Helene Minge Olsen

Bayesian Calibration for Modelling the Cardiovascular System Using the Two Element Windkessel Model

Master's thesis in Applied Physics and Mathematics
Supervisor: Ingelin Steinsland
Co-supervisor: Michail Spitieris
March 2022

Master's thesis

NTNU
Norwegian University of
Science and Technology

Helene Minge Olsen

Bayesian Calibration for Modelling the Cardiovascular System Using the Two Element Windkessel Model

NTNU

Norwegian University of
Science and Technology

*"I think it's much more interesting to live not knowing
than to have answers which might be wrong.
I have approximate answers and possible beliefs and different degrees of
uncertainty about different things,
but I am not absolutely sure of anything
and there are many things I don't know anything about,
such as whether it means anything to ask why we're here.
I don't have to know an answer.
I don't feel frightened not knowing things,
by being lost in a mysterious universe without any purpose,
which is the way it really is as far as I can tell."*

RICHARD P. FEYNMAN

Abstract

This thesis explores the properties of Windkessel Models based on synthetic simulation studies and is a part of the cross-disciplinary project "My Medical Digital Twin" at NTNU. The Windkessel models estimate the global arterial properties by relating measured (aortic) pressure and flow through linear differential equations. The aim is to make inferences about the physically interpretable parameters, namely, total vascular resistance and arterial compliance.

Inference is performed using an adaptive MCMC strategy. This allows us to successfully omit the use of an emulator for the computer model and is one of the main contributions of our work. As the second main contribution, we investigate the effect of dependent and independent noise in simulated pressure observations on the uncertainty and accuracy of parameter estimates. This is, to our knowledge, never done for the Windkessel models, nor other models based on linear differential equations.

We perform inference on models not accounting for the discrepancy, which confirms that not accounting for the model discrepancy leads to biased and unstable parameter estimates. Therefore, the inverse problem is set in a Bayesian calibration framework, where a Gaussian process models the discrepancy. Then, based on synthetically derived pressure observations, the inverse problem is solved, accounting for and quantifying the uncertainty in the model formulation and the parameter estimates.

Accounting for discrepancy allows us to recreate the true physical parameters successfully. Furthermore, this proposed framework yields a run time of estimating physical parameters for one person's pressure and flow observations to mere seconds. We also demonstrate the challenge of incorporating model discrepancy in confounding issues with physical parameters. This is solved through informative priors for observation noise and discrepancy parameters, yielding the most robust parameter estimates when subject to observations with time-dependent noise.

Sammendrag

Denne oppgaven utforsker egenskapene til Windkessel-modellene basert på syntetiske simuleringsstudier, og er en del av det tverrfaglige prosjektet "My Medical Digital Twin" ved NTNU. Windkessel-modellene estimerer de globale arterielle egenskapene ved å relatere målt (aorta) trykk og strømning gjennom lineære differensialligninger. Målet er å trekke slutninger om de fysisk tolkbare parameterne, nemlig total vaskulær motstand og arteriell etterlevelse.

Inferens utføres ved å bruke en adaptiv MCMC-strategi. Dette lar oss utelate bruken av en emulator for data-modellen og er et av hovedbidragene til dette arbeidet. Som det andre hovedbidraget undersøker vi effekten av avhengig og uavhengig støy i simulerte trykkobservasjoner på usikkerheten og nøyaktigheten til parameterestimater. Dette er, så vidt vi vet, aldri gjort for Windkessel-modellene, og heller ikke for andre modeller basert på lineære differensialligninger.

En enkel simuleringsstudie bekrefter at det å ikke ta hensyn til modellavviket fører til uriktige og ustabile parameterestimater. Derfor er det inverse problemet satt i et Bayesiansk kalibreringsrammeverk, der avviket er modellert av en Gauss-prosess. Basert på syntetisk utledede trykkobservasjoner løses det inverse problemet ved å redegjøre for og kvantifisere usikkerheten i modellformuleringen og parameterestimatene.

Å estimere avviket lar oss gjenskape de sanne fysiske parameterne på en vellykket måte. Videre gir dette foreslåtte rammeverket en kjøretid for å estimere fysiske parametere for én persons trykk- og strømningsobservasjoner på bare sekunder. Vi demonstrerer også utfordringen med å inkorporere modellavvik i identifiserings-problemer med fysiske parametere. Dette løses gjennom informative priorer for observasjonsstøy og avviksparametere, som også gir de mest robuste parameterestimatene ved observasjoner med tidsavhengig støy.

Preface

I think every master's student feels a bit overwhelmed by the responsibility and amount of work needed to be able to produce a master's thesis. The work is lonely, but I also believe all students have been lonely during the last two years of studying amid a pandemic.

One thing I am bringing with me from writing a master's is to trust the process. Going from complete chaos, turning into the first draft, becoming the quantasilliontht draft, from "I have all the time in the world" to "I am lucky if I can finish this before the deadline," into finally pressing that "submit" button and realize you've done it.

Two people definitely deserve praise for bringing me to this point: my supervisors, Ingelin Steinsland and Michial Spitieris. Thank you, Ingelin, for helping me put words to the mess in my head and structure, plan and prioritize simulation studies. Thank you, Michail, for guiding me through the world of Bayesian calibration and finding a bug or five in my code.

I also want to thank my family, my best friend and my boyfriend for always supporting and cheering for me, and of course; my coffee machine.

I hope you enjoy your reading!

Contents

Chapter 1

Introduction

Hypertension is a significant cause of premature death and is often referred to as "the silent killer" [WHO [1]]. This is because most people with hypertension are unaware of the problem, as it may have no warning signs or symptoms. Therefore, it is essential to measure blood pressure regularly and distribute information on preventing it from increasing.

"My Medical Digital Twin" is one of NTNU's Digital Transformation Projects. The Digital Twin is based on a broad combination of research areas, with the goal to develop a personalized twin for patient-specific hypertensive intervention [MyMDT [2]]. As every human is different and responds differently to treatment, patient-specific intervention with monitoring and predicting through a digital twin can contribute to changing the way we treat lifestyle diseases in the future.

As a part of the MyMDT project, this thesis analyzes the Windkessel Models commonly used to describe the cardiovascular system. The Windkessel models relate blood pressure $P(t)$ and blood flow $Q(t)$ in the aorta through linear differential equations and are referred to as lumped models [Westerhof, Lankhaar, and Westerhof [3]]. Models are a simplification of reality, and lumped models are a family of mathematical models for the load to the heart, in which the physics of the entire systemic arterial tree is represented by a few lumped parameters [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]]. By observing these parameters, we can describe an individual's medical condition, which in turn can provide information about possible treatment or response to treatment. Therefore, the main goal of this thesis is to obtain estimates of the physical parameters of the Windkessel models, quantify their

uncertainty and investigate how the estimates behave under different conditions related to noise.

The measurement of pulse pressure is recognized as an important factor in the prediction of cardiovascular mortality [Westerhof, Lankhaar, and Westerhof [3]]. We want to obtain a model that we know creates robust and unbiased parameter estimates. We can then observe how the digital twin's parameters change when subject to different medication. The models we consider are the two and three-element Windkessel models. The two-element Windkessel describes the hemodynamics of the arterial system in terms of two physically interpretable parameters; resistance R and compliance C. It can represent aortic pressure decay in diastole but falls short in systole where blood pressure reaches its maximum value [Westerhof, Lankhaar, and Westerhof [3]]. The terms diastole and systole refer to when the heart muscles relax and contract respectively [Silva [5]]. As the goal is to model the whole cardiac cycle, it is crucial to mend for the two-element Windkessel models' shortcomings in systole.

The three-element Windkessel model contains a third parameter interpreted as a tuning parameter. This makes the model capable of generating pressure wave profiles that closely resemble pressure waves measured in the arterial tree [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]]. One could argue for using the more complex model directly without any fuss. But, even though more complex models might yield an improved fit to data, they often bring identifiability issues of the parameters and/or loss of physical interpretability. As presented in [Spitieris, Steinsland, and Ingestrom [6]], the three-element Windkessel model fits observed data better but generally overestimates the total arterial compliance.

We are considering an inverse problem, that is, to learn about the values of parameters within the model from field observations, a process which is often called calibration [Brynjarsdóttir and O'Hagan [7]]. We are estimating the parameters R and C from the observations. In reality, the field observations would be a real person's medical history of flow and pressure measurements. However, these kinds of measurements have not been available for the work of this thesis. More importantly, we wish to keep the experiments within a controlled environment. Therefore we simulate noisy data from the three-element Windkessel model to obtain field observations. We know that the two-element Windkessel model is wrong, but by using

Bayesian calibration to get an accurate estimate of exactly how wrong the model is, we can produce robust inference of the parameters R and C.

In a digital twin setting, individual sensor data can be noisy. Therefore, part of our method's analysis revolves around investigating how noise in field observations affects the parameter estimates. As we do not live in a perfect world, the common assumption of the field observations containing independent and identically distributed noise does not hold in all cases. Since the future digital twin operates with real-time sensor acquired data from an individual, the noise is prone to dependencies over time, for example due to movement. Consequently, we investigate how the calibration procedure responds to simulation studies using synthetic field observations containing both independent and dependent noise. Furthermore, we only assume noise in simulated pressure observations, not in flow Q, which is unrealistic, but a common assumption.

Furthermore, to make appropriate use of observations of the physical system, it is vital to recognize model discrepancy, i.e., the difference between the simulation model and the computer model output. Hence, our primary focus lies in investigating the effect of model misspecifications and biased errors and seeing if Bayesian calibration can solve the issues.

The method is inspired by the general statistical formulation presented in [Higdon, Kennedy, Cavendish, Cafeo, and Ryne [8]] and Bayesian calibration of Windkessel models made by [Spitieris, Steinsland, and Ingestrom [6]]. [Spitieris, Steinsland, and Ingestrom [6]] outline a statistical approach for combining noisy field observations with the imperfect two-element Windkessel model to calibrate parameters and to characterize uncertainty in parameter estimates. By taking a Bayesian approach that closely follows that of [Kennedy and O'Hagan [9]], they can explicitly model uncertainty in model inputs and the discrepancy between the computer model simulator and the actual physical system. Furthermore, in line with suggestions of [Kennedy and O'Hagan [9]] (KOH), they incorporate underlying knowledge of the calibration parameters in the form of prior distributions based on prior knowledge about the processes, whereas a flexible Gaussian process model the discrepancy prior [Spitieris, Steinsland, and Ingestrom [6]]. The advantage of doing this in a Bayesian fashion is that the accuracy estimates of all predictions and parameter estimates are available, accounting for all of the modeled uncertainties [Berger and Smith [10]]. In addition, the Bayesian methodology produces input-dependent optimal values for the calibration

parameters, as well as characterizes the associated uncertainties via posterior distributions [Karagiannis, Konomi, and Lin [11]].

In many cases, the computer model is computationally expensive. To mend for this, [Spitieris, Steinsland, and Ingestrom [6]] impose in line with suggestions by KOH to apply a Gaussian process(GP) model as emulator(surrogate) for the computer model. KOH suggests inferring all unknowns for both emulator and calibration parameters simultaneously, using Markov Chain Monte Carlo (MCMC). [Gramacy [12], however, showed that fully Bayesian KOH calibration yields emulator model fits that can be unfaithful to computer model simulations, being biased by field data. Therefore [Spitieris, Steinsland, and Ingestrom [6]] follow a modularized approach proposed by [Bayarri, Berger, Paulo, Sacks, Cafeo, Cavendish, Lin, and Tu [13]]. This suggests estimating the parameters in two stages. First, they fit and estimate the hyperparameters of the emulator GP using only computer model simulations. These values are then fixed when continuing with the second stage of obtaining the posteriors for the parameters of interest.

Modularization helps because it limits flexibility somewhat through a more constrained prior, allowing only computer model runs to influence emulator fits. [Bayarri, Berger, Paulo, Sacks, Cafeo, Cavendish, Lin, and Tu [13]] showed that this can improve identifiability and mixing of the MCMC algorithm. Nevertheless, confounding issues are a concern, and by approximating the computer model with an emulator, we inarguably introduce more uncertainty into the estimation procedure [Gramacy [12]]. Furthermore, according to [Spitieris, Steinsland, and Ingestrom [6]], computer model outputs are usually functional, and specifically, the Windkessel models produce time-series outputs. This creates a computational challenge for the Bayesian calibration approach with the emulator due to the quadratic complexity of the GP models. In addition, mixing properties of the estimation procedure is slow when using an emulator and depends on the values of the real parameters.

As a result of arising issues from using an emulator, it becomes attractive to find computationally feasible alternatives. This is where adaptive MCMC enters the picture. The adaptive MCMC performs a Random Walk (RW) Metropolis-Hastings Algorithm, with a tuning of the proposal variance as the chain is running, in line with [Shaby and Wells [14]]. The method is made available through the R package NIMBLE [*NimbleUserManual* [15]],

and is an extension of the probabilistic programming language BUGS. One of its most significant advantages lies in its compiling scheme being implemented in C++. As a result of both NIMBLE's adaptive RW sampler and its compiling scheme, the simulation run-times and the burn-in phase of the chain are drastically reduced. In this thesis, we utilize this benefit of the adaptive MCMC scheme and can therefore deviate from previous work in [Spitieris, Steinsland, and Ingestrom [6]] by omitting the emulator and operating with the actual computer model itself.

There are two main contributions of this thesis. The first contribution is setting up and demonstrating the procedure of performing Bayesian calibration on the Windkessel models without fitting an emulator for the computer model. Secondly, we explore the properties of this approach with dependent and independent noise. This is, to our knowledge, not done before, neither for the Windkessel models nor for other models based on linear differential equations.

Hence, the simulation studies are conducted to answer the following research questions:

- Research Question 1: How does independent and dependent noise in field observations affect the inference of parameters?

- Research Question 2: Will performing Bayesian calibration without fitting an emulator yield well calibrated estimates from a misspecified computer model?

As for the rest of this thesis, Chapter 2 gives background information on the Windkessel Models, Bayesian Statistics, adaptive MCMC, and a brief introduction to Gaussian Processes. Next, we present the general framework and methodology in Chapter 2.3, which is tied to the Windkessel models in Chapter 3 where we introduce the simulation studies. Finally, Chapter 4 presents and discusses resulting parameter estimates with associated uncertainties.

Chapter 2

Background

The Windkessel models are described in Chapter 2.1. Bayesian inference and calibration are introduced in Sections 2.2 and 2.3. Inference is performed using Markov Chain Monte Carlo (MCMC) methods, introduced in Section 2.4. We use adaptive MCMC and the package NIMBLE by de Valpine [16] et.al. This is presented in Section 2.4.1. Finally, Gaussian Processes are briefly introduced in Section 2.5.

2.1   The Windkessel Models For Modelling the Cardiac Cycle

The cardiac cycle describes the physiological events associated with one single heartbeat and is illustrated in Figure 2.1. It is essentially split into systole (the contraction phase) and diastole (the relaxation phase). Each of these is then further divided into an atrial and ventricular component [Silva [5]. Following the grey area in Figure 2.1 named the ventricular ejection, we have the systole, where the ventricles, i.e., the two lower chambers of the heart, contract and eject blood out to the body through the aorta. As a result of the ventricles contracting, the aortic pressure rises and reaches its maximum at the systolic pressure. As the ventricles relax and allow fresh, oxygenated blood to fill the chambers, the pressure gradually drops and reaches the diastolic pressure [Klabunde [17]]. This is when pressure is at its lowest and occurs immediately before the ventricles again contract and eject blood into the aorta.

Together, diastolic and systolic pressure measurements make up the medical judgment of a patient's blood pressure. The current blood pressure categories, according to the guidelines of the European Society of
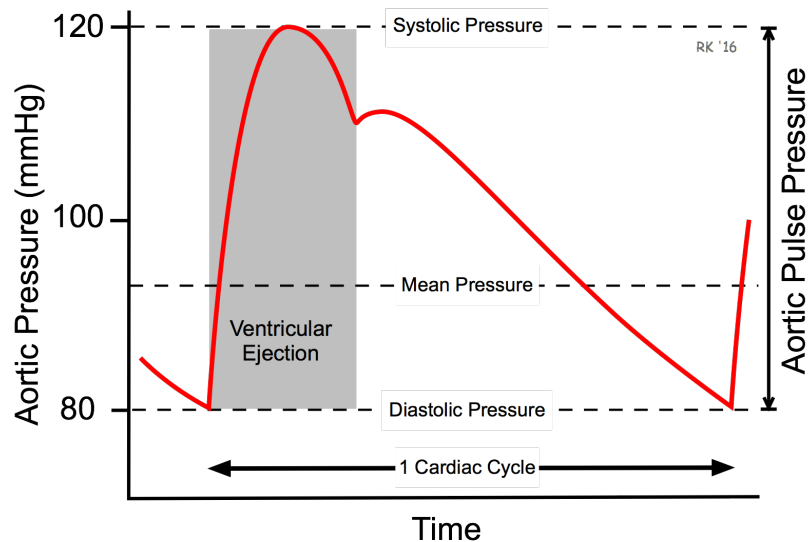
Figure 2.1: Arterial blood pressure over one Cardiac Cycle. Maximum point of the curve represents the Systolic pressure, whereas the minima the Diastolic pressyre. Image from Klabunde [17].

Category	Systolic (mmHg)		Diastolic (mmHg)
Optimal	<120	and	<80
Normal	120–129	and/or	80–84
High normal	130–139	and/or	85–89
Grade 1 hypertension	140–159	and/or	90–99
Grade 2 hypertension	160–179	and/or	100–109
Grade 3 hypertension	≥180	and/or	≥110
Isolated systolic hypertension[b]	≥140	and	<90

BP, blood pressure.
[a]BP category is defined according to seated clinic BP and by the highest level of BP, whether systolic or diastolic.
[b]Isolated systolic hypertension is graded 1, 2, or 3 according to systolic BP values in the ranges indicated. The same classification is used for all ages from 16 years.

Figure 2.2: The current blood pressure categories according to the guidelines of the European Society of Cardiology and the European Society of Hypertension, last updated in 2018. Image from [Williams et.al. [18]].

Cardiology and the European Society of Hypertension (ESC/ESH) updated 2018, are presented in Figure 2.2 [Williams et.al. [18]].

It is still an open question which model reflects the cardiac cycle the best and leads to the best estimates of arterial properties, but the Windkessel models prove as an eligible candidate [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]]. The Windkessel models relate blood pressure $P(t)$ and blood flow $Q(t)$ in the aorta and are so-called lumped models. Lumped models refer to a family of mathematical models for the load to the heart, in which the physics of the

entire systemic arterial tree is represented by a few lumped parameters [NTNU [19]]. These parameters are relevant in the context where we do not have access to a more fine-tuned model and can be used to monitor the effect of medications used to treat cardiovascular conditions such as hypertension.

$$wk_2 : Q(t) = \frac{1}{R}P(t) + C\frac{dP(t)}{dt} \tag{2.1}$$

$$wk_3 : Q(t)\left(1 + \frac{Z_c}{R}\right) + Z_c C\frac{dQ(t)}{dt} = \frac{1}{R}P(t) + C\frac{dP(t)}{dt} \tag{2.2}$$

Equation 2.1 presents the two-element Windkessel model, which describes the arterial system through two physically interpretable parameters: total peripheral resistance R and total arterial compliance C. We define flow $Q(t)$ as a function of time, and $P(t)$ represents the pressure. Total peripheral resistance R is often referred to as total vascular resistance and represents the resistance generated by the small arteries and arterioles. The compliance element C mimics the elastic properties and the buffering capacity of the large arteries [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4].

With the addition of a third parameter, aortic characteristic impedance Z_c, we have the three-element model, shown in Equation 2.2. Aortic characteristic impedance is interpreted as a tuning parameter and does not hold the same physical interpretation as R and C [Spitieris, Steinsland, and Ingestrom [6]]. The addition of characteristic aortic impedance can be seen as a link between the lumped Windkessel model and wave travel aspects of the arterial system [Westerhof, Lankhaar, and Westerhof [3]]. As a result, the three-element Windkessel model is capable of generating pressure wave profiles that closely resemble pressure waves measured in the arterial tree [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]].

The left part Figure 2.3 illustrates measured flow Q used as input to the Windkessel Models, which gives the pressure curve. Simulated pressure curves obtained from the two and three Windkessel models with fixed parameters, $R = 0.9$, $C = 1.3$, and $Z = 0.1$, are presented in the right panel of Figure 2.3. Compared to the real, measured pressure curve represented in Figure 2.1 we observe how the three-element Windkessel indeed mimics the pressure wave profile well. The two-element Windkessel, however, is
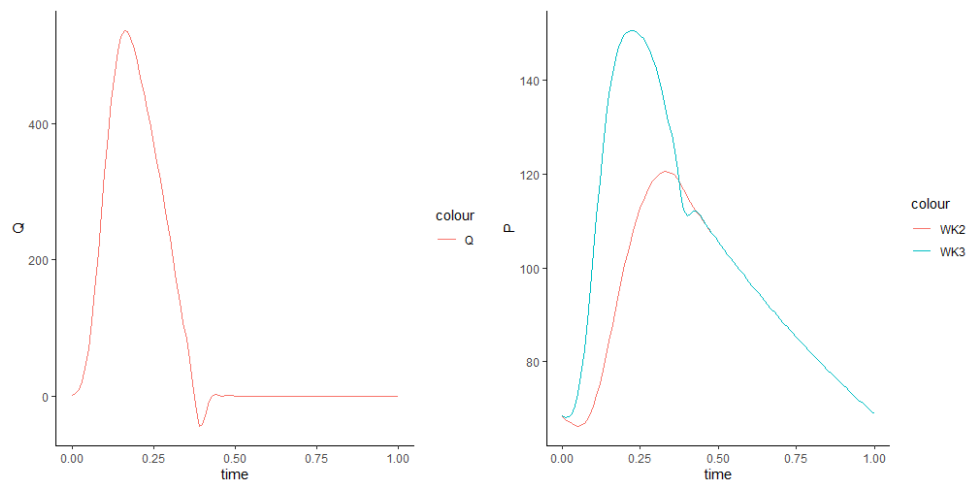
Figure 2.3: Left: Flow Q used as input for the WK models. Right: Estimated pressure curves from the two (red) and three (blue) element Windkessel model using flow Q in left plot and fixed parameter values $R = 0.9$, $C = 1.3$, and $Z = 0.1$.
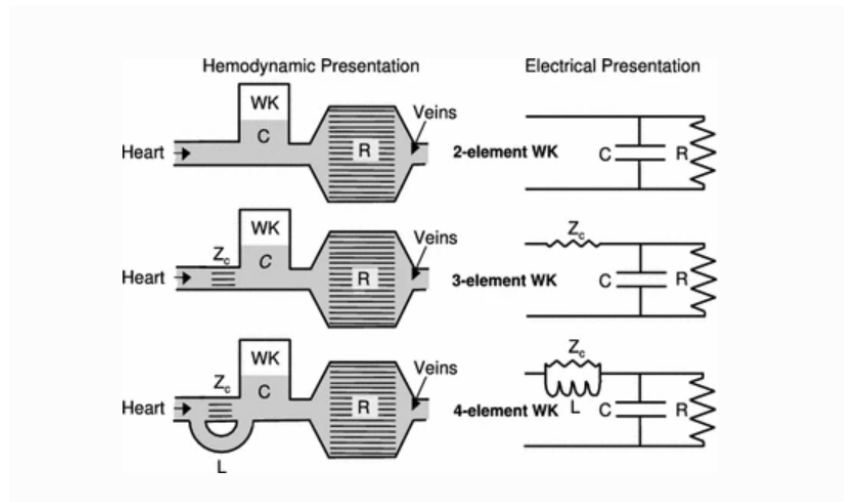

Figure 2.4: The Windkessel effect in both a hemodynamic and the equivalent electrical setting of the Windkessel models. Image from [Westerhof, Lankhaar, and Westerhof [3]].

capable of producing an exponential decay in diastolic pressure but does, as presented in [Westerhof, Lankhaar, and Westerhof [3]], fall short in systole.

It is common to present the Windkessel models as an electrical analogue, demonstrated in Figure 2.4 [Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]. Here we observe the WK effect in both a hemodynamic and the equivalent electrical analog. The original term 'Windkessel' means 'air chamber' in German. In the hemodynamic representation in Figure 2.4 one can observe blood running through the veins. The air reservoir marked 'WK' behaves as the WK effect. While blood is pumped through the veins, the air in the chamber is

compressed, pushing the blood out of the chamber. As a result of age or genetics, our veins become less compliant or elastic. This corresponds to a reduction in the size of the air reservoir. Less area to move yields larger flow; hence the pressure increases.

The electrical presentation in Figure 2.4 provides a more clear picture of the roles the different parameters play. Here $P(t)$ is the voltage that drives the circuit, and $Q(t)$ represents the current. Total arterial compliance C is represented as a conductor, which stores the energy, and the resistive element R resists the current. For the two-element Windkessel model, they are placed in parallel. When a second resistive element is placed in series with the RC model, namely the characteristic impedance Z_c, we have the three-element Windkessel model. As a result, the ratio of mean pressure and mean flow i.e., systemic vascular resistance, can be assumed to be $Z_c + R$ and not just R. Furthermore, the placement of Z_c underlines its role as a tuning parameter. The impedance will impact the blood flow arriving at the part identical to the two-element Windkessel model, introducing an amplitude loss and a potential flow shift (current).

To further demonstrate the impact the parameters have on the simulated pressure waveform, we observe Figure 2.5 for the two-element Windkessel and Figure 2.6 for the three-element Windkessel. Here we have calculated the pressure curve from the Windkessel models using the flow in Figure 2.3 and changing one parameter while keeping the others fixed. We commence with the two-element Windkessel model in Figure 2.5. To the left, we observe that the C parameter controls the amplitude of the pressure curve. However, the right plot illustrates the impact on the pressure curve when varying R. Here the shape of the pressure curve remains the same, but its intersection with the pressure axis increases intact with R. Hence R controls the diastole and C the systole.

Moving on to the three-element Windkessel in Figure 2.6, we observe the same behavior for R in the top right plot; however, the curve itself has a different shape. This is all due to the impedance Z, illustrated in the two bottom plots of 2.6. Here we are varying C with either $Z = 0.1$ or $Z = 0.01$. For the latter, we observe that small Z values yield pressure curves identical to the two-element model in Figure 2.5, whereas where $Z = 0.1$ we indeed obtain the characteristic "bump" in diastole as in Figure 2.1. In addition, increasing Z yields higher pressure, whereas increasing C yields lower pressure. As a result, we can conclude that both C and Z control the
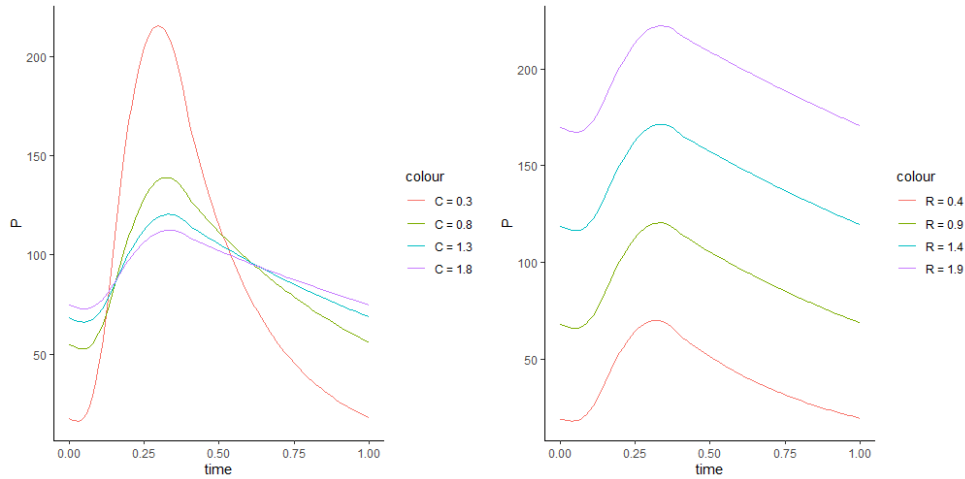
Figure 2.5: The output pressure curve from running the two element Windkessel Model for fixed $R = 0.9$ and different values of C (left), and fixed $C = 1.3$ and different values of R (right)

amplitude of the pressure curve in systole.

2.2 Bayesian Inference

We give a general introduction to Bayesian statistics and inference before diving into Bayesian calibration. The section is based on [van de Schoot, Depaoli, King, Kramer, Märtens, Tadesse, Vannucci, Gelman, Veen, Willemsen, and Yau [20]], [Robert [21]] and the master thesis of [Lam [22]].

Bayesian inference yields a probabilistic approach for inverse problems, i.e., the process of calculating the parameters θ by studying the causal factors behind them, like observations from a model. Bayes rule updates prior belief of an event with information extracted from data to make inferences. This is both done to estimate parameters and for prediction.

Parameter Estimation

Consider data points y, with y being a vector of field observations and θ a vector of parameters. Bayesian inference is interested in estimating the parameters θ given data y, which is performed through Bayes' theorem :

$$\pi(\theta|y) = \frac{L(y|\theta) \times p(\theta)}{p(y)} \propto L(y|\theta) \times p(\theta) \tag{2.3}$$

The posterior $\pi(\theta|y)$ describes our knowledge and uncertainty in θ given data y. The posterior can be used to find the point estimate of θ such as
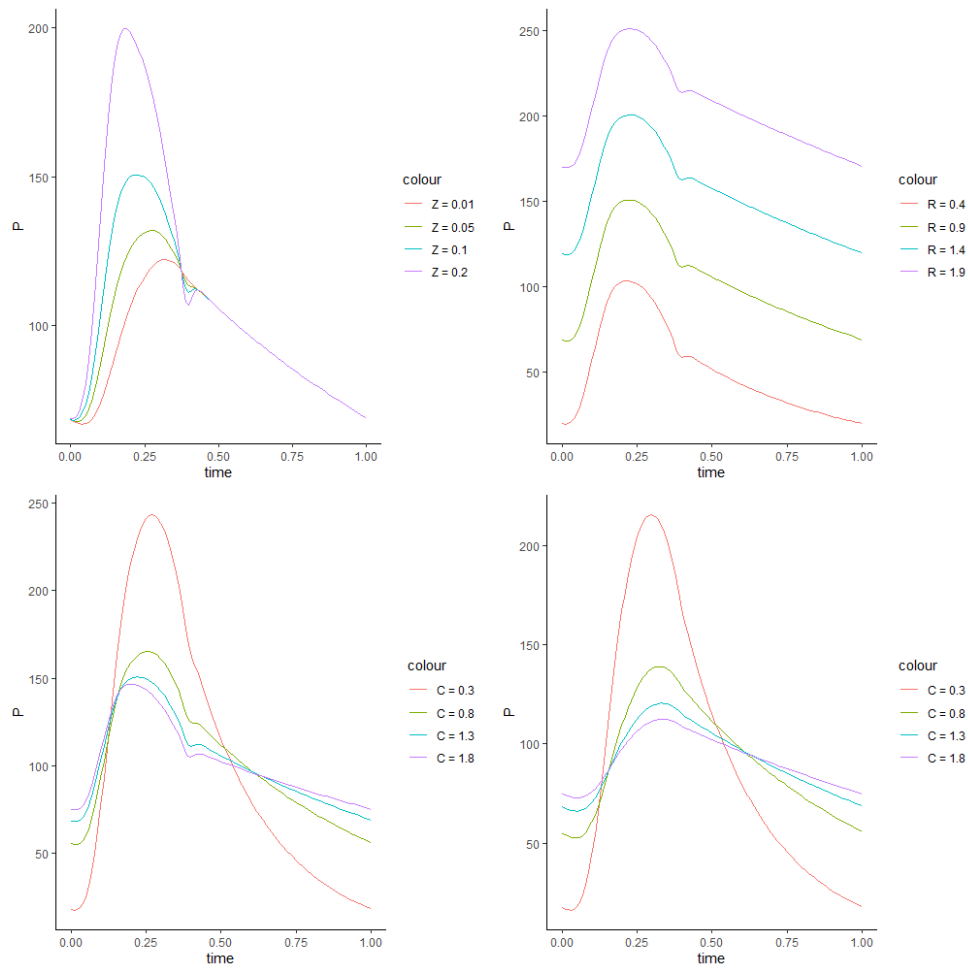
Figure 2.6: The output pressure curve from running the three element Windkessel Model with all but one parameter fixed. Top right plot show varied R and fixed $C = 1.3$ and $Z = 0.1$. The top left plot vary Z with $R = 0.9$, and $C = 1.3$. Bottom plots vary C with $R = 0.9$, and $Z = 0.1$ (left) and $Z = 0.01$(right).

posterior mean or its maximum a posteriori (MAP) value and characterize the associated uncertainties. As the posterior is generally not analytically tractable, we resort to sampling approaches to draw samples from an approximate distribution.

The prior distribution $p(\theta)$ represents prior information about the parameters and is specified mainly from prior data, literature, or expert opinion. The priors help keep inference well defined; for example, if the physical interpretation of the parameter limits it to take a specific range of values, e.g., be strictly non-negative or can be chosen to be vague.

The likelihood-function $L(y|\theta)$ is the key component that connects model inputs to the measured quantity of interest. It is often determined by the noise model, which describes a disagreement between field observations y and parameters of interest.

$p(y)$ in the denominator is considered a normalizing constant and is dropped from the equation as it does not depend on θ. Hence, the posterior becomes proportional to the product of the likelihood function and prior distribution. As a result, the posterior now contains both prior knowledge of what was thought to be the distribution of θ before observing the data, and information of the data itself. This represents the core of Bayesian statistics and how it is reflected in Bayes' theorem, i.e., merging prior knowledge and information from data to make inference. If not much data is available, the posterior will be strongly influenced by the prior. On the other side, data will dominate the influence of the posterior when there is too much available. It is, therefore, crucial to find the right balance between the two when one performs inference.

Prediction

To predict a new data point $\tilde{y}$, Bayesian statistics uses the posterior predictive distribution which has the following form:

$$\pi(\tilde{y}|y) = \int L(\tilde{y}|\theta)\pi(\theta|y)d\theta$$

Here we obtain the posterior predictive distribution $\pi(\tilde{y}|y)$ used to predict $\tilde{y}$ through the posterior distribution $\pi(\theta|y)$ found in Equation 2.3 and the likelihood function $L(\tilde{y}|\theta)$. As we integrate over the domain of θ, we are accounting for the uncertainty. Furthermore, since we are using the

posterior distribution $\pi(\theta|y)$ from Equation 2.3, both prior knowledge of θ and information from the data can be found in the posterior predictive distribution.

Credible Intervals (CI)

A credible interval is an interval that contains a parameter with a specified probability and is the Bayesian statistics' equivalent to the frequentist statistics' confidence interval. A $100(1-\alpha)\%$ credible interval is an interval covering $100(1-\alpha)\%$ of the posterior distribution. The bounds of the interval are the upper and lower percentiles of the parameter's posterior distribution and can be expressed for prediction as $(\tilde{y}_L, \tilde{y}_U)$, where

$$P(\tilde{y}_L < \theta < \tilde{y}_U|y) = \int_{\tilde{y}_L}^{\tilde{y}_U} \pi(\tilde{y}|y)d\theta = 1 - \alpha$$

For predictions in this thesis, the mean and the 90 % credible interval are used, meaning it has the upper and lower 5 % percentiles of the posterior distribution as its bounds.

2.3   Bayesian Calibration of Windkessel Models

In this Chapter we introduce Bayesian calibration and set up the statistical formulation of the computer model η accounting for discrepancy. The framework is based upon [Kennedy and O'Hagan [9]], [Bayarri, Berger, Paulo, Sacks, Cafeo, Cavendish, Lin, and Tu [13]] and [Higdon, Kennedy, Cavendish, Cafeo, and Ryne [8]] as well as the previous work of and [Spitieris, Steinsland, and Ingestrom [6]].

For an input vector x, field observations y are made of the real process γ

$$y(x_i) = \gamma(x_i) + \epsilon(x_i), \qquad i = 1, \dots, n \tag{2.4}$$

Where $\epsilon(x_i)$ denotes the observation error .

We want to simulate the real system γ through a computer model $\eta(x, \theta)$, and rewrite equation 2.4 as:

$$y(x_i) = \eta(x_i, \theta) + \epsilon(x_i), \qquad i = 1, \dots, n \tag{2.5}$$

Where $\eta(x,\theta)$ denotes the computer model output given the input (x,θ). The vector x contains observable and controllable inputs, whereas θ represents the un-observable calibration parameters required to run the code of the computer model. The process of fitting a computer model η to field observations y by adjusting parameters is known as calibration. Traditional calibration is typically made by simple fitting algorithms such as least squares(LS). After calibration, the model is used with the fitted parameter values to predict the system's future behavior.

It is common, however, that computer models do not fit the field observations perfectly. This leads to the model being systematically biased, which in turn yields biased calibration parameter estimates [Spitieris, Steinsland, and Ingestrom [6]].

Kennedy and O'Hagain proposed a statistical framework that improves the traditional approach by introducing a systemic bias function that links the computer model η with the real process γ in the following manner:

$$\gamma(x_i) = \eta(x_i,\theta) + \delta(x_i), \qquad i = 1,\ldots,n \qquad (2.6)$$

Here $\delta(x_i)$ accounts for discrepancy between computer model simulator $\eta(x,\theta)$ and reality $\gamma(x_i)$. The most important goal of the discrepancy is identifying flaws in the model, with the hope they can be corrected [Berger and Smith [10]].

We can now rewrite the field observations with a computer model simulator $\eta(x,\theta)$.

$$y(x_i) = \eta(x_i,\theta) + \delta(x_i) + \epsilon(x_i), \qquad i = 1,\ldots,n \qquad (2.7)$$

Kennedy and O'Hagans' approach improves the estimation procedure in two respects [Kennedy and O'Hagan [9]]. First, the predictions allow for all sources of uncertainty, including the remaining uncertainty over the fitted parameters. Second, it attempts to correct for any inadequacy of the model, which is revealed by a discrepancy between the observed data and the model predictions from even the best-fitting parameter values. This is done by adding the discrepancy as an unknown parameter in the estimation procedure.

This methodology allows one to adjust the predictions by the estimated

discrepancy and provides tolerance bounds for adjusted predictions. Depending on the size of the discrepancy, this can yield considerably more accurate predictions than can be achieved by using the model alone. Note, however, that this adjustment might have limited utility in extrapolation to new situations unless we are willing to make strong assumptions about how the bias extrapolates.

In this thesis, we refer to the term Bayesian Calibration as the process of performing inference with estimating the discrepancy.

2.4   Markov Chain Monte Carlo (MCMC)

This section is based on [Brooks [23]], [Roberts and Rosenthal [24]] and [Shaby and Wells [14]].

Posterior distributions in Bayesian statistics may often have a complex, high-dimensional form that does not belong to any known distribution family. This makes it impossible to calculate analytically. Therefore we resort to sampling techniques such as Markov Chain Monte Carlo(MCMC), where we sample from the posterior directly and obtain sample estimates of the quantities of interest.

The idea is: suppose we have some distribution $\pi(\theta), \theta \in E \subseteq \mathbb{R}^p$ known up to a normalization constant. We refer to this as the target distribution. If the target distribution $\pi(\theta)$ is very complex, we cannot sample from it directly. Therefore, we create an indirect sample method by constructing an a-periodic, irreducible Markov Chain with state-space E and whose stationary distribution is $\pi(\theta)$. Hence, if we run the chain for sufficiently long, simulated values can be treated as a dependent sample from the target distribution and used as the basis for summarizing important features of $\pi(\theta)$.

The main theorem underpinning the MCMC algorithm is that any aperiodic and irreducible chain will have a unique stationary distribution as $i \to \infty$, where i is a number of iterations [Roberts and Rosenthal [24]]. We manipulate the chain to let the target distribution π become the stationary distribution, and there are many updating and transition schemes to obtain this. We focus on the Random Walk Metropolis updating scheme.

