Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Selma Lerkerød Flo

# Physics-Informed Bayesian Calibration Accounting for Model Discrepancy in a Linearized Homogeneous Ordinary Differential Equation

## A Study of the Pendulum

Master's thesis in Applied Physics and Mathematics
Supervisor: Ingelin Steinsland
July 2023

NTNU
Norwegian University of
Science and Technology

Selma Lerkerød Flo

# Physics-Informed Bayesian Calibration Accounting for Model Discrepancy in a Linearized Homogeneous Ordinary Differential Equation

## A Study of the Pendulum

Master's thesis in Applied Physics and Mathematics
Supervisor: Ingelin Steinsland
July 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

# Abstract

This thesis investigates a physics-informed, fully Bayesian framework for parameter estimation and uncertainty analysis in linearized homogeneous ODEs. The incorporation of a discrepancy term to account for the often intractable, systematic deviation between the simplified model and the true noisy observation process is examined. A comparison is made between a model incorporating this discrepancy term and a model without it. The discrepancy term, modeled as a flexible Gaussian Process, is incorporated into the forcing function of the homogeneous ODE as a virtual force that counterbalances the deviation between the model and the process. Hamiltonian Monte Carlo is used for inference, with the pendulum as a case study.

Our empirical studies demonstrate that the discrepancy term effectively captures the systematic deviation in a test case with observations from a linear process with an added Gaussian Process, accurately recovering parameter estimates in contrast to the non-discrepancy model. Furthermore, the discrepancy model outperforms the non-discrepancy model in estimating noise in a damped process up to a certain level of observation noise. Our findings also suggest that the model's enhanced flexibility from the discrepancy term results in a more effective exploration of the posterior space during HMC iterations. These findings support the inclusion of the discrepancy term on the forcing function in the model. However, a fundamental challenge arises when applying this approach to processes where the true process closely resembles the model process with different parameters. This limitation becomes evident when utilizing the linearized model with a discrepancy term for a nonlinear process. However, these inaccurate estimates lead to an improved fit to the data and better predictions. Overall, this research improves our understanding of physics-informed Bayesian calibration accounting for discrepancy in simplified models and addresses challenges in capturing systematic discrepancies in linearized nonlinear homogeneous processes.

# Samandrag

Denne masteroppgåva utforskar eit fysikkinformert, fullstendig bayesiansk rammeverk for parameterestimering og usikkerheitsanalyse i lineariserte homogene ordinære differensiallikningar (ODE-ar). Vi undersøkjer korleis eit avviksledd kan inkorporerast for å ta omsyn til ofte analytisk ukvantifiserbare, systematiske avvik mellom den forenkla modellen og den faktiske observasjonsprosessen. Vi samanliknar ein modell som inkluderer dette avviksleddet med ein modell som ikkje gjer det. Avviksleddet, som er modellert som en fleksibel Gaussisk Prosess, blir integrert i den homogene ODE-en sin kraftfunksjon som ei virtuell kraft som motverkar avviket mellom modellen og prosessen. Vi brukar Hamiltonian Monte Carlo for inferens, med pendelen som eit case-studie.

Dei empiriske studiane våre viser at avviksleddet effektivt fangar opp systematiske avvik i ein test-case med observasjonar frå ein lineær prosess med ein addert Gaussisk Prosess. Den korrigerar parameterestimata i motsetning til modellen utan avviksledd. Vidare presterer modellen med avviksledd betre enn modellen utan dette leddet når støyen i ein dempa prosess skal estimerast, opp til ei viss grad av observasjonsstøy. Funna våre tyder også på at den auka fleksibiliteten frå avviksleddet i modellen resulterar i meir effektiv utforsking av posteriorirommet under HMC-iterasjonane. Desse funna støttar inkluderinga av avviksleddet i kraftfunksjonen til modellen. Imidlertid møter vi ei grunnleggjande utfordring når modellen brukast på avvikande prosessar der den faktiske prosessen liknar den forenkla modellen, men med ulike parametrar. Denne avgrensinga blir tydeleg når vi brukar den lineariserte modellen med avviksledd på en ikkje-lineær prosess. Likevel fører desse unøyaktige estimata til ei betre tilpassing til dataen og betre prediksjonar. Totalt sett bidreg denne studien til forståelsen vår av fysikkinformert bayesiansk kalibrering som tek omsyn til avvik i forenkla modellar, og adresserer utfordringar knytt til å fange opp systematiske avvik i lineariserte ikkje-lineære homogene prosessar.

# Preface

This master thesis was completed during the spring of 2023, the last semester of my masters in Applied Physics and Mathematics at the Norwegian University of Science and Technology (NTNU), within the Department of Mathematical Sciences.

This thesis is an extension of the groundwork laid during my project thesis in the fall of 2022. However, unforeseen challenges (unrelated to my studies) arose during the semester in which I was working on my project thesis, hindering my ability to grasp the theory thoroughly and resulting in a weak foundation. Consequently, although I incorporated some of my previous work, substantial revisions and refinements were necessary, driven by a deeper understanding of the subject matter.

I would like to extend my gratitude to my supervisor, Ingelin Steinsland, for her guidance and contributions to shaping the trajectory of this thesis. I am grateful for her support and valuable input throughout the process. I would also like to acknowledge the work of Michail Spitieris and Ingelin Steinsland in this field, which served as a foundation of my study. Their research which I often refer to as Spitieris and Steinsland (2022) was published in April this year (Spitieris and Steinsland, 2023). I want to thank Michail for taking the time to provide input in response to my inquiries about his code as I worked on adapting it for my own models.

The development of the thesis began with studying their framework and the careful selection of a case study. Along the way, various modifications were explored with the aim of discovering interesting aspects. This process provided a new, both challenging and rewarding, way of working compared to being given predefined problems to solve.

Lastly, I want to express my appreciation to my fellow classmates. Our engaging academic discussions have been a source of inspiration and enlightenment, and the supportive environment has greatly enriched my five years of study.

*Selma Lerkerød Flo, June 2023*

# Table of Contents

# Chapter 1

# Introduction

Inverse problems, where the goal is to determine the values of model parameters given the observed output or data, arise in a wide range of scientific fields. These problems typically occur in scenarios that involve remote sensing or indirect measurements, where the aim often is to find the internal characteristics of an inaccessible region by analyzing the output measurements obtained from outside sources. Inverse problems also involve reconstructing unobserved events using present-state measurements (Groetsch, 1993).

Examples of inverse problems include reconstructing medical images in CT and MRI from partial measurements (Song et al., 2022), analyzing ground motion caused by earthquake waves to explore the Earth's deep interior (Thurber and Ritsema, 2015), and estimating parameters in climate models that are difficult to measure directly, such as carbon dioxide absorption by oceans and ice melting rates (Flato et al., 2013).

To address inverse problems effectively, the Bayesian framework is commonly used. Methods within this framework use both prior knowledge and observed data to estimate parameters in the form of posterior distributions (Givens and Hoeting, 2012a). These posterior distributions also serve as a foundation for making predictions about unobserved data that capture and quantify the uncertainties in the estimations and predictions (Kennedy and O'Hagan, 2001).

Incorporating prior knowledge about the physics of a system can be highly advantageous when working with inverse problems. Physics-informed priors, often modeled as Gaussian Processes (GPs), contain structured prior knowledge about

the system's behavior (Raissi et al., 2017).

However, when the governing equations used in the physics-informed priors are imprecise, a deviation arises between the true observed physical process and the constructed model, potentially leading to biased and over-confident parameter estimates (Brynjarsdóttir and O'Hagan, 2014). The deviation can result from various factors, such as simplified or omitted physics in the simulator, numerical approximations required for computing simulator outputs, and the limited applicability of the simulator's assumptions in different real-world scenarios (Gardner et al., 2021).

A possible approach to address this issue is introducing a *discrepancy term* to our model, which aims to capture the deviation between the process and the model (Kennedy and O'Hagan, 2001). This term represents the uncertainty stemming from the, often intractable, systematic deviation between the model and the true observed process, distinct from the independent observation noise. Due to its desirable properties, a flexible Gaussian Process (GP) is typically used to model this discrepancy.

In 2001, Kennedy and O'Hagan (2001) introduced a Bayesian framework accounting for model discrepancy, which has been further developed in numerous studies, including Higdon et al. (2004), Bayarri et al. (2007) and Brynjarsdóttir and O'Hagan (2014). Spitieris and Steinsland (2022) combined this with the theory about physics-informed priors for linear differential equations (Raissi et al., 2017), in a fully Bayesian framework. The research conducted by Spitieris and Steinsland (2022) compares two models: one with a discrepancy term added to one of the processes in the covariance matrix and another without accounting for discrepancy. The study examines various kernel functions, computational complexity reduction, and big data approximations. They found that the model accounting for discrepancy can recover the true parameters of the physical models when the reality is more complex than the model. Their approach offers computational advantages over conventional Bayesian methods, resulting in faster analysis and inference. Their findings are illustrated through two examples; the heat flow process and the Windkessel model.

This thesis further investigates the impact of accounting for model discrepancy in parameter estimation. The study of Spitieris and Steinsland (2022) mainly focused on time-dependent ODEs and linear space-time dependent inhomogeneous PDEs, both linear. We investigate this approach on linearized nonlinear homogeneous ODEs and linear homogeneous ODEs.

While recent techniques propose physics-informed priors using *numerical* Gaussian process regressions at each timestep for modeling linearizations of nonlinear processes (Raissi et al., 2018), our study takes a simpler approach. We approximate

the nonlinear process by a linearized model with a discrepancy term represented by a flexible GP. This term aims to capture the deviation between the nonlinear process and the linear model, avoiding the complexity of physics-informed priors for nonlinear systems. We use the Markov Chain Monte Carlo (MCMC) method Hamiltonian Monte Carlo (HMC) extension No-U-Turn sampler (NUTS) in STAN to obtain estimates for the posterior distributions of the parameters of interest.

In contrast to Spitieris and Steinsland (2022), who incorporates the discrepancy term into the observed process of interest in the structure of the physics-informed prior, we take a different approach. We move the discrepancy term to the homogeneous side of the differential equation, often referred to as the force function, and interpret it as an external force that influences the process to conform to the model's behavior. We incorporate this external force as a discrepancy term in the physics-informed prior, assuming it follows a Gaussian Process. This is done with the expectation of facilitating the model's ability to identify discrepancies in a constant function.

We use the pendulum process as a case study to assess the differences between a model that accounts for discrepancy and a model that does not. For an idealized pendulum system characterized by a nonlinear homogeneous ODE describing the angular displacement, the angle can be approximated using linearization in the context of small angles. This linearization can be incorporated in the physics-informed priors, simplifying the analysis but introducing deviations to the original nonlinear process. The inclusion of a discrepancy term in the physics-informed priors aims to capture this deviation.

We use synthetically simulate noisy observations obtained from various pendulum processes to assess the model's performance. We then explore how the model responds to different types of systematic deviations between the model and the observed process. This investigation evaluates the model's ability to recover accurate parameter estimates in a linear pendulum process with a controlled systematic deviation, a nonlinear idealized process, and a nonlinear damped process.

By studying the pendulum system, we can gain insights that apply to other physical systems approximated by homogeneous linear ODEs, expanding our understanding to more complex systems with similar characteristics. The pendulum can represent problems that can only be imprecisely described by a linear homogeneous ODE, resulting in a disparity between the model and the true underlying process. Our investigation assesses the model's ability to account for and capture the discrepancy arising from the linearized approximation or other types of systematic discrepancies between the model and the observed process.

To explore the differences between the model accounting for discrepancy and the

model that does not, the following experiments are conducted:

**E1:** A comparative analysis is performed between the model that incorporates the discrepancy and the model that does not. The two models are compared across various behaviors of the deviation between the model and the actual process. The evaluation focuses on the model's capacity to capture and differentiate these deviations from the rest of the process.

**E2:** The sensitivities of the two models are compared and evaluated for the following factors:

- **(a)** The amount of observation noise in the observations.
- **(b)** The number of cycles in the observations.

Chapter 2 introduces the governing equations for the different pendulum processes and their differences. Chapter 3 provides the theoretical background for our model approach, while Chapter 4 presents the method of simulating observation and constructing the physics-informed Bayesian models of the pendulum, with and without a discrepancy term, including our shift of the discrepancy term. In Chapter 5, experimental configurations are presented, and Chapter 6 presents and discusses the results. A summary and conclusion of the study are found in Chapter 7.

# Chapter 2

# The Dynamics of the Pendulum and Difference Between Pendulum Models

This chapter examines the pendulum process and its approximation through linearization. This is later used as a case study to examine the performance of our models, as it serves as a simple model that can help understand other systems with similar characteristics. This chapter is a revised and extended version of the chapter introducing the pendulum process in my unpublished project thesis.

## 2.1 The Idealized Pendulum (TP)

An idealized pendulum consists of a weight attached to the end of a massless rod, only affected by gravity. A pendulum setup is illustrated in Figure 2.1. When released from an initial angle $u_0$ at time $t = 0$, the pendulum swings back and forth under the influence of gravity.

The motion of the idealized pendulum can be described by a homogeneous nonlinear differential equation, which can be derived using Newton's Second Law of Motion or Lagrangian mechanics (Owen, 2014):

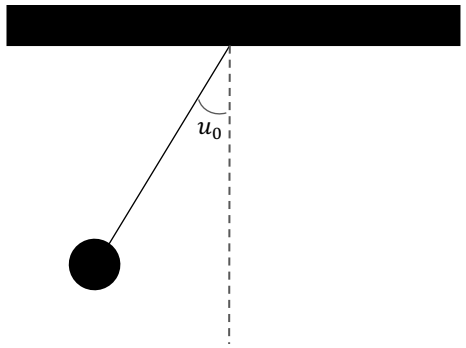



Figure 2.1: Illustration of the pendulum with initial angle $u_0$.

$$\frac{d^2u(t)}{dt^2} + \frac{1}{R}sin(u(t)) = f(t), \quad f(t) = 0. \tag{2.1}$$

Here, $u(t)$ is the angle at time $t$ and $\frac{1}{R} = \frac{g}{L}$, where $g$ is the gravitational constant and $L$ the length of the rod. Note that $\frac{1}{R}$ is commonly denoted as $\omega_0^2$ in literature, and that the unit is $\left[\frac{1}{R}\right] = \left[\omega_0^2\right] = \frac{1}{s^2}$. Equation (2.1), describing the true idealized process, will be referred to as TP.

The term $f(t)$ represents an external force applied to the pendulum. In a non-idealized process, this force could come from various sources, such as air resistance, friction in the pendulum's pivot point, or an external motor driving the pendulum's motion. $u(t)$ then describes the motion of the pendulum as it responds to this external force, with the first term on the left-hand side representing the pendulum's inertia and the second term representing the restoring force due to gravity. In the idealized pendulum, $f(t) = 0$, meaning that there is no external force acting on the pendulum.

There is no analytical solution to this equation, but we can solve it numerically with methods like RK4 (Butcher, 2008). In the case of simulating the angular displacement of the pendulum, we need constraints on the initial angle $u(0)$ and velocity $\frac{du(0)}{dt}$.

## 2.2 Linearization of the Pendulum (LP)

By looking at the Taylor expansion of the sine function, given by

$$sin(u) = u - \frac{u^3}{3!} + \frac{u^5}{5!} - \frac{u^7}{7!} + ..., \tag{2.2}$$

we see that for sufficiently small values of $u$, $sin(u) \approx u$ is a good approximation. Then the governing equation becomes

$$\frac{d^2u(t)}{dt^2} + \frac{1}{R}u(t) = f(t), \quad f(t) = 0. \tag{2.3}$$

We refer to the process of Equation (2.3) as the linearized process (LP). An advantage of this linearization is that it has an analytic solution for the angle displacement:

$$u(t) = u_0 \cos\left(\sqrt{\frac{1}{R}}t\right). \tag{2.4}$$

The error of this linearization is given by the higher-order terms in the Taylor series expansion in Equation (2.2).

We compare the linearized process (LP) to the true process (TP) for various initial angles in Figure 2.2. We observe a systematic displacement error between TP and LP, caused by the approximation $\sin u(t) \approx u(t)$ which amplifies with the initial angle $u_0$.

In Figure 2.3a, we see that the difference between LP and TP has a cyclic pattern that increases with time and continues to rise until LP deviates by more than half a cycle from TP. Then the error starts decreasing. Eventually, the LP's higher frequency allows it to catch up to the TP when it is one wavelength ahead, and the discrepancy becomes zero again. This cycle repeats, indicating that the discrepancy is a periodic shift. Figure 2.4a illustrates this discrepancy for a shorter time period.

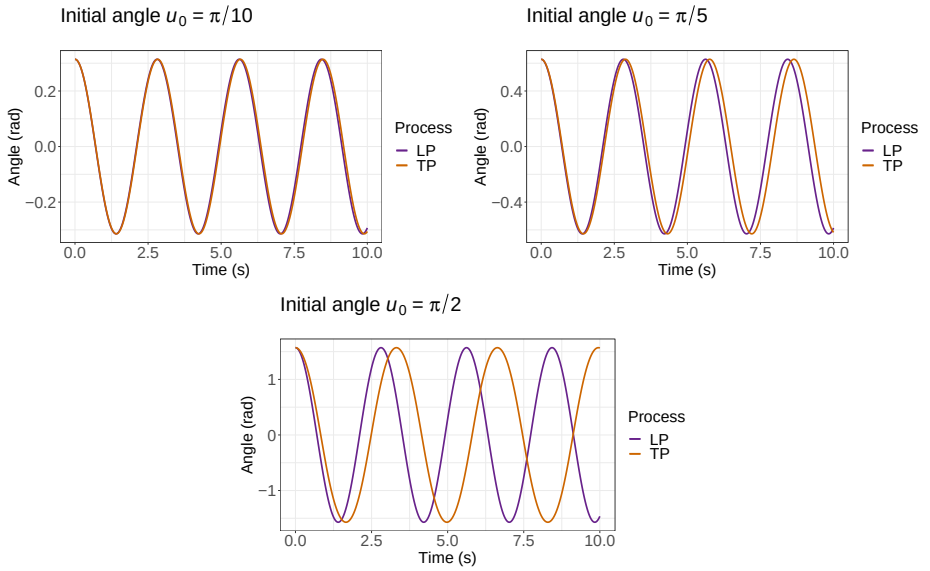
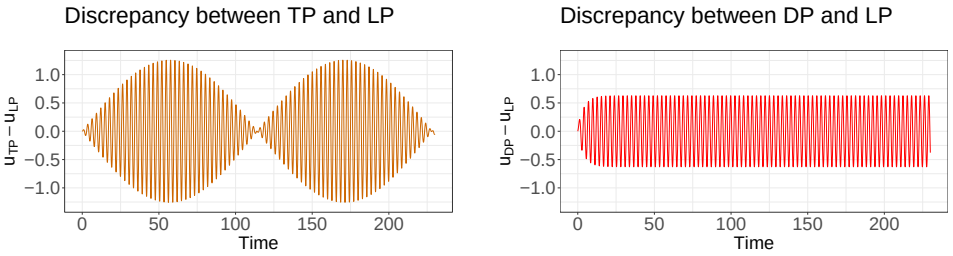
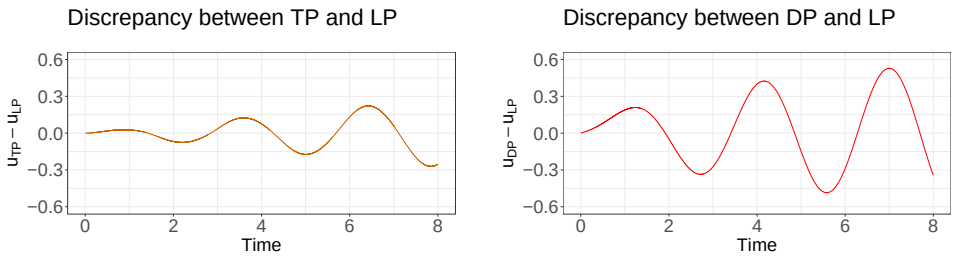



Figure 2.2: Plot of pendulum angle motion $u(t)$, with initial angles $u_0 = \pi/10$, $u_0 = \pi/5$ and $u_0 = \pi/2$ respectively. The purple line shows the simulated motion of the true idealized process (TP) from Equation (2.1) while red shows the simulated motion of the linearized approximation (LP) from Equation (2.3).




(a) $u_{TP} - u_{LP}$ (b) $u_{DP} - u_{LP}$

Figure 2.3: Difference between the angular motion of TP (true idealized process) and LP (linearized process), and DP (damped process) and LP, respectively, for 8 seconds of the pendulum motion.




(a) $u_{TP} - u_{LP}$ (b) $u_{DP} - u_{LP}$

Figure 2.4: Discrepancy between the angular motion of TP (true idealized process) and LP (linearized process), and DP (damped process) and LP, respectively, for 8 seconds of the pendulum motion.

## 2.3 Pendulum with Damping (DP)

The idealized pendulum (TP) discussed above would oscillate with a constant amplitude indefinitely. In the real world, energy dissipates, due to for instance friction and air resistance, causing the amplitude of the pendulum's motion to gradually decrease until it eventually comes to rest. A system like this, known as a damped pendulum, introduces additional complexity to the calculations by breaking energy conservation within the system. Following the formulation by Chasnov (2022), the equation of motion for a damped pendulum is given by:

$$\frac{d^2u(t)}{dt^2} + \lambda\frac{du(t)}{dt} + \frac{1}{R}sin(u(t)) = f(t), \quad f(t) = 0 \tag{2.5}$$

Here, $\lambda = \mu/m$ with $\mu$ representing the friction coefficient and $m$ denoting the pendulum's mass. We refer to this system as the damped process (DP).

A simulation of the damped pendulum, obtained using Euler's method, together with a simulation of LP with the corresponding physical parameters, is shown in Figure 2.5. The observed difference between the processes is attributed to both the linearization $\sin u(t) \approx u(t)$ and the damping effect caused by the term $\lambda\frac{du(t)}{dt}$ in DP that is not present in LP. Linearization introduces a shift in time ($t$), while damping influences the amplitude difference in the $y$-direction, both of which contribute to an increasing discrepancy as time progresses.

Figures 2.3b and 2.4b illustrate the discrepancy $u_{DP} - u_{LP}$ between the linear

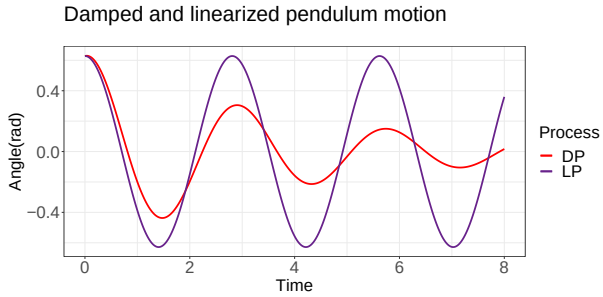



Figure 2.5: Simulations from the damped pendulum (DP) with initial angle $u_0 = \pi/5$, $R = 0.2$ and damping parameter $\lambda = 0.5$, together with the linearized pendulum LP. The pendulum's behavior is observed over a period of $\Delta t = 8s$.

process (LP) and the damped process (DP) over different time durations of 230 and 8 seconds, respectively. In the simulations, DP has a damping coefficient of $\lambda = 0.5$, which causes it to come to rest after only a few cycles gradually. In Figure 2.4b, we see that the error between the LP and DP increases over time. Conversely, Figure 2.3b demonstrates that once the pendulum reaches a state of rest, the error stops growing and exhibits constant periodic behavior opposite to LP. This is because the LP continues oscillating while DP remains constant at zero.

