Bojana Rosić, Jan Sýkora, Oliver Pajonk, Anna Kučerová and Hermann G. Matthies

Comparison of Numerical Approaches to Bayesian Updating
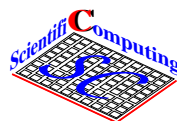
This document was created October 2014 using LaTeX 2$_\varepsilon$.

Institute of Scientific Computing
Technische Universität Braunschweig
Hans-Sommer-Straße 65
D-38106 Braunschweig, Germany

e-mail: wire@tu-bs.de
url: www.wire.tu-bs.de

Comparison of Numerical Approaches to Bayesian Updating

Bojana Rosić, Jan Sýkora, Oliver Pajonk, Anna Kučerová
and Hermann G. Matthies

Institute of Scientific Computing, TU Braunschweig
wire@tu-bs.de

This paper investigates the Bayesian process of identifying unknown model parameters given prior information and a set of noisy measurement data. There are two approaches being adopted in this research: one that uses the classical formula for measures and probability densities and one that leaves the underlying measure unchanged and updates the relevant random variable. The former is numerically tackled by a Markov chain Monte Carlo procedure based on the Metropolis-Hastings algorithm, whereas the latter is implemented via the ensemble/square root ensemble Kalman filters, as well as the functional approximation approaches in the form of the polynomial chaos based linear Bayesian filter and its corresponding square root algorithm.

The study attempts to show the principal differences between full and linear Bayesian updates when a direct or a transformed version of measurements are taken into consideration. In this regard the comparison of both strategies is provided on the example of a steady state diffusion equation with nonlinear and transformed linear measurement operators.

Keywords: polynomial chaos based linear Bayesian update, Markov chain Monte Carlo, square root algorithm, ensemble Kalman filter

3

Contents

1 Introduction

In many applications one would like to predict the behaviour of a physical system with the help of a mathematical model whose parameters are not directly observable. In order to obtain the desired information about the system state or to allow for a prediction of a system response beyond observed localities, the model parameters are estimated from noisy data gathered by measuring some of the observable system responses. Most of the existing studies have tackled the issue by tuning the model parameters such that the distance between the observed and predicted system responses is minimised in a certain norm (e.g. [31, 9]). However, the resulting optimisation problem is often ill-posed – the minimised function is multimodal, non-smooth or non-differentiable — and hence a regularisation procedure [6] or a soft-computing-based method [13] is required. Yet, such a fitting-based approach provides only a one point estimate and hence omits the related uncertainties in measurements, imperfections of the numerical model as well as the preliminary knowledge about the material parameters arising from their physical occurrence. Unlike point-estimation techniques, the probabilistic concept transforms the prior expert-based probability description to a posterior via the incorporation of observations. From a Bayesian point of view this further means that the unknown parameters are taken to be uncertain and are modelled with the help of random variables (RVs)/fields (RFs), whose probability descriptions are coming from expert knowledge and the maximum entropy law [27]. This prior knowledge is then updated to a posterior distribution via Bayes's rule given in terms of conditional probabilities. In this regard, the process of assimilating more information obtained via experiments becomes well-posed. As a final outcome, the posterior distribution summarizes all available information about the model parameters such as the mean value, variance, probability of occurrence etc.

The primary computational challenge in Bayesian inference consist in extracting information from the posterior by evolving the probability measure. The Markov chain Monte Carlo (MCMC) method [10] is one of the most commonly used techniques for this kind of parameter estimation. In MCMC methods, the Markov chain is constructed such that the asymptotic distribution of the chain is the Bayesian posterior distribution. The posterior is sampled by letting the Markov chain run for a sufficiently long time. With the intention of accelerating the MCMC method some authors (e.g. [15, 16, 14, 29]) have introduced stochastic spectral methods into the computation. Expanding the prior random process into a polynomial chaos (PCE) or a Karhunen-Loève expansion (KLE), the inverse problem becomes an inference on the weights of the KLE or PCE coefficients. Another solution is to combine polynomial chaos theory with the maximum likelihood estimation and to calculate the

1

parameter estimates in a recursive manner (see [20]), or to apply a local linearisation of the forward model to improve the acceptance probability of proposed moves [5]. However, the previously mentioned methods are all based on pure sampling procedures, or a combination of spectral approximations and MCMC. Therefore, they are slowly convergent and often computationally infeasible especially when one deals with large-scale problems.

To mitigate this issue, the authors of [21, 19, 22, 18] constructed a more efficient approach based on conditional expectation, which is an equivalent way to formulate the Bayesian update. The conditional expectation can be approximated by linear or higher order maps, which have to be found during the updating. In this way the Bayesian update (BU) is an algebraic formula, which can be computed in a purely analytical way as indicated in [21, 19, 22, 18]. In a simpler version, this idea appeared independently in [2], whereas in [23] it appears as a variant of the Kalman filter [12].

The aim of this study is to investigate the differences between full and linear Bayesian updates on a simple linear diffusion model with one uncertain parameter. For this purpose two scenarios are considered: one that features a direct—classical Bayesian update—and a second one that introduces a transformed measurement operator—a transformed Bayesian update. The transformation is studied from the mathematical and numerical point of view in order to better understand the linear Bayesian update. In addition, the paper proposes the use of the proxy modelling in the assimilation process in order to reduce the required computational time.

The paper is organised in the following way: Section 2 gives a short introduction into the model problem for which the Bayesian inference is presented in Section 3. Section 4 gives a brief overview of the available methods for the computation of a Bayesian update. The methods are analysed and compared in Section 5. Finally, the paper is concluded in Section 6.

2 Model problem

The problem considered in this paper is a steady state heat transfer expressed by an energy balance equation

$$-\operatorname{div}(\kappa \nabla u(x)) = f(x), \quad \forall x \in \mathscr{G} \subset \mathbb{R}^2, \tag{1}$$

2

in which the scalar conductivity coefficient $\kappa \in \mathscr{K}$ together with the loading $f(x)$ and initial conditions u_0 determine the system response u. In computational practice u is very often evaluated assuming that the conductivity parameter κ is known. This process is called the prediction of the system response, or the forward problem. However, the aim of this paper is to not to determine u but κ given the set of observation data z—the measurements of the system response in a few points of the physical domain of consideration.

Usually, the measurement set is mathematically formalised by an observation operator H, which relates the complete model response $u \in \mathscr{U}$ to an observation y in some vector space $\mathscr{Y}$ [21, 22]:

$$H : (\kappa, u(x)) \mapsto y(x) = H(\kappa; u(x)) \in \mathscr{Y}. \tag{2}$$

Since the modelled values y differ from the real data set z, the previous equation transforms to

$$y = z - \varepsilon = H(\kappa; u(x)), \tag{3}$$

in which the variable ε subsumes both model imperfections and the measurement error. A key aspect of this is that one may try to compute the non-observable thermal conductivity κ given z from Eq. (3). This further agrees with the inversion of the operator H, which in general may not be invertible or has a non-continuous inverse— the ill-posed problem. To ensure the existence and uniqueness of the solution, the previous issue can be resolved by, for example, taking the additional information into consideration. In a Bayesian point of view this corresponds to the prescription of a prior distribution on the model parameter κ.

3 Identification via Bayesian regularisation

The Bayesian inference treats the problem in Eq. (3) by acquiring additional knowledge on the parameter set next to the observation data. Such a description has two ingredients, the measurable function or random variable, and the probability measure. One group of methods updates the measure—the classical Bayesian updating [10], the other group changes the function—the linear Bayesian updating [22]. In this section we show connection between these methods as well as give their short description.

The prior information on κ comes from expert knowledge about its realistic values and can be modelled in a form of a prior probability density function $p(\kappa)$ with the

help of the maximum entropy approach [27]. In this manner, κ in Eq. (3) can be described by the finite-variance $\mathcal{K}$ -valued RV/ RF

$$\kappa(\omega) : \Omega \to \mathcal{K} \tag{4}$$

on a probability space $S := L_2(\Omega, \mathfrak{A}, \mathbb{P})$. Here, Ω denotes the space of elementary events ω, $\mathfrak{A}$ is the σ-algebra and $\mathbb{P}$ stands for the probability measure.

Once the prior is chosen, the posterior density can be obtained with the help of classical Bayes's rule:

$$\pi_a(\kappa) := p(\kappa|z) = \frac{p(z|\kappa)}{p(z)} p(\kappa) \propto L(\kappa) p(\kappa) \tag{5}$$

given in terms of the conditional probability density functions. Here, $\pi_a(\kappa)$ and $p(\kappa)$ stand for the posterior and prior density functions of κ, and $L(\kappa)$ is the likelihood giving a measure of how good the model is explaining the data z.

The law described in Eq. (5) is not the one used further. Namely, the random variable κ is restricted to the positive cone in the vector space, and hence requires a transformation, see [22]. By defining the bijective differentiable mapping

$$T_q : \quad \mathcal{K} \to \mathcal{Q} \tag{6}$$

from the model $\mathcal{K}$ to the assimilation $\mathcal{Q}$ space, κ is transformed to a random variable

$$q = T_q(\kappa) \tag{7}$$

which lives in a vector space. As a consequence, Bayes's rule in Eq. (5) obtains the form

$$\pi_q(q) = \frac{p(z|q)}{p(z)} p_q(q), \tag{8}$$

where $p_q(q)$ and $\pi_q(q)$ are the prior and posterior density of q, respectively. Once the assimilation is performed, the back-transformation to the model space is applied such that

$$\pi_a(\kappa) = \pi_q(q) \frac{dT_q(\kappa)}{d\kappa} \tag{9}$$

holds, where $dT_q(\kappa)/d\kappa$ denotes the Radon-Nikodým derivative of the assimilation measure with respect to the original measure. In this manner the process of computing $\pi_a(\kappa)$ is equivalent to the problem of evaluating the likelihood function $L(q) = p(z|q)$. The likelihood is incorporating the information from the data into the updating process, and hence, it is shaped by the measurement density [28]. For

4

the measurements that one assumes a Gaussian distribution, the likelihood takes the form

$$L(q) = \exp\left(-\frac{1}{2}(d)^{\mathrm{T}} C_\varepsilon^{-1}(d)\right), \tag{10}$$

in which d denotes the difference

$$d = z - y \tag{11}$$

between the forecast $y = Y(q, u)$ and the measurement z, whereas C_ε stands for the measurement covariance.

Following this, the evaluation of Eq. (8) corresponds to the simulation of the forward problem

$$-\operatorname{div}(T_q(q)\nabla u(x)) = f(x), \tag{12}$$

and the response forecast

$$y(x, \omega) = Y(q(\omega); u(x, \omega)). \tag{13}$$

Here, Y denotes the observation operator with respect to the transformed parameter q such that

$$z = H(\kappa) = Y(q), \quad Y = H \circ T_q^{-1} \tag{14}$$

holds.

Note that in a special case when the random variable κ follows a lognormal distribution the transformation in Eq. (7) coincides with the Gaussian anamorphosis [25, 24]. Indeed, by defining q as $\log(\kappa)$, the non-Guassian RV κ transforms to the Gaussian RV q.

