



ENGINEERING MATHEMATICS
AND COMPUTING LAB



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Learning model discrepancy of an electric motor with Bayesian inference

David John, Michael Schick, Vincent Heuveline

Preprint No. 2018-01

Preprint Series of the Engineering Mathematics and Computing Lab (EMCL)



<http://emcl.iwr.uni-heidelberg.de>



Preprint Series of the Engineering Mathematics and Computing Lab (EMCL)

ISSN 2191-0693

Preprint No. 2018-01

The EMCL Preprint Series contains publications that were accepted for the Preprint Series of the EMCL. Until April 30, 2013, it was published under the roof of the Karlsruhe Institute of Technology (KIT). As from May 01, 2013, it is published under the roof of Heidelberg University.

A list of all EMCL Preprints is available via Open Journal System (OJS) on <http://archiv.ub.uni-heidelberg.de/ojs/index.php/emcl-pp/>

For questions, please email to

info.at.emcl-preprint@uni-heidelberg.de

or directly apply to the below-listed corresponding author.

Affiliation of the Authors

David John^{a,b,1}, Michael Schick^b, Vincent Heuveline^{a, c}

^a*Engineering Mathematics and Computing Lab (EMCL), Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany*

^b*Corporate Research, Robert Bosch GmbH, Germany*

^c*Data Mining and Uncertainty Quantification Group (DMQ), Heidelberg Institute for Theoretical Studies (HITS) GmbH, Germany*

¹*Corresponding Author: David John, david.john@de.bosch.com*

Impressum

Heidelberg University

Interdisciplinary Center for Scientific Computing (IWR)

Engineering Mathematics and Computing Lab (EMCL)

Im Neuenheimer Feld 205,

69120 Heidelberg

Germany

Published on the Internet under the following Creative Commons License:

<http://creativecommons.org/licenses/by-nc-nd/3.0/de> .



<http://emcl.iwr.uni-heidelberg.de>

Learning model discrepancy of an electric motor with Bayesian inference

David John, Michael Schick, Vincent Heuveline

August 23, 2018

Abstract

Uncertainty Quantification (UQ) is highly requested in computational modeling and simulation, especially in an industrial context. With the continuous evolution of modern complex systems demands on quality and reliability of simulation models increase. A main challenge is related to the fact that the considered computational models are rarely able to represent the true physics perfectly and demonstrate a discrepancy compared to measurement data. Further, an accurate knowledge of considered model parameters is usually not available. E.g. fluctuations in manufacturing processes of hardware components or noise in sensors introduce uncertainties which must be quantified in an appropriate way. Mathematically, such UQ tasks are posed as inverse problems, requiring efficient methods to solve. A popular approach for UQ in inverse problems is Bayesian inference. This work investigates the influence of model discrepancies onto the calibration of physical model parameters and further considers a Bayesian inference framework including an attempt to correct for model discrepancy by an additional term. A Markov Chain Monte Carlo (MCMC) method is utilized to approximate the posterior distribution. A polynomial expansion with unknown coefficients is used to approximate and learn model discrepancy and system parameters simultaneously. This work extends by discussion and specification of a guideline on how to define the model discrepancy term complexity, i.e. here the maximum polynomial degree, based on the available measurement data. Furthermore, the suggested method is applied to an electric motor model with synthetic measurement data and evaluated by comparing the results to the reference. The example illustrates the importance and promising perspective of the method by good approximation of discrepancy and parameters.

1 Introduction

The increasing complexity of technical systems yields high demands on model quality and numerical accuracy. The advancements in computational power allow detailed models to represent these systems, but uncertainty is still present in most models. Quantification of these models under uncertainty is desired to make statements about reliability and accuracy. Consequently, there is a need for efficient solvers and new additional methods for analysis, which are able to cope with the soaring complexity of models and that take into account all sources of uncertainty.

Uncertainty Quantification (UQ) methods focus on understanding, quantifying and propagating uncertainty in computational simulation of models. Uncertainty arises from different sources and is often grouped in two categories: aleatoric and epistemic uncertainties. Aleatoric uncertainties are inherent to the model itself and are not due to insufficient knowledge. They are stochastic variations that cannot be reduced, e.g. by performing more measurements. Epistemic uncertainties are due to model assumptions, parameterizations and discretization. If a model for a system of interest is available, then epistemic uncertainty is subdivided into model form uncertainty and parametric uncertainty. The first category arises if reasonable doubts exist that the model itself is structurally correct and the latter, if the model reflects reality well enough, but the parameters are uncertain [19, 22].

To be more specific we formulate the general forward problem. Let $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{Y}$, for some spaces $\mathcal{X} \subseteq \mathbb{R}^d, \mathcal{Y} \subseteq \mathbb{R}^n, d, n \in \mathbb{N}$, be an operator representing a simulation model that maps some input $x \in \mathcal{X}$ to an output $y = \mathcal{G}(x)$. In general \mathcal{G} is nonlinear. If this x is uncertain, one might formulate it as a random variable X in a probability space (Ω, Σ, P) . With the random variable X the output $Y = \mathcal{G}(X)$ is a random variable as well, with an unknown probability distribution of Y . The task of forward uncertainty quantification methods is now to characterize the distribution of Y with appropriate methods, e.g. with the Monte Carlo method.

On the other hand the inverse problem describes the task of finding a $x \in \mathcal{X}$ for a given measurement $y \in \mathcal{Y}$ such that $y = \mathcal{G}(x)$. Generally, this equality does not hold as the measurements y are usually corrupted by observation noise. Hence, one considers

$$y = \mathcal{G}(x) + \eta, \quad (1)$$

where $\eta \in \mathcal{Y}$ represents the observation noise due to measurement noise or model error. Now, simply inverting \mathcal{G} is not possible as η is unknown and in general \mathcal{G}^{-1} is non-existent. A classical approach to solve this problem is by minimizing the data misfit, i.e. $\min_{x \in \mathcal{X}} \frac{1}{2} \|y - \mathcal{G}(x)\|_{\mathcal{Y}}^2$. However, this problem is typically ill-posed in the sense of Hadamard, i.e. multiple solutions might exist and stability might be a problem. In order to obtain a well-posed problem regularization is necessary. One approach is Tikhonov regularization, also known as ridge regression in statistics, where a regularization term $R(x)$ is added to the data misfit, i.e. $\min_{x \in \mathcal{X}} \frac{1}{2} \|y - \mathcal{G}(x)\|_{\mathcal{Y}}^2 + \lambda R(x)$, for a $\lambda > 0$ [10]. An example is $R(x) = \|\hat{x} - x\|^2$ for some norm $\|\cdot\|$ and a fixed $\hat{x} \in \mathcal{X}$. However, the choice of regularization is somewhat arbitrary [21]. The Bayesian approach to inverse problems yields a natural regularization of ill-posed problems [4, 10, 21] and is furthermore a popular approach for UQ in inverse problems. With the Bayesian approach the problem is formulated in terms of probability theory and one is interested in finding a probability measure $\mu^y(x)$ on \mathcal{X} with probability density $\pi(x|y)$ that expresses how likely a certain $x \in \mathcal{X}$, i.e. $\mathcal{G}(x)$, describes the given data y under consideration of the noise. The inverse problem in this setting is well-posed under slight assumptions and leads to a natural way of regularization due to the definition of prior distributions for unknown parameters, see [21]. Usually, the problem is formulated as $y = \mathcal{G}(x) + \eta$, where η is modeled as a realization of a zero mean random variable with probability density ρ , e.g. $\eta \sim N(0, \Sigma)$ with unknown covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. The probability of y given x has density $\pi_{\eta}(y|x) := \rho(y - \mathcal{G}(x))$ and is denoted as likelihood. Let $\pi_0(x)$ be the prior density expressing prior knowledge of the unknown $x \in \mathcal{X}$. Bayes' formula leads to

$$\pi(x|y) = \frac{\pi_{\eta}(y|x)\pi_0(x)}{\pi(y)}, \quad (2)$$

where $\pi(y) = \int_{\mathcal{X}} \pi(x, y) dx$ is called marginal or evidence. The evidence works as a normalizing constant and is usually omitted, i.e. only the relation $\pi(x|y) \sim \pi_{\eta}(y|x)\pi_0(x)$ is considered. Generally, the posterior distribution $\pi(x|y)$ is intractable and one can not sample from it directly, hence approximative methods are required. The approximation can be done by filtering, variational and sampling methods [5, 21]. An example for the latter is Metropolis-Hastings Markov-Chain Monte-Carlo (MH-MCMC) sampling, which is utilized in this work [5, 8, 17].

A difficulty in solving the inverse problem is to capture all sources of uncertainty and correctly identify them, often called "identification problem". Uncertainty sources are

- **parametric uncertainty**, e.g. either unknown model parameters (epistemic uncertainty) that are completely unknown or only known to a certain degree for instance by a probability distribution expressing an a-prior degree of believe, or variable / stochastic parameters (aleatoric uncertainty) or even both, unknown stochastic parameters,
- **structural model uncertainty**, e.g. lack of knowledge, missing physics, in modeling the system of interest,
- **solution method uncertainty**, e.g. errors introduced by implementation and also by numerical approximation schemes and
- **observation uncertainty**, e.g. noise and errors that are introduced by observing and recording quantities of interest of the system under consideration.

These categories are not necessarily selective and single uncertainty sources might not uniquely be associated to one single category.

For a full quantification of uncertainties in the inverse problem all sources should be considered. A main challenge is related to the fact that the considered computational models \mathcal{G} are rarely able to represent the true physics perfectly and demonstrate a discrepancy compared to measurement data. To be more specific the term model discrepancy in this work denotes the difference between simulation model for the true physical parameters and the true system, hence the structural uncertainty that might be due to missing physics in the formulation of the simulation model. It also comprises numerical errors. However, this model discrepancy is usually unknown. Considering model discrepancy in inverse problems is crucial in order to obtain realistic calibrations of unknown physical parameters [3] and to further obtain a quantification of model quality and reliability, which is important for most engineering tasks. In the following we give a brief overview of previous work considering model discrepancy in inverse problems.

The Kennedy and O’Hagan framework [11] is one of the first attempts to model and explicitly take account of all the uncertainty sources that arise in the calibration of computer models. In particular, model discrepancy is considered by an additional term in the Bayesian formulation of the inverse problem. Gaussian Processes are then used to model the model discrepancy and the simulator. One conclusion is that one only calibrates and receives the best fitting parameters according to the given data and defined error structure. These obtained parameter estimates do not necessarily correspond to the true physical values. Often the computer model does explain the data better for different values and restriction of some calibration parameters to the true physical value might lead to worse results. Following [11], Arendt et. al. [1, 2] suggest a modular Bayesian approach, where first the hyperparameters of the Gaussian processes for the simulator and the model discrepancy term are approximated by maximum likelihood estimates consecutively. Then the posterior distribution of the parameters conditioned on the data and the hyperparameter estimates are approximated. Further, the identification problem, i.e. the problem of distinguishing between effects of the calibration parameters and the model discrepancy, is discussed. Examples illustrate that sometimes this separation is possible under mild assumptions, e.g. smoothness of the model discrepancy, but also that it is not possible in other cases. In the companion work [2] they show an approach how to improve identifiability by using multiple responses and representing correlation between responses. Another work using multiple responses is [14]. Brynjarsdóttir and O’Hagan [3] state that with the Kennedy and O’Hagan framework [11] in order to infer physical parameters and model discrepancy simultaneously a sufficient prior distributions for at least one of those must be given. In an example they show that a constrained Gaussian Process prior for the model discrepancy, including best and most realistic prior information, yields good results for interpolation and learning about physical parameters, but still seems to be bad for extrapolation. Hence, simply introducing model discrepancy with weak prior information is not enough. However, Tuo and Wu [24, 25] showed that the choice of the model discrepancy prior has a permanent influence onto the parameter posterior distribution even in the large data limit. Plumlee [15] presented an approach to improve identifiability by defining a prior distribution of the model discrepancy that is orthogonal to the gradient of the model. Nagel et. al. [13] modeled the model discrepancy term by an polynomial expansion, assuming smoothness for the true underlying model discrepancy.

Following all these works we consider a Bayesian model with a term for measurement noise and a model discrepancy term in order to infer parameters of simulation model for given measurement data. This work adapts the idea of representing model discrepancy as a polynomial expansion [13]. The major contribution of this work is to provide answers on how to select a polynomial degree by keeping the complexity of the model low while still providing high accuracy in discrepancy modeling. This is shown in a practical guideline, which recommends how to select a sufficient maximum polynomial degree of the truncated polynomial expansion, based on the available data and the estimation of measurement noise. Furthermore, critical points conditioning the choice of the model discrepancy term are discussed in detail.

The framework is then applied to the calibration of a direct current (DC) electric motor model under consideration of model inadequacy, measurement uncertainty and parametric uncertainty in a synthetic setting. I.e. synthetic measurements are created and a modified electric motor model, containing an artificial model error are used to infer model parameters. Due to the synthetic setup of the example, reference parameter values and reference discrepancies are available and allow a quantitative evaluation of the considered methods performance and accuracy.