## 2.4 Difference Between Pendulum Models

Next, we examine the energy behavior of the systems. In a mechanical system (Goldstein et al., 2002),

$$E(t) = E_p(t) + E_k(t) \tag{2.6}$$

where the potential energy $E_p$ is given by

$$E_p(t) = mgL\,(1 - \cos(u(t))) \tag{2.7}$$

and the kinetic energy $E_k$ is given by

$$E_K(t) = \frac{1}{2}L^2 m \left(\frac{du(t)}{dt}\right)^2 \tag{2.8}$$

Figure 2.6 displays the energy plots for LP, TP, and DP. For TP, the energy remains constant. The small oscillations observed in the energy plot are due to errors in the numerical simulation, and vanish by choosing a sufficiently small step length.

The mechanical energy of LP exhibits periodicity, indicating periodic energy gain and loss. This periodicity can be interpreted as the corresponding TP being driven and decelerated by an external periodic force,

$$\frac{d^2u(t)}{dt^2} + \frac{1}{R}sin(u(t)) = f_p(t), \quad f_p(t) = -\frac{1}{R}\left(\frac{u(t)^3}{3!} - \frac{u(t)^5}{5!} + \frac{u(t)^7}{7!} - ...\right). \quad (2.9)$$

Lastly, we examine the mechanical energy in DP, shown in Figure 2.6. Here, the periodic curve also displays a downward trend, signifying energy dissipation over time. This energy dissipation can be interpreted as the corresponding TP with an external damping force,

$$\frac{d^2u(t)}{dt^2} + \frac{1}{R}sin(u(t)) = f_d(t), \quad f_d(t) = -\lambda\frac{du(t)}{dt}. \quad (2.10)$$

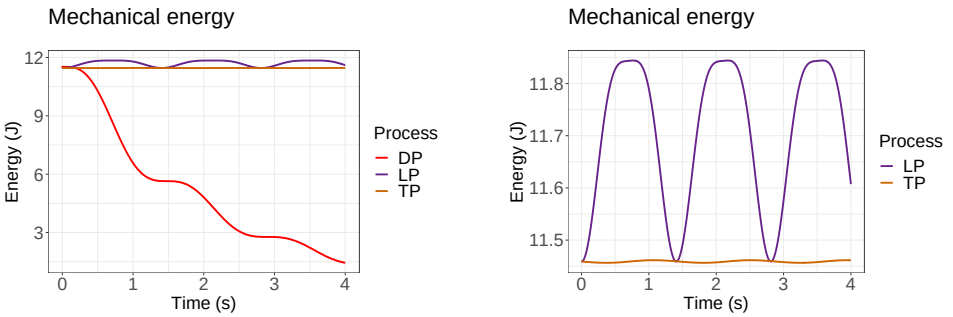



Figure 2.6: The left plot shows the mechanical energy in the pendulum processes $E(t) = E_P(t) + E_K(t)$ for TP (true idealized process), LP (linearized process) and DP (damped process) over a simulation period $\Delta t = 4s$. The right plot only shows TP and LP in order to see the differences between the two better.

# Chapter 3

# Background

This chapter provides an overview of the background theory that forms the foundation of the physics-informed Bayesian Calibration method that we later explore experimentally.

We first introduce Bayesian inference, a widely recognized methodology that enables the integration of prior knowledge and the quantification of uncertainty. We discuss the utilization of Markov Chain Monte Carlo (MCMC) methods to address computational challenges in Bayesian analysis, focusing on the mechanisms of MCMC and briefly presenting the Hamiltonian Monte Carlo (HMC) method. Additionally, we present convergence diagnostics used when assessing inferences' reliability in the MCMC process.

We then introduce physics-informed priors (Raissi et al., 2017) within the Bayesian calibration through the Gaussian Process assumption and an approach to account for model discrepancy (Kennedy and O'Hagan, 2001), and how these theories are combined and integrated into a fully Bayesian framework (Spitieris and Steinsland, 2022).

## 3.1 Bayesian Inference

In this section, we present the widely established and recognized theory on Bayesian Inference, with Gelman, Carlin et al. (2013) serving as our primary reference.

The Bayesian approach incorporates prior knowledge and observed data for infer-

ence, unlike Frequentist statistics that rely solely on observations. Bayesian data analysis can be understood as a three-step process: Model specification, posterior inference, and model assessment (and possibly iterating these steps if needed).

Let $\theta$ be the vector containing the unknown parameters. In the Bayesian inference framework, the elements of $\theta$ are treated as random variables, and the posterior distribution of the parameters given the observed data $y$ is calculated. This is done by Bayes Theorem, i.e., multiplying the prior distribution of $\theta$ by the likelihood of the data given the parameter and then normalizing by the marginal likelihood of the data.

Thus, the posterior distribution of $\theta$ given the observations $y$ is expressed as

$$\overbrace{p\left(\theta|y\right)}^{\text{posterior}} = \frac{\overbrace{p\left(y|\theta\right)}^{\text{likelihood}}\ \overbrace{p\left(\theta\right)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}} \tag{3.1}$$

Here, $p(\theta)$ is the prior distribution of $\theta$, which means it represents what is known about $\theta$ *without* knowledge of the data $y$. The posterior distribution $p\left(\theta|y\right)$, on the other hand, represents our revised knowledge about $\theta$ due to the data $y$. The likelihood function $p(y|\theta)$ (also denoted as $L(\theta|y)$) gives the likelihood of observing $y$ with the given hypothesis of $\theta$ (Givens and Hoeting, 2012b).

The denominator $p(y) = \int p(y|\theta)p(\theta)d\theta$ (or $p(y) = \sum p(y|\theta)p(\theta)$ if $\theta$ is discrete) is the marginal likelihood, or evidence, of the data. We can view the marginal likelihood of the data as a normalizing constant that ensures the posterior distribution $p(y|\theta)$ integrates or sums to one. Therefore, Bayes Theorem is commonly expressed as follows:

$$\text{posterior} \ \propto \ \text{likelihood} \ \times \ \text{prior} \tag{3.2}$$

To obtain the posterior distribution, one needs to specify priors, affecting the results. Choosing an appropriate prior may be challenging if little is known about the parameters of interest.

Obtaining the posterior density distribution of a single component $\theta^i$ (the marginal posterior) can also be challenging. This distribution is obtained through integration as follows:

$$p\left(\theta^i \mid y\right) = \int p\left(\theta \mid y\right) d\theta^1 \ldots d\theta^{i-1} d\theta^{i+1} \ldots d\theta^d \tag{3.3}$$

This is used to find the minimum mean squared error (MMSE) estimation of one component $\theta^i$,

$$\hat{\theta}^i_{MMSE} = E\left(\theta^i \mid y\right) = \int_{\theta_i} \theta^i p\left(\theta^i \mid y\right) d\theta^i, \tag{3.4}$$

and evaluate confidence of the estimation of $\theta^i$ by assessing its credible interval width.

However, in most real applications, the integral in Equation (3.3) is highly dimensional and may not have a closed form, which makes it difficult or impossible to solve analytically. Therefore, numerical methods such as Markov Chain Monte Carlo (MCMC) are used to approximate the posterior distribution.

## 3.2 Markov Chain Monte Carlo (MCMC)

MCMCs are widely used simulation-based techniques for solving complex problems in high-dimensional spaces when we cannot work it out analytically (Andrieu et al., 2003). MCMC methods are widely used in scientific analyses to numerically estimate model parameters, generate predictions and assess uncertainties using Bayesian inference, often in large dimensional spaces.

### 3.2.1 The Conceptual Framework of MCMC

The main idea behind Monte Carlo simulation methods is that we want to draw a set of independent and identically distributed (i.i.d.) samples $\{\theta^{(i)}\}_{i=1}^N$ from a target density $p(\theta)$ we cannot sample directly from, defined on a (possibly) high-dimensional space $\mathcal{X}$. The methods approximate the target density by obtaining a sufficient number of these samples. The set $\mathcal{X}$ can represent system configurations, posterior spaces, or feasible solution sets.

Markov Chain Monte Carlo (MCMC) draws serially correlated samples from a proposal distribution instead of independent ones. However, the samples form a Markov Chain of $N$ random variables where two draws $\theta^{(i)}$ and $\theta^{(i+n)}$ become increasingly independent as $n$ increases (Robert and Casella, 1999).

If the effective sample size ($N_{eff}$ or $ESS$), which indicates how many of the simulated samples are equivalent to independent samples, is sufficiently large, then a large portion of the sample's empirical distribution can serve as a reliable approximation of the target distribution.

The probabilistic Programming language for statistical inference, STAN, is commonly used as an MCMC sampler for Bayesian analysis.

### 3.2.2 Metropolis Hastings

The Metropolis Hastings (MH) method is a commonly used MCMC method for obtaining a sequence of random samples from the target distribution, where a new position $\theta^*_{i+1}$ is proposed by generating a sample from the proposal distribution $Q(\theta'_{i+1}|\theta_i)$ (Robert and Casella, 1999):

$$\theta'_{i+1} \sim Q(\theta'_{i+1}|\theta_i). \tag{3.5}$$

The transition probability $T(\theta'_{i+1}|\theta_i)$ given by

$$T(\theta'_{i+1}|\theta_i) = \min\left(1, \frac{p(\theta'_{i+1}|y)}{p(\theta_i|y)} \frac{Q(\theta_i|\theta'_{i+1})}{Q(\theta'_{i+1}|\theta_i)}\right) \tag{3.6}$$

determines whether the proposed step is accepted or rejected. If the step is accepted, we set $\theta_i = \theta'_{i+1}$ and continue iterating. If the step is rejected, the previous point is kept, and we propose a new step. This process is repeated many times to generate a large number of samples from the posterior distribution.

### 3.2.3 Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS)

Hamiltonian Monte Carlo (HMC) is an advanced version of the MH algorithm. HMC offers two notable advantages over standard MH methods: reduced sample correlation, resulting in a higher effective sample size (ESS), and improved detection of convergence issues (Honkela, 2020).

HMC efficiently explores the posterior distribution of a given density function using concepts from Hamiltonian dynamics and molecular dynamics (Duane et al.,

1987). By simulating Hamiltonian dynamics and incorporating gradient information, HMC aims to enhance chain convergence (Betancourt, 2018).

The No-U-Turn Sampler (NUTS) extends HMC by simulating the motion of a fictitious particle until it makes a U-turn, serving as a diagnostic for improper posteriors, making it even more efficient (Hoffman and Gelman, 2014). HMC-NUTS is the standard method in STAN. Instead of going into extensive technical explanations of HMC and HMC-NUTS, our focus lies on its implications for evaluating convergence diagnostics to avoid inaccurate interpretation of results.

### 3.2.4 Diagnostic Tools for Assessing MCMC Convergence

Markov Chain Monte Carlo (MCMC) algorithms aim to sample from the target posterior distribution by reaching convergence. However, due to finite sample sizes, assessing convergence in practical sampling is challenging. To evaluate MCMC convergence, several measures can be used, and the convergence diagnostics we present are among those presented in Givens and Hoeting (2012).

Trace plots visually represent the sampled parameter values over iterations. Stable and smooth patterns suggest convergence, while unstable behavior indicates mixing or convergence problems. Running multiple chains with different initial values helps explore all important characteristics of the target distribution. By comparing the behavior of these chains, we can identify issues and improve mixing if necessary.

Another diagnostic tool is the potential scale reduction factor $\hat{R}$. It compares within-chain and between-chain variances to assess convergence. Values close to 1 indicate convergence, while significantly larger values indicate lack of convergence.

Autocorrelation function plots (ACF) show the correlation between samples at different lags, indicating poor mixing if autocorrelation decays slowly. Cross-correlations between parameters can also indicate convergence issues.

The effective sample size (ESS) measures the number of independent samples obtained from $N$ samples of the Markov Chain, given by

$$ESS = \frac{N}{1 + 2\sum_{t=-\infty}^{\infty} \rho_t}. \tag{3.7}$$

Here, the sum over $t$ represents the summation of the ACF values $\rho_t$ across all possible lags $t$. Higher ESS values indicate better convergence and more efficient sampling.

## 3.3 Gaussian Processes

Having established the foundations of Bayesian inference, we now focus on Gaussian Processes (GPs), which are used in our physics-informed priors. Gaussian Processes provide a robust framework for modeling and predicting complex systems, offering a flexible and probabilistic approach to capture uncertainties in predictions (Raissi et al., 2017). This section presents relevant theory from Rasmussen and Williams (2006), Gaetan and Guyon (2010) and Gramacy (2022).

A Gaussian Process (GP) is a stochastic process where every set of realizations follows a multivariate normal distribution. This means that a GP denoted as $g(t)$ can be characterized by its mean vector $\mu : X \to \mathbb{R}$ and kernel $K : X \times X \to \mathbb{R}^+$. The kernel is also referred to as the covariance function. For a GP $g(t)$, we use the notation $g(t) \sim GP(\mu(t), K_{gg}(t, t'))$. The notation indicates that the GP $g(t)$ is drawn from a multivariate normal distribution with mean function $\mu(t)$ and covariance $K_{gg}(t, t') = Cov(g(t), g(t'))$ that gives the covariance between the GP $g(t)$ at times $t$ and $t'$.

The GP assumption relies on the premise that it is reasonable to capture the process of $g(\cdot)$ in a joint normal distribution. This assumption can be useful for making predictions about data based on prior knowledge and observations, as the predictions are given as probability distributions. Given their convenience, flexibility, and ability to provide realistic representations, GPs are widely employed in statistical theory and modeling.

Some commonly used kernels are the linear kernel, the squared exponential kernel, the periodic kernel, the Matérn kernel, the rational quadratic kernel, and the neural network kernel. In this work, we use the squared exponential kernel

$$K^{SE}(t, t') = \sigma^2 e^{-0.5\left(\frac{t-t'}{l}\right)^2}. \tag{3.8}$$

where $\sigma$ and $l$ are the hyperparameters determining the kernel's behavior. The squared exponential kernel is widely used due to its smoothness, interpretability of hyperparameters, and other desirable properties. Specifically, it is infinitely differentiable for all values of the parameters $\sigma$ and $l$.

The length scale hyperparameter $l$ determines the extent of correlation between points in the process. If $l$ is small, points are correlated over a short length scale, while a large $l$ indicates correlation over a longer length scale (observations that are more distant from each other still exhibit a detectable level of correlation).

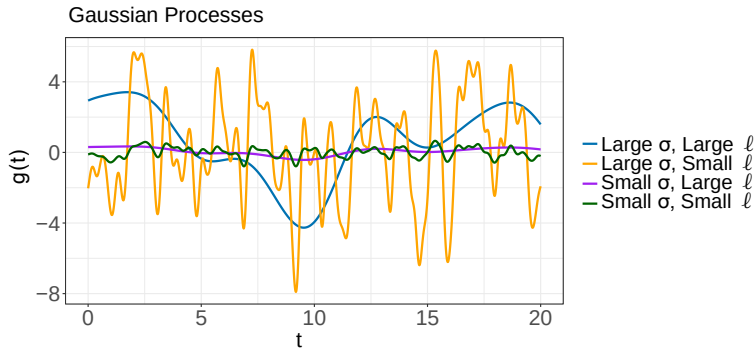



Figure 3.1: GPs for different magnitudes of scale parameters $\sigma$ and length scale parameters $\ell$. One realization of each process.

The output scale hyperparameter $\sigma$ determines the overall scaling of the output values, representing the average distance of the function from its mean. The variances at each time point $t$ are given by $\sigma^2$, which serves as the diagonal elements of the covariance matrix. The value of $K(t, t')$ should equal $\sigma^2$ when $t = t'$.

Figure 3.1 illustrates the relationship between these properties and the magnitude of the parameters $\sigma$ and $\ell$. Each parameter combination is represented by a single realization.

Two advantageous properties of GPs are as follows:

**Property 1.** *The sum of two GPs remains a GP.*

**Property 2.** *The derivatives of a GP remain a GP for a differentiable covariance function.*

Property 2 states that

$$\mathrm{Cov}\left(g(t), \frac{\partial g\left(t'\right)}{\partial t'}\right) = \frac{\partial K_{gg}\left(t, t'\right)}{\partial t'} \tag{3.9}$$

and

$$\mathrm{Cov}\left(\frac{\partial g(t)}{\partial t}, \frac{\partial g\left(t'\right)}{\partial t'}\right) = \frac{\partial^2 K_{gg}\left(t, t'\right)}{\partial t \partial t'} \tag{3.10}$$

for a differentiable covariance function.

## 3.4 The Linear Differential Operator

We present the linear differential operator $\mathcal{L}_t^{\mathbf{\Phi}}$, which enables us to derive expressions for the physics-informed covariance matrix (Raissi et al., 2017). We can express the underlying linear differential equation governing the physical process in the following form:

$$\mathcal{L}_t^{\mathbf{\Phi}} u(t) = f(t), \tag{3.11}$$

with

$$\mathcal{L}_t^{\Phi} = \Phi_n \frac{d^n}{dt^n} + \Phi_{n-1} \frac{d^{n-1}}{dt^{n-1}} + \ldots + \Phi_1 \frac{d}{dt} + \Phi_0 \tag{3.12}$$

Here, $\mathcal{L}$ represents the linear differential operator, the subscript $t$ denotes the derivative with respect to $t$, and $\Phi$ is a vector containing the unknown physical parameters involved in the equation. Note that elements of the vector $\Phi$ can take a value of zero.

## 3.5 Physics-Informed Priors

In their work, Raissi et al. (2017) introduced physics-informed multi-output Gaussian Process priors that aim to systematically integrate the structural properties of the underlying physical model into the covariance function. This section gives an overview of relevant theory presented in their research.

The motivation for this approach is to enhance the model by incorporating prior knowledge regarding the relationship between measurements, considering their spatial or temporal separation. This relationship is effectively captured within the covariance function discussed in Chapter 3.3. Raissi et al. (2017) employ a linear differential operator (Chapter 3.4) to construct a multi-output GP. They consider a system described by

$$\mathcal{L}_t^{\mathbf{\Phi}} u(t) = f(t), \tag{3.13}$$

Assuming that a $u(t)$ follows a GP, it can be expressed as

$$u(t) \sim GP(\mu_u(t), K_{uu}(t, t'|\theta)) \tag{3.14}$$

where $\theta$ are the hyperparameters of the kernel. For instance, for the squared exponential in Equation (3.8), $\theta$ would have consisted of $\sigma$ and $\ell$.

In cases involving multiple GPs that exhibit interdependence, such as $u(t)$ and $f(t)$ with their relationship described in Equation (3.13), we can construct a multivariate GP of $u(t)$ and $f(t)$ that represents the linear correlation between two GPs. Using Properties 1 and 2, we have that

$$\begin{aligned} u(t) &\sim GP(\mu_u(t), K_{uu}(t, t'|\theta)) \\ \mathcal{L}_t^{\Phi} u(t) = f(t) &\sim GP(\mu_f(t), K_{ff}(t, t'|\theta, \Phi)) \end{aligned} \tag{3.15}$$

where $\Phi$ is the vector containing the physical parameters of the system, $\mu_u(t)$ is the mean function of $u(t)$ and the kernel $K_{uu}(t, t') = Cov(u(t), u(t'))$ describes the dependency between $u(t)$ and $u(t')$, meaning how similar the angle magnitude at two different times $t$ and $t'$ are. Likewise, $K_{ff}(t, t') = Cov(f(t), f(t'))$, describes the relationship between $f(t)$ and $f(t')$. Note that $\Phi$ containing the physical parameters now has become a hyperparameter of the kernel $K_{ff}(t, t'|\theta, \Phi))$.

The covariance between the two GPs can be written as

$$\begin{aligned} Cov(u(t), f(t)) &= K_{uf}(t, t'|\theta, \Phi) \\ Cov(f(t), u(t)) &= K_{fu}(t, t'|\theta, \Phi) \end{aligned} \tag{3.16}$$

Combining Equations (3.15) and (3.16), the joint distribution of $u(t)$ and $f(t)$ can be written as

$$\begin{bmatrix} u(t) \\ f(t) \end{bmatrix} = GP\left(\mu(t) = \begin{bmatrix} \mu_u(t) \\ \mu_f(t) \end{bmatrix}, K(t, t') = \begin{bmatrix} K_{uu}(t, t'|\theta) & K_{uf}(t, t'|\theta, \Phi) \\ K_{fu}(t, t'|\theta, \Phi) & K_{ff}(t, t'|\theta, \Phi) \end{bmatrix}\right). \tag{3.17}$$

Here, $K(t, t')$ is $(N_u + N_f) \times (N_u + N_f)$-dimensional, where $N_u$ is the dimension of $u(t)$ and $N_f$ is the dimension of $f(t)$. Note that the assumption of a multivariate normal distribution might not accurately reflect the true nature of the data but is often a useful assumption.

Next, we aim to determine the expressions for the block elements $K_{uf}(t, t'|\theta, \Phi)$, $K_{fu}(t, t'|\theta, \Phi)$, and $K_{ff}(t, t'|\theta, \Phi)$ in the covariance matrix $K(t, t')$. The expression

of $K_{uu}(t, t'|\theta)$, is determined by the choice of kernel with hyperparameters $\theta$.

By Properties 1 and 2, we have that for any linear differential operator $\mathcal{L}_t^{\Phi}$ describing the relationship between $u(t)$ and $f(t)$, the covariance matrix expressions are given by (Särkkä, 2011):

$$K_{fu}(t, t'|\theta, \Phi) = \mathcal{L}_t^{\mathbf{\Phi}} K_{uu}(t, t'|\theta) \tag{3.18a}$$

$$K_{uf}(t, t'|\theta, \Phi) = \mathcal{L}_{t'}^{\mathbf{\Phi}} K_{uu}(t, t'|\theta) \tag{3.18b}$$

$$K_{ff}(t, t'|\theta, \Phi) = \mathcal{L}_t^{\mathbf{\Phi}} \mathcal{L}_{t'}^{\mathbf{\Phi}} K_{uu}(t, t'|\theta). \tag{3.18c}$$

The essential advantage of this approach is the construction of a multi-output GP that encompasses both $u(t)$ and $f(t)$, eliminating the necessity for a numerical solution of the governing differential equation (Spitieris and Steinsland, 2022).

## 3.6 Physics-Informed Bayesian Calibration

Model calibration involves updating the unknown parameters of a model to align its outputs with observed data or uncertainties (Higdon et al., 2004; Kennedy and O'Hagan, 2001). This calibration process allows for accurate predictions of the behavior of a specific process within a given context by estimating the appropriate inputs.

We put the theory of physics-informed priors into a fully Bayesian framework, following Spitieris and Steinsland (2022).