The transformations previously mentioned are especially important when the second version of Bayes's rule is applied—the one that upates the measurable function. This alternative formulation of Bayes's rule can be achieved by expressing the conditional probabilities in Eq. (8) in terms of conditional expectation. Following the mathematical derivation in [21, 22], this approach boils down to a quadratic minimisation problem:

$$q_a(\omega) = P_{\mathscr{Q}_{sn}} q = \arg\min_{\eta \in \mathscr{Q}_{sn}} \|q - \eta\|_{L_2}^2, \tag{15}$$

in which $P_{\mathscr{Q}_{sn}}$ is the orthogonal projection operator of q onto the space of the new information $\mathscr{Q}_{sn} := \mathscr{Q} \otimes S_n$. This space of $\mathscr{Q}$-valued random variables with finite variance is defined by the triplet $S_n := L_2(\Omega, \mathfrak{S}, \mathbb{P})$, where $\mathfrak{S} := \sigma(Y)$ denotes the sub-σ-algebra generated by Y. According to the Doob-Dynkin lemma [3], one may state

that $\eta := \phi \circ Y \circ q$, in which ϕ belongs to the space $L_0(\mathscr{Y}, \mathscr{Q})$ of measurable maps. Constraining the vector space $L_0(\mathscr{Y}, \mathscr{Q})$ to the subspace of linear maps $\mathscr{L}(\mathscr{Y}, \mathscr{Q})$, the minimisation problem in Eq. (15) leads to a unique solution K. This gives an affine approximation of Eq. (15)

$$q_a(\omega) = q_f(\omega) + K(z(\omega) - y_f(\omega)), \tag{16}$$

also known as a linear Bayesian posterior estimate. Here, q_f represents the prior random variable, q_a is the posterior, y_f is the forecasted measurement and K represents the very well-known Kalman gain

$$K := C_{q_f y_f} \left(C_{y_f} + C_\varepsilon \right)^{-1} \tag{17}$$

which can be easily evaluated if the appropriate covariance matrices $C_{q_f y_f}$, C_{y_f} and C_ε are known. We would like to emphasise that the Hilbert-space setting of $\mathscr{Q}$ and $\mathscr{Y}$ has made the formulation in Eq. (16) possible [12]. Therefore, the transformation in Eq. (7) was neccesary.

It is interesting to note that the projection in Eq. (16) is performed over a smaller space than $\mathscr{Q}_{sn}$. An implication of this is that available information is not completely used in the process of updating. It is therefore likely that the minimisation error remains larger. However, the computation of the projection becomes simpler. Another advantage of Eq. (16) compared to Eq. (8) is that the inference in Eq. (16) is given in terms of RVs instead of conditional densities. Namely, $q_a(\omega)$, $q_f(\omega)$, $z(\omega)$ and $y_f(\omega)$ denote the RVs used to model the posterior, prior, observation and forecasted observation, respectively.

Having in mind that Y is in general nonlinear, one may improve the estimation in Eq. (16) by linearising the measurement. In other words, one may apply the nonlinear transformation

$$z_t = T_z(z) = (T_z \circ H \circ T_q^{-1})(q) = G(q) \tag{18}$$

such that the transformed measurement z_t is linear in q. In our example in Eq. (1) one has that

$$\kappa \sim 1/z. \tag{19}$$

Furthermore, if κ follows lognormal distribution

$$\kappa \sim \exp(q), \tag{20}$$

the transformation in Eq. (18) reads
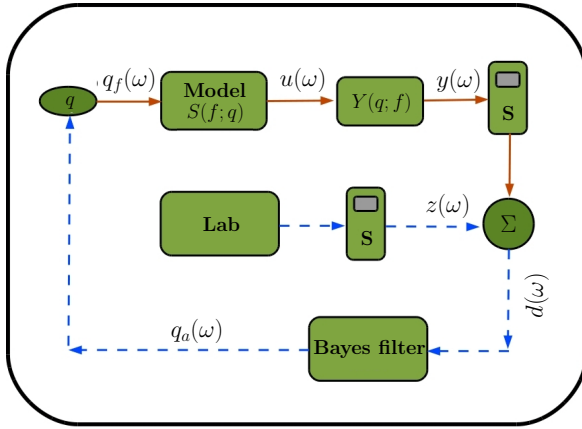
$$z_t = -\log(z) \sim q. \tag{21}$$

6

Figure 1: Schematic representation of Bayesian approach to identification

The last relation coincides with the Gaussian anamorphosis because the non-Gaussian z is transformed to a standard Gaussian z_t. In this example the transformation is easy to achieve as it is purely algebraic. When an algebraic transformation is not possible, one may apply empirical anamorphosis function as shown in [24]. Note that for the multivariate case, the transformation has to be applied to each of random variables individually (locally).

After the measurement transformation, Bayes's rule assimilates the measurement data z_t with the prior information q by means of the following formula:

$$p(q|z_t) = \frac{p(z_t|q)}{p(z_t)} p(q). \tag{22}$$

For a Gaussian measurement error, this means evaluation of the likelihood function

$$L(q) = \exp\left(-\frac{1}{2}(T_z(d))^{\mathrm{T}} C_{\varepsilon_t}^{-1}(T_z(d))\right), \tag{23}$$

in which the distance d between the measurement z and the forecast y is transformed to $T_z(d)$, as well as the covariance function from C_ε to C_{ε_t}. Hence, by transforming observations, the measurement errors also transform. Such a transformation can lead to an overestimation of the measurement error with respect to the transformed forecast error as observed in [25]. Therefore, one often advises to compute the variance of the measurement error in the assimilation space directly from the transformed

7

measurements.

The linear Bayes's rule corresponding to Eq. (22) is characterised by a solution belonging to $\mathcal{Q} \otimes S_n$, in which $S_n := L_2(\Omega, \mathfrak{S}, \mathbb{P})$, where $\mathfrak{S} := \sigma(G)$ denotes the sub-σ-algebra generated by G. In our case G describes the linear relation between q and z_t, and hence the linear Bayes's rule in Eq. (16) becomes optimal. However, note that the transformations introduce additional errors into the computation process. This is especially the case for the measurement errors, as already explained.

The schematic representation of the Bayesian inference is shown in Fig. 1. The scheme describes the closed loop of one Bayesian update. The loop starts by assuming the prior distribution $q_f(\omega)$, which is then propagated through the model $S(f;q)$ and the measurement operator Y to the forecasted (predicted) measurement $y(\omega)$ read by sensor S. The prediction is then substracted from noisy data $z(\omega)$ coming from real experiments, and the resulting difference is forwarded to the Bayesian filter, which further produces the posterior distribution $q_a(\omega)$, i.e. the updated value of $q_f(\omega)$.

4 Computational approaches

In recent years there has been an increasing amount of literature on computational approaches related to Bayesian inference. However, this paper reviews only the research conducted on MCMC [15, 10] and linear Bayesian filters [21, 19, 2]. The main aim of this work is to contrast the linear Bayesian methods to a full MCMC approach on a numerical example.

4.1 Markov chain Monte Carlo

Markov chain Monte Carlo is a sampling procedure used for the estimation of the posterior probability density function via Eq. (5) or Eq. (8), respectively. The method is very general as it does not require any model approximations in contrast to those further described. Instead, MCMC constructs a Markov chain with the posterior as an equilibrium distribution. The two most often used types of this algorithm are: the Gibbs sampling technique [26] and the Metropolis-Hastings algorithm [4], the one used in this paper.

8

As shown in Algorithm 1, the Metropolis scheme generates a sequence of samples (states) $q^{(i)}$ whose values solely depend on the previous sample in the chain. The value of the new state $q^{(i)}$ is generated with the help of the proposal distribution $g_k(q|q^{(i-1)})$, also known as the transition kernel. In practice the transition kernel is very often chosen to be symmetric such that $g_k(q^{(*)}|q^{(i)}) = g_k(q^{(i)}|q^{(*)})$. A typical example of such a kernel is the normal distribution centered at the previous state—the random walk chain. However, other types of kernels can be used in a similar manner, e.g. independence chains, rejection sampling chains, auto-regressive chains or grid-based chains, see [30]. Once the new sample is drawn it is either accepted with the probability r or rejected with probability $1-r$.

1: **procedure** MHA($p(q), g_k, L(q), N$)
2: draw initial value $q^{(0)}$ from prior $p(q)$
3: **for** *each* $i = 1 \rightarrow N$ **do**
4:      draw $q^{(*)}$ from proposal distribution
5:          $g_k(q^{(*)}|q^{(i-1)})$
6:      evaluate the probability of acceptance
7:          $r = \min\{1, \frac{\pi_a(q^{(*)})g_k(q^{(i-1)}|q^{(*)})}{\pi_a(q^{(i-1)})g_k(q^{(*)}|q^{(i-1)})}\}$
8:      accept the next state with probability r
9:          $q^{(i)} = q^{(*)}$
10:     or reject with probability $1-r$
11:         $q^{(i)} = q^{(i-1)}$
          end
12: **end procedure**

Algorithm 1: Metropolis-Hastings procedure

The advantage of the previous algorithm is that it does not need target probabilities but only ratios of target probabilities to work. In this manner the computation of the normalisation constant in Eq. (8) is avoided. However, the samples are not independent any more as they are drawn from the proposal distribution with the probability r. Even though this is the case, the obtained samples can still be used for the evaluation of integrals in a Monte Carlo fashion. However, this works only if the Markov chain is aperiodic, irreducible, and positive recurrent [32]. Under the previously mentioned regularity conditions the sampling sequence $q^{(i)}$ converges in distribution to our target posterior distribution (Theorem 3 in [30]) regardless of the starting point. However, the speed of the convergence greatly depends on the initial choice. Due to this, a few starting samples have to be excluded from the chain a so-called "burn in period".

Computationally, the Markov chain Monte Carlo algorithm is a very demanding pro-
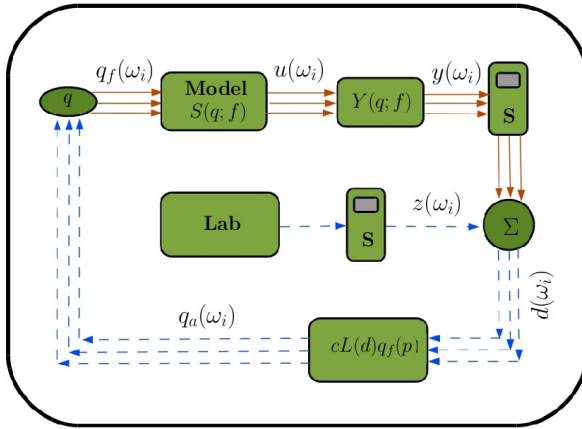
9

Figure 2: The algorithmic scheme of an inverse problem solved by MCMC filtering

cedure because one has to evaluate the system response for each new proposed sample, see Fig. 2. The evaluation starts with samples that are used as input in the deterministic model $S(f, q)$, which closely describes the system response u as a function of the external loading f and the input parameters q. Once the response is computed, the measurement operator Y is applied and the value of the observable quantity y is diagnosed by a sensor. The predicted measurements are then compared to the real data, which further results in the distance measure d entering the likelihood function, see Eq. (10). In this manner a posterior sample is obtained, and the process is repeated all over again for the next sample.

4.2 Proxy modelling

To speed up the assimilation process one may introduce a proxy model for the forecasted measurement. Usually, the proxy model is made with the help of a functional approximation of random variables/fields entering the process, and a stochastic Galerkin procedure [22]. To this end, both the predicted system response u_f and observation y_f can be represented in a polynomial chaos expansion form

10

[17, 35, 11, 33]

$$\begin{array}{rcl} \hat{u}_f(\omega) & = & \displaystyle\sum_{\mathscr{J}} u_f^{(\alpha)} H_\alpha(\omega), \\[2mm] \hat{y}_f(\omega) & = & \displaystyle\sum_{\mathscr{J}} y_f^{(\alpha)} H_\alpha(\omega), \end{array} \tag{24}$$

in which $H_\alpha(\omega)$ represent the generalised orthogonal polynomials and $\mathscr{J}$ stands for the set of all finite non-negative integer sequences, i.e. multi-indices α such that

$$\begin{array}{rcl} \mathscr{J} := \{ \alpha = (\alpha_1, ..., \alpha_j, ...) & | & \alpha_j \in \mathbf{N}_0, \\[2mm] & & |\alpha| := \displaystyle\sum_{k=1}^{\infty} \alpha_j < \infty \} \end{array} \tag{25}$$

holds. Due to computational reasons, only a finite subset of $\mathscr{J}$ is taken, i.e. the expansion in Eq. (24) is truncated to a finite number of terms. This results in another type of error which has to be added to the modelling error mentioned above.