Considering all sources of uncertainties increases the complexity of the problem one needs to solve and

consequently increases computational cost. Hence, efficient methods need to be developed to leverage this drawback. One possibility is to increase the efficiency of the posterior approximation method, e.g. by using additional information (first or second order derivatives, e.g. gradient, Hessian) of the problem [9, 18, 20]. Another option is to replace the simulation model by an surrogate to leverage computational cost. In this work surrogate modeling is omitted for the sake of simplicity, but we emphasize that in cases where simulation time is a bottleneck surrogate modeling, such as Polynomial Chaos expansions (PCE) [6, 27] or Gaussian processes (GPs) [11, 16], could leverage overall simulation time and speed up the inference.

This paper is structured in the following way: Section 2 formulates the considered example setup, by defining the electric motor model, an artificial model error and synthetic measurements. In Section 3 a Bayesian inference model considering model discrepancy and measurement noise is specified, followed by a discussion of the model discrepancy term complexity. The numerical results of the methods applied to the example are presented in Section 4. This work concludes in Section 5.

2 Electric motor model

The following introduces briefly the mathematical model of an electric motor and the corresponding numerical approximation. This is followed by the synthetic problem set up with synthetic measurement data creation and artificial model error.

2.1 Model differential equations

A direct current (DC) electric motor is a rotating electrical machine converting electrical energy into mechanical energy. It is used in cases where a DC voltage source is available and a variable motor speed operation is required. With the magnitude of the applied voltage the motor speed can be varied. Elements of a classical DC motors are a fixed stator and a movable rotor. One or more field windings on the stator generate a permanent magnetic field, in which the rotor is located. The rotor itself is surrounded by the armature winding consisting of several coils. When electrical current is applied to the armature coil a magnetic field is generated. Due to the permanent magnetic field caused by the stator a physical force, the Lorentz force, affects the rotor. This force causes the rotor to rotate. After a certain rotation, when the magnetic fields are aligned a commutator located on the rotor changes the poles such that the current and therefore the magnetic poles of the rotor reverse their direction. Thus, the rotor remains in its continuous torque. For controlling this system variables of interest are current and voltage on the electrical side and torque and speed on the mechanical side. The following two equations relate these variables [23]. Let I [A] denote the electric current and ω [rad/s] the angular velocity, i.e. the speed. For $t > 0$ the ordinary differential equations

$$L\dot{I}(t) = -RI(t) - c_m\omega(t) + U, \quad (3)$$

$$J\dot{\omega}(t) = c_g I(t) - d\omega(t) + T_L, \quad (4)$$

with initial conditions $I(0) = I_0 \in \mathbb{R}, \omega(0) = \omega_0 \in \mathbb{R}$, describe the electro-mechanical behavior of a DC electric motor. The parameters $R, c_m, c_g, U, d, L, J \in \mathbb{R}$ denote the cable harness resistance R [Ω], motor constants c_m [Vs/rad], c_g [Nm/A], voltage U [V], friction d [$kg\ m/s$], inductivity L [H], inertia J [$kg\ m^2$] and constant torque T_L [Nm] required by the load. For the following only the resistance R is considered uncertain and all other parameters are fixed. With the initial conditions $I_0 = 0$ and $\omega_0 = 0$ and $T_L = 0$ the model describes the start of an electric motor from resting position without load. For the following the parameter values are set to the following arbitrarily chosen values: $R_0 = 0.1, c_m = 0.01, c_g = 0.01, U = 24, d = 10^{-6}, L = 10^{-4}, J = 5 * 10^{-5}$.

2.2 Numerical approximation

In order to approximate a solution of the electric current I and the angular velocity ω for $t \in [0, T], T = 0.5[s]$, we define equidistant discrete time points $t_i := i\Delta t$ for $i = 1, \dots, M, M \in \mathbb{N}$ with step size $\Delta t := T/M$. Now, let $\hat{I}(t_i)$ and $\hat{\omega}(t_i)$ denote numerical approximations of I and ω at time points t_i for a given time integration scheme. For this work an explicit Runge-Kutta method of order 4(5) is used, in

particular dopri5 (Dormand and Prince), see [26]. Consequently, the order of the numerical approximation error is $\mathcal{O}(\Delta t^4) = \mathcal{O}(10^{-12})$, for $T = 0.5$ and $M = 501$. As the measurements are created synthetically numerical approximation error can be neglected for this work. For notational convenience we define the forward model operator $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{Y}$, with $\mathcal{X} \subseteq \mathbb{R}$ and $\mathcal{Y} \subseteq \mathbb{R}^{2M}$, by

$$\mathcal{G}(R) := [\hat{I}(t_1), \dots, \hat{I}(t_M), \hat{\omega}(t_1), \dots, \hat{\omega}(t_M)], \quad (5)$$

as the operator delivering numerical approximations $\hat{I}(t_i), \hat{\omega}(t_i)$ at the discrete time points $t_i, i = 1, \dots, M$ depending on the parameter $R \in \mathcal{X}$ and for fixed parameters $c_g, c_m, U, d, L, J \in \mathbb{R}$.

2.3 Synthetic measurement

In order to generate synthetic measurements a realization of zero mean Gaussian distributed noise ε is added to the numerical approximation $\mathcal{G}(R_0)$ for a fixed $R_0 = 0.1$, i.e.

$$\mathbf{y} := \mathcal{G}(R_0) + \varepsilon, \quad (6)$$

where $\varepsilon = [\varepsilon_I, \varepsilon_\omega] \sim N(0, \Sigma)$, with covariance matrix $\Sigma \in \mathbb{R}^{2M \times 2M}$. For independent and identically distributed measurement noise for current and angular velocity at each time step the covariance is defined as $\Sigma := \text{diag}(\Sigma_I, \Sigma_\omega)$, where $\Sigma_I, \Sigma_\omega \in \mathbb{R}^{M \times M}$ and $\Sigma_{I,\omega} := \text{diag}(\sigma_{I,\omega})$, with $\sigma_{I,\omega} > 0$. Figure 1 displays one measurement of I and ω in the time interval $[0, 0.5]$ seconds with $\sigma_{I,0} = 2, \sigma_{\omega,0} = 10$ and number of time steps $M = 501$.

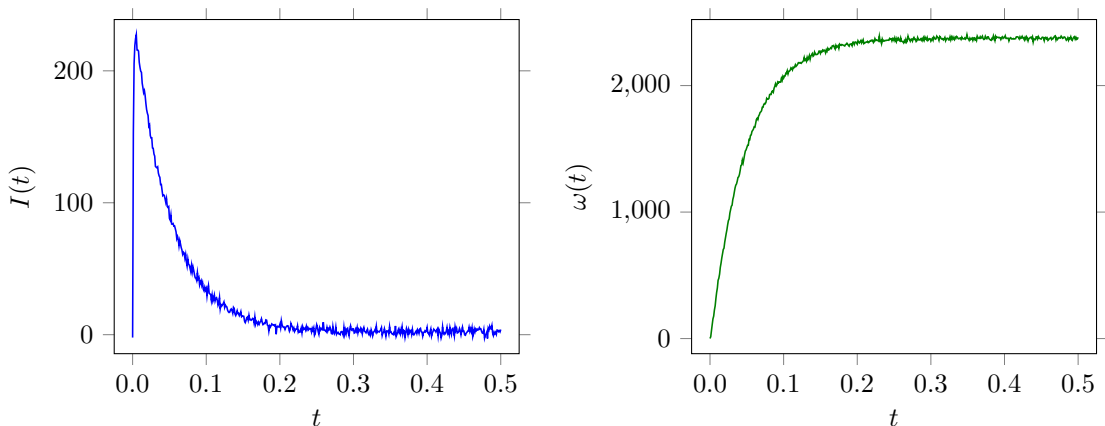


Figure 1: Noisy synthetic measurements of electric current $I(t)$ and angular velocity $\omega(t)$ at $M = 501$ equidistant time points $t_i, i = 1, \dots, M$ in the time interval $[0, 0.5]$ seconds.

2.4 Artificial model error

As the goal of this work is to learn about model discrepancy an artificial model error $\phi \in \mathbb{R}$ is introduced into the equations of the electric motor

$$L\dot{I}(t) = -RI(t) - c_m\omega(t) + U, \quad (7)$$

$$J\dot{\omega}(t) = \phi c_g I(t) - d\omega(t) + T_L. \quad (8)$$

Let $\mathcal{G}_\phi(R)$ denote the numerical approximation of the model for an artificial model error factor ϕ . Obviously, for $\phi = 1$ the identity $\mathcal{G} = \mathcal{G}_{\phi=1}$ holds, if the same numerical approximation method and the same time step is used. In the following, \mathcal{G}_ϕ is referred to as simulation model.

If $\phi \neq 1$, this can be interpreted as missing physics, i.e. by damping or amplifying a part of the equation. Or it can be interpreted as a misspecified model, if the parameter c_g is fixed. However, for this showcase it is not important how to interpret this artificial model error. For the following it is sufficient

that the approximation of the model with $\phi \neq 1$ has some discrepancy to the model with $\phi = 1$. And this is the case, if c_g is fixed to the same value for both models.

Define by

$$\mathbf{d}(R) = [\mathbf{d}_I, \mathbf{d}_\omega]^T(R) := \mathcal{G}(R_0) - \mathcal{G}_\phi(R), \quad (9)$$

$$\mathbf{d}^\varepsilon(R) = [\mathbf{d}_I^\varepsilon, \mathbf{d}_\omega^\varepsilon]^T(R) := \mathbf{y} - \mathcal{G}_\phi(R), \quad (10)$$

the discrepancy $\mathbf{d}(R) \in \mathbb{R}^{2M}$ and the noisy discrepancy $\mathbf{d}^\varepsilon(R) \in \mathbb{R}^{2M}$, depending on $R \in \mathcal{X}$. The discrepancy $\mathbf{d}(R)$ is the difference between model $\mathcal{G}(R_0)$ with reference value R_0 , used for generation of the synthetic measurements, and the simulation model $\mathcal{G}_\phi(R)$ for $R \in \mathcal{X}$. The noisy discrepancy $\mathbf{d}^\varepsilon(R)$ is the difference between measurements \mathbf{y} and the simulation model $\mathcal{G}_\phi(R)$ for $R \in \mathcal{X}$. The reference discrepancies $\mathbf{d}_0 := \mathbf{d}(R_0)$ and $\mathbf{d}_0^\varepsilon := \mathbf{d}^\varepsilon(R_0)$ for $\phi = 0.9$ and R_0 are displayed in Figure 2 for I and ω , respectively.

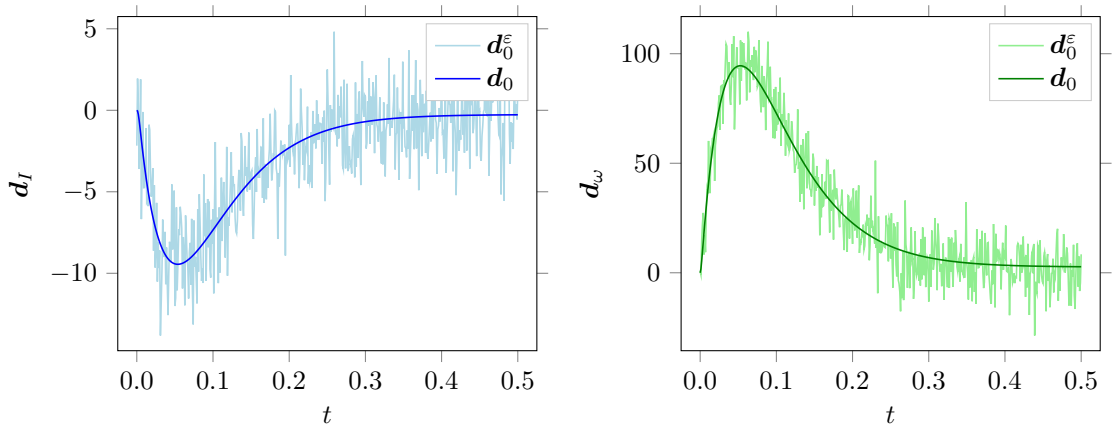


Figure 2: Reference discrepancy \mathbf{d}_0 and \mathbf{d}_0^ε of electric current $I(t)$ and angular velocity $\omega(t)$.

3 Bayesian inference solution process

This section considers methods addressing the inverse problem of finding an $R \in \mathcal{X}$ such that $\mathbf{y} = \mathcal{G}_\phi(R)$, where $\mathbf{y} \in \mathbb{R}^{2M}$ are noisy measurements and \mathcal{G}_ϕ is the simulation model containing an artificial model error ϕ . The goal here is not just to infer an optimal $R \in \mathcal{X}$, but also one that is as close as possible to the reference R_0 , i.e. the true physical value, which was used to create the measurements \mathbf{y} . Additionally we consider and learn about the model discrepancy during inference.