Consider noisy observations $y_{obs}$ given by

$$y_{obs} = \begin{bmatrix} u_{obs} \\ f_{obs} \end{bmatrix} = \begin{bmatrix} u(t_u) + \epsilon_u \\ f(t_f) + \epsilon_f \end{bmatrix}, \tag{3.19}$$

where $t_u$ and $t_f$ are the vectors containing the temporal or spatial values of which we have observations $u_{obs}$ and $f_{obs}$ with measurement noises $\epsilon_u \sim \mathcal{N}(0, \sigma_u^2)$ and $\epsilon_f \sim \mathcal{N}(0, \sigma_f^2)$, respectively. We have the multi-output GP

$$p(y_{obs}|\theta, \Phi, \sigma_u, \sigma_f) = \mathcal{N}(\mu, K) \tag{3.20}$$

where $\mu$ is the vector containing the mean functions and can generally be described as

$$\mu = \begin{bmatrix} \mu_u(t_u|\beta) \\ \mu_f(t_f|\beta, \Phi) \end{bmatrix} \tag{3.21}$$

(Spitieris and Steinsland, 2022), where $\beta$ comprises the parameters of the mean functions. However, unlike Spitieris and Steinsland (2022) we focus solely on scenarios with constant mean functions, so we only refer to the mean vector as

$$\mu = \begin{bmatrix} \mu_u \\ \mu_f \end{bmatrix}. \tag{3.22}$$

The covariance matrix takes the following form:

$$K = \begin{bmatrix} K_{uu}(t_u, t'_u|\theta) + \sigma_u^2 I_u & K_{uf}(t_u, t'_f|\theta, \Phi) \\ K_{fu}(t_f, t'_u|\theta, \Phi) & K_{ff}(t_f, t'_f|\theta, \Phi) + \sigma_f^2 I_f \end{bmatrix} \tag{3.23}$$

In Chapter 3.1, the parameters of interest was denoted by $\theta$ for generality. Here we we want to obtain estimates for all parameters $\alpha = (\theta, \Phi, \sigma_u, \sigma_f)$, where $\theta$ contains the model parameters, $\Phi$ contains the physical parameters and $\sigma_u^2$ and $\sigma_f^2$ are the measurement noise variances in $\epsilon_u \sim \mathcal{N}(0, \sigma_u^2)$ and $\epsilon_f \sim \mathcal{N}(0, \sigma_f^2)$, respectively. As emphasized by Kennedy and O'Hagan (2001), considering an input as unknown, despite knowing its true value, can enhance the computer model. The known parameter is then categorized within the set $\Phi$ as a prior distribution centered around the true value with a non-zero variance. This approach can allow for deviations from the true physical value and may result in an empirically improved computer model of reality.

This approach of Bayesian calibration with physics-informed priors eliminates the need to predefine values for the independent measurement noises, described by their variances $\sigma_u^2$ and $\sigma_f^2$. Instead, we assign priors to these variables, allowing the model to evaluate and determine posterior probabilities for their values. This flexibility is particularly beneficial when our knowledge about these parameters is uncertain or lacking.

Another advantage of this approach is that it does not require explicit specification of initial conditions. Instead, when analyzing the observations, our model relies on the underlying differential equation, which is the foundation for our physics-informed priors. This is especially advantageous when our knowledge regarding these initial magnitudes is limited.

### 3.6.1 Model Discrepancy

When the model is simplified compared to the true process, a discrepancy arises between the model and the observed data. This model-form uncertainty leads to an imperfect alignment between the model and the observed process (Kennedy and O'Hagan, 2001). Recognizing model discrepancy is essential in effectively utilizing observations from the physical system (Brynjarsdóttir and O'Hagan, 2014). In this section, an approach for accounting for model discrepancy is introduced.

Suppose that we observe the process $y(t)$, corrupted by noise

$$y_{obs}(t_i) = y(t_i) + \epsilon_i, \tag{3.24}$$

where $y(\cdot)$ is the true process, and $\epsilon_i$ represents the observation error for the $i$'th observation. Assuming we have the simplified model $y_M(\cdot, \cdot)$, Kennedy and O'Hagan (2001) added a functional discrepancy term $\delta(t)$ that is independent of $y_M(\cdot, \cdot)$. We propose the observed process can be expressed as (Spitieris and Steinsland, 2023):

$$y_{obs}(t) = y(t) + \epsilon = y_M(t, \Phi) + \delta(t) + \epsilon. \tag{3.25}$$

Here, the vector $\Phi$ comprises the physical parameters. The discrepancy term $\delta(t)$ is typically modeled using a flexible GP, and the independent noise term $\epsilon_t$ is usually modeled as independent and identically distributed (iid) normal variables with mean 0 and variance $\sigma_\epsilon^2$. This corresponds to

$$y(t) = y_M(t, \Phi) + \delta(t), \tag{3.26}$$

where $y_M(\cdot, \cdot)$ and $\delta(\cdot)$ are independent.

Differentiating between model discrepancy and the physical process in complicated calibration problems can be challenging (Brynjarsdóttir and O'Hagan, 2014). They illustrate the dangers of failing to acknowledge model discrepancy and emphasize that in many calibration problems, the objective is to estimate the values of parameters with limited a priori knowledge. Without reliable priors, it becomes challenging for the model to differentiate between process, the model discrepancy and other sources of uncertainty. This includes accurately differentiating between the systematic deviation $\delta(t)$ and the observation noise $\epsilon_t$ in the data.

Utilizing informative priors for the discrepancy term may help reduce the risk of misclassification of the deviation type (Brynjarsdóttir and O'Hagan, 2014). How-

ever, in real-world scenarios, our understanding of the underlying characteristics of this term is often limited. We aim to explore this aspect further through our experimental investigations.

## 3.7 Physics-Informed Bayesian Calibration Accounting for Model Discrepancy

To account for the imperfections of physical models as representations of reality, we adopt the approach of Kennedy and O'Hagan (2001) by incorporating a functional model discrepancy into our model formulation. We follow Spitieris and Steinsland (2022) and, for simplicity, assume the discrepancy only exists in the function $u(t)$. We call the process following the model $u_M(t)$. This leads to the following model formulation:

$$y(t) = \begin{bmatrix} u(t) \\ f(t) \end{bmatrix} = \begin{bmatrix} u_M(t) + \delta_u(t) \\ f(t) \end{bmatrix}, \tag{3.27}$$

where $\delta_u(t)$ is the difference between the process following the model, $u_M(t)$, and the actual process $u(t)$. We assume that $\delta_u(t)$ follows a GP, $\delta_u(t) \sim GP\left(0, K_{\delta_u}\left(t, t'\right)\right)$. We assume that $y(t)$ follows the multivariate GP

$$p\left(y(t)\right) = \mathcal{N}\left(\mu(t), K_{\text{disc}}(t, t')\right) \tag{3.28}$$

where $\mu(t) = [\mu_u(t), \mu_f(t)]^T$ and

$$K_{\text{disc}}(t, t') = \begin{bmatrix} K_{uu}\left(t, t' \mid \theta\right) + K_{\delta_u}\left(t, t' \mid \theta_{\delta_u}\right) & K_{uf}\left(t, t' \mid \theta, \Phi\right) \\ K_{fu}\left(t, t' \mid \theta, \Phi\right) & K_{ff}\left(t, t' \mid \theta, \Phi\right) \end{bmatrix}. \tag{3.29}$$

Considering noisy observations $y_{obs}$ of this process at times $t_u$ and $t_f$ respectively,

$$y_{obs} = \begin{bmatrix} u_{obs} \\ f_{obs} \end{bmatrix} = \begin{bmatrix} u_M(t_u) + \delta_u(t_u) + \epsilon_u \\ f(t_f) + \epsilon_f \end{bmatrix}, \tag{3.30}$$

where $\epsilon_u \sim \mathcal{N}(0, \sigma_u^2)$ and $\epsilon_f \sim \mathcal{N}(0, \sigma_f^2)$, we get that

$$p\left(y \mid \theta, \theta_{\delta_u}, \Phi, \sigma_u, \sigma_f\right) = \mathcal{N}\left(\mu, K_{\text{disc}}\right)$$

where $\Phi$ is the vector of physical parameters, $\theta$ is the vector of the model hyperparameters from the multi-output GP of $y$ and $\theta_{\delta_u}$ is the vector of the model parameters of the GP of $\delta_u(t)$.

The covariance matrix containing measurement noise is

$$K_{\text{disc}} = \begin{bmatrix} K_{uu}\left(t_u, t_u' \mid \theta\right) + K_{\delta_u}\left(t_u, t_u' \mid \theta_{\delta_u}\right) + \sigma_u^2 I_u & K_{uf}\left(t_u, t_f' \mid \theta, \Phi\right) \\ K_{fu}\left(t_f, t_u' \mid \theta, \Phi\right) & K_{ff}\left(t_f, t_f' \mid \theta, \Phi\right) + \sigma_f^2 I \end{bmatrix}, \tag{3.31}$$

where $t_u$ and $t_f$ represents the times at which the observations of $u(t)$ and $f(t)$ are done, $\beta$ is the vector of the mean function parameters (Spitieris and Steinsland, 2023).

## 3.8 Posterior Predictive Checks

Posterior distributions obtained from MCMC sampling provide information about parameter uncertainty, and if we know the true value, we can compare it to the posterior mean to assess estimation bias. Posterior predictive checks (PPCs) complement this analysis by comparing model predictions from the estimated parameters to observed data, facilitating further assessment of the compatibility between the model and the observed process. When we aim to test the model against a known process, we can compare the model's predictions to both the observed data and the known process. This comparison reveals insights into the model's performance and helps identify any discrepancies that may arise.

In a Bayesian model, represented as $p\left(y \mid \alpha\right)$ where $\alpha$ denotes the unknown parameters, we can determine the posterior predictive distribution (PPD) for new data, represented by $y^*$, given the observed data, $y$. This is achieved through the following expression (Gelman, Meng et al., 1996):

$$p\left(y^* \mid y, H\right) = \int p\left(y^* \mid \alpha, H\right) \cdot p(\alpha, H \mid y)\, d\alpha, \tag{3.32}$$

where $H$ is the model including the prior distributions $p(\alpha)$, $p\left(y^* \mid \alpha\right)$ is the sampling distribution, representing the likelihood of the new data $y^*$, also referred to as replications, given the parameters $\alpha$. The expression $p(\alpha, H \mid y)$ corresponds to the posterior distribution of the parameters $\alpha$ given the model $H$ and

the observed data $y$. The posterior predictive distribution encompasses two distinct sources of uncertainty: the sampling uncertainty of output $y$ given $\alpha$ and parametric uncertainties concerning $\alpha$.

In terms of calculating the replications $y^*$ in this approach, we consider the multivariate GP

$$y_{obs} \sim GP(\mu, K) \tag{3.33}$$

where $y_{obs} = (u_{obs}, f_{obs})$ is our observations at times $t_u$ and $t_f$ respectively, $\mu = (\mu_u, \mu_f)$, and $K$ is the covariance matrix, including measurement noise.

If we want to make predictions $u^*$ at new points $t_u^*$, we use that the distribution of the new point follows a multivariate Gaussian (Spitieris and Steinsland, 2022):

$$\begin{aligned} p\left(u_* \mid t_u^*, t, y_{obs}, \alpha\right) &= \mathcal{N}\left(\mu_u^*, \Sigma_u^*\right) \\ \mu_u^* &= \mu_u\left(t_u^*\right) + V_u^{*T} K^{-1}(y - \mu) \\ \Sigma_u^* &= K_{uu}\left(t_u^*, t_u^*\right) - V_u^{*T} K^{-1} V_u^*, \end{aligned} \tag{3.34}$$

where $t = (t_u, t_f)$, $y_{obs} = (u_{obs}, f_{obs})$, $\alpha$ is the model parameters and $V_u^{*T} = [K_{uu}\left(t_u^*, t_u\right) \quad K_{uf}\left(t_u^*, t_f\right)]$. The posterior distributions for new observations $f^*$ can be calculated correspondingly.

To improve computational efficiency, we can use some techniques in the calculations, following Betancourt (2017). Firstly, we use the Cholesky decomposition of the covariance matrix $K$, denoted as $L_K$. In the computation of $\mu_u^*$, we solve $K^{-1}y$ by solving the system:

$$tri(L_K)^{-1} y \; tri(L_K)^{-1}, \tag{3.35}$$

where

$$\operatorname{tri}(A)[m,n] = \begin{cases} A[m,n] & \text{if } m \geq n, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{3.36}$$

(Stan Development Team, 2022). These computations are divided into two parts and performed efficiently using the STAN-functions `mdivide_left_tri_low()` and `mdivide_right_tri_low()`.

Similarly, in the computation of $\Sigma_u^*$, we solve $V_u^T K^{-1} V_u^*$ by solving the system:

$$(tri(L_K)^{-1}\ V_u^{*T})^T (tri(L_K)^{-1}\ V_u^{*T}) \tag{3.37}$$

using `mdivide_left_tri_low()`.

# Chapter 4

# Method: Physics-Informed Bayesian Calibration for the Pendulum

In this chapter, we introduce the approach of a physics-informed Bayesian calibration framework for the pendulum process. The approach involves three steps:

1. Observation Data Simulation: We simulate noisy observations from four different pendulum processes: the linearized pendulum process (LP), the linearized pendulum process with an added Gaussian Process (LPGP), the true idealized process (TP), and the damped process (DP). This is referred to as noisy *observations* from the (LP/LPGP/TP/DP) *process*. A detailed description is provided in Chapter 4.1.
2. Model Construction: We develop two physics-informed Bayesian calibration models. The first model ($M1_{WOD}$) does not account for discrepancy, while the second model ($M2_{WDf}$) incorporates discrepancy. These are referred to as the *models*. The construction of these models can be found in Chapter 4.2.
3. Inference: We employ HMC-NUTS within STAN for inference, and the Rstan interface is used to post-process the posterior samples generated by the models.

In Chapter 6, we evaluate the performance of our models with the simulated obser-

vations as input. The experimental configurations employed in these experiments are outlined in Chapter 5.

## 4.1 Observation Data Simulations

In Spitieris and Steinsland (2022) and Chapter 3.6.1, discrepancy is introduced as the model's deviation from the observation process, which is the case in a physical context. However, during the assessment and discussion of our results, we adopt a different perspective: the observation process deviating from the model. This conceptual shift maintains the mathematical framework while offering a more intuitive explanation and evaluation of our approach. Consequently, it becomes meaningful to test the model on various processes that exhibit different types of deviations from the model.

The process of simulating observation data is illustrated in Figure 4.1.

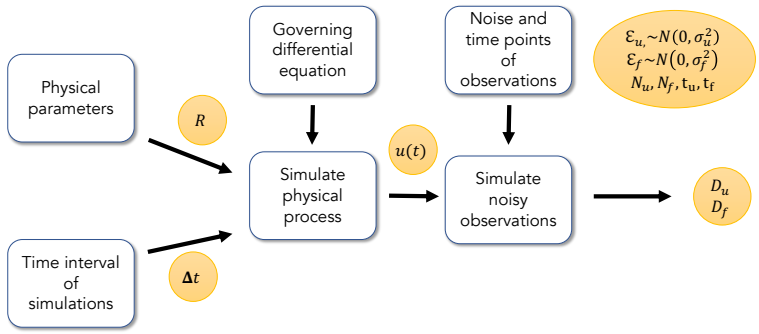



Figure 4.1: Procedure of simulating noisy observations $D_u$ and $D_f$. The governing differential equation is parameterized with physical and simulation parameters. Noisy observations are generated, resulting in observation data $D_u = \{(t_i, u_i)\}_{i=1}^{N_u}$ and $D_f = \{(t_j, f_j)\}_{j=1}^{N_f}$. Here, $t_u$ and $t_f$ represent the sets $t_u = \{t_i\}_{i=1}^{N_u}$ and $t_f = \{t_j\}_{j=1}^{N_f}$, the times where $u$ and $f$ are observed, respectively. $\Delta t$ represents the time interval of our observations, starting at $t_0 = 0$

We want to assess the performance of the two models on four different pendulum angle processes $u(t)$. For all simulations, we set the initial values as follows:

$$u(0) = u_0 \tag{4.1}$$

$$\frac{du(0)}{dt} = 0. \tag{4.2}$$

The four different pendulum angle simulations $D_u = \{(t_i, u_i)\}_{i=1}^{N_u}$ are obtained from the following processes:

1. **LP**: $u(t)$ is simulated from the solution for the linearized pendulum process (Equation (2.4)).
2. **LPGP**: $u(t)$ is simulated from the linearized pendulum process (Equation (2.4)) with the addition of a Gaussian process $\delta_{sim}(t) \sim \mathcal{N}(0, K_{\delta_{sim}}(t, t'))$, such that we simulate from

$$u(t) = u_0 \cos\left(\sqrt{\frac{1}{R}}t\right) + \delta_{sim}(t)$$

3. **TP**: $u(t)$ is simulated from the true idealized pendulum process (Equation (2.1)) using the Runge-Kutta 4 (RK4) method (Butcher, 2008) for approximating the solution of $u(t)$.
4. **DP**: $u(t)$ is simulated from the damped pendulum process (Equation (2.5)) and a numerical solution for $u(t)$ is obtained using the Euler method.

In LPGP,

$$K_{\delta_{sim}}(t, t') = \sigma_{\delta_{sim}} e^{\frac{-0.5(t-t')^2}{l_{\delta_{sim}}^2}}.$$

The GP allows us to introduce a systematically controlled discrepancy that we know fulfills the GP assumption, and observe the model's response. It is important to note that the initial seed influences the simulations for the GP, so we conduct multiple trials using different GP realizations. More on this in Section 5.1.3.

From each of these simulations of $u(t)$ we extract their value at $N_u$ equally distributed times $t_u = \{t_i\}_{i=1}^{N_u}$ and add noise $\epsilon_{u,i} \overset{i.i.d}{\sim} N(0, \sigma_u^2)$ to each observation. By this, we obtain our noisy angle observations, denoted by $D_u = \{(t_i, u_i)\}_{i=1}^{N_u}$.

We include the right-hand side of the pendulum equations discussed in Chapter 2 in the input data as noisy observations from the process $f(t) = 0$. Let $D_f = \{(t_j, f_j)\}_{j=1}^{N_f}$ represent the observations, where $f_j$ denotes the noisy observations $f_j = 0 + \epsilon_j$ at equally distributed times $t_f = \{t_j\}_{j=1}^{N_f}$, and $N_f$ is the total number of observations. Here, $\epsilon_{f,j} \overset{i.i.d}{\sim} N(0, \sigma_f^2)$.

## 4.2 Constructing the Models

Next, we construct the physics-informed Bayesian Calibration models $M1_{WOD}$ and $M2_{WDf}$. The outline of the models is shown in Figure 4.2. Here $M1_{WOD}$ does not include the parts with a dashed contour, whereas $M2_{WDf}$ does.

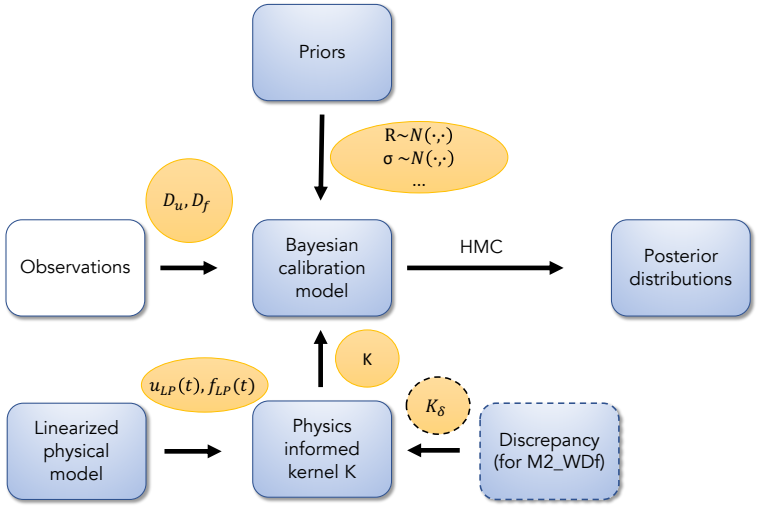



Figure 4.2: Outline of the physics-informed Bayesian calibration procedure. We base the model on a linear physical model, where its structure is incorporated in the physics-informed kernel K. The dashed line around the discrepancy $K_\delta$ that goes into the physics-informed kernel indicates that this is only incorporated in the Bayesian calibration model for $M2_{WDf}$. The kernel K of $M1_{WOD}$ does not contain the discrepancy term $K_\delta$. Then we assign priors for the parameters in the model, including the physical parameters, the noise parameters, and the model parameters. We run numerous Hamiltonian Monte Carlo (HMC) iterations to obtain posterior distributions.

### 4.2.1 $\mathbf{M1_{WOD}}$: Model That Does Not Account for Model Discrepancy

For the linearized pendulum (LP) from Equation (2.3), the linear differential operator of $\mathcal{L}_t^{\Phi} u(t) = f(t)$, becomes

$$\mathcal{L}_t^{\Phi} = \frac{d^2}{dt} + \frac{1}{R}, \tag{4.3}$$

Our model $M1_{WOD}$ is then based on

$$p(y_{obs}|\alpha) = \mathcal{N}(\mu, K) \tag{4.4}$$

where $\alpha = (\Phi, \theta, \sigma_u, \sigma_f)$ with $\Phi = (R)$ and $\theta = (\sigma, \ell)$. we assume that the mean function is given by $\mu_u(t) = 0$. This assumption is based on the fact that the pendulum oscillates around its equilibrium position. Furthermore, as we have $f(t) = 0$ in our processes, it has a zero mean for all $t$ (indicating that no other external forces are acting on the pendulum), so we set $\mu_f(t) = 0$. By this, we get

$$\mu = \begin{bmatrix} \mu_u \\ \mu_f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.5}$$

Note that in other cases, these mean functions can take the form of variable-dependent functions or functions of each other. This is exemplified through the Windkessel experiments discussed in Spitieris and Steinsland (2022).

Next, the kernel can be expressed by

$$K = \begin{bmatrix} K_{uu}(t_u, t'_u|\theta) + \sigma_u^2 I_u & K_{uf}(t_u, t'_f|\theta, \Phi) \\ K_{fu}(t_f, t'_u|\theta, \Phi) & K_{ff}(t_f, t'_f|\theta, \Phi) + \sigma_f^2 I_f \end{bmatrix} \tag{4.6}$$

where $I_u$ and $I_f$ are the $(N_u \times N_u)$ and $(N_f \times N_f)$ identity matrices, respectively, and, as obtained in my project thesis

$$K_{uu}(t,t'|\theta) = \sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}$$

$$\begin{aligned} K_{uf}(t,t'|\theta,\Phi) &= \mathcal{L}_{t'}^{\Phi} K_{uu}(t,t'|\theta) \\ &= -\frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^2} + \frac{\sigma^2 (t-t')^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^4} + \frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{R} \end{aligned}$$

$$\begin{aligned} K_{fu}(t,t'|\theta,\Phi) &= \mathcal{L}_{t}^{\Phi} K_{uu}(t,t'|\theta) \\ &= -\frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^2} + \frac{\sigma^2 (t-t')^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^4} + \frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{R} \end{aligned}$$

$$\begin{aligned} K_{ff}(t,t'|\theta,\Phi) &= \mathcal{L}_{t}^{\Phi} \mathcal{L}_{t'}^{\Phi} K_{uu}(t,t'|\theta) \\ &= \frac{3\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^4} - \frac{6\sigma^2 (t-t')^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^6} \\ &\quad + \frac{\sigma^2 (t-t')^4 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^8} \\ &\quad - \frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{R\, l^2} + \frac{\sigma^2 (t-t')^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{R\, l^4} \\ &\quad + \frac{1}{R}\left( \frac{-\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^2} + \frac{\sigma^2 (t-t')^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{l^4} + \frac{\sigma^2 \mathrm{e}^{-\frac{0.5(t-t')^2}{l^2}}}{R} \right) \end{aligned} \tag{4.7}$$

### 4.2.2 $\mathbf{M2_{WDf}}$: Model That Accounts for Discrepancy

Next, we look at the construction of $M2_{WDf}$.

As discussed, our model $M1_{WOD}$ is based on the linearized pendulum process (LP), such that

$$\mathcal{L}_t^{\Phi}(u_{LP}(t)) = f(t) = 0 \tag{4.8}$$

yields. However, as seen in Section 2, for TP and DP, we can have observations that

follow a process $u(t)$ that do not conform to the model's process $u_{LP}(t)$, implying that the model is imprecise for these processes.

