Bayesian Model Selection and Parameter Estimation for Strongly Nonlinear Dynamical Systems

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Applied Science

by

Philippe Bisaillon

Department of Civil and Environmental Engineering
Carleton University

Ottawa-Carleton Institute of Civil and Environmental Engineering

September 2013

Abstract

In this thesis, a Bayesian framework for model selection and parameter estimation is reported to tackle strongly nonlinear systems using a set of observational data. The model selection task involves estimating the posterior probabilities of each proposed model based on a set of observations. To evaluate the probability of a model, the evidence is computed using the Chib-Jeliazkov method that makes use of Metropolis-Hastings Markov Chain Monte Carlo samples of the posterior distribution of the parameters. The parameter estimation algorithm entails a state estimation procedure that is carried out by non-Gaussian filters. The probability density function of the system state is nearly Gaussian even for strongly nonlinear models when the measurements are dense. The Extended Kalman filter can then be used for state estimation. When the measurements are sparse and noisy, the nonlinearity in the model or measurement operator introduces non-Gaussian features in the system state. In this case, the performance of the Extended Kalman filter becomes unsatisfactory pointing out the need of non-Gaussian filters for state estimation. In this investigation, the Ensemble Kalman filter (EnKF) and Particle filter (that uses EnKF for the proposal density) handle the non-Gaussian state estimation problem. The methodology is illustrated with two numerical examples, namely (a) the Double-Well system having multiple equilibrium (fixed) points, (b) a nonlinear mass-spring-damper system having multiple types of nonlinearities such as freeplay, cubic stiffness and hysteresis.

Acknowledgments

I would like to thank my supervisors Professor Abhijit Sarkar and Professor Dominique Poirel for their support and guidance throughout my MASc degree. I would also like to thank Mohammad Khalil and Rimple Sandhu for all the help they have provided me. I must mention Professor Chris Pettit from the US Naval Academy for his many pertinent comments regarding the Double-Well system. I also thank the members of my thesis committee, Professor David Lau from Carleton University and Professor Ousmane Seidou from the University of Ottawa for their judicious observations. I am infinitely grateful for the support and encouragements of my family. I would like to give a special mention to Vanessa, for her generosity and love.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Definition
MH	Metropolis-Hastings
MLE	Maximum Likelihood Estimation
MCMC	Markov Chain Monte Carlo
DR	Delayed Rejection
DRAM	Delayed Rejection Adaptive Metropolis
MPI	Message Passing Interface
ARWMH	Adaptive Random Walk Metropolis-Hastings
pdf	probability density function
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
EKF	Extended Kalman filter
EnKF	Ensemble Kalman filter
PF	Particle filter
PF-EnKF	Particle filter using EnKF as proposal distribution

Chapter 1

Introduction

1.1 Motivation

In many engineering fields a robust model is necessary to provide accurate predictions. These models need to adequately capture the behaviour of the system that is often nonlinear in nature. For instance, the fluid-structure interactions are often treated with the assumption that both structure and aerodynamics behave linearly. Using the overarching principle of superposition, the assumption of linearity renders the computational simulations efficient, but may introduce serious deficiencies. This assumption implies conservatism in design. Ignoring nonlinear effects may have dramatic consequences. The solar powered Helios aircraft crash is one example [2]. Strong nonlinearities are present in deployable structures used in concerts and stadiums due to the looseness of joints. Such joint looseness induces frictional forces and clearances and may invalidate the (linear) model for structural dynamics introduced, for instance, due to crowd movement [3]. Limit cycle oscillations are present in aeronautical structures due to structural and aerodynamic nonlinearities [4, 5].

Advances in sensor technology like miniaturization and data compression algorithms open up the possibility of blending large amounts of observational data making Bayesian inference a general tool to quantify nonlinearity and reduce uncertainty.

The thesis involves understanding the interplay between nonlinearity and uncertainty. In modelling and simulation, the effect of modelling and measurement uncertainties must be considered for robust predictions. For instance, modelling uncertainty in a structure can arise from variations of the material properties, inaccurate modelling of the constitutive property of the material, imperfect knowledge of the boundary conditions and unmodelled physics [6]. Due to the uncertainty in the system, the model updating becomes a statistical inference problem.

1.2 Objectives

This thesis builds on the Bayesian model selection and parameter estimation framework developed by Khalil et al.; Sandhu (2012); Sandhu et al. (2013) [4, 7, 8]. The parameters of each proposed model are first estimated in the form of a joint probability distribution. The joint posterior distribution of the parameters is approximated by Metropolis-Hastings (MH) MCMC samples. The optimal MH proposal is obtained using an adaptive proposal distribution [9]. The evidence of each proposed model is estimated using Chib-Jeliazkov method [10]. The Bayesian framework includes a state estimation procedure handled by the Extended Kalman filter (EKF). The Bayesian framework has been validated both numerically and experimentally [7]. The Ensemble Kalman filter (EnKF) is used instead of EKF for parameter estimation of the Duffing system in [11]. The parameters are also estimated using a particle filter with an EnKF proposal (PF-EnKF) by augmenting the state vector with the parameters. The performance of EnKF with MCMC is superior to that of PF-EnKF alone. For accurate state estimates, EKF requires that the probability density function of the state is Gaussian or nearly Gaussian. A nonlinear model or measurement operator may lead to non-Gaussian features in the system state. The performance of EKF is satisfactory when the measurements are dense even in the case of strongly nonlinear

systems [4,7,8]. Due to the assimilation of dense observations, the state pdf becomes nearly Gaussian even for nonlinear systems. For sparse measurements, nonlinearities in the model and measurement operators lead to a non-Gaussian pdf of the system state. The objective of this thesis is to extend the Bayesian model selection and parameter estimation algorithm to handle the case of sparse data when the system states are non-Gaussian.

1.3 Thesis Overview

Chapter 2 provides a concise literature review of the common form of nonlinearities encountered in structural dynamics, system identification of nonlinear problems, Bayesian model selection and parameter estimation. Chapter 3 introduces the Bayesian framework for model selection and parameter estimation. Chapter 4 describes the non-Gaussian state estimation problem, including numerical illustration of the performance of various nonlinear filters. Chapter 5 demonstrates the applications of the Bayesian inference algorithm for two strongly nonlinear systems: (a) a Double-Well system having multiple fixed points and (b) a nonlinear oscillator having cubic, freeplay and hysteretic nonlinearities. Conclusion and future research directions are provided in Chapter 6.

1.4 Implementation

This section gives a brief overview of the implementation of the algorithms presented in this thesis. The code is first implemented using MATLAB then in C++ using the Amardillo C++ library [12]. To enhance the performance of the code, the algorithms are coded in a parallel fashion. The MCMC algorithm uses both shared and distributed memory systems using Message Passing Interface with Open MPI [13]. The

state estimation algorithm for EnKF and PF-EnKF is also parallelized using Open MPI to reduce the execution time.

The code is executed on a Linux Cluster running OpenSUSE 11.1 consisting of 22 nodes of two 3.0 Ghz Xeon processors having 8 cores each and 326 GB of memory per node.

Chapter 2

Literature Review

2.1 Nonlinearities in structural dynamics

This section reviews the typical sources of nonlinearities that commonly arise in the dynamics of civil and aerospace structures. The types of nonlinearities considered are geometric, freeplay and hysteretic nonlinearities (e.g. [14]).

Geometric nonlinearity arises when a structure undergoes large displacement. A simple pendulum provides an example of such nonlinearity. The equation of motion for a single degree of freedom pendulum is $\ddot{\theta} + \frac{g}{L} sin(\theta) = 0$. For small angular displacement, $\sin(\theta) \approx \theta$. However, for large displacement, this assumption no longer holds due to nonlinearity stemming from $sin(\theta)$ [15]. Such form of geometric nonlinearity can be modelled by a cubic spring in which the force developed by the gravity is $f_s = k_l u + k_{nl} u^3$. Freeplay nonlinearity belongs to the family of nonlinearity caused by boundary conditions. Freeplay is characterized by a bilinear force-displacement diagram and commonly encountered in aircraft structures due to loose hinges and attachments [3, 16]. The hysteretic nonlinearity arises, for instance, from the loose rivets or friction in the connection [17, 18].

2.2 Nonlinear system identification

In engineering science, robust predictive models that capture the relevant physical characteristics of the system are generally required. The process of calibrating the parameters of such models based on measurements is called system identification. An engineer should not model the data, but rather the information contained in the data by minimizing the effect of noise [19]. A high order Fourier series or a high order polynomial could fit the data perfectly, but be inadequate for robust predictions [19].

The general approach to nonlinear system identification consists of three steps. The first step is the detection which aims to determine if nonlinear behaviour is present. If such behaviour is present, the second stage consists or determining where nonlinearities occur, what type of nonlinearities are present and ultimately their respective form. For instance, the Duffing oscillator is an example of nonlinear restoring force in the form of a polynomial. The last step is to determine the parameters of the nonlinear expressions. To estimate these parameters, several methods are available such as least-square or nonlinear optimization algorithms [20].

Numerous system identification methods have been developed for nonlinear systems. Following [15], these methods can be classified into several categories, such as linearisation, time and frequency-domain methods, time-frequency analysis, black-box modelling. This classification is not exhaustive and further categorizations could be made such as parametric and non-parametric methods [15].

The first category of nonlinear system identification is the linearisation methods. The general approach of these methods is to use approaches developed for linear system. One such technique is *equivalent linearisation* for weakly nonlinear dynamic systems having random excitation [21]. The drawback of this method is that it cannot be applied to strongly nonlinear random oscillators.

The *restoring force surface* introduced by [22,23] in which the systems characteristics are expressed in terms of orthogonal functions. This method is simple but requires that the displacement, velocity and acceleration data are available at each degree of freedom. Recent extensions to this method have been reported for nonlinear system identification in the absence of input measurements by [24]. Another technique is *NARMAX* (Nonlinear autoregressive-moving-average with eXogeneous input) which provides a description of a stationary stochastic process with one polynomial for the auto-regression and another polynomial for the moving average [25, 26].

2.3 Bayesian Model Selection

At its heart, Bayesian model selection makes use of Bayes' theorem which blends prior knowledge of the system with observational data. In Bayesian statistical modelling, the prior distribution and the likelihood function contain the prior of the system and information contained in observational data. There are various approaches to prescribe the prior distribution depending on the level of knowledge such as diffuse prior, Jeffrey's prior, conjugate prior, informative prior, etc. [27] The prior distribution can have a major impact on any Bayesian estimation procedure. The effect of the prior distribution will be discussed in detail in this investigation. With the posterior distribution obtained from the combination of the likelihood function and the prior, one can make an inference on the model parameters.

Bayesian approach is also useful to compare the suitability of plausible models. The model selection problem defines a numerical measure of the evidence in favour of a model among other models [28]. The ratio of the evidence of two models is called the Bayes factor, providing a measure on how strong the data favours a model over another model. However, when comparing multiple models, another approach is to estimate the probability of each model in a given model set [28–30]. Furthermore,

this approach can adequately deal with uncertainties associated with our limited knowledge of the system.

2.4 Parameter Estimation

Given a model, parameter estimation is the task of evaluating those parameters. The frequentist approach estimates the parameters by minimizing a cost function. For example, the maximum likelihood estimation approach selects the parameters that maximize the likelihood of the data. In this approach, the point estimates of the parameters are used for prediction. The Bayesian approach, on the other hand, treats these parameters as random variables that are described by their joint probability density function. This joint distribution is obtained using Bayes' theorem. The major drawback of the Bayesian approach is that the evaluation of the posterior density may require a costly sampling method like MCMC.

Chapter 3

Bayesian Inference for Model Selection and Parameter Estimation

Bayesian inference is defined as the process of fitting a probability model to data where the results are in the form of a probability distribution of the parameters of the model and its predictions [31].

In this investigation, Bayesian inference is used for model selection and parameter estimation. Two categories of model selection are considered; (a) the information-theoretic approach and (b) the Bayesian approach. Following the introduction of model selection methodologies, the parameter estimation technique is described.

3.1 Model selection

The objective of this section is to introduce model selection. The focus of this chapter is to provide a brief overview of Bayesian model selection methodologies. A concise description of information-theoretic model selection approach is also provided in order to contrast Bayesian and information-theoretic approaches.

Model selection involves identifying one or more optimal models from a set of plausible models through a trade-off between the average data fit and the model

complexity [6–8, 28–30].

The Bayesian model selection involves computing the posterior probability of each proposed model. Clearly, this approach relies on the data and proposed models. If the observations used for model selection do not capture some relevant physics of the system, the selected models will not be adequate for understanding the system and for making predictions. There are several approaches to model selection.

3.1.1 Information theoretic approach

The information-theoretic method of model selection is a frequentist approach to compare the proposed models by assigning a score using the parameter estimates obtained by maximizing the likelihood method [32].

For each model $\mathcal{M}$, the maximum likelihood estimate (MLE) $\boldsymbol{\phi}_{MLE}$ of the parameters is obtained by maximizing the likelihood function $p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M})$ where $\boldsymbol{D}$ is the data. The most common approaches are the Akaike Information Criterion (AIC) [33] and the Bayesian Information Criterion (BIC). The AIC is non-Bayesian and can be seen as an extension of Fisher's maximum likelihood method [34]. BIC, developed by Schwarz [35], is an asymptotic approximation of the evidence $p(\boldsymbol{D}|\mathcal{M})$. It only holds good for large data sets and unique MLE of model parameters [34]. Kass and Raftery [36] showed that BIC tends toward simpler models while AIC favours more complex models. The next equations are the formal definition of AIC and BIC respectively when J data points are available [32, 33, 35].

$$\mathrm{AIC}_i = -2 \sum_{j=1}^{J} \ln \left(p(\boldsymbol{d}_j|\boldsymbol{\phi}_{MLE}, \mathcal{M}_i) \right) + 2k \tag{3.1}$$

$$\mathrm{BIC}_i = -2 \sum_{j=1}^{J} \ln \left(p(\boldsymbol{d}_j|\boldsymbol{\phi}_{MLE}, \mathcal{M}_i) \right) + 2k \ln(J) \tag{3.2}$$

The first term in the equations represents the data fit and the second term is a penalty term for model complexity. The penalty is based on the number of parameters k in the model for AIC and additionally the number of data points J for BIC. As mentioned previously, AIC and BIC criteria require a large data set. In this thesis, the effect of data sparsity is investigated making AIC and BIC impracticable for these cases.

3.1.2 Bayesian model selection

Another method of model selection is the Bayesian approach. Figure 1 presents the general approach to model selection. For each proposed model, the parameters are estimated in the form of a posterior distribution. From that posterior parameter probability distribution, the evidence of each model can be calculated. Finally the probability of each model is computed using the evidence.

Using a set of J available measurements $\boldsymbol{D} = \{\boldsymbol{d_1}, ..., \boldsymbol{d}_J\}$, the relative probability of each N models proposed is estimated. Models can be proposed based on the relevant physical laws that govern the system [28, 32, 37, 38]. Using Bayes' theorem, the probability of each model given the available data and the set of N proposed models $\boldsymbol{M}$ is defined by [34]

$$P(\mathcal{M}_i|\boldsymbol{D}, \boldsymbol{M}) = \frac{p(\boldsymbol{D}|\mathcal{M}_i)P(\mathcal{M}_i|\boldsymbol{M})}{\sum_{j=1}^{N} p(\boldsymbol{D}|\mathcal{M}_j)P(\mathcal{M}_j|\boldsymbol{M})}, \tag{3.3}$$

where $p(\boldsymbol{D}|\mathcal{M}_i)$ is the evidence (or the marginal likelihood) of model $\mathcal{M}_i$ given by the data $\boldsymbol{D}$ and $P(\mathcal{M}_i|\boldsymbol{M})$ is the prior probability of model $\mathcal{M}_i$. The prior can be taken as $\frac{1}{N}$ if no prior knowledge is available that would favour one model over another. To keep notation simple, the model set $\boldsymbol{M}$ is implicit in the following equations.

Using Bayes' theorem, the evidence of the model $\mathcal{M}_i$ can be computed as [29,30,39]

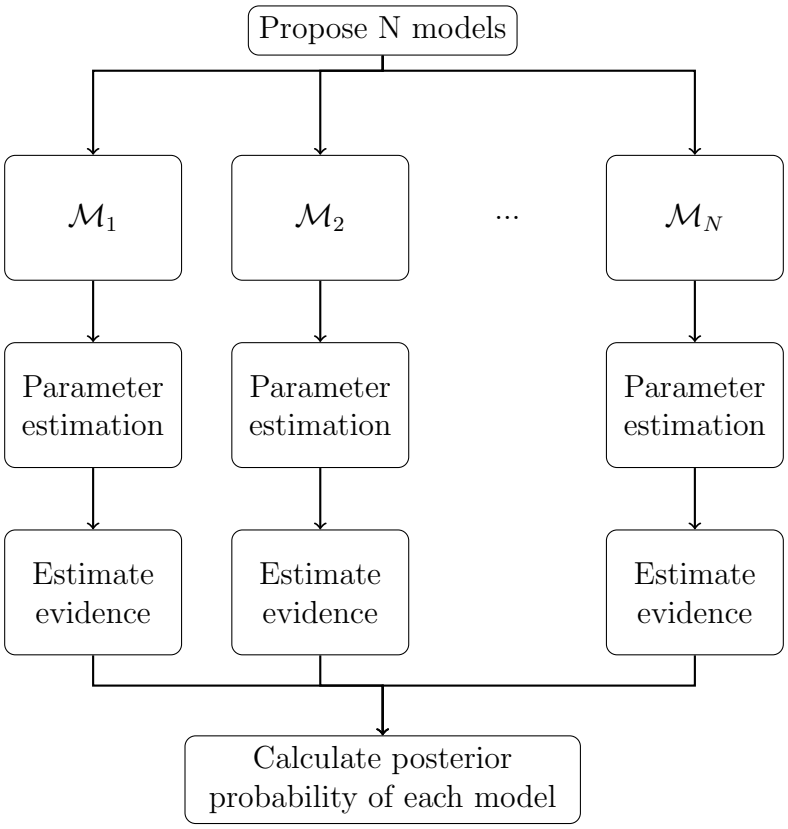


Figure 1: High-level flowchart of the model selection process.

$$p(\boldsymbol{D}|\mathcal{M}_i) = \frac{p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)p(\boldsymbol{\phi}|\mathcal{M}_i)}{p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)}, \tag{3.4}$$

where $p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)$ is the likelihood, $p(\boldsymbol{\phi}|\mathcal{M}_i)$ the parameters prior distribution and $p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)$ is the posterior probability density function (pdf) of the parameters.

Estimation of the evidence

A component of the model selection approach is the parameter estimation as evident from Eq. (3.4). The evidence can be expressed as

$$p(\boldsymbol{D}|\mathcal{M}_i) = \int p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)p(\boldsymbol{\phi}|\mathcal{M}_i)d\boldsymbol{\phi}. \tag{3.5}$$

In most cases, there are no analytical solutions to Eq. (3.5). If the number of parameters is high, a numerical solution may also be challenging. The integral of Eq. (3.5) can be approximated by the Monte Carlo method [30] in which the evidence becomes

$$p(\boldsymbol{D}|\mathcal{M}_i) = \frac{1}{N}\sum_{s=1}^{N} p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)p(\boldsymbol{\phi}|\mathcal{M}_i). \tag{3.6}$$

The N samples are drawn from the prior parameter distribution $p(\boldsymbol{\phi}|\mathcal{M}_i)$. It might not be possible to use this method if the parameter distribution is complex or noninformative. Furthermore, the region of high probability from which the samples are drawn often does not coincide with the high probability region of $p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)$, leading to a poor estimate [30].

Another solution is to use a multi-level sampling technique developed to estimate the evidence. A sequence of related pdfs is generated in which the first pdf is the prior $p(\boldsymbol{\phi}|\mathcal{M}_i)$ and the last pdf is the posterior parameter distribution. The samples are generated using the Transition MCMC method (TMCMC) [40]. The major drawback is the high number of samples required for each of the intermediate pdfs and the difficulty to handle noninformative priors.

Another solution is to compute the evidence by using Eq. (3.4) at a given $\boldsymbol{\phi} = \boldsymbol{\phi}^*$ using the Chib-Jeliazkov method [10]. Friel [41] compared different techniques to estimate the evidence. The fastest method is the Laplace approximation. This approximation requires that the posterior distribution must be adequately approximated by a Gaussian distribution. The Chib-Jeliazkov method is the second most efficient method reviewed. The posterior distribution does not need to be unimodal. In the case of sparse data, the posterior distribution tends to have non-Gaussian features. Hence, this method will be adopted to compute the evidence as it provides accurate estimates.

Effect of the parameters prior distribution

A critical aspect of the Bayesian approach is the inclusion of prior information. The results of model selection is affected by the parameter prior distribution $p(\boldsymbol{\phi}|\mathcal{M}_i)$ even when the number of measurements available is large [34, 36, 42]. For example, in the case of large diffuse priors, the simpler models will always be selected. On the other hand, using only the likelihood tends to select the more complex model leading to overfitting the data [30]. The prior penalizes complex models [30, 39, 43]. For each proposed models, prior distributions are needed for each parameters. In the case of a model where each parameter has a physical interpretation (i.e. the stiffness, the damping, the mass, etc.) one may often have prior information about these parameters. In the case of a statistical model where the physical meaning of parameters is not clear, it is more challenging. The effect of the prior on the evidence can be clearly seen in Eq. (3.4). The better the data-fit, the higher the evidence. Generally, as the model complexity increases, $p(\boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i)$ tends to be lower. As this value decreases, the evidence decreases and thus penalizes complex models. In the case of noninformative prior, complex models are not penalized.

Some work has been done to tackle this issue by Bishop [43] and Oh [39] to parametrize the priors in *automatic relevance determination*. The standard deviation of the prior distribution is selected to control the trade-off between the model complexity and the data-fit error.

If a noninformative prior density of model $\mathcal{M}_i$ $p(\boldsymbol{\phi}|\mathcal{M}_i)$ is defined by $\frac{h(\boldsymbol{\phi}|\mathcal{M}_i)}{c_i}$ where $c_i = \int h(\boldsymbol{\phi}|\mathcal{M}_i)d\boldsymbol{\phi}$. Therefore, $p(\boldsymbol{\phi}^*|\mathcal{M}_i) = \frac{1}{c_i}$. In that case, Eq. (3.4) becomes

$$p(\boldsymbol{D}|\mathcal{M}_i) = \frac{p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i)}{p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)} \frac{1}{c_i}. \tag{3.7}$$

If all the models have the same number of parameters, one can assume $c_i = C$. When calculating the probability of each model, the constant will not influence the

result. However, in the case of model with different number of parameters, it is problematic to make the same assumption. A correction factor can be added based on the number of parameters of a model and the dimension of the data. Following [44], the term selected for this paper is $c_i = J^{\frac{kn}{2}}$ where k is the number of parameters of a model, α controls the impact of the number of parameters and J is the number of data points considered. The correction factor is equivalent to changing the prior probability of the proposed models.