As already reasoned in the introduction the Bayesian approach to inverse problems has certain advantages to address the inverse problem. Regularization comes along naturally by defining eventually existing initial knowledge about unknown parameters as prior probability measure with density π_0 . Also measurement noise is considered separately. Furthermore, methods considering model discrepancy can be easily formulated within the Bayesian framework as it allows to define a prior distribution for the model discrepancy term. As we are interested in separating several sources of uncertainty it is natural to formulate the problem in terms of probability theory and use Bayesian inference.

Following [13] we introduce two Bayesian models that differ in their complexity. The simpler model only considers identically and independent distributed (i.i.d.) measurement noise, whereas the more complex model considers model discrepancy additionally.

In this work sampling methods are utilized to approximate the posterior distribution probability density function, in particular a Metropolis-Hastings Markov-Chain Monte-Carlo (MH-MCMC) method [5]. We use the MH-MCMC implementation of the Python package PyMC3 [17].

3.1 Bayesian model 1 (BM1): measurement noise

As the measurements \mathbf{y} are noisy, additive noise $\boldsymbol{\varepsilon}$ is added to the simulation model

$$\mathbf{y} = \mathcal{G}(R) + \boldsymbol{\varepsilon}, \quad (11)$$

where $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_I, \boldsymbol{\varepsilon}_\omega]^T \in \mathbb{R}^{2M}$. Here, $\boldsymbol{\varepsilon}_I, \boldsymbol{\varepsilon}_\omega$ are considered as zero-mean Gaussian random variables $\mathcal{N}(\mathbf{0}, \sigma_I^2 \mathbf{I}_M)$ and $\mathcal{N}(0, \sigma_\omega^2 \mathbf{I}_M)$ with unknown standard deviations $\sigma_I, \sigma_\omega > 0$, respectively. Consequently, the **Bayesian model 1 (BM1)** is

$$\mathbf{y} = \mathcal{G}(R) + \boldsymbol{\varepsilon}(\sigma_I, \sigma_\omega), \quad (12)$$

considering only identically and independent distributed (i.i.d.) Gaussian measurement noise. Hence, the likelihood, i.e. the distribution of the measurements \mathbf{y} conditioned on the parameters $R, \sigma_I, \sigma_\omega$ is

$$\mathbf{y}|R, \sigma_I, \sigma_\omega \sim \mathcal{N}(\mathcal{G}(R), \Sigma(\sigma_I, \sigma_\omega)), \quad (13)$$

with $\Sigma(\sigma_I, \sigma_\omega) = \text{diag}(\sigma_I^2 \mathbf{I}_M, \sigma_\omega^2 \mathbf{I}_M)$. By expressing a-priori knowledge of the parameters $R, \sigma_I, \sigma_\omega$ in terms of a prior probability density function $\pi(R, \sigma_I, \sigma_\omega)$, Bayes' formula yields for the posterior probability density

$$\pi(R, \sigma_I, \sigma_\omega | \mathbf{y}) \sim \pi(\mathbf{y}|R, \sigma_I, \sigma_\omega) \pi(R, \sigma_I, \sigma_\omega). \quad (14)$$

I.e. the posterior probability density is proportional to the product of likelihood and prior up to a normalization constant. Let the prior be given by $\pi(R, \sigma_I, \sigma_\omega) = \pi(R) \pi(\sigma_I) \pi(\sigma_\omega)$. The priors for the unknown standard deviations of the noise $\pi(\sigma_I), \pi(\sigma_\omega)$ are defined as uninformative Inverse Gamma distributions $\text{InvGamma}(1, 1)$. This is a common choice for conjugate priors of scale parameters in Bayesian statistics [5], in particular for a Gaussian likelihood with given mean. The inverse Gamma distribution $\text{InvGamma}(1, 1)$ ensures the positiveness of $\sigma_I, \sigma_\omega > 0$, see Figure 6 for a plot of the density function of $\text{InvGamma}(1, 1)$.

3.2 Bayesian model 2 (BM2): measurement noise and model discrepancy

The second model approach extends the first one by considering model discrepancy with an additive term $\boldsymbol{\delta}$, additionally to additive i.i.d. measurement noise. Hence, the considered model is

$$\mathbf{y} = \mathcal{G}(R) + \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (15)$$

where $\boldsymbol{\delta} = [\boldsymbol{\delta}_I, \boldsymbol{\delta}_\omega]^T \in \mathbb{R}^{2M}$. Note that this is already the discretized version and $\boldsymbol{\delta}_I = [\delta_I(t_0), \dots, \delta_I(t_M)]$, $\boldsymbol{\delta}_\omega = [\delta_\omega(t_1), \dots, \delta_\omega(t_M)]$ are vectors of the, yet to define, model discrepancy terms $\delta_I, \delta_\omega : D \rightarrow \mathbb{R}$ evaluated at the discrete time points $t_i, i = 1, \dots, M$. Omitting the superscripts for a moment, we now assume that $\delta \in L^2(D)$. The space of square integrable functions $L^2(D)$ is equipped with the inner product $\langle f, g \rangle = \int_D f g d\mu$, where μ is a measure on D and $f, g \in L^2(D), D = [0, T]$. Let $\{p_j\}_{j \in \mathbb{N}} \subseteq L^2(D)$ be a basis of functions $p_j : D \rightarrow \mathbb{R}$ dense in $L^2(D)$, i.e. $\text{span}(\{p_j\}_{j \in \mathbb{N}}) = L^2(D)$. Then for all $\delta \in L^2(D)$ there exists $\{a_j\}_{j \in \mathbb{N}} \subseteq \mathbb{R}$ with $\sum_{j \in \mathbb{N}} |a_j|^2 < \infty$ such that δ can be represented by the expansion

$$\delta(t) = \sum_{j=0}^{\infty} a_j p_j(t). \quad (16)$$

For practicability reasons the expansion is truncated after a $K \in \mathbb{N}$

$$\delta(t) \approx \delta^K(t) = \sum_{j=0}^K a_j p_j(t). \quad (17)$$

Such an truncated expansion was also used in [13]. With this let, for the truncation parameter $K_I, K_\omega \in \mathbb{N}$, the truncated functional expansions

$$\delta_I^{K_I}(t) = \sum_{j=0}^{K_I} a_j^{(I)} p_j(t), \quad \delta_\omega^{K_\omega}(t) = \sum_{j=0}^{K_\omega} a_j^{(\omega)} q_j(t), \quad (18)$$

be approximative models for the model discrepancy terms δ_I, δ_ω . The bases $\{p_j\}_{j=0,\dots,K_I}$ and $\{q_j\}_{j=0,\dots,K_\omega}$ are not necessarily identical. Note that each expansion could be truncated with own truncation parameters K_I, K_ω , but for notational convenience and due to later usage we stick to $K = K_I = K_\omega$ for the following. Let $\mathbf{a} = [\mathbf{a}^{(I)}, \mathbf{a}^{(\omega)}] = [a_0^{(I)}, \dots, a_K^{(I)}, a_0^{(\omega)}, \dots, a_K^{(\omega)}] \in \mathbb{R}^{2K+2}$ be the vector containing all coefficients and $\boldsymbol{\delta}_I^K(\mathbf{a}) = [\delta_I^K(t_0), \dots, \delta_I^K(t_M)]$, $\boldsymbol{\delta}_\omega^K(\mathbf{a}) = [\delta_\omega^K(t_1), \dots, \delta_\omega^K(t_M)]$ be the vectors of the model discrepancy terms evaluated at the discrete time points $t_i, i = 1, \dots, M$. Hence $\boldsymbol{\delta}^K(\mathbf{a}) = [\boldsymbol{\delta}_I^K, \boldsymbol{\delta}_\omega^K]^T(\mathbf{a}) \in \mathbb{R}^{2M}$ denotes the approximation of the true underlying model discrepancy $\boldsymbol{\delta}$, depending on the coefficients \mathbf{a} and the truncation parameter K . The **Bayesian model 2 (BM2)** with all parameter dependencies is

$$\mathbf{y} = \mathcal{G}(R) + \boldsymbol{\delta}^K(\mathbf{a}) + \varepsilon(\sigma_I, \sigma_\omega). \quad (19)$$

The number of unknown parameters depends on K and is now $3 + 2K + 2$. With the additional unknown coefficients \mathbf{a} , the prior probability density function is defined as

$$\pi(R, \sigma_I, \sigma_\omega, \mathbf{a}) = \pi(R)\pi(\sigma_I)\pi(\sigma_\omega)\pi(\mathbf{a}), \quad (20)$$

where $\pi(\mathbf{a}) = \prod_{j=0}^K \pi(a_j^{(I)})\pi(a_j^{(\omega)})$. The prior distributions for $\pi(\sigma_I)$ and $\pi(\sigma_\omega)$ are specified as above. Now, with the likelihood

$$\mathbf{y}|R, \sigma_I, \sigma_\omega, \mathbf{a} \sim \mathcal{N}(\mathcal{G}(R) + \boldsymbol{\delta}^K(\mathbf{a}), \Sigma(\sigma_I, \sigma_\omega)), \quad (21)$$

the posterior is given by

$$\pi^K(R, \sigma_I, \sigma_\omega, \mathbf{a}|\mathbf{y}) \sim \pi(\mathbf{y}|R, \sigma_I, \sigma_\omega, \mathbf{a})\pi(R, \sigma_I, \sigma_\omega, \mathbf{a}), \quad (22)$$

where π^K denotes the dependence on K .

3.2.1 Choosing basis functions

If knowledge about the discrepancy is available, this should be modeled accordingly by defining an appropriate prior distribution for $\boldsymbol{\delta}$, i.e. in the case of $\boldsymbol{\delta}^K(\mathbf{a})$ for \mathbf{a}, K and of course by an appropriate choice of basis functions. However, in general this knowledge is not available and some modeling assumptions need to be made. Following the assumption above, the basis needs to be dense in $L^2(D)$. With the additional assumption that $\boldsymbol{\delta}$ is rather smooth, polynomials are a reasonable choice. For this let $p_j : D \rightarrow \mathbb{R}$ be a polynomial with polynomial degree $\deg(p_j) = j$ and $\{p_j\}_{j \in \mathbb{N}}$ be an orthonormal polynomial basis, with the orthogonality conditions $\langle p_j, p_k \rangle = \int_D p_j p_k d\mu$, for a measure μ on D . Let $w : D \rightarrow \mathbb{R}_+$ be the weighting function of μ . Then, for instance, the

- (a) Legendre polynomials for $w(t) = \frac{1}{|D|}$, with $|D| = T$ and
- (b) Tschebyschew polynomials for $w(t) = \frac{1}{\sqrt{1-t^2}}$

are dense in $L^2(D)$ and thus possible choices for the expansion. Of course there are further bases that are dense in $L^2(D)$. However, for our assumptions Legendre polynomials with the constant weighting function are a reasonable choice for $\{p_j\}_{j=0,\dots,K}$. This choice was also made in [13]. The first five Legendre polynomials are displayed in Figure 3.

3.2.2 Choosing prior distributions for the coefficients

In [13] they argue that the choice of the prior distribution for the coefficients is a bit delicate and somewhat arbitrary, as they have no physical meaning. Since no knowledge about the model discrepancy $\boldsymbol{\delta}$ is available we assume $\boldsymbol{\delta}^K(\mathbf{a})$ to be as small as possible by specifying priors that are centered around zero. Every around zero centered probability distribution with decaying tails should be sufficient. However, following [13], we opt for zero mean Laplace distributions, as it assign the highest probability around zero and decays exponentially towards the tails. The probability density function of *Laplace*(a, b) is $f(x) = \frac{1}{2b} \exp(-\frac{|x-a|}{b})$ for $x \in \mathbb{R}$ [12]. Compared to a Gaussian distribution, Laplace distribution enforces the zero a bit more and has slightly fatter tails. For a visual comparison of Laplace versus Gaussian

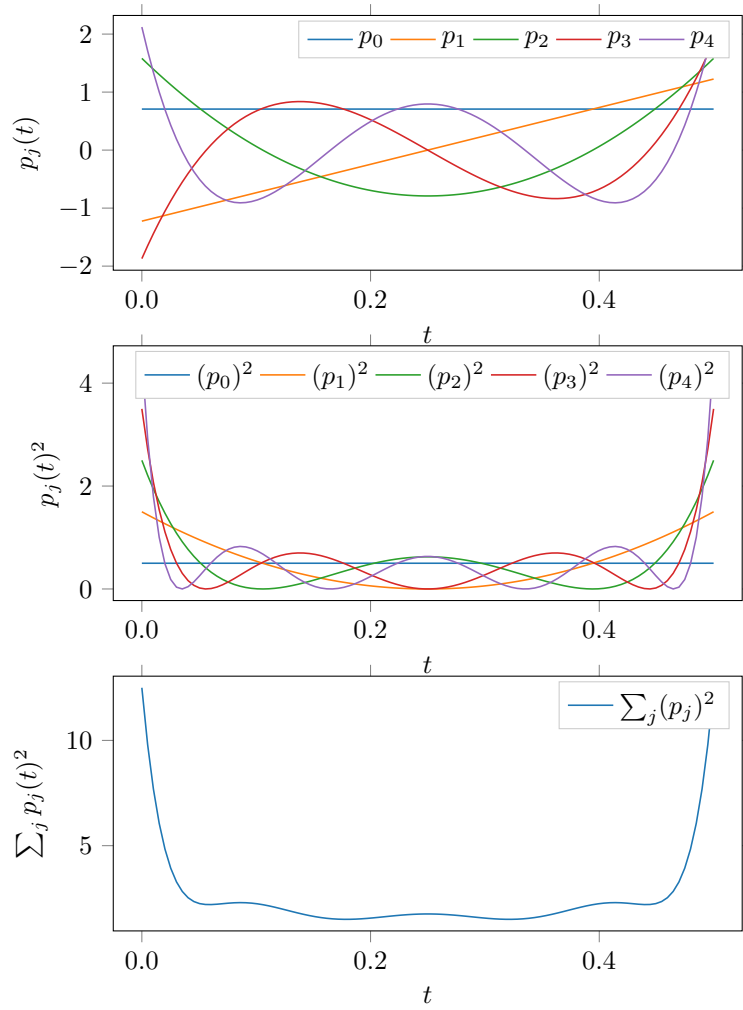


Figure 3: Left: Normalized Legendre polynomials. Middle: Squared polynomials. Right: Sum of squared polynomials.