To increase its efficiency, the MCMC cycle can be modified such that the forward model is substituted with a less accurate but computationally cheaper proxy model (see Fig. 3), as already reported in [14]. In this manner the forward model is not individually solved for each MCMC sample, but apriori, see Fig. 3. After evaluating the functional approximation of the measurement, the sampling occurs and the update loop proceeds in the same manner as described previously, see Fig. 2 and Fig. 3.

4.3 Linear Bayesian inference

The advantage of the methods described in the previous section is that they are model-independent. However, their main drawback is the slow convergence. The issue of high computational cost can be improved via recently developed Bayesian linear methods [21, 19, 22, 18] as shown in Eq. (16). Recalling that the RV $q_a(\omega)$ can be numerically represented by either sampling $q_a(\omega_i)$ or the functional approximation such as polynomial chaos expansion, one may distinguish at least two numerical approaches to the problem given in Eq. (16): the ensemble Kalman filter [7] and the polynomial chaos based update [21, 22].
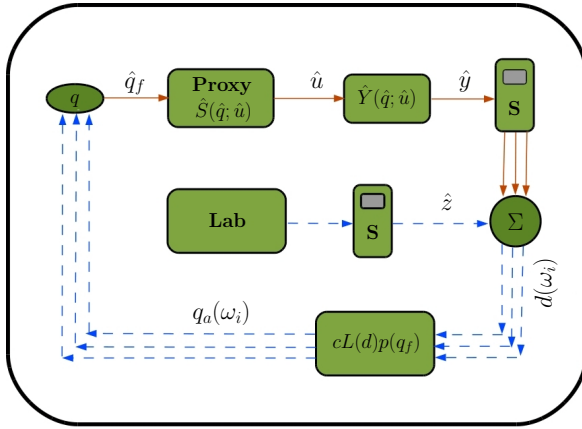
11

Figure 3: The algorithmic scheme of an inverse problem solved by proxy MCMC filtering

4.3.1 Ensemble Kalman filter

The simplest way to numerically estimate q_a is to sample Eq. (16) in a Monte Carlo fashion. Such a procedure starts by building ensembles of prior samples $\boldsymbol{Q}_f := [q_f(\omega_1), \ldots, q_f(\omega_Z)]$, forecasts $\boldsymbol{Y}_f := [y_f(\omega_1), \ldots, y_f(\omega_Z)]$ and measurements $\boldsymbol{Z}$, such that Eq. (16) can be formulated in a matrix notation as

$$\boldsymbol{Q}_a = \boldsymbol{Q}_f + \boldsymbol{K}(\boldsymbol{Z} - \boldsymbol{Y}_f), \tag{26}$$

in which $\boldsymbol{K}$ takes the form as in Eq. (16). Note that its corresponding covariances may be estimated from the ensemble, i.e.

$$\boldsymbol{C}_{q_f, y_f} \approx \frac{1}{Z-1} \tilde{\boldsymbol{Q}}_f \tilde{\boldsymbol{Y}}_f^T \quad \text{and} \quad \boldsymbol{C}_{y_f} \approx \frac{1}{Z-1} \tilde{\boldsymbol{Y}}_f \tilde{\boldsymbol{Y}}_f^T. \tag{27}$$

Here, $\tilde{\boldsymbol{Q}}_f = \boldsymbol{Q}_f - \bar{q}_f \mathbf{1}_Z^T$ and $\tilde{\boldsymbol{Y}}_f = \boldsymbol{Y}_f - \bar{y}_f \mathbf{1}_Z^T$ represent the fluctuation parts of corresponding RVs, where $\bar{q}_f = \frac{1}{Z} \sum_{z=1}^Z q_f(\omega_z)$ and $\bar{y}_f = \frac{1}{Z} \sum_{z=1}^Z y_f(\omega_z)$ are the estimated means and $\mathbf{1}_Z$ is a vector of ones of size Z.

This method is a Monte Carlo method, hence it also suffers from the slow convergence with increasing Z. On the other hand, it is fairly simple to implement: all it needs are random samples, see Fig. 4. In practice the number of samples is often low,

12
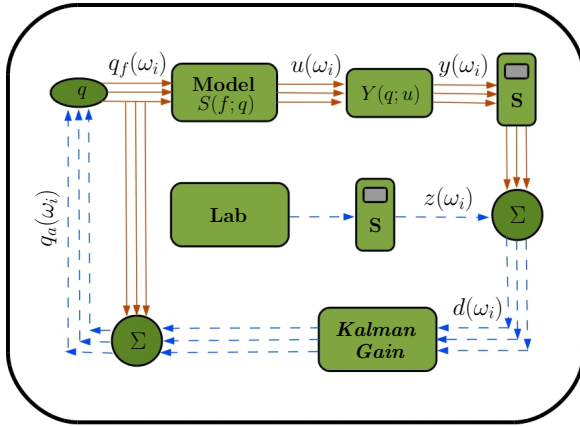
Figure 4: The algorithmic scheme of an inverse problem solved by the ensemble Kalman filter

and then special care is needed when computing the covariances and the Kalman gain K, see [8]. To reduce the computation time one may use the proxy model instead of a forward simulator in a similar way as it is done in the MCMC procedure. In this manner only the update formula in Eq. (16) is sampled.

4.3.2 Polynomial chaos based linear Bayesian update

To avoid the sampling procedure presented previously in a form of the ensemble Kalman filter (EnKF) algorithm, one may use the opportunity to functionally approximate the random variables (fields) in Eq. (16). In this light the linear Bayesian procedure can be reduced to a simple algebraic method. Starting from the functional representation of the prior

$$\hat{q}_f = \sum_{\alpha} q_f^{(\alpha)} H_\alpha(\omega) \tag{28}$$

and the proxy in Eq. (24), one may discretise Eq. (16) as:

$$\hat{q}_a = \hat{q}_f + K\big(\hat{\mathbf{z}} - \hat{\mathbf{y}}_f\big), \tag{29}$$

13
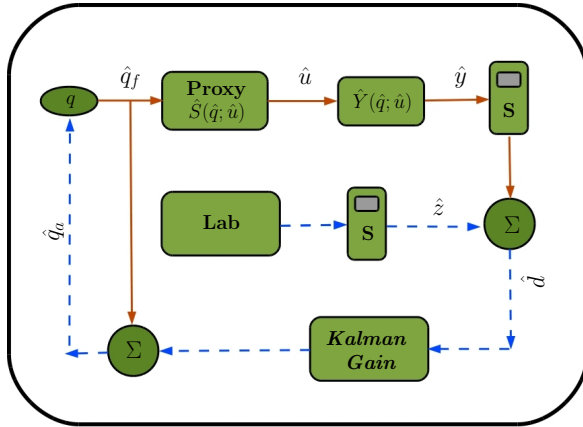
Figure 5: The polynomial chaos based linear Bayesian scheme

where $\hat{\mathbf{z}} \in \mathbb{R}^{L \times Z}$ is the PCE of the measurement. Here, K in Eq. (29) is the Kalman gain evaluated in an algebraic way knowing that

$$C_{q_f, y_f} = \sum_{\alpha > 0} \alpha! \, q_f^{(\alpha)} (\mathbf{y}_f^{(\alpha)})^T. \tag{30}$$

Note that in the numerical computation $\hat{q}_a \in \mathbb{R}^Z$, $\hat{q}_f \in \mathbb{R}^Z$, $\hat{\mathbf{y}}_f \in \mathbb{R}^{L \times Z}$ and $\hat{\mathbf{z}} \in \mathbb{R}^{L \times Z}$ are PCEs with cardinality Z determined by $(L+1)$ RVs and polynomial order p. Here, the number $(L+1)$ subsumes all the RVs describing the prior and the RVs $\{\theta_i\}_{i=1}^L$ used to model the measurement error ε.

The previous algorithm is shown in Fig. 5 where the whole update process can be represented by only one loop.

4.3.3 Square root polynomial chaos based linear Bayesian update

The idea of linear Bayesian inference is allowing the computation of the posterior in a quite efficient way, however the update requires the introduction of additional RVs—corresponding to ε— into the update process. This essentially may enlarge the dimension of the stochastic space one is working with, especially in case of sequential updating (for an illustration see [22, 18]). To avoid the presence of the

14

observation RVs and corresponding PCE in Eq. (16), one may follow the idea of a square root filter [1], as the authors already addressed in [19]. In such an algorithm the evaluation of the posterior consists of two phases:

1. the estimation of the posterior mean via

$$\mathbb{E}(\hat{q}_a) = \mathbb{E}(\hat{q}_f) + K(\mathbb{E}(\hat{z}) - \mathbb{E}(\mathbf{y}_f)), \tag{31}$$

2. the prediction of the varying part $\tilde{q}_a := \hat{q}_a - \mathbb{E}(\hat{q}_a)$ via

$$\tilde{q}_a = \frac{S_{q_a}}{\sqrt{\Delta}}. \tag{32}$$

Here, $\sqrt{\Delta} = \mathrm{diag}\,(\sqrt{\alpha!})$ stands for the square root of the Gram matrix $\Delta = \mathbb{E}(H_\alpha H_\beta)$, while S_{q_a} denotes the matrix square root of the posterior covariance $C_{q_a} = S_{q_a} S_{q_a}^T$. The latter one represents a linear transformation of the prior matrix square root S_{q_f}, i.e.

$$S_{q_a} = S_{q_f} V \sqrt{I - \Sigma^T \Sigma} V^T. \tag{33}$$

The transformation essentially comes from the definition of the covariance structure $(C_{y_f} + C_\varepsilon)$ and its decomposition

$$\left(C_{y_f} + C_\varepsilon\right) = B\Lambda B^T, \tag{34}$$

where B^T rotates the simulated measurements into directions aligned with the covariance structure $(C_{y_f} + C_\varepsilon)$, while $\Lambda^{-\frac{1}{2}}$ weights them accordingly. In this light matrices V and Σ are obtained by the singular value decomposition of

$$W = \Lambda^{-\frac{1}{2}} B^T H S_{q_f}, \tag{35}$$

for more details on mathematical derivation please see [19].

Eq. (34) is exactly the place where the additional information (in the form of C_ε) enters the update— C_ε specifies directions and magnitude of the uncertainty (variance) reduction induced by the observation. Thus, no additional random variables have to be included into the update. However, note that the square-root formulation is only equivalent to the standard linear filter form in case of Gaussianity. In addition, if the random variable is represented by an ensemble, the previous algorithm is of the ensemble square root type.

15

5 Numerical results

To test the numerical procedures described previously, two benchmark problems are introduced: a two point boundary value problem in one spatial dimension and a heat conduction problem for a rectangular plate.

5.1 One dimensional heat problem

The thin metal rod of unit length is exposed to the deterministic heat source $f(x) = 5(1-x)$ linearly dependent on the spatial coordinate x. The heat transfer in the rod is assumed to be steady, i.e. described by the one-dimensional heat equation:

$$-\operatorname{div} \kappa(\omega)u(x,\omega) = f(x,\omega), \tag{36}$$

with zero Dirichlet boundary conditions. Here, $\kappa(x,\omega)$ denotes the conductivity coefficient one is uncertain about and $u(x,\omega)$ is the temperature response.
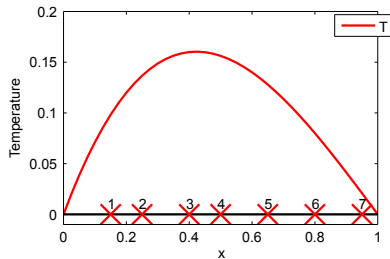
Figure 6: The virtual measurement: the temperature dependence on the point position x. The red crosses represent the sensor placements.