Bayesian model averaging

Bayesian model averaging can be used to make robust predictions [30, 34] using the results of model selection. The posterior robust predictive pdf is based on the posterior probability of each model $\mathcal{M}$ in the model set $\boldsymbol{M}$ containing N models. To predict the pdf of a quantity of interest $\mathbf{x}$, the Total Probability Theorem can be used as [30, 34]

$$p(\mathbf{x}|\boldsymbol{D}, \boldsymbol{M}) = \sum_{j=1}^{N} p(\mathbf{x}|\boldsymbol{D}, \mathcal{M}_j) P(\mathcal{M}_j|\boldsymbol{D}, \boldsymbol{M}) \tag{3.8}$$

where $P(\mathcal{M}_j|\boldsymbol{D}, \boldsymbol{M})$ is the posterior probability of model $\mathcal{M}_j$ found through the model selection process. The posterior pdf given by each model is weighted by its posterior probability. The posterior pdf of the value of interest can be found through a Monte Carlo simulation using the MH samples representing the posterior distribution of the parameters [34]

$$p(\mathbf{x}|\boldsymbol{D}, \mathcal{M}_j) \approx \frac{1}{G} \sum_{g=1}^{G} p(\mathbf{x}|\boldsymbol{\phi}_g, \boldsymbol{D}, \mathcal{M}_j). \tag{3.9}$$

3.2 Parameter Estimation

3.2.1 State Space Models

In this investigation, the mathematical representation of the systems is given by the stochastic ordinary differential equation. The discrete state space representation of a general nonlinear system is given by [1, 4, 45]

$$\mathbf{x}_{k+1} = g_k\left(\mathbf{x}_k, \boldsymbol{f}_k, \boldsymbol{q}_k\right) \tag{3.10}$$

where g_k is the model operator, $\boldsymbol{f}_k$ is a deterministic input, $\boldsymbol{q}_k$ is the stochastic input and $\mathbf{x}_k$ is the state vector of the system at time step k. The stochastic input may represent the random forcing, modelling error or a combination of both. Bayesian inference offers the ability to combine prior knowledge about the model with observations characterized by the following equation [4, 45]

$$\boldsymbol{d}_j = h_j(\mathbf{x}_{d(j)}, \boldsymbol{\varepsilon}_j) \tag{3.11}$$

where $\boldsymbol{d}_j$ is the j-th observation vector at time step $k = d(j)$ and ε_j is the measurement noise. In this thesis, the measurements are simulated. The uncertainty in the measurement model is represented by a measurement error The number of observations available may not be the same as the number of time integration points. The mapping function $d(\cdot)$ links the state $\mathbf{x}_k$ of the system with the corresponding observation. Figure 2 illustrates that relationship.

For instance, the second measurement $\boldsymbol{d}_2$ is available at time instant t_3 (i.e. $k = 3$,$d(2) = 3$).

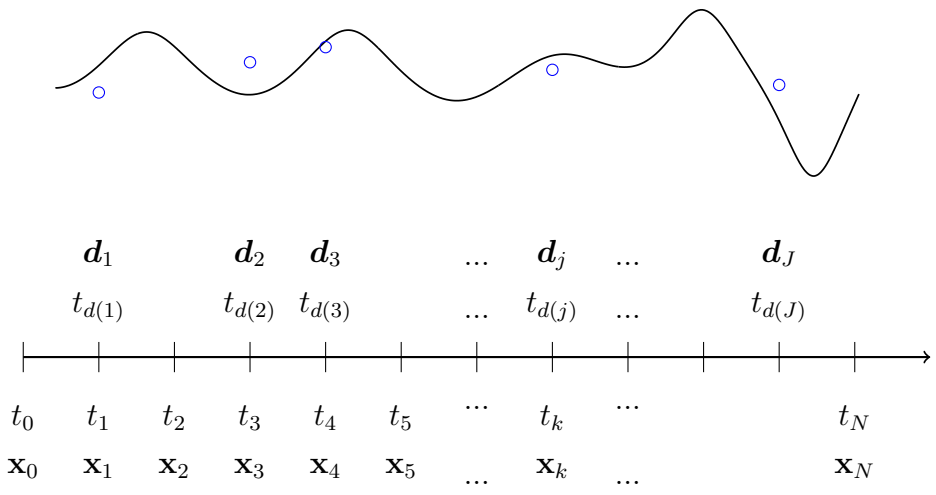


Figure 2: Discrete time domain, adapted from [1].

3.3 Bayesian Inference

To estimate the evidence, three quantities need to be evaluated: $p(\boldsymbol{D}|\boldsymbol{\phi}^*, \mathcal{M}_i)$, $p(\boldsymbol{\phi}^*|\mathcal{M}_i)$ and $p(\boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i)$. The prior distribution is either known or unknown. For unknown prior distribution, one may decide to apply a correction factor as discussed in the last chapter. In this section a strategy will be developed to evaluate those distributions.

Following [46], the joint posterior pdf of the state $\mathbf{x}$ and parameter $\boldsymbol{\phi}$ vector of a

given model $\mathcal{M}_i$ with a set of measurements $\boldsymbol{D} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_j, ..., \boldsymbol{d}_J\}$ is

$$p(\mathbf{x}_1, ..., \mathbf{x}_{d(J)}, ..., \mathbf{x}_k, \boldsymbol{\phi}|\boldsymbol{d}_1, ..., \boldsymbol{d}_J) \propto p(\boldsymbol{\phi}) \left[\prod_{k=1}^{d(1)} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi})\right] p(\boldsymbol{d}_1|\mathbf{x}_{d(1)}, \boldsymbol{\phi})$$
$$\vdots$$
$$\left[\prod_{k=d(J-1)+1}^{d(J)} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi})\right] p(\boldsymbol{d}_J|\mathbf{x}_{d(J)}, \boldsymbol{\phi})$$
$$\left[\prod_{k=d(J)+1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi})\right]. \tag{3.12}$$

The marginal distribution of the parameter vector is

$$p(\boldsymbol{\phi}|\boldsymbol{d}_1, ..., \boldsymbol{d}_J) \propto p(\boldsymbol{\phi}) \left[\prod_{k=1}^{d(1)-1} \int_{-\infty}^{\infty} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi}) d\mathbf{x}_k\right] \int_{-\infty}^{\infty} p(\mathbf{x}_{d(1)}|\mathbf{x}_{d(1)-1}, \boldsymbol{\phi}) p(\boldsymbol{d}_1|\mathbf{x}_{d(1)}, \boldsymbol{\phi}) d\mathbf{x}_{d(1)}$$
$$\vdots$$
$$\left[\prod_{k=d(J-1)+1}^{d(J)-1} \int_{-\infty}^{\infty} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi}) d\mathbf{x}_k\right] \int_{-\infty}^{\infty} p(\mathbf{x}_{d(J)}|\mathbf{x}_{d(J)-1}, \boldsymbol{\phi}) p(\boldsymbol{d}_J|\mathbf{x}_{d(J)}, \boldsymbol{\phi}) d\mathbf{x}_{d(J)}$$
$$\left[\prod_{k=d(J)+1}^{K} \int_{-\infty}^{\infty} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{\phi}) d\mathbf{x}_k\right]. \tag{3.13}$$

Introducing $\mathcal{M}_i$, the above equation can be concisely written as

$$p(\boldsymbol{\phi}|\boldsymbol{d}_1, ..., \boldsymbol{d}_J, \mathcal{M}_i) \propto p(\boldsymbol{\phi}|\mathcal{M}_i) \prod_{j=1}^{J} \int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i) p(\boldsymbol{d}_J|\mathbf{x}_{d(j)}, \boldsymbol{\phi}, \mathcal{M}_i) d\mathbf{x}_{d(j)}. \tag{3.14}$$

Combining Eqs. (3.14) and (3.4), an expression of the parameter posterior pdf

becomes

$$p(\boldsymbol{\phi}|\boldsymbol{D},\mathcal{M}_i) = \frac{1}{p(\boldsymbol{D}|\mathcal{M}_i)} p(\boldsymbol{\phi}|\mathcal{M}_i) \prod_{j=1}^{J} \int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\boldsymbol{\phi},\mathcal{M}_i) p(\boldsymbol{d}_J|\mathbf{x}_{d(j)},\boldsymbol{\phi},\mathcal{M}_i) d\mathbf{x}_{d(j)}. \tag{3.15}$$

The likelihood is therefore evaluated as follows

$$p(\boldsymbol{D}|\boldsymbol{\phi},\mathcal{M}_i) = \prod_{j=1}^{J} \int_{-\infty}^{\infty} p(\boldsymbol{d}_j|\mathbf{x}_{d(j)},\boldsymbol{\phi},\mathcal{M}_i) \underbrace{p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\boldsymbol{\phi},\mathcal{M}_i)}_{\text{state estimation}} d\mathbf{x}_{d(j)} \tag{3.16}$$

In equation (3.14), the posterior ordinate can be evaluated up to proportionality because the normalizing constant of the posterior pdf, namely the evidence, is not known. Hence, a Markov Chain Monte Carlo sampling technique will be used to sample from the posterior distribution.

To estimate the likelihood, a state estimation procedure is required. More details on the evaluation of this integral are presented in the next chapter. To evaluate the posterior ordinate at $\boldsymbol{\phi} = \boldsymbol{\phi}^*$, Chib-Jeliazkov [10] method is used. This method requires Metropolis-Hastings samples of the posterior distribution. The next figure shows the parameter estimation algorithm and how those samples are obtained.

As seen in the state estimation box, the MH algorithm requires the ratio of the posterior pdf evaluated at two different points. Since they share the same normalization constant, Eq. (3.14) is sufficient for MCMC. The next section will introduce the Markov Chain Monte Carlo algorithm used.

3.3.1 Markov Chain Monte Carlo

In the Bayesian framework, a high dimensional integral is required to obtain the posterior distribution. The dimension of that integral relates to the number of parameters. Depending on the distribution, no analytical solution may exist and the

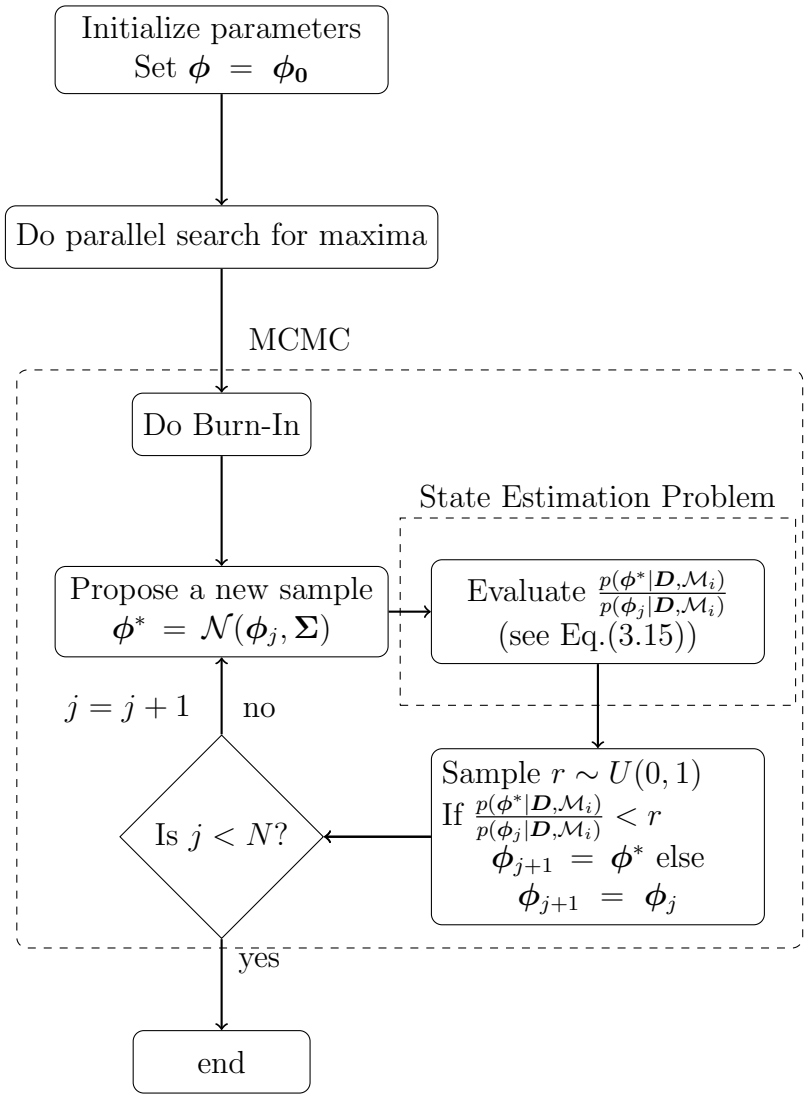


Figure 3: Flowchart of the parameter estimation procedure.

numerical evaluation of the integral is challenging. Markov Chain Monte Carlo methods are used to bypass this integral. These methods sample directly from the posterior distribution by constructing a Markov chain.

A sequence of random vector $\{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_N\}$ is called a first-order Markov chain if vector i only depends on the previous vector $i-1$ such that [47, 48]

$$p(\mathbf{x}_{i+1} = \boldsymbol{y}|\mathbf{x}_i = \boldsymbol{z}, ..., \mathbf{x}_0 = \boldsymbol{a}) = p(\mathbf{x}_{i+1} = \boldsymbol{y}|\mathbf{x}_i = \boldsymbol{z}). \tag{3.17}$$

The transition kernel, denoted $p(\mathbf{x}_{i+1} = \boldsymbol{y}|\mathbf{x}_i = \boldsymbol{z})$, is homogeneous when it is independent of i. In this case the transition kernel is denoted by

$$p(\mathbf{x}_{i+1} = \boldsymbol{y}|\mathbf{x}_i = \boldsymbol{z}) = A(\boldsymbol{y}, \boldsymbol{z}) \; \forall \; i. \tag{3.18}$$

The transition kernel must satisfy [49]

$$A(\boldsymbol{y}, \boldsymbol{z}) \geq 0 \, \forall \, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{S} \text{ and } \int A(\boldsymbol{y}, \boldsymbol{z}) d\boldsymbol{z} = 1. \tag{3.19}$$

There are three properties that Markov chain should meet to ensure their effectiveness: irreducibility, aperiodicity and reversibility [50, 51]. A Markov chain is said to be irreducible when all state of the chain can be reached from any starting position in any number of iterations. This requires that the transition kernel is greater than 0 for all state sequences. A Markov chain is aperiodic when there are no periodic oscillations between two states. The last property is reversibility. A Markov chain is reversible if the transition kernel of the chain and its respective reversed chain are the same. The probability of jumping from $\boldsymbol{y}$ to $\boldsymbol{z}$ is the same as jumping from $\boldsymbol{z}$ to $\boldsymbol{y}$.

$$p(\mathbf{x}_{i+1} = \boldsymbol{y}|\mathbf{x}_i = \boldsymbol{z}) = p(\mathbf{x}_i = \boldsymbol{z}|\mathbf{x}_{i+1} = \boldsymbol{y}) \tag{3.20}$$

A Markov chain satisfies the detailed balance condition if and only if it is a reversible Markov chain [51]

$$A(\boldsymbol{y}, \boldsymbol{z})\pi(\boldsymbol{y}) = A(\boldsymbol{z}, \boldsymbol{y})\pi(\boldsymbol{z}). \tag{3.21}$$

If the above equation is satisfied, the chain has a stationary distribution π and is reversible with respect to π.

The Markov chain must sample from the stationary distribution $\pi(\cdot)$. If a chain

satisfies the three aforementioned properties, the chain will reach the stationary distribution after a burn-in period regardless of the starting position [47, 48, 52]. The stationary distribution satisfies

$$\int \pi(\boldsymbol{z})A(\boldsymbol{z},\boldsymbol{y})d\boldsymbol{z} = \pi(\boldsymbol{y}). \tag{3.22}$$

Metropolis-Hastings Algorithm

In the problem of Bayesian inference, the posterior distribution is known up to proportionality if the normalizing constant is unknown. Metropolis-Hastings (MH) algorithm permits sampling of the target distribution when the distribution is known up to proportionality using a simple transition kernel. From a starting point $\boldsymbol{y}$, a candidate point $\boldsymbol{z}$ is proposed from a proposal pdf $q(\boldsymbol{y},\boldsymbol{z})$. The candidate point is accepted with probability

$$\alpha(\boldsymbol{y},\boldsymbol{z}) = \min\left(1, \frac{\pi(\boldsymbol{z})q(\boldsymbol{z},\boldsymbol{y})}{\pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{z})}\right). \tag{3.23}$$

If the proposal distribution is symmetric, $q(\boldsymbol{y},\boldsymbol{z}) = q(\boldsymbol{z},\boldsymbol{y})$, the acceptance probability becomes

$$\alpha(\boldsymbol{y},\boldsymbol{z}) = \min\left(1, \frac{\pi(\boldsymbol{z})}{\pi(\boldsymbol{y})}\right). \tag{3.24}$$

The transition kernel is

$$A(\boldsymbol{y},\boldsymbol{z}) = \alpha(\boldsymbol{y},\boldsymbol{z})q(\boldsymbol{y},\boldsymbol{z}). \tag{3.25}$$

To satisfy the detailed balance condition,

$$\begin{aligned} A(\boldsymbol{y},\boldsymbol{z})\pi(\boldsymbol{y}) &= A(\boldsymbol{z},\boldsymbol{y})\pi(\boldsymbol{z}) \\ \alpha(\boldsymbol{y},\boldsymbol{z})q(\boldsymbol{y},\boldsymbol{z})\pi(\boldsymbol{y}) &= \alpha(\boldsymbol{z},\boldsymbol{y})q(\boldsymbol{z},\boldsymbol{y})\pi(\boldsymbol{z}). \end{aligned} \tag{3.26}$$

The challenge in this method is to find a good proposal distribution $q(\cdot,\cdot)$ that maximizes the efficiency of the chain. An automatic process to find the optimal proposal is presented next.

Adaptive Random-Walk Metropolis-Hastings

A form of Random-Walk Metropolis-Hastings MCMC will be used in which the proposal distribution is a Gaussian distribution centred at the current position of the chain with a given variance. The adaptation provides the optimal variance of the Gaussian proposal distribution. The burn-in algorithm is presented in Fig. 4.

The adaptation is performed using the past history of the chain [9, 53]. The proposal of a new sample of the Markov chain is normally distributed with a mean being the previous sample and a covariance matrix based on the covariance of the previous chain. The length of each adapted chain is based on the number of parameters and the strength of the nonlinearity of the system. The current position of the chain is $\mathbf{x}_{a,k}$ where a indicates the chain number in the adaptation procedure. A new sample $\boldsymbol{y}$ is proposed from

$$\boldsymbol{y} \sim \mathcal{N}(\mathbf{x}_k, C_t). \tag{3.27}$$

The covariance of the proposal C_t is based on the covariance of the previous chain $t-1$

$$C_t = s_d COV(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N). \tag{3.28}$$

A scaling factor s_d is used [54]

$$s_d = \frac{2.38^2}{d} \tag{3.29}$$

where d is the dimension of the target distribution. If the posterior distribution is unimodal, it maximizes the efficiency of the chain [54].

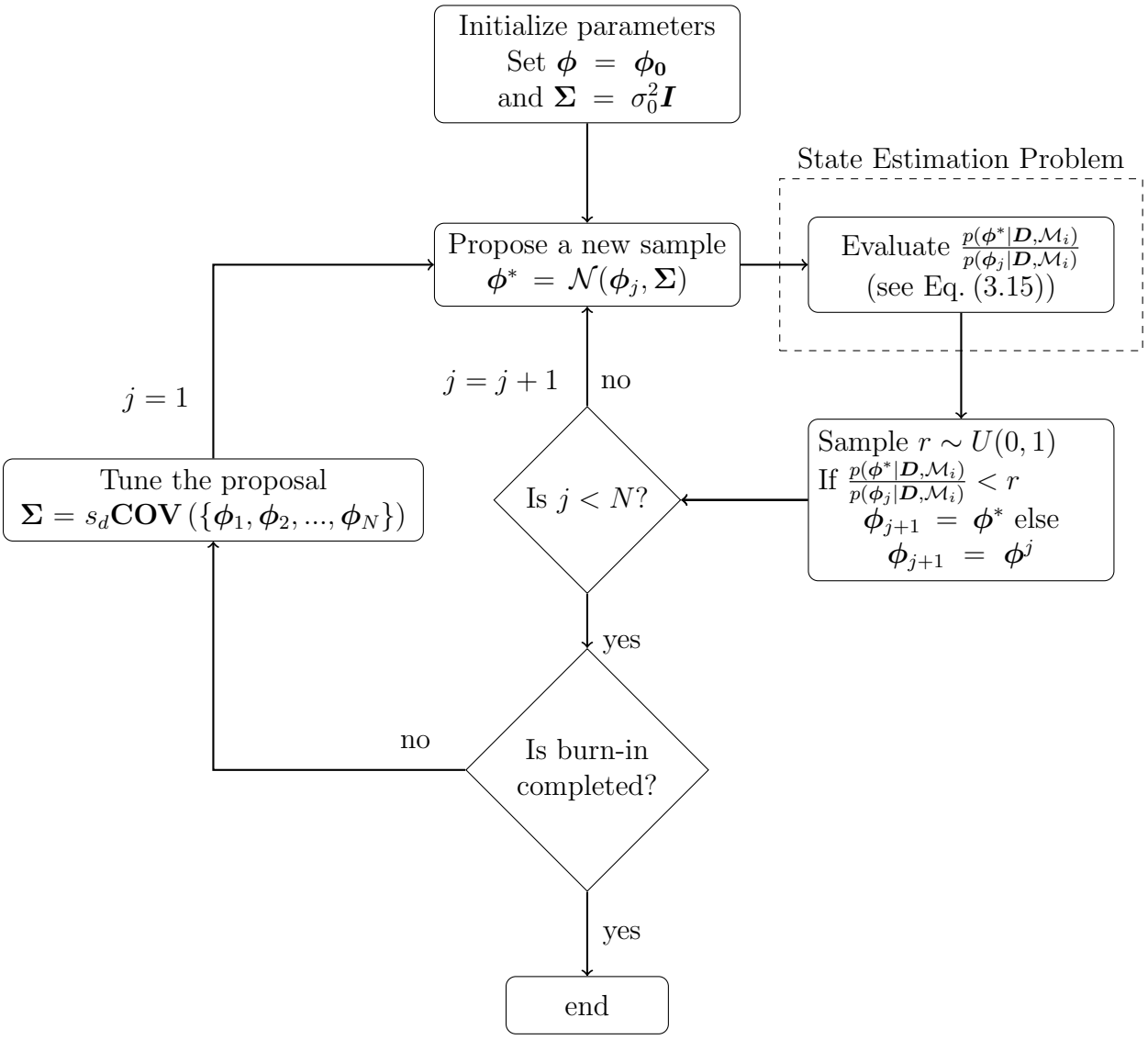


Figure 4: Flowchart of the burn-in procedure.