To understand the Random Walk Metropolis updating scheme, we start by briefly introducing its core - namely the Metropolis-Hastings update. The

Metropolis-Hastings algorithm allows us to sample from a generic target distribution, even if we do not know the normalizing constant. To do this, we construct, and sample from a Markov chain, whose stationary distribution is the target distribution we are looking for. It consists of picking an arbitrary starting value and iteratively accepting or rejecting candidate samples drawn from a candidate distribution. We want to sample from a target distribution $\pi(\theta)$, but we only know it up to some normalizing constant. Hence

$$\pi(\theta) \propto g(\theta) \tag{2.8}$$

Since we do not know the normalizing constant, because $g(\theta)$ might be hard to integrate, we only have $g(\theta)$ to work with. The Metropolis-Hastings algorithm will proceed as follows:

1. Select initial value for θ_0

2. For i in m (where m is a large number of iterations) repeat:

   (a) Draw a candidate θ^* from proposal distribution $\theta^* \sim q(\theta^*|\theta_{i-1})$

   (b) Compute:
   $$\alpha = \frac{g(\theta^*)q(\theta_{i-1}|\theta^*)}{g(\theta_{i-1})q(\theta^*|\theta_{i-1})} \tag{2.9}$$

   (c) Check if:
   $\alpha \geq 1$: accept candidate θ^* and set $\theta_i = \theta^*$
   $0 < \alpha < 1$: accept candidate θ^* and set $\theta_i = \theta^*$ with probability α OR reject candidate θ^* and set $\theta_i = \theta_{i-1}$ with probability $1 - \alpha$

Steps (b) and (c) act as a correction since the proposal distribution q is not the target distribution π. At each step in the chain, we draw a candidate and decide whether to move the chain there or remain where we are. If the proposed move to the candidate is advantageous, i.e., $\alpha > 1$, we move for sure. On the other hand, if the proposed move is not advantageous, we might still move, but only with probability α. Since our decision to move to the candidate only depends on where the chain currently is, this is a Markov Chain.

One careful choice we must make is with the candidate generating distribution q [Shaby and Wells [14]]. It may or may not depend on the previous iterations value of θ. One example where it does not depend on the previous value would be if $q(\theta^*)$ is always the same distribution. If we take

this option, $q(\theta)$ should be similar to $\pi(\theta)$ to approximate it. Another popular option, one that does depend on the previous iteration, is the Random Walk Metropolis-Hastings. Here the proposal distribution is centered around the previous iteration. For instance, it might be a normal distribution where the mean is our previous iteration θ_{i-1}. Because the normal distribution is symmetric around its mean, this example comes with another really nice advantage. Now the two proposal distributions $q(\theta_{i-1}|\theta^*) = q(\theta^*|\theta_{i-1})$, cancel each other out when we calculate the acceptance probability α, hence 2.9 in the Algorithm reduces to

$$\alpha = \frac{g(\theta^*)}{g(\theta_{i-1})} \tag{2.10}$$

Clearly, not all candidate draws are accepted, so our Markov Chain sometimes stays where it is. Possibly for many iterations. How often you want the chain to accept candidates depends on the type of algorithm. If you approximate π with q and always draw candidates from that distribution, accepting candidates often is a good thing, it means that q approximates π well. However, you still may want q to have a more considerable variance than π and see some rejection in candidates as an assurance that q is covering the space well. On the contrary, a high acceptance rate for Random Walk is not good. If the Random Walk takes too small steps, it will accept candidates often, but it will take a very long time to explore the posterior distribution fully. Furthermore, if the Random Walk is taking too large steps, many of its proposals will have low probability and low acceptance rate. This will cause us to waste many draws. Ideally, a Random Walk sampler should accept somewhat between 23 and 50 percent of candidates proposed.

The Random Walk Metropolis-Hastings algorithm is applicable whatever the dimension of θ. Finding the posterior of the joint distribution of $\theta_1, \theta_2, \theta_3, \theta_4$, is equivalent to finding the posterior distribution of Θ, where Θ is a 4×1 vector with $\Theta = [\theta_1, \theta_2, \theta_3, \theta_4]$. The simplest idea is to apply the same strategy to each of the components, and defining a proposal of independent components such that

$$q(\Theta^*|\Theta_{i-1}) = q(\theta_1^*|\theta_{1,i-1})q(\theta_2^*|\theta_{2,i-1})q(\theta_3^*|\theta_{3,i-1})q(\theta_4^*|\theta_{4,i-1})$$

with a specified variance in the different directions the algorithm is moving [Roberts and Rosenthal [24]].

2.4.1 Adaptive MCMC

As mentioned above, a vital but difficult task for Metropolis-Hastings algorithms is choosing a good proposal distribution q. If, for instance, we assume q to be a normal distribution, to obtain efficient mixing, it is crucial to make sure that parameters such as the variance, σ_q^2, of the proposal distribution are small enough to approximate the target distribution π, but large enough to be sure we do not remain in a tiny area of the posterior space. The adaptive MCMC attempts to deal with this challenge by learning better parameter values while the algorithm is running. [Roberts and Rosenthal [24]] showed that adaptive MCMC can be very successful at finding good parameter values with little user intervention. Here, good is defined as some appropriate measure of Markov chain mixing, such as the integrated autocorrelation of a functional of interest. Adaptive MCMCs are proven to not only converge correctly but often have significantly better mixing properties than comparable nonadaptive algorithms [Roberts and Rosenthal [24]].

NIMBLE

NIMBLE stands for Numerical Inference for statistical models for Bayesian and Likelihood Estimation. It adopts and extends the probabilistic programming language BUGS as a modeling language and allows for the usage of customized models programmed by the user. Other packages that use the BUGS language are only for MCMC; however, NIMBLE turns BUGS code into model objects which can be used for whatever algorithm desired. This includes algorithms provided with NIMBLE and algorithms written by the user using nimbleFunctions. NIMBLE extends BUGS by allowing multiple parameterizations for distributions, user-written functions, and distributions. The NIMBLE algorithms are written so they can adapt to different statistical models. For MCMC, NIMBLE can assign a default set of sampler choices, but it is possible to customize the samplers from R. For example, one can choose what parameters to sample in a block and then easily write customized samplers and include them. Another one of NIMBLE great advantages lies in its compiling scheme. The compiler generates C++ for the customized models and algorithms, compiles that in C++, and allows for usage from R without knowing anything about or writing anything in C++ [*NIMBLE – An R package for programming with BUGS models and compiling parts of R.* [25]].

We apply the "RW sampler" from the NIMBLE library, which executes Random Walk adaptive Metropolis-Hastings sampling with a normal proposal distribution. The adaptation routine is described in full detail in [Shaby and Wells [14]]. This sampler can be applied to any scalar continuous-valued stochastic node and can optionally sample on a log scale [de Valpine [16]]. The adaptive MCMC constructs the proposal distribution $q(\theta, \cdot)$ as a multivariate normal distribution:

$$q(\theta, \cdot) \sim \mathcal{N}(\theta, \sigma_q \Sigma_0)$$

For some positive scaling constant σ_q and some $d \times d$ positive definite matrix Σ_0, where d is the number of parameters to be approximated.

The sampling algorithm applies a popular adaptive strategy of running a single chain, which, at each iteration, uses previous states to generate a proposal. The complication is that this approach destroys the Markov property; however [Shaby and Wells [14]] implements "controlled MCMC" to make sure that the resulting processes are ergodic and hence converge to the desired target distributions. This strategy involves vanishing adaption, where the idea is to dampen the adaption process such that the proposal distribution becomes approximately constant for large i. The algorithm performs what [Shaby and Wells [14]] calls Log Adaptive Proposals (LAP) of the Random Walk algorithm. The intuition behind the LAP is quite simple. It takes a block of Random Walk Metropolis steps and then estimates that block's acceptance rate. If it accepts too often, it increases σ_q; if it accepts too rarely, it decreases σ_q.

As a result of both NIMBLE's adaptive RW sampler and its compiling scheme, the simulation run-times and the burn-in phase of the chain are drastically reduced.

2.5   Gaussian Processes (GP)

This section is based on the book by [Gramacy [12]] and definitions of the powered exponential and Matern covariance functions are obtained from [scikit-learn [26]]. We apply Gaussian processes in this thesis both when modeling the discrepancy in Section 2.3 as well as simulating both independent and dependent noise in the simulation studies in Chapter 3.

A general definition of a Gaussian Process is a prior over unknown functions where any finite collection of realizations (i.e., n observations) are modeled as having a multivariate normal (MVN) distribution. More specifically, an MVN distribution means that all characteristics of those realizations can be completely described by their mean μ and covariance matrix Σ. These are often referred to as functions, i.e. mean-function $\mu(x)$ and covariance function or covariance kernel function $C(x, x')$. Applying these functions at some specific input location $x_1, x_2, \cdots, x_n$ can explain the gathered characteristics of these input locations from their mean vector and resulting covariance matrix. Hence we can define the collection of realizations drawing from an n-variate distribution as:

$$\mathcal{Y} \sim \mathcal{N}(\mu_n, \Sigma_{n \times n})$$

The covariance function explains the majority of the characteristics of a GP. In an MVN, a covariance function must be positive (semi)definite. This means that when defining a covariance matrix based on evaluating pairs of n x-values, we must have that:

$$x^T \Sigma x \geq 0 \forall x$$

Positive definiteness is the multivariate extension of requiring a univariate Gaussian to have a positive variance parameter σ^2.

As an example, we consider a covariance function known as the squared exponential. It is parameterized by a length scale parameter ρ, which can either be a scalar or a vector with the same number of dimensions as the number of covariates in the distribution. The length scale represents the rate of decay of correlation, i.e. it determines how far one needs to move in input space for the function values to become uncorrelated. The smaller the value of ρ, the higher the correlation between design points. More specifically, a small length scale value means that function values can change quickly, whereas large values mean that the function values change slowly. In other words, the length scale determines how far we can reliably extrapolate from the training data.

The squared exponential covariance function is given by:

$$C(d) = \sigma_\delta \exp\left(-\frac{||d||^2}{2\rho^2}\right) \tag{2.11}$$

Where $||d||^2$ is the squared Euclidean distance and d represents the distance between point x and x'. Observe how covariance here decays exponentially fast as x, and x' grows farther away from each other in the input or x-space. σ_δ is a scaling factor and determines the variation of the function from its mean. Small σ_δ means that the function stays close to its mean value, whereas larger values mean more variation and allows for the function to chase outliers.

Another example is the class of Matern covariance functions, which is a generalization of the squared exponential covariance function. It has an additional parameter ν which controls the smoothness of the resulting function. The smaller the ν, the less smooth the approximated function is. As $\nu \to \infty$, the covariance function becomes equivalent to the squared exponential. When $\nu = 1/2$, the Matern covariance function becomes identical to the absolute exponential covariance function [scikit-learn [26]]. Important intermediate values are $\nu = 3/2$ (once differentiable functions) and $\nu = 5/2$ (twice differentiable functions).

The covariance function is given by:

$$C(d) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{\rho}||d||\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}}{\rho}||d||\right) \tag{2.12}$$

Where $||d||$ is the Euclidean distance, $K_\nu(\cdot)$ is a modified Bessel function and $\Gamma(\cdot)$ is the gamma function.

Chapter 3

Models and Simulation Studies

We organize the simulation studies by first introducing the simulation models, from which we simulate synthetic field observations. Then we present the inference models with corresponding priors. Furthermore, we introduce the inference methods and a small toy example to illustrate the motivation for estimating the discrepancy. Finally, we present the simulation studies and their specifications.

3.1 Simulation Models

The simulation models are defined in line with Equation 2.4 in Chapter 2.3. Field observations y are, in the remaining of the thesis, noisy pressure observations P_{obs}. We simulate P_{obs} from the true model γ, which is either the two or three element Windkessel model with some added observational noise ϵ. Hence we can rewrite Equation 2.4 as

$$P_{obs}(t_i) = wk^{\gamma}_{model}(Q_i, t_i, \theta^*) + \epsilon(Q_i, t_i), \qquad i = 1, \ldots, n \qquad (3.1)$$

Where Q and t are input vectors containing the flow and time observations respectfully. θ^* is a vector holding the true values of the calibration parameters required to simulate from the Windkessel models. When $wk^{\gamma}_{model} = wk^{\gamma}_{2}$, θ^* contains R_{true} and C_{true}, and when $wk^{\gamma}_{model} = wk^{\gamma}_{3}$, θ^* contains the physical parameters R_{true} and C_{true} as well as the tuning parameter Z_{true}.

We obtain the pressure P by solving the linear differential Equations 2.1 and 2.2 of the two and three-element Windkessel models in the same way as

Segers, Rietzschel, Buyzere, Stergiopulos, Westerhof, Bortel, Gillebert, and Verdonck [4]. Here the pressure response P is calculated in the frequency domain through the impedance Z_{model}. Flow is decomposed into a Fourier series ΣQ_n. The response for each harmonic (index n) is calculated and recomposed into ΣP_n to obtain pressure estimate P from the model. For each harmonic the following applies:

$$P_n = Z_{model}Q_n$$

For the three element Windkessel model, Z_{model} becomes:

$$Z_{WK3} = Z_c + \frac{R}{1 + i2\pi fRC}$$

Where f is the frequency of the corresponding discrete Fourier transform samples, i the complex constant, R and C the physical parameters, and Z_c characteristic impedance. As Z is not an input or a tuning parameter in the two-element Windkessel model, the impedance is calculated as follows:

$$Z_{WK2} = \frac{R}{1 + i2\pi fRC}$$

To simulate noisy field observations we introduce independent noise ϵ and time-dependent noise ϵ_t to the simulation model wk^γ_{model}. It is important to note that both independent and dependent noise $\epsilon(\cdot)$ and $\epsilon_t(\cdot)$ are *functions* of the input points flow Q and time t, but the t in subscript of $\epsilon_t(\cdot)$ implies the *dependency* in time.

The two noise models are defined as following:

3.1.1   Independent And Identically Distributed Noise ϵ

As a benchmark, we first model the noise ϵ in the field observations to be independent in time and identically distributed $\mathcal{N}(0, \sigma^2 = 3^2)$. We present one example of a realisation from this distribution in Figure 3.1.

3.1.2   Time Dependent Noise ϵ_t

To simulate time dependent noise we apply a Gaussian Process with a Matern 5-2 covariance function described in Equation 3.2, as presented in Chapter 2.5 with $\nu = 5/2$
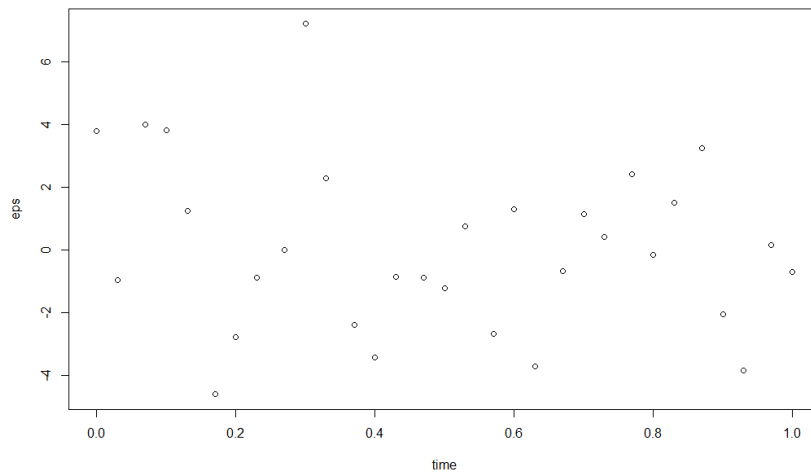
Figure 3.1: One realisation of independent noise ϵ.

$$C_{v=5/2}(d) = \sigma^2\left(1 + \frac{\sqrt{5}|d|}{\rho} + \frac{5d^2}{3\rho^2}\right)\exp\left(-\frac{\sqrt{5}|d|}{\rho}\right) \tag{3.2}$$

Where d represents distance between design points x and x'. The length scale ρ is set to 0.4. σ is an additional component of covariance when $x = x'$, and is set to 3 as for the independent noise. One realisation from this distribution is represented in Figure 3.2, where the errors clearly depict dependencies in time.

3.1.3   Simulated Inflow $Q(t)$

Input data of flow Q is plotted as a function of time t in Figure 3.3. To prevent overfitting and reduce numerical load, we sample $n = 50$ evenly distributed data points from Figure 3.3 and use this to simulate from the Windkessel model.

We want to keep the experiments within a controlled environment and use the same simulated flow Q as input for all experiments.
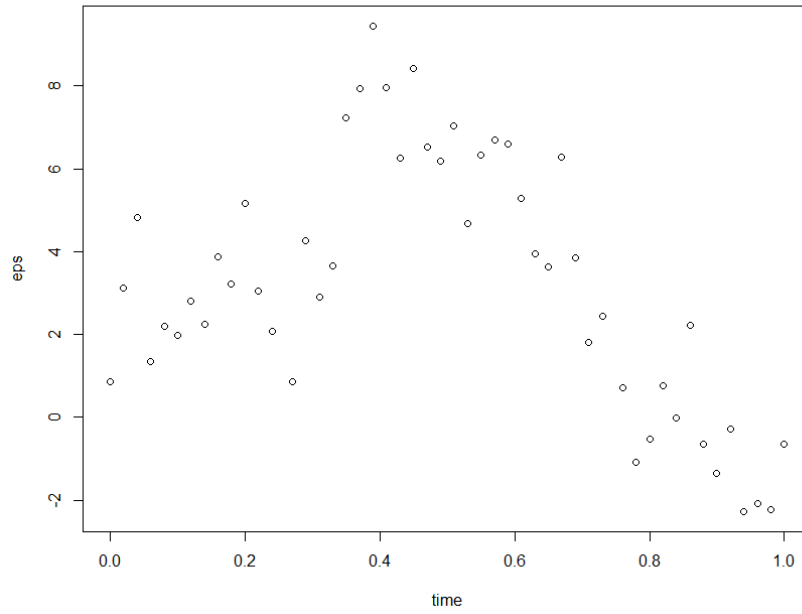
Figure 3.2: One realisation of dependent noise ϵ_t from a Gaussian Process with a Matern 5-2 kernel from 3.2 with length scale $\rho = 0.4$.
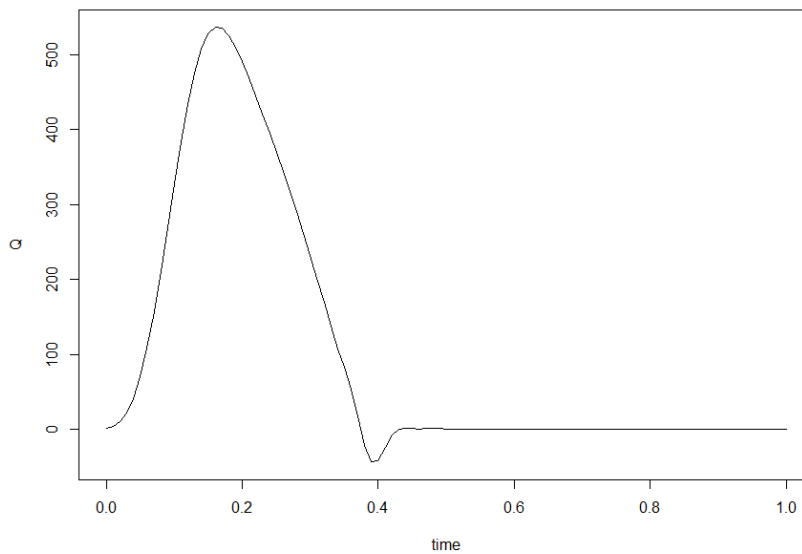
Figure 3.3: Flow $Q(t)$ used for the Windkessel models

3.2 Inference Models

3.2.1 Model Without Discrepancy

We combine the general definition of the computer model η simulating the real process γ without discrepancy from Equation 2.5 in Chapter 2.3 and

rewrite the equation as:

$$P_{obs}(t_i) = wk^{\eta}_{model}(Q_i, t_i, \theta) + \epsilon(Q_i, t_i), \qquad i = 1, \dots, n \qquad (3.3)$$

Where θ represents the calibration parameters we want to estimate. As for the simulation model, when $wk^{\eta}_{model} = wk^{\eta}_2$, θ represents R and C, and when $wk^{\eta}_{model} = wk^{\eta}_3$, θ represents the physical parameters R and C as well as the tuning parameter Z. In addition we want to estimate the standard deviance σ of noise ϵ.

Priors

The simulations are performed in a Bayesian framework presented in 2.2. Priors are set to the following:

$$p(R) \sim \mathcal{U}[0.5; 3]; p(C) \sim \mathcal{U}[0.5; 3];$$
$$p(\sigma) \sim \mathcal{U}[0, 30];$$

Where $\mathcal{U}$ denotes a uniform distribution between 0.5 and 3 for R and C. The choice of this range for the physical parameters is based upon previous work by [Spitieris, Steinsland, and Ingestrom [6]]. The prior for σ is set to be a uniform distribution between 0 and 30. For simulation studies where computer model is the three-element Windkessel model, Z is estimated and has the same prior as the physical parameters.

3.2.2 Model With Discrepancy

We now combine Equation 2.7 from 2.3, defining the computer model η simulating the real process γ accounting for discrepancy δ, and rewrite it as follows:

$$P_{obs}(t_i) = wk^{\eta}_{model}(Q_i, t_i, \theta) + \delta(Q_i, t_i) + \epsilon(Q_i, t_i), \qquad i = 1, \dots, n \qquad (3.4)$$

To consider the potential non-linearity of the discrepancy function, a Gaussian process prior with a powered exponential covariance kernel as presented in Chapter 2.5 is used to model the discrepancy. The powered exponential kernel is specified for our use case in Equation 3.5. As the input for the Windkessel model consists of both time t and flow Q, the length scale

is now a two-dimensional vector, where ρ_1 corresponds to flow Q and ρ_2 to time t.

$$C_\delta(d) = \sigma_\delta^2 \exp\left(-\frac{||d||^2}{2\rho_\delta^2}\right) + \sigma^2 \qquad (3.5)$$

We conduct the simulation studies for the inference model containing discrepancy by applying a full calibration framework, fitting model and discrepancy simultaneously, and two modularization frameworks, where we fit the model and discrepancy separately. We first present the priors for the full calibration framework, then the priors for fitting the discrepancy alone, and finally the two modularization approaches. The method is presented in 3.3.2

Priors For Full Calibration

In addition to R,C and σ we need to estimate the parameters ρ_1, ρ_2 and σ_δ related to the discrepancy δ. Keep in mind that the parameters of the discrepancy are non-identifiable and that we do not know the truth.

We want to construct priors of the length scales based on our knowledge of the data so far. We have transformed input data t and Q to be in the range between 0 and 1; consequently, there is no information in the data from which the likelihood can inform length scales above $\rho = 1$. This is because the likelihood is non-identified above the maximum covariate distance. Therefore we set the prior of the length-scales to a uniform prior between 0 and 1.

Priors for the physical parameters R and C remain the same and we set a uniform prior for both standard deviances as follows:

$$p(R) \sim \mathcal{U}[0.5,3]; p(C) \sim \mathcal{U}[0.5,3];$$
$$p(\sigma) \sim \mathcal{U}[0,10]; p(\sigma_\delta) \sim \mathcal{U}[0,500];$$
$$p(\rho_1) \sim \mathcal{U}[0,1]; p(\rho_2) \sim \mathcal{U}[0,1];$$

Priors For Fitting Discrepancy Alone

When we fit the discrepancy alone, we use the same argument as for the full calibration to set the limits of the prior for length scales. More specifically, we use the range of our input parameters time and flow. Since both t and Q are

transformed to be on a scale between 0 and 1, length scales are expected to be between 0 and 1.

However, this time, we are looking for a stricter prior for the length scale that places negligible mass below a lower length scale, l, and above an upper length scale, u. We follow the same approach from [Betancour [27]], where the choice falls on an inverse Gamma distribution due to its properties of a lighter tail towards zero, which more strongly constrains the posterior above l.

We want to translate the lower l and upper u length scales to a specified inverse Gamma distribution. To obtain the parameters shape a and scale b for this, we define the following criterion, which represent an inverse Gamma distribution that has 1% of prior mass above and below the limits of the length scales:

$$\int_0^l d\rho \text{Inv} - \mathscr{G}(\rho|a,b) = 0.01$$
$$\int_u^\infty d\rho \text{Inv} - \mathscr{G}(\rho|a,b) = 0.01$$

We obtain approximate values for a and b satisfying this criterion by making a Gaussian approximation to the tail conditions,

$$l \approx \mu - 3\sigma = \frac{b}{a-1} - 3\frac{b}{\sqrt{(a-1)^2(a-2)}}$$
$$u \approx \mu + 3\sigma = \frac{b}{a-1} + 3\frac{b}{\sqrt{(a-1)^2(a-2)}}$$

Hence, to obtain a prior distribution for length scales ρ, we solve the two equations above and set $l = 0$ and $u = 1$. This results in parameter approximations of shape a = 11 and scale b = 5, and an inverse Gamma distribution with the following specifications:

$$E(X) = \frac{b}{a-1} = \frac{5}{11-1} = 0.5$$
$$Var(X) = \frac{b^2}{(a-1)^2(a-2)} = \frac{5^2}{(11-1)^2(11-2)} = \frac{1}{36}$$

We set uniform priors for the variances σ and σ_δ. A reasonable range of values is assumed to be between 0 and 10 for σ, whereas σ_δ remains non-identifiable; consequently, we set a larger range between 0 and 500. This results in the following priors:

$$p(\sigma) \sim \mathcal{U}[0,10]; p(\sigma_\delta) \sim \mathcal{U}[0,500]$$
$$p(\rho_1) \sim Inv - \mathcal{G}[a=11; b=5]; p(\rho_2) \sim Inv - \mathcal{G}[a=11; b=5]$$

Priors For First Modular Framework

In the first modular framework, we fix the length scale parameters using the MAP estimates from fitting the discrepancy. Furthermore, we continue with uniform priors for σ, σ_δ and the physical parameters - leaving us with the following:

$$p(R) \sim \mathcal{U}[0.5,3]; p(C) \sim \mathcal{U}[0.5,3];$$
$$p(\sigma) \sim \mathcal{U}[0,10]; p(\sigma_\delta) \sim \mathcal{U}[0,500];$$

Priors For Second Modular Framework

In the second modular framework, we use the posterior from fitting the discrepancy as priors for the length scale parameters. We continue with an inverse Gamma distribution as prior, but this time we use the range of the posteriors as lower and upper bound. Consequently, $l = 0$ and $u = 1$ for ρ_1, whereas $l = 0$ and $u = 0.1$ for ρ_2, resulting in inverse Gamma distributions with the following specifications:

ρ_1 :

$$E(X) = \frac{b}{a-1} = \frac{5}{11-1} = 0.5$$

$$Var(X) = \frac{b^2}{(a-1)^2(a-2)} = \frac{5^2}{(11-1)^2(11-2)} = \frac{1}{36}$$

ρ_2 :

$$E(X) = \frac{b}{a-1} = \frac{0.5}{11-1} = 0.05$$

$$Var(X) = \frac{b^2}{(a-1)^2(a-2)} = \frac{0.5^2}{(11-1)^2(11-2)} \approx 0.278$$

With all priors summed up as:

$$p(R) \sim \mathcal{U}[0.5, 3]; p(C) \sim \mathcal{U}[0.5, 3];$$
$$p(\sigma) \sim \mathcal{U}[0, 10]; p(\sigma_\delta) \sim \mathcal{U}[0, 500];$$
$$p(\rho_1) \sim Inv - \mathcal{G}[a = 11,; b = 5]; p(\rho_2) \sim Inv - \mathcal{G}[a = 11; b = 0.5];$$

To ensure the robustness of the estimates when we conduct the simulations with observations containing dependent noise, we set strictly limiting priors for the σ's using an inverse Gamma distribution instead of uniform distributions. Hence, we change the prior for σ to be an inverse Gamma prior with $l = 2.9$ and $u = 3.1$, and for σ_δ, the lower bound $l = 0$ and upper $u = 100$. Yielding inverse Gamma distributions with the following specifications

σ :

$$E(X) = \frac{b}{a-1} = \frac{24303}{8102-1} = 3$$

$$Var(X) = \frac{b^2}{(8102-1)^2(8102-2)} = \frac{24303^2}{(11-1)^2(11-2)} = \frac{1}{900}$$

σ_δ :

$$E(X) = \frac{b}{a-1} = \frac{500}{11-1} = 50$$

$$Var(X) = \frac{b^2}{(a-1)^2(a-2)} = \frac{500^2}{(11-1)^2(11-2)} \approx 277.78$$

The new, more informative, priors are now:

$$p(R) \sim \mathcal{U}[0.5,3]; p(C) \sim \mathcal{U}[0.5,3];$$
$$p(\sigma) \sim Inv - \mathcal{G}[a = 8102, b = 24303]; p(\sigma_\delta) \sim Inv - \mathcal{G}[a = 11, b = 500];$$
$$p(\rho_1) \sim Inv - \mathcal{G}[a = 11, ; b = 5]; p(\rho_2) \sim Inv - \mathcal{G}[a = 11; b = 0.5];$$

3.3 Inference Methods

3.3.1 Samples From The MCMC

Posterior distributions are obtained using the adaptive Markov Chain Monte Carlo method discussed in 2.4.1. The results are evaluated from its parameter estimates and prediction uncertainty after running the adaptive MCMC one time with number of samples set to $N_{MCMC} = 10000$ and $N_{burn-in} = 5000$. The sampling is conducted with the same data but different initial points to ensure convergence. If the chains do not overlap, we run the chain longer until convergence.

3.3.2 Modularization

When performing Bayesian Calibration, we take a full calibration approach by fitting both model and discrepancy simultaneously. However, this might yield identifiability issues, and therefore we wish to pursue a modular approach by fitting discrepancy and model separately.

To obtain an initial approximation of the discrepancy, we conduct a fitting of the two-element Windkessel model and calculate the difference between the mean predictions and the field observations simulated from the three-element Windkessel. We then carry on by fitting the discrepancy and estimate its parameters ρ_1, ρ_2, σ_δ as well as noise in field observations σ with priors as presented in Chapter 3.2.2

After having fitted the discrepancy, we perform two types of modularization when returning to fitting the model:

1. Fix length scale priors to the MAP estimate of the posterior distribution in Figure 4.6a.

2. Use posteriors of length scales in Figure 4.6a as priors.

From now, we refer to the two approaches defined above as mod 1 and mod 2, respectively.

3.4   Motivational Toy Example

The motivation for including the discrepancy in the model is illustrated with a small toy example. We perform two simulation experiments where synthetic field observations are simulated with independent noise ϵ. We want to demonstrate a situation where the computer model can recreate the true process and when the computer model is biased. For both examples, we set the computer model to be the two-element Windkessel model wk_2^{η}. We then conduct two simulations where the simulation model is first defined as wk_2^{γ}. Then we perform a second simulation, but this time with simulation model wk_3^{γ}.

Equation 2.5 holds in the situation where there is assumed to be no discrepancy present, as the only difference between computer model wk_2^{η} and real system wk_{model}^{γ} should be the noise ϵ, which here is assumed to be independently distributed. We can therefore make a small test of whether or not there is discrepancy present by subtracting the expression on the right of Equation 2.5 from the expression on the left and plotting the result. We perform this on the two simulation setups, and results are depicted in Figure 3.4 and 3.5.

In the case of no discrepancy, one can expect a scatter plot resembling independent and identically distributed noise, randomly scattered around its mean with a constant variance. This is illustrated in figure 3.4, which is obtained by indeed the first experiment, where both simulation model and computer model are represented by the two-element Windkessel model. Hence, we can safely conclude that there is no discrepancy present.

Observing Figure 3.5 below, however, one can observe a pattern resembling the pressure curve. This plot is generated by performing the same test of subtracting the right side of Equation 2.5 from the expression on the left and plotting the result. In this example, the computer model is still represented by the two-element Windkessel model, but this time, we have defined the simulation model with the more complex three-element Windkessel model. The pattern of the noise in Figure 3.5 is a clear indicator that there is a discrepancy present that needs to be taken into account in the estimation procedure. Furthermore, Figure 3.5 confirms the shortcomings of the
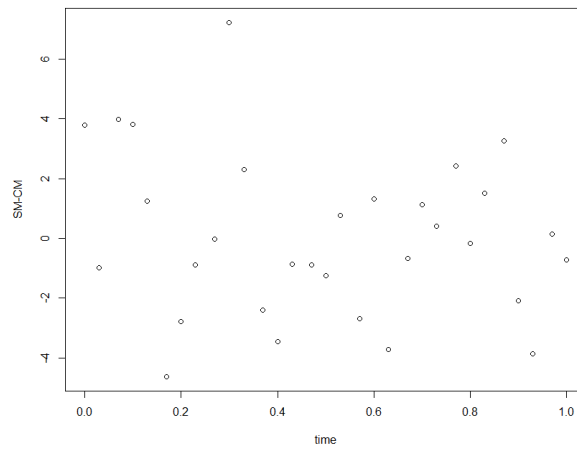
Figure 3.4: Discrepancy with no pattern
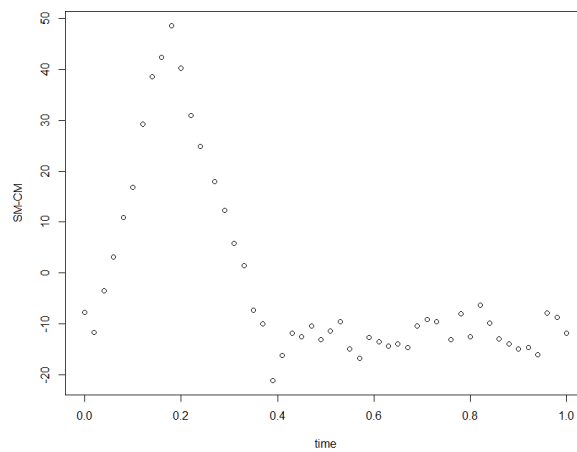
Figure 3.5: Discrepancy with pattern

two-element Windkessel model in systole, as we observe, this is where the discrepancy is at its largest.

3.5 Simulation Studies

Table 3.1 specifies all simulation studies in terms of the corresponding simulation and computer model, as well as the inference model. The independent noise is represented by ϵ, and ϵ_t represents dependent noise. For simplicity, time t, flow Q, and θ are omitted in the expressions. The full specifications can be found in Chapter 3.1 and 3.2.

Note that all experiments are set in a Bayesian framework, but in this thesis, the term Bayesian Calibration is only used for the experiments where the discrepancy is estimated.

We distinguish between cases where simulation and computer model are based on the same or different Windkessel models. Observe for instance, line 1 in the table above, where both simulation and computer models are represented as the two-element Windkessel, i.e., matching models. On the contrary, in line 3, where the simulation model is the three-element Windkessel, and the computer model the two-element Windkessel - hence they do not match. Therefore, we refer to the latter as the non-matching case and the former as a matching model case.

The column defining the inference model contains the models under investigation. Here we have omitted the parameters in the two and three-element Windkessel model, but as defined in Chapter 2.1 we estimate R and C for the two-element Windkessel model, and R, C and Z for the three-element Windkessel. Furthermore, we estimate the parameters contained in the independent and dependent noise models, defined in 3.1.1 and 3.1.2 respectively. Finally, where δ is included, we perform Bayesian Calibration and estimate the parameters of the discrepancy.

The column defining the inference method implies if we infer all parameters for the Bayesian Calibration at once or if we perform modularization by fitting the discrepancy and model separately. When we perform Bayesian Calibration and infer discrepancy and model parameters simultaneously, we refer to this as full calibration marked as "full calib" in the table. Furthermore, we make two modularization approaches, referred to as "Mod 1" and "Mod 2". For Simulation Study 1 and part of Simulation Study 3 we infer all parameters simultaneously but do not estimate the discrepancy. As we have restricted the usage of the term calibration for the case of estimating the discrepancy, we therefore mark these columns with a "-".

Model Combination And Simulation Study Summary				
Simulation Model	Computer Model	Inference Model	Simulation Study	Inference Method
$wk_2^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \epsilon$	1.1	-
$wk_3^\gamma + \epsilon$	wk_3^η	$wk_3^\eta + \epsilon$	1.2	-
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \epsilon$	1.3	-
$wk_2^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \epsilon$	1.4	-
$wk_3^\gamma + \epsilon_t$	wk_3^η	$wk_3^\eta + \epsilon$	1.5	-
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \epsilon$	1.6	-
$wk_2^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.1	Full Cal
$wk_3^\gamma + \epsilon$	wk_3^η	$wk_3^\eta + \delta + \epsilon$	2.2	Full Cal
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.3	Full Cal
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.4	Full Cal
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.5	Mod 1
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.6	Mod 1
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.7	Mod 2
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	2.8	Mod 2
$wk_2^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \epsilon$	3.1	-
$wk_3^\gamma + \epsilon_t$	wk_3^η	$wk_3^\eta + \epsilon$	3.2	-
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \epsilon$	3.3	-
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	3.4	Full Cal
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	3.5	Mod 1
$wk_3^\gamma + \epsilon_t$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	3.6	Mod 2
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	4.1	Full Cal
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	4.2	Mod 1
$wk_3^\gamma + \epsilon$	wk_2^η	$wk_2^\eta + \delta + \epsilon$	4.3	Mod 2

Table 3.1: Summary of all simulation studies

True Parameter Values		
Parameter	Matching Models	Non-Matching Models
$R = R_{wk_2}$	0.9	-
$R = R_{wk_3}$	0.9	-
$R = R_{wk_3} + Z$	-	1.0
Z	0.1	0.1
C	1.3	1.3
σ	3	3

Table 3.2: Summary of true parameter values used for simulation studies.

The main goal of these experiments is to investigate whether we can estimate physical parameters R and C from field observations. However, for us to be able to validate our results, we need to simulate from a model where the values of the true physical parameters are known. Note that in terms of interpretability, as described in 2.1, total vascular resistance R takes a different value when we fit the two-element Windkessel Model to the three-element Windkessel. On the other hand, the total arterial compliance C should be identical. Furthermore, we have no information about the true value of the parameters for the discrepancy but report the maximum posterior mean for all experiments. All true parameter values for simulation studies are defined in table 3.2.

Simulation Algorithm

Algorithm 1 presents the pseudocode for the general simulation algorithm used in all simulation studies.

Algorithm 1 Simulation Algoritm

1: **function** SIMALG(time,flow, N_{sims}, θ^*)
2:     Load Windkessel Model Simulators
3:     Transform flow to 0-1 scale
4:     **for** N_{sims} **do**
5:         **if** independent noise **then**
6:             Simulate P_{obs} from simulation model $wk^{\gamma}_{model} + \epsilon$
7:         **else**
8:             Simulate P_{obs} from simulation model $wk^{\gamma}_{model} + \epsilon_t$
9:         **end if**
10:         **if** Bayesian Calibration **then**
11:             Inference Model $= wk^{\eta}_{model} + \delta + \epsilon$
12:         **else**
13:             Inference Model $= wk^{\eta}_{model} + \epsilon$
14:         **end if**
15:         Define priors
16:         Run adaptive MCMC sampling on inference model
17:         Report mean, upper and lower 5% quantile of posteriors for parameters
18:     **end for**
19: **end function**

Code to simulate pressure observations from all the Windkessel Models were provided by the project [MyMDT [2]]. We use NIMBLE to define our models, with corresponding priors, covariance- and likelihood function. Furthermore , we conduct the MCMC sampling using NIMBLE. Complete code can be found in the following repository `https://github.com/helemo/Master-Thesis`.

3.5.1   Simulation Study 1: Inference Models Without Discrepancy

In Simulation Study 1, we fit matching and non-matching models. The parameters of the computer models are estimated using a Bayesian framework but without estimating the discrepancy. We produce 100 noisy data sets by using the simulation models corresponding to Simulation Study 1 in Table 3.1 and add both dependent and independent noise. We report the resulting 90% credible interval for each run by taking the 90th percentile of the sorted samples, and we estimate the average posterior mean of the

resulting 100 credible intervals. Furthermore, we report the coverage of the true value after simulating the posterior 100 times.

