}_k, \qquad (6.10)$$

implying that the experimental input and response records for both identification signals will be used simultaneously to identify a single set of parameters. The final model is likewise selected

through validation with respect to both signals. In approach C, the state vector is again defined as in Eq. (6.3) and the candidate models are generated by training with respect to signal 10. The candidate models are then validated, and a final model selected, with respect to signal 9. This method employs a cross-validation approach as opposed to the self-validation employed in the other methods.

The accuracy of the results obtained with the three identification approaches are compared by analyzing the ability of each approach to generate a final model which minimizes the MSE for the *evaluation signals* (signals 11-16, Table 4.1). In interpreting these results, it is important to remember that the MSE is equivalent to the log-likelihood of the measurements, given information about the states and the parameters $(\log(p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta})))$. The MSE provides a measure of the error with respect to the known sensor noise. It is therefore expected that the MSE evaluated for these models will be large due to the small values of σ_d^2 and σ_a^2.

6.3   Results and Discussion

The three identification approaches described in Fig. 6.2 are studied herein. The results of each approach are first presented in terms of the distribution of the candidate models, with the purpose of identifying how consistently the minimum MSE solution is identified. The selected NES models from each approach are further tested against the evaluation signals given in Table 4.1 to quantify their accuracy with respect to the MSE formulation given in Eq. (6.7).

6.3.1   Models Generated Separately Using Different Training Signals (Approach A)

The first stage of the analysis focuses on approach A, in which the training and validation phases are run separately for each identification signal. The distributions of the candidate models generated from the 5,000 identification trials using each of these response signals are provided in Fig. 6.3. It should be noted that some of these identification trials failed due to a numerical instability and are therefore not included in the analysis. Though instability in the UKF can be observed when only the dynamic states of the system are estimated, the potential for instability in its execution increases when the parameters are modeled as constant states, as is typically assumed in state augmentation for joint state and parameter identification [17]. This modeling assumption can generate numerical errors in the algorithm such that the covariance matrix $\mathbf{P}$ becomes singular.

These unstable cases appear during the identification process and corresponded to 15-20% of the total identification trials, depending on the signal used for training. All other identification trials exhibited stable mean values in the parameters throughout the final 25 sec of the identification signal and reduced in variance from the assumed prior distribution, with the majority of identification trials exhibiting a significant reduction in variance suggestive of full convergence. The trials experiencing only moderate reductions in variance, suggestive of partial convergence, are also included in the following analysis. The stability of the parameter estimates in those trials suggests that, should the identification progress using experimental responses at a similar level of information, the fully converged result would be in the vicinity of the current mean. As the scope of this analysis encompasses identification using limited response data, these partially converged results are included as estimates of the final convergence.
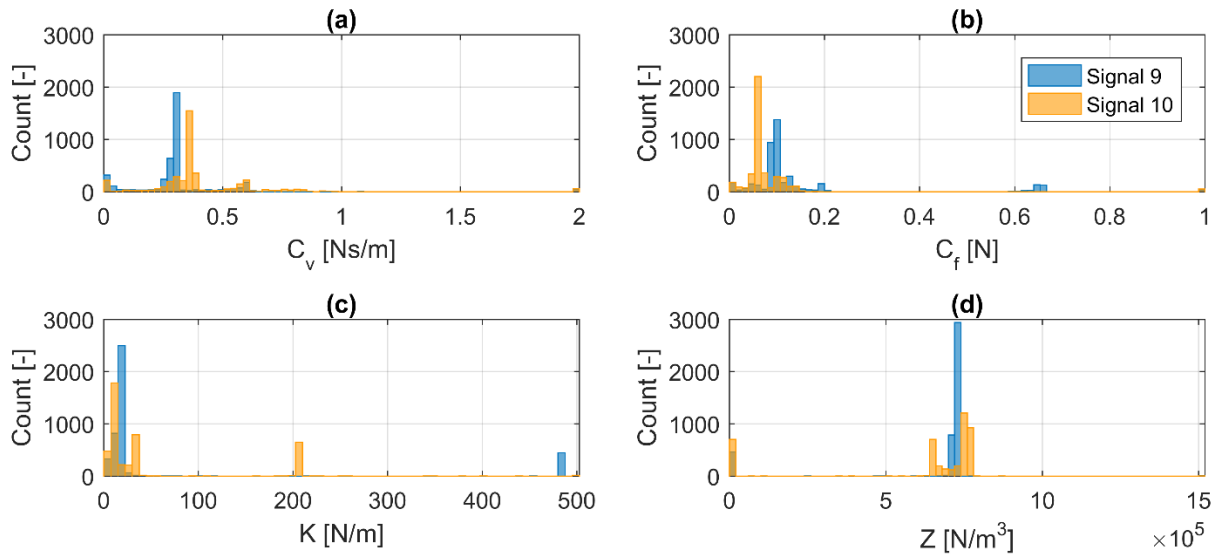
Fig. 6.3 – Distribution of models generated using identification approach A.

The results in Fig. 6.3 clearly show that working with each identification signal separately produces different interpretations of the NES model. Training and validating with signal 9 results in tight distributions for the parameters, with the only significant variation in the models occurring in the interpretation of the linear stiffness parameter, k. In contrast, training and validating with signal 10 generates multi-modal distributions for the nonlinear stiffness parameter, z, and for the linear stiffness parameter, k. Though both signals generate tight unimodal distributions for the

damping parameters, each of those distributions are offset from each other such that there is little overlap between the two models around their respective modes. The differences between the models identified with these two signals is a direct result of the variation in the information available in the training signals. However, the fact that these differences extend even to parameters which are identified with tight distributions indicates that the models generated from this approach may not be reliable.

At first glance, it appears that using signal 9 to generate the final NES model should produce a more accurate and robust result. To test this hypothesis, the final identified models with respect to each training signal are used to simulate the response of the NES to the experimental measurements of the evaluation signals given in Table 4.1. The comparison between the observed and simulated responses are shown in Fig. 6.4, where Fig. 6.4 (a-f) are generated with the model corresponding to signal 9, and Fig. 6.4 (g-l) are generated with the model corresponding to signal 10. The results show that although both models are able to capture some aspects of the device's behavior, particularly in the evaluation signals which are similar to identification signals, neither model is able to fit the entire range of NES responses. These observations, coupled with the goodness of fit of the identified models to the training signals (see Table 6.1), suggest that the UKF is over-fitting the models to the training signal. This behavior is especially evident for the model generated with signal 9. Despite the fact that this model is consistently identified with respect to signal 1, its ability to replicate the behavior expressed under other excitations is limited. This situation is also true for the model generated from signal 10. Though this model is able to more accurately replicate the behavior expressed in signals 11-15 than the model in Fig. 6.4(a-f), it fails to capture the behavior expressed in signal 16 (see Fig. 6.4(l)).

Table 6.1 provides results for evaluating this identification approach quantitatively. For each signal, three models are studied. *MSE model* refers to final identified model for each signal, as described in Section 6.2.2 and shown in Fig. 6.4. *Modal model* refers to a representative model selected from the mode of the candidate models, where the mode is defined over the parameters such that the ranges of c_v, c_f, k, and z in the candidate models are partitioned into increments of 0.04 Ns/m, 0.02 N, 10 N/m, and 10,000 N/m^3, respectively. This model is used give an indication of the most likely result of the identification approach should only a single prior distribution be selected at random from the set described in Section 6.2.2. *Sub-modal model* refers to a subset of the Modal models containing the minimum MSE model for each mean over the 10 initial variance
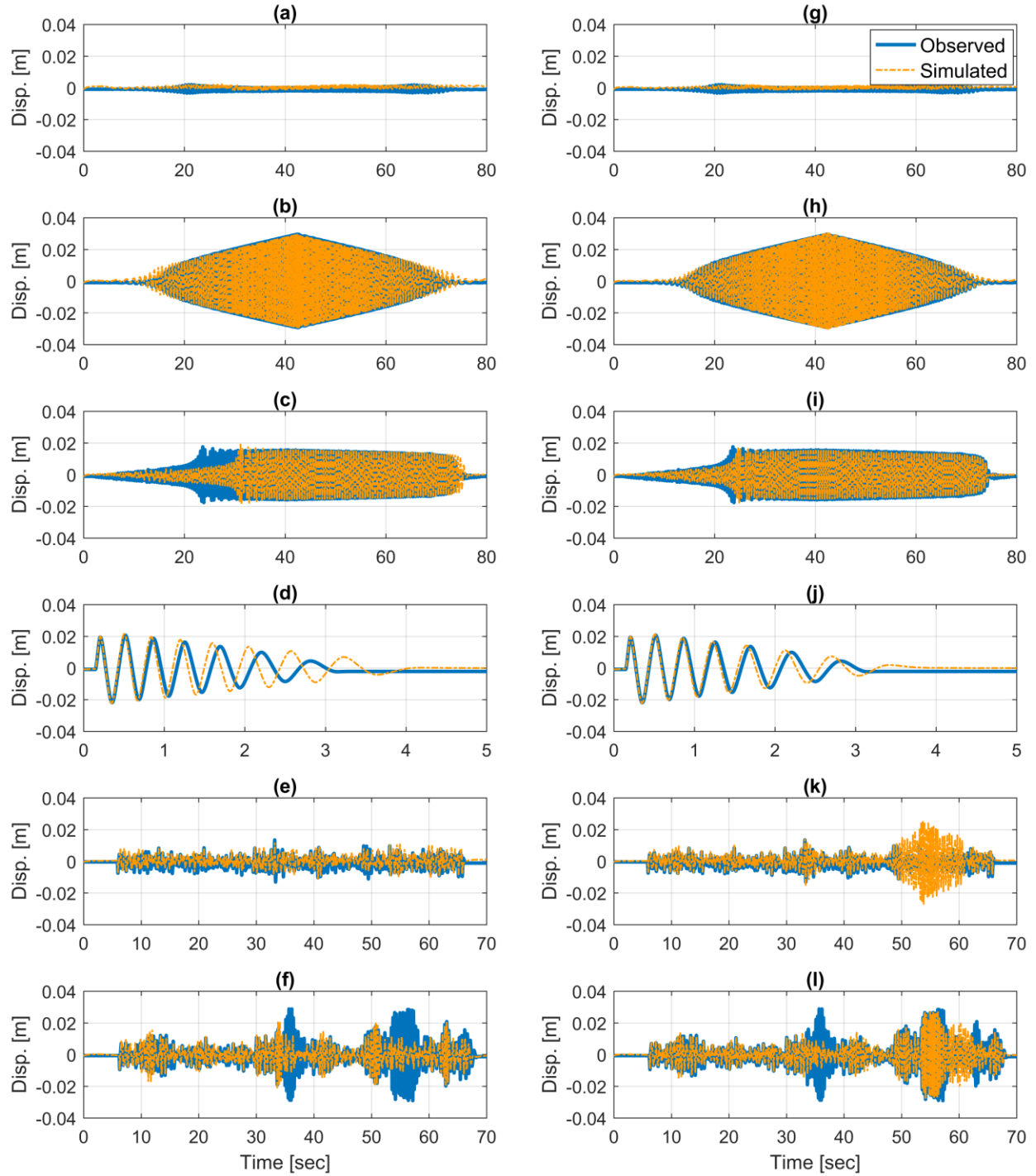
Fig. 6.4 – Response comparison for the models identified using approach A. Subplots (a-f) compare the simulated device responses from the model generated with signal 9 to the experimental device responses to signals 11-16. Subplots (g-l) compare the simulated device responses from the model generated with signal 10 to these same experimental device responses.

values, resulting in a set of 500 candidate models. The mode of these models is observed as it is representative of the MSE model that is most likely to be selected if the same identification process were to be performed on a smaller set of prior means, which would be desirable for a more rapid implementation of the method.

Table 6.1 – Evaluation results for models identified using identification approach A. The MSE for the selected models are given with respect to the validation signal, as well as with respect to the evaluation signals.

		Identified Models					MSE/10^8 for Evaluation Signals					
#	Selected Model	c_v $\left[\dfrac{Ns}{m}\right]$	c_f [N]	k $\left[\dfrac{Ns}{m}\right]$	$\dfrac{z}{10^5}$ $\left[\dfrac{N}{m^3}\right]$	$\dfrac{MSE}{10^8}$ [-]	11	12	13	14	15	16
1	MSE	0.330	0.050	19.8	7.12	1.3	1.8	1.3	15.7	45.0	4.4	33.8
	Modal	0.297	0.094	21.3	7.26	147.6	1.6	101.5	16.7	6.1	3.8	33.4
	Sub-Modal	0	0.650	484.2	0	132.9	1.2	75.6	42.2	52.3	13.9	54.6
2	MSE	0.351	0.064	37.8	6.39	1.6	1.2	0.6	4.6	12.4	18.7	28.0
	Modal	0.375	0.060	9.6	7.59	22.7	2.8	102.8	20.6	7.8	6.6	66.9
	Sub-Modal	0.290	0.104	32.6	6.55	9.4	1.3	102.3	12.4	8.6	1.8	57.4

The results in Table 6.1 give us some insight into the robustness of the identification process with respect to the signals used. In each case, some differences exist in the selected models. The MSE and Modal models identified using signal 9 are fairly consistent, but in the Sub-modal model the viscous damping, c_v, and nonlinear stiffness, z, are entirely eliminated. It is interesting to note that despite the minor variations between the MSE and Modal models, the Modal model shows a large increase in the MSE value generated from the validation step, as well as in the case where the MSE is evaluated against signal 12, which is similar in form to signal 9. The particular form this error takes is a simulated low amplitude response similar to the experimental response to signal 11. This outcome indicates that small variations in model parameters can significantly shift the input energy threshold of the NES model. The MSE and Modal models identified with signal 10 differ primarily with respect to the nonlinear stiffness parameter, z. However, both the Modal and Sub-modal models produce very poor characterizations of the NES behavior in comparison with the MSE model, particularly with respect to signal 12. Again, this result indicates that the differences between these models and the MSE model increase the input energy threshold

118

of the device, yielding a low amplitude displacement for a case where high amplitude displacements were experimentally observed.

The analysis of these models and their response to the evaluation signals demonstrates that the identified results are not as robust as they first appear. Not only do the models appear to over-fit to the signal they are trained with, but variations between the MSE and mode-based models show stark differences in error when evaluated against the signals 11-16. Due to the uncertainties involved in experimental model identification, a more consistent and accurate identification approach is required to enable a reduced number of identification trials while still providing a reasonable guarantee on the accuracy of the results.

6.3.2   Models Generated Using Training Signals in Parallel (Approach B)

The second stage of the analysis focuses on identification approach B, in which the identified NES model is generated simultaneously using signals 9 and 10. This approach is facilitated by adopting the state vector given in Eq. (6.8) for use with the UKF. The models generated with this process are shown in Fig. 6.5 for all 5,000 identification trials. This identification approach generates a tight distribution of candidate models, with mean values similar to those generated in approach A with signal 9. These distributions do guarantee a certain level of similarity between the selected MSE model and the models generated from the modes of the identified parameter sets, which can be seen in Table 6.2. However, the data also show that there are large differences in the degree to which these seemingly similar models are able to represent the experimental NES responses, indicating a lack of robustness in the MSE model generated for this case.

The results from identification approach A in Section 6.3.1 suggest that the identified model tends to over-fit to the behavior in the training signal, particularly when that training signal is highly sensitive to variations in the parameters, as is the case with signal 9 (see Section 5.3). It is therefore expected that training with two signals, both for which the parameters have exhibited moderate to high levels of sensitivity, will generate a model which accommodates the behavior present in both training signals. Such a model would then minimize the error for both evaluation signal 12, which has the same form as signal 9, and evaluation signal 13, which has the same form as signal 10. It is evident from Fig. 6.6(b,c) that this is not the case. Rather, the results in Fig. 6.6 (b,c), and indeed the MSE model itself, appear very similar to the identification results using only signal 9, as given in Fig. 6.4(a-f) and Table 6.1. The Modal and Sub-modal models likewise obtain
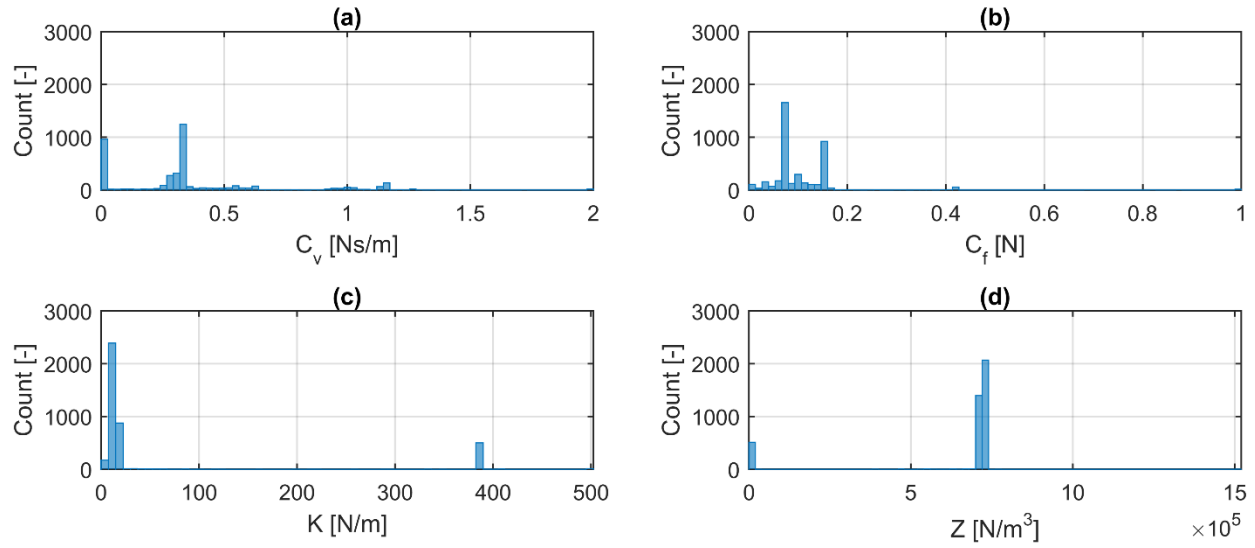
Fig. 6.5 - Distribution of models generated using identification approach B.

similar results as training with only signal 9, particularly in MSE values resulting from the validation step. These results imply that simultaneously performing the identification with respect to two training signals tunes the parameters toward the signal to which the parameters exhibit the highest sensitivity, which in this case is signal 9.

Table 6.2 – Evaluation results for models identified using identification approach B. The MSE for the selected models are given with respect to the validation signal, as well as with respect to the evaluation signals.

Identified Models						MSE/10^8 for Evaluation Signals					
Selected Model	c_v $\left[\frac{Ns}{m}\right]$	c_f [N]	k $\left[\frac{Ns}{m}\right]$	$\frac{z}{10^5}$ $\left[\frac{N}{m^3}\right]$	$\frac{MSE}{10^8}$ [-]	11	12	13	14	15	16
MSE	0.337	0.053	20.7	7.09	14.0	1.7	1.1	14.5	36.1	3.9	64.3
Modal	0.324	0.074	14.7	7.22	148.4	2.1	102.3	19.3	5.5	5.4	36.7
Sub-Modal	0.283	0.102	18.1	7.13	168.2	1.7	100.6	20.4	10.5	4.6	35.4

These observations can be understood by considering the square error loss function of the Kalman filter. At each time step, the Kalman gain is determined such that the square error between the predicted states and the true states will be minimized, where observations of the device serve as an indicator of the true state. Variations of the parameters in the augmented UKF state vector

120

generate model responses with a wider range of error values for device responses which are highly sensitive to parameter variations than for those which exhibit a lower sensitivity. The UKF can therefore most effectively minimize the mean square error by minimizing the error in response to those highly sensitive signals. In this analysis, this means that the information for training the NES model available in signal 10 is overshadowed by the information in signal 9, which is more sensitive to parameter variations.
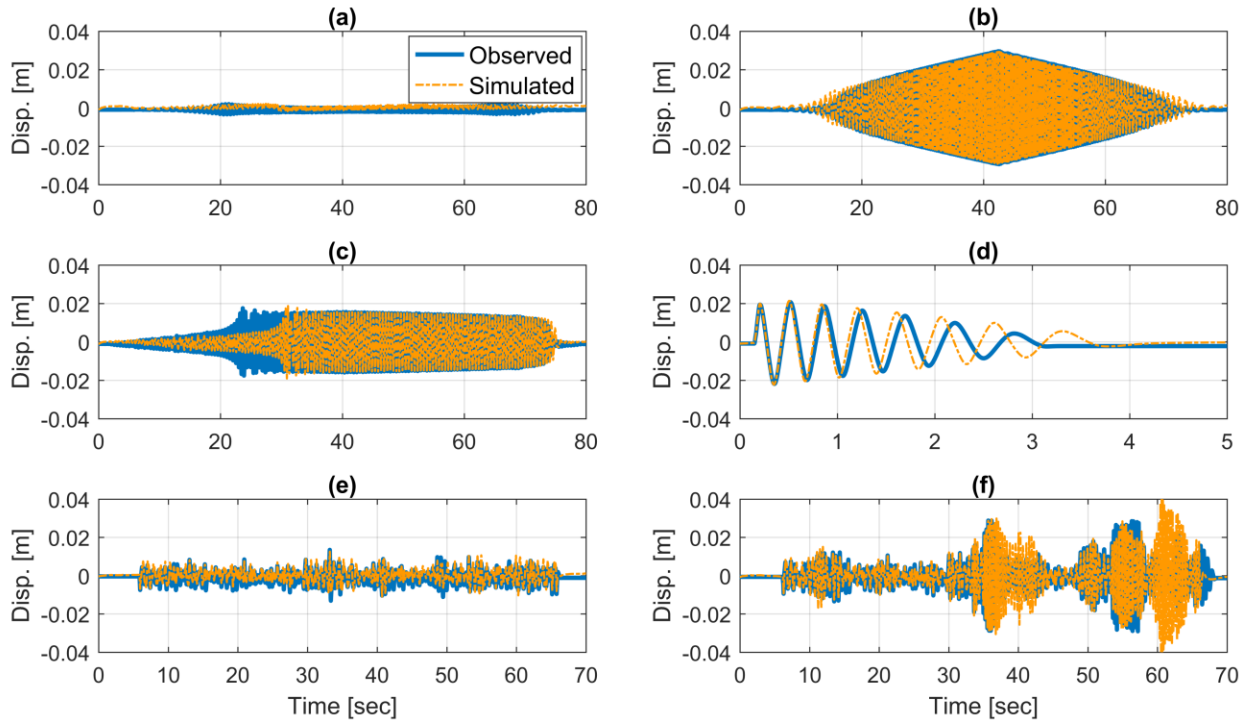
Fig. 6.6 – Response comparison for the models identified using approach B to the experimental device responses to signals 11-16, given in subplots (a)-(f) respectively.

The results from identification approach B demonstrate that the tendency of the UKF to over-fit the NES model to the signal used for training cannot be solved by simultaneously using multiple signals exhibiting different device behaviors in the training and validation processes. A similar approach which would appear to have the potential to counteract these issues would be to train, and likewise validate, the model sequentially on both identification signals. To achieve this, 5,000 candidate models and their covariance matrices are generated from the 5,000 prior distributions using a single identification signal. The candidate models and their respective covariance matrices then serve as the prior distributions for UKF identification with the other

121

training signal. The evaluation of this approach showed no particular advantage over approach B, regardless of the ordering of the signals used in the sequential identification process. The similarity of the results between this method and those of approach B, suggests that the sensitivity of the parameters to the information available in signal 9 is the governing factor in any dual identification approach using these signals. An approach is instead required which will take advantage of the information available in the training signals without over-fitting the identified model to suit those particular responses.

6.3.3   Models Generated Using Sequential Training Signals (Approach C)
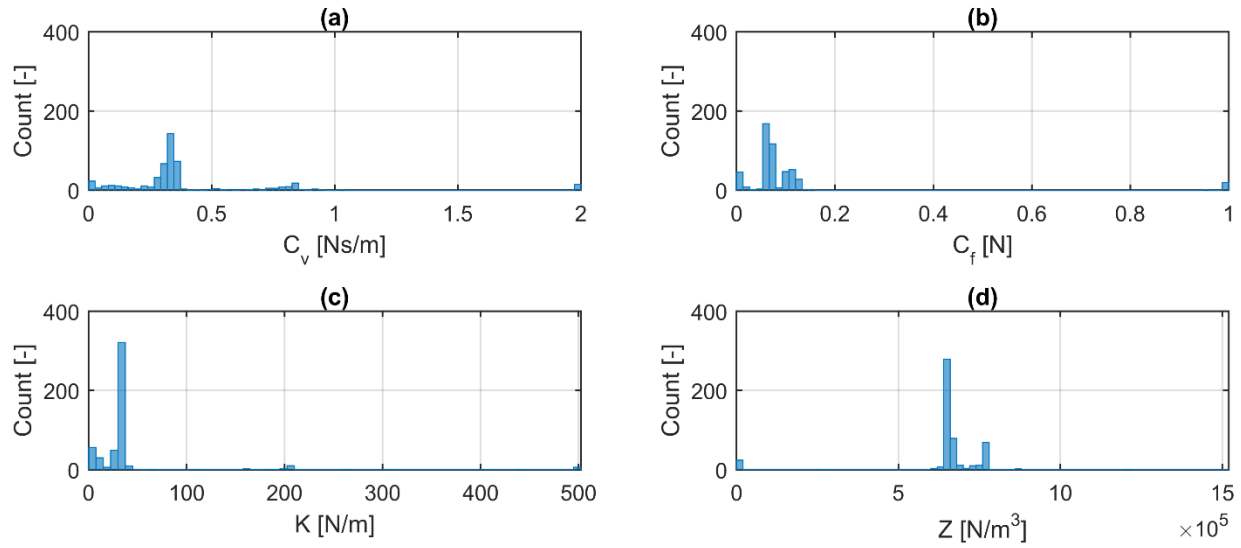
Fig. 6.7 - Distribution of the minimum MSE models for each prior mean on the parameters, generated using identification approach C.