Delayed Rejection

Delayed rejection is modification made to MH where if a sample is rejected, a new sample can be proposed. Consider a Markov chain at the position $\boldsymbol{x}_k = \boldsymbol{y}$. The candidate point $\boldsymbol{z}_1$ generated from $q_1(\boldsymbol{y}, \cdot)$ and accepted with probability

$$\alpha_1(\boldsymbol{y}, \boldsymbol{z}_1) = min\left(1, \frac{\pi(\boldsymbol{z}_1) q_1(\boldsymbol{z}_1, \boldsymbol{y})}{\pi(\boldsymbol{y}) q_1(\boldsymbol{y}, \boldsymbol{z}_1)}\right) = min\left(1, \frac{N_1}{D_1}\right). \tag{3.30}$$

In the case of rejection, instead of setting $\boldsymbol{x}_{k+1} = \boldsymbol{y}$, a second candidate point $\boldsymbol{z}_2$ is proposed. The second state depends on the current position of the chain $\boldsymbol{y}$ but also on the value that was rejected. The acceptance probability of the second stage is

$$\alpha_2(\boldsymbol{y}, \boldsymbol{z}_1, \boldsymbol{z}_2) = min\left(1, \frac{\pi(\boldsymbol{z}_2)q_1(\boldsymbol{z}_2, \boldsymbol{z}_1)q_2(\boldsymbol{z}_2, \boldsymbol{z}_1, \boldsymbol{y})[1 - \alpha_1(\boldsymbol{z}_2, \boldsymbol{z}_1)]}{\pi(\boldsymbol{y})q_1(\boldsymbol{y}, \boldsymbol{z}_1)q_2(\boldsymbol{y}, \boldsymbol{z}_1, \boldsymbol{z}_2)[1 - \alpha_1(\boldsymbol{y}, \boldsymbol{z}_1)]}\right) = min\left(1, \frac{N_2}{D_2}\right) \tag{3.31}$$

where q_2 is another proposal density. Delaying rejection can be iterated up to i-th times and the acceptance probability of the i-th stage is [55]

$$\begin{aligned}\alpha_i(\boldsymbol{y}, \boldsymbol{z}_1, ..., \boldsymbol{z}_i) =& min(1, \frac{\pi(\boldsymbol{z}_i)q_1(\boldsymbol{z}_i, \boldsymbol{z}_{i-1})q_2(\boldsymbol{z}_i, \boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2})...q_i(\boldsymbol{z}_i, \boldsymbol{z}_{i-1}, ..., \boldsymbol{y})}{\pi(\boldsymbol{y})q_1(\boldsymbol{y}, \boldsymbol{z}_1)q_2(\boldsymbol{y}, \boldsymbol{z}_1, \boldsymbol{z}_2)...q_i(\boldsymbol{y}, \boldsymbol{z}_1, ..., \boldsymbol{z}_i)} \\ & \frac{[1 - \alpha_1(\boldsymbol{z}_i, \boldsymbol{z}_{i-1})][1 - \alpha_2(\boldsymbol{z}_i, \boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2})]...[1 - \alpha_{i-1}(\boldsymbol{z}_i, ..., \boldsymbol{z}_1)]}{[1 - \alpha_1(\boldsymbol{y}, \boldsymbol{z}_1)][1 - \alpha_2(\boldsymbol{y}, \boldsymbol{z}_1, \boldsymbol{z}_2)]...[1 - \alpha_{i-1}(\boldsymbol{y}, \boldsymbol{z}_1, ..., \boldsymbol{z}_i)]}) \\ & = min\left(1, \frac{N_i}{D_i}\right)\end{aligned} \tag{3.32}$$

where q_i denotes the proposal density of the i-th stage. The acceptance probability preserves the reversibility property of the chain. Delayed rejection can be stopped at any step.

Multiple strategies can be used when implementing DR. The number of trials can be fixed or a p-coin can be tossed after each rejection and if the outcome is head another point is then proposed [55].

Concluding remarks on MCMC methods

ARWMH and DR were used for model selection. Other techniques like Adaptive Metropolis [53] and Delayed Rejection Adaptive Metropolis [55] were implemented by the writer. It was found that there was no performance gain in using DRAM over ARWMH for the problems considered in this investigation. Delayed rejection was

sometimes necessary during burn-in when the proposal was too large to avoid the case where the rejection ratio was 100% making it impossible to adapt. These findings are reported in the literature [55]. In dimensions up to 20, no major differences are noticed in the results between these algorithms. ARWMH is used in this thesis for this reason, in addition to its simplicity in implementing Chib-Jeliazkov for evidence computation as discussed next.

3.3.2 Chib-Jeliazkov method

The evidence is estimated using Chib-Jeliazkov method [10]. The posterior ordinate is estimated using Metropolis-Hastings samples of the posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)$. From the parameter estimation procedure, M posterior samples have been generated $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_M\}$ with a proposal density $q(\boldsymbol{\phi}', \boldsymbol{\phi})$.

Using Eq. (3.23), the acceptance probability is

$$\alpha(\boldsymbol{\phi}, \boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i) = \min\left(1, \frac{p(\boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i)q(\boldsymbol{\phi}', \boldsymbol{\phi})}{p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)q(\boldsymbol{\phi}, \boldsymbol{\phi}')}\right). \tag{3.33}$$

From Eq. (3.25), the transition kernel is

$$p(\boldsymbol{\phi}, \boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i) = \alpha(\boldsymbol{\phi}, \boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i)q(\boldsymbol{\phi}, \boldsymbol{\phi}'). \tag{3.34}$$

To satisfy the detailed balance condition, Eq. (3.34) must respect

$$p(\boldsymbol{\phi}, \boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i)p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) = p(\boldsymbol{\phi}', \boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)p(\boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i). \tag{3.35}$$

Substituting Eq. (3.34) into (3.35), the detailed balance equation becomes

$$\alpha(\boldsymbol{\phi}, \boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i)q(\boldsymbol{\phi}, \boldsymbol{\phi}')p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) = \alpha(\boldsymbol{\phi}', \boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i)q(\boldsymbol{\phi}', \boldsymbol{\phi})p(\boldsymbol{\phi}'|\boldsymbol{D}, \mathcal{M}_i). \tag{3.36}$$

Eq. (3.36) is valid for any $\boldsymbol{\phi}'$. In that case, one can select $\boldsymbol{\phi}' = \boldsymbol{\phi}^*$. Substituting $\boldsymbol{\phi}' = \boldsymbol{\phi}^*$ and integrating both sides of the equation, one gets [7, 56]

$$\int \alpha(\boldsymbol{\phi}, \boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i) q(\boldsymbol{\phi}, \boldsymbol{\phi}^*) p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) d\boldsymbol{\phi} = \int \alpha(\boldsymbol{\phi}^*, \boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) q(\boldsymbol{\phi}^*, \boldsymbol{\phi}) p(\boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i) d\boldsymbol{\phi}. \tag{3.37}$$

The posterior ordinate $p(\boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i)$ can be taken out of the integral because it is a constant. Writing Eq. (3.37) in terms of the posterior ordinate, one gets

$$p(\boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i) = \frac{\int \alpha(\boldsymbol{\phi}, \boldsymbol{\phi}^*|\boldsymbol{D}, \mathcal{M}_i) q(\boldsymbol{\phi}, \boldsymbol{\phi}^*) p(\boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) d\boldsymbol{\phi}}{\int \alpha(\boldsymbol{\phi}^*, \boldsymbol{\phi}|\boldsymbol{D}, \mathcal{M}_i) q(\boldsymbol{\phi}^*, \boldsymbol{\phi}) d\boldsymbol{\phi}}. \tag{3.38}$$

The numerator integral is in terms of the parameters from the posterior distribution while the denominator integral is in terms of the parameters from the fixed proposal distribution $q(\boldsymbol{\phi}^*, \boldsymbol{\phi})$.

A pdf can be approximated by N samples with

$$p(\mathbf{x}) \approx \frac{1}{N} \sum_{s=1}^{N} \delta(\mathbf{x} - \mathbf{x}_s). \tag{3.39}$$

Applying this definition to the posterior distribution and the fixed proposal distribution in Eq. (3.38),

$$p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i) \approx \frac{\int \alpha(\boldsymbol{\phi},\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi},\boldsymbol{\phi}^*)\frac{1}{N}\sum_{s=1}^{N}\delta(\boldsymbol{\phi}-\boldsymbol{\phi}_s)d\boldsymbol{\phi}}{\int \alpha(\boldsymbol{\phi}^*,\boldsymbol{\phi}|\boldsymbol{D},\mathcal{M}_i)\frac{1}{M}\sum_{j=1}^{M}\delta(\boldsymbol{\phi}-\boldsymbol{\phi}_j)d\boldsymbol{\phi}} \tag{3.40}$$

$$\approx \frac{\frac{1}{N}\sum_{s=1}^{N}\int \alpha(\boldsymbol{\phi},\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi},\boldsymbol{\phi}^*)\delta(\boldsymbol{\phi}-\boldsymbol{\phi}_s)d\boldsymbol{\phi}}{\frac{1}{M}\sum_{j=1}^{M}\int \alpha(\boldsymbol{\phi}^*,\boldsymbol{\phi}|\boldsymbol{D},\mathcal{M}_i)\delta(\boldsymbol{\phi}-\boldsymbol{\phi}_j)d\boldsymbol{\phi}} \tag{3.41}$$

$$\approx \frac{\frac{1}{N}\sum_{s=1}^{N}\alpha(\boldsymbol{\phi}_s,\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi}_s,\boldsymbol{\phi}^*)}{\frac{1}{M}\sum_{j=1}^{M}\alpha(\boldsymbol{\phi}^*,\boldsymbol{\phi}_j|\boldsymbol{D},\mathcal{M}_i)}. \tag{3.42}$$

Using the definition of the acceptance probability in Eq. (3.23), the previous equation becomes

$$p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i) \approx \frac{\frac{1}{N}\sum_{s=1}^{N}\min\left(1,\frac{p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi}^*,\boldsymbol{\phi}_s)}{p(\boldsymbol{\phi}_s|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi}_s,\boldsymbol{\phi}^*)}\right)q(\boldsymbol{\phi}_s,\boldsymbol{\phi}^*)}{\frac{1}{M}\sum_{j=1}^{M}\min\left(1,\frac{p(\boldsymbol{\phi}_j|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi}_j,\boldsymbol{\phi}^*)}{p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)q(\boldsymbol{\phi}^*,\boldsymbol{\phi}_j)}\right)}. \tag{3.43}$$

Eq. (3.43) can be simplified. In this work, the number of draws for the denominator and numerator is the same ($N = M$). Furthermore, the proposal distribution $q(\cdot,\cdot)$ is Gaussian and symmetric. The simplified equation used to estimate the posterior ordinate is

$$p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i) \approx \frac{\sum_{s=1}^{N}\min\left(1,\frac{p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)}{p(\boldsymbol{\phi}_s|\boldsymbol{D},\mathcal{M}_i)}\right)q(\boldsymbol{\phi}_s,\boldsymbol{\phi}^*)}{\sum_{j=1}^{N}\min\left(1,\frac{p(\boldsymbol{\phi}_j|\boldsymbol{D},\mathcal{M}_i)}{p(\boldsymbol{\phi}^*|\boldsymbol{D},\mathcal{M}_i)}\right)} \tag{3.44}$$

where the ratios of posterior ordinates are evaluated using Eq. (3.14). To evaluate this ratio, a state estimation procedure is carried out as discussed in the next chapter.

Chapter 4

State estimation

4.1 Introduction

In the previous chapter, the parameter estimation procedure is formulated. It was shown that the parameters posterior distribution includes a state estimation procedure. In this chapter, different state estimation algorithms are described. An example demonstrates the performance of the filters against the strength of nonlinearities.

From the previous chapters, we restate the modelling and measurement equations as

$$\mathbf{x}_{k+1} = g_k\left(\mathbf{x}_k, \boldsymbol{f}_k, \boldsymbol{q}_k\right), \tag{4.1}$$

$$\boldsymbol{d}_j = h_j(\mathbf{x}_{d(j)}, \boldsymbol{\varepsilon}_j). \tag{4.2}$$

The filtering procedure is divided into two main steps. The first step is the forecast. In the forecast step, the pdf of the system state is marched in time using a model representing the physics of the system of the form seen in Eq. (4.1). The update step consists of blending an incoming observation of the system with the current knowledge using Eq.(4.2). A general overview of the state estimation procedure is shown in Fig. 5.

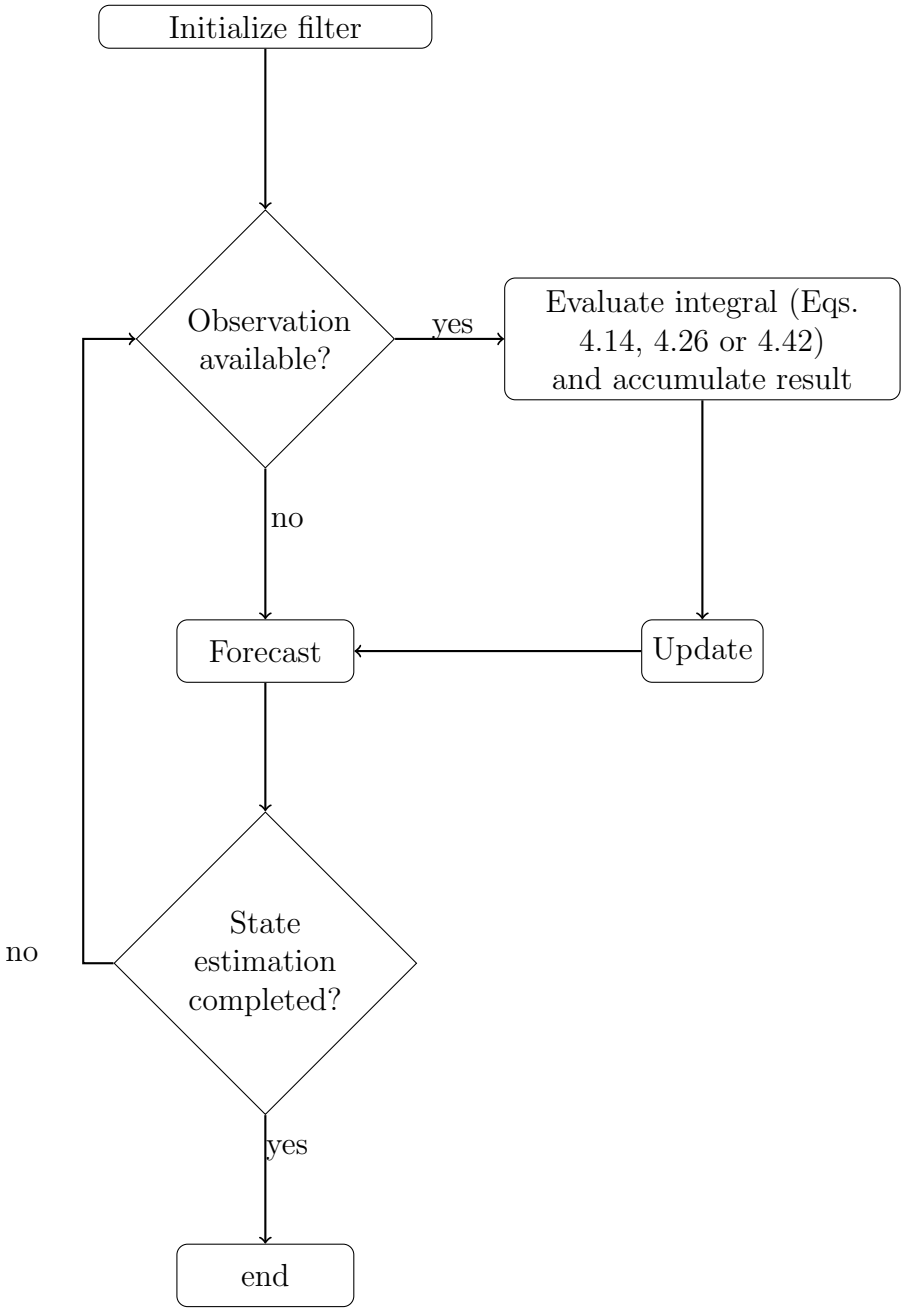


Figure 5: Flowchart of the state-estimation procedure.

For the nonlinear state estimation, the Extended Kalman filter (EKF), Ensemble Kalman filter (EnKF) and Particle filter using EnKF for its proposal density are considered in this investigation.

4.2 Extended Kalman Filter

Details of the EKF are widely available (i.e. [1, 4, 45, 57–60]) but a brief description will be presented next for completeness. EKF linearizes the nonlinear model and measurement operators in Eqs. (4.2,4.1). In the forecast step, the physics of the system is considered through the model operator. The next state of the system is predicted using the previous state and the proposed model. In the update stage, the predicted state is combined with an incoming observation. The Kalman gain matrix achieves the balance between the predicted state and the incoming observation. For the following, the measurement noise is $\epsilon_j \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_j)$ and the model noise is $q_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_k)$.

When an observation is available, the update step takes place [59, 60]

$$\boldsymbol{C}_j = \left.\frac{\partial h(\mathbf{x}_{d(j)}, \boldsymbol{\varepsilon}_j)}{\partial \mathbf{x}_{d(j)}}\right|_{\mathbf{x}_{d(j)}=\mathbf{x}^f_{d(j)}, \boldsymbol{\varepsilon}_j=\mathbf{0}} \tag{4.3}$$

$$\boldsymbol{D}_j = \left.\frac{\partial h(\mathbf{x}_{d(j)}, \boldsymbol{\varepsilon}_j)}{\partial \boldsymbol{\varepsilon}_j}\right|_{\mathbf{x}_{d(j)}=\mathbf{x}^f_{d(j)}, \boldsymbol{\varepsilon}_j=\mathbf{0}} \tag{4.4}$$

$$\boldsymbol{K}_{d(j)} = \boldsymbol{P}^f_{d(j)} \boldsymbol{C}_j^T \left[\boldsymbol{C}_j \boldsymbol{P}^f_{d(j)} \boldsymbol{C}_j^T + \boldsymbol{D}_j \mathbf{\Gamma}_j \boldsymbol{D}_j^T\right]^{-1} \tag{4.5}$$

$$\mathbf{x}^a_{d(j)} = \mathbf{x}^f_{d(j)} + \boldsymbol{K}_{d(j)}(\boldsymbol{d}_j - \boldsymbol{h}(\mathbf{x}^f_{d(j)}, \mathbf{0})) \tag{4.6}$$

$$\boldsymbol{P}^a_{d(j)} = (\boldsymbol{I} - \boldsymbol{K}_j \boldsymbol{C}_j) \boldsymbol{P}^f_{d(j)}. \tag{4.7}$$

The forecast is given by [59, 60]

$$\boldsymbol{A}_k = \left.\frac{\partial g(\mathbf{x}_k, \boldsymbol{f}_k, \boldsymbol{q}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\mathbf{x}^f_k, \boldsymbol{q}_k=\mathbf{0}} \tag{4.8}$$

$$\boldsymbol{B}_k = \left.\frac{\partial g(\mathbf{x}_k, \boldsymbol{f}_k, \boldsymbol{q}_k)}{\partial \boldsymbol{q}_k}\right|_{\mathbf{x}_k=\mathbf{x}^f_k, \boldsymbol{q}_k=\mathbf{0}} \tag{4.9}$$

$$\mathbf{x}^f_{k+1} = g_k(\mathbf{x}_k, \boldsymbol{f}_k, \mathbf{0}) \tag{4.10}$$

$$\boldsymbol{P}^f_{k+1} = \boldsymbol{A}_k \boldsymbol{P}^a_k \boldsymbol{A}_k^T + \boldsymbol{B}_k \boldsymbol{Q}_k \boldsymbol{B}_k^T. \tag{4.11}$$

The integral of Eq. (3.16) can be performed analytically since the integral involves the product of the likelihood $p(\boldsymbol{d}_j|\mathbf{x}_{d(j)}, \boldsymbol{\phi}, \mathcal{M}_i)$ and the current pdf of the system state $p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i)$. The likelihood can be written as follows

$$p(\boldsymbol{d}_j|\mathbf{x}_{d(j)}, \boldsymbol{\phi}, \mathcal{M}_i) = \mathcal{N}(\boldsymbol{d}_j|h_j(\mathbf{x}_{d(j)}, \mathbf{0}), \boldsymbol{\Gamma}_j). \tag{4.12}$$

The current pdf of the system state is also normally distributed

$$p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i) = \mathcal{N}(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{P}^f_{d(j)}). \tag{4.13}$$

The likelihood function is Eq. (3.16) becomes

$$p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i) = \prod_{j=1}^{J} \int_{-\infty}^{\infty} \mathcal{N}(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{P}^f_{d(j)}) \mathcal{N}(\boldsymbol{d}_j|h_j(\mathbf{x}_{d(j)}, \mathbf{0}), \boldsymbol{\Gamma}_j). \tag{4.14}$$

When the conditional pdf of the state is weakly non-Gaussian or Gaussian, the Extended Kalman filter (EKF) can efficiently estimate $p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i)$ [4]. However, EKF fails when the distribution becomes strongly non-Gaussian, for instance, exhibiting multiple modes due to the presence of multiple equilibrium points in the dynamical model. A non-Gaussian filter like the Ensemble Kalman filter or a particle filter is necessary.

