Addressing challenges in nonlinear structural dynamics: Bayesian methods for input, system, and output identification.

A Thesis submitted to the University of Sheffield

for the degree of Doctor of Philosophy in the Faculty of Engineering

by

Joe D. Longbottom

Department of Mechanical Engineering

University of Sheffield

October 2024

Acknowledgements

This thesis would not have been possible without the support, guidance, and friendship of many individuals. I am grateful to each and every one of them.

Firstly, my deepest gratitude goes to Tim. His patience and generosity with his time have been invaluable throughout this journey, allowing me to explore ideas freely while grounding me when I was repeatedly convinced that everything was broken. His guidance has been instrumental not only in overcoming challenges but also in shaping the person I am today. Beyond academia, his companionship—through chess, diplomacy, and his invaluable advice—has been an unwavering source of support and inspiration. For this, I am truly grateful.

Special thanks go to Brandon for his friendship; he has been a constant presence, sitting beside me through countless hours until (nearly) the very end, helping me check my sanity when it often seemed like none remained. I am also grateful to Max, whose mentorship and guidance have been invaluable throughout my research, and to Ewan for helping me hold down the fort along the way. My thanks also go to Lizzy for her support and for checking in with me throughout the years.

I am deeply thankful to the members of the DRG, whose brilliant and inspiring community made pursuing a PhD feel more like a privilege than a challenge. Their camaraderie and enthusiasm have enriched this time of my life in countless ways. I am also grateful to many other friends who have stood by me for years, bringing endless joy and support through all the twists and turns.

Finally, to Rosie: for reminding me, even in the hardest moments, that brighter days lie ahead. Your unwavering support, patience, and understanding have been a constant source of strength, and for this, I am forever grateful.

Abstract

The field of structural dynamics is essential for predicting how systems respond to external excitations. This is crucial for optimizing performance, mitigating damage, and guiding maintenance and operational decisions. As engineering projects grow increasingly ambitious, the complexity of these systems intensifies, resulting in more pronounced nonlinear and non-stationary behaviours, as well as greater uncertainty in the physical understanding of these structures. This compounding of factors necessitates the integration of data-driven, uncertainty-aware approaches with physical models to effectively capture the dynamic behaviour. This thesis seeks to addresses these challenges by developing new Bayesian methodologies towards system identification, prediction, and input estimation in nonlinear dynamic systems.

Input identification plays a critical role in structural dynamics, involving the recovery of external forces from output-only measurements. This process is essential for characterising latent forces, providing insight into operational loads, and aiding prognosis through direct fatigue load analysis. The Gaussian process latent force model (GPLFM) has emerged as a powerful tool for input identification, allowing for the recovery of distributions over temporal latent forcing functions in the presence of measurement uncertainty. However, the application of GPLFM faces several significant challenges.

Existing approaches for joint input and state identification using GPLFM have been effective when system nonlinearities are static; however, they have not yet been applied to systems with dynamic nonlinearities. These systems not only introduce complex, time-varying nonlinearities into the system's transition function but also necessitate additional hidden states, which are not simply derivatives or integrals of other states. To address these challenges, this thesis investigates the used of the

GPLFM for joint input-state identification in systems with hysteretic nonlinearities.

Furthermore, a limitation of the standard GPLFM is the assumption that the estimated force can be modelled *a priori* as a stationary process. This assumption can become inadequate when dealing with non-stationary forces, such as those encountered in operational loads such as wind, wave, traffic, or seismic loads. To address this, a new hierarchical state-space formulation of the GPLFM is developed, designed to effectively capture smooth, non-stationary forcing functions.

The identification of nonlinearities in dynamic systems remains a challenge across engineering disciplines, serving to uncover governing physical phenomena and predict system responses to new inputs. To meet these objectives, the GPLFM can also be applied to the identification of latent restoring forces. Latent restoring force identification provides a pathway to infer critical, often nonlinear, internal mechanisms that restore equilibrium. Assuming input-output measurements are available, existing methods for latent restoring force identification have been shown to be effective. However, for many structural applications, access to input-output data is not feasible. Therefore, this work presents a new methodology to perform latent restoring force identification using output-only measurements. By combining the GPLFM latent restoring force framework with novel post-processing techniques, it is demonstrated that it is possible to jointly recover the temporal functions of the latent states, the latent restoring force, and the latent input force, as well as a GP representation of the underlying restoring force surface.

In addition to latent force identification, this thesis explores the under-exploited interface of probabilistic numerics and Bayesian (parametric) system identification. In engineering, accurately modelling nonlinear dynamic systems from data contaminated by noise is both essential and complex. Established Sequential Monte Carlo (SMC) methods, used for the Bayesian identification of these systems, facilitate the quantification of measurement uncertainty in the parameter identification process. However, these methods require numerical integration of nonlinear continuous-time ordinary differential equations (ODEs) to align theoretical models with discretely sampled data. For most nonlinear systems, the absence of closed-form solutions necessitates numerical approximations for this step, which introduces numerical uncertainty into the parameter evaluation process. This thesis develops a new methodology to efficiently identify latent states and system parameters from noisy measurements while simultaneously incorporating probabilistic solutions to ODEs into the identification process.

By addressing these challenges, this thesis bridges critical gaps in Bayesian system identification prediction, and input estimation, advancing both the theoretical foundations and practical applications of uncertainty-aware tools for the structural dynamicist.

Table of Contents

Chapter 1

Introduction

From wind turbines and milling machines to buildings and bridges, all structures move. This motion may be integral to their primary function, such as the rotation of a wind turbine blade that generates electricity, or it may be an excitation that the design seeks to mitigate, such as a resonance that causes fatigue of the blade. In both cases, the field of structural dynamics is essential for understanding how structures move, which in turn aids in optimising performance, mitigating damage, targeting maintenance, and informing operational decisions. In pursuit of these goals, the aims of a structural dynamicist can be viewed through three fundamental tasks:

- **System Identification**: the construction of a mathematical model that represents the essential features and behaviours of a system.
- **Output Identification/Prediction/Simulation**: The evaluation of a structure's response using a mathematical model, from inputs and initial conditions.
- **Input Identification**: the inference of the inputs to a structure from an observed dynamic response.

1.1 System identification

A system can be loosely defined as any set of interrelated components that work together to achieve a specific function or behaviour. In the context of structural dynamics, a system is typically understood as a temporal function that relates inputs (forces) to outputs (structural responses). System identification is the task of learning a mathematical model that represents this input-output relationship as a functional map.

Models are essential in structural dynamics as they enable the prediction and analysis of structural responses under various loading conditions, which is fundamental to designing safe and efficient structures. They facilitate design optimisation and the evaluation of potential performance issues, enhancing the understanding of the underlying physical phenomena. Furthermore, models are critical for structural health monitoring (SHM) [1, 2], allowing online condition assessment, and for digital twins (DT) [3, 4], which support advanced simulations and predictive maintenance. In fact, the utility of an accurate and *trusted* model to an engineer is almost endless. The challenge lies in building them.

System identification can conventionally be considered to follow two paths: physics-based (white-box) and data-driven (black-box).

White-box models are derived from fundamental principles and laws of nature—such as Newton's first law and Hooke's law—leveraging the engineer's insight into the physics of the structure to causally describe the observable behaviours of systems.

When a system exhibits linear behaviour and input-output observations are available, the task of system identification is well understood. In structural dynamics, *modal analysis* (MA) is the standard method for identifying linear dynamics. MA provides a compact, decoupled representation of the dynamics and offers physical insight through natural frequencies and modeshapes [5].

However, almost all systems exhibit some degree of nonlinearity, ranging from minor inconveniences to significant impacts. Unlike linear systems, nonlinear systems do not obey the principle of superposition and cannot be decoupled by linear transformations, rendering linear modal analysis inapplicable. Standard techniques commonly involve linearisation, but this is inadequate when nonlinearity is significant in the systems dynamics.

As engineering and technology advance, leading to increasingly ambitious projects and the use of advanced materials in extreme environments, the prevalence and importance of understanding nonlinearities in structural dynamics are becoming more pronounced. The ability to model and identify nonlinear systems accurately is thus becoming crucial. Nonlinearities in structures can manifest in various ways, from causing amplitude-dependent resonance frequencies to more complex phenomena such as bifurcations [6], jump phenomena [7], and chaos [8].

Nonlinearities in engineering systems arise from various sources. Material properties, such as non-homogeneous composites [9] and the nonlinear response of materials beyond their elastic limit, play a significant role. Hysteresis [10], or the dependence of a material's response on its deformation history, results in energy dissipation and lag between applied forces and displacements. Geometry-induced nonlinearities lead to complex load-deformation relationships [7]. Interfaces and boundary conditions, such as friction and contact nonlinearity, including Coulomb friction [11] or slip regimes in bolted joints [12] and soil-structure interaction [13] further complicate the behaviour of the system. Moreover, linear analysis cannot adequately address phenomena such as rocking [14], buckling [15, 16], and structural failure [17], which are inherently nonlinear.

If the mathematical model of the nonlinear system can be derived from first principles, parameter estimation techniques—where specific parameters are tuned using input-output measurements—can provide a powerful framework for system identification within a white-box framework. However, constructing accurate models from first principles is challenging due to the complex dynamics and interactions of interrelated components, and nonlinearities only further complicate this process.

When systems exhibit unknown dynamics or nonlinearity, the white-box approach–which employs engineering insights to describe system behaviors from fundamental principles–can be restrictive. Conversely, black-box models rely solely on empirical input-output measurements to infer a functional map that reflects observed correlations. However, these models do not inherently ensure that the inferred correlations correspond to causal relationships, potentially limiting their interpretability and reliability and undermining an engineer's trust in the model.

Grey box models combine elements of both white box and black box models. These models incorporate partial physical knowledge of the system, described from fundamental principles, and empirical data-driven components, allowing for improved

accuracy and interpretability in scenarios where complete physical understanding is not available.

Each approach – physics-based, data-based, and hybrid – possesses distinct merits and drawbacks [18], which will be discussed in more detail later. However, a common requirement for all these approaches is the need for observed data. Data-based methods rely heavily on measurements to infer correlations, while physics-based models necessitate data for validation purposes and grey box as a hybrid can require both.

The acquisition of observational data is inherently fraught with uncertainties due to limitations in the data acquisition process. A significant source of uncertainty is incomplete information, also referred to as epistemic uncertainty. Epistemic uncertainty can arise for several reasons: (1) sparse measurements that fail to provide a comprehensive depiction of system behaviour; (2) the impracticality of achieving full state observation; (3) indirect measurements of the variable of interest; and (4) the lack of comprehensive data on structural behaviours in operational, extreme environmental, and damage conditions. Obtaining the necessary data to reduce these uncertainties is often prohibitively expensive or impossible due to physical and safety constraints.

Another layer of uncertainty is the inherent stochasticity present in the identification, referred to as aleatoric uncertainty. Measurement noise is a primary example of this, as it can obscure the true state of the system, thereby complicating the modelling process.

Summary

Accurate modelling of nonlinear systems is essential for the continued design, construction, and maintenance of advanced structures in a cost-effective and environmentally sustainable manner. As engineering projects become more ambitious, the complexity of these systems increases, leading to greater uncertainty in the physical knowledge of the systems and more pronounced nonlinear behaviours. This increased uncertainty and nonlinearity necessitate the development of advanced data-based methods to supplement traditional modelling approaches. Given that data is inherently uncertain, the models constructed from this data must also be uncertain. It is imperative, therefore, to quantify these uncertainties to inform engineering deci-

sions. By providing a more comprehensive assessment of uncertainty, engineers can consider the probabilistic nature of their models, thereby making risk-aware choices in complex engineering applications.

1.2 Prediction and simulation

In an engineering context, prediction refers to forecasting a system's behavior one step ahead, whereas simulation involves forecasting its behavior over an extended time horizon. While both tasks are crucial in structural dynamics, prediction is strictly a subset of simulation, with accurate simulation posing a significantly greater challenge.

Simulation of structural responses to operational loads is an essential aspect of design optimisation and validation [19, 20]. The foundation of an accurate and reliable model enables engineers to evaluate structures under diverse conditions, including scenarios that are impractical or prohibitively expensive to recreate experimentally. This capability allows for the optimisation of design parameters in a more efficient and cost-effective manner. In addition, simulation is essential for advanced decision support and predictive maintenance in DTs.

Prediction and simulation are also essential to the system identification process itself. Prediction plays a critical role in system identification, when iterative model simulations are necessary to evaluate likelihoods. Such as required for parameter estimation [21, 22], Bayesian filtering and smoothing [23], and Markov Chain Monte Carlo (MCMC) methods [24, 25].

Simulation plays a crucial role in model validation, where it is used to evaluate model performance on a validation dataset. This process ensures that the model adequately captures the system dynamics and generalises well to new data.

Additionally, simulation can be necessary for hyperparameter optimisation. By evaluating different configurations through performance metrics, simulation helps in identifying the optimal hyperparameters that enhance model generalisation. This iterative process of assessing various hyperparameter settings ensures that the model performs well on unseen data.

A significant challenge in prediction and simulation arises when uncertainty is con-

sidered. Uncertainty can originate from various sources: the system's initial conditions may be uncertain, the functional form or parameters of the model may be imprecisely known, and in cases where closed-form solutions are unavailable, numerical methods must be employed. These methods are inherently approximate [26], introducing further uncertainty into the integration process.

This uncertainty in numerical integration is particularly pronounced in nonlinear systems. While solutions to linear models are generally exact and solvable in closed form, this is rarely the case for nonlinear models. The absence of closed-form solutions in nonlinear systems necessitates the use of numerical methods, which are inherently approximate and, therefore, introduce uncertainty into simulation and prediction.

Whether uncertainty arises from initial conditions, the system model, or from approximations and the inherent uncertainty in numerical integration, it can significantly affect various critical areas of engineering applications. This uncertainty propagates through predictions and simulations, impacting the reliability of these methods. In design optimisation and validation, uncertainties can undermine the integrity of design assessments. For model validation, uncertainties obscure the true reliability of models, making it difficult to accurately assess model performance and evaluate its validity. In system identification, uncertainty directly influences the evaluation of likelihoods, introducing additional and often unaccounted-for uncertainty into the model identification process.

Summary

Prediction and simulation are essential tasks in structural dynamics. When uncertain initial conditions, uncertain models or numerical solutions are employed in the prediction or simulation of nonlinear systems, the process inherently introduces uncertainty, which propagates through both predictions and simulations, influencing any applications that rely on these methods. Consequently, the accurate quantification and management of this combined uncertainty is critical for ensuring the reliability and robustness of any conclusions drawn from these processes.

In the context of numerical uncertainty, there is no unique solution to the system identification problem when numerical integration is involved. Instead, the 'optimal' model identification is influenced by the numerical errors present in the simulation.

As a result, system identification for nonlinear systems is inherently uncertain, necessitating a probabilistic approach to properly account for this uncertainty. This principle underpins the probabilistic perspective on numerical methods, referred to as Probabilistic Numerics (PN) [27, 28]. By explicitly addressing uncertainty within the system identification process, a more comprehensive evaluation of overall uncertainty can be achieved.

1.3 Input identification

In the context of structural dynamics, the term input refers to external forces or excitations applied to a structural system that induce a dynamic response. These inputs can be deterministic–such as harmonic loads or impulsive forces–or stochastic–such as random vibrations. The task of input identification is concerned with recovering the inputs from the output only measurements.

Input identification is essential in structural dynamics for robust design. It provides precise characterisation of forces and excitations, clarifying operational loads. This understanding enables the development of accurate models, optimises design parameters, and ensures structural resilience and performance under extreme conditions.

Moreover, input identification can have applications in SHM [29]. Prognosis is a primary challenge in SHM [30, 31]. It involves predicting future structural conditions and performance based on current health assessments and historical data. By estimating the remaining useful life and potential failure modes of a structure, prognosis analyses trends in degradation and damage accumulation. An accurate prognosis enables timely maintenance and repair, optimising the lifespan and safety of the structure while minimising downtime and costs.

An approach to the prognosis in SHM is through direct fatigue load analysis, which relies on the estimation of input loads [32–36]. Latent force modelling can be used to help identify these loads, providing a potential path to an accurate fatigue load analysis. Furthermore, by quantifying operational loads and assessing the current damage state, it becomes possible to predict the expected future damage. This can facilitate timely maintenance and improve remaining life predictions.

Input identification is similar to system identification and relies heavily on data, specifically data from structures that contain both epistemic and aleatoric uncer-

tainty. Inputs identified from uncertain data must also carry inherent uncertainty. This uncertainty should be quantified to effectively inform engineering decisions.

One model type increasingly recognised within the structural dynamics community for input identification, or as it is otherwise known, latent force modelling, is the state-space Gaussian process latent force model (GPLFM). The GPLFM allows for the recovery of a distribution over the time series of the forcing function, given the uncertainty in the measurements. This probabilistic approach provides a robust framework for handling the inherent uncertainties in the identification.

Despite its strengths, which will be discussed in more detail later (Chapter 3), it has some limitations and drawbacks. For one inference with the GPLFM becomes more challenging when the system is nonlinear due to the lack of a closed-form solution for the inference process. Moreover, the state-space GPLFM inherently assumes that the unknown forces are stationary *a priori*, a presumption that can hinder accurate force recovery when the underlying temporal functions are non-stationary. This is often the case for environmental loads such as wind [37], wave [38], earthquake [39], and traffic loads [40]. Non-stationary forces require models that can adapt to changing statistical properties over time to ensure accurate force estimation and subsequent structural analysis.

Summary

Input identification is a critical process in structural dynamics, involving the recovery of external forces from output-only measurements. This process is essential for characterising latent forces to clarify operational loads. Moreover, input identification can contribute to SHM by aiding prognosis through direct fatigue load analysis. The state-space GPLFM has emerged as a powerful tool for latent force modelling, allowing for the recovery of a distribution over the forcing function time series in the presence of measurement uncertainty. However, the application of GPLFM faces challenges in nonlinear systems due to the absence of closed-form inference solutions, and in scenarios with non-stationary forces, such as environmental loads. Working to reduce these limitations is crucial for further developing the GPLFM and its application to structural dynamics.

1.4 Joint identification

One challenge that has been largely undiscussed so far is the effect and interlinkage of the three fundamental aspects of structural dynamics: system identification, prediction, and input identification. These aspects are intrinsically connected, forming parts of the same overarching problem – starting from the input, the system model maps to the outputs. Consequently, solving any one of these challenges conventionally requires knowledge of the other two:

- System identification requires inputs and outputs.
- Prediction requires inputs and the system model.
- Input identification requires the system model and outputs.

This interdependency poses a significant problem for both system identification and input identification, especially when only output measurements are available. Moreover, even when outputs are measured, they often provide an incomplete description of the states, featuring sparse or indirect measurements of the variables of interest – acceleration being the most common due to the availability of low-cost piezoelectric accelerometers [5]. This problem is only further complicated by measurement noise.

Therefore, there is a need to jointly identify both the hidden states and the system model, or both the hidden states and the system input from noisy output-only measurements. In the ideal case, one would recover the hidden states, the system model, and the system input from noisy output-only measurements, which presents a significantly more challenging task. The difficulty in joint identification arises from the lack of unique solutions – numerous combinations of states, inputs, and models can yield the same output.

Summary

Joint identification is essential to ensure that system identification and input identification techniques can be applied to real-world structural dynamics problems where only incomplete output measurements are available. However, this can be particularly challenging due to the fundamental unidentifiability of the general problem.

1.5 Thesis contribution

Throughout this introduction, several challenges arising from the fundamental tasks of the structural dynamicist – system identification, prediction, and input identification – have been discussed. These challenges primarily emerge due to complications introduced by nonlinearity, uncertainty, and non-stationarity, as well as the difficulties associated with joint identification. As it is well beyond the scope of this thesis to fully address or completely solve these challenges, a more focused approach to expand the state-of-the-art must be taken. The primary aim of this thesis has been to advance Bayesian methods for input, system, and output identification in nonlinear dynamic systems. Specifically, four core contributions are presented:

1. A new implementation of the GPLFM for nonlinear joint input-state estimation is introduced. This contribution extends the approach to systems with dynamic nonlinearities, such as those exhibiting hysteresis. These systems pose a significant challenge as they not only introduce dynamic nonlinearities into the system's transition function but also require additional hidden states, which are not simply time derivatives or integrals of other hidden states.

2. A new methodology to perform latent restoring force identification using output-only measurements is presented. In this work, the GPLFM latent restoring force approach is combined with novel post-processing steps within an Operational Modal Analysis (OMA) framework, by assuming the system input is a white noise process. From this, it is demonstrated that the temporal functions of the latent states, latent restoring force, and latent input force can be recovered, along with a GP representation of the underlying restoring force surface.

3. The State-Space GPLFM has proven effective in recovering latent input and restoring forces, as well as uncovering the functional forms of nonlinearity in dynamic systems. Additionally, it provides a practical approach for jointly identifying latent forces, system states, and model parameters. One limitation of the GPLFM is that the estimated force is modeled *a priori* as a stationary process. This limitation is particularly evident when dealing with non-stationary forces, which display time-varying characteristics that defy simple modeling techniques. A primary difficulty lies in identifying latent input forces with evolving spectra and latent restoring forces, where nonlinearity affects

the system's stiffness and consequently changes the frequency content of the restoring force. Recognizing this gap, this contribution addresses these challenges through the development of a non-stationary GPLFM by introducing a time-varying length scale into the GPLFM, which is modeled as an additional GP.

4. The application of a probabilistic numerical method for solving ordinary differential equations (ODEs) in the joint parameter-state identification of nonlinear dynamic systems within a sequential Monte Carlo (SMC) framework is presented. This approach efficiently identifies latent states and system parameters from noisy measurements, simultaneously incorporating probabilistic solutions to the ODEs in the identification challenge. The primary advantage of this methodology lies in its capability to produce posterior distributions over system parameters, thereby representing the inherent uncertainties in both the data and the identification process.

Chapter 2

Background

This chapter provides an overview of the background and theoretical foundations necessary to develop a framework for the research presented in this thesis. A brief review of some relevant literature is included to establish a context for subsequent discussions. However, a more thorough and detailed examination of the pertinent literature–which identifies gaps in the current state of the art and motivates the research in relation to the four objectives of this thesis–is provided in Chapters 3 to 7.

2.1 System identification

To take any action on a system with intent, an understanding of the system is required. Intentional action relies on the ability to predict, control, and fundamentally to understand the dynamics of a system. Systems can be understood through the construction of models. Models serve as a representation of a system, designed to capture its essential features and behaviours for the purpose of understanding, predicting, and controlling it. Examples range from the subconscious—with internal neural models that allow us to run, walk, and jump by combining sensory inputs with motor commands [41]—to the grand—with models of celestial mechanics that enable scientists to predict planetary positions and design trajectories for spacecraft and satellites [42]. Understanding systems through the lenses of a model is essential in all areas of life, making the construction of these models equally important.

Models, whether white-box or black-box, can be constructed through two primary approaches. The first approach involves gathering empirical data by observing and measuring a system's interactions with its environment. This data is then used to develop mathematical formulations that capture the system's essential features and behaviours. The second approach involves refining or extending pre-existing models, based on new data, or combining them to describe more complex systems. In both cases, the foundation of all models is empirical observation.

This principle applies not only to the foundational elements of mechanics, such as Newton's laws of motion [43], which describe the causal effects of forces on the motion of mass, but also extends to 'opaque', high-dimensional models like transformer neural networks [44], which are widely used in large language models and are often characterised by a correlational nature.

Observing and measuring the world to build models that can then be used to understand and influence it is a core pursuit of science. One field that addresses this challenge directly for dynamic phenomena is *system identification*.

Due to the variety of approaches to model building across different scientific disciplines, the lines between system identification, machine learning, statistics, and other related fields are often blurred [45]. In addition, system identification techniques have been applied in fields ranging from control [46, 47] to machine learning [48], mechanical engineering [49–51], and biology [52], with the potential for cross-fertilisation [53] between these areas of research.

This thesis will focus on system identification as it pertains to the learning of mathematical representations of structural dynamics systems from both physical insight and measured data, while also quantifying the inherent uncertainties in the identification process. To this end, tools from across the fields of science will be considered.

2.1.1 A spectrum of models for system identification

All models are fundamentally rooted in data, yet their construction varies significantly. Although the conventional distinction between white-box and black-box models provides a useful starting point, a more contemporary perspective conceptualises these approaches as the two extremes of a continuous spectrum. This framework acknowledges that the construction of models varies significantly, allowing

them to be positioned along a continuum based on the degree to which they incorporate *a priori* physical principles versus empirical data.

At one end of this spectrum are the white-box models, or physics-based models, which are explicitly grounded in fundamental physical principles and laws of nature. This principled approach in their construction provides a transparency that is especially valuable in engineering, where understanding the underlying mechanisms is critical to building trust. Moreover, by capturing causal input-output relationships, these models can enable robust extrapolation beyond the range of observed data, making them particularly effective in scenarios where predictive accuracy and interpretability are paramount.

However, the applicability of white-box models can be limited when dealing with complex, real-world systems. In such cases, the simplifications and assumptions required to formulate a physics-based model can lead to inaccuracies. Capturing the full range of behaviours in a complex system might necessitate an unmanageable level of detail or result in a model that is too computationally expensive to be practical. Additionally, when the underlying physical processes are not well understood or are influenced by numerous interacting factors, constructing a white-box model that accurately reflects reality can be challenging, if not impossible.

On the other hand, black-box or data driven models primarily rely on data to identify patterns without explicitly incorporating causality or underlying physical laws. The principal advantage of black-box models is their capacity to recover complex mappings between inputs and outputs, assuming the availability of sufficient and high-quality data. When the dataset extensively covers the relevant domain and is sufficiently informative, black-box models can effectively capture complex dynamics that are challenging to derive from first principles. This capability is particularly valuable in contexts where the underlying physical processes are either highly complex, poorly understood, or involve numerous interacting factors, rendering the development of a reliable white-box model impractical or infeasible. In such scenarios, black-box models provide a powerful alternative.

However, the limitations of black-box models are significant. One of the most critical challenges is that engineering data, while sometimes abundant, often fail to comprehensively capture all behaviours of interest. The available data might consist of indirect measurements of these behaviours, and in many operational settings, the data may be noisy, corrupted, or incomplete. Such deficiencies in data quality and

coverage can severely restrict the model's ability to generalize beyond the training set, thereby reducing its reliability in predicting unseen scenarios. Furthermore, black-box models are inherently less interpretable as they do not explicitly encode physical laws or causal relationships. This lack of transparency can lead to issues of trust, especially in high-stakes applications where understanding the rationale behind a model's predictions is critical. Additionally, the absence of an embedded causal structure often results in poor extrapolation to new, unseen conditions, making predictions outside the training domain potentially unreliable [1].

Between these two extremes lie grey-box models, which combine elements of both approaches. Grey-box models incorporate physical insights into data-driven frameworks, balancing the extrapolative power, interpretability, and trust associated with white-box models with the flexibility and expressive power characteristic of black-box models. This hybrid approach offers a more nuanced solution in situations where purely data-driven or physics-based methods might be insufficient.

Understanding the diverse approaches to modelling is crucial to selecting the appropriate method for a given problem. An effective visualisation of the spectrum of physics-informed models is presented by Cross et al. [55], which illustrates how these models can be positioned on axes representing the data-driven and physics-based problem settings.

The choice of modelling approach in system identification often hinges on the balance between how much of the system can be described by known or modelled physics (within a given computational budget) versus how much can be characterised by available data. This relationship can be visualised in Figure 2.1, which maps different scenarios and the corresponding optimal modelling approaches.

Figure 2.1 depicts a plot divided into five distinct sections. Section A represents a scenario where a significant amount of the system's physics is well understood, but little data is available, or the data that is available is not very informative. In this context, a white-box approach, which relies heavily on physical laws and principles, is preferable.

In contrast, Section C illustrates the opposite situation, where there is minimal knowledge of the underlying physics, but a large quantity of high-quality data is

[1] Although some correlations may correspond to causal relationships, this is not guaranteed. Ensuring causality in data-driven models requires causal inference algorithms, which determine the direction and nature of these relationships. For further details, see [54]

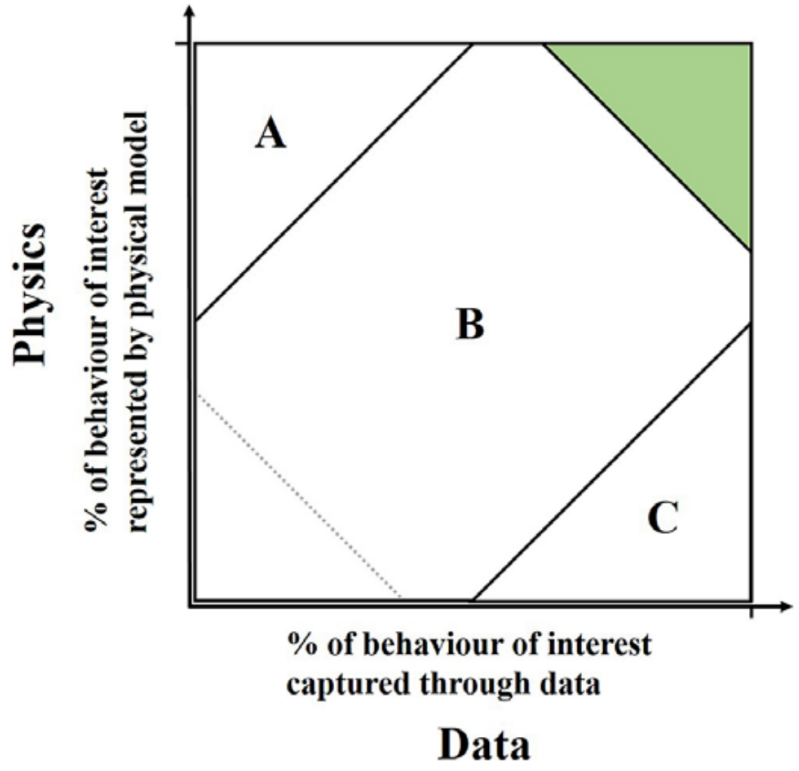

Figure 2.1: The mapping of problem settings based on the knowledge available from physical insight and data [55].

available. Here, black-box approaches, which depend primarily on data to infer complex relationships, are most effective.

The green area in the figure represents an ideal scenario in which both the physics of the system is well understood and there is an abundance of high-quality data. In this region, any method—be it white-box, black-box, or a hybrid—can be applied effectively.

In the bottom left corner, the section depicts a situation where little is known about the system, either from a physical or data perspective. In this case, the system cannot be accurately modelled using existing methods. More research or the collection of additional data is needed to move out of this region and enable effective system identification.

Finally, Section B occupies the middle ground. This area represents scenarios where both physical insight and data are available, but neither is sufficiently complete to independently identify the system optimally. In Section B, a grey-box approach, which combines elements of both white-box and black-box models, is most suitable. By integrating partial physical knowledge with available data, grey-box models can leverage the strengths of both approaches to achieve better accuracy and reliability than either method could achieve alone.

To illustrate the practical application of these concepts across different modelling approaches, Figure 2.2 visualises how various methods within the field of system identification align with the spectrum of available physical knowledge and data. Figure 2.2 provides a broad mapping of several widely recognised methods within the system identification field onto the axes of Figure 2.1. It is important to note that the position of each method on this spectrum may vary considerably depending on the specific model type and the context in which it is employed.

2.1.2 Nonlinear system identification

One of the paramount challenges in system identification is the accurate modelling of nonlinear systems. Nonlinear system identification has emerged as a vibrant area of research, with methods employed across the spectrum—from white-box to black-box approaches. This diversity is well documented in several comprehensive review articles [56–58], which highlight the evolution and breadth of techniques developed

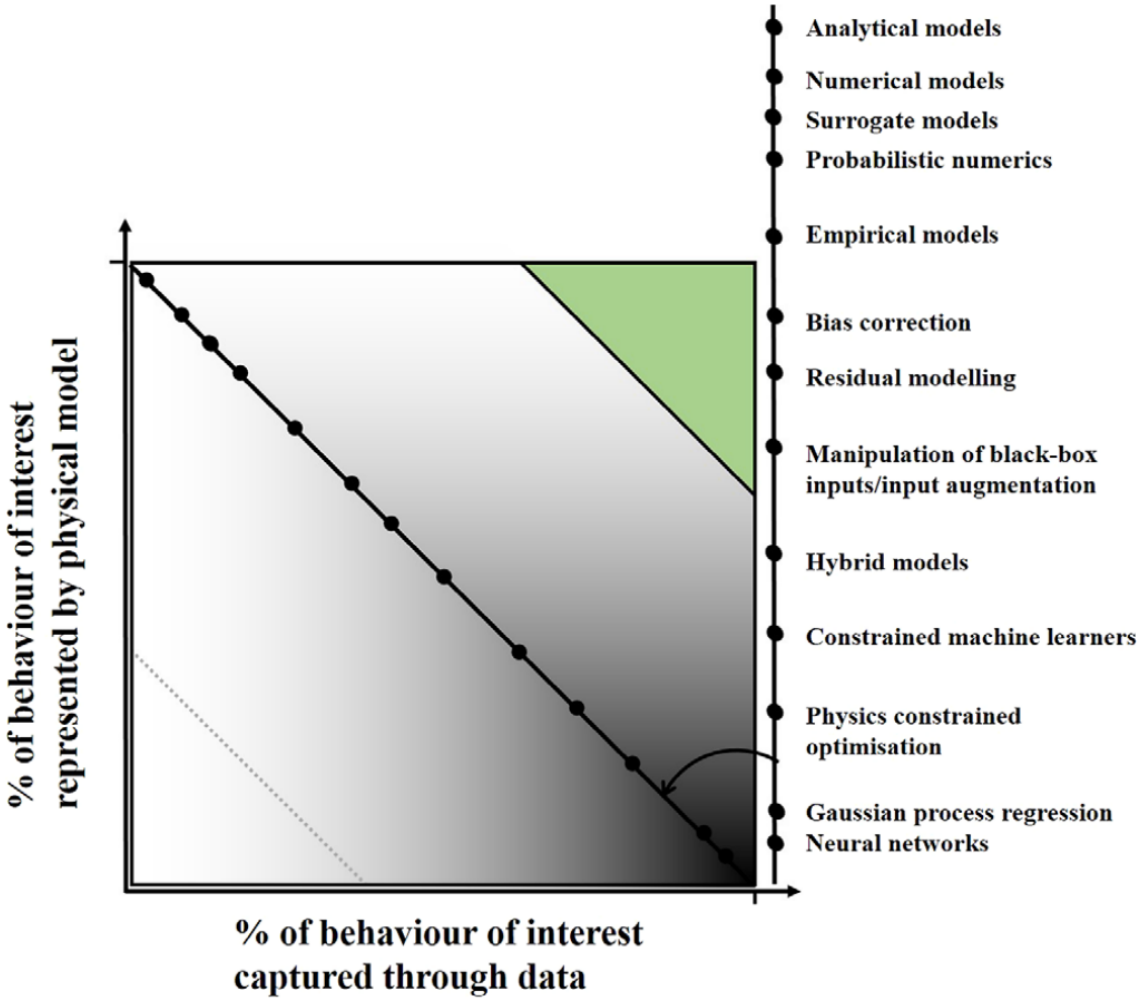

Figure 2.2: A non-exhaustive list of modeling approaches loosely aligned with the data/physics problem setting, as depicted in Figure 2.1 [55]

to tackle nonlinear dynamics.

Given the expansive nature of the field, a complete discourse on all existing methods is beyond the scope of this thesis. Instead, the following sections will delve into selected works that are particularly relevant to the objectives of this study.

At the lighter end of the spectrum–where gaining a physically meaningful representation of system dynamics is a priority–the nonlinear system identification process can often be considered as a progression through three stages [56]:

1. **Detection**: the determination of whether nonlinearity is present in the structural behavior.
2. **Characterisation**: the localization of the nonlinearity, determination of its type, and selection of an appropriate functional form.
3. **Parameter estimation**: the estimation of the nonlinearity's functional parameters and the evaluation of their uncertainty.

Under this framework detection is the initial step in nonlinear system identification, with various methods proposed for this purpose [5, 7, 59, 60]. Upon confirmation of nonlinearity, the process advances to characterisation and parameter estimation.

One classic approach to characterisation involves the use of residual methods. These methods combine a physical component with an additional data-driven component, which is employed to compensate for errors in the physical model. A well-known technique within residual methods is the restoring force surface method introduced by Masri and Caughey [61], with similar work independently developed under the name of force-state mapping by Crawley and Aubert [62] and Crawley and O'Donnell [62].

In the seminal work of Masri and Caughey, the behaviour of a single degree-of-freedom (SDOF) system was analysed using the restoring force surface method [61]. The 'restoring force' refers to the internal forces that act to return the system to equilibrium, governed by an unknown, potentially nonlinear function.

Masri et al. demonstrated that, given the mass, acceleration, and applied external force of the system - available through experimental data - the equation of motion can be rearranged to isolate and identify this unknown restoring force. This approach

aligns with Newton's first law, ensuring that the internal forces balance the external forces acting on the system.

To estimate the restoring force, the original study employed Chebyshev orthogonal polynomials as a modelling tool. By fitting these polynomials to the observed data, the nonlinear behaviour of the system can be characterised, enabling predictions of the system's response to new external forces or providing insight into the physical nature of the restoring force.

Probabilistic approaches to residual modelling have also been explored [63–65], including the use of the Gaussian process nonlinear AutoRegressive model with eXogenous inputs (GP-NARX) [29], which is a dynamic extension of the widely recognised Gaussian process (GP) regression method [66]. This grey-box approach has been used to model the discrepancy between measured data and a physics-derived model [51, 67, 68]. In the finite element modelling community, probabilistic residual modeling using GPs has also been investigated [69, 70].

Building on the concept of the restoring force surface method, GPs have also been applied to latent input force modelling using the Gaussian process latent force model (GPLFM) [71]. While the restoring force surface method focusses on identifying the missing internal forces that act to return the system to equilibrium, latent input force modelling seeks to determine the missing external forces that produce an observed system output. State space variants [72, 73] of the GP have also been applied to the GPLFM for latent force modelling [74–78]. In these works a state-space model of the system be it linear or nonlinear is derived from first principles, and a state-space GP is used to infer the latent input.

Further, the GPLFM has been extended to address the latent restoring force problem [79], providing an alternative probabilistic approach to recovering internal restoring forces and thus deriving distribution over a functional form of a latent nonlinear restoring force surface.

In parallel, GP state-space models have found application within a grey-box framework of probabilistic numerics, where the GP is used to capture the uncertain in numerical integration [27, 28].

If the functional form of a nonlinear system can be characterised or described directly, the problem of system identification can often be reduced to parameter estimation. Parameter estimation is central to nonlinear system identification, where

the goal is to determine the model parameters that accurately represent the dynamic behaviour of a system. This process typically involves optimising the model to minimise the difference between its output and observed data.

Non-probabilistic techniques, such as evolutionary optimisation, have been effectively applied to parameter estimation in nonlinear systems. Worden et al. have extensively used these techniques in various nonlinear benchmarks [51], including the identification of hysteretic systems [80] and experimental structures [81].

More recently, Bayesian methods have gained prominence in the parameter estimation process, as researchers have increasingly recognised the importance of quantifying uncertainties in nonlinear system identification. The Bayesian framework offers a robust probabilistic approach to this challenge.

Beck and Katafygiotis [82] introduced a Bayesian statistical framework for updating models and their uncertainties. Muto and Beck [83] further enhanced this framework by integrating stochastic simulation for Bayesian updating and model class selection, particularly in the context of hysteretic systems. In parallel, Wills et al. [84] and Schön et al. [85] explored parameter estimation in nonlinear dynamic systems using a maximum likelihood framework.

To address the challenges of uncertainty in parameter estimation, Bayesian methods have been combined with Markov chain Monte Carlo (MCMC) techniques, as explored by Green and Worden [86], and with Sequential Monte Carlo (SMC) methods, as demonstrated by Schön et al. [87]. Ebrahimian et al. [88] applied the Extended Kalman Filter (EKF) for material parameter estimation in nonlinear finite element models, showcasing the effectiveness of direct differentiation methods within a Bayesian updating framework. Additionally, particle filters have emerged as powerful tools for Bayesian state and parameter estimation in uncertain dynamical systems. Ching, Beck and Porter [89] used particle filters in this context. Semi-parametric methods have also been employed in the identification of Wiener systems, where a Gibbs sampling method was used to recover the distribution over the parameters and the state of a linear dynamic system, while a non-parametric GP was used to capture the static nonlinearity in the observation model [90].

2.2 Ordinary differential equations

Structures in the physical world display dynamic behaviours as their internal energy states evolve over time. For instance, a suspension bridge subjected to wind and traffic loads sways and flexes, continually converting kinetic energy into potential energy and back again. Understanding these dynamic behaviours is crucial for modelling the response of structures to various external forces.

To achieve this understanding, a mathematical model that accurately describes the temporal evolution of structures is required. Differential equations, which inherently express changes in quantities, are particularly well-suited for this purpose. This section explores the dynamics of structures through the framework of Ordinary Differential Equations (ODEs).

An ODE [91] is an equation that consists of one or more functions of a single independent variable and their derivatives, which provides a means to describe temporal evolution. The general form of a temporal ODE is expressed as:

$$\frac{d^n y(t)}{dt^n} + a_{n-1}(t)\frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + a_1(t)\frac{dy(t)}{dt} + a_0(t)y(t) = u(t), \tag{2.1}$$

where $y(t)$ is the function representing the system's response over time, $a_i(t)$ are the coefficient functions that may vary with time, and $u(t)$ represents external forces or inputs acting on the system. In this context, dt denotes an infinitesimally small increment in time.

2.2.1 Second-order ODEs in structural dynamics

In structural dynamics, the most common ODEs are second-order, as they naturally arise from Newton's second law of motion. According to this fundamental principle, the force $u(t)$ acting on a mass m causes an acceleration $\ddot{y}$, which can be mathematically expressed as:

$$u(t) = m\ddot{y}, \tag{2.2}$$

where each over-dot indicating a derivative with respect to time. Thus, when a force is applied, a structure accelerates in the direction of the force.

Structures inherently resist deformation, and this resistance manifests as a restoring

force that attempts to return the structure to its equilibrium position. For materials that exhibit linear elasticity, this restoring force is described by Hooke's law:

$$u(t) = ky, \tag{2.3}$$

where k denotes the stiffness of the structure and x is the displacement from the equilibrium position.