3.5.2 Simulation Study 2: Bayesian Calibration

In Simulation Study 2 we take the same approach as in Simulation Study 1 in 3.5.1, as we will investigate the case of both independent and dependent noise and produce 100 noisy data sets. This time, however, we wish to estimate the discrepancy as well.

We assume the bias to be normally distributed with a squared exponential kernel. Hence, to model the discrepancy we need to estimate its kernel parameters lengthscale ρ_δ and scaling parameter σ_δ presented in Equation 3.5. We also estimate noise in field observations σ.

We commence with a full calibration, fitting both model and discrepancy simultaneously, with priors defined as in 3.2.2. Furthermore, we perform modularization as described in 3.3.2.

Matching Models

The main focus lies on performing the simulation studies where the computer model is biased, i.e., non-matching models as presented in Table 3.1. Nevertheless, we experiment on the matching models for completeness, where there is no discrepancy between the computer and simulation models. We perform full calibration and simulate the noisy data-set with independent noise, specified in Table 3.1 as simulation studies 2.1 and 2.2. The priors are set to the same as for the full calibration in simulation study 2.3.

3.5.3 Simulation Study 3: Dependencies In ϵ_t

Due to simulation studies 1 and 2 showing instabilities in the presence of dependent noise, we want to investigate this further. We do so by rerunning all simulation studies in 1 and 2 with reduced and increased dependencies in the noise. As presented in 2.5, length scale is the parameter controlling the dependencies. Hence, we perform the simulation studies with length scales between $\rho = [0.1, 0.4, 0.8, 1, 5]$. As explained in 2.5, the smaller the length scale, the more correlated. Figures 3.6a, 3.6b, 3.6c, 3.6d and 3.6e depict corresponding realisations of the noise with different values of length scale. Note that the previous simulation studies are performed with a

dependent noise of length scale set to 0.4. It is also included here for comparison. Furthermore, observe that the closer we get to 1, the less correlated the noise gets, but then go above the value of 1 for the length scale, the noise again takes a slight dependent shape. This is again caused by the fact that our input to generate the noise, time, is on a 0-1 scale; hence there is no information in the data to support larger length scale values.

From the simulation models presented in Table 3.1, we simulate 50 noisy data-sets for each length scale and make inference using the corresponding inference method. From the resulting credibility intervals, we identify outliers in the sense of very narrow or very large credible intervals. We want to investigate whether possible mixing problems of the chain cause the size of the credibility interval or if there is a problem with the model itself. Hence we run the corresponding chain for 1 million iterations with a burn-in of 500 000.
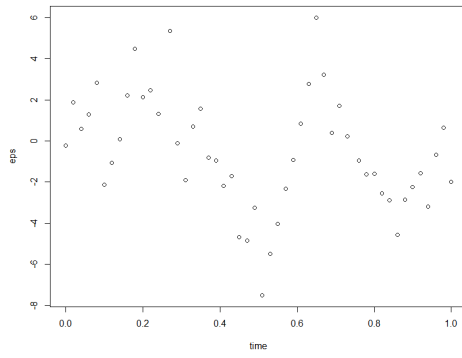
(a) Length scale $\rho = 0.1$
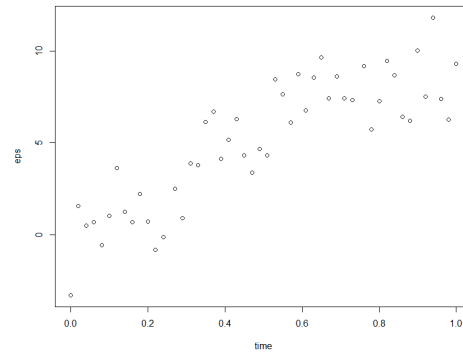
(b) Length scale $\rho = 0.4$
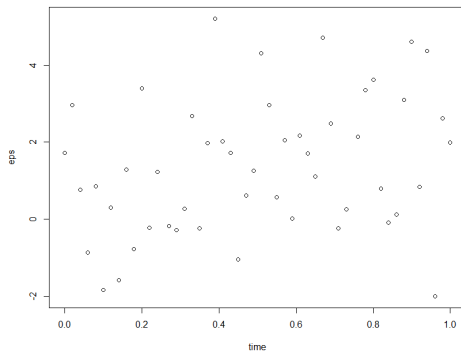
(c) Length scale $\rho = 0.8$
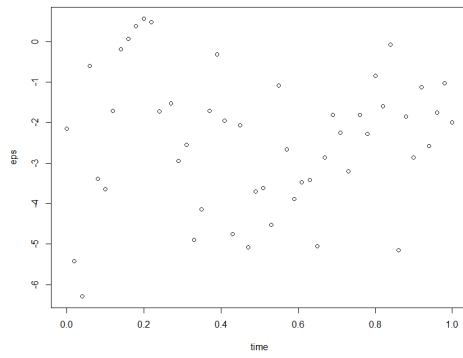
(d) Length scale $\rho = 1.0$
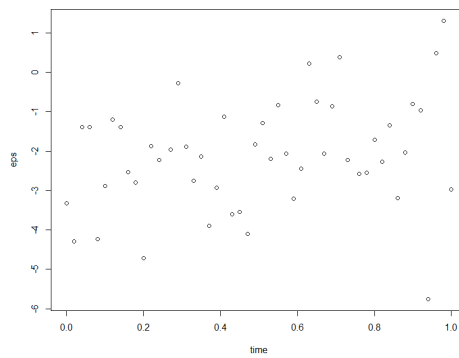
(e) Length scale $\rho = 5.0$

Figure 3.6: Simulation Study 3: One realisation of the noise with different dependencies.

3.5.4 Simulation Study 4: Computer Model Discrepancy

[Kennedy and O'Hagan [9]] presented that a flawed model which tries to describe complex data will show its uncertainty in its posteriors. Therefore, we want to perform simulation studies where we increase uncertainty by increasing Z when simulating data from the simulation model. Furthermore, the solution presented by [Kennedy and O'Hagan [9]] states that in the case of large uncertainty in the posteriors, we incorporate the discrepancy function to reduce the uncertainty. Therefore, we only conduct this simulation study on inference models where we estimate the discrepancy, as it would be hard to separate if the uncertainty in the posteriors occur because we are not estimating the discrepancy or, in fact, due to the increase in discrepancy.

As presented in Chapter 2.1, larger Z increases the amplitude of the three-element Windkessel model; hence we obtain a more considerable discrepancy when we have the two-element Windkessel model as a computer model. To investigate this effect on our model set up, we perform the simulation study on 5 simulated data sets generated with 5 different values of $Z = [0.01, 0.05, 0.1, 0.15, 0.2]$ and independent noise. $Z = 0.1$ is the parameter value we have used to generate the data-sets for all simulation studies in this thesis and is included for comparison. The resulting simulated pressure curves from the three-element Windkessel model are illustrated in Figure 3.7, where the black stapled line represents the two-element Windkessel model used as a computer model, and we observe how the discrepancy is increasing as we increase Z.
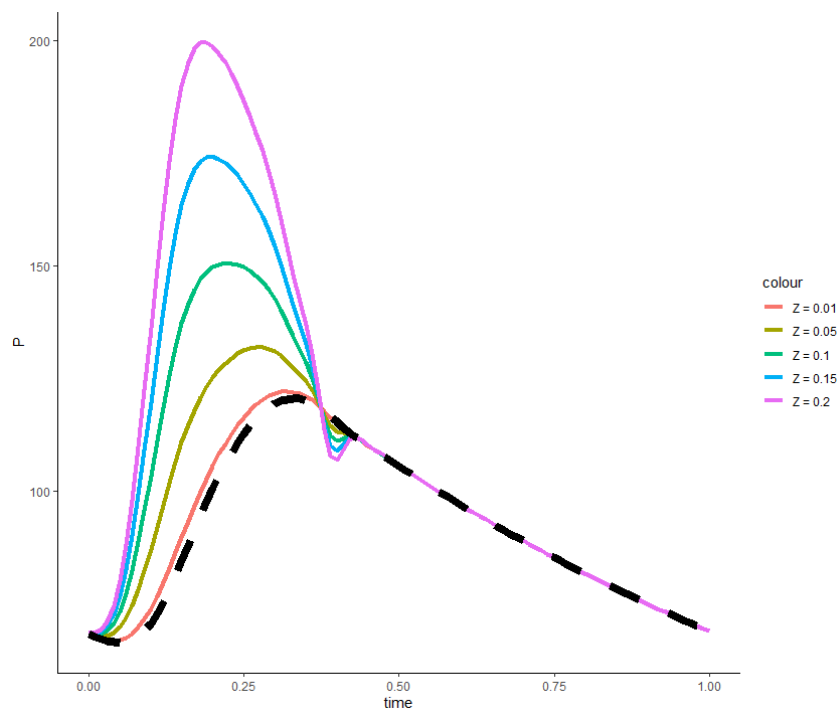
Figure 3.7: Simulation Study 4: Resulting pressure curves from increased Z in the three element Windkessel model. The thick, black stapled line represents the two element Windkessel model used as computer model for all experiments.

Results

We summarize the results for each simulation study by reporting the bias, mean 90% credible interval(CI), and the total coverage of the true mean for all 90% CI's. The mean CI is calculated from averaging the lower and upper bound of the 100 90% CI's. Together with the mean MAP estimate, we can investigate the biasedness and uncertainty in the posteriors. In addition, we present the total coverage of the true value to draw conclusions of the stability of the credible intervals for each simulation run. For example, one might observe an excellent average MAP estimate and narrow credible intervals, but the coverage tells us how much the posterior is "jumping" for each simulated data-set.

We focus on presenting results related to the physical parameters and observation noise σ, whereas results related to the discrepancy's parameters and all results in full can be found in the Appendix.

Finally, we present MCMC diagnostics to assess convergence of the chains.

For a summary of simulation studies and their corresponding computer, simulation and inference model, we refer to Table 3.1.

4.1 Results Simulation Study 1: Inference Models Without Discrepancy

We present mean MAP, CI, and coverage of the results for the physical parameters and observational noise from simulation study 1, and refer to the Appendix 6.1 for the entire presentation of the results.

We commence with C and observe mean MAP estimates with corresponding

mean 90% CI's in Figure 4.1. The MAP estimates are accurate for simulation studies 1.1, 1.2, 1.4, and 1.5 (matching models) with narrow mean 90% credible intervals. Nevertheless, when we look at the coverage depicted in Figure 4.5, we see that just around half of the simulations gave 90% CIs covering the true mean. Hence, we have instability of the posterior for each model run. To the contrary, results from simulation studies 1.3 and 1.6 (the non-matching models) report that all CI's cover the true mean. Going back to Figure 4.1, however, we see large uncertainties in the average CI. This implies a stable but uncertain estimate of the posterior.

The mean CI's for R are depicted in Figure 4.2. As for C, the mean CI's from simulation studies 1.3 and 1.6(non-matching models) show a biased MAP with an increase in uncertainty for the posterior. Note that the true value of R changes between the matching and non-matching results as explained in Chapter 2.1. Furthermore, we observe that the results for simulation studies 1.4 and 1.5 corresponding to data generated from the simulation model with dependent noise are very narrow, but the average MAP estimates are biased and lie outside the average credible intervals. To understand why this happens, we look at the total coverage in Figure 4.5, where we do observe an insufficient coverage of the true mean for these simulation studies. This implies that the credible intervals are "jumping" around, leading to biased estimates of the average MAP. We confirm this by plotting all resulting 90% CI's from the simulation studies 1.4,1.5 and 1.6 in Figures 4.3a-c), which indeed depicts large variability in estimates and credibility intervals between each simulated data-set. This is a good example of the case where we have posterior distributions with a small uncertainty, but the estimate is not to be trusted for one data-set as we might be far from the true parameter value.

For simulation studies 1.1, 1.2, 1.4 and 1.5 we observe a good and stable fit for σ, with narrow average CI's and high total coverage in Figure 4.4. On the contrary, the observation noise σ is highly overestimated for simulation studies 1.3 and 1.6.

The biased mean MAP and increased uncertainty in the mean CI's for the parameters in simulation studies 1.3 and 1.6 are results of the discrepancy between computer and simulation model that we do not estimate in simulation study 1.

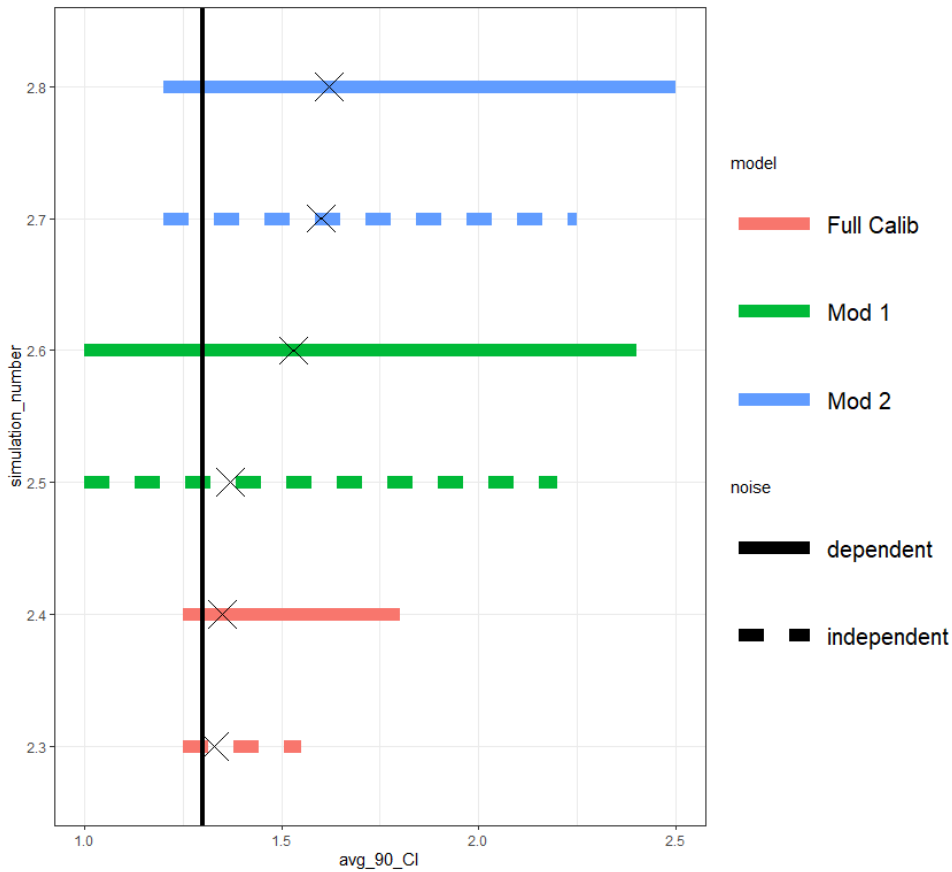We summarize the total coverage of the true parameter value for the
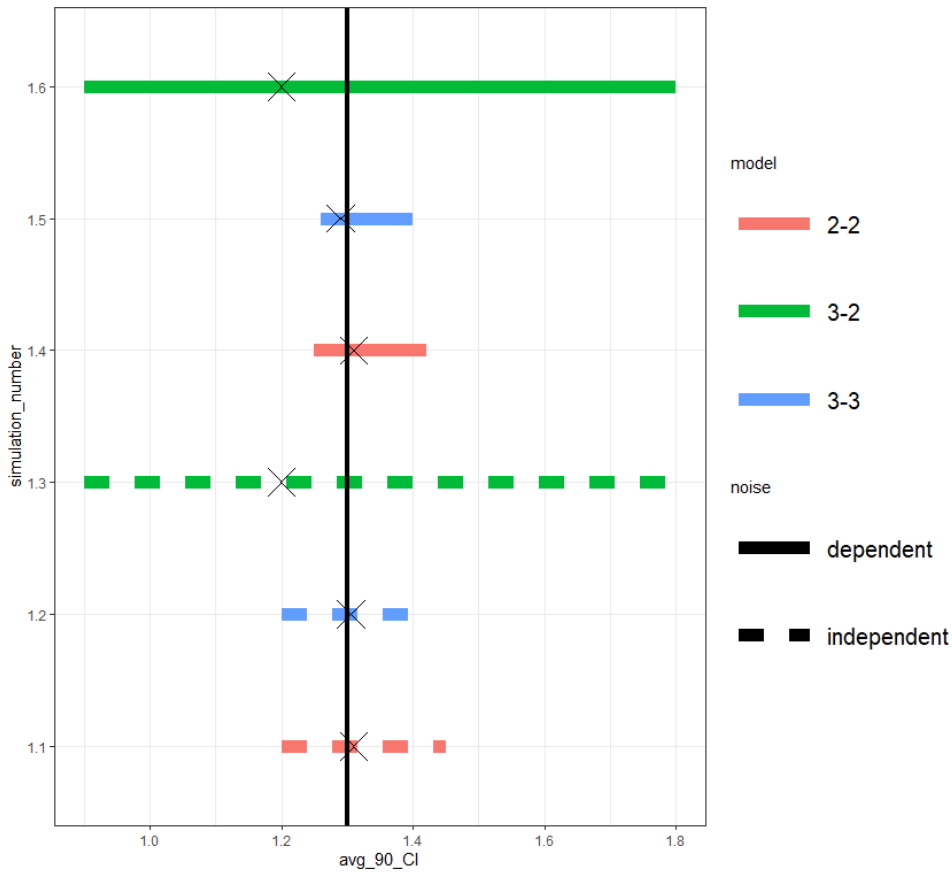
Figure 4.1: Simulation Study 1: Average Credible Interval from simulating the posterior of C 100 times. The corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and the vertical thick black line represents the true parameter value. Legend "Model" refers to which Windkessel model used as simulation model(To the left of the dash -) and computer model (to the right of the dash -).

physical parameters R, C and observations noise σ in Figure 4.5. The coverage is indeed much higher for independent noise compared to dependent noise. Nevertheless, we have high coverage for simulation study 1.3 and 1.6 for both physical parameters, except for observation noise σ, which is zero. However, we have seen that the high coverage of R and C results from the increased uncertainty in the credible intervals for simulation studies 1.3 and 1.6.

4.2   Results Simulation Study 2: Bayesian Calibration

We now present the results from performing Bayesian calibration where we estimate the discrepancy in addition to the physical parameters and observational noise. Results are presented through the mean MAP, the mean
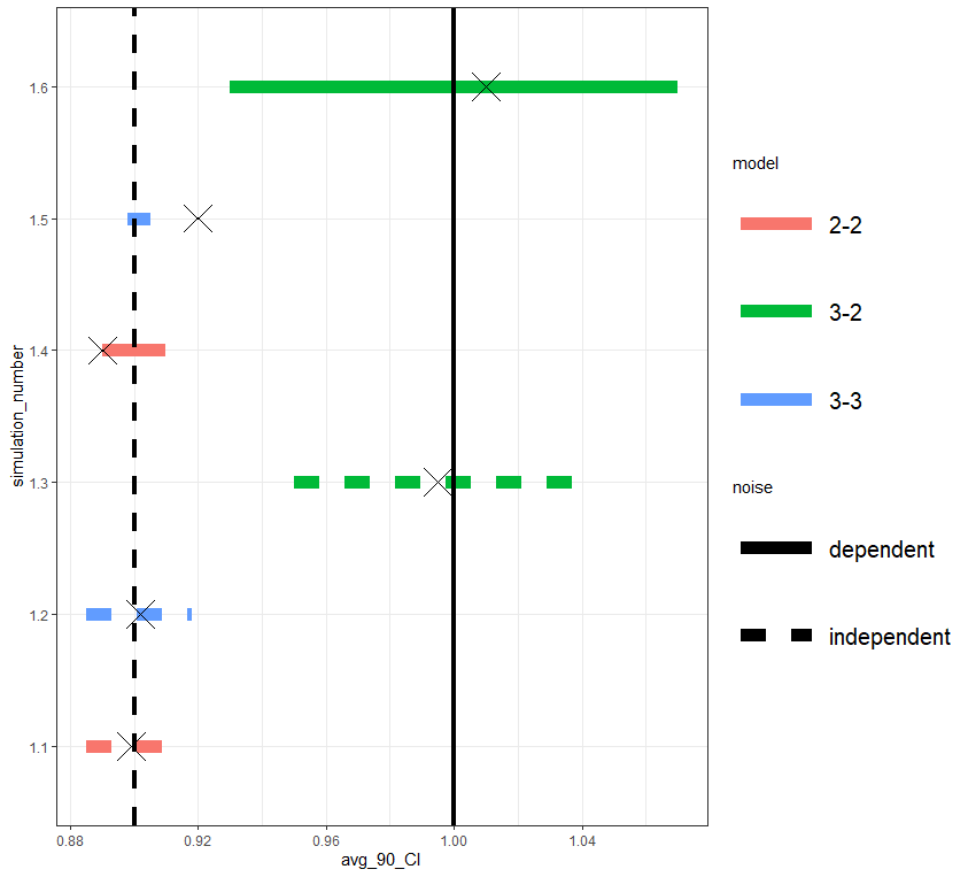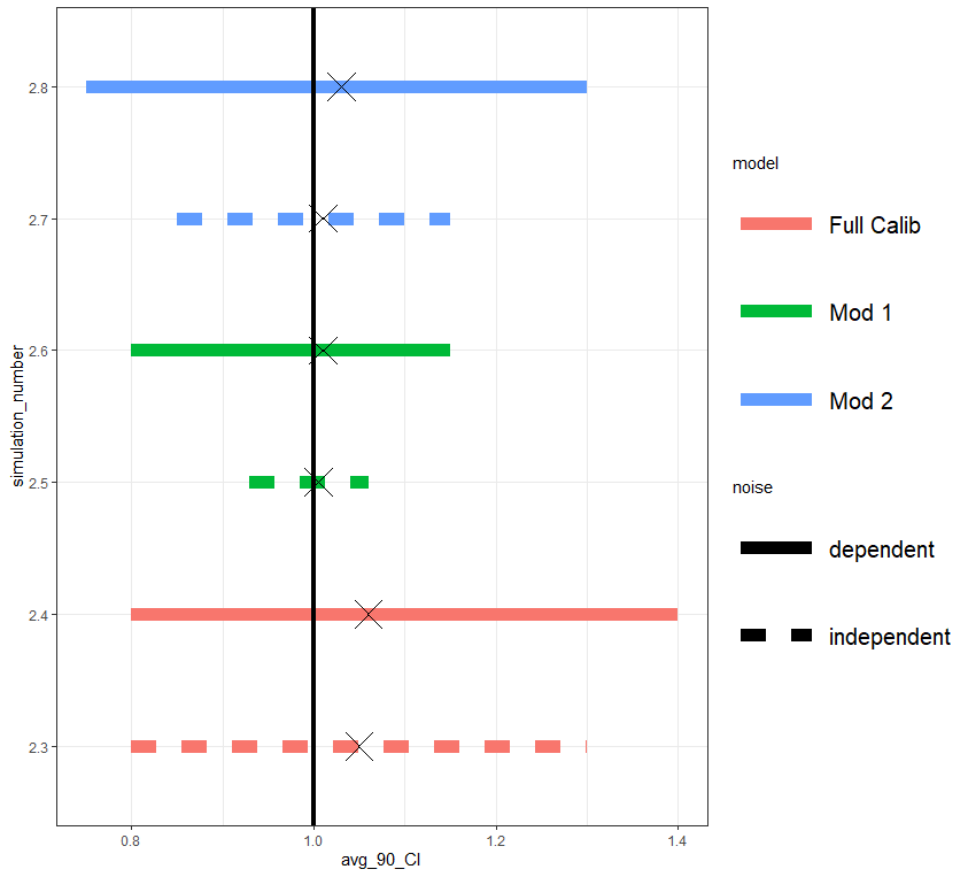
Figure 4.2: Simulation Study 1: Average Credible Interval from simulating the posterior of R 100 times. The corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and the vertical thick black line represents the true parameter value for simulation studies 1.3 and 1.6 and dashed, vertical line true parameter value for simulation studies 1.1,1.2,1.4,1.5. Legend "Model" refers to which Windkessel model used as simulation model(To the left of the dash -) and computer model (to the right of the dash -).
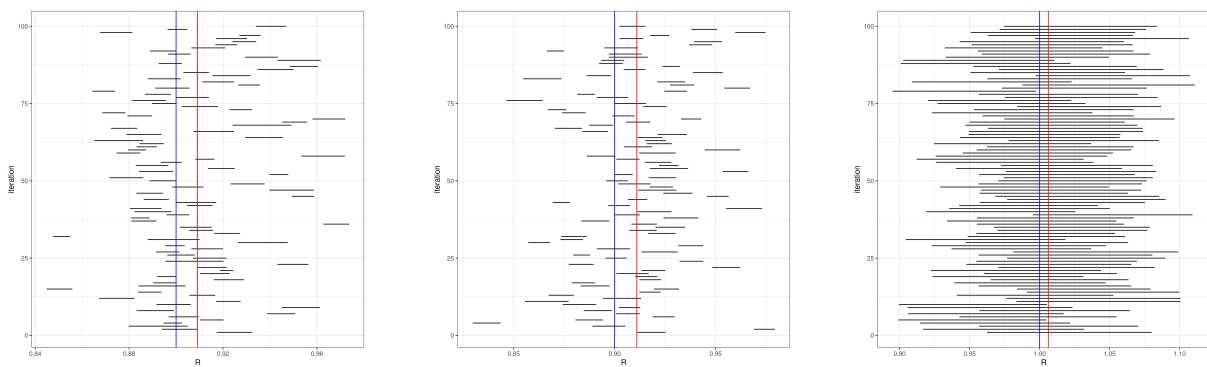
(a) Simulation Study 1.4: 2-2 - dependent noise ϵ_t

(b) Simulation Study 1.5: 3-3 - dependent noise ϵ_t

(c) Simulation Study 1.6: 3-2 - dependent noise ϵ_t

Figure 4.3: Mean MAP estimate with corresponding 90% credible intervals of R from 100 simulations of the posterior distribution with dependent noise ϵ_t. The blue line represents the true parameter value and the red line the mean MAP.
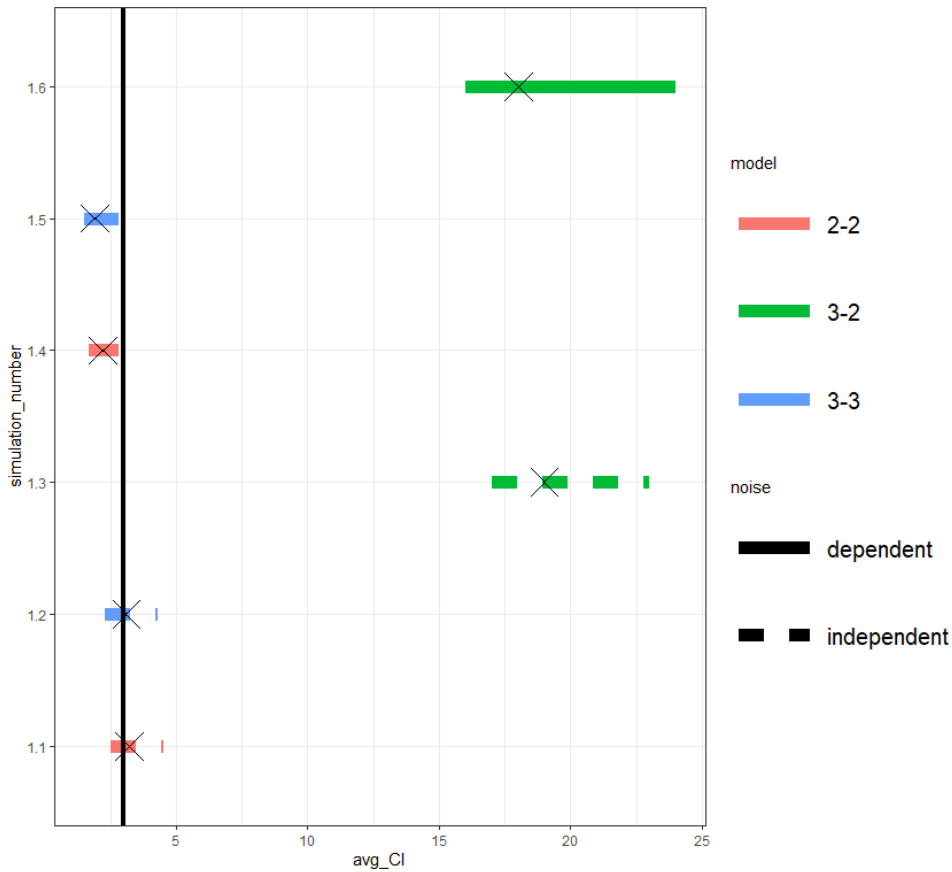
Figure 4.4: Simulation Study 1: Average Credible Interval from simulating the posterior of σ 100 times. The corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and the vertical thick black line 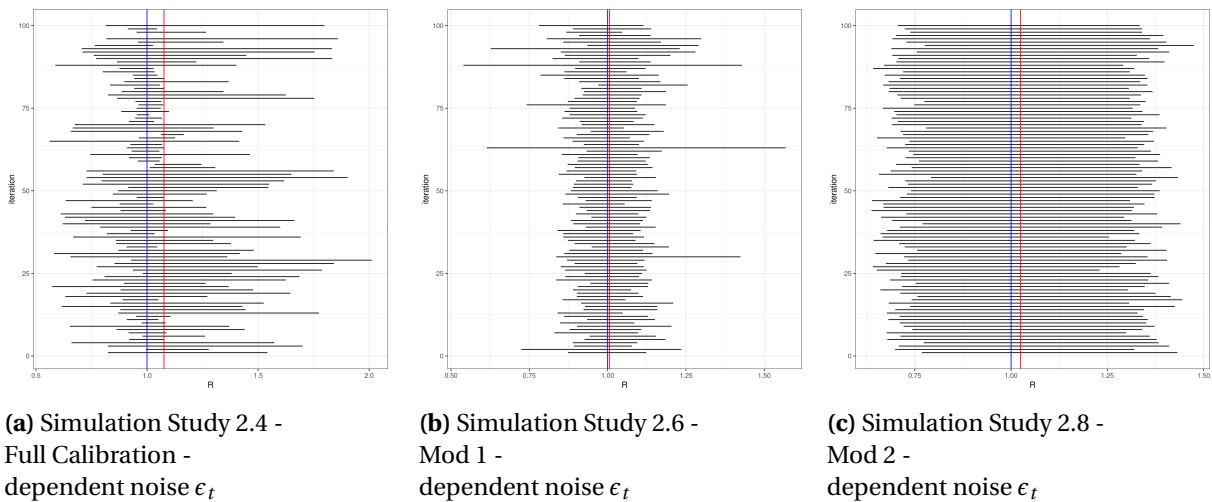represents the true parameter value. Legend "Model" refers to which Windkessel model used as simulation model(To the left of the dash -) and computer model (to the right of the dash -).

CI, and coverage of the true value. We focus on the results for the physical parameters and observational noise in this Chapter and refer to Appendix 6.3 and 6.4 for the entire presentation of the results for simulation study 2.

4.2.1 Results From Fitting the Discrepancy

We commence with the results from fitting the discrepancy alone, as this creates the baseline for the upcoming modular approaches. The resulting posteriors are depicted in Figure 4.6a. The correlation between parameters are shown in Figure 4.6b, where all correlation coefficients are of high significance(the larger number of $*'s$ behind coefficient, the more significant), but of moderate size except for the one between length scale ρ_1 and σ_δ. The first modular approach utilizes the MAP estimate of the length scale parameters, hence $\rho_1 = 0.50$ and $\rho_2 = 0.075$, whereas the second
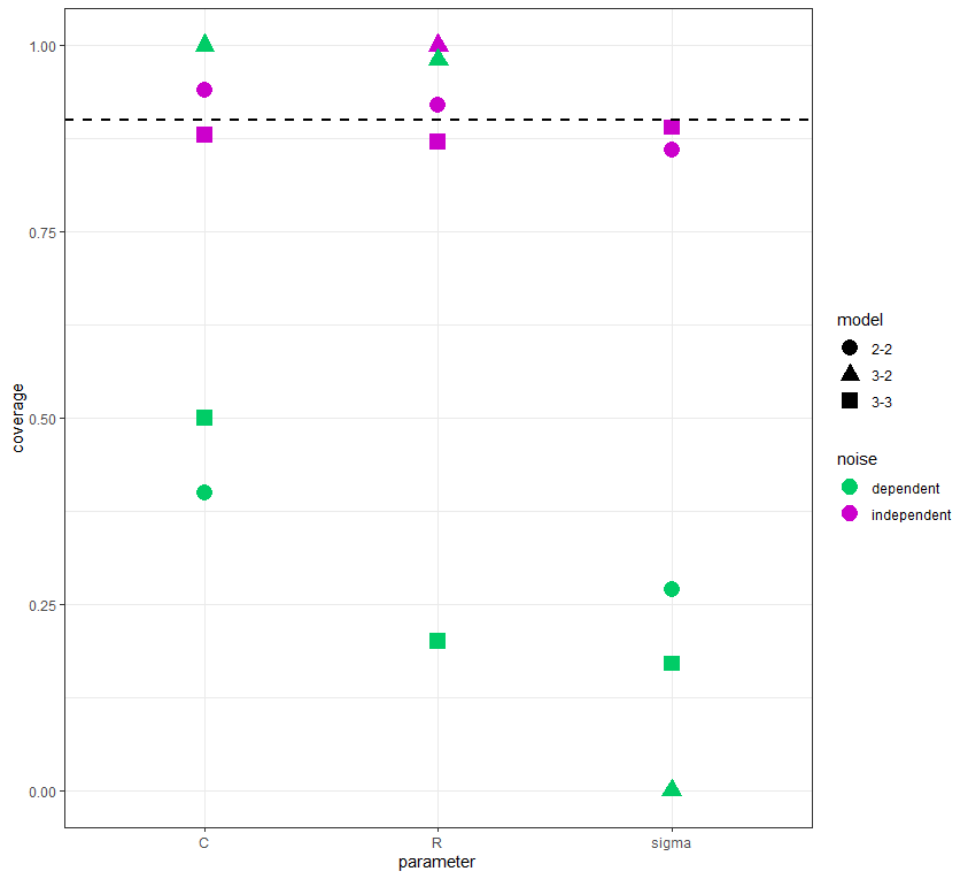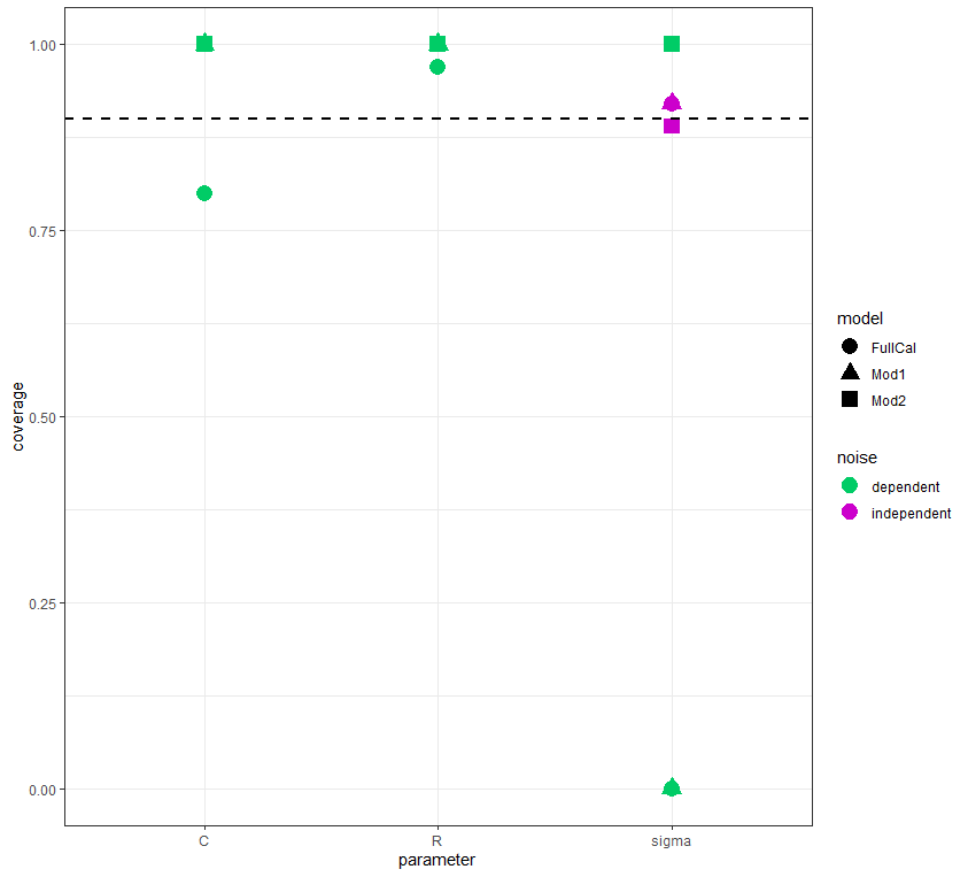
Figure 4.5: Simulation Study 1: Coverage of true mean of resulting 90% credible intervals for 100 simulations of the posterior of the physical parameters R and C, and observation noise σ. From Table 3.1: 2-2 corresponds to simulation studies 1.1 and 1.4, 3-3 to simulation study 1.2 and 1.5, and 3-2 simulation study 1.3 and 1.6. Dashed line is at 0.9 and represents the expected coverage from a 90% credible interval.

modular approach takes the limits of the posterior and forms a prior for the length scales as presented in Chapter 3.

4.2.2   Results Bayesian Calibration

We have performed Bayesian calibration through a full calibration approach, fitting discrepancy and model simultaneously, and two modular approaches, fitting discrepancy and model separately.

Figure 4.7 shows the mean MAP and credible intervals for C. Here we observe larger uncertainties for the mean CI's of the two modular approaches compared to the full calibration. In addition, the mean MAP estimate for simulation studies 2.6, 2.7, and 2.8 overestimates the true value of C. Furthermore, the introduction of dependent noise is prominent in the results, with a further increase of uncertainty when compared to independent noise. The high coverage in Figure 4.11 for the two modular approaches implies that the CI's are stable but with large uncertainty for all data-sets. The full calibration also depicts a high coverage in Figure 4.11, but is coupled with narrower average CIs, and less biased average MAP estimates in Figure 4.7 than the two modular frameworks. Nevertheless, with a tail towards the right, which is further increased for the dependent noise.

Compared to the mean MAP and CI's for C, R overall has less biased MAP estimates and narrower mean CI's in Figure 4.8. This is further backed up by high coverage of the true parameter value in Figure 4.11. Whereas C showed the largest uncertainties for the modularization, R, on the other hand, shows the largest average CI for the full calibration with biased MAP estimates in Figure 4.8. For comparison to simulation study 1, we plot all resulting 90% CI's for simulation studies 2.4, 2.6 and 2.8 of R in Figures 4.9a-c). Even though we do not observe the same instability of the credible intervals, we now observe outliers in the sense of very large or tiny credible intervals for simulation studies 2.4 and 2.6. We investigate what might cause this behavior further in simulation study 3.