4.3 Ensemble Kalman Filter

The Ensemble Kalman filter addresses two major limitations of EKF. The first one is the use of an approximate closure and the second limitation is the computational cost of carrying the error covariance matrix in large dimensional space [1]. The closure stems from the fact that higher-order moments in the error covariance are discarded leading to an unbound variance growth in some cases. In the Ensemble Kalman filter

(EnKF), the initial state ensemble matrix is generated by drawing samples from the prior pdf of the initial state vector. When a measurement is available, the update takes place as [46]

$$\boldsymbol{d}_{j,i} = h(\boldsymbol{x}^f_{d(j),i}, \boldsymbol{\epsilon}_{j,i}) \tag{4.15}$$

$$\bar{\boldsymbol{d}}_j = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{d}_{j,i} \tag{4.16}$$

$$\bar{\mathbf{x}}^f_{d(j)} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}^f_{d(j),i} \tag{4.17}$$

$$\boldsymbol{P}_{xd} = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{x}^f_{d(j),i} - \bar{\mathbf{x}}^f_{d(j)})(\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}}_j)^T \tag{4.18}$$

$$\boldsymbol{P}_{dd} = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}}_j)(\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}}_j)^T \tag{4.19}$$

$$\boldsymbol{K}_{d(j)} = \boldsymbol{P}_{xd}\boldsymbol{P}^{-1}_{dd} \tag{4.20}$$

$$\mathbf{x}^a_{d(j),i} = \mathbf{x}^f_{d(j),i} + \boldsymbol{K}_{d(j)}(\boldsymbol{d}_j - \boldsymbol{d}_{j,i}). \tag{4.21}$$

Each sample is then propagated in time using the full nonlinear model in Eq. (3.10)

$$\mathbf{x}^f_{k+1,i} = g(\mathbf{x}^a_{k,i}, \boldsymbol{f}_k, \boldsymbol{q}_{k,i}). \tag{4.22}$$

The forecast mean of the state vector is

$$\hat{x}^f_k = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}^f_{k,i}. \tag{4.23}$$

The conditional distribution of the state can be approximated by [61, 62]

$$p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i) \approx \frac{1}{N}\sum_{s=1}^{N} \delta(\mathbf{x}_k - \mathbf{x}^f_{k,s}). \tag{4.24}$$

Using Eq. (4.24), the likelihood function in Eq. (3.16) becomes [46]

$$
\begin{aligned}
p(\boldsymbol{D}|\boldsymbol{\phi},\mathcal{M}_i) &= \prod_{j=1}^{J}\int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)}|\mathbf{x}_{d(j)-1},\boldsymbol{\phi},\mathcal{M}_i)p(\boldsymbol{d}_j|\mathbf{x}_{d(j)},\boldsymbol{\phi},\mathcal{M}_i)d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J}\int_{-\infty}^{\infty} \frac{1}{N}\sum_{s=1}^{N}\delta(\mathbf{x}_k-\mathbf{x}_{k,s}^{f})p(\boldsymbol{d}_j|\mathbf{x}_{d(j)},\boldsymbol{\phi},\mathcal{M}_i)d\mathbf{x}_{d(j)} \\
&\approx \prod_{j=1}^{J}\left[\frac{1}{N}\sum_{s=1}^{N} p(\boldsymbol{d}_j|\mathbf{x}_{d(j),s}^{f},\boldsymbol{\phi},\mathcal{M}_i)d\mathbf{x}_{d(j)}\right]. \qquad (4.25)
\end{aligned}
$$

(4.26)

EnKF demonstrates superiority over the Extended Kalman filter for strongly non-Gaussian systems as it avoids the linearization of the model and measurement operators. For strongly non-Gaussian systems, the Gaussian update step adopted from the Extended Kalman filter shown in Eq. (4.20) may sometimes lead to unwieldy results in EnKF [46].

4.4 Particle Filter using an EnKF proposal

The particle filters are fully non-Gaussian [62]. For EnKF, each sample has a weight of 1/N. For the particle filter (PF), the particles have varying weights which correspond to their likelihood, the proposal distribution and their previous position. The prediction step is the same for the particle filter and the EnKF filter. The difference is the update step. In the update step, the weight of each particle is modified according to the following relation [62, 63]

$$
w_{d(j),i} \propto w_{d(j)-1,i}\frac{p(\boldsymbol{d}_j|\mathbf{x}_{d(j),i})p(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i})}{q(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i},\boldsymbol{d}_j)} \qquad (4.27)
$$

where $p(\boldsymbol{d}_j|\mathbf{x}_{d(j),i})$ is the likelihood, $p(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i})$ is the prior and $q(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i}, \boldsymbol{d}_j)$ is the proposal distribution. Various proposal distributions are reported in the literature [46, 62]. In this thesis, two different distributions are presented. If the proposal distribution is the same as the prior distribution, the particle filter is called bootstrap filter [62, 64]. The weight update equation becomes

$$w_{d(j),i} \propto w_{d(j)-1,i} p(\boldsymbol{d}_j|\mathbf{x}_{d(j),i}). \tag{4.28}$$

As $q(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i}, \boldsymbol{d}_j) = p(\mathbf{x}_{d(j),i}|\mathbf{x}_{d(j)-1,i})$, the effect of the last measurement $\boldsymbol{d}_j$ is not used to update the state of the particle in the bootstrap filter. Another option is to select the proposal given by an EnKF filter. The proposal retains some nonlinearity in contrast to a Gaussian proposal from EKF or UKF. Furthermore, EnKF, depending on the ensemble size, can be more efficient than EKF since all the particles share the same proposal. For the EnKF based proposal, the particles are nudged using the new observation before updating the particle weights.

The update step of the particle filter with EnKF proposal is [46]

$$\boldsymbol{d}_{j,i} = h(\boldsymbol{x}^f_{d(j),i}, \boldsymbol{\epsilon}_{j,i}) \tag{4.29}$$

$$\bar{\boldsymbol{d}_j} = \sum_{i=1}^{N} w_{d(j)-1,i} \boldsymbol{d}_{j,i} \tag{4.30}$$

$$\bar{\mathbf{x}}^f_{d(j)} = \sum_{i=1}^{N} w_{d(j)-1,i} \mathbf{x}^f_{d(j),i} \tag{4.31}$$

$$\boldsymbol{P}_{xd} = \frac{1}{1 - \sum_{i=1}^{N} w^2_{d(j)-1,i}} \sum_{i=1}^{N} w_{d(j)-1,i} (\mathbf{x}^f_{d(j),i} - \bar{\mathbf{x}}^f_{d(j)})(\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}_j})^T \tag{4.32}$$

$$\boldsymbol{P}_{dd} = \frac{1}{1 - \sum_{i=1}^{N} w^2_{d(j)-1,i}} \sum_{i=1}^{N} w_{d(j)-1,i} (\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}_j})(\boldsymbol{d}_{j,i} - \bar{\boldsymbol{d}_j})^T \tag{4.33}$$

$$\boldsymbol{K}_{d(j)} = \boldsymbol{P}_{xd} \boldsymbol{P}^{-1}_{dd} \tag{4.34}$$

$$\mathbf{x}^a_{d(j),i} = \mathbf{x}^f_{d(j),i} + \boldsymbol{K}_{d(j)} (\boldsymbol{d}_j - \boldsymbol{d}_{j,i}) \tag{4.35}$$

$$\tilde{w}_{d(j),i} = w_{d(j)-1,i} p(\boldsymbol{d}_j | \boldsymbol{x}^a_{d(j),i}) \frac{\sum_{l=1}^{N} w_{d(j)-1,l} \frac{1}{|\boldsymbol{H}_f|} K(\boldsymbol{H}_f^{-1}(\mathbf{x}^f_{d(j),l} - \mathbf{x}^a_{d(j),i}))}{\sum_{l=1}^{N} \frac{1}{N} \frac{1}{|\boldsymbol{H}_a|} K(\boldsymbol{H}_a^{-1}(\mathbf{x}^a_{d(j),l} - \mathbf{x}^a_{d(j),i}))} \tag{4.36}$$

$$w_{d(j),i} = \frac{\tilde{w}_{d(j),i}}{\sum_{i=1}^{N} \tilde{w}_{d(j),i}}. \tag{4.37}$$

where $K(\cdot)$ is a Gaussian kernel and $\boldsymbol{H}_a$ and $\boldsymbol{H}_f$ are the bandwidth matrix for the proposal and the prior respectively. Following [65], these matrix are proportional to the square root of the ensemble covariance matrix and the constant of proportionality

is the multivariate form of Scott's rule [66]. The bandwidth matrix are defined as

$$\boldsymbol{H}_a = N^{\frac{-1}{n+4}} \left[\frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}^a_{d(j),i} - \bar{\mathbf{x}}^a_{d(j)})(\mathbf{x}^a_{d(j),i} - \bar{\mathbf{x}}^a_{d(j)})^T \right]^{\frac{1}{2}} \tag{4.38}$$

$$\boldsymbol{H}_f = N_{eff}^{\frac{-1}{n+4}} \left[\frac{1}{1 - \sum_{i=1}^{N} w^2_{d(j)-1,i}} \sum_{i=1}^{N} w_{d(j)-1,i} (\mathbf{x}^f_{d(j),i} - \bar{\mathbf{x}}^f_{d(j)})(\mathbf{x}^f_{d(j),i} - \bar{\mathbf{x}}^f_{d(j)})^T \right]^{\frac{1}{2}} \tag{4.39}$$

where n is the dimension of the state vector and N_{eff} is the effective ensemble size defined in Eq.(4.43). For forecast, the full nonlinear model in Eq. (3.10) is applied to each particle

$$\mathbf{x}^f_{k+1,i} = g(\mathbf{x}^a_{k,i}, \boldsymbol{f}_k, \boldsymbol{q}_{k,i}). \tag{4.40}$$

The probability density function is approximated by the weighted particles as

$$p(\mathbf{x}_{d(j)} | \mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i) \approx \sum_{s=1}^{N} w_{d(j),s} \delta(\mathbf{x}_{d(j)} - \mathbf{x}_{d(j),s}). \tag{4.41}$$

By using the previous equation, the integral of Eq. (3.16) can be evaluated as follows [4]

$$\begin{aligned} p(\boldsymbol{D}|\boldsymbol{\phi}, \mathcal{M}_i) &= \prod_{j=1}^{J} \int_{-\infty}^{\infty} p(\mathbf{x}_{d(j)} | \mathbf{x}_{d(j)-1}, \boldsymbol{\phi}, \mathcal{M}_i) p(\boldsymbol{d}_j | \mathbf{x}_{d(j)}, \boldsymbol{\phi}, \mathcal{M}_i) d\mathbf{x}_{d(j)} \\ &\approx \prod_{j=1}^{J} \int_{-\infty}^{\infty} \sum_{s=1}^{N} w_{d(j),s} \delta(\mathbf{x}_{d(j)} - \mathbf{x}_{d(j),s}) p(\boldsymbol{d}_j | \mathbf{x}_{d(j)}, \boldsymbol{\phi}, \mathcal{M}_i) d\mathbf{x}_{d(j)} \\ &\approx \prod_{j=1}^{J} \sum_{s=1}^{N} w_{d(j),s} p(\boldsymbol{d}_j | \mathbf{x}_{d(j),s}, \boldsymbol{\phi}, \mathcal{M}_i). \end{aligned} \tag{4.42}$$

Degeneracy is a common problem with particle filters [62]. It is the phenomenon where the particles are clustered in the high probability region. The particles in the low probability region have zero weight. A measure of this phenomenon is the

effective ensemble size described with [67]

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} w_{d(j),i}^2}. \tag{4.43}$$

One way to reduce the degeneracy is to increase the number of particles which may not always be possible. Another way is to resample when the effective ensemble size is below a certain threshold. Several resampling algorithms are available [63, 68]. In this work, the multinomial resampling [68, 69] has been successfully used.

4.4.1 Resampling algorithm

The multinomial resampling algorithm is as follows [46, 68, 69]:

1. Draw N independent samples from the uniform distribution $\mathcal{U}(0, 1)$
2. Obtain the cumulative density of the weights

$$v_j = \sum_{l=1}^{j} w_{k,l} \tag{4.44}$$

3. Move the particles

for $m = 0$ **to** N **do**

Find n such that $v_n \leq u_m < v_{n+1}$

Copy the nth particle into the new ensemble

end for

4. Reset all the weights to $1/N$

The resampling algorithm can be executed after each update iteration or when the effective sample size shown in Eq. (4.43) falls below a certain threshold. A common one is $N_{eff}/3$ [70]. If all weights are equal, N_{eff} is the same as the ensemble size N.

The effective ensemble size decreases as the particles becomes more degenerate. The resampling step is necessary for the good operation of the particle filter. Use of the particles after a resampling operation lead to a poor estimate, as the higher weighted particles are duplicated. This results in a loss of diversity that overestimates the high probability regions [68].

4.4.2 Performance

The major drawback of the PF-EnKF is its computational cost. To reduce the execution time, the distributed implementation of the filter is critical in this investigation to exploit high performance computing. The next table summarizes the simulation time needed for one evaluation of the posterior pdf in Eq. (3.14) for a particle filter with 1000 particles. There are 2001 predict operations, 101 update operations, 101 integrals evaluation and the necessary resampling.

Table 1: Execution time of the state estimation procedure using PF-EnKF

Number of cores	Particles per cores	Time [s]
1	1000	12.49
2	500	6.36
4	250	3.35
5	200	2.95
8	125	1.82
10	100	1.57
20	50	1.09
25	40	0.95
40	25	0.77
50	20	0.87
100	10	1.08
125	8	1.24

The simulation time decreases with an increasing number of cores. When the number of particles per core is smaller than 25, the simulation time increases with the number of cores. This is due to the communication cost between all the cores.

The results of parameter estimation and model selection are tied to the state estimation procedure of the different filters. The next section presents a problem where the impact of nonlinearity is easily realized and serves as a benchmark to compare the different filters.

4.5 Example: Hidden state estimation

The following example is based on a problem presented by Evensen in [1]. Consider a system defined by the following model

$$x = g(\alpha, q) \text{ where } q \sim \mathcal{N}(0, \boldsymbol{Q}). \tag{4.45}$$

The state of the system is x and is function of the unknown parameter α. The algorithm described in the previous chapters can be used to estimate α by sampling from the parameter posterior distribution using a MCMC chain. There is another way to estimate this parameter. By augmenting the state vector with the parameter, one could use a filter to estimate both the state and the parameter at the same time. This second approach is faster but may not be as robust. The state vector augmented with the parameter is

$$\boldsymbol{u} = \begin{Bmatrix} x \\ \\ \alpha \end{Bmatrix}. \tag{4.46}$$

The goal of this example is to gain insight into the update steps of various nonlinear

filters. The parameter of the system has a Gaussian error α'

$$\alpha = \alpha_o + \eta \text{ where } \eta \sim \mathcal{N}(0, \boldsymbol{P}^f). \tag{4.47}$$

Two cases are considered for the function $g(\alpha, q)$. The first case is linear and the second case is nonlinear.

1. $x = \alpha + q$

2. $x = cos(\alpha) + q$

The state x is measured using $d = h(x, \varepsilon)$ where $\varepsilon \sim \mathcal{N}(0, \Gamma)$. Two measurements operators will be considered:

1. $d = x + \varepsilon$

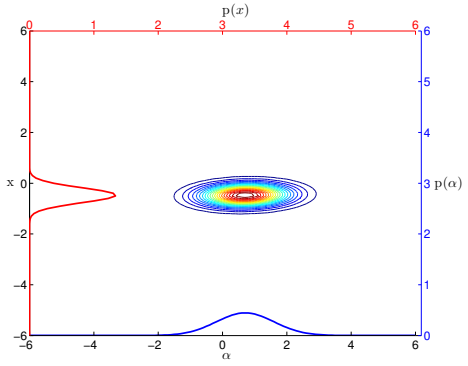
2. $d = cos(x) + \varepsilon$

The goal of this example is to compare the probability density function of the state given a measurement of $d = -0.5$ provided by a full Bayesian analysis and the filters previously introduced in this chapter. Using Bayes' theorem, the posterior pdf is

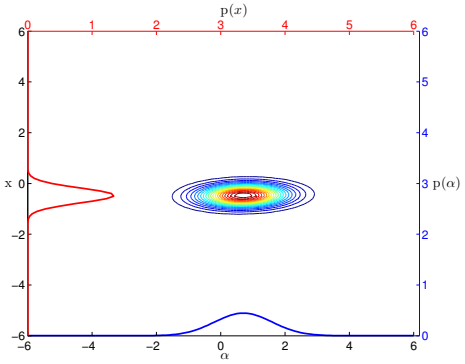
$$p(\boldsymbol{u}|d) \propto e^{-0.5\left[\frac{(d-h(x,0))^2}{\Gamma} + \frac{(\alpha-\alpha_i)^2}{\boldsymbol{P}^f} + \frac{(x-g(\alpha,0))^2}{\boldsymbol{Q}}\right]} \tag{4.48}$$

where the normalizing constant can be estimated by evaluating the volume under the surface.

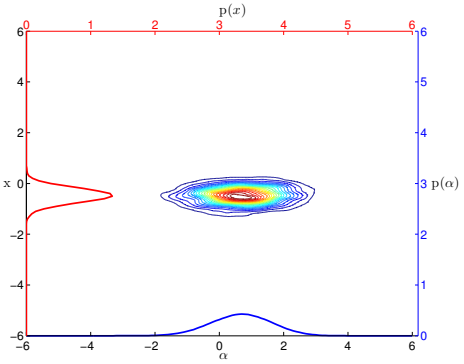
The results of the parameter and state estimation are shown in Figs. 6-9. The EnKF, the bootstrap filter and PF-EnKF use 2000 particles. Figure 6 presents the joint posterior pdf of $\boldsymbol{u}$ and the marginal distribution of x and α when the model and the measurement operators are linear. The results are quite similar no matter which filters are used. The bootstrap filter estimate has a higher variance than its PF-EnKF counterpart. This can be explained by the fact that the update step of



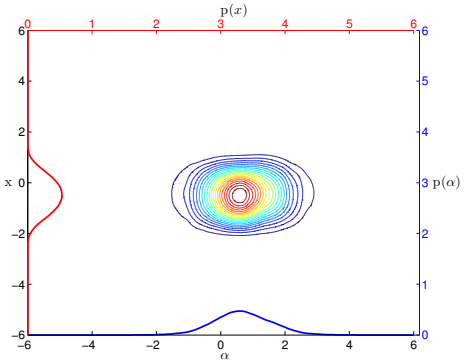
(a) Full Bayesian



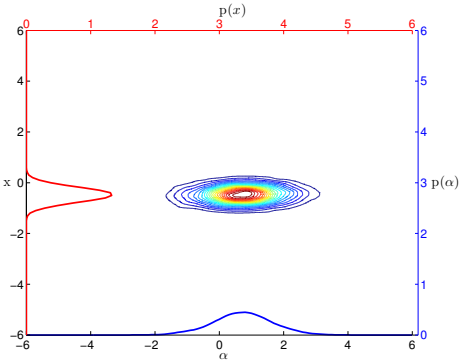
(b) EKF filter



(c) EnKF filter



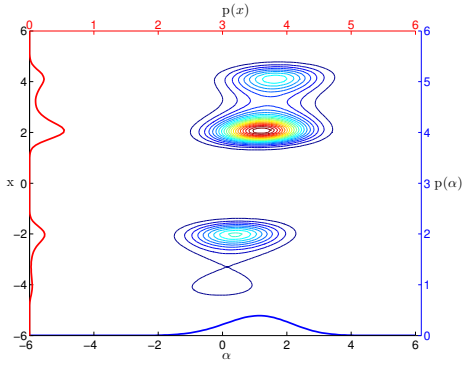
(d) Bootstrap filter



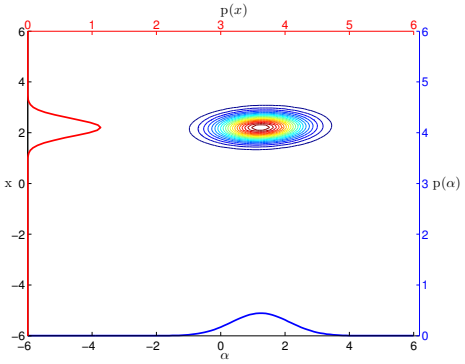
(e) PF-EnKF filter

Figure 6: Posterior pdf using various filters for case I-I.

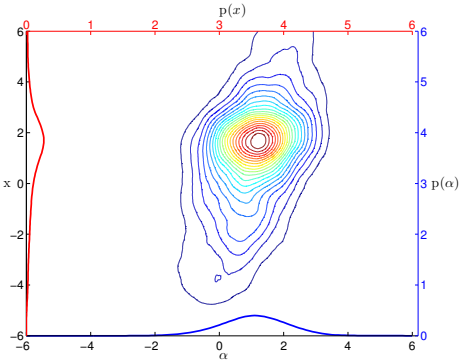
the proposal distribution of the bootstrap filter does not include the measurement $d = -0.5$. Hence, a higher number of particles are required to have the same level of accuracy as PF-EnKF. In Fig. 7, the parameters are estimated when the measurement operator is nonlinear and the model operator remains linear.



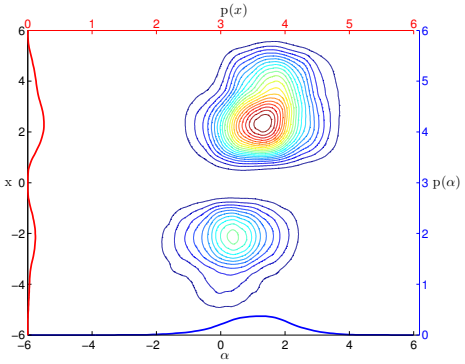
(a) Full Bayesian



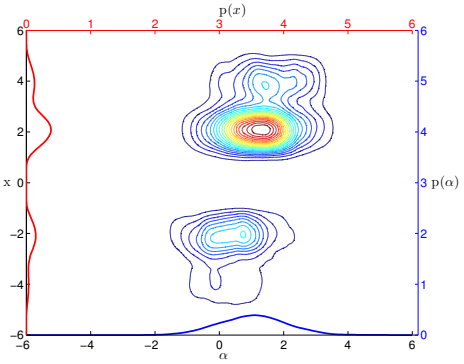
(b) EKF filter



(c) EnKF filter



(d) Bootstrap filter

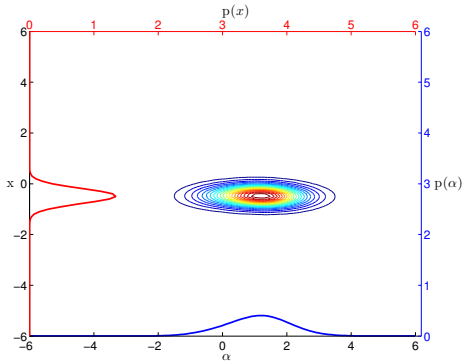


(e) PF-EnKF filter

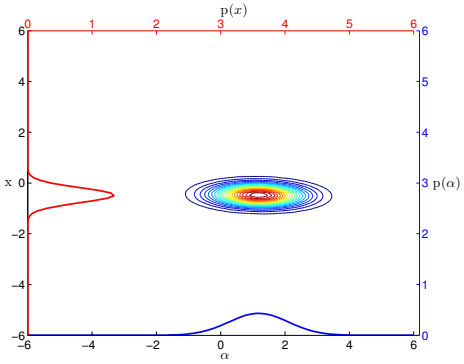
Figure 7: Posterior pdf using various filters for case I-II.