Equating these forces results in a second-order homogeneous equation differential that describes the dynamics of a simple oscillator. The simple-harmonic equation of motion is,

$$m\ddot{y} - ky = 0 \tag{2.4}$$

Under realistic physical conditions where m and k are positive, all non-trivial solutions to this equation are periodic. Consequently, the structure will display oscillatory motion during free vibration, with a period defined by the system's natural frequency.

Under this framework a structure when excited would oscillate forever. In practical scenarios, it is essential to consider the mechanisms through which energy is dissipated in a structure. One common model for energy dissipation is linear viscous damping, where the damping force is proportional to the velocity of the mass. This damping force can be expressed as,

$$u(t) = c\dot{y} \tag{2.5}$$

where c is the damping coefficient and $\dot{y}$ is the velocity.

By combining these forces—restorative and damping—with the inertial force, the equation of motion for a SDOF system subjected to an external force $u(t)$ can be formulated as:

$$m\ddot{y}(t) - c\dot{y}(t) - ky(t) = u(t). \tag{2.6}$$

Nonlinear second-order ODEs in structural dynamics

Eq.(2.6) provides the second-order ODE for a linear SDOF system. While linear second-order ODEs provide a foundation for understanding structural dynamics, many real-world systems exhibit nonlinear behaviours that cannot be accurately captured by linear models.

The general form of a nonlinear second-order ODE in structural dynamics can be expressed as:

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) + f(\ddot{y}, \dot{y}, y) = u(t). \tag{2.7}$$

where $f(\ddot{y}, \dot{y}, y)$ is some nonlinear function.

2.2.2 State-space models

Throughout this thesis, second-order ODEs are predominantly considered within the equivalent state-space representation [92]. This representation projects the second-order ODE into a set of first-order ODEs by focussing on the evolution of the *state* of the system.

In the context of state-space models (SSM), the term state is defined as the complete set of variables that encapsulate all necessary information to describe the system's current situation, thus determining its future behaviour independently of past inputs. Correspondingly, the *state vector* $\mathbf{x}(t)$ is a column vector where each element represents one of the essential *state variables*,

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_d(t) \end{bmatrix}$$

where $x_1(t), x_2(t), \ldots, x_d(t)$ represent the state variables and d is the dimension of the vector.

The first half of the SSM is the transition function, which mathematically characterises how the state of a system evolves over time. The linear time-invariant

continuous-time transition function is typically expressed as:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2.8}$$

where, $\dot{\mathbf{x}}(t)$ represents the derivative of the state vector $\mathbf{x}(t)$ with respect to time, indicating the instantaneous rate of change of the state. The matrix $\mathbf{A}$, known as the transition matrix, defines the intrinsic dynamics of the system, independent of any external inputs, while the matrix $\mathbf{B}$ determines how the input vector $\mathbf{u}(t)$ influences the evolution of the state.

A significant advantage of the state-space formulation is its ability to incorporate an observation model, which is crucial in practical scenarios. Typically, when measurements are taken from a structure, not all states of the system are directly measured or even measurable. The state-space model addresses this by explicitly modelling the internal states of the system and including an observation model that links these states to measured quantities. The linear continuous-time observation model is expressed as:

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \tag{2.9}$$

where $\mathbf{y}(t)$ denotes the output or observed vector, $\mathbf{C}$ is the observation matrix and $\mathbf{D}$ often represents the direct transmission path from input to output.

When dealing with nonlinear systems, the state-space formulation can be generalised to accommodate the complexities that arise from nonlinearity. In such cases, the system dynamics and the observation model are described by nonlinear functions. The general formulation for a nonlinear continuous-time state-space model is given by the following equations:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t)) \tag{2.10}$$

$$\mathbf{y}(t) = g(\mathbf{x}(t), \mathbf{u}(t)) \tag{2.11}$$

The function $f(\mathbf{x}(t), \mathbf{u}(t))$ describes the nonlinear relationship between the current state $\mathbf{x}(t)$, the input $\mathbf{u}(t)$, and the rate of change of the state. Similarly, the observation model is given by $\mathbf{y}(t) = g(\mathbf{x}(t), \mathbf{u}(t))$ represents the nonlinear mapping from the state vector and the input to the observed outputs.

Summary

SSMs will play a foundational role throughout this thesis. They are preferred over second-order ODEs for several reasons. Firstly, the incorporation of an observation model, as previously mentioned, allows for a direct connection between the theoretical model and real-world measurements. Secondly, SSMs facilitate straightforward simulation techniques: linear systems can be simulated using the matrix exponential, while nonlinear systems can be painlessly implemented into numerical methods. Third, SSMs provide a robust framework for handling uncertainty, both in the measurements and in the transition model itself. Fourth, the framework supports optimal estimation techniques, including filtering and smoothing, which are critical for accurate state estimation in the presence of noise and other uncertainties. Finally, the modular nature of SSMs allows for simple extensions to multi-degree-of-freedom (MDOF) systems and other augmentations essential for efficient latent force modelling. An introduction to these topic will be provided later in the chapter.

2.3 Initial value problems

Prediction and simulation underpin much of structural dynamics, enabling the analysis of structures and their behaviours under a range of expected operating conditions. Moreover, they are indispensable for system identification, facilitating techniques such as parameter estimation and latent force modelling.

The simulation and analysis of dynamic systems are often rooted in the concept of Initial Value Problems (IVPs) [93]. An IVP involves solving an ODE with specified initial conditions, providing a framework for predicting the evolution of a system's state over time. Mathematically, an IVP can be expressed as:

$$\dot{\mathbf{v}} = f(t, \mathbf{v}(t)), \quad \mathbf{v}(t_0) = \mathbf{v}_0. \tag{2.12}$$

Here, f represents a vector-valued function that defines the system's dynamics, $\mathbf{v}(t) \in \mathbb{R}^d$ is the solution vector and $\mathbf{v}_0$ denotes the initial state at time t_0. Engineers often extend this framework to dynamic systems under forced excitation. The SSM transition function Eq.(2.10) can be used to accommodate such scenarios when an initial state vector is known $\mathbf{x}(t_0) = \mathbf{x}_0$. This continuous-time equation allows for

the exact evaluation of the derivatives of the state vector at any instance in time. Therefore, the solution to the states at any time t is the integral of $f(\mathbf{x}(t), \mathbf{u}(t))$ between t_0 and t:

$$\mathbf{x}(t) = \int_{t_0}^{t} f(\mathbf{x}(t), \mathbf{u}(t))\, dt, \tag{2.13}$$

where, dt is the change in time such that $dt = t - t_0$

When the underlying dynamic system is linear and time-invariant and under free vibration such that $\mathbf{u}(t) = 0 \quad \forall t$, the solution to this integral is available in closed form and can be calculated as a matrix exponential:

$$x(t) = e^{Adt} x_0, \tag{2.14}$$

where e^{Adt} is the matrix exponential. This solution provides a direct method to determine the state of the system at any time t. This framework is extended to the forced case as,

$$x(t) = e^{Adt} x_0 + \int_{t_0}^{t} e^{A(dt)} Bu(t)\, dt, \tag{2.15}$$

here, the integral term accounts for the contribution of the input over the interval from t_0 to t, and may require numerical approximation.

Moreover, for almost all nonlinear systems, a closed-form solution to the integral in Equation (2.13) is not available. Solving such IVPs requires numerical methods, as analytical solutions are typically infeasible.

2.3.1 Numerical methods for solving IVPs

Many numerical methods have been developed to approximate solutions to IVPs when no closed-form solution is available [94]. The foundation of many methods is the Taylor series [95, 96], an infinite sum expressed as:

$$\mathbf{x}(t+h) = \mathbf{x}(t) + h\frac{d\mathbf{x}(t)}{dt} + \frac{h^2}{2!}\frac{d^2\mathbf{x}(t)}{dt^2} + \frac{h^3}{3!}\frac{d^3\mathbf{x}(t)}{dt^3} + \cdots + \frac{h^n}{n!}\frac{d^n\mathbf{x}(t)}{dt^n} \tag{2.16}$$

This expansion is fundamental to understanding how numerical methods progress

the solution of an IVP from time t to $t + h$ and even forms the bias for the matrix exponential $e^{\mathbf{A}h}$ used to solve linear time-invariant systems.

Euler's method

Euler's method [97, 98] is one of the simplest numerical methods for solving IVPs. The relationship between Euler's method and the Taylor series is direct: Euler's method is derived by truncating the Taylor series after the first derivative term,

$$x(t + h) \approx x(t) + h \cdot f(x(t), u(t)). \tag{2.17}$$

Euler's method has local error proportional to h^2 and global error proportional to h. As h approaches zero, the solution of Euler's method approaches the true solution of the integral.

Euler's method can also be seen as the simplest form of the Runge-Kutta family of methods, specifically as the first-order Runge-Kutta method.

General Runge-Kutta methods

While Euler's method is a first-order Runge-Kutta method, higher-order Runge-Kutta [99–102] methods improve accuracy by considering additional evaluations of the function $\mathbf{f}(t, \mathbf{x}(t))$ at intermediate points within each time step. Perhaps the most commonly used is the fourth-order Runge-Kutta method (RK4),

$$\begin{aligned}
k_1 &= h \cdot \mathbf{f}(t, \mathbf{x}(t)), \\
k_2 &= h \cdot \mathbf{f}(t + \frac{h}{2}, \mathbf{x}(t) + \frac{k_1}{2}), \\
k_3 &= h \cdot \mathbf{f}(t + \frac{h}{2}, \mathbf{x}(t) + \frac{k_2}{2}), \\
k_4 &= h \cdot \mathbf{f}(t + h, \mathbf{x}(t) + k_3), \\
\mathbf{x}(t + h) &\approx \mathbf{x}(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).
\end{aligned} \tag{2.18}$$

Limitations of numerical integration methods

Although numerical integration methods provide powerful tools for solving IVPs, they are not without limitations. Key challenges include:

- **Stability**: Numerical methods can be unstable, especially for stiff equations, where small time steps are required to maintain accuracy.
- **Computational Cost**: Higher-order methods and smaller time steps increase computational cost.
- **Error Accumulation**: Truncation errors accumulate over time, potentially leading to significant deviations from the true solution.

2.4 Probability

Uncertainty is an intrinsic aspect of structural dynamics that profoundly impacts the design, construction, and maintenance of structures. Uncertainty arises from multiple sources, each contributing to the unpredictability encountered in engineering practice.

Uncertainty in data acquisition is a key challenge, as measurement noise can obscure the true state of a system, and sparse or indirect measurements can provide incomplete information about a structure's behaviour. In almost all practical scenarios, the full state of a system is not directly observable, further complicating efforts to accurately capture the dynamics of a structure.

In addition, the variability inherent in physical processes contributes to uncertainty. Variations in material properties, the effects of construction methods, and interactions between different structural components introduce unpredictability in the actual behaviour of structures. These factors lead to discrepancies between the predicted outcomes of the models and the observed performance of the structure in practice.

The process of representing physical systems through mathematical models also introduces uncertainty. Simplifications, assumptions, and approximations inherent in the modelling process can lead to errors in predictions, affecting the reliability of

these models. As a result, there is often a gap between the theoretical behaviour predicted by the models and the actual performance of the structure.

Numerical uncertainty in simulations further complicates the predictive accuracy of models. This type of uncertainty arises from the discretisation of continuous processes, the numerical methods employed, and the finite precision of computations. Numerical errors can propagate through simulations, leading to deviations between simulated results and real-world behavior, particularly in complex or nonlinear systems.

To account for these uncertainties, engineers have traditionally employed conservative design and maintenance principles, which incorporate substantial safety factors. Although this approach increases safety margins, it often comes at the cost of efficiency and leads to over-engineered structures with higher material costs and greater environmental impacts.

Bayesian probability offers a powerful framework for addressing uncertainty in structural dynamics. By systematically updating the likelihood of hypotheses about structural behaviour using prior knowledge and empirical data, Bayesian methods enable a more refined and data-driven approach to uncertainty management. Rooted in Bayes' theorem [103, 104], this approach allows for the continuous revision of probabilities as new data becomes available, providing a robust method for reasoning under uncertainty.

2.4.1 Bayes' Theorem

At the core of Bayesian methods lies Bayes' theorem, which states:

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)} \tag{2.19}$$

where $p(A \mid B)$ is the posterior probability of event A given event B, $p(B \mid A)$ is the likelihood of event B given event A, $p(A)$ is the prior probability of event A, and $p(B)$ is the marginal likelihood of event B.

The prior probability, $p(A)$, represents the initial belief about the event A before observing any data. It enables the incorporation of existing knowledge or expert opinions into the analysis, providing a more comprehensive understanding of the

problem when data are limited in quality or in coverage over the domain. The likelihood, $p(B \mid A)$, quantifies the probability of observing the data given that event A is true and facilitates updating of probabilities as new evidence is acquired. The marginal likelihood, also known as the model evidence, $p(B)$, serves as a normalisation constant that ensures that the probabilities sum to one. It is given by the integral,

$$p(B) = \int p(B \mid A)p(A)\, dA. \tag{2.20}$$

The posterior probability, $p(A \mid B)$, is computed by combining the prior and the likelihood, reflecting an updated belief about the event A after incorporating the evidence B. This iterative process of updating beliefs is fundamental to Bayesian inference, enabling continuous refinement as new data become available.

Evaluating $p(A \mid B)$ however, can be challenging primarily due to the difficulties associated with calculating the marginal likelihood $p(B)$. One of the main obstacles in applying Bayes' theorem is the computation of $p(B)$ (Eq. (2.20)). This calculation involves integrating the likelihood over the entire prior distribution, which makes the result sensitive to regions where the likelihood is low but the prior distribution has significant support. Consequently, accurately estimating $p(B)$ becomes particularly difficult in complex, high-dimensional spaces. In many cases, this integral is analytically intractable, necessitating the use of approximation techniques such as Monte Carlo integration to evaluate the posterior $p(A \mid B)$.

2.4.2 Monte Carlo

Monte Carlo methods encompass a broad class of computational algorithms that rely on repeated random sampling to solve numerical integration problems.

Monte Carlo integration is a specific application of Monte Carlo methods aimed at estimating the expected value of some function of a random variable. Formally, the expected value $\mathbb{E}[f(X)]$ of a function $f(X)$ with respect to a probability distribution $p(x)$ can be expressed as:

$$\mathbb{E}[f(X)] = \int f(x)p(x)\, dx \tag{2.21}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, and $p(x)$ is the target distribution of X.

In low-dimensional spaces, numerical integration techniques can be employed to compute the integral. This process involves discretising the interval into n points. At each point on the grid, the function $f(x)$ and the probability density $p(x)$ are evaluated. The products $f(x_i)p(x_i)$ are then summed and weighted by the grid spacing, which approximates the expected value $\mathbb{E}[f(X)]$. However, these techniques do not scale well to higher dimensions. An alternative approach is to draw multiple random samples, $x_n \sim p(x)$, and then compute,

$$\mathbb{E}[f(x)] \approx \frac{1}{N_s} \sum_{n=1}^{N_s} f(x_n). \tag{2.22}$$

This technique is referred to as Monte Carlo integration. It offers an advantage over numerical integration methods by concentrating function evaluations in regions with significant probability, thereby avoiding the necessity of uniformly covering the entire space. The error in Monte Carlo integration decreases as $O(1/\sqrt{N_s})$ and is theoretically independent of the dimensionality of x, based solely on the number of samples N_s.

However, a primary challenge in Monte Carlo integration, particularly in high-dimensional spaces, is the difficulty of obtaining representative samples from the target distribution $p(x)$. In high-dimensional spaces, the volume of the distribution may be concentrated in small regions. This concentration makes it difficult to efficiently sample from the areas that contribute significantly to the integral. As a result, despite the method's theoretical robustness, the accuracy and reliability of the estimation can be adversely affected in practice when sampling is not sufficiently representative of the target distribution.

2.4.3 Importance sampling

Here $\pi(x)$ will be considered to be the target distribution. This in the context of Bayesian methods this if often a conditional distribution or posterior distribution of the form $\pi(x) = p(x \mid y)$.

The fundamental idea of importance sampling is to sample from a different distribution, known as the proposal distribution $q(x)$, which is more convenient to sample

from than the target $\pi(x)$ and provides a better representation of the $\pi(x)$ in regions where the integrand has significant contributions. By choosing a proposal distribution $q(x)$ that closely resembles $\pi(x)$ or places more weight on regions where $f(x)$ is large, the variance of the estimator can be significantly reduced. However, it is crucial to select an appropriate proposal distribution. The poor choice of $q(x)$ can lead to increased variance and inefficient sampling.

Let $q(x)$ be the proposal distribution from which the samples are drawn. The expected value $\mathbb{E}[f(X)]$ can be reformulated as:

$$\mathbb{E}[f(X)] = \int f(x)\pi(x)\,dx = \int f(x)\frac{\pi(x)}{q(x)}q(x)\,dx \tag{2.23}$$

Here, $q(x)$ must be chosen such that $q(x) > 0$ whenever $f(x)\pi(x) \neq 0$. This reformulation allows the use of samples drawn from the proposal distribution $q(x)$, transforming the problem into an estimation of the expected value under $q(x)$.

Using N_s samples $x_n \sim q(x)$, the expected value $\mathbb{E}[f(X)]$ can be approximated by:

$$\mathbb{E}[f(X)] \approx \frac{1}{N_s}\sum_{n=1}^{N_s} f(x_n)\frac{\pi(x_n)}{q(x_n)} = \frac{1}{N_s}\sum_{i=n}^{N_s} w_n f(x_n). \tag{2.24}$$

where the importance weights,

$$w_n = \frac{\pi(x_n)}{q(x_n)} \tag{2.25}$$

adjust the contributions of the proposal samples $x_n \sim q(x)$, ensuring that they adequately represent the target distribution $\pi(x)$.

Self-normalised importance sampling

In situations where the normalisation constant of the target distribution is unknown, self-normalised importance sampling provides a practical solution. Consider a target distribution $\pi(x)$ known only up to a normalisation constant Z, such that:

$$\pi(x) = \frac{\tilde{\pi}(x)}{Z}, \tag{2.26}$$

where $\tilde{\pi}(x)$ is the unnormalised target distribution and $Z = \int \tilde{\pi}(x)\, dx$ is the unknown normalisation constant.

The unnormalised weights $\tilde{w_n}$ can be normalised by ensuring that the N_s samples sum to 1:

$$\hat{w_n} = \frac{\tilde{w_n}}{\sum_{n=1}^{N_s} \tilde{w_n}}, \tag{2.27}$$

The normalized weights $\hat{w_n}$ provide an estimate that approximates the true weights w_n/Z. The expectation of a function $f(x)$ under the target distribution $p(x)$ can then estimated by:

$$\mathbb{E}[f(x)] \approx \sum_{n=1}^{N_s} \hat{w_n} f(x_n). \tag{2.28}$$

2.5 Stochastic differential equations

This chapter has thus far introduced ODEs as a mathematical tool to describe the temporal evolution of dynamic systems. Additionally, it has been discussed that the dynamics of systems and their observations are inherently uncertain. To account for this uncertainty in the characterisation of dynamic systems, a mathematical framework is required to describe the stochastic temporal evolution of these systems. In this section, stochastic differential equations (SDE) [105, 106] are introduced.

SDEs are a class of differential equations that incorporate random processes, allowing for the modelling of systems subject to inherent uncertainties and random fluctuations. Unlike ODEs, which provide a deterministic description of system dynamics, SDEs integrate stochasticity through the inclusion of terms that represent random perturbations. The general form of an SDE can be expressed as:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{L}(\mathbf{x}(t), t)\mathbf{w}(t), \tag{2.29}$$

where $\mathbf{x}(t)$ represents the state vector of the system at time t, while $\mathbf{f}(\mathbf{x}(t), t)$ is a deterministic function that describes the drift of the system, analogous to the deterministic component in ordinary differential equations. The function $\mathbf{L}(\mathbf{x}(t), t)$ modulates the intensity of the stochastic effects, and $\mathbf{w}(t)$ is a zero mean white noise

process, representing the stochastic component of the system.

The solution of an SDE provides a probabilistic description of the state of the system over time. The study of SDEs is essential in fields such as finance [105, 107, 108], physics [109], biology [110], and engineering [111, 112].

2.5.1 Probabilistic state-space models

Probabilistic State-Space Models (PSSMs) describe dynamical systems where the true states are not directly observable – also referred to as *hidden* states – and only noisy, indirect measurements are available. In continuous time, PSSMs can be represented by SDEs to model the underlying state dynamics, while the observation process is typically modeled probabilistically to account for measurement noise. The probabilistic nature of these models allows for the incorporation of uncertainty in both the state evolution and the observations. In the case of a linear time-invariant system, a continuous-time PSSM consists of the following equations:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{q}(t) \tag{2.30a}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) + \mathbf{v}(t) \tag{2.30b}$$

and for a nonlinear system:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{q}(t)) \tag{2.31a}$$

$$\mathbf{y}(t) = g(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \tag{2.31b}$$

where, $\mathbf{q}(t)$ represents the process noise, which accounts for the uncertainty in the evolution of the system's state due to unmodeled dynamics or external disturbances, while $\mathbf{v}(t)$ represents the measurement noise, which characterises inaccuracies or uncertainties in observed data due to sensor errors or environmental influences.

From this point onward, all SSMs discussed will be probabilistic, so the term 'SSM' will be used to refer to PSSM.

2.6 Bayesian filtering and smoothing

Bayesian filtering and smoothing are key techniques for estimating the state of systems modelled by SSMs. These methods use a Bayesian framework to update prior distributions over the state vector $\mathbf{x}(t)$ evaluated from an uncertain state transfer function (2.31a) and refine the state estimates based on noisy observations $y(t)$ through a uncertain measurement model (2.31b) , incorporating both the uncertain dynamics of the system and the uncertainty inherent in the measurements into an optimal estimate of state[2].

In a Bayesian setting, the SSM equations (2.31) can be equivalently considered in discrete time as drawing samples from probability distributions:

$$\mathbf{x}_0 \sim p(\mathbf{x}_0), \tag{2.32a}$$

$$\mathbf{x}_t \sim p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \tag{2.32b}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{u}_t). \tag{2.32c}$$

where $\mathbf{x}_t \triangleq \mathbf{x}(t)$. The prior distribution $p(\mathbf{x}_0)$ encapsulates initial knowledge of the state before any measurements. The transition distribution $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ models the evolution of the state from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ while accounting for system uncertainties. The measurement model $p(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{u}_t)$ represents how the observed data $\mathbf{y}_t$ manifests given the underlying state $\mathbf{x}_t$. The SSM is shown graphically in Figure 2.6.

SSMs are typically used for three primary tasks: prediction (or simulation), filtering, and smoothing. These tasks involve estimating the distributions of the hidden states $\mathbf{x}_{1:t}$ using different amounts of information about the observations $\mathbf{y}$. The objective of prediction is to estimate the distribution of future states, in the absence of measurements $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{u}_{1:t})$. Filtering focusses on determining the state distribution based on observations up to the present time, represented by $p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \mathbf{u}_{1:t})$. Smoothing, on the other hand, involves estimating the state distributions by using the complete sequence of observations $p(\mathbf{x}_t \mid \mathbf{y}_{1:T}, \mathbf{u}_{1:T})$, where T is the final time step.

When the transition and observation distributions are linear functions with addi-

[2]optimal for linear systems that can be solved with a Kalman filter

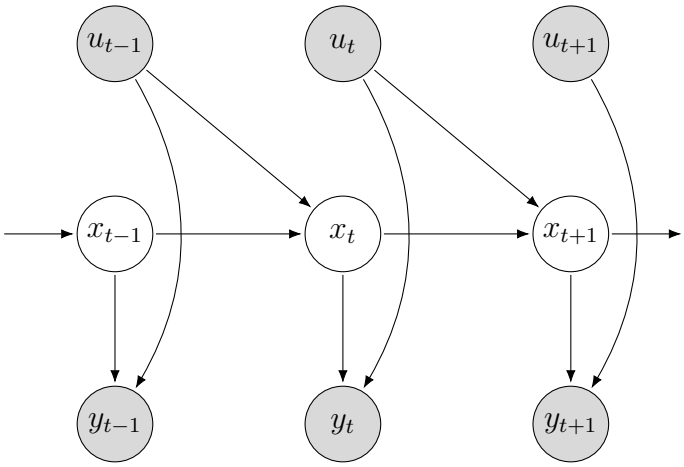

Figure 2.3: A graphical illustration of the dependencies in a state-space model. The latent states x_t evolve over time, influenced by the previous state x_{t-1} and the current input u_t. The observations y_t are generated from the current state x_t and potentially the current input u_t.

tive Gaussian noise, solutions for the filtering and smoothing distributions can be found in closed form. These solutions are provided by the well-known Kalman filter [113] and the Rauch-Tung-Striebel (RTS) smoother [114] respectively. However, for nonlinear systems or when the noise is not additive Gaussian, the filtering and smoothing distributions become analytically intractable.

To handle nonlinear SSMs, various approximation techniques have been proposed. Notable among these are the Extended Kalman Filter (EKF) [23] and the Unscented Kalman Filter (UKF) [115]. These methods, along with others [23], approximate the nonlinear system by presuming a Gaussian distribution at each time step. Although these techniques can be effective, they may struggle with accurately representing complex distributions, such as multimodal distributions, in the states. An alternative method involves using importance sampling, as seen in the Particle Filter (PF) [116]. This section will introduce these methods, but for further details the excellent texts by Sarkka [23] and by Doucet and Johansen [116] are recommended.

2.6.1 Kalman filter

The Kalman Filter (KF) is a widely used algorithm for estimating the state of a linear dynamic system from a series of noisy measurements. It provides an optimal

solution for linear Gaussian state-space models. The KF consists of two main steps: prediction and update.

Prediction step:

$$\mathbf{x}_t^- = \mathbf{A}_{t-1}\mathbf{x}_{t-1}, \tag{2.33a}$$

$$\mathbf{P}_t^- = \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}^\top + \mathbf{\Sigma}_{t-1}. \tag{2.33b}$$

In the prediction step, the state estimate $\mathbf{x}_t^-$ and the covariance estimate $\mathbf{P}_t^-$ are calculated based on the previous state $\mathbf{x}_{t-1}$ and covariance $\mathbf{P}_{t-1}$, the discrete state transition matrix $\mathbf{A}_{t-1}$, and the discrete process noise covariance $\mathbf{\Sigma}_{t-1}$. The prior prediction is updated via:

Update step:

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^-, \tag{2.34a}$$

$$\mathbf{S}_t = \mathbf{H}_t\mathbf{P}_t^-\mathbf{H}_t^\top + \mathbf{R}_t, \tag{2.34b}$$

$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{H}_t^\top\mathbf{S}_t^{-1}, \tag{2.34c}$$

$$\mathbf{x}_t = \mathbf{x}_t^- + \mathbf{K}_t\mathbf{v}_t, \tag{2.34d}$$

$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t\mathbf{S}_t\mathbf{K}_t^\top. \tag{2.34e}$$

The update step involves measurement $\mathbf{y}_t$, measurement matrix $\mathbf{H}_t$, and measurement noise covariance $\mathbf{R}_t$. The Kalman gain $\mathbf{K}_t$ is calculated, which is then used to update the state estimate $\mathbf{x}_t$ and the covariance estimate $\mathbf{P}_t$ to incorporate the observation.

2.6.2 Extended Kalman filter

The Extended Kalman Filter (EKF) extends the KF to nonlinear systems by linearising the state and observation models around the current estimate. The first-order EKF approximates the nonlinear functions using a first-order Taylor expansion.

Prediction step:

$$\mathbf{x}_t^- = \mathbf{f}(\mathbf{x}_{t-1}), \tag{2.35a}$$
$$\mathbf{P}_t^- = \mathbf{F}_\mathbf{x}(\mathbf{x}_{t-1})\mathbf{P}_{t-1}\mathbf{F}_\mathbf{x}^\top(\mathbf{x}_{t-1}) + \mathbf{\Sigma}_{t-1}. \tag{2.35b}$$

Update step:

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{h}(\mathbf{x}_t^-)\mathbf{x}_t^-, \tag{2.36a}$$
$$\mathbf{S}_t = \mathbf{H}_\mathbf{x}(\mathbf{x}_t^-)\mathbf{P}_t^-\mathbf{H}_\mathbf{x}^\top(\mathbf{x}_t^-) + \mathbf{R}_t, \tag{2.36b}$$
$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{H}_\mathbf{x}^\top(\mathbf{x}_t^-)\mathbf{S}_t^{-1}, \tag{2.36c}$$
$$\mathbf{x}_t = \mathbf{x}_t^- + \mathbf{K}_t\mathbf{v}_t, \tag{2.36d}$$
$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t\mathbf{S}_t\mathbf{K}_t^\top. \tag{2.36e}$$

where $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$ are the nonlinear state transition and observation functions, respectively, and $\mathbf{F}_\mathbf{x}$ and $\mathbf{H}_\mathbf{x}$ are the Jacobians of $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$, respectively.

2.6.3 Particle filter

The Particle Filter (PF) is an importance sampling-derived Sequential Monte Carlo (SMC) method used for estimating the state of nonlinear and non-Gaussian dynamic systems. Unlike filtering methods that rely on linearisation or Gaussian approximations, the PF approximates the posterior distribution of the state by using a set of weighted samples, also known as particles. A general framework for the PF is discussed in four steps:

1. Sample: $\mathbf{x}_t^{(i)} \sim q(\mathbf{x}_t)$

The target is to evaluate the filtering distribution $\pi(\mathbf{x}_t) \triangleq p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \mathbf{u}_{1:t})$. However, direct sampling from this distribution is not feasible. Instead, a set of N particles $\{\mathbf{x}_t^{(i)}\}_{i=1}^N$ is drawn from a proposal distribution $q(\mathbf{x}_t)$, according to the importance sampling method. Ideally, the proposal distribution should closely approximate the target distribution. When this is not easily achievable, samples can be drawn from the prior distribution (state transition model) $\mathbf{x}_t^{(i)} \sim p(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)}, \mathbf{u}_{t-1})$, as it ensures that $q(\mathbf{x}_t) > 0$ whenever $p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \mathbf{u}_{1:t}) \neq 0$. This approach is known as the

Bootstrap Particle Filter.

2. Weight: $\widetilde{w}^{(i)} = \widetilde{\pi}(\mathbf{x}_t^{(i)}) / q(\mathbf{x}_t^{(i)})$

After propagation, each particle $\mathbf{x}_t^{(i)}$ is assigned an unnormalised weight $\widetilde{w}^{(i)}$ based on how well it explains the observation $\mathbf{y}_t$ at time t. The weight is typically proportional to the likelihood $p(\mathbf{y}_t \mid \mathbf{x}_t^{(i)}, \mathbf{u}_{1:t})$.

3. Normalise weight: $\hat{w}^{(i)} = \widetilde{w}^{(i)} / \sum_{j=1}^{N} \widetilde{w}^{(j)}$

The weights are then normalised as per Eq.(2.27).

4. Resampling

The Particle Filter is a sequential method in which the same set of particles is propagated through time using a state transition model. As the algorithm progresses through the prediction and reweighting steps (by repeating steps 1 to 3), it gathers more information about the quality of the particles $\{\mathbf{x}_t^{(i)}\}_{i=1}^{N}$. This process often leads to particle degeneracy, where most particles have negligible weights, and only a few carry the majority of the weight. To address this imbalance, resampling is implemented. Resampling duplicates particles with higher weights and discards those with lower weights, ensuring that the particle set better approximates the target distribution, independent of the initial weights. After resampling, the weights $\{\hat{w}_t^{(i)}\}_{i=1}^{N}$ are uniformly reset to $1/N$. There after, the particles are propagated forward in time by repleating steps 1 through 3 untill resmapling is need again.

Algorithm

These steps are performed sequentially from $t = 0$ to $t = T$ shown in Algorithm 1.

Algorithm 1 Particle Filter (for $i = 1, \ldots, N$)

1: **Initialization** ($t = 0$):

(a) Sample $\mathbf{x}_0^{(i)} \sim q(\mathbf{x}_0)$.

(b) Set initial weights: $\hat{w}_0^{(i)} = \frac{1}{N}$.

2: **for** $t = 1$ **to** T **do**

(a) **Resample**

(b) **Propagate:** Sample $\mathbf{x}_t^{(i)} \sim p(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{a_t^i})$.

(c) **Weight:** Compute $\widetilde{w}_t^{(i)} = p(\mathbf{y}_t \mid \mathbf{x}_t^{(i)})$ and normalize:

$$\hat{w}_t^{(i)} = \widetilde{w}_t^{(i)} \Big/ \sum_{j=1}^{N} \widetilde{w}_t^{(j)}.$$

2.6.4 RTS smoothing

Rauch-Tung-Striebel (RTS) smoothing is an algorithm for computing the smoothed state estimates in linear Gaussian state-space models. It complements the Kalman filter by performing a backward recursion to refine the state estimates.

Backward recursion:

$$\mathbf{x}_{t+1}^{-} = \mathbf{A}_t \mathbf{x}_t, \tag{2.37a}$$

$$\mathbf{P}_{t+1}^{-} = \mathbf{A}_t \mathbf{P}_t \mathbf{A}_t^{\top} + \mathbf{\Sigma}_t, \tag{2.37b}$$

$$\mathbf{G}_t = \mathbf{P}_t \mathbf{A}_t^{\top} \left(\mathbf{P}_{t+1}^{-}\right)^{-1}, \tag{2.37c}$$

$$\mathbf{x}_t^s = \mathbf{x}_t + \mathbf{G}_t \left(\mathbf{x}_{t+1}^s - \mathbf{x}_{t+1}^{-}\right), \tag{2.37d}$$

$$\mathbf{P}_t^s = \mathbf{P}_t + \mathbf{G}_t \left(\mathbf{P}_{t+1}^s - \mathbf{P}_{t+1}^{-}\right) \mathbf{G}_t^{\top}. \tag{2.37e}$$

where $\mathbf{x}_t^s$ and $\mathbf{P}_t^s$ are the smoothed state estimate and covariance, respectively.

2.6.5 Extended RTS smoothing

The Extended RTS Smoother (ERTS) extends RTS smoothing to nonlinear systems by linearising the state and observation models. It uses the EKF for the forward pass and an extended version of the RTS smoother for the backward pass.

Backward recursion:

$$\mathbf{x}_{t+1}^{-} = \mathbf{f}(\mathbf{x}_t), \tag{2.38a}$$

$$\mathbf{P}_{t+1}^{-} = \mathbf{F}_{\mathbf{x}}(\mathbf{x}_t)\mathbf{P}_t\mathbf{F}_{\mathbf{x}}^{\top}(\mathbf{x}_t) + \mathbf{Q}_t, \tag{2.38b}$$

$$\mathbf{G}_t = \mathbf{P}_t\mathbf{F}_{\mathbf{x}}^{\top}(\mathbf{x}_t)\left(\mathbf{P}_{t+1}^{-}\right)^{-1}, \tag{2.38c}$$

$$\mathbf{x}_t^s = \mathbf{x}_t + \mathbf{G}_t\left(\mathbf{x}_{t+1}^s - \mathbf{x}_{t+1}^{-}\right), \tag{2.38d}$$

$$\mathbf{P}_t^s = \mathbf{P}_t + \mathbf{G}_t\left(\mathbf{P}_{t+1}^s - \mathbf{P}_{t+1}^{-}\right)\mathbf{G}_t^{\top}. \tag{2.38e}$$

where $\mathbf{F}_{\mathbf{x}}(\mathbf{x}_t)$ is the Jacobian of the state transition function $\mathbf{f}(\cdot)$ evaluated at $\mathbf{x}_t^{-}$.

2.7 Gaussian processes

Throughout this thesis, GPs will be employed predominantly in the equivalent state-space/SDE formulation. However, before introducing this formulation, it may be useful to cover some of the foundations of GPs.

GPs serve as a fundamental approach in nonparametric Bayesian regression, enabling the encapsulation of uncertainty in input-output mappings through a probabilistic lens [66]. A GP, denoted as $p(f)$ for a prior over functions f, transitions to a posterior $p(f|D)$ upon incorporating data D, via Bayes' rule. This framework posits that each data point, given a specific input, emerges as a sample from a univariate Gaussian conditioned on that input. Mathematically, a GP is fully specified by its mean function $m(x)$ and the covariance kernel $k(x, x')$,

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')), \tag{2.39}$$

where,

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{2.40}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))], \tag{2.41}$$

such that $\mathbb{E}[\cdot]$ denotes the expectation operator, $f(\mathbf{x})$ is a real-valued function, $\mathbf{x}$ and $\mathbf{x}'$ are input vectors, and $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ are the mean and covariance functions, respectively.

In this work, as is common, it will be assumed that the prior mean function is zero, leaving the covariance function to encode the assumptions that are to be embedded *a priori*.

The choice of covariance function, or kernel, is crucial, as it directly influences how closely the prior aligns with the posterior distribution. By selecting a specific kernel, practitioners can incorporate assumptions about the smoothness, periodicity, and other characteristics of the function of interest. Commonly used kernels include the Radial Basis Function (RBF), the Matérn, and the exponential kernels, each reflecting different prior beliefs.

One popular set of covariance kernel is the Matern class of covariance kernels given,

$$k_{\text{Matérn}}(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{||\mathbf{x} - \mathbf{x}'||}{\ell} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{||\mathbf{x} - \mathbf{x}'||}{\ell} \right), \tag{2.42}$$

where $\Gamma(\cdot)$ is the gamma function, $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind, $||\mathbf{x} - \mathbf{x}'||$ is the Euclidean distance between input vectors $\mathbf{x}$ and $\mathbf{x}'$, ℓ and ν are non-negative parameters. More intuitively ν can be considered to encode the prior "roughness" of the GP and ℓ is the length scale which sets the "distance" over which the function values are correlated. For $\nu = 0.5$, the Matérn kernel reduces to the exponential kernel, and as $\nu \to \infty$, it approaches the RBF kernel.

2.8 Evaluation metrics

In this thesis, two evaluation metrics are employed to assess the performance and accuracy of the proposed models: Root Mean Square Error (RMSE) and Normalised Mean Square Error (NMSE). This section will detail these methods for later use.

2.8.1 Root mean square error

The RMSE is a commonly used metric to evaluate the accuracy of model predictions. The RMSE provides an absolute measure of the difference between predicted and observed values, with lower values indicating better model performance. Mathematically, the RMSE is expressed as:

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2} \tag{2.43}$$

Here, y_i represents the observed value at the i-th point, $\hat{y}_i$ represents the predicted value at the i-th point, and M is the total number of observations. The RMSE is particularly sensitive to large errors, as these are squared during calculation, thus providing a strong indication of the presence of outliers or significant discrepancies between the model and the observed data.

2.8.2 Normalized mean square error

The NMSE is a evaluation metric that accounts for the scale of the observed data by normalising the error. This normalisation facilitates comparison across different datasets or models by expressing the error relative to the variance of the observed data. The NMSE is defined as:

$$NMSE = \frac{100}{M\sigma_y^2}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 \tag{2.44}$$

In this context, σ_y^2 denotes the variance of the observed data y. The NMSE metric ranges from zero, representing a perfect model fit, and increases as the fit quality declines. Anecdotally, the classification of NMSE performance can be considered as:

- **Excellent**: NMSE $< 1\%$
- **Very Good**: $1\% \leq$ NMSE $< 5\%$
- **Good**: $5\% \leq$ NMSE $< 10\%$

Chapter 3

Stationary latent force models

The latent force model (LFM) is a framework that incorporates latent variables – unobserved or hidden variables – to identify unmodelled forces from observed data. There are two types of latent force model. The first is the latent input force (LIF) model, which is essential for recovering unobserved inputs to a system, facilitating damage assessment, fatigue prediction, and robust design by offering insight into loading conditions. The second is the latent restoring force (LRF) model that infers often nonlinear, internal mechanisms that restore equilibrium, crucial for accurately capturing the dynamics of a system.

One model type that has gained increasing recognition within the structural dynamics community for its application in both LIF and LRF identification is the Gaussian process latent force model (GPLFM). The GPLFM is particularly valued for its ability to recover not only the latent forces but also to provide a posterior distribution over these forces, accounting for the uncertainty inherent in the identification process. The GPLFM has been explored in the literature, with numerous authors highlighting its effectiveness as a robust framework. However, several significant challenges remain unresolved.

The aim of this chapter is to provide a comprehensive introduction to the GPLFM for joint input-state identification. This will include an overview of the existing literature on the use of GPLFMs for LIF identification, as well as a derivation of the state-space GP from the mean and covariance function of a stationary temporal GP. Moreover, the development of an established formulation of a state-space

GPLFM from a state-space GP will be provided. Finally, an example study will be presented to illustrate the application of the GPLFM, along with a discussion on how the selection of observed state variables influences the performance of the GPLFM in joint input-state estimation. This work will serve as a foundation for understanding the fundamentals of the GPLFM and will lay the groundwork for the new advancements introduced later in subsequent chapters: Chapters 4 to 6.

3.1 Related work

Alvarez et al. [71] first introduced the use of a GPLFM as a method to identify unknown input forces in linear dynamical systems, using a zero-mean temporal GP with a stationary kernel by formulating the problem,

$$m\ddot{x} + c\dot{x} + kx = U(t), \tag{3.1a}$$

$$U(t) \sim \mathcal{GP}(0, k(t, t')). \tag{3.1b}$$

and was further extended by Alvarez et al. to encompass Multi-Degree-of-Freedom (MDOF) systems. A critical challenge highlighted within their approach was the computational intensity required to solve the model.

GPs are powerful tools for modelling and predicting time-series, due to their flexibility and ability to quantify uncertainty. However, their practical application is hindered by significant computational challenges, particularly the $\mathcal{O}(N^3D^3)$ complexity, where N is the number of data points and D represents the dimensionality or degrees of freedom of the system. This high computational cost can be prohibitive when dealing with large time series datasets such as found in structural dynamics and latent force modelling.

To mitigate this computational burden, sparse approximations of GPs have been developed [117, 118]. Alvarez et al [119] demonstrates that the computational load of a multiple-output GP can be significantly reduced by implementing a partially independent training conditional (PITC) sparse approximation. This method reduces the computational complexity from $\mathcal{O}(N^3D^3)$ to $\mathcal{O}(N^3D)$, making it more feasible for high-dimensional applications. However, even with this reduction, the compu-

tational demands can remain substantial, especially in cases where the number of data points vastly exceeds the degrees of freedom.