In the final identification approach analyzed herein, approach C, the NES device model is identified by training on one signal and validating on another. This process is intended to take advantage of the varying levels of information in the identification signals to generate increasingly precise models of the NES device. The selection of which signal to use for training and which to use for validation is made through Sobol' sensitivity analysis. The variability in the candidate models is initially preserved by selecting a signal with lower sensitivity values for training. In this case, signal 10 is selected. A higher sensitivity signal, in this case signal 9, is then used to refine

the candidate models during the validation phase. This process aims to generate a robust model which is representative of the global behavior of the system instead of being preemptively tuned to a single response to which the parameters are highly sensitive.

Due to the nature of identification approach C, an analysis of the full distribution of models identified by the UKF yields no additional information as to the robustness of the method. More insight is found using the distribution of the 500 models identified as having the lowest MSE for each mean of the prior distribution on the parameters. These results, shown in Fig. 6.7, capture the effect of the two-stage process in concentrating the full distribution identified for signal 10 in Fig. 6.3 to emphasize the solution which best fits both training responses.
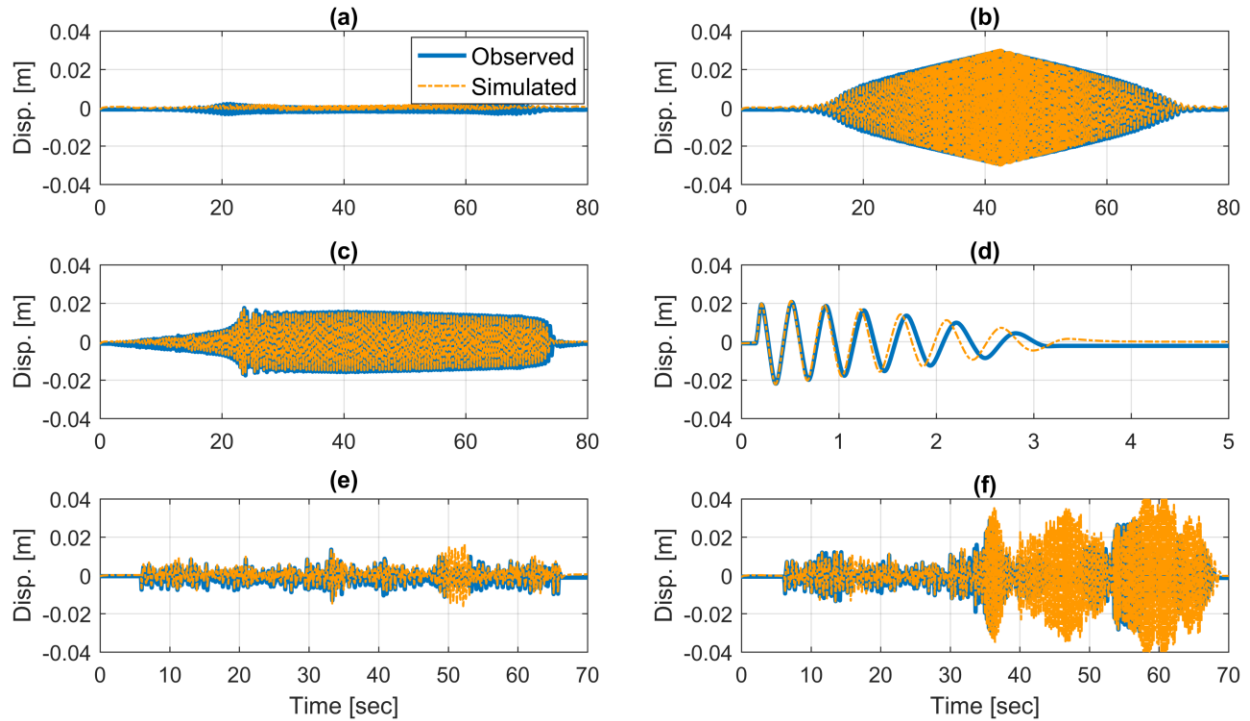
Fig. 6.8 - Response comparison for the models identified using approach C to the experimental device responses to signals 11-16, given in subplots (a)-(f) respectively.

The MSE and Sub-modal models identified with this method are given in Table 6.3, with the evaluation of the MSE model against the testing signals graphically displayed in Fig. 6.8. Some variation does exist in the parameter values identified between the two models, the largest being a 23% difference in the values identified for the linear stiffness parameter. Despite these differences, the models produce very similar, low-error results in both the validation phase and when compared

with the evaluation signals. This suggests that the range of variation expressed by these models is able to produce a stable identification of the NES device behavior, including the input energy threshold. In contrast with the results identified from the other identification approaches, which yield widely varied evaluation results with parameter variations on a similar level as this sequential identification method, the results of this method are remarkably robust. Although there are some responses for which the MSE is still fairly high, such as with signal 16, the responses collectively report much lower error values than those determined for the models generated by the other identification approaches. The MSE model response to signal 16 displays the limitations of the method. Though the cross-validation approach mitigates errors due to overfitting of the model, it cannot completely eliminate them, particularly for models such as the NES, where small adjustments in the parameters can shift the bifurcation in the response. The result of this inability to perfectly tune the model is particularly evident in the response of the high amplitude BLWN signal, as it randomly activates both facets of the NES response. The MSE value for this signal is notably quite high in all cases, though it happens to be particularly high for this case. However, despite having an elevated MSE value, the estimated response overestimates the displacements of the device, as shown in Fig. 6.8, thereby giving a conservative estimate of the response behavior.

Table 6.3 – Evaluation results for models identified using identification approach C. The MSE for the selected models are given with respect to the validation signal, as well as with respect to the evaluation signals.

Identified Models						MSE/10^8 for Evaluation Signals					
Selected Model	c_v $\left[\dfrac{\text{Ns}}{\text{m}}\right]$	c_f [N]	k $\left[\dfrac{\text{Ns}}{\text{m}}\right]$	$\dfrac{z}{10^5}$ $\left[\dfrac{\text{N}}{\text{m}^3}\right]$	$\dfrac{\text{MSE}}{10^8}$ [-]	11	12	13	14	15	16
MSE	0.296	0.075	43.2	6.23	0.7	1.2	0.6	1.1	17.1	5.6	113.6
Sub-Modal	0.344	0.064	33.1	6.54	0.9	1.3	0.7	6.9	10.5	2.0	68.6

The assumption that sensitivity should drive the selection of signal order for identification approach C can be further validated by the substitution of signal 10 in the training phase with another signal identified as having mid-range sensitivity, such as a band-limited white noise (BLWN) signal generated at the same amplitude as signal 16 but from a different random noise sequence. The parameter sensitivities for c_v, c_f, k, and z in response to this signal are 0.22, 1.00,

0.34, and 0.62, respectively. The results of this analysis are given in Table 6.4 and Fig. 6.9. The models generated and the comparisons with the evaluation signals are similar to those obtained when signal 10 is used in the training phase. The only significant differences in the results are in the MSE determined by a comparison to signal 14 and 16. For signal 14, the MSE increased in comparison to that expressed in the models generated from signal 10, which is shown in Fig. 6.9(d) as by a phase shift in the latter half of the response. For signal 16, particularly when the Sub-modal model is used, the error of the selected model in response to these signals is much smaller than that expressed with the models generated using signal 10 in the training phase. This is likely due to the similarity of the training signal to the evaluation signal. In general, as shown in Fig. 6.9 for the MSE model, the generated models still consistently express the overall trends in NES behavior.
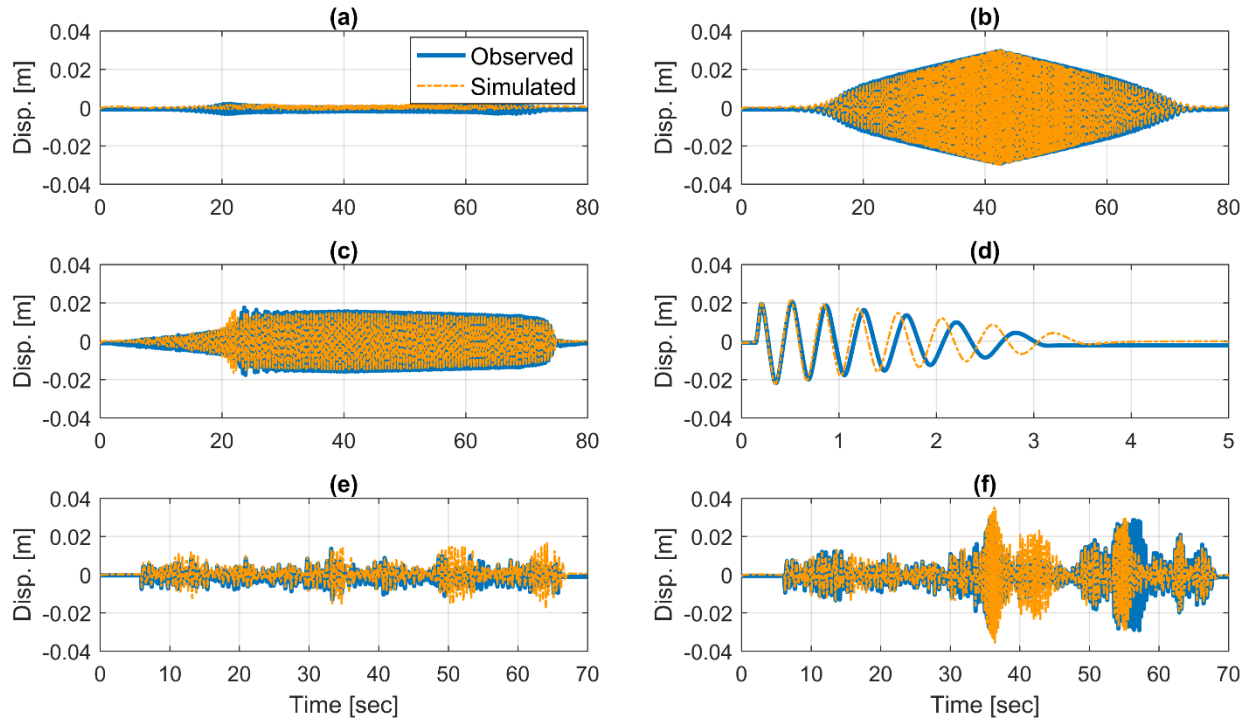
Fig. 6.9 - Response comparison for the models identified using approach C to the experimental device responses to signals 11-16, given in subplots (a)-(f) respectively. Alternate BLWN signal is used to identify the final simulated model.

It is important to note that although identification approach C is the most consistent and robust approach overall, it may not be as effective when the training phase incorporates signals with low sensitivity. Here, based on empirical evidence, I tentatively define low sensitivity to be

signals in which the majority of the parameters have a maximum sensitivity index of less than 20%. When that is the case, the candidate models generated by the training phase are not of high enough quality to be adequately representative of the signal for which they are trained. The validation phase is therefore equally ineffective at drawing together a large enough sample of high-quality models to guarantee a robust result. I likewise suggest that training signals be selected from those in which the maximum sensitivity of at least half of the parameters is between 20% and 80%. Validation signals, which should be of high sensitivity, are suggested to have a maximum sensitivity of above 80% in at least half of the parameters.

Table 6.4 – Evaluation results for identification approach C when a BLWN signal is used for training. The MSE for the selected models are given with respect to the validation signal, as well as with respect to the evaluation signals.

Identified Models						MSE/10^8 for Evaluation Signals					
Selected Model	c_v $\left[\dfrac{\text{Ns}}{\text{m}}\right]$	c_f [N]	k $\left[\dfrac{\text{Ns}}{\text{m}}\right]$	$\dfrac{z}{10^5}$ $\left[\dfrac{\text{N}}{\text{m}^3}\right]$	$\dfrac{\text{MSE}}{10^8}$ [-]	11	12	13	14	15	16
MSE	0.316	0.066	46.1	6.29	0.7	1.1	0.5	4.0	35.6	10.3	39.0
Sub-Modal	0.311	0.066	29.5	6.73	0.8	1.4	0.7	8.6	22.6	2.4	29.1

To compare these identification approaches I have considered 5,000 prior distributions. However, in practical identification scenarios, I ideally would aim to apply this approach using a limited number of identification trials. To determine whether this approach would still be valid for that scenario, a new sample of just five means and their ten associated coefficients of variation is drawn from the prior distribution of the parameters independent of the original 5,000 samples, according to the same process described in Section 6.2.2. Identification approach C is implemented with these 50 samples using signal 10 in the training phase and signal 9 in the validation phase, resulting in an MSE model in SI units of

$$0.664\ddot{x} + 0.344\dot{x} + 0.065\tanh(200\dot{x}) + 34x + 651{,}000x^3 = -0.664\ddot{x}_g, \qquad (6.11)$$

which generates MSE values $\{1.3, 0.7, 6.5, 10.9, 2.1, 79.1\} \times 10^8$ in response to evaluation signals 11-16, respectively. The model generated, and the comparison with the evaluation signals, are nearly identical to those of the Sub-modal model generated with all 5,000 identification trials, with the exception of the response to evaluation signal 16, which has a lower MSE result than is shown in Table 6.3, but is still elevated in comparison with the other evaluation signals. In this case, an

126

elevated MSE occurs due to an overestimation by the selected model of the response behavior in the final 10 sec of the BLWN response. However, the behavior during the majority of the response is well-represented. Overall, these results demonstrate that identification approach C can be used to generate robust models of the NES device which are representative of the device behavior.

6.4   Conclusions

Herein, I have developed and demonstrated a method for robustly identifying the model of a NES device using the UKF and experimental data. The NES has complex behavior, and exhibits dual stability about a particular input energy threshold. My approach leverages information available from a sensitivity analysis on the parameters, performed in advance of experimentation, to select device responses to be used in training and validating a model of the device. By comparing three approaches to identifying the model, a robust and consistent procedure is suggested and validated. This approach is shown to fit the response behavior of the identified model to the full range of responses expected from the experimental system, as opposed to tuning the model toward the training signal only. The method is also shown to provide a robust means to generate models for this class of devices with complex chaotic behavior, and enable the experimental identification of a model which can accurately estimate the desired range of NES behavior with just a few identification trials.

Through the evaluation of several options for identification approaches I have also demonstrated that, although the UKF is a powerful tool for identification, it must be used with care and judgement. In every identification approach, it is shown that small variations on the prior distributions on the parameters may produce significant variability in the identified models. These models may mischaracterize the behavior of the device completely, especially if training and validation steps are not properly implemented. It is only when appropriate consideration is given to the sensitivity of the training signals to the parameters that a robust method for model identification could be generated. This consideration is especially critical here due to the complexity of the dynamic behavior of the NES device.

The procedure developed herein provides an effective solution to the problem of over-fitting during UKF identification in experimental dynamic systems. Other potential methods to combat this effect exist and have been evaluated numerically, such as the introduction of a regularization term into the algorithm [15]. As these identification algorithms progress beyond

127

experiments to implementations, such methods must be further explored to enhance the robustness and accuracy of the results. Further exploration on this topic, through variations on the Bayesian approximation algorithm, is the topic of the next chapter.

6.5   References

1.   Chatzi, E. N., Smyth, A. W. & Masri, S. F. Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty. *Struct. Saf.* **32,** 326–337 (2010).

2.   Song, W. & Dyke, S. Development of a cyber-physical experimental platform for real-time dynamic model updating. *Mech. Syst. Signal Process.* **37,** 388–402 (2013).

3.   Song, W. & Dyke, S. J. Real-time dynamic model updating of a hysteretic structural system. *J. Struct. Eng.* **140,** 1–14 (2014).

4.   Shao, X., Mueller, A. & Mohammed, B. A. Real-Time hybrid simulation with online model updating: methodology and implementation. *J. Eng. Mech.* **142,** 1–19 (2015).

5.   Kontoroupi, T. & Smyth, A. W. Online noise identification for joints state and parameter estimation of nonlinear systems. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2,** 1–12 (2016).

6.   Lund, A., Dyke, S. J., Song, W. & Bilionis, I. Identification of an experimental nonlinear energy sink device using the unscented Kalman filter. *Mech. Syst. Signal Process.* **136,** (2020).

7.   Saltelli, A. *et al. Global Sensitivity Analysis. The Primer*. (John Wiley & Sons, 2008).

8.   Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55,** 271–280 (2001).

9.   Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **145,** 280–297 (2002).

10.   Lund, A., Silva, C., Dyke, S. J., Song, W. & Bilionis, I. Response of a prototype nonlinear energy sink device to various base excitations. *DesignSafe-CI* (2019). doi:10.17603/ds2-nh34-e513

11.   Wan, E., van der Merwe, R. & Nelson, A. T. Dual estimation and the unscented transformation. in *Advances in Neural Information Processing Systems* 666–672 (2000).

12. Sivia, D. S. & Skilling, J. *Data analysis: a Bayesian tutorial*. (Oxford University Press, 2006).

13. Muto, M. & Beck, J. L. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *J. Vib. Control* **14,** 7–34 (2008).

14. De, S., Johnson, E. A., Wojtkiewicz, S. F. & Brewick, P. T. Computationally efficient Bayesian model selection for locally nonlinear structural dynamic systems. *J. Eng. Mech.* **144,** (2018).

15. Bishop, C. M. *Pattern recognition and machine learning*. *Pattern Recognition* (Springer, 2006). doi:10.1117/1.2819119

16. Worden, K. Data processing and experiment design for the restoring force surface method, part I: integration and differentiation of measured time data. *Mech. Syst. Signal Process.* **4,** 295–319 (1990).

17. Sarkka, S. *Bayesian filtering and smoothing*. *Cambridge University Press* (Cambridge University Press, 2013). doi:10.1017/CBO9781139344203

7. COMPARISON OF VARIATIONAL INFERENCE AND UNSCENTED KALMAN FILTER FOR NONLINEAR STRUCTURAL SYSTEMS

Although several approaches exist to implement Bayesian inference on practical identification and monitoring problems, research in the predictive modeling of civil engineering structures has typically focused on methods which operate from either an analytical or a sampling perspective. As discussed in Chpts. 2 and 3, analytical inference techniques refer to those linear and nonlinear identification methods inspired by the Kalman filter [1], which was developed in the 1960s as the optimal filter for linear systems whose model and observation uncertainties can be described as Gaussian. Kalman filter methods have the common benefit of computational speed, allowing for near real-time structural identification, but are limited by the assumptions used to generate their analytical framework. Sampling techniques, such as particle filters [2] or sequential Monte Carlo algorithms [3], were developed to remove the barriers imposed by this analytical framework by drawing the necessary statistical information from a set of weighted sample points, or *particles*, to enable the use of any stochastic model of the system. However, this enhanced representation of the system comes at the cost of increased computational time and limited scalability to larger problems, which create additional barriers to the application of these methods with practical structural systems [4]–[7].

Optimization-based techniques to Bayesian predictive modeling have recently been developed which show great potential for balancing the computational speed, accuracy, and scalability necessary for practical implementation to civil engineering structures. These approaches, collectively referred to as *variational inference*, are well-recognized in the statistics community as a computationally efficient alternative to Monte Carlo (MC) methods for the inference of the hidden states, initial conditions, and physical parameters of a system, $\mathbf{z}$, from observations of its behavior, $\mathbf{y}$ [8]. The base variational inference algorithm, as well as the state-of-the-art automatic differentiation variational inference (ADVI), have been defined in Chpt. 3.

In this chapter, I apply ADVI to the identification of a simulated, single-degree-of-freedom Bouc-Wen system subject to base vibration. This system was selected as an introductory study of the algorithm in comparison with techniques more common to Bayesian structural identification, and is a common case study in developing similar comparisons between analytical and sampling methods [9],[10]. In particular, I draw a comparative analysis between the ADVI and the UKF

algorithms to address the relative accuracy of the identification methods, repeatability of the results given variations on the prior information, robustness to varying levels of measurement noise, and resilience to incorrect assumptions on the process noise between the two approaches. By so doing, I simulate many of the issues faced by experimentalists in applying these inference approaches and demonstrate the relative performance and flexibility of the two algorithms.

The chapter is organized as follows: Section 7.1 defines the simulated case study used to characterize the accuracy and reliability of the two inference approaches; and Section 7.2 discusses the influence of process noise assumptions and measurement noise intensities on the reliability of the results from each approach. The main conclusions are summarized in Section 7.3. The content of this chapter has been adapted from the published work of the author [11].

