Bayesian Inverse Problems and Kalman Filters

Oliver G. Ernst, Björn Sprungk, Hans-Jörg Starkloff

Abstract We provide a brief introduction to Bayesian inverse problems and Bayesian estimators emphasizing their similarities and differences to the classical regularized least-squares approach to inverse problems. We then analyze Kalman filtering techniques for nonlinear systems, specifically the well-known Ensemble Kalman Filter (EnKF) and the recently proposed Polynomial Chaos Expansion Kalman Filter (PCE-KF), in this Bayesian framework and show how they relate to the solution of Bayesian inverse problems.

1 Introduction

In recent years the interest and research activity in uncertainty quantification (UQ) for complex systems modelled by partial differential equations (PDEs) has increased significantly. This is due both to growing available computing resources as well as new efficient numerical methods for high-dimensional problems, which together make the solution of UQ problems associated with PDEs feasible. The motivation driving UQ is the simple fact that, in practical applications, we usually do not know parameters, coefficients or even boundary conditions for the PDE model under consideration exactly. A typical example are material properties such as conductivity. At the same time, we may still have some knowledge about possible values for these

Oliver G. Ernst
Department of Mathematics, TU Chemnitz, Reichenhainer Str. 41, 09126 Chemnitz, Germany, e-mail: oliver.ernst@mathematik.tu-chemnitz.de

Björn Sprungk
Department of Mathematics, TU Chemnitz, Reichenhainer Str. 41, 09126 Chemnitz, Germany, e-mail: bjoern.sprungk@mathematik.tu-chemnitz.de

Hans-Jörg Starkloff
Fachgruppe Mathematik, University of Applied Sciences Zwickau, 08012 Zwickau, Germany, e-mail: hans.joerg.starkloff@fh-zwickau.de

uncertain input data, e.g., the hydraulic conductivity of layered clay may be between 10^{-6} and 10^{-4} cm/s. A careful simulation would take into account the uncertainty in the input data and quantify the resulting uncertainty in the output of the physical or PDE model. Although there are also other mathematical techniques for modeling uncertainty such as fuzzy set theory or interval arithmetic, we focus here on the probabilistic approach.

Initially, the main interest has been in solving the *forward problem*, in which one is given the probability law of uncertain data $u \sim \mu$ with the goal of computing the corresponding law of a quantity of interest $\phi = F(u)$, where F represents the composition of solving a PDE and evaluating a functional of its solution. Current numerical methods for this task include, e.g., multilevel Monte Carlo, stochastic Galerkin and stochastic collocation methods, proper orthogonal decomposition, and Gaussian process emulators.

Within UQ, the more fundamental task is to develop a good probability law for the unknown quantity u reflecting our (possibly subjective) knowledge of u, since this determines the outcome. In general, transforming expert knowledge and physical reasoning into a probability distribution is a subtle and quite difficult task. Moreover, incorporating any available information about the unknown into the probability law is desirable, since this will, in general, reduce our uncertainty and lead to improved models. For this reason the *inverse problem* has received increased attention in the UQ community.

Specifically, given noisy data $z = G(u) + \varepsilon$, the task is to either identify u or make inferences, i.e., refine an initial model of u. Here we want to distinguish between identification, i.e., determining a value u which best explains the data, and inference, i.e., updating our understanding or belief about u based on the new information z.

The latter is more interesting for UQ purposes, since adjusting prior probability models of the unknown according to indirect data yields an improved uncertainty model for u, whereas identification would merely provide a certain best estimate with no indication of how well this estimate is determined.

In the probabilistic setting, incorporating new information into a given prior model (i.e., a prior random variable or probability measure), is performed by *conditioning* this model on the available information, resulting in a conditional measure. The procedure of conditioning, and thus also the conditional measure or distribution, are rooted in Kolmogorov's fundamental concept of *conditional expectation*. In particular, *Bayes' rule* provides an analytic expression for the conditional measure in terms of the prior measure and provides the main tool in Bayesian inference as well as Bayesian inverse problems.

Since Bayesian inverse problems have gained much attention in the scientific computing community in the last few years, numerous algorithms and numerical methods have been proposed for their solution. We provide a short overview of existing methods and focus on the *Kalman Filter* and two of its variants, namely the *Ensemble Kalman Filter* [16] and the *Polynomial Chaos Expansion Kalman Filter* [35], which have recently proposed for UQ in association with inverse problems. In particular, we investigate what these Kalman filtering methods are actually computing and how they relate to Bayesian inverse problems and Bayes estimators. Thus,

our main purpose is to clarify which quantities Kalman filters can and cannot approximate.