To address this, we start by describing the actual process as $u(t) = u_{LP}(t) + \Delta u(t)$. The term $\Delta u(t) = u(t) - u_{LP}(t)$ represents the deviation between the actual process we are considering and the corresponding linear process on which the model is based. Figure 2.4 illustrates two examples of how the deviation $\Delta u(t)$ can look.

### Shifting the discrepancy term from $\mathbf{u(t)}$ to $\mathbf{f(t)}$

If our observations from $u(t)$ deviate from the model $\mathcal{L}_t^\Phi$ that we use in our physics-informed priors, we need to account for this when we apply our model to the observations $u(t) \neq u_{LP}(t)$. For a linear differential operator, it holds that

$$\mathcal{L}_t^\Phi(u_{LP}(t) + \Delta u(t)) = \mathcal{L}_t^\Phi(u_{LP}(t)) + \mathcal{L}_t^\Phi(\Delta u(t)), \tag{4.9}$$

and inserting Equation (4.8), gives

$$\mathcal{L}_t^\Phi(u_{LP}(t) + \Delta u(t)) = f(t) + \mathcal{L}_t^\Phi(\Delta u(t)). \tag{4.10}$$

Suppose the difference between our observations and the corresponding model process, $\Delta u(t)$, follows a GP. According to the GP properties, the transformed process $\mathcal{L}_t^\Phi(\Delta u(t))$ will also follow a GP. We define $\delta(t) = \mathcal{L}_t^\Phi(\Delta u(t))$, where $\delta(t) \sim GP(0, K_\delta(t, t'))$. With $u(t) = u_{LP}(t) + \Delta u(t)$, we can express the equation as:

$$\mathcal{L}_t^\Phi(u(t)) = f(t) + \delta(t) \tag{4.11}$$

Here, the term $\delta(t)$ should capture the deviation between the model and the process. This shift is done with the expectation that it is easier to identify the systematic error in the model on the side of which the model is expecting a constant function than to the process the linear operator is applied to.

**What $\delta(\mathbf{t})$ Means in Terms of TP and DP**

This shift becomes more intuitive when we explore the role of the discrepancy term $\delta(t)$ in capturing the discrepancy between the LP-based model and the specific processes of TP and DP. We look at Equations (2.9) and (2.10), which explains LP and DP as TPs, but with external forces $f_p(t)$ and $f_d(t)$, respectively, acting on it.

Starting with LP and TP, LP represents TP affected by an external force $f_p(t)$. We now turn it around and say that TP equals a corresponding LP but with an opposing external force, given by

$$f_{-p}(t) = -f_p(t) = \frac{1}{R}\left(\frac{u(t)^3}{3!} - \frac{u(t)^5}{5!} + \frac{u(t)^7}{7!} - ...\right). \tag{4.12}$$

This means that $u_{TP} = u_{LP} + \Delta u_{(TP-LP)}$ with $f(t) = 0$ can be written as

$$\mathcal{L}_t^{\Phi}(u_{TP}(t)) = \mathcal{L}_t^{\Phi}(u_{LP} + \Delta u_{(TP-LP)}(t)) = f_{-p}(t) \tag{4.13}$$
$$\tag{4.14}$$

with $f_{-p}(t) \neq 0$. As we see, this does not align with the model $\mathcal{L}_t^{\Phi}(u(t)) = 0$, hence we add a discrepancy term that should aim to satisfy

$$f(t) = f_{-p}(t) + \delta(t) = 0 \tag{4.15}$$

Consequently, the discrepancy term $\delta(t)$ aims to capture the behavior of $f_p(t)$ as $f_{-p} + f_p(t) = 0$.

We employ a similar approach for DP, where we previously viewed it as a TP with an external force $f_d(t)$. Now, we view it as a LP with an external force $f_{-dp}(t)$, where

$$f_{-dp}(t) = -f_d(t) - f_p(t) = \lambda\frac{du(t)}{d(t)} + \frac{1}{R}\left(\frac{u(t)^3}{3!} - \frac{u(t)^5}{5!} + \frac{u(t)^7}{7!} - ...\right), \tag{4.16}$$

so for $u_{DP} = u_{LP} + \Delta u_{(DP-LP)}(t)$ with $f(t) = 0$, we have

$$\mathcal{L}_t^{\Phi}(u_{DP}(t)) = \mathcal{L}_t^{\Phi}(u_{LP} + \Delta u_{(DP-LP)}(t)) = f_{-dp}(t) \tag{4.17}$$
$$\tag{4.18}$$

where $f_{-dp}(t) \neq 0$. This does not fit the model, hence we add a discrepancy term that aims to satisfy

$$f(t) = f_{-dp}(t) + \delta(t) = 0 \tag{4.19}$$

Here, the discrepancy term $\delta(t)$ aims to capture the behavior of $f_{dp}(t)$ as $f_{-dp} + f_{dp}(t) = 0$.

Incorporating additional knowledge into the discrepancy term would be simple in this case. However, in other cases, limited information may be available regarding this discrepancy, motivating us to model it as a flexible GP, treating it as a process we do not know.

### Incorporating the Discrepancy Term in the Model

Given the assumption that $\delta(t) \sim GP(0, K_\delta(t, t'))$, the expression for $K_\delta(t, t')$ when employing the squared exponential kernel can be expressed as follows:

$$K_\delta(t, t') = \sigma_\delta^2 e^{-0.5\left(\frac{t-t'}{l_\delta}\right)^2}, \tag{4.20}$$

with scale hyperparameter $\sigma_\delta$ and length scale hyperparameter $l_\delta$.

We consider our observation process $y(t)$ to be represented as:

$$y(t) = \begin{bmatrix} u(t) \\ f(t) \end{bmatrix} = \begin{bmatrix} u_{LP}(t) + \Delta u(t) \\ f_{-force}(t) + f_{force}(t) \end{bmatrix} = \begin{bmatrix} u_{LP}(t) + \Delta u(t) \\ f_{-force}(t) + \delta(t) \end{bmatrix}, \tag{4.21}$$

where $f(t) = 0$, $f_{-force}(t)$ represents the virtual force we get from treating the actual process with the linearized model, and $f_{force}(t)$ is the virtual force that offsets the total observation process $f(t)$ back to zero.

In order to capture the discrepancy, we use Equation (4.11), and get the model $M2_{WDf}$,

$$p(y_{obs}|\alpha_{disc}) = \mathcal{N}(\mu, K_{disc}) \tag{4.22}$$

where $\alpha_{disc} = (\Phi, \theta, \sigma_u, \sigma_f, \sigma_\delta, l_\delta)$. For $\mu$, the same holds as for $M1_{WOD}$. The kernel is given by

$$K_{disc} = \begin{bmatrix} K_{uu}(t_u, t'_u|\theta) + \sigma_u^2 I_u & K_{uf}(t_u, t'_f|\theta, \Phi) \\ K_{fu}(t_f, t'_u|\theta, \Phi) & K_{ff}(t_f, t'_f|\theta, \Phi) + K_\delta(t_f, t'_f|\sigma_\delta, l_\delta) + \sigma_f^2 I_f \end{bmatrix} \tag{4.23}$$

with elements from Equation (4.7), in addition to $K_\delta(t_f, t'_f|\sigma_\delta, l_\delta)$ from Equation (4.20).

As discussed, we propose placing the discrepancy term on the $f$ side of the equation. The shift in the discrepancy term is intended to help the model identify the systematic deviation between the model and the process. This shift is made based on the expectation that it may be easier to identify systematic errors in the model when focusing on the side where the model expects a constant function, rather than within the more complex process on the left-hand side.

# Chapter 5

# Experimental Configurations

In this chapter, we define the parameter values of the simulations and the priors of the models described in Chapter 4. Then we present how we do repeated experiments in E1, and the modifications of E2a and E2b. These parameters and configurations are subsequently employed in the experiments conducted in Chapter 6.

## 5.1 Experiment 1 (E1) Configurations

This section specifies the details and parameters for Experiment 1. These configurations also serve as the basic configurations for Experiment 2, enabling a systematic evaluation of individual variations.

### 5.1.1 Simulation Parameters

For our pendulum process simulations, described in Chapter 4.1, we utilize the parameter values listed in Table 5.1. Figure 5.1a provides a visualization of observations generated using these parameters, showing that the observations span approximately 1.5 cycles.

The computational cost of inference increases with the number of measurement points in the input data, as noted by Spitieris and Steinsland (2022). This consideration is important when determining the quantity of measurement points $N_u$

and $N_f$. Additionally, practical constraints and limitations in measurement device capabilities can restrict the frequency and number of measurements that can be obtained.

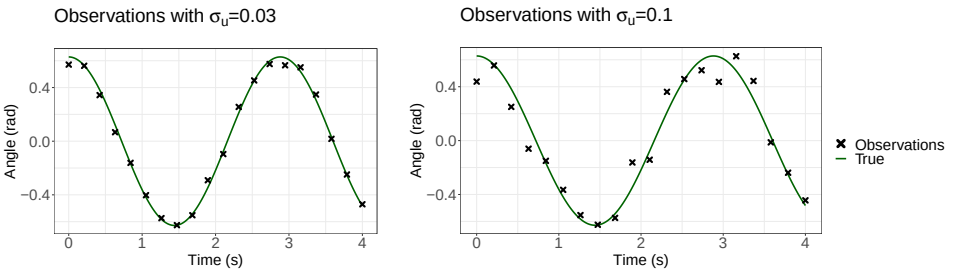
In their study, Spitieris and Steinsland (2022) utilized three replicates for each measurement point, which can help distinguish between independent and systematic errors. However, implementing such practices in real-life scenarios may not be feasible, as it often requires the ability to perform simultaneous or synchronized measurements. Therefore, for our investigation, we will use a single set of measurements for each time point.

| Parameter | Value |
|---|---|
| $u_0$ | $\frac{\pi}{5}$ |
| $R$ | $0.2s^2$ |
| $\lambda$ | $0.5\frac{1}{s}$ |
| $\sigma_u^2$ | $0.03^2$ |
| $\sigma_f^2$ | $0.03^2$ |
| $N_u$ | 20 |
| $N_u$ | 20 |
| $\Delta t$ | $4s$ |
| $\sigma_{\delta_{sim}}$ | $\frac{\pi}{30}$ |
| $l_{\delta_{sim}}$ | 1 |

Table 5.1: Parameters used in the Pendulum angle simulations for E1. $u_0$ is the initial angle, $R = \frac{L}{g}$ is the physical parameter deciding the frequency, $\lambda$ is the damping parameter, $\sigma_u^2$ is the measurement noise variance of the observed angle measurements $u$, $\sigma_f^2$ is the measurement noise variance of the observed $f$, $N_u$ is the number of angle observations, $N_f$ is the number of observations of $f$, $\Delta t$ is the time period of the observations (starting at $t_0 = 0$) and $\sigma_{\delta_{sim}}$ and $l_{\delta_{sim}}$ are the hyperparameters of the added GP in LPGP.

### 5.1.2 Priors

In Chapter 3.1, we discussed the importance of reasonable priors in Bayesian analysis. We want to choose priors that imitate a real-life experiment, so we do not set them too informative. Since none of the parameters take negative values, we follow Spitieris and Steinsland (2022) and impose a lower limit of 0 on all parameters, and obtain improper truncated prior distributions. The physical priors we use in the models are chosen as follows:




(a) $\epsilon_u \sim \mathcal{N}(0, 0.03^2)$ (b) $\epsilon_u \sim \mathcal{N}(0, 0.1^2)$

Figure 5.1: An example of observations of the pendulum angle motion from LP with measurement noise variances $\sigma_u^2 = 0.03^2$ and $\sigma_u^2 = 0.1^2$, respectively. In green are simulations from the true idealized process $u(t)$ for a continuous scale of $t$ of the pendulum with initial angle $u_0 = \pi/5$.

$$\begin{aligned} R &\sim \mathcal{N}(0.2, 0.1^2) \\ \sigma_u &\sim \text{Half-}\mathcal{N}(0, 0.05^2) \\ \sigma_f &\sim \text{Half-}N(0, 0.1^2), \end{aligned} \tag{5.1}$$

while the model parameter priors are

$$\begin{aligned} \sigma &\sim \mathcal{N}(1, 1^2) \\ l &\sim \mathcal{N}(1, 1^2) \end{aligned} \tag{5.2}$$

For the model with discrepancy, $M2_{WDf}$, additional priors are introduced for the parameters of the discrepancy $\delta(t)$:

$$\begin{aligned} \sigma_\delta &\sim \text{Half-}\mathcal{N}\left(0, 1^2\right) \\ l_\delta &\sim \mathcal{N}(1, 1^2) \end{aligned} \tag{5.3}$$

These priors are illustrated in Figures 5.2, 5.3, and 5.4 to get a better understanding of their distributions.

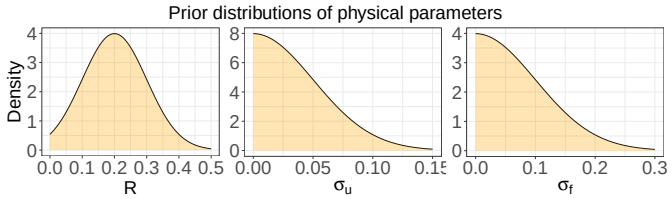



Figure 5.2: Priors for the process parameters used in the model. Note that these are truncated and improper.

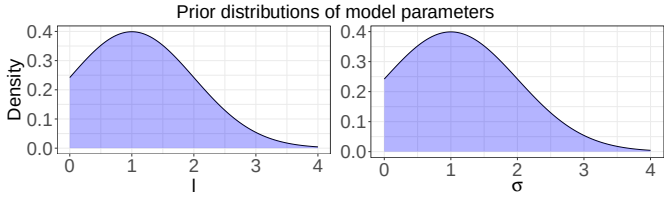



Figure 5.3: Truncated improper priors for the covariance matrix in the model.

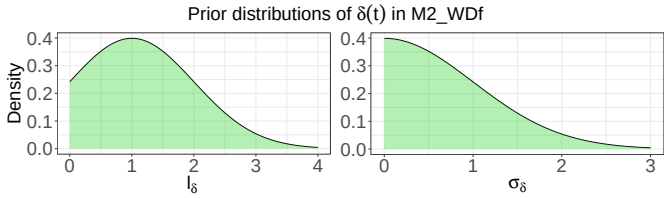



Figure 5.4: Truncated improper priors used for the covariance matrix of the discrepancy term in the model.

### 5.1.3 GP Realizations for LPGP

In order to investigate various discrepancies between the model and the underlying process in the data simulation, we employ five distinct realizations of GPs for experimental evaluation with values of $\sigma_\delta$ and $l_\delta$ from Table 5.1. These GPs are shown in Figure 5.5.

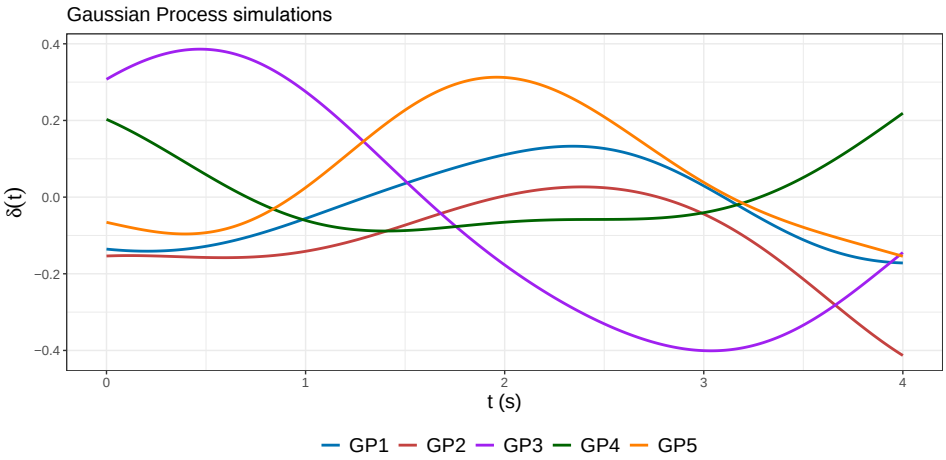



Figure 5.5: GP realizations that are added to LP in the simulations of the LPGP process.

### 5.1.4 Repeated Experiments

We employ a methodical approach to mitigate the influence of randomness in stochastic experiments on our conclusions due to favorable or unfavorable measurement noise placements. By conducting repeated experiments with different realizations, we can investigate the effect of the variability introduced by the placement of the noise. Algorithm 1 outlines this procedure.

**Algorithm 1** Repeated experiments

Simulate $u(t)$ from the process that we want to investigate
**for** $i = 1 :$ num_experiments **do**
  **for** k = 1:$N_u$ **do**
    $\epsilon_{u,k} \sim \mathcal{N}(0, \sigma_u^2)$
    $u_{i,k} = y(t_{i,k}) + \epsilon_{u,k}$
  **end for**
  **for** j = 1:$N_f$ **do**
    $\epsilon_{f,j} \sim \mathcal{N}(0, \sigma_f^2)$
    $f_{j,k} = 0 + \epsilon_{f,j}$
  **end for**
  Fit model to obtain posteriors for $\alpha_i$
**end for**
Plot histograms of each parameter of $\alpha_i$ in the same coordinate system for all num_experiments to compare

## 5.2 Experiment 2a (E2a) Configurations: Measurement Noise

In this experiment, we investigate the model's response to varying levels of measurement noise by adjusting the parameter $\sigma_u$ in the simulations. We explore a range of values for $\sigma_u$ and adjust the priors accordingly. The experiment follows the procedure outlined in Algorithm 2.

In Experiment 2, we consider a range of 10 equally distributed values of $\sigma_u$ between 0.02 and 0.15:

$$\sigma_{u,sim} = (0.02, 0.344, ..., 0.15)$$

Concurrently, we adjust the priors of $\sigma_u$ to accommodate these changes. The prior distribution for $\sigma_u$ is now given by:

$$\sigma_u \sim \mathcal{N}(\sigma_{u,sim}, 0.05^2),$$

still truncated with a lower limit of 0, as the standard deviation cannot be negative.

**Algorithm 2** Repeated experiments with different values for the measurement noise variance $\sigma_u$

Set $\sigma_u = (\sigma_{u,min}, ..., \sigma_{u,max})$
Simulate $u(t)$ from the process that we want to investigate
**for** (i=1:length($\sigma_u$)) **do**
  **for** k = 1:$N_u$ **do**
    $\epsilon_{u,k} \sim \mathcal{N}(0, \sigma_u^2)$
    $u_{i,k} = y(t_{i,k}) + \epsilon_{u,k}$
  **end for**
  **for** j = 1:$N_f$ **do**
    $\epsilon_{f,j} \sim \mathcal{N}(0, \sigma_f^2)$
    $f_{j,k} = 0 + \epsilon_{f,j}$
  **end for**
  Fit model to obtain posteriors of $\alpha_j$
  Calculate $R_{bias,i} = abs(R_{\text{true}} - R_{\text{posterior mean}})$
  Calculate $\sigma_{u,bias,i} = abs(\sigma_{u\text{true}} - \sigma_{u\text{posterior mean}})$
  $R_{\text{upper, i}}$ = Upper 95% quantile value of $R_i$
  $R_{\text{lower, i}}$ = Lower 95% quantile value of $R_i$
  $CI_{\text{width, i}} = R_{\text{upper}} - R_{\text{lower}}$
**end for**
Plot ($\sigma_u$, $R_{bias}$)
Plot ($\sigma_u$, $\sigma_{u,bias}$)
Plot ($\sigma_u$, $CI_{width}$)

## 5.3 Experiment 2b (E2b) Configurations: Number of Cycles

In this experiment, we investigate the impact of varying the number of observed cycles on the model's performance. The objective is to determine whether increasing the number of cycles leads to improved estimates, and how the model responds to increasing the observation period.

We conduct two experiments with an extended measurement duration, resulting in an increased number of cycles compared to Experiment 1 (E1). The measurement noise is restored to the baseline value of $\sigma_u = 0.03$, and the observation period is increased from $\Delta t = 4s$, where we have approximately 1.5 cycles, to $\Delta t = 8s$, where we have approximately 3 cycles.

The first experiment (E2b-1) keeps the same number of measurements as for $\Delta t = 4$, $N_u = N_f = 20$ measurements. The second experiment (E2b-2) consists of $N_u = N_f = 40$ measurements, meaning that we maintain the same measurement density as for $\Delta t = 4$.

# Chapter 6

# Results and Discussion

All experiments in this chapter can be reproduced by running the codes in the following GitHub repository: `https://github.com/selmalerke/MasterThesis.git`.

The results are presented and discussed sequentially in this chapter, with more complex findings having dedicated sections for detailed discussion after presenting the experiments. Each experiment concludes with a summary, showing the corresponding PPCs. A summary of the noteworthy findings can be found Chapter 7.

## 6.1 Results from Experiment 1 (E1): A Comparative Analysis of the Models $\mathbf{M1_{WOD}}$ and $\mathbf{M2_{WDf}}$

When assessing model performance in parameter estimation, bias and the width of the credible interval are key considerations. The aim is to achieve unbiased results with appropriately sized credible intervals, avoiding excessively wide or overly confident predictions.

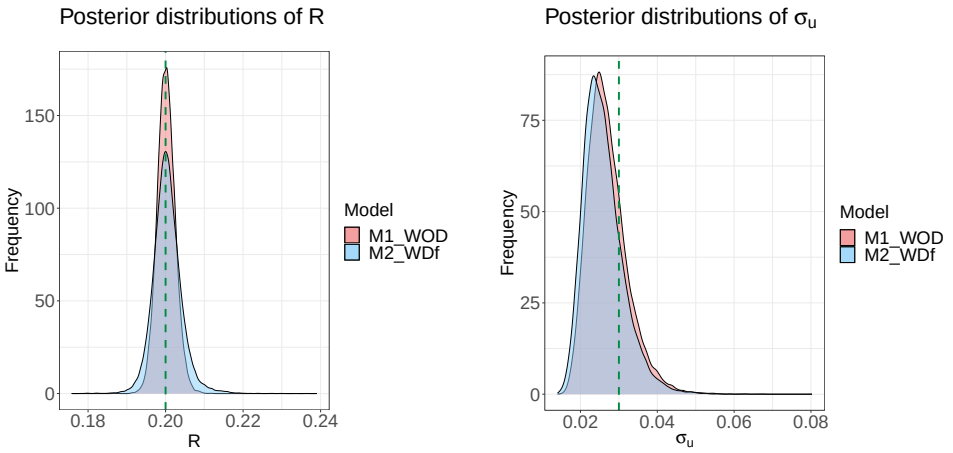
In this experiment, we compare the model results between $M1_{WOD}$ and $M2_{WDf}$ defined in Chapter 4.2, for the four different processes presented in Chapter 4.1. Our focus is on examining the estimates of the physical parameter $R$ and the measurement noise parameter $\sigma_u$, the parameters of most interest in a real-life problem. We refer to the posterior means of the two parameters as $\hat{R}$ and $\hat{\sigma}_u$, respectively. Knowing their true values from the simulation process allows us to evaluate the

model results. While the posterior distributions don't directly provide estimated amplitude and start angle, we can assess their estimates through posterior predictive checks.

Experimental configurations for this experiment are described in Chapter 5.1.

### 6.1.1 Simulated Observations from the Linearized Process (LP)

First, we test our two models on observations simulated from the linearized process (LP), i.e., the same process on which the model $M1_{WOD}$ is based. The posterior distributions of $R$ and $\sigma_u$ for one realization of LP observations are displayed in Figure 6.1, along with their true values represented by dashed green lines.