7.1   Case Study: Single-Degree-of-Freedom Bouc-Wen System

The relative performance of the ADVI and UKF algorithms is herein evaluated through the identification of the states $\mathbf{z} = [\mathbf{x}, \dot{\mathbf{x}}, \mathbf{r}]$ and the parameters, $\boldsymbol{\theta} = [c, k, \beta, n, \gamma]$, from the Bouc-Wen system

$$m\ddot{x}(t) + c\dot{x}(t) + kr(t) = -m\ddot{x}_g(t), \tag{7.1}$$

where the term $\ddot{x}_g$ is representative of an arbitrary base acceleration input and the states $\mathbf{x}$ and $\dot{\mathbf{x}}$ describe the physical displacement and velocity of the system, respectively. The state $\mathbf{r}$ describes the non-physical Bouc-Wen hysteretic component of the motion, the dynamics of which are expressed by

$$\dot{r}(t) = \dot{x}(t) - \beta|\dot{x}(t)||r(t)|^{n-1}r(t) - \gamma\dot{x}(t)|r(t)|^{n}. \tag{7.2}$$

A schematic diagram of this system is shown in Fig. 7.1. For this simulated case study, the true parameters of the system are set in accordance with those used in [9], such that the physical parameters of mass, stiffness, and damping are given values of $m = 1$ kg, $c = 0.3$ Ns/m, and $k = 9$ N/m, respectively. The non-physical Bouc-Wen shape parameters are assigned the values of $\beta = 2$ m^{-2}, $n = 2$, and $\gamma = 1$ m^{-2}. I choose to observe only the absolute acceleration of the system, which is shown in Fig. 7.2(c) and given as

$$y(t) = -\frac{1}{m}\left(c\dot{x}(t) + kr(t)\right). \tag{7.3}$$
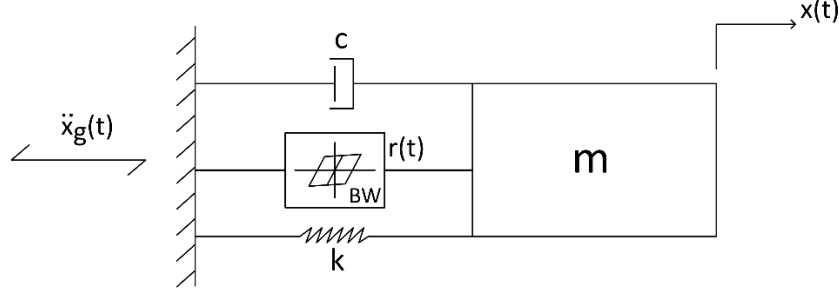
Fig. 7.1. Schematic Diagram of the SDOF Bouc-Wen System

Nondimensionalizing the equations of motion can ease the computation of the approximate posterior for many inference approaches, as it often puts the states and parameters on a similar scale and allows the inference algorithm to operate in a reduced space of potential solutions (see Chpt. 5) . I therefore nondimensionalize Eq. (7.1) – Eq. (7.3) using the time scale $t_c = \sqrt{k/m} = \omega_n = 3$ and the length scale $x_c = 0.05$ m, yielding transition and observation equations

$$\ddot{\bar{x}}(\tau) + 2\xi\dot{\bar{x}}(\tau) + \bar{r}(\tau) = -\frac{\ddot{x}_g(\tau/\omega_n)}{x_c\omega_n^2}, \tag{7.4}$$

$$\dot{\bar{r}}(\tau) = \dot{\bar{x}}(\tau) - \beta x_c^n|\dot{\bar{x}}(\tau)||\bar{r}(\tau)|^{n-1}\bar{r}(\tau) - \gamma x_c^n\dot{\bar{x}}(\tau)|\bar{r}(\tau)|^n, \tag{7.5}$$

$$\bar{y}(\tau) = -2\xi\dot{\bar{x}}(\tau) - \bar{r}(\tau), \tag{7.6}$$

where $\xi = c/2m\omega_n = 0.05$ is the damping ratio. The length scale is selected as representative of the expected maximum amplitude of response for the Bouc-Wen system, exemplifying a value an experimentalist may select for nondimensionalization when no other obvious relationships are present to leverage in the nondimensionalization process. The problem is now transformed such that the states and parameters we wish to infer are $\bar{\mathbf{z}} = [\bar{\mathbf{x}}, \dot{\bar{\mathbf{x}}}, \bar{\mathbf{r}}] = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ and $\bar{\boldsymbol{\theta}} = [\xi, \omega_n, \beta, n, \gamma] = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$. The transition dynamics can therefore be expressed in state-space form as

$$\begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \dot{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_2 \\ -2\theta_1\mathbf{x}_2 - \mathbf{x}_3 - \dfrac{\ddot{x}_g(\tau/\theta_2)}{x_c\theta_2^2} \\ \mathbf{x}_2 - \theta_3 x_c^n|\mathbf{x}_2||\mathbf{x}_3|^{\theta_4-1}\mathbf{x}_3 - \theta_5 x_c^n\mathbf{x}_2|\mathbf{x}_3|^{\theta_4} \end{bmatrix}. \tag{7.7}$$

Inference on this continuous system is performed using an Euler discretization scheme at a sampling frequency of $f_s = 128$ Hz. Although some methods are available for inference on

continuous systems, this strategy was selected to preserve a practical comparison of common applications of the UKF and variational inference methods.
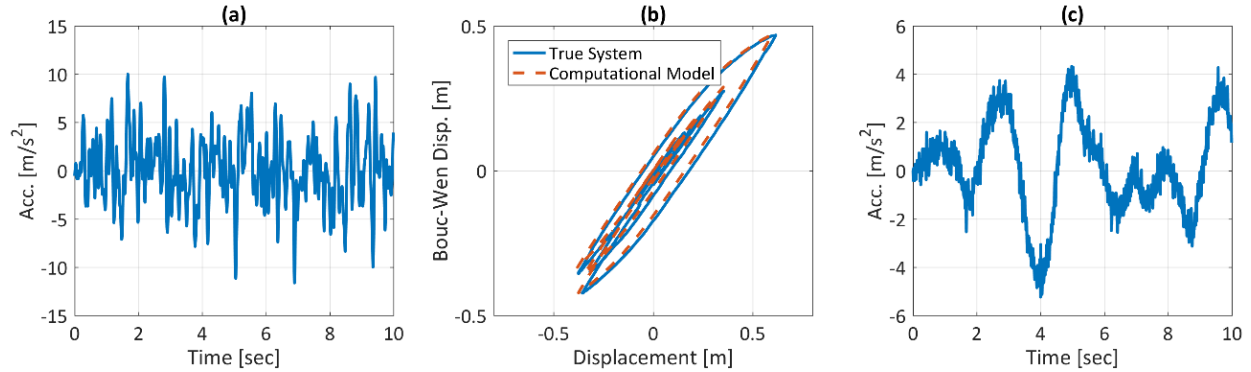
Fig. 7.2. Base acceleration and structural responses used in this case study. (a) base acceleration (b) Bouc-Wen hysteresis (c) noise-contaminated response acceleration

For the purpose of identification, a base excitation should be selected such that the resulting system response contains sufficient information to identify all parameters of the system. This condition is a particularly important consideration for nonlinear systems, as variations in the input signal may allow the system to express different aspects of its nonlinear behavior, as discussed in detail in Chpt. 5. In accordance with these considerations, I select a band-limited white noise (BLWN) base excitation signal with a maximum amplitude of ~ 10 m/s^2 and a frequency cut-off of 10 Hz, as shown in Fig. 7.2(a). The selection of the BLWN excitation is primarily made based on the indications of parameter identifiability from a Sobol' sensitivity analysis, see Chpt. 5, conducted using the python library SALib [12]. Of the variety of signals tested, the BLWN excitation produced a response with the highest level of sensitivity to all parameters, and particularly to the nonlinear shape parameters. The results of the Sobol' analysis for the BLWN signal are given in Fig. 7.3, which shows that the response of the Bouc-Wen system to this input signal is sensitive to variations in all parameters, though it is significantly more sensitive to the parameters ξ and ω_n. These results indicate that both algorithms should be able to identify all parameters of the system, though the identification will be less reliable, or more likely to result in different identified parameters given variations on the prior, for the nonlinear parameters. Further indications of system identifiability are given in Fig. 7.2(b). The response in this figure shows that

the BLWN input excites the system into its nonlinear range of response, and should therefore contain information concerning all parameters of interest.
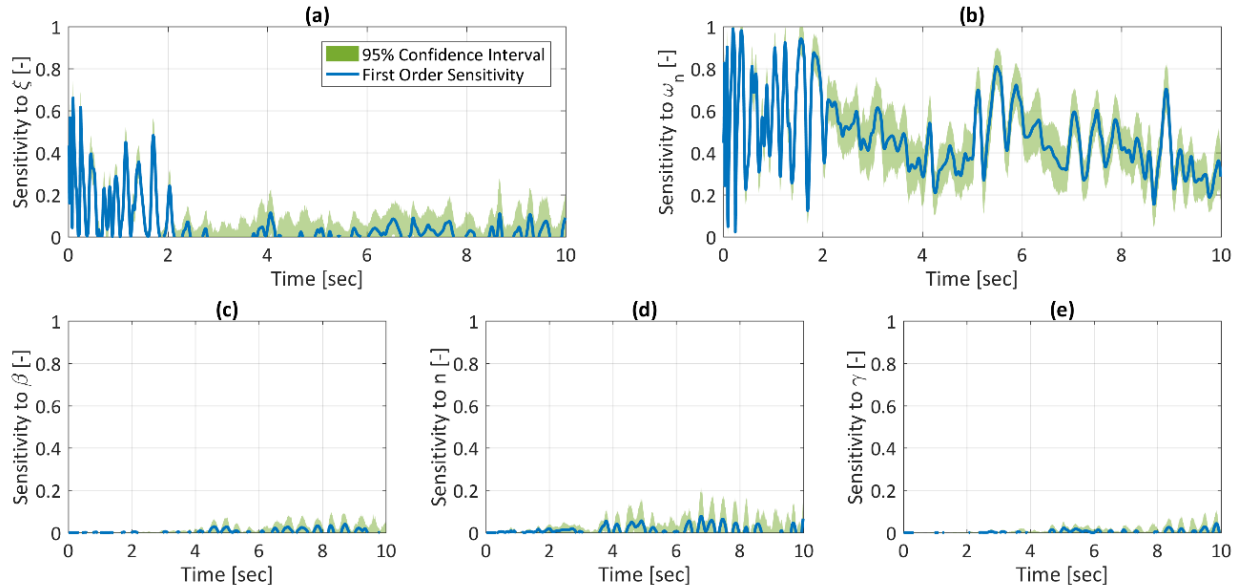
Fig. 7.3. First order sensitivity of the BLWN response to the system parameters

7.1.1 Parametric Studies

Understanding the performance of an inference algorithm in the face of model uncertainty and measurement error allows an experimentalist to determine where various algorithms can be applied for greatest effect. These practical issues pervade experiments at all scales and are therefore specifically evaluated in this comparative study of the UKF and variational inference techniques. Model uncertainty describes the difference between the computational model used for inference and the true experimental system from which data is being collected. It is often difficult to characterize this uncertainty robustly in an experimental setting and as such, the model uncertainty is often over- or under-estimated, which can have a large impact on the accurate inference of the system model.

To study this effect in simulation, the 'true' system model is developed such that it includes an additive, zero-mean Gaussian model uncertainty term on the states, given as $\mathbf{w}_t \sim N(0, \mathbf{Q})$, where $\mathbf{Q}$ is the process noise covariance matrix. The standard deviation of this term with respect to each of the states is selected such that the imposed noise represents a 1%, 2%, and 2% root-

mean-square noise-to-signal ratio (RMS-NSR) on the displacement, velocity, and Bouc-Wen displacement, respectively. This scaling factor on the imposed RMS noise is expressed by the vector $\mathbf{w}_{\text{RMS-NSR}} = [0.01, 0.02, 0.02]$. The additive model uncertainty term is then given by

$$\mathbf{w}_t \sim N\left(0, \text{diag}\left(\sqrt{\Delta\tau} \cdot \mathbf{w}_{\text{RMS-NSR}} \cdot \text{RMS}(\bar{\mathbf{z}})\right)^2\right), \tag{7.8}$$

where the $\sqrt{\Delta\tau}$ scaling factor is included to represent the uncertainty associated with discretizing the continuous ODE [13]. This noise is then propagated through the equation of motion for the computational model to generate the 'true' system response. The difference between the true response and the response of the computational model can be seen in Fig. 7.2(b). To study the impact of over- and under-estimating the true process uncertainty, the modeling assumption on the process noise covariance is varied according to

$$\mathbf{Q} = \text{diag}\left(\sqrt{\Delta\tau} \cdot \mathbf{w}_{\text{RMS-NSR}} \cdot \text{RMS}(\bar{\mathbf{z}}) \cdot \lambda_w\right)^2, \tag{7.9}$$

where $\lambda_w \in [80, 90, 100, 110, 120]\%$ expresses the degree to which the true process noise in Eq. (7.9) is over- or under-estimated. This process noise covariance matrix, modified by the appropriate λ_w, is assumed known for all parameter studies and is taken as constant throughout the inference process.

Measurement uncertainty describes the variations in our observations of the system which are not representative of its true behavior and is commonly attributed to noise in the sensors or testing environment. Although this uncertainty is simpler to estimate experimentally, it can be difficult to control, leading to observation signals with very high NSRs. To study the impact of measurement uncertainty on the selected inference algorithms, the measurement noise covariance of the true response data used for inference is varied according to

$$\mathbf{R} = \left(v_{\text{RMS-NSR}} \cdot \text{RMS}(\bar{\mathbf{y}})\right)^2, \tag{7.10}$$

where $v_{\text{RMS-NSR}} \in [0.1, 5, 10, 20, 30, 40, 50]\%$. This measurement noise covariance matrix is assumed known for all parameter studies conducted herein and is taken as constant throughout the inference process.

For each parameter study, one uncertainty measure is held constant while the other is varied. The base example to connect these two parameter studies is selected as the case in which $\lambda_w = 100\%$ and $v_{\text{RMS-NSR}} = 20\%$, representing the case in which the model uncertainty is correctly

characterized and the measurement uncertainty is moderately high. The results from these parameter studies are explored in detail in Section 7.2.

7.1.2 Implementation of the Selected Inference Methods

Knowing the two approaches to have distinct differences in their means of approximating the inference process, I seek to make an equitable comparison between them to assess their relative proficiency in predictive modeling for structural systems. Understanding the reliability of these algorithms with respect to reasonable variations on the selected prior is a key aspect of this assessment, as the choice of prior distributions on the states and parameters can have a large impact on the results obtained from the UKF and variational inference methods (see Chpts. 5 and 6). In practice, low variance prior distributions can often be determined for the states, but prior knowledge of the parameters is usually far less explicit. The robustness of these methods to variations in the priors on the parameters is therefore evaluated by selecting 50 distributions for the parameter priors which are representative of likely assumptions that an experimentalist might make in defining this problem. These 50 prior distributions are used with each case in the parameter studies. In all cases, prior distributions on the dynamic states are uniformly set as $x_i(0) \sim N(0, 0.25^2)$, as the system is known to be at rest prior to excitation.

The prior distributions on the parameters are selected such that the representation of the priors is equivalent between the two algorithms. The 50 preliminary mean values are first selected using a Latin hypercube (LH) sample on $\xi \in [0.1,1]$, $\omega_n \in [0,10]$, $\beta \in [0,25]$, $n \in [2,6]$ and $\gamma \in [0,25]$. Different distributional representations are then selected for each algorithm to accommodate their implementation. In the case of the UKF, all prior distributions must be Gaussian. To accommodate this constraint while also reflecting the existence of the parameters in the positive domain, inference is performed on the logarithm of the parameters, resulting in the prior distributions $\ln(\xi) \sim N(\ln(\mu_\xi), \sigma_\xi^2)$, $\ln(\omega_n) \sim N(\ln(\mu_{\omega_n}), \sigma_{\omega_n}^2)$, $\ln(\beta) \sim N(\ln(\mu_\beta), \sigma_\beta^2)$, $\ln(n) \sim N(\ln(\mu_n), \sigma_n^2)$, and $\ln(\gamma) \sim N(\ln(\mu_\gamma), \sigma_\gamma^2)$, where the means μ_k are members of the LH sample and the variance terms σ_k^2 are hand-selected to represent the uncertainty level a typical experimentalist might assign to the parameter. Due to the flexibility of variational inference, the priors on the parameters can be expressed directly with a lognormal distribution as $\xi \sim LogNormal(\ln(\mu_\xi), \sigma_\xi^2)$, $\omega_n \sim LogNormal(\ln(\mu_{\omega_n}), \sigma_{\omega_n}^2)$, $\beta \sim LogNormal(\ln(\mu_\beta), \sigma_\beta^2)$,

$n \sim LogNormal(\ln(\mu_n), \sigma_n^2)$, and $\gamma \sim LogNormal(\ln(\mu_\gamma), \sigma_\gamma^2)$. Inference then proceeds on the parameters directly instead of on their natural logarithm. It should be noted that the support of the prior distributions selected for VI are constrained to a subset of $\mathbb{R}^K$. To allow for full support in $\mathbb{R}^K$ when implementing the ADVI algorithm (see Section 3.2), the parameters are transformed within the algorithm according to

$$\hat{\boldsymbol{\theta}} = \boldsymbol{T}_{LN}(\boldsymbol{\theta}) = \ln(\boldsymbol{\theta}). \tag{7.11}$$

The emission and transmission probability models are represented slightly differently between the two inference algorithms. For the UKF these densities are described primarily by their means, which are given by the deterministic functions

$$\mathbf{f}(\bar{\mathbf{z}}_k, \ln(\bar{\boldsymbol{\theta}}), \ddot{x}_{g,k}) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{k-1} + \Delta\tau \begin{bmatrix} x_2 \\ -2\bar{\theta}_1 x_2 - x_3 - \dfrac{\ddot{x}_g((\tau-1)/\bar{\theta}_2)}{x_c \bar{\theta}_2} \\ x_2 - \bar{\theta}_3 x_c^{\bar{\theta}_4}|x_2||x_3|^{\bar{\theta}_4-1}x_3 - \bar{\theta}_5 x_c^{\bar{\theta}_4} x_2 |x_3|^{\bar{\theta}_4} \end{bmatrix}_{k-1} , \tag{7.12}$$

$$\mathbf{h}(\bar{\mathbf{z}}_k, \ln(\bar{\boldsymbol{\theta}})) = -2\bar{\theta}_{1,k} x_{2,k} - x_{3,k}. \tag{7.13}$$

The covariance associated with these mean terms is given by the process noise and measurement noise distributions, as defined in Eq. (7.9) and Eq. (7.10). Note that the structure of the UKF algorithm phrases the augmented state as a multivariate Gaussian, where dependence is implied across the states and parameters. Using variational inference, the emission and transmission probabilities for this system can be understood by

$$p(x_{1,k}|\bar{\mathbf{z}}_{k-1}, \hat{\boldsymbol{\theta}}) = N(x_{1,k}| x_{1,k-1} + \Delta\tau(x_{2,k-1}), w_1^2), \tag{7.14}$$

$$p(x_{2,k}|\bar{\mathbf{z}}_{k-1}, \hat{\boldsymbol{\theta}}) =$$
$$N\left(x_{2,k} \mid x_{2,k-1} + \Delta\tau\left(-2e^{\hat{\theta}_1} x_{2,k-1} - x_{3,k-1} - \dfrac{\ddot{x}_g(\tau_{k-1}/e^{\hat{\theta}_2})}{x_c e^{\hat{\theta}_2}} \right), w_2^2 \right), \tag{7.15}$$

$$p(x_{3,k}|\bar{\mathbf{z}}_{k-1}, \hat{\boldsymbol{\theta}}) = N\big(x_{3,k} \mid x_{3,k-1}$$
$$+\Delta\tau\left(x_{2,k-1} - e^{\hat{\theta}_3} x_c^{e^{\hat{\theta}_4}}|x_{2,k-1}||x_{3,k-1}|^{e^{\hat{\theta}_4}-1} x_{\tau-1,3} - e^{\hat{\theta}_5} x_c^{e^{\hat{\theta}_4}} x_{2,k-1}|x_{3,k-1}|^{e^{\hat{\theta}_4}} \right), w_3^2 \big), \tag{7.16}$$

$$p(y_k|\bar{\mathbf{z}}_k, \hat{\boldsymbol{\theta}}) = N\big(y_k \mid -2e^{\hat{\theta}_1} x_{2,k} - x_{3,k}, v^2 \big), \tag{7.17}$$

where the assumption of independence across the states and parameters is enforced by the separation of the distributions expressed in Eq. (7.14)-Eq. (7.17).

7.2 Results and Discussion

As discussed in Section 7.1.1, the goal of this case study is to ascertain the relative effectiveness of the UKF and variational inference methods in identifying the hidden states $\bar{\mathbf{z}}$ and parameters $\bar{\boldsymbol{\theta}}$ from the responses of the nondimensionalized Bouc-Wen system when measurement noise levels and process noise assumptions are varied. To gain a preliminary understanding of the results, I first examine the case in which the measurement uncertainty factor is set to $v_{RMS-NSR} = 0.1\%$ and the process uncertainty factor is set to $\lambda_w = 100\%$. This scenario represents the most ideal case of those studied herein, as the selected uncertainty factors reflect a situation in which the difference between the model and the true system can be defined with certainty and the measurements can be trusted explicitly.

For each inference method, 50 identification trials are conducted corresponding to the 50 prior distributions on the parameters defined in Section 7.1.2. For each identification trial, the identified model is selected as the one which maximizes the posterior probability on the parameters, or in other words, the model which minimizes the 0-1 loss [14]. The resulting 50 candidate models are then used to re-simulate the response of the inferred system to the given BLWN base excitation. The system models inferred by each algorithm which respectively result in the minimum RMS error on the states are shown in Fig. 7.4 in terms of the prior and posterior distributions on the parameters. It should be noted that the modeling approach for the parameters in the UKF allows for correlations between them which are not expressed in the variational inference method, where the parameters are explicitly assumed independent through the definition of their variational family. Fig. 7.4 therefore shows the marginal posteriors of the respective parameters for the UKF to allow for a general comparison with the variational inference approach. The results suggest that both models express similar levels of certainty and accuracy in their posterior approximations of the linear parameters ξ and ω_n. The two algorithms also appear to produce similar results in their expression of the nonlinear parameters $\beta, n,$ and γ, though the results from variational inference express greater certainty. Although slight differences exist between the identified models, they result in nearly identical approximations of the system response, as shown in Fig. 7.5. Here, the true response of the Bouc-Wen system is compared with model responses which are re-simulated from the posterior modes given in Fig. 7.4. These re-simulated responses show minimal variation from the true response of the system, suggesting that both of the inferred models are adequate solutions.
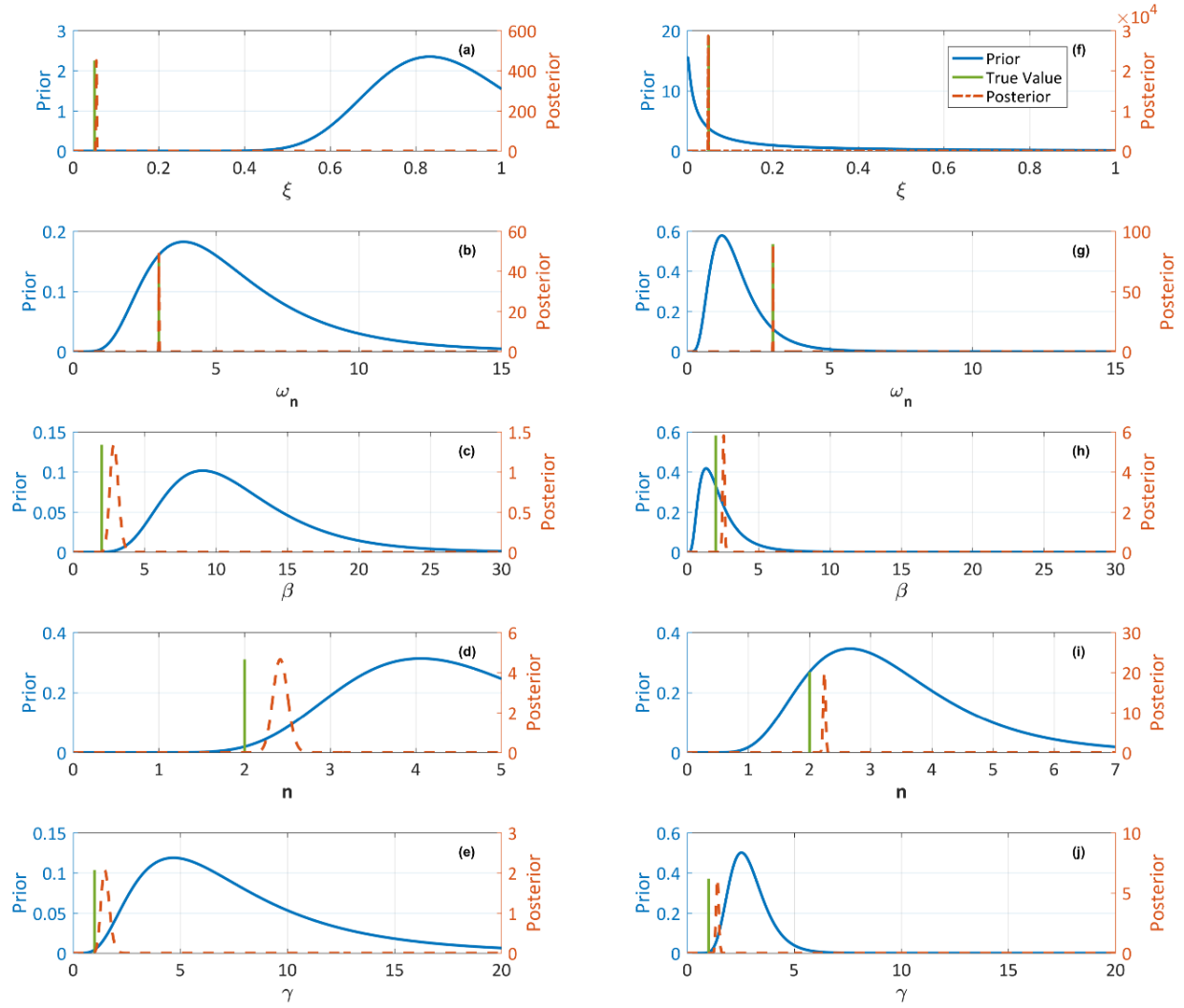
Fig. 7.4. Priors and posteriors of the minimum RMS error identification trials. (a-e) UKF identification trials (f-j) variational inference identification trials.
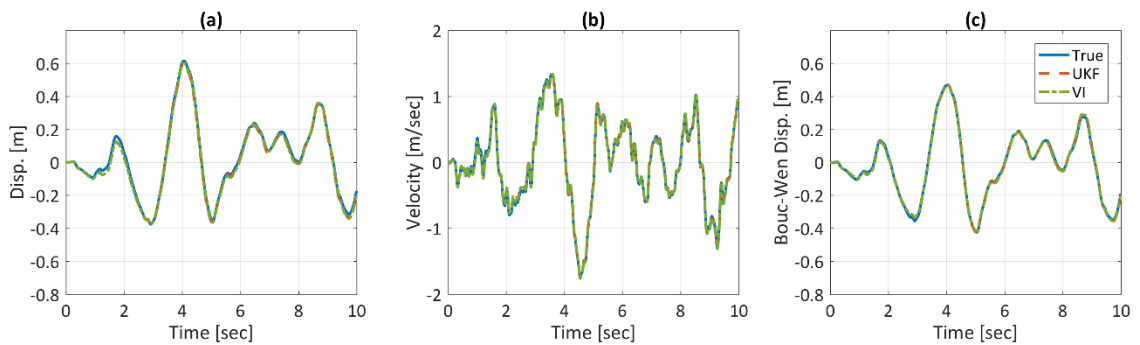
Fig. 7.5. Comparison of the true system response with model responses re-simulated from the minimum RMS error identification trials. (a) displacement response (b) velocity response (c) Bouc-Wen displacement response