The remainder of this paper is organized as follows: Section 2 briefly recalls the deterministic and Bayesian approaches to inverse problems and provides a short overview of computational methods. In Section 3 we consider Kalman filtering methods and analyze these in the light of Bayes estimators. In particular, we show that these filtering methods approximate a random variable which is, in general, not distributed according to the desired posterior measure. Moreover, we illustrate the performance of Kalman filters and the difference between their output and the solution of the Bayesian inverse problem for a simple 1D boundary value problem in Section 4. A summary and conclusions are given in Section 5.

2 Bayesian Approach to Inverse Problems

In this section we introduce the setting and notation for the inverse problem and recall the basic concepts of the classical regularized least-squares and the Bayesian approaches.

Throughout the article, $|\cdot|$ shall denote the Euclidean norm on $\mathbb{R}^k$, $\|\cdot\|$ the norm on a general separable Banach space $(\mathscr{X}, \|\cdot\|)$, $\mathscr{X}^*$ the topological dual of $\mathscr{X}$ and $\mathscr{Y}$ a second separable Banach space.

We consider the abstract inverse problem of identifying an unknown $u \in \mathscr{X}$ given finite-dimensional but noisy observations $z \in \mathbb{R}^k$ according to the model

$$z = G(u) + \varepsilon \tag{1}$$

containing an observation operator $G : \mathscr{X} \to \mathbb{R}^k$ and measurement noise $\varepsilon \in \mathbb{R}^k$.

Example 1 (Elliptic PDE). Consider the problem of determining the logarithm $\kappa \in C(D)$ of the conductivity $\exp(\kappa)$ of an incompletely known porous medium occupying a bounded domain $D \subset \mathbb{R}^d$ given observations of the pressure head p at several locations in the domain of a fluid in stationary flow through the medium. The relation between κ and p can be modelled by, e.g.,

$$-\nabla \cdot (\mathrm{e}^{\kappa} \nabla p) = f \text{ on } D, \qquad p|_{\partial D} = 0. \tag{2}$$

Here the unknown is $u = \kappa$ and the observation operator G is the mapping $\kappa \mapsto (p(x_1), \ldots, p(x_k))$ for given measurement locations $x_i \in D$, $i = 1, \ldots, k$.

Example 2 (Discrete dynamics). Consider a discrete-time dynamical system $\{y_n\}_{n \in \mathbb{N}_0}$ with state evolution equation

$$y_{n+1} = h_n(y_n), \qquad y_0 = x \in \mathbb{R}^N,$$

where $h_n : \mathbb{R}^N \to \mathbb{R}^N$ governs the (deterministic) dynamics driving the system at step n. Suppose we observe J noisy states

$$z_{n_j} = y_{n_j} + \varepsilon_j, \qquad j = 1,\dots,J, \quad 0 < n_1 < \cdots < n_J,$$

and wish to infer from these the unknown initial state $u = x$. Setting $G_j = h_0 \circ \cdots \circ h_{n_j-1}$ and $G := (G_1,\dots,G_J)$, we arrive at a problem of the form (1). By extending the unknown u to the vector $(y_{n_0}, y_{n_1},\dots,y_{n_J})$ one may also infer J additional states.

Remark 1. Identification problems for dynamical systems with sequentially arriving data call for special, efficient sequential methods for solving (1). These are methods for computing the solution for $z = (z_{n_1},\dots,z_{n_J})^\top$ based only on the solution for $(z_{n_1},\dots,z_{n_{J-1}})^\top$ and the current observation z_{n_J}. For brevity, we omit considerations of sequentiality in this work.

2.1 Deterministic Identification for Inverse Problems

Solving (1) by determining $u = G^{-1}(z)$ is usually not an option since $\varepsilon \neq 0$ generally results in $z \notin G(\mathscr{X})$. Moreover, the more general least-squares formulation $u = \operatorname{argmin}_{v\in\mathscr{X}} |z - G(v)|^2$ is typically ill-posed, as u may depend discontinuously on z and is often heavily underdetermined. Making (1) mathematically tractable is usually achieved by some form of *regularization*, which, generally speaking, involves the incorporation of additional *prior* information on u and ε. A comprehensive introduction to the regularized least-squares approach to inverse problems is given in [10]. We briefly summarize this approach for nonlinear G here.