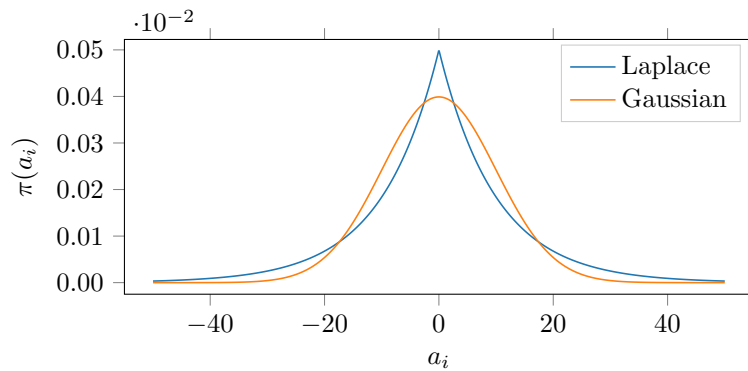


Figure 4: The probability density function of a Laplace distribution $Laplace(0, 10)$ compared to a Gaussian distribution $\mathcal{N}(0, 10^2)$.

distribution for the same specification, see Figure 4. Consequently, the $\pi(a_j^{(I)}), \pi(a_j^{(\omega)})$ are modeled as Laplace distributions with zero mean and variance 10, according to [13]. As the coefficients $a_j^{(I)}, a_j^{(\omega)}$ have independent zero mean priors the resulting distributions for $\delta_I^K, \delta_\omega^K$ are centered around zero as well. For the variance we obtain

$$V[\delta_I^K(t)] = \sum_{j=0}^K V[a_j^{(I)}] p_j^2(t), \quad V[\delta_\omega^K(t)] = \sum_{j=0}^K V[a_j^{(\omega)}] q_j^2(t). \quad (23)$$

Note that for the particular choice of the Legendre polynomials the variance of $\delta^K(t), \delta^K(t)$ is non-constant and varies with t , as displayed in Figure 3.

3.2.3 Choosing the truncation parameter K

The remainder of this section addresses the open question on how to specify an appropriate value for the truncation parameter K . With the choice of Legendre polynomials K now corresponds to the maximum polynomial degree K . This K determines the complexity of the model discrepancy term $\delta^K(\mathbf{a})$. Important factors in order to specify the model discrepancy term, or more specifically its complexity, are

- accuracy,
- computational costs,
- bias-variance tradeoff (underfitting, overfitting) and
- non-identifiability.

These points are important for a general definition of the model discrepancy term. Here, with the already made choice of Legendre polynomials and Laplace distribution for the polynomial coefficients, only the maximum polynomial degree K can be modified to adjust the complexity of $\delta^K(\mathbf{a})$. Hence, the factors listed above are explained in terms of the maximum polynomial degree K , which will serve as a synonym for the term model complexity. An optimal K should be large enough such that $\delta^K(\mathbf{a})$ is accurate enough to approximate the underlying discrepancy correct, but at the same time it should be as small as possible since a large K increases the number of unknown parameters and consequently computational costs. Furthermore, a large K might yield overfitting and non-identifiability of parameters. Overfitting occurs when for a large K the high degree polynomials start to reproduce oscillations of the measurement noise. With increasing K the discrepancy term $\delta^K(\mathbf{a})$ yields an increased flexibility and is able to approximate a growing class of functions. However, a remedy of this increased flexibility is a loss of information in the prior of $\delta^K(\mathbf{a})$. I.e. the prior gets non-informative, this might yield non-identifiability of all unknown parameters, depending on the information of the remaining model parameter prior distributions. In [3] they state that in order to infer model discrepancy and model parameters at the same time, at least for one of those an informative prior must be given. The term bias-variance tradeoff, illustrated in Figure 5, describes the fact that with increasing model complexity K the bias, i.e. the difference between the mean of the estimator and the reference value ($\mathbb{E}[R|\mathbf{y}] - R_0$), decreases, but at the same time the variance of the estimator $V[R|\mathbf{y}]$ increases with the model complexity [7]. Consequently, the mean square error (MSE) of R , which is the sum of squared bias and variance $MSE(R) = (\mathbb{E}[R|\mathbf{y}] - R_0)^2 + V[R|\mathbf{y}]$, is minimal for an optimal model complexity K_{opt} . With this optimal model complexity the contributions of bias and variance are somehow balanced. Models with an model complexity over this optimum are called overfitting (explaining the measurement data to good) and below underfitting (explaining the measurement data to bad).

Finally, taking all these factors into account the approach in this work on how to find an optimal K is following: Start with an initial $K = K_0 \in \mathbb{N}$ and increase K iteratively until the marginal posterior distribution $\pi(\sigma_I, \sigma_\omega|\mathbf{y})$ of the noise standard deviation stabilizes. I.e. that some distance $D(\pi^K(\sigma_I, \sigma_\omega|\mathbf{y}), \pi^{K+L}(\sigma_I, \sigma_\omega|\mathbf{y})) < tol$ for a given tolerance $tol > 0$ and $L = 1, 2, 3, \dots$. Then select the K such that this condition holds. Why is this sufficient? If a model discrepancy is present in BM1, then the noise term is the only instance to capture it. As the noise term is modeled with zero mean, the standard deviation might be overestimated consequently. By adding the model discrepancy term in BM2, this term, which is allowed to have a non-zero mean, captures, depending on its flexibility and degrees of freedom, a fraction of the model discrepancy. As a consequence, the noise term $\boldsymbol{\varepsilon}$ needs to represent only

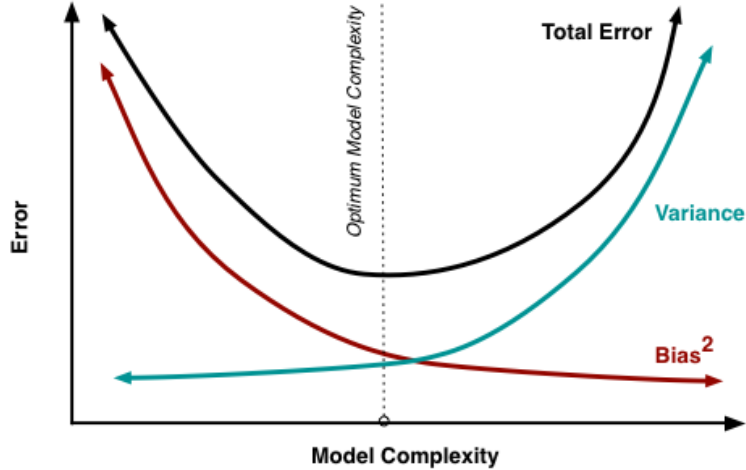


Figure 5: Bias-variance tradeoff: This figure shows the contribution of bias and variance to the total error. From <http://scott.fortmann-roe.com/docs/BiasVariance.html>

the remaining discrepancy and is estimated by a smaller value. Due to the smoothness assumption of $\delta^K(\mathbf{a})$, ε collects also everything that is not smooth enough. Now, if the estimated standard deviation of the noise does not change anymore (within the tolerance) from K to $K+L$ for $L = 1, 2, 3, \dots$ the smallest sufficient K is found. For this K the model discrepancy term $\delta^K(\mathbf{a})$ should represent the underlying model discrepancy appropriately and a separation of model uncertainty and measurement uncertainty is achieved.

4 Numerical results

In this section the numerical results for the two Bayesian models, BM1 and BM2, applied to the electric motor model are presented and compared to the reference, which is available due to the fully synthetic problem set up. Reference parameter values are $R_0 = 0.1$, $\sigma_{I,0} = 2$ and $\sigma_{\omega,0} = 10$. The reference discrepancy \mathbf{d}_0 and the noisy reference discrepancy \mathbf{d}_0^ε as defined in Section 2.4 are displayed in Figure 2. The prior distribution for the unknown parameter is chosen as Gaussian distribution $R \sim \mathcal{N}(R_0, 0.2R_0)$, with prior mean R_0 and variance $0.2R_0$. Variations of the prior mean by $\pm 15\%$ did not influence the upcoming results, thus simply the reference value was chosen as mean of the prior. The results are obtained by approximating the posterior distributions with MH-MCMC samples. The approximated marginal posterior distributions displayed in the following plots are kernel density estimates of the corresponding marginal MH-MCMC samples. Consequently, the marginal posterior moments are empirically approximated using Monte Carlo integration with a certain number of the obtained MH-MCMC samples.

4.1 Results Bayesian model 1

This section presents the numerical results obtained with Bayesian model 1 (BM1) for an artificial model error $\phi = 0.9$. Figure 6 displays the marginal MH-MCMC samples and corresponding posterior distributions for parameters R , σ_I and σ_ω in comparison to the reference values. With respect to a burn in phase, only the last 700 samples of three independently sampled Markov chains, each of total length 1500, are used to obtain the results. The chains stabilize, as can be seen in Figure 6, this confirms convergence. However, the sampling behavior is poor as the effective sample size is low. This is due to the highly informative data and the so-called concentration effect, where the posterior distribution concentrates on a small region, which makes sampling inefficient, see [18, 20] for further reading.

Figure 6 shows that the marginal posterior distribution for σ_ω is close to the reference, but for R and

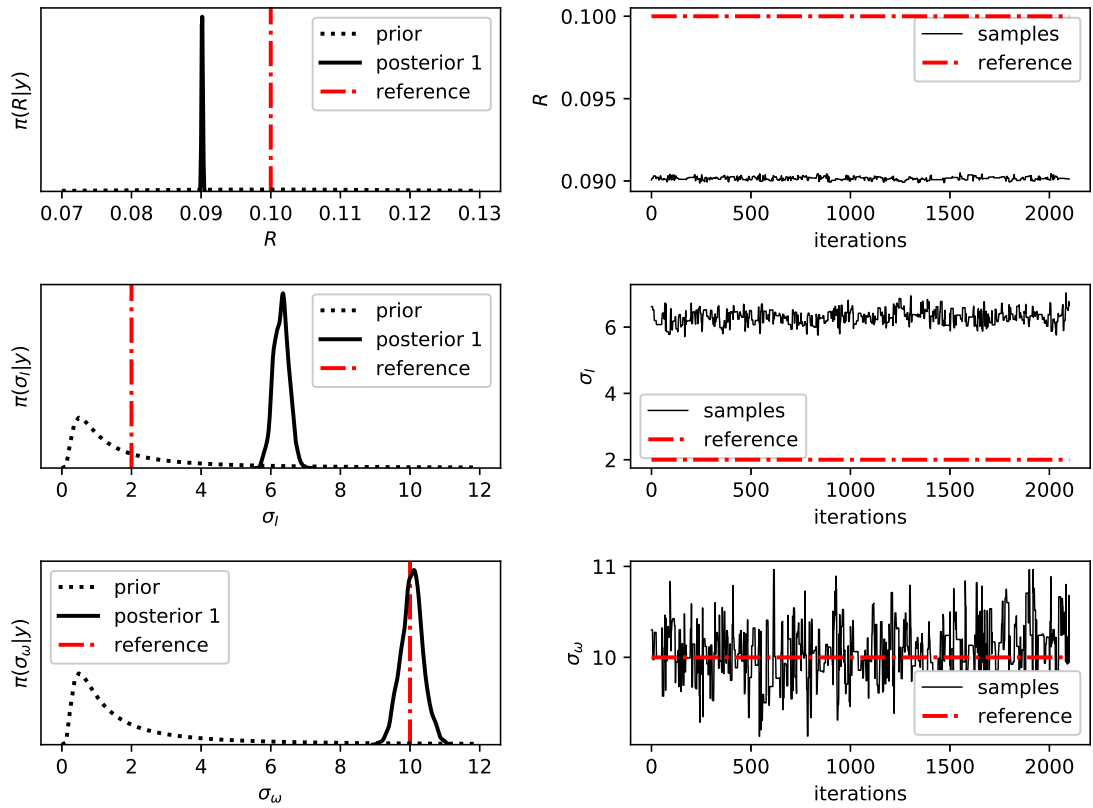


Figure 6: Posterior distributions and samples from BM1: The figures on the right hand side show marginal MH-MCMC samples of the parameters R , σ_I and σ_ω (from top to bottom). The figures on the left hand side show the kernel density estimates of the corresponding samples as an approximation of the marginal posterior distributions. The prior distributions are plotted as dotted black lines and the reference values as dashed red lines.

