Journal of Applied and Computational Mechanics

Shahid Chamran University of Ahvaz

Research Paper

Variational Inference for Nonlinear Structural Identification

Alana Lund[1] , Ilias Bilionis[2] , Shirley J. Dyke[3]

[1] Doctoral Candidate, Lyles School of Civil Engineering, Purdue University, 1040 S. River Road, West LaFayette, IN, 47907, USA, Email: alund15@purdue.edu

[2] Associate Professor, School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, 47907, USA, Email: ibilion@purdue.edu

[3] Professor, Lyles School of Civil Engineering and School of Mechanical Engineering, , Purdue University, 585 Purdue Mall, West Lafayette, 47907, USA, Email: sdyke@purdue.edu

Abstract. Research interest in predictive modeling within the structural engineering community has recently been focused on Bayesian inference methods, with particular emphasis on analytical and sampling approaches. In this study, we explore variational inference, a relatively unknown class of Bayesian inference approaches which has potential to realize the computational speed, accuracy, and scalability necessary for structural health monitoring applications. We apply this method to the predictive modeling of a simulated Bouc-Wen system subject to base vibration and compare the performance of this inference approach to that of the unscented Kalman filter. From this investigation, we find that though variational inference is more computationally intensive than the unscented Kalman filter, it exhibits superior performance and flexibility.

Keywords: System identification; Predictive modeling; Bayesian inference; Unscented Kalman filter; Nonlinear systems.

1. Introduction

Predictive modeling allows civil engineers to expand the basis of their health monitoring decisions from observations of the current behavior of a structure to projections of its performance in future hazardous events. Bayesian inference methods give a unique and helpful perspective on this problem as they are able to quantify the inherent epistemic uncertainties that arise due to observations of the system which are both finite in length and limited in the information they contain regarding the states and parameters of interest. 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.

Analytical inference techniques refer to those 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. Several approximate forms of the Kalman filter, such as the extended (EKF) [2], [3] and unscented (UKF) [4], [5] adaptations, have been developed for nonlinear inference and have been found to be effective in a variety of experimental applications [6]–[8]. Recent advances in these methods have addressed improvements to joint state-parameter identification [9]–[12] and joint input-state identification [13]–[15], among other topics of concern for practical monitoring scenarios [16]–[18] . These 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 [19] or sequential Monte Carlo algorithms [20], were developed to remove the barriers imposed by this analytical framework. Instead of operating on a limited set of probability models with a tractable analytical form, sampling approaches enable the use of any stochastic model of the system by drawing the necessary statistical information from a set of weighted sample points, or *particles*, which are propagated through the model. As the number of particles and the length of observation approach infinity, these techniques are guaranteed to converge in expectation to the true system posterior. Comparative studies with analytical techniques have indeed shown that sampling techniques provide a much more informative quantification of the structural monitoring parameters and their uncertainty [10], [21]. 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 [22]–[25]. A more extensive summary of these perspectives and the various techniques which embody them can be found in [26].

A third, optimization-based approach to predictive modeling with Bayesian inference has recently been developed which shows great potential for balancing the computational speed, accuracy, and scalability necessary for practical implementation to civil engineering structures. This approach, referred to as *variational inference*, is 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, $\mathbf{y}$ [27]. The algorithm begins with the definition of the stochastic model of the system $p(\mathbf{y}, \mathbf{z})$, which asserts any prior beliefs about the hidden states and the relationship of these states with observations of the system. A distributional form, $\mathbf{Q}$, is then proposed which serves as an approximation of the posterior density, $p(\mathbf{z}|\mathbf{y})$. This distributional form, called the *variational family* or *guide*, represents a family of distributions whose members,

$q(\mathbf{z})$, can be specified by tuning the distributional parameters to different values. 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

$$
\begin{aligned}
\hat{q} &= \underset{q \in \mathbf{Q}}{\arg\min} \, \mathrm{KL}\big(q(\mathbf{z}) \, || \, p(\mathbf{z}|\mathbf{y})\big) \\
&= \underset{q \in \mathbf{Q}}{\arg\min} \, \big(\mathbb{E}_{q(\mathbf{z})}[\log q(\mathbf{z})] - \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{y}, \mathbf{z})] + \log p(\mathbf{y})\big),
\end{aligned}
\tag{1}
$$

a measure of the information lost by approximating $p(\mathbf{z}|\mathbf{y})$ with $q(\mathbf{z})$ [28]. Although it is possible to propose a guide of any distributional form, proposing a simple and flexible guide allows for efficient computation of a posterior approximation which is well representative of the true posterior.

Variational inference emerged as an alternative Bayesian inference strategy in the mid-90s as a result of the adaptation of mean-field theory from statistical physics [29]–[32], with the first comprehensive statement of the method introduced in 1999 [33]. 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 [27]. Though the impact of the method was expanded by the incorporation of additional probabilistic models into the set of allowable posterior approximations [34]–[36], its true power was made available by the development of stochastic variational inference [37], which introduced stochastic optimization on mini-batches of data, and black-box variational inference [38], which introduced gradient optimization on Monte Carlo (MC) samples of previously intractable expectations. These innovations allow for the scalable and flexible implementation of the variational inference approach, respectively. Automatic differentiation variational inference (ADVI) represents the current state-of-the-art in variational inference, combining the stochastic and black-box techniques with distributional transformations to enable the automatic implementation of the method on a wide range of probabilistic models [39]. Current research seeks to extend the method by introducing variations on the selected loss function [40], [41] and adapting the method for recursive inference [42], [43]. Variational inference has been used extensively in the fields of linguistics [44], [45], image processing [46], and computational biology [47], [48], among others, but has yet to be adapted for applications in predictive modeling for civil structures.

In this paper, we apply ADVI to the identification of a simulated, single-degree-of-freedom Bouc-Wen system subject to base vibration. This study extends the comparative analysis between the ADVI and the UKF algorithms performed by the authors in [49] 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, we 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 paper is organized as follows: Section 2 explores the implementation of the ADVI and UKF algorithms; Section 3 defines the simulated case study used to characterize the accuracy and reliability of the two inference approaches; and Section 4 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 5.

2. Inference Methodology and Implementation

Bayesian inference methods for system identification use noisy observations of the behavior of a system, $\mathbf{y}_{1:T}$, over time $t = 1 \cdot \Delta t, 2 \cdot \Delta t, 3 \cdot \Delta t, \ldots, T$ to discern its hidden states, $\mathbf{z}_{1:T}$, initial conditions, $\mathbf{z}_0$, and physical parameters, $\boldsymbol{\theta}$. These methods assume that the form of the system model is known and can be considered to be a sufficient, but imperfect, representation of the physical system. Structural identification is typically described in terms of stochastic dynamical systems, which are characterized by a Markovian transmission probability $p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta})$ and an emission probability $p(\mathbf{y}_t|\mathbf{z}_t, \boldsymbol{\theta})$. Note that though the states vary with respect to time, the parameters are herein assumed constant. With this understanding of the stochastic model, we can discern the hidden states and parameters using either a batch method or a filtering method. Batch methods infer $p(\mathbf{z}_{0:T}, \boldsymbol{\theta}|\mathbf{y}_{1:T})$, the full joint posterior on the states and parameters, from all data points simultaneously, such that

$$
p(\mathbf{z}_{0:T}, \boldsymbol{\theta}|\mathbf{y}_{1:T}) = \frac{\prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{z}_t, \boldsymbol{\theta}) \prod_{t=1}^{T} p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) \, p(\mathbf{z}_0, \boldsymbol{\theta})}{p(\mathbf{y}_{1:T})},
\tag{2}
$$

where $p(\mathbf{z}_0, \boldsymbol{\theta})$ is the joint prior distribution on the initial conditions of the dynamic states and system parameters and $p(\mathbf{y}_{1:T})$ is the model evidence. Filtering methods operate recursively by instead determining $p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t})$, the marginal posterior of states and parameters at the current time step. The marginal posterior is constructed by first determining the marginal prior through the Chapman-Kolmogorov equation,

$$
p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) = \int p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) p(\mathbf{z}_{t-1}, \boldsymbol{\theta}|\mathbf{y}_{1:t-1}) \, d\mathbf{z}_{t-1},
\tag{3}
$$

and then developing the marginal posterior with Bayes' Theorem,

$$
p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{z}_t, \boldsymbol{\theta}) p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}.
\tag{4}
$$

If necessary, the full posterior, $p(\mathbf{z}_{0:T}, \boldsymbol{\theta}|\mathbf{y}_{1:T})$, can then be constructed by iterating backwards in time with a smoothing algorithm [9]. However, even without the implementation of a smoothing algorithm, the marginal posterior on the parameters at time T, $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$, is equivalent between the batch and filtering interpretations due to the treatment of $\boldsymbol{\theta}$ as constant in eq. (3).