However, the effectiveness of an inference approach lies not only in its best performance, but in the repeatability of that performance given realistic variations on the assumptions used in the model. I therefore analyze the relative accuracy of the UKF and variational inference methods over all 50 identification trials, as shown in Fig. 7.6 and Fig. 7.7. Fig. 7.6 gives the distribution of the posterior modes for all identification trials. In large part, these data show what we might expect from such an ideal case in which our measurement error is low and our modeling error is known. The distributions of the modes around all parameters are fairly well concentrated, particularly for the linear parameters ξ and ω_n to which the sensitivity of the response of the system is known to be high. There are no significant outliers, which indicates that both the UKF and variational inference trials did not begin to diverge in their search for the true system model. It is interesting to note, however, that the relative spread of posterior modes in the nonlinear parameters resulting from the UKF is wider than those identified with variational inference. These results suggest that the variational inference method is better able to consistently identify parameters with relatively low levels of practical identifiability than the UKF.
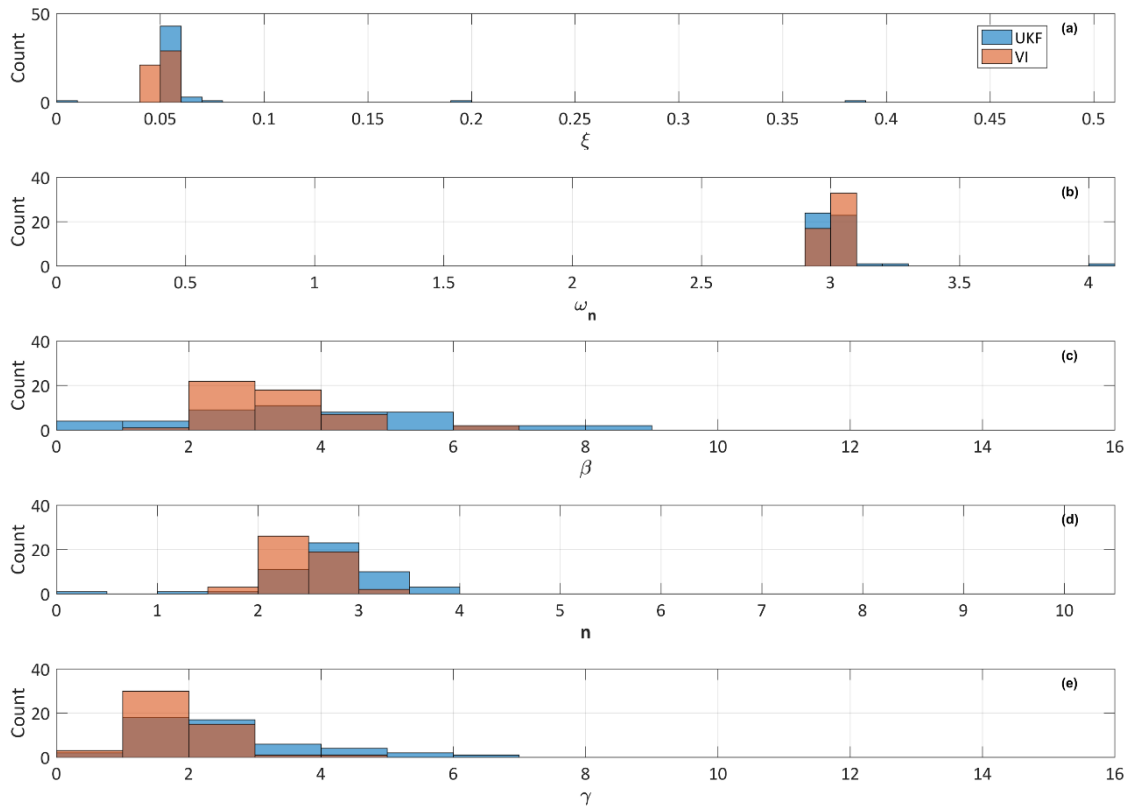
Fig. 7.6. Distribution of the posterior modes on the parameters for all 50 identification trials

A final comparison of the two inference methods in this ideal case is given in Fig. 7.7. Here, the RMS error in the states for each inference trial is plotted in a case-by-case comparison between the UKF or variational inference methods. The data clearly show that, regardless of the variations in the parameters, variational inference consistently provides a low-error response with respect to the true states. The UKF is able to match this performance for the majority of the identification trials, but experiences larger variations in error due to the variations in the parameter posteriors shown in Fig. 7.6. These results suggest that even though the posteriors may not precisely match the true parameters, the variational inference approach is more adept at consistently finding parameter combinations which locally minimize the error between the true response and the re-simulated model.
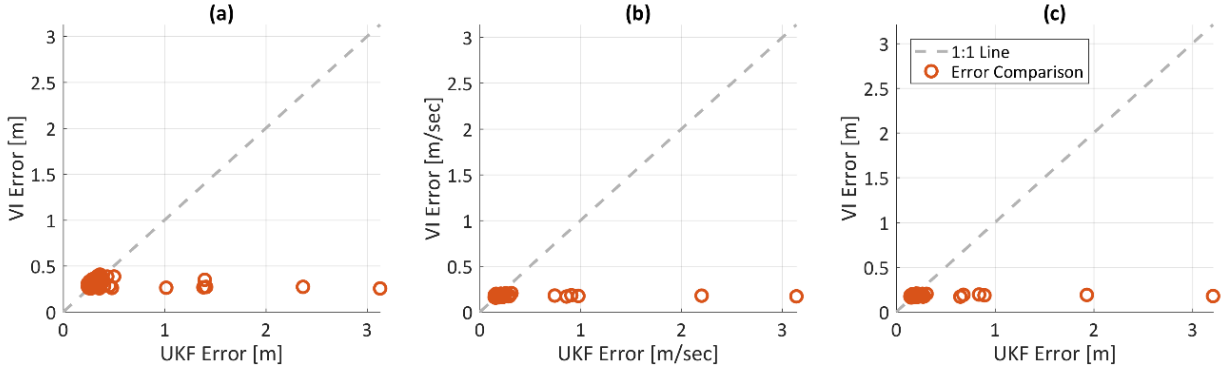
Fig. 7.7. Comparison of the RMS error on the states for the 50 identification trials. (a) displacement comparison (b) velocity Comparison (c) Bouc-Wen displacement comparison

The results from this preliminary study focusing on a particular case of model and measurement uncertainty allow us to approach the results from the parameter studies on these noise terms from a more informed perspective.

7.2.1   Influence of Process Noise Assumptions

In this study, I evaluate the effect of process noise assumptions during inference on the resulting identified models. To accomplish this, I take the measurement uncertainty factor as constant at $v_{RMS-NSR} = 20\%$, representing a moderately high level of measurement noise, and allow the process uncertainty factor to vary according to $\lambda_w \in [80, 90, 100, 110, 120]\%$. The results of this study are given in Fig. 7.8 and Fig. 7.9.

Fig. 7.8 shows the accuracy of the posterior models identified for each inference method at each level of variation in the assumptions on process noise. The information in this graph can be viewed as a consolidated form of the information presented in the histograms of Fig. 7.6. For each value of λ_w, the percentage of models for which all parameters exhibit less than a 10% error from their true values are recorded by the black line. The percentage of models for which the parameters exhibit other levels of error are recorded similarly, in varying shades of grey. The results show a similar level of accuracy and consistency between the two inference methods, though the increase in suitable models with error tolerance for the variational inference approach suggests a larger concentration of models close to, if not exactly meeting, the correct solution. One can additionally observe from these plots that the UKF is relatively indifferent to process noise in terms of the accuracy of the models produced, and settles at around 5% of models having an accuracy of 50% or better regardless of the assumption on the process noise covariance. The variational inference method, in contrast, appears to benefit from an underestimation of the process noise covariance, yielding 10% of models having an accuracy of 50% or better when the process noise covariance is assumed at 80% of its true value, as opposed to the 6% of models with accuracy 50% or better achieved when the process noise is assumed to be 120% of its true value.

Fig. 7.8. Percentage of inference trials yielding models whose parameters are all less than a certain percentage error, shown as the process noise uncertainty factor (λ_w) is varied. (a) UKF Models (b) VI Models.

Further insight into the robustness of these methods with respect to incorrect process noise assumptions is given in Fig. 7.9. Similar to a box-and-whisker plot, this figure shows the minimum value, first quartile, median, third quartile, and maximum value of the RMS error in the states at each level of λ_w. The results show that, in addition to the nominally higher degree of model accuracy suggested by Fig. 7.8, variational inference also produces models which more

consistently achieve low errors in comparison with the true response of the system. This behavior is particularly evident through a comparison with the UKF results, as the third quartile of the variational inference model responses lie consistently below the median of the UKF model responses for all system states. It is also significant to note the large number of outliers produced by the UKF in comparison with variational inference. In combination with the results from Fig. 7.8, this result suggests that the UKF experiences a greater number of outliers during model generation which do not generate locally optimal solutions in relation to the true system response. Variational inference, in contrast, seems to generate models which are more representative of the true behavior of the system, despite having a similar rate of parameter error to the UKF models. Even when an accurate model is not produced, this method finds locally optimal models which reduce the error of the inferred model relative to the true response. Overall, variational inference appears to produce superior results when assumptions on the process noise are varied.

Fig. 7.9. Statistics of the RMS error in the states given the identified models for each inference method when the process noise uncertainty factor (λ_w) is varied.

7.2.2 Influence of Measurement Noise Level

For the second parameter study, I evaluate the impact of measurement noise in the response data on the resulting identified models. To study this effect, I take the process uncertainty factor as constant at $\lambda_w = 100\%$, representing a correct assumption on the process noise covariance, and allow the measurement noise to vary according to $v_{RMS-NSR} \in [0.1, 5, 10, 20, 30, 40, 50]\%$. The results of this study are given in Fig. 7.10 and Fig. 7.11.
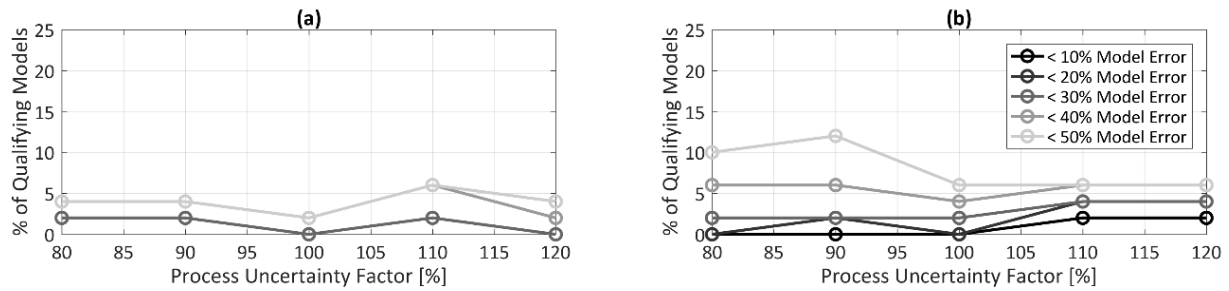
Fig. 7.10. Percentage of inference trials yielding models whose parameters are all less than a certain percentage error, shown as the measurement noise uncertainty factor ($\mathbf{v}_{\mathbf{RMS-NSR}}$) is varied. (a) UKF Models (b) VI Models.

Fig. 7.10 shows the accuracy of the posterior models identified for each inference method at each level of variation in the measurement noise. The results express the relative levels of accuracy and rates of degradation in accuracy exhibited by each approach. The UKF achieves a peak level of accuracy and consistency at 5% measurement noise with only around 15% of models having an accuracy of 50% or better, while variational inference peaks at 0.1% measurement noise with around 45% of models having an accuracy of 50% or better. This lack of performance in the UKF is compensated by a slower rate of degradation in performance. For measurement noise levels with $v_{RMS-NSR} = 30\%$ or greater, variational inference does not produce any models with at least 50% accuracy, whereas approximately 5% of UKF models can still claim to be within 50% of the true parameters. These results suggest that variational inference is a very effective tool in cases of moderate measurement uncertainty, but that the UKF is much more reliable for cases where measurement uncertainty is extremely high.

These results are in agreement with those given in Fig. 7.11. This plot shows that the error in the states as a result of the inferred models increases significantly for the variational inference cases when measurement noise levels are greater than or equal to 30%. The error resulting from the UKF inference cases, in contrast, increases slowly and consistently across all measurement noise levels tested. It is only for cases of low to moderate measurement noise, where $v_{RMS-NSR} \in [0.1, 5, 10, 20]\%$, that the performance of the variational inference method exceeds that of the UKF. However, as was mentioned in the previous parameter study, the inferred UKF responses show a greater number of outliers than those of the variational inference method. The consistency of this result between the case studies suggests that the variational inference method is more adept at consistently finding parameter combinations which locally minimize the error between the true

144

response and the re-simulated model, even though the local minima may shift away from the true solution in the presence of excessive measurement noise.
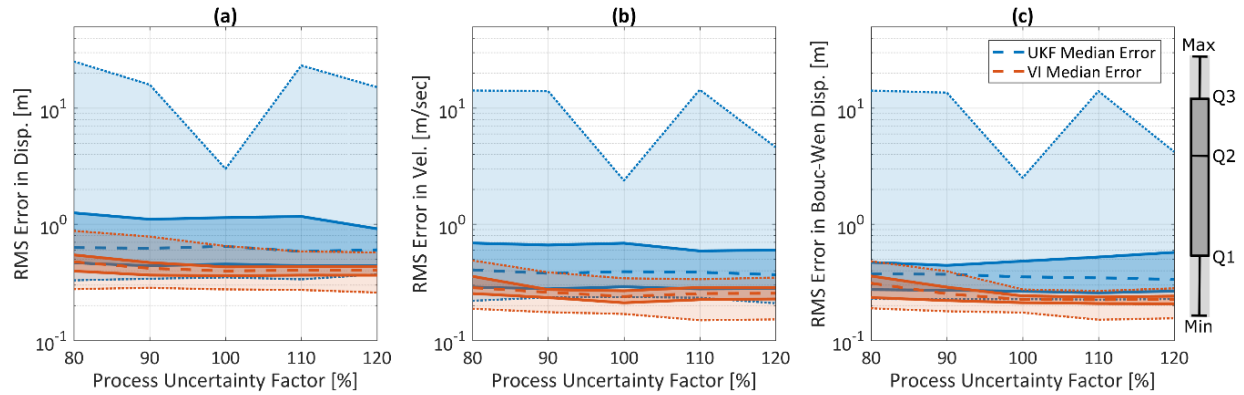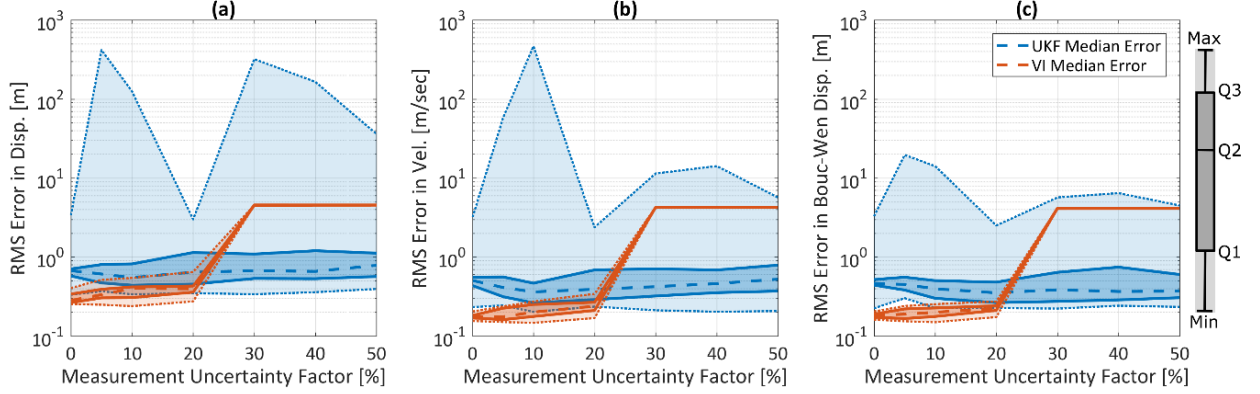

Fig. 7.11. Statistics of the RMS error in the states given the identified models for each inference method when the measurement noise uncertainty factor ($v_{RMS-NSR}$) is varied.

7.2.3 Comparison of Variational Inference Results with Those Using More Informative Priors

The results expressed in previous sections give a one-to-one comparison of the UKF and variational inference algorithms, expressing the problem formulation for both algorithms in terms compatible with the more constrained UKF representation. However, part of the value of the variational inference approach is its ability to offer a more precise expression of both the prior and posterior uncertainty on the states and parameters. In this section, I demonstrate the impact of this flexibility by reimplementing the variational inference approach with prior distributions on the parameters which take full advantage of the level of knowledge that would be available to an experimentalist faced with this identification problem. Particularly, the parameters ξ and n parameters have clear domains of $\xi \epsilon [0,1]$ and $n \epsilon [1,6]$ for this case study. With variational inference, preliminary knowledge concerning these parameters can be described by the prior distributions by $\xi \sim Uniform(0, \mu_\xi)$ and $n \sim Uniform(0, \mu_n)$, where μ_ξ and μ_n are as described in Section 7.1.2. This expression of the prior probability requires the definition of the transformation

$$\widehat{\boldsymbol{\theta}}_U = T_U(\boldsymbol{\theta}_U) = \ln\left(\frac{\boldsymbol{\theta}_U - a}{b - \boldsymbol{\theta}_U}\right), \tag{7.18}$$

145

to allow for full support in $\mathbb{R}^K$ when implementing ADVI, where $\boldsymbol{\theta}_U$ refers to the parameters with uniform prior distributions and a and b refer to the respective lower and upper bounds on the uniform prior. All other parameters are given the same prior distributions as expressed in Section 7.1.2. Inference using these prior distributions is then repeated on the same parameter studies examined in previous sections.
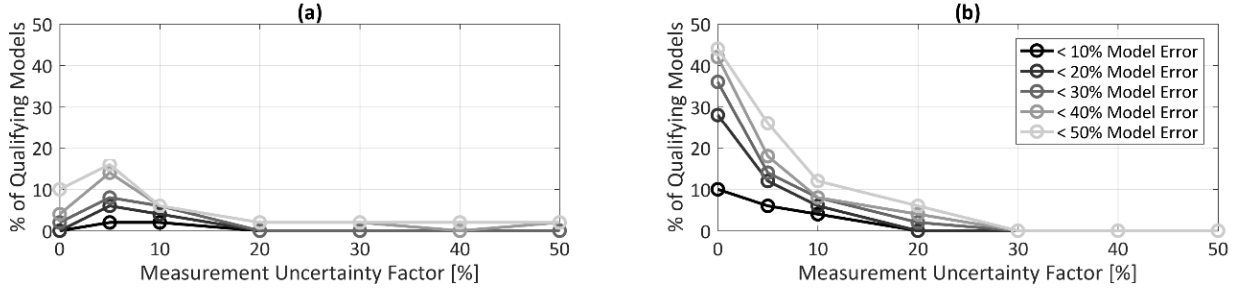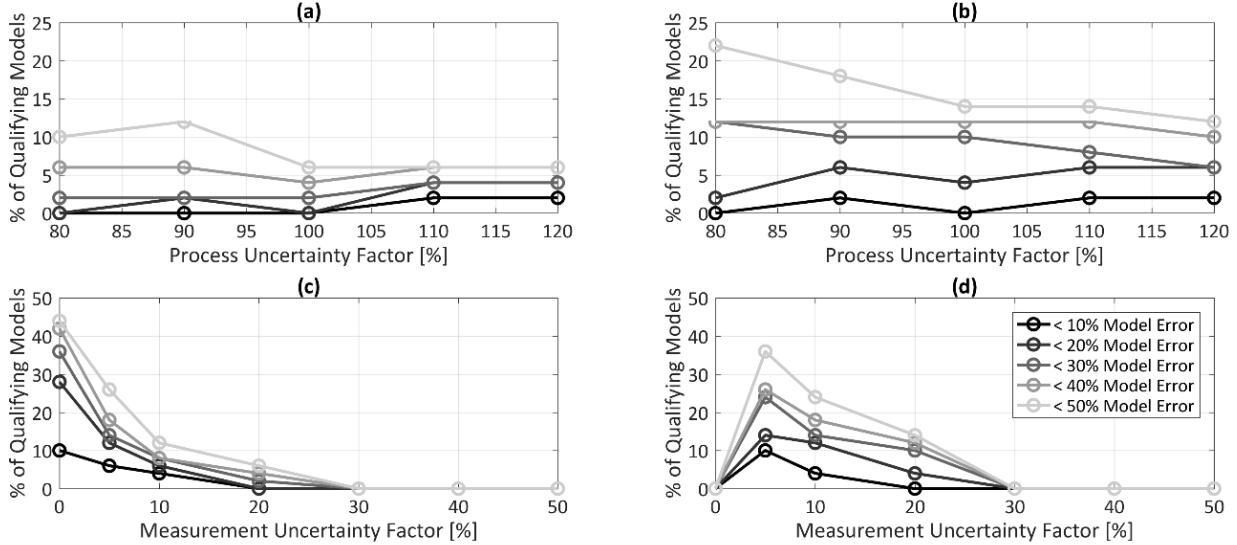

Fig. 7.12 Percentage of inference trials yielding models whose parameters are all less than a certain percentage error. (a) λ_w is varied for VI models with lognormal priors, (b) λ_w is varied for VI models with lognormal priors incorporating uniform priors, (c) $v_{RMS-NSR}$ is varied for VI models with lognormal priors, (d) $v_{RMS-NSR}$ is varied for VI models incorporating uniform priors.

The results of these inference trials are reported in Fig. 7.12 and Fig. 7.13 in comparison with the variational inference results obtained using solely lognormal priors on the parameters. It is significant to note the key differences in the two sets of variational inference results. Overall, the variational inference case with more informed parameter priors performed better than the case with solely lognormal priors, which indicates the value of the flexibility of variational inference to clearly specify the prior information known to the problem. One key exception to this trend is shown in Fig. 7.12(c,d), which shows the results of the measurement noise study. Here, the informed variational inference trial expresses no accurate models at $v_{RMS-NSR} = 0.1\%$, whereas the variational inference trial with less informative priors yields the largest number of accurate

146

models at this same measurement noise level. The precise reason for this behavior is unclear, but it is likely associated with a distortion of the local optimum of the ELBO due to the low level of noise and the particular expression of the prior on the parameters. Indeed, all parameters of the model for this case are identified with high levels of accuracy, except for ξ which is consistently identified at ~0.02. This result validates research avenues in the field of alternative loss functions for variational inference as it demonstrates that though the ELBO is an extremely useful loss function, it is not infallible.
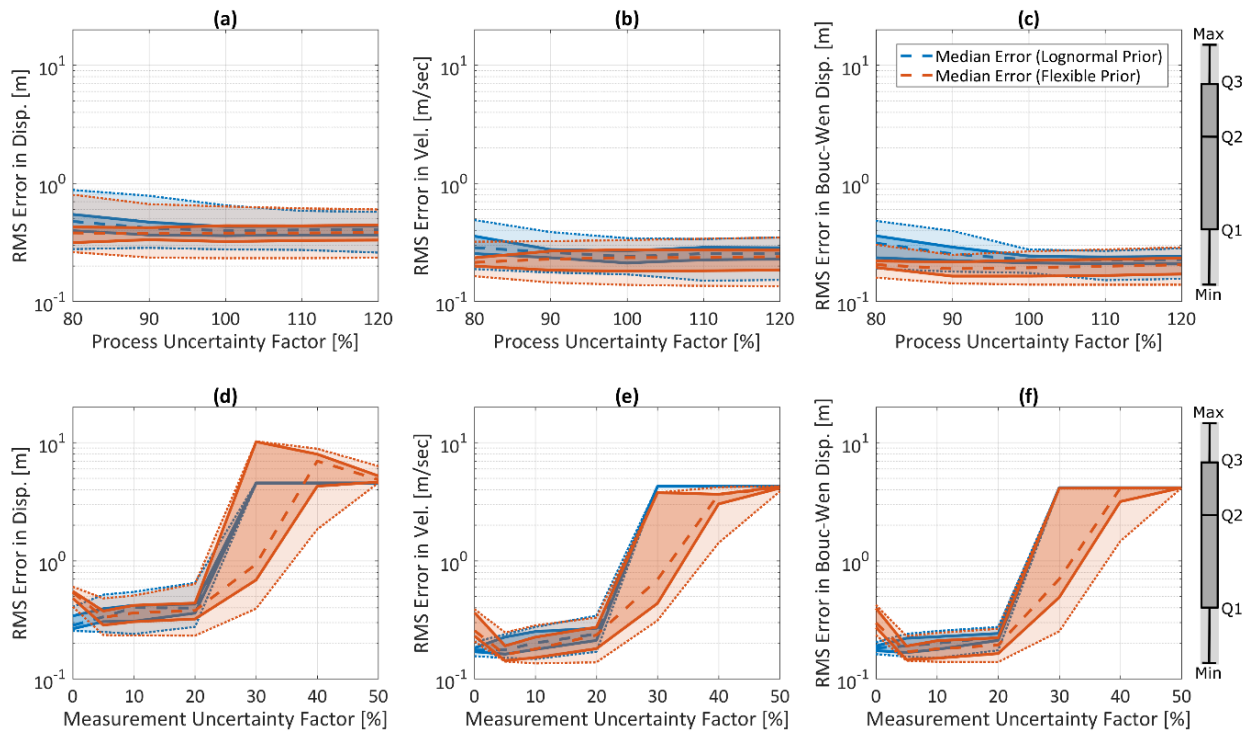
Fig. 7.13. Statistics of the RMS error in the states given the identified VI models when the (a-c) process noise uncertainty factor (λ_w) is varied, (d-f) measurement noise uncertainty factor ($v_{RMS-NSR}$) is varied.