The parameter α is unimodal and appropriately estimated by all filters even though the joint distribution is multimodal. The state x is best estimated by PF-EnKF. The estimate of EnKF is unimodal and the estimate of the bootstrap filter

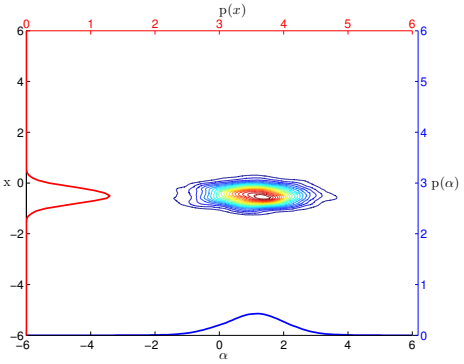
does not correctly model the tail of the distribution. Contrary to EKF, EnKF estimate encompasses the true distribution of x. In the Fig 8, the same exercise is repeated but with linear measurement and a nonlinear model operator.



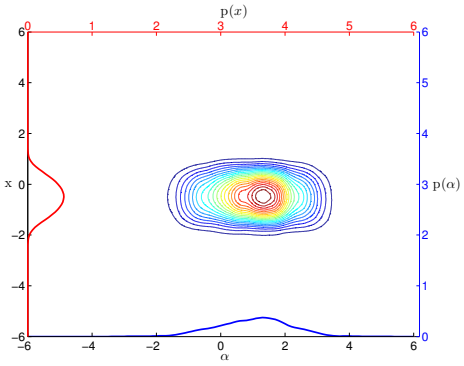
(a) Full Bayesian



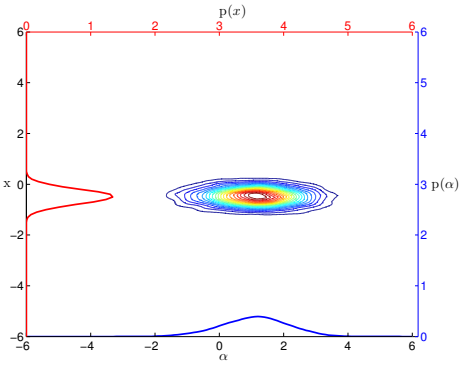
(b) EKF filter



(c) EnKF filter



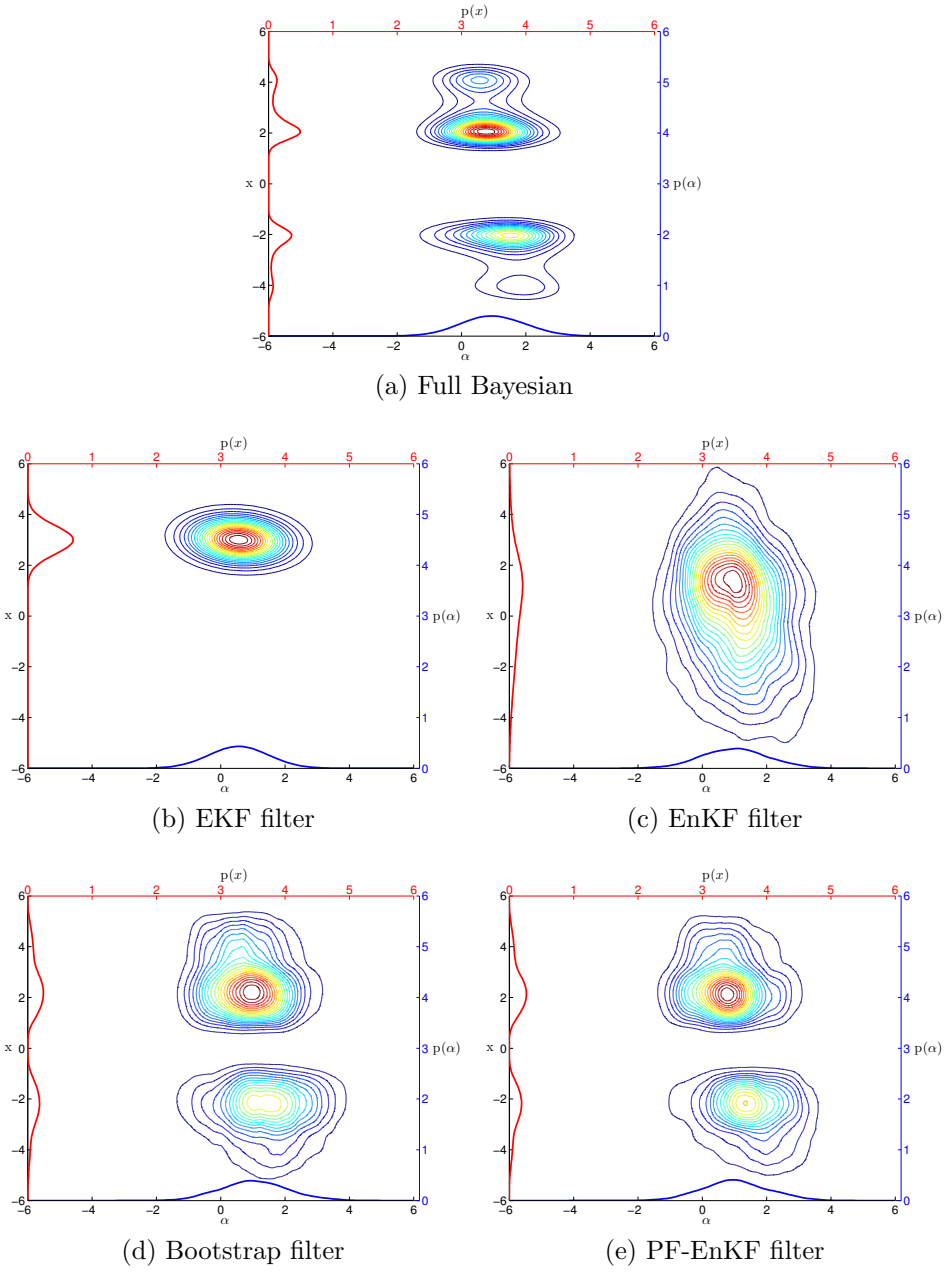
(d) Bootstrap filter



(e) PF-EnKF filter

Figure 8: Posterior pdf using various filters for case II-I.

Even though the model operator is nonlinear, the joint pdf of the state and parameter distribution is unimodal and nearly Gaussian. Finally, Fig. 9 presents the results when the model and measurement operators are nonlinear.



(a) Full Bayesian

(b) EKF filter

(c) EnKF filter

(d) Bootstrap filter

(e) PF-EnKF filter

Figure 9: Posterior pdf using various filters for case II-II.

The same observations can be made when the model operator is linear and the

measurement operator is nonlinear. EKF tracks the highest mode and EnKF blends multiple modes together. EnKF is well suited as a proposal for the particle filter since the posterior distribution support encompasses the true distribution of the parameters.

Chapter 5

Numerical application

5.1 Validation

The state estimation procedure will be validated in a simple linear system with additive Gaussian model and measurement noises where the Extended Kalman filter, the Ensemble Kalman filter and the particle filter using EnKF proposal results will be compared.

Consider the linear dynamical system driven by Gaussian white noise $W(t)$ of the form

$$\dot{x} = W(t). \tag{5.1}$$

Eq (5.1) can be rewritten as

$$x_{k+1} = x_k + \sqrt{\Delta_t} q_k \text{ where } q_k \sim \mathcal{N}(0, 1). \tag{5.2}$$

If x_k can be represented by a Gaussian pdf, x_{k+1} is also normally distributed as the addition of two Gaussian random variables is also Gaussian. The Extended Kalman filter, having the same form as the Kalman filter in this case, will provide optimal state estimation and can be used to validate the other filters.

The measurement equation is

$$d_j = x_{d(j)} + \epsilon_j \text{ where } \epsilon_j \sim \mathcal{N}(0, 0.05). \tag{5.3}$$

The log of the evidence for the same set of data is recorded in table 2.

Table 2: Evidence of simple linear system

State Estimation	$\ln(p(\boldsymbol{D}\|\mathcal{M}))$
EKF (KF)	126.69
PF-EnKF	126.62
EnKF	126.61

Next we carry out a simple model selection problem. The effect of the number of measurements on the evidence is investigated. Using the system shown in Eq. (5.2) and the measurement equation of (5.3), the proposed models are

$$\mathcal{M}_1 : \dot{x} = \sigma W(t), \tag{5.4}$$

$$\mathcal{M}_2 : \dot{x} = a_1 + \sigma W(t), \tag{5.5}$$

$$\mathcal{M}_3 : \dot{x} = a_2 + \sigma W(t), \tag{5.6}$$

$$\mathcal{M}_4 : \dot{x} = a_1 + a_2 x + \sigma W(t), \tag{5.7}$$

$$\tag{5.8}$$

where $W(t)$ is a Gaussian white noise process. The discretized form of each proposed

model is

$$\mathcal{M}_1 : x_{k+1} = x_k + \sigma\sqrt{\Delta t} q_k, \tag{5.9}$$

$$\mathcal{M}_2 : x_{k+1} = x_k + a_1 \Delta t + \sigma\sqrt{\Delta t} q_k, \tag{5.10}$$

$$\mathcal{M}_3 : x_{k+1} = x_k + a_2 x_k \Delta t + \sigma\sqrt{\Delta t} q_k, \tag{5.11}$$

$$\mathcal{M}_4 : x_{k+1} = x_k + a_1 \Delta t + a_2 x_k \Delta t + \sigma\sqrt{\Delta t} q_k, \tag{5.12}$$

$$\tag{5.13}$$

where $\Delta t = 0.01$ and $q_k \sim \mathcal{N}(0, 1)$.

There is no prior distribution available for the parameters. In other words $p(\boldsymbol{\phi}|\mathcal{M}_i) \propto 1$. The effect of the correction factor for noninformative prior will also be studied. The correction factor is $\frac{1}{J^{1.5k_i}}$ where k_i is the number of parameters of $\mathcal{M}_i$ and J is the number of data points available. The natural logarithm of the evidence for different number of data points is shown in Fig. 10.

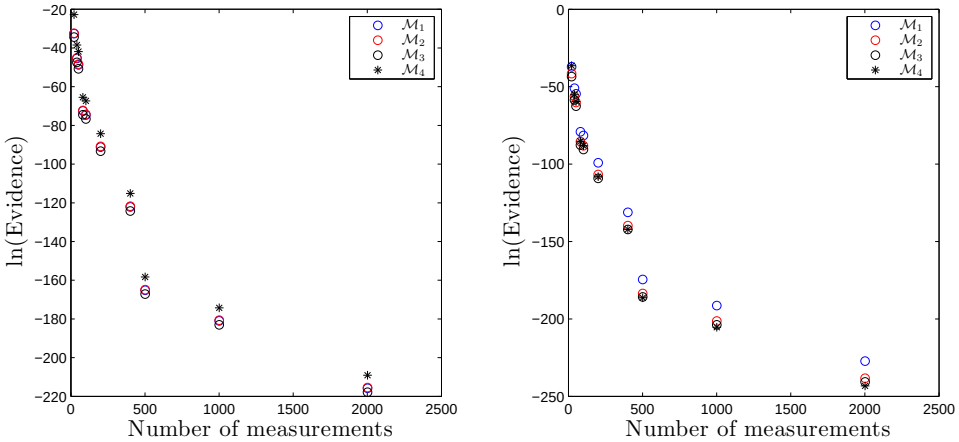


Figure 10: Natural logarithm of the evidence vs the number of data points for noninformative prior (left) and noninformative prior with the correction factor (right)

The selected model for the case of noninformative prior with no correction factor is $\mathcal{M}_4$ with probability close to 1 which is independent of the number of measurements

points. For the case of noninformative prior with a correction factor, the selected model when the number of measurement points is below 21 is $\mathcal{M}_4$, otherwise the correct model ($\mathcal{M}_1$) is selected with high probability. There are two conclusions that can be drawn from this exercise. The evidence strongly depends on the number of data points and the use of an noninformative prior without a correction factor tends to select the complex model.

5.2 Double-Well System

For numerical illustration, we consider a double-well system excited by white noise $W(t)$ given by the following equation: [71–73]

$$\dot{x}(t) = 4x(t) - 4x(t)^3 + 0.6W(t). \tag{5.14}$$

The system has two stable fixed points at $x = \pm 1$ and an unstable fixed point at $x = 0$. Its motion is similar to that of a particle oscillating between two wells (see Fig. 11). The system is also analogous to a massless shallow arch that can snap between two configurations. Using the Euler-Maruyama method [74], the discretized form of Eq. (5.14) is

$$x_{k+1} = x_k + 4 \cdot \Delta t(x_k - x_k^3) + 0.6\sqrt{\Delta t} q_k, \tag{5.15}$$

where $q_k \sim N(0, 1)$. The following linear measurement equation is used

$$d_j = x_d(j) + \varepsilon_j \quad \text{where } \varepsilon_j \sim N(0, 0.08). \tag{5.16}$$

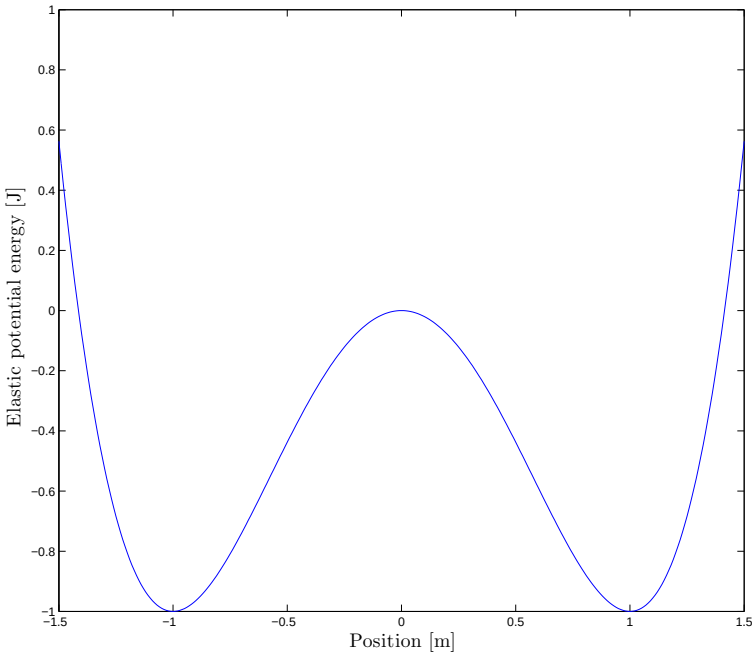


Figure 11: Double well potential.

For the model selection exercise, the candidate models are defined as follows

$$\mathcal{M}_1 : \dot{x} = \alpha x + 0.6W(t) + \sigma_m W_m(t) \tag{5.17}$$

$$\mathcal{M}_2 : \dot{x} = \alpha x + \beta x^3 + 0.6W(t) + \sigma_m W_m(t) \tag{5.18}$$

$$\mathcal{M}_3 : \dot{x} = \alpha x + \beta x^3 + \gamma x^5 + 0.6W(t) + \sigma_m W_m(t) \tag{5.19}$$

where $W_m(t)$ defines a unit white noise process that represent the modelling error. $W(t)$ and $W_m(t)$ are mutually independent. The optimal model will be selected using the model selection algorithm which involves estimating the parameters α, β, γ and σ_m for a given model. The magnitude of the stochastic forcing is known.

Three cases using different set of data will be considered. In all cases, the system is simulated for 20 seconds with $\Delta t = 0.01\,\text{s}$. In Case 1, 101 data points are available and measurements contain data in both well. In Case 2, 101 points are available

but the measurements are only available in one well. Finally in Case 3, 21 measurements points are available in both wells. EnKF and PF-EnKF have 1250 particles initially sampled from a Gaussian distribution $\mathcal{N}(1.0, 0.5)$. The stationary MCMC chain contains $100,000$ samples.

5.2.1 Case 1

Figure 12 shows one realisation with the measurements used for model selection. The real signal represents the position of the particle. The measurements track the position of the particle. It also presents the result of a state estimation procedure using the filters with the given parameters $\alpha = 4$, $\beta = -4$, $\gamma = 0$ and $\sigma_m = 0$ in Eq. (5.15).

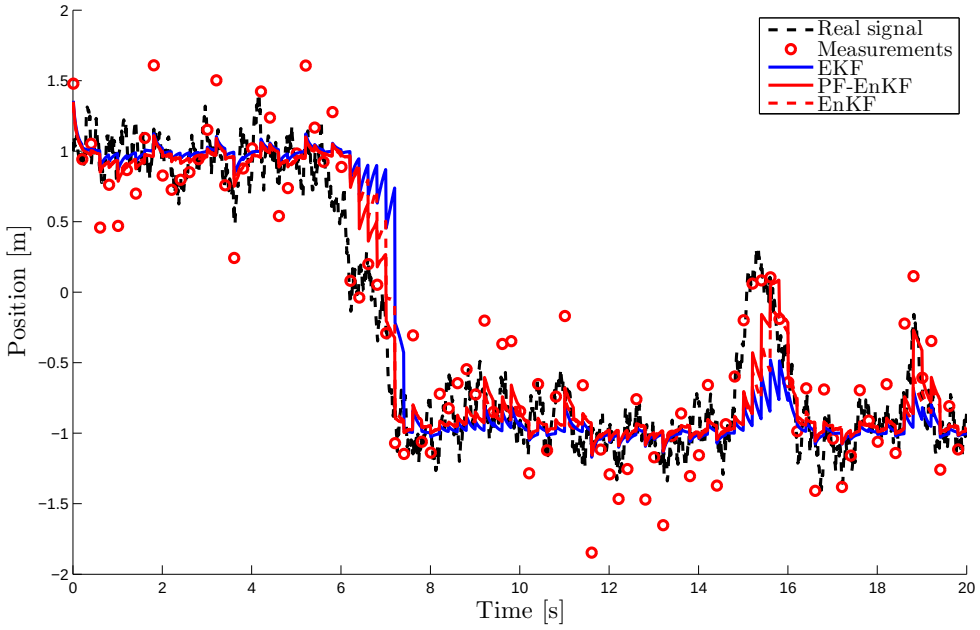


Figure 12: State estimation with the measurements used for model selection for Case 1.

The EKF, EnKF and PF-EnKF can all track the signal given the amount of data available. The histogram of the measurements is shown in Fig. 13. Evidently, the

data capture the dynamics from both wells.

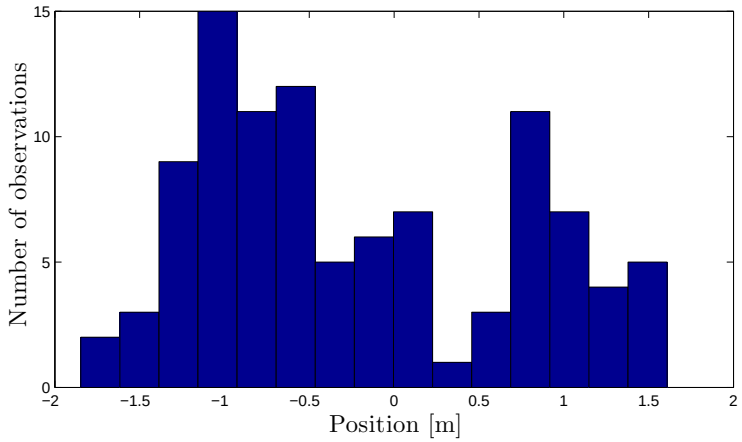


Figure 13: Histogram of the measurements used for model selection for Case 1.

The marginal distributions of the parameters of each model are plotted in Fig. 14. The panels in the left, middle and right columns display the parameter posteriors when EKF, EnKF and PF-EnKF are used as the state estimators respectively. The dashed black line represents the true value of the parameters ($\alpha = 4$, $\beta = -4$, $\gamma = 0$) used to generate the signal. The strength of the noise is always positive.

The variance of the parameter posteriors increases with the model complexity. The posterior parameter pdfs obtained by EnKF and PF-EnKF are similar. Due to the Gaussian approximation, EKF results differ from those obtained by the non-Gaussian filters.

In table 3, the 95% highest posterior density intervals (HPDI) are displayed. For $\mathcal{M}_2$ the true value of α and β is contained in the HPDI for EnKF and PF-EnKF but not for EKF.

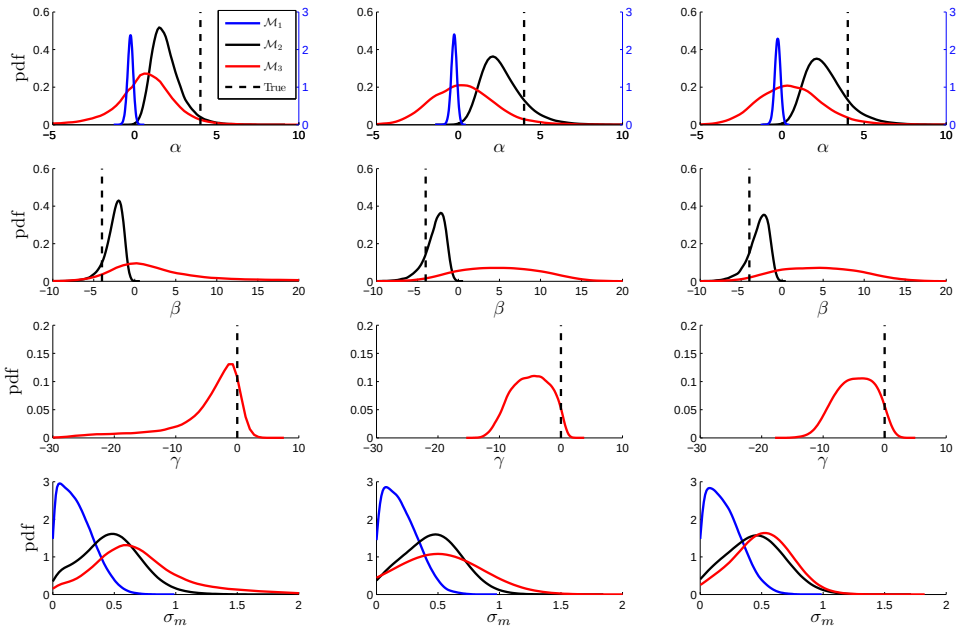


Figure 14: Marginal parameter pdf for Case 1: (a) the left column with EKF state estimator, (b) the middle column with EnKF state estimator, (c) the right column with PF-EnKF state estimator.