A significant advancement in the efficient computation of GPs with temporal inputs was introduced by Hartikainen and Särkkä [72]. They proposed that GPs with stationary covariance functions can be equivalently expressed as linear time-invariant (LTI) stochastic differential equations (SDEs). This formulation enables efficient computation of the posterior via the Kalman filter and RTS smoother. The smoother provides estimates of the states conditioned on all past and futures observations ($p(x_t|y_{1:T})$), which are necessary to replicate the full posterior inference of a standard batch GP regression.

The state-space representation of GPs not only maintains the expressive power and uncertainty quantification inherent in GPs but also dramatically reduces the computational complexity to $\mathcal{O}(ND^3)$. This reduction is especially beneficial when the number of data points, N, is large relative to the state dimension, D. As a result, this approach provides a practical and efficient means of applying GPs to large-scale time-series datasets, making it particularly well-suited for applications in latent force modelling.

This state-space formulation forms the foundation for much of the subsequent work in GPLFMs and addresses the computational challenges associated with the GPLFM initially defined by Alvarez et al. [71]. Hartikainen and Särkkä [73] were the first to implement this approach within a latent force framework, proposing an augmented SSM that integrates a state-space GP to infer the LIF directly into the SSM governing the system dynamics.

Recent studies have extended the application of state-space GPLFMs for force identification across various mechanical systems. These investigations include the use of GPLFMs for joint input-state estimation in linear multi-degree-of-freedom (MDOF) systems, with hyperparameters optimized via maximum likelihood estimation [74], as well as their application in nonlinear systems, such as the Duffing oscillator, where Particle Gibbs and Markov Chain Monte Carlo (MCMC) methods are employed for hyperparameter tuning [75].

Further contributions have demonstrated the use of GPLFMs in joint input-state parameter identification for both linear SDOF and MDOF systems, employing MCMC to determine distributions over system and hyperparameters [77, 78].

The adoption of GPLFMs in experimental settings has been increasingly recognised, with contributions including wind load estimation for long-span suspension bridges [120], onshore wind turbines [121], and in-situ validation of predicted subsoil strain responses in offshore wind turbines [122]. This further underscores the versatility and effectiveness of the model in real-world applications.

The state-space formulation of GPs has become a foundational framework in the advancement of latent force modeling using GPLFMs. By offering a significant reduction in computational complexity—from $\mathcal{O}(N^3D^3)$ in traditional GP approaches to $\mathcal{O}(ND^3)$—it addresses the scalability issues that are critical when working with large time-series datasets prevalent in structural dynamics. This efficiency gain has been essential for the practical application and adoption of GPLFMs in recent research. Recognizing its importance, the derivation of the state-space representation of GPs will be presented in the following section.

3.2 Derivation of state-space Gaussian processes

The starting point for deriving a state-space model from a GP is a linear time-invariant SDE that governs the behavior of a continuous-time random process $f(t)$. Such a process can be described by an SDE of order m as follows:

$$\frac{d^m f(t)}{dt^m} + a_{m-1}\frac{d^{m-1} f(t)}{dt^{m-1}} + \cdots + a_1\frac{df(t)}{dt} + a_0 f(t) = w(t), \tag{3.2}$$

where the coefficients $a_0, a_1, \ldots, a_{m-1}$ are constants, and $w(t)$ represents a white noise process with spectral density $S_w(\omega) = q$. The goal is to express this SDE in a first-order state-space form.

To achieve this transformation, a state vector $\mathbf{x}(t)$ is introduced. This vector encapsulates the process $f(t)$ and its derivatives up to order $(m-1)$, and is written as:

$$\mathbf{x}(t) = \begin{bmatrix} f(t) \\ \frac{df(t)}{dt} \\ \vdots \\ \frac{d^{m-1}f(t)}{dt^{m-1}} \end{bmatrix}. \tag{3.3}$$

By constructing the state vector in this manner, the original m-th order SDE is reduced to a system of first-order differential equations,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{L}w(t), \tag{3.4}$$

where $\mathbf{F}$ is the transition matrix, and $\mathbf{L}$ is the input vector that determines how the noise $w(t)$ enters the system. The matrices are explicitly given by:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{m-1} \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.5}$$

The transition matrix $\mathbf{F}$ encodes the deterministic dynamics of the process, while the input vector $\mathbf{L}$ introduces stochasticity through the white noise input $w(t)$.

The system described by the SDE in Equation (3.2) can also be understood in the frequency domain through its transfer function, which is obtained via the Fourier transform of the system's differential equation. This transfer function, $G(i\omega)$, relates the input $w(t)$ to the output $f(t)$ in the frequency domain. The transfer function $G(i\omega)$ is expressed as:

$$G(i\omega) = \frac{b_0}{(i\omega)^m + a_{m-1}(i\omega)^{m-1} + \cdots + a_1(i\omega) + a_0}, \tag{3.6}$$

where b_0 is a constant such that $q = |b_0|^2$, q is the spectral density of the white noise process, and the denominator is derived from the characteristic polynomial of the differential operator in Equation (3.2). This transfer function can be used to

compute the spectral density $S(\omega)$ of the process $f(t)$, which is given by:

$$S(\omega) = |G(i\omega)|q|G(i\omega)|^{-1}, \tag{3.7}$$

Therefore, by using $\mathbf{H} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^\top$ it is possible to extract the first component of the state vector $\mathbf{f}(t)$ from $\mathbf{x}(t)$,

$$\mathbf{f}(t) = \mathbf{H}\mathbf{x}(t), \tag{3.8}$$

such that the spectral density of the state-space model can be calculated using the following expression:

$$S(\omega) = \mathbf{H}\left(\mathbf{F} + i\omega\mathbf{I}\right)^{-1}\mathbf{L}\,q\,\mathbf{L}^\top\left[\left(\mathbf{F} - i\omega\mathbf{I}\right)^{-1}\right]^\top\mathbf{H}^\top, \tag{3.9}$$

where $\mathbf{I}$ denotes the identity matrix.

Importantly, it is possible to reverse this approach by starting with a covariance function for a temporal GP. By determining the spectral density from this, it is possible to obtain a stable Markov process as long as the covariance of the GP is stationary.

3.2.1 Matérn kernel as an SDE

Recall the Matérn covariance function introduced in the previous chapter:

$$k(r) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,r}{\ell}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\,r}{\ell}\right), \tag{3.10}$$

In this case, the expression is restricted to the temporal domain, where $r = |t - t'|$ represents the temporal distance between two points.

The spectral density of a stationary process with a Matérn covariance can be derived by taking the Fourier transform of the covariance function. The spectral density is given by:

$$S(\omega) = \sigma^2 \frac{2\pi^{1/2}\Gamma(\nu + \frac{1}{2})}{\Gamma(\nu)} \lambda^{2\nu} \left(\lambda^2 + \omega^2\right)^{-(\nu+\frac{1}{2})}, \tag{3.11}$$

where $\lambda = \sqrt{2\nu}/\ell$. When $\nu = p + \frac{1}{2}$, where p is a non-negative integer, the spectral density simplifies to:

$$S(\omega) \propto \left(\lambda^2 + \omega^2\right)^{-(p+1)}. \tag{3.12}$$

This can also be expressed as:

$$S(\omega) \propto (\lambda + i\omega)^{-(p+1)} (\lambda - i\omega)^{-(p+1)}, \tag{3.13}$$

From this, the transfer function $G(i\omega)$ of the corresponding LTI system can be extracted as:

$$G(i\omega) = (\lambda + i\omega)^{-(p+1)}, \tag{3.14}$$

and the process can be modeled as driven by white noise with spectral density q, which is given by:

$$q = \frac{2\sigma_f^2 \pi^{1/2} \lambda^{(2p+1)} \Gamma(p+1)}{\Gamma(p + \frac{1}{2})}. \tag{3.15}$$

For instance, when $p = 1$, the corresponding state-space representation of the process is:

$$\frac{d\mathbf{x}(t)}{dt} = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix} \mathbf{x}(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t), \tag{3.16}$$

and with p = 2,

$$\frac{d\mathbf{x}(t)}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda^3 & -3\lambda^2 & -3\lambda \end{pmatrix} \mathbf{x}(t) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w(t), \tag{3.17}$$

It is then possible to solve this this continuous time state-space model with a Kalman filter and RTS-smoother by first converting the model to discrete time (details seen later in the chapter).

3.3 A framework for joint input-state estimation as a latent force problem

The fundamental concept behind the LFM for LIF recovery is that, if the equation of motion is well-characterised, the parameters are known, and all system states are measured, the total internal restoring force of the system can be calculated (e.g., the left-hand side of Eq. (3.1a)). According to Newton's laws of motion, this restoring force must be equal to the external force applied to the system (e.g., the right-hand side of Eq. (3.1a)). When measurement noise introduces uncertainty in the states, a distribution over the LIF can still be recovered, reflecting the uncertainty in the identification process.

However, for this approach to be applied without modification, it requires direct measurements of all internal states of the system. However, this is rarely feasible, even in controlled laboratory environments, and is even less likely in real-world engineering scenarios, particularly under operational conditions. Therefore, effective latent force modeling must be capable of performing reliably, even when only partial state information is available. This challenge highlights the crucial need for joint input-state estimation.

This section details the process of augmenting a continuous-time LTI SSM with a continuous-time LTI state-space GP to form a continuous-time GPLFM. The augmentation establishes the framework necessary for joint input-state estimation in latent force problems. Additionally, the discretisation of continuous-time LTI GPLFMs will be discussed. Once in this form, discrete time Bayesian filtering and smoothing techniques can be applied. Moreover, this section will explore how the specific state variable that is observed influences the performance of the GPLFM in joint input-state estimation.

3.3.1 Stochastic differential equations for latent force modelling

Previously it has been discussed how a dynamic system can be formulated as a coninuious time SSM (Section 2.5.1), such that,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \tag{3.18a}$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t)) \tag{3.18b}$$

In this work an alternative perspective to the transition function of the SSM is taken (Eq.(3.18a)). The transition function of a SSM can be equivalently considered as a special case of an SDE. This alternative view point will prove useful as this work will draw heavily upon SDE theory to develop and solve both the stationary GPLFM and, in particular, the non-stationary GPLFM in Chapter 6. A continuous-time LTI SDE equation can be expressed as:

$$\boldsymbol{\mathcal{X}}(0) \sim \mathcal{N}(\mathbf{x}(0), \mathbf{P}(0)), \tag{3.19a}$$

$$\dot{\boldsymbol{\mathcal{X}}}(t) = \mathbf{F}\boldsymbol{\mathcal{X}}(t) + \mathbf{L}w(t), \tag{3.19b}$$

where $\mathbf{x}(0)$ and $\mathbf{P}(0)$ are the mean and the covariance that describe the Gaussian distribution over the initial state $\boldsymbol{\mathcal{X}}(0)$. $\mathbf{F}$ is the continuous-time state transition matrix, $\mathbf{L}$ is the diffusion matrix and $w(t)$ is a white noise process.

In this work, the focus will be on the propagation of the statistics of states through time, that is, the mean $\mathbf{x}(t)$ and covariance $\mathbf{P}(t)$, rather than samples from the distribution specified in Eq(3.19). The linear continuous time derivatives of the mean and covariance are provided as follows,

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t), \tag{3.20a}$$

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{F}^T + \mathbf{F}\mathbf{P}(t) + \mathbf{L}\mathbf{Q}\mathbf{L}^T \tag{3.20b}$$

where $\mathbf{P}(t)$ is the state covariance and $\mathbf{Q}$ is the noise covariance or white noise spectral density.

3.3.2 Constructing the GPLFM for a linear SDOF system

The construction of the GPLFM will initially be demonstrated for a linear SDOF system,

$$m\ddot{x} + c\dot{x} + kx = U(t). \tag{3.21}$$

However, it is feasible to extend this framework to nonlinear and/or MDOF systems. The construction of the state-space GPLFM for a linear SDOF system begins with the consideration of the second-order differential equation Eq.(3.21) in state-space form with the continuous time mean and covariance defined by Eq.(3.20) and parameterised by the matrices,

$$F_{\text{dyn}} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix}, \quad B_{\text{dyn}} = \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}, \tag{3.22}$$

$$L_{\text{dyn}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad P_{0_{\text{dyn}}} = \begin{bmatrix} \sigma^2_{x(0)} & 0 \\ 0 & \sigma^2_{\dot{x}(0)} \end{bmatrix}, \quad Q_{\text{dyn}} = q \tag{3.23}$$

The subscript *dyn* indicates that the arrays define the internal dynamic states of the system. The subscripts $x(0)$ and $\dot{x}(0)$ denote the variances in displacement and velocity, respectively, at $t = 0$. The term q represents the process noise in the system dynamics.

The next step is to consider the SDE representation of the GP the practitioner has chosen for the analysis. Here, as introduced in Section 2.7 the Matérn 3/2 kernel is used. The continuous time mean and covariance are defined by the arrays,

$$F_{\text{GP}} = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{bmatrix}, \tag{3.24}$$

$$L_{\text{GP}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad P_{\infty_{\text{GP}}} = \begin{bmatrix} \sigma^2_{u(0)} & 0 \\ 0 & 2\lambda\sigma^2_{u(0)} \end{bmatrix} \quad Q_{\text{GP}} = 4\lambda^3\sigma^2_u, \tag{3.25}$$

where, subscript *GP* indicates that the arrays define the states of the GP, $\lambda = \sqrt{3}/\ell$, and ℓ and σ^2_u denote the length scale and variance hyperparameters, respectively.

By adopting the assumption from Eq. (3.1b), which posits that the input to the dynamic SDE can be modeled by a temporal GP, it becomes possible to augment the dynamic SSM Eq. (3.22) such that the latent input is treated as an internal state of the system, modelled by the SDE GP. This allows the transition of the mean states to be represented as:

$$\underbrace{\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \\ \dot{u}(t) \\ \ddot{u}(t) \end{bmatrix}}_{\dot{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k}{m} & -\frac{c}{m} & \frac{1}{m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\lambda^2 & -2\lambda \end{bmatrix}}_{F_{\text{GPLFM}}} \underbrace{\begin{bmatrix} x(t) \\ \dot{x}(t) \\ u(t) \\ \dot{u}(t) \end{bmatrix}}_{\mathbf{x}(t)} \tag{3.26}$$

Furthermore, the covariance over the states through time is captured by,

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)F^T_{\text{GPLFM}} + F_{\text{GPLFM}}\mathbf{P}(t) + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}}_{L_{\text{GPLFM}}} \underbrace{\begin{bmatrix} q & 0 \\ 0 & 4\lambda^3\sigma^2_u \end{bmatrix}}_{Q_{\text{GPLFM}}} \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}^T}_{L^T_{\text{GPLFM}}} \tag{3.27a}$$

$$P_{0_{\text{GPLFM}}} = \begin{bmatrix} \sigma^2_{x(0)} & 0 & 0 & 0 \\ 0 & \sigma^2_{\dot{x}(0)} & 0 & 0 \\ 0 & 0 & \sigma^2_{u(0)} & 0 \\ 0 & 0 & 0 & 2\lambda\sigma^2_{u(0)} \end{bmatrix} \tag{3.27b}$$

Eq. (3.26) can be visualized as a block matrix system, where the upper-left block captures the system dynamics, and the lower-right block defines the Gaussian process prior. This GP prior models the latent force, while the upper-right block defines

how the external force affects the internal states, through the term $\frac{1}{m}$, ensuring that the force is appropriately scaled when entering the dynamics via the acceleration term. The lower-left block, which is zero, ensures that the dynamic states do not affect the GP input – a reflection of causality in the loading.

The covariance structure of the system is given in Eq. (3.27). The term $P_{0_{\text{GPLFM}}}$ represents the initial covariance of the system states, with diagonal elements corresponding to the variances of displacement, velocity, and GP states. The process noise covariance, Q_{GPLFM}, reflects the uncertainty in the model in both the system dynamics and the latent force modelled by the GP.

3.3.3 Discretization for implementation with Bayesian filtering and smoothing

For practical implementation, the continuous-time state-space GPLFM derived in Eq. (3.26) and Eq. (3.27) must be discretized to allow for the application of recursive estimation techniques such as the Kalman filter and RTS smoother.

As this system is LTI, the discrete-time transition matrix $A(x_{t-1})$ is computed using the matrix exponential $\exp(F(x_{t-1})dt)$, which benefits from robust numerical methods.

In contrast, calculating the discrete-time noise covariance $\Sigma(x_{t-1})$ is more complex and numerically demanding. This complexity can be mitigated using matrix fraction decomposition (MFD)[1] [106], which simplifies the computation of $\Sigma(x_{t-1})$ to evaluating a matrix exponential:

$$\begin{bmatrix} A & \Sigma A^T \\ 0 & A^{-T} \end{bmatrix} = \exp\left(\begin{bmatrix} F & LQL^T \\ 0 & -F^T \end{bmatrix} dt \right) \tag{3.28}$$

[1] also known as Van loan discretisation [123]

Table 3.1: Comparison of NMSE across all states across different observations for the SDOF multisine example example, relative to simulated data.

Observation:	Displacement	Velocity	Acceleration
$x\ (m)$	4.58×10^{-3}	2.80×10^{-2}	9.23×10^{0}
$\dot{x}\ (ms^{-1})$	4.91×10^{-1}	7.18×10^{-3}	1.03×10^{-1}
$\ddot{x}\ (ms^{-2})$	1.15×10^{1}	3.38×10^{-1}	2.41×10^{-2}
$U(t)\ (N)$	1.41×10^{1}	4.20×10^{-1}	1.83×10^{0}

3.4 Example study

To further illustrate this latent force approach, the GPLFM constructed in Section 3.3.2 will be employed in a example study. In this example study, an SDOF system with parameters $m = 1$, $c = 0.1$, and $k = 50$ is subjected to a random phase multisine loading defined by:

$$U(t) = \sum_{i=1}^{N} A_i \sin(\omega_i t + \phi_i)$$

where A_i represents the amplitude, ω_i denotes the angular frequency, and ϕ_i is a random phase uniformly distributed in $[0, 2\pi]$. Here, $A_i = 10 \quad \forall i$, and ω_i is a vector ranging from 0.5 Hz to 10.5 Hz in 10 linearly spaced intervals such that $N = 10$.

To provide data for the analysis the system is discretised and simulated from zero initial conditions at a sample frequency of 2048Hz. This simulated data set will form the baseline the performance of the GPLFM can be compared against.

For the example study, three scenarios are presented. In each scenario, a different state variable is made available as the observation. These measurements are corrupted by artificial noise through the addition of i.i.d. samples from a zero-mean Gaussian distribution with 5% of the root mean squared (RMS) of the observed signal, such that the signal-to-noise ratio remains constant regardless of the amplitude of the actual signal.

The results of the identification are presented in three figures, and the NMSE (see Section 2.8.2) for all states and the recovered latent force across all three case studies is summarized in Table 3.1. Overall, the recovery of dynamic states and latent force varies significantly depending on which state is observed. The latent force is most accurately recovered when velocity is observed, followed by acceleration, and then

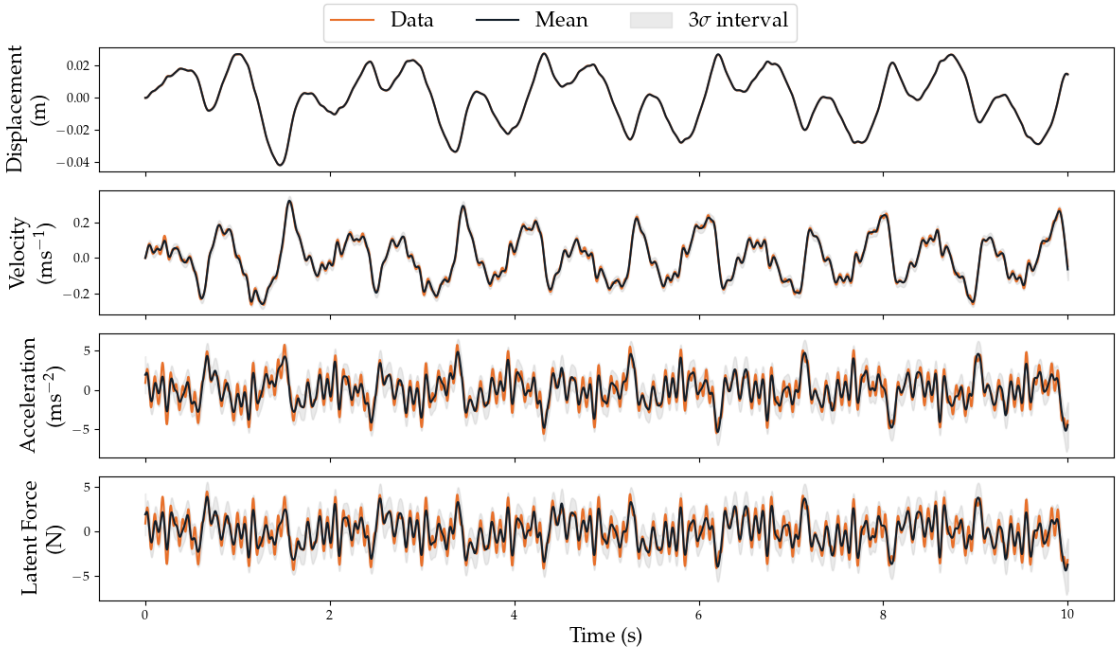

Figure 3.1: Comparison of ground truth (orange) and GPLFM predictions (black) for displacement, velocity, acceleration, and latent force when **displacement** is observed. The shaded region represents the 3σ confidence interval.

displacement.

In the first scenario, displacement is observed, and the results are presented in Figure 3.1. Both displacement and velocity are recovered with high accuracy. However, the recovery of acceleration and the latent force is less accurate, exhibiting a significant increase in error as evidenced both visually in Figure 3.1 and numerically in Table 3.1. This is accompanied by increased uncertainty in these states, as shown in Figure 3.1. Despite the elevated errors, the ground truth consistently falls within the 3σ uncertainty bounds, indicating that the uncertainty has been correctly estimated.

Furthermore, it is important to note that acceleration is derived from displacement, velocity, and the latent force. Since the recovery of the latent force exhibits greater error and uncertainty compared to displacement and velocity, the errors and uncertainties in acceleration are primarily influenced by the latent force[2].

The second scenario examines velocity as the observed variable, with results presented in Figure 3.2. In this case, the recovery of all latent states and the latent

[2]For clarity, acceleration is not part of the state vector $\mathbf{x}(t)$ and is instead calculated from $\ddot{x} = u - kx - c\dot{x}$ after the identification is completed. Therefore, the error in $\ddot{x}$ is a sum of the errors in the latent force u and the states x and $\dot{x}$.

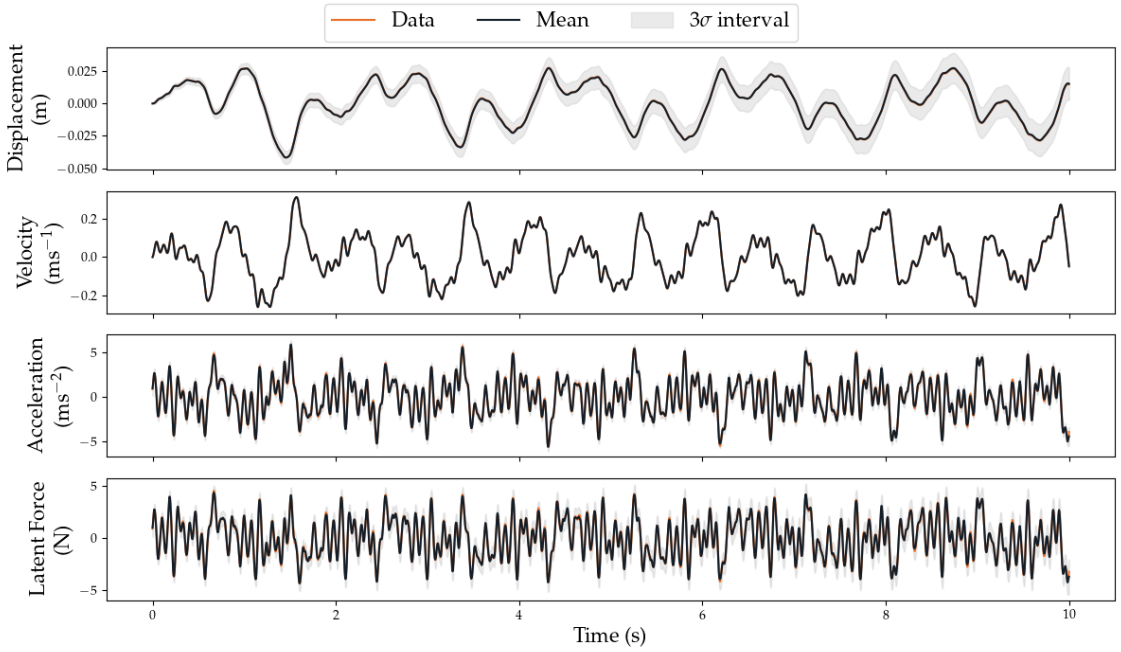

Figure 3.2: Comparison of ground truth (orange) and GPLFM predictions (black) for displacement, velocity, acceleration, and latent force when **velocity** is observed. The shaded region represents the 3σ confidence interval.

force is excellent. Notably, when velocity is observed, the latent force is recovered with the highest accuracy. However, there is a slight increase in the NMSE for displacement compared to the previous scenario. This increase is attributed to the fact that displacement is no longer directly observed.

In the third scenario, acceleration is observed, and the results are depicted in Figure 3.3. Compared to the scenario where velocity is observed, there is a slight decrease in overall performance. Specifically, acceleration and velocity are recovered excellently, displacement recovery is good, and latent force recovery is very good. This indicates that while observing acceleration introduces some challenges, the recovery of key state variables and the latent force remains highly accurate.

However, it can be seen that there is a low-frequency drift in the recovered displacement state. As the displacement begins to drift, the uncertainty bounds widen, reflecting increasing uncertainty in the displacement estimates. This widening of uncertainty bounds serves as a strong indicator that the posterior distribution, not just the mean, is important and essential for fully understanding the system's behavior. This drift phenomenon will be discussed in detail in later.

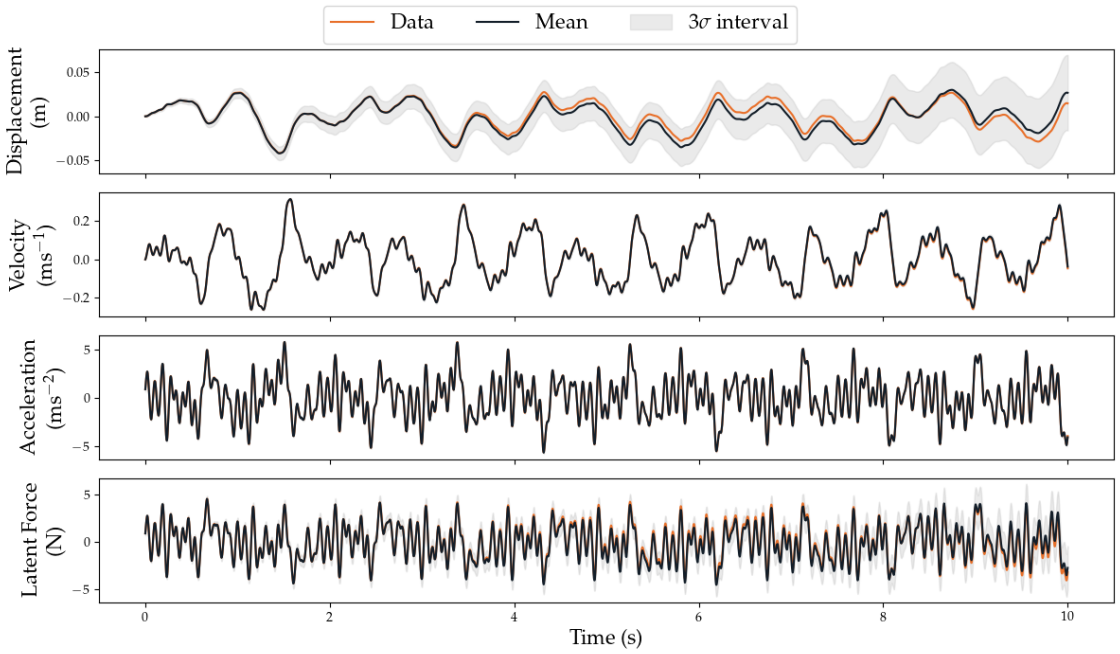

Figure 3.3: Comparison of ground truth (orange) and GPLFM predictions (black) for displacement, velocity, acceleration, and latent force when **acceleration** is observed. The shaded region represents the 3σ confidence interval.

3.4.1 Discussion

The results demonstrated that the GPLFM provides a powerful and effective framework for joint input-state identification. However, the accuracy of the recovery is highly dependent on which state variable is observed. This observation raises two primary questions: Why is joint input-state identification possible when there are multiple latent variables, and why does the observed state variable have such a significant effect on the quality of the recovery? This section seeks to address these two points.

Feasibility of joint-input state estimation

It is reasonable to first question why joint input-state estimation is possible. Even for this simple SDOF example, when one state variable is measured, it leaves at least one state variable and the input unobserved. Therefore, it might be expected that there would be no unique solution to this problem. For instance, if displacement is observed, there exists an infinite set of possible velocities and inputs that would satisfy the system's equations of motion. Consequently, without additional

constraints or information, one would not anticipate being able to uniquely and accurately recover the unobserved states and input from limited observations.

It might be expected that access to other state variables, beyond the one observed, could be achieved through numerical differentiation or integration. However, this is generally not feasible due to the challenges associated with numerical differentiation and integration on signals corrupted by significant noise, as these processes tend to amplify noise or induce drift.

Furthermore, Bayesian filtering and smoothing techniques, which leverage probabilistic models to optimally combine prior information with observed data and are effective at recovering latent states from noisy observations, cannot traditionally be applied here as they rely on known inputs to propagate the state estimates forward in time (see Section 2.6).

The problem of joint input-state identification is solved by constructing the GPLFM using two separate SSM. The first SSM defines the prior over the dynamics of the system, while the second SSM represents the GP and extends the state-space by introducing additional dimensions defining a prior over the input. The variation in the prior over the dimensions of the GPLFM that represents the dynamic states and the latent forces, encodes sufficient structure to make the problem both separable and solvable.

In particular, a strong prior is imposed over the latent states, derived from the system dynamics (Eq.(3.22)), to ensure accurate recovery of the derivative relationships between the states. This prior enforces the known derivative relationships between the observed data and the latent states, such that acceleration is the derivative of velocity and velocity is the derivative of displacement. In the example study, low variance in this prior is ensured by setting $Q_{\text{GPLFM}_{11}}$ to zero for the dimensions associated with the system dynamics, implying no process noise in the dynamics. This reflects the assumption that the dynamic model is correct[3] [4].

In contrast, the Matérn kernel, used as a prior for the latent force, is more flexible

[3]Note that variance still enters the dynamic states due to uncertainty in the measurements and the latent force, and is reflected in the posterior covariance.

[4]The process noise on the dynamic states does not need to be zero for the GPLFM to correctly identify the states and latent force. However, the process noise must be sufficiently small. If the process noise is too large, it weakens the prior, implying little knowledge about the system's dynamic structure. As a result, it becomes difficult to separate the contributions of the system dynamics from those of external inputs.

but less informative, encoding only that the latent force is a realization of a stochastic process with damped oscillatory components. This allows for a wide range of possible functions to be modeled. The smoothness of the process is controlled by the hyperparameter ν, the correlation length by ℓ, and the variance by σ^2, providing a flexible prior. Together, they define the process noise as $Q_{\text{GPLFM}_{22}} = 4\lambda^3\sigma^2$, where $\lambda = \sqrt{3}/\ell$. As σ^2 increases and ℓ decreases, the process noise on the Matérn kernel increases. Here, $\ell = 0.05$ and $\sigma^2 = 3$, resulting in $Q_{\text{GPLFM}_{22}} = 498831$, ensuring that $Q_{\text{GPLFM}_{11}} << Q_{\text{GPLFM}_{22}}$.

The contrast in the strength and therefore variance of the priors over the latent states and latent force ensures that the Matérn GP over the latent force remains flexible enough to learn a wide range of inputs. At the same time, the stronger prior with lower variance over the latent states ensures accurate state estimation. This balance prevents the GP for the latent force from overshadowing the contributions of the latent states, thereby preserving the integrity of the state representation.

The GPLFM prior can then be updated from noisy observations via Bayesian filtering and smoothing. This process avoids the need for direct numerical differentiation or integration of the states by providing an optimal estimate of the state variables (for the linear case), while minimizing the impact of noise. Consequently, it reduces the noise amplification that differentiation can cause and the drift that integration might induce during the identification process. This approach is made possible by modeling the input as a GP, which augments the state-space and allows state estimates to be propagated forward in time.

Impact of the observed variable on the quality of recovery

The results of the example study demonstrate that the selection of the observed state variable significantly affects the accuracy of the recovery of both the latent states and the latent force. Specifically, the most accurate recovery occurs when velocity is observed, followed by acceleration, and finally displacement.

As previously discussed, it is possible to view the latent force identification problem as a challenge to recover total internal restoring force of the system, which is defined by the left-hand side of the equation of motion:

$$F_{\text{int}}(t) = m\ddot{x}(t) + c\dot{x}(t) + kx(t) \tag{3.29}$$

According to Newton's laws, the external input force is equal and opposite to the total internal restoring force, i.e.,

$$U(t) = F_{\text{int}}(t) \tag{3.30}$$

Therefore, the observed state that allows for the most accurate recovery of the internal restoring force will also lead to the most accurate estimation of the latent input force.

When only one state variable is observed, the quality of the recovery depends fundamentally on how well the observed state helps to reconstruct the total internal restoring force. There are two primary factors to consider in this context, assuming that it is possible to observe each state and that the signal-to-noise ratio (SNR) of the observations is equal for each observable state.

Propagation of measurement noise through the system model The first factor concerns the derivative or integral relationship between the observed variable and the internal latent states[5]. When a noisy observation is integrated or differentiated through the observation or state transition matrix as part of the update step in the filter or smoother, additional uncertainty is introduced on top of the original uncertainty in the observation.

While the filtering and smoothing steps in the GPLFM framework significantly reduce the amplification of high-frequency noise and mitigate drift by incorporating prior knowledge from the physical model, the presence of noise—and thereby uncertainty—in the observations means that some information is fundamentally lost. This loss cannot be completely recovered in a joint input-state identification context, even with optimal estimation techniques.

For instance, when displacement is observed, estimating velocity and acceleration requires differentiating the noisy displacement measurements. This differentiation can lead to inaccuracies in the estimated system state. Low-amplitude but high-frequency oscillations in velocity—caused by significant accelerations—may have a

[5]Note that in this context, integration and differentiation refer specifically to the calculation of the derivative or integral of a state variable, not through numerical techniques, but by passing the variable through the measurement and system SSM as part of Bayesian filtering and smoothing techniques.

minimal effect on the displacement, making them easily obscured by noise. This discrepancy makes it difficult to accurately capture the system's acceleration, introducing further uncertainty in force recovery, particularly affecting the contribution from the inertia term $m\ddot{x}(t)$. This phenomenon is illustrated in Figure 3.1. When displacement is observed in the example study, although velocity recovery is relatively accurate, small errors cause high-frequency components in the recovered acceleration and latent input force to be either missed or smoothed out.

Conversely, when acceleration is observed, velocity and displacement must be estimated by integrating the noisy acceleration measurements. This process can introduce errors into the state estimation. Low-frequency changes in velocity, caused by small variations in acceleration, are easily masked by noise. The accumulated errors in velocity can cause substantial drift in the displacement state, affecting the contribution from the stiffness term $kx(t)$ and introducing significant inaccuracies in identifying the latent input force. This behavior is illustrated in Figure 3.3. When acceleration is observed in the example study, although velocity recovery remains relatively accurate, small errors progressively accumulate, resulting in considerable drift in the displacement state. This drift, in turn, induces a corresponding and opposite drift in the recovered latent force.

When velocity is observed, only a single integration or differentiation step is required to estimate displacement or acceleration, respectively. This reduces the compounding errors seen when two integration or differentiation steps are necessary. In the case study example where velocity is observed, an excellent recovery of all dynamic states, as well as the most accurate recovery of the latent input force, is achieved.

It should be noted that improved results across all observed states could be achieved if the hyperparameters of the GP used to infer the latent force were optimized. In this example study, the hyperparameters were selected and fixed across the three scenarios. This was done partly to highlight the effect of the observed state on identification and partly because the optimal hyperparameters would vary depending on the observed state. Fixing them allows for a more straightforward comparison.

Contribution to the total internal restoring force The second factor is the proportion of the total internal restoring force that is contributed by the observed state. Observing a state directly (even if noisily) is likely to produce the best estimation of that particular state. Therefore, observing the state that has the

largest contribution to the total internal restoring force is likely to provide the best estimate of $F_{\text{int}}(t)$, assuming noise effects are consistent across observed state varaibles.

In the context of the linear SDOF system, the contributions of displacement, velocity, and acceleration to $F_{\text{int}}(t)$ are given by $kx(t)$, $c\dot{x}(t)$, and $m\ddot{x}(t)$, respectively. The dominant contributor to the total internal restoring force is primarily determined by the characteristics of the input force, such as its frequency content.

For instance, in systems subjected to high-frequency inputs (relative to the dynamical properties of the system), the inertial term $m\ddot{x}(t)$ becomes more significant. Conversely, for low-frequency inputs, the stiffness term $kx(t)$ may dominate. In the example study, the input force has significant high-frequency components, making the acceleration term a larger contributor to $F_{\text{int}}(t)$.

In the example study, the inertia term $m\ddot{x}(t)$ contributes a larger proportion to the total internal restoring force than the stiffness term $kx(t)$. This explains why the latent force is recovered more accurately when acceleration is observed compared to when displacement is observed. Although significant drift occurs in the displacement when acceleration is observed, this has only a limited effect on the accuracy of the latent force recovery. In contrast, when displacement is observed, high-frequency content in the acceleration is significantly smoothed, leading to a corresponding smoothing of high-frequency components in the latent force hindering the latent force recovery.

Combined effect of both factors The combined effect of these two factors explains why observing different state variables significantly impacts latent force recovery. It is important to note that the optimal state to observe depends on the specific system under consideration. Additionally, these interactions become more complex when dealing with MDOF systems. However, more importantly, the choice of which state variable to observe for identification is primarily determined by the practicalities of measuring physical systems. Due to the widespread availability of affordable piezo-electric accelerometers, acceleration data is commonly used in most practical structural dynamics testing [5]. However, this does not undermine the importance of understanding how the observation interacts with the model.

3.5 Summary

The chapter presented an introduction to the GPLFM for joint input-state identification. Key topics covered include a review of relevant literature on GPLFMs, the derivation of the GP based on the mean and covariance functions of a stationary temporal GP, and the formulation of a state-space GPLFM from a state-space GP. A example study was also discussed to demonstrate the application of the GPLFM, emphasizing how the choice of observed state variables affects its performance in joint input-state estimation. This chapter serves as a basis for understanding the fundamentals of GPLFM, providing the groundwork for the advancements introduced in subsequent chapters.

Chapter 4

Stationary nonlinear GPLFMs for input force identification for Hysteretic Systems

In the previous chapter, the GPLFM was introduced as an effective method for joint input-state estimation. This method was demonstrated through a linear SDOF example study, which investigated the influence of different observed variables on the identification process. It was observed that the derivative and integral relationships between the uncertain observed variable and the latent states significantly affect the quality of latent force recovery.

However, a challenge that has not yet been addressed is the scenario where latent states of the system are not merely the time derivatives or integrals of other hidden states. Additionally, these hidden states do not correspond to the time derivatives or integrals of the observed signal.

To address this gap, this chapter aims to explore a challenging case: assessing whether the joint latent force approach can be used to recover both latent inputs and latent internal states from noisy observations in nonlinear hysteretic systems.

4.1 Bouc Wen model of hysteresis

Hysteresis can be considered a nonlinear memory effect where the state of a system depends not only on its current input but also on its history of past inputs. Hysteresis can be used to model internal friction, deformation of rubber, and shape memory alloys, to name a few examples. The equation of motion for the Bouc-Wen model of hysteresis [124] is given as,

$$m\ddot{x} + c\dot{x} + kx + z(x, \dot{x}) = u(t) \tag{4.1a}$$

$$\dot{z}(x, \dot{x}) = \alpha\dot{x} - \beta\left(\gamma|\dot{x}||z|^{\nu-1}z + \delta\dot{x}|z|^{\nu}\right) \tag{4.1b}$$

where, x is displacement, $\dot{x}$ is velocity, m is mass, c is viscous damping, k is linear stiffness and z encodes the nonlinear hysteresis memory effect. The rate of change of z, $\dot{z}$ is defined by α, β, γ, δ and ν which are used to tune the shape and smoothness of the hysteresis loop.

4.2 Nonlinear GPLFMs

In the previous chapter, it was demonstrated that a linear GPLFM can be solved analytically using a Kalman filter and RTS smoother. However, for latent input force identification in nonlinear systems, the GPLFM must also become nonlinear. This extension introduces significant challenges, as Bayesian filtering and smoothing for nonlinear systems often lack analytical solutions. Additionally, accurately representing and propagating uncertainty through nonlinear transformations is complex, as these transformations typically result in non-Gaussian distributions.

Rogers et al. introduced a methodology for nonlinear input-state estimation using a nonlinear GPLFM [75]. This methodology employs a particle Gibbs with ancestor sampling (PGAS) [125] inference scheme combined with the GPLFM to produce a Monte Carlo approximation of the smoothing distribution. It is used to infer both the latent forces and hidden internal states from noisy acceleration measurements in a nonlinear Duffing oscillator. The results demonstrate a high accuracy in the

recovery process.

However, a limitation of this work so far is that it is restricted to systems where all latent internal states are either integrals or derivatives of the noisy observations. This constraint limits the applicability of the method to more complex systems exhibiting nonlinear dynamics that cannot be solely described by such relationships.

In this chapter, an extension of the framework proposed by Rogers et al. [75] is developed. A nonlinear GPLFM capable of joint input-state estimation for nonlinear systems, including those with latent internal states not restricted to integrals or derivatives of the noisy observations, is presented. Specifically, the methodology is applied to systems that exhibit Bouc Wen hysteresis, a challenging form of nonlinearity.

The importance of this extension stems from the fact that numerous real-world systems exhibit intricate nonlinear behaviours. By broadening the applicability of the GPLFM, this work enables the joint estimation of input forces and internal states in such systems, thereby enhancing the scope of GPLFM for input identification.