Perhaps the most critical feature of these results, however, is their consistency. Though some specific features of the results change, and often improve, when a more informed prior is used, the underlying trends in the data remain the same. Fig. 7.12(c,d) shows this effect clearly with respect to the trends in model accuracy as measurement noise is increased. Regardless of the prior distribution on the parameters, the variational inference does not produce any models with

147

an accuracy of at least 50% for measurement noise levels with $v_{RMS-NSR} = 30\%$ or greater. These same trends are expressed in Fig. 7.13(d-f), where the RMS error on the states increases significantly, and nearly equivalently, between the two variational inference trials when $v_{RMS-NSR} = 30\%$ or greater. The response of the variational inference trials with respect to variations in the process noise is also similar, with both exhibiting consistently high levels of model accuracy and corresponding low levels of RMS error in the states. These results suggest that the variational inference method is fairly robust to even significantly different interpretations on the prior uncertainty in the parameters, which make it ideal for experimental settings where the limited knowledge is available concerning the parameters in advance of experimentation. In contrast, the UKF results in the previous sections express the dependence of this algorithm on the selection of the parameter priors, without any clear indication of which priors will be well-performing prior to inference.

7.3   Conclusions

In this chapter I have compared the ability of the UKF and variational inference methods to identify the hidden states and parameters of a simulated single degree-of-freedom Bouc-Wen system excited by a BLWN base motion. These inference approaches were used to conduct two parameter studies which assessed their accuracy and reliability with respect to variations on the assumptions of modeling error in the inference model and the level of measurement noise present in system response. An equivalent comparison of the algorithms in these parameter studies revealed that the UKF generates models with a greater number of outliers than the variational inference approach. Results from the modeling error study indicate that both algorithms result in similar rates of model accuracy when assumptions on the process noise are varied. When measurement noise levels are varied, variational inference only produces superior models to the UKF for moderate levels of measurement noise. Models produced using the UKF are more likely to give accurate results when measurement noise levels are extremely high. When the variational inference approach was reimplemented with more informative and flexible expressions of the prior distributions on the parameters, its overall performance was shown to increase, though the trends in its behavior remained consistent. Overall, the variational inference method was shown to be robust to the typical experimental conditions of moderate levels of measurement noise and general uncertainty in the level of model error.

The accuracy and robustness of the variational inference method does come at the cost of some increased computational time. Whereas the UKF executes on the order of seconds, variational inference requires execution times on the order of hours. One way in which this issue could be addressed is through the development of a filtering approach to variational inference, which would act on much smaller segments of data at a given time and therefore increase the optimization speed overall. This concept is explored in the following chapter.

7.4 References

1. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82,** 35–45 (1960).

2. Gordon, N. J., Salmond, D. J. & Smith, A. F. M. Novel approach to nonlinear / non-Gaussian Bayesian state estimation. *IEE Proc. (Radar Signal Process.* **140,** 107–113 (1993).

3. Doucet, A., Godsill, S. & Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10,** 197–208 (2000).

4. Snyder, C., Bengtsson, T., Bickel, P. & Anderson, J. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136,** 4629–4640 (2008).

5. Chatzi, E. & Smyth, A. Particle filter scheme with mutation for the estimation of time-invariant parameters in structural health monitoring applications. *Struct. Control Heal. Monit.* **20,** 1081–1095 (2013).

6. Olivier, A. & Smyth, A. W. Particle filtering and marginalization for parameter identification in structural systems. *Struct. Control Heal. Monit.* **24,** 1–25 (2017).

7. Eftekhar Azam, S. & Mariani, S. Dual estimation of partially observed nonlinear structural systems: A particle filter approach. *Mech. Res. Commun.* **46,** 54–61 (2012).

8. Blei, D. M., Kucukelbir, A. & Mcauliffe, J. D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **112,** 859–877 (2017).

9. Wu, M. & Smyth, A. W. Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Struct. Control Heal. Monit.* **14,** 971–990 (2007).

10. Chatzi, E. N. & Smyth, A. The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Struct. Control Heal. Monit.* **16,** 99–123 (2009).

11. Lund, A., Bilionis, I. & Dyke, S. J. Variational Inference for Nonlinear Structural Identification. *J. Appl. Comput. Mech.* (2020). doi:10.22055/jacm.2020.32626.2049

12. Herman, J. & Usher, W. SaLib: An open-source Python library for sensitivity analysis. *J. Open Source Softw.* **2,** (2017).

13. Øksendal, B. *Stochastic Differential Equations*. (Springer, 2003).

14. Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. (MIT Press, 2012).

8. AUTOMATIC DIFFERENTIATION VARIATIONAL FILTERING FOR NONLINEAR STRUCTURAL SYSTEM IDENTIFICATION

Though variational inference is a promising option for predictive structural modeling, its full potential in this application area cannot be realized until a variational filtering approach is developed for real-time inference on continuous data streams. Several theoretical approaches have been proposed to further the application of variational inference for online learning, but specific adaptations must be considered to suit the practical needs of the structural health monitoring community.

8.1 Review of Current Variational Filtering Approaches

The primary focus of variational inference for online learning of dynamical systems has been with respect to the Kalman filter and smoother. Ghahramani and Beal [1],[2] introduced an initial approach to variational Kalman smoothing on linear dynamical systems in 2001. In their approach, which they showed is analogous to the expectation-maximization algorithm, they identify the hidden states, parameters, and measurement noise by developing conjugate-exponential priors on the system terms and iteratively executing the variational optimization algorithm to develop the approximate posterior. Sarrka and Nummenmaa expanded on these ideas to extend the standard Kalman filtering approach to be noise adaptive [3]. In their method the predictive step of the filter remains analytical, and the corrector step is replaced with a variational inference update to allow for the inference of time-varying noise terms. Auvinen et al. [4] approached the generation of a variational Kalman filtering scheme by replacing the analytical expression for the posterior with the optimization of a maximum a posteriori (MAP) estimate. This approach has some similarity to that of Sarrka and Nummenmaa, but with the replacement of the standard variational objective, the ELBO, with the MAP estimate. Ait-El-Fquih and Hoteit [5] integrated variational inference directly with the prediction step of the Kalman filter, using it to accommodate diagonal and tri-diagonal covariance approximations for high-dimensional state-space models. Friston et al., in recognition of the constraints on the variational filtering approaches for complex dynamical systems, proposed a generalized theoretical framework called dynamic expectation maximization, which operates online on generalized system models to approximate the optimal posterior [6].

Though the method does have the advantage of being generalized, it was developed prior to the advent of black-box inference, and therefore requires significant skill from the user for implementation.

Other approaches to variational filtering have been proposed and shown to operate successfully on a number of theoretical examples. Most notably, Smidl and Quinn [7] introduced a generalized theoretical framework for variational Bayesian filtering in 2005, which applied variational inference in the Bayesian update step to infer the hidden states of generalized dynamical systems. Later work from these authors [8] introduced a variety of methods, with varying levels of computational speed and accuracy, to couple the variational Bayesian filter with the particle filter. Friston [9] approached variational filtering for dynamical systems from an energy perspective, by expressing the system dynamics in terms of the flow on a variational energy manifold. Most recently, Ait-El-Fquih and Hoteit [10] have adapted variational Bayesian filtering to couple with the multiple particle filtering approach, thereby increasing the accessibility of these methods to high-dimensional state-space systems. Though these methods have had some success, they do not take advantage of recent advances in black box and stochastic variational inference and have not been applied in experimental systems with high levels of uncertainty. In Section 8.2, I will adapt and expand upon these methods to generate an efficient, adaptable, and robust algorithm for real-time variational filtering on structural systems.

Outside of the specific focus on filtering methods for dynamical systems, several researchers have been addressing the challenges of incorporating continuous data streams into variational inference. This topic was introduced by Broderick et al. [11] who asserted that SVI could be adapted to a continuous data setting by returning to the classical view of Bayesian updating to produce a sequence of posteriors instead of determining a sequence of approximations to a fixed posterior on a fixed data set. Later works highlighted an issue with this approach, in that it fails to compensate for the tendency of variational inference to deliver overconfident posteriors. One of the strengths of variational inference is that it allows for the optimization of a simplified assertion of the posterior instead of calculating the true posterior exactly. This simplification can lead to artificially small variance estimates in the approximate posterior [12]. In a continuous data setting, this overconfidence can inhibit the adaptability of the inference to data which indicates changes to the condition of the system. McInerny et al. [13] and Jihan et al. [14] proposed some ideas to combat this issue. McInerny et al. approached the problem by refining the variational

152

objective to avoid overconfident posteriors. This 'population' posterior represents a reformulation of SVI to be scaled over the overall number of data points in the population of data instead of over the data in a fixed segment [13]. Jihan et al. approached the problem by propagating only the information concerning the expectation to future inference steps. The variance is kept constant at a predetermined value to allow for the posterior to adapt to drift in the system over time. Additionally, the authors scaled the batch based on the total number of samples seen at that time, thereby reducing the influence of future data and avoiding large swings in the posterior due to noise, instead of system drift [14]. These approaches have shown some success, particularly in operating with independent, identically distributed data sets. Further consideration is required for conditionally dependent data that would be likely in structural monitoring scenarios.

I use key ideas generated in these works to propose a simple variational filtering approach for the online identification of lumped-parameter structural systems. My approach is similar to current methods in that it clearly distinguishes between the predictor phase, which I achieve through Monte Carlo (MC) sampling for uncertainty propagation, and the corrector phase, which is I achieve through variational inference. In contrast with current approaches, my filter relaxes any constraints on the stochastic model of the system through the adoption of the distributional transformations used in automatic differentiation variational inference. Additionally, my approach introduces flexible filtering parameters such as the inference batch size, batch overlap length, and data downsampling ratio (d), to allow the user to efficiently adapt the filter to the particular structural health monitoring application. Though overconfident posteriors are not directly addressed by my approach, the flexible filtering parameters allow the user to mitigate some of the issues associated with them. The filter is explained in greater detail in Section 8.2 and applied to a simple numerical case study in Section 8.3 to develop a practical understanding of its efficacy for predictive structural modeling.

8.2 Novel Variational Filtering Approach for Structural Systems

I frame the problem of variational filtering from the perspective of inferring from small batches of data in a continuous stream. The size of the batch, $S \geq 1$, is a variable in the construction of the filter, but is assumed constant throughout its implementation. It is likewise assumed that the filter progresses to accommodate new data by moving from batch to batch with an overlap of $W \geq 0$ elements with the previously encountered data. Given data progressing at a fixed rate of $t_k = k\Delta t$

where $k = 1, 2, 3, \ldots, K$, the inference batch $\mathcal{B}_i$ represents the ordered list of time indices $\big(i \cdot (S - W), \; i \cdot (S - W) + 1, \ldots, (\; i \cdot (S - W) + S)\big)$, where $i = 0, 1, 2, \ldots, N$ and $N = floor\big(K - {}^S\!/_{S-W}\big) + 1$ is the total number of batches required to pass through the data set.

We begin the derivation by recalling the Bayesian filtering equations. As discussed in Chpt. 2, the purpose of Bayesian filtering is to compute the marginal posterior of the states $\mathbf{x}_k$ and parameters $\boldsymbol{\theta}$ at each time step given all the data available up to that point, $p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k})$. In each step, the marginal prior satisfies:

$$p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1}, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{k-1}) d\mathbf{x}_{k-1}, \quad (8.1)$$

where time-invariance is enforced on the parameters. The marginal posterior can then be stated as

$$p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k}, \mathbf{u}_{1:k}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k) p(\mathbf{x}_k, \boldsymbol{\theta} | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k})}. \quad (8.2)$$

An equivalent logical structure must be present in a variational filtering approximation. To accomplish this succinctly, I first develop a MC approximation of the marginal prior (predictor phase) and use the resulting distributional parameters to inform the starting point for the variational inference of the posterior (corrector phase). This general structure for the filtering approximation is shown in Fig. 8.1.

Approximate Marginal Prior   **Minimize Posterior Loss**
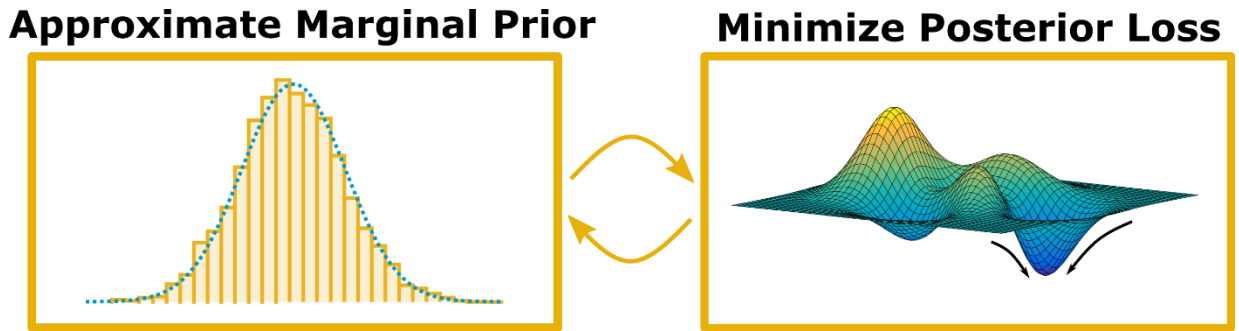
Fig. 8.1 – The predictive phase in variational filtering is accomplished through an MC sampling approximation. The corrector phase is achieved through variational inference, with the predicted marginal prior as a starting point.

Adopting an overlapping batch framework for progressing through the filter allows the user some flexibility in manipulating the respective inference accuracy and optimization speed. For example, larger batch sizes typically yield greater accuracy in per-batch inference operations, due

to greater access to data. However, they can require longer optimization times due to the increased number of states being inferred simultaneously. Likewise, larger overlaps between batches may allow for greater predictive accuracy among the overlapped states, which are partially optimized in the previous corrector phase, and the new states, which are projected forward in time from the results of the previous corrector phase. Projecting fewer new states forward in time reduces the potential for large deviations from the true state in the predictor phase, thereby speeding the optimization process in the corrector phase. This flexibility in filter construction allows the user to tune the operation of the filter to problem-specific needs, while still maintaining the speed and accuracy required by the application.

      With these preliminaries in place, I define the variational filtering algorithm. The algorithm is initialized by asserting priors on the initial condition of the system states $p(\mathbf{x}_0)$, the values of the global system parameters $p(\boldsymbol{\theta})$, the transition model $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1})$, and the emission model $p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k)$. The transition and emission models are assumed to remain constant throughout the inference process. Though these models can take any form, it is convenient and descriptive to represent them as normally distributed, according to

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}) = N\left(\mathbf{x}_k|\mathbf{x}_{k-1} + \int_{t_{k-1}}^{t_k} F\big(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)\big)\, dt, \mathbf{Q}_{k-1}\right) \tag{8.3}$$

$$= N(\mathbf{x}_k|f(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}), \mathbf{Q}_{k-1}) \tag{8.4}$$

$$p(\mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k) = N(\mathbf{y}_k|h(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_k), \mathbf{R}_k), \tag{8.5}$$

where $F(\cdot)$ is the deterministic ODE describing the transition dynamics on the states, $f(\cdot)$ is the deterministic model for the assumed progression of the states in continuous time, $h(\cdot)$ is the deterministic observation model, $\mathbf{Q}_k$ is the covariance on the transition probability, and $\mathbf{R}_k$ is the covariance on the emission probability. It is common to use $f(\cdot)$ to approximates the integral in Eq. (8.3) numerically, through approaches such as Euler or Runge Kutta integration. Herein, the numerical integration approach is deliberately left unspecified, as different accuracies may be required for different applications. It is important to note however, that the discretization step, Δt, of the integration approach will have a significant impact on the speed of the resulting filter. In some instances, it may be beneficial to subsample the data for use with the corrector phase, which is the variational inference portion of the analysis. Further details on this filter adaptation are given in Section 8.2.2.

Given the prior information, the marginal priors over all states in the filtering batch, $\mathbf{x}_{\mathcal{B}_i}$, and the parameters, $\boldsymbol{\theta}$, can be defined as

$$p\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) = p\left(\mathbf{x}_{\mathcal{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) p\left(\boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) \quad (8.6)$$

where $\mathcal{B}_{i-1}(S)$ refers to the final element in the ordered list represented by $\mathcal{B}_{i-1}$. The vectors $\mathbf{u}_{1:\mathcal{B}_{i-1}(S)}$ and $\mathbf{y}_{1:\mathcal{B}_{i-1}(S)}$ therefore respectively represent all control inputs and all observation data from the beginning of the observation history up through the end of batch $\mathcal{B}_{i-1}$. The marginal prior on the states can be further specified as

$$p\left(\mathbf{x}_{\mathcal{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) =$$

$$\int \prod_{k=\mathcal{B}_i(1)}^{\mathcal{B}_i(S)} p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}) \, p\left(\mathbf{x}_{\mathcal{B}_{i-1}(S-W)} | \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) d\mathbf{x}_{\mathcal{B}_{i-1}(S-W)}. \quad (8.7)$$

The marginal posterior is then given by

$$p\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) =$$

$$\frac{p\left(\mathbf{y}_{\mathcal{B}_i} | \mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}, \mathbf{u}_{\mathcal{B}_i}\right) p\left(\mathbf{x}_{\mathcal{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) p\left(\boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right)}{\iint p\left(\mathbf{y}_{\mathcal{B}_i} | \mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}, \mathbf{u}_{\mathcal{B}_i}\right) p\left(\mathbf{x}_{\mathcal{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) p\left(\boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) d\mathbf{x}_{\mathcal{B}_i} d\boldsymbol{\theta}}. \quad (8.8)$$

Equations (8.6) and (8.8) respectively form the predictor and corrector equations for the batch filter. As noted in Chpt. 2, the computational difficulties that arise in the integration promote the use of approximate techniques to develop the marginal prior and posterior. In this case, the approximate marginal posterior is developed through variational inference. I therefore assert a generalized variational family, or guide, over the batch of

$$g_{\mathcal{B}_i}\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi}\right) = g_{\mathcal{B}_i}\left(\mathbf{x}_{\mathcal{B}_i}; \boldsymbol{\phi}_x\right) g_{\mathcal{B}_i}(\boldsymbol{\theta}; \boldsymbol{\phi}_\theta), \quad (8.9)$$

such that the states and parameters are assumed independent. The guide is then optimized using automatic differentiation variational inference (ADVI) as described in Section 3.2, such that approximate marginal posterior is

$$\hat{g}_{\mathcal{B}_i}\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi}\right) = \underset{g \in G}{\arg\min} \, \mathrm{KL}\left(g_{\mathcal{B}_i}\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi}\right) \| p\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right)\right). \quad (8.10)$$

where the variational parameters $\boldsymbol{\phi}$ are initialized according to distributional parameters determined from the approximate marginal priors on the states and parameters. The selection of a guide, $g_{B_i}\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi}\right)$, that is sufficiently flexible to capture the true posterior is a critical component of this optimization. For structural predictive modeling, a normal distribution has been

found to be a suitable selection when coupled with the distributional transformations on the parameters afforded by ADVI, which can constrain the optimization on the parameter space to physical values. Other distributional forms may allow even greater flexibility, such as the ability to capture multi-modal posteriors, and are a promising topic for further study. The number of optimization steps, O, is the other key variable in this process. As O increases, the approximate posterior, $\hat{g}_{\mathcal{B}_i}(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi})$, will converge to a local optimum of the solution space. If O is too small, $\hat{g}_{\mathcal{B}_i}(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi})$ will project a sub-optimal approximation of the marginal posterior, which will then be propagated to subsequent filtering steps and could quickly lead to divergence in the filter. Large values of O can also be problematic, sometimes leading to computational inefficiency. However, some strategies, including variance reduction techniques, exist which can reduce the need for large values of O and therefore increase the efficiency of the algorithm. The user should select an O that is sufficiently small for accurate, online variational filtering in the chosen application.

The marginal prior for each batch is developed by extracting and extending the information available in the marginal posterior through sampling. Due to the assumed independence of the states and parameters, and the nature of the parameters in this problem as constants, I define the terms $p\left(\mathbf{x}_{\boldsymbol{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right)$ and $p\left(\boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right)$ separately. The approximate marginal prior over the parameters for the batch, $\mathcal{B}_i$, can be expressed simply as

$$\hat{p}\left(\boldsymbol{\theta} | \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) = \hat{g}_{\mathcal{B}_{i-1}}(\boldsymbol{\theta}; \boldsymbol{\phi}_\theta). \tag{8.11}$$

To develop the marginal prior over the states, I construct a multivariate normal approximation of the distribution of the states over the batch which I fit by sampling and propagating uncertainty from the posterior of the previous batch. The approximate marginal prior over the states can then be given the representative form

$$\hat{p}\left(\mathbf{x}_{\boldsymbol{B}_i} | \boldsymbol{\theta}, \mathbf{u}_{1:\mathcal{B}_{i-1}(S)}, \mathbf{y}_{1:\mathcal{B}_{i-1}(S)}\right) = N\left(\mathbf{x}_{\boldsymbol{B}_i} | \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}, \widehat{\boldsymbol{\Sigma}}_{\mathcal{B}_i}\right). \tag{8.12}$$

As discussed earlier in this section, an overlap of $W \geq 0$ is included between batches to allow for the potential of increased inference accuracy and speed as the filter progresses. Batch $\mathcal{B}_i$ therefore includes W states from the previous batch, $\mathcal{B}_{i-1}$, which have already been corrected with refence to their corresponding observations, and $(S - W)$ states that have as yet only been seen in the current batch and have not undergone the corrector phase of inference. The mean of the approximate marginal prior, $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}$, can therefore most effectively take advantage of the available

information through a partition of the mean between the overlapped states, $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(1:W)$, which can be translated directly from the marginal posterior as

$$\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(j) = \mathbb{E}\left[\hat{g}_{\mathcal{B}_{i-1}}\left(\mathbf{x}_{\mathcal{B}_i(S-W+(j-1))}; \boldsymbol{\phi}\right)\right], \tag{8.13}$$

with $j = 1, \ldots, W$ and the new states, $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(W+1:S)$, which are developed by propagating MC samples from $\hat{g}_{\mathcal{B}_{i-1}}\left(\mathbf{x}_{\mathcal{B}_i(S-1)}; \boldsymbol{\phi}\right)$ to form a prediction on the remaining states in the current batch, $\mathbf{x}_{\mathcal{B}_i(W+1:S)}$. The mean on the new states in the current batch can therefore be expressed as

$$\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(j) = \frac{1}{P}\sum_{p=1}^{P} f\left(\hat{\mathbf{x}}_{\mathcal{B}_i(j-1)}^{(p)}, \widehat{\boldsymbol{\theta}}^{(p)}, \mathbf{u}_{\mathcal{B}_i(j-1)}\right), \tag{8.14}$$

with $j = W+1, \ldots, S$, where P samples from $\hat{g}_{\mathcal{B}_{i-1}}\left(\mathbf{x}_{\mathcal{B}_i(S-1)}; \boldsymbol{\phi}\right)$ and $\hat{g}_{\mathcal{B}_{i-1}}(\boldsymbol{\theta}; \boldsymbol{\phi}_\theta)$ are propagated through the state transition equations to develop the mean of the approximate marginal prior. Note that in the case of the first batch, $\mathcal{B}_0$, no data from a previous corrector phase is available and the marginal prior on the states will be approximated solely from the forward propagation of MC samples on the priors $p(\mathbf{x}_0)$ and $p(\boldsymbol{\theta})$.

The statement of the covariance approximation is directly related to the choice of structure of the assumed variational family, $g_{\mathcal{B}_i}\left(\mathbf{x}_{\mathcal{B}_i}, \boldsymbol{\theta}; \boldsymbol{\phi}\right)$. Though mean-field variational families, which use a diagonal covariance structure, are common to many variational inference applications, tri-diagonal covariance structures have been found to better represent the Markovian assumptions in dynamical systems [15],[16], and therefore lead to more accurate inference. Each of these structures scale well with time, as shown in Fig. 8.2(a-b), leading to equivalent representations of the posterior uncertainty regardless of batch size, window overlap, or data length. Approximations to the full covariance matrix are also possible with variational filtering, though they do not necessarily scale well in time. For instance, a sparse approximation of the full covariance can be made over each batch by enforcing a decomposition, **L**, of the covariance matrix which contains only the first column and main diagonal. The resulting approximation scales linearly with batch size, similar to the tri-diagonal approximation, but its density in specifying the full covariance of the states over time is dependent on batch size and window overlap, as shown in Fig. 8.2(c).