The UKF is a well-recognized filtering method which has been used to successfully identify the dynamic states and parameters of a variety of nonlinear structural systems, including the Bouc-Wen system which is the focus of this study [7], [50]. Variational inference, in contrast, traditionally operates in batch form. The details of these inference approaches are discussed in the following sections.

2.1 Unscented Kalman Filter

The Kalman filter is designed to give the optimal, analytical solution to eq. (3) and eq. (4) for linear systems with Gaussian-distributed transmission and emission probability models [1]. However, practical engineering problems are often found to be nonlinear, either due to the configuration of the system or the need to identify parameters from the model using experimental observations. The UKF was introduced to accommodate these situations as an approximate form of the Kalman filter for highly nonlinear systems. To accomplish this, the UKF takes advantage of the concept of the unscented transform, which allows for the

approximation of the posterior from the statistics of a set of deterministically selected sigma points, $\boldsymbol{\mathcal{X}}^{(t)}$, which are passed through the nonlinear system equations [5].

The filter first requires the definition of the prior distribution on the hidden states and parameters, $[\mathbf{z_0} \quad \boldsymbol{\theta}] \sim N(\boldsymbol{\mu}_0, \mathbf{P}_0)$, and the nonlinear transmission and emission probabilities on those states, which can be formed from the generalized transition and observation models

$$\mathbf{z}_t = \mathbf{f}(\mathbf{z}_{t-1}, \boldsymbol{\theta}, \mathbf{u}_{t-1}) + \mathbf{w}_{t-1} \tag{5}$$

$$\mathbf{y}_t = \mathbf{h}(\mathbf{z}_t, \boldsymbol{\theta}, \mathbf{u}_t) + \mathbf{v}_t \tag{6}$$

where $\mathbf{u}_t$ refers to the input excitation which drives the system and $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ and $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R}_t)$ respectively represent the process and measurement noise terms, which add uncertainty to the otherwise deterministic system. The transmission and emission distributions can then be expressed as $p(\mathbf{z}_t|\mathbf{f}(\mathbf{z}_{t-1}, \boldsymbol{\theta}, \mathbf{u}_{t-1}), \mathbf{Q}_{t-1})$ and $p(\mathbf{y}_t|\mathbf{h}(\mathbf{z}_t, \boldsymbol{\theta}, \mathbf{u}_t), \mathbf{R}_t)$, respectively.

Each filter iteration then begins with the deterministic selection of a weighted set of sigma points which describe the distribution of the states and parameters, $[\mathbf{z}_t \quad \boldsymbol{\theta}]$. The selection of the sigma points can be accomplished in a number of ways to emphasize different aspects of the generating distribution. The sigma point set used herein is constructed using an augmented system state, $\mathbf{z}_t^a = [\mathbf{z}_t \quad \boldsymbol{\theta} \quad \mathbf{w}_t \quad \mathbf{v}_t]^T$, such that $\boldsymbol{\mathcal{X}}^{(t)} = [\boldsymbol{\mathcal{X}}^{(\mathbf{z}_t)} \quad \boldsymbol{\mathcal{X}}^{(\boldsymbol{\theta})} \quad \boldsymbol{\mathcal{X}}^{(\mathbf{w}_t)} \quad \boldsymbol{\mathcal{X}}^{(\mathbf{v}_t)}]^T$, as described in [50]. These sigma points are then propagated through the state transition model given by eq. (5) to estimate $p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t-1})$, which is the predictive distribution of the states and parameters given prior information about the system and its measurement. Next, the predictive distribution is projected through the observation function, eq. (6), to estimate the distribution on the predicted measurement. This distribution is then compared with the true measurement to generate the marginal posterior distribution on the hidden states, $p(\mathbf{z}_t, \boldsymbol{\theta}|\mathbf{y}_{1:t})$. For a more comprehensive explanation of the UKF algorithm, the reader is referred to [9].

Parameter identification is typically incorporated in the UKF through joint state and parameter identification. This method augments the parameters to the state vector, as in $\mathbf{z}_t^a$, and imposes on their transmission a dynamic model with a mean of $\boldsymbol{\theta}$ and a process variance near $\mathbf{0}$. When the parameters are known to be constant throughout the course of the identification, as has been assumed in this derivation, there is no process variance on the parameters. The transmission dynamics can instead be viewed purely through the states, such that the prior on the parameters given the observations $\mathbf{y}_{1:t-1}$ simply informs the propagation of the states to the next time step, $\mathbf{z}_t$. This treatment of the parameters asserts that the only update to their posterior distribution comes from new information available in the observations of the system, yielding an estimate of the parameters that is not prone to process drift. The identified parameters can then be reasonably interpreted from the final posterior at $t = T$. Though the state augmentation approach has been shown to work well in practice for a variety of structural identification problems, it has been observed to increase the frequency of the divergence of the algorithm [9].

2.2 Automatic Differentiation Variational Inference

ADVI generalizes and automates the variational inference method by combining the scalability of stochastic variational inference and the flexibility of black-box variational inference with distributional transforms that standardize the problem domain for compatibility with a generalized solution approach. Though ADVI incorporates elements from a variety of techniques, the procedure itself is quite simple and can be understood in terms of the four steps shown in Fig. 1. The derivation of the algorithm expressed herein is adapted from the work of Kucukelbir et al. [39]. 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.

In the first step, defining the problem domain, we express the stochastic dynamical system in terms of prior distributions on the initial conditions $p(\mathbf{z}_0)$ and the physical parameters $p(\boldsymbol{\theta})$ as well as through the transmission probabilities $p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta})$ and the emission probabilities $p(\mathbf{y}_t|\mathbf{z}_t, \boldsymbol{\theta})$. Given these probabilities, we can express the joint probability of the hidden states, initial conditions, parameters, and observations as

$$p(\mathbf{y}_{1:T}, \mathbf{z}_{0:T}, \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{z}_t, \boldsymbol{\theta}) \prod_{t=1}^{T} p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta}) \prod_{m=1}^{M} p(z_{0,m}) \prod_{k=1}^{K} p(\theta_k), \tag{7}$$

where M is the dimension of the state space and K is the dimension of the parameter space. The factorization of the prior distributions in eq. (7) 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.

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

$$p(\mathbf{y}_{1:T}, \hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}) = p\left(\mathbf{y}_{1:T}, [\mathbf{z}_{0:T}, \boldsymbol{\theta}] = T^{-1}([\hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}])\right) \left|\det J_{T^{-1}}([\hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}])\right|, \tag{8}$$

where $T(\cdot)$ is a one-to-one differentiable function which transforms $[\mathbf{z}_{0:T}, \boldsymbol{\theta}]$ to have full support in $\mathbb{R}^{T+M+K}$ and $J_{T^{-1}}(\cdot)$ is the Jacobian of the inverse of T [39].

In the second step, proposing a variational family, we specify the form of the approximate posterior model that we believe will adequately represent the system. As we have transformed the hidden states and parameters to have support in $\mathbb{R}^{T+M+K}$, the Gaussian distribution is a simple and effective choice for the variational family. We represent the approximate posterior of the system states to be mutually independent and capture the Markovian transitions within each state through a tri-diagonal covariance matrix. This yields the variational families

$$q(\hat{\mathbf{z}}; \boldsymbol{\phi}_{\hat{\mathbf{z}}}) = N(\hat{\mathbf{z}}|\boldsymbol{\mu}_{\hat{\mathbf{z}}}, \mathbf{L}_{\hat{\mathbf{z}}}\mathbf{L}_{\hat{\mathbf{z}}}^T) \tag{9}$$
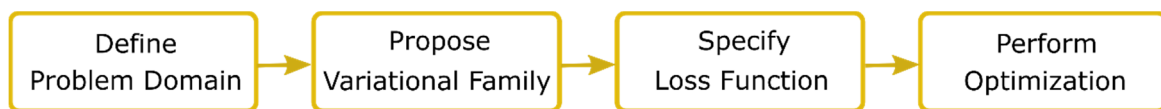
Fig. 1. Key steps of the ADVI algorithm [49]

which are parameterized by the variational parameters $\boldsymbol{\phi}_{\hat{\mathbf{z}}} = (\boldsymbol{\mu}_{\hat{\mathbf{z}}}, \mathbf{L}_{\hat{\mathbf{z}}})$, where $\mathbf{L}_{\hat{\mathbf{z}}}$ is a lower triangular matrix with non-zero elements only on the two primary diagonals. These variational parameters are unconstrained in $\mathbb{R}^{M(3T-1)}$. Likewise, we propose a Gaussian variational family on the hidden parameters which enforces their assumed mutual independence through a diagonal covariance matrix, yielding

$$q\big(\hat{\boldsymbol{\theta}}; \boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}\big) = N\left(\hat{\boldsymbol{\theta}} | \boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}}, \mathrm{diag}(\exp(\boldsymbol{\rho}_{\hat{\boldsymbol{\theta}}})^2)\right) \tag{10}$$

which is parameterized by $\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}} = (\boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}}, \boldsymbol{\rho}_{\hat{\boldsymbol{\theta}}})$ and produces a set of variational parameters which are unconstrained in $\mathbb{R}^{2K}$. The full model then requires optimization on $\mathbb{R}^{M(3T-1)+2K}$.