(a) Posterior distribution of R

(b) Posterior distribution of $\sigma_u$

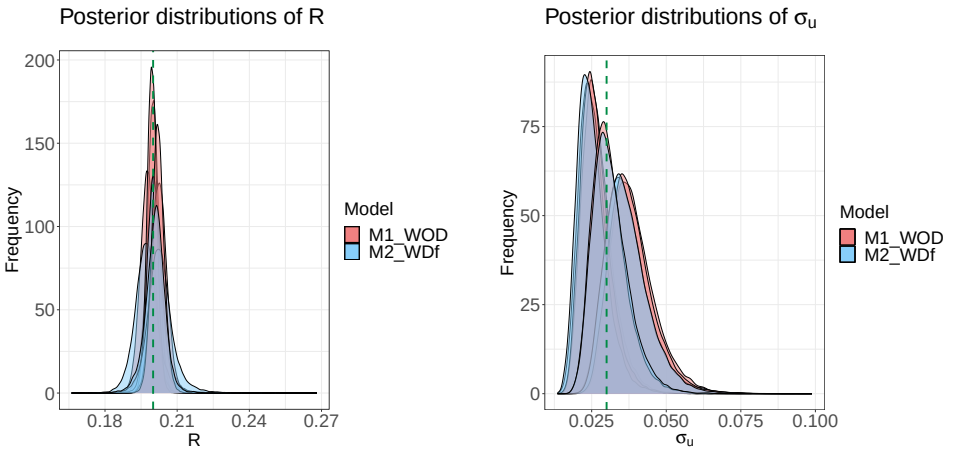
Figure 6.1: Comparison of marginal posterior distributions from computer models on observations from LP with $N_u = 20, N_f = 20$ and noisy observations equally distributed from 0 to 4 seconds.

The trace plots in Figure Appendix B.1 display the exploration of the parameter space by the chains during iterations, offering insights into convergence and stability in the analysis (although not ensuring them). Additionally, summary statistics for the STAN fits of the two models can be found in Appendix B.2.

As expected, both models' posteriors cover the true value. Notably, the posterior distribution of $R$ from $M2_{WDf}$ appears wider than that of $M1_{WOD}$. This difference can be attributed to the additional discrepancy term in $M2_{WDf}$ intended to account for a non-existent systematic error in the simulated process. Consequently, $M2_{WDf}$ exhibits unnecessary flexibility, resulting in a less confident posterior distribution.

The posterior mean estimate of $\sigma_u$, $\hat{\sigma}_u$, in Figure 6.1b shows a slight leftwards bias. This bias is likely influenced by the arrangement of the observed data points. With only 20 observations of an angle affected by noise, the positioning of these observations can lead to underestimation or overestimation of $\sigma_u$.

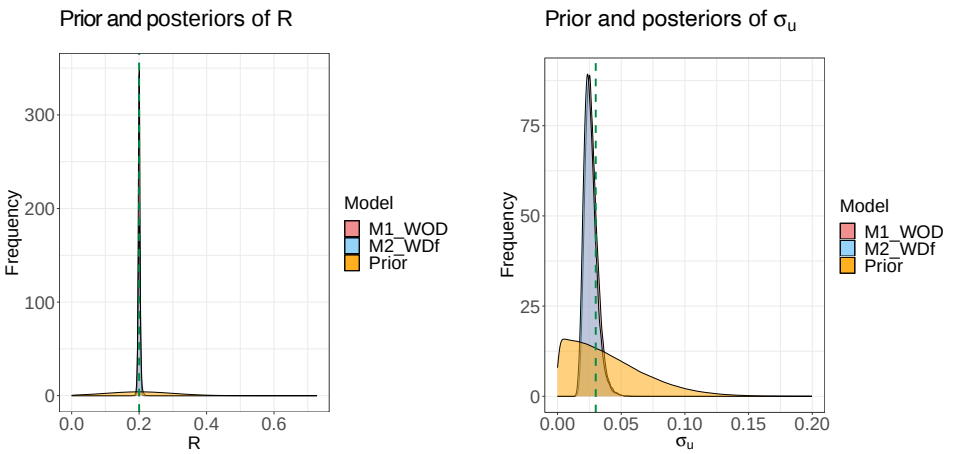
To test this hypothesis, we simulate five different observation realizations from the LP process and obtain posterior distributions for each realization. Figure 6.2 illustrates variations in the observed posteriors of $\sigma_u$ across these sets, indicating that relying on 20 observations may not accurately capture the measurement noise in the posterior mean estimation. However, the posterior distribution consistently covers the true value in all five realizations. Furthermore, Figure 6.2 demonstrates similar behavior in the posterior distributions across different realizations: the posterior distribution of $R$ is slightly wider for $M2_{WDf}$ compared to $M1_{WOD}$, while the posteriors of $\sigma_u$ are quite similar for both models in each realization.




(a) Posterior distributions of R

(b) Posterior distributions of $\sigma_u$

Figure 6.2: Comparison of posterior distributions from $M1_{WOD}$ and $M2_{WDf}$ on observations following the LP process with $N_u = 20, N_f = 20$ equally distributed noisy observations from 0 to 4 seconds.




(a) Prior and posterior distributions of $R$ (b) Prior and posterior distributions of $\sigma_u$

Figure 6.3: Prior distributions and posterior distributions from the two models on $N_u = N_f = 20$ noisy observations equally distributed between 0 and 4 seconds from LP. Note that the priors are truncated at 0 as we know that the values of both $R$ and $\sigma_u$ are non-negative.

Figure 6.3 visually compares the posterior and prior distributions, illustrating the increased precision and decreased uncertainty achieved through the models.

### 6.1.2 Simulated Observations from the Linear Process with an Added Gaussian Process (LPGP)

Secondly, we incorporate a Gaussian Process (GP) into the angle observation, as described in Chapter 4.1. For this experiment, the observations follow the same process as $M2_{WDf}$.

We begin by examining the first LPGP process, LPGP1, which incorporates GP1 (the blue realization in Figure 5.5). Figure 6.4 compares the model results for the observations from this process. In terms of $\hat{R}$, $M1_{WOD}$ exhibits a slight bias, while the distribution of $M2_{WDf}$ centers around the true value. Regarding $\sigma_u$, $M2_{WDf}$ effectively captures the noise, while the posterior distribution of $M1_{WOD}$ deviates significantly from the true value. The bottom row of Figure 6.4 displays the posterior distributions for five different observation realizations of this process, illustrating consistent differences between $M1_{WOD}$ and $M2_{WDf}$ across different noise

placements. This suggests that $M1_{WOD}$ attributes the discrepancy to measurement noise, as it lacks a discrepancy term to account for the additional Gaussian Process in the observations, while $M2_{WDf}$ successfully separates the systematic deviation and independent noise from the process.

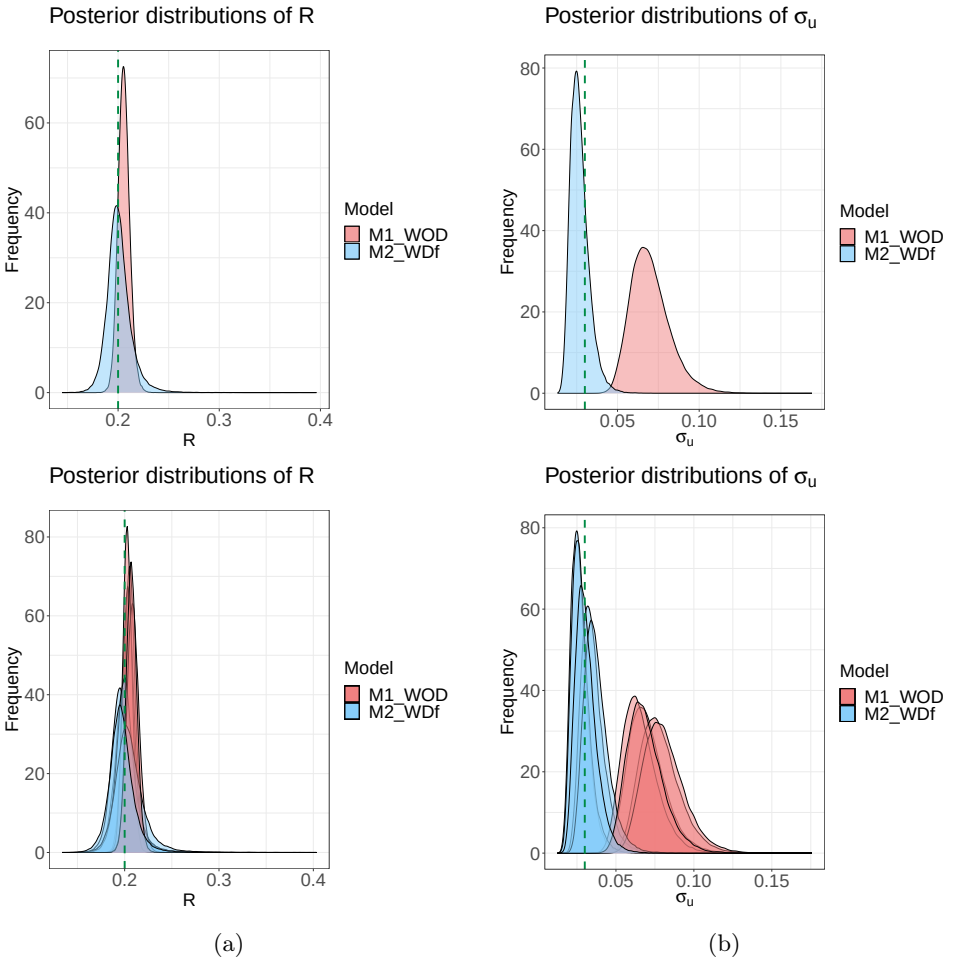
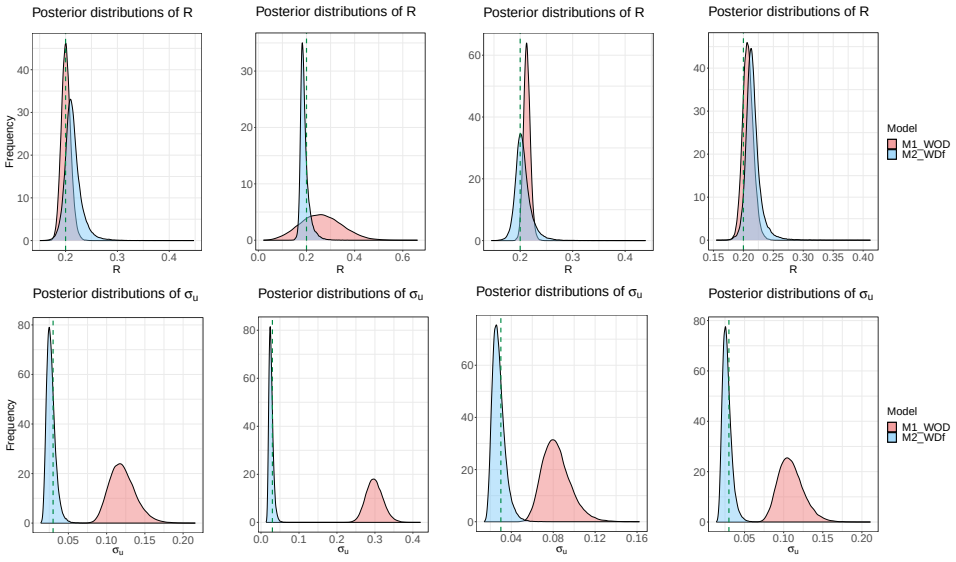



Figure 6.4: Comparison of posterior distributions from observations following the LPGP1-process with $N_u = 20, N_f = 20$ and equally distributed noisy observations from 0 to 4 seconds. The first row displays results from one observation realization, while the second row shows posteriors from five different observation realizations.

To explore the performance of the models on observations with different systematic discrepancies, we evaluate model results on different GP realizations. The GP realizations are depicted in Figure 5.5, and the corresponding posterior distributions are shown in Figure 6.5.




(a) GP2 (b) GP3 (c) GP4 (d) GP5

Figure 6.5: Comparison of posterior distributions from simulated data following different LPGP processes with $N_u = 20, N_f = 20$ and noisy observations equally distributed from 0 to 4 seconds.

Notably, for LPGP2 and LPGP5, $M1_{WOD}$ provides slightly more accurate posterior mean estimates of $R$ compared to $M2_{WDf}$. However, $M1_{WOD}$ consistently struggles to accurately estimate the noise across all LPGP iterations, while $M2_{WDf}$ consistently successfully distinguishes the noise from the rest of the process.

For LPGP3, $M1_{WOD}$ greatly overestimates the noise, resulting in a posterior distribution of $R$ that closely resembles the prior distribution. This phenomenon is discussed in Chapter 6.1.6. For this process, $M2_{WDf}$ manages to recover a good estimate $\hat{R}$, and better captures the true value of $\sigma_u$ in its posterior distribution, despite the slight bias of its mean.

### 6.1.3 Simulated Observations from the True Idealized Process (TP)

Next, we examine the data simulated from TP, where a discrepancy arises between the model and the observation process due to linearization. The posterior distributions in Figure 6.6 show that both models indicate a value of $R$ around 0.21, deviating from the true value of 0.2. As seen in Chapter 2, the linearization affects the frequency, which is explained by the parameter $R$. This means that the discrepancy term in $M2_{WOD}$ fails to capture the systematic discrepancy between the model and the process that comes from the linearization of the model.

Chapter 6.1.5 explores this issue further.

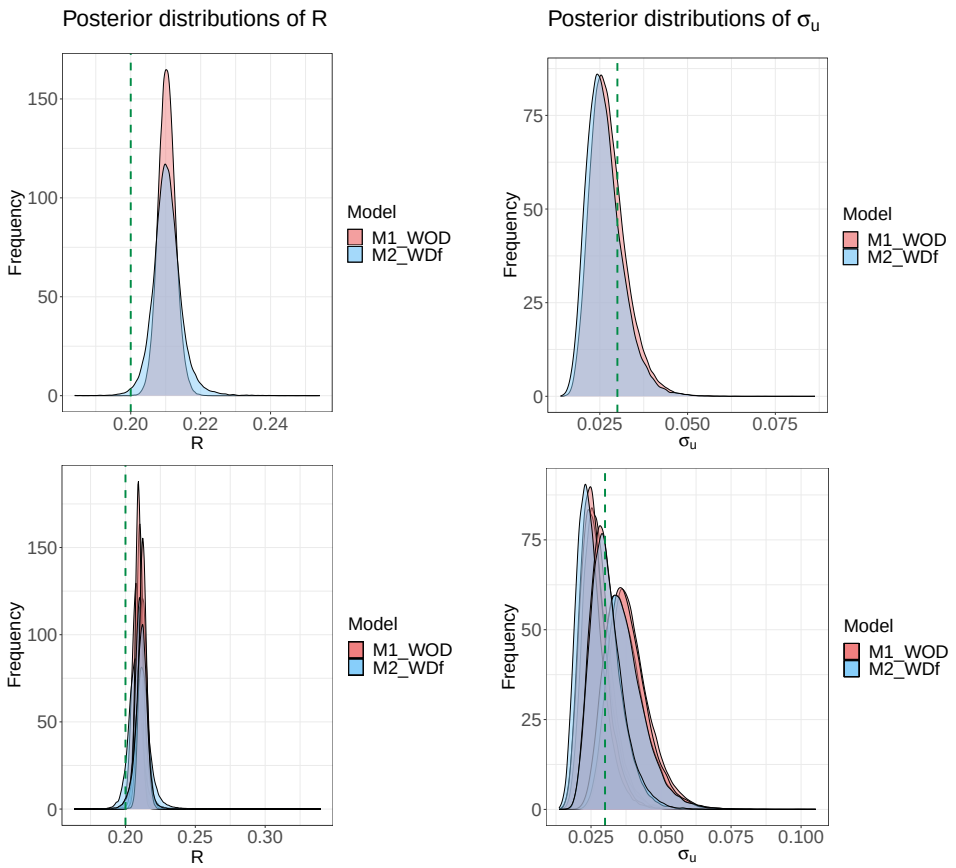


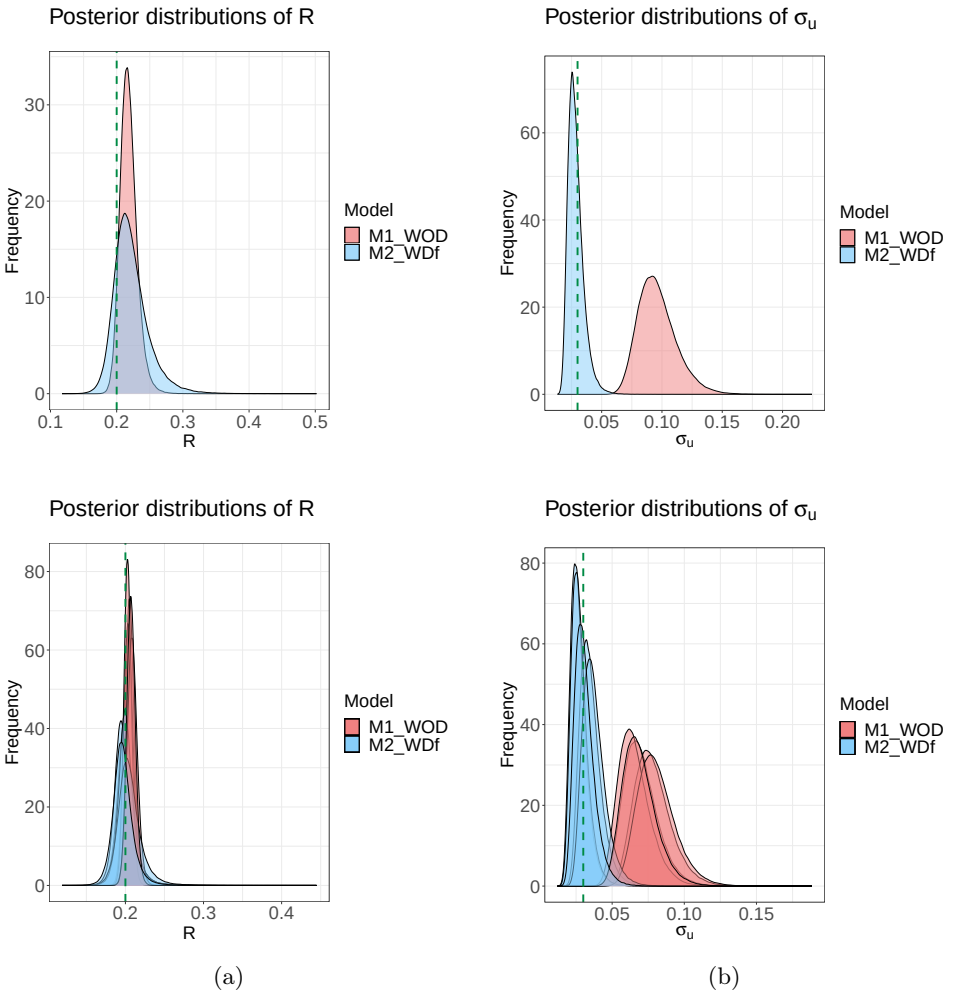

Figure 6.6: Comparison of posterior distributions from observations following TP with $N_u = 20, N_f = 20$ and equally distributed noisy observations from 0 to 4 seconds. The first row displays results from one observation realization, while the second row shows posteriors from five different observation realizations.

### 6.1.4 Simulated Observations from the Damped Process (DP)

In the case of analyzing the posterior distributions from DP, it is reasonable to simplify the analysis by assuming that the linearization in the model influences the parameter $R$ (frequency), while the damping mechanism in the process affects the parameter $\sigma_u$ (measurement noise).

Similarly to TP, both models exhibit a rightwards bias in the posterior distribution of $R$ (Figure 6.7), as discussed in Chapter 6.1.5.

However, in Figure 6.7, $M2_{WDf}$ successfully estimates the posterior of $\sigma_u$, whereas $M1_{WOD}$ fails to do so. The increasing discrepancy in $y$-direction is interpreted as measurement noise by $M1_{WOD}$, while $M2_{WDf}$ captures this systematic deviation by its discrepancy term, effectively distinguishing it from the rest of the process.




(a) (b)

Figure 6.7: Comparison of posterior distributions from observations following DP with $N_u = 20, N_f = 20$ and equally distributed noisy observations from 0 to 4 seconds. The first row displays results from one observation realization, while the second row shows posteriors from five different observation realizations.

### 6.1.5 When the Model Accounting for Discrepancy Fails to Capture Deviation Due to Linearization

In Chapter 6.1.3, the model $M2_{WDf}$ fails to capture the systematic discrepancy between the LP-based model and the TP-based process due to linearization. To investigate this further, we analyze the mean squared error (MSE) between TP and LP. By continuously simulating TP without noise using the true value of $R = 0.200$, we calculate the deterministic MSE between TP and continuously simulated LPs for different values of $R$. The resulting MSE values are plotted in red in Figure 6.8, where the minima indicate the $R$ values of LP corresponding to the lowest MSE with TP.

We observe that the minimum MSE does not align with the corresponding LP value, but rather occurs at $R = 0.210$ with an MSE of $2.19 \cdot 10^{-06}$. This finding suggests that the model fits the data better for $R = 0.210$ than its true value of $R = 0.200$. In other words, a TP with $R = 0.200$ closely resembles an LP with $R = 0.210$, more so than an LP with $R = 0.200$. Consequently, this explains the prominent concentration of posterior distributions around $R = 0.210$ for both $M1_{WOD}$ and $M2_{WDf}$. In essence, $M2_{WDf}$ fails to identify the systematic discrepancy arising from the linearization in the model because it more closely resembles a model process with a different value of $R$.

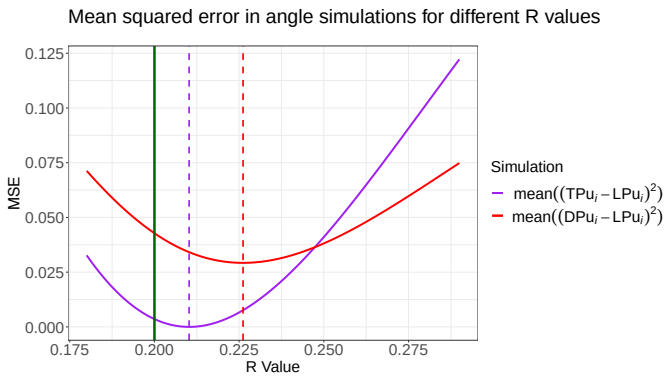



Figure 6.8: Mean squared error in angle simulations. The green line is the true value of $R$, while purple and red dashed lines are the minimum mean squared error between the angle measurements of LP-TP and LP-DP, respectively, for different values of R.

Moving on to the estimation of $R$ for DP, we find that the minimum MSE between

LP and DP occurs at $R = 0.226$ with an MSE of 0.0293, deviating from the true value of $R = 0.200$. This observation explains the rightward bias observed in the posterior distributions depicted in Figure 6.7. Although the bias may appear less prominent compared to that of TP, it is due to the wider overall posterior distribution.

### 6.1.6 When the Model Interprets the Process as a Zero-Amplitude, Noisy Process

Specific combinations of observation data and models can lead to a misinterpretation of the observations as independent noise centered around zero. Consequently, the model overestimates the magnitude of the noise, resulting in an estimated value of $\hat{\sigma_u} \approx 0.3$, which significantly deviates from the true value. As a consequence, the posterior distribution of parameter $R$ remains similar to the prior distribution. An illustrative example of this scenario can be observed in LPGP3 in Figure 6.5b, where the true value of $\sigma_u$ is 0.03, but the posterior distribution centers around 0.3. The models the results as shown in Figure 6.9. Here, the red dotted lines represent the predictions of $M1_{WOD}$, while the blue dotted lines represent the predictions of $M2_{WDf}$. The corresponding 95% credible intervals are shown in the same color as their respective predictions. The black line represents the underlying simulation process (LPGP3), and the black "x" marks represent the noisy input observations used for inference.