σ_I the marginal posteriors concentrate at values different to the reference values and do not even assign a significant probability to the reference values.

Definition 4.1. Define the relative error for a parameter estimate \hat{x} as

$$\epsilon_{rel}(x) = \frac{|\hat{x} - x_0|}{|x_0|}, \quad (24)$$

where x_0 is the reference value. \hat{x} might be the empirical approximation of the marginal posterior mean $\mathbb{E}[x|\mathbf{y}]$, i.e. $\hat{x} = \frac{1}{L} \sum_{l=1}^L x_l$, where $x_l, l = 1, \dots, L$ are a selection of MH-MCMC samples of the posterior density $\pi(x|\mathbf{y})$.

The marginal posterior mean values of R, σ_I and σ_ω and their respective relative error with respect to the reference values $R_0 = 0.1, \sigma_{I,0} = 2$ and $\sigma_{\omega,0} = 10$ are displayed in Table 1. The relative error $\epsilon_{rel}(\hat{R})$ is around 10%. Inserting the marginal posterior mean $\hat{R} \approx \mathbb{E}[R|\mathbf{y}]$ of R in the simulation model \mathcal{G}_ϕ yields

	marginal posterior mean			relative error		
	\hat{R}	$\hat{\sigma}_I$	$\hat{\sigma}_\omega$	$\epsilon_{rel}(\hat{R})$	$\epsilon_{rel}(\hat{\sigma}_I)$	$\epsilon_{rel}(\hat{\sigma}_\omega)$
BM1	9.03e-02	6.29e+00	1.03e+01	9.66e-02	2.14e+00	2.79e-02
BM2 ($K = 9$)	9.92e-02	2.02e+00	9.95e+00	8.48e-03	1.03e-02	4.89e-03

Table 1: Marginal posterior mean and relative error of parameters R, σ_I and σ_ω for BM1 and BM2 with $K = 9$, respectively. The relative errors are with respect to the reference values $R_0 = 0.1, \sigma_{I,0} = 2$, and $\sigma_{\omega,0} = 10$.

the noisy discrepancy $\mathbf{d}^\epsilon(\hat{R}) = [\mathbf{d}_I^\epsilon, \mathbf{d}_\omega^\epsilon]^T(\hat{R}) = \mathbf{y} - \mathcal{G}_\phi(\hat{R})$, displayed in Figure 7.

Remark 4.1. For simplicity the mean $\mathbb{E}[\mathbf{d}^\epsilon(R)|\mathbf{y}] = \mathbb{E}[\mathbf{y} - \mathcal{G}_\phi(R)|\mathbf{y}] = \mathbf{y} - \mathbb{E}[\mathcal{G}_\phi(R)|\mathbf{y}] \approx \mathbf{y} - \mathcal{G}_\phi(\mathbb{E}[R|\mathbf{y}])$ is approximated by assuming linearity for \mathcal{G}_ϕ . Thus only the mean of R needs to be computed. This does not cause large approximation errors as the posterior distribution of R has very small support and consequently linearity for \mathcal{G}_ϕ needs to be assumed in a small neighborhood of \hat{R} only.

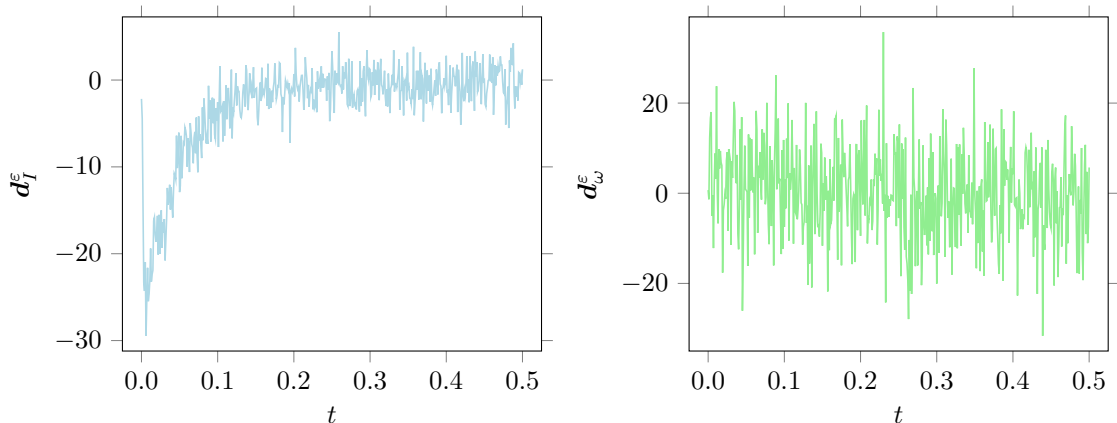


Figure 7: These plots show the remaining discrepancy $\mathbf{d}^\epsilon(\hat{R}) = [\mathbf{d}_I^\epsilon, \mathbf{d}_\omega^\epsilon]^T(\hat{R}) = \mathbf{y} - \mathcal{G}_\phi(\hat{R})$, where \hat{R} is the marginal posterior mean obtained with BM1.

The noisy discrepancy $\mathbf{d}_\omega^\epsilon(\hat{R})$ for ω is basically the measurement noise, since $\mathbb{E}[\mathbf{d}_\omega^\epsilon] \approx 0$ and $V[\mathbf{d}_\omega^\epsilon]^{1/2} = 10.18 \approx \sigma_{\omega,0} \approx \hat{\sigma}_\omega$. But for \mathbf{d}_I^ϵ there is some discrepancy, at least in the first half of the time interval,

different to measurement noise. Calculating the empirical standard deviation for a fixed zero mean $\sigma_I^\dagger := \frac{1}{M-1} \sum_{i=1}^M (\mathbf{d}_I^\varepsilon(\hat{R})_i)^2$, where $\mathbf{d}_I^\varepsilon(\hat{R})_i$ denotes the i -th component of $\mathbf{d}_I^\varepsilon(\hat{R})$, i.e. at time point t_i , one obtains $\sigma_I^\dagger = 6.23 \approx \hat{\sigma}_I$. The obtained value corresponds to the overestimated marginal posterior distribution of σ_I that centers around a similar value, see Figure 6 and Table 1. Obviously, by only considering i.i.d. zero mean measurement noise, BM1 leads to biased and overconfident parameter estimates. Furthermore, the calibrated simulation model does not represent the data accurately enough. However, for BM1 and the model \mathcal{G}_ϕ with the artificial model error $\phi = 0.9$, this is the best result one can get.

4.2 Results Bayesian model 2

This section presents the numerical results obtained with Bayesian model 2 (BM2) for an artificial model error $\phi = 0.9$. Now, the additional model discrepancy term δ is expected to cover, depending on its definition, at least a fraction of the underlying model discrepancy and consequently yields an improved parameter estimation. As introduced in Section 3 the model discrepancy term δ is represented by a polynomial expansion $\delta^K(\mathbf{a})$ with maximum polynomial degree K and unknown coefficients $\mathbf{a} = [\mathbf{a}^{(I)}, \mathbf{a}^{(\omega)}] = [a_0^{(I)}, \dots, a_K^{(I)}, a_0^{(\omega)}, \dots, a_K^{(\omega)}] \in \mathbb{R}^{2K+2}$. Figure 8 displays moments of the marginal posterior distributions, i.e. mean, standard deviation and a 95% confidence interval of R, σ_I and σ_ω , obtained via BM2 for varying maximum polynomial degree $K = 0, \dots, 20$. For each K the last 40000 samples of three independently sampled Markov chains of length 100000 are used to approximate the moments. In contrast to BM1 and due to the increasing number of unknowns in BM2, a larger number of MH-MCMC samples is required in order to obtain convergence. For better comparison and interpretation of the results, the index *BM1* in the figure denotes the previous results without δ , obtained via Bayesian model 1. Following

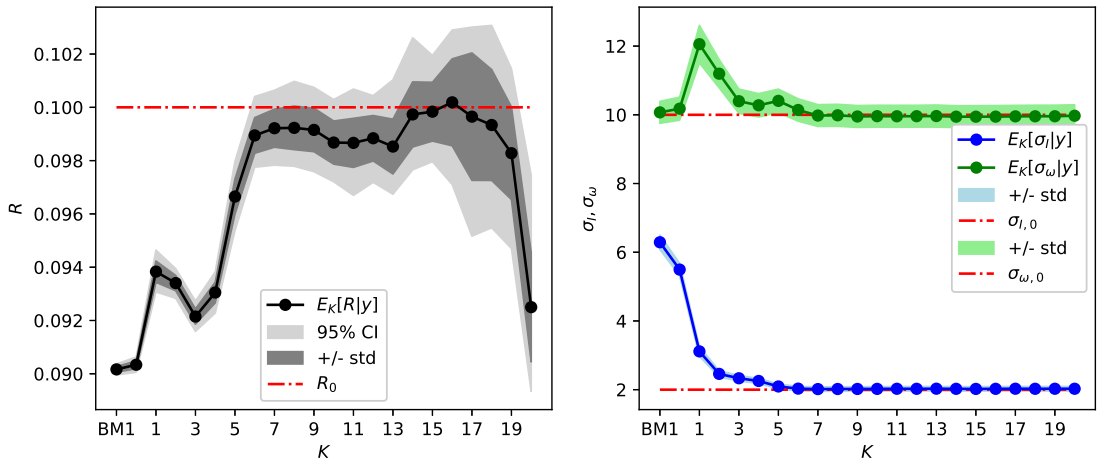


Figure 8: This figures show results obtained with BM2 for varying polynomial degree $K = 0, \dots, 20$. Moments (mean, mean \pm standard deviation and 95% confidence interval) of the marginal posterior distributions of R are displayed on the left hand side and moments (mean and mean \pm standard deviation) of σ_I, σ_ω on the right hand side. Reference values are plotted as dashed red lines. The index *BM1* denotes the previous results without δ , obtained via Bayesian model 1.

the guideline specified in Section 3 the marginal posterior distributions of the noise standard deviations of σ_I and σ_ω in Figure 8 are considered to pick an appropriate K . Both marginal posterior distributions $\pi^K(\sigma_I|\mathbf{y})$ and $\pi^K(\sigma_\omega|\mathbf{y})$ stabilize for $K \geq 7$ and are almost identical for $K \geq 9$. I.e. mean and standard deviation of the marginal posterior distribution are almost identical for $K \geq 9$. This indicates that $K = 9$ is a sufficient polynomial degree and increasing K further does not improve the results with the current specification of the model discrepancy term. Detailed results for $K = 9$ are presented later on.

Now, assessing the marginal posterior distribution of R for varying K in Figure 8 one observes that for $K = 6, \dots, 19$ the marginal posterior mean of R roughly stabilizes in some kind of a plateau close

to the reference R_0 , but the posterior standard deviation increases with K . For $K = 0, \dots, 5$ the model discrepancy term $\delta^K(\mathbf{a})$ is not flexible enough to approximate the underlying discrepancy appropriate enough. As a consequence, the standard deviations of the measurement noise are overestimated and the estimation of the parameter R is still biased for those values of K . But, compared to the results from BM1, already $K = 1$ improves the inference.

In order to compare and evaluate the solutions we define error measures for the parameters and the model discrepancy term.

Definition 4.2. Let x with posterior probability density $\pi(x|\mathbf{y})$ be an estimation for an unknown $x_0 \in \mathbb{R}$, for given data \mathbf{y} . Denote x_0 as reference parameter. The mean square error (MSE) of x is given as

$$MSE(x) := \mathbb{E}[(x - x_0)^2|\mathbf{y}] = Bias(x)^2 + V[x|\mathbf{y}], \quad (25)$$

with bias $Bias(x) = (\mathbb{E}[x|\mathbf{y}] - x_0)$. The mean is with respect to $\pi(x|\mathbf{y})$. In case of the posterior density $\pi^K(x|\mathbf{y})$ depending on K write $MSE_K(x) = \mathbb{E}_K[(x - x_0)^2|\mathbf{y}] = Bias_K(x)^2 + V_K[x|\mathbf{y}]$.