For the purpose of predictive structural modeling, I have found that a tridiagonal structure provides a sufficiently accurate representation of the uncertainty. As with the mean of the marginal prior, $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}$, the covariance of the marginal prior, $\widehat{\boldsymbol{\Sigma}}_{\mathcal{B}_i}$, can be partitioned to reflect the W states

inferred in the previous batch, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mathcal{B}}_i}(1{:}W, 1{:}W)$, and the $(S-W)$ remaining states that have not yet been corrected against observed data, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mathcal{B}}_i}(W+1{:}S, W+1{:}S)$. The covariance of the marginal prior over the initial subset can therefore be expressed as

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mathcal{B}}_i}(1{:}W, 1{:}W) = \mathbb{V}\big[\hat{g}_{\boldsymbol{\mathcal{B}}_{i-1}}\big(\mathbf{x}_{\mathcal{B}_i(S-W{:}S-1)}; \boldsymbol{\phi}\big)\big], \tag{8.15}$$

where $\mathbb{V}[\cdot]$ denotes the variance operator. The covariance of the second subset of data is determined through sequential sampling approximations of the covariance, as

$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mathcal{B}}_i}(j-1{:}j, j-1{:}j) =$

$$\frac{1}{P}\sum_{p=1}^{P} \otimes\left(\begin{bmatrix} F\left(\hat{\mathbf{x}}^{(p)}_{\mathcal{B}_i(j-1)}, \widehat{\boldsymbol{\theta}}^{(p)}, \mathbf{u}_{\mathcal{B}_i(j-1)}\right) - \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(j-1) \\ F\left(\hat{\mathbf{x}}^{(p)}_{\mathcal{B}_i(j)}, \widehat{\boldsymbol{\theta}}^{(p)}, \mathbf{u}_{\mathcal{B}_i(j)}\right) - \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(j) \end{bmatrix}\right), \quad j = W+1, \dots, S \tag{8.16}$$

where the operator $\otimes(\cdot)$ denotes the outer product of the vector with itself. It is important to note that the defined covariance structure on the uncorrected states in the current batch, $\boldsymbol{\mathcal{B}}_i(W+1{:}S)$, ensures a coherent, positive-definite structure as well as a tri-diagonal covariance approximation. Other options exist by which one could obtain a tridiagonal covariance structure from a batch of samples, but I have found these alternative truncation techniques to be either more computationally intensive or less representative to the true uncertainty. As with the mean of the marginal prior for the first batch, $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\mathcal{B}}_0}$, the covariance of the marginal prior over the first batch, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\mathcal{B}}_0}$, is of necessity entirely developed through forward propagation of MC samples on the priors $p(\mathbf{x}_0)$ and $p(\boldsymbol{\theta})$.
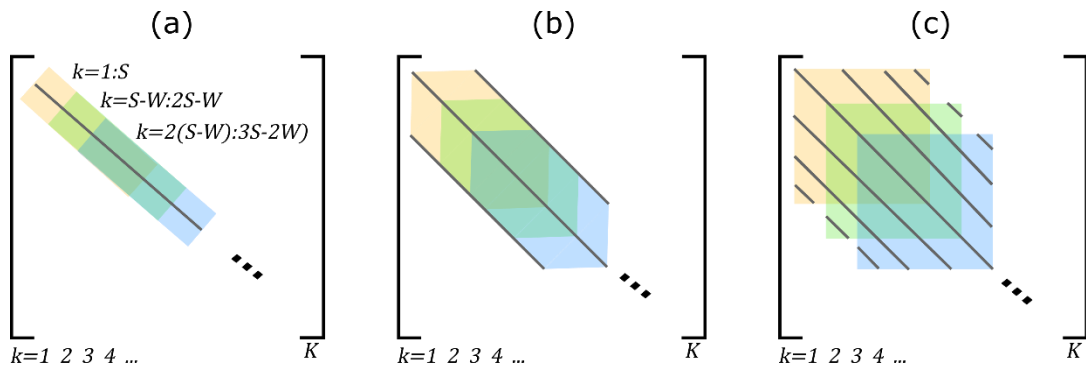
Fig. 8.2 – Development of the covariance structure of a single state across filtering steps. (a) Diagonal covariance (b) Tridiagonal covariance (c) Full batch covariance

159

8.2.1 Distributional Transformation in the Filtering Context

For the majority of initial prior distributions on the states and the parameters, the progression of the filter is fairly straightforward. However, the introduction of distributional transformations, as discussed in Section 3.2, can add some additional complexity to the marginal priors across filtering steps. Take, for example, the assumption of a uniform prior over a particular parameter ξ, which is asserted as independent from the remaining states and parameters. This parameter would have the initial prior distribution

$$p_{\boldsymbol{B}_0}(\xi) = \frac{1}{b-a} 1_{[a,b]}(\xi), \tag{8.17}$$

where $1_{[a,b]}(\xi)$ is a function that outputs a value of 1 when $\xi \in [a,b]$ and 0 otherwise, thereby representing the support of the distribution. The constants a and b are the left and right bounds on the uniform distribution, respectively. The parameter is then transformed according to

$$\hat{\xi} = \Upsilon(\xi) = \ln\left(\frac{\xi-a}{b-\xi}\right) \tag{8.18}$$

to have full support on the real numbers, such that the transformed parameter ϕ_ξ can be easily optimized. This results in a prior distribution on ϕ_ξ of

$$p_{\boldsymbol{B}_0}(\hat{\xi}) = \frac{e^{\phi_\xi}}{\left(1 + e^{\phi_\xi}\right)^2}. \tag{8.19}$$

Following the procedure in Section 3.2, a normal distribution is used to represent $g_{\boldsymbol{B}_0}(\hat{\xi}; \phi_{\hat{\xi}})$ for optimization, with the variational parameters ϕ_ξ describing the mean μ_ξ and the standard deviation σ_ξ. The optimal distribution $\hat{g}_{\boldsymbol{B}_0}(\xi; \phi_\xi)$ can then be described by an inverse transformation $\Upsilon^{-1}(\phi_\xi)$ into the support of the prior

$$\hat{g}_{\boldsymbol{B}_0}(\xi; \phi_\xi) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(-\frac{\left(\ln\left(\frac{\xi-a}{b-\xi}\right) - \mu_\xi\right)^2}{2\sigma_\xi^2}\right) \frac{b-a}{(\xi-a)(b-\xi)} 1_{[a,b]}(\xi). \tag{8.20}$$

Note that this approximate posterior is characterized by four parameters, a, b, μ_ξ, and σ_ξ, but only μ_ξ and σ_ξ are optimized as a and b are facets of the prior, $p_{\boldsymbol{B}_0}(\xi)$. In subsequent filtering steps, this posterior will become the new prior on the parameter ξ, $\hat{p}_{\boldsymbol{B}_i}(\xi) = \hat{g}_{B_{i-1}}(\xi; \phi_\xi)$, be transformed to have full support on the states, $\hat{p}_{\boldsymbol{B}_i}(\lambda_\xi) = N(\phi_\xi | \mu_\xi, \sigma_\xi)$, be optimized to update

the distributional parameters μ_ξ, and σ_ξ, and finally be transformed back to the support of the original prior, $\xi \in [a, b]$, with an equivalent distributional form to Eq. (8.20). Through this example, we can see that though the distributional transformations suggested by the ADVI implementation [12] do initially increase the complexity of the filter execution, the approximate posterior both retains the support of the prior throughout the filter implementation and stabilizes in complexity after the second filtering step.

8.2.2 Downsampling to Increase Filter Speed

The inference rate implied by the filtering algorithm given in Section 8.2 is sufficient for systems requiring a relatively low sampling rate. In this case, the speed of the optimization is fairly well balanced by the flow of incoming data. When higher sampling rates are required, the filtering method can be adapted to maintain a throughput balanced with incoming data rates. One option that has the potential to address this issue is the reduction in inference parameters at each filtering batch through multi-step integration in the calculation of the ELBO.

Following the assumptions which led to the decomposition of the probability model in Eq. (3.2), the components of the ELBO from Eq. ((8.21)3.8), can be more explicitly expressed as

$$
\begin{aligned}
\text{ELBO}(g) &= \sum_{k=1}^{K} \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log p\left(\mathbf{y}_k | \Upsilon^{-1}\left(\left[S_{\phi_{\hat{\mathbf{x}}_k}}^{-1}(\eta_{\hat{\mathbf{x}}_k}), S_{\phi_{\hat{\theta}}}^{-1}(\eta_{\hat{\theta}}) \right] \right), \mathbf{u}_k \right) \right] \\
&+ \sum_{k=1}^{K} \left(\mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log p\left(\Upsilon_{\phi_{\hat{\mathbf{x}}_k}}^{-1}\left(\left[S_{\phi_{\hat{\mathbf{x}}_k}}^{-1}(\eta_{\hat{\mathbf{x}}_k}) \right] \right) | \Upsilon_{\phi_{\hat{\mathbf{x}}_{k-1}}}^{-1}\left(\left[S_{\phi_{\hat{\mathbf{x}}_{k-1}}}^{-1}(\eta_{\hat{\mathbf{x}}_{k-1}}) \right] \right), \Upsilon_{\phi_{\hat{\theta}}}^{-1}\left(\left[S_{\phi_{\hat{\theta}_n}}^{-1}(\eta_{\hat{\theta}_n}) \right] \right), \mathbf{u}_k \right) \right] \right. \\
&+ \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log \left| \det J_{\Upsilon_{\phi_{\hat{\mathbf{x}}_k}}^{-1}}\left(\left[S_{\phi_{\hat{\mathbf{x}}_k}}^{-1}(\eta_{\hat{\mathbf{x}}_k}) \right] \right) \right| \right] \Big) \\
&+ \sum_{m=1}^{M} \left(\mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log p\left(\Upsilon_{\phi_{\hat{\mathbf{x}}_{0,m}}}^{-1}\left(\left[S_{\phi_{\hat{\mathbf{x}}_{0,m}}}^{-1}(\eta_{\hat{\mathbf{x}}_{0,m}}) \right] \right) | \mathbf{u}_0 \right) \right] + \right. \\
&\mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log \left| \det J_{\Upsilon_{\phi_{\hat{\mathbf{x}}_{0,m}}}^{-1}}\left(\left[S_{\phi_{\hat{\mathbf{x}}_{0,m}}}^{-1}(\eta_{\hat{\mathbf{x}}_{0,m}}) \right] \right) \right| \right] \Big) \\
&+ \sum_{n=1}^{N} \left(\mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log p\left(\Upsilon_{\phi_{\hat{\theta}_n}}^{-1}\left(\left[(\eta_{\hat{\theta}_n}) \right] \right) \right) \right] + \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)} \left[\log \left| \det J_{\Upsilon_{\phi_{\hat{\theta}_n}}^{-1}}\left(\left[S_{\phi_{\hat{\theta}_n}}^{-1}(\eta_{\hat{\theta}_n}) \right] \right) \right| \right] \right) \\
&- \mathbb{E}_{g(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})} \left[\log g(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi}) \right]. \quad (8.21)
\end{aligned}
$$

Looking at the components of Eq. (8.21), it can be seen that the expectation of the log likelihood, $\mathcal{L}(\mathbf{y}_k)$, is given by the first summation, the expectation of the log transition probability, $\mathcal{T}(\mathbf{x}_k, \mathbf{u}_k)$, is given by the second summation and subsequent expectation, the expectation on the log priors,

$\mathcal{P}(\boldsymbol{\theta})$, is given by the third summation and fourth summations and subsequent expectations, and the entropy, $\mathbb{H}(\mathbf{x}, \boldsymbol{\theta})$ is given by $-\mathbb{E}_{g(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})}\left[\log g(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})\right]$. To increase the efficiency of the optimization process while still retaining the same data availability in each batch, the number of dynamic states in the batch can be reduced through downsampling. This process only affects the computation of the expectations of the log likelihood, $\mathcal{L}(\mathbf{y}_k)$, and log transition probability, $\mathcal{T}(\mathbf{x}_k, \mathbf{u}_k)$. In their reduced state form, these terms can be expressed as

$$\mathcal{L}^{(d)}(\mathbf{y}_k) = \sum_{k=1}^{K/d} \mathbb{E}_{N(\boldsymbol{\eta}; \mathbf{0}, I)}\left[\log p\left(\mathbf{y}_{dk} | \Upsilon^{-1}\left(\left[S^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}(\eta_{\hat{\mathbf{x}}_{dk}}), S^{-1}_{\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}}(\eta_{\hat{\boldsymbol{\theta}}})\right]\right)\right)\right] \tag{8.22}$$

and

$$\mathcal{T}^{(d)}(\mathbf{x}_k, \mathbf{u}_k) =$$

$$\sum_{k=1}^{K/d} \mathbb{E}_{N(\boldsymbol{\eta}; \mathbf{0}, I)}\left[\log p\left(\Upsilon^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}\left(\left[S^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}(\eta_{\hat{\mathbf{x}}_{dk}})\right]\right) | \Upsilon^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{d(k-1)}}}\left(\left[S^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{d(k-1)}}}\left(\eta_{\hat{\mathbf{x}}_{d(k-1)}}\right)\right]\right), \Upsilon^{-1}_{\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}}([(\eta_{\hat{\boldsymbol{\theta}}})])\right)\right]$$

$$+ \mathbb{E}_{N(\boldsymbol{\eta}; \mathbf{0}, I)}\left[\log\left|\det J_{\Upsilon^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}}\left(\left[S^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}(\eta_{\hat{\mathbf{x}}_{dk}})\right]\right)\right|\right], \tag{8.23}$$

where d is an integer representing the down-sampling ratio on the states. To preserve the accuracy of the computation of expectation of the log-transition probability with respect to the known data and process noise covariance, which is a function of several factors including the sampling rate, I perform multi-step integration over the states, taking advantage of the excitation data available at a higher sampling, $\mathbf{u}_k$, rate while only generating states for inference every d time steps.

Recall that $\mathbf{x}_{dk} = \Upsilon^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}\left(\left[S^{-1}_{\boldsymbol{\phi}_{\hat{\mathbf{x}}_{dk}}}(\eta_{\hat{\mathbf{x}}_{dk}})\right]\right)$ and $\boldsymbol{\theta}$. My approach is to redefine the transition probability from Eq. (8.3) as

$$p\left(\mathbf{x}_{dk} | \mathbf{x}_{d(k-1)}, \boldsymbol{\theta}, \mathbf{u}_{d(k-1):dk-1}\right) = N\left(\mathbf{x}_{dk} | F\left(\mathbf{x}_{d(k-1)}, \boldsymbol{\theta}, \mathbf{u}_{d(k-1):dk-1}\right), \mathbf{Q}_{d(k-1)}\right) \tag{8.24}$$

where

$$F\left(\mathbf{x}_{d(k-1)}, \boldsymbol{\theta}, \mathbf{u}_{d(k-1):dk-1}\right) = \int_{d \cdot t_{k-1}}^{d \cdot t_k} f\left(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)\right) dt + F\left(\mathbf{x}_{d(k-1)}, \boldsymbol{\theta}, \mathbf{u}_{d(k-1):dk-2}\right);$$

$$F\left(\mathbf{x}_{d(k-1)}, \boldsymbol{\theta}, \mathbf{u}_{d(k-1)}\right) = \int_{d \cdot t_{(k-1)}}^{d \cdot t_{(k-1)} + \Delta t} f\left(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)\right) dt + \mathbf{x}_{d(k-1)}. \tag{8.25}$$

This recursive function can be integrated with any sufficiently accurate numerical scheme. For the examples examined in this study a fixed-step Runge-Kutta method is used, operating at the same sampling rate as the data collection. Note that the transition covariance, $\mathbf{Q}_{d(k-1)}$, is not augmented

162

in this computation, as the accuracy of the integration is retained. The emission probability calculation exists solely at the lower sampling rate, with respect to the transition samples $\mathbf{x}_{dk}$, as

$$p(\mathbf{y}_{dk}|\mathbf{x}_{dk}, \boldsymbol{\theta}, \mathbf{u}_{dk}) = N(\mathbf{y}_{dk}|h(\mathbf{x}_{dk}, \boldsymbol{\theta}, \mathbf{u}_{dk}), \mathbf{R}_{dk}). \tag{8.26}$$

The measurement noise covariance, $\mathbf{R}_{dk}$, is likewise left equivalent to $\mathbf{R}_k$, as the measurement noise is assumed to be unaffected by down-sampling during post-processing.

To support these inference operations, the computation of the marginal prior must develop the uncertainty over the states at the same reduced sampling rate. This can be achieved while ensuring the computational accuracy afforded by evaluating the data at the sampling rate used for monitoring. The computation of the mean of the marginal prior $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}$ is initially performed equivalently as in Eq. (8.13) and Eq. (8.14). The down-sampled mean, $\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i/d}$, is then generated by selecting sample means at the instances k imposed by the lower sampling rate, such that

$$\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i/d} = \left[\widehat{\boldsymbol{\mu}}_{\mathcal{B}_i(0)}, \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i(d)}, \dots, \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i(S)}\right], \tag{8.27}$$

resulting in a batch size of length S/d. The covariance over these states is then computed according to Eq. (8.16), such that $\widehat{\boldsymbol{\Sigma}}_{\mathcal{B}_i/d}$ is generated as

$$\widehat{\boldsymbol{\Sigma}}_{\mathcal{B}_i/d}(j-1:j,j-1:j) =$$

$$\frac{1}{P}\sum_{p=1}^{P} \otimes \left(\begin{bmatrix} F\left(\widehat{\mathbf{x}}_{\mathcal{B}_i(dj-d)}^{(p)}, \widehat{\boldsymbol{\theta}}^{(p)}, \mathbf{u}_{\mathcal{B}_i(dj-d)}\right) - \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(dj-d) \\ F\left(\widehat{\mathbf{x}}_{\mathcal{B}_i(dj)}^{(p)}, \widehat{\boldsymbol{\theta}}^{(p)}, \mathbf{u}_{\mathcal{B}_i(j)/d}\right) - \widehat{\boldsymbol{\mu}}_{\mathcal{B}_i}(dj) \end{bmatrix}\right), \tag{8.28}$$

with $j = W+1, \dots, S/d$. An analysis and discussion of this approach to increasing inference speed is included in Section 8.3.4.

8.3 Numerical Study: Linear Single-Degree-of-Freedom System

A simple numerical case study serves well to illustrate the properties of the variational filtering approach. Here, I use a simulated single-degree-of-freedom (SDOF) shear frame structure subject to base excitation to demonstrate the method and investigate the impact of batch size, batch overlap length, and down-sampling ratio. The selected system, excitation, and response are described in Section 8.3.1. Section 8.3.2 describes a base comparison between variational filtering and variational inference with respect to this system. The remainder Section 8.3 is used to evaluate the impact of the various filtering parameters.

8.3.1 System Definition

The SDOF oscillator used for this illustrative example can be described by the equation of motion

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = -m\ddot{x}_g(t), \tag{8.29}$$

where $m = 25$ kg, $c = 50$ Ns/m, and $k = 1 \times 10^4$ N/m. As in Chpt. 7, the system is excited through ground motion, represented by the acceleration $\ddot{x}_g(t)$, and its response is described by its displacement, $x(t)$, and velocity, $\dot{x}(t)$. For the purposes of this example, I assume that only the absolute acceleration is observed, such that

$$y(t) = -\frac{1}{m}\left(c\dot{x}(t) + kr(t)\right). \tag{8.30}$$

To aid in algorithm convergence, the system is nondimensionalized using the time and length scales $t_c = 3$ Hz and $x_c = 0.05$ m, respectively. These scales are chosen as fixed constants within feasible realm of system properties which are assumed unknown to the user, where t_c is a reflection of a possible natural frequency and x_c is a reflection of the maximum displacement. The nondimensionalized transition and observation equations can therefore be given as

$$\ddot{\bar{x}}(\tau) + \frac{2\omega_n\xi}{\omega_c}\dot{\bar{x}}(\tau) + \frac{\omega_n^2}{\omega_c^2}\bar{x}(\tau) = -\frac{\ddot{x}_g(\tau/\omega_c)}{x_c\omega_c^2}, \tag{8.31}$$

$$\bar{y}(t) = -\frac{2\omega_n\xi}{\omega_c}\dot{\bar{x}}(\tau) - \frac{\omega_n^2}{\omega_c^2}\bar{x}(\tau), \tag{8.32}$$

where $\omega_n = 20$ rad/sec (3.18 Hz) and $\xi = 0.05$ are the 'unknown' parameters that we wish to infer. The inference problem then becomes one of inferring the states $\mathbf{z} = [\bar{\mathbf{x}}, \dot{\bar{\mathbf{x}}}] = [\mathbf{x}_1, \mathbf{x}_2]$ and the parameters $\boldsymbol{\theta} = [\omega_n, \xi] = [\theta_1, \theta_2]$. The transition and emission probabilities on this system are set as fixed for this example, resulting in the stochastic dynamical equations

$$p(x_1(\tau)|\mathbf{z}(\tau-1), \boldsymbol{\theta}) = N\left(x_1(\tau)| \text{RK4}\left(x_2(\tau-1)\right), \mathbf{Q}_1\right), \tag{8.33}$$

$$p(x_2(\tau)|\mathbf{z}(\tau-1), \boldsymbol{\theta}) =$$

$$N\left(x_2(\tau)\,|\text{RK4}\left(-\frac{2\theta_1\theta_2}{\omega_c}x_2(\tau-1) - \frac{\theta_1^2}{\omega_c^2}x_1(\tau-1) - \frac{\ddot{x}_g(\tau/\omega_c)}{x_c\omega_c^2}\right), \mathbf{Q}_2\right). \tag{8.34}$$

$$p(y(\tau)|\mathbf{z}(\tau), \boldsymbol{\theta}) = N\left(y(\tau)| -\frac{2\theta_1\theta_2}{\omega_c}x_2(\tau) - \frac{\theta_1^2}{\omega_c^2}x_1(\tau), \mathbf{R}\right), \tag{8.35}$$

where

$$\mathbf{Q}_i = \left(\sqrt{\Delta\tau} \cdot 0.01 \cdot \text{RMS}(\mathbf{z}_i)\right)^2, \tag{8.36}$$

representing a 1% root-mean-square (RMS) uncertainty in the transition dynamics coupled with the uncertainty associated with discretizing the continuous ODE [17],

$$\mathbf{R} = \left(0.05 \cdot \text{RMS}(\bar{\mathbf{y}})\right)^2, \tag{8.37}$$

representing a 5% RMS uncertainty in the observation signal, and RK4$(\cdot)$ represents the numerical integration of the transition dynamics from $(\tau - 1)$ to τ using the fixed-step, fourth order Runge Kutta method at a time step of $\Delta\tau$.