The true value of the thermal conductivity κ_t is taken to be one realisation of a log-normal random variable described independently from the a priori distribution—so called truth. The corresponding data set–observations–are then obtained with the help of the deterministic finite-difference (FD) approach. The temperature is estimated on a uniform mesh of 41 points, from which $L = 7$ randomly chosen nodes are used to place the measurement sensors, see Fig. 6. Each measurement is subjected to

16

Gaussian noise with zero mean and covariance $C_\varepsilon = \sigma_\varepsilon^2 I$ (I is the identity matrix of size L).

For the purpose of measurement prediction, the prior conductivity $\kappa_f(\omega)$ is assumed to be a lognormally distributed RV with mean 2.3 and standard deviation 0.3. Once the prior is chosen, the predicted measurement is evaluated on the uniform mesh of 21 spatial points with the help of the stochastic Galerkin approach, see Eq. (36). Note that the spatial mesh is taken to be different than the one chosen for the computation of the "virtual truth". This is done in order to avoid the "inverse crime" problem [34].

As depicted in Fig. 7 a), the "virtual truth" is taken to lie in

- C_1: high probability ($\kappa_t = 2$),

- C_2: 2σ ($\kappa_t = 1.7$),

- and C_3: low probability ($\kappa_t = 1.4$)

regions of the prior. While the first case scenario (i.e. when $\kappa_t = 2$) represents a reliable assumption of the prior, the other two case scenarios are describing situations in which one cannot have precise expert knowledge on the value of the parameter κ.

Following this, Fig. 7 b) compares the predicted measurement and the observation including the measurement noise for each of the previously mentioned scenarios. These results argue how much of an impact the prior distribution has on the distance between the measured and observed data sets. Namely, if the prior is such that the truth lies in the high probability region then the distance is small, and vice versa.

5.1.1 Nonlinear measurement

The experimental set up as described previously introduces the temperature observations into the identification process. However, one may note that this kind of measurement is inconsistent with the statement made in Section 3 where the linear dependence between the measurement and the parameter has been assumed. According to Fig. 8, this hypothesis does not hold. Even though this is the case, the objectives of the following numerical computations are to determine whether and up to which degree the truth can be identified from the nonlinear measurement data.

17
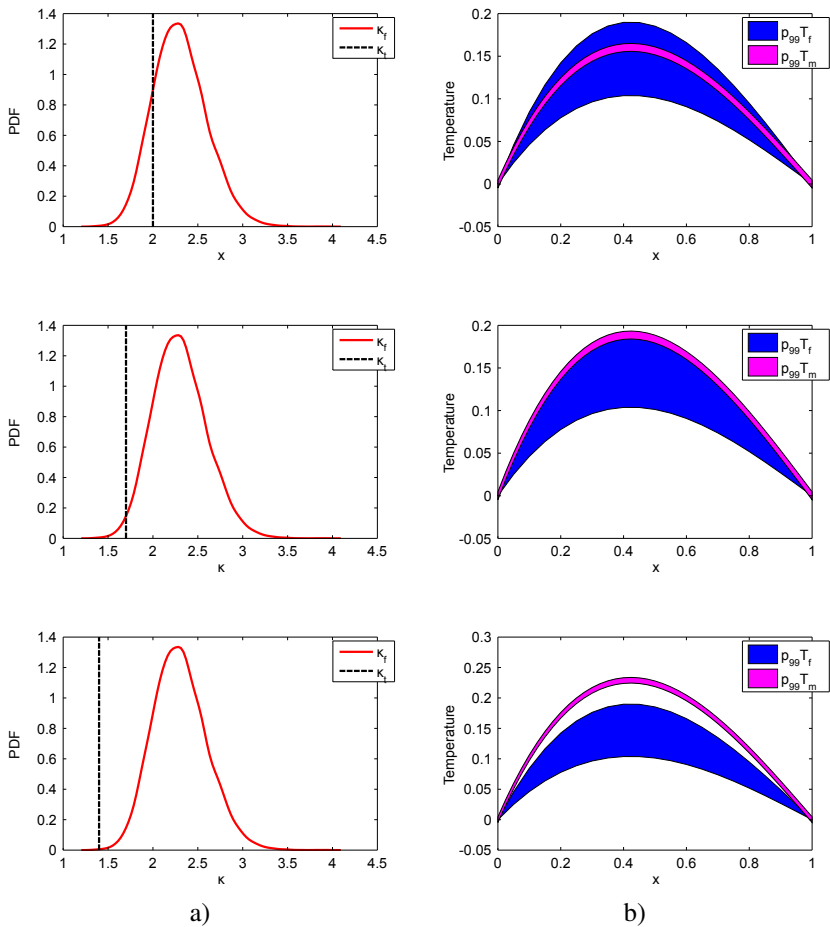
Figure 7: a) Three case scenarios for the true value of the conductivity coefficient κ b) the predicted vs. the virtual measurement: 99% probability regions of the predicted temperature T_f and 99% probability regions of the measurement T_m
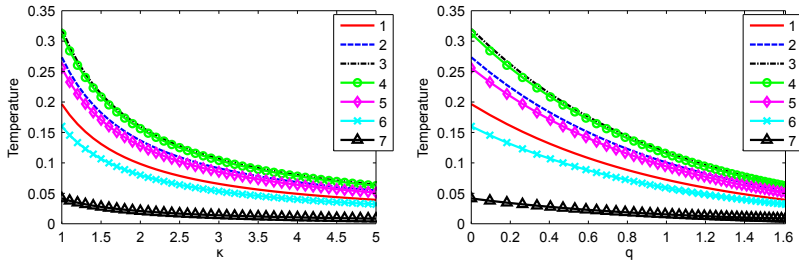
18

Figure 8: The dependence of temperature on the parameter set κ (left figure) and its transformation q (right figure)

The six identification procedures—one random variable based (reduced) linear Bayesian update (RLBU)[1], full linear Bayesian update (FLBU), the square root update (SQRT), ensemble Kalman filter (EnKF) with 1000 samples, square root ensemble Kalman filter (EnKFS) with 1000 samples, and the full Bayesian MCMC update with 10^5 samples—are used to estimate the value of the parameter κ given seven temperature observations. The update process is performed only once using the complete measurement data. The results obtained, as shown in Table 1 and Table 2, indicate that the MCMC procedure is the only one which can identify the truth in all three case scenarios. In contrast to this, the linear approximants are able to estimate the truth only in the first case scenario although with an overestimated standard deviation. The overestimation appears to be stronger in case of the square root posterior, as well as posteriors obtained from the ensemble data (EnKF-kind of procedures). Since the square root estimation is not equivalent to the linear Bayesian; and since the ensemble Kalman filter estimates strongly depend on the chosen seed (here 1000 samples), this finding was expected. Contrary to the expectations, the one random variable linear Bayesian update is underestimating the posterior variance. It seems that the underestimation happens due to constraints put on the basis on which the posterior is projected. Namely, one random variable linear Bayesian update is neglecting (projecting out) the additional random variables coming from the measurement data in the process of updating.

These findings are consistent with the plots of posterior probability density functions given in Fig. 9 and Fig. 10, where one may clearly notice that the MCMC procedure

---

[1] random variables describing the measurement error are not taken into consideration—they are projected out during the update process

19

Table 1: Comparison of modes and standard deviations for the posterior κ obtained by different update procedures

Parameter	Method	Mode			Std		
	Case:	C_1	C_2	C_3	C_1	C_2	C_3
	Truth	2	1.7	1.4	0	0	0
	MCMC	1.9979	1.6952	1.4102	0.0131	0.0123	0.0163
κ	RLBU	1.9934	1.6338	1.2294	0.0117	0.0096	0.0072
	FLBU	1.9976	1.6373	1.2314	0.0184	0.0151	0.0113
	SQRT	1.9814	1.6236	1.2224	0.0274	0.0225	0.0169
	EnKF	1.9994	1.6328	1.2341	0.0262	0.0202	0.0158
	EnKFS	1.9859	1.6237	1.2311	0.0262	0.0207	0.0160

Table 2: Comparison of modes and standard deviations for the posterior q obtained by different update procedures

Parameter	Method	Mode			Std		
	Case:	C_1	C_2	C_3	C_1	C_2	C_3
	Truth	0.6931	0.5306	0.3365	0	0	0
	MCMC	0.6921	0.5278	0.3437	0.0066	0.0069	0.0098
q	RLBU	0.6899	0.4909	0.2065	0.0058	0.0058	0.0058
	FLBU	0.6928	0.4935	0.2095	0.0092	0.0092	0.0092
	SQRT	0.6839	0.4849	0.2004	0.0134	0.0134	0.0134
	EnKF	0.6910	0.4928	0.2064	0.0119	0.0126	0.0124
	EnKFS	0.6858	0.4892	0.2010	0.0121	0.0126	0.0120
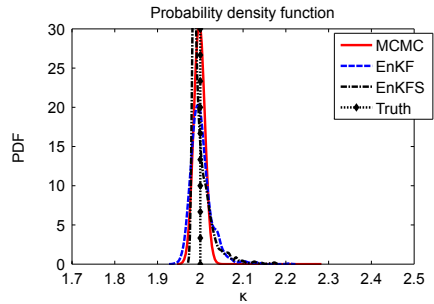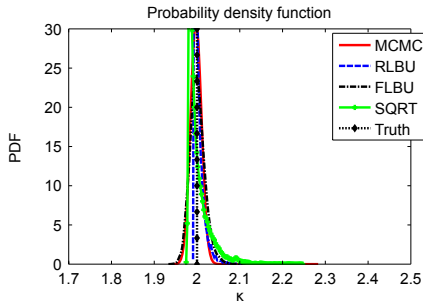
20

a) The truth lies in the one sigma region of the prior
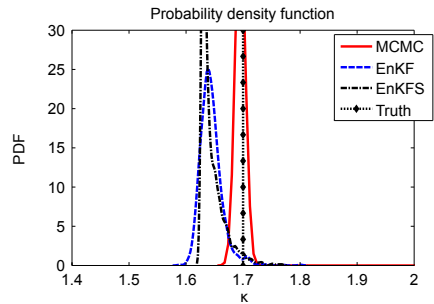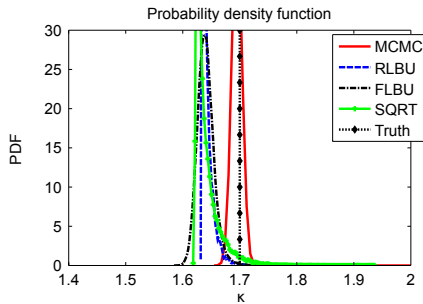
b) The truth lies in the two sigma region of the prior
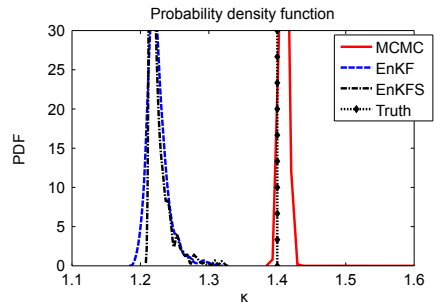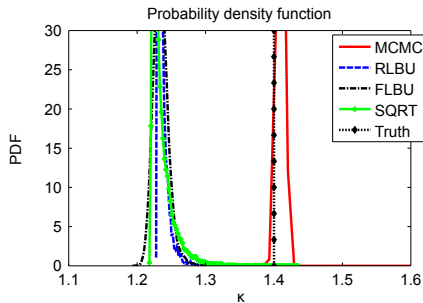
c) The truth lies in the three sigma region of the prior