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

$$\begin{aligned} \mathrm{ELBO}(q) &= -\mathrm{KL}\big(q(\mathbf{z}) \,||\, p(\mathbf{z}|\mathbf{y})\big) + \log p(\mathbf{y}) \\ &= \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}[\log p(\mathbf{y}_{1:T}|\mathbf{z}_{0:T}, \boldsymbol{\theta})] + \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}[\log p(\mathbf{z}_{0:T}, \boldsymbol{\theta})] - \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}[\log q(\mathbf{z}_{0:T}, \boldsymbol{\theta})] \\ &= \mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}[\log p(\mathbf{y}_{1:T}|\mathbf{z}_{0:T}, \boldsymbol{\theta})] - \mathrm{KL}(q(\mathbf{z}_{0:T}, \boldsymbol{\theta}) \,||\, p(\mathbf{z}_{0:T}, \boldsymbol{\theta})), \end{aligned} \tag{11}$$

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, $p(\mathbf{y})$, which can be difficult to determine. We therefore select the optimal member of the variational family $\boldsymbol{Q}$ 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, $\mathbb{E}_{q(\mathbf{z},\boldsymbol{\theta})}[\log p(\mathbf{y}_{1:T}|\mathbf{z}_{0:T}, \boldsymbol{\theta})]$, and encouraging densities which stay close to the prior, $\mathrm{KL}(q(\mathbf{z}_{0:T}, \boldsymbol{\theta}) \,||\, p(\mathbf{z}_{0:T}, \boldsymbol{\theta}))$ [27]. 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. (8) as

$$\mathrm{ELBO}(q) = \mathbb{E}_{q(\hat{\mathbf{z}},\hat{\boldsymbol{\theta}};\boldsymbol{\phi})}\left[\log p\left(\mathbf{y}_{1:T}, T^{-1}([\hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}])\right) + \log\big|\det J_{T^{-1}}([\hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}])\big|\right] - \mathbb{E}_{q(\hat{\mathbf{z}},\hat{\boldsymbol{\theta}};\boldsymbol{\phi})}[\log q(\hat{\mathbf{z}}_{0:T}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})]. \tag{12}$$

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, we introduce an additional transformation, referred to as *elliptical standardization*. This transformation can be expressed as $\eta_{\hat{\mathbf{z}}} = S_{\boldsymbol{\phi}_{\hat{\mathbf{z}}}}(\hat{\mathbf{z}}) = \mathbf{L}_{\hat{\mathbf{z}}}^{-1}(\hat{\mathbf{z}} - \boldsymbol{\mu}_{\hat{\mathbf{z}}})$ for the states and $\eta_{\hat{\boldsymbol{\theta}}} = S_{\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}}(\hat{\boldsymbol{\theta}}) = \mathrm{diag}(\exp(\boldsymbol{\rho}_{\hat{\boldsymbol{\theta}}}))^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}})$ for the parameters, yielding the modified ELBO

$$\mathrm{ELBO}(q) = \mathbb{E}_{N(\boldsymbol{\eta};\mathbf{0},I)}\left[\log p\left(\mathbf{y}, T^{-1}\left([S_{\boldsymbol{\phi}_{\hat{\mathbf{z}}}}^{-1}(\eta_{\hat{\mathbf{z}}}), S_{\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}}^{-1}(\eta_{\hat{\boldsymbol{\theta}}})]\right)\right) + \log\left|\det J_{T^{-1}}\left([S_{\boldsymbol{\phi}_{\hat{\mathbf{z}}}}^{-1}(\eta_{\hat{\mathbf{z}}}), S_{\boldsymbol{\phi}_{\hat{\boldsymbol{\theta}}}}^{-1}(\eta_{\hat{\boldsymbol{\theta}}})]\right)\right|\right] - \mathbb{E}_{q(\hat{\mathbf{z}},\hat{\boldsymbol{\theta}};\boldsymbol{\phi})}[\log q(\hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}; \boldsymbol{\phi})]. \tag{13}$$

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 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. In this study, we implement ADVI using the ADAM stochastic optimization algorithm [51] as part of the python library pyTorch [52]. For further information about variational inference and ADVI, the reader is referred to [27] and [39].

3. Case Study – Single Degree-of-Freedom Bouc-Wen System

To evaluate the relative performance of the ADVI and UKF algorithms, we propose to identify 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

$$m\ddot{x}(t) + c\dot{x}(t) + kr(t) = -m\ddot{x}_g(t), \tag{14}$$

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

$$\dot{r}(t) = \dot{x}(t) - \beta|\dot{x}(t)||r(t)|^{n-1}r(t) - \gamma\dot{x}(t)|r(t)|^n. \tag{15}$$

A schematic of this system is shown in Fig. 2. For this simulated case study, the true parameters of the system are set in accordance with those used in [50], 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}. We choose to observe only the absolute acceleration of the system, which is shown in Fig. 3(c) and given as
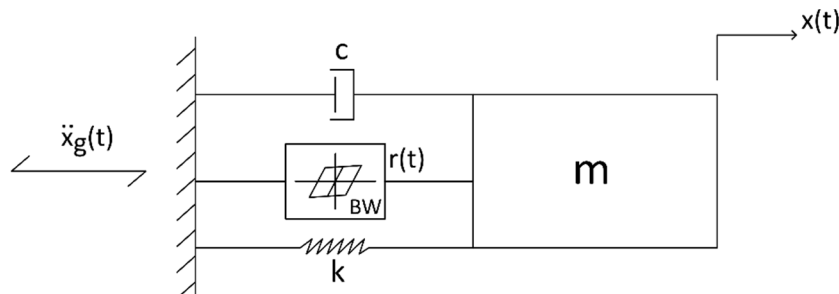
$$y(t) = -c\dot{x}(t) - kr(t). \tag{16}$$

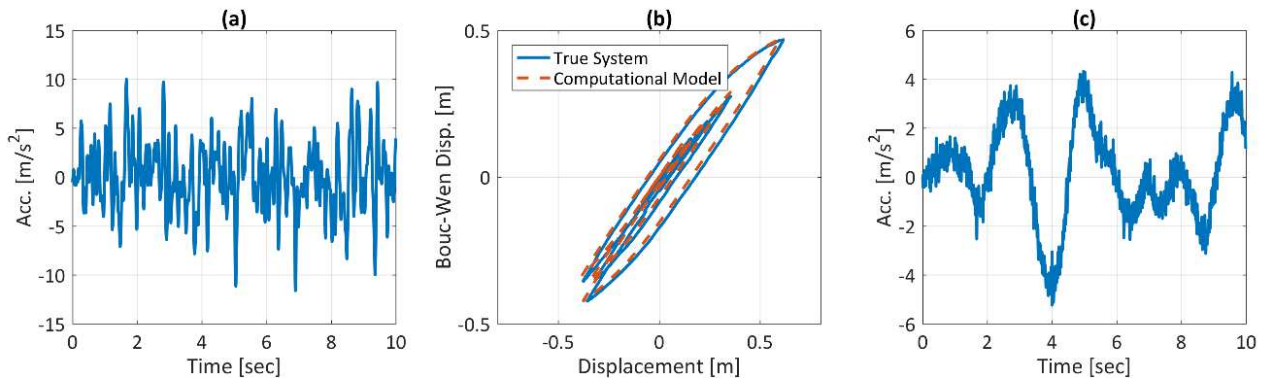**Fig. 2.** Schematic Diagram of the SDOF Bouc-Wen System

Fig. 3. Base acceleration and structural responses used in this case study.
(a) base acceleration (b) Bouc-Wen hysteresis (c) noise-contaminated response acceleration [49]