The resulting mean CI's for the observation noise σ are presented in Figure 4.10. Here we observe minor uncertainties and accurate average MAP estimates for all simulation studies with independent noise. However, the second modular approach is the only framework able to remain robust against the dependent noise. This is expected due to the highly informative
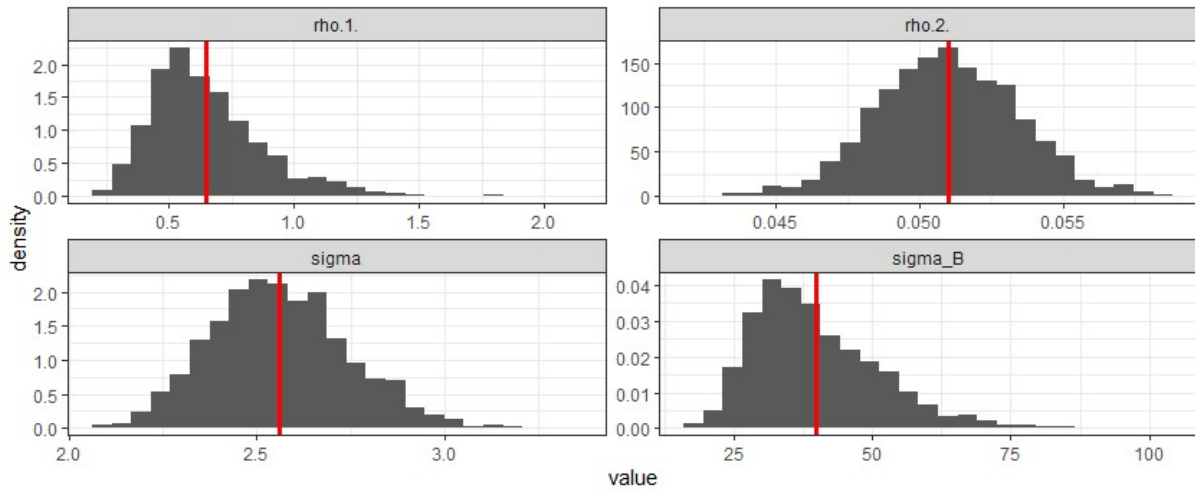
(a) Simulation Study 2: Resulting posterior distributions from estimating the discrepancy from one simulated data-set. Red line represents estimated MAP of parameter.
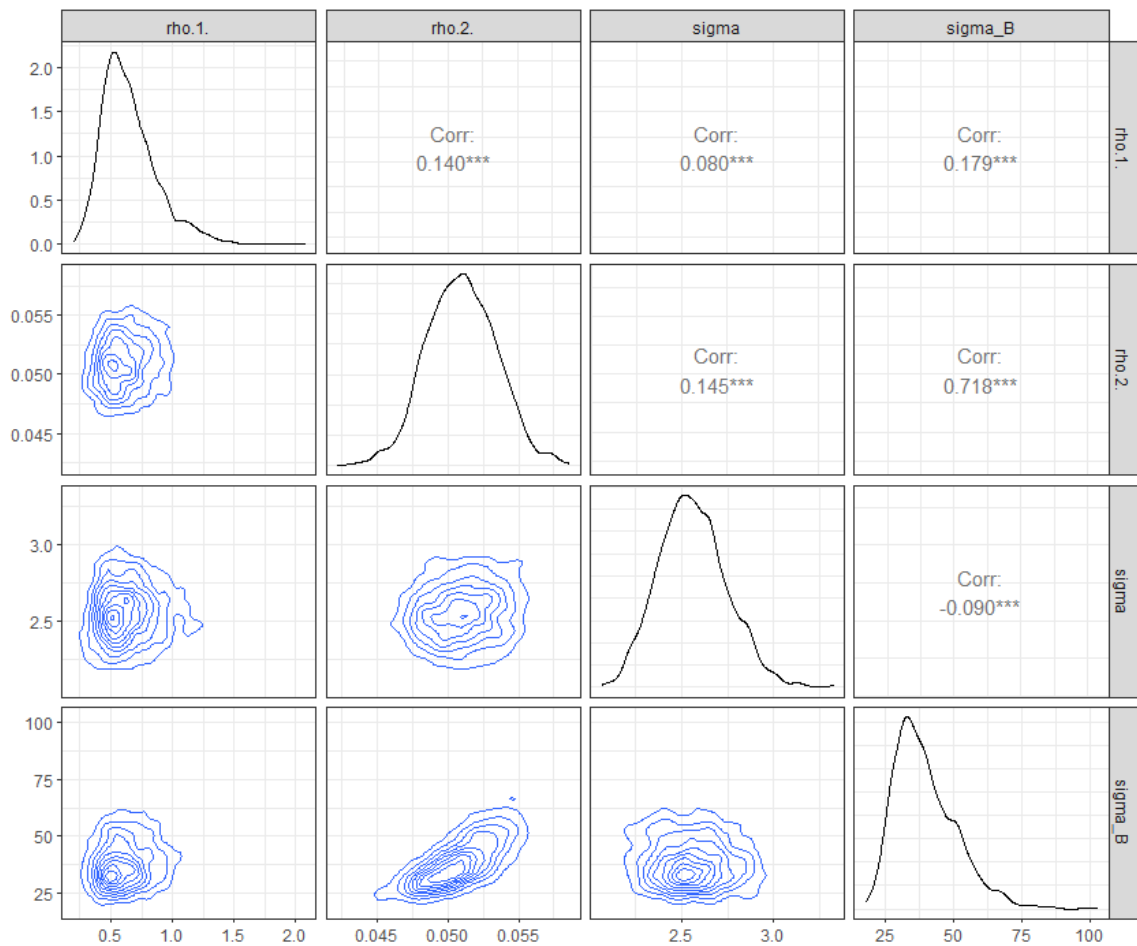
(b) Simulation Study 2: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 4.6: Discrepancy parameters estimate.

Figure 4.7: Simulation Study 2: Average Credible Interval from simulating the posterior of C 100 times. Corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and vertical thick black line represents the true parameter value.

and strictly limiting priors we incorporate here compared to the first modular and full calibration framework.

The total coverage of the true parameter value for the physical parameters R, C and observations noise σ is summarized in Figure 4.11. We observe that the coverage in total is high, indicating stable, credible intervals. Nevertheless, we observe a coverage of 0 for σ in simulation studies 2.4 and 2.5. In addition, C also suffers from a reduction in coverage for simulation study 2.4(when we perform the simulations in the full calibration framework with dependent noise). However, as we have observed in Figure 4.7, the uncertainty in the credible intervals is large for C for particularly the second modular approach. R, on the other hand, yields average credible intervals with less uncertainty for both dependent and independent noise and consistent high coverage of the true value.

Combining the results observed from the average credible intervals and MAP

Figure 4.8: Simulation Study 2: Average Credible Interval from simulating the posterior of R 100 times. Corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and vertical thick black line represents the true parameter value.

(a) Simulation Study 2.4 - Full Calibration - dependent noise ϵ_t

(b) Simulation Study 2.6 - Mod 1 - dependent noise ϵ_t

(c) Simulation Study 2.8 - Mod 2 - dependent noise ϵ_t

Figure 4.9: Average MAP estimate with corresponding 90% credible intervals of R from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. The blue line represents the true mean and the red line the average of the 100 maximum posterior mean values.
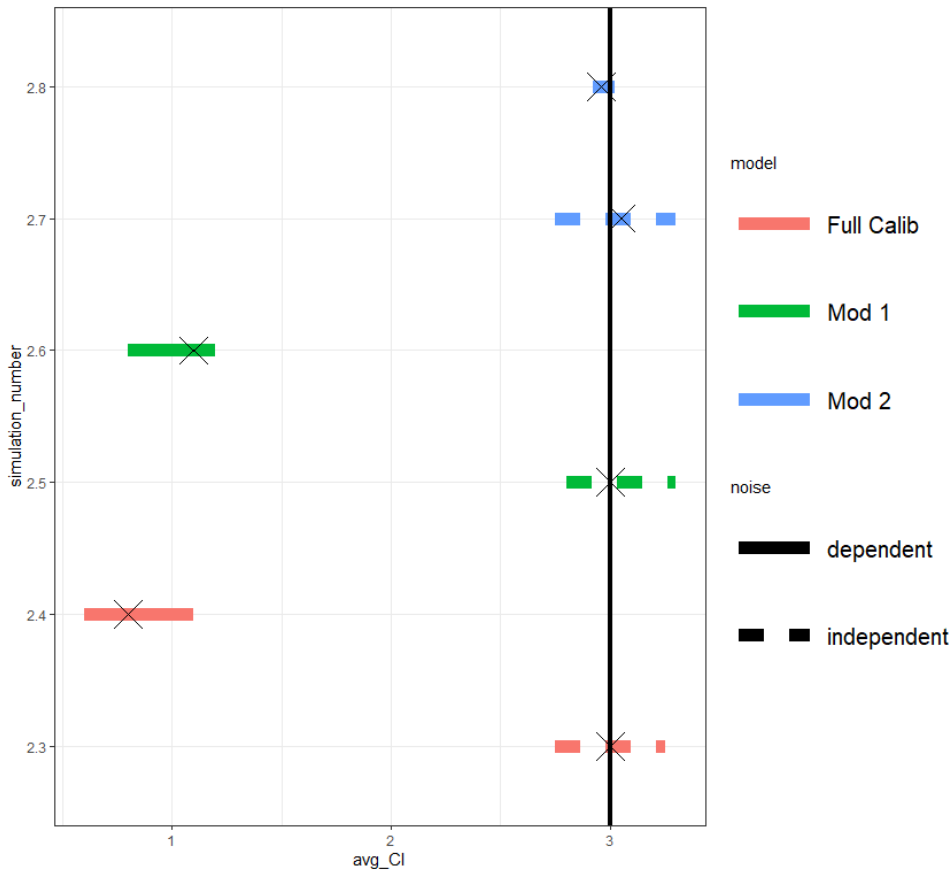
Figure 4.10: Simulation Study 2: Average Credible Interval from simulating the posterior of σ 100 times. Corresponding simulation study number as presented in Table 3.1 is depicted on the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectfully. X marks the average MAP estimate, and vertical thick black line represents the true parameter value.

estimate with the total coverage, we conclude that the most stable and robust results are obtained for the second modular approach, at the cost of slightly larger uncertainties in the posteriors for physical parameters C and R. .

4.3   Results Simulation Study 3: Dependencies in ϵ_t

We have constructed noise with length scales $\rho = (0.1, 0.4, 0.8, 1.0, 5.0)$ to investigate the effect of different dependencies in ϵ_t. For each level of dependent noise, we generated 50 noisy data sets, from which we simulated the posteriors for all simulation studies.

Results from simulation study 3.1, 3.2, and 3.3, showed behaviour in line with what we observed for simulation study 1, i.e., a large instability in the credible intervals for each data-set causing the posteriors to "jump" around. Furthermore, all credibility intervals behaved similarly, growing narrower as

Figure 4.11: Simulation Study 2: Total coverage of true mean of resulting 90% credible intervals for 100 simulations of the posterior of physical parameters R and C and observations noise σ. Dashed line is at 0.9 and represents the expected coverage from a 90% credible interval.

we increased the length scale. Therefore we do not pursue to investigate the credible intervals further, as we did not identify outliers of interest.

Simulation studies 3.4, 3.5, and 3.6, however, showed behaviour with what we observed in simulation study 2. This is illustrated in Figures 4.12a, 4.12b and 4.12c, which present resulting 90% credible intervals of the physical parameter R with from simulation studies 3.4, 3.5 and 3.6 with length scale $\rho = 0.8$. The second modular approach (Figure 4.12c) is once again robust when introduced to all kinds of dependent noise. This is no surprise since this framework is set up with the most informative and limiting priors. Consequently, the resulting credible intervals from simulating the posterior for different dependencies remained stable across the board, with Figure 4.12c presented as one example.

However, we observe a different behavior for the full calibration and first modular approach. In line with what we observed in simulation study 2, Figures 4.12a and 4.12b show credible intervals with some very small and

huge ranges. We investigated outliers of these two frameworks further by choosing one small and one large, credible interval and ran the corresponding chain much longer.

We present results from conducting the experiments with length scale $\rho = 0.8$ as we encountered numerical issues with the sampling algorithm of the full calibration for all other length scales. In addition, the full calibration framework showed substantial mixing issues for the smallest observed credible intervals, for instance, nr 5 from the bottom in Figure 4.12a. An example of this is presented in Appendix 6.5. Therefore, we chose to run the chain longer for medium-sized credible intervals, such as simulation 2 in Figure 4.12a, and can already conclude that there are issues with the full calibration framework.

We present corresponding noise for the chosen large and small credible intervals in Figures 4.14a and 4.14b and 4.13a, 4.13b.

Resulting posteriors are presented in Figures 4.15a-d) for the small CI's, and Figures 4.16a-d) for the big CI's. We observe no mixing issues when comparing the posteriors from running the chain longer to the ones obtained from the shorter chain. The only difference seems to be that the posteriors become smoother. What is interesting, however, is that the MAP estimate of the marginal variance σ_δ is much larger for the big CI's in Figures 4.16a-d) compared to the small CI's in Figures 4.15a-d). Hence, we can conclude that the instability in the credible intervals for the first modular framework is not caused by mixing issues. This also holds for the full calibration when we investigated credibility intervals of the same size or larger than the first modular approach. However, the smallest credibility intervals observed in Figure 4.12a showed large mixing issues.

4.4 Results Simulation Study 4: Computer Model Discrepancy

We have performed simulation studies to investigate the effect of increasing the discrepancy. Posterior distributions of simulating from one data-set with the same fixed independent noise for all experiments are depicted for both modularization approach 1 and 2 in the Figures 4.18 and 4.19. Here it remains clear, as for the previous results of simulation study 2, that C incorporates the increased discrepancy as uncertainty in its posterior for both modular

(a) Simulation Study 3.4 -
Full Calibration

(b) Simulation Study 3.5 -
Mod 1

(c) Simulation Study 3.6 -
Mod 2
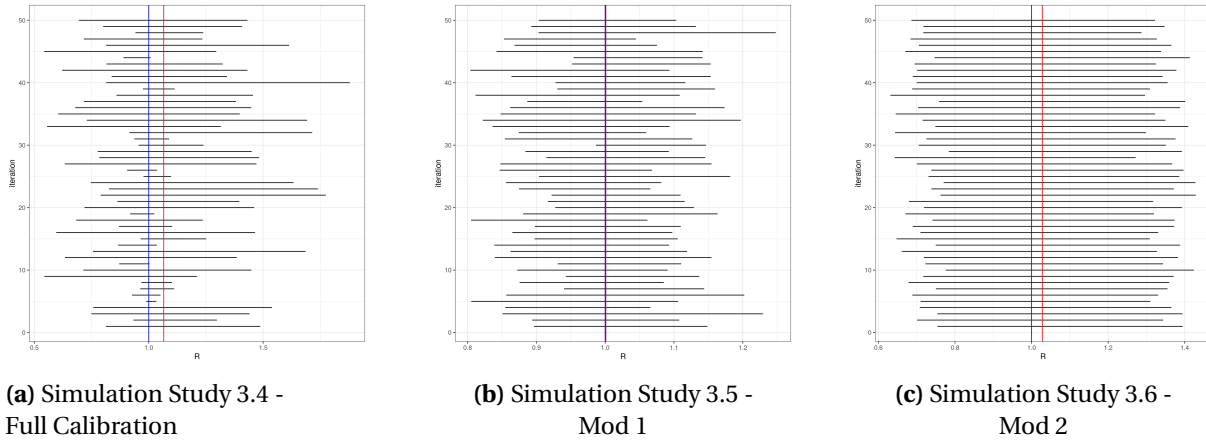
Figure 4.12: 90% credible intervals of R from 50 simulated noisy data-sets with dependent noise and length scale 0.8 ϵ_t. Blue line is true parameter value, red is average MAP estimate.

(a) Full Calibration:
simulation 2 from Figure 4.12a

(b) Mod 1:
simulation 38 from Figure 4.12b
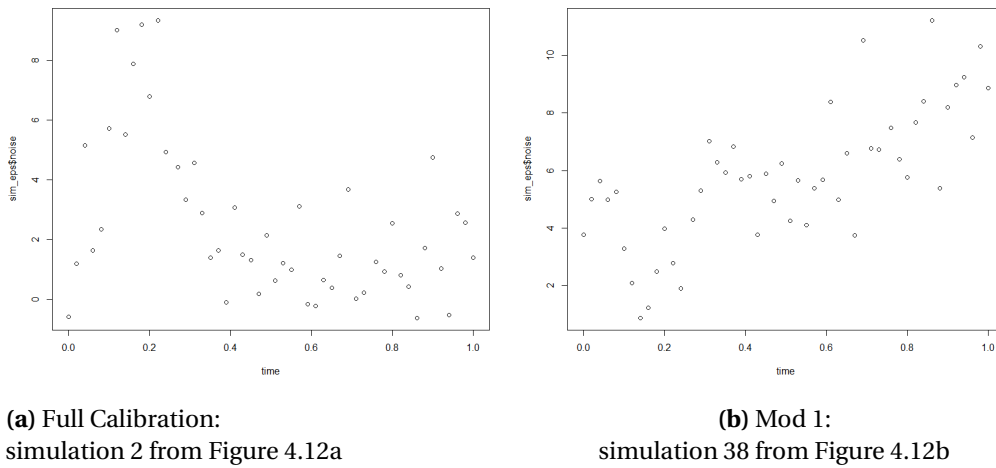
Figure 4.13: Simulation Study 3: Corresponding dependent noise ϵ_t to small credible interval.

(a) Full Calibration:
simulation 40 from Figure 4.12a

(b) Mod 1:
simulation 50 from Figure 4.12b
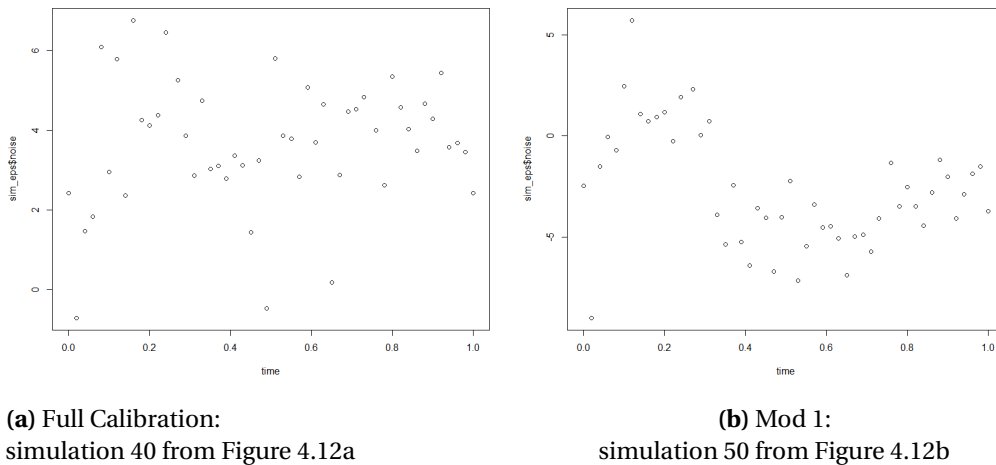
Figure 4.14: Simulation Study 3: Corresponding dependent noise ϵ_t to big credible interval.

frameworks. The reason for C taking on the uncertainty can be explained by its role as the parameter that controls the amplitude for the two-element Windkessel model.

In Figure 4.17, however, the full calibration yields increased uncertainty in the posteriors for σ_δ and ρ_2 as well as growing uncertainty in the posterior for R. It is reasonable that R incorporates the increased discrepancy resulting from an increase in Z. This is because Z is interpreted as a resistive element, and the sum of R and Z make up the total vascular resistance when the two-element Windkessel is fitted to the three-element Windkessel model. Nevertheless, when compared to the two modular frameworks, we would have expected C to incorporate a larger amount of uncertainty compared to R in the full calibration framework too.

4.5   Convergence of MCMC

To ensure convergence, we performed a test by running 4 MCMC chains on the same generated data-set with different initial values and observed whether the mixing plots and posteriors were overlapping. This was conducted for each simulation study and inference model. All chains showed signs of convergence; therefore, we only present two cases in this Chapter. We have chosen simulation study 1.1, with independent noise in simulated observations, and the second one from simulation study 2.8, with dependent noise.

Figure 4.20 shows good mixing for simulation study 1.1, where we do not estimate the discrepancy. Furthermore, Figure 4.21 depicts good mixing for simulation study 2.8. Note the large uncertainty in the mixing for parameter C, which is reflected in the average credibility intervals presented in Figure 4.7.
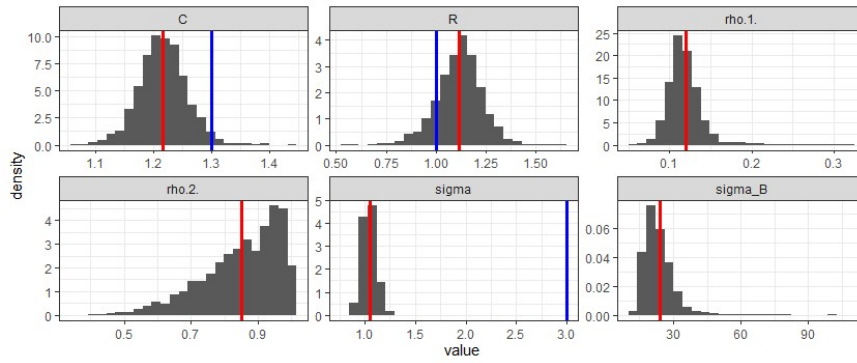
(a) Simulation Study 3.4 - Full Calibration
Small CI - short chain
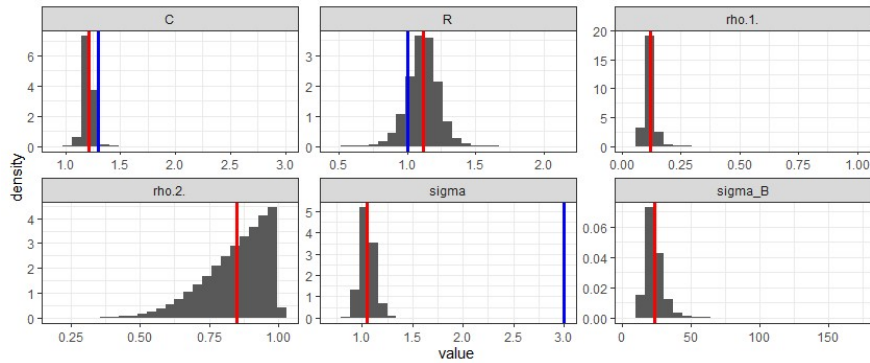
(b) Simulation Study 3.4 - Full Calibration
Small CI - long chain
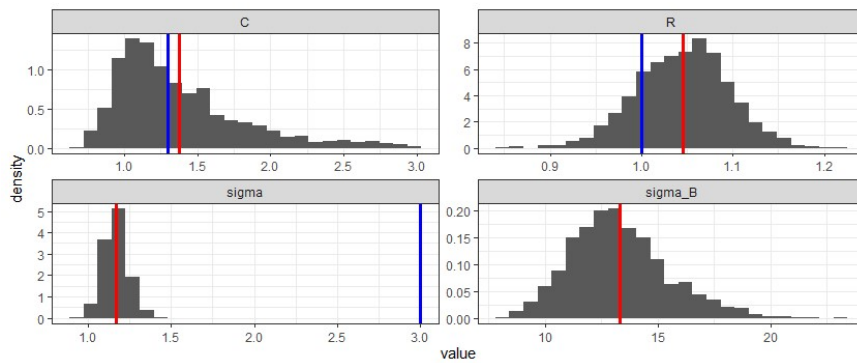
(c) Simulation Study 3.5- Mod 1
Small CI - short chain
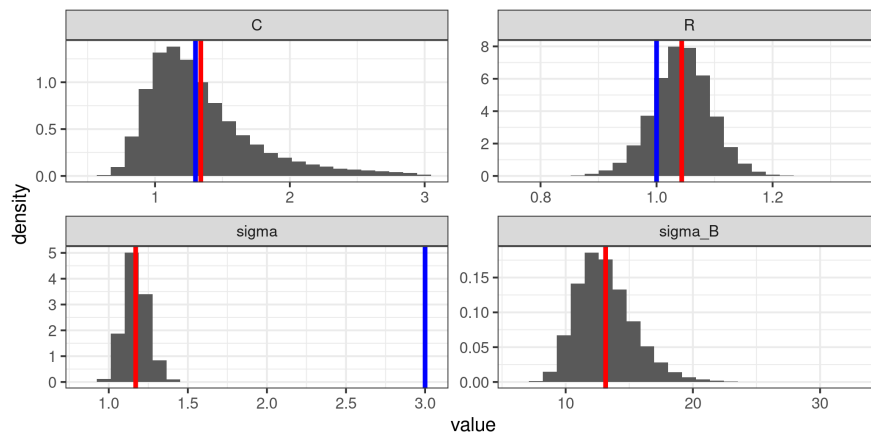
(d) Simulation Study 3.5- Mod 1
Small CI - long chain

Figure 4.15: Simulation Study 3: Resulting posteriors for running the chain short and long for small credible intervals.
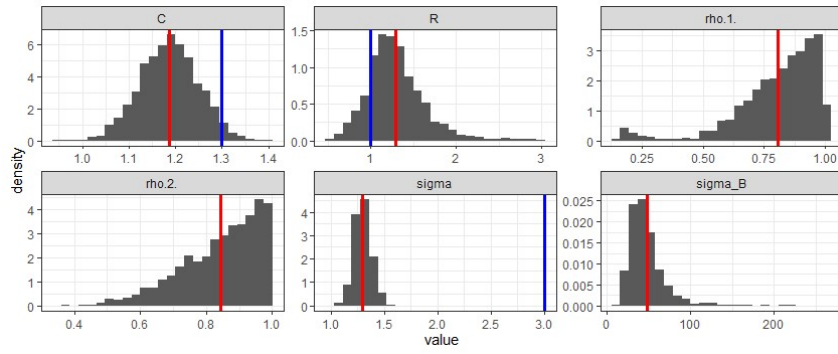
(a) Simulation Study 3.4 - Full Calibration
Big CI - short chain
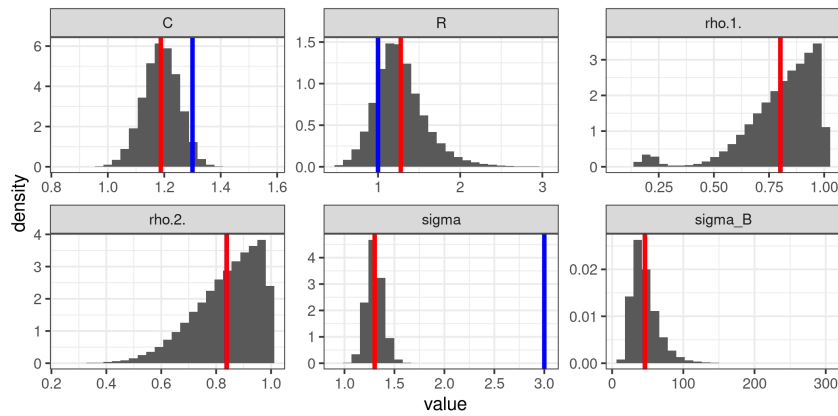
(b) Simulation Study 3.4 - Full Calibration
Big CI - long chain
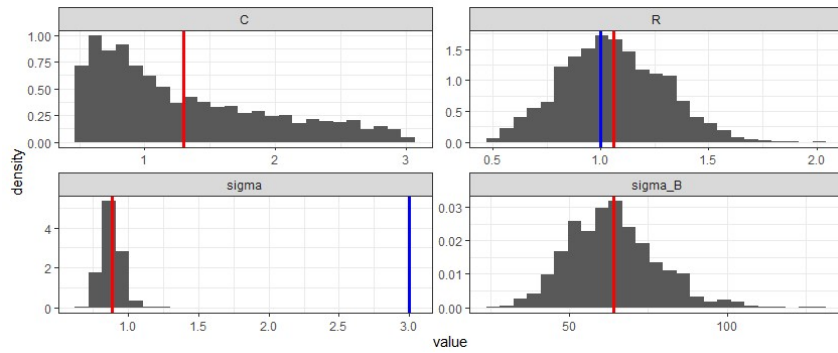
(c) Simulation Study 3.5- Mod 1
Big CI - short chain
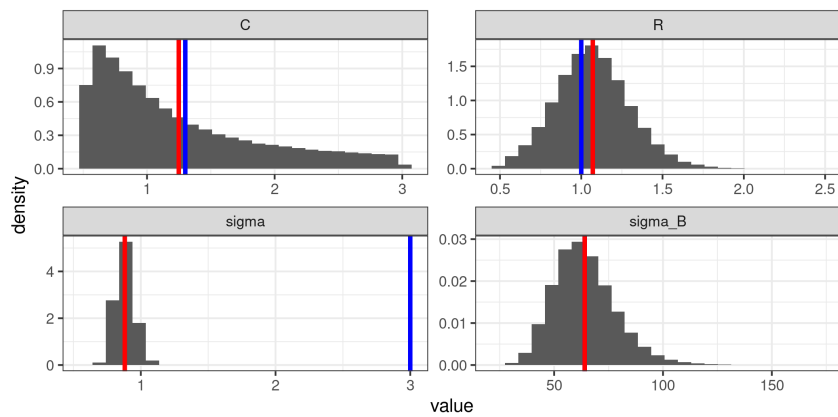
(d) Simulation Study 3.5- Mod 1
Big CI - long chain

Figure 4.16: Simulation Study 3: Resulting posteriors for running the chain short and long for large, credible intervals.
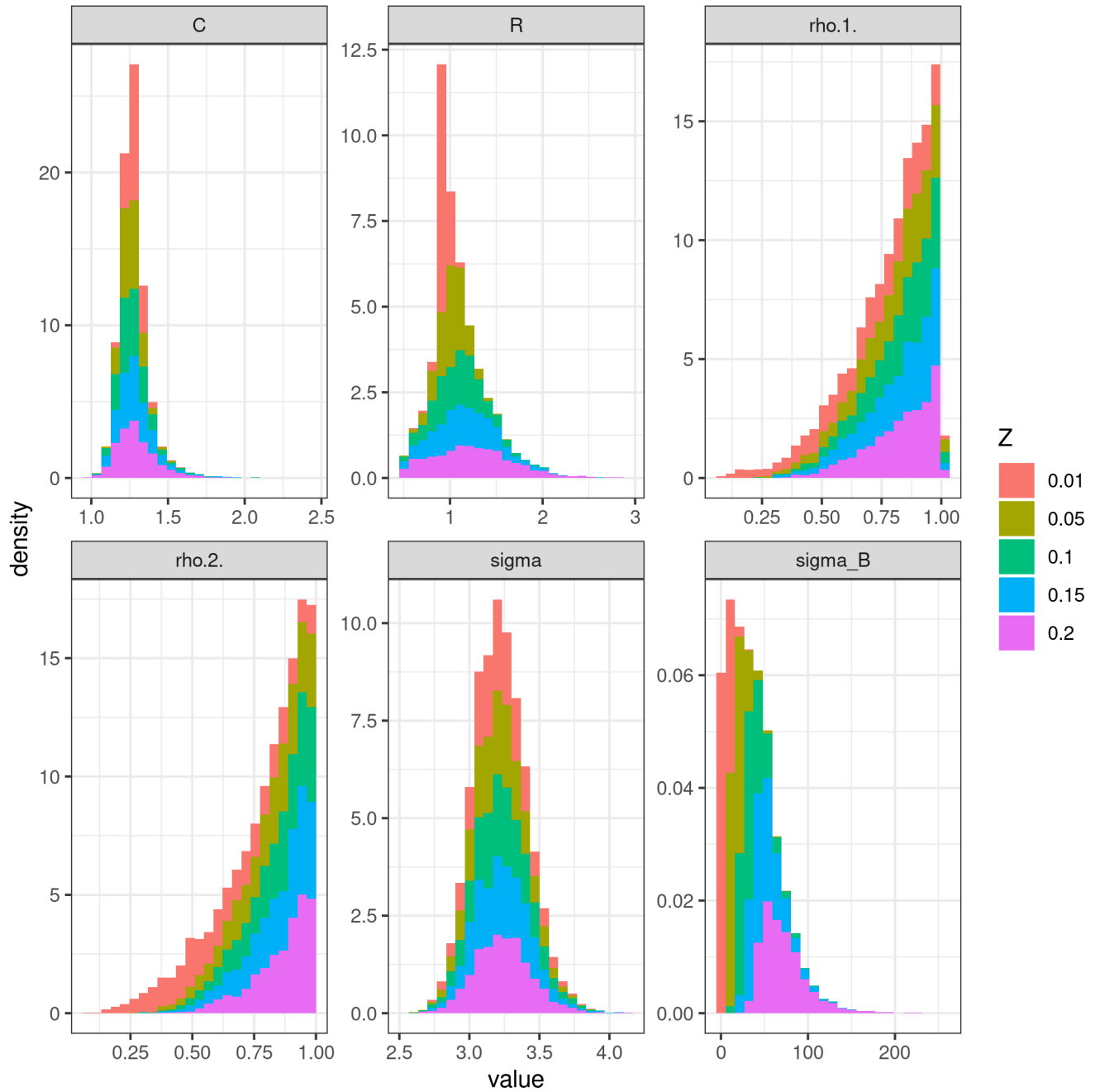
Figure 4.17: Simulation Study 4.1: Posterior distributions from simulating from one data-set with independent noise and increasing Z in the full calibration framework.
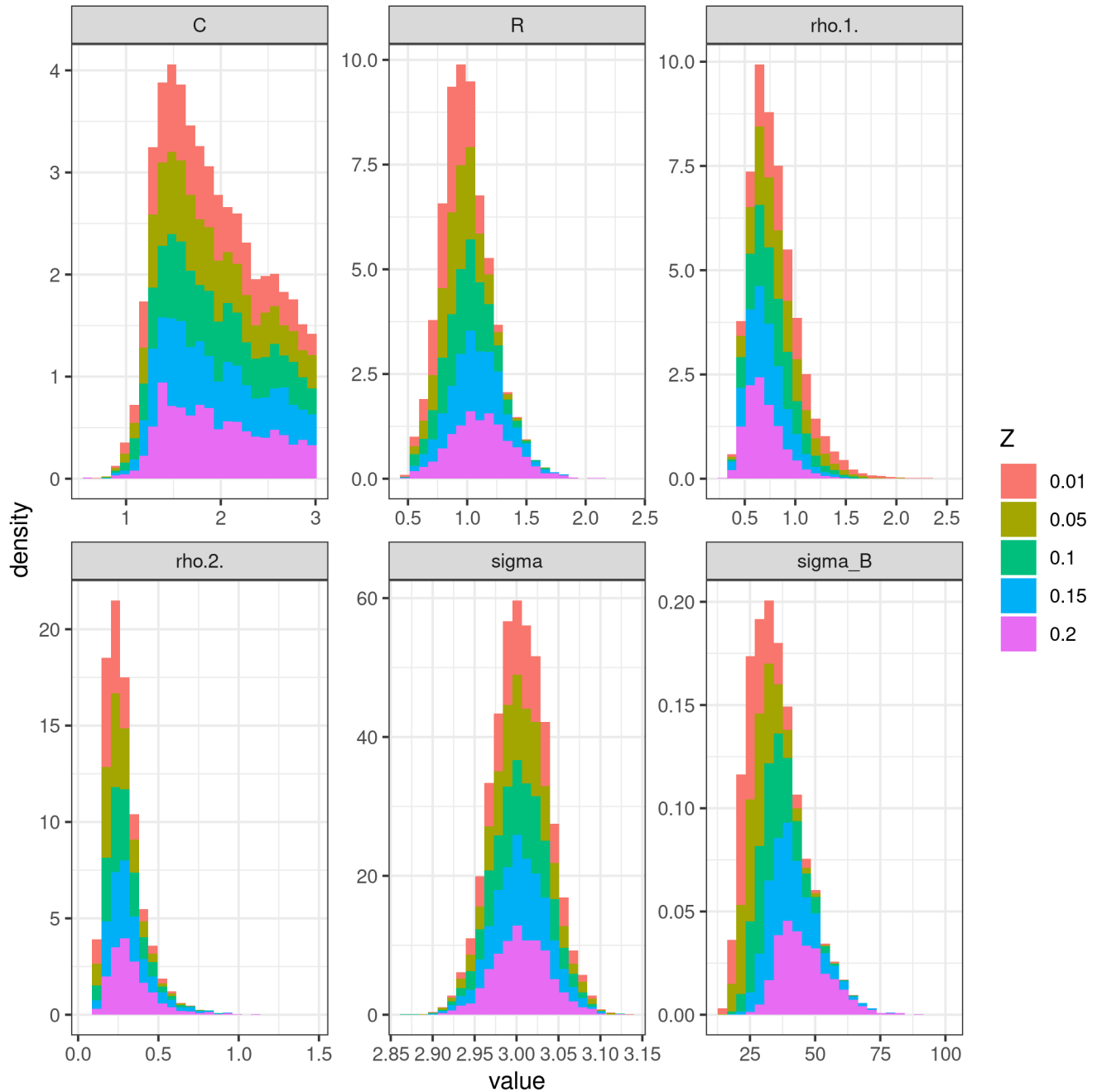
Figure 4.18: Simulation Study 4.2: Posterior distributions from simulating from one data-set with independent noise and increasing Z in the first modular framework(Mod 1).
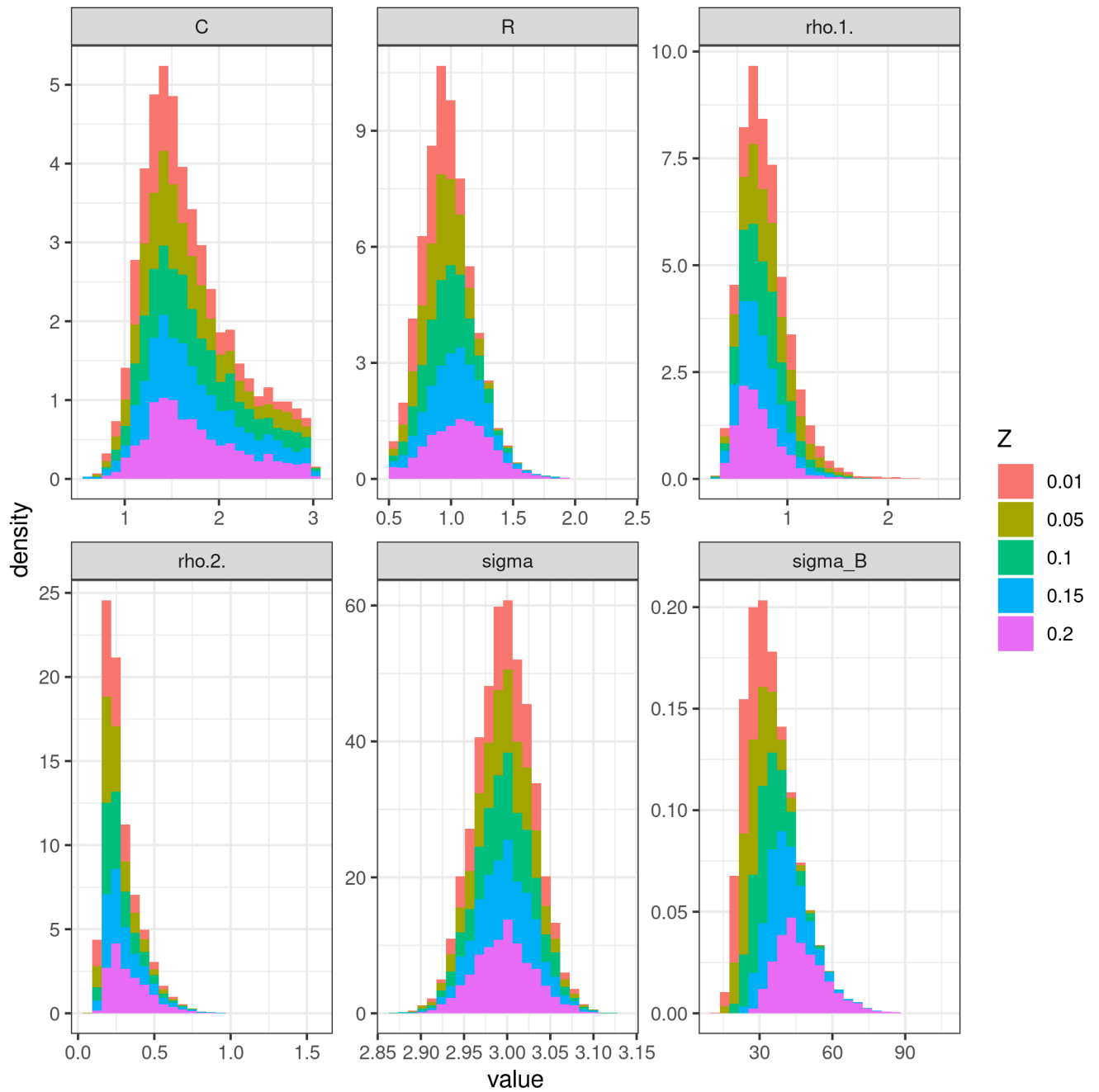
Figure 4.19: Simulation Study 4.3: Posterior distributions from simulating from one data-set with independent noise and increasing Z in the second modular framework(Mod 2)
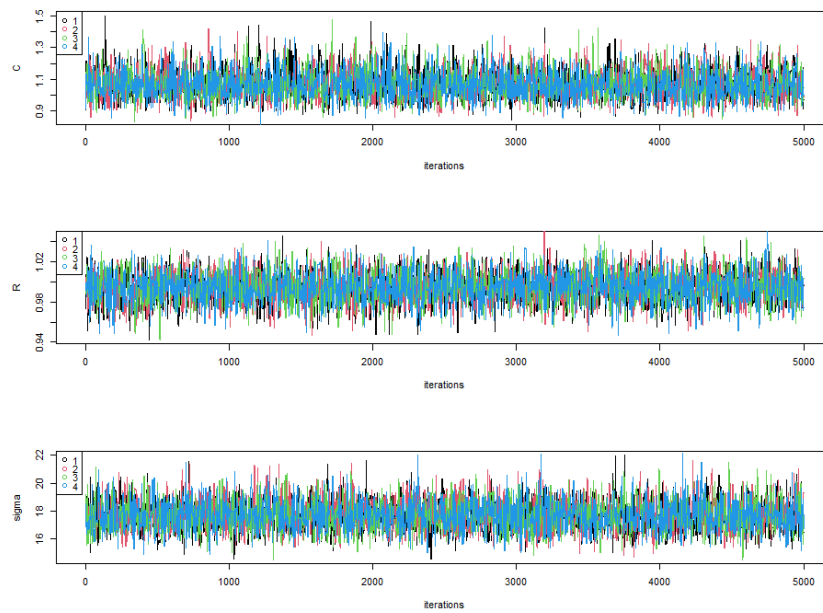
Figure 4.20: Mixing after conducting 4 chains with different initial values and one simulated data-set. Inference model is non-matching models with independent noise in simulation study 1.
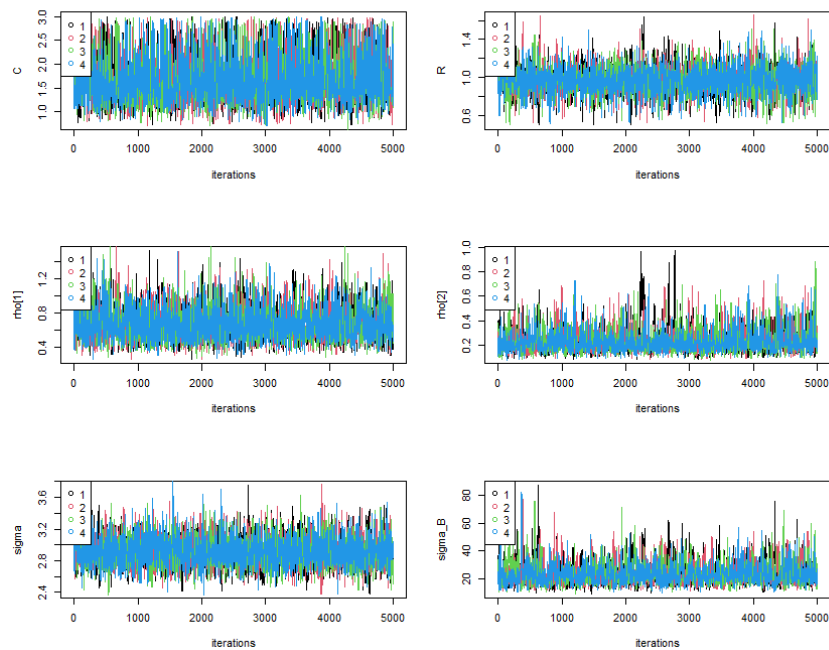
Figure 4.21: Mixing after conducting 4 chains with different initial values and one simulated data-set. Inference model is modular approach 2 and dependent noise in simulation study 2.