Definition 4.3. In order to have something similar for the model discrepancy estimations $\delta^K(\mathbf{a}) = [\delta_f^K(\mathbf{a}), \delta_\omega^K(\mathbf{a})] \in \mathbb{R}^{2M}$ with the posterior density function $\pi^K(\mathbf{a}|\mathbf{y})$ and with respect of the reference discrepancy \mathbf{d}_0 we construct for $\delta_* \in \{\delta_f^K(\mathbf{a}), \delta_\omega^K(\mathbf{a})\}$ the normed MSE

$$nMSE_K(\delta_*) := \|MSE_K(\delta_*)\| = \|Bias_K(\delta_*)^2 + V_K(\delta_*)\|, \quad (26)$$

where $\|\cdot\|$ is the $L^2(\mathbb{R}^{2M})$ norm. With triangle and Cauchy Schwarz' inequality this can be bounded by

$$nMSE_K(\delta_*) \leq nBias_K(\delta_*)^2 + nV_K(\delta_*) =: \overline{nMSE}_K(\delta_*). \quad (27)$$

Here the normed bias $nBias_K(\delta_*) := \|\mathbb{E}_K[\delta_*|\mathbf{y}] - \mathbf{d}_0\|$ and variance $nV_K(\delta_*) := \|V_K[\delta_*|\mathbf{y}]\|$. The mean is with respect to $\pi^K(\mathbf{a}|\mathbf{y})$.

Figure 9 displays $Bias_K(R)$, variance $V_K[R|\mathbf{y}]$ and mean square error $MSE_K(R)$ of R with respect to R_0 , depending on the model complexity, i.e. maximum polynomial degree K of the discrepancy term $\delta^K(\mathbf{a})$. The MSE of R is minimal around $K = 7, 8, 9$ and $K = 15$, indicating those values as an optimal

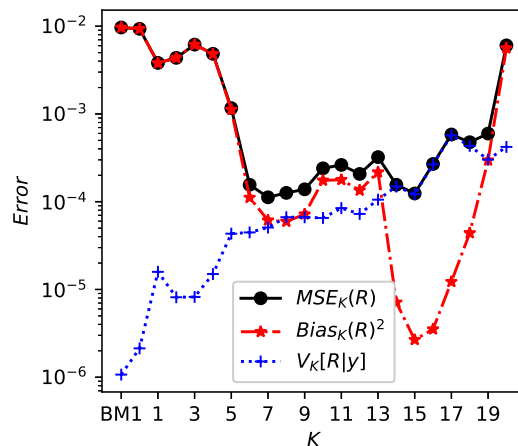


Figure 9: Bias-variance tradeoff: Bias $Bias_K(R)$, variance $V_K[R|\mathbf{y}]$ and mean square error $MSE_K(R)$ of R with respect to R_0 , depending on K , for $K = 0, \dots, 20$ from BM2. The index BM1 denotes the previous results without δ , obtained via Bayesian model 1.

model complexity and backing up the decision of $K = 9$ as sufficient polynomial degree. Figure 9 largely corresponds to the bias-variance tradeoff, since the variance increases with K and the bias decreases with K , at least until $K = 15$. The variance increases due to the increased number of parameters, i.e. more degrees of freedom. This results in more uncertainty, i.e. variance, for each parameter by using the same amount of data. Furthermore, for large K overfitting starts, i.e. $\delta^K(\mathbf{a})$ starts to oscillate and reproduce parts of the measurement noise. The bias decreases with increasing K , but only until $K = 15$, then it increases as another phenomena, namely non-identifiability, occurs additionally.

Figure 10 displays bias, variance and MSE of σ_I and σ_ω with respect to the reference values $\sigma_{I,0}, \sigma_{\omega,0}$,

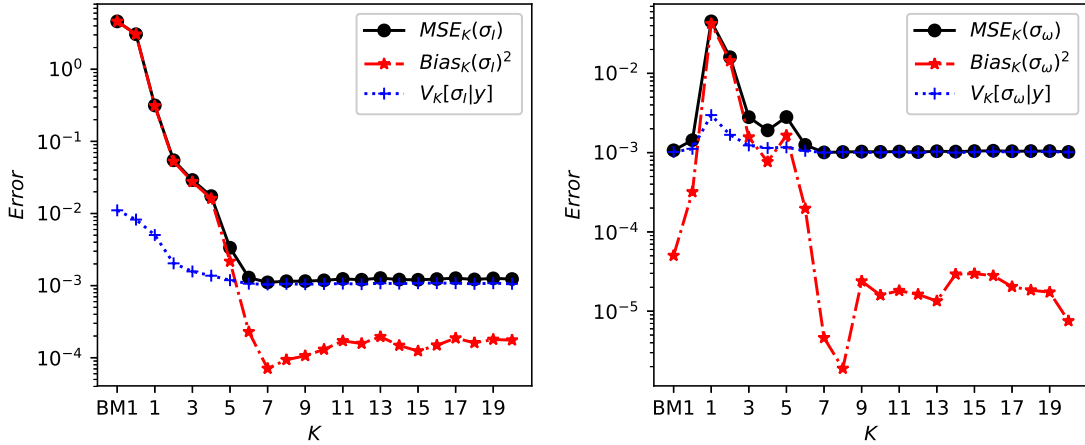


Figure 10: Bias, variance and MSE of σ_I and σ_ω with respect to the reference values $\sigma_{I,0}, \sigma_{\omega,0}$, depending on K , for $K = 0, \dots, 20$ from BM2. The index BM1 denotes the previous results without δ , obtained via Bayesian model 1.

for $K = 0, \dots, 20$ from BM2. For $K \geq 7$ the MSE for both stays on the same level mainly as the variance stays on the same level. As for each K the same data \mathbf{y} is used there is a point where the estimation of $\pi^K(\sigma_I|\mathbf{y})$ and $\pi^K(\sigma_\omega|\mathbf{y})$ can not be improved further, without considering new additional data.

Figure 11 displays the normed $nBias_K(\delta_*)$, $nV_K(\delta_*)$, $nMSE_K(\delta_*)$ and $\overline{nMSE}_K(\delta_*)$ for the model discrepancy approximations $\delta_* \in \{\delta_I^K(\mathbf{a}), \delta_\omega^K(\mathbf{a})\}$ depending on K , for $K = 0, \dots, 20$ from BM2.

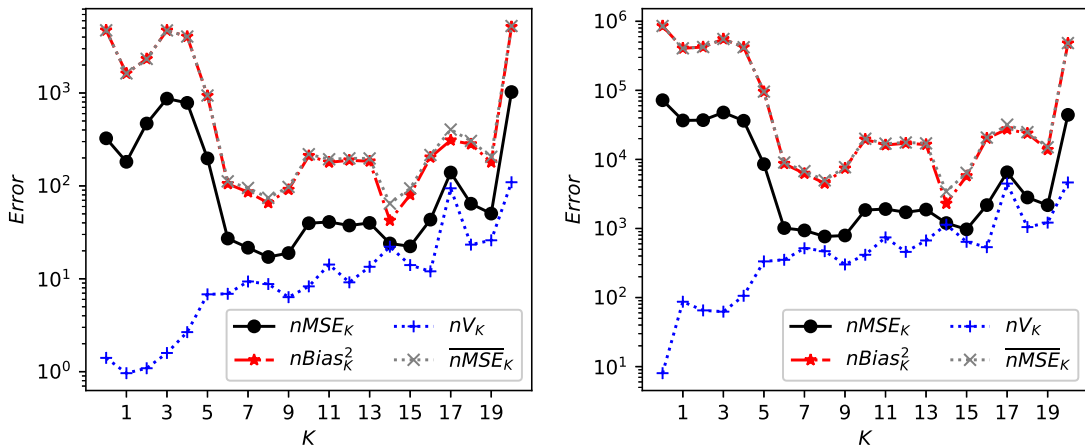


Figure 11: Bias-variance tradeoff: $nBias_K(\delta_*)$, $nV_K(\delta_*)$, $nMSE_K(\delta_*)$ and $\overline{nMSE}_K(\delta_*)$ for the model discrepancy approximations $\delta_* \in \{\delta_I^K(\mathbf{a}), \delta_\omega^K(\mathbf{a})\}$ depending on K , for $K = 0, \dots, 20$ from BM2.

crepancy approximations $\delta_* \in \{\delta_I^K(\mathbf{a}), \delta_\omega^K(\mathbf{a})\}$ depending on K , for $K = 0, \dots, 20$ from BM2. $nBias_K(\delta_*)$ is minimal for $K = 8$. Here again, as in Figure 9, the effects of Bias-variance tradeoff and non-identifiability can be observed.

Especially for $K = 20$ overfitting and non-identifiability can be observed. For $K = 20$ the marginal posterior samples of R in Figure 12 do not converge any more, even not for an increased sample size of three independent chains with 500.000 samples each. Consequently, the posterior distribution is spread out and biased. Also the posterior distribution of $\delta^K(\mathbf{a})$ for polynomial degree $K = 20$, displayed in

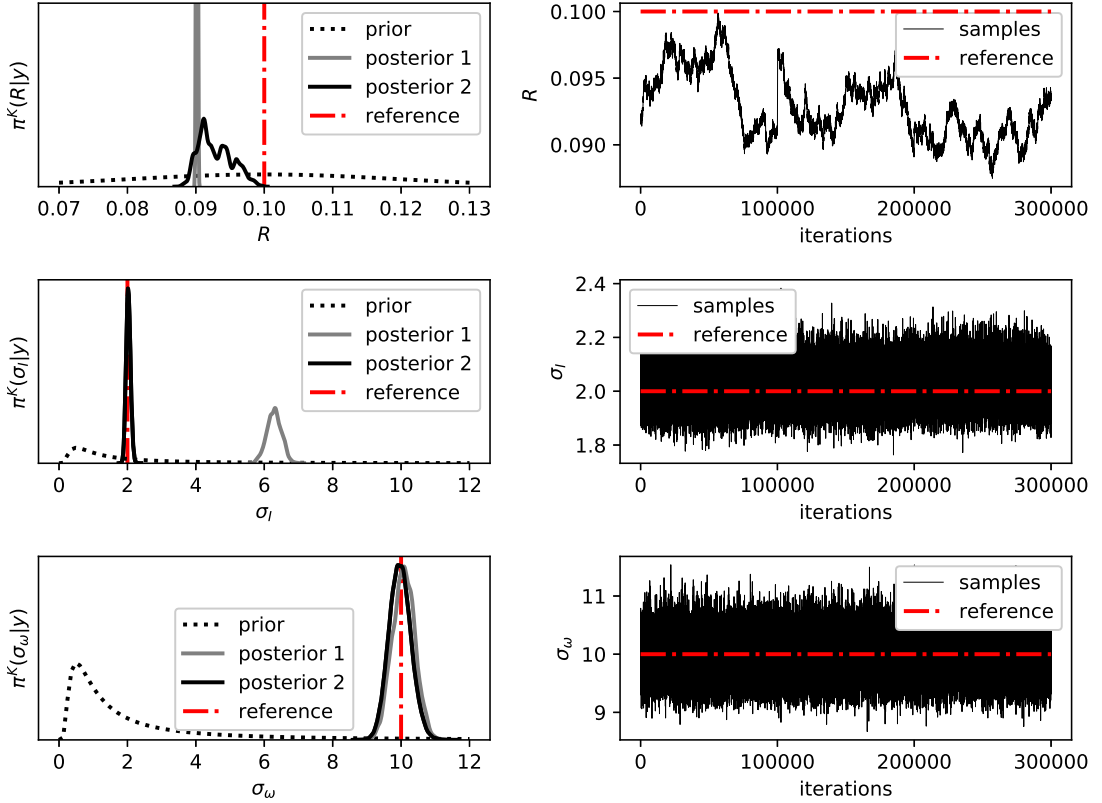


Figure 12: Posterior distributions and samples from BM1 and BM2 with $K = 20$: The figures on the right hand side show marginal MH-MCMC samples of the parameters R, σ_I and σ_ω (from top to bottom). The figures on the left hand side show the kernel density estimates of the corresponding samples as an approximation of the marginal posterior distributions. The results without model discrepancy term from BM1 are plotted in grey and the results with model discrepancy term from BM2 and polynomial degree $K = 20$ are black. The prior distributions are plotted as dotted black lines and the reference δ values as dashed red lines.

Figure 13, is spread out and the reference discrepancy is not captured anymore. Only the bounds of the posterior distribution of $\delta^K(\mathbf{a})$ tend towards the reference. This posterior distribution corresponds more to the discrepancy $\mathbf{d}(\hat{R})_I$ that was still present after inference with BM1, see Figure 7. Furthermore, overfitting can be observed as the model discrepancy term tries to capture the high oscillations of the measurement noise. Looking at the posterior distributions of the polynomial coefficients in Figure 14 one observes that for I the coefficients a_0^I and a_1^I of the constant and linear polynomials are somehow specific and have small support, but for a_2^ω, a_3^ω and a_4^ω it is spread out and they are non-identifiable. All other concentrate around zero. For ω one observes flat posterior distributions for a_0^ω and a_1^ω , i.e. the coefficients of the constant and linear polynomials are non-identifiable. All other posterior distributions of the coefficients a_j^ω for $j = 2, \dots, 20$ concentrate around zero with decreasing support.