Table 3: 95% highest posterior density interval of the marginal distribution of the parameters for Case 1.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	[−0.60,0.06]	[0.40,3.84]	[−3.26,4.16]
	$\beta(-4)$		[−4.90,−0.72]	[−7.44,19.79]
	γ			[−20.85,2.46]
	σ_m	[0.00,0.46]	[0.00,0.88]	[0.00,1.47]
EnKF	$\alpha(4)$	[−0.60,0.06]	[0.65,5.35]	[−3.45,4.04]
	$\beta(-4)$		[−5.58,−0.80]	[−3.87,14.07]
	$\gamma(0)$			[−10.31,0.33]
	σ_m	[0.00,0.45]	[0.00,0.81]	[0.01,0.84]
PF-EnKF	$\alpha(4)$	[−0.61,0.06]	[0.64,5.35]	[−3.08,4.49]
	$\beta(-4)$		[−5.60,−0.77]	[−4.23,13.84]
	$\gamma(0)$			[−10.19,0.45]
	σ_m	[0.00,0.46]	[0.00,0.81]	[0.01,0.86]

The next table displays the Maximum a posteriori estimates (MAP) of the

marginal distribution of the parameters. The MAP of the parameters is similar for EKF, EnKF and PF-EnKF except for the third model. The MAP estimates of the second model are less biased compared to the other proposed models.

Table 4: MAP of the marginal distribution of the parameters for Case 1.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	-0.26	1.49	0.58
	$\beta(-4)$		-1.97	0.23
	$\gamma(0)$			-1.41
	σ_m	0.06	0.47	0.61
EnKF	$\alpha(4)$	-0.25	2.10	0.15
	$\beta(-4)$		-2.16	4.67
	$\gamma(0)$			-4.36
	σ_m	0.08	0.48	0.50
PF-EnKF	$\alpha(4)$	-0.28	2.10	0.31
	$\beta(-4)$		-2.27	4.55
	$\gamma(0)$			-3.78
	σ_m	0.07	0.45	0.52

Using the parameter posterior samples, the evidence is estimated for each model. The corresponding probability of each model is calculated for two cases of noninformative prior: (a) without the correction factor and (b) with the correction factor.

Table 5: Model probability using EKF for state estimation

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	9.90%	90.10%
noninformative (with correction factor)	12.86%	86.37%	0.77%

Table 6: Model probability using EnKF for state estimation

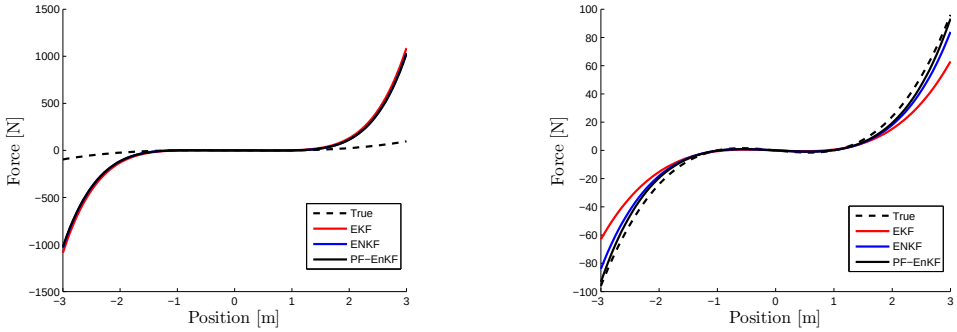
Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	3.98%	96.02%
noninformative (with correction factor)	9.55%	88.35%	2.10%

Table 7: Model probability using PF-EnKF for state estimation

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	3.04%	96.96%
noninformative (with correction factor)	7.21%	89.96%	2.83%

When no correction factor is applied, the complex model is selected. Using the correction factor, the probability of the first and second model increases. The second model is the most probable, being the true model that generated the data.

The expected spring force is estimated using Bayesian model averaging. In Fig. 15, the different estimates of the spring force versus the displacement are plotted when different state estimators are used. Note the scales are different in (a) and (b). When no correction factor is applied, the estimates of the spring force have a large bias for displacements greater than 2. Using the results of model selection with a correction factor provides a better estimate of the spring force.



(a) No correction factor (b) With correction factor

Figure 15: Expected spring force of the Double-Well system for Case 1 using the model selection results without correction (a) and with correction (b).

5.2.2 Case 2

The effect of data distribution is studied in this case. The same number of data points is used as in Case 1. In contrast to the first case, the data contains dynamics around the right well (i.e. the fixed point at 1). State estimation is satisfactory with all the filters as shown in Fig. 16.

The histogram in Fig. 17 again shows that all the measurements are near the equilibrium point at $x = 1$.

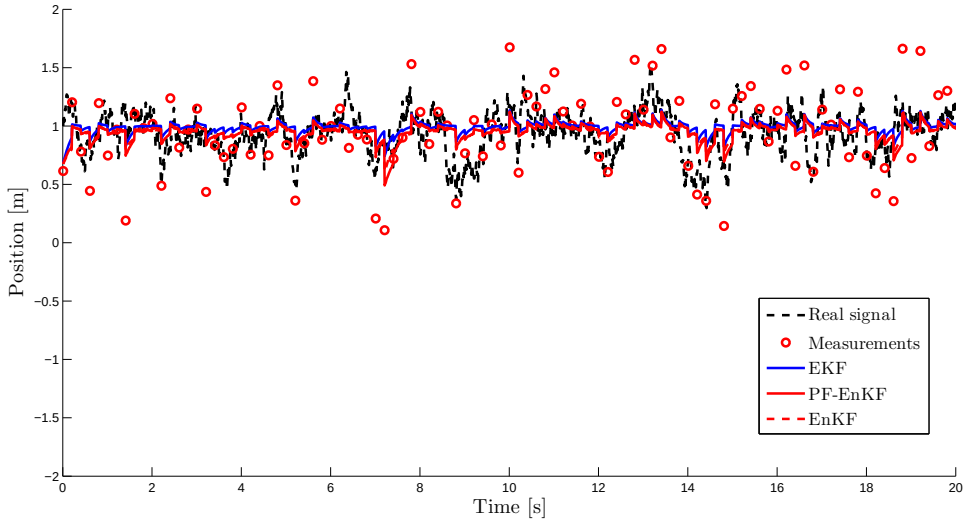


Figure 16: State estimation with the measurements used for model selection for Case 2

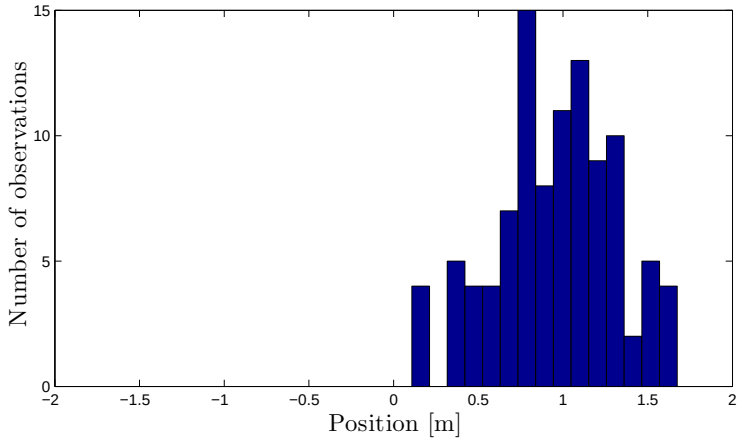


Figure 17: Histogram of the measurements used for model selection for Case 2.

The parameter pdfs are shown in Fig. 18. Compared to Case 1, the pdfs are more biased and uncertain as the measurements do not capture all the dynamics of the second equilibrium point at $x = -1$.

Similarly to Case 1, the HDPI and the MAP are shown in the tables 8 and 9 respectively.

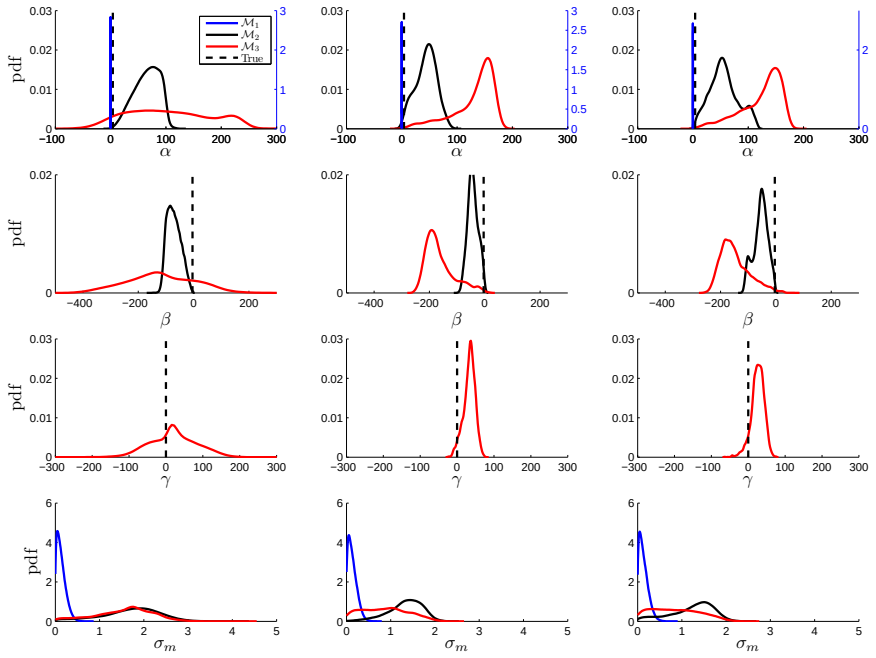


Figure 18: Marginal parameter pdf for Case 2: (a) the left column with EKF state estimator, (b) the middle column with EnKF state estimator, (c) the right column with PF-EnKF state estimator.

Table 8: 95% highest posterior density interval of the marginal distribution of the parameters for Case 2.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	[−0.45,0.11]	[23.69,100.21]	[−11.42,238.15]
	$\beta(-4)$		[−113.17,−22.79]	[−356.98,102.45]
	$\gamma(0)$			[−100.93,145.02]
	σ_m	[0.00,0.33]	[0.14,2.75]	[0.02,2.52]
EnKF	$\alpha(4)$	[−0.46,0.10]	[5.91,75.74]	[44.77,175.59]
	$\beta(-4)$		[−75.86,−6.08]	[−238.36,−41.84]
	$\gamma(0)$			[0.05,62.14]
	σ_m	[0.00,0.32]	[0.52,1.97]	[0.00,1.74]
PF-EnKF	$\alpha(4)$	[−0.45,0.13]	[11.68,108.91]	[43.74,174.73]
	$\beta(-4)$		[−106.32,−9.33]	[−231.57,−31.39]
	$\gamma(0)$			[−11.94,57.22]
	σ_m	[0.00,0.33]	[0.11,1.95]	[0.00,1.78]

The HDPI has increased compared to Case 1. The true value of the parameters

does not reside in these intervals except for γ. The MAP estimates are highly biased.

Table 9: MAP of the marginal distribution of the parameters for Case 2.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	-0.15	77.30	72.11
	$\beta(-4)$		-83.73	-130.07
	$\gamma(0)$			15.76
	σ_m	0.04	1.92	1.76
EnKF	$\alpha(4)$	-0.17	49.17	156.22
	$\beta(-4)$		-47.80	-193.72
	$\gamma(0)$			37.20
	σ_m	0.05	1.42	1.01
PF-EnKF	$\alpha(4)$	-0.15	52.02	147.83
	$\beta(-4)$		-51.46	-182.86
	$\gamma(0)$			26.82
	σ_m	0.05	1.51	0.32

The next tables present the results of model selection.

Table 10: Model probability using EKF for state estimation.

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	0.41%	99.59%
noninformative (with correction factor)	0.00%	80.57%	19.43%

Table 11: Model probability using EnKF for state estimation.

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	2.22%	97.78%
noninformative (with correction factor)	0.00%	95.85%	4.15%

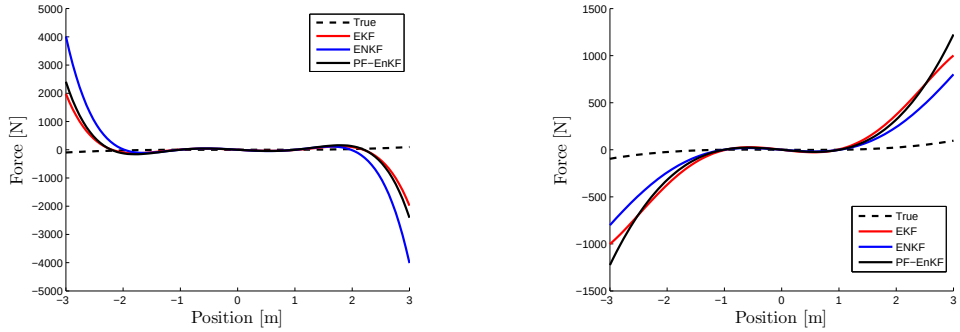
Table 12: Model probability using PF-EnKF for state estimation.

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	4.52%	95.48%
noninformative (with correction factor)	0.00%	97.96%	2.04%

When there is no correction factor, the more complex model is selected. When the correction factor is applied, the correct model is selected albeit with a lower certainty

for EKF. Model selection approach still selected the correct model. However, the parameter estimates have significant inaccuracy compared to the previous case. This fact highlights the importance of the quality of the measurements.

In Fig. 19, the estimate of the spring force is shown using the model selection results of Case 2. Note the scales are different in (a) and (b). The estimates have a larger bias than in the previous case. The effect of the correction factor is more drastic since the estimated force without the correction factor changes direction at large deflections.



(a) No correction factor

(b) With correction factor

Figure 19: Expected spring force of the Double-Well system for Case 2 using the model selection results without correction (a) and with correction (b).

5.2.3 Case 3

The last case considered is analogous to Case 1 but contains fewer data points (i.e. 21 data points). Figure 20 presents the measurements used and highlights the effect of data sparsity on state estimation.

The EKF fails to track the jump from one well to the other. Given that state estimation fails, one can reasonably expect that the parameter estimates and model selection algorithm will also fail. PF-EnKF and EnKF can successfully track the jump despite sparse measurements. Fig. 21 shows that the data capture the dynamics

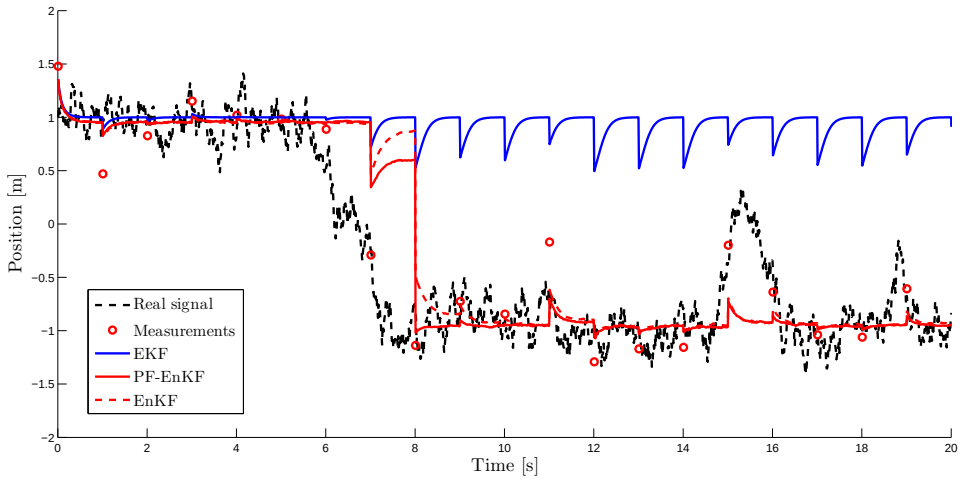


Figure 20: State estimation with the measurements used for model selection for Case 3.

around both stable equilibrium points at $x \pm 1$.

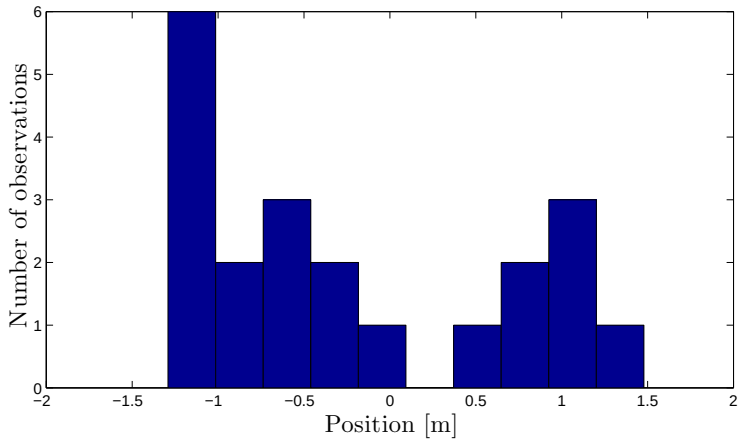


Figure 21: Histogram of the measurements used for model selection for Case 3.

The marginal distributions of the parameters are shown in Fig. 22. The parameter estimates with EKF as the state estimator are highly biased and have a large uncertainty. The true value of the parameters does not reside in the support of the corresponding marginal pdf. Given that the state estimation is erroneous, the parameter estimates are not representative of the information contained in the data. When

state estimation is performed by the non-Gaussian filters, the variance of estimates are smaller and the true value of the parameters resides in the support of their pdf.

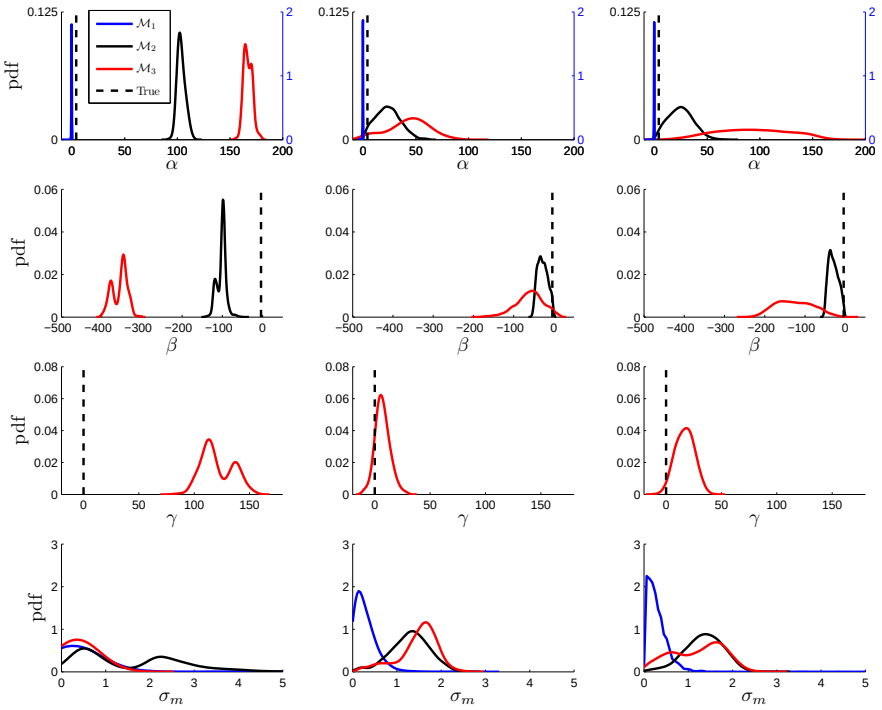


Figure 22: Marginal parameter pdf for Case 3: (a) the left column with EKF state estimator, (b) the middle column with EnKF state estimator, (c) the right column with PF-EnKF state estimator.

The HPDI is shown in table 13. The range of those values is higher than the one shown in Table 3. This is expected because there is more uncertainty in the analysis given that less information can be extracted from the measurements.

Table 13: 95% highest posterior density interval of the marginal distribution of the parameters for Case 3.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	[−0.87,0.12]	[99.41,110.98]	[160.04,174.47]
	$\beta(-4)$		[−123.50,−83.19]	[−389.31,−324.48]
	$\gamma(0)$			[97.53,146.75]
	σ_m	[0.00,0.84]	[0.00,3.65]	[0.00,0.68]
EnKF	$\alpha(4)$	[−0.79,0.09]	[2.37,44.11]	[1.24,75.96]
	$\beta(-4)$		[−46.79,−3.42]	[−123.52,10.05]
	$\gamma(0)$			[−4.62,21.60]
	σ_m	[0.00,0.76]	[0.44,2.17]	[0.43,2.23]
PF-EnKF	$\alpha(4)$	[−0.76,0.10]	[2.05,45.34]	[31.30,155.52]
	$\beta(-4)$		[−49.20,−6.36]	[−209.17,−38.26]
	$\gamma(0)$			[0.34,33.07]
	σ_m	[0.00,0.73]	[0.43,2.16]	[0.07,2.13]

The MAP estimates do not coincide with the true values of the parameters and are biased.

Table 14: MAP of the marginal distribution of the parameters for Case 3.

		$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
EKF	$\alpha(4)$	−0.26	102.30	164.62
	$\beta(-4)$		−98.59	−345.72
	$\gamma(0)$			113.26
	σ_m	0.24	0.53	0.34
EnKF	$\alpha(4)$	−0.25	22.44	46.81
	$\beta(-4)$		−33.94	−52.85
	$\gamma(0)$			5.16
	σ_m	0.14	1.37	1.65
PF-EnKF	$\alpha(4)$	−0.29	25.08	89.06
	$\beta(-4)$		−37.02	−156.78
	$\gamma(0)$			18.47
	σ_m	0.06	1.40	1.63

Finally, the results of model selection for Case 3 are presented.

Table 15: Model probability using EKF for state estimation.

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	99.94%	0.06%
noninformative (with correction factor)	5.23%	94.77%	0.00%

Table 16: Model probability using EnKF for state estimation.