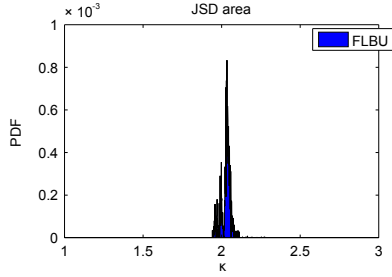Figure 9: Comparison of posterior probability density functions describing κ
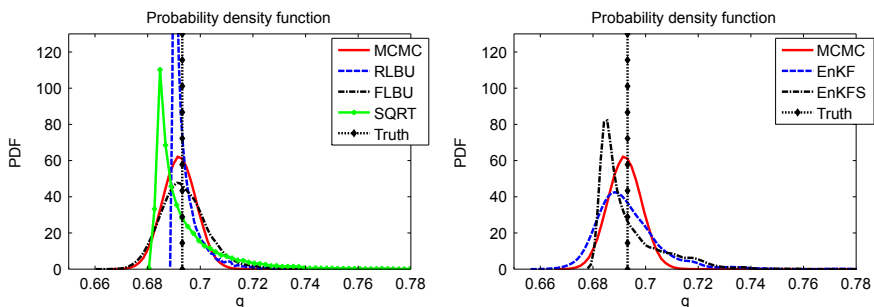
a) The truth lies in the one sigma region of the prior
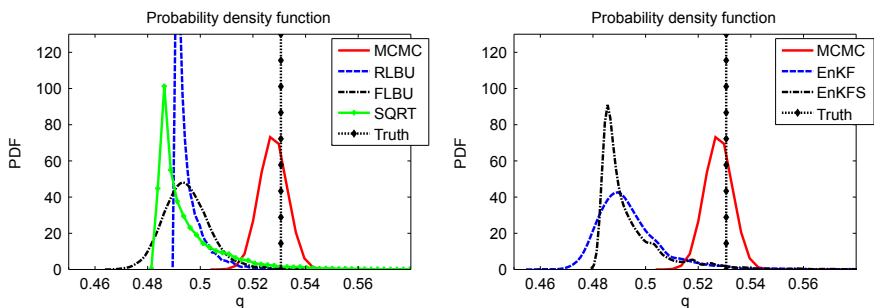
b) The truth lies in the two sigma region of the prior
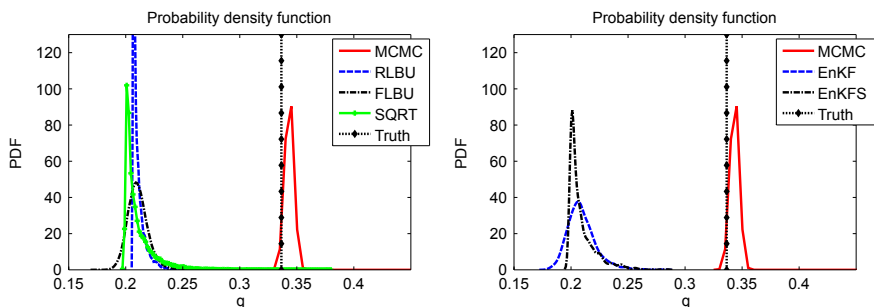
c) The truth lies in the three sigma region of the prior

Figure 10: Comparison of posterior probability density functions describing q

Figure 11: The JSD area measuring the distance of the FLBU and MCMC posteriors describing κ

outperforms other numerical techniques. Moreover, the linear Bayesian methods "overshoot" when the truth is assumed to be $\kappa_t = 1.7$ or $\kappa_t = 1.4$. This issue can be explained in part by a nonlinearity of the measurement operator. However, there are other possible explanations, such as the prior which is "wrongly" assumed.

To determine the similarity between the MCMC (π_1) and other posterior distributions (π_i), the Jensen-Shannon divergence (JSD):

$$\mathrm{JSD}(\pi_i \| \pi_1) = \frac{1}{2} D(\pi_i \| \frac{1}{2}(\pi_i + \pi_1)) + \frac{1}{2} D(\pi_1 \| \frac{1}{2}(\pi_i + \pi_1)) \tag{37}$$

is estimated. This metric is a smooth and symmetrised version of the Kullback-Leibler divergence $D(\pi_i \| \pi_1)$ defined as

$$D(\pi_i \| \pi_1) = \int_{-\infty}^{\infty} \ln\left(\frac{\rho_i(x)}{\rho_1(x)}\right) \rho_i(x) \, \mathrm{d}x, \tag{38}$$

where ρ_i and ρ_1 represent the densities of the quantity obtained by one of the Bayesian linear approximation methods and MCMC methods, respectively. In the first case scenario of the truth, the JSD value between the MCMC and FLBU probability distributions is equal to the total area under a curve in Fig. 11. This value corresponds to the error obtained by accumulation of sampling errors and the error caused by a nonlinearity of the measurement operator.

23

5.1.2 Linear measurement

To get an adequate understanding of the conclusions drawn in the previous section, one has to consider the experiments in which the measurement operator is linear. Since the relationship between the parameter and the observation is explicitly known

$$\kappa \sim \exp(q) \sim 1/z, \tag{39}$$

one may linearise the measurement operator via the following transformation

$$\log\kappa \sim q \sim \log(1/z). \tag{40}$$

In such a case the estimation parameter q linearly depends on $\log(1/z)$. Following this, the numerical analysis is repeated as in the previous section, only this time with the new version of the measurement. Note that the nonlinear transformation is applied on the measurement data solely without the measurement error.

This study produced results which confirm the findings of a great deal of previous works in this field, see [21, 19, 22, 18]. Namely, as results in Table 3 and Table 4 show, the methods based on the linear Bayesian formula are able to identify the truth in all three case scenarios without strong overestimations of variance. However, this is not the case for the one random variable based linear Bayesian update. This method underestimates the posterior variance similarly to the case study already discussed in the previous section. Therefore, the use of the one random variable based linear Bayesian update is not advised in practice. Furthermore, the MCMC procedure shows a slightly different behaviour than in the nonlinear case. The nonlinear transformation of the predicted measurement and observation data in a polynomial chaos form have resulted in a poor posterior estimation in the worst case scenario, when the truth takes the value in the low probability region of the prior. The problem appears due to large numerical errors caused by both sampling and transformation. The previous findings are in agreement with the probability density plots shown in Fig. 12 and Fig. 13, where one may clearly observe the described behaviour of the linear procedures. The difference in the posterior estimates of the linear Bayesian update and full MCMC procedure for the first case scenario can be seen in Fig. 14, where the JSD area is plotted.

By having run several MCMC computations for the worst case scenario, we came to the conclusion   that the posterior distribution converges to the linear Bayes's posterior with the increase of the polynomial order of the proxy model. Increasing the polynomial order from four to six, the MCMC posterior distribution approaches the one obtained by the full linear Bayesian update of fourth order, as shown in

24

Table 3: Comparison of modes and standard deviations for the posterior κ obtained by different update procedures

Parameter	Method	Mode			Std		
		C_1	C_2	C_3	C_1	C_2	C_3
	MCMC	1.9982	1.7109	1.4878	0.0342	0.0269	0.0209
	RLBU	2.0020	1.7044	1.4062	0.0054	0.0039	0.0034
κ	FLBU	2.0023	1.7046	1.4064	0.0332	0.0281	0.0232
	SQRT	2.0023	1.7046	1.4064	0.0339	0.0288	0.0241
	EnKF	2.0024	1.7038	1.4046	0.0340	0.0294	0.0237
	EnKFS	2.0022	1.7043	1.4054	0.0339	0.0290	0.0245

Table 4: Comparison of modes and standard deviations for the posterior κ obtained by different update procedures

Parameter	Method	Mode			Std		
		C_1	C_2	C_3	C_1	C_2	C_3
	Truth	0.6931	0.5306	0.3365	0	0	0
	MCMC	0.6921	0.5369	0.3972	0.0171	0.0155	0.0209
	RLBU	0.6942	0.5341	0.3407	0.0023	0.0027	0.0022
q	FLBU	0.6942	0.5341	0.3407	0.0165	0.0165	0.0165
	SQRT	0.6942	0.5341	0.3407	0.0167	0.0174	0.0168
	EnKF	0.6948	0.5339	0.3410	0.0167	0.0175	0.0170
	EnKFS	0.6940	0.5346	0.3410	0.0166	0.0173	0.0167
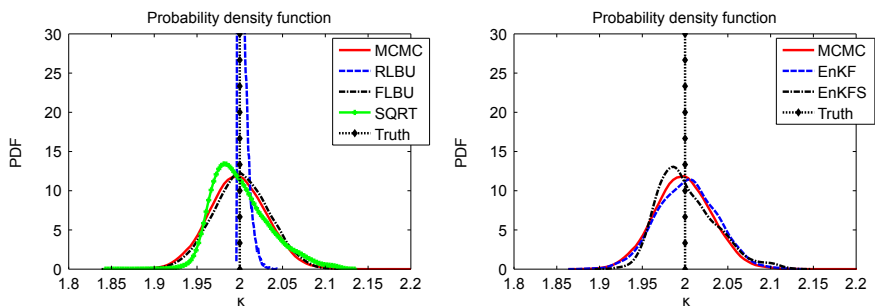
a) The truth lies in the one sigma region of the prior
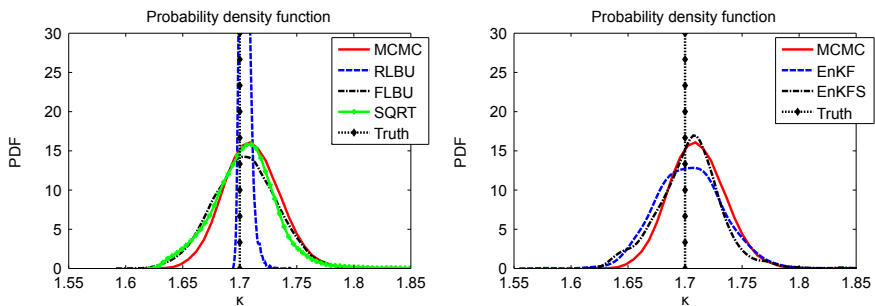
b) The truth lies in the two sigma region of the prior
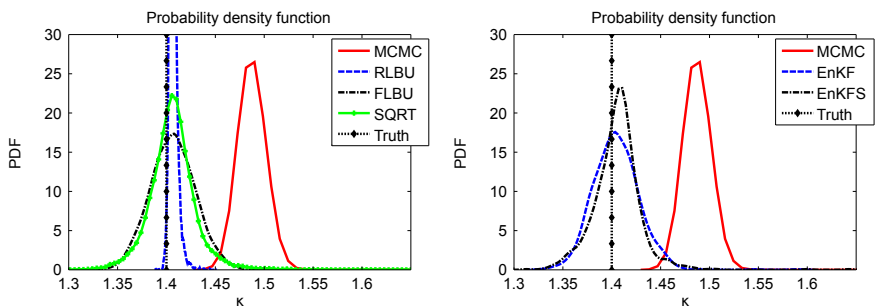
c) The truth lies in the three sigma region of the prior