The conceptual starting point for the deterministic approach is the noise-free model $z^\dagger = G(u)$, i.e., $z = z^\dagger + \varepsilon$. Since we want to identify the element $u \in \mathscr{X}$ which led to the observations z, it is reasonable to assume that the "true", unpolluted data $z^\dagger$ lies in the range of G. Thus we assume the existence of $u^\dagger \in \mathscr{X}$ such that $G(u^\dagger) = z^\dagger$. This is sometimes called the *attainability assumption* [11]. Next, we introduce a penalty or regularizing functional $\mathrm{R} : \mathscr{X} \to [0,\infty]$ and define an R-minimizing solution to $z^\dagger = G(u)$ to be any element $u^* \in \mathscr{X}$ which satisfies

$$\mathrm{R}(u^*) = \min\left\{\mathrm{R}(u) : u \in \mathscr{X},\ G(u) = z^\dagger\right\}. \tag{3}$$

Note that u^* need not be unique. Furthermore, the choice of R is significant and reflects prior assumptions about u. Often R is taken to be convex. A common choice for R is, e.g., $\mathrm{R}(u) = \|u - u^{\mathrm{ref}}\|^2$, where $u^{\mathrm{ref}} \in \mathscr{X}$ is a given reference state known to lie in the vicinity of the solution. For a broader discussion of different penalty functionals we refer to [36].

However, since only polluted data $z = z^\dagger + \varepsilon$ is available, we can only ask for an approximation of u^* which should improve with diminishing noise ε. This approximation is the regularized solution $\hat{u}_\alpha$ given by

$$\hat{u}_\alpha = \operatorname*{argmin}_{u\in\mathscr{X}} |z - G(u)|^2 + \alpha \mathrm{R}(u), \tag{4}$$

where $\alpha \in [0,\infty)$ serves as a regularization parameter to be chosen wisely. If further smoothness assumptions on u^* and G are satisfied and if α is chosen as a suitable function $\alpha = \alpha(\delta)$ of the noise level $|\varepsilon| \leq \delta$, then convergence rate bounds such as

$$\|\hat{u}_{\alpha(\delta)} - u^*\| = O(\sqrt{\delta}) \qquad \text{as } \delta \to 0$$

can be obtained [11]. These rates are typically based on explicit error estimates such as $\|\hat{u}_{\alpha(\delta)} - u^*\| \leq C(\alpha)\sqrt{\delta}$ for the above result. For further analysis of the smoothness requirements on u^* and related convergence rates see, e.g., [23] and, for appropriate choices $\alpha = \alpha(\delta)$, see, e.g., [1] and the references therein.

2.2 The Bayesian Inverse Problem

Recall that, in order to regularize the usually ill-posed least-squares formulation of the inverse problem (1), we incorporated additional prior information about the desired u into the (deterministic) identification problem by way of the regularization functional R. A further possibility for regularization is to restrict u to a subset or subspace $\tilde{\mathscr{X}} \subset \mathscr{X}$, e.g., by using a stronger norm of $u - u^{\text{ref}}$ as the regularization functional. Speaking very broadly, the Bayesian approach stems from yet another way of modelling prior information on u and adding it to the inverse problem. In this case we express our prior belief about u through a probability distribution μ_0 on the Banach space $\mathscr{X}$, by which a quantitative preference of some solutions u over others may be given by assigning higher and lower probabilities. However, the goal in the Bayesian approach is not the identification of a particular $u \in \mathscr{X}$, but rather inference on u, i.e., we would like to *learn from the data* in a statistical or probabilistic fashion by *adjusting our prior belief* μ_0 about u in accordance with the newly available data z. The task of identification may also be achieved within the Bayesian framework through *Bayes estimates* and *Bayes estimators*, which are discussed in Section 2.3.

The Bayesian approach to the inverse problem (1) thus differs conceptually from the regularized least-squares approach as summarized above in that its objective is inference rather than identification. As stated in [24], the Bayesian approach[1] is based on the following four principles:

1. All quantities occurring in (1) are modelled as random variables.
2. The randomness describes our degree of information concerning their realizations.
3. This degree of information concerning these values is encoded in probability distributions.
4. The solution of the inverse problem is the posterior probability distribution.