Chapter 5

Discussion and Conclusion

5.1 Discussion

We have presented a Bayesian calibration framework for the arterial Windkessel models. Our primary interest is to obtain estimates of two physically interpretable parameters (total arterial compliance, C and vascular resistance, R) and observe the behaviour of the estimates when subject to independent and dependent noise. We chose the two-element Windkessel model (2.1) as a computer model, and simulated synthetic field observations from the three-element Windkessel model (2.2) with added noise. Posterior distributions of the estimates of the physical parameters were obtained by performing Bayesian inference, both accounting for and not accounting for model discrepancy.

For all inference models we considered 100 noisy blood pressure simulations from the simulation model with both independent and dependent noise to evaluate the robustness of the estimates. To investigate the results, we compare the mean MAP and credible interval from the 100 simulations and assess them in combination with the total coverage of the true parameter value for each data-set.

When we performed inference on the models not accounting for the discrepancy, physical parameters R and C were well-calibrated when fitted to field observations generated with independent noise. On the contrary, the introduction of dependent noise was reflected in considerable variability in estimates and credibility intervals between each simulated data-set. Nevertheless, the mean MAP was able to approximate the true parameter value. Still, this is a flawed model because we in practice, want to estimate

the parameters using observations from one person, and the results are unreliable for one data-set. Furthermore, when we fit the two element Windkessel model to field observations generated from the three element Windkessel model without accounting for discrepancy, the measurement noise σ, acted inappropriately as a tuning parameter, shifting away from its true value to try to correct for model discrepancy. This was remedied by introducing Bayesian calibration.

Inference on the model accounting for discrepancy yielded non-identifiable posteriors when simultaneously estimating the discrepancy and model. Therefore, the need for modularization was imminent. When the simulation model was coupled with independent noise to generate field observations, we obtained well-calibrated posteriors from fitting the model using fixed values of the length scales or setting priors according to the posteriors from fitting the discrepancy. In the case of dependent noise, however, the framework with estimated priors for the length scales proved to be the only method robust enough to yield well-calibrated results. This framework also included the most informative priors for the σ's. Hence, it demonstrates the importance of choosing suitable priors when incorporating model discrepancy in statistical inverse problems.

The highly informative priors we used for the observational noise σ can be argued as restricting the applicability to situations where we have to know the actual value of σ. However, it is realistic to obtain such information even when we do not initially know the true value. This is because when conducting a modular approach, we get knowledge of the range of the observation noise when fitting the discrepancy. Therefore we have available information to perform this limitation in future experiments.

The full calibration and first modular approach with fixed values for the discrepancy length scales yielded irregular credibility intervals for the simulations with dependent noise. Therefore, we conducted simulation studies to investigate the impact of different dependencies in the noise. The full calibration framework showed severe mixing issues. On the contrary, the first modular framework with fixed length scale values depicted no mixing issues, but the results indicated an overfitted bias model. For the first modular approach, we have estimated and fixed the length scale of the discrepancy to $\rho_2 = 0.07$. Hence, we observed that as the length scales go towards zero, the effect of the discrepancy goes to the marginal variance σ_δ and the physical parameters.

Furthermore, we investigated the impact of increasing the discrepancy, which resulted in posteriors with large uncertainties for both physical parameters. [Brynjarsdóttir and O'Hagan [7]] state that the challenge with incorporating model discrepancy in statistical inverse problems is being confounded with calibration parameters, which will only be resolved with meaningful priors. In other words, to reduce uncertainty in the case of large discrepancies, the framework proposed in this thesis needs additional information.

We wanted to keep the experiments within a controlled environment and used the same simulated flow as input for all experiments. Further, we assumed no noise in measured flow for all simulation studies, as integrating this into the models was beyond the scope of this thesis. Nevertheless, exploring other inflow curves is an interesting topic for future investigation, as noise in flow will cause issues with uncertainty in inference.

Furthermore, using the three-element Windkessel model as the real model is not realistic because real pressure data obtained from a person might follow a much more complex model. We use the three-element Windkessel, as it can yield results not far from reality. However, most importantly, it is used to investigate properties that can be hard to deduce from an even more complex model.

An important note is the speed of the algorithm. Estimating physical parameters from one person's flow data using the adaptive MCMC and modular framework takes only mere seconds, which scales well in a digital twin setting.

5.2   Conclusion

The difference in inference for independent and dependent noise shows itself in the stability and uncertainty in the posteriors. Furthermore, we successfully obtain estimates of the physical parameters using the Bayesian calibration framework without an emulator. Coupled with highly informative priors we obtain robust posteriors, nevertheless at the cost of an increased uncertainty for the physical parameters compared to inference models without estimating the discrepancy.

Chapter 6

Appendix

6.1 Results Simulation Study 1: Independent noise

6.1.1 Results For One Simulated Data Set

Figures 6.1a, 6.1b and 6.1c display the resulting posterior distributions from simulation studies 1.1,1.2 and 1.3 with independent noise ϵ. Apart from the posterior for σ for simulatoin study 1.3, the results yield overall well calibrated parameter estimates.

Figure 6.2a, 6.2b and 6.2c represent the same posterior distributions as Figures 6.1a, 6.1b and 6.1c, but with the corresponding correlation between the parameters. Observe for simulation study 1.2 in Figure 6.2b, how R and Z are negatively correlated. This corresponds to theory, where we know that for the three element Windkessel, total arterial resistance becomes the sum of R and Z. In the same Figure, one can also observe a slight positive correlation between C and Z.

6.1.2 Results For Simulation Study

The reported 90% credible intervals of the 100 simuulations of posterior of C is plotted in Figures 6.3a and 6.3b for simulation study 1.1 and 1.2. Here, one can observe the resulting 94% coverage of the true mean for the two element Windkessel, and an 88% coverage for the three element Windkessel. Similarily, Figure 6.4a depicts a 92% coverage of the true mean for R and Figure 6.4b and 87% coverage. Whereas Figure 6.5a depicts a 86% coverage of the true mean for σ and Figure 6.5b and 89% coverage.

We move on to simulation study 1.3. Here the parameters C and R, have a
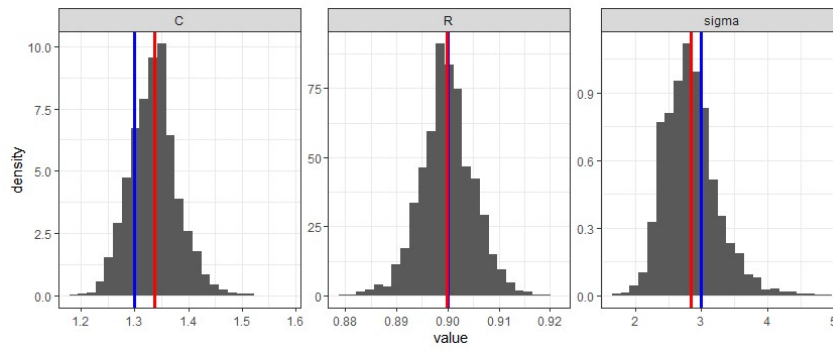
(a) Simulation Study 1.1: WK2 - iid noise ϵ
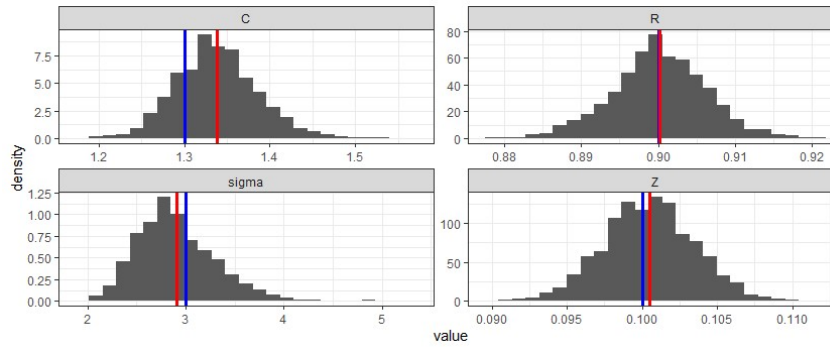
(b) Simulation Study 1.2: WK3 - iid noise ϵ
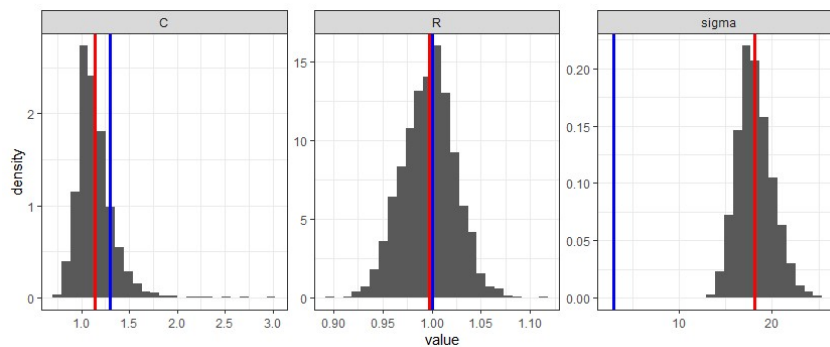
(c) Simulation Study 1.3: iid noise ϵ

Figure 6.1: Posterior distributions for calibration parameters and standard deviance σ from one simulated data-set and independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.

coverage of 100%, displayed by Figures 6.3c and 6.4c. In contrast to simulation study 1.1 and 1.2, where we observe narrower credible intervals, for simulation study 1.3 they look almost identical for each simulation. Looking closer, C in Figure 6.3c almost spans the prior entirely with a range $\in [0.8; 2]$. This is most likely a result of the model discrepancy being introduced and not accounted for, consequently introducing more uncertainty in the estimates. At first glance, one could expect the same for R in 6.4c, but here we observe the span of the credible intervals being $\in [0.95; 1.05]$, hence very narrow credible intervals. In addition to this, the

overestimation of sigma in Figure 6.10c indicates that the variable is inaccurately behaving as a tuning parameter to fix problems with the model.

6.2 Results Simulation Study 1: Dependent noise

6.2.1 Results For One Simulated Data Set

We present the posterior distributions after simulating from one data-set with dependent noise ϵ_t in Figures 6.6a, 6.6b and 6.6c. Observe a similar behaviour for all posteriors of C compared to the case of independent noise. R on the other hand, now depicts a biased estimate, nevertheless on a small scale. In addition, the posterior distribution for σ is again shifting to the end of the range of its prior for the non-matching models, becoming highly overestimated. Furthermore, correlation between parameters in Figure **??**, **??** and 6.7c depict an almost identical behaviour as for the results of simulating from one data-set for independent noise.

6.2.2 Results For Simulation Study

Figure 6.8a, 6.8b and 6.8c display the resulting 90% credible intervals of C for 100 independent simulations with dependent noise. Compared to the independent noise case, we observe much narrower credible intervals for the matching models. This implies less uncertainty in the estimate itself, however in Figure 6.8a the coverage of the true value lies around 40%, whereas for 6.8b it is a bit more than 50%. Hence, all over a less coverage of the true mean for the matching models when compared to the independent noise case, but the average of the posterior mean remains more or less unbiased. This implies that in the presence of dependent noise the resulting posteriors when simulating from one data-set are uncertain, and might present bias. Nevertheless, reasonable estimates can be obtained from the average MAP estimate after several simulation runs. To the contrary, when we look at the result from the non-matching models presented in Figure 6.8c, there is more similarity with the result from the independent noise case. Here coverage is all the way up to 100% and posteriors are spanning a large chunk of the prior. This implies more uncertainty in the posterior itself.

Both Figures 4.3a and 4.3b for the 90% credible intervals of R, show behaviour in accordance with the results of C for matching models with

dependent noise, i.e. considerably more narrow credible intervals for each simulation. The coverage of the true mean is less than 20 % for both the two and three element Windkessel. This is an even larger drop in coverage than the C estimates when compared to the independent noise case, and underlines the care needed when using parameter estimates from only one simulation run in the presence of dependent noise. However, the span of the values is much more narrow compared to C, where we have a range of values between $\in [1.1; 1.6]$, but for R the range lies between $\in [0.84; 0.96$, and $R_{true} = 0.9$. Figure 4.3c shows a similar behaviour to the non-matching models with independent noise, but not the same identical shape of the credible intervals and the coverage is reduced to 98%. In addition, the span of the credible intervals are slightly larger. (from $\in [0.92; 1.05]$ to $\in [0.9; 1.10]$

As the measurement noise σ is no longer the only parameter defining the kernel, which Equation 3.2 describes, it can explain the underestimation for the matching models depicted in Figure 6.10a and 6.10b. The measurement noise still has the same true value of $\sigma = 3$, but the coverage of the true mean of the credible intervals is around 27% for the two element Windkessel. It is further reduced for the three element Windkessel in Figure 6.10b to a coverage around 17%. For non-matching models however, the discrepancy is again overruling the estimate, hence the coverage is similar to the non-matching models with independent noise, with 0 % coverage of the true value and an average MAP estimate around the same value as for independent noise.

6.3   Results Simulation Study 2: Independent noise

We first present the results from fitting the discrepancy alone, which forms the baseline for the two modular approaches. Then we present results from the full calibration, where we infer all parameters simultaneously, and then results from the two modularization strategies for one simulated data-set and for 100 simulated data-sets with independent noise .

True values of the discrepancy parameters ρ_1, ρ_2 and σ_δ are not known, hence we report MAP estimates from the posterior distributions and credible intervals.

6.3.1   Results From Fitting The Discrepancy

The resulting posteriors from estimating the discrepancy form the baseline for the modular approaches one and two and are depicted in Figure 4.6a. The correlation between parameters are shown in Figure 4.6b, where all correlation coefficients are of high significance(the larger number of $^{*\prime}s$ behind coefficient, the more significant), but of moderate size except for the one between length scale ρ_1 and σ_δ. The first modular approach utilizes the MAP estimate of the length scale parameters, whereas the second modular approach takes the limits of the posterior and forms a prior for the length scales.

6.3.2   Results For One Simulated Data Set

Full Calibration

We present the results from the full calibration (simulation study 2.3 in Table **??**) with independent noise ϵ. As presented in 3.5.2 We have conducted the full calibration with two different uniform priors for the length scales.

First we have the results from conducting the simulations with a uniform prior between 0 and 1 in Figure 6.12a. The MAP estimates of physical parameters R and C show a slight overestimation when compared to the true mean, and there is uncertainty in the long tails of both posterior distributions. The observation noise σ shows a slight underestimation, nevertheless a well calibrated posterior compared to the results from Simulation Study 1 with non-matching models, where σ is heavily overestimated.

To the contrary, we can observe that the posteriors for the length scales are accumulating at the end of the range of their priors, indicating that they have not converged. Due to our input covariates time and flow being transformed to a 0-1 scale, there should not be information in the data pushing the length scales to larger values than 1. Out of curiosity, however, we conduct the experiment with a larger range in the priors to observe if the posteriors of length scales do converge at one point. Nevertheless, correlation between parameters are moderate in Figure 6.12b, which shows relatively small values. Note, however, how the correlation between C and allother parameters are reported as highly significant, with three stars * behind. This level of significance is also reported for the correlation between the length scales, and ρ_1 and σ_δ.

The resulting posteriors from increasing the range of priors for the lengths scales are presented in 6.13a. While the length scale ρ_1, corresponding to flow, seems to accumulate somewhere below 50, ρ_2 corresponding to time does not seem to converge, not even within a range of 0 to 5000. This is interesting, as mentioned above, there is no information in the data that would draw values of the length scales up to such extremes. In addition, allowing ρ_1 to explore a larger posterior has resulted in an increase correlation with σ_δ as presented in Figure 6.13b. Most importantly, the results yield less accurate posteriors in Figure 6.13a for the physical parameters, which are the main stars of the simulation studies. Furthermore, we observe an increase in correlation between R and C in Figure 6.13b. Therefore, we conclude that we continue the full calibration approach for the rest of the simulation studies with a uniform prior for length scales between 0 and 1, and turn to the modularization framework to be able to identify posteriors for the length scales.

Mod 1: Fixed Values For Length Scale Parameters

We present the results from estimating the parameters from one data-set and utilizing the first modularization approach, where we fix the length scales to the MAP estimates of the posteriors from fitting the discrepancy. The resulting posteriors are presented in Figure 6.14a. Here C looks to be a bit overestimated, but what indicates uncertainty in the posterior is the long tail, which looks to span the whole prior range between [0,5;3]. R is slightly over estimated, but with a well calibrated posterior. In addition σ is also slightly over estimated, but with the true value contained in the posterior. Furthermore, we observe very small correlation between parameters Figure 6.14b, where all correlation coeficients are reported as highly significant except for the ones between σ and the physical parameters. Noteworthy would be the correlation between C and σ_δ which is considerably larger if we compare it to the correlation with other parameters, and might be a factor causing the uncertainty in the posterior for C.

Mod 2: Priors For Length Scale Parameters

We present the results simulating from one data-set with the second modular approach of taking the estimated posteriors from fitting the discrepancy as priors for length scales ρ_1 and ρ_2. The resulting posteriors in Figure 6.15a show a slight overestimation of R, and a slightly large range in the posterior. C

on the other hand is overestimated, with a large tail in its posterior, indicating uncertainty in the estimate. Furthermore, the length scales are larger than when we conducted the fitting of the discrepancy alone. In addition the MAP of σ_δ has increased from 15 for the first modular approach to 25 here in the second modular approach.

We again observe moderate degrees of correlation between parameters in Figure 6.15b. Noteworthy is, that again compliance C is significantly correlated with all the other parameters. Furthermore, the two length scales are significantly correlated with the scaling parameter σ_δ of the discrepancy, in line with what we observed when we fitted the discrepancy alone. Compared to the results in Figure 4.6a, however, this correlation seems to be reduced.

Matching Models

We finalize the simulations with one data-set and independent noise with presenting the results from conducting full calibration on the matching models as presented in Table 3.1.

Performing Bayesian calibration when there in theory is no discrepancy present, yields identifiability issues, and both Figure 6.16a and 6.17a illustrate this for the discrepancy parameters. Both the two and three element Windkessel model look like they are in fact estimating a discrepancy that does not exist, as we obtain a posterior for σ_δ with a MAP around 3 and the length scales around 0.5. The physical parameters, however, are close to unbiased in their MAP estimate with small uncertainty in the posteriors.

6.3.3   Results For Simulation Study

We now move on to presenting the results from simulating the posterior from 100 synthetically derived noisy data-sets. We compare the full calibration with the two modular approaches in terms of the 90% credible intervals total coverage of the true mean, as well as the stability of the CI's between each run.

We commence with the 90% credible intervals of C presented in 6.18a, 6.18b and 6.18c. The average MAP estimate becomes more and more overestimated as we move from full calibration, the first modular approach and finally to the second modular approach. Furthermore, the credible

intervals for the two modular approaches seem more stable, but in line with the results from one data-set, the uncertainty is much larger compared to the full calibration. Looking at the full calibration in Figure 6.18a, we observe narrower credible intervals. Nevertheless, we observe some sudden jumps resulting in large spread of the CI. The reason for this is unclear, but could be due to the fact that we simulate noise randomly for each data-set. Sudden large simulated noise might cause sudden large uncertainties in the posteriors.

The results for the resistance R is presented in Figures 6.19a, 6.19b and 6.19c. Here, the CIs seem to be stable for each data-set. Nevertheless, the least results are presented for the full calibration, with an overestimated average MAP as well as the largest uncertainty in the resulting CIs. Average MAPs for both the modular approaches are nearly unbiased, but with the first modular approach showing slighly narrower Confidence intervals.

These are much more accurate and stable results when compared to the results for C, and might connected to the correlation plot we observed for the results when simulating from one data-set. Here C depicted significant correlation with all parameters, whereas R did not hold that same level of correlation, hence we observe the lower level of uncertainty in the Credible intervals. Furthermore, we know that C is directly linked to the amplitude of the pressure curve, and with a fitting of the two element Windkessel to the three element Windkessel, C is the parameter to make up for the shortcomings in the amplitude of the two element Windkessel.

The results for σ presented in Figures 6.20a, 6.20b and 6.20c are close to identical in both the stability and range of the credible intervals, as well as the unbiasedness of the average MAP estimate.

Results for both length scales are presented in Figures 6.22b, 6.22b and 6.23a, 6.23b for ρ_1 and ρ_2 respectively. As expected, their credible intervals are accumulating around 1 for full calibration, but remain stable for the second modular approach. It is interesting, however, that ρ_2 is systematically larger, with an average MAP 0.25 for the second modular approach compared to around 0.07 when we fitted the discrepancy alone.

Credible intervals for σ_δ are reported in Figures 6.21a, 6.21b and 6.21c. Here we observe that both the average MAP and range of credible intervals are at its largest of σ_δ for the full calibration. This might be caused by the posteriors for the length scale parameters not converging, but rather accumulating at

the end of the range of their prior at 1. To the contrary, the average MAP and CI's are at their lowest for the first modular approach, where we do not estimate, but fix the length scales. As we have observed, length scales and σ_δ are significantly correlated, and this might explain why the estimates are at their lowest for the first modular approach.

6.4   Results Simulation Study 2: Dependent noise

We now present the results from the full calibration, where we infer all parameters simultaneously, and then results from the two modularization strategies for one simulated data-set and for 100 simulated data-sets with dependent noise .

Again, true values of the discrepancy parameters ρ_1, ρ_2 and σ_δ are not known, hence we report MAP estimates from the posterior distributions and credible intervals.

6.4.1   Results For One Simulated Data Set

Full Calibration

In line with the results presented for the full calibration approach with independent noise, we have performed inference on the parameters with uniform priors between 0 and 1 for the length scales. Resulting posteriors are presented in Figure 6.24a. Compared to the case of independent noise, C shows an underestimated MAP value with a larger range in the posteriors, whereas R is overestimated with a tail indicating uncertainty in the posterior. Furthermore, σ is now very underestimated, while σ_δ looks almost identical to the independent noise. In addition, the posteriors for the length scales show the same behaviour of accumulating to the end of the range of their prior. As a result, we observe that the physical parameters and noise σ are the ones trying to correct for the dependent noise. Hence, full calibration is not a robust inference method in this case.

Mod 1: Fixed Values For Length Scale Parameters

The resulting posteriors from estimating the parameters with fixed length scales are presented in Figure 6.25a. Here, the uncertainty of the posterior for C is extreme, and the MAP estimate strongly overestimated. Surprisingly, R is very well calibrated, with an unbiased MAP estimate as well as a narrow

range in the posterior. The noise σ is very underestimated, whereas σ_δ has a slightly increased MAP estimate when compared to the posterior from the data-set with independent noise.

Mod 2: Priors For Length Scale Parameters

When we introduce dependent noise and estimate the parameters using the second modular approach we run in to some numerical issues. This is depicted in Figure 6.26a. Surprisingly the calibration of the physical parameters look good, but the sampling is barely conducted for most of the parameters of the discrepancy as well as the noise σ.

With an attempt of improving this, we modify the priors of the σ's as explained in 3.5.2 to be even more restrictive. Therefore we replace the uniform priors with an inverse Gamma distributions and conduct the experiments again. Results from one simulated data-set are presented in Figures 6.27a and 6.27b. Interestingly enough, at the cost of improving the calibration for the discrepancy parameters and noise σ, we have introduced more uncertainty in the posterior for R. Furthermore, the posterior for C still has an overestimated MAP estimate and a large tail that spans the range of the entire prior between [0.5;3].

Correlation coefficient and its significance between C and other parameters, however, is reduced compared to results from the full calibration and first modular approach for both dependent and independent noise. This is depicted in Figure 6.27a, where in fact correlation in general seems to be reduced.

6.4.2 Results For Simulation Study

We now present the results from simulating the posterior from 100 synthetically derived noisy data-sets with dependent noise. We compare the full calibration with the two modular approaches in terms of the 90% credible intervals total coverage of the true mean, as well as the stability of the CI's between each run.

The results for the estimation of compliance C are presented in Figures 6.28a, 6.28b and 6.28c. Observe for the full calibration how there is a very large variation in uncertainty for each data-set, as the width of the credible intervals go from unrealistically narrow to extremely wide. Nevertheless, the average MAP estimate is almost unbiased. For the modular approaches, we

have an overestimated average MAP estimate. The credible intervals are wider, and more stable than for the full calibration, however, the results for fixed length scales seem to be a bit more unstable than for the results where we estimate the length scales. In addition, overall the credible intervals show more uncertainty for the two modular approaches, and the MAP estimate is slightly more overestimated for the second modular approach.

Moving on to the results for total resistance R presented in Figures 4.9a, 4.9b and 4.9c. We again observe more unstable credible intervals for each data-set for the full calibration, and an overestimated MAP estimate. Furthermore, the first modular approach show the best estimate of the average MAP, which is nearly unbiased with little uncertainty in the CI's, however for R we also observe this sudden jumps in ranges of credible intervals as we did for C in the full calibration framework. For the second modular approach we have the most stable CI's, which look close to stationary for each data-set, and are narrower than for the full calibration, nevertheless wider than for the first modular approach.

Furthermore, we observe the results for the observations noise σ in Figures 6.30a, 6.30b and 6.30c. Here we observe a large underestimation for both the full calibration and first modular approach, which both have uniform priors for σ. In addition, we have the same behaviour very narrow credible intervals and sudden jumps of large uncertainty for the full calibration. The credible intervals for the first modular approach are also unstable from data-set to data-set, but remain narrow for all runs. For the second modular approach however, we an inverse Gamma distribution as a prior, which is strictly limiting the noise. This is reflected in 6.30c, where we observe stable and narrow credible intervals.

The results for σ_δ in Figures 6.31a, 6.31b and 6.31c show similar behaviour for the results of σ, with unstable credible intervals and large jumps for the full calibration. This is also represented for the the first modular approach, nevertheless on a smaller scale, and an average MAP estimate approximately around 24, whereas the second modular approach is stable around an average MAP of 34. This can be blamed on the more inverse Gamma distribution being a more robust prior when introduced to dependent noise, compared to a uniform distribution.

The instability of the full calibration is further presented for both length scale parameters in Figures 6.32a and 6.33a. Furthermore, the MAP estimate for

the second modular approach are larger than before, with $\rho_1^{\bar{MAP}} = 0.7$ and $\rho_2^{\bar{MAP}} = 0.3$

6.5 Results Simulation Study 3

6.5.1 Mixing Issues of The Full Calibration

We present results from conducting Simulation Study 3 with dependent noise generated with length scale 1 to demonstrate how the full calibration runs in to numerical issues when we run the chain for longer. This can be observed in Figure 6.40b
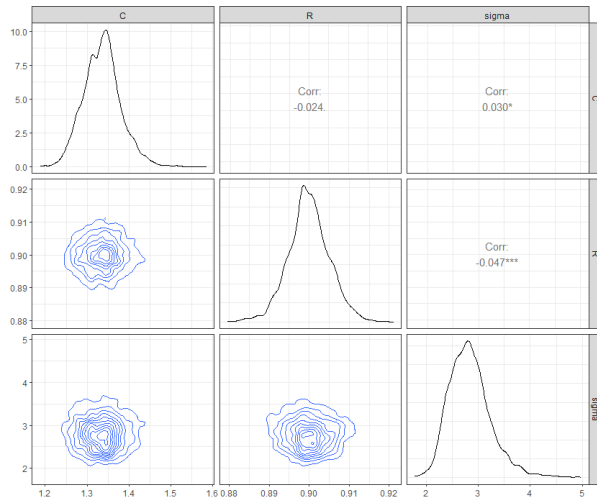
(a) Simulation Study 1.1: WK2 - Matching-models - iid noise ϵ
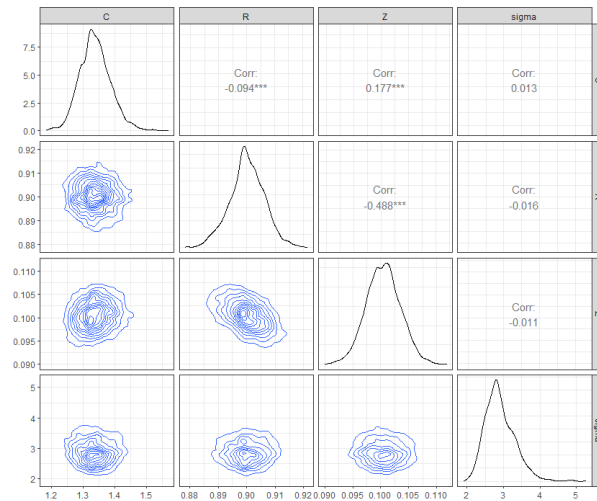
(b) Simulation Study 1.2: WK3 - Matching-models - iid noise ϵ
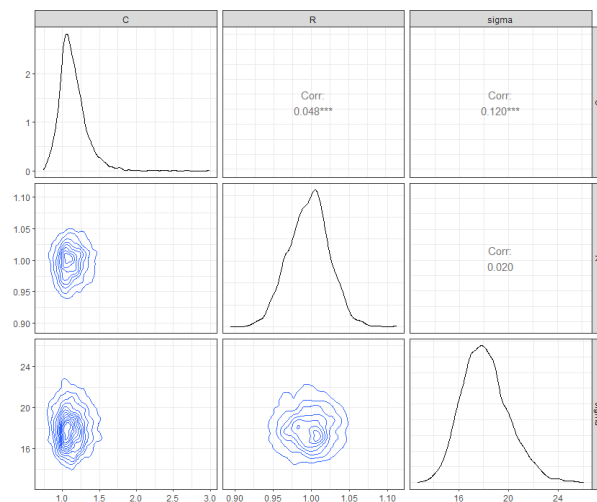
(c) Simulation Study 1.3: iid noise ϵ

Figure 6.2: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.
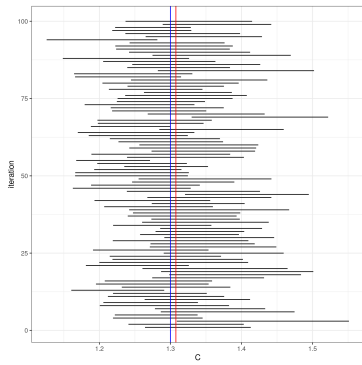
(a) Simulation Study 1.1: WK2 - Matching-models - iid noise ϵ
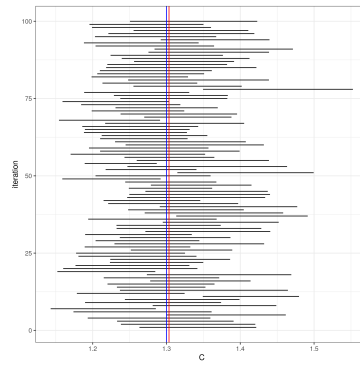
(b) Simulation Study 1.2: WK3 - Matching-models - iid noise ϵ
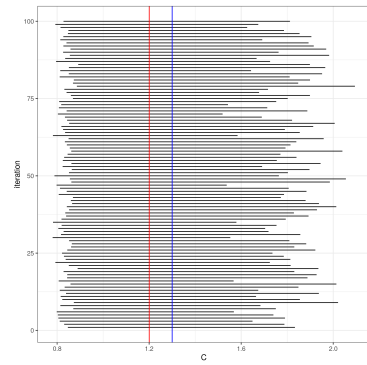
(c) Simulation Study 1.3: iid noise ϵ

Figure 6.3: Average MAP estimate with corresponding 90% credible intervals of C from 100 independent simulations of the posterior distribution with iid noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
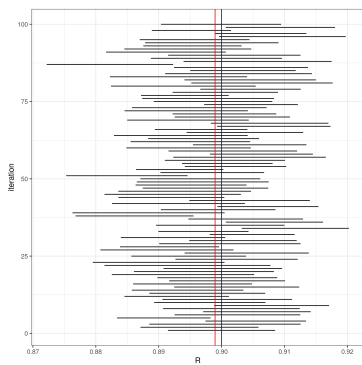

(a) Simulation Study 1.1: WK2 - Matching-models - iid noise ϵ
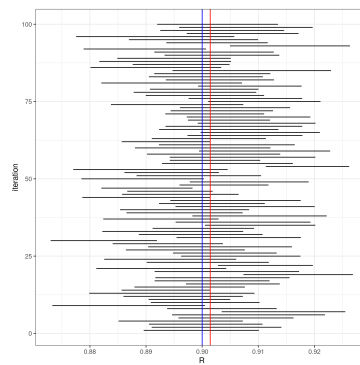
(b) Simulation Study 1.2: WK3 - Matching-models - iid noise ϵ
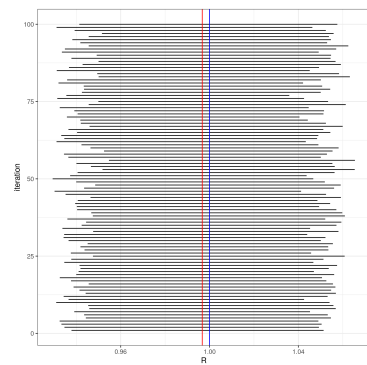
(c) Simulation Study 1.3: iid noise ϵ

Figure 6.4: Average MAP estimate with corresponding 90% credible intervals of R from 100 independent simulations of the posterior distribution with iid noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
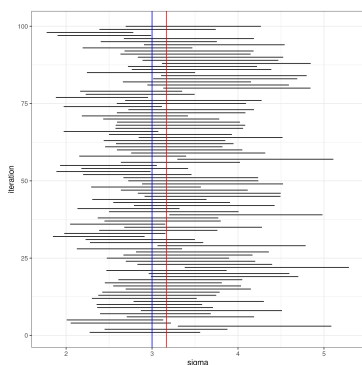
(a) Simulation Study 1.1: WK2 - Matching-models - iid noise ϵ
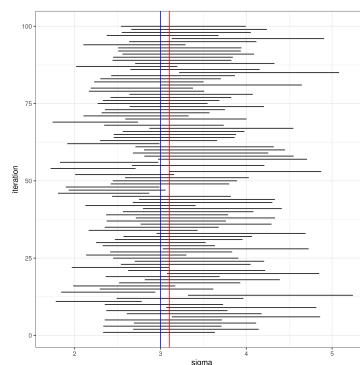
(b) Simulation Study 1.2: WK3 - Matching-models - iid noise ϵ
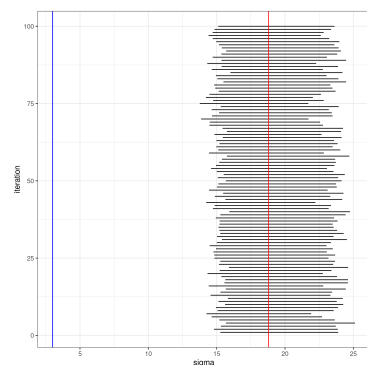
(c) Simulation Study 1.3: iid noise ϵ