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 [53]. We therefore nondimensionalize eq. (14) – eq. (16) 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

$$\ddot{\bar{x}}(\tau) + 2\xi\dot{\bar{x}}(\tau) + \bar{r}(\tau) = -\frac{\ddot{x}_g(\tau/\omega_n)}{x_c\omega_n^2}, \tag{17}$$

$$\dot{\bar{r}}(\tau) = \dot{\bar{x}}(\tau) - \beta x_c^n |\dot{\bar{x}}(\tau)||\bar{r}(\tau)|^{n-1}\bar{r}(\tau) - \gamma x_c^n\dot{\bar{x}}(\tau)|\bar{r}(\tau)|^n, \tag{18}$$

$$\bar{y}(\tau) = -2\xi\dot{\bar{x}}(\tau) - \bar{r}(\tau), \tag{19}$$

where $\xi = c/2m\omega_n = 0.05$ is the damping ratio. 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

$$\begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \dot{\mathbf{x}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_2 \\ -2\theta_1\mathbf{x}_2 - \mathbf{x}_3 - \dfrac{\ddot{x}_g(\tau/\theta_2)}{x_c\theta_2^2} \\ \mathbf{x}_2 - \theta_3 x_c^n |\mathbf{x}_2||\mathbf{x}_3|^{\theta_4-1}\mathbf{x}_3 - \theta_5 x_c^n\mathbf{x}_2|\mathbf{x}_3|^{\theta_4} \end{bmatrix}. \tag{20}$$

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 the practical comparison of the UKF and variational inference methods.

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. In accordance with these considerations, we select a band-limited white noise (BLWN) base excitation signal with a maximum amplitude of ~10 m/s² and a frequency cut-off of 10 Hz, as shown in Fig. 3(a). The selection of the BLWN excitation is primarily made based on the indications of parameter identifiability from a Sobol' sensitivity analysis [54] conducted using the python library SALib [55]. 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. 4, 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 [53]. Further indications of system identifiability are given in Fig. 3(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.

3.1 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, we develop our 'true' system model such that it includes an additive, zero-mean Gaussian model uncertainty term on the states, given as $\mathbf{w}_t$ in eq. (5). 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 $\mathbf{w}_{\text{RMS}-\text{NSR}} = [0.01, 0.02, 0.02]$. The additive model uncertainty term is then given by

$$\mathbf{w}_t \sim N\left(0, \text{diag}\left(\sqrt{\Delta\tau} \cdot \mathbf{w}_{\text{RMS}-\text{NSR}} \cdot \text{RMS}(\bar{\mathbf{z}})\right)^2\right), \tag{21}$$
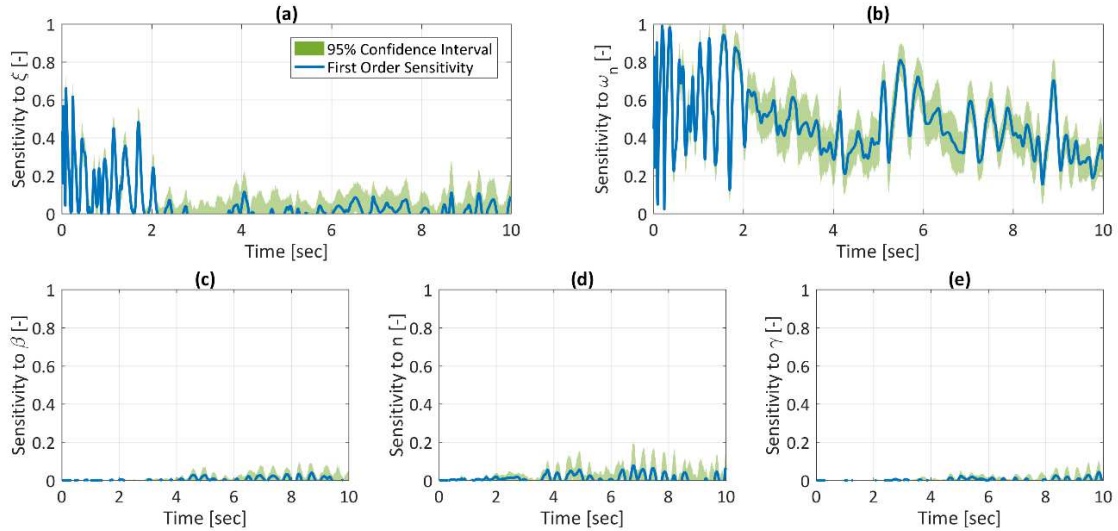
Fig. 4. First order sensitivity of the BLWN response to the system parameters

where the $\sqrt{\Delta\tau}$ scaling factor is included to represent the uncertainty associated with discretizing the continuous ODE [56]. This noise is then propagated through the equation of motion to generate the true response. The difference between the true response and the response of the computational model can be seen in Fig. 3(b). To study the impact of over- and under-estimating the true process uncertainty, we vary the modeling assumption on the process noise covariance according to

$$\mathbf{Q} = \text{diag}\left(\sqrt{\Delta\tau} \cdot \mathbf{w}_{\text{RMS-NSR}} \cdot \text{RMS}(\bar{\mathbf{z}}) \cdot \lambda_w\right)^2, \tag{22}$$

where $\lambda_w \in [80, 90, 100, 110, 120]\%$ expresses the degree to which the true process noise in eq. (21) 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, we vary the measurement noise covariance of the true response data used for inference according to

$$\mathbf{R} = \left(v_{\text{RMS-NSR}} \cdot \text{RMS}(\bar{\mathbf{y}})\right)^2, \tag{23}$$

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, we hold one uncertainty measure constant while the other is varied. We select the base example to connect these two parameter studies 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 4.

3.2 Implementation with Selected Inference Methods

Knowing the two approaches to have distinct differences in their means of approximating the inference process, we 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 [12], [53]. In practice, low variance prior distributions can often be determined for the states, but prior knowledge of the parameters is usually far less explicit. We therefore evaluate the robustness of these methods to variations in the priors on the parameters 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 $\xi \,\epsilon\, [0.1, 1]$, $\omega_n \,\epsilon\, [0, 10]$, $\beta \in [0, 25]$, $n \,\epsilon\, [2, 6]$ 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, the parameters are transformed within the algorithm according to

$$\hat{\boldsymbol{\theta}} = \boldsymbol{T}_{LN}(\boldsymbol{\theta}) = \ln(\boldsymbol{\theta}). \tag{24}$$
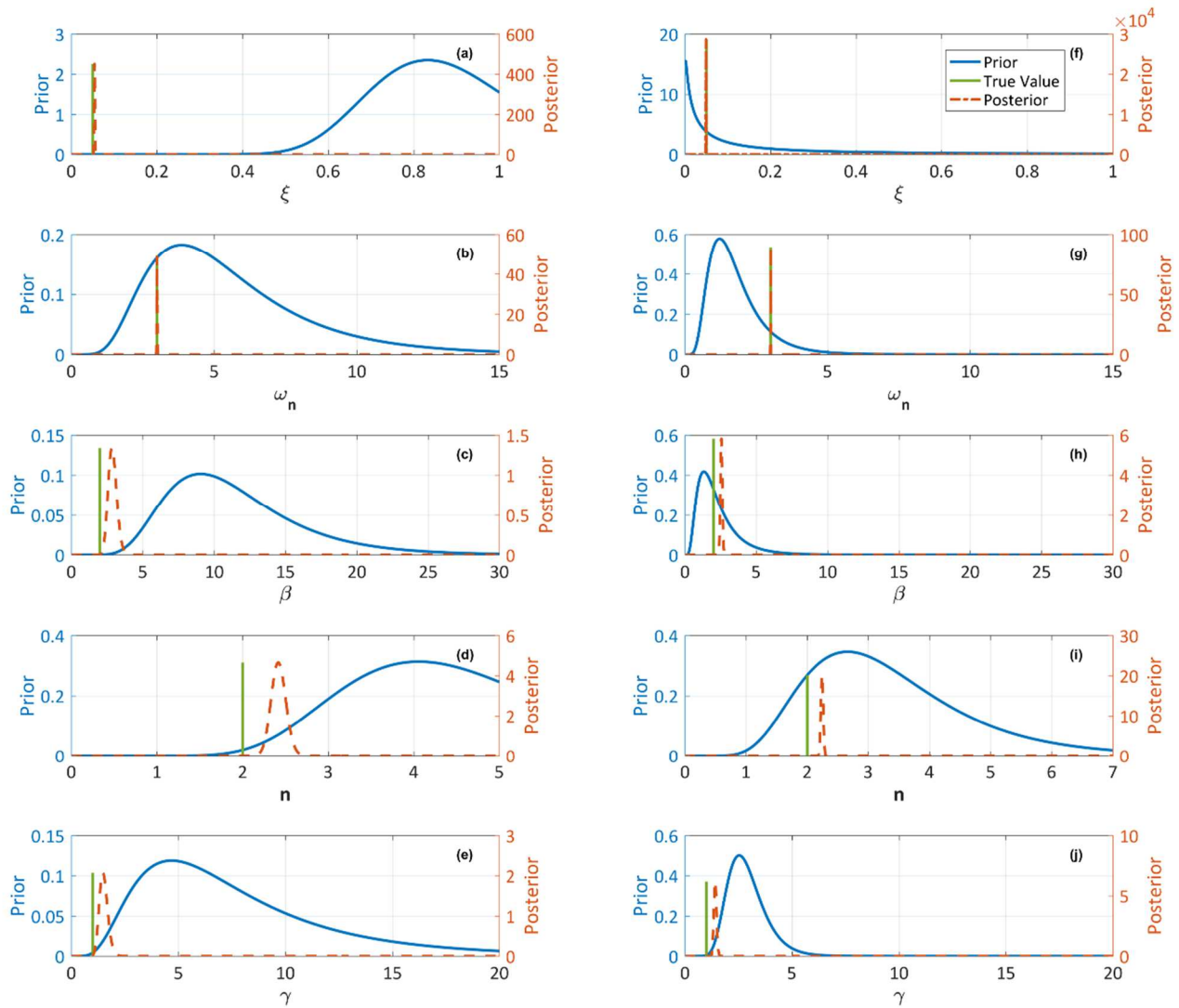
Fig. 5. Priors and posteriors of the minimum RMS error identification trials. (a-e) UKF identification trials (f-j) variational inference identification trials.