In the Bayesian setting we therefore replace our model (1) in the following with

[1] This is referred to in [24] as the *statistical inversion approach*.

$$Z = G(U) + \varepsilon, \tag{5}$$

where ε and hence Z are both random variables on $\mathbb{R}^k$ while U is a random variable on $\mathscr{X}$ whose posterior probability distribution given the available observations $Z = z$ is to be determined. Before giving a precise definition of the posterior distribution we require some basic concepts from probability theory.

2.2.1 Probability Measures and Random Variables

Let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a probability space. We denote by $\mathscr{B}(\mathscr{X})$ the Borel σ-algebra of $\mathscr{X}$ generated by the open sets in $\mathscr{X}$ w.r.t. $\|\cdot\|$. A measurable mapping $X : (\Omega, \mathscr{F}) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$ is called a random variable (RV) and the measure $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$, i.e., $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$ for all $A \in \mathscr{B}(\mathscr{X})$, defines the distribution of X as the push-forward measure of $\mathbb{P}$ under X. Conversely, given a probability measure μ on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, then $X \sim \mu$ means $\mathbb{P}_X = \mu$. By $\sigma(X) \subset \mathscr{F}$ we denote the σ-algebra generated by X, i.e., $\sigma(X) = \{X^{-1}(A) : A \in \mathscr{B}(\mathscr{X})\}$.

The Bochner space of p-integrable $\mathscr{X}$-valued RVs, i.e., the space of RVs $X : \Omega \to \mathscr{X}$ such that $\int_\Omega \|X(\omega)\|^p \, \mathbb{P}(\mathrm{d}\omega) < \infty$, is denoted by $L^p(\Omega, \mathscr{F}, \mathbb{P}; \mathscr{X})$ or simply $L^p(\mathscr{X})$ when the context is clear.

An element $m \in \mathscr{X}$ is called the *mean* of a RV X if for any $f \in \mathscr{X}^*$ there holds $f(m) = \mathbb{E}[f(X)]$. Here and in the following $\mathbb{E}$ denotes the expectation operator w.r.t. $\mathbb{P}$. If $X \in L^1(\Omega, \mathscr{F}, \mathbb{P}; \mathscr{X})$ then its mean is given by $m = \mathbb{E}[X] = \int_\Omega X(\omega) \, \mathbb{P}(\mathrm{d}\omega)$. An operator $C : \mathscr{Y}^* \to \mathscr{X}$ is called the *covariance* of two RVs $X : \Omega \to \mathscr{X}$ and $Y : \Omega \to \mathscr{Y}$ if it satisfies $f(Cg) = \mathbb{E}\big[f(X - \mathbb{E}[X]) \, g(Y - \mathbb{E}[Y])\big]$ for all $f \in \mathscr{X}^*$ and $g \in \mathscr{Y}^*$. We denote the covariance of X and Y by $\mathrm{Cov}(X, Y)$ and, if $X = Y$, simply by $\mathrm{Cov}(X)$.

Besides normed vector spaces of RVs we will also work with metric spaces of probability measures. One notion of distance between measures is the *Hellinger metric* d_H: given two probability measures μ_1 and μ_2 on the Banach space $\mathscr{X}$, it is defined as

$$d_H(\mu_1, \mu_2) := \left[\int_{\mathscr{X}} \left(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}(u)} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}(u)} \right)^2 \nu(\mathrm{d}u) \right]^{1/2},$$

where ν is a dominating measure of μ_1 and μ_2, e.g., $\nu = (\mu_1 + \mu_2)/2$. Note that the definition of the Hellinger metric is independent of the dominating measure. For relations of the Hellinger metric to other probability metrics such as total variation distance or the Wasserstein metric, we refer to [17].

In the following, we will use upper case latin letters such as X, Y, Z, U to denote RVs on Banach spaces and lower case latin letters like x, y, z, u for elements in these Banach spaces or realizations of the associated RVs, respectively. Greek letters such as ε will be used to denote RVs on $\mathbb{R}^k$ as well as their realizations.

2.2.2 Conditioning

Bayesian inference consists in updating the probability distribution encoding our prior knowledge on the unknown U to a new probability distribution reflecting a gain in knowledge due to new observations. There are certain subleties associated with the probabilistic formulation of this transition from prior to posterior measure, and we take some care in this section to point these out.