Figure 6.5: Average MAP estimate with corresponding 90% credible intervals of σ from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
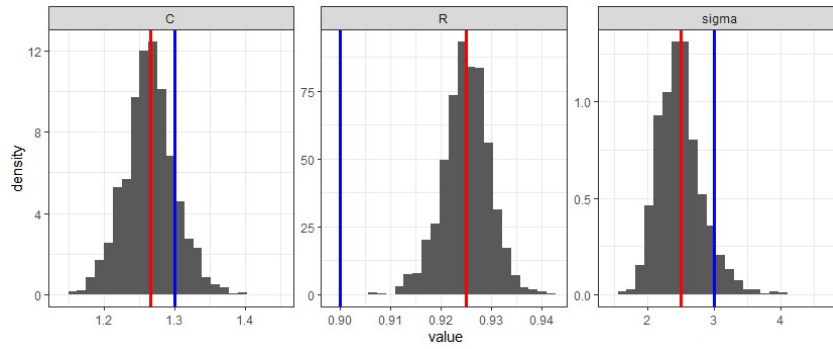
(a) Simulation Study 1.4: WK2 - Matching-models - dependent noise ϵ_t
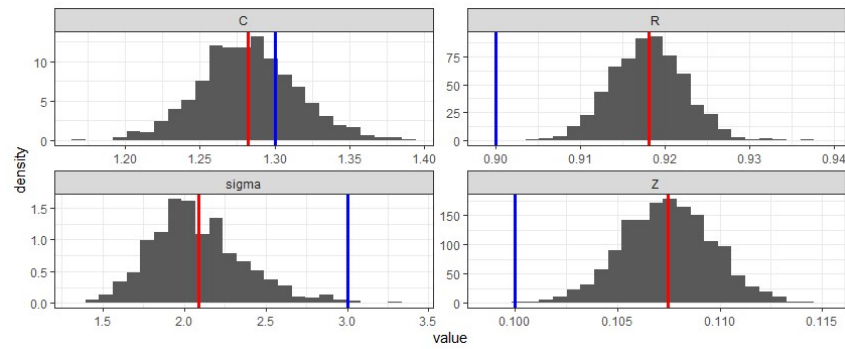
(b) Simulation Study 1.5: WK3 - Matching-models - dependent noise ϵ_t
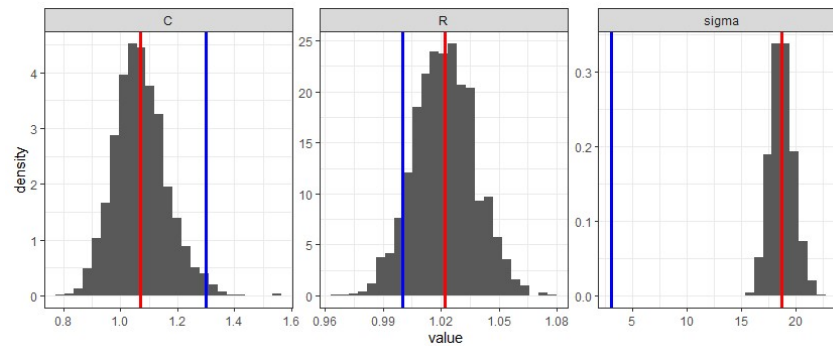
(c) Simulation Study 1.6: dependent noise ϵ_t

Figure 6.6: Posterior distributions for calibration parameters and standard deviance σ from one simulated data-set and dependent noise ϵ_t. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
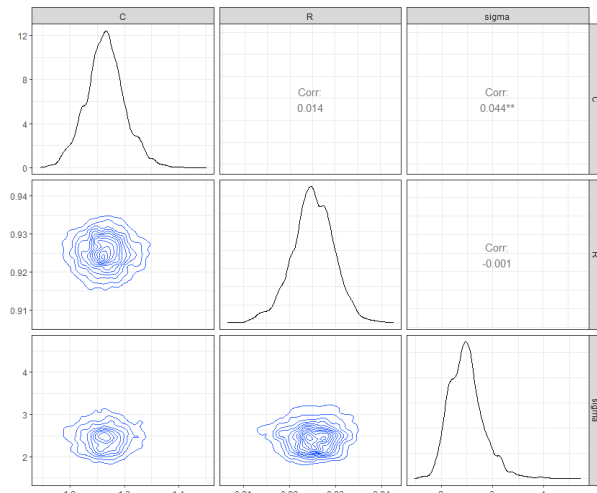
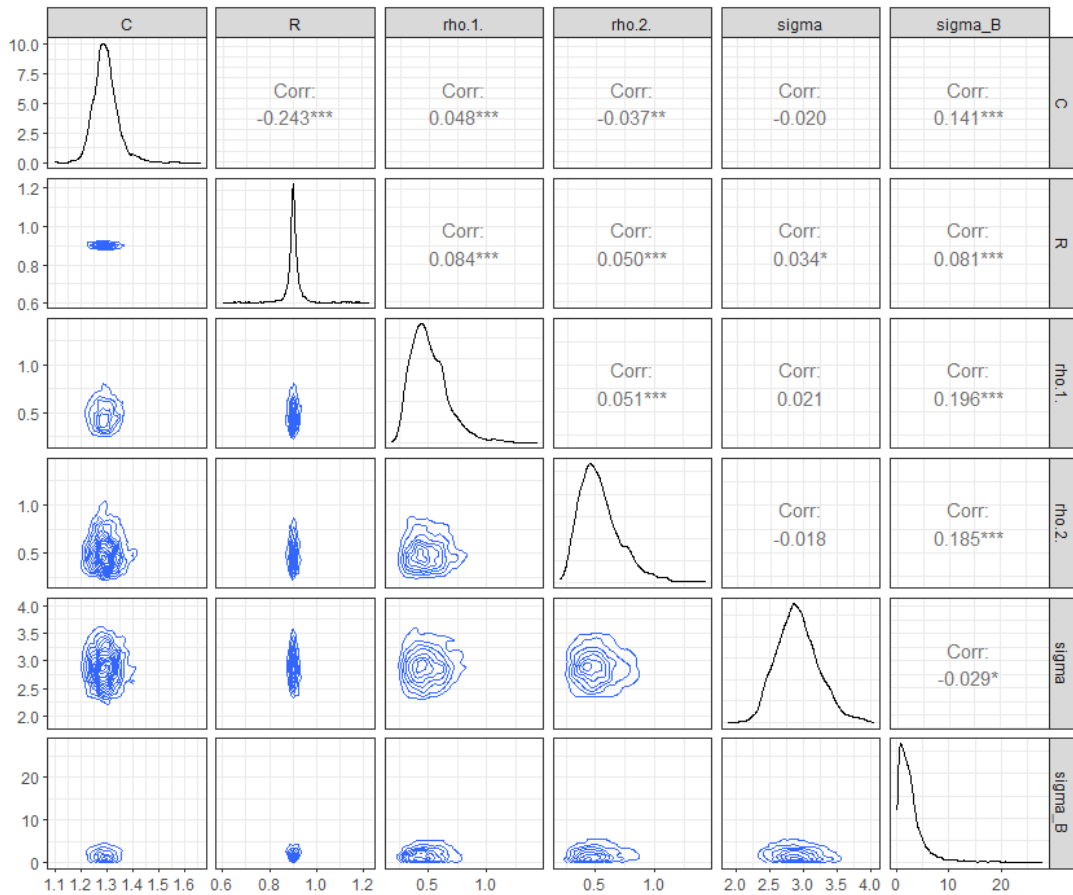**(a)** Simulation Study 1.4: WK2 - Matching-models - dependent noise ϵ_t
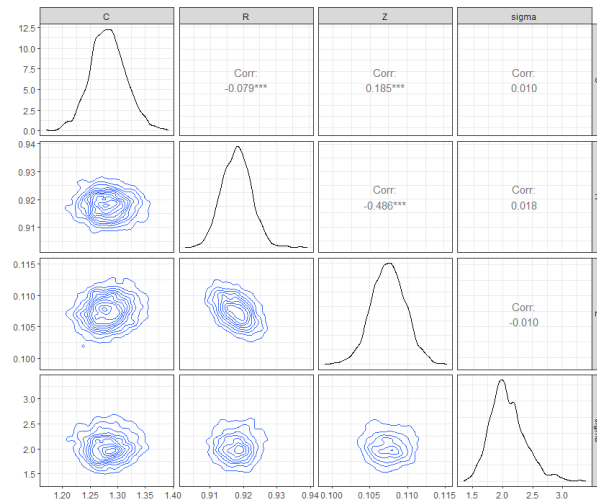
(b) Simulation Study 1.5: WK3 - Matching-models - dependent noise ϵ_t
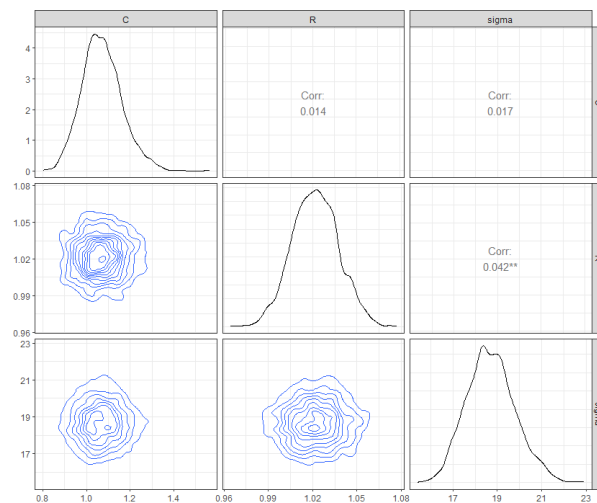
(c) Simulation Study 1.6: dependent noise ϵ_t

Figure 6.7: Posterior distributions with correlation plots for calibration parameters and standard deviance σ from running the MCMC sampling once with dependent noise ϵ_t.
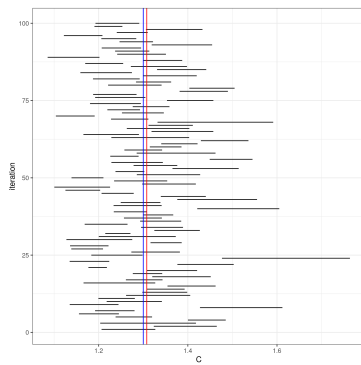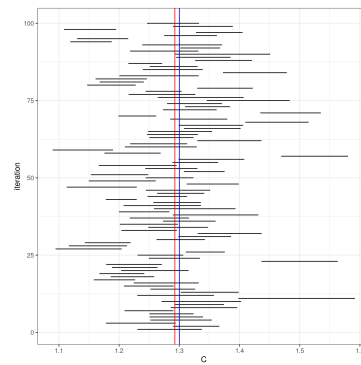
(a) Simulation Study 1.4: WK2 - Matching-models - dependent noise ϵ_t

(b) Simulation Study 1.5: WK3 - Matching-models - dependent noise ϵ_t
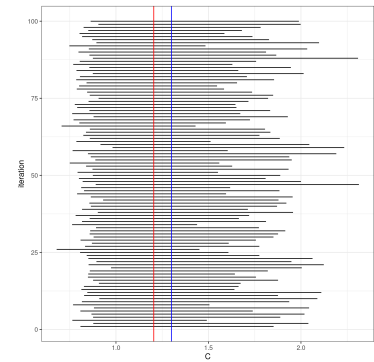
(c) Simulation Study 1.6: dependent noise ϵ_t

Figure 6.8: Average MAP estimate with corresponding 90% credible intervals of C 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
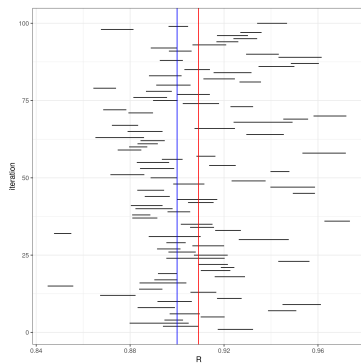
(a) Simulation Study 1.4: WK2 - Matching-models - dependent noise ϵ_t
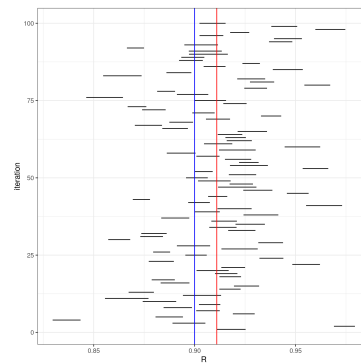
(b) Simulation Study 1.5: WK3 - Matching-models - dependent noise ϵ_t
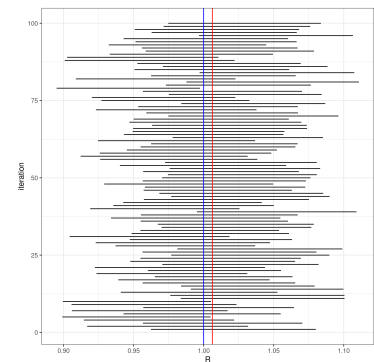
(c) Simulation Study 1.6: dependent noise ϵ_t

Figure 6.9: Average MAP estimate with corresponding 90% credible intervals of R from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values

(a) Simulation Study 1.4: WK2 - Matching-models - dependent noise ϵ_t

(b) Simulation Study 1.5: WK3 - Matching-models - dependent noise ϵ_t

(c) Simulation Study 1.6: dependent noise ϵ_t
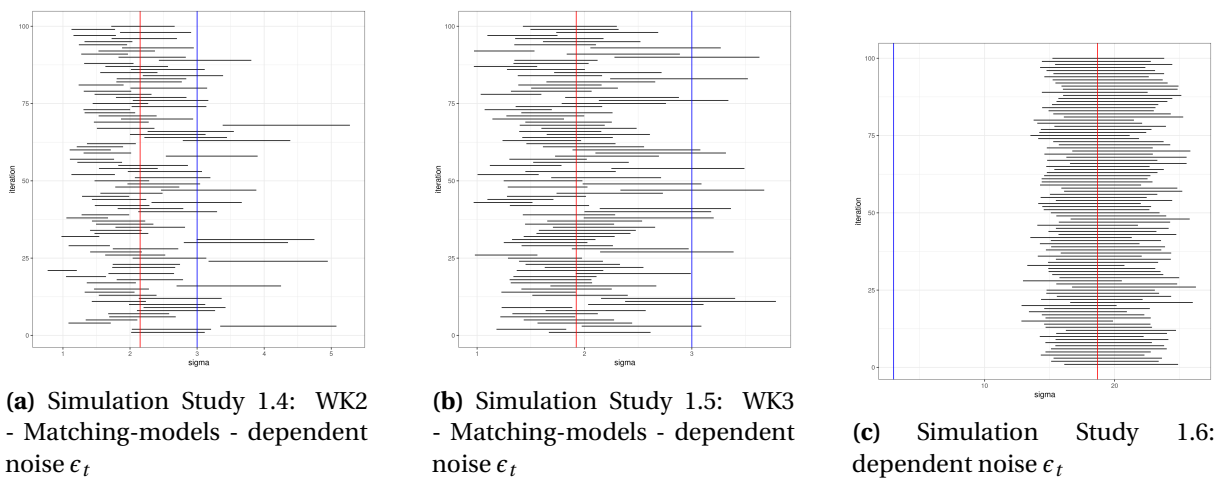
Figure 6.10: Average MAP estimate with corresponding 90% credible intervals of σ from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
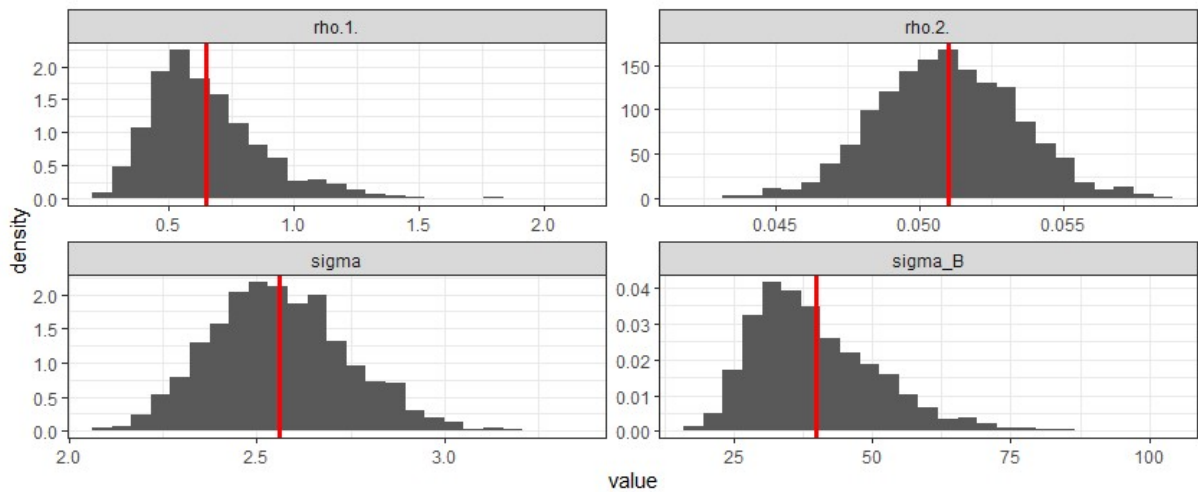
(a) Discrepancy: Resulting posterior distributions from estimating the discrepancy from one simulated dataset. Red line represents estimated MAP of parameter.
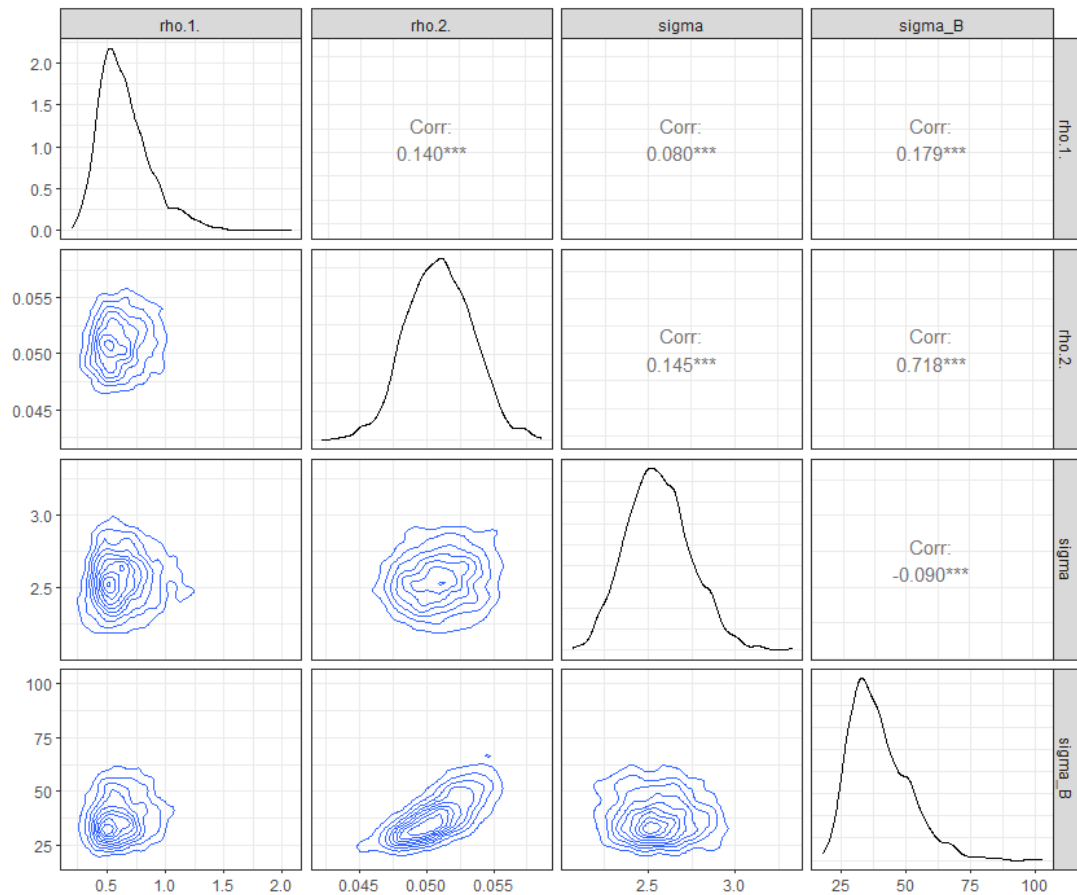
(b) Discrepancy: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.11: Discrepancy parameters estimate from one simulated dataset
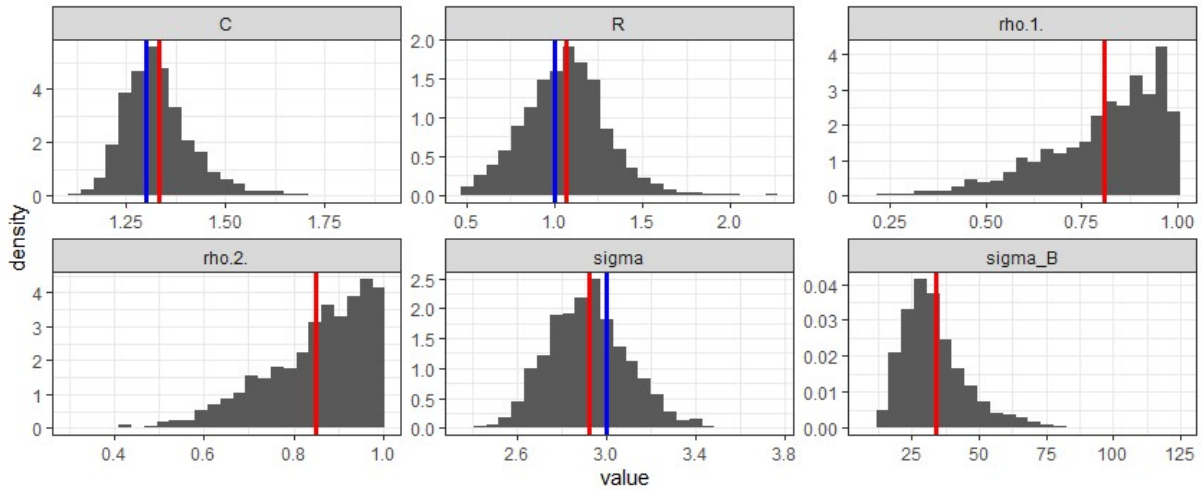
(a) Simulation Study 2.3 - Full Calibration: Resulting posterior distributions from one simulated data-set with priors for ρ's $\sim \mathcal{U}[0,1]$ and independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
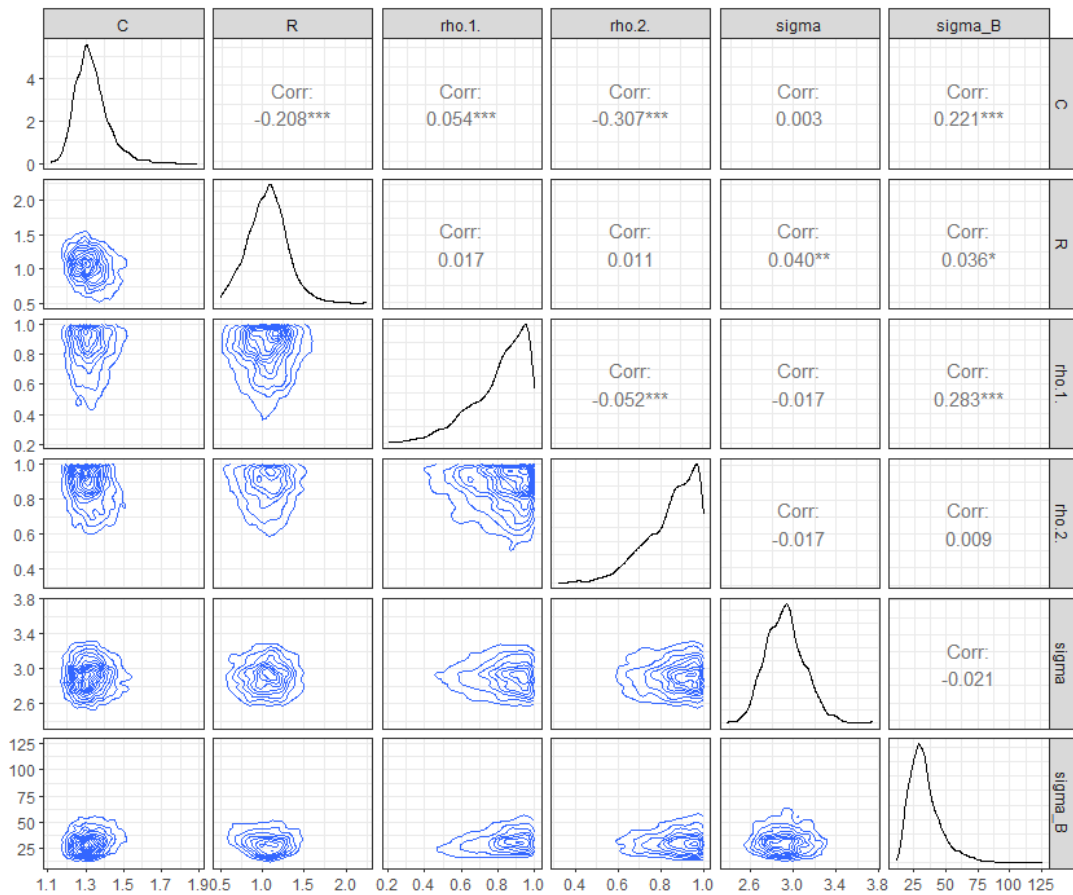
(b) Simulation Study 2.3 - Full Calibration: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.12: Posterior distributions and correlation plots for calibration parameters, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from full calibration with iid noise ϵ. Priors for ρ's are uniform with a range of 0 to 1.
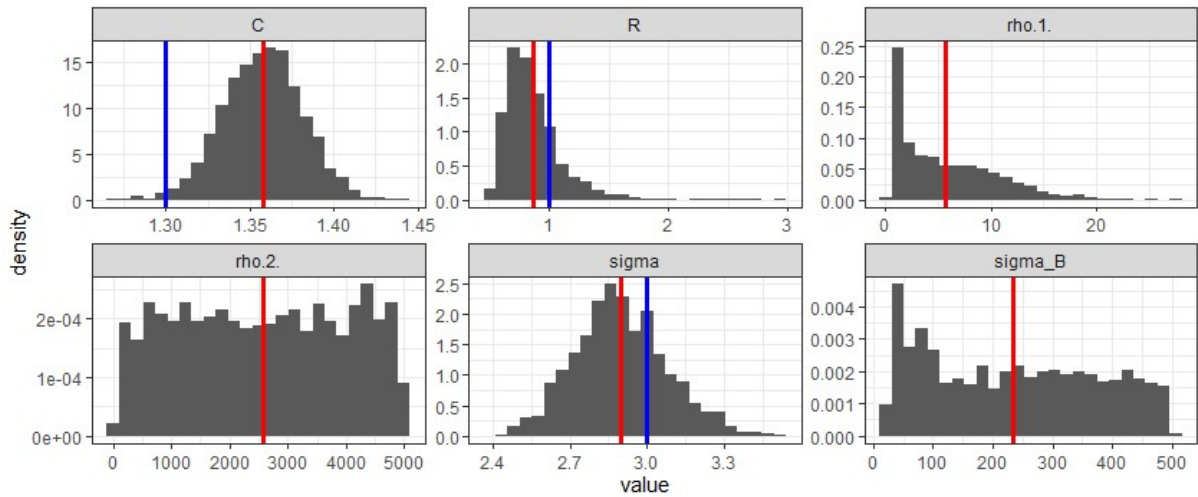
(a) Simulation Study 2.3 - Full Calibration: Resulting posterior distributions from one simulated data-set with priors for ρ's $\sim \mathscr{U}[0, 5000]$ and independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
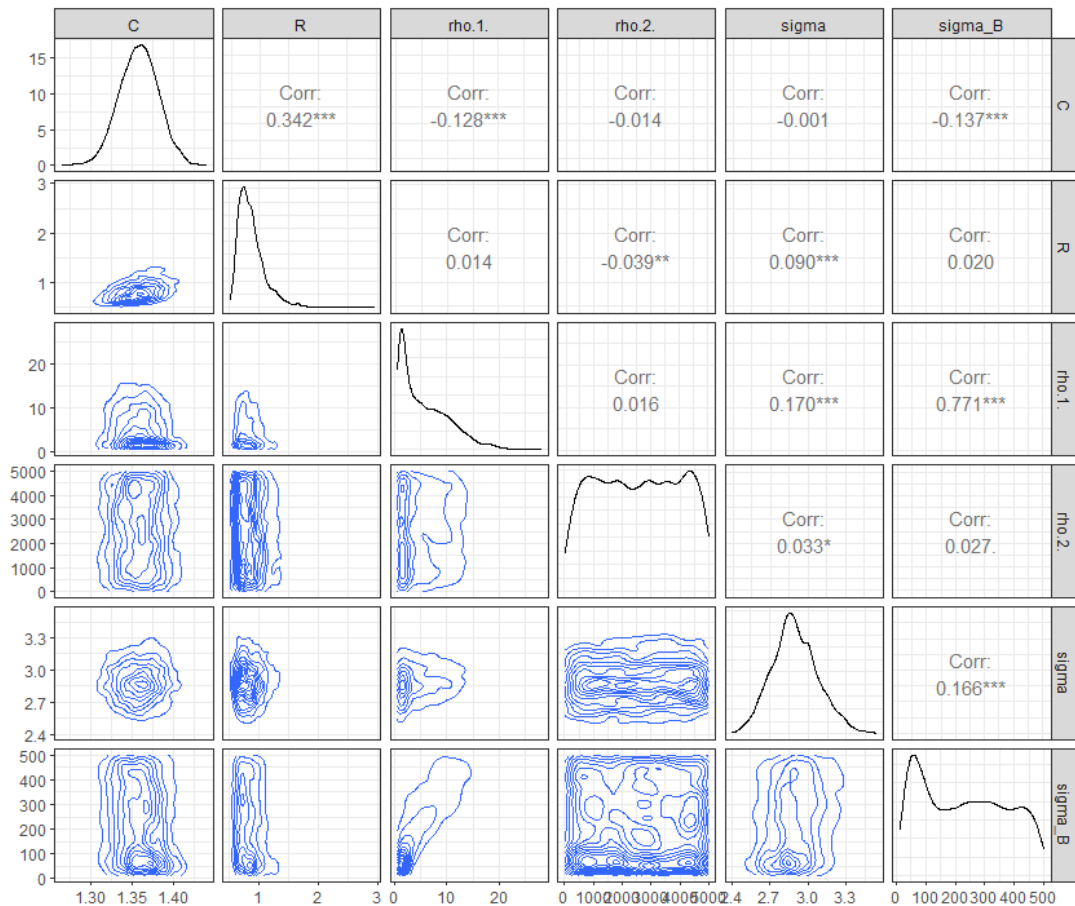
(b) Simulation Study 2.3 - Full Calibration: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.13: Posterior distributions and correlation plots for calibration parameters, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from full calibration with iid noise ϵ. Priors for ρ's are uniform with a range of 0 to 5000.
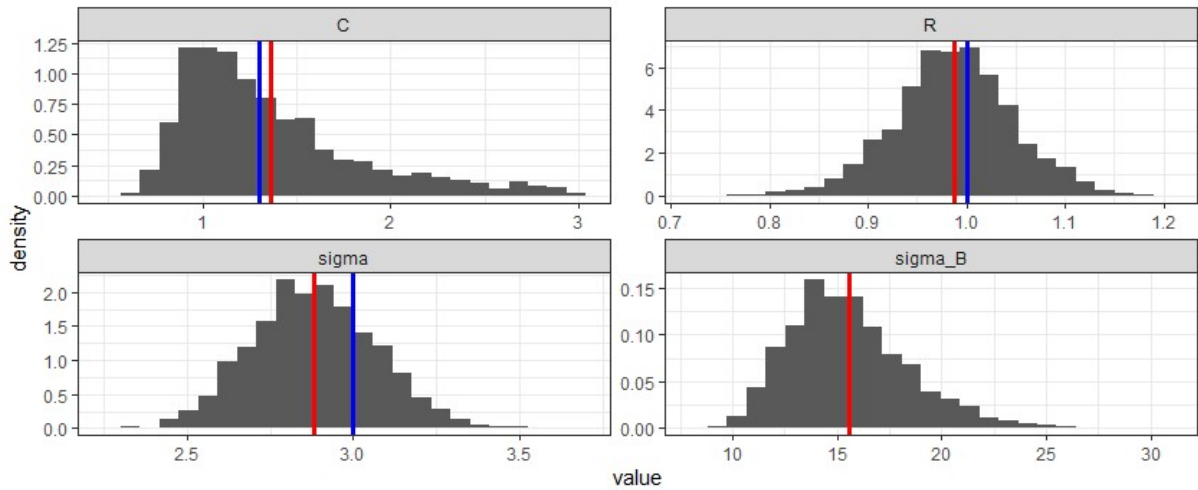
(a) Simulation Study 2.5 - Mod 1: Resulting posterior distributions from one simulated data-set with fixed MAP values for length scales ρ and independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
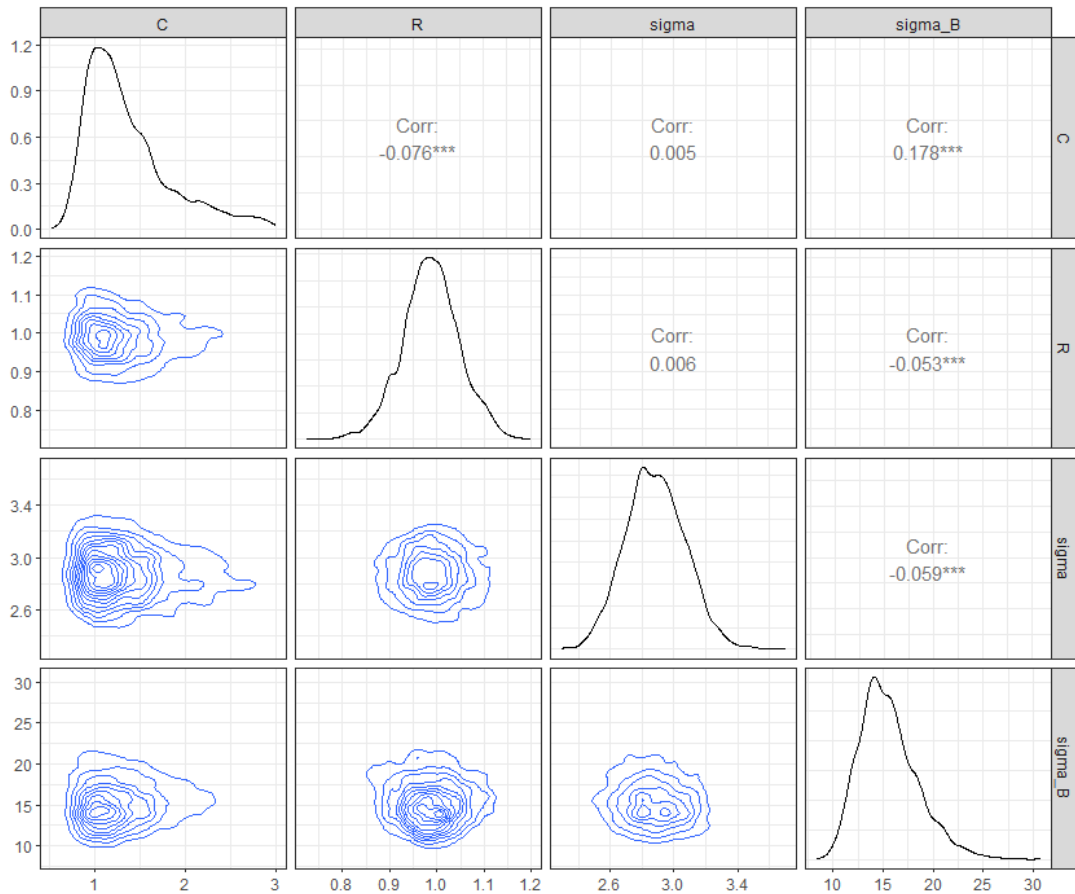
(b) Simulation Study 2.5 - Mod 1: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.14: Posterior distributions and correlation plot for physical parameters R and C, standard deviance σ and bias parameter σ_δ from using estimated fixed values for length scales ρ with independent noise ϵ.
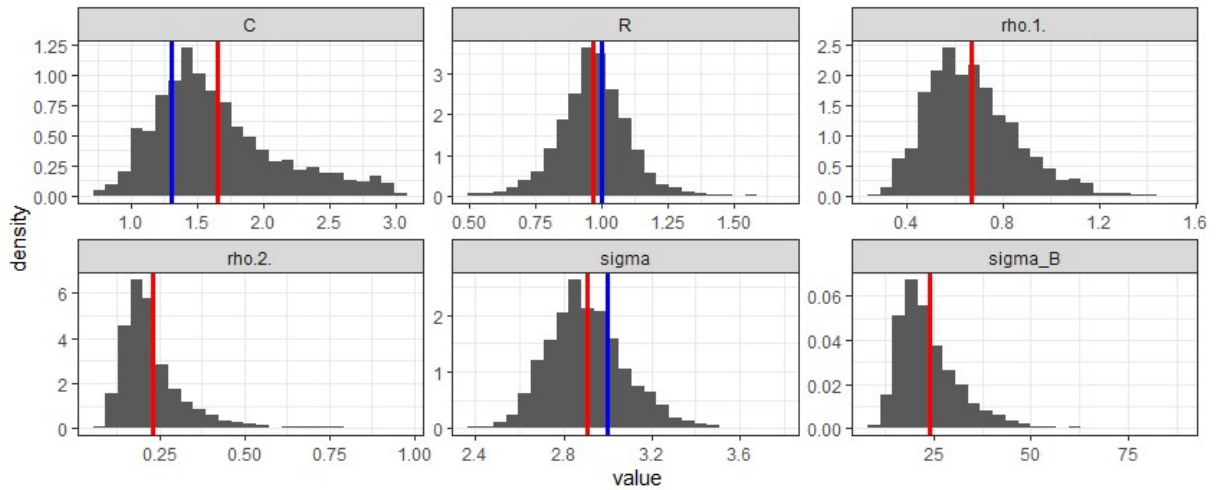
(a) Simulation Study 2.7 - Mod 2: Resulting posterior distributions from one simulated dataset with priors for length scales ρ and independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
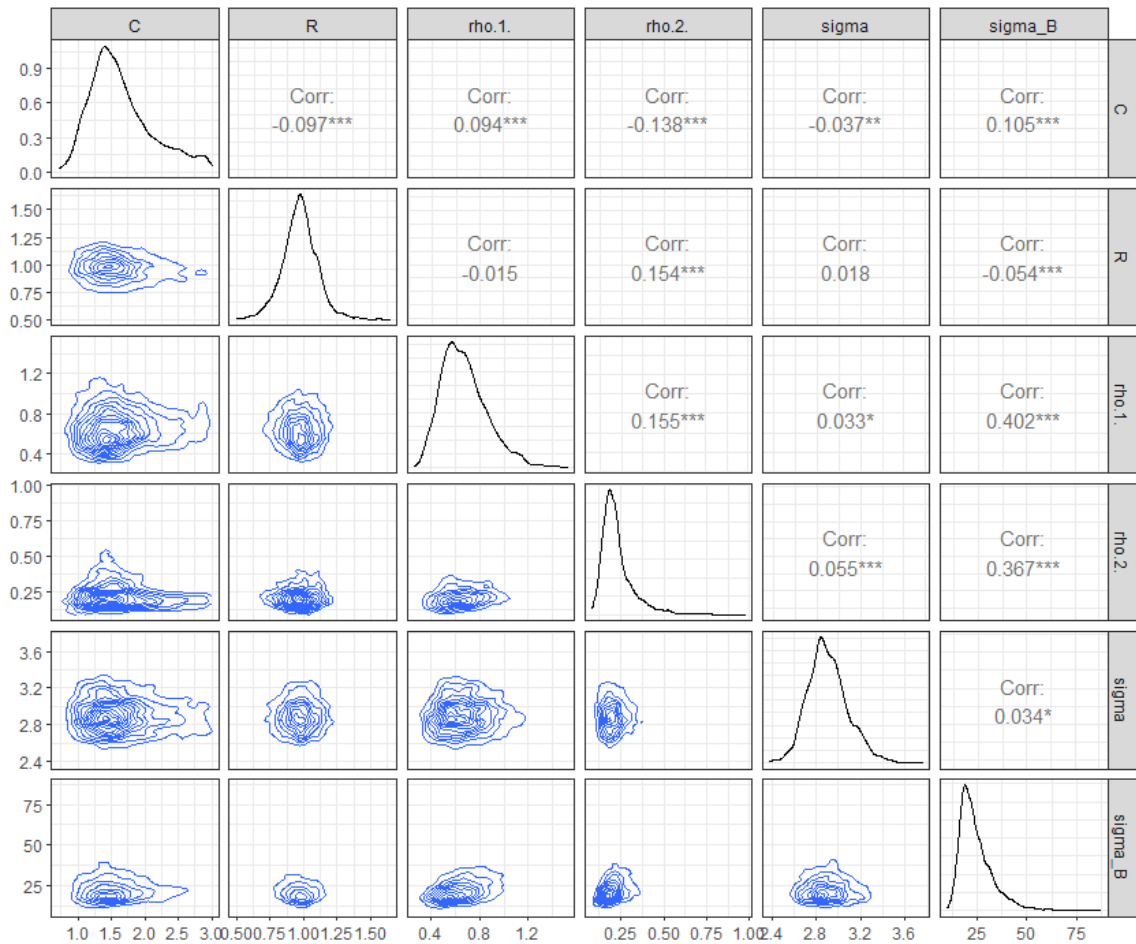
(b) Simulation Study 2.7 - Mod 2: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.15: Posterior distributions and correlation plots for physical parameters R and C, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from using estimated posteriors as priors for length scales ρ with independent noise ϵ.
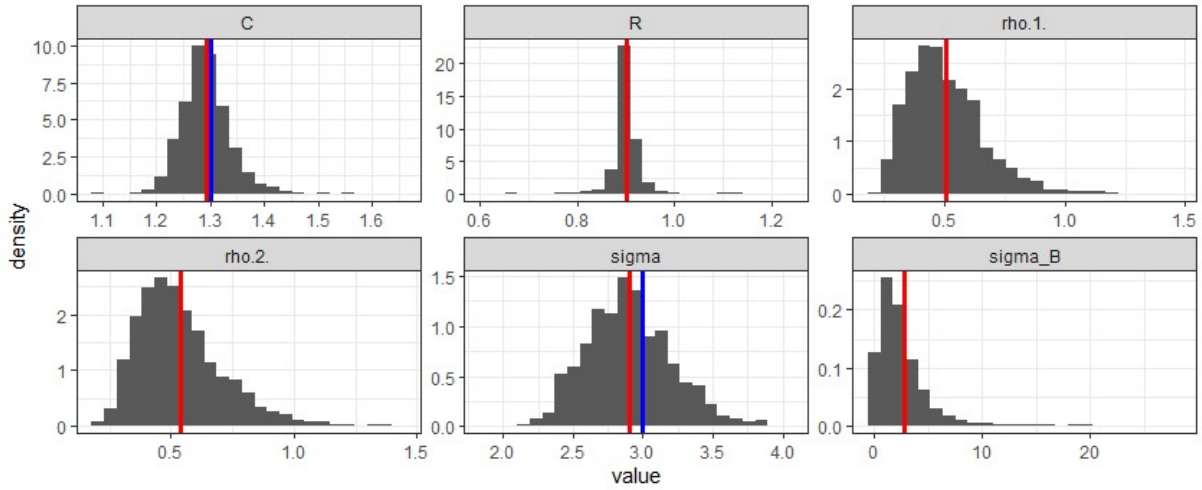
(a) Simulation Study 2.1 - WK2 - Matching Models: Resulting posterior distributions from performing full calibration on one simulated data-set with independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.