4.3 Constructing a Nonlinear GPLFM

To develop the nonlinear GPLFM for joint input-state identification in systems exhibiting Bouc-Wen hysteresis, the first step is to define the nonlinear transition model for the mean state of the system as a set of first-order ODEs:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ -\frac{c}{m}\dot{x} - \frac{k}{m}x - \frac{1}{m}z + \frac{1}{m}u(t) \\ \alpha\dot{x} - \beta\left(\gamma|\dot{x}||z|^{\nu-1}z + \delta\dot{x}|z|^{\nu}\right) \end{bmatrix} \tag{4.2}$$

Then the nonlinear transition model can be augmented such that the input is treated as an internal state of the system, modelled by a state-space GP. This approach enables the representation of the transition of the mean states of the nonlinear GPLFM as follows:

$$\underbrace{\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{z} \\ \dot{u} \\ \ddot{u} \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} \dot{x} \\ -\frac{c}{m}\dot{x} - \frac{k}{m}x - \frac{1}{m}z + \frac{1}{m}u(t) \\ \alpha\dot{x} - \beta\left(\gamma|\dot{x}||z|^{\nu-1}z + \delta\dot{x}|z|^{\nu}\right) \\ \dot{u} \\ -\lambda^2 u - 2\lambda\dot{u} \end{bmatrix}}_{F(x)} \tag{4.3}$$

Furthermore, the covariance over the states is captured by:

$$L = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad P_0 = \begin{bmatrix} \sigma_{\dot{x}}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\ddot{x}}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{z}}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 2\lambda\sigma^2 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & 4\lambda^3\sigma^2 \end{bmatrix} \tag{4.4}$$

This formulation introduces an additional dimension beyond that seen for the GPLFM formulation in Eq.(3.26) and Eq.(3.27) to characterise the temporal mean and covariance of the internal state z.

4.4 Inference

The main challenge in moving from linear GPLFM to nonlinear is in the additional complications within the inference. Since the system is nonlinear, there is no longer a closed-form solution to the filtering and smoothing distribution of the GPLFM. Therefore, approximations need to be employed.

One effective method for recovering the filtering distributions over nonlinear SSM is the particle filter. The particle filter is employed to approximate distributions of the kind $p(x_{1:t} \mid y_{1:t})$. This process involves estimating the entire sequence of states up to the current time, conditioned on the observations available up to that point. This is done as a Monte Carlo method where a collection of weighted point masses (or particles) evolves over time, and through repeated proposals, weighting, and resampling, they create an importance sampling approximation to the filtering distribution at each time step. A brief introduction to the particle filter is found in

Section 2.6.3.

However, in this work, access is needed to the smoothing distribution. The reason is that the smoothing distribution is essential for correctly replicating the full posterior of a temporal GP in state-space form. The smoothing distribution estimates the probability distribution of the state at a given time, conditioned on all observations collected up to a future point in time. More concretely, smoothing is concerned with generating samples from the joint distribution $p(x_{1:T} \mid y_{1:T})$ or evaluating marginal distributions $p(x_t \mid y_{1:T})$, where $t = 1, \ldots, T$.

One potential approach for accessing the marginals $p(x_t \mid y_{1:T})$ involves first computing the joint distribution $p(x_{1:T} \mid y_{1:T})$ using a particle filter and the marginals can be obtained by integrating out the variables $x_{1:t-1}$ and $x_{t+1:T}$. However, this strategy is typically impractical when the number of particles n is much smaller than the length of the observation sequence T, as the particle filter tends to suffer from particle degeneracy and sample impoverishment over long sequences, leading to inaccurate marginal estimates.

There are a number of methods that can instead be applied to recovering the smoothing distribution. A naïve approach to estimate this distribution might be to use a Markov chain Monte Carlo (MCMC) method. However, this approach has two main limitations. First, the likelihood of the system is not available in a closed-form expression. Second, the smoothing distribution is often highly dimensional, which complicates the construction of efficient proposals within traditional MCMC techniques, such as the Metropolis-Hastings algorithm [126].

To address these limitations, Lindsten et al. developed a Particle MCMC (PMCMC) technique known as Particle Gibbs with Ancestor Sampling (PGAS) [127]. This algorithm builds upon the previously derived PMCMC methods developed by Andrieu et al. [125]. Andrieu et al. demonstrated that by incorporating a Sequential Monte Carlo (SMC) scheme within an MCMC framework, it is possible to make the inference on the high-dimensional distributions found in nonlinear state-space models more efficient.

The approach by Andrieu et al. leverages the marginal likelihood estimate provided by SMC, denoted as $\hat{p}_\theta(x_{1:T}|y_{1:T})$, to provide a valid kernel that can be used within a Metropolis-Hastings framework. For this SMC Markov kernel to be valid, it must be ergodic with a unique stationary distribution. To ensure this,

Andrieu et al. proposed a modification to the standard particle filter where one particle trajectory is predetermined *a priori*. This predetermined path, represented as $x'_{1:T} = (x'_1, \ldots, x'_T)$, functions as a reference trajectory, guiding the simulated particles towards a specific region of the state space. This gives rise to the Particle Gibbs (PG) algorithm.

The core of the SMC algorithm of the PG operates similarly to the standard PF and shares many of the same steps. However, there is a difference in that one particle trajectory is specified *a priori* as a reference trajectory $x'_{1:T} = (x'_1, \ldots, x'_T)$. This is implemented by setting the trajectory of the Nth (final) particle and its ancestry a^N deterministically for all time.

When propagating the particles through time using the move kernel in the PG SMC algorithm, the particle proposals are conditioned on the retention of the reference trajectory $x'_{1:T}$, After a complete pass of the SMC algorithm from $t = 1$ to $t = T$, the Markov kernel can be defined by the probability distribution $\mathbb{P}(x^*_{1:T} = x^i_{1:T}) \propto w^i_T$.

Although the addition of a reference trajectory $x'_{1:T}$ ensures that the kernel is ergodic with a unique stationary distribution, it has been noted by Lindsten and Schön [128] and Chopin and Singh [129] that it can lead to significant path degeneracy, causing poor mixing efficiency of the kernel. To address this problem, Lindsten et al. introduced an *ancestor sampling* step, formulating the PGAS algorithm. In this ancestor sampling step, a new ancestor for the reference trajectory is sampled at each time step. The process involves determining which particles at time $t-1$ could have generated the reference trajectory observed at time t based on the proposal probabilities.

The aim of the ancestor sampling step is to, for each time step $t \geq 2$ artificially assign a history to the partial path of the reference trajectory $x'_{t:T}$ by assigning its history as one of the particles from the set $\{x^i_{1:t-1}\}_{i=1}^N$. This reassigning of the historical path of the reference trajectory is facilitated by sampling a new value for the index variable $a^N_t \in \{1, \ldots, N\}$, which links $x'_{t:T}$ to one of the particles. This selection in made based upon the weights which can be calculated from,

$$\widetilde{w}^i_{t-1|T} \triangleq w^i_{t-1} \frac{p_T((x^i_{1:t-1}, x'_{t:T}))}{p_{t-1}(x^i_{1:t-1})} \tag{4.5}$$

for $i = 1, \ldots, N$ and where, $(x^i_{1:t-1}, x'_{t:T})$ denotes the concatenated particle path

created by assigning $x'_{t:T}$ with an alternate path history. The calculation of these weights then allows for a_t^N to be sampled with probability proportional to $\widetilde{w}^i_{t-1|T}$.

Algorithm 2 outlines the PGAS procedure. For notational clarity, the algorithm is presented assuming fixed model parameters, θ, and thus this dependence is suppressed in this work. For instance, the initialization density is written as $r_1(x_1)$ instead of the more complete $r_1(x_1|\theta)$. This simplification reflects the study's focus on state-only inference. However, the PGAS kernel is naturally suited for joint parameter and state estimation when used within a Gibbs framework, as illustrated in the context of GPLFM identification [77, 78].

In Algorithm 2 M_t is the proposal kernel,

$$M_t(a_t, x_t) = \frac{w_{t-1}^{a_t}}{\sum_{j=1}^N w_{t-1}^j} r_t(x_t|x_{1:t-1}^{a_t}) \tag{4.6}$$

W_t is the weight function for $t \geq 0$,

$$w_t^i = W_t(x_{1:t}^i) = \frac{\gamma_t(x_{1:t}^i)}{\gamma_{t-1}(x_{1:t-1}^i) r_t(x_t^i|x_{1:t-1}^i)} \tag{4.7}$$

and γ_t is the target density all of a standard SMC sampler (see Section 2.5.1 or [130] for more).

By constructing the Markov kernel in Algorithm 2, it is possible to draw samples from the smoothing distribution $p(x_{1:T}|y_{1:T})$ using Algorithm 3 allowing or a Monte Carlo approximation of the smoothing distribution.

4.5 Results

For this case study, two distinct loading conditions were investigated. In the first scenario, the load was defined as a random sample drawn from a GP in time. This scenario aimed to simulate the response using a sample from the GP prior. In the second scenario, the system was loaded with a 30 Hz sine wave excitation, which is not explicitly drawn from the GP prior over the forcing. The decision to use

Algorithm 2 PGAS Markov kernel

Require: Reference trajectory $x'_{1:T} \in \mathcal{X}^T$.
Ensure: Sample $x^*_{1:T} \sim P^N(x'_{1:T}, \cdot)$ from the PGAS Markov kernel.
1: Draw $x_1^i \sim r_1(x_1)$ for $i = 1, \ldots, N-1$.
2: Set $x_1^N = x'_1$.
3: Set $w_1^i = W_1(x_1^i)$ for $i = 1, \ldots, N$.
4: **for** $t = 2$ to T **do**
5: Draw $\{a_t^i, x_t^i\} \sim M_t(a_t, x_t)$ for $i = 1, \ldots, N-1$.
6: Set $x_t^N = x'_t$.
7: Compute $\{\tilde{w}^i_{t-1|T}\}_{i=1}^N$ according to Eq.(4.5).
8: Draw a_t^N with $\mathbb{P}(a_t^N = i) \propto \tilde{w}^i_{t-1|T}$.
9: Set $x^i_{1:t} = (x^{a_t^i}_{1:t-1}, x_t^i)$ for $i = 1, \ldots, N$.
10: Set $w_t^i = W_t(x^i_{1:t})$ for $i = 1, \ldots, N$.
11: **end for**
12: Draw k with $\mathbb{P}(k = i) \propto w_T^i$.
13: **return** $x^*_{1:T} = x^k_{1:T}$.

Algorithm 3 MCMC smoother

Require: $x_{1:T}[0]$ (Initial state trajectory)
Ensure: $x_{1:T}[1], \ldots, x_{1:T}[K]$ (K samples from the Markov chain)
1: **for** $k = 1, \ldots, K$ **do**
2: Run the PGAS (Algorithm 2) conditional on $x_{1:T}[k-1]$ to obtain $x_{1:T}[k]$.
3: **end for**

a sine wave was driven by its widespread application in vibrational testing and its capability to produce distinct and measurable resonant responses in mechanical systems.

For both case studies, data for the analysis was generated by discretising and simulating the system from zero initial conditions at a sample frequency of 4096 Hz. This simulated dataset serves as the baseline for evaluating the performance of the GPLFM. For identification, only true (non-gravitational) acceleration measurements were made available. These measurements were corrupted by artificial noise through the addition of i.i.d. samples from a zero-mean Gaussian distribution with a standard deviation of 2% of the RMS value of the signal. To recover the latent input force from the noise-corrupted measurements, it is assumed that the initial conditions, the functional form of the model, and the system parameters are known *a priori*.

In each case study, the unknown latent force was modeled using a GP with a Matérn 5/2 kernel, which augments the nonlinear SSM with three additional states.

4.5.1 Loaded with GP

Latent force recovery was performed using the GPLFM and the PGAS Algorithm. The measured acceleration response to this load is shown in Figure 4.1. The results of the estimated states from the proposed inference scheme are shown are shown in Figure 4.2. A enlarged view of the latent force recovery can be seen in Figure 4.3.

In the left panel of Figure 4.2, the paths sampled using MCMC are depicted. The Markov chain was executed 5000 over iterations. There was a burn-in of an initial 1000 iterations that were discarded to ensure convergence to the target distribution. This is crucial because, at the start of the chain, the samples may be far from the typical set of the target distribution, potentially biasing the results. After the burn-in period, the samples are more representative of the stationary distribution. After that the samples were thinned by a factor of two. The particle filter was implemented as a bootstrap filter with 15 particles. Even with this relatively small number of particles, the smoothing distribution appears to be well approximated. When constructing a Gaussian approximation of the smoothing distribution from these samples–as illustrated in the right column of the figure–it can be seen that all the true states fall within a three-sigma confidence interval.

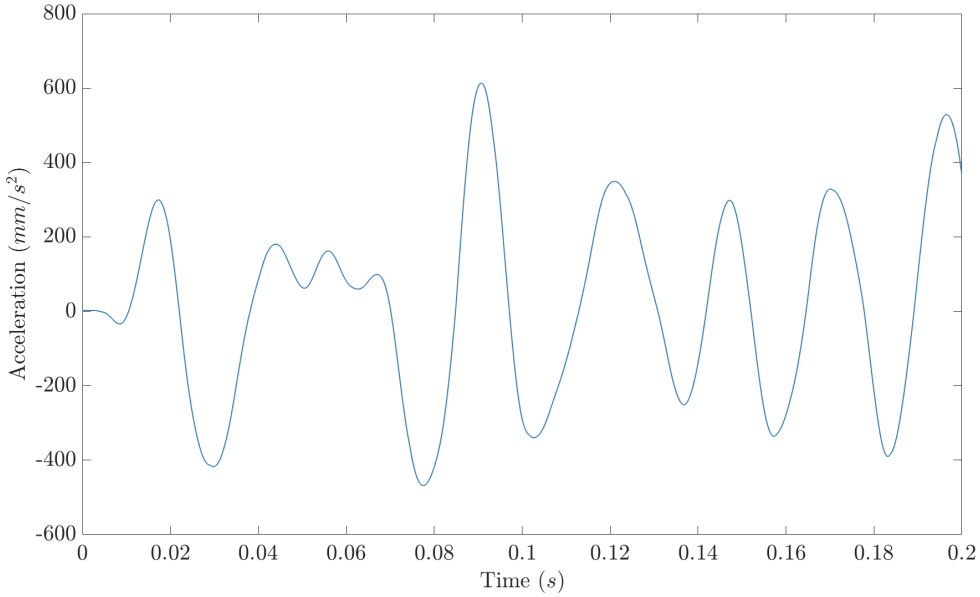

Figure 4.1: Acceleration measured from the Bouc-Wen oscillator driven by a random sample from a Gaussian process with a Matern 5/2 kernel.

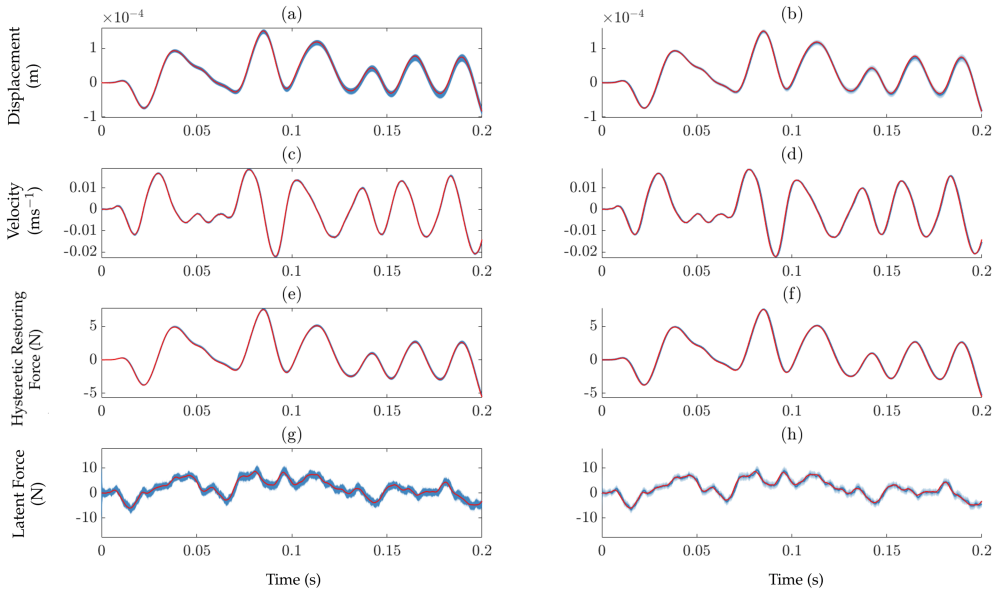

Figure 4.2: Estimated states were inferred using the PGAS sampling of the smoothing distributions. The left column displays samples acquired through the MCMC scheme, while the right column presents Gaussian approximations to the distributions, including one, two, and three sigma intervals shaded. The four rows represent the first four states of the model: displacement, velocity, $z(x, \dot{x})$, and the latent force. In each plot, the ground truth is depicted in red.

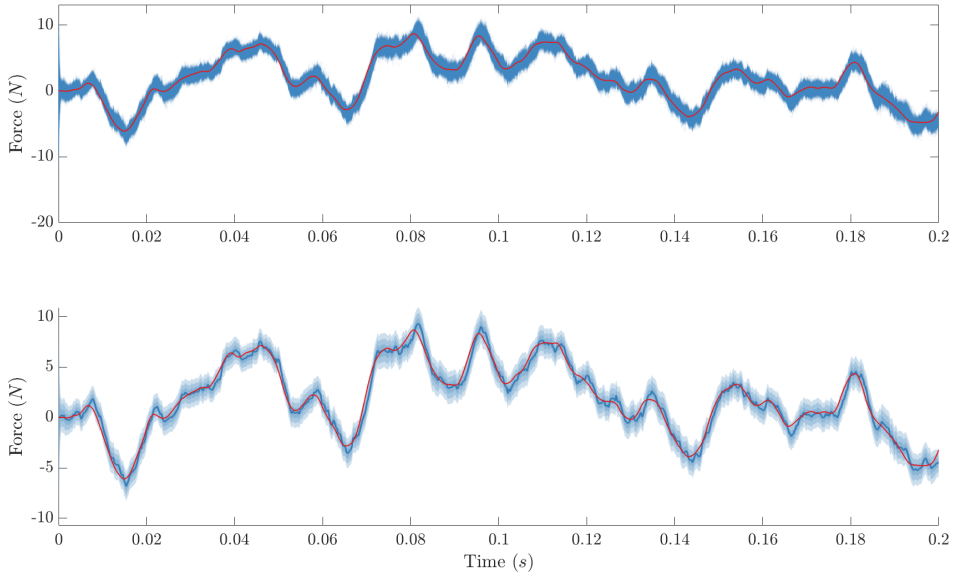

Figure 4.3: Enlarged plot from Figure 4.2 of the estimation of loading of the random sample drawn from the GP in time for the Bouc-Wen oscillator using a nonlinear latent force model. The actual ground truth is represented in red. The top section illustrates various sampled loading signals, while the bottom section shows the mean estimate in blue accompanied by shaded regions indicating three-sigma intervals from the approximated Gaussian distribution.

Table 4.1: NMSE for the recovery of the states and the latent force when the hysteretic nonlinear system is loaded by a random sample drawn from a Gaussian process in time.

State	NMSE
x	0.38
$\dot{x}$	0.33
$z(x, \dot{x})$	0.25
U	2.18

To provide a quantitative comparison, the NMSE was calculated for the mean of the samples across all states, including the latent force. A summary of these results can be found in Table 4.1.

Throughout the plots in Figure 4.2 and the NSMEs in Table 4.1 it is seen that this GPLFM is able to recover the dynamic states of the system excellently. This is seen throughout the dynamic states including the hidden state of the hysteresis $z(x, \dot{x})$.

There is an increase in NMSE when considering the forcing that can be attributed to two primary factors. First, the 2% noise in the measurements obscures high-frequency, low-amplitude oscillations. Second, the high-frequency components of the force are averaged out when the mean of the samples is taken. Increasing the number of particles in the particle filter could alleviate this issue by improving the resolution of the state distribution, thus enabling a more accurate evaluation of the expectation over both the dynamic states and the latent force.

Despite the increase in NMSE and the limitations in capturing high frequencies in the latent force, the estimate of the forcing remains very good (Section 2.8.2). Furthermore, since the uncertainty in the forcing is also quantified and shown in the figure, it is evident that the true forcing consistently falls within the three-sigma confidence interval.

4.5.2 Loaded with Sine Wave

In the last case study, the structure of the latent force model was matched exactly with the actual forcing signal applied to the system. To examine a more complex and realistic scenario, the current case study considers a situation in which the loading signal is not drawn from a GP. Instead, the loading signal is represented as a deterministic function specifically a sine wave mathematically expressed as:

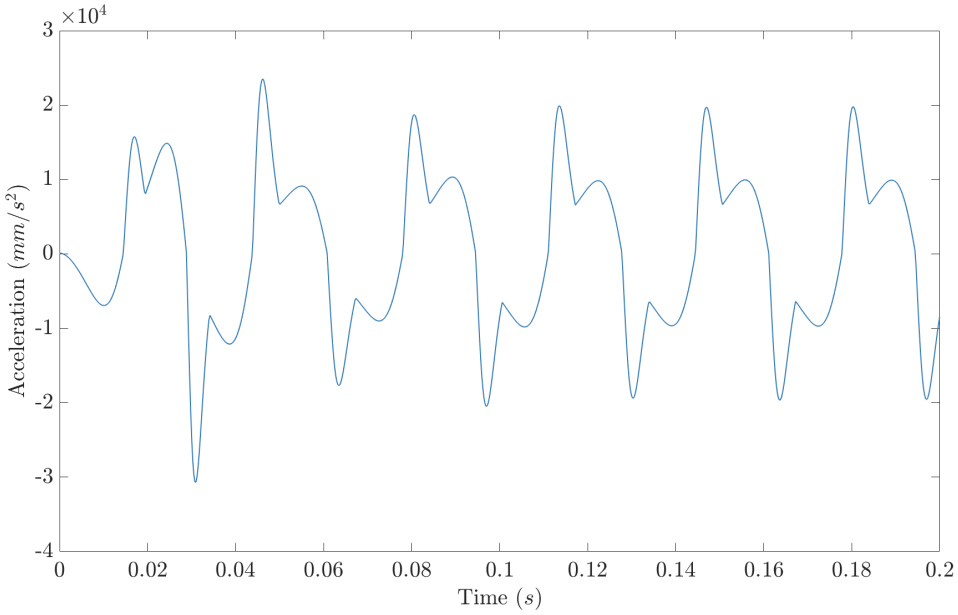

Figure 4.4: Acceleration measured from the Bouc-Wen oscillator driven by 30 Hz sine wave.

$$u(t) = 120 \sin(30 \cdot 2\pi \cdot t). \tag{4.8}$$

Figure 4.4 illustrates the system's response to the sinusoidal forcing.

If the functional form of the input was known *a priori* to be a sine wave, simpler methods would suffice to recover the signal. However, in most operational conditions—such as wind or wave loading on offshore structures—the functional form of the loading remains unknown. To address this uncertainty, the GPLFM is employed. Consequently, this case study investigates whether the Matérn 5/2 kernel can appropriately serve as a prior for the loading function in a hysteric system when its functional form is unknown.

As per the previous case study, the identification was performed using the nonlinear SSM of the Bouc Wen system. This system was augmented with a state-space GP with a Matérn 5/2 kernel and the latent force was recovered using the PGAS algorithm and noisy acceleration measurements. Initially, 5000 samples of the states were obtained, from which the first 1000 samples were discarded as burn-in. After that the samples were thinned by a factor of two. 15 particles were used in the bootstrap particle filter. This sampling procedure enabled the construction of the

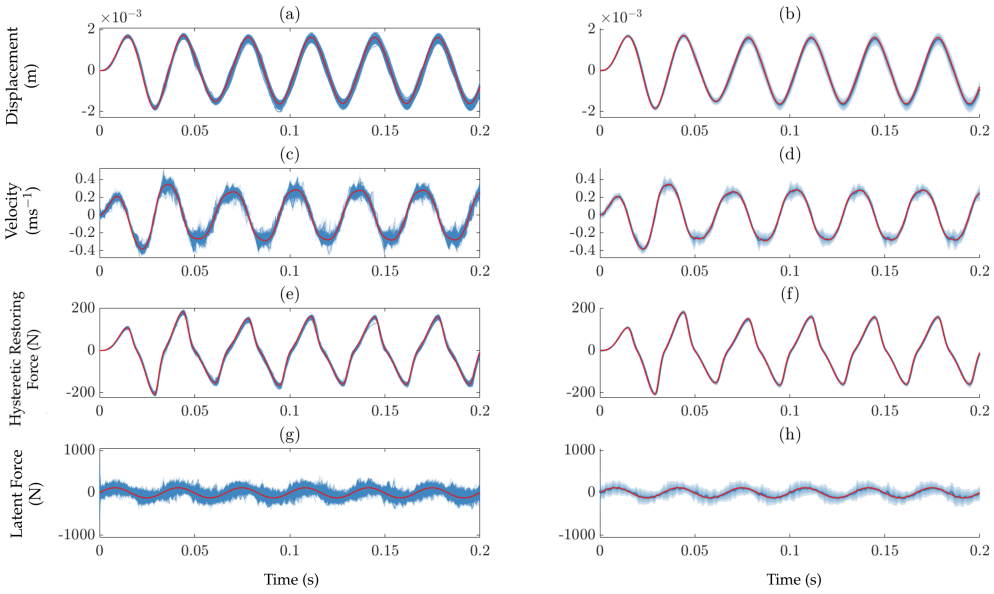

Figure 4.5: Estimated states were inferred using the PGAS sampling of the smoothing distributions. The left column displays samples acquired through the MCMC scheme, while the right column presents Gaussian approximations to the distributions, including one, two, and three sigma intervals shaded. The four rows represent the first four states of the model: displacement, velocity, $z(x, \dot{x})$, and the latent force. In each plot, the ground truth is depicted in red.

results in Figure 4.5.

In Figure 4.5, the left-hand column displays the samples obtained from MCMC in blue, while the ground truth is represented in red. The right-hand column depicts Gaussian assumed densities, created by calculating the expectation and variance of the samples at each time point. The rows of Figure 4.5 correspond to different states: displacement, velocity, $z(x, \dot{x})$, and force.

As per the previous case study the NMSE was also calculated to a quantitative comparison of the mean of the samples of each state to the ground truth. The results can be found in Table 4.2.

It can be seen from Table 4.2 that there is a slight increase in prediction error across all states when compared to the first case study. This is accompanied by a notable rise in estimated uncertainties for all states, as seen in Figure 4.5. The increased uncertainty and error can be primarily attributed to the misalignment between the loading signal and the prior functional form of the GPLFM. Since the prior is less informative about the underlying latent force, the accuracy of recovery decreases,

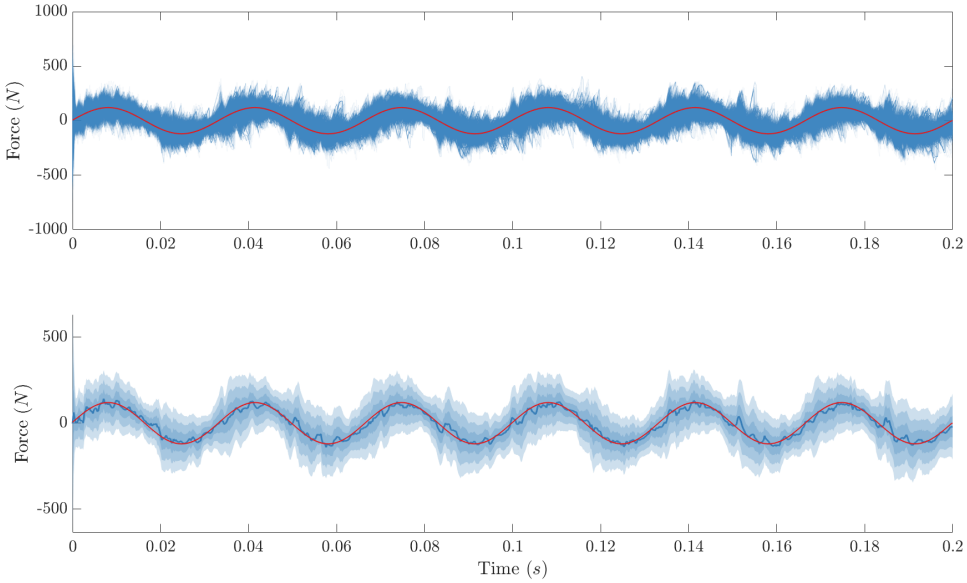

Figure 4.6: Sine wave loading estimation for the Bouc-Wen oscillator using a nonlinear latent force model. The actual ground truth is represented in red. The top section illustrates various sampled loading signals, while the bottom section shows the mean estimate in blue accompanied by shaded regions indicating three-sigma intervals from the approximated Gaussian distribution.

Table 4.2: NMSE for the recovery of the states and the latent force when the hysteretic nonlinear system is loaded by a sine wave.

State	NMSE
x	0.48
$\dot{x}$	0.36
$z(x, \dot{x})$	0.31
U	5.22

leading to a broader spread in the posterior samples and, hence, greater overall uncertainty.

For the latent force this can be seen more clearly in the enlarged view in Figure 4.6. The top half of the figure shows the uncertainty represented as a set of Monte Carlo samples where as the bottom half shows the uncertainty as an approximated Gaussian of the smoothing distribution. From this it can be seen that forcing state exhibits the most significant increase in uncertainty, especially when compared to Figure 4.3 from the previous case study. However, by looking at the expectation of the samples also in the bottom half of the Figure in blue it can be seen that despite a increase in NMSE the recovery of the main trend of the sine wave is evident, with the mean of the sampled states aligning well with the ground truth.

It can be noted that most of the error is in the erroneous addition of high-frequency content to the recovered signal. It is likely this is caused by the limited number of particle in the filter. However, it is also possible that a further tuning of the hyperparameters might help address this issue.

The increase NMSE and uncertainty is not just limited to the latent force, it feeds through into the dynamic states as well. This is seen particularly in the velocity state, which, demonstrates a marked increase in uncertainty at the response's peaks.

However, from Table 4.2 it can be understood that the recovery of the dynamics states can be considered excellent and the recovery of the latent force can be considered good verging on very good (Section 2.8.2). This provides strong evidence that the proposed approach for joint input-state estimation in nonlinear systems is effective, even in the presence of hysteretic nonlinearities. The additional states introduced in the model (z and $\dot{z}$) do not lead to significant observability issues, and the estimates of both the states and the latent force remain highly accurate.

However, the methodology is not without its shortcomings. A significant computational burden is one major drawback, primarily due to the necessity of running multiple particle filters, which considerably increases the computation time relative. To address this limitation, more efficient inference schemes should be investigated in future work.

Alternatively, employing a different approximation method for the nonlinear system's smoothing distribution, such as a Gaussian smoother that approximates the nonlinear dynamics, may be beneficial. This approach is expected to be particularly

advantageous when the system's nonlinearity is weak. By exploring these alternatives, the computational efficiency of the current methodology can be substantially improved, facilitating its application to a broader range of practical problems.

4.6 Summary

This chapter investigated the performance of the GPLFM in a joint input-state identification approach, which was used to recover both latent inputs and latent internal states from noisy observations. In this investigation, the latent states of the system were not the time derivatives or integrals of other hidden states or the observed variables. Specifically, the use of a GPLFM for joint input-state recovery was examined in nonlinear hysteretic systems.

By employing the nonlinear GPLFM method in conjunction with Particle Gibbs and Ancestor Sampling to approximate the smoothing distribution over latent states and input forces, the results indicated highly accurate input-state estimation for this complex dynamical system.

This analysis illustrates that the effectiveness of the GPLFM is not contingent on the hidden states being derivatives or integrals of one another or the observed state. Instead, as long as the GPLFM is constructed such that the transition function and observation model can adequately map between all states and observations, the GPLFM remains applicable to such systems.

Chapter 5

Stationary GPLFMs for Output-Only Restoring Force Identification

In the previous chapters, the GPLFM was introduced as an effective method for joint input-state estimation in both linear and nonlinear systems. However, the application of the GPLFM is not limited to this form of problem.

The identification and characterisation of nonlinear dynamic systems remains an ongoing challenge across various engineering disciplines. This process serves two primary purposes: to gain insight into the physical phenomena governing a system's behavior, and to predict the system's response to new inputs. To meet these objectives, the GPLFM can be applied to the identification of latent restoring forces.

Latent restoring force identification provides a pathway to infer, critical, often nonlinear, internal mechanisms that restore equilibrium. This step is crucial to ensure that a model accurately represents the system's dynamics, particularly when these dynamics are not known *a priori*. By employing the GPLFM in this context, models can be learned that offer a more accurate representation of complex systems while also quantifying the uncertainty in the identification process. This enhances both the understanding of the system and its predictive capability, as well as clarifying its limitations as a result of the uncertainty in the identification.

Despite significant advancements in latent input force modeling through GPLFMs,

the exploration of GPLFMs for latent restoring force identification remains relatively underdeveloped. Rogers and Friis [79] made a notable contribution to this area by developing a GPLFM for nonlinear restoring force estimation. Their work, focusing on an SDOF Duffing oscillator, demonstrated the capability of the GPLFM in jointly estimating parameters, states, and restoring forces. Initially, the GP model recovered the time series of the restoring force, followed by a nonlinear restoring force surface reconstruction through polynomial curve fitting. More recently, Marino and Cicirello [131] extended this work by developing a switching GPLFM model capable of capturing non-smooth, non-stationary, nonlinear internal restoring forces, such as those generated by frictional contact in mechanical systems.

However, both approaches assume access to both input and output observations, albeit noisy. In controlled laboratory environments, where input-output measurements can be carefully monitored, this assumption holds. Data collection in such settings can facilitate reliable identification of latent restoring forces using the established methods.

When dealing with large-scale infrastructures, however, replicating these controlled conditions becomes significantly more challenging. Variability in structures, environmental influences, and the sheer scale of the systems introduces significant challenges. As a result, controlled experiments on large-scale infrastructures are often infeasible due to both physical constraints and financial limitations. As such, access to direct input measurement is often severely limited.

Therefore, there is a need for latent restoring force identification methods that can be performed using output-only measurements. This requirement introduces significant challenges, as it necessitates the joint estimation of input forces, latent restoring forces, and internal states, thereby substantially increasing the non-identifiability of the system. The absence of direct input measurements complicates the inference process, making it difficult to accurately disentangle the contributions of latent restoring forces from the observed outputs.

To address these challenges, a novel methodology is introduced that combines a stationary linear GPLFM with additional post-processing steps including Gaussian process regression to recover a restoring force surface representation of the unknown nonlinearity. This approach is designed to facilitate accurate latent restoring force identification and nonlinear restoring surface recovery from output-only data by leveraging the strengths of the GPLFM framework while mitigating the complexities

associated with increased non-identifiability.

5.1 Related work

The need for output-only system identification has long been recognized in the field of structural dynamics and engineering. In many practical situations, especially in large-scale structures or systems subjected to environmental or operational loads, direct input measurements are either difficult or impossible to obtain. To address this challenge, Operational Modal Analysis (OMA) was developed as a field for identifying the dynamic properties of structures using only output measurements [132]. Unlike more traditional experimental Modal Analysis, which requires controlled excitation inputs, OMA relies solely on the structure's response to ambient or operational forces, making it a widely applicable technique for systems where input information is unavailable or impractical to measure. By leveraging statistical methods to extract modal parameters from these outputs, OMA has become an essential tool for identifying natural frequencies, damping ratios, and mode shapes in a variety of complex systems.

OMA relies solely on output data, assuming the system is excited by unknown environmental forces—commonly modelled as a white noise process due to its stationary, uncorrelated, flat spectrum nature. While substantial research supports OMA for linear systems [133, 134], its application to nonlinear systems remains less explored but promising.

Recent studies have begun to bridge this gap. For instance, Friis et al. [135] have proposed a method for approximating nonlinear systems as equivalent linear systems using output-only data. Furthermore, Vesterholm et al. [136] used a random decrement technique combined with principal component analysis to detect nonlinearity. This work was compared to earlier attempts by Macías et al. [137], which combined principal component analysis with a finite element model for detection of nonlinear dynamic behaviour.

Building on these advancements, this chapter introduces a novel approach for identifying nonlinear dynamic systems when the functional form of the nonlinearity and the input forces are unknown. The primary objective is to recover a temporal function of latent input force and a functional representation for the unknown nonlinear

component of the dynamic system.

5.2 Unifying framework for latent force identification

So far it has been seen how the GPLFM can be constructed for latent input force problems. The aim of this section is to detail the equivalence of solving latent input force (LIF) and latent restoring force (LRF) problems for mass-spring-damper dynamical systems in the state-space framework. In doing so, it will be shown how a state-space formulation of a GP designed to learn temporal functions can be applied to both problems.

For an SDOF LIF problem characterised by a latent force $U(t)$, the state-space model (SSM) can be expressed as:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{U(t)}{m} - \frac{k}{m}x - \frac{c}{m}\dot{x} \end{bmatrix} \tag{5.1}$$

whereas for an SDOF LRF problem, with a known input force $u(t)$ and a latent restoring force $f(x, \dot{x})$, the SSM is depicted as:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{u(t)}{m} - \frac{k}{m}x - \frac{c}{m}\dot{x} - f(x, \dot{x}) \end{bmatrix} \tag{5.2}$$

where m is mass, c is the damping coefficient, k stiffness.

In both the LIF and LRF scenarios, the core challenge goes beyond simply inferring the system's internal state to learning a latent temporal component that influences the acceleration $\ddot{x}$. In the LIF case, this involves a latent force $U(t)$, whereas for LRF, the focus is on a latent force represented as $f(x, \dot{x})$. The critical insight is that the physical manifestation of the force—whether it originates from within the system or from an external excitation—does not alter the underlying mathematical challenge: to learn the missing component of acceleration $\ddot{x}$ at discrete time steps t as a temporal function. Consequently, methods devised for learning the latent forces within LIF problems are applicable to LRF problems.

5.3 GPLFM for Output-Only Nonlinear System Identification

With the equivalence of solving LIF and LRF problems established, it may be useful to formalise the exact problem that this chapter addresses. Consider the second-order differential equation for an SDOF system:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) + f(\dot{x}(t), x(t)) = u(t), \tag{5.3}$$

where $f(\dot{x}(t), x(t))$ represents a unknown nonlinear function of the states. This state dependent function can equivalently be considered as a temporal function or a temporal latent restoring force $R(t)$:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) + R(t) = U(t), \tag{5.4}$$

The challenge lies in recovering the temporal functions of the latent input force $U(t)$, latent restoring force $R(t)$, and recovering a nonlinear mapping for $f(\dot{x(t)}, x(t))$ from incomplete noisy observation of the system state.

It can be noted that there exist infinitely many solutions where $R(t) + U(t)$ is equal to a constant. Consequently, the contributions to the system's response from the nonlinear dynamics and the unknown input cannot be separated without additional constraints on the analysis. In this context, $U(t)$ is assumed to be modelled as a white noise process, $U(t) \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is the variance of the white noise process. This assumption is particularly common in OMA [133, 138], where unknown ambient excitation is often idealised as a broadband stochastic process to enable system identification from output-only data, without which identification would not be possible. While this represents a simplification of complex, unmodelled real-world excitations, it serves as a common and necessary assumption. However, if the true input deviates significantly from a white noise process, its characteristics could be erroneously attributed to the nonlinear dynamics ($R(t)$), potentially biasing the separated contributions and affecting the overall accuracy of the identification

Under this assumption, a solution to this problem using the GPLFM can be achieved

by first modelling both the latent input force $U(t)$ and the latent restoring force $R(t)$ as GPs:

$$R(t) \sim \mathcal{GP}(0, k_r(t,t')), \quad U(t) \sim \mathcal{GP}(0, k_u(t,t')), \tag{5.5}$$

where $k_u(t,t')$ is the kernel of the input process $U(t)$ and $k_r(t,t')$ of the unknown restoring force $R(t)$.

As the state-space GPLFM is *a priori* a Markov process meaning that the state at one time step depends only on the state at the previous time step it will not be possible to separate $U(t)$ and $R(t)$ in an online fashion. Rather, separation of contributions to a total resorting force,

$$\hat{R}(t) = R(t) - U(t), \tag{5.6}$$

will have to take place in post processing steps. As such, an advantageous characteristic of the GP is that the addition of two GPs can be expressed in a closed form, so that,

$$\hat{R}(t) = (R(t) - U(t)) \sim \mathcal{GP}(0, k_u(t,t') + k_r(t,t')). \tag{5.7}$$

Given this, the identification of $\hat{R}(t)$ can be reformulated to a standard latent input force identification problem as seen in Chapter 3,

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = -\hat{R}(t), \quad \hat{R}(t) \sim \mathcal{GP}\left(0, k_r(t,t') + k_u(t,t')\right). \tag{5.8}$$

Therefore, it is now possible to augment the dynamic SSM equation such that the latent input is treated as an internal state of the system, modelled by the state-space GP. This allows the transition of the mean states to be represented as:

$$\underbrace{\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \\ \dot{\hat{R}}(t) \\ \ddot{\hat{R}}(t) \end{bmatrix}}_{\dot{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{k}{m} & -\frac{c}{m} & \frac{1}{m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\lambda^2 & -2\lambda \end{bmatrix}}_{\mathbf{F}_{\text{GPLFM}}} \underbrace{\begin{bmatrix} x(t) \\ \dot{x}(t) \\ \hat{R}(t) \\ \dot{\hat{R}}(t) \end{bmatrix}}_{\mathbf{x}(t)} \tag{5.9}$$

Furthermore, the covariance over the states is captured by,

$$L_{\text{GPLFM}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad P_{0_{\text{GPLFM}}} = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_{\dot{x}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\hat{R}}^2 & 0 \\ 0 & 0 & 0 & 2\lambda\sigma_{\hat{R}}^2 \end{bmatrix}, \quad Q_{\text{GPLFM}} = \begin{bmatrix} q & 0 \\ 0 & 4\lambda^3\sigma^2 \end{bmatrix} \tag{5.10}$$

The GPLFM Bayesian filtering and smoothing solution to Eq. (5.8) provides the posterior distribution $p(\mathbf{x}_{1:T} \mid y_{1:T})$ over the internal states $x(t)$, $\dot{x}(t)$, and the total latent restoring force $\hat{R}(t)$, given the noisy measurements $y_{1:T}$. This is achieved despite the incorrect latent force structure in Eq. (5.8) because the GPLFM encodes the correct derived relationship between the observations and the latent states. However, the estimates for the total latent force $\hat{R}_{1:T}$ will still contain contributions from the random excitation $U_{1:T}$ and the internal restoring force $R_{1:T}$, which must be separated.