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

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

(25)

$$\mathbf{h}\left(\bar{\mathbf{z}}_\tau, \ln(\bar{\boldsymbol{\theta}}_\tau)\right) = -2\bar{\theta}_{\tau,1}x_{\tau,2} - x_{\tau,3}.$$

(26)

The covariance associated with these mean terms is given by the process noise and measurement noise distributions, as defined in eq. (22) and eq. (23). 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

$$p\left(x_{\tau,1}|\bar{\mathbf{z}}_\tau, \bar{\boldsymbol{\theta}}\right) = N\left(x_1(\tau)|\, x_{\tau-1,1} + \Delta\tau \cdot x_{\tau-1,2}, w_1^2\right),$$

(27)

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

(28)

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

(29)

$$p\left(y_\tau|\bar{\mathbf{z}}_\tau, \bar{\boldsymbol{\theta}}\right) = N\left(y_\tau \,|\, -2\bar{\theta}_1 x_{\tau,2} - x_{\tau,3}, v^2\right),$$

(30)

where the assumption of independence across the states and parameters is enforced by the separation of the distributions expressed in eq. (27)-eq. (30).
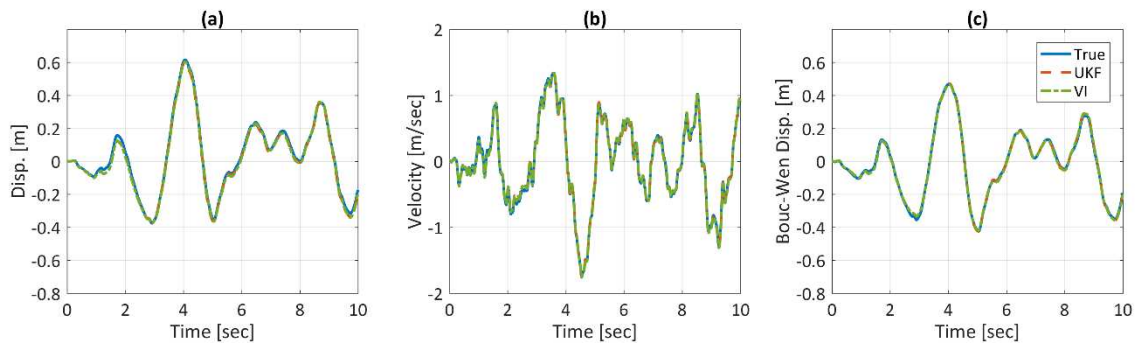
Fig. 6. Comparison of the true system response with model responses re-simulated from the minimum RMS error identification trials.
(a) displacement response (b) velocity response (c) Bouc-Wen displacement response.
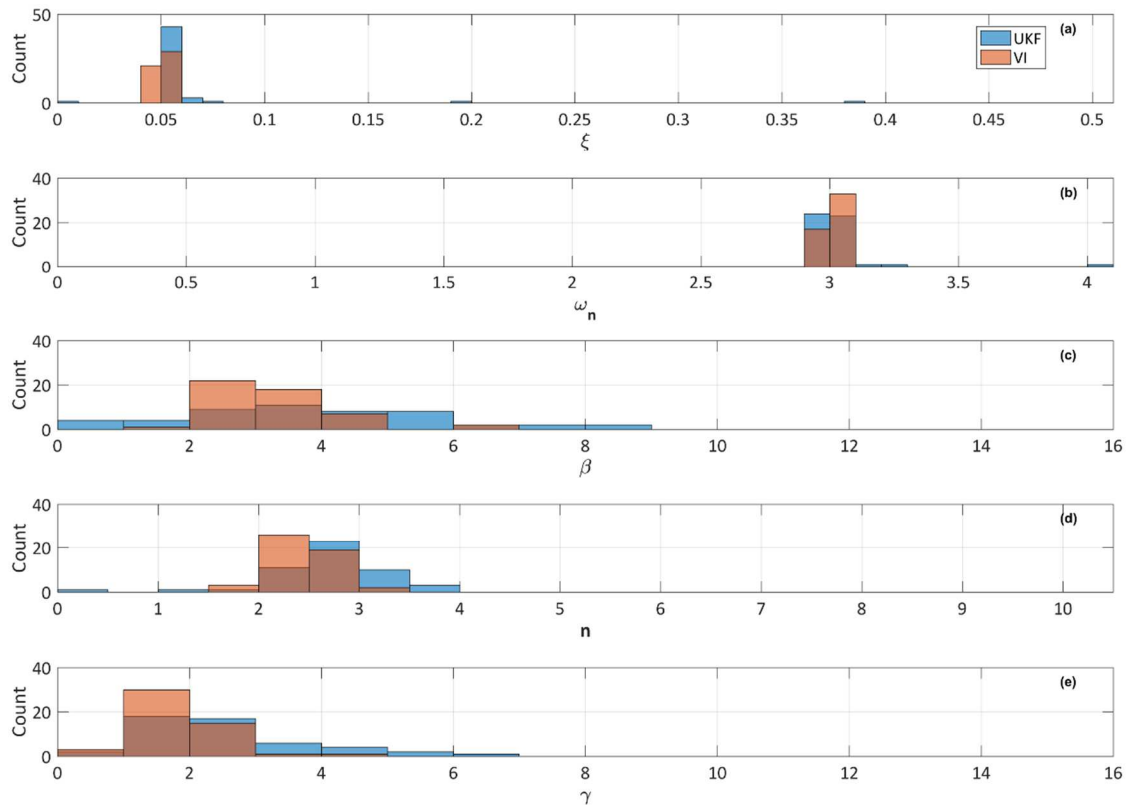

Fig. 7. Distribution of the posterior modes on the parameters for all 50 identification trials.

4. Results and Discussion

As discussed in Section 3, the goal of this case study is to ascertain the relative effectiveness of the UKF and variational inference methods in identifying the hidden states $\bar{z}$ and parameters $\bar{\theta}$ 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, we first examine the case in which the measurement uncertainty factor is set to $v_{\text{RMS}-\text{NSR}} = 0.1\%$ 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 3.2. 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 [57]. 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. 5 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. 5 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. 6. 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. 5. These re-simulated responses show minimal variation from the true response of the system, suggesting that both of the inferred models are adequate solutions.

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. We therefore analyze the relative accuracy of the UKF and variational inference methods over all 50 identification trials, as shown in Fig. 7 and Fig. 8. Fig. 7 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.

A final comparison of the two inference methods in this ideal case is given in Fig. 8. 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. 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.