(b) Simulation Study 2.1 - WK2 - Matching Models: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.16: Posterior distributions and correlation plots for calibration parameters, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from performing full calibration with independent noise ϵ.
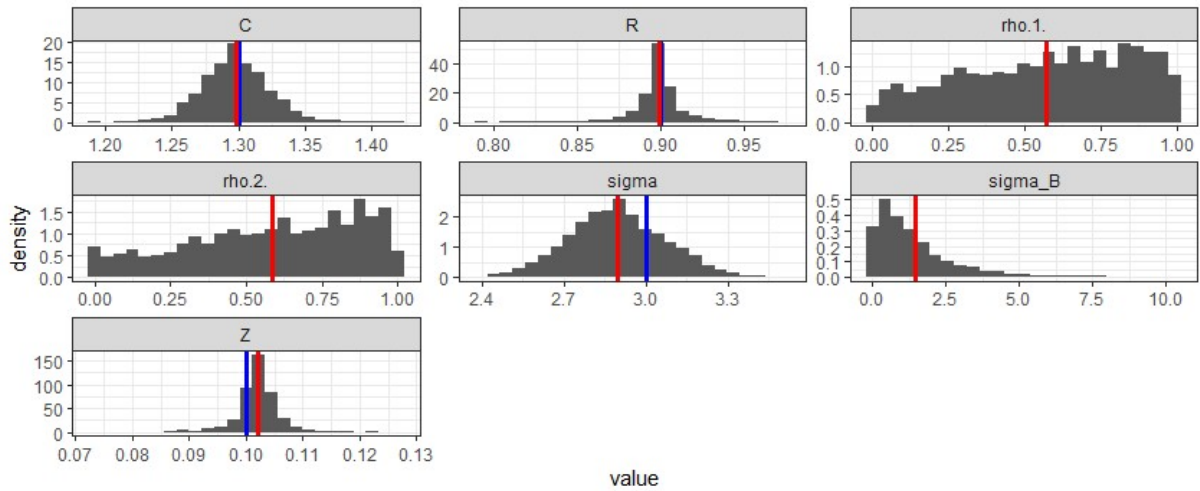
(a) Simulation Study 2.2 - WK3 - Matching Models: Resulting posterior distributions from performing full calibration on one simulated data-set with independent noise ϵ. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
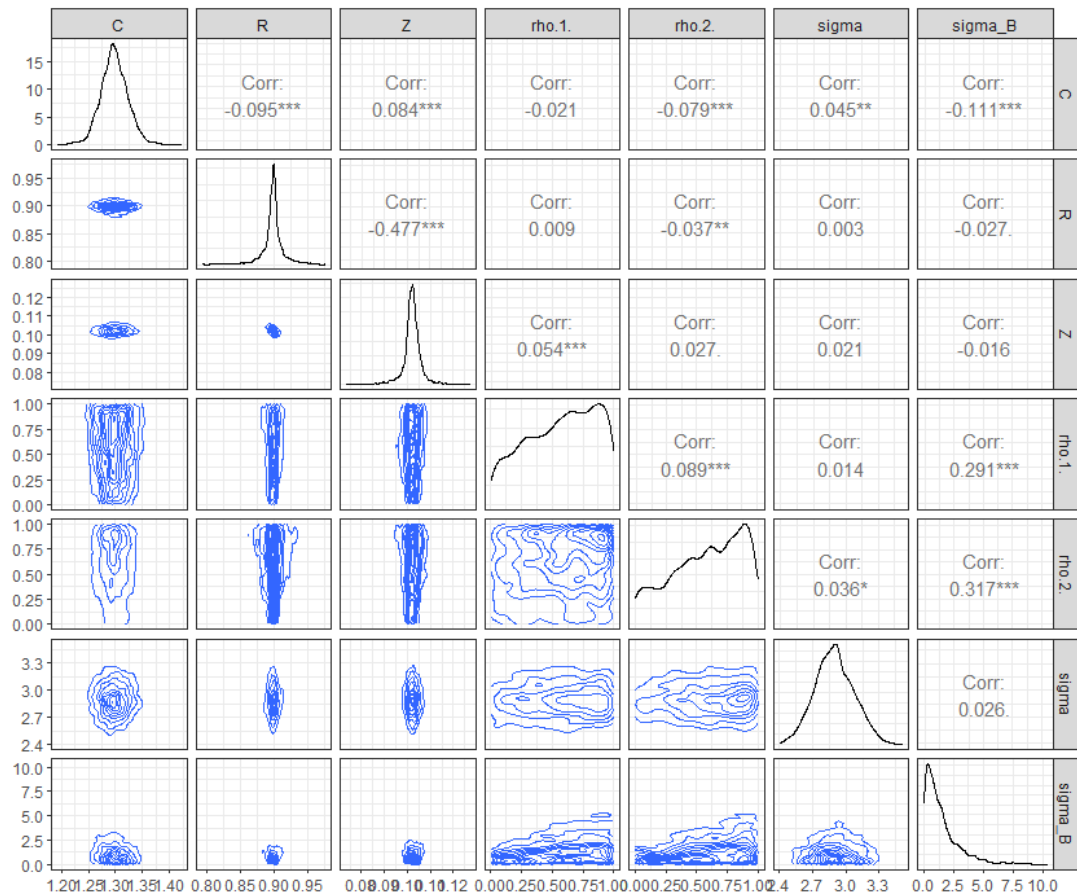
(b) Simulation Study 2.2 - WK3 - Matching Models: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.17: Posterior distributions and correlation plots for calibration parameters, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from performing full calibration with independent noise ϵ.
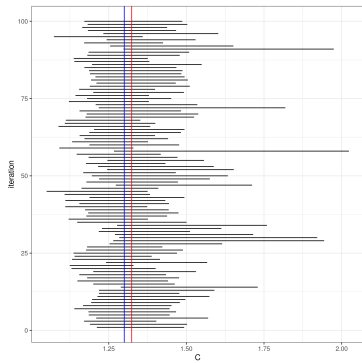
(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
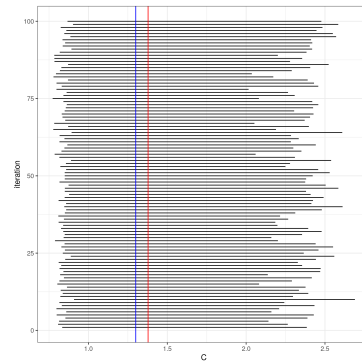
**(b)** Simulation Study 2.5 - Mod 1 - iid noise ϵ
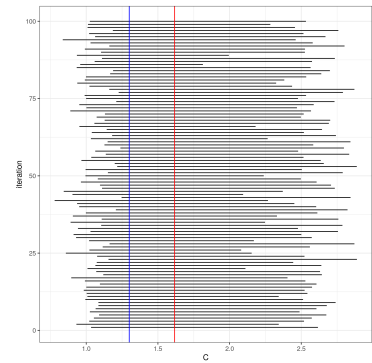
(c) Simulation Study 2.7 - Mod 2 - iid noise ϵ

Figure 6.18: Average MAP estimate with corresponding 90% credible intervals of C from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
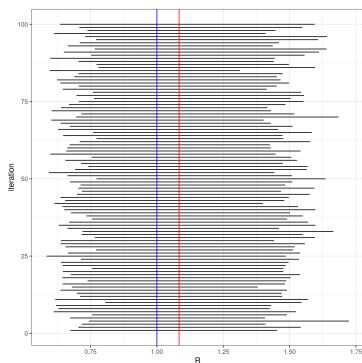

(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
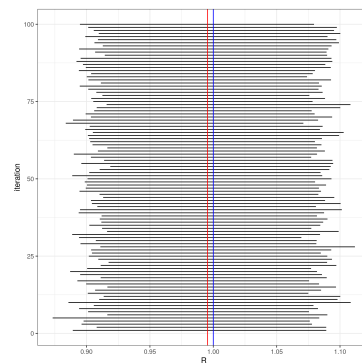
(b) Simulation Study 2.5 - Mod 1 - iid noise ϵ
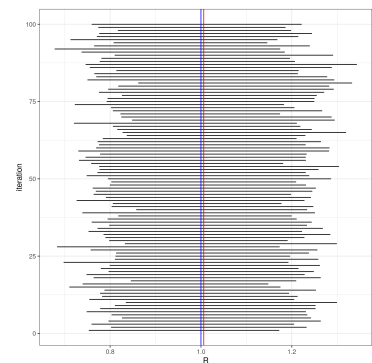
(c) Simulation Study 2.7 - Mod 2 - iid noise ϵ

Figure 6.19: Average MAP estimate with corresponding 90% credible intervals of R from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
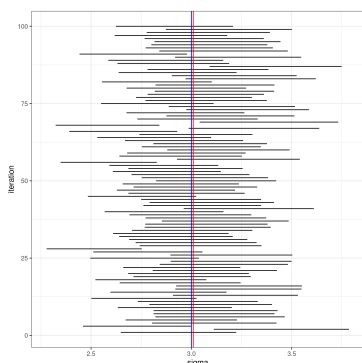
(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
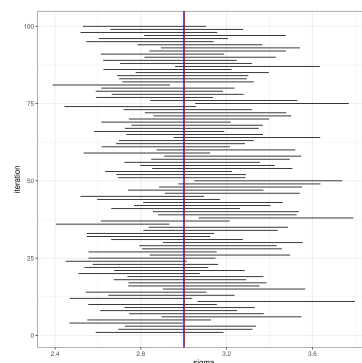
(b) Simulation Study 2.5 - Mod 1 - iid noise ϵ
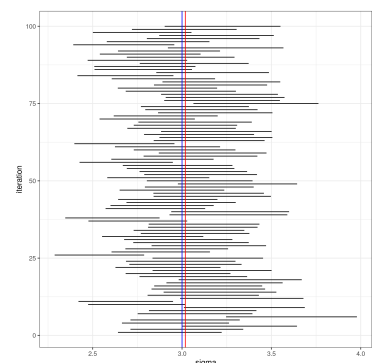
(c) Simulation Study 2.7 - Mod 2 - iid noise ϵ

Figure 6.20: Average MAP estimate with corresponding 90% credible intervals of σ from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
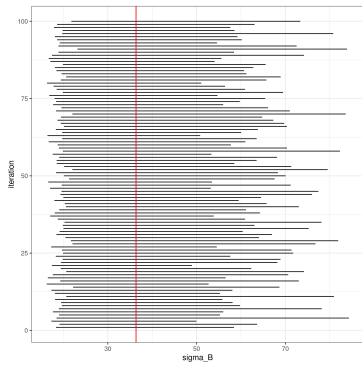
(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
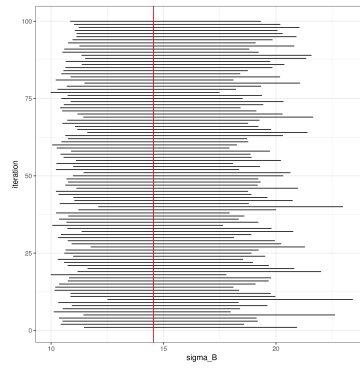
(b) Simulation Study 2.5 - Mod 1 - iid noise ϵ
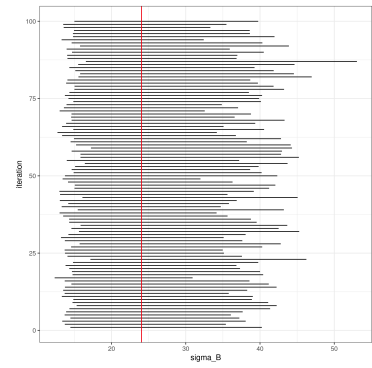
(c) Simulation Study 2.7 - Mod 2 - iid noise ϵ

Figure 6.21: Average MAP estimate with corresponding 90% credible intervals of σ_δ from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
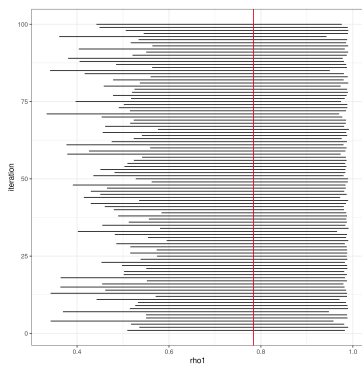
(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
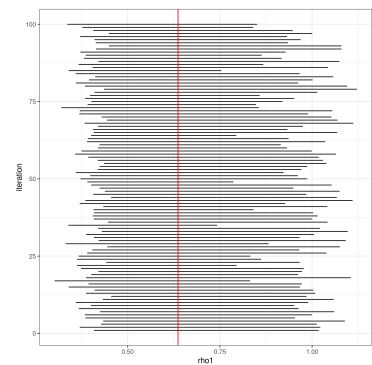
(b) Simulation Study 2.7 - Mod 2- iid noise ϵ

Figure 6.22: Average MAP estimate with corresponding 90% credible intervals of ρ_1 from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
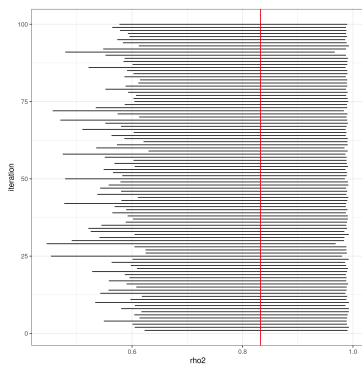
(a) Simulation Study 2.3 - Full Calibration - iid noise ϵ
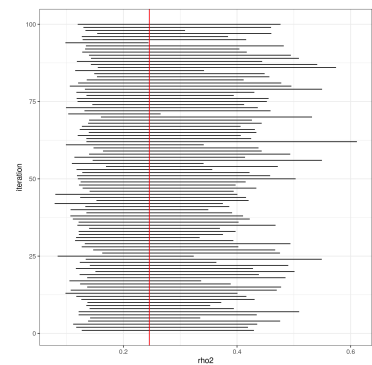
(b) Simulation Study 2.7 - Mod 2- iid noise ϵ

Figure 6.23: Average MAP estimate with corresponding 90% credible intervals of ρ_2 from 100 independent simulations of the posterior distribution with independent noise ϵ. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
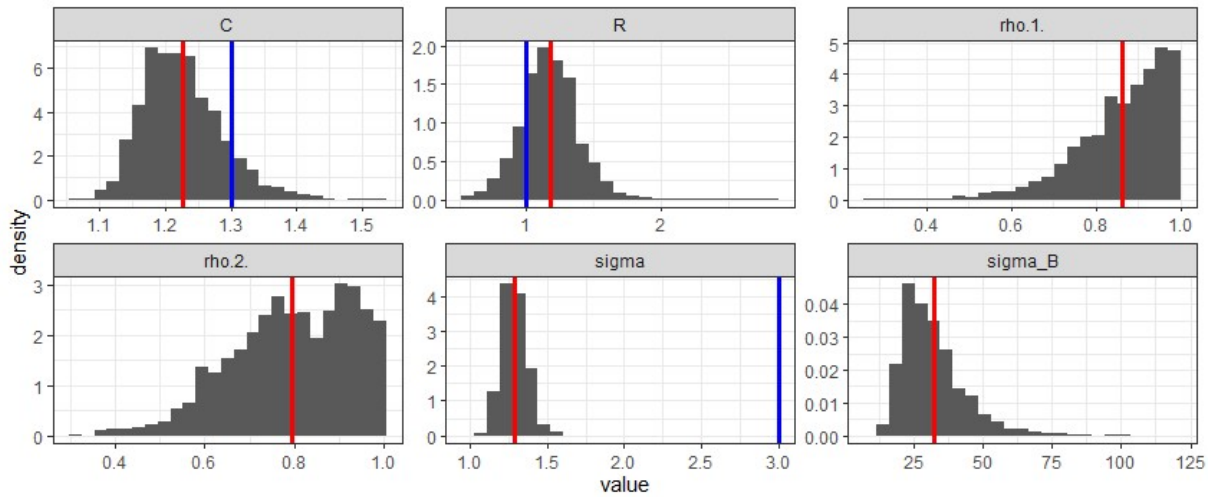
(a) Simulation Study 2.4 - Full Calibration: Resulting posterior distributions from one simulated data-set with dependent noise ϵ_t. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
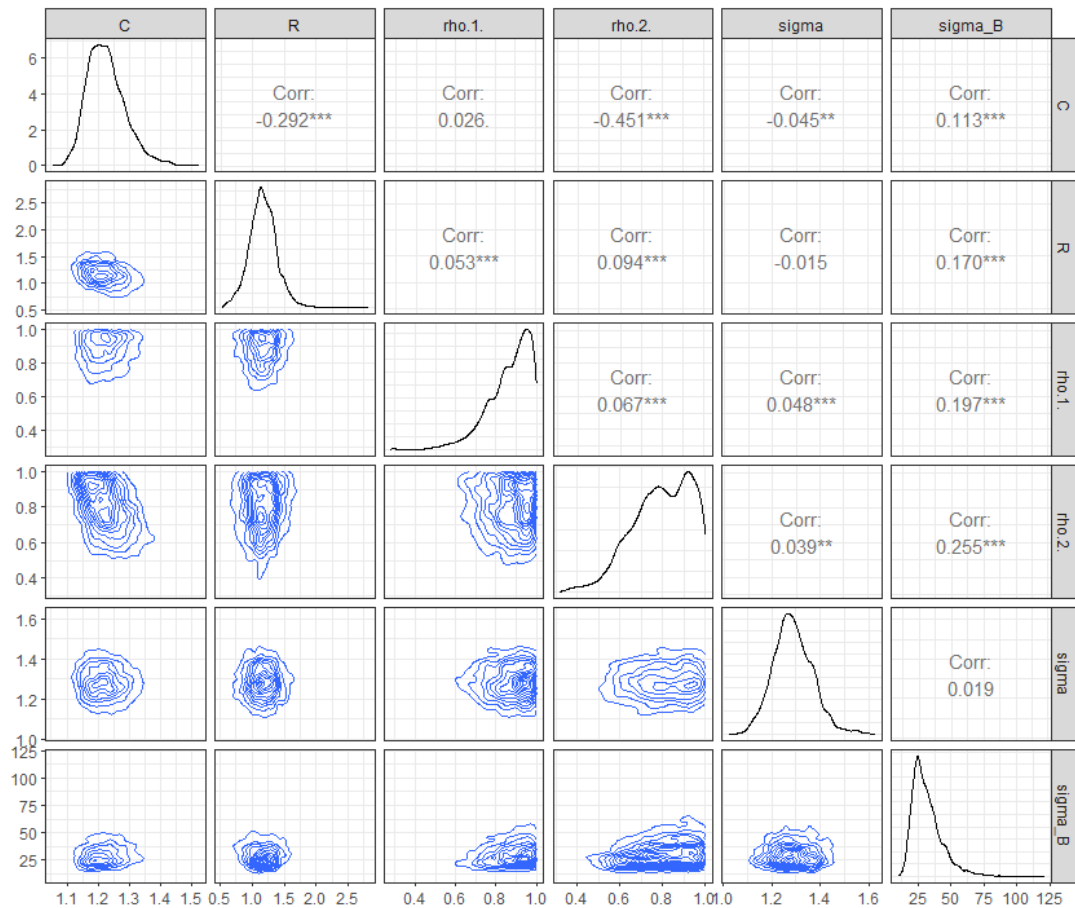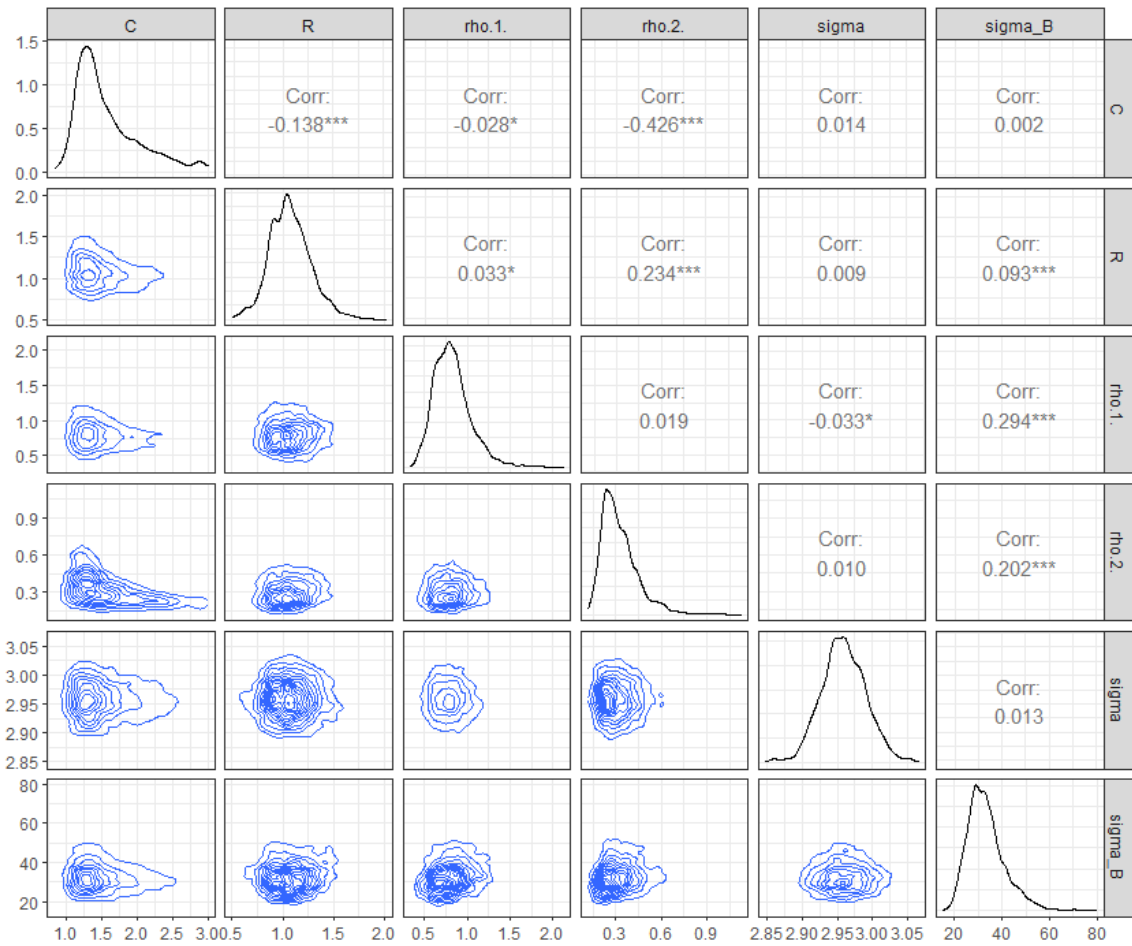

(b) Simulation Study 2.4 - Full Calibration: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.24: Posterior distributions with correlation and mixing plots for calibration parameters, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2 from full calibration with dependent noise ϵ_t.
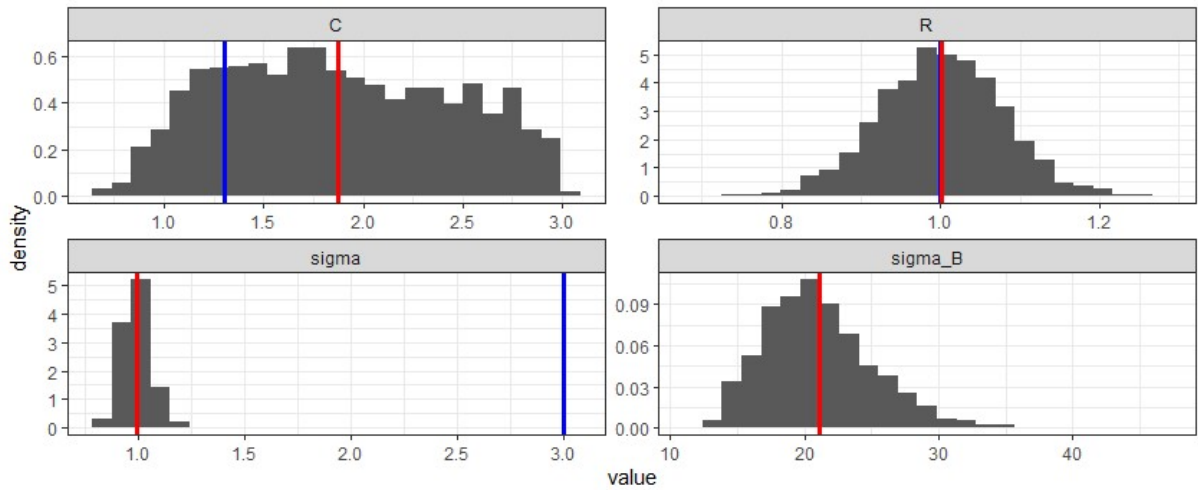
(a) Simulation Study 2.6 -Mod 1: Resulting posterior distributions from one simulated data-set with fixed MAP values for length scales ρ and dependent noise ϵ_t. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
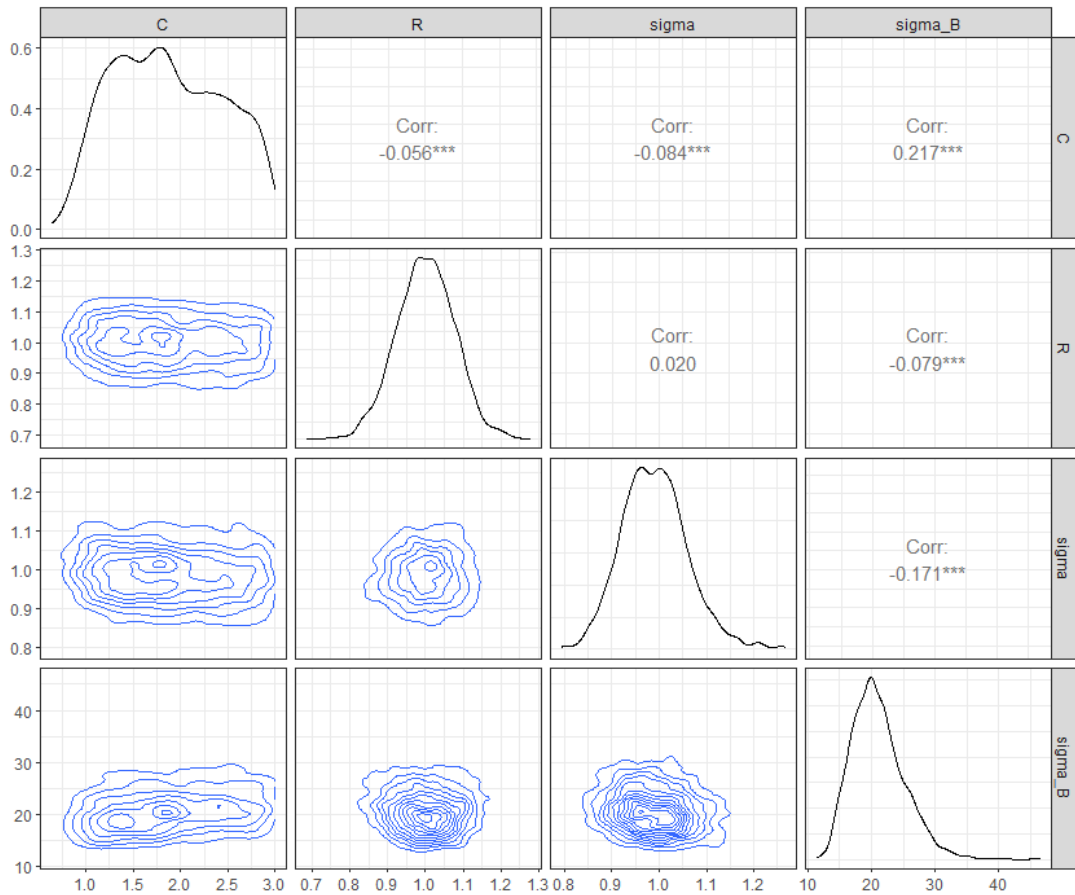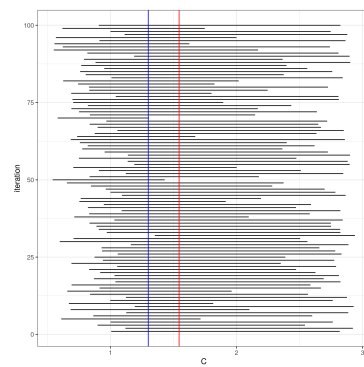

(b) Simulation Study 2.6 -Mod 1: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.25: Posterior distributions and correlation plot for physical parameters R and C, standard deviance σ and bias parameter σ_δ from using estimated fixed values for length scales ρ with dependent noise ϵ_t.
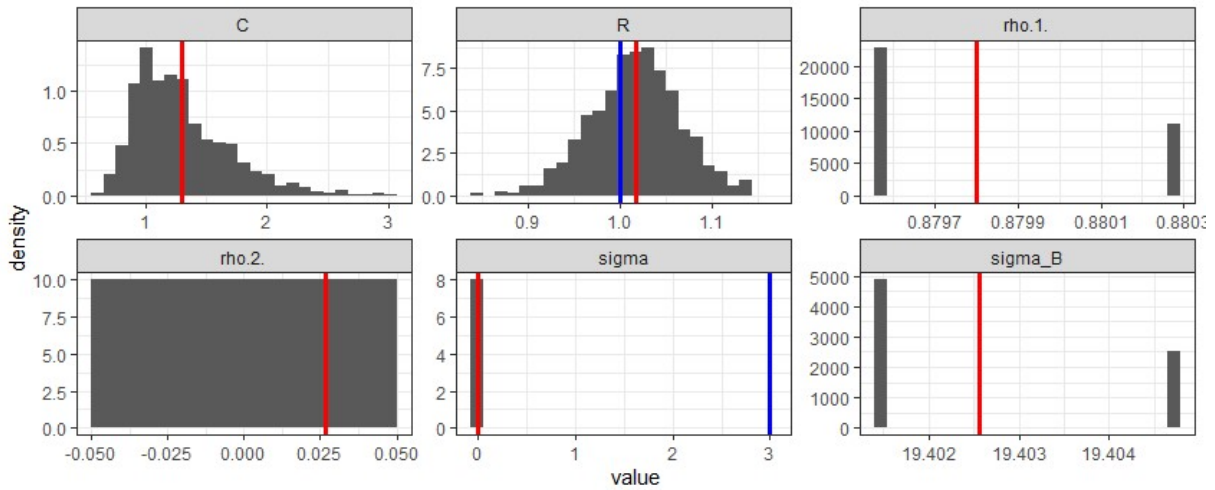
(a) Simulation Study 2.8 - Mod 2: Resulting posterior distributions from one simulated dataset with priors for length scales ρ and dependent noise ϵ_t and uniform priors for σ's. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.
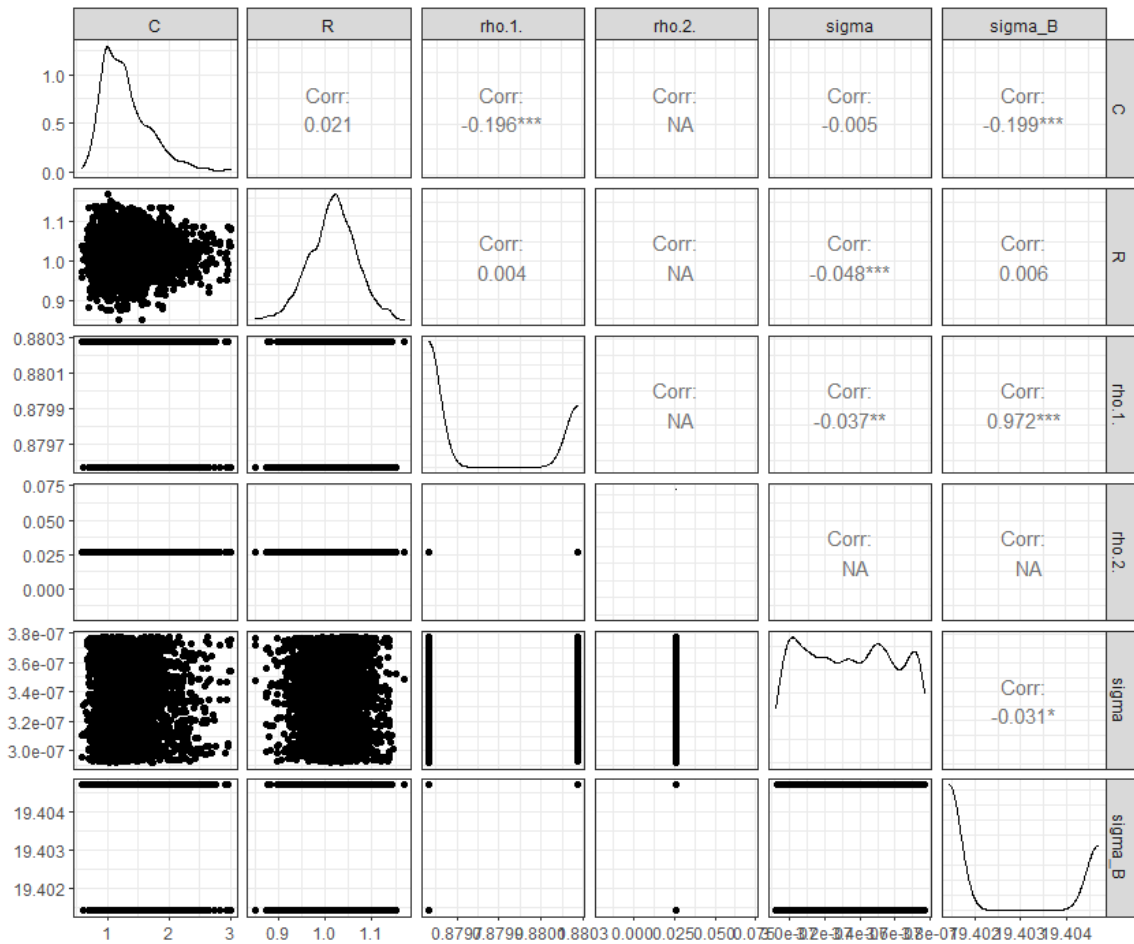
(b) Simulation Study 2.8 - Mod 2: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.26: Posterior distributions and correlation plots for physical parameters R and C, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2. We use estimated posteriors as priors for length scales ρ and uniform priors for σ's with dependent noise ϵ_t.
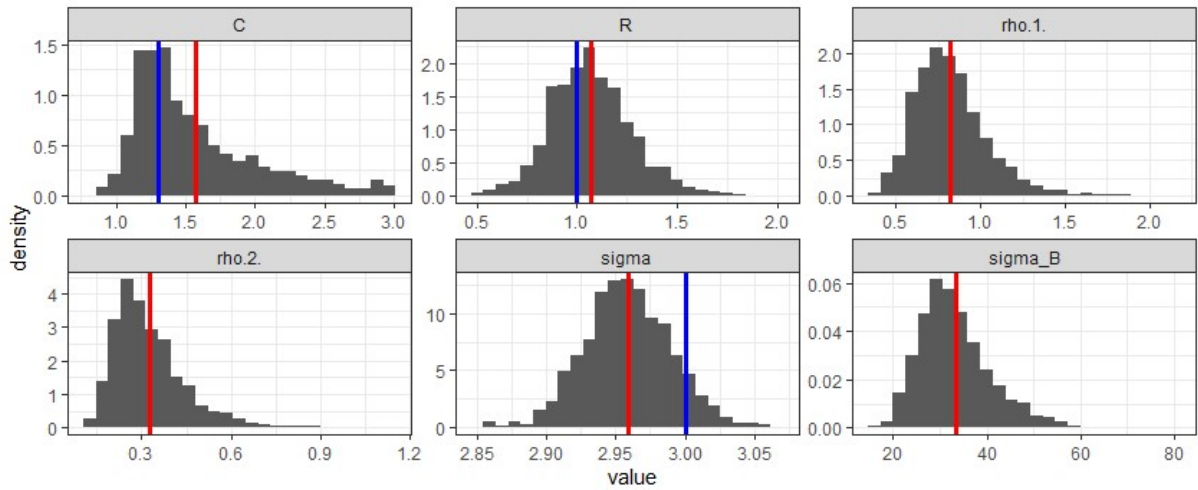
(a) Simulation Study 2.8 - Mod 2: Resulting posterior distributions from one simulated dataset with priors for length scales ρ and dependent noise ϵ_t and inverse gamma priors for σ's. Red line represents estimated MAP of parameter, whereas blue line represents the known, true value.

(b) Simulation Study 2.8 -Mod 2: Correlation plot of the parameters. Contourplots of each pair of numeric variable are drawn on the left part of the figure. Pearson correlation is displayed on the right. Variable distribution is displayed on the diagonal.