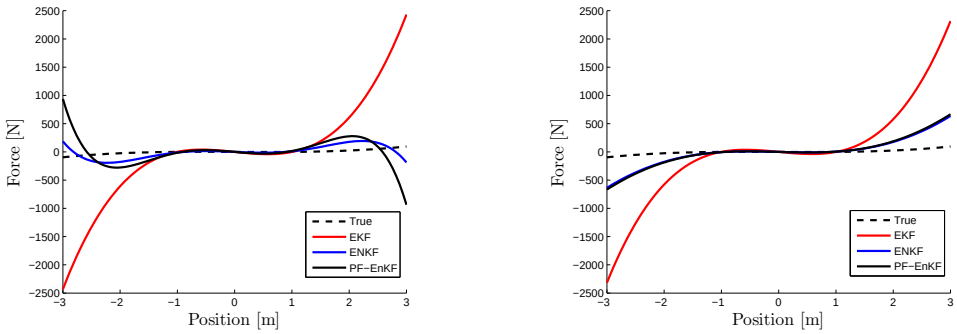
Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	1.74%	98.26%
noninformative (with correction factor)	0.06%	94.69%	5.25%

Table 17: Model probability using PF-EnKF for state estimation.

Prior	$Pr(\mathcal{M}_1\|\boldsymbol{D})$	$Pr(\mathcal{M}_2\|\boldsymbol{D})$	$Pr(\mathcal{M}_3\|\boldsymbol{D})$
noninformative (without correction factor)	0.00%	3.12%	96.88%
noninformative (with correction factor)	0.06%	96.97%	2.97%

In this case EKF selects the correct model whether a correction factor is applied or not. As the state estimation procedure shown in Fig. 20 is erroneous, the parameter estimates in Fig. 22 are incorrect. Hence model selection results are not meaningful in this case of sparse data. In the case of EnKF and PF-EnKF, the correct model is selected when the correction factor is applied.

In Fig. 23, the estimated spring force is shown for Case 3. The use of EKF as state estimator does not provide good estimate of the spring force. Compared to Case 1, these estimates are highly biased.



(a) No correction factor (b) With correction factor

Figure 23: Expected spring force of the Double-Well system for Case 3 using the model selection results without correction (a) and with correction (b).

5.3 Nonlinear Mass-Spring-Damper System

In this section the parameter estimation algorithm will be applied to a mass-spring-damper system containing various types of nonlinearities.

5.3.1 Nonlinearities

The types of nonlinearities considered are cubic stiffness, freeplay (backlash) and hysteresis [3, 16–18]. Cubic stiffness has the following mathematical representation

$$f_c = k_l u + k_{nl} u^3 \tag{5.20}$$

where f_c is the restoring force, u is the displacement and k_l and k_{nl} are the linear and nonlinear stiffness coefficients. The nonlinear stiffness k_{nl} can be positive or negative. If it is positive, at high displacement the restoring force due to the nonlinear term will be greater than its linear counterpart. As the displacement increases, the effect of the nonlinear part becomes significant. Such system is said to have hardening characteristics. Examples of such systems are plates and beams undergoing large deflection [3]. The nonlinear stiffness term is negative when the effective stiffness decreases with the displacement. Such system is said to exhibit softening stiffness characteristics. The buckling of beams and plates may be modelled by a softening spring over a limited range [3]. The next figure is an idealized representation of an hardening spring.

Freeplay is a specific form of piecewise linear stiffness. In aeroelastic systems, freeplay nonlinearity can occur due to worn-out hinges and loose rivets [16]. The combination of freeplay nonlinearity with other types of structural and aerodynamic nonlinearities in the structure and in the aerodynamic can change the system response.

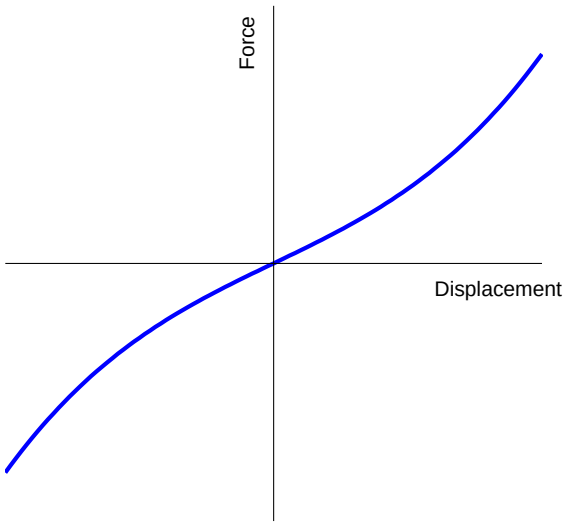


Figure 24: Idealized hardening cubic stiffness force-displacement diagram.

The mathematical representation of freeplay [3, 16–18] is

$$f_f = \begin{cases} k_f(u - \delta_f) & \text{if } u > \delta_f \\ 0 & \text{if } -\delta_f \leq u \leq \delta_f \\ k_f(u + \delta_f) & \text{if } u < -\delta_f \end{cases} . \tag{5.21}$$

Figure (25) is the graphical representation of freeplay.

The last form of nonlinearity considered in this investigation is hysteresis [17, 18]. It is defined by three variables, δ_h, a_h and f_0. Contrary to the cubic stiffness and freeplay, this type of nonlinearity is dependent on the direction of the deflection. For

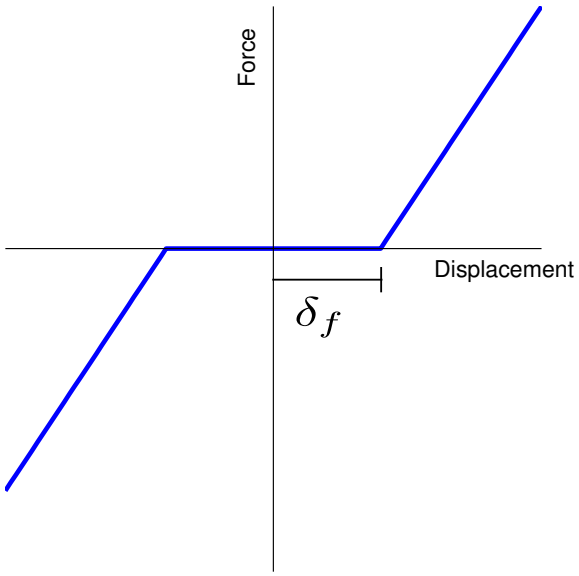


Figure 25: Idealized freeplay force-displacement diagram.

positive deflection, the restoring force is given by

$$f_h = \begin{cases} k_h(u - a_h) + f_0 & \text{if } u < a_h \\ f_0 & \text{if } a_h \leq u \leq (\delta_h + a_h) \\ k_h(u - a_h - \delta_h) + f_0 & \text{if } u > (\delta_h + a_h) \end{cases} . \tag{5.22}$$

For negative deflection, the restoring force becomes

$$f_h = \begin{cases} k_h(u + a_h) - f_0 & \text{if } u > -a_h \\ -f_0 & \text{if } -(a_h + \delta_h) \leq u \leq -a_h \\ k_h(u + a_h + \delta_h) - f_0 & \text{if } u < -(a_h + \delta_h) \end{cases} \tag{5.23}$$

where $f_0 = \frac{1}{2} k_h(2a_h + \delta_h)$. Figure 26 is the graphical representation of hysteresis.

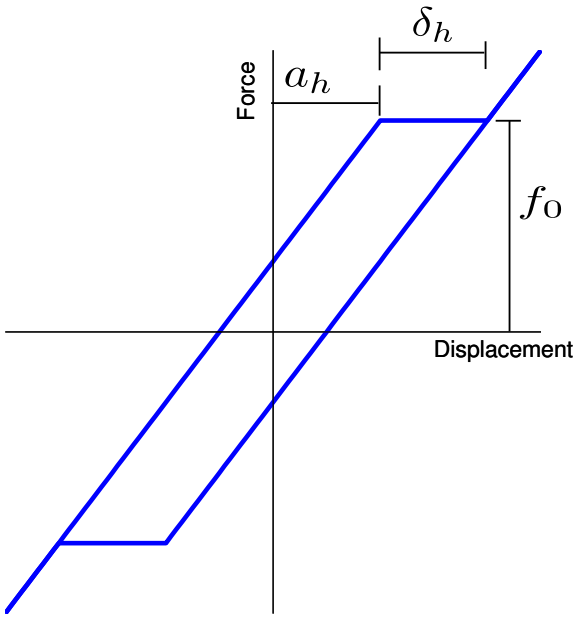


Figure 26: Idealized hysteresis force-displacement diagram.

In Fig. 27, a force-displacement diagram shows the combination of the three nonlinearities.

The force-displacement diagrams for all three types of nonlinearities and their combination (as in Fig. 27) are also shown in Fig. 28.

A single degree of freedom exhibiting the aforementioned nonlinearities is investigated next. The dynamics of the nonlinear system is defined by

$$m\ddot{u} + c\dot{u} + f_s(u) + f_f(u) + f_h(u, \dot{u}) = \mathrm{Tcos}(\omega \mathrm{t}) + \sigma \mathrm{W(t)} \tag{5.24}$$

where m and c are the mass and damping coefficient, T and ω are the amplitude and frequency of the deterministic forcing, σ is the strength of the modelling noise and $W(t)$ is the white noise process with unit strength. The discrete state-space

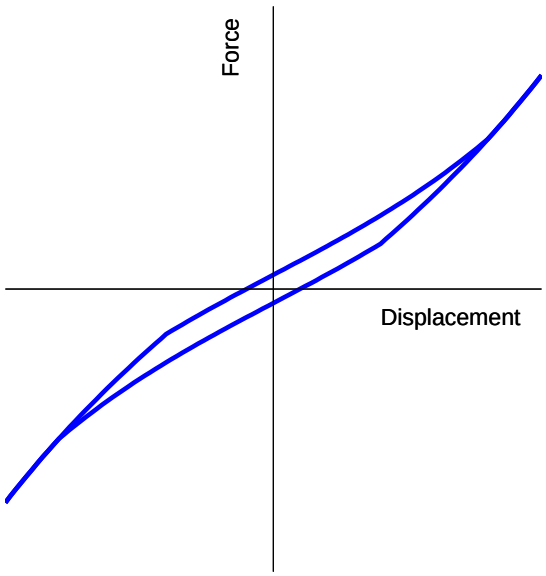


Figure 27: Idealized combination of hysteresis, cubic stiffness and freeplay force-displacement diagram.

representation of this system is

$$\mathbf{x}_{k+1} = \begin{Bmatrix} \mathbf{x}_{1,k} + \Delta t \mathbf{x}_{2,k} \\ \mathbf{x}_{2,k} - \dfrac{\Delta t}{m}\left(f_s + f_f + f_h + c\mathbf{x}_{2,k}\right) \end{Bmatrix} + \frac{\Delta t}{m} \begin{Bmatrix} 0 \\ \mathrm{T}\cos(\omega \mathrm{t_k}) \end{Bmatrix} + \frac{1}{m}\boldsymbol{q}_k \tag{5.25}$$

where

$$\mathbf{x}_k = \begin{bmatrix} u_k \\ \dot{u}_k \end{bmatrix} \tag{5.26}$$

and

$$\boldsymbol{q}_k \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{Q}_k\right) \text{ and } \boldsymbol{Q}_k = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2 \Delta t \end{bmatrix} \tag{5.27}$$

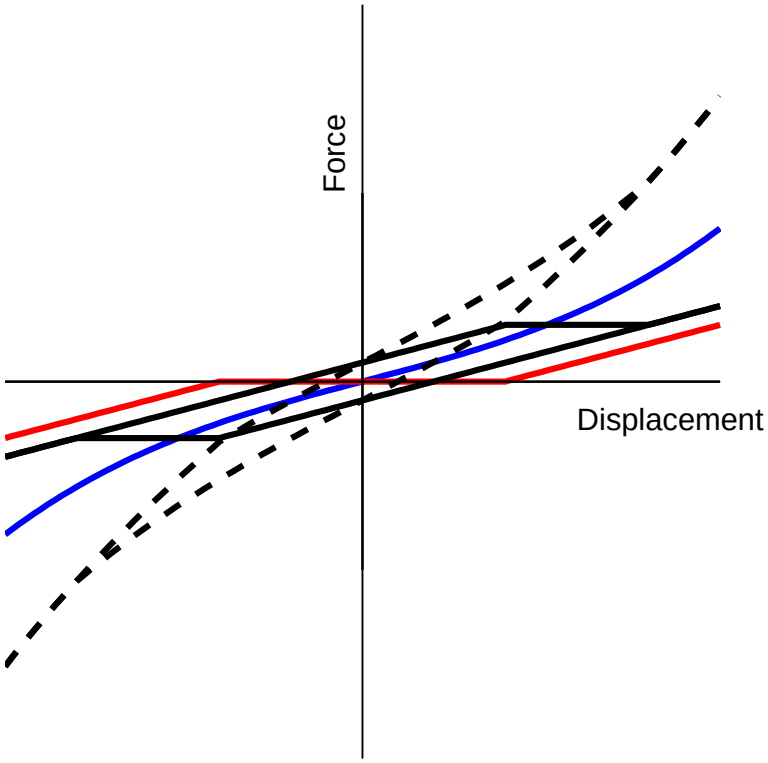


Figure 28: Force-displacement diagram of all nonlinearities.

5.3.2 Case 1

For the first experiment, the response is generated using Eq. (5.24) with $\Delta t = 0.01\,\mathrm{s}$, $m = 1\mathrm{kg}$, $w = 1.5\mathrm{rad/s}$, $\mathrm{T} = 0.5\mathrm{N}$, $k_l = 1\mathrm{N/m}$, $k_{nl} = 1\mathrm{N/m^3}$, $k_h = \mathrm{N/m}$, $k_f = \mathrm{N/m}$, $c = 0.3\,\mathrm{Ns/m}$, $\delta_h = 1.5\,\mathrm{m}$, $\delta_f = 0.2\,\mathrm{m}$ and $\sigma = 0$. The position of oscillator is recorded but not its velocity.

$$d_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_k + \varepsilon_k \text{ where } \varepsilon_k \sim N(0, \Gamma) \tag{5.28}$$

The response is simulated for 30 seconds. Fig. 29 presents a case where the signal to noise ratio is high and the measurements are sparse. The variance of the

measurements is $\Gamma = 0.05$. We assume that $k_l = k_h = k_f$.

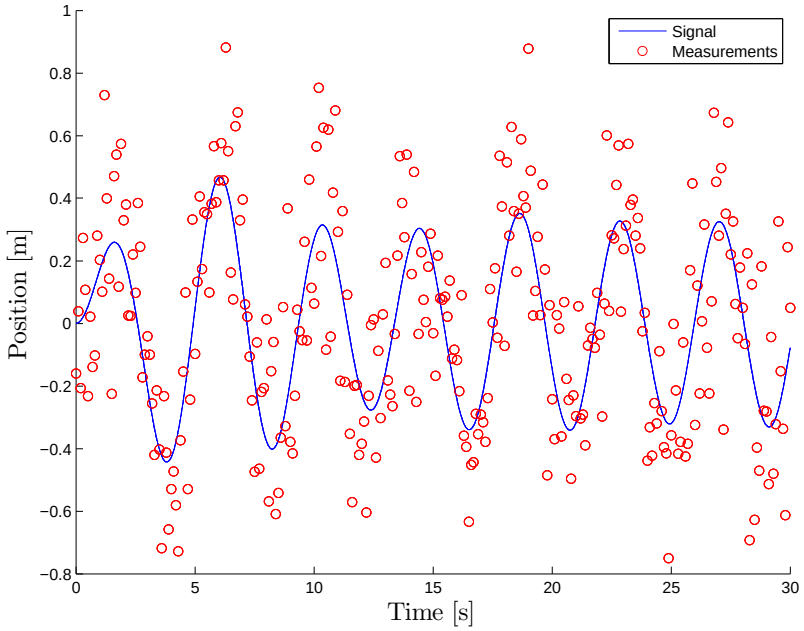


Figure 29: One realisation of the mass-spring-damper system with the generated measurements.

The parameters of the system are estimated using the measurements presented in Fig. 29. The posterior pdfs of the parameters are shown in Fig. 30. The state estimation procedure is tackled by a PF-EnKF using 1250 particles. The stationary MCMC chain contains 50,000 samples.

The true values of the parameters fall inside the supports of the pdfs. The bias in the estimates is evident for all parameters except the cubic stiffness. The size of the plateau δ_f for freeplay is bimodal.

Next we consider the cases when k_l,k_h and k_f are no longer restricted to be equal. The strength of measurement noise is varied in order to investigate its influence on the parameter estimation.

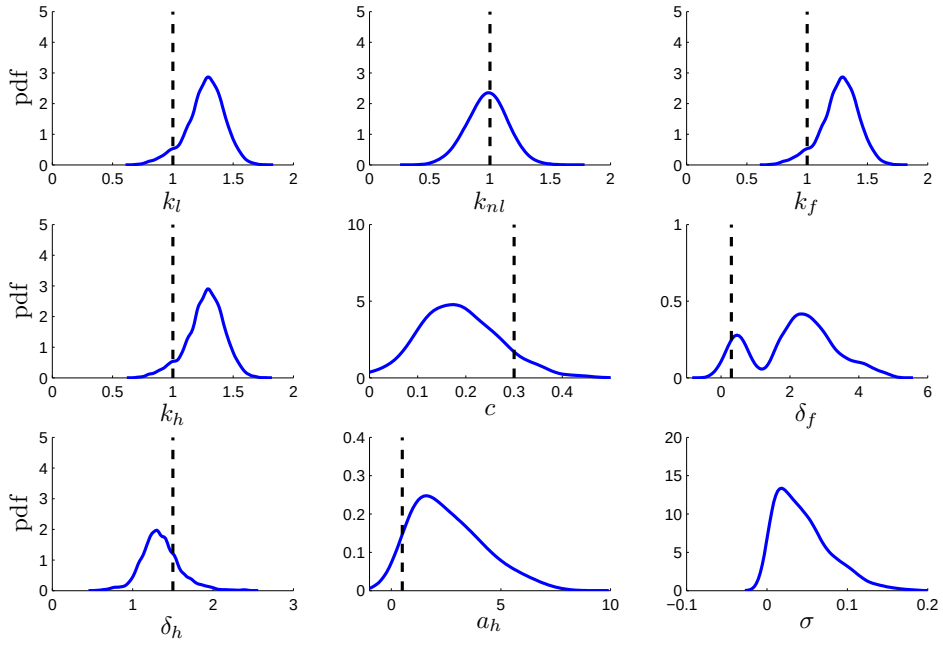


Figure 30: Posterior pdfs of the parameters of the true model when state estimation is tackled by PF-EnKF.

5.3.3 Case 2

In the first case, Γ is assumed to be 0.005. Figure (31) shows one realization of the system with the corresponding measurements. The strength of the forcing T is assumed to be 3.

In Fig. 32, the posterior pdfs of the parameters are shown.

Due to low noise in the measurements, the bias of the estimates is significantly reduced. The marginal posterior pdf of k_l has some non-Gaussian features, but its true value falls within the high probability region. The pdf of the cubic stiffness k_{nl} is Gaussian and its mean coincides with the true value of k_{nl}. Similarly to k_l, the posterior pdf of the freeplay stiffness k_f has some non-Gaussian features, but the true value also lies in the high probability region. The pdf of δ_f has two modes and the true value is close to one mode. The pdf of the hysteresis stiffness is Gaussian and has

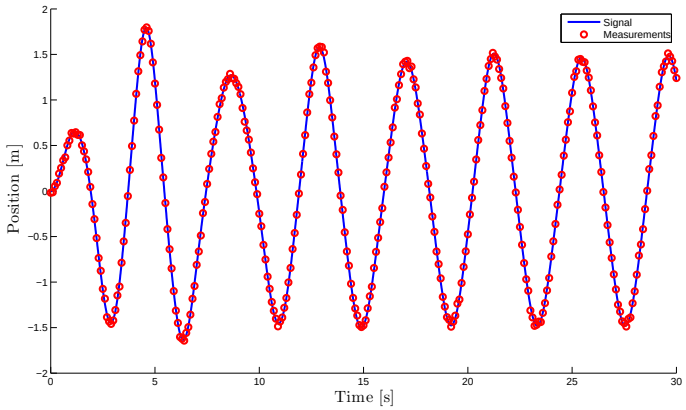


Figure 31: One realisation of the mass-spring-damper system with the generated measurements with a variance of 0.005.

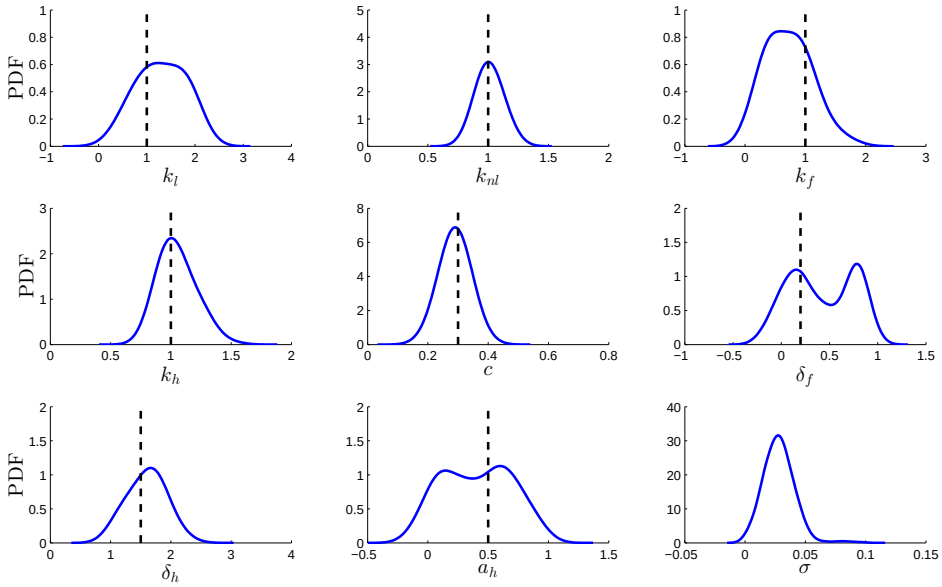


Figure 32: Posterior pdfs of the parameters of the true model when state estimation is tackled by PF-EnKF and the variance of the measurements is 0.005.

almost no bias. The marginal pdf of δ_h has some non-Gaussian properties and its mode is close to the true value. The marginal pdf of a_h has two modes and the true value is close to the second mode. Finally the damping is adequately estimated with low bias and uncertainty. To evaluate the effect of measurement sparsity, the parameters

of the model are estimated using 101 measurement points. The measurements are shown in Fig. 33.