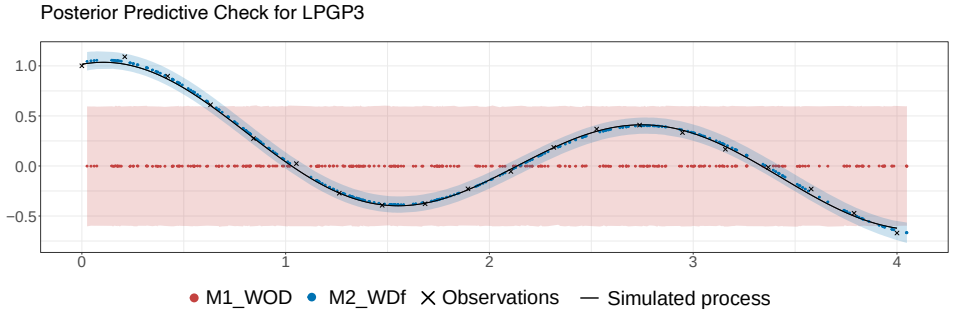



Figure 6.9: Posterior predictive of LPGP3. The red and blue dotted lines show the predictions from $M1_{WOD}$ and $M2_{WDf}$, respectively. The blurred red and blue shows the corresponding 95% credible interval for $M1_{WOD}$ and $M2_{WDf}$, respectively (E1).

As illustrated, the model $M1_{WOD}$ interprets this situation as a process with no amplitude ($u_0 = 0$), as shown in Figure 6.9 indicating that the equation holds true for all values of $R$, as

$$u(t) = u_0 \cos\left(\sqrt{\frac{1}{R}}t\right), \quad u_0 = 0 \tag{6.1}$$

does not depend on $R$.

To explain this, we first look at Figure 6.10, which illustrates the wide range of potential frequency values captured by the prior distribution of $R$. Additionally, it is important to note that the model possesses no prior knowledge regarding the initial angle or amplitude of the process, apart from the information provided by the observations together with the prior on the measurement noise.

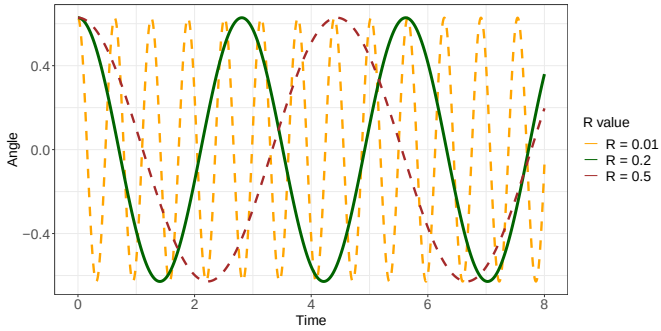



Figure 6.10: Illustration that shows the range of frequencies encompassed by the prior distribution of $R$ in the pendulum motion. The green line corresponds to the specific value $R = 0.2$ we use in the simulations.

The true solution involves a combination of $R$ and $\sigma_u$. The model considers the parameter priors, physics-informed priors, and observed data to determine the probabilities of different solutions.

However, the solution space for $R$ is relatively small, as depicted in Figure 6.11. While the specific case in Figure 6.11, which exhibits the solution space of $R$ for LP with $\sigma_u = 0.03$, successfully converges to the correct solution, it highlights the challenges and limitations of finding a solution that strictly adheres to the constraints of $M1_{WOD}$. In $M1_{WOD}$, the model attempts to explain the observations as a combination of the underlying process and independent noise, without incor-

porating an additional term to account for other deviations, as $M2_{WDf}$ does. This makes $M1_{WOD}$ more sensitive to deviations that do not originate from the process or the independent noise. Consequently, for some deviating processes, the only viable solution within the framework of $M1_{WOD}$ is a noisy zero-amplitude process.

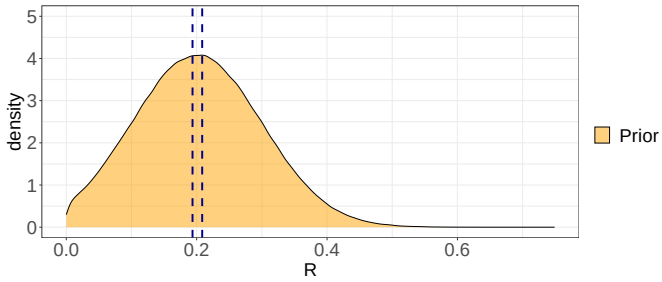



Figure 6.11: Valid solutions of $R$ for LP with $\sigma_u = 0.03$ are in between the two dark blue dashed lines. The yellow distribution is the prior distribution of $R$.

### 6.1.7 Posterior Predictive Checks and Takeaways from Experiment 1

In order to gain a more intuitive assessment of our results, we proceed by simulating replicated data based on the fitted models and comparing them to the observed data.

Figure 6.12 shows the posterior predictive checks (PPCs) for $M1_{WOD}$ and $M2_{WDf}$ applied to each simulation process in E1. It is important to note that in regions where only the blue PPCs are visible, the red PPCs are almost identical, with the blue PPCs effectively covering them.

In Figure 6.12, both models show similar and accurate predictions for LP and TP, with reasonable $CI$ widths. Note that for TP, the PPC aligns with the process, even though the estimations of $R$, as shown in Chapter 6.1.3, are biased. These observations are consistent with the scenario emphasized by Kennedy and O'Hagan (2001), who discusses that the true values of physical quantities do not necessarily coincide with $\Phi$ and that the physically true value of a calibration parameter can result in a poorer fit and less accurate predictions than an alternative value.

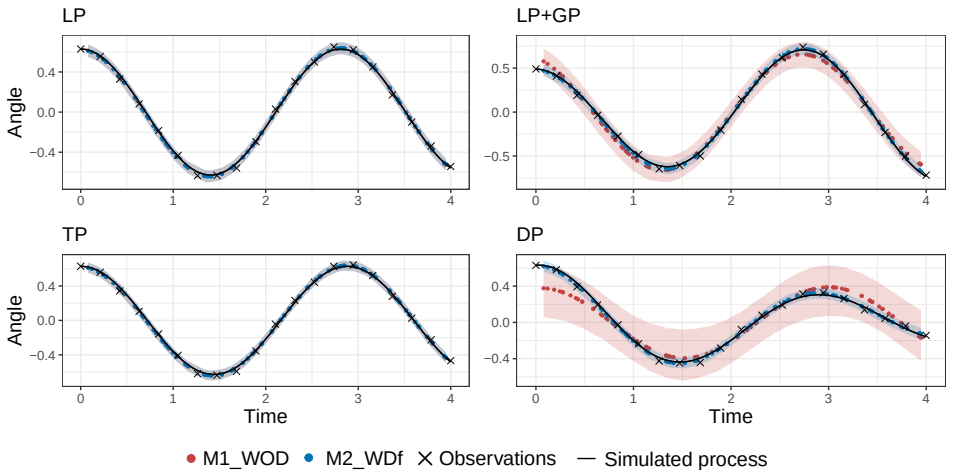



Figure 6.12: Posterior predictive checks for E1. The red and blue dotted lines show the predictions from $M1_{WOD}$ and $M2_{WDf}$, respectively. The blurred red and blue shows the corresponding 95% credible interval for $M1_{WOD}$ and $M2_{WDf}$, respectively. LP+GP denotes the process LPGP1. Note that for LP and TP, the posterior predictive curves for the two models exhibit almost identical behavior. Consequently, the results obtained for $M2_{WDf}$ cover the results obtained for $M1_{WOD}$.

However, for LPGP, $M1_{WOD}$ deviates from the true process near the maximums and has a wider 95% credible interval, indicating higher measurement uncertainty than $M2_{WDf}$. This aligns with the findings in Figure 6.4, where $M1_{WOD}$ tends to overestimate the measurement noise. In contrast, $M2_{WDf}$ effectively predicts unknown measurements and closely matches the actual data. As discussed, the discrepancy is attributed to the relative flexibility of the models, with $M1_{WOD}$ being less adaptable in handling process deviations.

$M1_{WOD}$ misinterprets the damping as measurement noise in an undamped process when examining the data simulated from DP. This results in an underestimation of the amplitude at the beginning of the process and an overestimation towards the end, and an overestimation of the noise as shown in Figure 6.7. On the other hand, $M2_{WDf}$ more accurately predicts the unknown measurements due to its extra flexibility in the discrepancy term.

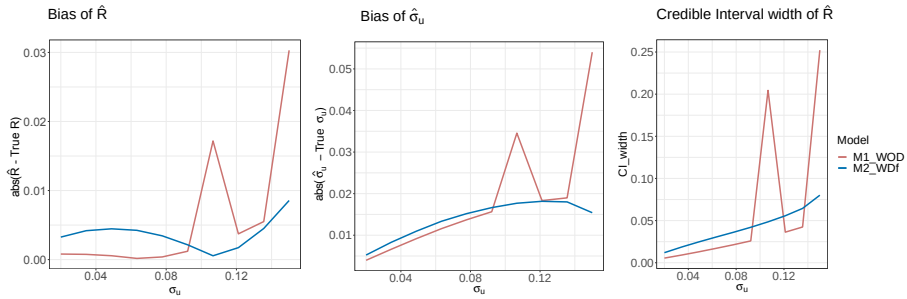
The PPCs in Figure 6.12 highlight that $M2_{WDf}$ accurately predicts all four different processes, while $M1_{WOD}$ only succeeds in accurately predicting LP and TP.

## 6.2 Resuts from Experiment 2 (E2)

In Experiment 2, we examine two variations: E2a, which involves varying the noise parameter $\sigma_u$, and E2b, where we increase the number of observed cycles. Detailed configurations can be found in Chapter 5. These results are evaluated in conjunction with those from Chapter 6.1, which includes the posterior distributions for $\sigma_u = 0.03$ and $\Delta t = 4s$.

### 6.2.1 Experiment 2a (E2a): Measurement Noise

This experiment examines the performance of the models under varying levels of measurement noise variance, as described in Chapter 5.2. Each run consists of 3 chains of $200,000$ HMC iterations. The objective is to compare the models' responses to different noise levels. Note that the $y$-axis in the plots varies depending on the process, as the main focus is comparing the two models for each process.



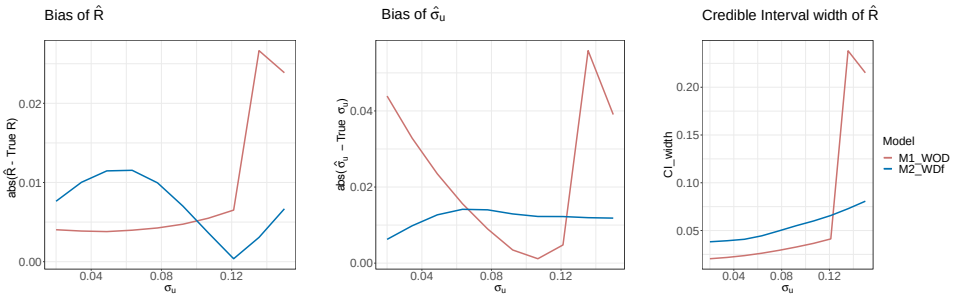

(a) Absolute bias of $\hat{R}$ (b) Absolute bias of $\hat{\sigma}_u$ (c) CI width of $\hat{R}$

Figure 6.13: Absolute bias of $\hat{R}$, absolute bias of $\hat{\sigma}_u$, and width of credible interval for LP at varying levels of measurement noise determined by $\sigma_u$.

Figure 6.13 shows the bias and credible interval width for $R$ and $\sigma_u$ across different values of $\sigma_u$. We observe that as the noise level increases, the stability of $M1_{WOD}$ in estimating $R$ and $\sigma_u$ diminishes compared to $M2_{WDf}$. Notably, there are peaks in the biases and credible interval width of $R$ for $M1_{WOD}$ at $\sigma_u = 0.11$ and $\sigma_u = 0.15$. This suggests that accurate estimates with $M1_{WOD}$ may require more iterations or yield incorrect results for the given noisy observations, highlighting its lower flexibility compared to $M2_{WDf}$. In Appendix C.1, a similar experiment with $300,000$ HMC iterations demonstrates the absence of the peak at $\sigma_u = 0.11$

for $M1_{WOD}$. This indicates that increasing the number of iterations can improve the accuracy of estimates with $M1_{WOD}$.

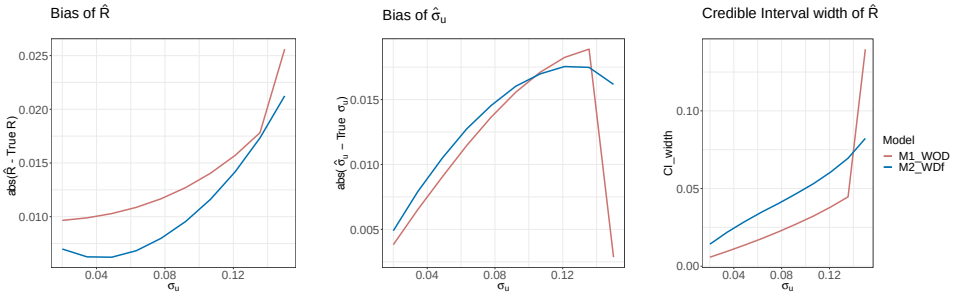
Figure 6.14 shows the same diagnostics, but for LPGP1 observations. For $\sigma_u < 0.06$, the results align with Figure 6.4. In Figure 6.14b, $M1_{WOD}$ exhibits decreasing bias of $\sigma_u$ until a peak is reached at $\sigma_u = 0.12$. In $M2_{WOD}$, the bias of $R$ shifts from leftward to rightward at $\sigma_u = 0.12$. These findings reinforce the greater stability of $M2_{WDf}$ in converging to the correct outcome.




(a) Absolute bias of $\hat{R}$ (b) Absolute bias of $\hat{\sigma}_u$ (c) CI width of $\hat{R}$

Figure 6.14: Absolute bias of $\hat{R}$, absolute bias of $\hat{\sigma}_u$, and width of credible interval for LPGP1 at varying levels of measurement noise determined by $\sigma_u$.

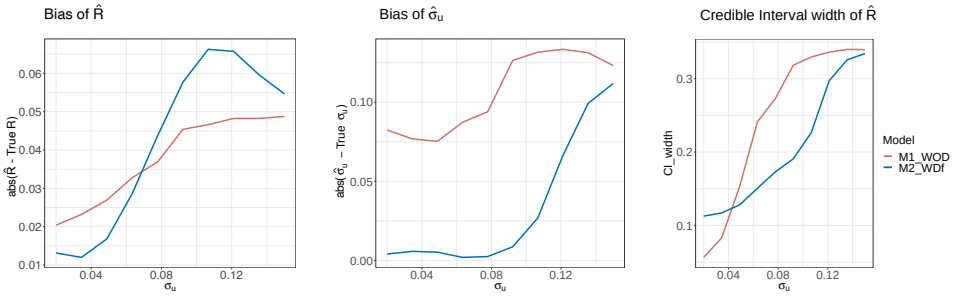
Figure 6.15 displays the bias and CI-width for TP. A general increasing trend is observed. However, a notable finding emerges: a dip is observed for $M1_{WOD}$ at $\sigma_u = 0.15$, accompanied by a peak in both the bias and CI-width of $R$. This suggests that the model has identified a posterior distribution of $R$ that is both incorrect and uncertain while still providing a reasonable estimation of $\sigma_u$. Despite previous observations suggesting that $M2_{WDf}$ failed to capture the systematic discrepancy between LP and TP, a closer examination of the bias of $R$ in Figure 6.15a reveals consistently slightly improved estimates by $M2_{WDf}$ compared to $M1_{WOD}$. However, this improvement comes at the expense of a slightly wider credible interval for $R$, as depicted in Figure 6.15c. While the complete systematic deviation between the model and TP remains unaccounted for, it is evident that $M2_{WDf}$ captures a portion of the discrepancy, resulting in slightly less biased results compared to $M1_{WOD}$. As discussed earlier, the presence of remaining bias actually results in a superior fit to the data and enhanced predictive performance, compared to the scenario where the estimates are unbiased.




(a) Absolute bias of $\hat{R}$ (b) Absolute bias of $\hat{\sigma}_u$ (c) CI width of $\hat{R}$

Figure 6.15: Absolute bias of $\hat{R}$, absolute bias of $\hat{\sigma}_u$, and width of credible interval for TP at varying levels of measurement noise determined by $\sigma_u$.

Figure 6.16 shows the performance of $M2_{WOD}$ compared to $M1_{WOD}$ for DP. $M2_{WOD}$ demonstrates the ability to distinguish between measurement noise and damping at lower noise levels, but this becomes more challenging as the noise level increases, leading to bias and wider confidence intervals beyond a certain noise threshold. The consistent overestimation of noise in the damped process by $M1_{WOD}$ is due to the absence of a term capturing systematic damping. As the noise level increases, the estimates of $R$ from $M2_{WDf}$ are more biased than for $M1_{WOD}$. This indicates that the flexibility of the discrepancy term in a highly noisy process makes it challenging to discern between systematic deviation and measurement noise, resulting in an incorrect estimate of the frequency, represented by $R$.




(a) Absolute bias of $\hat{R}$ (b) Absolute bias of $\hat{\sigma}_u$ (c) CI width of $\hat{R}$

Figure 6.16: Absolute bias of $\hat{R}$, absolute bias of $\hat{\sigma}_u$, and width of credible interval for DP at varying levels of measurement noise determined by $\sigma_u$.

### Posterior Predictive Checks for E2a

Figure 6.17 shows the PPCs for the four processes with a noise parameter of $\sigma_u = 0.1$. Similar to E1 with $\sigma_u = 0.03$, the predictions for LP and TP remain close to the true values, but with wider 95% CIs due to increased noise. The differences between the models are less pronounced compared to the lower noise level.

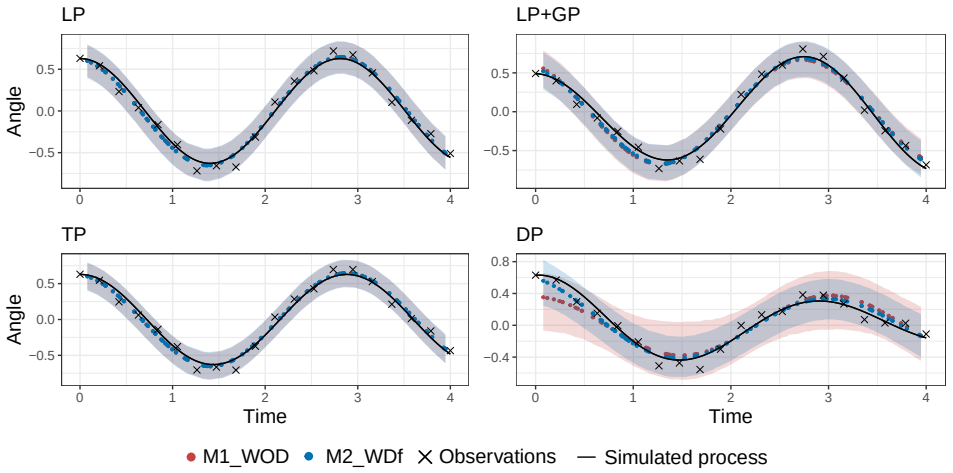



Figure 6.17: Posterior predictive for E2a with noise parameter of $\sigma_u = 0.1$. The red and blue dotted lines show the predictions from $M1_{WOD}$ and $M2_{WDf}$, respectively. The blurred red and blue show the corresponding 95% credible interval for $M1_{WOD}$ and $M2_{WDf}$, respectively. Note that for LP and TP, the posterior predictive curves for the two models exhibit almost identical behavior. Consequently, the results obtained for $M2_{WDf}$ cover the results obtained for $M1_{WOD}$ (E2a).

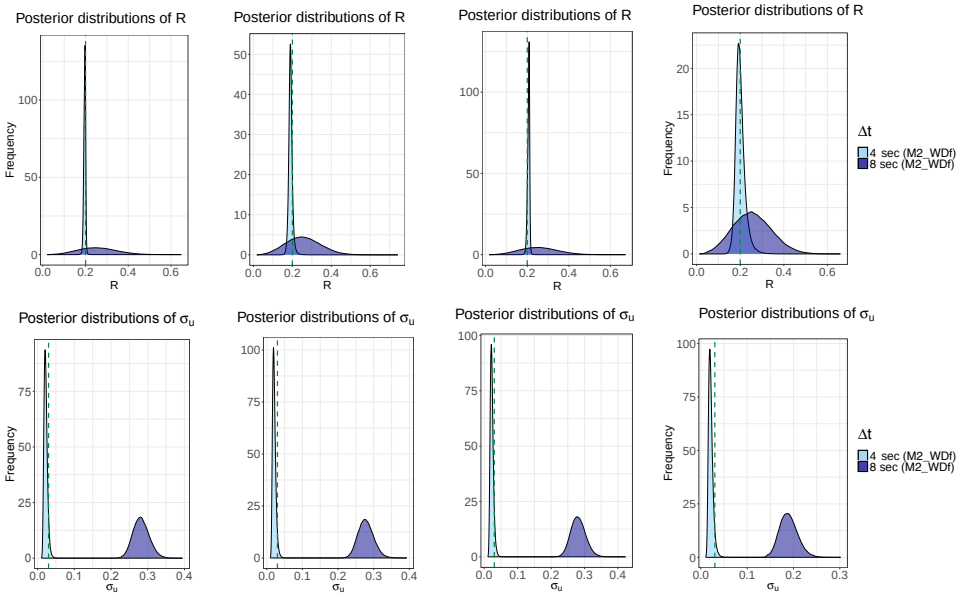
In particular, for LPGP1, the models perform almost similarly, contrasting the results for $\sigma_u = 0.03$. This aligns with the findings in Figure 6.14, where $M2_{WDf}$ does not outperform $M1_{WOD}$ for noise levels approaching and exceeding $\sigma_u \approx 0.1$.

However, there is still a visible distinction in the PPCs for DP, with $M1_{WOD}$ exhibiting inaccuracies in estimating the decaying amplitude and consistently overestimating the noise, consistent with the results in Figure 6.16b.

### 6.2.2 Experiment 2b (E2b): Number of Cycles

This experiment investigates the effect of doubling the number of observed cycles. It is conducted in two variations: E2b-1 maintains the same number of measurements, resulting in half the measurement density as before, while E2b-2 doubles the measurement number to maintain the same measurement density as previous experiments. Each run consists of 5 chains of 100,000 HMC iterations. The rest of the configurations remain as before.

#### Doubled Number of Cycles, Halved Measurement Density (E2b-1)




(a) LP (b) LPGP1 (c) TP (d) DP

Figure 6.18: Posterior distributions from $M2_{WDf}$ from processes with $N_u = N_f = 20$, and true value of $R = 0.2$ as the green dashed line. Dark blue has the doubled amount of cycles in the input data, as the observations are equally spread over $\Delta t = 8s$, whereas light blue has an observation period of $\Delta t = 4s$ (E2b-1).