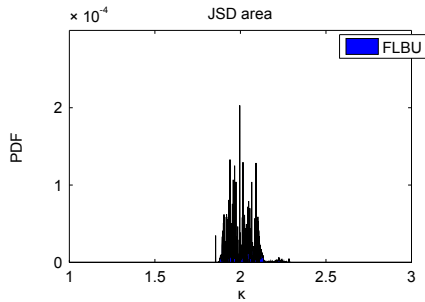Figure 12: Comparison of posterior probability density functions describing κ
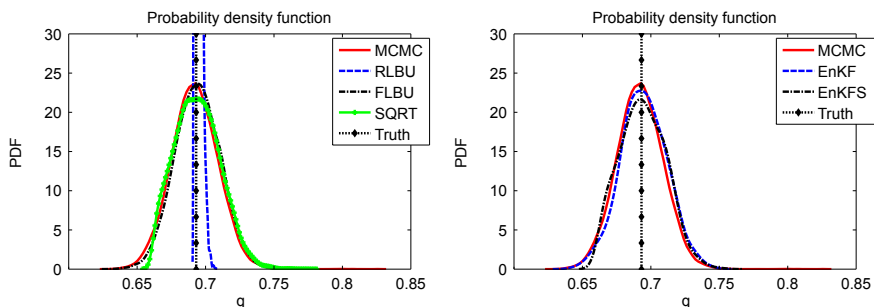
a) The truth lies in the one sigma region of the prior
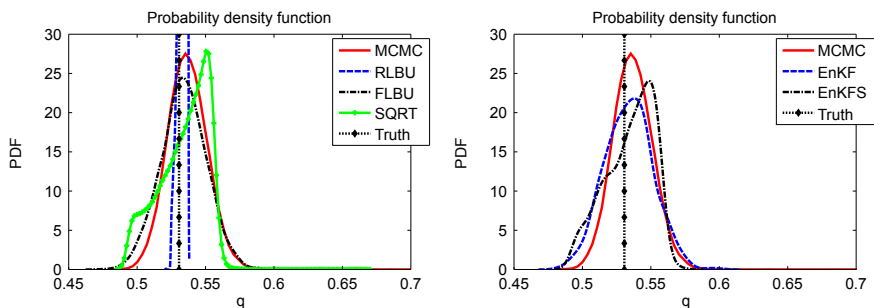
b) The truth lies in the two sigma region of the prior
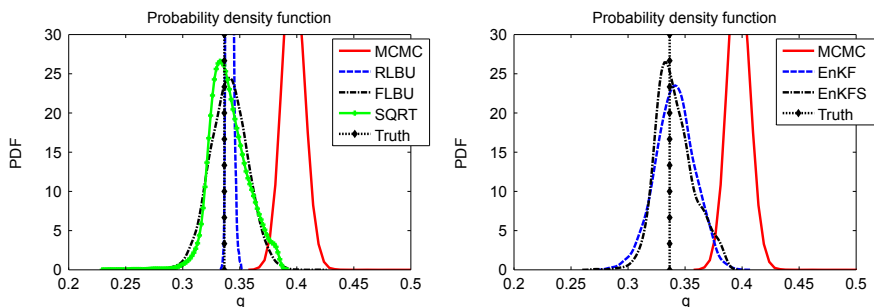
c) The truth lies in the three sigma region of the prior

Figure 13: Comparison of posterior probability density functions describing *q*

Figure 14: The JSD area measuring the distance of the FLBU and MCMC posterior describing κ
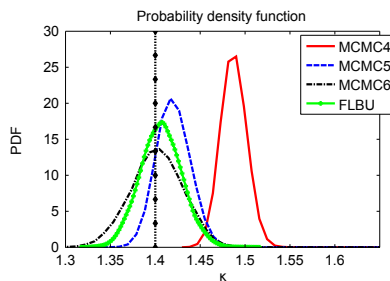
Figure 15: The improvement of posterior distribution obtained by the MCMC procedure with the increase of the polynomial order from 4 (MCMC4) to 6 (MCMC6). The reference is the full linear Bayesian update of polynomial order 4 (FLBU).

28
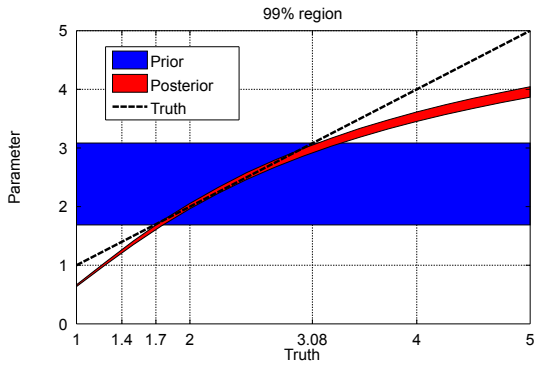
a)
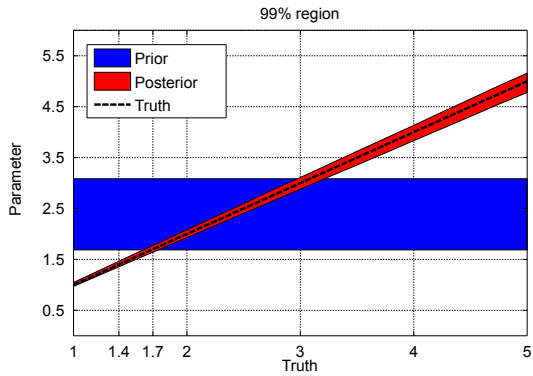
b)

Figure 16: Comparison of the upper 99% bounds of posterior and prior of the parameter κ for: a) nonlinear measurement b) linear measurement

29

Fig. 15. This further means that the modelling error can have a huge influence on the MCMC result in the low probability regions of the prior.

Having the previous results in mind, the update procedure is repeated for all possible "truth scenarios" —for the values of κ_t between one and five—and the prior as described in the beginning of this section. As shown in Fig. 16, the truth is inside the 99% region of the posterior (red area) in case of linear measurement. However, the same line crosses the 99% region of the posterior (red area) in case of nonlinear measurement. For the former scenario, the posterior contains the truth only in a small region around the 2.3 value (prior mean), where the truth line appears to be the tangent. On the other hand, in the linear case the posterior better estimates the truth although the variance can be over- or underestimated. This leads to the conclusion that only the measurement operators with slight nonlinearities can be handled with the linear Bayes procedure.

5.2 Two dimensional heat problem

In order to improve the previous qualitative analysis, the steady diffusion problem is also examined on a two-dimensional rectangular domain, see Fig. 17. The boundary conditions consist of a heat flux $q = 100 W m^{-2}$ prescribed on the left boundary and a constant temperature of twenty degrees Celsius imposed on the right boundary. The computational domain is discretised with the help of 204 irregular finite elements and $N = 124$ nodes.
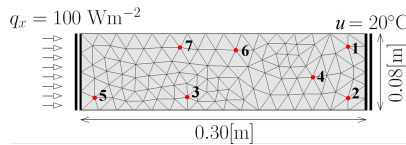

Figure 17: Experimental setup

For the virtual truth, the thermal conductivity κ_t is taken to be one realisation of a lognormal random variable described independently from the a priori distribution. The temperature is evaluated with the help of a deterministic finite element (FE) method, but only the values in 7 randomly chosen points (FE nodes highlighted by red dots in Fig. 17) are taken into consideration. For reasons of simplicity, these points (sensors) are not optimally placed, even though this can be achieved with
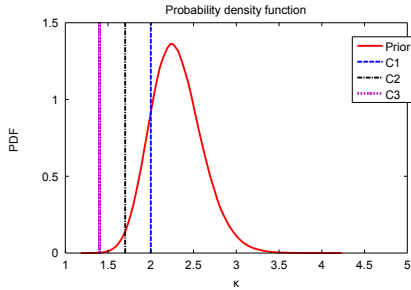
30

Figure 18: The prior distribution and three case scenarios for the truth C_1-C_3

the help of the optimisation theory. The measured data are additionally disturbed by Gaussian noise with zero mean and covariance $C_\varepsilon = \sigma_\varepsilon^2 I$ in order to simulate realistic data.

For this study, the prior thermal conductivity κ_f is designed with the help of the maximum entropy approach which takes all available information about the conductivity parameter in the process of model selection. Thanks to the positive definiteness of κ, the prior is taken to be a lognormal random variable with the mean $\bar{\kappa}_f = 2.3$ and standard deviation $\sigma_f = 0.3$. In a similar manner as before, the value of κ_t is adopted such that the truth places in the one, two or three sigma region of the prior, see Fig. 18. The reasons for this are the same as described before in Section 5.1.

5.3 Forward problem

For prediction purposes the stochastic diffusion problem described by uncertain conductivity coefficient is solved with the help of the stochastic Galerkin method [22]. These results are verified with the help of a pure Monte Carlo approach with one million samples, see Fig. 19 for its convergence in mean and variance. For further investigation only one surrogate model is selected through a validation process: that is the polynomial chaos expansion of order 4. Compared to the MC reference solution this approximation results in 0.3029e-4 for the relative error in the mean, and 0.0011 for the relative error in variance. The mean value and variance of such an approximated solution are shown in Fig. 20. As expected, the mean is a linear and the variance is a nonlinear function of the coordinates.
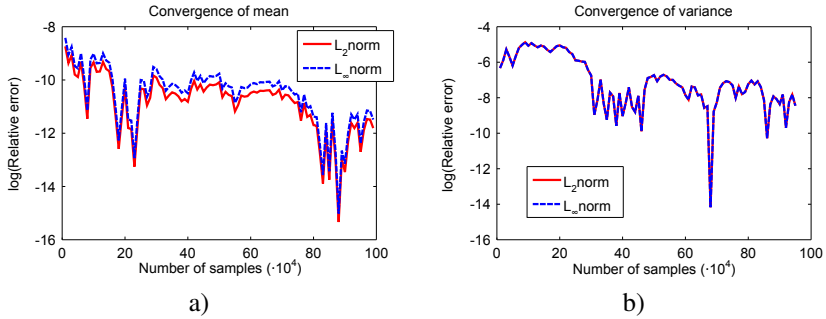
31

Figure 19: Convergence of the a) mean and b) variance of the Monte Carlo method with 10^6 samples

5.4 Identification

As in the previous example, the experimental analysis is run by measuring the temperature, i.e. the nonlinear function of the conductivity parameter. For comparison purposes, several computational strategies as described in Section 5.1 are implemented and tested: the one random variable based linear Bayesian update (RLBU), full linear Bayesian update (FLBU), the square root update (SQRT), ensemble Kalman filter (EnKF) with 1000 samples, square root ensemble Kalman filter (EnKFS) with 1000 samples and the full Bayesian MCMC update with 10^5 samples. The last procedure is declared as the reference solution. Its convergence with respect to the number of samples can be seen in Fig. 21, where the relative errors of the mean and variance

$$\varepsilon_m = \frac{\|\bar{\kappa}_a^N - \bar{\kappa}_a^R\|}{\|\bar{\kappa}_a^R\|}, \quad \varepsilon_v = \frac{\|\text{var } \kappa_a^N - \text{var } \kappa_a^R\|}{\|\text{var } \kappa_a^R\|}, \quad (41)$$

are plotted, respectively. Here, $\bar{\kappa}_a^N, \text{var } \kappa_a^N$ stand for the mean and variance of the posterior distribution obtained with N samples, whereas $\bar{\kappa}_a^R, \text{var } \kappa_a^R$ denote the mean and variance of the posterior distribution as a result of 10^5 runs. According to these plots the relative errors are slowly converging with the number of samples, as expected. This results in an accuracy of ca. 1e-12 for the mean conductivity and 1e-8 for the conductivity variance.