First, it will be useful to define an expectation with respect to time, conditioned on the states, for some general function of time $G(t)$:

$$\mathbb{E}_t \left[G(t) \mid x(t), \dot{x}(t) \right] = \int G(t) p(G(t) \mid x(t), \dot{x}(t)) \, dt. \tag{5.11}$$

Now consider the expectation of the input which is assumed to be Gaussian with respect to time, conditioned upon the states:

$$\mathbb{E}_t \left[U(t) \mid x(t), \dot{x}(t) \right] = \mathbb{E}_t \left[U(t) \right] \approx 0. \tag{5.12}$$

The input to the system is a white noise process so is statistically independent of

the system states and depends only on the stochastic process itself resulting in an expectation of zero [1].

Moreover, expectation of the internal restoring force with respect to time, conditioned upon the states, is determined by the nonlinear function:

$$\mathbb{E}_t \left[R(t) \mid x(t), \dot{x}(t) \right] \approx f(x(t), \dot{x}(t)). \tag{5.13}$$

Thus, by taking the expectation of the total restoring force with respect to time and conditioned on the states, it is possible to write:

$$\mathbb{E}_t \left[\hat{R}(t) \mid x(t), \dot{x}(t) \right] = \mathbb{E}_t \left[R(t) \mid x(t), \dot{x}(t) \right] - \mathbb{E}_t \left[U(t) \mid x(t), \dot{x}(t) \right]. \tag{5.14}$$

Consequently,

$$\mathbb{E}_t \left[\hat{R}(t) \mid x(t), \dot{x}(t) \right] \approx \mathbb{E}_t \left[R(t) \mid x(t), \dot{x}(t) \right] \approx f(x(t), \dot{x}(t)), \tag{5.15}$$

implying that the expectation of the total restoring force, with respect to the states, is simply the nonlinear restoring force surface as the contribution to the expectation from the external forcing white noise process is zero.

To evaluate the expectation in Eq. (5.15), this work proposes fitting a second GP in the phase-space in a post-processing step such that:

$$f(x(t), \dot{x}(t)) \sim \mathcal{GP}\left(0, k(\mathbf{x}(t), \mathbf{x}(t)')\right). \tag{5.16}$$

This GP performs a regression to learn a mapping from the estimated states ($x(t)$ and $\dot{x}(t)$) to the total latent force, $\hat{R}(t)$, all of which are estimated by the GPLFM. The phase-space GP can be trained using the recovered time series data $(\mathbf{x}_{1:T}, \hat{\mathbf{R}}_{1:T})$, where the inputs to the GP are the 2-dimensional state vectors $\mathbf{x}(t) = [x(t), \dot{x}(t)]$ and the outputs are the corresponding total latent force values $\hat{R}(t)$.

[1] Note the expectation is approximate as the integral must be computed as a Monte Carlo approximation from discrete samples.

The expectation of the distribution over the function learned by the GP approximates $f(x(t), \dot{x}(t))$, and the variance corresponds to the uncertainty in the recovery of the total restoring force as well as the variance in the white noise process input $U(t)$. Unfortunately, the contributions to the variance are challenging to separate. This will be discussed in more detail later.

With the representation of the unknown nonlinear restoring force recovered, it becomes possible to retrieve the time-series of the internal restoring force $R_{1:T}$ and the input force $U_{1:T}$. $R_{1:T}$ can be recovered by evaluating the GP representation of the unknown nonlinear restoring force at the recovered values of $x_{1:T}$ and $\dot{x}_{1:T}$. Thereafter, $U_{1:T}$ is recovered by computing:

$$R_{1:T} - \hat{R}_{1:T} = U_{1:T}. \tag{5.17}$$

5.4 Results

To assess the effectiveness of the proposed approach, this case study will examine the Duffing oscillator,

$$m\ddot{x} + c\dot{x} + kx + k_3x^3 = U(t), \tag{5.18}$$

where $k_3 = 1 \times 10^9$ denotes the cubic stiffness coefficient. The substantial value of k_3 ensures that nonlinearity significantly influences the system at the applied forcing level. The linear system parameters are defined as $m = 1\,\text{kg}$, $c = 2$, and $k = 1 \times 10^4$.

For the analysis, the Duffing oscillator was excited with a white noise process modeled as a multisine, including every 0.2 Hz between the ranges of 0.2 Hz to 2000 Hz [2]. The system was simulated using a fourth-order Runge-Kutta scheme over $T = 5\,\text{s}$ with a sampling rate of 4096 Hz. To evaluate the effectiveness of the GPLFM in identifying the latent restoring force, the functional form of the Duffing oscillator's nonlinearity was hidden, resulting in a model of the form:

[2]This will evalaute the efficacy of the proposed methodology given that the Gaussian input assumption holds true.

Table 5.1: NMSE results for the output-only latent restoring force approach

Variable	**NMSE**
x	2.56×10^{-1}
$\dot{x}$	1.56×10^{-2}
$\ddot{x}$	1.96×10^{-3}
$\hat{R}_R(t)$	7.32×10^{1}
$\hat{R}_{\hat{R}}(t)$	2.03×10^{-2}
$R(t)$	9.87×10^{-1}
$U(t)$	1.66

$$m\ddot{x} + c\dot{x} + kx + f(x, \dot{x}) = U(t). \tag{5.19}$$

Furthermore, to test the GPLFM in the OMA case, the input force $U(t)$ was hidden but assumed to be a white noise process *a priori*. Only acceleration measurements, perturbed by adding independent and identically distributed (i.i.d.) samples from a zero-mean Gaussian distribution accounting for 1% of the root mean square (RMS) of the signal, were provided for identification. The states x and $\dot{x}$ were hidden. Note it is assumed that the linear system's functional form, and the system parameters are known *a priori*.

The hyperparameters for the GPLFM were set at $\ell = 1 \times 10^{-3}$ and $\sigma^2 = 30$. The performance of the model prior to the post processing steps is shown in Figure 5.1 with an enlarged view of the recovered latent force visible in Figure 5.2. The states and acceleration were recovered excellently, as supported by the NMSE calculations presented in Table 5.1. However, the recovery of the LRF ($\hat{R}(t)$) was poor with respects to the true $R(t)$ both visually and numerically indicated as $\hat{R}_R(t)$ in Table 5.1 with a NMSE of 7.32×10^1, due to contributions from the LIF $U(t)$. However, the recovery of $\hat{R}(t)$ with respects to the true $R(t) - U(t)$ was excellent with an NMSE of 2.03×10^{-2} indcated in in Table 5.1 as $\hat{R}_{\hat{R}(t)}$.

With good estimates of x, $\dot{x}$, and $\hat{R}(t)$ in hand, a map for the nonlinearity was identified non-parametrically by fitting a GP with a Matérn 5/2 kernel from the estimated input x and $\dot{x}$ to the output $\hat{R}(t)$. The Matérn 5/2 kernel was selected due to its smoothness properties and flexibility. This makes it a good choice of prior when the functional form of the nonlinearity is unknown *a priori*.

To reduce computational burden, the GP was trained on a subset of 2048 sequential data points taken between 2.5 and 3.0 seconds. This subset was selected due to

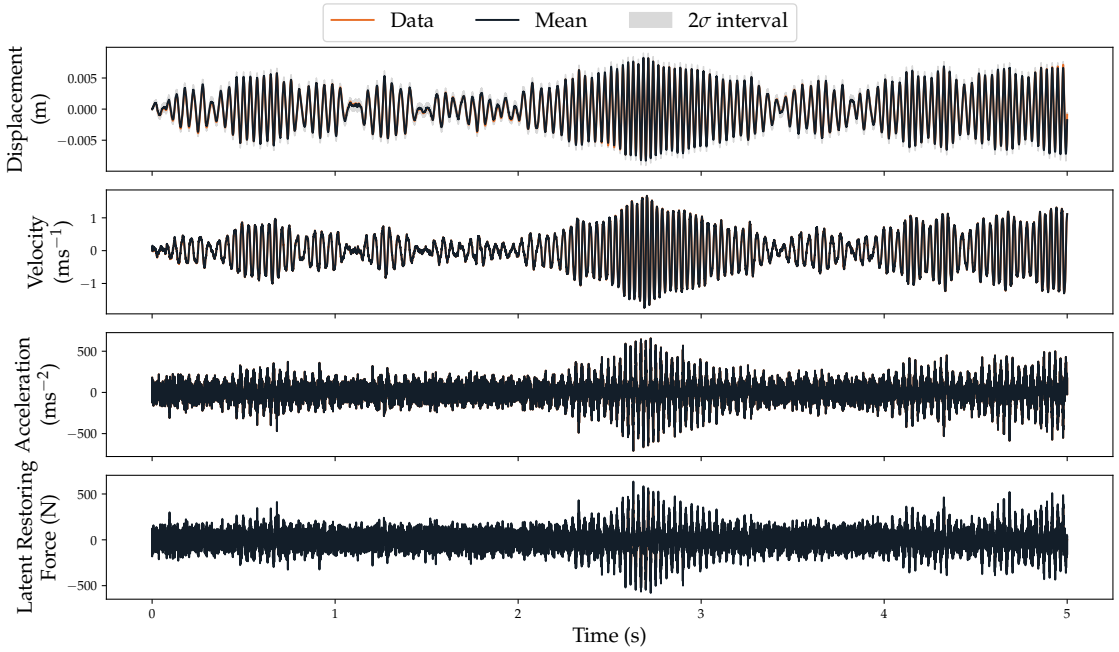

Figure 5.1: State estimates recovered by the smoother before the post processing step, showing the expected values (in black) against the ground truth (orange) with the 2σ shown in shaded grey.

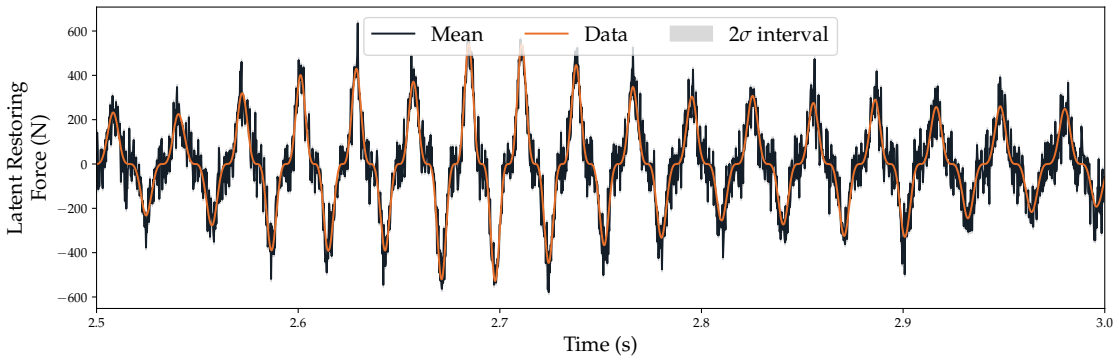

Figure 5.2: An enlarged view of the latent restoring force recovered by the RTS smoother before the post processing step, showing the expected values (in black) against the ground truth (orange) with the 2σ shown in shaded grey.

the relative prevalence of high displacement data points within this region. This selection aimed to provide better training around areas of high displacement and thereby higher restoring force. Figure 5.3 shows the GP fit for the states x and $\dot{x}$ against $\hat{R}(t)$. The subset of data points used for training is depicted in green.

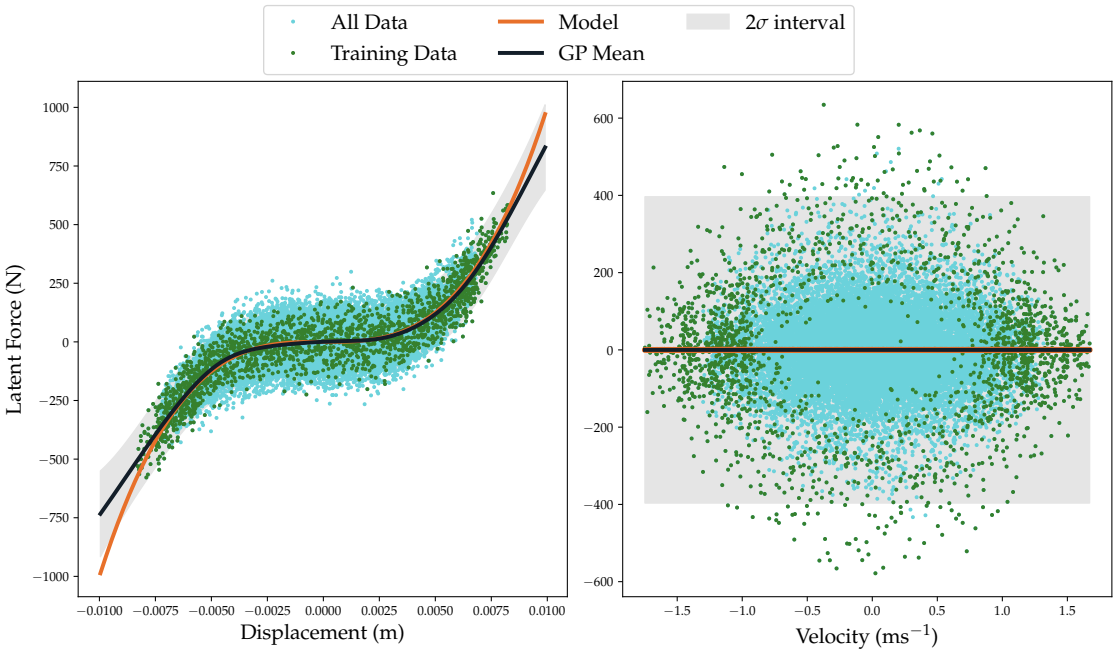

Figure 5.3: Recovered restoring force surfaces, the expected restoring force is shown in black and the true restoring force in orange. All of the data is shown in a light blue with the training data shown as the green dots.

Examining the displacement latent force plot, it can be seen that the GP mean fits the true nonlinearity generally well. However, it can be observed that the fit is not perfect, particularly at the edges of the training data. This misalignment has three potential sources. First, if the recovery of displacement or the total latent restoring force is incorrect, it would bias the GP. However, as seen from the NMSE, the recovery in the states and the total latent restoring force is excellent. The second potential source is the GP prior. A more informative prior would increase the accuracy of the identification [3]. However, this case study is designed to demonstrate a scenario when the practitioner has no prior information about the functional form of the nonlinearity other than it is perhaps a non-infinitely smooth continuous nonlinearity. The third and actual source of the error is based on the number of data points. The recovery of the nonlinear map is dependent on the expectation of the

[3]For example if the practitioner is expecting the system to express a polynomial nonlinearity a polynomial kernel could be chosen to embed this prior knowledge.

LIF being zero. As the expectation $\mathbb{E}_t\left[\hat{R}(t) \mid x(t), \dot{x}(t)\right]$ is calculated as a Monte Carlo approximation, it requires a sufficient number of samples to converge to the underlying mean. This is particularly prevalent in high displacement and high LRF regions of the nonlinear map due to these data points being relatively rare within the entire dataset due to the tall narrow peaks typical of the restoring force of the Duffing oscillator at high displacement. Therefore, an increase in the accuracy of the latent force map could be achieved by increasing the number of time steps used in the GP as it corresponds to using more Monte Carlo samples when computing $\left[\hat{R}(t) \mid x(t), \dot{x}(t)\right]$. However, an increase in computational complexity required for training the GP needs to be contended with. Whilst, the state-space GPLFM used to recover $x(t)$, $\dot{x}(t)$ and $\hat{R}(t)$ scales linearly with the number of time steps ($\mathcal{O}(t)$) the phase space GP used to learn the restoring force surface scales cubic ($\mathcal{O}(t^3)$). To counteract this, sparse approximations to the GP or alternatives to the regression methods could be explored.

Outside of the training data, the mean and the ground truth begin to diverge, which is accompanied by the typical behaviour of the GP: the variance in the estimation grows when making predictions beyond the observed data. Positively, the long length scale of the GP helps it make a reasonable extrapolation.

On the right-hand side of Figure 5.3, the relationship between the velocity and the total latent restoring force is presented. It can be seen that this data presents no correlation, with the mean of the GP being zero for all displacements [4]. Overall, this correctly identifies that the contribution of the LRF $f(x, \dot{x})$ is entirely a function of the displacement x, which is identified to be approximately cubic.

With the nonlinear map recovered, it is now possible to separate the time series of the internal LRF and the LIF from the total LRF $\hat{R}(t)$. This is done by first evaluating the posterior mean and variance of the GP for the states at each time step to provide the time series for the internal LRF $R_{1:T}$. Secondly, the mean and variance of the LIF $U_{1:T}$ are recoverd by removing the LRF $R_{1:T}$ contributions from $\hat{R}(t)$, enabling the recovery of the stochastic input's time series. Plots for the post processed LIF and LRF are presented in Figure 5.4, with a zoomed-in view in Figure 5.5. Moreover, the NMSE for the internal LRF and the LIF are presented in Table 5.1.

Overall, Figures 5.4 and 5.5 show that the mean for the LRF and the LIF has

[4]The distribution of the samples across the states is not uniform

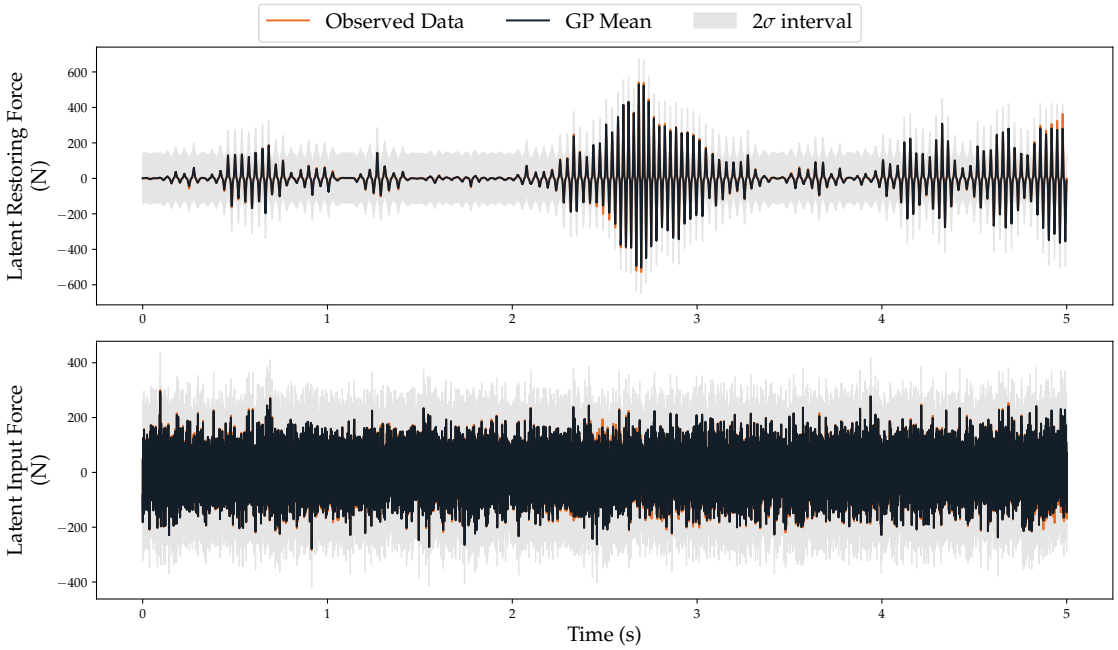

Figure 5.4: Latent input force and latent restoring force recovered from post processing step.

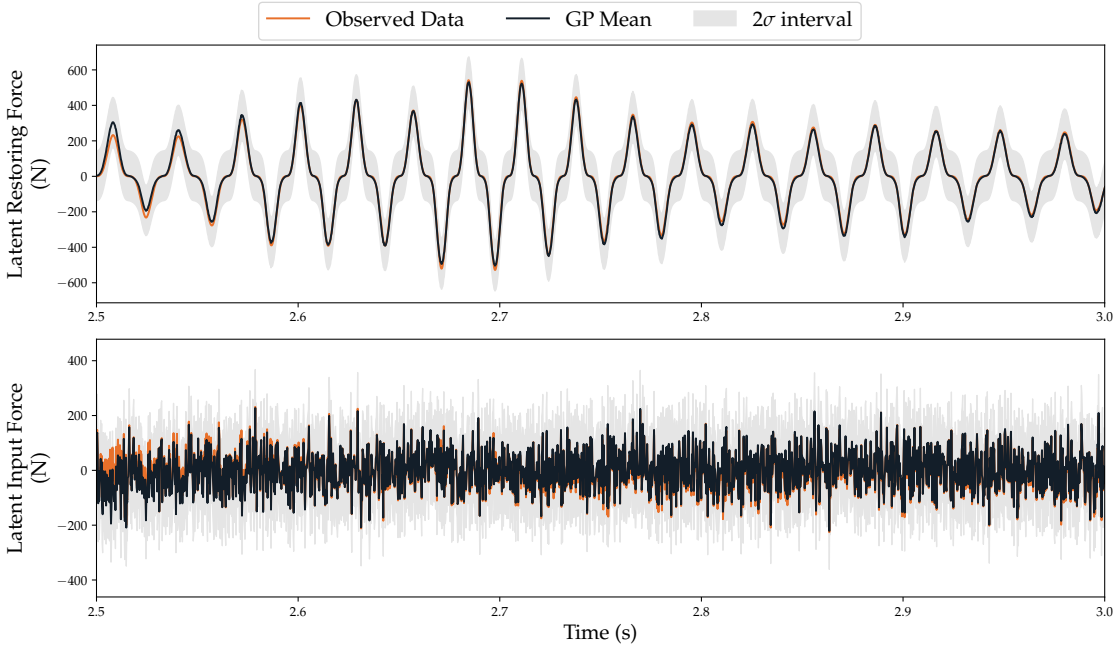

Figure 5.5: Enlarged view of the latent input force and latent restoring force recovered from post processing step.

been recovered well, and their contributions towards $\hat{R}(t)$ have been successfully separated. This is confirmed by NMSEs that score as excellent for the LRF $R(t)$ and very good for the LIF $U(t)$. However, small discrepancies in the time-series can be observed. This error is a result of the error in the latent restoring force recovery. This would mostly likely be improved in increasing the number of time steps in the Monte Carlo approximation.

A notable characteristic of both the recovery of the internal LRF and the LIF is the high 2σ interval, particularly so due to the low 2σ of the recovered total LRF $\hat{R}(t)$. This is a direct feed-through from the recovered nonlinear map. The variance in the GP nonlinear map is a combination of the uncertainty in total LRF recovery and variance in the white noise input force. Since both the uncertainty in the total LRF recovery and the variance in the white noise input force are Gaussian distributed, separating the contributions toward the uncertainty in the internal latent force map and thereby the LRF and the LIF is challenging. Therefore, the total uncertainty in the GP nonlinear map is propagated through to the recovery of both latent forces in Figure 5.4.

5.5 Summary

This chapter has presented a novel methodology for latent restoring force identification using output-only measurements. By integrating a GPLFM with post-processing techniques within an OMA framework by assuming a white noise input, the temporal functions of the latent states, latent restoring forces, and latent input forces were successfully reconstructed. Additionally, a GP representation of the underlying restoring force surface was recovered, providing deeper insight into the system's nonlinear dynamics. The accurate estimation of latent states, forces, and functions is particularly notable given the challenges associated with the non-identifiability of the system.

Chapter 6

Non-stationary Gaussian Process Latent Force Models

In the previous chapters, an introduction to GPLFMs based on an SDE formulation of a GP was provided. The implementation of this framework was demonstrated for applications in latent input force (LIF) recovery for linear and nonlinear systems. These methods were also shown to be effective for latent restoring force (LRF) identification in a output-only nonlinear system identification context.

For practical applications of the GP to latent force modeling, the SDE formulation is often essential, as it significantly reduces computational complexity from $\mathcal{O}(N^3)$ to $\mathcal{O}(N)$, here N denotes the number of time steps in a time-series dataset. This reduction is crucial for handling large time-series datasets common in engineering applications. However, the SDE formulation of the GP, as derived by Hartikainen and Särkkä [72], is predicated on the assumption that the GP covariance function is stationary. This is a necessary condition to express the GP equivalently as a LTI SDE. Consequently, SDE-based GPs inherently have the assumption embedded that a unknown latent function is stationary *a priori*. This creates a fundamental conflict, as the dynamics of many nonlinear, time-invariant systems result in non-stationary time series, directly violating the model's stationarity assumption.

Although a stationary GP can be employed to learn non-stationary temporal functions, its performance is generally suboptimal for capturing the full complexity of non-stationary dynamics. This limitation becomes particularly evident when deal-

ing with non-stationary forces, which exhibit time-varying characteristics that are difficult to model with simple techniques. In LIF tasks, non-stationary loads are commonly encountered in structures subjected to environmental conditions, such as wind [37], waves [38], earthquakes [39], and traffic loading [40]. Additionally, in the context of LRF modeling, any nonlinearity—such as large deflections, bifurcations [6], and hysteresis [10]— introduces state-dependent internal restoring forces. These nonlinearities can result in significant non-stationarity in the temporal response of the internal restoring force, further complicating the assumption of stationarity in latent force identification.

Despite significant research efforts toward the development of non-stationary kernels [66, 139–142], their application to the formulation of SDE GPs remains unexplored. Recent work by Marino and Cicirello [131] has made progress in this area through the development of a switching GPLFM model. This model adeptly captures non-smooth, non-stationary, and nonlinear internal restoring forces, such as those arising from frictional contacts in mechanical systems. The approach employs multiple GPs with distinct length scales, where each GP captures a stationary component of the LRF. The GPs are dynamically switched depending on the identification requirements.

However, this switching model is designed for learning discontinuous nonlinearities, and is therefore not well-suited for capturing the continuous non-stationary LIFs commonly observed in environmental loads, as well as non-stationary LRFs that arise in systems with smooth state-dependent dynamics. Therefore, a significant challenge remains: to develop an SDE formulation for a GPLFM within the state-space designed to learn *smooth*, non-stationary functions. This will be achieved through the development of a novel formulation of a non-stationary GPLFM that incorporates a time-varying length scale into the SDE GP. The introduction of a time-varying length scale substantially improves the model's ability to adapt to the non-stationarity of the underlying latent functions. Additionally, this length scale is dynamically learned through a secondary GP, also formulated within the state-space framework. This new hierarchical or "deep" GP configuration offers a more flexible and adaptive modeling approach, where the primary GP captures the latent forces, while the secondary GP adjusts the length scale dynamically to account for non-stationary behavior.

6.1 Development of the Non-Stationary GPLFM

To form a foundation for the development of the temporal non-stationary GPLFM the Matérn class of kernels is selected for latent force recovery due to the classes flexibility and the feasibility of expressing their SDE formulation in closed form. Specifically the Matérn kernel with a roughness of 3/2 is selected. Recall, it's continuous time mean and covariance formulation is specified by:

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t), \tag{6.1a}$$

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t)\mathbf{F}^T + \mathbf{F}\mathbf{P}(t) + \mathbf{L}\mathbf{Q}\mathbf{L}^T, \tag{6.1b}$$

where,

$$F = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix} \quad L = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad P_\infty = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\lambda\sigma^2 \end{pmatrix} \tag{6.2}$$

and where $Q = 4\lambda^3\sigma^2$, $\lambda = \sqrt{3}/\ell$, and ℓ and σ^2 denote the length scale and variance hyperparameters, respectively.

The length scale parameter, denoted as ℓ, is a pivotal attribute of kernel functions in a GP, serving as a measure of "distance" over which the function values are significantly correlated. Intuitively, it can be thought of as defining the temporal radius within which observations influence each other. A smaller length scale implies that function values change rapidly over short distances. In contrast, a larger length scale suggests that function values are correlated over greater distances. With the GP in the state-space formulation it is evident that ℓ effects both the mean of the GP through the transition matrix F and the covariance through prior covariance P_∞ and the process noise Q.

Adjusting the length scale allows the model to capture the underlying function's behaviour more accurately, striking a balancing between over-fitting and over-smoothing. If the length scale is too short the model will over-fit, capturing noise as if it were signal. If the length scale is too long, the model will over-smooth and gloss over essential variations in the signal.

In the context of non-stationary functions, the challenge is further pronounced. Non-stationary functions have statistical properties that vary over time. Therefore, fixing the statistical properties with a fixed length scale, regardless of its value, compels a compromise: at certain points within the data, it may induce overfitting, interpreting random fluctuations as meaningful patterns, while at other points, it may lead to underfitting, missing important dynamics of the signal. This inherent dilemma underscores that any single length scale, when applied uniformly, is insufficient to optimally model non-stationary functions.

Remark. For clarity, it is noted that a stationary GP, configured with well-chosen hyperparameters, can model any functions accurately, including non-stationary ones. However, this is pivotal on access to a sufficient quantity of high-quality data. The ability of a stationary GP to adapt to non-stationary functions is attributed to the fact that, with the accumulation of high-quality data, the posterior distribution converges away from the prior and towards actual data, thereby lessening the impact of the prior. A significant challenge in much of Bayesian inference arises in scenarios where high-quality data is not abundantly available. In such cases, the selection of an informative prior becomes essential, ensuring that the inference remains robust and precise, even in the face of scarce or lower-quality data.

Given this backdrop, this work advocates that for non-stationary functions, an optimal length scale, ℓ, should inherently be dynamic, adjusting over time to the evolving characteristics of the latent force. This leads to the advocacy for conceptualizing ℓ not as a static parameter but as a temporal function that can be represented as a GP,

$$\ell \sim \mathcal{GP}(0, k(t, t')). \tag{6.3}$$

To accommodate this dynamic nature of ℓ, the GPLFM is expanded to include ℓ as an augmented hidden state that represents the unknown length scale of the GP that infers the latent force. As a consequence, the temporal derivative $\dot{\ell}$ is also tracked. The transition function of the non-stationary GP SDE is given as,

$$\underbrace{\begin{bmatrix} \dot{u} \\ \ddot{u} \\ \dot{\hat{\ell}}_1 \\ \ddot{\hat{\ell}}_1 \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ -\left(\frac{\sqrt{3}}{G(\hat{\ell}_1)}\right)^2 & -2\left(\frac{\sqrt{3}}{G(\hat{\ell}_1)}\right) & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\lambda_2^2 & -2\lambda_2 \end{bmatrix}}_{F(x)} \underbrace{\begin{bmatrix} u \\ \dot{u} \\ \hat{\ell}_1 \\ \dot{\hat{\ell}}_1 \end{bmatrix}}_{x} \tag{6.4}$$

such that the covariance of the GP is can be calculated from,

$$L = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \quad P_\infty = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & 2\left(\frac{\sqrt{3}}{G(\hat{\ell}_1)}\right)\sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & 2\lambda_2\sigma_2^2 \end{pmatrix} \quad Q = \begin{pmatrix} 4\left(\frac{\sqrt{3}}{G(\hat{\ell}_1)}\right)^3\sigma_1^2 & 0 \\ 0 & 4\lambda_2^3\sigma_2^2 \end{pmatrix} \tag{6.5}$$

This non-stationary GPLFM is constructed of one GP that is parameterised by a length scale that is a function of a second GP. Such that $G(\hat{\ell}_1)$ is the length scale of the first GP and $\lambda_2 = \sqrt{3}\ell_2$ where ℓ_2 defines the length scale of the second GP. The hyperparameters σ_1^2 and σ_2^2 are the variance of the first and second GP, respectively. When incorporated with a dynamic system and measurement data, this framework enables the simultaneous inference of the latent force's length scale and the latent force itself through Bayesian filtering and smoothing techniques.

The function $G(\cdot)$ denotes a nonlinear mapping from $\mathbb{R}$ to $\mathbb{R}^+$, ensuring the positivity of ℓ_1 such that $\ell_1 = G(\hat{\ell}_1)$. This mapping is pivotal, as the Matérn 3/2 GP kernel requires ℓ_1 to be strictly positive. However, bounding ℓ_1 to $\mathbb{R}^+$ is not enough to ensure numerical stability when practically implementing the model. It can be seen that $Q_{11} \propto (1/G(\ell_1))^3$ is such that if $G(\ell_1)$ becomes sufficiently large or small the numerical solver can become very quickly numerically unstable. Therefore, an upper and lower bound on ℓ must be enforced. With this in mind, a modified $arctan(\cdot)$ function is selected[1]:

[1]Note, the function $G(\hat{\ell}_1)$ is a choice for the practitioner any function can be used so long as it maps $\mathbb{R}$ to $\mathbb{R}^+$ and is numerically stable for practical implementation.

$$\ell_1 = G(\hat{\ell}_1) = \frac{(\ell_{1,\max} - \ell_{1,\min})}{\pi} \left(\arctan\left(\hat{\ell}_1\right) + \frac{\pi}{2} \right) + \ell_{1,\min} \tag{6.6}$$

for bounds $[\ell_{\min}, \ell_{\max}]$. This function is chosen as it is easy to written down its derivative in a manor that is numerically stable to compute. This will be important for solving the model which will be discussed later.

In the development of the non-stationary GPLFM, the matrices $F(x)$ and $Q(x)$ are now nonlinear because they have become a function of the state x. However, it should be expected that some computational convenience would be lost in capturing this additional complexity.

It should be noted that, despite the GPLFM becoming nonlinear, a matrix representation of the model is maintained. This is unconventional as the transition of states and the update of covariances can no longer be directly computed through a straightforward discretisation and matrix-vector product (see Section 3.3.3). Therefore, an alternative approach will be presented in the next section.

Despite this divergence, the representation of the non-stationary GPLFM in matrix form is intentionally preserved for several reasons. Firstly, it facilitates an easier comparison with stationary GPLFMs. Secondly, this matrix representation aids in visualising how latent variables influence the dynamics of the model, offering intuitive insights into the model's behaviour. Third, it mirrors the practical implementation within the coding framework. Lastly, and most importantly, maintaining this form directly supports the discretisation and linearisation processes essential for solving the non-stationary GPLFM.

For a latent force problem described by the second-order differential equation,

$$m\ddot{x} + c\dot{x} + kx = U(t) \tag{6.7}$$

where $U(t)$ is non-stationary it is proposed that the full system would be,

$$\underbrace{\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{U} \\ \ddot{U} \\ \dot{\ell} \\ \ddot{\ell} \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k}{m} & -\frac{c}{m} & -\frac{1}{m} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\left(\frac{\sqrt{3}}{G(\hat{\ell})}\right)^2 & -2\left(\frac{\sqrt{3}}{G(\hat{\ell})}\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\lambda^2 & -2\lambda \end{bmatrix}}_{F(x)} \underbrace{\begin{bmatrix} x \\ \dot{x} \\ U \\ \dot{U} \\ \hat{\ell} \\ \dot{\hat{\ell}} \end{bmatrix}}_{x} \tag{6.8}$$

where $U(t)$ is the augmented hidden state which represents the unknown forcing and $\hat{\ell}$ is the augmented hidden state which tracks the unknown length scale of the GP that learns $U(t)$.

6.2 Filtering and Smoothing for Non-Stationary GPLFMs

For discrete-time linear systems, the Gaussian filtering and smoothing problem can be solved in closed form [23]. Unfortunately, the model defined in Eq.(6.4) does not conform to these criteria. This section aims to detail the process of discretising a linear system and subsequently linearising a nonlinear system. Through this approach, it will be demonstrated how an approximate discrete linear model can be derived from a continuous nonlinear system. Consequently, this facilitates the computation of an approximate solution to the filtering and smoothing problem for the non-stationary GPLFM using Gaussian assumed density filtering and smoothing.

For the reader's convenience, the Kalman filter equations are repeated below, as they will be needed as reference throughout this section. Here it is made explicit that the discrete-time state transition matrix, $A(x_{t-1})$ and the discrete process noise covariance $\Sigma(x_{t-1})$ are state-dependent:

Prediction Step:

$$x_t^- = A(x_{t-1})x_{t-1}, \tag{6.9a}$$

$$P_t^- = A(x_{t-1})P_{t-1}A(x_{t-1})^T + \Sigma(x_{t-1}). \tag{6.9b}$$

Update Step:

$$v_t = y_t - H_t x_t^-, \tag{6.10a}$$
$$S_t = H_t P_t^- H_t^T + R_t, \tag{6.10b}$$
$$K_t = P_t^- H_t^T S_t^{-1}, \tag{6.10c}$$
$$x_t = x_t^- + K_t v_t, \tag{6.10d}$$
$$P_t = P_t^- - K_t S_t K_t^T. \tag{6.10e}$$

The matrices $A(x_{t-1})$ and $\Sigma(x_{t-1})$ are derived from the discretisation of linear continuous-time differential equations. For linear systems, the calculation of $A(x_{t-1})$ through matrix exponential $\exp(F(x_{t-1})dt)$ is straightforward, benefiting from robust numerical techniques. In contrast, calculating $\Sigma(x_{t-1})$ is more complex and that is more numerically demanding. This complexity can be mitigated by employing matrix fraction decomposition (MFD) [106], which simplifies the computation of $\Sigma(x_{t-1})$ to evaluating a matrix exponential:

$$\begin{bmatrix} A & \Sigma A^T \\ 0 & A^{-T} \end{bmatrix} = \exp\left(\begin{bmatrix} F & LQL^T \\ 0 & -F^T \end{bmatrix} dt \right) \tag{6.11}$$

To implement this discretisation for the nonlinear system, the system must first be linearised. An approximate solution to a nonlinear SDE can be calculated by approximately integrating the following continuous-time differential equations from the initial conditions $m(0) = \mathbb{E}[x(0)]$ and $P(0) = \mathrm{Cov}[x(0)]$ to a target time t, as shown below:

$$\frac{dx}{dt} = f(x,t), \tag{6.12a}$$
$$\frac{dP}{dt} = PF_x(x_t)^T + F_x(x_t)P + LQ(x_t)L^T \tag{6.12b}$$

where $F_x(x_{t-1})$ is the Jacobian of $F(x_{t-1})$.

The local linearisation technique used here involves fixing $F(x_{t-1})$, $F_x(x_{t-1})$, and $Q(x_{t-1})$ at $t-1$ for the integral between $t-1$ and t. This enables the discretisation

of the Eq.(6.12) in a manner analogous to the linear system. Except for this time in Eq.(6.9) $F(x_{t-1})$ is traded out for $F_x(x_{t-1})$ as per Eq.(6.12) such that the prediction equations for the discrete-time linearised nonlinear system are given as,

$$x_t^- = A(x_{t-1})x_{t-1}, \tag{6.13a}$$

$$P_t^- = A_x(x_{t-1})P_{t-1}A_x(x_{t-1})^T + \Sigma(x_{t-1}). \tag{6.13b}$$

Here, $A(x_{t-1})$ is calculated as $\exp(F(x_{t-1})dt)$, $A_x(x_{t-1})$ as the discretised Jacobian $\exp(F_x(x_{t-1})dt)$, and $\Sigma(x_{t-1})$ from MFD Eq.(6.11).

These calculations are performed with the matrices evaluated and fixed from $t-1$ to t. It is noted that this linearisation is repeated in each timestep, re-evaluating $A(x_{t-1})$, $A_{x_{t-1}}$, and $\Sigma(x_{t-1})$ as states transition from x_{t-1} to x_t. Given the relatively slow performance of many numerical matrix exponential implementations for higher-dimensional matrices, this work proposes the use of the Taylor series approximation:

$$e^X \approx \sum_{k=0}^{n} \frac{1}{k!} X^k \tag{6.14}$$

This approximates the matrix exponential as an n ordered Taylor series expansion such that if n $= \infty$ Eq.(6.14) is equal to the true matrix exponential. Here n is the user determined depending on the nonlinearity of the system and the step size. Note that if $n = 1$, then the proposed prediction step is equivalent to the Euler approximation of the nonlinear system.

Algorithm 4 Non-stationary GPLFM

Initialisation: Set the initial system state x_0 and covariance P_0.
Initialise $Q(x_0)$ as per Eq. (6.5).
for $t = 1$ to T **do**
    Filter Prediction Step:
    Evaluate $F(x_{t-1})$, $F_x(x_{t-1})$, and $Q(x_{t-1})$ at $t-1$.
    Discretise $F(x_{t-1})$ and $Q(x_{t-1})$ using matrix fraction decomposition Eq. (6.11).
    Discretise $F_x(x_{t-1})$ with the matrix exponential.
    Propose x_t^- and P_t^- using Eq. (6.13) and the discretised matrices $A(x_{t-1})$, $A_x(x_{t-1})$, and $\Sigma(x_{t-1})$.
    Filter Update Step:
    Update x_t^- and P_t^- using observations y_t via Eq. (6.10).
    return x_t, P_t
end for
for $t = T$ down to 1 **do**
    Smoother Prediction Step:
    Evaluate $F(x_t)$, $F_x(x_t)$, and $Q(x_t)$ at t.
    Discretise $F(x_t)$ and $Q(x_t)$ using matrix fraction decomposition Eq. (6.11).
    Discretise $F_x(x_t)$ with the matrix exponential.
    Propose x_{t+1}^- and P_{t+1}^- using Eq. (6.13) and the discretised matrices $A(x_t)$, $A_x(x_t)$, and $\Sigma(x_t)$.
    Smoother Update Step:
    Update x_t and P_t see Section (2.6.4).
    return x_t^s, P_t^s
end for
Output: $x_{0:T}^s$, $P_{0:T}^s$.

Remark. A potential limitation of the proposed approach arises from the way the model is discretised and linearised for computation. Specifically, this relates to the handling of the process noise covariance matrix, $\Sigma(x)$, which quantifies the intrinsic uncertainty of the system's dynamics. In the developed non-stationary model, this matrix is state-dependent, as its value changes with the estimated length scale, $\hat{\ell}_1$. To make the calculations tractable, a "zero-order hold" is employed for this matrix. This means the value of Σ is calculated using the state at the beginning of a time-step (x_{t-1}), after which that value is assumed to remain constant for the entire duration of the step. The limitation here is that the state, including $\hat{\ell}_1$, is changing during the time step. By holding Σ constant, the changing in the uncertainty over the time step in the length scale itself is ignored when the model's overall state and uncertainty are updated. In other words, how small variations in $\hat{\ell}_1$ within the interval would affect the system's predicted behaviour is not fully accounted for. This means that the effects of variations in $\hat{\ell}_1$ on $\Sigma(x_{t-1})$, and thus on the posterior distribution over the states $p(x \mid y)$, are not included in the filtering and smoothing updates. The development of a method to fully incorporate these intra-step variations remains an open challenge for future research. However, it is important to note that despite this theoretical simplification, excellent recovery of the latent states is demonstrated by the model in the presented case studies.