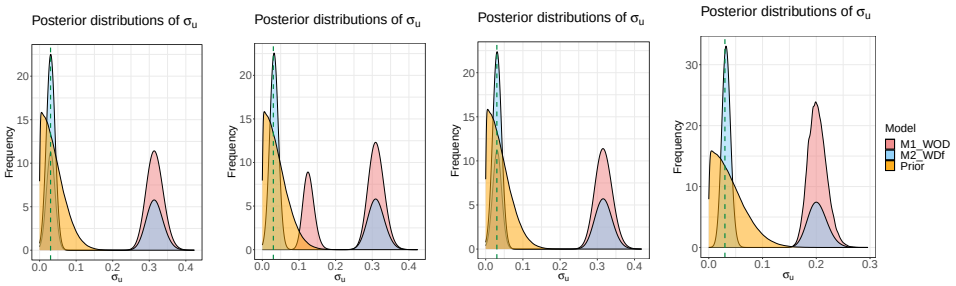
Figure 6.18 shows comparisons of posterior distributions from $M2_{WDf}$ of the parameters for the four processes using $M2_{WDf}$ with $N_u = N_f = 20$ measurement

points and different lengths of the observation period: $\Delta t = 4s$ (light blue) and $\Delta t = 8s$ (dark blue). We observe that spreading the same number of observations over a larger number of cycles leads the model to estimate a process with negligible amplitude and high noise as the most probable outcome. This phenomenon is discussed in Chapter 6.1.6 for LPGP3. These results shows that selecting a time interval spanning multiple cycles does not improve parameter estimation accuracy within this number of measurements. In Appendix C.2, we see that the same yields for $M1_{WOD}$. Intuitively, this observation makes sense, as detecting patterns becomes more challenging with lower observation density.

In Appendix C.3, the posterior distributions for the halved number of cycles ($\Delta t = 2$), with $N_u = N_f = 20$, i.e., the doubled measurement density, are compared to the standard settings ($\Delta t = 4$). The comparison shows that the posteriors are nearly identical between the two scenarios.

### Doubled Number of Cycles, Maintained Measurement Density (E2b-2)

Next, we explore the results obtained for $\Delta t = 4s$, $N_u = 20$, $N_f = 20$ and $\Delta t = 8s$, $N_u = 40$, $N_f = 40$, where the aim is to explore the impact of doubling the observation period while maintaining the same measurement density. The posterior distributions of $\sigma_u$, along with their prior distributions, are shown in Figure 6.19. Interestingly, we observe bimodal posteriors for the variable $M2_{WDf}$ in all processes, while for $M1_{WOD}$, bimodal posteriors are obtained for all processes except DP where the posterior centers around the noisy zero-amplitude solution. This is further discussed in Chapter 6.2.3.



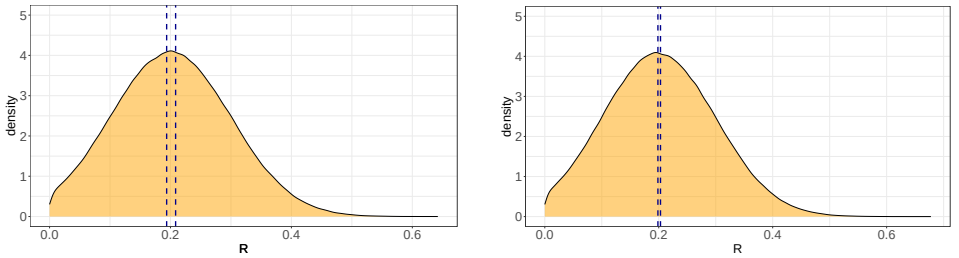

(a) LP (b) LPGP (c) TP (d) DP

Figure 6.19: Posterior distributions from $M2_{WDf}$ from processes with the true value of $R = 0.2$ as the green dashed line. $N_u = 40$ and $= N_f = 40$ observations are equally spread over $\Delta t = 8s$ (E2b-2).

### 6.2.3 Bimodality in Posterior Distributions for More Cycles in the Observations

We see that contrary to the expectation of improved results for an increased observation interval with the same measurement density, we find that the opposite is true. In the results of E2b-2, we observed bimodality between the estimates converging to the correct estimate or to the same phenomenon as discussed in Chapter 6.1.6, where the posterior distribution of $R$ closely resembles the prior distribution, assuming a process with zero amplitude and noisy measurements.

This finding can have two potential explanations. Firstly, it is possible that some chains have not converged despite conducting the same number of MCMC iterations compared to the corresponding $\Delta t = 4s$ scenario. Secondly, the bimodal posteriors suggest the existence of two distinct solutions with different probabilities, indicating that the chain has converged to a bimodal distribution. In Appendix C.4 a trace plot for such a case, specifically for LPGP from Figure 6.18b. The plot reveals the absence of transitions between the modes, indicating that the chain is trapped and unable to move freely between the modes.




(a) Prior and solution space of $R$ for $\Delta t = 4s$ when $\sigma_u = 0.03$.

(b) Prior and solution space of $R$ for $\Delta t = 8s$ for LP when $\sigma_u = 0.03$.

Figure 6.20: Valid solutions of $R$ for LP with $\sigma_u = 0.03$ are in between the two dark blue dashed lines, $\Delta t = 4s$ and $\Delta t = 8s$, respectively. The yellow distribution is the prior distribution of $R$.

The reason why this phenomenon occurs for longer measurement periods but not for the corresponding shorter ones, even with the same number of iterations, can be attributed to the fact that for a smaller fraction of the process, the solution space of $R$ is larger due to its ability to vary to some extent while still conforming to the process. However, when dealing with more cycles, the solution space of $R$

becomes narrower. This is because the frequency must accommodate all the data points within the cycles, leaving less room for deviation since any errors in the frequency estimation would propagate and lead to inaccurate predictions. In the illustration in Figure 6.20, we see that the values of $R$ the possible solution can take is significantly narrower for $\Delta t = 8s$ than for $\Delta t = 4s$. This undermines our explanation for the difficulty in achieving a correct convergence of solutions with more cycles in the observations. Consequently, it increases the likelihood of becoming trapped in the zero-amplitude solution.

### Posterior Predictive Checks for E2b-2

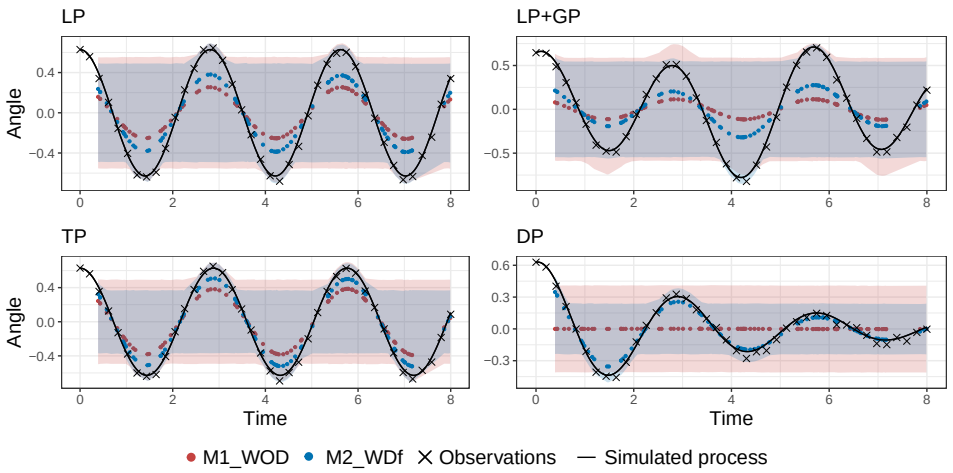
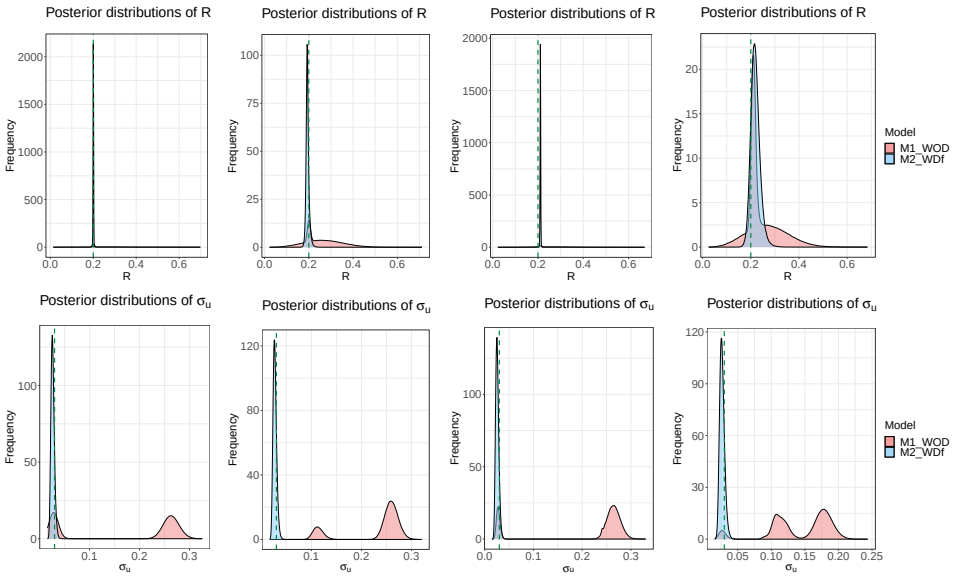



Figure 6.21: Posterior predictive for E2b. The red and blue dotted lines show the predictions from $M1_{WOD}$ and $M2_{WDf}$, respectively. The blurred red and blue shows the 95% credible interval for $M1_{WOD}$ and $M2_{WDf}$, respectively. Prior on $\sigma_u \sim \mathcal{N}(0, 0.05^2)$.

Next, we look at the PPCs. Figure 6.21, displays the PPCs for all four processes with $\Delta t = 8s$ and $N_u = N_f = 40$. We see that this is a middle of the bimodal posterior distributions obtained for the experiments, and we also see that for DP, where the posterior of $M1_{WOD}$ was not bimodal, the model predicts a zero-amplitude noisy process.

## More Informative Priors on Noise

In our scenario, the prior distribution of $\sigma_u$ contains more information than that of $R$. As a result, a significant improvement in model performance would require a substantial increase in the information content of $R$ relative to the impact of increasing the information content of $\sigma_u$.




(a) LP (b) LPGP (c) TP (d) DP

Figure 6.22: Posterior distributions with prior on $\sigma_u \sim \mathcal{N}(0, 0.03^2)$ instead of $\sigma_u \sim \mathcal{N}(0, 0.05^2)$. Observations from from processes with $N_u = N_f = 40$, and true value of $R = 0.2$ as the green dashed line.

We investigate the models using a more informative prior on $\sigma_u$, specifically $\sigma_u \sim \mathcal{N}(0, 0.03^2)$ instead of $\sigma_u \sim \mathcal{N}(0, 0.05^2)$. The posterior distributions are depicted in Figure 6.22, while the PPCs can be observed in Figure 6.23. It is evident that due to the increased flexibility, $M2_{WDf}$ better captures the underlying process, while $M1_{WOD}$ still gets bimodal posteriors, indicating two solutions with different probabilities or that the chains have not yet converged. This bimodality is also observed for the posteriors of LP from $M1_{WOD}$, although the model aligns with the process of LP.

This observation suggests that with greater flexibility, $M2_{WDf}$ explores the posterior space more effectively and is less prone to becoming trapped in an incorrect mode. Nevertheless, as observed in E2b-2 with less informative priors on the noise, it requires either a larger number of iterations than the ones conducted in that particular case, or more informative priors, as investigated here. Note that it appears that $M2_{WDf}$ slightly underestimates the noise with this informative prior that is biased to the left, placing a bit too much trust in observations.

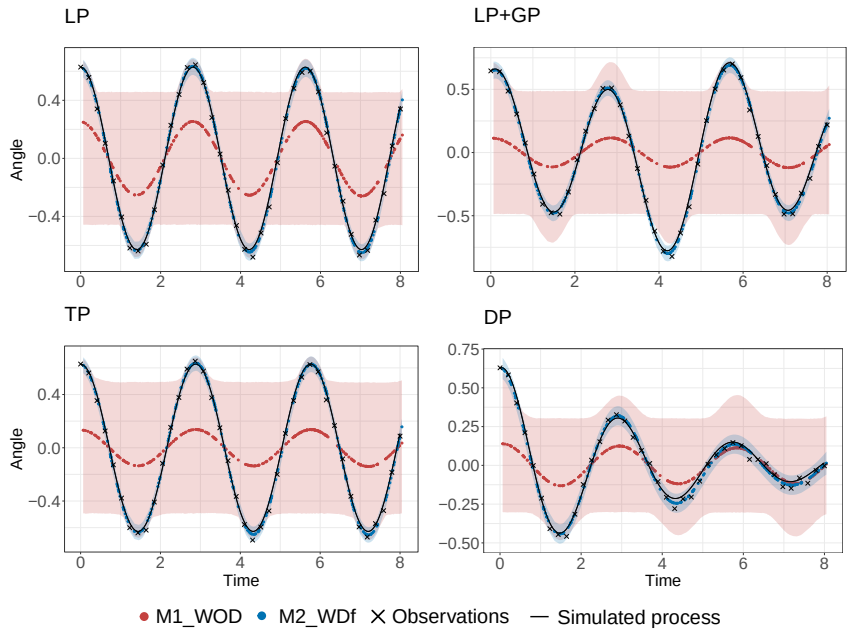



Figure 6.23: Posterior predictive for E2b with more informative priors. The red and blue dotted lines show the predictions from $M1_{WOD}$ and $M2_{WDf}$, respectively. The blurred red and blue shows the 95% credible interval for $M1_{WOD}$ and $M2_{WDf}$, respectively. Prior on $\sigma_u \sim \mathcal{N}(0, 0.03^2)$ instead of $\sigma_u \sim \mathcal{N}(0, 0.05^2)$.

The set of experiments in E2b shows that the model struggles when dealing with more cycles, but that $M2_{WDf}$ tends to explore the posterior space more effecively, and is more affected by more informative priors. In the case of several cycles in the input data, a different kernel choice is worth considering, as indicated by the promising results of the periodic kernel presented in Spitieris and Steinsland (2022). However, it is clear that a shorter fraction of the input data cycles is desired for the current kernel to perform effectively.

# Chapter 7

# Summary and Conclusion

In this thesis, we have explored the effectiveness of incorporating a discrepancy term in a physics-informed Bayesian Calibration model, specifically for a homogeneous linearized ODE. We use a model that follows a homogeneous linear differential equation $\mathcal{L}_t^{\Phi}(u(t)) = f(t) = 0$, where $\mathcal{L}_t^{\Phi}$ is the linear differential operator with respect to $t$, and $\Phi$ is the physical parameters. We constructed one model that does not account for discrepancy between the model and the observation process, and one that does.

Setting up the models requires initial effort and knowledge of the physics of the process. The inclusion of a discrepancy term offers a flexibility that aims to distinguish the systematic discrepancy from the rest of the observed process.

In contrast to Spitieris and Steinsland (2022), which incorporated the discrepancy on the process that is assumed to include a systematic deviation, which in this case is $u(t)$, we included the discrepancy term on the opposite side, $f(t)$. This is interpreted as follows: We represent a process following the model as $u_M(t)$. However, if the actual observation process is $u(t) = u_M(t) + \Delta u(t), \Delta u(t) \neq 0$, the model is incorrect. In this case, we consider it as if the process originates from the model process $u_M(t)$ but is affected by an external force $f_{force}(t)$ that compels the model process to resemble the actual process $u(t) = u_M(t) + \Delta u(t)$. Hence, we introduce a discrepancy term $\delta(t)$ into $f(t)$ to nullify this force $f_{force}(t)$. This results in $\mathcal{L}_t^{\Phi}(u(t)) = f_{force}(t) + \delta(t) = 0$.

We have investigated the models' behavior compared to a model without a discrepancy term for noisy observations from four processes: (1) when there is no

systematic deviation between the model and the observed process, (2) when the process is nonlinear so that the deviation arises from the linearization (3) when the deviation follows a Gaussian Process (GP) and (4) when the process involves an additional damping term that is not accounted for in the model.

When analyzing the models' performance on a process with a controlled deviation from the model, which follows the GP assumption, we observe that the model containing a discrepancy term successfully distinguishes the systematic deviation from the actual process for a linear process. It recovers the true values in its estimates, unlike the model without discrepancy, which either categorizes the deviation as measurement noise or fails to estimate the process entirely, resulting in estimates of a process with zero amplitude and high levels of independent and identically distributed noise.

However, the linearized model fails to accurately capture deviations from the nonlinear process due to linearization in its discrepancy term. This is because the nonlinear process more closely resembles a linearized process with a different parameter. This incorrect estimate provides a better fit to the data and leads to improved predictions, a scenario discussed by Kennedy and O'Hagan (2001).

On the other hand, when estimating the noise for a damped process, the model with a discrepancy term outperformed the model without a discrepancy term. However, when the observations contain high noise levels, the model with a discrepancy term struggles to distinguish systematic deviations from independent noise.

Additionally, we observed indications that the model with a discrepancy term exhibited more stability in terms of convergence: it consistently needed fewer HMC iterations to converge to the solution compared to the model without a discrepancy term, which became clear when we had more noisy observations and when the input data consisted of more cycles. This suggests that the increased flexibility in the model with the discrepancy term enables more effective exploration of the posterior space, reducing the likelihood of getting trapped in suboptimal modes.

With an increased number of cycles in the observations, the solution space narrows, rendering both models more sensitive to deviations from the process. The model without a discrepancy term becomes particularly sensitive, lacking the flexibility to distinguish deviations from the process and only being able to account for observations originating from the process or independent noise. Conversely, even when the observations stem from the same process as the model without discrepancy, the model with the additional discrepancy term exhibits a higher likelihood of converging to the correct distributions. In the study by Spitieris and Steinsland (2022), the performance of three different covariance functions for the kernel was examined. Notably, the periodic kernel demonstrated promising results in

capturing periodic behaviors. The potential benefits of applying this kernel when dealing with observations spanning multiple cycles are worth considering. Further investigation into the application of the periodic kernel in such scenarios may lead to improved outcomes, including modeling the discrepancy with a periodic kernel when the deviation between the model and the process exhibits periodic behavior.

## Conclusion

In conclusion, incorporating a discrepancy term as an external force in the physics-informed Bayesian Calibration model offers promising results in identifying systematic deviations and improving estimation accuracy. Evidently, the added flexibility provided by the discrepancy term allows for more effective exploration of the posterior space and increases the likelihood of finding the correct solution for a given number of HMC iterations. However, a fundamental problem arises when applying this approach to a deviating process closely resembling the model process with different parameters. This situation becomes evident when using a linearized model with a discrepancy term on a nonlinear process. In such instances, it becomes impossible for the discrepancy term to accurately capture the deviation completely, hindering the model's ability to recover the true parameter estimates. However, despite the incorrect parameter estimates, this can result in a better fit of the data, leading to improved predictions.

# References


Andrieu, C. et al. (2003). 'An introduction to MCMC for machine learning'. In: *Machine Learning* 50.1, pp. 5–43. DOI: 10.1023/A:1020281327116. URL: https://doi.org/10.1023/A:1020281327116.

Bayarri, M. J. et al. (2007). 'A Framework for Validation of Computer Models'. In: *Technometrics* 49.2, pp. 138–154. DOI: 10.1198/004017007000000092. URL: https://doi.org/10.1198/004017007000000092.

Betancourt, M. (2017). *Robust Gaussian Processes in Stan*. URL: https://betanalpha.github.io/assets/case_studies/gp_part3/part3.html#1_initial_setup.

— (2018). *A Conceptual Introduction to Hamiltonian Monte Carlo*. arXiv: 1701.02434 [stat.ME].

Brynjarsdóttir, J. and A. O'Hagan (2014). 'Learning about physical parameters: the importance of model discrepancy'. In: *Inverse Problems* 30.11, p. 114007. DOI: 10.1088/0266-5611/30/11/114007. URL: https://dx.doi.org/10.1088/0266-5611/30/11/114007.

Butcher, J. C. (2008). *Numerical Methods for Ordinary Differential Equations*. New York: John Wiley & Sons. ISBN: 978-0-470-72335-7.

Chasnov, J. R. (2022). *11. The Damped, Driven Pendulum*. URL: https://batch.libretexts.org/print/url=https://math.libretexts.org/Bookshelves/Scientific_Computing_Simulations_and_Modeling/Scientific_Computing_(Chasnov)/II%3A_Dynamical_Systems_and_Chaos/11%3A_The_Damped%2C_Driven_Pendulum.pdf.

Duane, S. et al. (1987). 'Hybrid Monte Carlo'. In: *Physics Letters B* 195.2, pp. 216–222. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)91197-X. URL: https://www.sciencedirect.com/science/article/pii/037026938791197X.

Flato, G. et al. (2013). 'Evaluation of Climate Models'. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. F. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Gaetan, C. and X. Guyon (2010). *Spatial Statistics and Modeling*. Springer. URL: https://link.springer.com/content/pdf/10.1007/978-0-387-92257-7.pdf.

Gardner, P. et al. (2021). 'Learning model discrepancy: A Gaussian process and sampling-based approach'. In: *Mechanical Systems and Signal Processing* 152, p. 107381. ISSN: 0888-3270. DOI: https://doi.org/10.1016/j.ymssp.2020.107381. URL: https://www.sciencedirect.com/science/article/pii/S0888327020307676.

Gelman, A., J. B. Carlin et al. (2013). *Bayesian Data Analysis*. Taylor & Francis Ltd. Chap. 1.

Gelman, A., X. Meng and H. Stern (1996). 'Posterior Predictive Assessment of Model Fitness via Realized Discrepancies'. In: *Statistica Sinica*, pp. 733–760. URL: http://www.stat.columbia.edu/~gelman/research/published/A6n41.pdf.

Givens, G. H. and J. A. Hoeting (2012a). *Computational Statistics, Second Edition*. Wiley, pp. 1–17. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118555552.

— (2012b). 'Markov Chain Monte Carlo'. In: *Computational Statistics*. 1st. John Wiley & Sons, Inc. Chap. 1.5. ISBN: 9780470533314. DOI: 10.1002/9781118555552.

— (2012c). 'Markov Chain Monte Carlo'. In: *Computational Statistics*. 1st. John Wiley & Sons, Inc. Chap. 7.3.1. ISBN: 9780470533314. DOI: 10.1002/9781118555552.

Goldstein, H., Charles Poole and J. Safko (2002). *Classical Mechanics*. 3rd. Addison-Wesley.

Gramacy, R. B. (2022). *Surrogates*. CRC Press, Taylor & Francis Group. URL: https://bookdown.org/rbg/surrogates/.

Groetsch, C. W. (Jan. 1993). *Inverse Problems in the Mathematical Sciences*. Originally published by Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden. Springer Fachmedien Wiesbaden. ISBN: 3-528-06545-1. DOI: 10.1007/978-3-322-99202-4.

Higdon, D. et al. (2004). 'Combining Field Data and Computer Simulations for Calibration and Prediction'. In: *SIAM J. Scientific Computing* 26, pp. 448–466. DOI: 10.1137/S1064827503426693.

Hoffman, M. D. and A. Gelman (2014). 'The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo'. In: *Journal of Machine Learning Research* 15.47, pp. 1593–1623. URL: http://jmlr.org/papers/v15/hoffman14a.html.

Honkela, A. (22nd Sept. 2020). *Computational Statistics I*. URL: https://www.cs.helsinki.fi/u/ahonkela/teaching/compstats1/book/hamiltonian-monte-carlo-hmc.html (visited on 17th Apr. 2023).

Kennedy, M. C. and A. O'Hagan (2001). 'Bayesian calibration of computer models'. In: *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*. URL: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00294.

Owen, F. (2014). *Simple pendulum via Lagrangian mechanics*. URL: http://www.aoengr.com/Dynamics/LagrangianMechanicsPendulum.pdf.

Raissi, M., P. Perdikaris and G. E. Karniadakis (2017). 'Machine Learning of Linear Differential Equations Using Gaussian Processes'. In: *Journal of Computational Physics* 348. DOI: https://doi.org/10.1016/j.jcp.2017.07.050.