A comparison of identification results in Fig. 22 reveals that the MCMC procedure is the only one able to identify the truth in all three assumed cases. The methods based

32
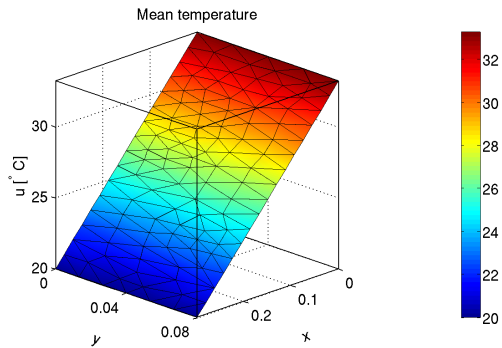
Mean temperature

a) the mean value of temperature

Temperature variance
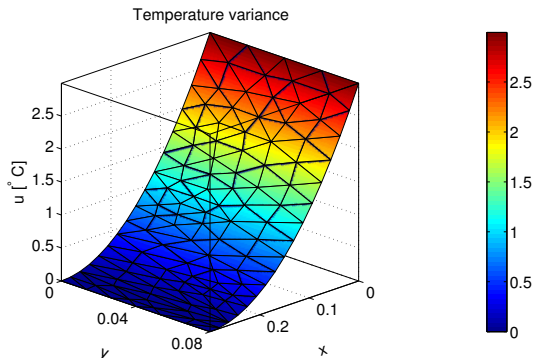
b) the temperature variance

Figure 20: The second order statistics obtained with the help of the stochastic Galerkin method

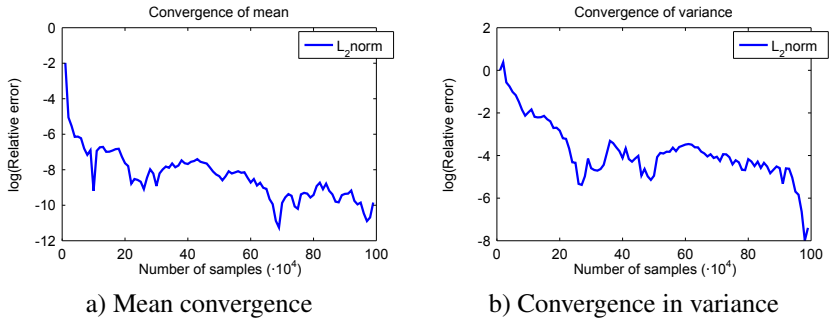a) Mean convergence   b) Convergence in variance

Figure 21: The Markov chain Monte Carlo convergence of posterior κ_a for the Gaussian measurement error described by a standard deviation 0.3

on the linear approximation behave well in cases when the truth lies in the high probability region, otherwise an "overshooting" occurs. This lends weight to the previously given argument that the error of the linear approximant strongly depends on the nonlinearity of the measurement operator, as well as on the prior assumption.

The previous results are also supported by the plots of posterior 99% confidence intervals in Fig. 23. Initially, before the measurements are carried out, the 99% confidence interval is assumed to be broad in order to "catch" the truth. With every new successive measurement the probability region narrows down such that the interval becomes almost deterministic after seven performed measurements. Even though the truth is assumed to be deterministic, the posterior 99% confidence interval does not disappear due to the measurement and model errors, as well as the error due to the nonlinearity of the measurement operator. Additionally, the results, as seen in Fig. 23, indicate that the full linear Bayesian update (FLBU), as well as the EnKF almost match the MCMC results in the first scenario, whereas the one random variable linear Bayesian update (RLBU) underestimates the posterior variance. On the other hand, the square root filter and EnKF square root filter deliver similar results with slightly shifted median.

The issues previously described can be resolved at the expense of improving the prior description. This can be done by moving the mean of the prior distribution towards the truth. Similar to the Kalman way of updating, we may obtain more information about the prior mean using the existing measurement data. Once this has been done, one may alternate the old prior with the newly obtained mean value
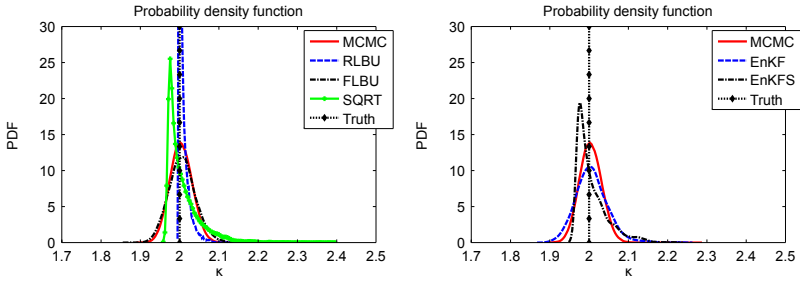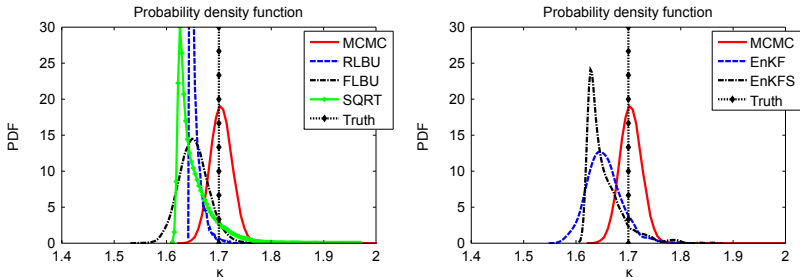
34

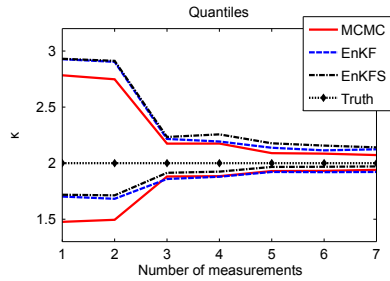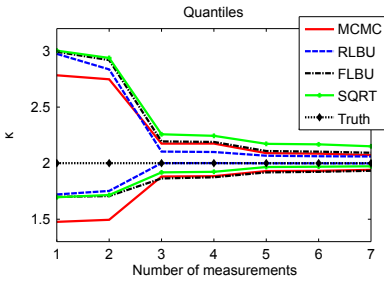a) The truth lies in the one sigma region of the prior

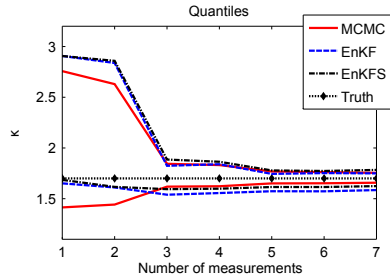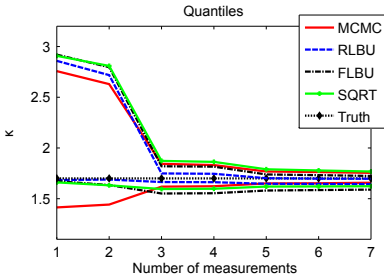b) The truth lies in the two sigma region of the prior
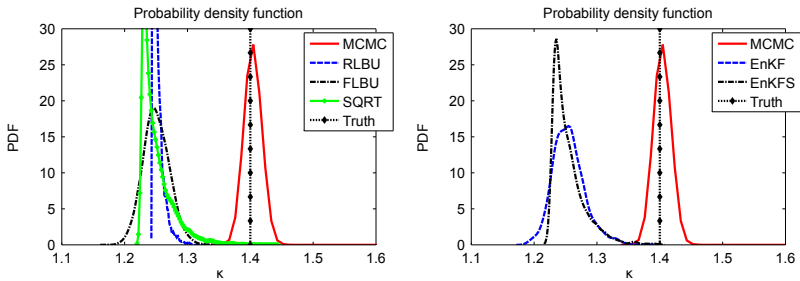
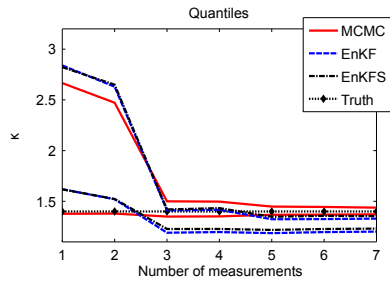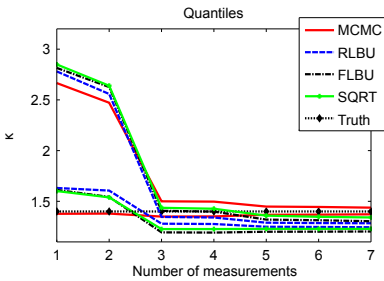c) The truth lies in the three sigma region of the prior

Figure 22: Comparison of posterior probability density functions describing κ

35

a) The truth lies in the one sigma region of the prior

b) The truth lies in the two sigma region of the prior

c) The truth lies in the three sigma region of the prior

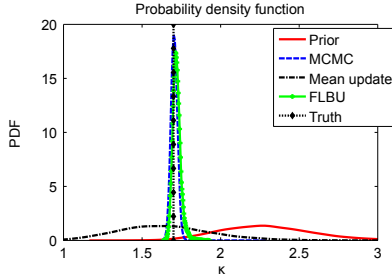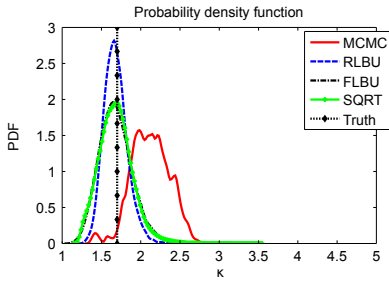Figure 23: Comparison of posterior probability density functions describing κ for different number of measurement points
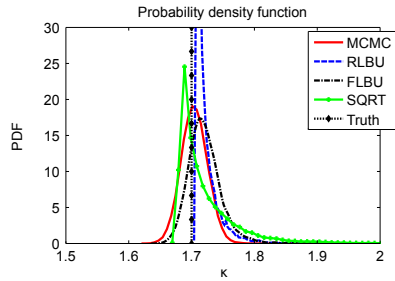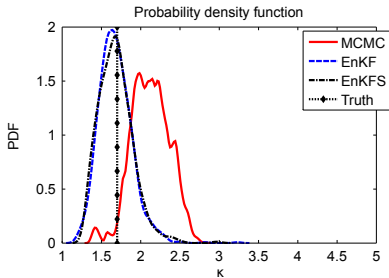
Figure 24: The process of updating by first moving the prior mean

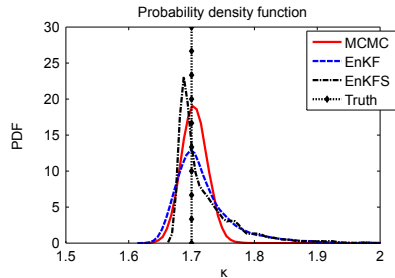

a) Two measurements

b) Seven measurements

c) Two measurements

d) Seven measurements

Figure 25: Comparison of probability density functions of the posterior of κ for different numbers of measurement points after moving the prior

37
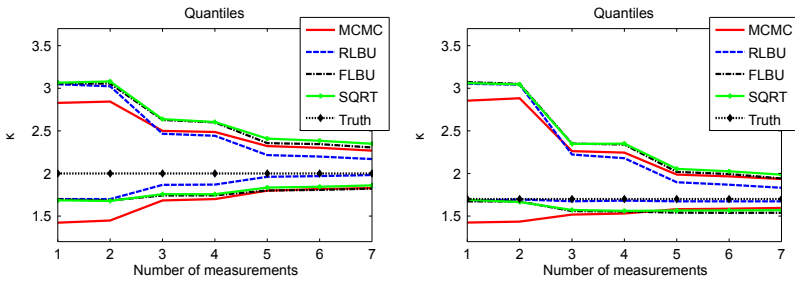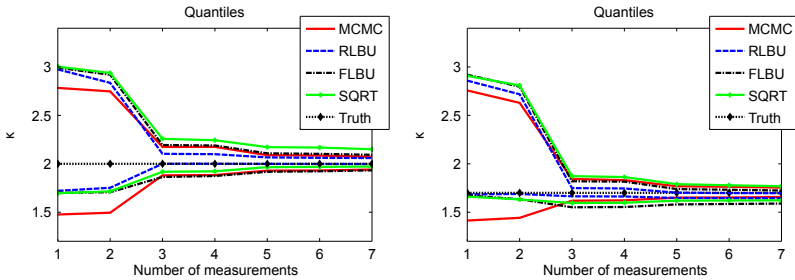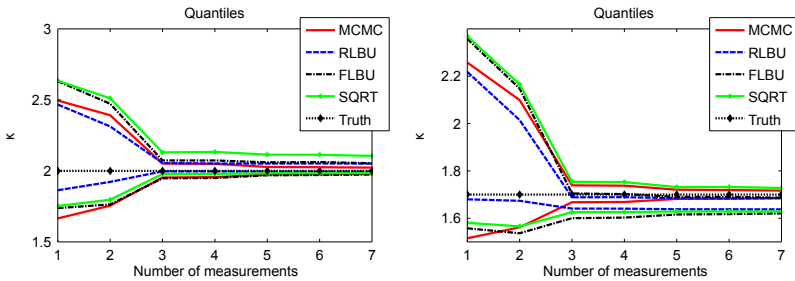
Figure 26: The influence of the measurement error on the posterior quantiles: left is the truth in C_1 scenario and right in C_2 scenario