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.
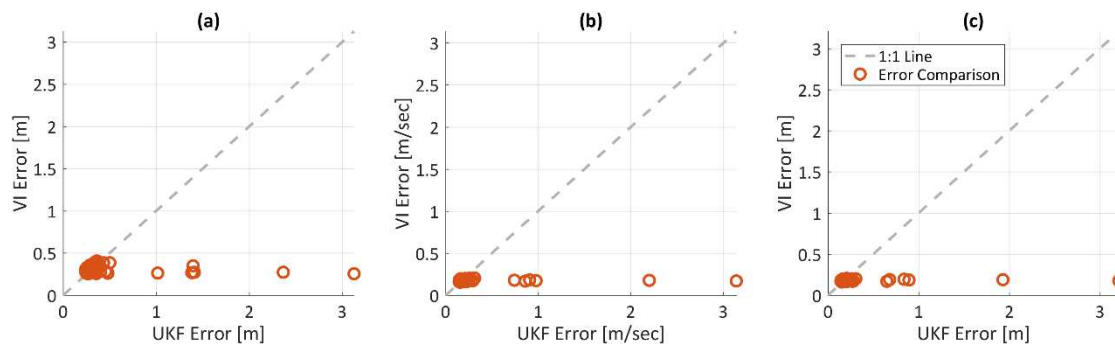

Fig. 8. Comparison of the RMS error on the states for the 50 identification trials.
(a) displacement comparison (b) velocity Comparison (c) Bouc-Wen displacement comparison
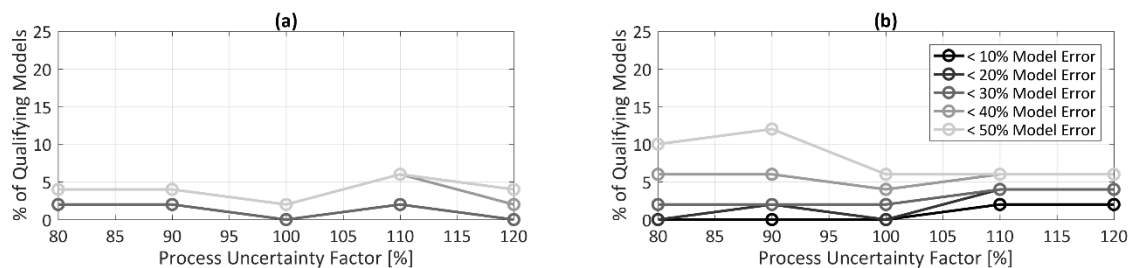

Fig. 9. Percentage of inference trials yielding models whose parameters are all less than a certain percentage error, shown as the process noise uncertainty factor (λ_w) is varied. (a) UKF Models (b) VI Models.
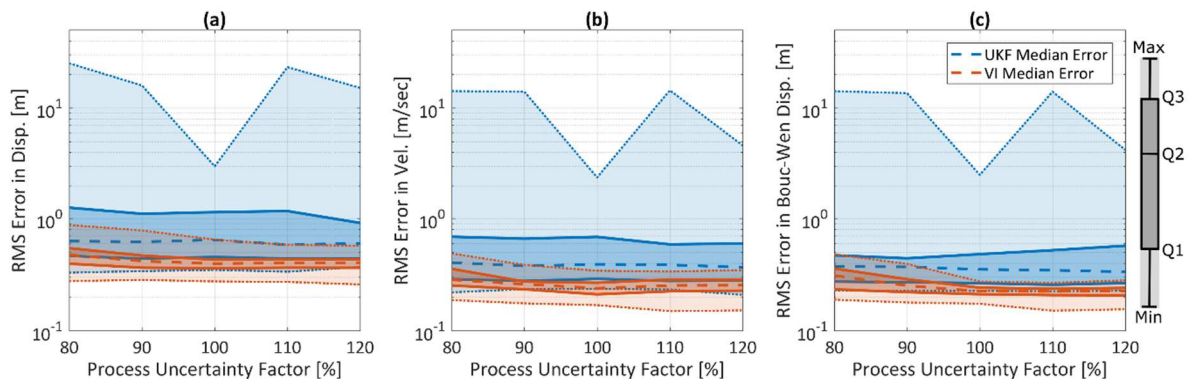

Fig. 10. Statistics of the RMS error in the states given the identified models for each inference method when the process noise uncertainty factor (λ_w) is varied.

4.1 Influence of Process Noise Assumptions

In this study, we evaluate the effect of process noise assumptions during inference on the resulting identified models. To accomplish this, we 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. 9 and Fig. 10.

Fig. 9 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. 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.

Further insight into the robustness of these methods with respect to incorrect process noise assumptions is given in Fig. 10. 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. 9, 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. 9, 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.

4.2 Influence of Measurement Noise Level

For the second parameter study, we evaluate the impact of measurement noise in the response data on the resulting identified models. To study this effect, we 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. 11 and Fig. 12.

Fig. 11 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. 12. 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 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.

4.3 Comparison to More Informative Variational Inference 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, we 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 3.2. This expression of the prior probability requires the definition of the transformation

$$\widehat{\boldsymbol{\theta}}_U = T_U(\boldsymbol{\theta}_U) = \ln\left(\frac{\boldsymbol{\theta}_U - a}{b - \boldsymbol{\theta}_U}\right), \tag{31}$$

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 3.2. Inference using these prior distributions is then repeated on the same parameter studies examined in previous sections.
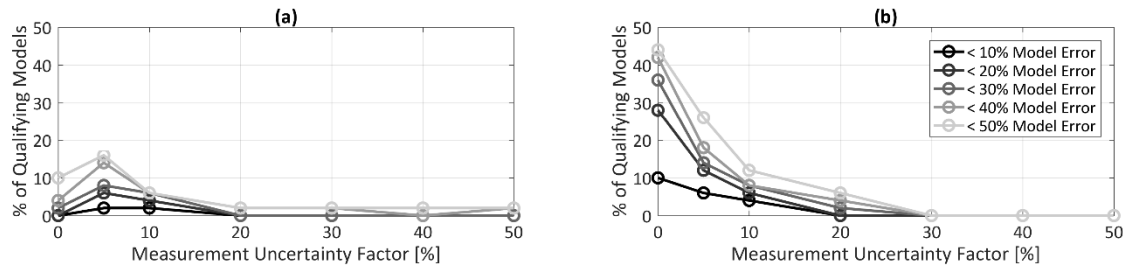
Fig. 11. Percentage of inference trials yielding models whose parameters are all less than a certain percentage error, shown as the measurement noise uncertainty factor ($v_{RMS-NSR}$) is varied. (a) UKF Models (b) VI Models.
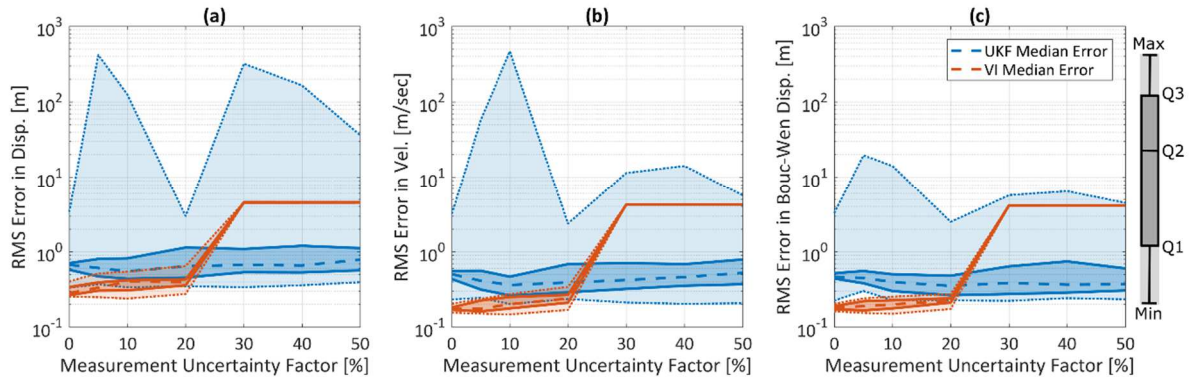