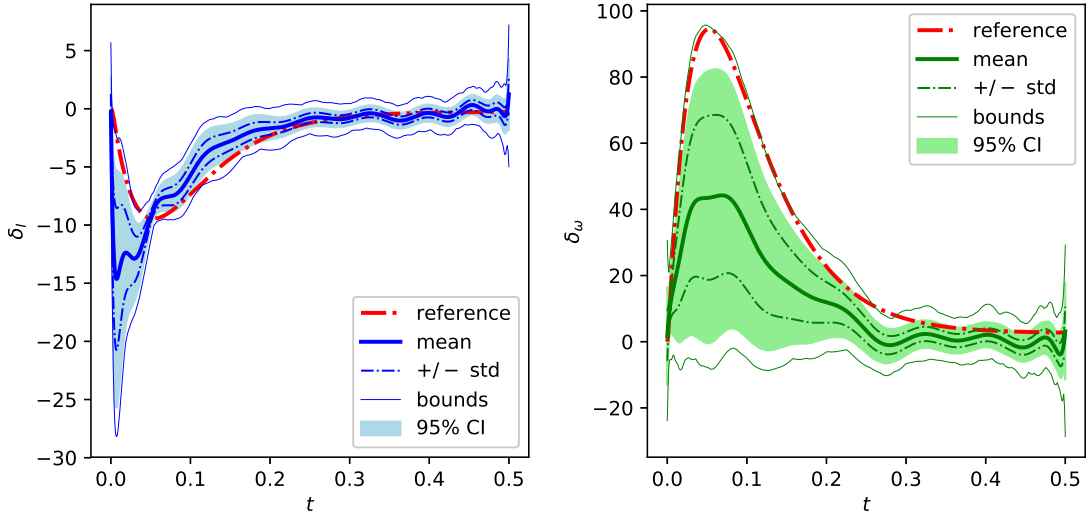


Figure 13: Posterior distribution of model discrepancy $\delta^K(\mathbf{a}) = [\delta_I^K, \delta_\omega^K](\mathbf{a})$ for $K = 20$.

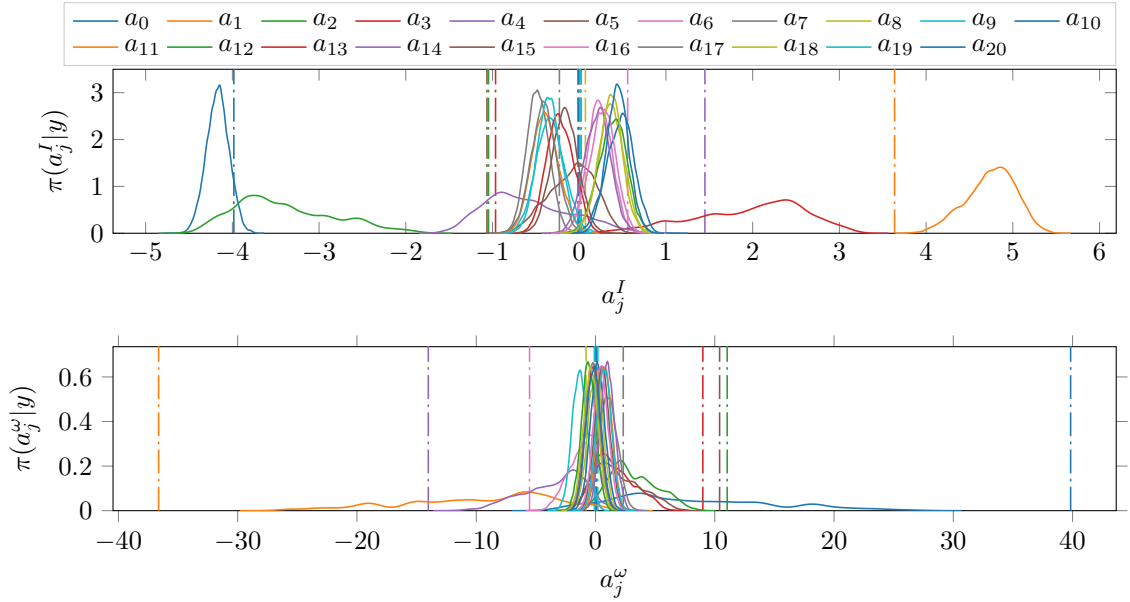


Figure 14: Marginal posterior distributions of polynomial coefficients \mathbf{a} of $\delta^K(\mathbf{a})$ for $K = 20$. The upper plot displays the marginal posterior distributions of $a_j^{(I)}$ and the lower of $a_j^{(\omega)}$, $j = 0, \dots, K$. The vertical lines are the coefficients \mathbf{a} obtained from least squares fitting of $\delta^K(\mathbf{a})$ to the reference discrepancy \mathbf{d}_0 .

However, the posterior distributions of σ_I and σ_ω (see Figure 12) concentrate at the reference values. This indicates that albeit the posterior distribution of R and $\delta^K(\mathbf{a})$ are spread out, the sum $\mathcal{G}_\phi(R) + \delta(\mathbf{a}) + \varepsilon(\sigma_{I,\omega})$ still represents the data \mathbf{y} well.

Finally, considering all these observations in Figure 8, and omitting the reference for a while, it makes sense to consider a polynomial degree $K = 9$. $K = 7$ would be already sufficient, but in order to allow a bit more flexibility for $\delta^K(\mathbf{a})$ $K = 9$ is considered as a sufficient polynomial degree. Also by considering the reference $K = 9$ is confirmed to be a good choice. However, $K = 15$ might yield a better marginal posterior mean and has a comparable MSE, but the variance is also increased and the posterior distribution of $\delta^{15}(\mathbf{a})$ already shows an oscillating behavior. Furthermore, in order to keep computational cost low a small K is preferred.

Now, for BM2 and $K = 9$ the number of parameters to sample from is $3 + 2(K + 1) = 23$. Figure 15 displays the last 40.000 samples of three independently sampled Markov chains of total length 100.000. The marginal posterior distribution of R concentrates around the reference value and the relative error of

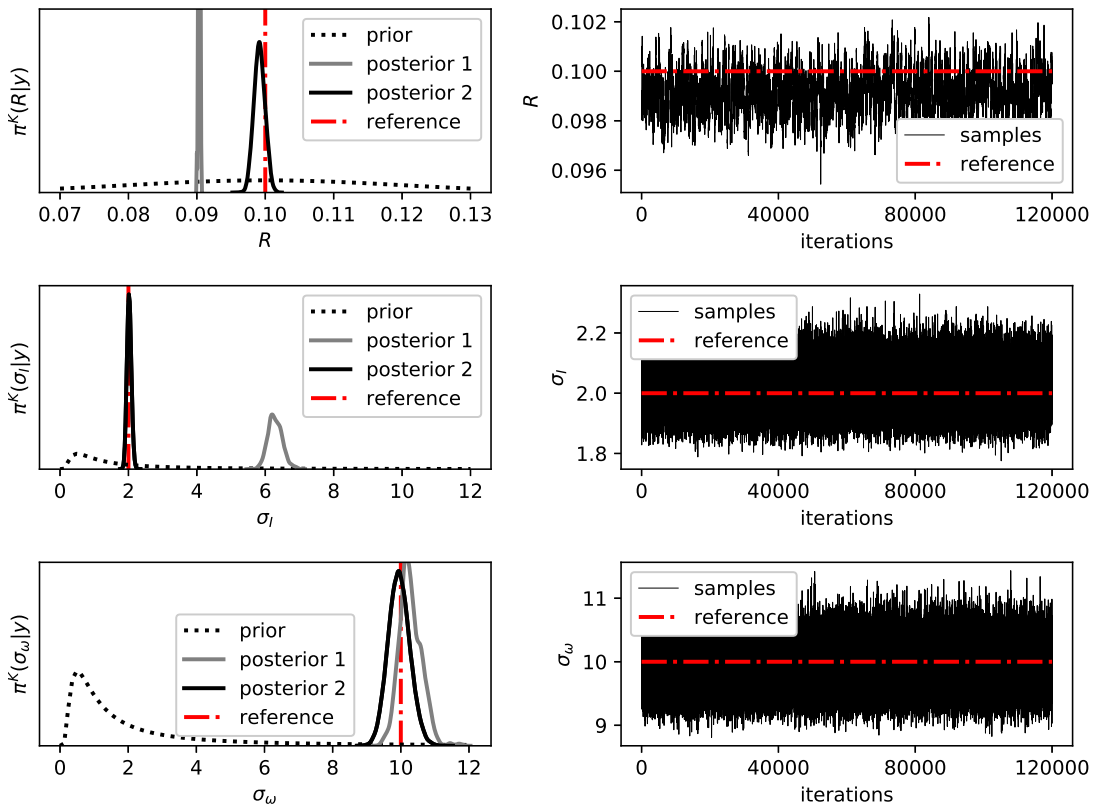


Figure 15: Posterior distributions and samples from BM1 and BM2 with $K = 9$: The figures on the right hand side show marginal MH-MCMC samples of the parameters R, σ_I and σ_ω (from top to bottom). The figures on the left hand side show the kernel density estimates of the corresponding samples as an approximation of the marginal posterior distributions. The results without model discrepancy term from BM1 are plotted in grey and the results with model discrepancy term from BM2 and polynomial degree $K = 9$ are black. The prior distributions are plotted as dotted black lines and the reference values as dashed red lines.

the posterior mean reduces around one order of magnitude compared to the result of BM1, see Table 1. Also, both noise standard deviations concentrate around the reference and significantly improve in relative errors compared to BM1. Figure 16 displays the concentration of the posterior distribution of $\delta^K(\mathbf{a}) = [\delta_I^K, \delta_\omega^K]^T(\mathbf{a})$ around the reference discrepancy \mathbf{d}_0 . The relative error $nBias_K(\delta_*^K - d_{0,*})/\|d_{0,*}\|$ with

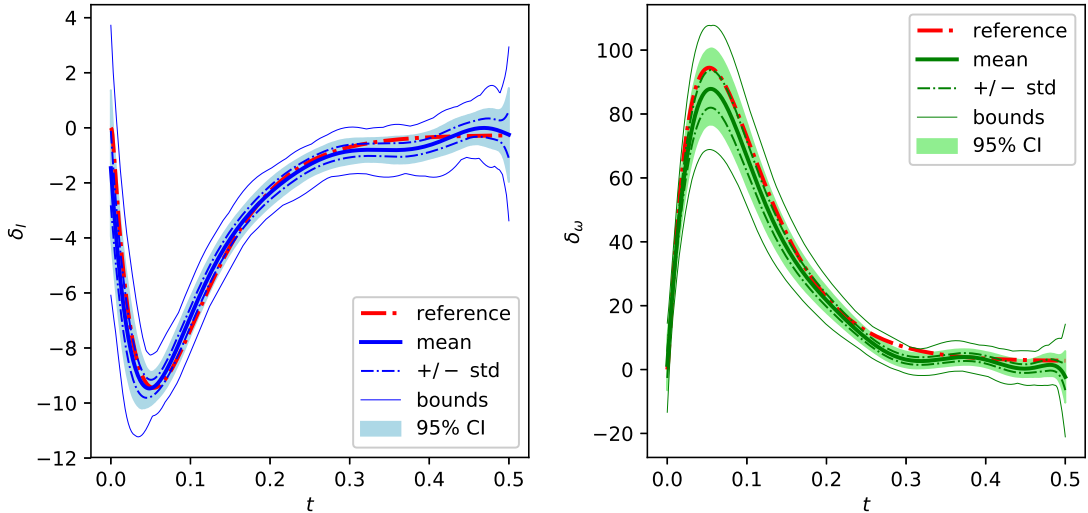


Figure 16: Posterior distribution of model discrepancy $\delta^K(\mathbf{a}) = [\delta_I^K, \delta_\omega^K](\mathbf{a})$ for $K = 9$.

respect to the L^2 -norm, is $1.0\text{e-}01$ for I and $9.2\text{e-}02$ for ω . Also the marginal posterior distributions of the polynomial coefficients \mathbf{a} in Figure 17 correspond to the reference coefficients. Overall, the posterior distribution of $\mathcal{G}_\phi(R) + \delta^K$ yields a good approximation of the measurements \mathbf{y} with only a small variance band.

5 Conclusion

In this work a method to infer model parameters and model discrepancy is considered and applied to synthetic measurement data of an electric motor. The suggested Bayesian model 2 (BM2) considers measurement noise and model discrepancy, in order to separate these two sources of uncertainty and improve physical parameter estimation. The model discrepancy term is modeled as a truncated polynomial expansion $\delta^K(\mathbf{a})$ with unknown coefficients \mathbf{a} and an maximum polynomial degree K . The smoothness assumption, implied by the polynomials, restricts the flexibility of $\delta^K(\mathbf{a})$ depending on K and introduces specific prior knowledge about the true underlying model discrepancy. This work discusses and then defines a guideline on how to define appropriate model discrepancy term complexity, i.e. here the maximum polynomial degree K of $\delta^K(\mathbf{a})$ based on the marginal posterior distribution of measurement noise standard deviation.