Using this discretisation and linearisation, it is possible to implement an RTS smoother (seen in Section 2.6) using the described prediction Eq.(6.13). The pseudo code for the proposed non-stationary GPLFM is presented in Algorithm 4 which outlines the step-by-step computational procedure.

6.3 Case Studies

Identifying non-stationary internal and external forces represents a crucial challenge in engineering. To evaluate the effectiveness of the non-stationary GPLFM developed in this research, this section presents two case studies. The first case study investigates an SDOF system subjected to a sine sweep load, which tests the model's ability to model LIFs with non-stationary spectra. The second case study focuses on a bursting Duffing oscillator, chosen to demonstrate the model's capability in handling systems where nonlinearity significantly alters the system dynamics and LRF spectra. These case studies provide practical insights into the performance and

applicability of the proposed model. The case studies are performed for both the non-stationary GPLFM and the stationary GPLFM so that the performance can be compared.

6.3.1 Sine Sweep

This case study explores an SDOF system subjected to an exponential sine sweep. The governing equation for the system's response is represented by the following set of equations:

$$m\ddot{x} + c\dot{x} + kx = U(t), \tag{6.15a}$$

$$U(t) = \sin\left(K\left(e^{t/L} - 1\right)\right), \quad K = \frac{T\omega_1}{\ln\left(\frac{\omega_2}{\omega_1}\right)}, \quad L = \frac{T}{\ln\left(\frac{\omega_2}{\omega_1}\right)}, \tag{6.15b}$$

where $T = 1.5$s represents the duration of the sine sweep, $\omega_1 = 0.05$Hz is the initial frequency, and $\omega_2 = 25$Hz is the final frequency. The model parameters are set with $m = 0.5$ kg, $k = 20$ N/m, and $c = 0.1$ Ns/m.

To provide data for the analysis the system is discretised and simulated from zero initial conditions at a sample frequency of 1024Hz. This simulated data set will form the baseline the stationary and non-stationary GPLFMs performance can be compared against.

For identification, only velocity measurements are made available. These measurements are corrupted artificial noise through the addition of i.i.d. samples from a zero mean Gaussian distribution with 1% of the RMS of the signal.

To recover the latent input force from the noise-corrupted measurements, it is assumed that the initial conditions, functional form of the model and the system parameters are known *a priori*. Latent force recovery was performed using the non-stationary GPLFM, as described in Algorithm 4. The hyperparameters used in this identification are summarised in Table 6.1 and the results are shown in Figure 6.1. The non-stationary GPLFM demonstrated excellent force recovery capabilities, with the identified input force closely tracking the true force.

The length scale of the non-stationary GPLFM is learnt during the identification, starting at a value of 1.87 and finishing at 0.16. This change corresponds to the

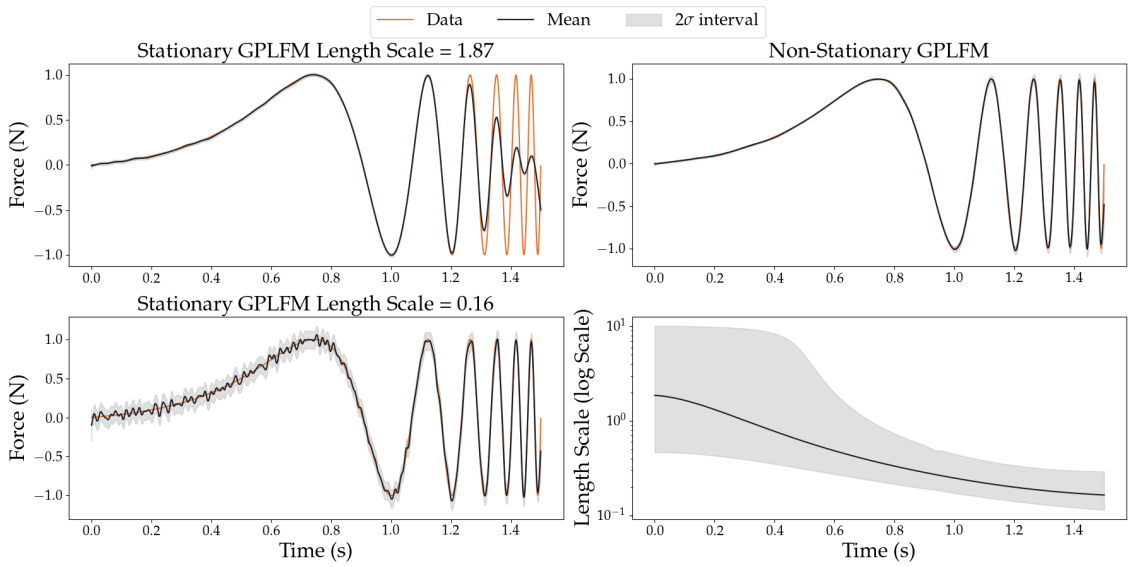

Figure 6.1: Comparison of latent force recovery using stationary and non-stationary GPLFMs for the sine sweep case study. Top left: Force recovery using a stationary GPLFM with length scale 1.87. Bottom left: Force recovery using a stationary GPLFM with length scale 0.16. Top right: Latent force recovery for the non-stationary GPLFM. Bottom right: Dynamic adaptation of the length scale in the non-stationary GPLFM.

variation in the input force frequency from 0.05 Hz to 25 Hz, showcasing the model's ability to learn the length scale of the underlying latent forces effectively. It is noteworthy that the initial length scale ℓ_1 of the filter was set at 5; the different initial length scale observed in the results arises due to the backward recursion of the smoother.

Despite the high accuracy and low variance in the recovered latent forces, a larger variance was observed on the length scale. This is an unfortunate side effect of propagating uncertainty through the nonlinear function described in Eq.(6.6). This problem could be counter acted with the use of Monte Carlo and SMC methods to evaluate the smoothing distribution. However, the application of these methods have to be considered against the additional computational complexity they incur.

To evaluate the performance of the non-stationary GPLFM comprehensively, a comparison was also performed using a standard stationary GPLFM model (implemented as seen in Chapter 3). The hyperparameters used by the stationary GPLFMs are summarised in Table 6.2 and the results can be seen in Figure 6.1.

For the stationary GPLFM, the variance σ^2 was set at 10, identical to the first

GP in the non-stationary GPLFM used to learn the latent force. The comparison revealed that when the stationary GPLFM was configured with a length scale of 1.89 (matching the initial length scale of the non-stationary GPLFM), it performed adeptly at lower frequencies but failed to accurately recover the input force as the frequency increased, due to over-smoothing. Conversely, with a shorter length scale of 0.16 (matching the end configuration of the non-stationary GPLFM), the stationary model initially introduced erroneous high-frequency content when the underlying frequency of the input was low. However, its performance improved as the input frequency increased.

However, a noticeable discrepancy was observed in the recovered latent forces towards the end of the dataset for both the stationary and non-stationary GPLFMs. This mismatch can be attributed to the limitations of the smoother, as fewer future measurements are available for conditioning at later time steps. As a result, as time t increases, the distribution over the latent force tends to converge more closely to the filtering distribution, which in turn reduces the accuracy of the force estimates. It is worth noting that if a longer time series of measurements were provided, such a significant error at 1.5 s would likely not be observed across all models.

To provide a quantitative comparison, NMSE was calculated for each GPLFM model across all states, including the latent force. The summary of these results is presented in Table 6.3. It was observed that the non-stationary GPLFM consistently exhibited the lowest NMSE across all states.

The NMSE values for the observed state velocity $\dot{x}$ and its integral displacement x are low for both models. This reflects the smoothing and filtering algorithms' ability to effectively use the structure of the ODE to reduce measurement noise and accurately integrate/differentiate the observed states. However, since acceleration $\ddot{x}$ cannot be modelled by an ODE independent of the input force $U(t)$, the identification of these states is intrinsically linked and the stationary GPLFMs perform significantly worse than the non-stationary GPLFM. These discrepancies are due to the ability of the non-stationary GPLFM to adapt to the non-stationary frequency of the load.

6.3.2 Bursting Duffing Oscillator

The Duffing oscillator, characterised by its nonlinear cubic stiffness term, is explored under conditions that induce the bursting phenomenon. This phenomenon involves

Table 6.1: Hyperparameters used by the non-stationary GPLFM for the sine sweep case study.

Sine Sweep: Non-Stationary GPLFM	
Hyperparameter	Value
ℓ_{min}	$1\text{x}10^{-3}$
ℓ_{max}	11
ℓ_1	5
σ_1^2	10
ℓ_2	2
σ_2^2	10

Table 6.2: Hyperparameters used by stationary GPLFM 1 and 2 for the sine sweep case study.

Sine Sweep: Stationary GPLFM 1	
Hyperparameter	Value
ℓ	1.87
σ^2	10

Sine sweep: Stationary GPLFM 2	
Hyperparameter	Value
ℓ	0.16
σ^2	10

Table 6.3: Comparison of NMSE across all states and models for the sine sweep case study, relative to simulated data.

GPLFM Model:	Non-Stationary	Length scale = 0.16	Length scale = 1.87
$x\ (m)$	2.00×10^{-4}	2.05×10^{-4}	2.83×10^{-4}
$\dot{x}\ (ms^{-1})$	2.11×10^{-4}	4.22×10^{-4}	2.06×10^{-2}
$\ddot{x}\ (ms^{-2})$	7.90×10^{-2}	2.04×10^{-1}	3.84
$U(t)\ (N)$	1.99×10^{-1}	5.11×10^{-1}	9.65

rapid transitions between different dynamic states, prompted by changes in external forces, leading to abrupt, high-amplitude oscillations. The dynamic equation of the Duffing oscillator is given by:

$$m\ddot{x} + c\dot{x} + kx + k_3x^3 = u(t), \tag{6.16}$$

where k_3 represents the cubic stiffness coefficient. As displacement increases, the system stiffness similarly increases, resulting in a rising natural frequency and a non-stationary spectrum of the temporal restoring force. This characteristic makes the bursting Duffing oscillator an excellent challenge for testing the capabilities of the non-stationary GPLFM.

The system parameters are set with $m = 1$, $c = 0.1$, $k = 1\times10^4$. For the analysis, the equation is excited with a white noise process ($\mu = 0$, $\sigma^2 = 1$) scaled by $\gamma = 100$ and filtered with a fifth-order Butterworth filter. The Duffing oscillator is simulated using a fourth-order Runge-Kutta scheme over $T = 20$s at a sampling rate of 20480Hz. However, only displacement measurements, downsampled to 2048Hz and corrupted by the addition of i.i.d. samples from a zero-mean Gaussian distribution accounting for 1 percent of the RMS of the signal, are provided for identification.

To evaluate the non-stationary GPLFM's application to latent restoring force identification, the functional form of the Duffing oscillator's nonlinearity is hidden, providing a model of the form:

$$m\ddot{x} + c\dot{x} + kx + f(x, \dot{x}) = u(t) \tag{6.17}$$

Identification of the latent restoring force $f(x, \dot{x})$ from noise-corrupted measurements assumes that the initial conditions are known, as well as the functional form of the linear system, the linear system parameters, and the input $u(t)$. Given the system's excitation by white noise, identification can proceed using an OMA scheme without knowledge of the input measurements (see Chapter 5 for this extension).

The identification is run for both stationary and non-stationary GPLFMs. The stationary hyperparameters (those not being estimated by the GP) for these models were selected using an implementation[2] of the Self-Adaptive Differential Evolution

[2]https://github.com/MDCHAMP/FreeLunch

(SADE) heuristic optimiser [143] with the aim of minimising the approximate negitive marginal likelihood [23],

$$\phi_{t+1}(\theta) \approx \frac{1}{2}\log|2\pi S_{t+1}(\theta)| + \frac{1}{2}v_{t+1}^{T}(\theta)S_{t+1}^{-1}(\theta)v_{t+1}(\theta) \tag{6.18}$$

where ϕ denotes the negative log incremental likelihood. The terms $S_{t+1}(\theta)$ and $v_{t+1}(\theta)$ can be calculated using the intermediate steps of the Kalman filter Eq.(6.10a) and Eq.(6.10b) .

The SADE algorithm was executed six times with default parameters, each with a group size of 20, over 50 generations. The best set of hyperparameters across all runs and groups, along with the mean performance for each group, are depicted in Figure 6.2.

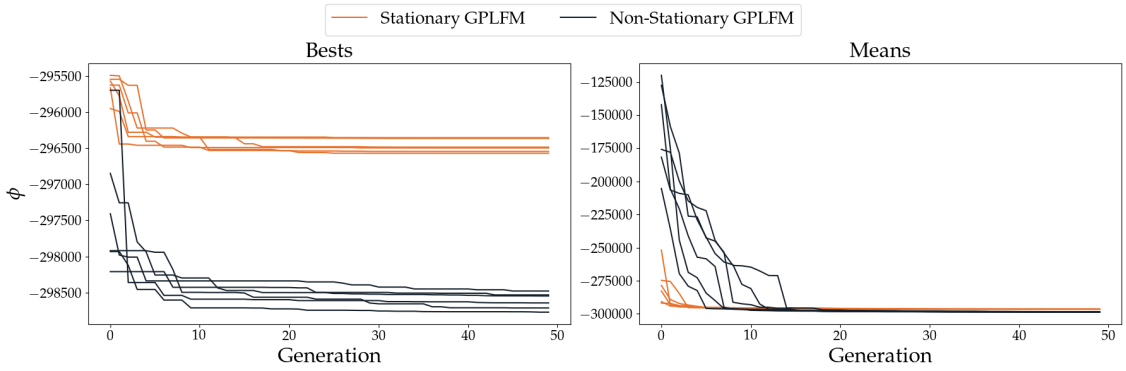

Figure 6.2: Convergence plots of the hyperparameter optimisation for the Duffing oscillator using the SADE algorithm. The convergence plot for the best set of hyperparameters across all runs and groups is present on the left. The Right hand plot shows the mean performance for each group.

It can be observed that the best set of hyperparameters across all runs and groups has not fully converged and still exhibits some variance. However, given that both models have undergone the same number of optimisation runs and the proposed non-stationary GPLFM has a larger optimisation space with four additional hyperparameters, this level of convergence is considered appropriate for comparing the two models. Despite the larger search space all of the best sets of hyperparameters for the non-stationary GPLFM consistently show a lower marginal likelihood compared to those of the stationary GPLFM. This is also true for the mean performance for each group with the non-stationary GPLFM outperforming the stationary GPLFM. This provides an initial indication that the non-stationary GPLFM is better suited

Table 6.4: Optimal hyperparameters for the stationary GPLFM determined using the SADE algorithm.

Bursting Duffing: Stationary GPLFM	
Hyperparameter	Value
ℓ	$8.566\text{x}10^{-3}$
σ^2	$4.861\text{x}10^{1}$

Table 6.5: Optimal hyperparameters for the non-stationary GPLFM determined using the SADE algorithm.

Busting Duffing: Non-Stationary GPLFM	
Hyperparameter	Value
ℓ_{min}	$5.000\text{x}10^{-4}$
ℓ_{max}	2.568
ℓ_1	$7.726\text{x}10^{-1}$
σ_1^2	$1.234\text{x}10^{2}$
ℓ_2	$7.652\text{x}10^{-2}$
σ_2^2	$1.988\text{x}10^{1}$

to this identification problem.

The optimal hyperparameters for the stationary GPLFM and the non-stationary GPLFM are presented in Table 6.4 and Table 6.5 respectively. For the non-stationary GPLFM, the optimal minimum length scale parameter ℓ_{min} was found to be 5.000×10^{-4}. This value represents the lower limit in the optimisation process, as smaller values often lead to numerical instability.

Figure 6.3 illustrates the identification results for the bursting Duffing oscillator using both the stationary and non-stationary GPLFMs with hyperparameters optimised during the study.

It is observed that the stationary GPLFM accurately learns the latent force when the forcing amplitude and frequency are high. However, it encounters difficulties in capturing the latent force at lower amplitudes and frequencies, introducing significant noise to the signal. This phenomenon is more apparent in the zoomed-in view on the left of Figure 6.3. It is evident that the stationary GPLFM does not adequately adapt to the non-stationarity of this identification task with the provided measurements.

It can be understood therefore that the optimisation of the hyperparameters has favoured selecting hyperparameters that learn the latent restoring force well when

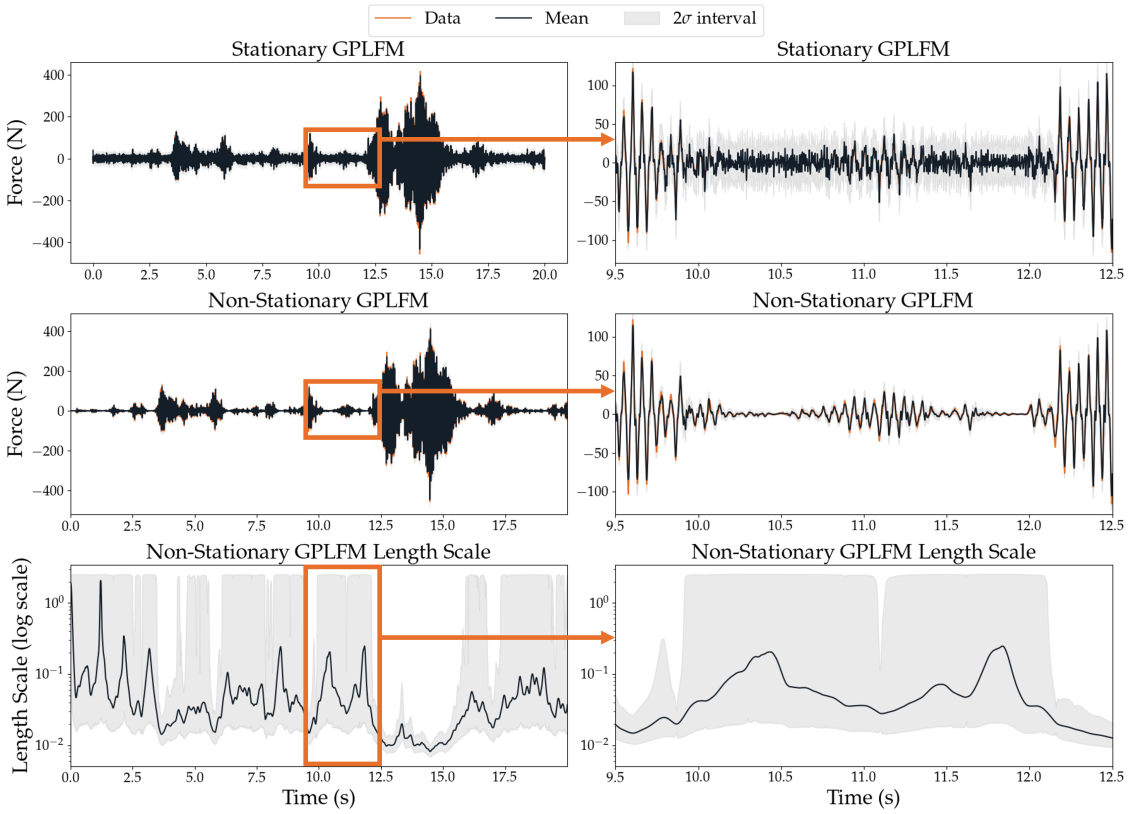

Figure 6.3: Performance comparison of stationary and non-stationary GPLFMs for latent restoring force recovery for the bursting Duffing case study. Top left: Stationary GPLFM showing overall force recovery. Middle left: Non-stationary GPLFM showing enhanced force recovery. Top right: Zoomed view of force recovery from 9.5 to 12.5 seconds for the stationary GPLFM. Middle right: Corresponding zoomed view for the non-stationary GPLFM, highlighting reduces noise over stationary GPLFM. Bottom panels: Length scale dynamics of the non-stationary GPLFM plotted, illustrating adaptation over time with the non-stationarity of the latent force.

Table 6.6: Comparison of NMSE across all states and both models for the bursting Duffing case study, relative to simulated data.

GPLFM Model:	Non-Stationary	Stationary
$x\ (m)$	4.98×10^{-3}	9.56×10^{-4}
$\dot{x}\ (ms^{-1})$	2.54×10^{-2}	2.67×10^{-2}
$\ddot{x}\ (ms^{-2})$	9.58×10^{-2}	1.94×10^{-1}
$f(x, \dot{x})\ (N)$	1.17	2.78

the amplitude and frequency are high, i.e. a shorter length scale and compromised on the identification when the latent force amplitude and frequency is lower. This is expected as over-smoothing the higher amplitude content will lead to a greater error than extra noise in the lower amplitude content.

The performance of the non-stationary GPLFM, as depicted in Figure 6.3, markedly surpasses that of the stationary model. The non-stationary GPLFM adapts well to varying frequencies and amplitudes, correctly identifying high-amplitude and frequency components without over-smoothing, and lower-amplitude and frequency components without introducing non-existent noise. This flexibility is owed to the second GP that learns that length scale. It is clearly seen that the length scale has been appropriately learned, with the length scale shortening as the frequency and amplitude increase and lengthening as both factors decrease.

For a numerical comparison, the NMSE for stationary and non-stationary GPLFMs were calculated across all states, and the results are presented in Table 6.6. A similar pattern is observed as in the sine sweep case study. The NMSEs for displacement x and velocity $\dot{x}$ are minimal for both models and predominantly influenced by the observations and their interaction with the linear SSM and the filtering and smoothing algorithms.

Notably the stationary GPLFM does perform better than the non-stationary GPLFM for recovering the observed state of displacement. The reasoning for this is not well understood. However, the NMSE for both of the models is very low and the recovery of displacement from both models can be considered excellent.

A more significant difference is observed in the acceleration $\ddot{x}$ and the latent restoring force $f(x, \dot{x})$, where the non-stationary GPLFM shows markedly better performance. This indicates that the uniform application of a single length scale is insufficient for optimally modeling non-stationary functions such as the bursting Duffing oscilla-

tor. Rather, when the temporal latent force exhibits a non-stationary frequency spectrum, the length scale ℓ of the GP that learns the latent force should not be a static parameter but should be modeled as a temporal function. This function can be represented as a GP that dynamically adapts to the non-stationary statistical properties of the underlying latent force.

6.4 Summary

In this chapter, a new formulation for a non-stationary GPLFM was proposed. This non-stationary GPLFM was formulated in a hierarchical or "deep" fashion by introducing a time-varying length scale into the GPLFM, which is modeled as an additional GP. Moreover, a routine to solve for the Bayesian filtering and smoothing distribution over the joint latent force-state identification was proposed.

Two case studies were conducted to benchmark the proposed non-stationary GPLFM against the established stationary GPLFM. The first involved an SDOF system subjected to a sine sweep load to assess the handling of non-stationary latent input forces. The second examined a bursting Duffing oscillator to evaluate performance in recovering non-stationary latent restoring forces.

Comparisons between stationary and non-stationary GPLFMs, through graphical representations and NMSE calculations, demonstrated that the stationary GPLFM, with a fixed length scale, either over-fit or over-smoothed, leading to poor latent force recovery under non-stationary conditions. In contrast, the non-stationary GPLFM, by learning a dynamic length scale, adeptly adapted to variations in force frequency, accurately recovering both low- and high-frequency components without over-smoothing or over-fitting. This study demonstrated that a non-stationary GPLFM prior is a more suitable choice than a stationary GPLFM prior when the underlying latent force exhibits significant non-stationarity.

Chapter 7

Probabilistic Numeric SMC Sampling for Bayesian Nonlinear System Identification in Continuous Time

In the previous chapters, the application of the GPLFM for joint state and latent force identification was demonstrated. This approach has been applied not only to recover latent states, input forces, and restoring forces but also provides the distribution over these variables, accounting for uncertainty in the observations used for identification. However, the uncertainty in the model, the states, and the observations is not the only source of uncertainty in systems where the ODE lacks a closed-form solution such as for most nonlinear systems. In such cases, numerical methods must be employed, which are inherently approximate and introduce additional uncertainty into the identification process.

In recent years there has been a growing interest in the uncertainty associated with numerical methods commonly employed across science and engineering. This interest has generated a research area known as *probabilistic numerics* [27, 28]. It has been considered how uncertainty in solvers of linear systems arises [130, 144] or how uncertainty quantification and Bayesian correction can be included in finite element analysis methods [145].

In this chapter, the currently under-exploited interface of probabilistic numerics and (Bayesian, parametric) system identification will be investigated and a novel method for parameter inference respecting the numerical uncertainty in solutions to intractable ODEs will be developed. This methodology will build up several existing approaches notably those for probabilistic solutions to initial value problems of ODEs [146] and sequential Monte Carlo inference for static parameter posteriors [147]. The combination of these approaches presents a new and powerful framework for performing inference in dynamic systems where posteriors over parameters of a continuous but intractable ODE are the objects of interest.

Drawing upon this previous work, a new methodology is introduced to efficiently identify latent states and system parameters from noisy measurements while simultaneously incorporating probabilistic solutions to the ODEs into the identification process. This holistic approach offers several significant benefits.

First, a proper and full treatment of uncertainty is provided by accounting for all sources of uncertainty in the model. Posterior distributions over system parameters are produced, representing both the prior beliefs about the parameters and the inherent uncertainties in the data and in the solution to intractable ODEs.

Second, the complexities associated with non-Gaussian transition densities for non-linear systems in filtering problems and SMC methods are avoided. Moreover, a $q+1$-times integrated Wiener process (IWP) is used in the probabilistic integration process, which extrapolates using polynomial splines of degree $q+1$, seamlessly allowing for higher-order integration.

Finally, systems learned as part of a system identification process are often embedded within a pipeline, followed by subsequent analysis — for example to predict the system's response to new inputs or as part of anomaly detection in SHM. The new methodology allows for a propagation of the full uncertainty in identification.

7.1 Introduction

One of the foremost challenges in system identification involves fusion of theoretical knowledge with measured data to create models that not only embody theoretical understanding but also correspond with the subtleties of real-world data. When the mathematical form of a system is known and experimental data is available, a

practitioner can use parameter estimation as a means to merge these insights into a calibrated model.

In mechanical engineering, structural dynamic systems are typically modeled by second-order differential equations derived from Newton's second law. These can be readily transformed into first-order ODEs by considering the state-space vector dynamics, forming a SSM, as outlined in Chapter 2. Here, the general SSM in continuous time is extended to explicitly show parameter dependence:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_f), \tag{7.1a}$$

$$\mathbf{y}(t) = g(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_g), \tag{7.1b}$$

where, f and g govern the transitions of the state $\mathbf{x}$ and observation $\mathbf{y}$ respectively, parameterised by $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$. In the parameter identification setting, it is assumed that the functional form of f and g are known.

In parametric identification, the parameters $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$ are typically determined through an optimisation procedure aimed at minimising the discrepancy between the predicted and observed system states. For linear, time-invariant SSMs, well-established techniques exist for determining unknown parameters from measured data, as detailed in [148, 149]. Due to the loss of superposition and challenges in analytical integration, the difficulty of parameter estimation is significantly increased for nonlinear systems [56, 58, 150].

Whether the system is linear or nonlinear, a critical challenge in parameter estimation arises from noise in experimental data. Such noise introduces uncertainties that complicate the optimisation process, potentially leading to biased, overconfident, or mis-specified models. If these uncertainties are not adequately addressed, they can degrade the model's precision and predictive accuracy. Bayesian parameter estimation offers a rigorous framework for probabilistic modelling, incorporating uncertainty into the estimation process, thereby enhancing model reliability in noisy environments.

Bayesian parameter estimation is well-established in structural dynamics, where it is used to quantify uncertainty by identifying probability distributions over model parameters, given the noise in the measurements. In the Bayesian framework, model

parameters are treated as random variables. The SSM incorporates these parameters as conditional components, alongside a prior distribution over the parameters $p(\boldsymbol{\theta})$. This prior enables practitioners to incorporate engineering insights into the identification process. The model can be expressed as the following Hidden Markov model:

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \tag{7.2a}$$

$$\mathbf{x}_1 \sim p(\mathbf{x}_1|\boldsymbol{\theta}), \tag{7.2b}$$

$$\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}), \tag{7.2c}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta}). \tag{7.2d}$$

The 'full' Bayesian solution requires computing the joint posterior $p(\mathbf{x}_{1:T}, \boldsymbol{\theta} \mid \boldsymbol{y}_{1:T})$, a task which is almost always intractable. Recursive algorithms offer a viable alternative, providing a means to approximate the marginal posterior $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$. When parameters $\boldsymbol{\theta}$ are held constant, Bayesian recursive filtering yields the distribution:

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto p(\boldsymbol{\theta}) \prod_{t=1}^{T} p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}), \tag{7.3}$$

Under this framework, a parameter estimation method was proposed by Chopin in the form of Iterated Batch Importance Sampling (IBIS) [147]. This approach involves the sequential sampling and reweighting of $\boldsymbol{\theta}$ values, guided by the likelihood increments $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta)$ as defined in Eq.(7.3), with the $\boldsymbol{\theta}$-particles being updated via resampling and MCMC techniques [151]. Chopin extends this methodology to the sequential Monte Carlo squared (SMC^2) algorithm, catering to scenarios where likelihood increments are intractable in closed form [152]. This method propagates multiple particle filters within the x-space, in tandem with SMC using MCMC moves in the θ-space. These methods lay the foundation for the parameter estimation techniques employed in this research.

A critical step in all these methodologies is the evaluation of trial parameters through the unnormalised posterior likelihood, which involves comparing predicted states, contingent on the parameters, against observed data. For most nonlinear systems,

the absence of a closed-form solution necessitates numerical approximation for this prediction step. Since the numerical solution to the system state prediction must be approximate, so must be the evaluation of the quality of the parameters made from the prediction. As a result, any 'optimal' parameter set is not truly unique, but is instead conditional on the specific errors introduced by the chosen numerical integration method. Hence, the identification of parameters for nonlinear systems is intrinsically uncertain, necessitating a probabilistic approach to capture this uncertainty. Viewing numerical approximations to integration as fundamentally probabilistic forms one of the foundations of *Probabilistic Numerics* (PN) [28].

PN explores the intersection of numerical analysis and probability theory, advocating for a statistical approach to computational problems. It operates on the premise that many numerical tasks—such as solving problems in linear algebra [130], quadrature [153], optimisation [154], ODEs [155], and PDEs [156]—lack exact solutions and therefore introduce uncertainty. PN addresses this by treating numerical solvers as agents that can quantify uncertainty with probability measures, allowing for richer outputs than traditional point estimates. This probabilistic approach enables smarter, uncertainty-aware decisions within algorithms and allows for the encoding of less-than-certain expectations into solvers.

Notably, within the field of PN Tronarp et al. [146] developed a method for probabilistic solutions to ODE. In their work, the integration of an ODE is reformulated as a nonlinear Bayesian filtering and smoothing problem. Here, the proposal $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is modelled as an integrated Wiener process, and the likelihood is defined by conditioning on a *pseudo-measurement* $\mathbf{Z}_t = \mathbf{z}(t) \stackrel{\triangle}{=} 0$, $\forall t$ that captures the derivative relationship of the ODE. Consequently, the posterior $p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{Z}_t)$ represents the probability of the state given the noise in the measurements and the uncertainty in the numerical integration, thus incorporating the integration uncertainty into the state estimation process.

Despite much work by many authors across the fields of uncertainty quantification in parameter estimation and numerical methods, little work has been done to combine these two areas. Schmidt et al [157] developed a probabilistic SSM for joint inference from ODES and data able to infer the system state and latent function as a temporal Gaussian Process (GP) given the uncertainty in the data and numerical methods in a single filer update. Tronarp et al [158] developed the Fenrir algorithm for a reframing of the state and parameter estimation into a Gauss–Markov process. This methodology initiates by refining a standard Gauss-Markov prior into a physics-

informed prior via PN. Subsequently, a posterior distribution is calculated using Gauss-Markov regression. Parameter estimates for the dynamical system are then extracted by maximising the marginal likelihood, resulting in a calibrated posterior distribution.

The aim of this work is to unify Bayesian parameter estimation and PN to establish a comprehensive probabilistic framework for parameter estimation in nonlinear systems. By explicitly incorporating the uncertainty in both measurement data and numerical computations, this unified approach will identify the posterior over the states given the uncertainty in the measurement and numerical integration as $p(\boldsymbol{\theta}|\mathbf{y}_t, \mathbf{Z}_t)$ which has not been done before. The framework is designed to facilitate informed decision-making by providing a more complete assessment of uncertainty, thereby enabling engineers to consider the probabilistic nature of their models and to make risk-aware choices in complex engineering applications.

7.2 Probabilistic Solutions to Ordinary Differential Equations

In this section, a methodology for obtaining probabilistic solutions to ODEs is introduced by formulating the initial value problem (IVP) as a nonlinear Bayesian filtering problem. Through this approach, a framework is established to recover the distribution over the states, accounting for the uncertainty inherent in solving the IVP. It will later be demonstrated how this framework can be used to propagate the uncertainty in numerical integration into a Bayesian parameter identification scheme.

7.2.1 From Ordinary to Stochastic Differential Equations

The IVP was introduced in Section 2.3. As discussed earlier, an IVP describes the evolution of a system governed by a differential equation, where the system's state is specified at an initial time. The task is to predict the system's future behaviour by solving the differential equation, starting from this known initial condition and following the dynamics prescribed by the ODE. For a system described by a deterministic ODE, the solution to the state evolution can be expressed as an integral,

$$\mathbf{x}(t) = \int_{t_0}^{t} f(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_f)\, dt, \tag{7.4}$$

However, for almost all nonlinear systems, a closed-form solution to the integral in Equation (7.4) is not available. Solving such IVPs requires numerical methods, as analytical solutions are typically infeasible.

One such method is Euler's method, which discretises time into small increments h for state progression:

$$\mathbf{x}(t+h) \approx \mathbf{x}(t) + h \cdot f(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_f). \tag{7.5}$$

Euler's method is a first order method with local error proportional to the square of h and global error linearly proportional to h such that as h approaches zero the solution to Euler's method approaches the true solution to the integral [159]. However, due to computational limitations or constraints imposed by the sampled frequency of $\mathbf{u}(t)$ it is not always possible to reduce step size to reduce error to an negligible levels. Whilst, this can in part be combated by the use of higher order methods it can sometimes not be enough or too challenging to implement. Under these circumstances, the resulting drift in state predictions can become significant, leading to uncertainty about the true solution of the integral Eq.(7.4).

It maybe natural therefore, to consider numerical integration as a linear SDE composed of a linear ODE to represent local linearisation, augmented with a random variable accounting for the unknown error in the integration process. The continuous-time SDE equation is expressed as:

$$\mathbf{X}(1) \sim \mathcal{N}(\boldsymbol{\mu}(1), \boldsymbol{\Sigma}(1)), \tag{7.6a}$$

$$d\mathbf{X}(t) = [\mathbf{F}\mathbf{X}(t) + \mathbf{u}(t)]dt + \mathbf{L}d\boldsymbol{\beta}(t), \tag{7.6b}$$

where $\boldsymbol{\mu}(1)$ and $\boldsymbol{\Sigma}(1)$ are the mean and the covariance that describe the Gaussian distribution over the initial conditions $\mathbf{X}(1)$. $\mathbf{F}$ is the state transition matrix, $\mathbf{u}(t)$

the force, $\mathbf{L}$ is the diffusion matrix and $\boldsymbol{\beta}(t)$ is defined as a vector of the standard Wiener processes. $X(t)$ is a vector of $\mathbf{X}^{(1)}(t)$ and $q+1$ derivatives such that $\mathbf{X}^{i+1}(t)$ is the derivative of $\mathbf{X}^{i}(t)$,

$$\mathbf{X}(t) = \begin{bmatrix} \mathbf{X}^{(1)}(t) \\ \mathbf{X}^{(2)}(t) \\ \vdots \\ \mathbf{X}^{(q+1)}(t) \end{bmatrix} \tag{7.7}$$

It has been established in Chapter 3 that filtering and smoothing on SDEs of this nature are equivalent to GP regression, where the SDE effectively forms the GP prior [106]. In this context, any given realisation of the SDE corresponds to a sample from the GP prior.

To fit the SDE in a Bayesian filtering framework measurements are required. For solving IVPs, Tronarp [146] proposes that a measurement can be defined upon the known derivative relationship between the states i.e. $\mathbf{F}$, $\mathbf{u}$ and $\mathbf{L}$ must be set such that $\mathbf{X}^{i+1}(t)$ is always be the derivative of $\mathbf{X}^{i}(t)$. This provides the following *pseudo-measurement*

$$\mathbf{Z}(t) = \begin{bmatrix} \mathbf{X}^{(2)} \\ \mathbf{X}^{(3)} \\ \vdots \\ \mathbf{X}^{(q+1)} \end{bmatrix} - f\left(\begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(q)} \end{bmatrix} \right) = \mathbf{0} \tag{7.8}$$

In this framework, the SDE is conditioned on the residual relationship, as expressed in Eq.(7.8), equating to zero. For example, at any moment, the velocity must equal the derivative (as calculated from the continuous SSM Eq.(7.1)) of displacement. Any deviation from this relationship indicates an integration error. However, conditioning the process $\mathbf{X}(t)$ on $\mathbf{z}(t) = 0$ for $t \in [1, T]$ is intractable in continuous time, necessitating the adoption of a discrete time approach.

7.2.2 The Discrete Time Solution

The discrete formulation only attempts to condition the process $\mathbf{X}(t)$ on $\mathbf{Z}(t) = \mathbf{z}(t) \triangleq 0$ at a set of discrete time-points, $\{t_{1:T}\}$. Under this Bayesian filtering framework the inference problem becomes

$$\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \tag{7.9a}$$

$$\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathcal{N}(\mathbf{A}(h)\mathbf{X}_t + \boldsymbol{\xi}(h), \mathbf{Q}(h)), \tag{7.9b}$$

$$\mathbf{Z}_t|\mathbf{X}_t \sim \mathcal{N}(\dot{\mathbf{C}}\mathbf{X}_t - f(\mathbf{C}\mathbf{X}_t, \mathbf{u}(t)), \mathbf{R}), \tag{7.9c}$$

$$\mathbf{z}_t \triangleq \mathbf{0}, \quad t = 1, \ldots, T, \tag{7.9d}$$

where h is the step size and $\mathbf{R}$ represents the measurement variance. In this work and in general when solving Eq.(7.9) with Gaussian filtering $\mathbf{R} = 0$. $\mathbf{C}$ is the observation matrix and $\dot{\mathbf{C}}$ is it's derivative such that $\mathbf{C} = \begin{bmatrix}\mathbf{I} & 0 & \cdots & 0\end{bmatrix}$ and $\dot{C} = \begin{bmatrix}0 & \mathbf{I} & \cdots & 0\end{bmatrix}$. That is, $\mathbf{C}\mathbf{X}_t = \mathbf{X}_t^{(1)}$ and $\dot{\mathbf{C}}\mathbf{X}_t = \mathbf{X}_t^{(2)}$. It is important to observe that the likelihood model $p(\mathbf{Z}_t \mid \mathbf{X}_t)$ exhibits nonlinearity, which adds complexity to the filtering solution. Additionally, $\mathbf{z}_t$ denotes the realisation of $\mathbf{Z}_t$, and the state transition $\mathbf{A}(h)$, process noise $\boldsymbol{\xi}(h)$, and process noise covariance $\mathbf{Q}(h)$ are defined as follows:

$$\mathbf{A}(h) = \exp(\mathbf{F}h), \tag{7.10a}$$

$$\boldsymbol{\xi}(h) = \int_1^h \exp(\mathbf{F}(h-\tau))\mathbf{u}d\tau, \tag{7.10b}$$

$$\mathbf{Q}(h) = \int_1^h \exp(\mathbf{F}(h-\tau))\mathbf{L}\mathbf{L}^T \exp(\mathbf{F}^T(h-\tau))d\tau. \tag{7.10c}$$

7.2.3 Building the Model

A commonly adopted prior for ODE solvers is the IWP. Specifically, the $q+1$ times IWP, denoted as IWP$(q+1)$, is used due to its capability to extrapolate using

polynomial splines of degree $q+1$ [28]. This particular choice, IWP$(q+1)$, facilitates the computation of integrals in Eq.(7.10) in closed form, as follows:

$$\mathbf{A}(h) = \mathbf{A}^{(1)}(h) \otimes \mathbf{I}, \tag{7.11a}$$

$$\boldsymbol{\xi}(h) = \mathbf{0}, \tag{7.11b}$$

$$\mathbf{Q}(h) = \mathbf{Q}^{(1)}(h) \otimes \boldsymbol{\Gamma}, \tag{7.11c}$$

where $\otimes$ is the Kronecker product, $\mathbf{I} \in \mathbb{R}^{d \times d}$ is the identity matrix and $\boldsymbol{\Gamma} \in \mathbb{R}^{d \times d}$ is a hyperparameter that calibrates the covariance in $\mathbf{Q}(h)$. $\mathbf{A}^{(1)}(h)$ and $\mathbf{Q}^{(1)}(h)$ are given by

$$\mathbf{A}(h)_{ij}^{(1)} = \mathbb{I}_{i \leq j} \frac{h^{j-i}}{(j-i)!} \tag{7.12a}$$

$$\mathbf{Q}(h)_{ij}^{(1)} = \frac{h^{2q+3-i-j}}{(2q+3-i-j)(q+1-i)!(q+1-j)!} \tag{7.12b}$$

where $\mathbb{I}$ is the indicator function.

7.2.4 Filtering Solution

Eq.(7.9) defines a Bayesian filtering problem. In Section 2.6, the Kalman filter and Extended Kalman filter were introduced. Here, the system is linear in the prediction step and nonlinear in the observation step, which necessitates a slightly different implementation. The linear prediction step is given by:

$$\boldsymbol{\mu}_{t+1}^{P} = \mathbf{A}(h)\boldsymbol{\mu}_{t}^{F} + \boldsymbol{\xi}(h), \tag{7.13a}$$

$$\boldsymbol{\Sigma}_{t+1}^{P} = \mathbf{A}(h)\boldsymbol{\Sigma}_{t}^{F}\mathbf{A}^{T}(h) + \mathbf{Q}(h), \tag{7.13b}$$

where $boldsymbol\mu_t^F$ and $\boldsymbol{\Sigma}_t^F$ are the first and second filtering moments and $\boldsymbol{\mu}_t^P$ and

$\boldsymbol{\Sigma}_t^P$ are the first and second predictive moments.