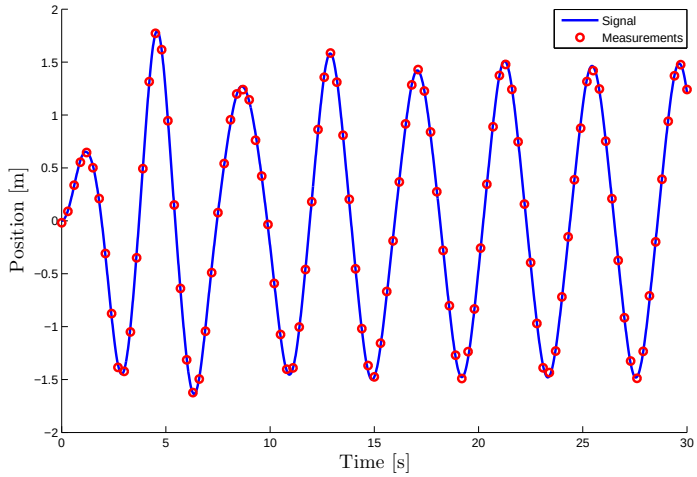


Figure 33: One realisation of the mass-spring-damper system with the generated measurements with a variance of 0.005.

The results of parameter estimation are shown in Fig. 34.

The uncertainty and the bias in the estimates are larger due to the decreasing amount of data available. The estimates of the stiffness parameters have some non-Gaussian features but are unimodal. The strength of the modelling error σ has doubled. Reducing the number of measurement points and increasing the measurement uncertainty have similar effects.

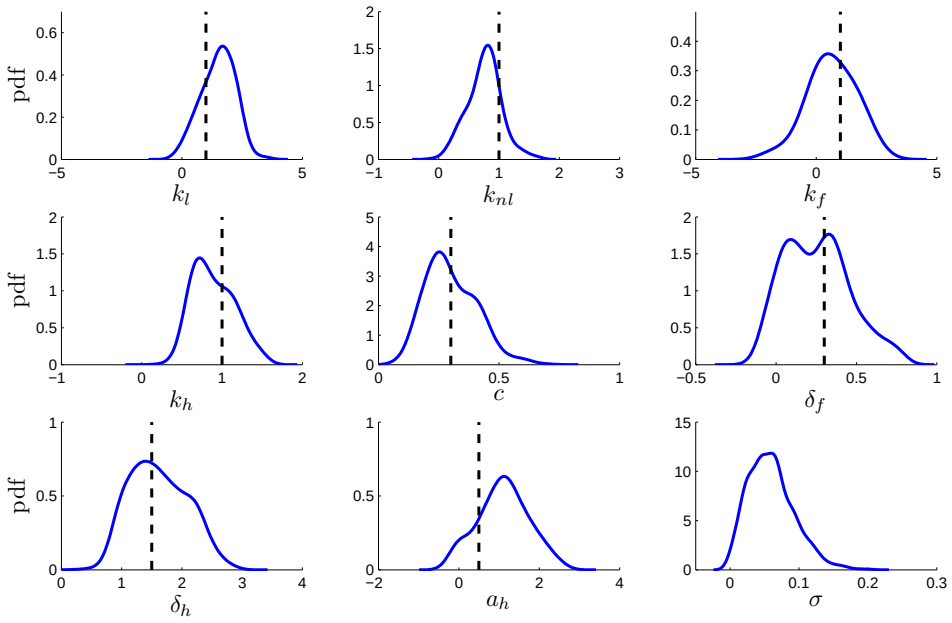


Figure 34: Posterior pdfs of the parameters of the true model when state estimation is tackled by PF-EnKF and the variance of the measurements is 0.005.

5.3.4 Case 3

The measurement noise is increased to $\Gamma = 0.02$. The measurements are shown in Fig. 35. The marginal posterior pdfs of the parameters are shown in Fig. 36.

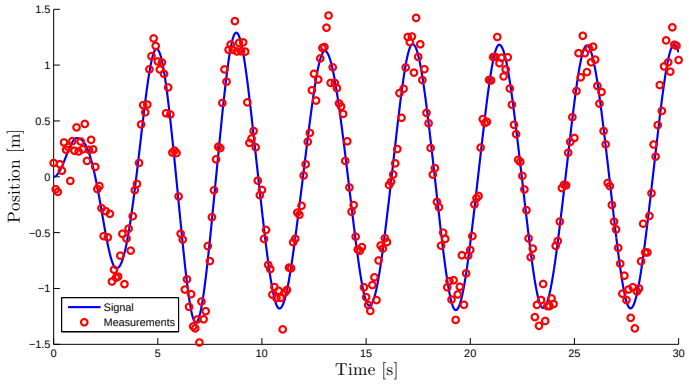


Figure 35: One realisation of the mass-spring-damper system with the generated measurements with a variance of 0.02.

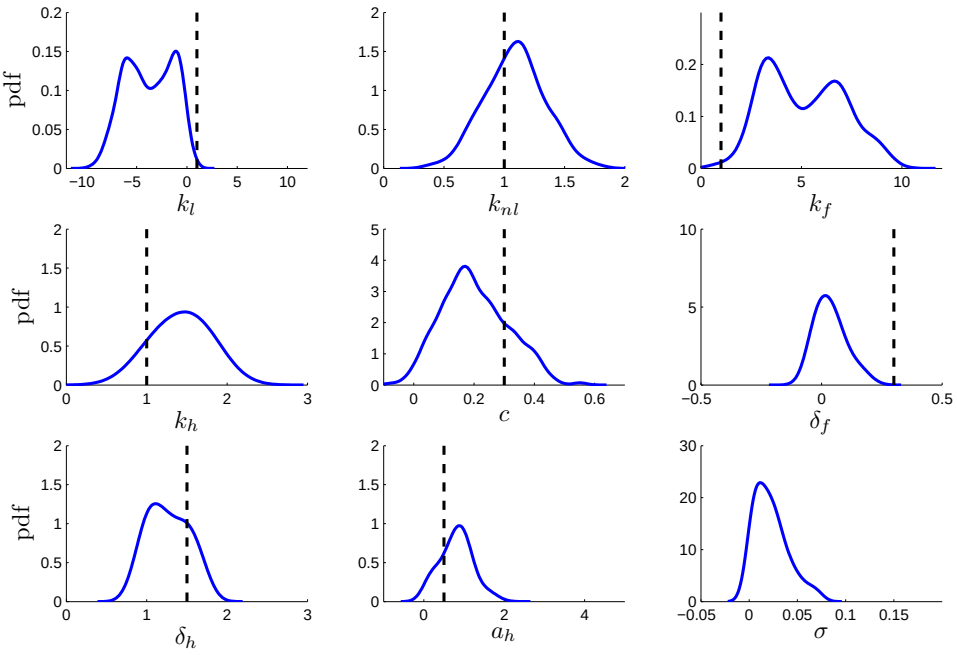


Figure 36: Posterior pdfs of the parameters of the true model when state estimation is tackled by PF-EnKF and the measurement variance is 0.02.

The observations used for parameter estimation contain more noise than the previous case. This fact leads to unsatisfactory results. The linear stiffness k_l, the freeplay stiffness k_f are bimodal and their true value are close to the tail of the pdfs. The estimate of the cubic stiffness k_{nl} has higher bias than the previous case, but still falls within the region of high probability. The hysteresis stiffness k_h pdf is unimodal. The true value lies in the support of the pdf. The damping coefficient has more uncertainty than the previous case. The freeplay plateau δ_f is poorly estimated as the true value is outside the support of the pdf. Finally, δ_h and a_h marginal pdfs are non-Gaussian but their true value is in the support of the pdf. The variance of the modelling noise is stronger than in the previous case. To increase the quality of the estimates, more measurements are needed.

Model selection is not carried out for this system due to extremely high computational time required using the available high performance computational facilities used in this investigation. Some areas of the algorithm that require fine-tuning to shorten the simulation time are briefly summarized in the conclusion.

Chapter 6

Conclusion

In this thesis, the Bayesian framework for model selection and parameter estimation developed by Khalil et al.; Sandhu (2012); Sandhu et al. (2013) [4, 7, 8] has been extended to handle strong nonlinearities in the state pdf of the system. In the framework proposed by Khalil and Sandhu, EKF is used for state estimation. There are two major limitations of using EKF. First, the state of the system is assumed to be Gaussian. In the presence of dense data, the state of the system remains nearly Gaussian even in the presence of strong nonlinearities. However, in the case of sparse data, the pdf of the system state cannot be approximated by a Gaussian distribution in the presence of nonlinearities. Second, EKF can only handle additive Gaussian noise in the model and measurement operators. In this thesis, the state estimation is carried out using EnKF or PF-EnKF. These non-Gaussian filters do not suffer from the two aforementioned limitations of EKF. Although EnKF cannot handle strong nonlinearities like the PF-EnKF, it has a lower computational cost due to a simpler update step. EnKF outperforms EKF as it can handle stronger nonlinearity by avoiding the linearization of the model and measurement operators. The choice of the filter depends on the sparsity of data and the strength of the nonlinearities in the system.

The contributions of this work can be summarized as follows. The algorithmic contribution lies in the application of EnKF and PF-EnKF for the state estimation

procedure in the general Bayesian framework of model selection and parameter estimation. These non-Gaussian filters are necessary for Bayesian model selection and parameter estimation for strongly nonlinear systems when the available data is sparse and noisy. The thesis also contributes to the application of the Bayesian model selection and parameter estimation framework. The first application contribution involves a validation of the framework for model selection and parameter estimation of the Double-Well system. This system exhibits three equilibrium points. The use of EnKF and PF-EnKF as state estimators for parameter estimation is necessary due to the non-Gaussian features of the system state. When the measurements are sparse, EKF fails to provide accurate state estimates due to strongly non-Gaussian system state. The second contribution lies in the successful estimation of the parameters of a single degree of freedom nonlinear mass-spring-damper system in the presence of cubic, freeplay and hysteretic nonlinearities.

The major limitation of the Bayesian framework using EnKF and PF-EnKF filters is the computational cost. The cost is reduced by a parallel implementation of PF-EnKF and EnKF. The MCMC algorithm is embarrassingly parallel as each core can run an independent Markov chain. Further tuning of the code is necessary to further improve the computational efficiency. This code tuning may include the use of GPUs (Graphical processing units). Once the computation cost is sufficiently reduced, the model selection of the mass-spring-damper presented in Chapter 5 will be performed. Some fine-tuning of the kernel density estimate algorithm used to estimate the prior and proposal distributions during the update step of PF-EnKF may be necessary. It is critical element of the update step of the particle filter. In high dimensional space, the cost can become prohibitive. Finally, model selection and parameter estimation of a multi-degree-freedom airfoil using experimental data from a wind tunnel experiment will be carried out. The system will include both structural and aerodynamic nonlinearities.

References

[1] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer. ISBN 9783642037108 (2009).

[2] T. E. Noll, J. M. Brown, M. E. Perez-davis, S. D. Ishmael, G. C. Tiffany, and M. Gaier. "Investigation of the Helios Prototype Aircraft Mishap Volume I Mishap Report." Technical Report January (2004).

[3] Worden. "Nonlinearity in Structural Dynamics." (2001).

[4] M. Khalil, D. Poirel, and A. Sarkar. "Probabilistic Parameter Estimation of a Fluttering Aeroelastic System in the Transitional Reynolds Number Regime." *Journal of Sound and Vibration* .

[5] D. Poirel and W. Yuan. "Aerodynamics of Laminar Separation Flutter at a Transitional Reynolds Number." *Journal of Fluids and Structures* .

[6] J. Beck and L. Katafygiotis. "Updating Models and Their Uncertainties. I: Bayesian Statistical Framework." *Journal of Engineering Mechanics* **124**(4), 455–461. ISSN 0733-9399 (1998).

[7] R. Sandhu. *Bayesian Model Selection and Parameter Estimation for a Nonlinear Fluid-structure Interaction Problem*. Master's thesis, Carleton University (2012).

[8] R. Sandhu, M. Khalil, and A. Poirel D. Sarkar. "Model Selection Methods for Nonlinear Aeroelastic Systems using Wind-Tunnel Data." In "54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference," (2013).

[9] H. Haario and E. Saksman. "Adaptive proposal distribution for random walk Metropolis algorithm." *Computational Statistics* **4**, 1–32 (1999).

[10] S. Chib and I. Jeliazkov. "Marginal Likelihood from the Metropolis Hastings Output." *Journal of the American Statistical Association* **96**(453), 270–281 (2001).

[11] M. Khalil, A. Sarkar, S. Adhikari, and D. Poirel. "The Estimation of Time-Invariant Parameters of Noisy Nonlinear Oscillatory Systems." *Journal of Sound and Vibration (submitted)* (2013).

[12] C. Sanderson. "Armadillo : An Open Source C ++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments." Technical report (2011).

[13] E. G. Bosilca, G. E. Fagg, George, , Jeffrey, T. Angskun, J. J. Dongarra, M. S. Kambadur, V. Sahay, Prabhanjan, , Ralph, B. Barrett, A. Lumsdaine, H. C. Graham, D. J. Daniel, R. L., Woodall, and T. S. "Open {MPI}: Goals, Concept, and Design of a Next Generation {MPI} Implementation." In "Proceedings, 11th European PVM/MPI Users' Group Meeting," pages 97—-104 (2004).

[14] D. S. Woolston, H. L. Runyan, T. A. Byrdsong, and U. S. N. A. C. for Aeronautics. *Some Effects of System Nonlinearities in the Problem of Aircraft Flutter*. Technical note. National Advisory Committee for Aeronautics (1955).

[15] G. Kerschen, K. Worden, A. F. Vakakis, and J.-C. Golinval. "Past, present and future of nonlinear system identification in structural dynamics." *Mechanical Systems and Signal Processing* **20**(3), 505–592. ISSN 0888-3270 (2006).

[16] A. Abdelkefi, R. Vasconcellos, F. D. Marques, and M. R. Hajj. "Modeling and identification of freeplay nonlinearity." *Journal of Sound and Vibration* **331**(8), 1898–1907. ISSN 0022-460X (2012).

[17] C. Popescu, Y. Wong, and B. Lee. "System Identification for Nonlinear Aeroelastic Models." In "46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference," Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics (2005).

[18] Y. Gu and Z. Yang. "Aeroelastic Analysis of an Airfoil with a Hysteresis Non-Linearity." In "47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference," Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics (2006).

[19] D. R. Anderson. *Model Based Inference in the Life Sciences: A Primer on Evidence*. Springer Science + Business Media. ISBN 9780387740751 (2008).

[20] K. Worden and G. R. Tomlinson. *Nonlinearity in Structural Dynamics: Detection, Identification and Modelling*. Taylor & Francis. ISBN 9781420033823 (2010).

[21] T. K. Caughey. "Equivalent Linearization Techniques." *The Journal of the Acoustical Society of America* **35**(11), 1706–1711 (1963).

[22] S. F. Masri and T. K. Caughey. "A Nonparametric Identification Technique for Nonlinear Dynamic Problems." *Journal of Applied Mechanics* **46**(2), 433–447. ISSN 0021-8936 (1979).

[23] S. F. Masri, T. K. Caughey, and H. Sassi. "Nonparametric Identification of Nearly Arbitrary Nonlinear Systems." *Journal of Applied Mechanics* **49**(3), 619–628. ISSN 0021-8936 (1982).

[24] M. Haroon, Y. W. Luk, and D. E. Adams. "A Technique for Estimating Linear Parameters Using Nonlinear Restoring Force Extraction in the Absence of an Input Measurement." *Journal of Vibration and Acoustics* **127**(5), 483–492. ISSN 0739-3717 (2005).

[25] I. J. Leontaritis and S. A. Billings. "Input-output parametric models for non-linear systems Part I: deterministic non-linear systems." *International Journal of Control* **41**(2), 303–328 (1985).

[26] I. J. Leontaritis and S. A. Billings. "Input-output parametric models for non-linear systems Part II: stochastic non-linear systems." *International Journal of Control* **41**(2), 329–344 (1985).

[27] Statisticat LLC. "Bayesian Inference." Technical report.

[28] L. Wasserman. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* **44**(1), 92–107. ISSN 0022-2496 (2000).

[29] J. Beck and K. Yuen. "Model Selection Using Response Measurements: Bayesian Probabilistic Approach." *Journal of Engineering Mechanics* **130**(2), 192–203. ISSN 0733-9399 (2004).

[30] J. L. Beck. "Bayesian system identification based on probability logic." *Structural Control and Health Monitoring* **17**(7), 825–847. ISSN 1545-2263 (2010).

[31] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC. ISBN 158488388X (2003).

[32] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag (2002).

[33] H. Akaike. "A new look at the statistical model identification." (1974).

[34] S. H. Cheung and J. L. Beck. "Calculation of Posterior Probabilities for Bayesian Model Class Assessment and Averaging from Posterior Samples Based on Dynamic System Data." *Computer-Aided Civil and Infrastructure Engineering* **25**(5), 304–321. ISSN 1467-8667 (2010).

[35] G. Schwarz. "Estimating the Dimension of a Model." *The Annals of Statistics* **6**(2), 461–464. ISSN 00905364 (1978).

[36] R. E. Kass and A. E. Raftery. "Bayes factors." *Journal of the american statistical association* **90**(430), 773–795 (1995).

[37] S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling (Springer Series in Statistics)*. Springer. ISBN 0387718869 (2007).

[38] J. O. Berger, L. R. Pericchi, J. K. Ghosh, T. Samanta, F. D. Santis, J. O. Berger, and L. R. Pericchi. "Objective Bayesian Methods for Model Selection: Introduction and Comparison." *Lecture NotesMonograph Series* **38**(2001), 135–207. ISSN 07492170 (2001).

[39] C. Oh, J. Beck, and M. Yamada. "Bayesian Learning Using Automatic Relevance Determination Prior with an Application to Earthquake Early Warning." *Journal of Engineering Mechanics* **134**(12), 1013–1020. ISSN 0733-9399 (2008).

[40] J. Ching and Y.-c. Chen. "Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging." *Journal of Engineering Mechanics* **133**(7), 816–832 (2007).

[41] N. Friel and J. Wyse. "Estimating the evidence a review." *Statistica Neerlandica* **66**(3), 288–308. ISSN 1467-9574 (2012).

[42] W. Zucchini. "An Introduction to Model Selection." *Journal of Mathematical Psychology* **44**(1), 41–61. ISSN 0022-2496 (2000).

[43] C. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer. ISBN 0387310738 (2007).

[44] H. K. van Strachan, R.W. and Dijk. "Improper priors with well defined Bayes Factors." Technical report, Econometric Institute Report (2004).

[45] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Number v. 63 in Mathematics in Science and Engineering. Academic Press. ISBN 9780123815507 (1970).

[46] M. Khalil and A. Sarkar. "Data Assimilation for Large-Scale Computational Models." In "54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference," Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics (2013).

[47] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC Texts in Statistical Science Series. Taylor & Francis. ISBN 9781584885870 (2006).

[48] J. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer. ISBN 0387952306 (2008).

[49] M. Khalil. *Bayesian Inference for Complex and Large-Scale Engineering Systems*. Ph.D. thesis, Carleton University (2013).

[50] C. Andrieu and J. Thoms. "A tutorial on adaptive MCMC." *Statistics and Computing* **18**(4), 343–373. ISSN 0960-3174 (2008).

[51] A. O'Hagan and J. Forster. *Kendall's advanced theory of statistics, Volume 2B, Bayesian inference*. Arnold, 338 Euston Road, London NW1 3BH (2004).

[52] S. R. W.R. Gilks and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC. ISBN 0412055511 (1995).

[53] H. Haario, E. Saksman, and J. Tamminen. "An adaptive Metropolis algorithm." *Bernoulli* pages 223–242 (2001).

[54] A. Gelman, G. O. Roberts, and W. R. Gilks. "Efficient {M}etropolis jumping rules." In "Bayesian statistics, 5 (Alicante, 1994) SE - Oxford Sci. Publ.", pages 599–607. Oxford Univ. Press, New York (1996).

[55] H. Haario, M. Laine, A. Mira, and E. Saksman. "DRAM: efficient adaptive MCMC." *Statistics and Computing* **16**(4), 339–354. ISSN 0960-3174 (2006).

[56] S. Chib and E. Greenberg. "Understanding the Metropolis-Hastings Algorithm." *American Statistician* **49**(4), 327–335. ISSN 00031305 (1995).

[57] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall (1979).

[58] Q. Chen. *Approximate Kalman Filtering*. Approximations and Decomposition Series. World Scientific. ISBN 9789810213596 (1993).

[59] S. Haykin. *Kalman Filtering and Neural Networks*. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley. ISBN 9780471369981 (2001).

[60] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Number v. 160 in Applied Mathematical Sciences. Springer. ISBN 9780387220734 (2004).

[61] M. Bocquet, L. Wu, and C. Pires. "Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation." *Monthly Weather Review* (2010).

[62] Z. Chen. "Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond." *Statistics* pages 1–69 (2003).

[63] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. "A Tutorial on Particle Filters for Online Nonlinear/non-Gaussian Bayesian Tracking." *IEEE Transactions on Signal Processing* **50**(2), 174–188 (2002).

[64] N. Papadakis, E. Mémin, A. Cuzol, and N. Gengembre. "Data Assimilation with the Weighted Ensemble Kalman Filter." *Tellus Series A : Dynamic Meteorology and Oceanography* **62**(5), 673–697 (2010).

[65] W. Häardie and M. Müller. "Multivariate and Semiparametric Kernel Regression." In "Smoothing and Regression," pages 357–391. John Wiley & Sons, Inc. ISBN 9781118150658 (2000).

[66] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall (1986).

[67] A. Kong, J. S. Liu, and W. H. Wong. "Sequential Imputations and {Bayesian} Missing Data Problems." *Journal of the American Statistical Association* **89**(425), 278–288 (1994).

[68] R. Douc, O. Cappe, and E. Moulines. "Comparison of Resampling Schemes for Particle Filtering." In "Proceedings of ISPA 2005, the 4th International Symposium on Image and Signal Processing and Analysis," pages 64–69. Zagreb, Croatia (2005).

[69] N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski. "An overview of sequential Monte Carlo methods for parameter estimation in general state-space models." In "Proceedings of the 15th IFAC Symposium on System Identification (SYSID)," Ml (2009).

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

[71] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Field*. Springer-Verlag (1983).

[72] R. N. Miller, M. Ghill, and F. Gauthiez. "Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems." *Journal of the Atmospheric Sciences* **51**(8), 1037–1056 (1994).

[73] S. Lynch. *Dynamical Systems with Applications using MATLAB*. Birkhäuser (2004).

[74] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics. Springer-Verlag. ISBN 9783540540625 (1992).