38
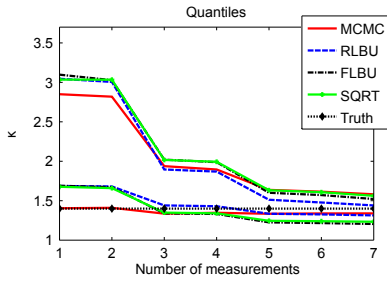
a) $\sigma_\varepsilon = 1$
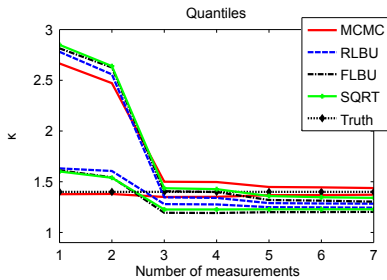
b) $\sigma_\varepsilon = 0.3$
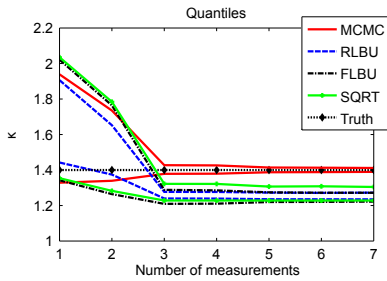
c) $\sigma_\varepsilon = 0.1$

Figure 27: The influence of the measurement error on the posterior quantiles for case scenario C_3

and continue the estimation as described previously. Even though the method just described represents an oversimplification (especially in case of nonlinear functions), it does work for slightly nonlinear functions (measurement operators). Such a case is depicted in Fig. 24, where the process of identifying the truth in 2σ region is illustrated. Here, the red line denotes the prior distribution, while the dashed black line is the newly adopted prior (same statistics besides mean). From the resulting plots of the posterior one may see that this time all procedures return similar results. This point emphasizes the importance of the prior assumption. Also, one may note in Fig. 25 that linear Bayesian updates are better than the MCMC update when the number of measurement points is small.

As mentioned earlier, the estimation of the conductivity coefficient greatly depends on the number of the measurement points, as well as on the measurement (model) errors. The analysis and simulation of the updating procedures for different levels of the measurement errors are shown in Fig. 26. From these figures, it is apparent that the posterior estimate better complies the truth for smaller values of error. The same is valid for second two case scenarios, see Fig. 27. However, these plots also reveal that the bigger measurement error can regularise the estimation process such that the 99% confidence interval of posterior includes the truth value.

6 Conclusions

This contribution aimed to present and compare different numerical approaches to Bayesian estimations of non-observable model parameters from noisy measurement data. The model parameter stands for the thermal conductivity and is represented by a random variable with a non-Gaussian prior distribution. The numerical findings suggest that Markov chain Monte Carlo sampling of the posterior distribution is a reliable way of computing the Bayesian update. However, the model simulation – often time-consuming – has to be evaluated for each sample in the chain, which makes the whole procedure computationally expensive. To overcome this problem, the stochastic Galerkin method is employed in order to construct a polynomial chaos based approximation of the model response. This is then used within the sampling procedure, instead of full model simulations. From these numerical findings it is evident that the proxy model may jeopardize the accuracy of the Markov chain Monte Carlo procedure due to a combination of both approximation and sampling errors. For reasons of efficiency the updating procedure is recast into an alternative linear Bayesian form thereby enabling a direct algebraic way of computing the posterior distribution. While the sampling version of the linear filter—also known as ensemble

40

Kalman filter—needs a considerably smaller number of samples than the MCMC procedure, this approach is severely underestimating or overestimating the residual uncertainty. On the other hand, the polynomial chaos linear Bayesian methods do not seem to suffer from any of previously mentioned issues. They deliver more reliable results than EnKF procedures and are better than proxy MCMC in case of linear measurements. However, LBU may suffer from larger residual errors when applied in nonlinear cases.

While the initial findings are promising, further research is necessary. Therefore, future analysis will be needed to validate the mentioned numerical behaviour of the presented computational procedures for the more complex diffusion problem when the conductivity parameter is modelled in a form of random field. In addition, the adaption of the polynomial chaos based linear Bayes filter has to be made in order to handle the assimilation from the noisy nonlinear measurements.

7 Acknowledgement

References

[1] A. Andrews. A square root formulation of the Kalman covariance equations. *AIAA Journal*, 22(6):1165–1166, June 1968.

[2] E. D. Blanchard. *Polynomial chaos approaches to parameter estimation and control design for mechanical systems with uncertain parameters*. PhD thesis, Department of Mechanical Engineering, VirginiaTech University, 2010.

[3] A. Bobrowski. *Functional analysis for probability and stochastic processes: an introduction*. Cambridge University Press, 2005.

[4] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[5] J. A. Christen and C. Fox. MCMC using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795 – 810, 2005.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Dordrecht: Kluwer, 2000.

[7] G. Evensen. *Data Assimilation, The Ensemble Kalman Filter*. Springer, Berlin, June 2007.

[8] Geir Evensen. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29:82–104, 2009.

[9] M. Friswell and J. E. Mottershead. *Finite element model updating in structural dynamics*. Kluwer Academic Publishers, Dordrecht, Neitherlands, 1995.

[10] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: stochastic simulation for Bayesian Inference*. Chapman and Hall/CRC, 2006.

[11] R. Ghanem and P. D. Spanos. *Stochastic finite elements—A spectral approach*. Springer-Verlag, New York, 1991.

[12] R. E. Kálmán. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering (Ser D)*, 82:35–45, 1960.

[13] A. Kučerová. *Identification of nonlinear mechanical model parameters based on softcomputing methods*. PhD thesis, Ecole Normale Supérieure de Cachan, Laboratoire de Mécanique et Technologie, 2007.

[14] A. Kučerová, J. Sýkora, B. Rosić, and H. G. Matthies. Acceleration of uncertainty updating in the description of transport processes in heterogeneous materials. *Journal of Computational and Applied Mathematics*, 236(18):4862–4872, 2012.

[15] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, June 2007.

[16] Y. M. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physics*, 6(4):826–847, 2009.

[17] H. G. Matthies. Stochastic finite elements: computational approaches to stochastic partial differential equations. *Zeitschrift für Angewandte Mathematik und Mechanik (ZAMM)*, 88(11):849–873, 2008.

42

[18] O. Pajonk, B. Rosić, A. Litvinenko, and H. G. Matthies. A deterministic filter for non-Gaussian Bayesian estimation. *Physica D: Nonlinear Phenomena*, 241(7):775–788, 2012.

[19] O. Pajonk, B. Rosić, and H. G. Matthies. Sampling-free linear Bayesian updating of model state and parameters using a square root approach. *Computers and Geosciences*, 55:70–83, 2012.

[20] B.L. Pence, H.K. Fathy, and J.L. Stein. A maximum likelihood approach to recursive polynomial chaos parameter estimation. In *Proceedings of American Control Conference (ACC)*, pages 2144–2151, 2010.

[21] B. Rosić, A. Kučerová, J. Sýkora, O. Pajonk, A. Litvinenko, and H. G. Matthies. Parameter identification in a probabilistic setting. *Engineering Structures*, 50:179–196, 2013.

[22] B. Rosić, O. Pajonk, A. Litvinenko, and H. G. Matthies. Sampling-free linear Bayesian update of polynomial chaos represenations. *Journal of Computational Physics*, 231(17):5761–5787, 2012.

[23] G. Saad and R. Ghanem. Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resources Research*, 45(W04417):–, 2009.

[24] A. Schöniger, W. Nowak, and H.-J. Hendricks Franssen. Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, 48(4):n/a–n/a, 2012.

[25] E. Simon and L. Bertino. Application of the Gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science*, 5(4):495–510, 2009.

[26] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1):3–23, 1993.

[27] C. Soize. Construction of probability distributions in high dimensions using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices. *International Journal for Numerical Methods in Engineering*, 76(10):1583–1611, 2008.

[28] L. Stone, R. L. Streit, and T. L. Corwin. *Bayesian multiple target tracking*. Artech House Inc, 2014.

[29] H. A. Tchelepi, H. Bazargan, and M. A. Christie. Efficient Markov chain Monte Carlo sampling using polynomial chaos expansion. In *Proceedings of the SPE Reservoir Simulation Symposium*, The Woodlands, Texas, United States, 2013. online.

[30] L. Tiernay. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994.

[31] A. N. Tikhonov. Regularisation of incorrectly posed problems. *Soviet Mathematics Doklady*, 4(4):1624–1627, 1963.

[32] R. L. Tweedie. Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and their Applications*, 3(4):385 – 403, 1975.

[33] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

[34] A. Wirgin. The inverse crime. Technical report, arXiv:math-ph(0401050), 2004.

[35] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24:619–644, 2002.

Technische Universität Braunschweig
Informatik-Berichte ab Nr. 2012-02

2012-03	J. Rang	An analysis of the Prothero–Robinson example for constructing new DIRK and ROW methods
2012-04	S. Kolatzki, M. Hagner, U. Goltz and A. Rausch	A Formal Definition for the Description of Distributed Concurrent Components - Extended Version
2012-05	M. Espig, W. Hackbusch, A. Litvinenko, H. G. Matthies and P. Wähnert	Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats
2012-06	S. Mennike	A Petri Net Semantics for the Join-Calculus
2012-07	S. Lity, R. Lachmann, M. Lochau, I. Schaefer	Delta-oriented Software Product Line Test Models - The Body Comfort System Case Study
2013-01	M. Lochau, S. Mennicke, J. Schroeter und T. Winkelmann	Extended Version of 'Automated Verification of Feature Model Configuration Processes based on Workflow Petri Nets'
2013-02	S. Lity, M. Lochau, U. Goltz	A Formal Operational Semantics of Sequential Function Tables for Model-based SPL Conformance Testing
2013-03	L. Giraldi, A. Litvinenko, D. Liu, H. G. Matthies, A. Nouy	To be or not to be intrusive? The solution of parametric and stochastic equations – the "plain vanilla" Galerkin case
2013-04	A. Litvinenko, H. G. Matthies	Inverse problems and uncertainty quantification
2013-05	J. Rang	Improved traditional Rosenbrock–Wanner methods for stiff ODEs and DAEs

2013-06	J. Koslowski	Deterministic single-state 2PDAs are Turing-complete
2014-01	B. Rosić, J. Diekmann	Stochastic Description of Aircraft Simulation Models and Numerical Approaches
2014-02	M. Krosche, W. Heinze	A Robustness Analysis of a Preliminary Design of a CESTOL Aircraft
2014-03	J. Rang	Apdative timestep control for fully implicit Runge–Kutta methods of higher order
2014-04	S. Mennicke, J.-W. Schicke-Uffmann, U. Goltz	Free-Choice Petri Nets and Step Branching Time
2014-05	A. Martens, C. Borchert, T. O. Geissler, O. Spinzcyk, D. Lohmann, R. Kapitza	Exploiting determinism for efficient protection against arbitrary state corruptions
2014-06	J. Rang	An analysis of the Prothero–Robinson example for constructing new adaptive ESDIRK methods of order 3 and 4
2014-07	J. Rang, R. Niekamp	A component framework for the parallel solution of the incompressible Navier- Stokes equations with Radau-IIA methods
2014-08	J. Rang	The Prothero and Robinson example: Convergence studies for Runge–Kutta and Rosenbrock–Wanner methods
2014-09	J. Wayetens, B. V. Rosić	Comparison of deterministic and probabilistic approaches to identify the dynamic moving load and damages of a reinforced concrete beam
2014-10	B. V. Rosić, J. Sykora, O. Pajonk, A. Kucerova, H. G. Matthies	Comparison of Numerical Approaches to Bayesian Updating

46