The distribution of the RV U, characterized by the probabilities $\mathbb{P}(U \in B)$ for $B \in \mathscr{B}(\mathscr{X})$, quantifies in stochastic terms our knowledge about the uncertainty associated with U. When new information becomes available, such as knowing that the event $Z = z$ has occurred, this is reflected in our quantitative description as the "conditional distribution of U given $\{Z = z\}$", denoted $\mathbb{P}(U \in B|Z = z)$. Unfortunately, $\mathbb{P}(U \in B|Z = z)$ cannot be defined in an elementary fashion when $\mathbb{P}(Z = z) = 0$, in which case the conditional distribution is defined by an integral relation. The key concept here is that of *conditional expectation*.

Given RVs $X \in L^1(\Omega, \mathscr{F}, \mathbb{P}; \mathscr{X})$ and $Y : \Omega \to \mathscr{Y}$, we define the conditional expectation $\mathbb{E}[X|Y]$ of X given Y as any mapping $\mathbb{E}[X|Y] : \Omega \to \mathscr{X}$ with the following two properties:

1. $\mathbb{E}[X|Y]$ is $\sigma(Y)$-measurable.
2. For any $A \in \sigma(Y)$ there holds

$$\int_A \mathbb{E}[X|Y]\, \mathbb{P}(\mathrm{d}\omega) = \int_A X\, \mathbb{P}(\mathrm{d}\omega).$$

Note that, since it is defined by an integral relation, the RV $\mathbb{E}[X|Y]$ is determined only up to sets of $\mathbb{P}$-measure zero and is thus understood as an equivalence class of such mappings. By the Doob-Dynkin Lemma (cf. [25, Lemma 1.13]) there exists a measurable function $\phi : \mathscr{Y} \to \mathscr{X}$ such that $\mathbb{E}[X|Y] = \phi(Y)$ $\mathbb{P}$-almost surely. Again, we note that this does not determine a unique function ϕ but an equivalence class of measurable functions, where $\phi_1 \sim \phi_2$ iff $\mathbb{P}(Y \in \{y \in \mathscr{Y} : \phi_1(y) \neq \phi_2(y)\}) = 0$. For a specific realization y of Y (and a specific ϕ), we also denote the function value by

$$\mathbb{E}[X|Y = y] := \phi(y) \in \mathscr{X}.$$

Setting $X = \mathbf{1}_{\{U \in B\}}$, one can, for each fixed $B \in \mathscr{B}(\mathscr{X})$, define

$$\mathbb{E}[\mathbf{1}_{\{U \in B\}}|Z = z] =: \mathbb{P}(U \in B|Z = z) \tag{6}$$

as an equivalence class of measurable functions $\mathbb{R}^k \to [0,1]$. One would like to view this, conversely, as a family of probability measures with the realization z as a parameter, giving the posterior distribution of U resulting from having made the observation $Z = z$. Unfortunately, this construction need not, in general, yield a probability measure for each fixed value of z (cf. [33]). In case $\mathscr{X}$ is a separable Banach space, a function

$$Q : \mathscr{B}(\mathscr{X}) \times \mathbb{R}^k \to \mathbb{R}$$

can be shown to exist (cf. e.g., [33]) such that

(a). For each $z \in \mathbb{R}^k$, $Q(\cdot, z)$ is a probability measure on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$,
(b). for each $B \in \mathscr{B}(\mathscr{X})$ the function

$$\mathbb{R}^k \ni z \mapsto Q(B,z)$$

is a representative of the equivalence class (6), i.e., it is measurable and there holds

$$\mathbb{P}(U \in B, Z \in A) = \int_A Q(B,z)\, \mathbb{P}_Z(\mathrm{d}z) \qquad \forall A \in \mathscr{B}(\mathbb{R}^k).$$

Such a function Q, also denoted by $\mu_{U|Z}$, is called the *regular conditional distribution* of U given Z and is defined uniquely up to sets of z-values of $\mathbb{P}_Z$-measure zero. We have thus arrived at a consistent definition of the posterior probability $\mathbb{P}(U \in B | Z = z)$ as $\mu_{U|Z}(B,z)$.

It is helpful to maintain a clear distinction between *conditional* and *posterior* quantities: the former contain the – as yet unrealized – observation as a parameter, while in the latter the observation has been made. Specifically, $\mu_{U|Z}$ is the conditional measure of U conditioned on Z, whereas $\mu_{U|Z}(\cdot, z)$ denotes the posterior measure of U for the observation $Z = z$.

2.2.3 Bayes' Rule and the Posterior Measure

We make the following assumptions for the model (5).