Fig. 12. Statistics of the RMS error in the states given the identified models for each inference method when the measurement noise uncertainty factor ($v_{RMS-NSR}$) is varied.
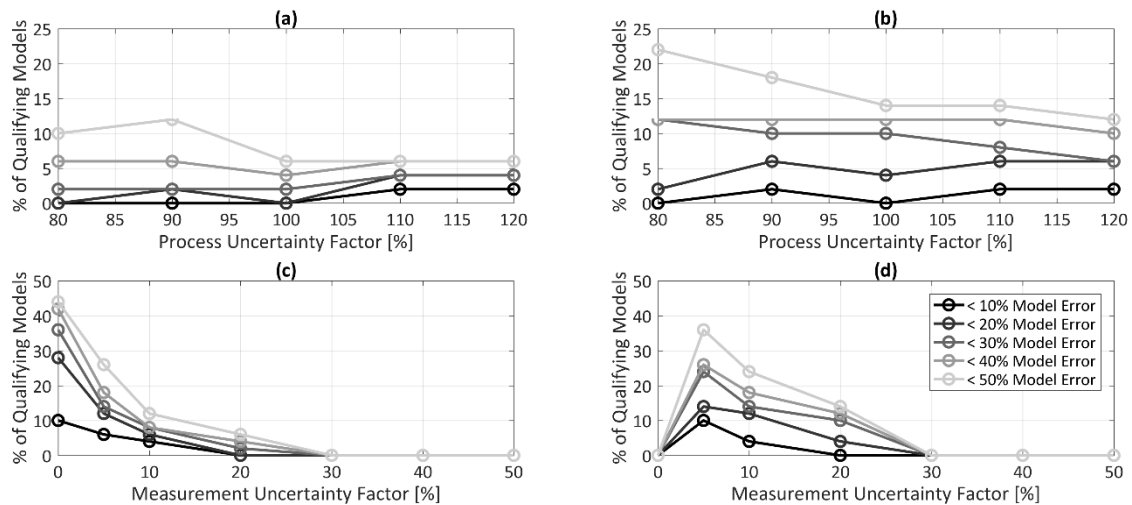

Fig. 13 Percentage of inference trials yielding models whose parameters are all less than a certain percentage error.
(a) λ_w is varied for VI models with lognormal priors, (b) λ_w is varied for VI models with lognormal priors incorporating uniform priors, (c) $v_{RMS-NSR}$ is varied for VI models with lognormal priors, (d) $v_{RMS-NSR}$ is varied for VI models incorporating uniform priors.

The results of these inference trials are reported in Fig. 13 and Fig. 14 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. 13(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 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.

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. 13(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 an accuracy of at least 50% for

measurement noise levels with $v_{RMS-NSR} = 30\%$ or greater. These same trends are expressed in Fig. 14(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.
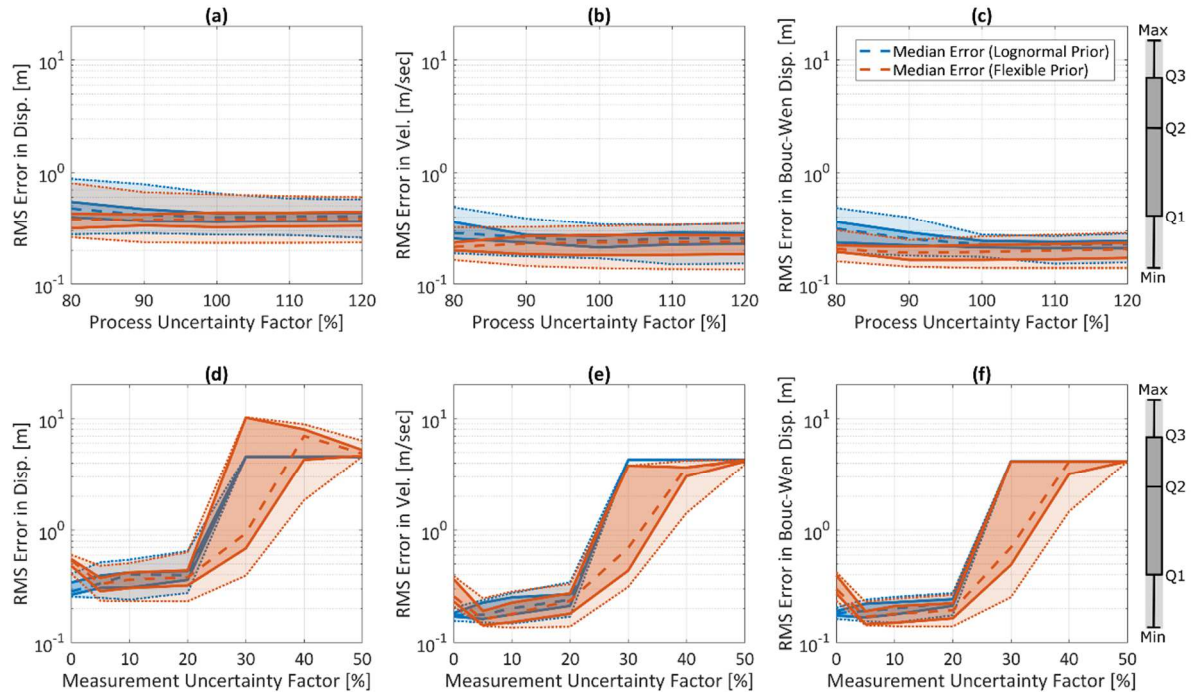
Fig. 14. Statistics of the RMS error in the states given the identified VI models when the (a-c) process noise uncertainty factor (λ_w) is varied, (d-f) measurement noise uncertainty factor ($v_{RMS-NSR}$) is varied.

5. Conclusions

In this paper we 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. As this issue is addressed in future implementations of variational inference, it will become an even more valuable option for structural health monitoring applications.

Author Contributions

A. Lund developed and implemented the mathematical framework, conducted the simulated experiments, analyzed the empirical results, and developed the inference comparison presented herein; I. Bilionis contributed to the mathematical development and examined the inference methods and their comparison; S.J. Dyke planned the scope of the work, initiated the project and suggested the simulated experiments. The manuscript was written through the contribution of all authors. All authors discussed the results, reviewed and approved the final version of the manuscript.

Acknowledgments

Conflict of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

Nomenclature

c	Damping [N·s/m]	$\mathbf{x}$	Displacement relative to ground [m]
f_s	Sampling frequency [Hz]	$\dot{\mathbf{x}}$	Velocity relative to ground [m/sec]
k	Stiffness [N/m]	x_c	Characteristic length scale [m]
m	Mass [kg]	$\ddot{x}_g$	Base acceleration [m/sec^2]
n	Bouc-Wen shape parameter [-]	β	Bouc-Wen shape parameter [m^{-n}]
$\mathbf{r}$	Bouc-Wen displacement relative to ground [m]	γ	Bouc-Wen shape parameter [m^{-n}]
t, τ	Dimensional [sec] and non-dimensional [-] time	ξ	Damping ratio [-]
t_c	Characteristic time scale [1/sec]	ω_n	Natural frequency [rad/sec]

References

[1] Kalman, R.E., A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82(1), 1960, 35–45, doi: 10.1115/1.3662552.

[2] Schmidt, S.F., The Kalman filter - Its recognition and development for aerospace applications, *Journal of Guidance and Control*, 4(1), 1981, 4–7, doi: 10.2514/3.19713.

[3] Sorenson, H.W., Kalman filtering techniques, *Advances in Control Systems*, 3, 1966, 219–292, doi: 10.1016/b978-1-4831-6716-9.50010-2.

[4] Julier, S.J., Uhlmann, J.K., Durrant-Whyte, H.F., A new approach for filtering nonlinear systems, *Proceedings of American Control Conference*, 1995, 3, 1628–1632, doi: 10.1109/ACC.1995.529783.

[5] Wan, E.A., Van Der Merwe, R., The unscented Kalman filter for nonlinear estimation, *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, 153–158.

[6] Zhou, L., Wu, S., Yang, J.N., Experimental study of an adaptive extended Kalman filter for structural damage identification, *Journal of Infrastructure Systems*, 14(1), 2008, 42–51.

[7] Chatzi, E.N., Smyth, A.W., Masri, S.F., Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty, *Structural Safety*, 32, 2010, 326–337, doi: 10.1016/j.strusafe.2010.03.008.

[8] Song, W., Dyke, S.J., Real-time dynamic model updating of a hysteretic structural system, *Journal of Structural Engineering*, 140(3), 2014, 1–14, doi: 10.1061/(ASCE)ST.1943-541X.0000857.

[9] Sarkka, S., *Bayesian filtering and smoothing*, 1st ed. Cambridge University Press, Cambridge, United Kingdom, 2013.

[10] Olivier, A., Smyth, A.W., On the performance of online parameter estimation algorithms in systems with various identifiability properties, *Frontiers in Built Environment*, 3(14), 2017, 1–18, doi: 10.3389/fbuil.2017.00014.

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

[12] Lund, A., Dyke, S.J., Song, W., Bilionis, I., Identification of an experimental nonlinear energy sink device using the unscented Kalman filter, *Mechanical Systems and Signal Processing*, 136, 2020, doi: https://doi.org/10.1016/j.ymssp.2019.106512.