Figure 6.27: Posterior distributions and correlation plots for physical parameters R and C, standard deviance σ and bias parameters σ_δ, ρ_1 and ρ_2. We use estimated posteriors as priors for length scales ρ and inverse Gamma priors for σ's with dependent noise ϵ_t.
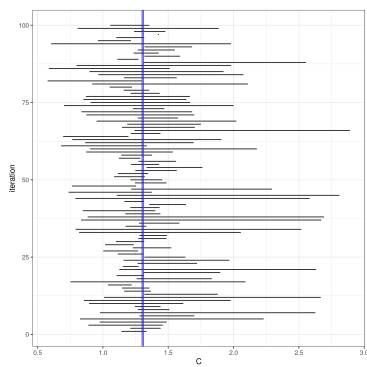
(a) Simulation Study 2.4 - Full Calibration - dependent noise ϵ_t

(b) Simulation Study 2.6 - Mod 1 - dependent noise ϵ_t
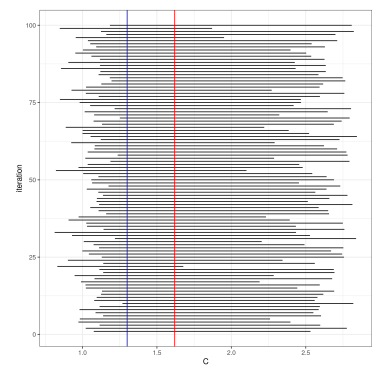
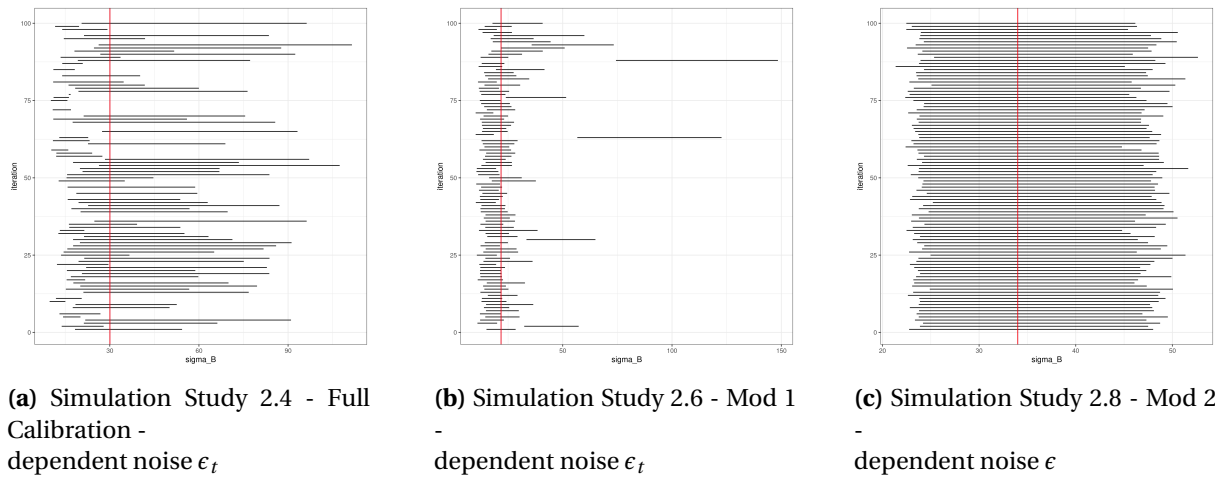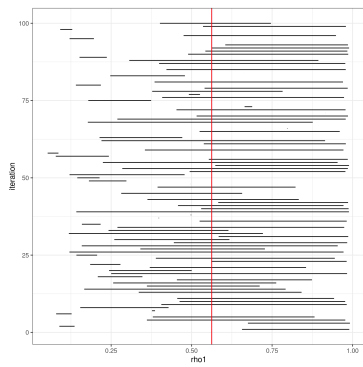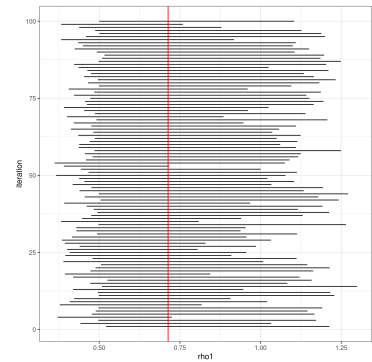**(c)** Simulation Study 2.8 - Mod 2 - dependent noise ϵ_t

Figure 6.28: Average MAP estimate with corresponding 90% credible intervals of C from 100 dependent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
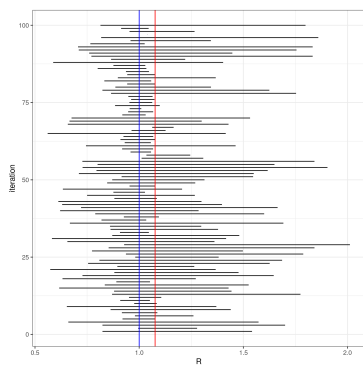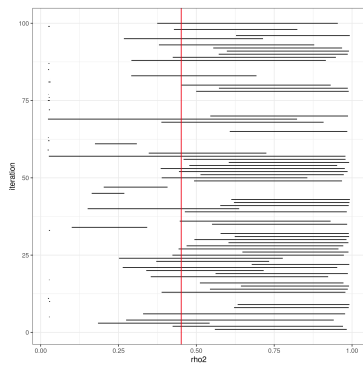

**(a)** Simulation Study 2.4 - Full Calibration - dependent noise ϵ_t
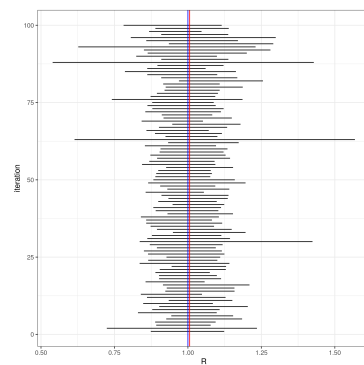
(b) Simulation Study 2.6 - Mod 1 - dependent noise ϵ_t
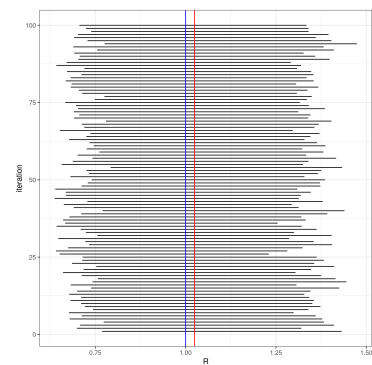
(c) Simulation Study 2.8 - Mod 2 - dependent noise ϵ_t

Figure 6.29: Average MAP estimate with corresponding 90% credible intervals of R from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values

(a) Simulation Study 2.4 - Full
Calibration -
dependent noise ϵ_t

(b) Simulation Study 2.6 - Mod 1
-
dependent noise ϵ_t

(c) Simulation Study 2.8 - Mod 2
-
dependent noise ϵ_t
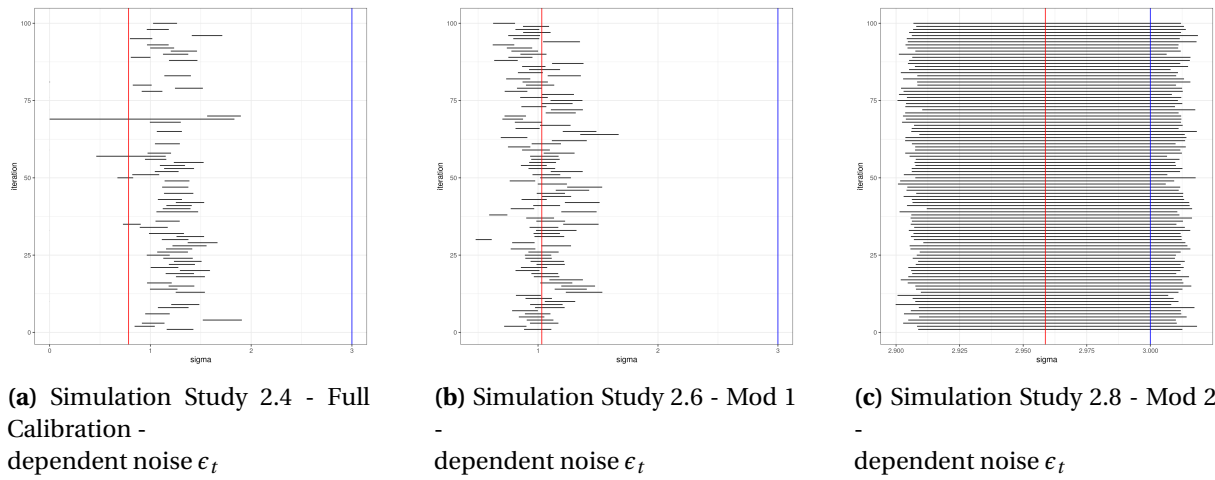
Figure 6.30: Average MAP estimate with corresponding 90% credible intervals of σ from 100
independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents
the true mean and red line the average of the 100 maximum posterior mean values

(a) Simulation Study 2.4 - Full
Calibration -
dependent noise ϵ_t

(b) Simulation Study 2.6 - Mod 1
-
dependent noise ϵ_t

(c) Simulation Study 2.8 - Mod 2
-
dependent noise ϵ

Figure 6.31: Average MAP estimate with corresponding 90% credible intervals of σ_δ from 100
independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents
the true mean and red line the average of the 100 maximum posterior mean values

(a) Simulation Study 2.4 - Full
Calibration -
dependent noise ϵ_t

(b) Simulation Study 2.8 -Mod 2 -
dependent noise ϵ_t

Figure 6.32: Average MAP estimate with corresponding 90% credible intervals of ρ_1 from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values

(a) Simulation Study 2.4 - Full
Calibration -
dependent noise ϵ_t
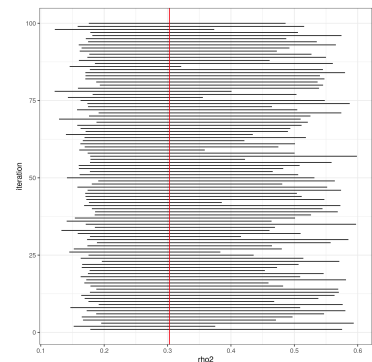
(b) Simulation Study 2.8 - Mod 2-
dependent noise ϵ_t

Figure 6.33: Average MAP estimate with corresponding 90% credible intervals of ρ_2 from 100 independent simulations of the posterior distribution with dependent noise ϵ_t. Blue line represents the true mean and red line the average of the 100 maximum posterior mean values
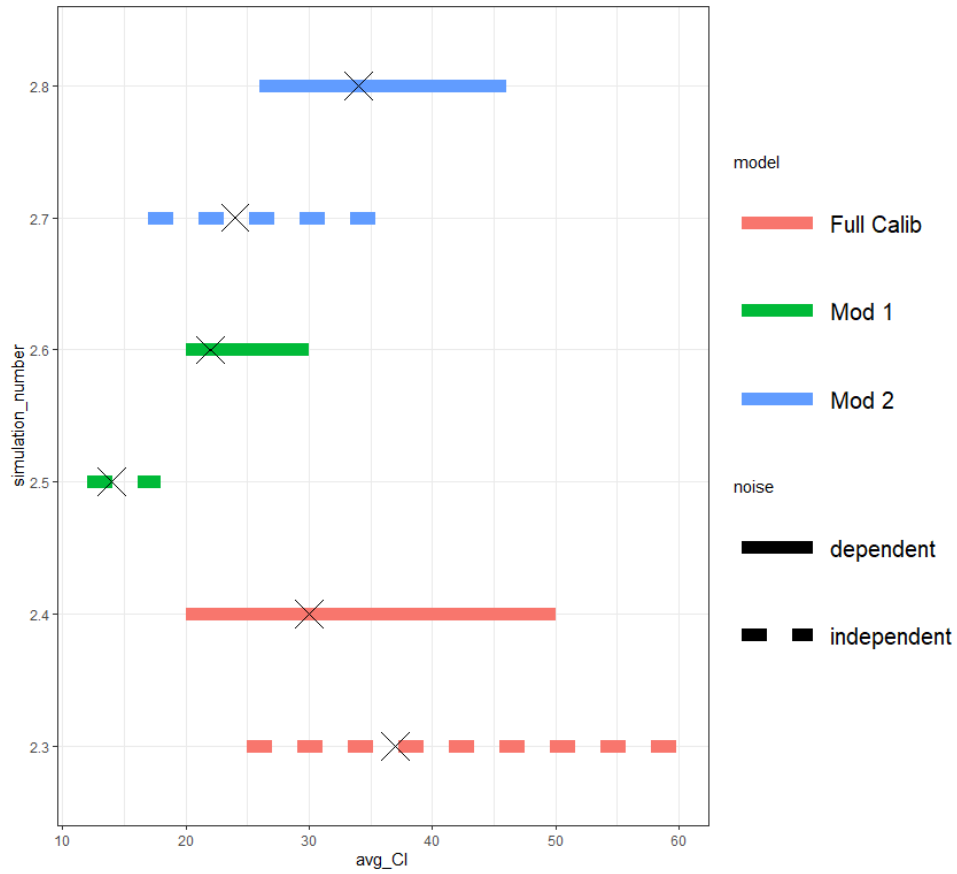
Figure 6.34: Simulation Study 2: Summary of the mean credibility intervals for length scale parameter σ_δ. Corresponding Simulation Study as defined in Table 3.1 can be deduced from the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectively.
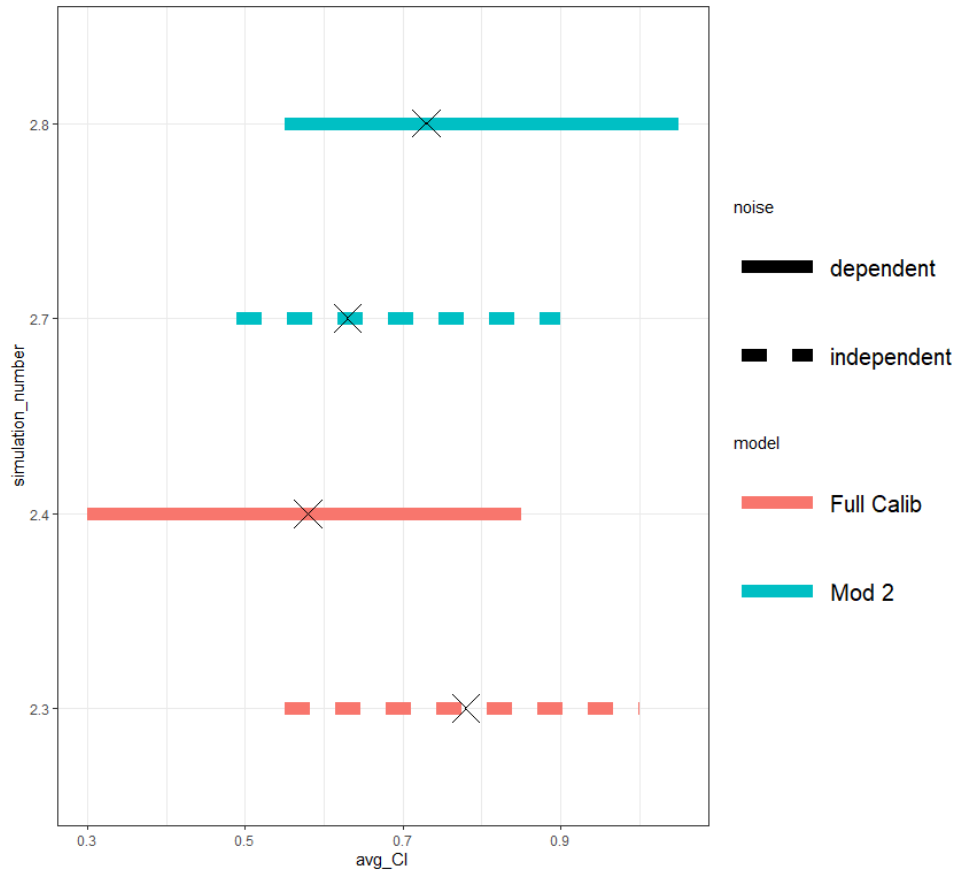
Figure 6.35: Simulation Study 2: Summary of the mean credibility intervals for length scale parameter ρ_1. Corresponding Simulation Study as defined in Table 3.1 can be deduced from the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectively.
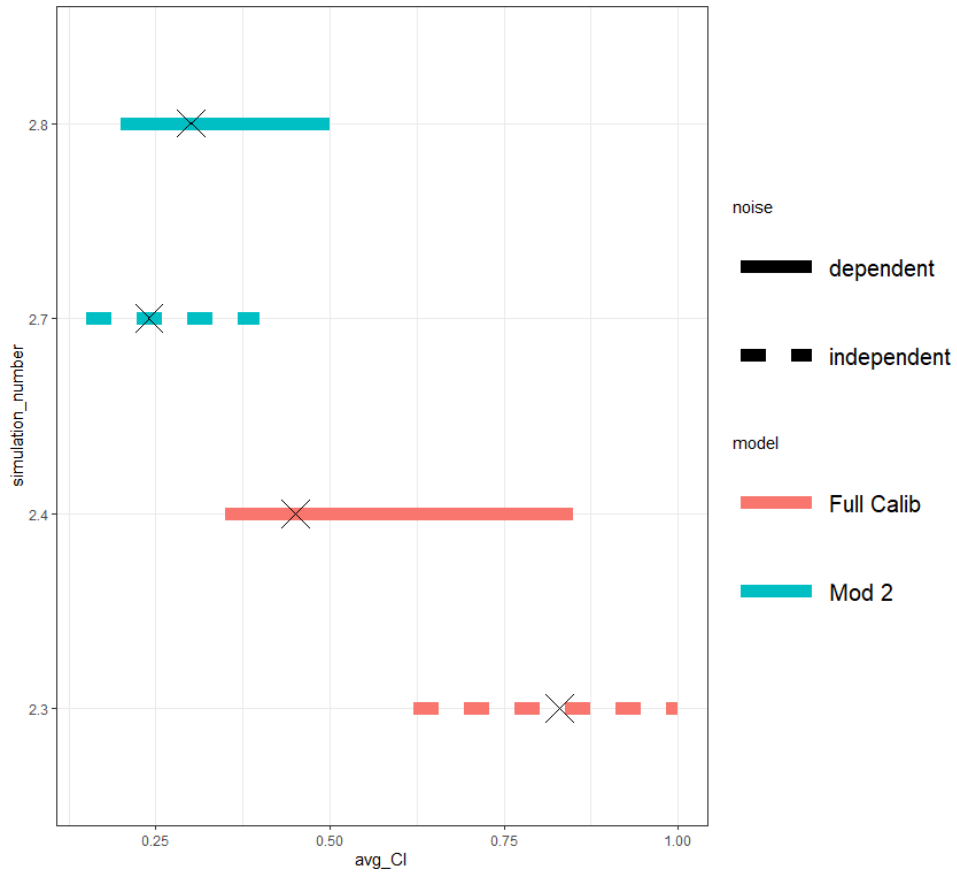
Figure 6.36: Simulation Study 2: Summary of the mean credibility intervals for length scale parameter ρ_2. Corresponding Simulation Study as defined in Table 3.1 can be deduced from the y-axis. Fully drawn and dashed lines correspond to dependent and independent noise respectively.
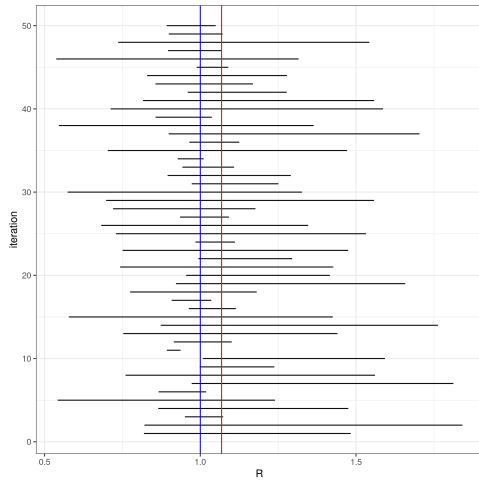
(a) Simulation Study 3: Full Calibration -
R for dependent noise with theta 1 ϵ_t
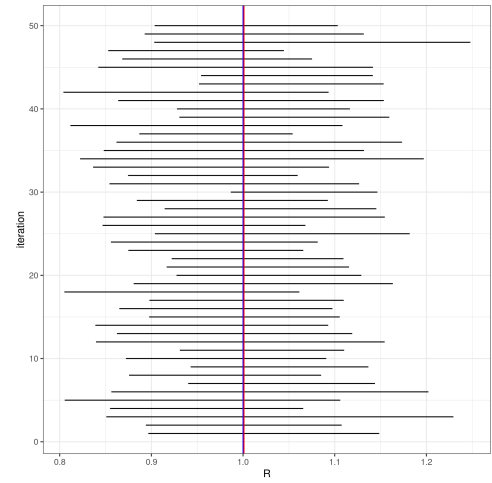
(b) Simulation Study 3: Mod 1-
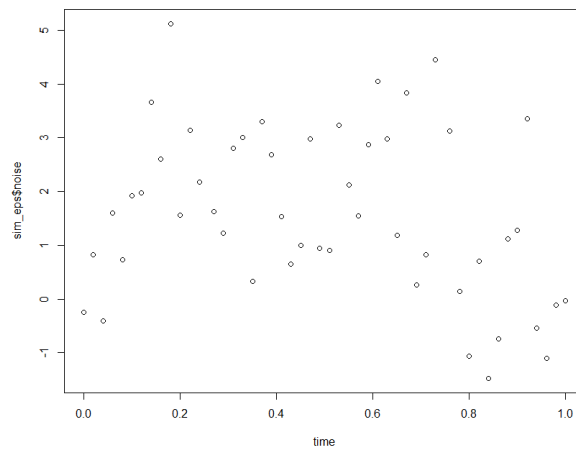R for dependent noise with theta 1 ϵ_t
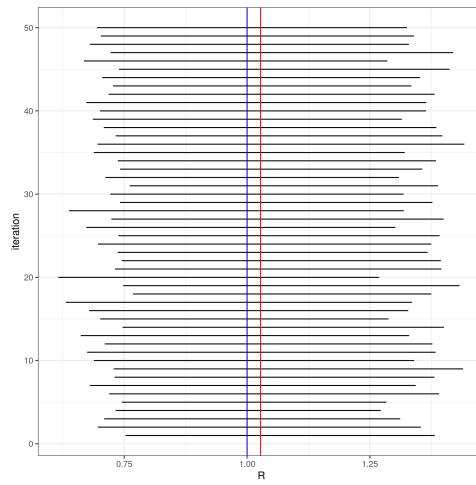
(c) Simulation Study 3: Mod 2-
R for dependent noise with theta 1 ϵ_t
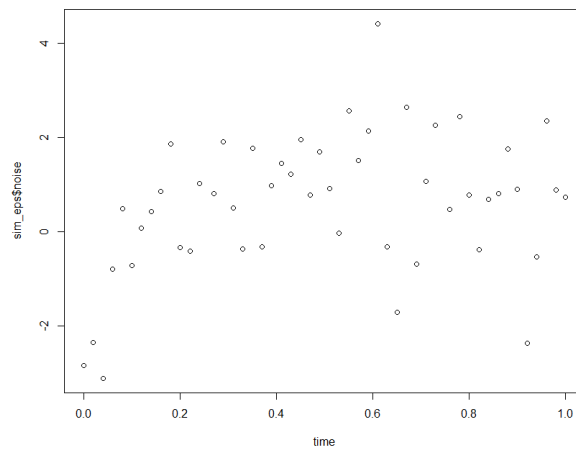
Figure 6.37

(a) Simulation Study 3: Full Calibration -
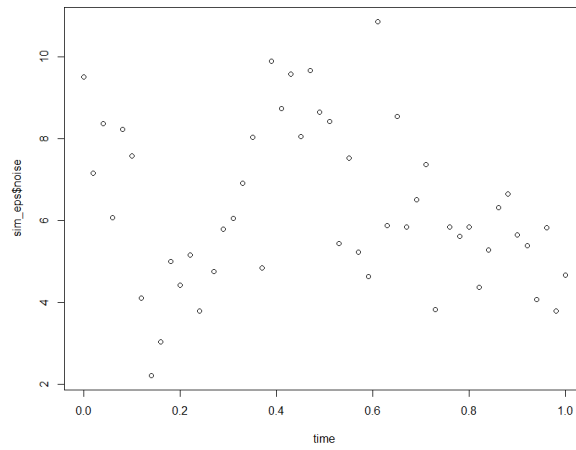Big CI for dependent noise ϵ_t - iteration 2
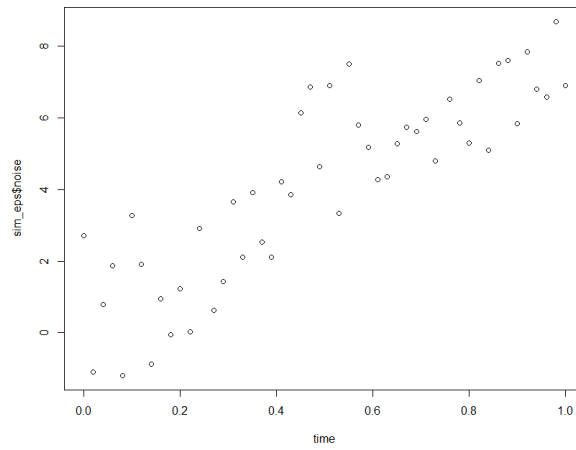
(b) Simulation Study 3: Mod 1-
Big CI dependent noise ϵ_t - iteration 48
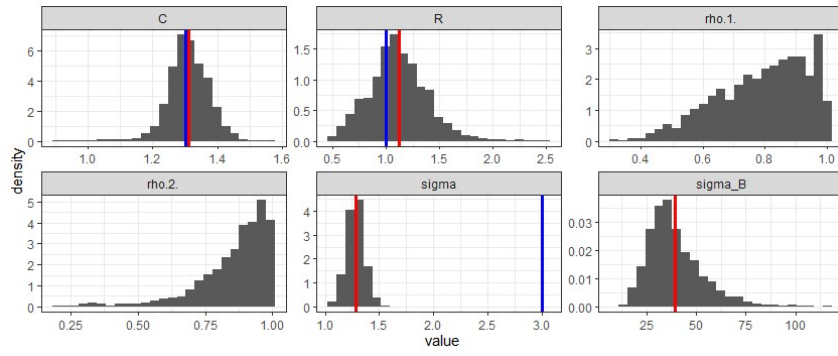
Figure 6.38

(a) Simulation Study 3: Full Calibration -
Small CI for dependent noise ϵ_t - iteration 3
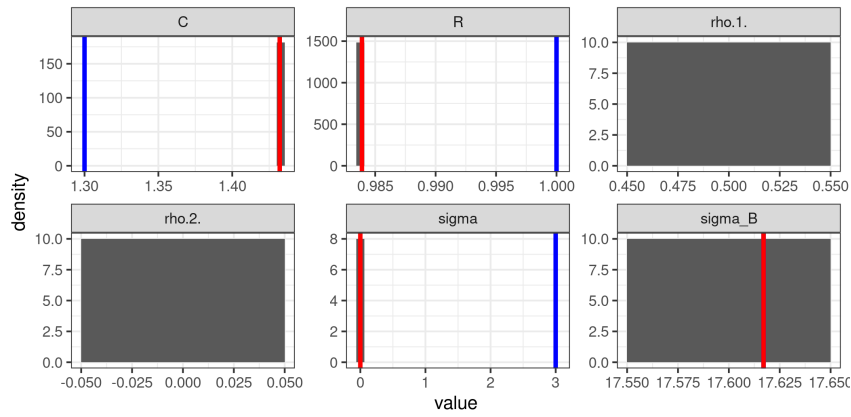

(b) Simulation Study 3: Mod 1-
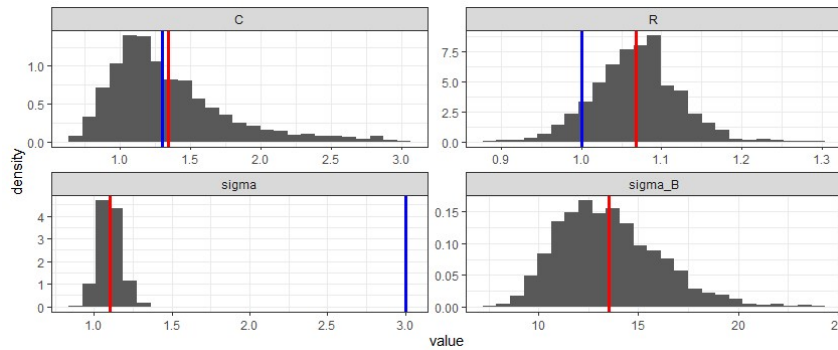Small CI dependent noise ϵ_t - iteration 30
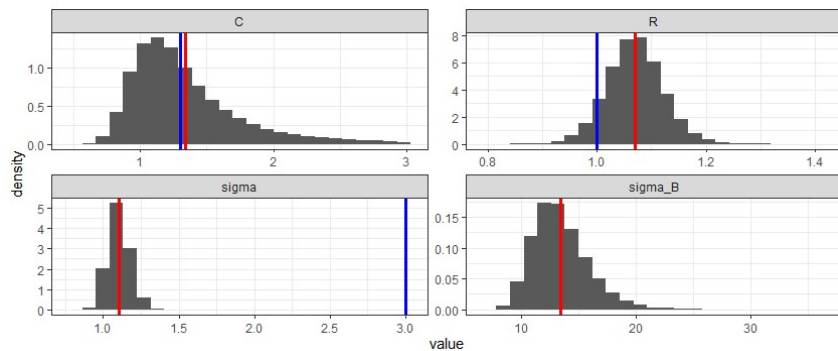
Figure 6.39

(a) Simulation Study 3: Full Calibration -
Small CI for dependent noise ϵ_t - short - theta 1

(b) Simulation Study 3: Full Calibration -
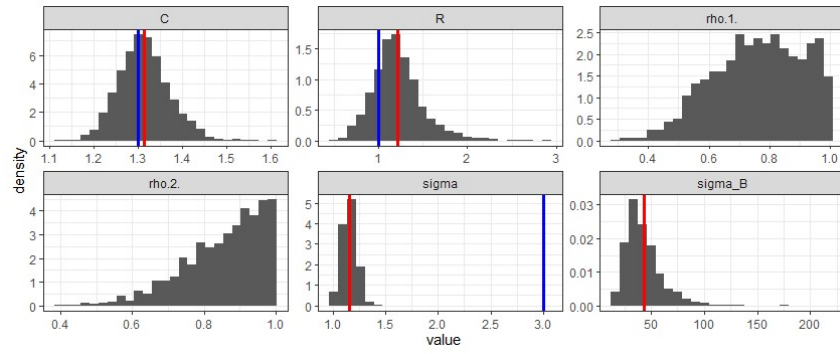Small CI for dependent noise ϵ_t - long - theta 1

(c) Simulation Study 3: Mod 1-
Small CI dependent noise ϵ_t - short - theta1
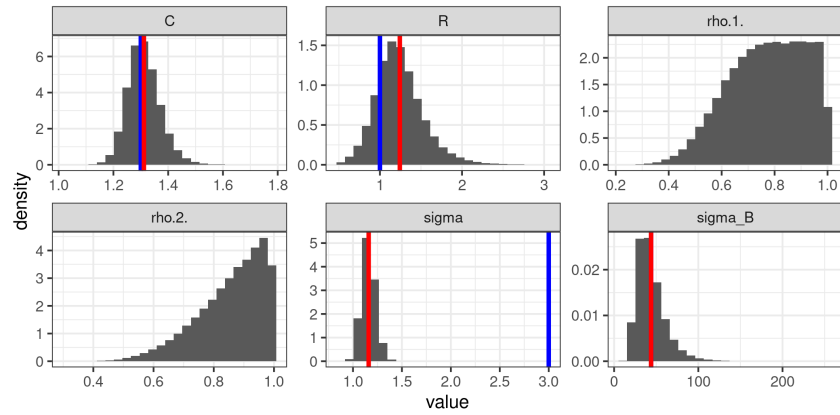

(d) Simulation Study 3: Mod 1-
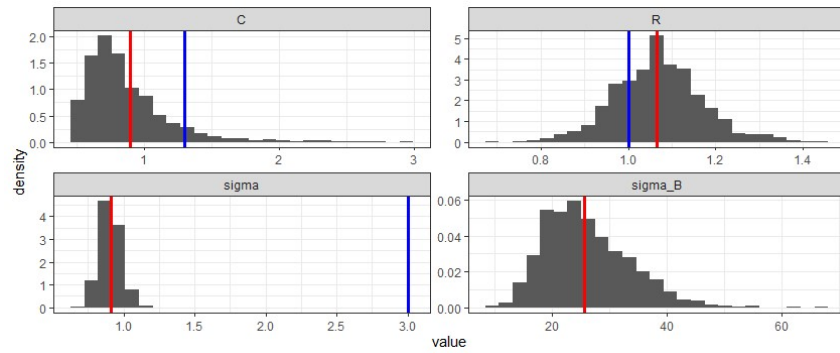Small CI dependent noise ϵ_t - long - theta1

Figure 6.40

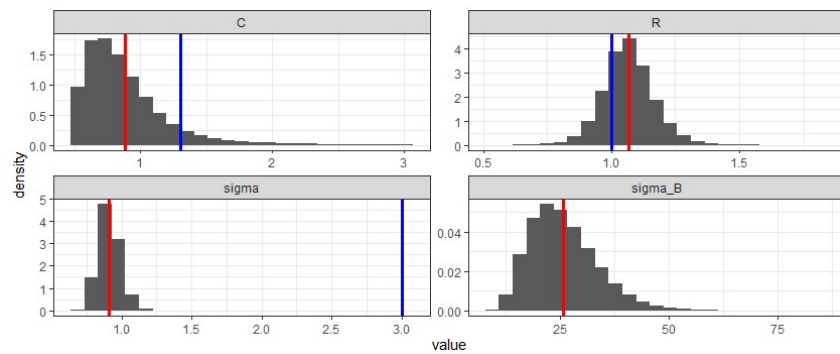**(a)** Simulation Study 3: Full Calibration -
Big CI for dependent noise ϵ_t - short - theta 1

(b) Simulation Study 3: Full Calibration -
Big CI for dependent noise ϵ_t - long - theta 1

(c) Simulation Study 3: Mod 1-
Big CI dependent noise ϵ_t - short - theta 1

(d) Simulation Study 3: Mod 1-
Big CI dependent noise ϵ_t - long - theta 1

Figure 6.41

References

[1] WHO. Hypertension. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension (visited on 04/30/2021).

[2] MyMDT. Mymdt. 2021. URL: https://www.ntnu.edu/digital-transformation/twin (visited on 04/30/2021).

[3] N. Westerhof, J.-W. Lankhaar, and B. E. Westerhof. The arterial Windkessel. en. *Medical & Biological Engineering & Computing*, 47(2):131–141, Feb. 2009. ISSN: 0140-0118, 1741-0444. DOI: 10 . 1007 / s11517 - 008 - 0359 - 2. URL: http://link.springer.com/10.1007/s11517-008-0359-2 (visited on 01/20/2021).

[4] P. Segers, E. R. Rietzschel, M. L. D. Buyzere, N. Stergiopulos, N. Westerhof, L. M. V. Bortel, T. Gillebert, and P. R. Verdonck. Three- and four-element Windkessel models: Assessment of their fitting performance in a large cohort of healthy middle-aged individuals. en. 222:14, 2008.

[5] J. C. Silva. Diastole vs. systole: What is the difference? en, Apr. 2018. (Visited on 01/10/2022).

[6] M. Spitieris, I. Steinsland, and E. Ingestrom. Bayesian calibration of Arterial Windkessel Model. *arXiv:2201.06883 [stat]*, Jan. 2022. URL: http://arxiv.org/abs/2201.06883 (visited on 03/04/2022). arXiv: 2201.06883.

[7] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, Oct. 2014. DOI: 10 . 1088 / 0266 - 5611 / 30 / 11 / 114007. URL: https://doi.org/10.1088/0266-5611/30/11/114007.

[8] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne. Combining Field Data and Computer Simulations for Calibration and Prediction. en. *SIAM Journal on Scientific Computing*, 26(2):448–466, Jan. 2004. ISSN: 1064-8275, 1095-7197. DOI: 10 . 1137 / S1064827503426693. URL: http://epubs.siam.org/doi/10.1137/S1064827503426693 (visited on 11/30/2021).

[9] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. en:40.

[10] J. O. Berger and L. A. Smith. On the Statistical Formalism of Uncertainty Quantification. en. *Annual Review of Statistics and Its Application*, 6(1):433–460, Mar. 2019. ISSN: 2326-8298, 2326-831X. DOI: 10 . 1146 / annurev - statistics - 030718 - 105232. URL: https :

//www.annualreviews.org/doi/10.1146/annurev-statistics-030718-105232 (visited on 01/20/2021).

[11] G. Karagiannis, B. A. Konomi, and G. Lin. On the Bayesian calibration of expensive computer models with input dependent parameters. *Spatial Statistics*, 34:100258, Dec. 2019. ISSN: 22116753. DOI: 10.1016/j.spasta.2017.08.002. URL: http://arxiv.org/abs/1708.09824 (visited on 12/16/2021). arXiv: 1708.09824.

[12] R. B. Gramacy. *Surrogates*. Taylor Francis Group, LLC, Virginia Tech, 2020.

[13] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A Framework for Validation of Computer Models. en. 49(2):18, 2007.

[14] B. Shaby and M. T. Wells. Exploring an Adaptive Metropolis Algorithm. en:31.

[15] *NimbleUserManual*. URL: https://r-nimble.org/html_manual/cha-mcmc.html (visited on 11/30/2021).

[16] P. e. de Valpine. Samplers: MCMC Sampling Algorithms in nimble: MCMC, Particle Filtering, and Programmable Hierarchical Modeling. en. URL: https://rdrr.io/cran/nimble/man/samplers.html (visited on 12/17/2021).

[17] R. E. Klabunde. CV Physiology | Arterial Blood Pressure. URL: https://www.cvphysiology.com/Blood%5C%20Pressure/BP002 (visited on 01/14/2022).

[18] B. Williams et.al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. en-US. *Journal of Hypertension*, 36(10):1953–2041, Oct. 2018. ISSN: 0263-6352. DOI: 10.1097/HJH.0000000000001940. URL: https://journals.lww.com/jhypertension/Fulltext/2018/10000/2018_ESC_ESH_Guidelines_for_the_management_of.2.aspx (visited on 01/13/2022).

[19] NTNU. Cardiovascular biomechanics. URL: https://folk.ntnu.no/leifh/teaching/tkt4150/._main026.html (visited on 01/13/2022).

[20] R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau. Bayesian statistics and modelling. en. *Nature Reviews Methods Primers*, 1(1):1–26, Jan. 2021. ISSN: 2662-8449. DOI: 10.1038/s43586-020-00001-2. URL: https://www.nature.com/articles/s43586-020-00001-2 (visited on 01/14/2022). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Scientific community;Statistics Subject_term_id: scientific-community;statistics.

[21] C. P. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. en. Springer texts in statistics. Springer, New York, 2nd ed edition, 2007. ISBN: 978-0-387-71598-8.

[22] H.-T. Lam. *Bayesian Calibration and Inference for Multiple Machines*, 2019.

[23] S. Brooks. Markov chain Monte Carlo method and its application. en. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):69–100, 1998. ISSN: 1467-9884. DOI: 10.1111/1467-9884.00117. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9884.00117 (visited on 01/21/2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9884.00117.

[24] G. O. Roberts and J. S. Rosenthal. Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009. ISSN: 1061-8600. URL: `https://www.jstor.org/stable/25651249` (visited on 12/16/2021). Publisher: [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America].

[25] NIMBLE – An R package for programming with BUGS models and compiling parts of R. en-US. URL: `https://r-nimble.org/` (visited on 03/07/2022).

[26] scikit-learn. Sklearn.gaussian_process.kernels.Matern. en. URL: `https://scikit-learn/stable/modules/generated/sklearn.gaussian_process.kernels.Matern.html` (visited on 03/07/2022).

[27] M. Betancour. Robust Gaussian Processes in Stan, 2017. URL: `https://betanalpha.github.io/assets/case_studies/gp_part3/part3.html` (visited on 02/10/2022).

Helene Minge Olsen

Master's Thesis

NTNU
Norwegian University of
Science and Technology