Assumption 1

1. *$U \sim \mu_0$, $\varepsilon \sim \mu_\varepsilon$ and $(U, \varepsilon) \sim \mu_0 \otimes \mu_\varepsilon$, i.e., U and ε are independent.*
2. *$\mu_\varepsilon = \rho(\varepsilon)\,\mathrm{d}\varepsilon$ where $\rho(\varepsilon) = C\mathrm{e}^{-\ell(\varepsilon)}$ with $C > 0$ and $\ell : \mathbb{R}^k \to \mathbb{R}_0^+$ measurable and nonnegative. Here $\mathrm{d}\varepsilon$ denotes Lebesgue measure on $\mathbb{R}^k$.*
3. *$G : \mathscr{X} \to \mathbb{R}^k$ is continuous.*

Throughout we assume $\mu_0(\mathscr{X}) = 1$ and $\mu_\varepsilon(\mathbb{R}^k) = 1$. By Assumption 1, the distribution μ_Z of Z in (5) is determined as $\mu_Z = C\pi(z)\mathrm{d}z$ where $C > 0$ and

$$\pi(z) := \int_{\mathscr{X}} \mathrm{e}^{-\ell(z - G(u))}\, \mu_0(\mathrm{d}u).$$

Note that $\pi(z)$ is well-defined since $|\mathrm{e}^{-\ell(z-G(u))}| \le 1$ and $\pi \in L^1(\mathbb{R}^k)$ due to Fubini's theorem [25, Theorem 1.27]. In particular, we have that $(U, Z) \sim \mu$ with $\mu(\mathrm{d}u, \mathrm{d}z) = C\mathrm{e}^{-\ell(z-G(u))}\, \mu_0(\mathrm{d}u) \otimes \mathrm{d}z$ where $\mathrm{d}z$ again denotes Lebesgue measure on $\mathbb{R}^k$. Further, we define the *potential*

$$\Phi(u; z) := \ell(z - G(u))$$

and assume the following to be satisfied.

Assumption 2

1. *The potential Φ is continuous w.r.t. z in mean-square sense w.r.t. μ_0, i.e, there exists an increasing function $\psi : [0,\infty) \to [0,\infty)$ with $\lim_{s\to 0} \psi(s) = \psi(0) = 0$ such that*
$$\int_X |\Phi(u;z) - \Phi(u;z')|^2 \, \mu_0(\mathrm{d}u) \leq \psi(|z-z'|).$$
For instance, there may exist a function $\theta \in L^2(\mathscr{X}, \mathscr{B}(\mathscr{X}), \mu_0; \mathbb{R})$ such that
$$|\Phi(u;z) - \Phi(u;z')| \leq \theta(u)\, \psi(|z-z'|).$$
2. *There holds $\pi(z) > 0$ for all $z \in \mathbb{R}^k$.*

Before stating the abstract version of Bayes' Rule in Theorem 1, we recall the finite-dimensional case $\mathscr{X} \simeq \mathbb{R}^n$ where it can be stated in terms of densities: here $\mu_0(\mathrm{d}u) = \pi_0(u)\mathrm{d}u$ and *Bayes' rule* takes the form
$$\pi^z(u) = \frac{1}{\pi(z)} \exp(-\Phi(u;z))\, \pi_0(u)$$
where $\mathrm{e}^{-\Phi(u;z)} = \mathrm{e}^{-\ell(z-G(u))}$ represents the *likelihood* of observing z when fixing u. The denominator $\pi(z)$ can be interpreted as a *normalizing constant* such that $\int_{\mathscr{X}} \pi^z(u)\,\mathrm{d}u = 1$. We now show that, in the general setting, Bayes' rule yields (a version of) the (regular) conditional measure $\mu_{U|Z}$ of U w.r.t. Z.

Theorem 1 (cf. [42, Theorems 4.2 and 6.31]). *Let Assumptions 1 and 2 be satisfied and define for each $z \in \mathbb{R}^k$ a probability measure on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ by*
$$\mu^z(\mathrm{d}u) := \frac{1}{\pi(z)} \exp(-\Phi(u;z))\, \mu_0(\mathrm{d}u). \tag{7}$$
Then the mapping $Q : \mathscr{B}(\mathscr{X}) \times \mathbb{R}^k$ given by
$$Q(B,z) := \mu^z(B) \qquad \forall B \in \mathscr{B}(\mathscr{X})$$
is a regular conditional distribution of U given Z. We call μ^z the posterior measure (of U given $Z = z$)*. Moreover, μ^z depends continuously on z w.r.t. the Hellinger metric, i.e., for any $z_1, z_2 \in \mathbb{R}^k$ with $|z_1 - z_2| \leq r$ there holds*
$$d_H(\mu^{z_1}, \mu^{z_2}) \leq C_r(z_1)\, \psi(|z_1 - z_2|),$$
where $C_r(z_1) = C(1 + \min\{\pi(z') : |z_1 - z'| \leq r\}^3)^{-1} < +\infty$.