The nonlinear observation model can be approximated using Taylor series methods. The zeroth order extended Kalman filter (EKF0) used in this work because global uncertainty quantification can be very changing for higher order methods when the ODE is multidimensional such that $d > 1$ (see Section 7.4.1 for an example). The filter update for the EKF0 is given as;

$$\mathbf{S}_t \approx \dot{\mathbf{C}}\boldsymbol{\Sigma}_t^P\dot{\mathbf{C}}^T + \mathbf{R}, \tag{7.14a}$$

$$\mathbf{K}_t \approx \boldsymbol{\Sigma}_t^P\dot{\mathbf{C}}^T\mathbf{S}_t^{-1}, \tag{7.14b}$$

$$\hat{\mathbf{z}}_t \approx \dot{\mathbf{C}}\boldsymbol{\mu}_t^P - f(\mathbf{C}\boldsymbol{\mu}_t^P, t_t), \tag{7.14c}$$

$$\boldsymbol{\mu}_t^F \approx \boldsymbol{\mu}_t^P + \mathbf{K}_t(\mathbf{z}_t - \hat{\mathbf{z}}_t), \tag{7.14d}$$

$$\boldsymbol{\Sigma}_t^F \approx \boldsymbol{\Sigma}_t^P - \mathbf{K}_t\mathbf{S}_t\mathbf{K}_t^T. \tag{7.14e}$$

It is important to note that for enhanced numerical stability, the implementation of these filters using the square root formulation is recommended. A derivation of the square root Kalman filter is available in [160].

7.2.5 Calibration

The validity of the posterior distribution over the system states depends not only on its mean but also crucially on its variance. In the probabilistic ODE solver, this variance is governed by the hyperparameters of the prior. For the IWP prior used here, the key hyperparameter is $\boldsymbol{\Gamma}$ as it governs the process noise covariance $\boldsymbol{Q}(h)$ (see Eg.(7.13b)).

The calibration of $\boldsymbol{\Gamma}$ is crucial for producing a reliable posterior. Since $\boldsymbol{\Gamma}$ in the EKF0 does not depend on the vector field of the function $f(\cdot)$ of the ODE it is proposed by Bosch et al. [161] that $\boldsymbol{\Gamma}$ should be a diagonal matrix $\boldsymbol{\Gamma} = \text{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ so that the variance σ^2 can be individually calibrated for each dimension d.

$\boldsymbol{\Gamma}$ can be optimised by maximising the marginal likelihood, which involves selecting σ^2 that maximise the evidence for the observed data. The marginal likelihood is

given by:

$$p(\mathbf{z}_{1:T}|\sigma^2) = p(\mathbf{z}_1|\sigma^2) \prod_{t=2}^{Y} p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \sigma^2). \tag{7.15}$$

However, computing this is as costly as solving the ODE, thus necessitating an approximation. Bosch et al. [161] proposed a quasi-ML estimator for multidimensional ODEs in the case of the EKF0, expressed as:

$$\hat{\Gamma}_{ii} = \frac{1}{T} \sum_{t=1}^{T} \frac{\left(\hat{\mathbf{z}}_t\right)_i^2}{\breve{s}_t}, \quad i \in \{1, \ldots, d\}. \tag{7.16}$$

where $\mathbf{S}_t = \mathbf{\Gamma}\breve{s}_t$ and $\mathbf{S}_t$ is the innovation covariance matrix from the EKF0 equations Eq.(7.14). This estimator, while an approximation, offers efficient calibration for the variance of the posterior and can be easily Incorporated into the EKF0 for almost no additional cost.

7.3 Sequential Monte Carlo for parameter estimation

In the previous section, it was shown how a probabilistic solutions to ODEs can be formulated to account for the inherent uncertainty arising from the need to approximate solutions to nonlinear ODEs—where no closed-form solution exists. However, for a full and proper treatment of uncertainty in parameter identification, a methodology must be developed that incorporates the numerical uncertainty from solving the ODE, the prior distribution over the parameters and the uncertainty in the measurements. To this end, Monte Carlo-based methods can be adapted to form a holistic identification approach.

Monte Carlo-based methods are numerical techniques that leverage repeated random sampling to approximate probability distributions [151]. Among these, the Iterated Batch Importance Sampling (IBIS) method [147] is designed to estimate posterior distributions of the form $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$. The IBIS algorithm comprises two

principal stages: firstly, evaluating the efficacy of a set of trial parameters $\boldsymbol{\theta}_{1:n}$, and secondly, proposing new parameters based on the assessed quality of the previous set. The innovation in IBIS is to use *importance sampling* and *particle rejuvenation* to explore partial distributions $p(\boldsymbol{\theta}|y_{1:t})$ (where $t < T$) to efficiently explore the parameter space and evaluate the posterior $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$. This section outlines the IBIS methodology, which evaluates a posterior distribution over parameters given the uncertain measurements $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$, and demonstrates how it can be used to develop a new approach that also accounts for the uncertainty in numerical integration to evaluate posteriors of the form $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{Z})$. This approach is outlined in Algorithm 1.

Algorithm 5 Particle System Algorithm

Generate a particle system $\{\boldsymbol{\theta}_n, w_n\}$ that targets the prior distribution $p(\boldsymbol{\theta}_n)$
for $t = 1$ to T **do**
  for $n = 1$ to N **do**
    Evaluate $p(\mathbf{X}_{t+1,n} \mid \mathbf{Z}_t, \boldsymbol{\theta}_n)$ according to Eq. (7.9)
    Compute $\phi_n(\boldsymbol{\theta}_n)$ according to the energy function in Eq. (7.17)
    Update the particle system weights in $\{\boldsymbol{\theta}_n, w_n\}$ as per Eq. (7.18)
  end for
  if ESS, as in Eq. (7.19), $< \epsilon$ **then**
    Sample a new particle system $\{\tilde{\boldsymbol{\theta}}_m, \tilde{w}_m\}_{m=1:N}$ using an Independent Metropolis Hastings proposal in Eq. (7.20)
    for $s = 1$ to t **do**
      for $m = 1$ to N **do**
        Evaluate $p(\mathbf{X}_{s+1,m} \mid \mathbf{Z}_s, \boldsymbol{\theta}_m)$ according to Eq. (7.9)
        Compute $\tilde{\phi}_m(\tilde{\boldsymbol{\theta}}_m)$ according to the energy function in Eq. (7.17)
        Update the particle system weights in $\{\tilde{\boldsymbol{\theta}}_m, \tilde{w}_m\}$ as per Eq. (7.18)
      end for
      Accept/reject $\{\tilde{\boldsymbol{\theta}}_m\}$ with probability given in Eq. (7.22)
    end for
  end if
end for

This inference scheme initiates at t_1, where the parameter distribution is represented by the prior $p(\boldsymbol{\theta})$, as informed by the practitioner's engineering knowledge. Direct conditioning of $p(\boldsymbol{\theta})$ on $\mathbf{y}_{1:T}$ is computationally intractable. Therefore, $p(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ will be approximated by conditioning on a discrete set of random samples $\boldsymbol{\theta}_{1:N} \sim p(\boldsymbol{\theta})$ (Algorithm 1.1). Using these sampled particles, initial conditions $\mathbf{X}_{t_1,1:N}$, and the probabilistic numerical integrator (referenced in Section 7.2), the distribution over the system states given the prior over the parameters and the uncertainty in the numerical integration at subsequent time steps $(\mathbf{X}_{t+1} \mid \mathbf{Z}_t, \boldsymbol{\theta}_n)_{n=1:N}$ can be

determined (Algorithm 1.3).

The quality of the trial parameters can be assessed by sequentially comparing states at $\mathbf{X}_{t+1}$ predicted by $f(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_n)$ and evaluated by the probabilistic ODE solver to the noisy measurements $\mathbf{y}_{t+1}$ through the state space measurement model Eq.(7.1b) [1]. The comparison is made by computing the approximate energy function [23] (Algorithm 1.4),

$$\phi_{t+1}(\boldsymbol{\theta}) = \frac{1}{2}\log|2\pi\mathbf{S}_{t+1}(\boldsymbol{\theta})| + \frac{1}{2}\mathbf{v}_{t+1}^T(\boldsymbol{\theta})\mathbf{S}_{t+1}^{-1}(\boldsymbol{\theta})\mathbf{v}_{t+1}(\boldsymbol{\theta}) \tag{7.17}$$

where ϕ denotes the negative log incremental likelihood or negative log incremental weight $p(\mathbf{y}_{t+1}|\mathbf{y}_t, \boldsymbol{\theta}_n, \mathbf{Z})$. The terms $\mathbf{S}_{t+1}(\boldsymbol{\theta})$ and $\mathbf{v}_{t+1}(\boldsymbol{\theta})$ can be calculated using the intermediate steps of the Kalman filter [23]. The energy function encapsulates the quality of the proposed parameters, considering the integration uncertainties and the influence of noisy observations. The unnormalised marginal likelihood $p(\mathbf{y}_{t+1}|\boldsymbol{\theta}_n, \mathbf{Z})$ can be approximated from

$$\log w_{t+1}(\boldsymbol{\theta}_n) = \log w_t(\boldsymbol{\theta}_n) - \phi_{t+1}(\boldsymbol{\theta}_n) \tag{7.18}$$

where w is the approximate the unnormalised weight and $w_{t=1} = 1/N$ such that the update transitions the unnormalised weight from $p(\mathbf{y}_t|\theta_n, \mathbf{Z}_t)$ to $p(\mathbf{y}_{t+1}|\boldsymbol{\theta}_n, \mathbf{Z}_{t+1})$ (Algorithm 1.5) and $w_T(\boldsymbol{\theta}_n) \propto p(\mathbf{y}_{1:T}|\boldsymbol{\theta}_n, \mathbf{Z}_{1:T})$.

The particle system represents $p(\theta|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ through a weighted set of particles, rather than directly yielding a set of particles whose distribution intrinsically approximates $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ itself. As the algorithm progresses through its prediction and reweighting steps (as outlined in Algorithms 1.3 to 1.5), it accumulates more information about the quality of the proposed particles $\boldsymbol{\theta}_{1:N}$. This process causes the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ to diverge from the initial distribution $p(\boldsymbol{\theta}|\mathbf{y}_{t=0})$. A common side effect of this process is that the particles become degenerate. This degeneracy primarily arises because, with each sequential observation, the likelihood for most particles diminishes significantly compared to a few that align closely with

[1] Note that the measurement $\mathbf{y}_{t+1}$ is not used in the filtering step of the probabilistic ODE solver as this would impede the ability to test the quality of the parameters $\boldsymbol{\theta}_n$.

the observed data, resulting in an imbalance where a few particles end up with the majority of the weight. To counteract this issue, resampling and rejuvenation steps can be implemented. Resampling effectively duplicates particles with higher weights and eliminating those with lower weights so that the set of unweighted particles approximate the distribution $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ its self. Therefore after resampling each particle weight $w_{1:N}$ is set to $1/N$.

Rejuvenation, in contrast, involves the introduction of a move step to the particles post-resampling to avert the loss of diversity and facilitate the exploration of the parameter space. This process entails slightly altering each particle based on a move kernel. This move is then accepted or rejected according to a criterion that ensures the overall particle set continues to approximate $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ accurately.

Efficiency in rejuvenation is critical as it is often computationally demanding, requiring a complete browsing of all past observations (Algorithm 1.9 to 1.11). Efficiency here comes from two parts. First, only rejuvenating the particles $\boldsymbol{\theta}_{1:N}$ when $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ is not well represented by the particles. To ensure this a standard degeneracy criterion in used,

$$ESS = \frac{(\sum_{j=1}^{N} w_j)^2}{\sum_{j=1}^{N} w_j^2} \tag{7.19}$$

where ESS is the effective sample size. The ESS can be compared to a user determined threshold ϵ (Algorithm 1.6). If the ESS drops below the threshold the particles are determined to be functionally degenerate and will be resampled and rejuvenated . Therefore, the threshold needs to be set at some compromise between propagating particles that can continue to infer useful information about the proposed parameters and not resampling too often and incurring unnecessary computational costs.

The second factor to consider when managing the computational cost of the algorithm is the efficiency of the move kernel itself. Since a theoretical guarantee of convergence does not guarantee that particles will explore the search space efficiently it is key to select a move kernel with this in mind.

For the move kernel to have a good efficiency it is important that the acceptance rate of new parameters is high. This could be easily achieved artificially via a random walk where a small perturbation to current particles is applied. However, this fails

to actually rejuvenate the particles as although particles are distinct they are also very similar and therefore very highly correlated. This approach leaves degeneracy high but less detectable.

Therefore, a move kernel should be chosen that only proposes new particles that weakly depend on the previous values. For this the independent Metropolis Hasting (IMH) kernel is chosen [152]. This ensures that the acceptance rate becomes better indicator of rejuvenation.

Since the $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ is approximated by the weighted particles, resampling and rejuvenation can be performed by sampling from that distribution. A rough approximation of the location of the mass of $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:t}, \mathbf{Z}_{1:t})$ is given by the expectation $\hat{\mathbf{E}}$ and variance $\hat{\mathbf{V}}$ of the particle system,

$$\hat{\mathbf{E}} = \frac{\sum_{j=1}^{N} w_j \boldsymbol{\theta}_j}{\sum_{j=1}^{N_\theta} w_j}, \quad \hat{\mathbf{V}} = \frac{\sum_{j=1}^{N} w_j \{\boldsymbol{\theta}_j - \hat{\mathbf{E}}\}\{\boldsymbol{\theta}_j - \hat{\mathbf{E}}\}'}{\sum_{j=1}^{N_\theta} w_j}, \tag{7.20}$$

such that samples can be drawn (Algorithm 1.7),

$$\tilde{\boldsymbol{\theta}} \sim \mathcal{N}(\hat{\mathbf{E}}, \hat{\mathbf{V}}) \tag{7.21}$$

After new parameters are proposed a complete browsing of observations from the initial time step up to the current must be performed so that the performance of the newly proposed parameters can be compared to that of the old parameters (Algorithm 1.9 to 1.11). This involves calculating the distribution over the parameters of the system given the uncertainty in the measured data and the numerical integration iteratively from time t_1 to t_{t+1}.

With the weights of the proposed particles calculated the move is accepted with probability (Algorithm 1.12),

$$\alpha = min\left(1, \frac{p(\tilde{\boldsymbol{\theta}})p(\mathbf{y}_t \mid \tilde{\boldsymbol{\theta}}, \mathbf{Z}_t)}{p(\boldsymbol{\theta})p(\mathbf{y}_t \mid \boldsymbol{\theta}, \mathbf{Z}_t)}\right) \tag{7.22}$$

The employment of the IMH kernel within SMC methods serves a dual purpose: resampling and rejuvenation of the particle set. This dual functionality stems from the kernel's capability to sample from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}\mathbf{Z}_{1:t})$ as approximated by the weighted particles. Such sampling not only ensures that the distribution of particles closely approximates the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{1:t}, Z_{1:t})$ itself but also facilitates an effective exploration of the parameter space by the particles. Finally, the newly resampled and rejuvenated particle system is input back into the outer loop and the particle weights are reset to $1/N$ so that the process of evaluation given new observations can continue until rejuvenation is once again required. This process is repeated for each of the the discrete time steps in the data set (Algorithm 1.12).

7.4 Case studies

In this section, three case studies are presented to evaluate the effectiveness of the proposed methodology. The first case study involves a simulation of the Bouc-Wen model of hysteresis. Here, the parameter estimation by the proposed method is compared against a known ground truth, facilitating a clear assessment of its accuracy.

The latter two case studies are based on experimental datasets: the Silver Box and an Electromechanical System (EMS). These present a greater challenge in quantifying identification accuracy, due to the absence of a predetermined ground truth. However, these cases can be considered more representative of real-world system identification challenges. Their inclusion in this study is intended to demonstrate the practical applicability and adaptability of the proposed methodology in a experimental settings.

7.4.1 Bouc Wen

The Bouc Wen model of hysteresis was intoduced in Section 4.1. As a recap the equation of motion for the Bouc Wen model of hysteresis [124] is given as,

Table 7.1: Prior parameter distributions for the Bouc Wen system.

Prior	**Distribution**
$p(m)$	$\mathcal{N}(2.1, 0.011)$
$p(c)$	$\mathcal{N}(8.8, 6.97)$
$p(k)$	$\mathcal{N}(5.9 \times 10^4, 2.18 \times 10^8)$
(α)	$\mathcal{N}(4.4 \times 10^4, 1.74 \times 10^8)$
$p(\beta)$	$\mathcal{N}(8.6 \times 10^2, 6.66 \times 10^4)$
$p(\gamma)$	$\mathcal{N}(0.93, 0.0541)$
$p(\delta)$	$\mathcal{N}(1.3, 0.1056)$

$$m\ddot{x} + c\dot{x} + ky + z(x, \dot{x}) = u(t) \tag{7.23a}$$

$$\dot{z}(x, \dot{x}) = \alpha\dot{x} - \beta\left(\gamma|\dot{x}||z|^{\nu-1}z + \delta\dot{x}|z|^{\nu}\right) \tag{7.23b}$$

where, x is displacement, $\dot{x}$ is velocity, m is mass, c is viscous damping, k is linear stiffness and z encodes the nonlinear hysteresis memory effect. The rate of change of z, $\dot{z}$ is defined by α, β, γ, δ and ν which are used to tune the shape and smoothness of the hysteresis loop.

For training data generation, a system based on Eq.(7.23) is simulated using a fourth-order Runge-Kutta scheme. The parameters for this simulation are detailed in Table 7.1. A random phase multisine load, with frequencies ranging from 0.5 Hz to 100 Hz across 2000 uniformly spaced steps, is applied to the system. The load's amplitude is initially increased linearly over the first 10% of the simulation time, and subsequently has maximum amplitude of 208N. The simulation is conducted for 3 seconds at a sampling frequency of 131072 Hz ensuring minimal integration error. However, the only data that will be used for the identification will be acceleration measurements down sampled to 4096 Hz to provide 12288 data points. The data is then corrupted with measurement noise through the addition of *i.i.d.* samples from a Gaussian distribution with zero mean and a standard deviation of 5% of the root mean square of the acceleration.

Using the framework for probabilistic parameter identification outlined in this work the posterior parameter distribution is identified from a prior distribution over the parameters, the nonlinear SSM, a known forcing and a noisy measurement of the acceleration state. Figure 7.1 shows the prior and posterior distributions over the

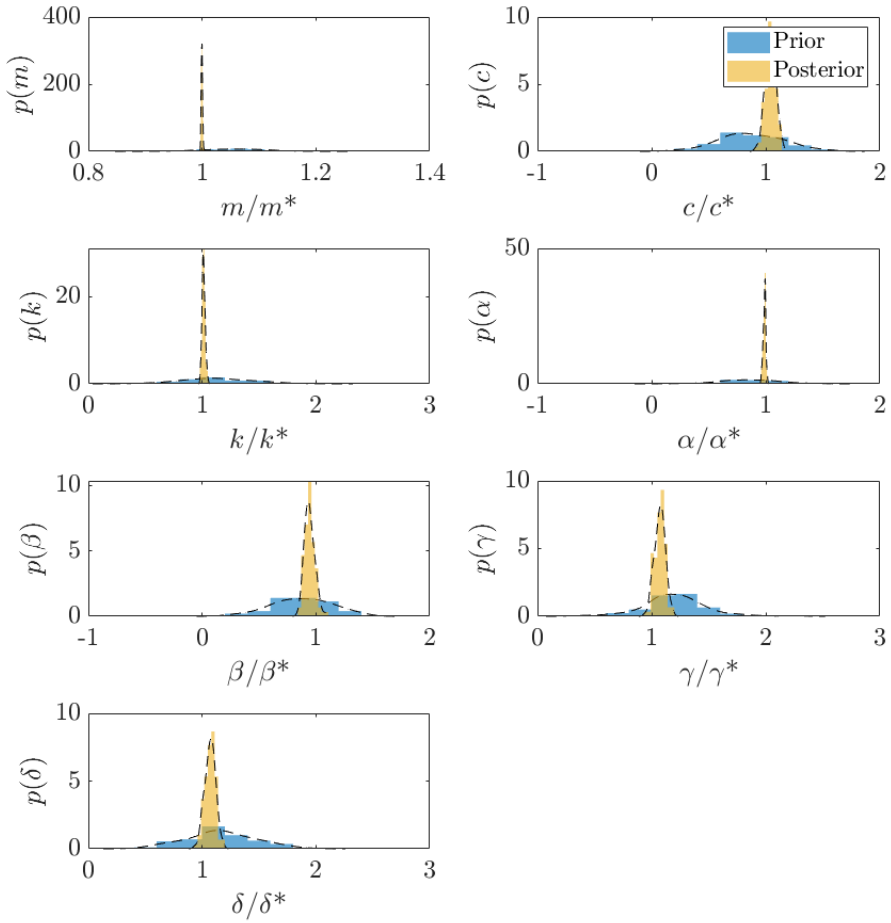

Figure 7.1: Prior and posterior histograms and PDFs normalised by the true parameter values for the Bouc Wen system (denoted by $(\cdot)^*$) .

parameters before and after training. It should be noted that Figure 7.1 shows the parameters normalised by the ground truth so that the quality of the identification can be more easily interpreted.

The prior means and variances can be found in Table 7.1. The prior was defined by perturbing the true values. Large variances are given to each of the parameters to imitated a scenario where a practitioner has low confidence in the prior values defined.

Figure 7.1 illustrates a discernible variation in both the accuracy and uncertainty of the posterior parameter estimates. Specifically, the parameters m, k, and α

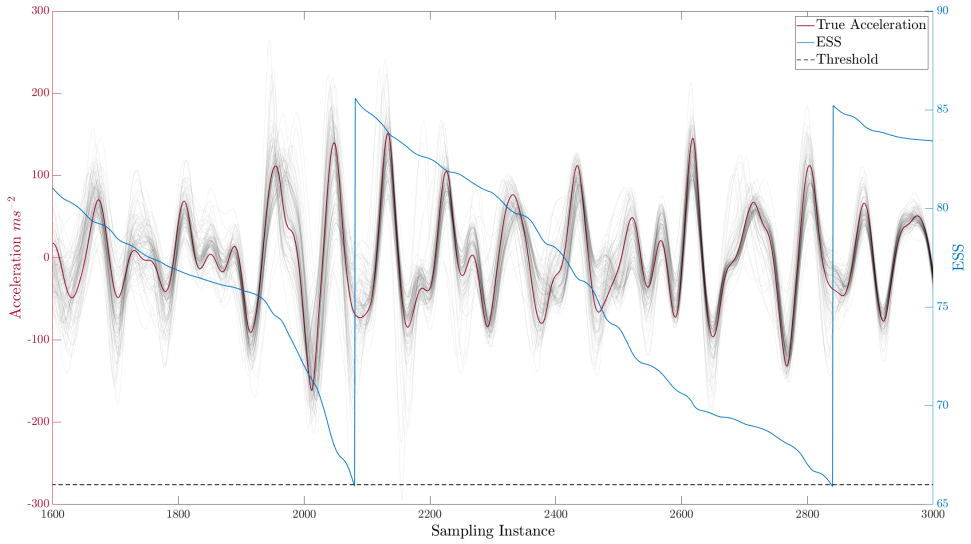

Figure 7.2: Acceleration samples from the filtering distribution, EES and threshold for time instances 1600 to 3000.

exhibit more precise estimations with lower variance compared to the parameters c, β, γ, and δ. This disparity is attributed to the differing extents to which each parameter influences the system's state. Consequently, it can be inferred that, under the loading conditions presented in the training data, the parameters m, k, and α play a more substantial role in governing the system's dynamics than c, β, γ, and δ. A more detailed discussion on this observation is provided later in this section.

Figure 7.2 serves to illustrate the phenomenon of particle degeneracy and the subsequent rejuvenation process. In this figure, the black plots shows samples from the distribution over the acceleration given the integration uncertainty and the parameters. The effective sample size (ESS), depicted in blue, is crucial for monitoring particle degeneracy. A decline in ESS below a predefined threshold indicates functional degeneracy of parameters, necessitating their resampling.

It is important to note that the figure selectively displays only a fraction of the total resampling instances, prioritizing readability. However, the omitted data conforms to the same pattern as that which is shown. As the sampling instances progress, errors in the proposed parameters become increasingly evident, leading to a reduction in ESS until the threshold is reached. Upon resampling, the new parameter set generally exhibits closer tracking of the system's true states, with a reduced variance across the sample set. This process repeats until a complete browsing of the training

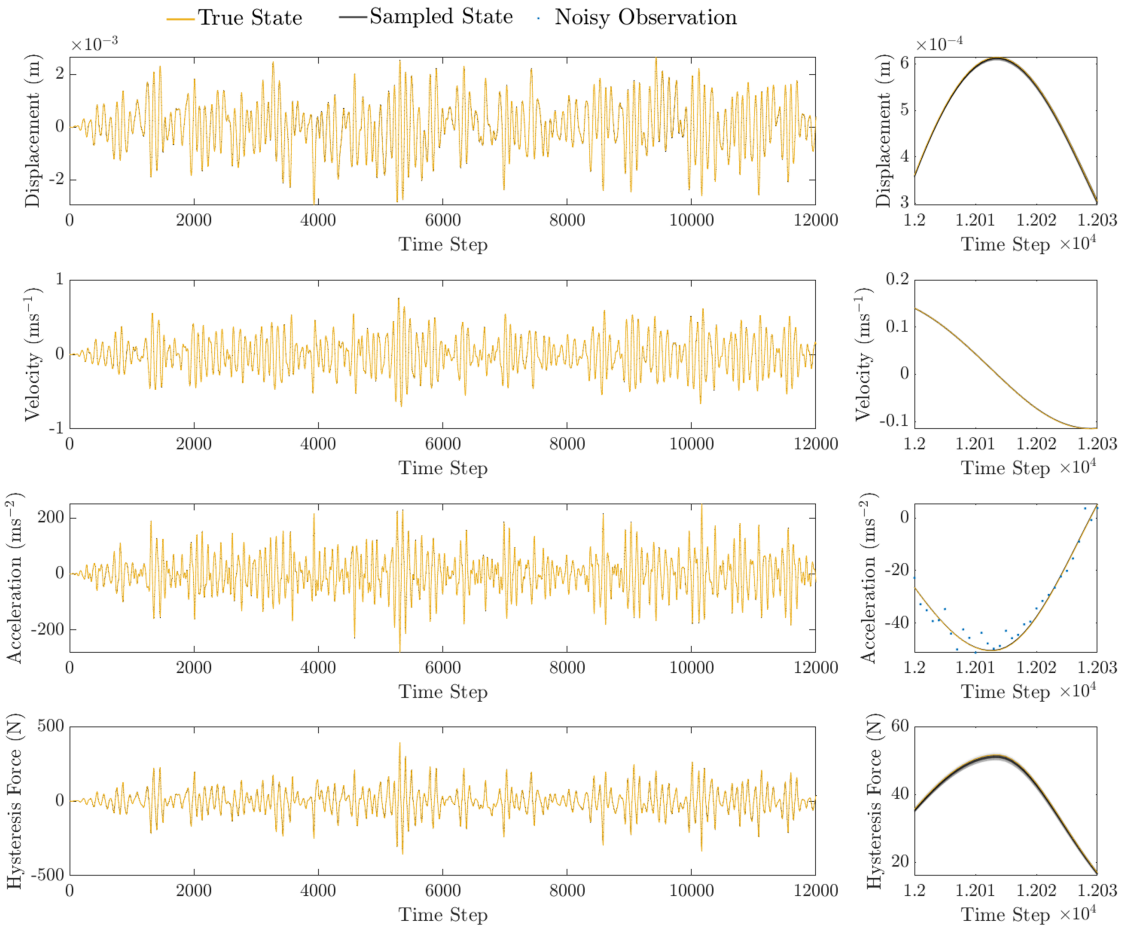

Figure 7.3: Samples from the posterior over the states for the Bouc Wen system plotted with the true states. An enlarge view of each state is shown to the left. Observations are plotted only for acceleration, as observations for the other states were not made available during the analysis.

data has been completed and a posterior distribution over the parameters given the entire data set is reached.

Figure 7.3 presents the true states of the Bouc-Wen system alongside the states generated from the sampled parameters of the posterior distribution. In the zoomed-in sections, it is evident that the states resulting from the posterior parameter distribution closely follow the ground truth. Additionally, it is observed that the variance of the posterior in the acceleration state is markedly lower than that of the noisy observations. This finding indicates that the posterior distribution over the parameters accurately captures the dynamics of the Bouc-Wen system under the specified training load.

Parameter estimation is achieved through the assessment of parameter quality as an ability to predict the state of the system. Therefore, parameters can only be optimised so long as a change in a parameter has an influence in the accuracy of the predicted state. This is limited by the ability of the training data to encapsulate the dynamics of the modelled system. This is limited by the level of noise in the data, any fundamental uncertainly in the identification process such as integration uncertainty and the degree to which the data expresses the systems nonlinearity.

To see a significant further reduction in the variance of the posterior distribution further training data must be provided. This new training data must be such that a system simulated with parameters sampled from the current posterior would have a signification increase in variance of the posterior across the states. Simply put, training data must be provided that forces the states of the system outside of the current noise floor. Two different approaches may be used to achieve this. The first it to increase the number of time steps. As the number of time steps increases small inaccuracies in the parameters are more likely to cause drift in the predicted states from the true states. The second is to used training data with a greater forcing amplitude as it will make the nonlinear effect on the system dynamics more dominant. However, both of these approaches must be balanced against increasing integration uncertainty. For a fixed time step size increasing the number of time steps allows for integration error to accumulate and increasing forcing magnitudes increases the gradients of the states and makes the makes linear approximations within the integrator less valid.

Here in lies a particular strength of the proposed methodology. When the uncertainly in the integration is accounted for in the posterior over the parameters it prevents the estimator from becoming overly confident in parameters that are biased due to numerical integration errors.

For additional validation of the identified parameters, the identified parameters will be used in simulation and compared in performance to the ground truth for a benchmark testing data set [162]. The benchmark data set contains noise free input and output measurements for both a sine-sweep and multisine loading condition. The data is simulated at 15000 Hz and down sampled to 750 Hz and consists of 8192 samples for both loading conditions and is performed using a Newmark integration method [163]. The sine-sweep data starts with zero initial conditions so is not steady state. The forcing amplitude is 40 N and the frequency bands rangers from 20 to 50 Hz with a sweep rate of 10 Hz/min. The random phase multi sine dataset is at

steady state and the frequencies range from 5 - 150 Hz with an RMS input value of 50 N. Note that this testing data set was at no point using in the train process.

To evaluate the performance of the parameter estimation the RMSE was calculated for each particle for both the sine-sweep and a multisine loading conditions. For the sine-sweep the maximum RMSE was 6.6416×10^{-6} the minimum was $4.6313\text{e}\times10^{-6}$ and the mean RMSE of all particles was 5.4017×10^{-6}. For the mutisine the maximum RMSE was 6.2220×10^{-6} the minimum was 7.1967×10^{-7} and the mean RMSE of all particles was 2.4772×10^{-6}.

7.4.2 Silverbox

The Silverbox [164] is an electrical circuit designed to emulate the behavior of the Duffing oscillator, a SDOF mass-spring-damper system characterised by a cubic spring term. The equation of motion for the Duffing oscillator is expressed as:

$$m\ddot{x} + c\dot{x} + kx + k_3x^3 = u(t) \tag{7.24}$$

In this equation, x represents displacement, $\dot{x}$ denotes velocity, m is the mass, c signifies viscous damping, k is the linear stiffness, k_3 the cubic stiffness, and $u(t)$ the forcing function. While the Silver Box does not perfectly replicate the theoretical Duffing oscillator, it serves as a highly accurate approximation.

The Silver Box benchmark [165] encompasses two datasets. The dataset predominantly analysed, and the one chosen for analysis in this study, features an input time signal that resembles an arrow, as shown in Figure 7.4. In the Silver Box, both input and output are in the form of voltage. The input voltage simulates the Duffing oscillator's forcing function, while the output voltage's response to this input mimics the displacement response of the Duffing oscillator to a forcing function $u(t)$.

The input voltage shown in Figure 7.4 is comprised of two distinct functions. The first is a Gaussian white noise of linearly increasing amplitude filtered by a 9^{th} order Butterworth filter. This forms the head of the arrow and is comprised of 40,000 samples. The remaining samples shows 10 realisations of an odd random phase multisine samples at fixed amplitude. All data is recorded with a sample frequency of 610.35 Hz.

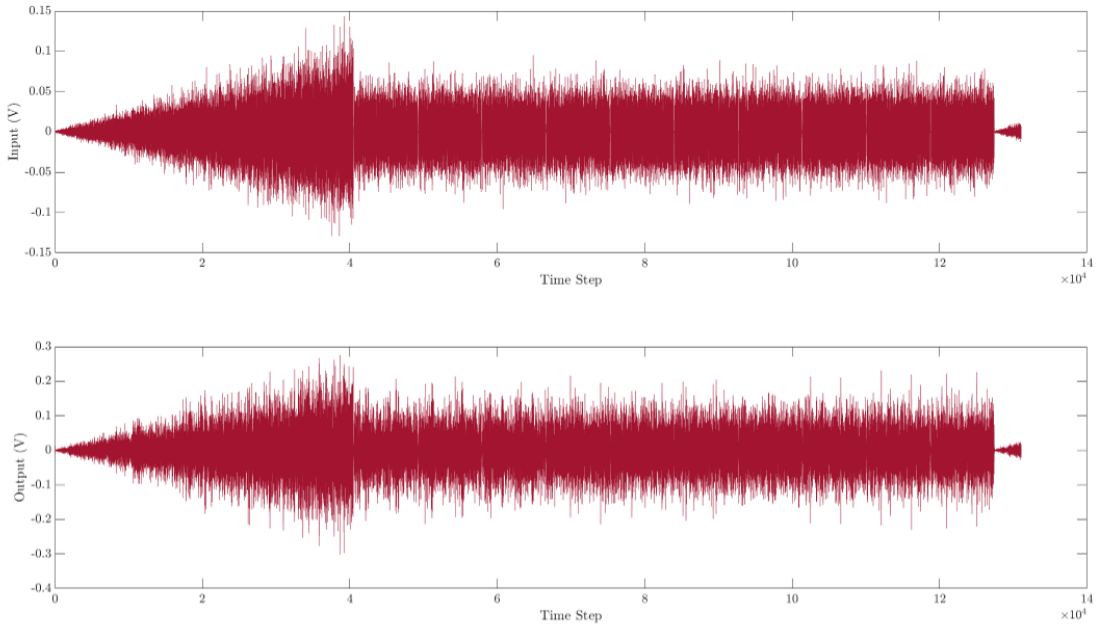

Figure 7.4: Measured input and response of the silver box benchmark.

The model is trained on data from the odd random phase multisine and tested on the data form the arrow head. Specifically the model is trained on 3072 observations ranging from data point 49,278 to 52,350. The prior and posterior distributions over the parameters can be found in Figure 7.5. It can be noted that all samples of parameters are taken from a log distribution to enforce the prior that all the parameters as defined in the Duffing equation must be positive for the identification to be physically meaningful.

Figure 7.5 illustrates the evolution of parameter distributions in the Duffing equation, transitioning from a broad prior to a more defined posterior distribution. This transition encapsulates the full extent of uncertainty inherent in the identification process. Unlike the analogous plot for the Bouc-Wen system, the parameters in this instance are not normalised against a known ground truth, as no such truth exists for this experimental system. However, an analysis of the variance ranges in the posterior distributions allows for the inference that parameters m, c, and k exert a more pronounced influence on the training data compared to the cubic stiffness parameter k_3.

Figure 7.6 displays the observed output voltage alongside the displacement state derived from parameters sampled from the posterior distribution. The focus is solely on the displacement state, as no observations for other states are available

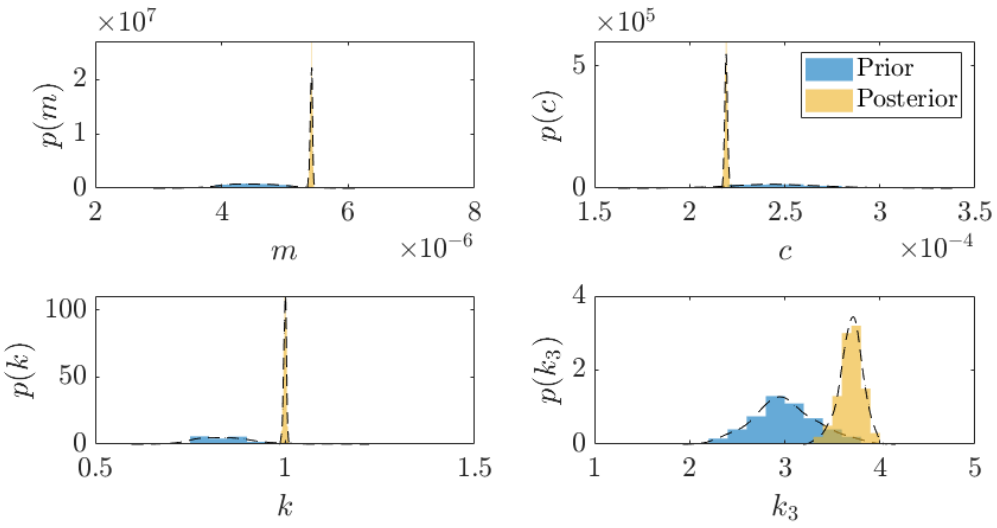

Figure 7.5: Prior and posterior histograms and PDFs for the silver box system.

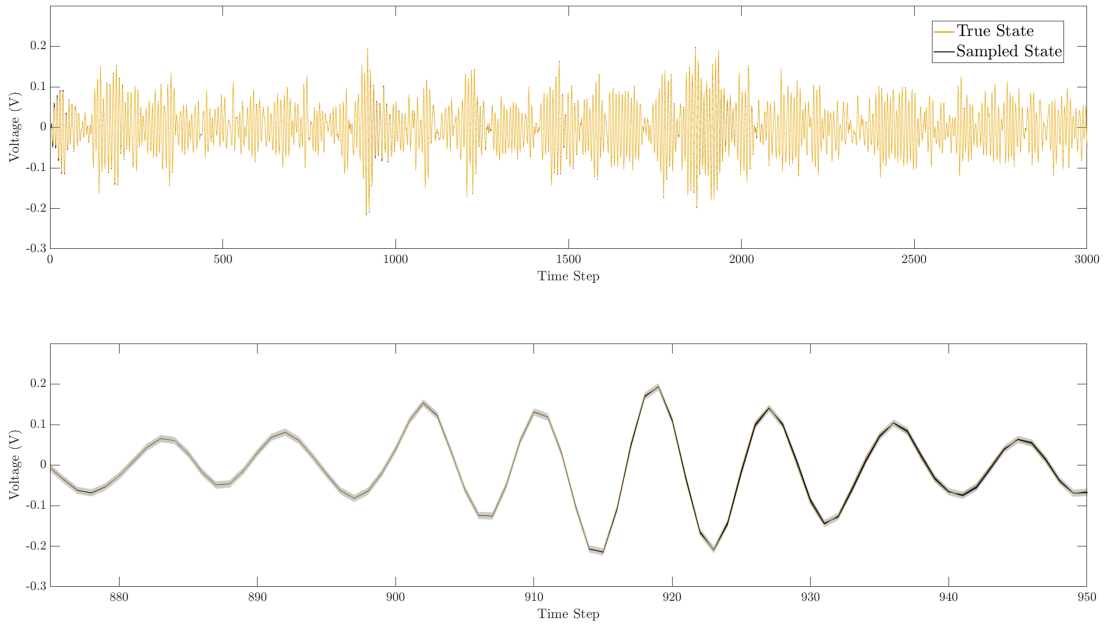

Figure 7.6: Samples from the posterior over the voltage for the silver box plotted with the measured output voltage. An enlarged view of the voltage is shown beneath.

for comparison. This plot reveals that the sampled parameters generally align well with the observed displacement, with most observed data points residing within the sampled distribution. However, a closer examination, particularly in the zoomed-in sections, reveals instances where the observed data deviates slightly from the sampled distribution.

This deviation can be attributed to the primary sources of uncertainty in this parameter estimation. Notably, the signal-to-noise ratio for the Silver Box is sufficiently high, rendering measurement noise negligible. Consequently, during identification, the measurement noise is set to an order of magnitude around 10^{-6}. This leaves integration uncertainty as the dominant source of error. The observed data's lack of smoothness, as evident in Figure 7.6, indicates that integration errors are likely to contribute significantly to the uncertainty in the states, and consequently, in the parameter quality assessment.

Now consider how the uncertainty in the integration is calibrated. The uncertainty in the integration is calibrated using Γ a quasi-MLE that approximates the average uncertainty from t=1 to t=T. This means at some points when the gradient of the states are small Γ will be an over estimate and when the gradient of the states are at their largest Γ will be an underestimate of the uncertainty in the integration. As such, it is not unexpected that at some extreme points the observed state falls outside the sampled states. However, even considering this it should be noted that the observed states do always fall comfortably inside 2σ of the second moment of the posterior. The second moment of the posterior is shown in faded gray in Figure 7.6.

Since, no ground truth is available to evaluate the quality of the identified parameters an alternative method must be used. For this the identified parameters will be used in simulation and compared to the observations for the arrowhead section of the dataset. The first 1000 time steps were excluded from this simulation to remove the transient so that the simulation was run from time step 1000 to 40,000. For this simulation a RMSE was calculated for each particle. The maximum RMSE was 2.9516×10^{-3} the minimum was 1.0567×10^{-3} and the mean RMSE of all particles was 1.8249×10^{-3}.

Figure 7.7: Image of the EMPS [166]

7.4.3 Electro-Mechanical Positioning System

The EMPS (Figure 7.7) is a standard configuration of drive system for prismatic joints of robots or machine tools. It is comprised of a DC motor equipped with an incremental encoder and a high-precision low-friction ball screw drive positioning unit. The EMPS also features an incremental encoder and accelerometer, but their data is excluded from the benchmark, mirroring the common absence of such measurements in industrial robots [167]. The nonlinear equation of motion for the EMPS is given as,

$$\tau_{idm}(t) = M\ddot{x}(t) + F_v\dot{x}(t) + F_c sign(\dot{x}(t)) + offset \tag{7.25}$$

where τ_{idm} is the joint torque/force, M is the inertia of the arm, F_v is the viscous friction and F_c is the Coulomb friction.

For full details of this benchmark see [166]. The EMPS provides two datasets. The first is for training and the second is for testing. The training dataset consists of 25s of motor force and motor position measurements sampled at 1000 Hz. When generating the training data the EMPS was excited with bang-bang accelerations. Training is performed based on 24576 data points and the prior over the parameters is shown if Figure 7.8. To ensure that M, F_v and F_c remain positive new proposals of these parameters will be sampled in the log space.