The framework applied to the electric motor showed promising perspectives by an improved estimation of the model parameters. Furthermore a good approximation of the a-priori unknown model discrepancy is learned during BM2 for a sufficient maximum polynomial degree $K = 9$. For too small K the flexibility of $\delta^K(\mathbf{a})$ is not sufficient and estimation is just slightly improved. Additionally, the results showed that if K is too large the prior contains to less information about the smoothness of the reference discrepancy and consequently the posterior distribution does not converge anymore. Consequently, in order to identify both, the underlying parameter value and the reference discrepancy of the test scenario, it is important to formulate at least for one of those unknowns a prior containing some information about the reference. Otherwise the problem is not identifiable.

For this example the model discrepancy δ was rather smooth. This might be not the case in other examples. If one knows that δ is non-smooth one might consider other representations for δ , such as radial basis functions, wavelets, Gaussian Processes or even neuronal networks. Here the question arises, if more flexibility implies non-identifiability. However, in our example, even for small K estimation could be improved, and in the case of no model discrepancy results do not degrade, if one uses an model discrepancy term. This might lead to the conclusion that using a model discrepancy term, even if it is not sufficiently modeled, does not degrade the results.

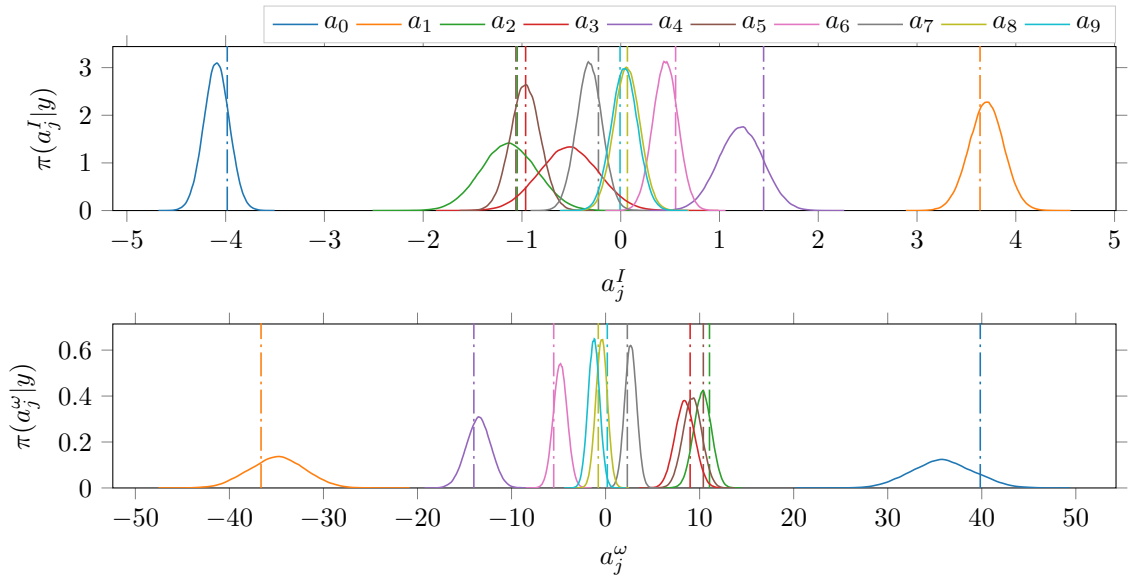


Figure 17: Marginal posterior distributions of polynomial coefficients \mathbf{a} of $\delta^K(\mathbf{a})$ for $K = 9$. The upper plot displays the marginal posterior distributions of $a_j^{(I)}$ and the lower of $a_j^{(\omega)}$, $j = 0, \dots, K$. The vertical lines are the coefficients \mathbf{a} obtained from least squares fitting of $\delta^K(\mathbf{a})$ to the reference discrepancy \mathbf{d}_0 .

This work showed that with an additional term for model discrepancy inference results can be improved and realistic information about the discrepancy can be collected, if some information about the parameters or the discrepancy is specified in the priors. However, computational costs increase drastically with the number of parameters for the model discrepancy term. As previously suggested, one might consider surrogate models to leverage this or improve sampling.

According to the definition of δ the resulting prior has a time dependent variance with high values at the boundaries. The posterior distribution of $\delta^K(\mathbf{a})$ shows a similar behavior at the boundary T . This might be due to the prior definition. In order to get rid of the wide posterior distribution at T it might be necessary to define a prior for $\delta^K(\mathbf{a})$ with (at least almost) constant variance over time. The solution proposed for the next problem, might also help in this case.

With the polynomial approach it is difficult to define the right polynomial degree K a-priori. If the maximal polynomial degree is too low, then it is not possible for $\delta^K(\mathbf{a})$ to approximate the reference discrepancy \mathbf{d}_0 appropriately. If the polynomial degree is too high, $\delta^K(\mathbf{a})$ tends to overfit and tries to capture, in addition to \mathbf{d}_0 , the highly oscillatory behavior of the measurement noise as well. The coefficients belonging to higher degree basis polynomials do not tend to zero as they actually should. A solution to this problem could be to model this in the definition of the polynomial coefficients prior distributions, i.e. by reducing the variance with increasing polynomial degree, e.g. $V[a_i] \sim 1/i, i = 1, \dots, N$.

Next steps are to apply this approach to higher dimensional problems and further to real field data to test its capabilities. Then, as real data implies a more complex simulator, surrogate modeling will be mandatory to leverage computational expenses. Additionally, evidence approximation could be another criteria to select a Bayesian model, which will be considered in future work.

References

- [1] Paul D. Arendt, Daniel W. Apley, and Wei Chen. Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability. *Journal of Mechanical Design*, 134(10):100908, 2012.
- [2] Paul D. Arendt, Daniel W. Apley, Wei Chen, David Lamb, and David Gorsich. Improving Identifiability in Model Calibration Using Multiple Responses. *Journal of Mechanical Design*, 134(10):100909, 2012.
- [3] Jenný Brynjarsdóttir and Anthony O’Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.
- [4] Masoumeh Dashti and Andrew M. Stuart. The Bayesian Approach to Inverse Problems. In Roger Ghanem, David Higdon, and Houman Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 311–428. Springer International Publishing, Cham, 2017.
- [5] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin, and Hal Steven Stern. *Bayesian Data Analysis // Bayesian data analysis*. Texts in Statistical Science Series. CHAPMAN & HALL/CRC and CRC Press, Boca Raton, 3 // third edition edition, 2013.
- [6] Philipp Glaser, Michael Schick, Kosmas Petridis, and Vincent Heuveline. Comparison between a polynomial chaos surrogate model and markov chain monte carlo for inverse uncertainty quantification based on an electric drive test bench. *ECCOMAS Congress 2016*, 2016.
- [7] Trevor Hastie, Robert Tibshirani, and J. H. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. Springer series in statistics. Springer New York and Springer, New York, NY, 2nd ed. edition, 2009.
- [8] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97, 1970.
- [9] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [10] Jari Kaipio and Erkki Somersalo. *Statistical and Computational Inverse Problems*. Applied Mathematical Sciences. Springer Science+Business Media, Inc, New York, NY, 2005.
- [11] Marc C. Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [12] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, 2012.
- [13] Joseph B. Nagel, Jörg Rieckermann, and Bruno Sudret. Uncertainty quantification in urban drainage simulation: Fast surrogates for sensitivity analysis and model calibration. 2017.
- [14] Rui Paulo, Gonzalo García-Donato, and Jesús Palomo. Calibration of computer models with multivariate output. *Computational Statistics & Data Analysis*, 56(12):3959–3974, 2012.
- [15] Matthew Plumlee. Bayesian Calibration of Inexact Computer Models. *Journal of the American Statistical Association*, 112(519):1274–1285, 2016.
- [16] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian process for machine learning*. MIT press, 2006.
- [17] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2(2):e55, 2016.
- [18] Claudia Schillings and Christoph Schwab. Scaling limits in computational bayesian inversion. *ESAIM: Mathematical Modelling and Numerical Analysis*, 50(6):1825–1856, 2016.
- [19] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. SIAM Computational Science & Engineering Series, Philadelphia, PA, USA, 2013.
- [20] Björn Sprungk. *Numerical Methods for Bayesian Inference in Hilbert Spaces*. Universitätsverlag der TU Chemnitz, Chemnitz, 1. auflage edition, 2018.
- [21] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

- [22] T. J. Sullivan. *Introduction to uncertainty quantification*, volume volume 63 of *Texts in applied mathematics*, 0939-2475. Springer, 2015.
- [23] Hamid A. Toliyat, editor. *Handbook of electric motors*, volume 120 of *Electrical and computer engineering*. Dekker, New York and Basel, 2. ed., rev. and exp edition, 2004.
- [24] Rui Tuo and C. F. Jeff Wu. A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795, 2016.
- [25] Rui Tuo and C. F. Jeff Wu. Efficient calibration for imperfect computer models. *Ann. Statist.*, 43(6):2331–2352, 2015.
- [26] G. Wanner and E. Hairer. *Solving ordinary differential equations I*, volume 1. Springer-Verlag, Berlin, 1991.
- [27] Dongbin Xiu and George Em Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

Preprint Series of the Engineering Mathematics and Computing Lab

recent issues

- No. 2017-07 Martin Baumann, Fabian Gebhart, Oliver Mattes, Sotirios Nikas, Vincent Heuveline: Development and implementation of a temperature monitoring system for HPC systems
- No. 2017-06 Simon Gawlok, Philipp Gerstner, Saskia Haupt, Vincent Heuveline, Jonas Kratzke, Philipp Lösel, Katrin Mang, Mareike Schmidtobreck, Nicolai Schoch, Nils Schween, Jonathan Schwegler, Chen Song, Martin Wlotzka: HiFlow3 – Technical Report on Release 2.0
- No. 2017-05 Nicolai Schoch, Vincent Heuveline: Towards an Intelligent Framework for Personalized Simulation-enhanced Surgery Assistance: Linking a Simulation Ontology to a Reinforcement Learning Algorithm for Calibration of Numerical Simulations
- No. 2017-04 Martin Wlotzka, Thierry Morel, Andrea Piacentini, Vincent Heuveline: New features for advanced dynamic parallel communication routines in OpenPALM: Algorithms and documentation
- No. 2017-03 Martin Wlotzka, Vincent Heuveline: An energy-efficient parallel multigrid method for multi-core CPU platforms and HPC clusters
- No. 2017-02 Thomas Loderer, Vincent Heuveline: New sparsing approach for real-time simulations of stiff models on electronic control units
- No. 2017-01 Chen Song, Markus Stoll, Kristina Giske, Rolf Bendl, Vincent Heuveline: Sparse Grids for quantifying motion uncertainties in biomechanical models of radiotherapy patients
- No. 2016-02 Jonas Kratzke, Vincent Heuveline: An analytically solvable benchmark problem for fluid-structure interaction with uncertain parameters
- No. 2016-01 Philipp Gerstner, Michael Schick, Vincent Heuveline, Nico Meyer-Hübner, Michael Suriyah, Thomas Leibfried, Viktor Slednev, Wolf Fichtner, Valentin Bertsch: A Domain Decomposition Approach for Solving Dynamic Optimal Power Flow Problems in Parallel with Application to the German Transmission Grid
- No. 2015-04 Philipp Gerstner, Vincent Heuveline, Michael Schick : A Multilevel Domain Decomposition approach for solving time constrained Optimal Power Flow problems
- No. 2015-03 Martin Wlotzka, Vincent Heuveline: Block-asynchronous and Jacobi smoothers for a multigrid solver on GPU-accelerated HPC clusters
- No. 2015-02 Nicolai Schoch, Fabian Kißler, Markus Stoll, Sandy Engelhardt, Raffaele de Simone, Ivo Wolf, Rolf Bendl, Vincent Heuveline: Comprehensive Pre- & Post-Processing for Numerical Simulations in Cardiac Surgery Assistance
- No. 2015-01 Teresa Beck, Martin Baumann, Leonhard Scheck, Vincent Heuveline, Sarah Jones: Comparison of mesh-adaptation criteria for an idealized tropical cyclone problem
- No. 2014-02 Christoph Paulus, Stefan Suwelack, Nicolai Schoch, Stefanie Speidel, Rüdiger Dillmann, Vincent Heuveline: Simulation of Complex Cuts in Soft Tissue with the Extended Finite Element Method (X-FEM)
- No. 2014-01 Martin Wlotzka, Vincent Heuveline: A parallel solution scheme for multiphysics evolution problems using OpenPALM

Preprint Series of the Engineering Mathematics and Computing Lab (EMCL)