Given the stochastic dynamical equations, the prior distributions on the inferred states are set to

$$x_i(0) \sim N(0, 0.25^2), \tag{8.38}$$

reflecting the knowledge that the system is at rest prior to excitation. The priors on the parameters are set as

$$\xi \sim Uniform(0,1), \tag{8.39}$$

$$\omega_n \sim Lognormal(3.6, 0.5^2), \tag{8.40}$$
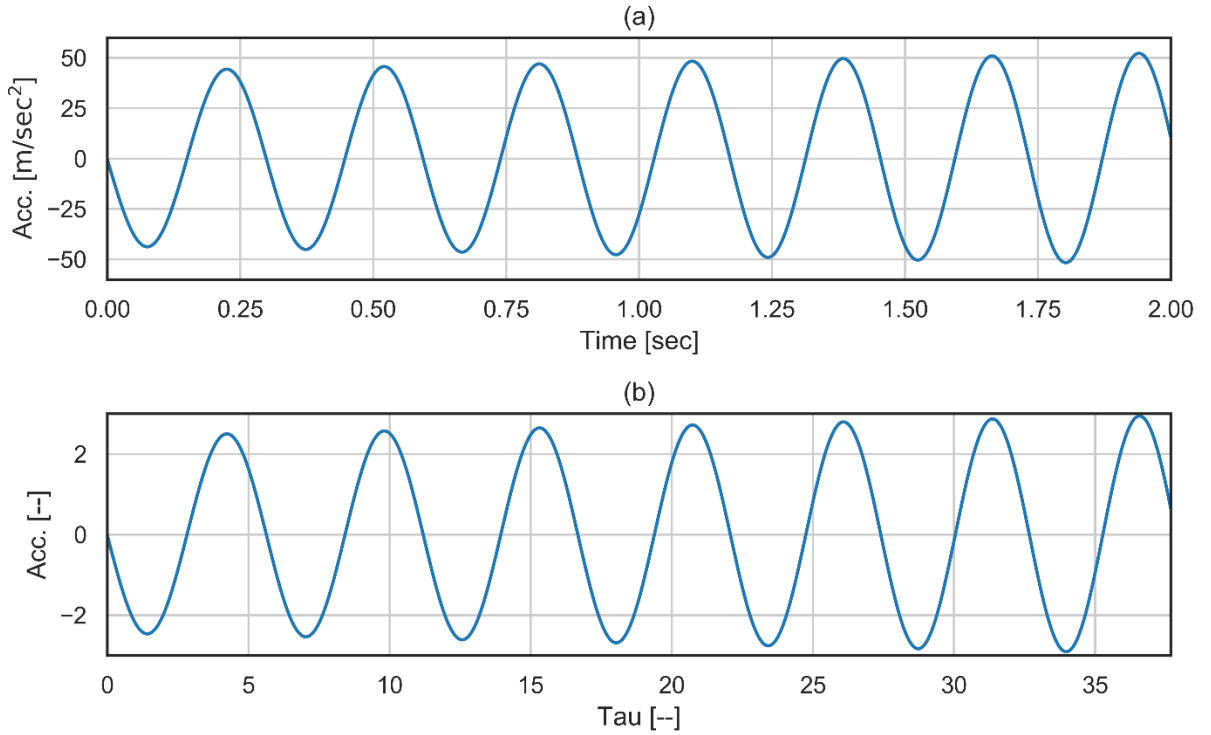
Fig. 8.3 – Ground motion acceleration input to SDOF structure. (a) Physical acceleration (b) Nondimensionalization of the input.

to assert a level of uncertainty commensurate with a typical experimental setting. As the priors on the parameters suggest a constrained optimization on ξ an ω_n, automatic differentiation variational inference (ADVI) is used to transform the distributions of these priors for unconstrained optimization in the variational filtering algorithm. This process is described in detail in Section 3.2 with respect to batch systems, and can be applied equivalently, save for the modifications discussed in Section 8.2.1, in a filtering context.
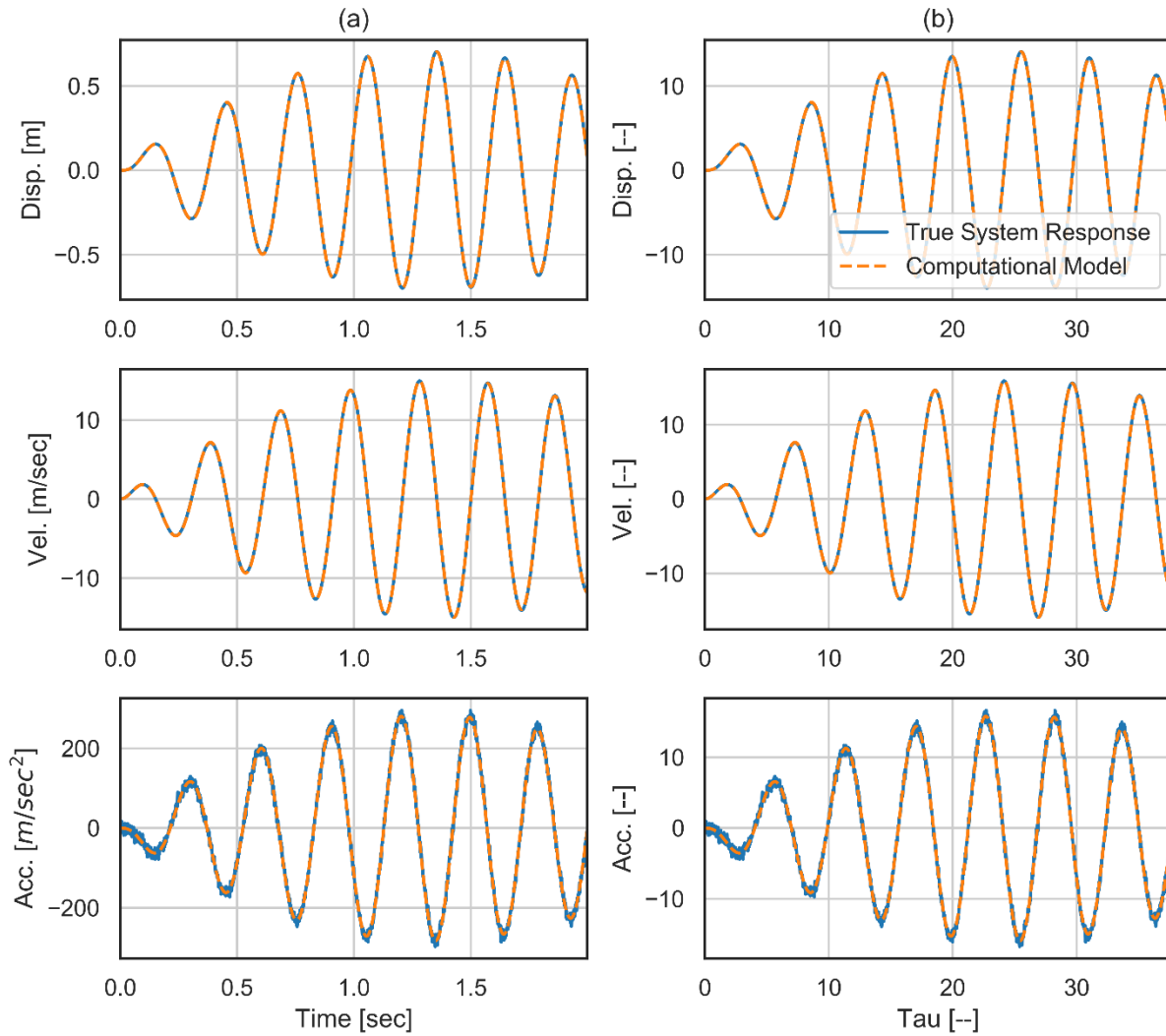
Fig. 8.4 – Response of the linear SDOF oscillator to a sine sweep excitation. (a) Physical response (b) Nondimensionalization of the response.

A sine sweep base excitation near the natural frequency of the system was selected for this example to generate a high amplitude dynamic response with a clear correspondence to the variations on the input, and therefore give the greatest opportunity for parameter identification. The excitation signal, shown in Fig. 8.3, is varied linearly in frequency from 3.33 to 3.67 Hz and is observed at a sampling frequency of 1024 Hz (physical time). The physical signal is given in Fig. 8.3(a) and the nondimensional form of the input is given in Fig. 8.3(b). This short, but information dense, input to the system has the advantage of requiring relatively few states for identification, which allows for a direct comparison between the established batch method for ADVI and the filtering method developed in this chapter (See Section 8.3.2).

The response of the system to the excitation is given in Fig. 8.4 for the physical system (a) and transformed to the nondimensional space for identification (b). The figures display the both the computational model (no uncertainty) and the true response of the system (process and measurement uncertainty), to give a qualitative description to the uncertainty given in Eqs. (8.36)-(8.37) A helpful feature of the nondimensional response is its ability to consolidate the magnitudes of the parameters and system responses, as shown in Fig. 8.4(b), to create a more concise feature space for optimization.

8.3.2   Batch Variational Inference vs. Base Filtering Case

To generate an equitable comparison case between the batch ADVI and the filtering ADVI approach described in Section 8.2, equivalent stochastic models are used between the two cases. Algorithmic parameters, in particular the batch length, batch overlap, downsampling ratio, and optimization iteration count are tuned specifically to each approach to generate a reasonable result. Whereas batch ADVI is set to examine the full signal in one batch at 5×10^5 optimization iterations, the variational filtering approach is tuned to use a total of 28×10^5 optimization iterations with a batch length of 10τ-sec at 90% batch overlap and a downsampling ratio of 1. As expected, the two approaches have very similar levels of performance, as shown in Fig. 8.5 and Fig. 8.6. Table 8.1 provides further confirmation of these results, showing the error in the mode of the marginal posterior on the parameters with respect to the true parameters, and the root-mean-square error (RMSE) on the states remodeled in accordance with the inferred parameters with respect to the true states.

The marginal posterior distribution on the system states given in Fig. 8.5 shows little difference between the states inferred by either method, save in the first ~10 sec of the signal, in which the divergence of the inferred state from the true solution is greater for the variational filtering approach than it is for batch ADVI. This result is expected due to the difference in posterior approximations between the two methods. Whereas batch ADVI generates an approximation to the full posterior on the states given all the data, $\mathbf{y}_{1:K}$, variational filtering will only generate a marginal posterior on the current states given data up to the point of evaluation, $\mathbf{y}_{1:k}$. As the filter is exposed to more and more of the data, the difference in approximation between the batch and filtering approach will decrease. However, as shown in Fig. 8.5, it is expected that the variational filter will diverge from the true state, and therefore from the batch approximation, to a greater degree at the beginning of the signal when fewer data points are available.
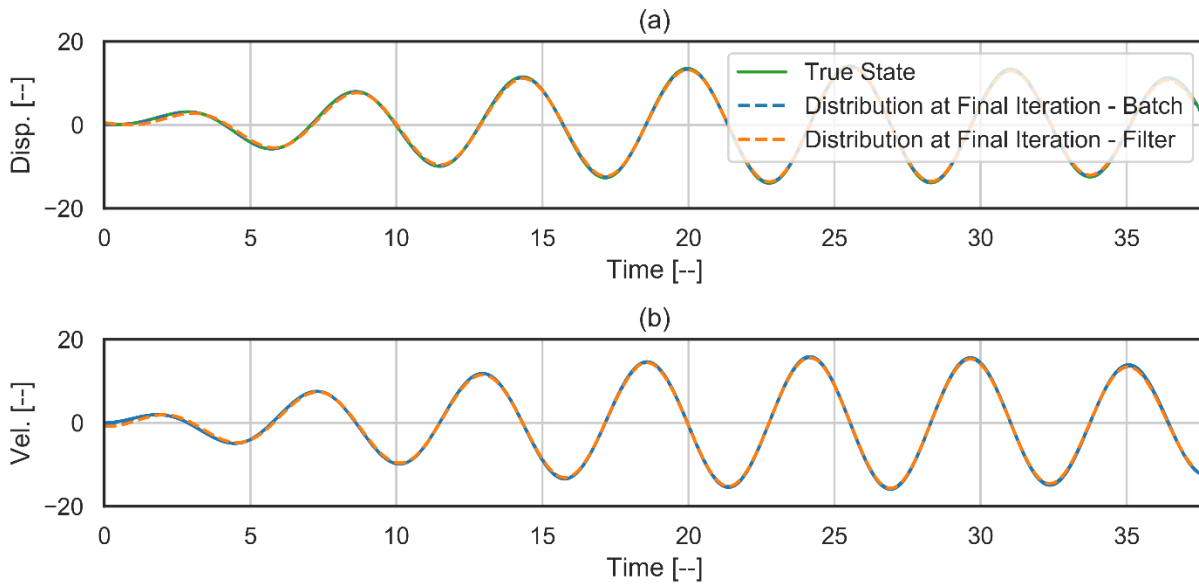
Fig. 8.5 – Posterior distribution on the inferred states as compared to the true states

The marginal posteriors on the parameters are displayed in Fig. 8.6, as well as a consolidated view of their optimization history. Though both parameters are estimated with reasonable accuracy in both cases, it should be noted that ω_n is inferred more quickly and with greater accuracy in both cases than is ξ. This result is fairly common in inverse problems, as ω_n is more easily tied to artifacts in the response data in comparison with ξ. Indeed, for both the batch

and filtering approaches, ω_n is accurately identified within 10% of the optimizations required for the filtering case. In contrast, ξ is identified quite differently between the batch ADVI case and the filtering approach. Though both of these cases require their respective full number of optimization iterations to achieve a relatively accurate result, that iteration count is much higher for the filtering case. However, the beginning of the inference history is relatively similar for the two approaches. Both get into the general neighborhood of the true solution, and then the speed of the inference for the filtering case drops off drastically, indicating a smaller gradient with respect to this parameter for the filtering solution as opposed to the batch solution. It is possible that this difference in the gradient occurs due to the relative availability of the data between the cases, which can increase the peakedness of the likelihood surface for optimization.

Table 8.1 – Summary of inference results for batch and filtering ADVI cross-comparison

Type	Total Opt. Iterations [x10^5]	State RMSE		Parameter Posteriors				Relative Run Time
		Disp. [--]	Vel. [--]	ξ [--]	Error [%]	ω_n [--]	Error [%]	
Batch	5	0.086	0.096	0.0503	0.60	20.02	0.08	1.00
Filter	28	0.994	1.084	0.0610	22.00	20.08	0.40	3.96

The relative run time in Table 8.1, which gives the computational time of each algorithm with respect to the computational time of the batch ADVI method, sheds additional light on the respective utility of the two inference approaches. Though each method ultimately yields similar inference results, the filtering approach takes approximately four times as long to achieve the same quality of results as the batch case, largely due to the smaller gradients in both the states and the damping ratio (ξ) as discussed previously. Though the filtering approach is still quite useful for extremely large data, as the batch case is limited in the number of states it can ultimately infer, and has been shown in this example to achieve similar accuracy to the batch case, additional consideration needs to be given to the computational efficiency of the filtering algorithm. In the following sections, I discuss some variables in the filter code which could alleviate this issue, including the batch size, the batch overlap length, and the downsampling ratio.
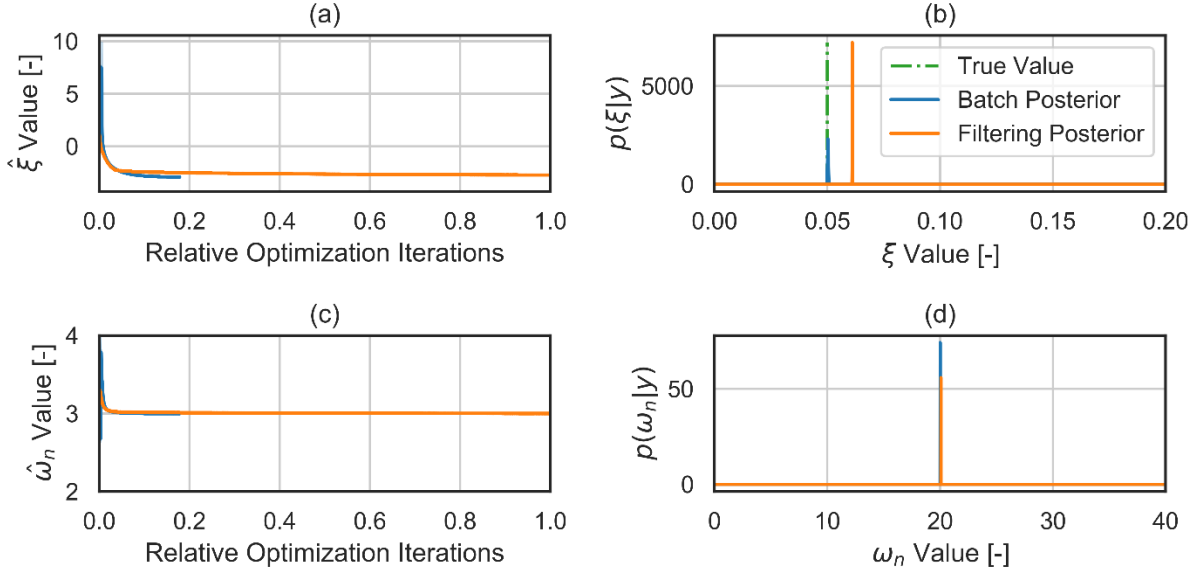
Fig. 8.6 – Inference results on the parameters for the cross-comparison case (a) $\hat{\xi}$ optimization history (b) Parameter posterior on ξ (c) ω_n optimization history (d) Parameter posterior on ω_n

8.3.3 Batch Size

In this parameter study, batch size is varied from 1 τ-sec to 14 τ-sec, encompassing response data representing a quarter period of the response up to 2 periods of the response. For each selected batch size, the number of optimization iterations per batch is varied from 1000 to 500,000, as shown in Table 8.2. The results from this study are reported similarly to Table 8.1, in terms of the RMSE on the remodeled states and the mode of the marginal posterior on the parameters.

The results in Table 8.2 indicates that inference accuracy increases with batch size, at least up through the size of the two response cycles studied herein. For extremely small batch size s (less than half of a cycle of the response data), the optimization problem will fail to converge to an accurate result, regardless of the number of optimization iterations used. There is simply not enough data in the batch that can be leveraged by the loss function to provide actionable information for inference. As the batch size gets larger, reaching up to two response cycles, the optimization problem retains greater and greater accuracy without getting too bogged down with the additional inference terms. Indeed, the run-times of the larger batch sizes are actually decreased with respect to those of the smaller batch sizes as few total batches are needed to be able to walk through the data. It can also be seen from the Batch Size – 14 cases in Table 8.2 that the relative

run time is actually slightly decreased in comparison to the filtering case in Section 8.3.2 for cases with error levels similar to that base filtering case (results in Table 8.1). This result demonstrates the flexibility of the filtering method, though the resultant run times are still much larger than would be practical for a real world filtering scenario.

Table 8.2 – Summary of inference results examining the cross-section of batch size and number of optimization iterations per batch

Batch Size [τ-sec]	Opt. Iter. per Batch	State RMSE		Parameter Posteriors				Relative Run Time
		Disp. [--]	Vel. [--]	ξ [--]	Error [%]	ω_n [--]	Error [%]	
1	1000	7.34	8.21	0.60	1099.00	22.20	11.01	0.41
1	5000	6.99	7.80	0.70	1294.54	16.09	-19.53	2.04
1	10000	6.99	7.80	0.62	1137.01	17.94	-10.28	4.05
1	50000	7.24	8.10	0.79	1476.36	18.31	-8.47	20.94
1	100000	8.28	9.33	0.21	324.87	28.66	43.31	39.86
2	1000	7.41	8.29	0.33	554.83	24.21	21.03	0.23
2	5000	6.97	7.78	0.31	521.70	22.47	12.37	1.16
2	10000	6.72	7.48	0.31	526.18	21.55	7.77	2.27
2	50000	6.32	7.01	0.33	550.99	19.67	-1.65	11.12
2	100000	5.95	6.59	0.28	454.74	19.32	-3.38	22.24
2	500000	6.21	6.89	0.31	520.41	19.56	-2.21	108.32
6	1000	7.85	8.80	0.25	402.87	25.74	28.70	0.08
6	5000	7.06	7.86	0.22	338.90	22.96	14.82	0.38
6	10000	5.67	6.27	0.16	222.65	21.28	6.38	0.78
6	50000	6.47	7.30	0.05	5.93	21.57	7.85	3.62
6	100000	6.33	7.27	0.04	-22.37	21.33	6.67	7.50
6	500000	0.86	0.98	0.05	-0.17	20.21	1.03	36.96
10	1000	7.94	8.92	0.25	395.20	26.25	31.27	0.05
10	5000	6.64	7.39	0.21	317.18	22.22	11.10	0.23
10	10000	5.66	6.25	0.18	252.58	21.08	5.40	0.48
10	50000	2.65	2.90	0.09	72.65	20.25	1.23	2.19
10	100000	0.94	1.03	0.06	20.55	20.08	0.42	4.25
10	500000	0.15	0.16	0.05	-0.27	19.97	-0.15	22.21
14	1000	8.19	9.22	0.24	373.24	28.54	42.69	0.03
14	5000	7.09	7.90	0.19	284.25	22.97	14.86	0.14
14	10000	5.49	6.06	0.14	171.24	21.34	6.72	0.30
14	50000	2.58	2.83	0.08	60.59	20.37	1.83	1.51
14	100000	1.42	1.57	0.06	26.98	20.22	1.08	2.89
14	500000	0.15	0.16	0.05	1.99	19.98	-0.08	13.91

Further clarity is given to these results through the visualization in Fig. 8.7, which shows the displacement RMSE as a function of the batch size and optimization iterations. The results show a definitive preference for larger batch sizes, with consistently high errors in the smaller batch sizes despite increased optimization iterations. It is assumed that this trend will degrade as the batch size becomes too large for efficient optimization. Prior experience from the author on this particular case study suggests that this degradation would occur after the number of inferred states per batch exceeds ~5,000. However, the number of inferred states required to make large batch optimization inefficient is problem dependent. Rather it seems prudent to limit the inference states to a sufficient large batch to incorporate at least one full cycle of response data, without committing extremely large numbers of states for inference at once.
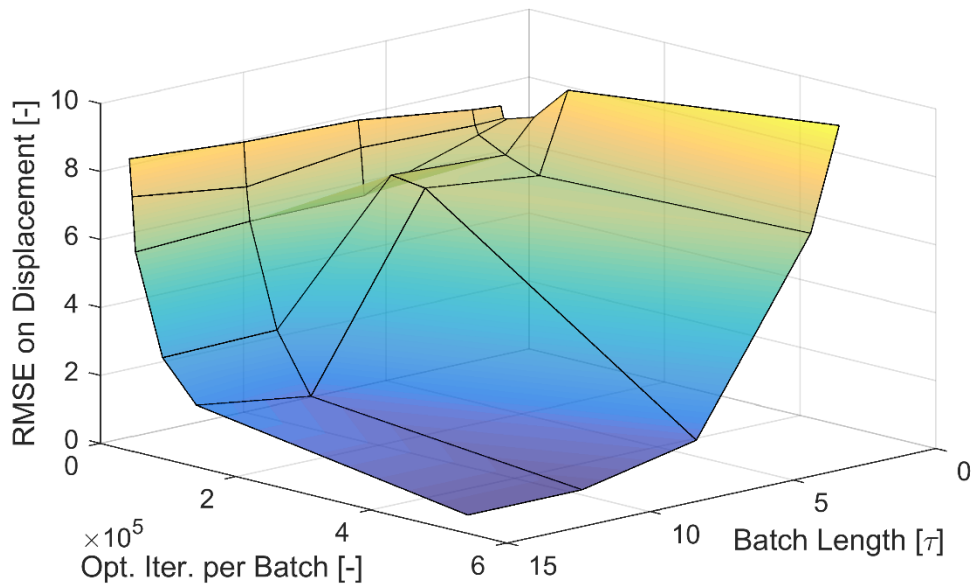
Fig. 8.7 – Displacement RMSE surface dependent on the cross section of batch size and number of optimization iterations per batch

8.3.4 Batch Overlap Length

In this parameter study, the amount of overlap between batches is varied from 0.2% (only one sample in common) to 99.9% (all but one sample in common) to study the effect of inferring some states across multiple batches on the accuracy of the inference. This technique essentially lends greater confidence to the prior on the states for the portion where the marginal posterior is known

from the previous inference step. These states then have the opportunity to be further corrected based on data from the following step, rather than being left with only the data in the current batch. For each selected batch overlap length, the number of optimization iterations per batch is varied from 1000 to 500,000, as shown in

Table 8.3.