[13] Dertimanis, V.K., Chatzi, E.N., Eftekhar Azam, S., Papadimitriou, C., Input-state-parameter estimation of structural systems from limited output information, *Mechanical Systems and Signal Processing*, 126, 2019, 711–746, doi: 10.1016/j.ymssp.2019.02.040.

[14] Yuen, K.V., Kuok, S.C., Online updating and uncertainty quantification using nonstationary output-only measurement, *Mechanical Systems and Signal Processing*, 66, 2016, 62–77, doi: 10.1016/j.ymssp.2015.05.019.

[15] Lei, Y., Xia, D., Erazo, K., Nagarajaiah, S., A novel unscented Kalman filter for recursive state-input-system identification of nonlinear systems, *Mechanical Systems and Signal Processing*, 127, 2019, 120–135, doi: 10.1016/j.ymssp.2019.03.013.

[16] Solonen, A., Hakkarainen, J., Ilin, A., Abbas, M., Bibov, A., Estimating model error covariance matrix parameters in extended Kalman filtering, *Nonlinear Processes in Geophysics*, 21(5), 2014, 919–927, doi: 10.5194/npg-21-919-2014.

[17] Erazo, K., Sen, D., Nagarajaiah, S., Sun, L., Vibration-based structural health monitoring under changing environmental conditions using Kalman filtering, *Mechanical Systems and Signal Processing*, 117, 2019, 1–15, doi: 10.1016/j.ymssp.2018.07.041.

[18] Song, M., Astroza, R., Ebrahimian, H., Moaveni, B., Papadimitriou, C., Adaptive Kalman filters for nonlinear finite element model updating, *Mechanical Systems and Signal Processing*, 143, 2020, 106837, doi: 10.1016/j.ymssp.2020.106837.

[19] Gordon, N.J., Salmond, D.J., Smith, A.F.M., Novel approach to nonlinear / non-Gaussian Bayesian state estimation, *IEE Proceedings (Radar and Signal Processing)*, 140(2), 1993, 107–113.

[20] Doucet, A., Godsill, S., Andrieu, C., On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, 10(3), 2000, 197–208.

[21] Chatzi, E.N., Smyth, A., The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing, *Structural Control and Health Monitoring*, 16, 2009, 99–123, doi: 10.1002/stc.

[22] Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., Obstacles to high-dimensional particle filtering, *Monthly Weather Review*, 136(12), 2008, 4629–4640, doi: 10.1175/2008MWR2529.1.

[23] Chatzi, E., Smyth, A., Particle filter scheme with mutation for the estimation of time-invariant parameters in structural health monitoring applications, *Structural Control and Health Monitoring*, 20, 2013, 1081–1095, doi: 10.1002/stc.

[24] Olivier, A., Smyth, A.W., Particle filtering and marginalization for parameter identification in structural systems, *Structural Control and Health Monitoring*, 24, 2017, 1–25, doi: 10.1002/stc.1874.

[25] Azam, S.E., Mariani, S., Dual estimation of partially observed nonlinear structural systems: A particle filter approach, *Mechanics Research Communications*, 46, 2012, 54–61, doi: 10.1016/j.mechrescom.2012.08.006.

[26] Haug, A.J., *Bayesian Estimation and Tracking: A Practical Guide*, 1st ed. John Wiley & Sons, Hoboken, New Jersey, 2012.

[27] Blei, D.M., Kucukelbir, A., Mcauliffe, J.D., Variational inference: A review for statisticians, *Journal of the American Statistical Association*, 112(518), 2017, 859–877, doi: 10.1080/01621459.2017.1285773.

[28] Bishop, C.M., *Pattern recognition and machine learning*, 1st ed. Springer, Cambridge, UK, 2006.

[29] Saul, L.K., Jaakkola, T., Jordan, M.I., Mean field theory for sigmoid belief networks, *Journal of Artificial Intelligence Research*, 4, 1996, 61–76.

[30] Saul, L.K., Jordan, M.I., Exploiting Tractable Substructures in Intractable Networks, *Advances in Neural Information Processing Systems*, 1996, 486–492.

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

[32] Hinton, G.E., Van Camp, D., Keeping Neural Networks Simple by Minimizing the Description Length of the Weights, *6th Annual Conference on Computational Learning Theory*, 1993.

[33] Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., An introduction to variational methods for graphical models, *Machine Learning*, 37(2), 1999, 183–233, doi: 10.1007/978-94-011-5014-9_5.

[34] Blei, D.M., Ng, A.Y., Jordan, M.I., Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 2003, 993–1022.

[35] Barber, D., Wiegernick, W., Tractable Variational Structures for Approximating Graphical Models, *Advances in Neural Information Processing Systems*, 1999, 183–199.

[36] Beal, M.J., Variational Algorithms for Approximate Bayesian Inference, Ph. D. Thesis, University College, London, 2003.

[37] Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J., Stochastic variational inference, *Journal of Machine Learning Research*, 14(1), 2013, 1303–1347.

[38] Ranganath, R., Gerrish, S., Blei, D.M., Black box variational inference, *Artificial Intelligence and Statistics*, 2014, 814–822.

[39] Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., Blei, D.M., Automatic differentiation variational inference, *Journal of Machine Learning Research*, 18(1), 2017, 430–474.

[40] Gardner, P., Lord, C., Barthorpe, R.J., A unifying framework for probabilistic validation metrics, *Journal of Verification, Validation, and Uncertainty Quantification*, 2019, doi: 10.1115/1.4045296.

[41] Bamler, R., Zhang, C., Opper, M., Mandt, S., Tightening Bounds for Variational Inference by Revisiting Perturbation Theory, *Journal of Statistical Mechanics: Theory and Experiment*, (12), 2019, doi: https://doi.org/10.1088/1742-5468/ab43d3.

[42] McInerney, J., Ranganath, R., Blei, D., The population posterior and Bayesian modeling on streams, *Advances in Neural Information Processing Systems*, 2015, 1153–1161.

[43] Jihan, N., Jayasinghe, M., Perera, S., Streaming stochastic variational Bayes; An improved approach for Bayesian inference with data streams, *PeerJ Preprints*, 2019, doi: 10.7287/peerj.preprints.27790v2.

[44] Blei, D.M., Lafferty, J.D., A Correlated Topic Model of Science, *The Annals of Applied Statistics*, 1(1), 2007, 17–35, doi: 10.1214/07-AOAS114.

[45] Cohen, S.B., Smith, N.A., Covariance in unsupervised learning of probabilistic grammars, *Journal of Machine Learning Research*, 11, 2010, 3017–3051.

[46] Likas, A., Galatsanos, N.P., A variational method for Bayesian blind image deconvolution, *International Conference on Image Processing*, 2004, 52(8), 2222–2233, doi: 10.1109/ICIP.2003.1246846.

[47] 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(1), 2010, 58–70.

[48] Raj, A., Stephens, M., Pritchard, J.K., fastSTRUCTURE: variational inference of population structure in large SNP data sets, *Genetics*, 197(2), 2014, 573–589, doi: 10.1534/genetics.114.164350.

[49] Lund, A., Bilionis, I., Dyke, S.J., Approximate Bayesian approach to rapid structural identification, *Proceedings of the 17th World Conference on Earthquake Engineering*, 2020.

[50] Wu, M., Smyth, A.W., Application of the unscented Kalman filter for real-time nonlinear structural system identification, *Structural Control and Health Monitoring*, 14, 2007, 971–990, doi: 10.1002/stc.

[51] Kingma, D.P., Ba, J., Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations*, 2014, 1–15.

[52] Paszke, A. *et al.*, PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems*, 2019, 8024–8035.

[53] Lund, A., Dyke, S.J., Song, W., Bilionis, I., Global sensitivity analysis for the design of nonlinear identification experiments, *Nonlinear Dynamics*, 98(1), 2019, 375–394, doi: 10.1007/s11071-019-05199-9.

[54] Saltelli, A. *et al.*, *Global Sensitivity Analysis. The Primer*, 1st ed. John Wiley & Sons, West Sussex, England, 2008.

[55] Herman, J., Usher, W., SaLib: An open-source Python library for sensitivity analysis, *Journal of Open Source Software*, 2(9), 2017, doi: doi:10.21105/joss.00097.

[56] Øksendal, B., *Stochastic Differential Equations*, 5th ed. Springer, Heidelberg, Germany, 2003.

[57] Murphy, K.P., *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Massachusetts, 2012.

ORCID iD

Alana Lund https://orcid.org/0000-0002-5786-4059
Ilias Bilionis https://orcid.org/0000-0002-5266-105X
Shirley J. Dyke https://orcid.org/0000-0003-3697-992X