— (2018). 'Numerical Gaussian Processes for Time-Dependent and Nonlinear Partial Differential Equations'. In: *SIAM Journal on Scientific Computing* 40.1, A172–A198. DOI: 10.1137/17M1120762. eprint: https://doi.org/10.1137/17M1120762. URL: https://doi.org/10.1137/17M1120762.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press. ISBN: 026218253X. URL: http://www.GaussianProcess.org/gpml.

Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Särkkä, S. (2011). 'Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression'. In: *Artificial Neural Networks and Machine Learning – ICANN 2011*. Ed. by Timo Honkela et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 151–158. ISBN: 978-3-642-21738-8. URL: https://doi.org/10.1007/978-3-642-21738-8_20.

Song, Y. et al. (2022). *Solving Inverse Problems in Medical Imaging with Score-Based Generative Models*. arXiv: 2111.08005 [eess.IV].

Spitieris, M. and I. Steinsland (2022). 'Bayesian Calibration of imperfect computer models using Physics-informed priors'. In: arXiv: 2201.06463v3. URL: https://arxiv.org/abs/2201.06463.

— (2023). 'Bayesian Calibration of Imperfect Computer Models using Physics-Informed Priors'. In: URL: https://www.researchgate.net/publication/357925983_Bayesian_Calibration_of_Imperfect_Computer_Models_using_Physics-Informed_Priors.

Stan Development Team (2022). *Stan Functions Reference*. URL: https://mc-stan.org/docs/functions-reference/linear-algebra-functions-and-solvers.html.

Thurber, C. and J. Ritsema (2015). '1.10 - Theory and Observations - Seismic Tomography and Inverse Methods'. In: *Treatise on Geophysics (Second Edition)*. Ed. by Gerald Schubert. Elsevier, pp. 307–337. DOI: https://doi.org/10.1016/B978-0-444-53802-4.00009-9. URL: https://www.sciencedirect.com/science/article/pii/B9780444538024000099.

# Appendix A

# Abbreviations

| | |
|---|---|
| ACF | Autocorrelation Function |
| CI | Credible Interval |
| DP | Damped Pendulum Process |
| $M1\_WOD$ | Model 1 Without Discrepancy |
| $M2\_WDf$ | Model 2 With Discrepancy on f |
| GP | Gaussian Process |
| HMC | Hamiltonian Monte Carlo |
| NUTS | No-U-Turn Sampler |
| LP | Linearized Pendulum Process |
| LPGP | Linearized Pendulum Process with added Gaussian Process |
| MCMC | Markov Chain Monte Carlo |
| MSE | Mean Squared Error |
| $N_{eff}$ or ESS | Effective Sample Size |
| ODE | Ordinary Differential Equation |
| PPC | Posterior Predictive Check |
| PPD | Posterior Predictive Distribution |
| RK4 | Runge-Kutta 4 |
| TP | True Idealized Pendulum Process |

# Appendix B

# Experiment 1 (LP) Diagnostics

## B.1 STAN Fit Summaries from Experiment 1 (LP)

Table B.1: Summary statistics of one run of model $M2_{WDf}$ with simulated data from LP ($N = 100000$ HMC iterations) reveal that $N_{eff}$ (ESS) is large, indicating numerous uncorrelated samples and suggesting convergence ($Rhat = 1$).

| | mean | $se_{mean}$ | sd | 2.5% | 25% | 50% | 75% | 97.5% | $N_{eff}$ | Rhat |
|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | 1.15 | 0.00 | 0.11 | 0.93 | 1.08 | 1.16 | 1.23 | 1.37 | 17023 | 1 |
| $\sigma$ | 1.38 | 0.00 | 0.49 | 0.65 | 1.02 | 1.30 | 1.65 | 2.55 | 18169 | 1 |
| $\sigma_u$ | 0.03 | 0.00 | 0.01 | 0.02 | 0.02 | 0.03 | 0.03 | 0.04 | 19276 | 1 |
| $\sigma_f$ | 0.03 | 0.00 | 0.01 | 0.02 | 0.03 | 0.03 | 0.04 | 0.05 | 20590 | 1 |
| $R$ | 0.20 | 0.00 | 0.00 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 22129 | 1 |
| lp | 74.22 | 0.01 | 1.64 | 70.11 | 73.39 | 74.55 | 75.41 | 76.37 | 12507 | 1 |

Table B.2: Summary statistics of one run of model $M2_{WDf}$ with simulated data from LP ($N = 100000$ HMC iterations) reveal that $N_{eff}$ (ESS) is large, indicating numerous uncorrelated samples and suggesting convergence ($Rhat = 1$). Additionally, the non-zero value of the discrepancy hyperparameter $\sigma_\delta$ indicates that the model has captured some of the non-existent discrepancy between the model and the process. However, the distribution for $\sigma_\delta$ is wide and uncertain due to its nonnegative constraint. This contributes to the extensive width of the posterior distribution of $R$ for $M2_{WDf}$.

| | Mean | $se_{mean}$ | sd | 2.5% | 25% | 50% | 75% | 97.5% | $N_{eff}$ | Rhat |
|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | 1.15 | 0.00 | 0.11 | 0.93 | 1.07 | 1.15 | 1.23 | 1.36 | 16444 | 1 |
| $\sigma$ | 1.37 | 0.00 | 0.49 | 0.65 | 1.01 | 1.29 | 1.65 | 2.53 | 18036 | 1 |
| $\sigma_u$ | 0.03 | 0.00 | 0.01 | 0.02 | 0.02 | 0.03 | 0.03 | 0.04 | 19045 | 1 |
| $\sigma_\delta$ | 0.11 | 0.00 | 0.13 | 0.01 | 0.04 | 0.07 | 0.14 | 0.46 | 23230 | 1 |
| $l_\delta$ | 1.60 | 0.01 | 0.78 | 0.34 | 1.01 | 1.54 | 2.12 | 3.28 | 21451 | 1 |
| $\sigma_f$ | 0.03 | 0.00 | 0.01 | 0.02 | 0.03 | 0.03 | 0.04 | 0.05 | 21558 | 1 |
| $R$ | 0.20 | 0.00 | 0.00 | 0.19 | 0.20 | 0.20 | 0.20 | 0.21 | 16678 | 1 |
| lp | 70.60 | 0.02 | 2.04 | 65.73 | 69.48 | 70.95 | 72.08 | 73.52 | 9567 | 1 |

## B.2 Trace Plots from Experiment 1 (LP)

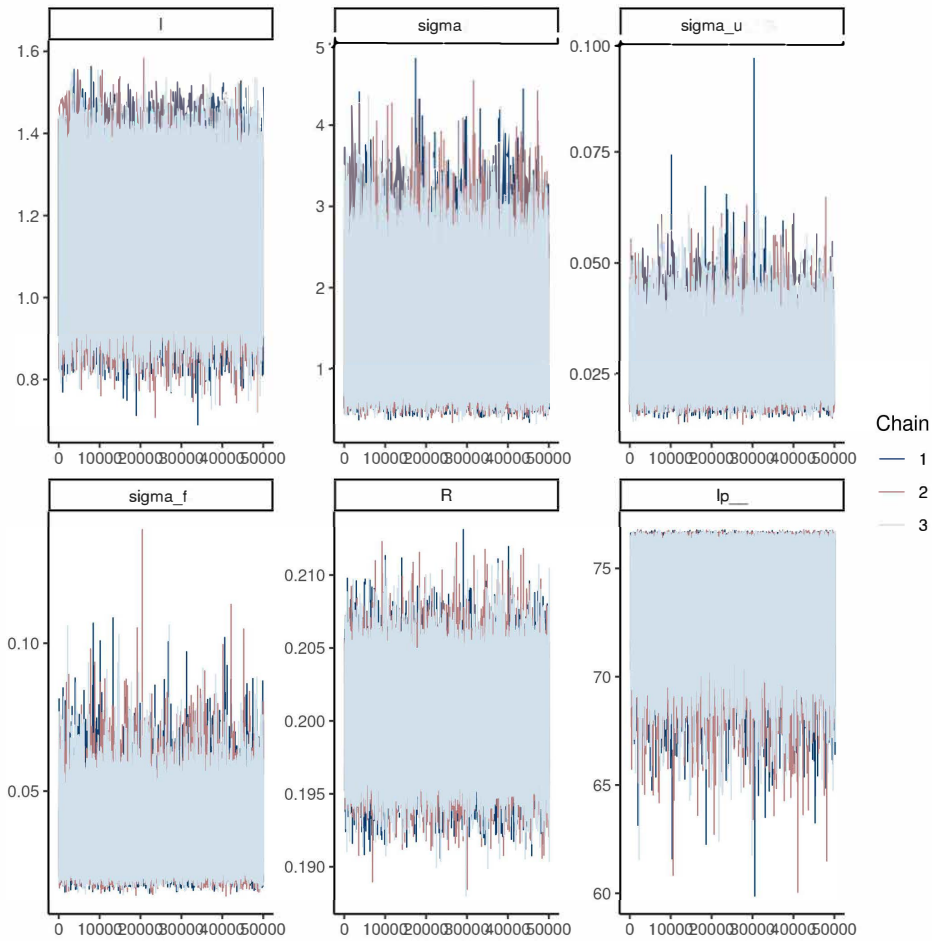



Figure B.1: Trace plots from three chains with 100000 HMC iterations of $M1_{WOD}$ for observations from LP. Note that scales on the $y$-axis vary for each parameter. The observed mixing pattern, resembling a belt, indicates convergence of the MCMC chains (E1, (LP)).

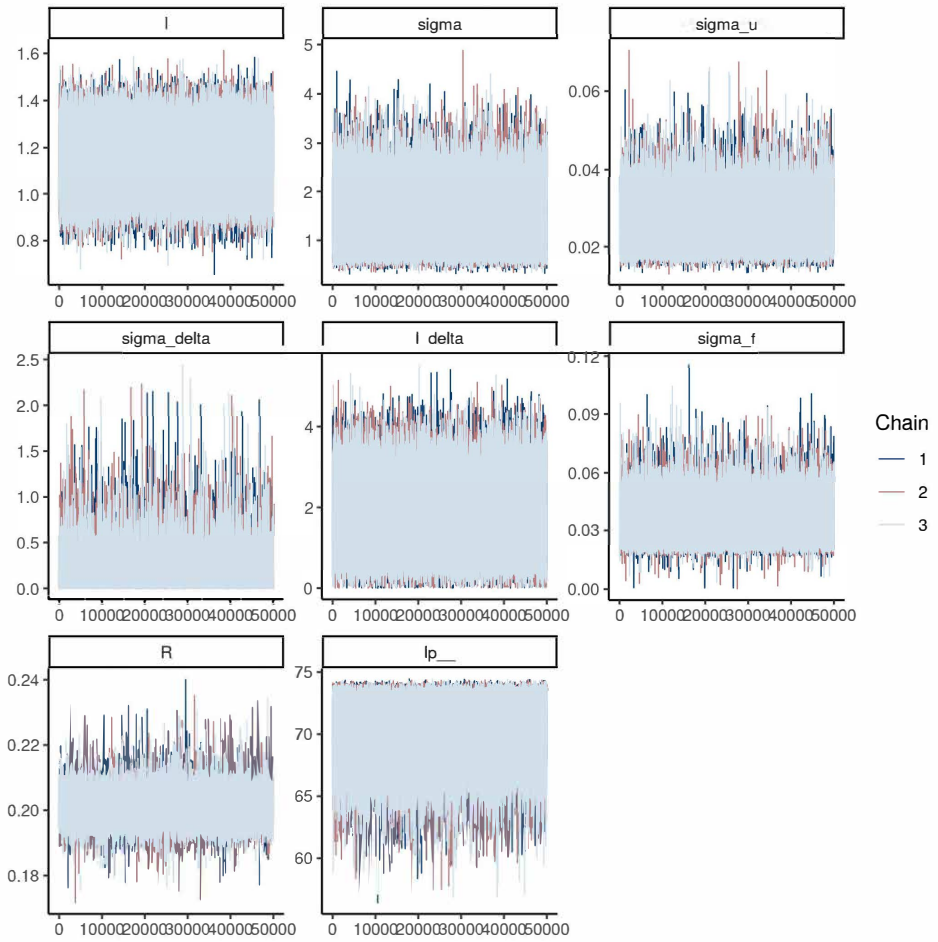
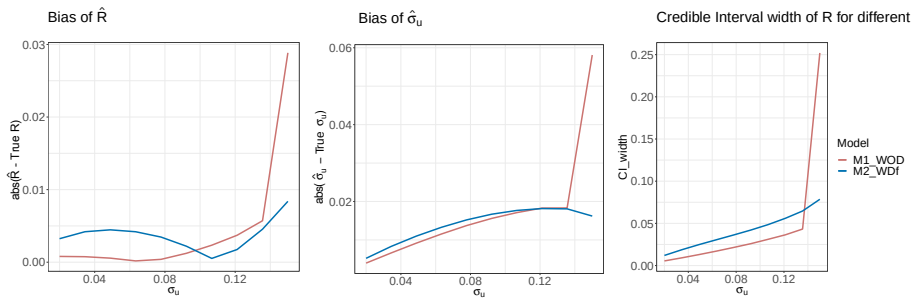



Figure B.2: Trace plots from three chains with 100000 HMC iterations of $M2_{WDf}$ for observations from LP. Note that scales on the $y$-axis vary for each parameter. The observed mixing pattern, resembling a belt, indicates convergence of the MCMC chains (E1, (LP)).

# Appendix C

# Additional Results from Experiment 2

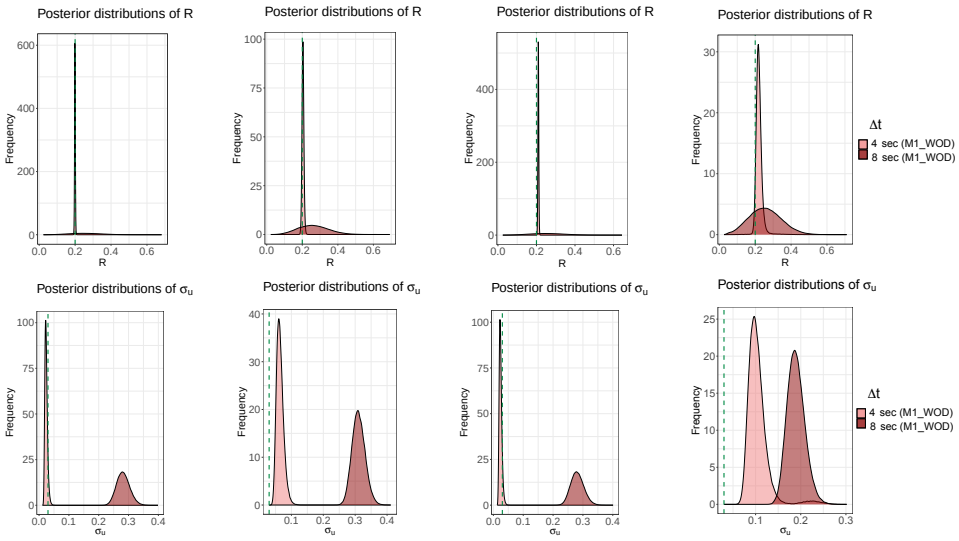
## C.1 Bias and Credible Interval Width for E2 (LP)




(a) Bias of $\hat{R}$ (b) Bias of $\hat{\sigma}_u$ (c) CI width of $\hat{R}$

Figure C.1: Absolute value of bias of $\hat{R}$, absolute value of bias of $\hat{\sigma}_u$ and with of credible interval for LP for different levels of measurement noise, determined by $\sigma_u$. This is the results from 300000 HMC iterations.

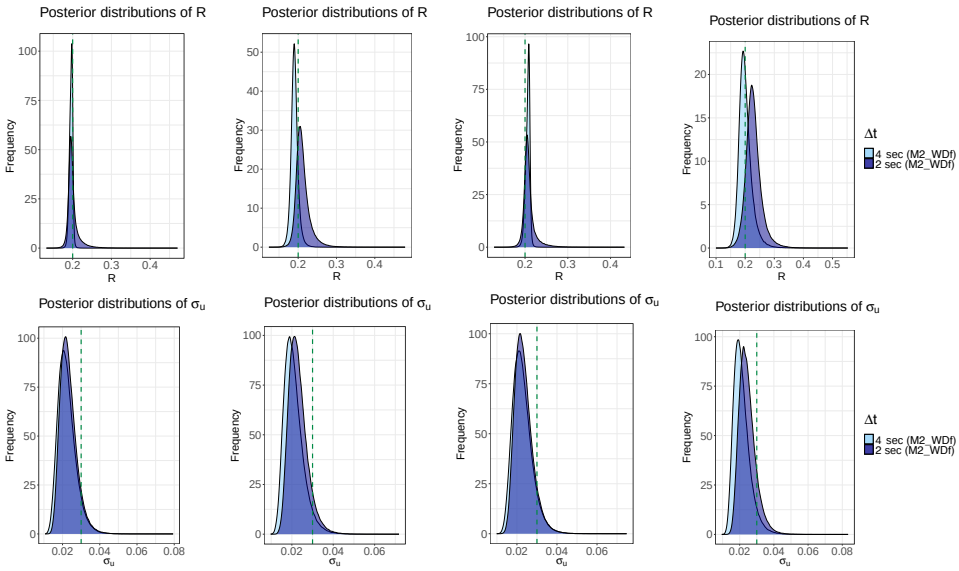
## C.2 Doubled Number of Cycles, Halved Measurement Density for $\mathbf{M1_{WOD}}$




(a) LP (b) LPGP (c) TP (d) DP

Figure C.2: Same number of measurements and number of iterations, $N_u = N_f = 20$, but with $\Delta t = 4$(light red) and $\Delta t = 8$(dark red) for $M1_{WOD}$.

## C.3 Halved Number of Cycles, Doubled Measurement Density for $\mathbf{M2_{WDf}}$




(a) LP (b) LPGP (c) TP (d) DP

Figure C.3: Same number of measurements $N_u = N_f = 20$ and number of iterations, but with $\Delta t = 4$(light blue) and $\Delta t = 2$(dark blue).

## C.4 Trace Plots for LPGP (E2b)

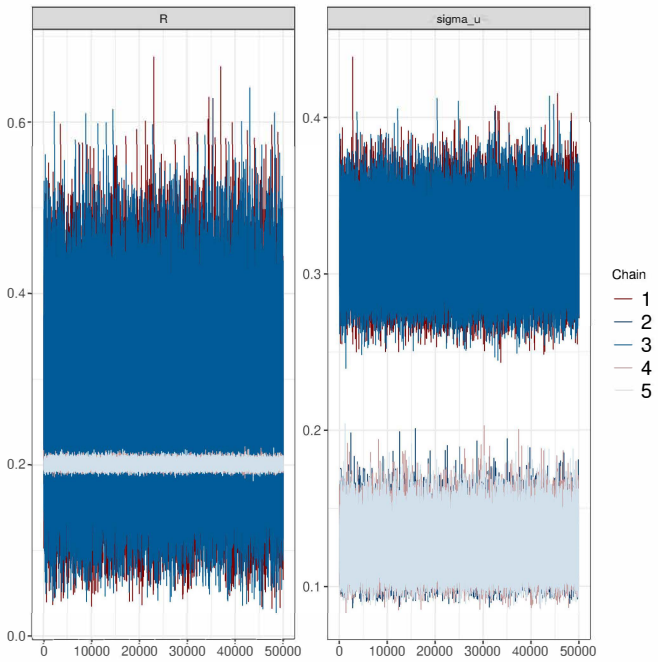
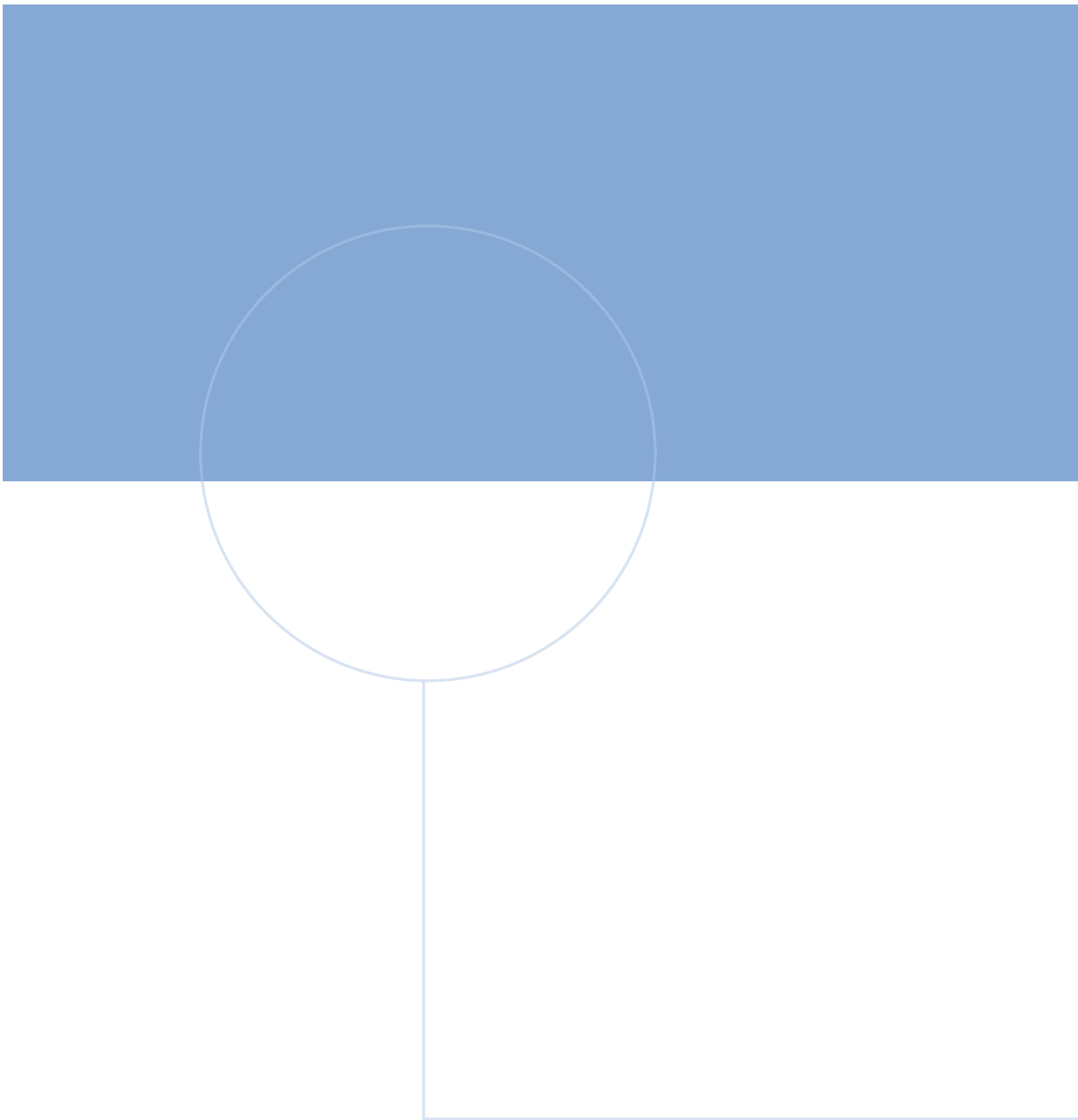



Figure C.4: Trace plot for $M1_{WOD}$ with observations from LPGP in E2b, showing non-mixing chains, showing a bimodality in the posteriors, suggesting mixing problems.



NTNU

Norwegian University of
Science and Technology