Table 8.3 – Summary of inference results examining the cross-section of batch overlap length and number of optimization iterations per batch

Batch Overlap [%]	Opt. Iter. per Batch	State RMSE		Parameter Posteriors				Relative Run Time
		Disp. [--]	Vel. [--]	ξ [--]	Error [%]	ω_n [--]	Error [%]	
0.2	1000	8.64	9.75	0.13	165.27	28.78	43.90	0.00
0.2	5000	8.82	9.97	0.11	119.83	27.90	39.48	0.02
0.2	10000	8.69	9.77	0.12	139.87	25.04	25.19	0.05
0.2	50000	3.60	3.96	0.09	82.27	20.67	3.37	0.24
0.2	100000	1.20	1.34	0.06	18.47	20.22	1.09	0.49
0.2	500000	0.06	0.06	0.05	-0.18	20.00	-0.02	2.41
10.0	1000	10.42	11.83	0.02	-55.74	25.62	28.08	0.01
10.0	5000	8.44	9.49	0.15	204.24	25.86	29.31	0.03
10.0	10000	7.85	8.76	0.14	178.08	23.61	18.03	0.07
10.0	50000	3.08	3.38	0.09	75.49	20.47	2.36	0.33
10.0	100000	1.01	1.13	0.06	16.17	20.17	0.87	0.66
10.0	500000	0.07	0.07	0.05	0.46	20.00	-0.01	3.22
50.0	1000	9.15	10.37	0.06	11.49	28.31	41.56	0.01
50.0	5000	8.55	9.61	0.14	179.39	25.75	28.76	0.05
50.0	10000	7.73	8.61	0.13	152.74	23.22	16.09	0.10
50.0	50000	3.22	3.53	0.09	84.63	20.45	2.27	0.52
50.0	100000	1.18	1.31	0.06	19.18	20.20	1.02	0.96
50.0	500000	0.08	0.09	0.05	0.28	19.99	-0.05	4.85
90.0	1000	7.94	8.92	0.25	395.20	26.25	31.27	0.05
90.0	5000	6.64	7.39	0.21	317.18	22.22	11.10	0.22
90.0	10000	5.66	6.25	0.18	252.58	21.08	5.40	0.44
90.0	50000	2.65	2.90	0.09	72.65	20.25	1.23	2.18
90.0	100000	0.94	1.03	0.06	20.55	20.08	0.42	4.32
90.0	500000	0.15	0.16	0.05	-0.27	19.97	-0.15	22.27
99.9	1000	6.41	7.12	0.26	421.34	21.29	6.47	2.32
99.9	5000	5.17	5.70	0.18	251.59	20.38	1.88	12.73
99.9	10000	11.82	13.86	0.00	-99.80	21.29	6.43	24.38
99.9	50000	1.97	2.15	0.07	48.89	20.18	0.91	113.83

The results demonstrate that the accuracy of the identification is fairly constant for different levels of batch overlap, with the number of optimization iterations playing a much more significant role in the identification accuracy than the overlap length. When the number of optimization iterations is small, it can be seen that the influence of the batch overlap tends to reduce the error in the estimation as overlap length increases. However, when the optimization iterations are sufficiently large (50,000 optimization iterations or greater), the identification accuracy is essentially equivalent between cases, which can also be seen very clearly in Fig. 8.8. Indeed, near equivalent identification accuracy to the base filtering case in Section 8.3.2 can be achieved using a 0.2% batch overlap at 100,000 optimization iterations per batch, resulting in a computation time that is half that of the batch case. These trends can be leveraged to the distinct advantage of the user to reduce computation time. Overall, the results indicate that the advantage of overlapping identification windows is not nearly as distinct as simply increasing the optimization iterations for the current window, and therefore generating a more accurate posterior approximation.
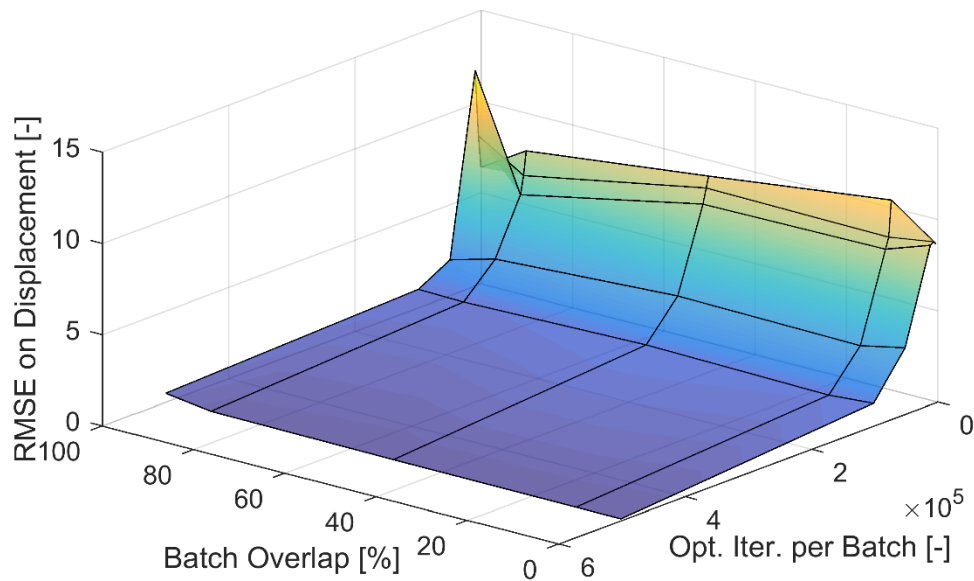

Fig. 8.8 – Displacement RMSE surface dependent on the cross section of batch overlap length and number of optimization iterations per batch

One feature of the data, shown as the peak in the error surface at 99.9% overlap and 10,000 optimization iterations in Fig. 8.8, merits further discussion. This peak is a direct result of the difficulty in obtaining the mode of the posterior distribution on ξ, which is calculated numerically instead of analytically due to the multiple transformations on the posterior distribution. The inferred states for this case present very low error, following the trends shown for the remodeled states in previous cases of this parameter study. However, the inference on the ξ parameter yielded a posterior distribution with such a small variance that it is incredibly difficult to numerically pinpoint the mode of the posterior, as the posterior has essentially become a dirac delta function. This overconfidence in the posterior occurs as a specific result of the high batch overlap and large number of optimization iterations used in this parameter study. When overlap occurs between the batches, the marginal prior asserted on the parameters per Section 8.2 incorporates not only information gleaned from previous observations, but also from observations which are used for the current batch. By 'double counting' the data, particularly for a large number of optimization iterations in the batch, the successive iterations of marginal posteriors become increasingly overconfident, eventually resulting in the case seen in Table 8.3, where the mode cannot feasibly be numerically extracted from the data. 'Double counting' the data appears to exacerbate the already existing issue of overconfidence in the approximate posterior, which has been noted in the variational inference literature [12]. This difficulty has been a particular focus on variational inference approaches for data streams, as noted in Section 8.1, with no definitive strategies yet proposed to effectively mitigate the issue for generalized filtering problems. For this particular filter, an adapted prior artificially increases the variance of the marginal filtering priors to account for the batch overlap size may be sufficient to combat the overconfident posteriors. More adaptive guide distributions may also serve to alleviate this issue, and represent a promising topic for future study. Given the results of this parameter study, it is currently recommended to avoid overlapping inference batches, particularly in excess of 50%.

8.3.5   Down-sampling Ratio

In this parameter study, the sampling rate, and therefore the number of data points and inferred states, is modified according to a factor termed the downsampling ratio (d). For a d of 1, all states are inferred and all data is used in the inference. For a d of $X \neq 1$, every X^{th} state and data point is used in the inference problem, though all excitation measurements are used to facilitate a multi-

step, or in this case X-step, integration between inferred states. For this parameter study, I investigate d values ranging from 1, representing a sampling frequency of $f_s = 1024$ Hz, to 16, representing a sampling frequency on the inferred states of 64 Hz. In theory, this approach would increase computational speed of the filter by greatly reducing the number of states required to be inferred for each case. In practice, this is not necessarily the case, as shown in Table 8.4.

Table 8.4 - Summary of inference results examining the cross-section of downsampling ratio and number of optimization iterations per batch

d	Opt. Iter. per Batch	State RMSE		Parameter Posteriors				Relative Run Time
		Disp. [--]	Vel. [--]	ξ [--]	Error [%]	ω_n [--]	Error [%]	
1	1000	7.94	8.92	0.25	395.20	26.25	31.27	0.05
1	5000	6.64	7.39	0.21	317.18	22.22	11.10	0.23
1	10000	5.66	6.25	0.18	252.58	21.08	5.40	0.45
1	50000	2.65	2.90	0.09	72.65	20.25	1.23	2.22
1	100000	0.94	1.03	0.06	20.55	20.08	0.42	4.38
1	500000	0.15	0.16	0.05	-0.27	19.97	-0.15	21.92
2	1000	7.22	8.07	0.53	969.45	21.72	8.61	0.05
2	5000	6.85	7.64	0.47	842.09	19.67	-1.63	0.25
2	10000	6.37	7.08	0.36	622.81	18.87	-5.67	0.51
2	50000	4.29	4.70	0.14	186.75	19.75	-1.23	2.42
2	100000	1.91	2.07	0.08	52.76	20.01	0.04	4.78
2	500000	0.25	0.27	0.05	4.40	19.97	-0.13	25.16
4	1000	7.43	8.32	0.81	1520.25	21.12	5.59	0.07
4	5000	7.28	8.14	0.73	1364.10	19.72	-1.42	0.36
4	10000	6.89	7.68	0.57	1042.18	17.73	-11.33	0.74
4	50000	5.11	5.62	0.19	286.63	19.33	-3.37	3.48
4	100000	2.69	2.93	0.09	84.69	19.95	-0.24	6.83
4	500000	0.54	0.57	0.06	11.06	19.94	-0.28	32.76
8	1000	7.43	8.32	0.87	1644.63	20.31	1.56	0.12
8	5000	7.33	8.21	0.79	1488.85	19.66	-1.69	0.59
8	10000	7.01	7.82	0.65	1209.76	17.33	-13.35	1.15
8	50000	5.47	6.03	0.23	352.03	19.00	-4.98	5.61
8	100000	2.89	3.14	0.10	94.49	19.78	-1.10	11.23
8	500000	0.57	0.61	0.06	12.78	19.98	-0.10	54.90
16	1000	7.61	8.53	0.92	1740.06	23.70	18.48	0.20
16	5000	7.50	8.40	0.88	1659.50	21.54	7.71	0.98
16	10000	7.45	8.35	0.91	1719.03	20.30	1.52	2.00
16	50000	5.93	6.57	0.28	464.41	18.94	-5.29	9.37

| 16 | 100000 | 3.40 | 3.71 | 0.11 | 121.53 | 19.89 | -0.57 | 19.21 |
| 16 | 500000 | 0.76 | 0.82 | 0.06 | 13.30 | 19.87 | -0.66 | 92.59 |

The results from this case study show that as the d increases, the inference accuracy decreases. Fig. 8.9 displays this effect most clearly, showing the largest drop-off in inference accuracy as d is changed from 1 to 2, with decreasing steps of accuracy drop-off in subsequent d. These results demonstrate that representing the true observations with an artificially sparse set of data points can cause computational issues in the inference. As the data sparsity increases, the log-likelihood surface becomes more varied, resulting in pockets of local maxima throughout the solution space which are not representative of the true system parameters.
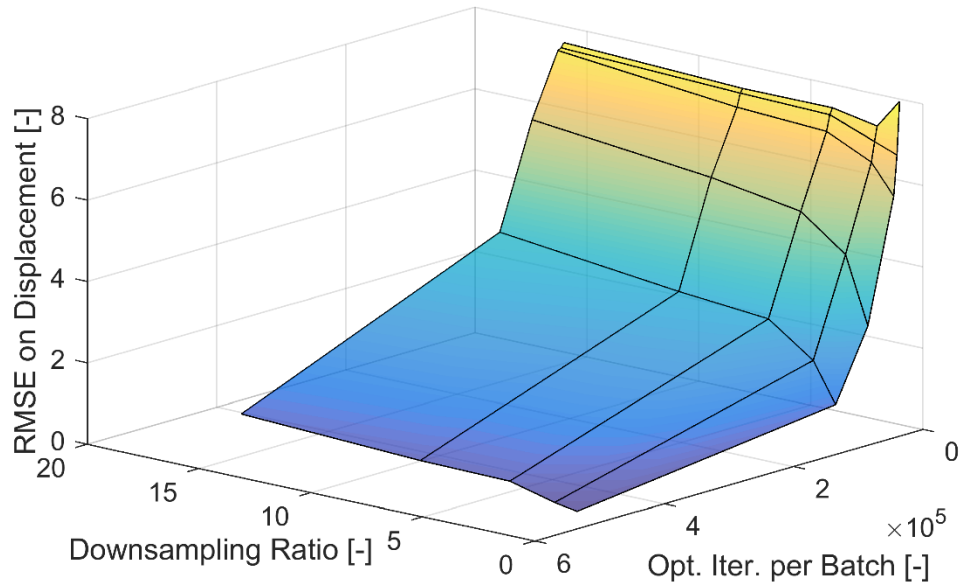

Fig. 8.9 – Displacement RMSE surface dependent on the cross section of batch overlap length and number of optimization iterations per batch

An example of this phenomenon is shown in Fig. 8.10 with respect to the inference of the natural frequency and damping ratio from a single degree-of-freedom oscillator. As d increases from 1 to 64, the likelihood surface develops several highly peaked local maxima. Optimization over this solution space, whether through variational inference or through other Bayesian filter techniques operating on similar likelihood maximization principals, can result in false solutions

which express high confidence in their identification, therefore masking their error from the user. As the tendency of the solution to fall into these false maxima is dependent on the initialization and prior distributions of the filter, there is no way to universally guarantee that the global maximum will be found. Instead, the user must determine the appropriate down-sampling ratio to achieve a relatively uniform likelihood surface, a task which is highly problem dependent and worthy of further study in the future work of the author. The current work will instead focus on defining these techniques and validating their effectiveness on a number of simulated and experimental case studies, as will be shown in the following chapters.

In addition to the issues caused in the likelihood surface by downsampling the data, Table 8.4 shows that the expected gains in computation time have also not been realized. This slowdown in computation with higher d occurs due to the necessity of multi-step integration in the computation of the ELBO. Multi-step integration is used to preserve near equivalent integration accuracy in the computation of the transition probability, and therefore preserve the known model error covariance for the system. Despite optimization of this computation for speed, multi-step integration still requires more computation time than a single-step integration approach, resulting in a buildup of computation time that exceeds the gains afforded by limiting the number of states inferred. Overall, the results of this parameter study recommend against the use of downsampling in the inference problem.
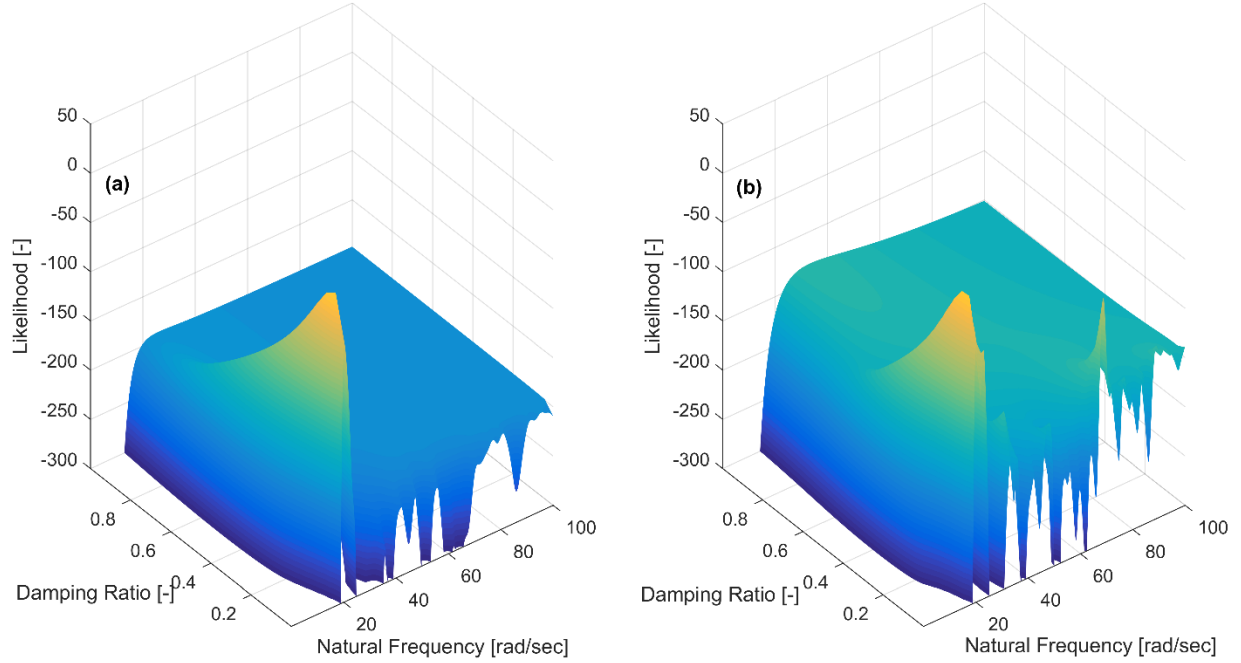
Fig. 8.10 – Comparison of log-likelihood surface at a down-sampling ratios
(a) $d = 1$ (b) $d = 64$

8.4 Conclusions

In this chapter, I have introduced and explored a novel variational filtering technique for the predictive modeling of structural systems. Variational inference stands distinct from other Bayesian inference methods due to its focus on the optimization of an approximate posterior. This approach allows for increased flexibility in the selection of the form of the approximate posterior, and has been shown in the literature to allow the user to reach a suitable understanding of the posterior at significantly less computational cost than sampling methods, and yet with more accuracy than analytical methods.

The developed approach to variational filtering was shown to produce near equivalent results to the batch optimization case, with a computational cost dependent on the tuning parameters used. Variations on these parameters were explored to determine how this cost could best be mitigated. These parameter studies resulted in the understanding that filtering cases incorporating larger batch sizes (between 1 and 2 system response cycles) and low batch overlaps (as low as only 1 sample overlapping) have the potential to produce accurate inference on the system at a low computational cost relative to the batch case. Investigation of the downsampling

approach introduced in Section 8.2.2 recommended against this strategy for use with variational filtering.

The purpose of developing a filtering approach for variational inference was to tap into the benefits of variational inference to tackle the issue of scalable predictive modeling for large-scale civil engineering structures. The work herein represents a starting point, which necessarily uses small scale, low dimensional systems to determine the effectiveness of the developed approach. Lessons learned from this illustrative example can then be extended for adaptation of the method to high dimensional systems. Primary among the lessons learned from this example is that the computational efficiency of the filtering case can be tuned to exceed that of the standard batch approach to variational inference. Though the efficiency of the method is not yet feasible for real time implementation, this tunability, coupled with the incorporation of further strategies to increase the computational efficiency (e.g. amortization, variance reduction, normalizing flows, etc. See Section 9.1) suggests that fruitful implementations of a large-scale filter can be developed. As the dimensionality of the problem increases, two main computational limitations are projected which could hinder the application of the filter: the numerical integration of the transition model and the increasing number of states to be optimized. The first of these issues can be addressed in a number of ways, such as by increasing the efficiency of the computational algorithm with respect to the particular implementation, by introducing parallel computing, or through dimensionality reduction in the states. The challenge of the increased dimensionality of the problem can be addressed through model order reduction techniques or through amortization (see Chpt. 9). It should be noted however that the batch and filtering cases for variational inference demonstrate a high tolerance to the dimensionality of the inferred state in time. By reducing the batch size and increasing the dimensionality of the problem, these concerns can be balanced. The final important lesson learned from this example is that though the stochastic model selected herein is simple, the method can easily be used to accommodate more complex systems, with accuracy dependent primarily on the specification of uncertainty by the user (see Chpt. 6). The variational inference family of approaches is therefore flexible enough to approach the practical uncertainties involved in modeling civil engineering infrastructure. Overall, this illustrative example serves as an effective starting point for the development of a practical variational filtering approach for application in large-scale civil engineering systems.

8.5 References

1. Ghahramani, Z. & Beal, M. J. Propagation algorithms for variational Bayesian learning. *Adv. Neural Inf. Process. Syst.* (2001).

2. Beal, M. J. & Ghahramani, Z. *The variational Kalman smoother*. *Gatsby University Technical Reports* (2001).

3. Sarkka, S. & Nummenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Trans. Automat. Contr.* **54,** 596–600 (2009).

4. Auvinen, H., Bardsley, J. M., Haario, H. & Kauranne, T. The variational Kalman filter and an efficient implementation using limited memory BFGS. *Int. J. Numer. Methods Fluids* **64,** 314–335 (2010).

5. Ait-El-Fquih, B. & Hoteit, I. Fast Kalman-like filtering for large-dimensional linear and Gaussian state-space models. *IEEE Trans. Signal Process.* **63,** 5853–5867 (2015).

6. Friston, K. J., Trujillo-Barreto, N. & Daunizeau, J. DEM: A variational treatment of dynamic systems. *Neuroimage* (2008). doi:10.1016/j.neuroimage.2008.02.054

7. Quinn, A. & Smidl, V. in *The variational Bayes method in signal processing* 145–157 (Springer, 2005).

8. Šmídl, V. & Quinn, A. Variational Bayesian filtering. *IEEE Trans. Signal Process.* **56,** 5020–5030 (2008).

9. Friston, K. J. Variational filtering. *Neuroimage* **41,** 747–766 (2008).

10. Ait-El-Fquih, B. & Hoteit, I. A variational Bayesian multiple particle filtering scheme for large-dimensional systems. *IEEE Trans. Signal Process.* **64,** 5409–5422 (2016).

11. Broderick, T., Boyd, N., Wibisono, A., Wilson, A. C. & Jordan, M. I. Streaming variational bayes. in *Advances in Neural Information Processing Systems* 1727–1735 (2013).

12. Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. & Blei, D. M. Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18,** 430–474 (2017).

13. McInerney, J., Ranganath, R. & Blei, D. The population posterior and Bayesian modeling on streams. in *Advances in Neural Information Processing Systems* 1153–1161 (2015).

14. Jihan, N., Jayasinghe, M. & Perera, S. Streaming stochastic variational Bayes; An improved approach for Bayesian inference with data streams. *PeerJ Prepr.* (2019). doi:10.7287/peerj.preprints.27790v2

15. Archer, E., Park, I. M., Buesing, L., Cunningham, J. & Paninski, L. Black box variational

inference for state space models. in *6th International Conference on Learning Representations* 1–11 (2015).

16.   Bamler, R. & Mandt, S. Structured black box variational inference for latent time series models. in *International Conference on Machine Learning* 1–5 (2017).

17.   Øksendal, B. *Stochastic Differential Equations*. (Springer, 2003).

9. CONCLUSIONS AND OPPORTUNITIES FOR FUTURE STUDY

Through this dissertation, I have explored the practical application of Bayesian inference methods for predictive modeling to structural systems. My work has resulted in several contributions to the field, including a novel global sensitivity analysis approach to the determination of practical identifiability in structural system models, the development of a robust method for unscented Kalman filter identification of systems exhibiting high levels of nonlinearity, and the development and application of variational filtering to structural systems.

The key thread in my contributions, which I have continuously emphasized through my work, is that the accuracy of stochastic methods for predictive modeling are highly dependent on the users accounting for the practical uncertainties inherent in both the structural system as well as the algorithmic approximation to the Bayesian inference solution. My work has generated several novel approaches to combat these issues, beginning with the adoption of Sobol' sensitivity analysis for practical identifiability and its adoption, coupled with a train-validate scheme, to ensure the reliability of the UKF algorithm. This investigation culminated in the exploration of variational inference as a solution approach to predictive modeling for structural systems, which has heretofore remained unknown to the structural engineering community. From this investigation, a novel variational filter has been generated and applied to a simple structural identification problem, showing accurate and robust identification of the system, as well as great flexibility in model development. Overall, my contributions have both highlighted these key concerns with applying Bayesian inference solutions to practical predictive modeling problems, as well as providing practical solutions to ensure that non-expert users can confidently and successfully employ Bayesian approaches for system identification to drive the real, critical decisions that are required in a structural health monitoring context.

9.1 Topics for Future Study

There exist several opportunities to extend the impact of this work through further study. Primary among these opportunities is the potential to increase the efficacy of the variational filter through advances in stochastic modeling and variational inference. Current issues with the variational filtering approach can be summarized in two main topics: (1) too large of a set of inferred states

for efficient inference; and (2) inflexible statements of the variational family leading to overconfident posteriors. Several ideas in the literature may be of help in alleviating these issues.

The number of states to infer can be reduced by the adaptation of amortization to the sequential state modeling case [1]. This approach reduces the number of inference parameters by generating a function that parameterizes the states in terms of the system observations, thereby eliminating the large number of local variables (system states) in the inference problem. Adapting this approach to the case of structural identification has the potential to greatly increase the inference speed.

To address the inflexible statement of the variational family, Liu et al. have developed Stein variational gradient descent, which postulates the variational family as a set of particles driven toward the true posterior, rather than setting a distinct computational form for the guide [2]–[4]. In addition to capturing the true variance of the posterior with greater accuracy, this approach has the distinct benefit of allowing for multi-modal posteriors, which is allows for a much more realistic view of the real-world system. Alternatively, normalizing flows can be used to increase the accuracy of the posterior [5]–[7]. This approach allows the user to parameterize the variational family to be flexible enough to capture any distribution, by using neural networks to define an invertible transformation on the guide to allow it to fit to any posterior. I anticipate that by exploring these and other promising methods and adapting them to the practical concerns faced in structural health monitoring and predictive structural modeling, variational will become a prominent and practical tool in the field of damage detection and structural health monitoring.

9.2 References

1. Zhang, C., Butepage, J., Kjellstrom, H. & Mandt, S. Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41,** 2008–2026 (2019).
2. Wang, D. & Liu, Q. Learning to draw samples: with application to amortized MLE for generative adversarial learning. in *International Conference on Learning Representations* 1–13 (2017).
3. Liu, Q. Stein variational gradient descent as gradient flow. in *Advances in Neural Information Processing Systems* 3116–3124 (2017).

4.   Liu, Q. & Wang, D. Stein variational gradient descent: A general purpose Bayesian inference algorithm. in *Advances in Neural Information Processing Systems* 2378–2386 (2016).

5.   Kobyzev, I., Prince, S. & Brubaker, M. Normalizing flows: An introduction and review of current methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–1 (2020). doi:10.1109/tpami.2020.2992934

6.   Altieri, N. & Duvenaud, D. Variational Inference with Gradient Flows. *NIPS Work.* **37,** 3–6 (2015).

7.   Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S. & Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *arXiv Prepr.* 1–60 (2019).