Figure 7.8 shows the prior and posterior parameter distribution for the EMPS benchmark. It can be seen that the diffuse prior converges to a posterior with a narrow variance. Excellent convergence is seen across all four identified parameters. Figure 7.9 shows the observed displacement plotted together with the displacement state generated from parameters sampled from the posterior distribution. Only the

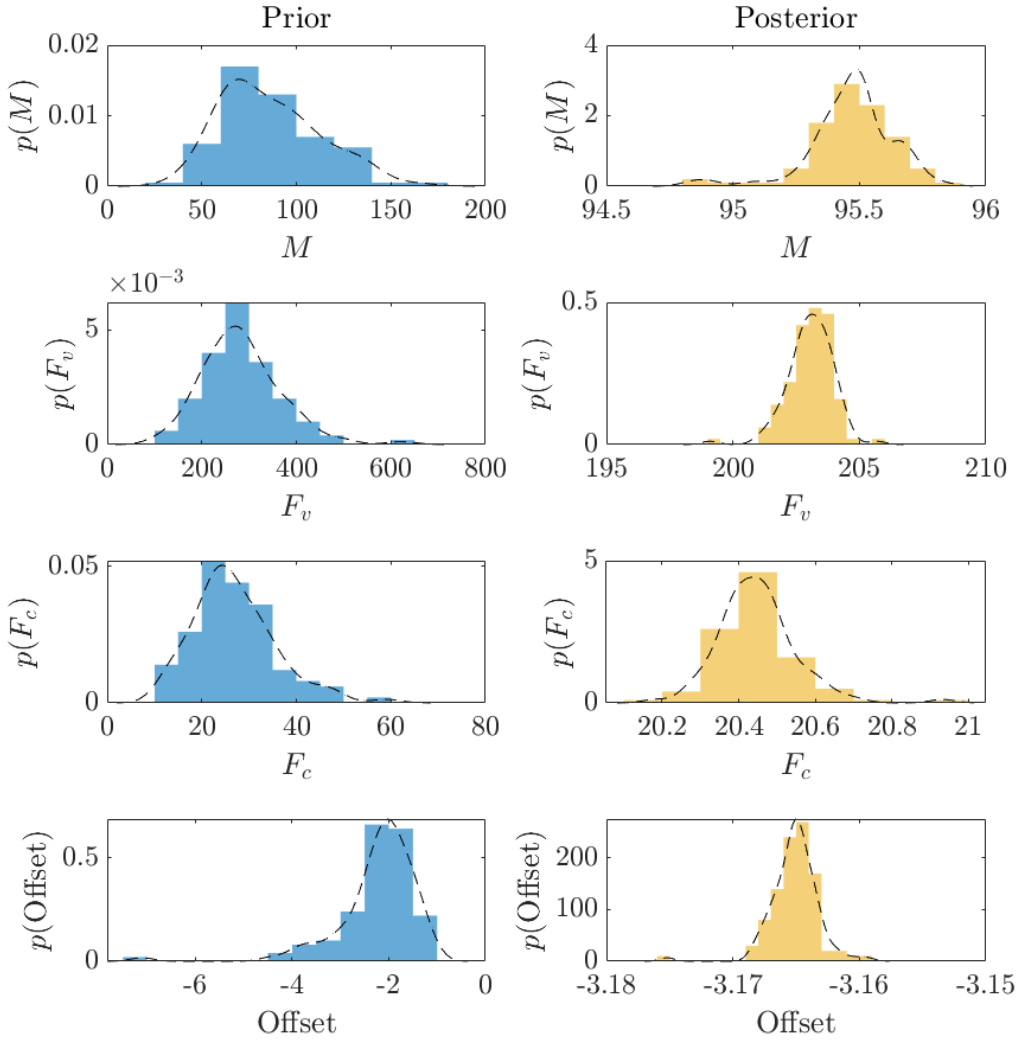

Figure 7.8: Prior and posterior histograms and PDFs for the EMPS shown separately due to the large reduction in variance from prior to posterior.

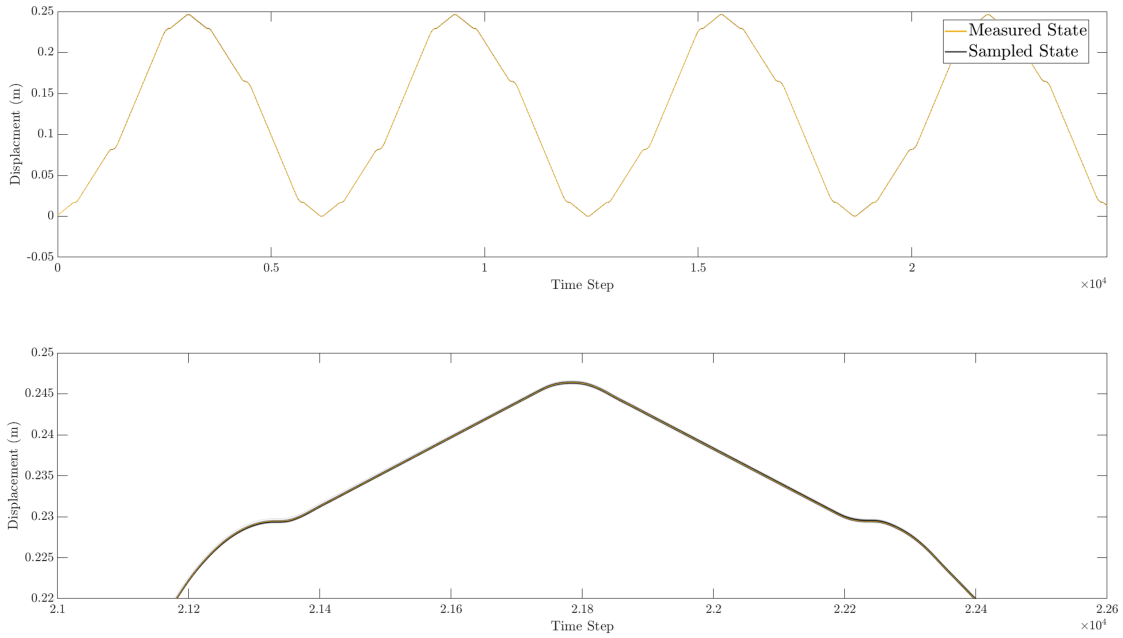

Figure 7.9: Samples from the posterior over the Displacement for the EMPS plotted with the measured output Displacement. An enlarge view of the displacement is shown beneath.

Table 7.2: Summary of Case Study Results

Case Study:	**B.W. Sinesweep**	**B.W. Multisine**	**Silver Box**	**EMPs**
Unit:	RMS (ms^{-2})	RMS (ms^{-2})	RMS (V)	RMS (m)
Min Particle	4.631×10^{-6}	7.197×10^{-7}	1.057×10^{-3}	2.950×10^{-4}
Max Particle	6.642×10^{-6}	6.222×10^{-6}	2.952×10^{-3}	8.990×10^{-4}
Mean Particle	5.402×10^{-6}	2.477×10^{-6}	1.825×10^{-3}	5.202×10^{-4}

displacement states is shown as no observations exist for the other states. The parameters sampled from the posterior distribution correctly track the measurements with the measured states always falling within the sampled distribution.

To evaluate the quality of the identified parameters. The identified parameters will be used in simulation and compared to the observations for the testing dataset. The testing dataset consists of 25s of motor force and motor position measurements sampled at 1000 Hz. When generating the testing data the EMPS is again excited with bang-bang accelerations however this time there is an additional pulse component to the loading.

The identified parameters are evaluated for the for the testing dataset and compared

against the measurements. For this simulation a RMSE was calculated for each particle. The maximum RMSE was 8.9900x10^{-4} the minimum was 2.9500x10^{-4} and the mean RMSE of all particles was 5.2018x10^{-4}.

7.5 Summary

This chapter introduced a Bayesian parameter estimation method that unifies SMC and Probabilistic Numerics to establish a comprehensive probabilistic framework for parameter estimation in nonlinear systems. Building on existing techniques for probabilistic ODE solutions and SMC-based parameter inference, the proposed approach efficiently identifies latent states and system parameters from noisy observations. It provides a comprehensive treatment of uncertainty by generating posterior distributions that encapsulate prior beliefs and the intrinsic uncertainties of both the data and the nonlinear ODE solutions. Through 3 case studies, it was demonstrated that the proposed procedure could be realised effectively. Notable results were achieved across both simulated and experimental datasets with low RMSE.

Chapter 8

Conclusion and Further work

Structural dynamics is essential for understanding how structures respond to external excitation, which is critical for optimising performance, mitigating damage, targeting maintenance, and informing operational decisions. The aim of this thesis was to address fundamental challenges in structural dynamics associated with nonlinearity, uncertainty, non-stationarity, and joint identification. These challenges arise in the three fundamental tasks of a structural dynamicist: system identification, prediction and simulation, and input identification.

- **Chapter 1** introduced some of the core challenges in structural dynamics, focusing on system identification, prediction, and input identification. It highlighted the importance of building accurate models for dynamic systems, particularly in the context of nonlinearities and uncertainties. The chapter established the need for advanced methods to address the limitations of traditional approaches, laying the motivations for the methodologies developed in this thesis.

- **Chapter 2** presented an overview of the background and theoretical foundations essential for developing the framework used in this thesis. Additionally, a concise review of relevant literature was included to establish the context for the later chapters.

- **Chapter 3** presented a comprehensive introduction to the Gaussian Process Latent Force Model (GPLFM) for joint input-state identification. The existing literature on the application of GPLFMs for latent input force identification

was reviewed. The derivation of the state-space Gaussian Process (GP) from the mean and covariance functions of a stationary temporal GP was outlined. Furthermore, an established formulation of a state-space GPLFM was constructed, based on this state-space GP representation. An example study demonstrated the practical application of the GPLFM, offering insights into the influence of observed state variable selection on the performance of the GPLFM in joint input-state estimation.

- In **Chapter 4**, the nonlinear joint input-state estimation methodology was extended for the first time to accommodate systems with dynamic nonlinearities, particularly those exhibiting hysteresis. This extension marks a significant advancement, as such systems introduce complex nonlinearities into the system's transition function and necessitate the estimation of additional hidden states that are not simply derivatives or integrals of observable states.

  The proposed methodology successfully addressed these challenges by inferring both latent forces and hidden internal states from noisy acceleration measurements in a nonlinear oscillator with hysteresis. By applying the nonlinear GPLFM, solved with Particle Gibbs with Ancestor Sampling (PGAS), highly accurate input-state estimations were achieved for this challenging system. State estimation Normalised Mean Squared Errors (NMSE) were consistently below 1%, even for the additional internal state $z(x, \dot{x})$.

  Furthermore, the force estimates exhibited low errors, with 2.18% error when a GP was used as the loading signal, and 5.22% error with sine wave forcing. In addition to these low error rates, the methodology effectively quantified the uncertainty in the state estimates, thereby enhancing the reliability of the system identification process.

- In **Chapter 5**, a new methodology was proposed for using a GPLFM to identify the latent restoring force surface of a unknown nonlinearity in an output-only context. This approach was implemented within a operational Modal Analysis (OMA) framework by modeling the system's input as a white noise process.

  The methodology demonstrated the capability to jointly recover the temporal behaviors of latent states, latent restoring forces, and latent input forces, while also deriving a GP representation of the underlying restoring force function. The methodology was benchmarked against a Duffing oscillator, which was excited by a white noise process. In this case study, the functional form of

the Duffing oscillator's nonlinearity, the input to the system, and the system's displacement and velocity states were all hidden. Only noisy acceleration observations were available during the analysis.

Although it was assumed that the linear system's functional form and parameters were known *a priori*, in the general case this problem is highly unidentifiable. However, by assuming the input to the system was a white noise process, it became possible to recover the latent restoring force surface. This was achieved by taking the expectation of the recovered total latent force with respects to time and conditioned upon the states. Once the restoring force surface was identified, the time series of the latent input force could then be recovered.

- In **Chapter 6**, to address the challenges of capturing smooth, non-stationary latent forces, a novel non-stationary state-space formulation of the GPLFM was developed. This formulation incorporates a time-varying length scale within the state-space GPLFM framework, enabling the model to adapt dynamically to the non-stationarity of the latent functions. The time-varying length scale is learned through a secondary GP, that augments the state-space structure, establishing a hierarchical or "deep" GP configuration. This advancement provides an effective approach for modeling complex, non-stationary behavior in latent force estimation.

  Techniques for linearizing and discretizing the model were introduced, allowing the application of Bayesian filtering and smoothing to solve for the latent forces. Two case studies were conducted to benchmark the performance of the proposed non-stationary GPLFM against the stationary GPLFM. The first case study involved a single degree of freedom (SDOF) system subjected to a sine sweep load to evaluate its ability to handle non-stationary latent input forces. The second study examined a bursting Duffing oscillator, aimed at assessing its effectiveness in recovering non-stationary latent restoring forces.

  Comparisons between the stationary and non-stationary GPLFMs, using graphical analyses and NMSE calculations, revealed that the stationary GPLFM, with a fixed length scale, either over-fitted or over-smoothed the data, leading to poor recovery of latent forces under non-stationary conditions. In contrast, the non-stationary GPLFM, by learning a dynamic length scale, successfully adapted to the varying frequency content, recovering both low- and high-frequency components accurately without over-smoothing or

over-fitting.

- Finally, **Chapter 7** introduced a novel Bayesian parameter estimation technique that combines Sequential Monte Carlo (SMC) methods with Probabilistic Numerics to create a unified probabilistic framework for parameter estimation in nonlinear systems.

  The authors posited that parameter estimation in nonlinear dynamic systems can be conceptually divided into three sequential stages. The first stage involves selecting a set of candidate parameters. In the second stage, these trial parameters, along with the model's functional form and initial conditions, are used to solve the initial value problem (IVP), thereby determining the system's state at a future time point.

  The third stage compares the predicted state from the IVP solution to the observed measured state at the corresponding time, optimizing the model parameters by minimizing the discrepancy between the predicted and measured states. This study employed a comprehensive Bayesian methodology across all three stages. In the first stage, a prior distribution was assigned over the parameter space, effectively reducing the search space and enhancing the efficiency of the parameter selection.

  During the second stage, a probabilistic ODE solver was used to explicitly account for the uncertainties inherent in solving the IVP for nonlinear dynamic systems, incorporating these uncertainties into the identification process. In the final stage, the posterior distribution over the parameters was evaluated, incorporating uncertainties from both the numerical integration and the measurements.

  Through three case studies, the proposed procedure was effectively demonstrated, achieving notable results with low Root Mean Square Error (RMSE) across both simulated and experimental datasets.

8.1 Further Work

Although the developed methodologies represent significant advancements, they also present certain limitations and challenges. This section highlights these limitations and outlines potential directions for future research.

- In **Chapter 4**, the PGAS method was used to approximate the smoothing distribution for the GPLFM. PGAS offers key advantages over Gaussian approximations by capturing complex, nonlinear relationships through Monte Carlo techniques, enabling more accurate modeling of non-Gaussian latent forces.

  However, PGAS incurs a significant computational burden due to the need for multiple particle filters. To reduce this load, a small number of particles was used in this work. This likely resulted in artifacts in the recovered signals, compromising the accuracy of latent force estimation.

  Future research should investigate the effect of increasing the number of particles on the smoothing distribution, as this may reduce artifacts and improve accuracy. Despite the reduced particle count, the computational burden remains high. To address this, more efficient inference methods or alternative approximations—such as Gaussian filters for weak nonlinearities or variational inference or SMC^2—could be explored to balance computational load and recovery accuracy.

- In **Chapter 5**, an output-only GPLFM was presented for the identification of a Duffing oscillator, a relatively simple SDOF nonlinearity. Extending this approach to multi-degree of freedom (MDOF) systems is a potential direction for further investigation.

  Such an extension is straightforward as long as only one nonlinearity is present and the state-space model (SSM) of the linear component of the MDOF system can be formulated. However, when multiple nonlinearities are involved, separating their contributions presents a significant challenge that warrants further investigation.

  Furthermore, more complex types of nonlinearities, such as hysteresis, could be explored. In particular, modeling the phase space of a Bouc-Wen hysteretic system when recovering the restoring force surface using a GP poses substantial difficulties, especially when it is not known *a priori* that additional states are required to represent the nonlinearity. This presents an interesting and complex problem for future work.

  In the current study, it was assumed that the linear parameters of the system are known. Future work could investigate whether informative priors placed over the functional form of the nonlinearity could allow for the simultaneous

recovery of both the nonlinear structure and the linear parameters as part of the identification process.

Moreover, this work assumes that the unobserved input is a white noise process. However, this assumption may not hold when the true underlying input is narrow-band, as is often the case in offshore structures. It is believed that both the nonlinear structure and a temporal distribution over the latent input force could still be recovered, provided the input is modeled as a stationary function.

Investigating this approach could significantly expand the applicability of the methodology to more realistic loading conditions, making it a promising direction for further research.

- In **Chapter 6**, a non-stationary GPLFM was developed within the state-space framework and applied to two SDOF case studies. As with the output-only GPLFM work in Chapter 5, this methodology could be extended to MDOF systems.

  Such an extension is relatively straightforward when only one latent force is present and the SSM of the linear component of the MDOF system is available. However, if multiple latent forces are involved, distinguishing their contributions presents a significant challenge that warrants further investigation.

  It is also likely that the non-stationary GPLFM could be adapted to the output-only case using the techniques presented in Chapter 5. In this study, the developed GPLFM was benchmarked against simulated datasets, which allows for comparison against ground truth data. However, evaluating the model's performance against real-world or experimental non-stationary datasets would provide further validation and highlight its practical applicability.

  One identified challenge in this work is the computational burden associated with re-discretizing the system at each time step. Investigating algebraic methods for this process could alleviate the need to approximate the matrix exponential at every step, improving computational efficiency.

  Additionally, the noise covariance, $Q(t)$, was updated using a zero-order hold due to the difficulties of higher-order approximations. Exploring higher-order approximations for the integral of $Q(t)$ could enhance the accuracy of the uncertainty recovery, further refining the model's performance.

Furthermore, it is proposed that not only the length scale but also the second hyperparameter of the Matérn kernel, the variance σ^2, should be modeled as a GP. This would allow the non-stationary GPLFM to account for systems where the latent force exhibits non-stationary amplitude.

For instance, this approach could improve the modeling of seismic forces, capturing the varying amplitudes between primary shocks and aftershocks, or wind loading, where gusts require precise representation of amplitude fluctuations.

- In **Chapter 7**, a Bayesian parameter estimation method was introduced, combining Sequential Monte Carlo (SMC) and Probabilistic Numerics to establish a unified probabilistic framework for parameter estimation in nonlinear systems.

  This approach addresses the challenges posed by numerical integration and the uncertainty it introduces into system identification. This problem extends beyond just parameter estimation. Future research should explore how probabilistic techniques for numerical integration can be applied to other system identification methods, such as GPLFMs.

  By accounting for the uncertainty introduced by numerical integration, these methods can improve the overall accuracy and reliability of system identification across a wider range of applications.

8.2 Closing Remarks

The primary aim of this thesis has been to advance Bayesian methods for input, system, and output identification in nonlinear dynamic systems. As modern engineering projects become increasingly ambitious, the complexity of the systems involved continues to grow. This escalation necessitates the development of advanced, data-based approaches to complement traditional techniques for building models and understanding system inputs. With the growing availability of large datasets, increased access to computational resources, and significant advancements in machine learning, the potential for data-driven approaches to better understand complex systems has never been more promising. However, enthusiasm for these methods must be tempered by the need to ensure that the models developed are not only representative of the data but also reflective of the underlying physical reality

that the data seeks to describe. In response to this, the work presented in this thesis has explored the integration of physical principles into data-driven methodologies. The goal has been to ground these models in real-world physics, enhancing their reliability and interpretability. Additionally, considerable attention has been given to the quantification of uncertainty, ensuring that the outputs of these models provide not only predictions but also an informed measure of confidence in those predictions. Looking forward, the combination of data-driven approaches with principled, physics-based modeling holds tremendous potential for further breakthroughs. As computational power and the sophistication of algorithms continue to grow, the future of physics informed system identification and uncertainty quantification is both exciting and full of promise, offering the potential for more robust and insightful solutions to the increasingly complex challenges faced by engineers and scientists.

Bibliography

[1] C. R Farrar and K Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, 2007.

[2] C. R Farrar and K Worden. *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012.

[3] M Liu, S Fang, H Dong, and C Xu. Review of digital twin about concepts, technologies, and industrial applications. *Journal of manufacturing systems*, 58:346–361, 2021.

[4] F Tao, H Zhang, A Liu, and A. Y Nee. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics*, 15(4):2405–2415, 2018.

[5] D. J Ewins. *Modal testing: theory, practice and application*. John Wiley & Sons, 2009.

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

[7] K Worden. *Nonlinearity in structural dynamics: detection, identification and modelling*. CRC Press, 2019.

[8] T. S Parker and L. O Chua. Chaos: A tutorial for engineers. *Proceedings of the IEEE*, 75(8):982–1008, 1987.

[9] P. P Castaneda and P Suquet. Nonlinear composites. *Advances in applied mechanics*, 34:171–302, 1997.

[10] R Bouc. Forced vibrations of mechanical systems with hysteresis. In *Proc. of the Fourth Conference on Nonlinear Oscillations, Prague, 1967*, 1967.

[11] E Popova and V. L Popov. The research works of coulomb and amontons and generalized laws of friction. *Friction*, 3:183–190, 2015.

[12] R. A Ibrahim and C Pettit. Uncertainties and dynamic problems of bolted joints and other fasteners. *Journal of sound and Vibration*, 279(3-5):857–936, 2005.

[13] M Lou, H Wang, X Chen, and Y Zhai. Structure–soil–structure interaction: Literature review. *Soil dynamics and earthquake engineering*, 31(12):1724–1731, 2011.

[14] M Chatzis and A Smyth. Robust modeling of the rocking problem. *Journal of Engineering Mechanics*, 138(3):247–262, 2012.

[15] A Gomes and J Appleton. Nonlinear cyclic stress-strain relationship of reinforcing bars including buckling. *Engineering Structures*, 19(10):822–826, 1997.

[16] B Budiansky and J. W Hutchinson. A survey of some buckling problems. *Journal of spacecraft and rockets*, 40(6):918–923, 2003.

[17] H Ozdemir. *Nonlinear transient dynamic analysis of yielding structures*. University of California, Berkeley, 1976.

[18] E. J Cross, S. J Gibson, M. R Jones, D. J Pitchforth, S Zhang, and T. J Rogers. Physics-informed machine learning for structural health monitoring. *Structural Health Monitoring Based on Data Science Techniques*, pages 347–367, 2022.

[19] V. V Toropov. Simulation approach to structural optimization. *Structural Optimization*, 1:37–46, 1989.

[20] X Han and J Liu. *Numerical simulation-based design*. Springer, 2017.

[21] T. K Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47–60, 1996.

[22] I. J Myung. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100, 2003.

[23] S Särkkä and L Svensson. *Bayesian filtering and smoothing*, volume 17. Cambridge university press, 2023.

[24] W. R Gilks, S Richardson, and D Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

[25] S Brooks, A Gelman, G Jones, and X.-L Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[26] J. C Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.

[27] P Hennig, M. A Osborne, and M Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150142, 2015.

[28] P Hennig, M. A Osborne, and H. P Kersting. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press, 2022.

[29] T. J Rogers. *Towards Bayesian system identification: with application to SHM of offshore structures*. PhD thesis, University of Sheffield, 2019.

[30] A Rytter. *Vibrational based inspection of civil engineering structures*. PhD thesis, Aalborg University, 1993.

[31] K Worden and J. M Dulieu-Barton. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring*, 3(1):85–98, 2004.

[32] N Perišić and U. T Tygesen. Cost-effective load monitoring methods for fatigue life estimation of offshore platform. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 45431, page V04BT02A005. American Society of Mechanical Engineers, 2014.

[33] N Perišić, P. H Kirkegaard, and U. T Tygesen. Load identification of offshore platform for fatigue life estimation. In *Structural Health Monitoring, Volume 5: Proceedings of the 32nd IMAC, A Conference and Exposition on Structural Dynamics, 2014*, pages 99–109. Springer, 2014.

[34] N Noppe, A Iliopoulos, W Weijtjens, and C Devriendt. Full load estimation of an offshore wind turbine based on scada and accelerometer data. In *Journal of Physics: Conference Series*, volume 753, page 072025. IOP Publishing, 2016.

[35] L Ziegler, S Voormeeren, S Schafhirt, and M Muskulus. Design clustering of offshore wind turbines using probabilistic fatigue load estimation. *Renewable Energy*, 91:425–433, 2016.

[36] K Maes, A Iliopoulos, W Weijtjens, C Devriendt, and G Lombaert. Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms. *Mechanical Systems and Signal Processing*, 76:592–611, 2016.

[37] A Kareem. Numerical simulation of wind effects: A probabilistic perspective. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10-11):1472–1497, 2008.

[38] S Chakrabarti. *Handbook of Offshore Engineering (2-volume set)*, volume 1. Elsevier, 2005.

[39] G Deodatis. Non-stationary stochastic vector processes: seismic ground motion applications. *Probabilistic engineering mechanics*, 11(3):149–167, 1996.

[40] U. D of Transportation. Beyond traffic: 2045 final report, 2016.

[41] M Kawato. Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, 9(6):718–727, 1999.

[42] C. D Murray and S. F Dermott. *Solar system dynamics*. Cambridge university press, 1999.

[43] I Newton, I. B Cohen, and A Whitman. *The Principia: mathematical principles of natural philosophy*. Univ of California Press, 1999.

[44] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A. N Gomez, Ł Kaiser, and I Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[45] L Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[46] J Decuyper, P Dreesen, J Schoukens, M. C Runacres, and K Tiels. Decoupling multivariate polynomials for nonlinear state-space models. *IEEE Control Systems Letters*, 3(3):745–750, 2019.

[47] V Cerone, V Razza, and D Regruto. A unified framework for the identification of a general class of multivariable nonlinear block-structured systems. *International Journal of Robust and Nonlinear Control*, 31(15):7344–7360, 2021.

[48] A Chiuso and G Pillonetto. System identification: A machine learning perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2 (1):281–304, 2019.

[49] S. A Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[50] M Schoukens and K Worden. Evolutionary identification of block-structured systems. In *Dynamics of Coupled Structures, Volume 4: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017*, pages 359–366. Springer, 2017.

[51] K Worden, R Barthorpe, E Cross, N Dervilis, G Holmes, G Manson, and T Rogers. On evolutionary system identification with applications to nonlinear benchmarks. *Mechanical Systems and Signal Processing*, 112:194–232, 2018.

[52] M. J Korenberg and I. W Hunter. The identification of nonlinear biological systems: Wiener kernel approaches. *Annals of Biomedical Engineering*, 18: 629–654, 1990.

[53] J Noël and M Schoukens. Cross-fertilising research in nonlinear system identification between the mechanical, control and machine learning fields: Editorial statement, 2019.

[54] P Spirtes. Introduction to causal inference. *Journal of Machine Learning Research*, 11(5), 2010.

[55] E. J Cross, T. J Rogers, D. J Pitchforth, S. J Gibson, S Zhang, and M. R Jones. A spectrum of physics-informed gaussian processes for regression in engineering. *Data-Centric Engineering*, 5:e8, 2024.

[56] 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, 2006.

[57] J.-P Noël and G Kerschen. 10 years of advances in nonlinear system identification in structural dynamics: A review. In *ISMA 2016-International Conference on Noise and Vibration Engineering*, 2016.

[58] J.-P Noël and G Kerschen. Nonlinear system identification in structural dynamics: 10 more years of progress. *Mechanical Systems and Signal Processing*, 83:2–35, 2017.

[59] K Wyckaert. Development and evaluation of detection and identification schemes for the nonlinear dynamical behaviour of mechanical structures. 1994.

[60] G Gloth and D Goge. Handling of non-linear structural characteristics in ground vibration testing. *The Shock and Vibration Digest*, 38(5):460–461, 2006.

[61] S Masri and T Caughey. A nonparametric identification technique for nonlinear dynamic problems. 1979.

[62] E Crawley and K O'Donnell. Identification of nonlinear system parameters in joints using the force-state mapping technique. In *27th Structures, Structural Dynamics and Materials Conference*, page 1013, 1986.

[63] P. D Arendt, D. W Apley, and W Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. 2012.

[64] J Brynjarsdóttir and A O'Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse problems*, 30(11):114007, 2014.

[65] P Gardner, T Rogers, C Lord, and R Barthorpe. Learning model discrepancy: A gaussian process and sampling-based approach. *Mechanical Systems and Signal Processing*, 152:107381, 2021.

[66] C. K Williams and C. E Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[67] T. J Rogers, G. R Holmes, E Cross, and K Worden. On a grey box modelling framework for nonlinear system identification. In *Special Topics in Structural Dynamics, Volume 6: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017*, pages 167–178. Springer, 2017.

[68] D. J Pitchforth, T. J Rogers, U. T Tygesen, and E. J Cross. Grey-box models for wave loading prediction. *Mechanical Systems and Signal Processing*, 159: 107741, 2021.

[69] H.-P Wan and W.-X Ren. A residual-based gaussian process model framework for finite element model updating. *Computers & Structures*, 156:149–159, 2015.

[70] H.-P Wan and W.-X Ren. Stochastic model updating utilizing bayesian approach and gaussian process model. *Mechanical Systems and Signal Processing*, 70:245–268, 2016.

[71] M Alvarez, D Luengo, and N. D Lawrence. Latent force models. In *Artificial Intelligence and Statistics*, pages 9–16. PMLR, 2009.

[72] J Hartikainen and S Särkkä. Kalman filtering and smoothing solutions to temporal gaussian process regression models. In *2010 IEEE international workshop on machine learning for signal processing*, pages 379–384. IEEE, 2010.

[73] J Hartikainen and S Sarkka. Sequential inference for latent force models. *arXiv preprint arXiv:1202.3730*, 2012.

[74] R Nayek, S Chakraborty, and S Narasimhan. A gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems and Signal Processing*, 128:497–530, 2019.

[75] T. J Rogers, K Worden, and E. J Cross. Bayesian joint input-state estimation for nonlinear systems. *Vibration*, 3(3):281–303, 2020.

[76] T. J Rogers, J. D Longbottom, K Worden, and E. J Cross. Nonlinear gaussian process latent force models for input estimation in hysteretic systems. In *Proceedings of the 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021)*, 2021. doi: 10.7712/120221.8017.18937.

[77] T Rogers, K Worden, G Manson, U Tygesen, and E Cross. A bayesian filtering approach to operational modal analysis with recovery of forcing signals. In *Proceedings of ISMA*, volume 2018, pages 5181–5194, 2018.

[78] T Rogers, K Worden, and E Cross. On the application of gaussian process latent force models for joint input-state-parameter estimation: With a view to bayesian operational identification. *Mechanical Systems and Signal Processing*, 140:106580, 2020.

[79] T. J Rogers and T Friis. A latent restoring force approach to nonlinear system identification. *Mechanical Systems and Signal Processing*, 180:109426, 2022.

[80] K Worden and G Manson. On the identification of hysteretic systems. part i: Fitness landscapes and evolutionary identification. *Mechanical Systems and Signal Processing*, 29:201–212, 2012.

[81] K Worden, O Tiboaca, I Antoniadou, and R Barthorpe. System identification of an mdof experimental structure with a view towards validation and verification. In *Topics in Modal Analysis, Volume 10: Proceedings of the 33rd IMAC, A Conference and Exposition on Structural Dynamics, 2015*, pages 57–65. Springer, 2015.

[82] J. L Beck and L. S Katafygiotis. Updating models and their uncertainties. i: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4): 455–461, 1998.

[83] M Muto and J. L Beck. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control*, 14(1-2):7–34, 2008.

[84] A Wills, T. B Schön, and B Ninness. Parameter estimation for discrete-time nonlinear systems using em. *IFAC Proceedings Volumes*, 41(2):4012–4017, 2008.

[85] T. B Schön, A Wills, and B Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39–49, 2011.

[86] P. L Green and K Worden. Bayesian and markov chain monte carlo methods for identifying nonlinear systems in the presence of uncertainty. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2051):20140405, 2015.

[87] T. B Schön, F Lindsten, J Dahlin, J Wågberg, C. A Naesseth, A Svensson, and L Dai. Sequential monte carlo methods for system identification. *IFAC-PapersOnLine*, 48(28):775–786, 2015. doi: 10.1016/j.ifacol.2015.12.161.

[88] H Ebrahimian, R Astroza, and J. P Conte. Extended kalman filter for material parameter estimation in nonlinear structural finite element models using direct differentiation method. *Earthquake Engineering & Structural Dynamics*, 44 (10):1495–1522, 2015.

[89] J Ching, J. L Beck, and K. A Porter. Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic engineering mechanics*, 21(1): 81–96, 2006.

[90] F Lindsten, T. B Schön, and M. I Jordan. Bayesian semiparametric wiener system identification. *Automatica*, 49(7):2053–2063, 2013.

[91] V. I Arnold. *Ordinary Differential Equations*. Springer, 3rd edition, 1992. ISBN 9783540548133.

[92] J Durbin and S. J Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.

[93] E. L Ince. *Ordinary differential equations*. Courier Corporation, 1956.

[94] J. D Lambert et al. *Numerical methods for ordinary differential systems*, volume 146. Wiley New York, 1991.

[95] B Taylor. *Methodus Incrementorum Directa et Inversa*. Gulielmus Innys, London, 1715.

[96] J Stewart. *Calculus: Early Transcendentals*. Cengage Learning, Boston, MA, 8th edition, 2015. ISBN 9781285741550.

[97] L Euler. *Institutionum Calculi Integralis*, volume 1. Academiae Scientiarum Petropolitanae, St. Petersburg, 1768.

[98] J. C Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, 2nd edition, 2008. ISBN 9780470753753.

[99] C Runge. Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, 1895.

[100] M. W Kutta. Beitrag zur näherungsweisen integration totaler differentialgleichungen. *Zeitschrift für Mathematik und Physik*, 46:435–453, 1901.

[101] J. C Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, 1st edition, 2003. ISBN 9780471967586.

[102] A Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, 2nd edition, 2009. ISBN 9780521734905.

[103] T Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[104] W. M Bolstad and J. M Curran. *Introduction to Bayesian statistics*. John Wiley & Sons, 2016.

[105] B Oksendal. *Stochastic Sifferential Equations: An Introduction with Applications*. Springer Science & Business Media, 2013.

[106] S Särkkä and A Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.

[107] F Black and M Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[108] T Sauer. Numerical solution of stochastic differential equations in finance. In *Handbook of computational finance*, pages 529–550. Springer, 2011.

[109] A Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.

[110] S Särkkä, A Solin, A Nummenmaa, A Vehtari, T Auranen, S Vanni, and F.-H Lin. Dynamic retrospective filtering of physiological noise in bold fmri: Drifter. *NeuroImage*, 60(2):1517–1527, 2012.

[111] S Särkkä, V Tolvanen, J Kannala, and E Rahtu. Adaptive kalman filtering and smoothing for gravitation tracking in mobile systems. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–7. IEEE, 2015.

[112] M. S Grewal and A. P Andrews. *Kalman filtering: Theory and Practice with MATLAB*. John Wiley & Sons, 2014.

[113] R. E Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[114] H. E Rauch, F Tung, and C. T Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[115] E. A Wan and R Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000.

[116] A Doucet, A. M Johansen, et al. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

[117] J Quinonero-Candela and C. E Rasmussen. A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

[118] R. E Turner. A unifying framework for sparse gaussian process approximation using power expectation propagation. 2017.

[119] M Alvarez and N Lawrence. Sparse convolved gaussian processes for multi-output regression. *Advances in neural information processing systems*, 21, 2008.

[120] Ø Petersen, O Øiseth, and E Lourens. Wind load estimation and virtual sensing in long-span suspension bridges using physics-informed gaussian process latent force models. *Mechanical Systems and Signal Processing*, 170:108742, 2022.

[121] J Bilbao, E.-M Lourens, A Schulze, and L Ziegler. Virtual sensing in an onshore wind turbine tower using a gaussian process latent force model. *Data-Centric Engineering*, 3:e35, 2022.

[122] J Zou, E.-M Lourens, and A Cicirello. Virtual sensing of subsoil strain response in monopile-based offshore wind turbines via gaussian process latent force models. *Mechanical Systems and Signal Processing*, 200:110488, 2023.

[123] C Van Loan. Computing integrals involving the matrix exponential. *IEEE transactions on automatic control*, 23(3):395–404, 1978.

[124] Y.-K Wen. Method for random vibration of hysteretic systems. *Journal of the engineering mechanics division*, 102(2):249–263, 1976.

[125] C Andrieu, A Doucet, and R Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342, 2010.

[126] K. P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[127] F Lindsten, T Schön, and M Jordan. Ancestor sampling for particle gibbs. *Advances in Neural Information Processing Systems*, 25, 2012.

[128] F Lindsten, T. B Schön, et al. Backward simulation methods for monte carlo statistical inference. *Foundations and Trends® in Machine Learning*, 6(1): 1–143, 2013.

[129] N Chopin and S. S Singh. On particle gibbs sampling. 2015.

[130] P Hennig. Probabilistic Interpretation of Linear Solvers. *SIAM J on Optimization*, 25, 2015.

[131] L Marino and A Cicirello. A switching gaussian process latent force model for the identification of mechanical systems with a discontinuous nonlinearity. *Data-Centric Engineering*, 4:e18, 2023.

[132] R Brincker and C Ventura. *Introduction to operational modal analysis*. John Wiley & Sons, 2015.

[133] B Peeters and G De Roeck. Stochastic system identification for operational modal analysis: a review. *J. Dyn. Sys., Meas., Control*, 123(4):659–667, 2001.

[134] E Reynders. System identification methods for (operational) modal analysis: review and comparison. *Archives of Computational Methods in Engineering*, 19:51–124, 2012.

[135] T Friis, M Tarpø, E. I Katsanos, and R Brincker. Equivalent linear systems of nonlinear systems. *Journal of Sound and Vibration*, 469:115126, 2020.

[136] K. K Vesterholm, R Brincker, and A Brandt. Random decrement technique for detection and characterization of nonlinear behavior. *Mechanical Systems and Signal Processing*, 143:106841, 2020.

[137] E. G Macıas, R Castro-Triguero, R Gallego, J Carretero, and M Gómez-Casero. Operational modal analysis and detection of non-linear structural behavior of bowstring arch bridge. In *International Operational Modal Analysis Conference*, 2015.

[138] R Brincker, L Zhang, and P Andersen. Modal identification of output-only systems using frequency domaindecomposition. *Smart materials and structures*, 10(3):441, 2001.

[139] M. N Gibbs. *Bayesian Gaussian processes for regression and classification*. PhD thesis, Citeseer, 1998.

[140] D. J MacKay et al. Introduction to gaussian processes. *NATO ASI series F computer and systems sciences*, 168:133–166, 1998.

[141] P. D Sampson and P Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

[142] C Paciorek and M Schervish. Nonstationary covariance functions for gaussian process regression. In S Thrun, L Saul, and B Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003.

[143] A. K Qin and P. N Suganthan. Self-adaptive differential evolution algorithm for numerical optimization. In *2005 IEEE congress on evolutionary computation*, volume 2, pages 1785–1791. IEEE, 2005.

[144] J Wenger and P Hennig. Probabilistic linear solvers for machine learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[145] A Abdulle and G Garegnani. A probabilistic finite element method based on random meshes: A posteriori error estimators and bayesian inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 384:113961, 2021.

[146] F Tronarp, H Kersting, S Särkkä, and P Hennig. Probabilistic solutions to ordinary differential equations as non-linear bayesian filtering: A new perspective. 2019.

[147] N Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3):539–552, 2002.

[148] M Viberg. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 31(12):1835–1851, 1995.

[149] L Ljung. *System Identification: Theory for the User*. Pearson Education, 1998.

[150] K Worden, G. R Tomlinson, and K Yagasaki. Nonlinearity in structural dynamics: detection, identification and modeling. *Appl. Mech. Rev.*, 55(2):B26–B27, 2002.

[151] C. M Bishop and N. M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[152] N Chopin, P Jacob, O Papaspiliopoulos, et al. Smc2: A sequential monte carlo algorithm with particle markov chain monte carlo updates. *arXiv preprint arXiv:1101.1528*, 2011.

[153] F.-X Briol, C. J Oates, M Girolami, M. A Osborne, and D Sejdinovic. Probabilistic integration: A role in statistical computation?, 2017.

[154] P Hennig and M Kiefel. Quasi-newton methods: A new direction. *The Journal of Machine Learning Research*, 14(1):843–865, 2013.

[155] M Schober, S Särkkä, and P Hennig. A probabilistic model for the numerical solution of initial value problems, 2017.

[156] J Wang, J Cockayne, O Chkrebtii, T. J Sullivan, and C. J Oates. Bayesian numerical methods for nonlinear partial differential equations. *Statistics and Computing*, 31:1–20, 2021.

[157] J Schmidt, N Krämer, and P Hennig. A probabilistic state space model for joint inference from differential equations and data. *Advances in Neural Information Processing Systems*, 34:12374–12385, 2021.

[158] F Tronarp, N Bosch, and P Hennig. Fenrir: Physics-enhanced regression for initial value problems, 2023.

[159] P. J Davis and P Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.

[160] A Wills, T. B Schön, F Lindsten, and B Ninness. Estimation of linear systems using a gibbs sampler. *IFAC Proceedings Volumes*, 45(16):203–208, 2012.

[161] N Bosch, P Hennig, and F Tronarp. Calibrated adaptive probabilistic ode solvers. In *International Conference on Artificial Intelligence and Statistics*, pages 3466–3474. PMLR, 2021.

[162] J.-P Noel and M Schoukens. Hysteretic benchmark with a dynamic nonlinearity. In *Workshop on nonlinear system identification benchmarks*, pages 7–14, 2016.

[163] N. M Newmark. A method of computation for structural dynamics. *Journal of the engineering mechanics division*, 85(3):67–94, 1959.

[164] R Pintelon and J Schoukens. *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

[165] T Wigren and J Schoukens. Data for benchmarking in nonlinear system identification. *Technical Reports from the department of Information Technology*, 6:2013–006, 2013.

[166] A Janot, M Gautier, and M Brunot. Data set and reference models of emps. In *Nonlinear System Identification Benchmarks*, 2019.

[167] W Khalil and E Dombre. *Modeling identification and control of robots*. CRC Press, 2002.