Proof. Continuity with respect to the Hellinger metric is a slight generalization of [42, Theorem 4.2] and may be proved in the same way with obvious modifications. To show that Q is a regular conditional distribution we verify the two properties (a) and (b) given in Section 2.2.2. The first follows from the construction of μ^z. For the second property, note that measurability follows from continuity. The continuity of

μ^z w.r.t. z in the Hellinger metric implies also that $\mu^z(B)$ depends continuously on z due to the relations between Hellinger metric and total variation distance (see [17]). Finally, we have for any $A \in \mathscr{B}(\mathbb{R}^k)$ and $B \in \mathscr{B}(\mathscr{X})$ that

$$\begin{aligned}\mathbb{P}(U \in B, Z \in A) &= \int_{A\times B} \mu(\mathrm{d}u, \mathrm{d}z) = \int_A \int_B C\mathrm{e}^{-\ell(z-G(u))}\, \mu_0(\mathrm{d}u)\,\mathrm{d}z \\ &= \int_A C\pi(z) Q(B,z)\,\mathrm{d}z = \int_A Q(B,z)\, \mathbb{P}_Z(\mathrm{d}z)\end{aligned}$$

which completes the proof. □

Remark 2. We wish to emphasize that Theorem 1 and Assumption 2 show in detail the connection between the smoothness of the potential $\Phi(u;z) = \ell(z - G(u))$ and the continuity of the posterior μ^z w.r.t. z for a general prior μ_0 and an additive error ε with Lebesgue density proportional to $\mathrm{e}^{-\ell(\varepsilon)}$. Roughly speaking, the negative log-likelihood ℓ and the posterior μ^z share the same local modulus of continuity. This generalizes the results in [42] in that we allow for non-Gaussian priors μ_0 and errors ε.

Thus, under mild conditions, the Bayesian inverse problem is well-posed. It is also possible to prove continuity of μ^z w.r.t. to the forward map G, see [42, Section 4.4], which is crucial when the forward map G is realized by numerical approximation.

To give meaning to the mean and covariance of $U \sim \mu_0$ and $Z = G(U) + \varepsilon$, we make the further assumption that all second moments exist:

Assumption 3 *There holds*

$$\int_{\mathscr{X}} (\, \|u\|^2 + |G(u)|^2\,)\, \mu_0(\mathrm{d}u) < +\infty \quad \textit{and} \quad \int_{\mathbb{R}^k} |\varepsilon|^2\, \mu_\varepsilon(\mathrm{d}\varepsilon) < +\infty.$$

2.3 Bayes Estimators

Although the posterior measure μ^z is by definition the solution to the Bayesian inverse problem, it is, in general, by no means easy to compute in practice. In special cases, e.g., when G is linear and μ_0 and μ_ε are Gaussian measures, closed-form expressions for μ^z are available, but in general μ^z can only be computed in an approximate sense, see also Section 2.4. Moreover, when the dimension of $\mathscr{X}$ is large or infinite, visualizing, exploring or using μ^z for postprocessing are demanding tasks.

Other, more accessible quantities from Bayesian statistics, [3] which are also more similar to the result of deterministic parameter identification procedures than the posterior measure, are point estimates for the unknown u. In the Bayesian setting a point estimate is a "best guess" $\hat{u}$ of u based on posterior knowledge. Here "best" is determined by a *cost function* $c : \mathscr{X} \to \mathbb{R}_+$ satisfying $c(0) = 0$ and $c(u) \le c(\lambda u)$ for any $u \in \mathscr{X}$ and $\lambda \ge 1$. This cost function describes the loss or costs $c(u - \hat{u})$ incurred

when $\hat{u}$ is substituted for (the true) u for post processing or decision making. Note that also more general forms of a cost function are possible, see, e.g., [2, 3].

For any realization $z \in \mathbb{R}^k$ of the observation RV Z we introduce the *(posterior) Bayes cost* of the estimate $\hat{u}$ w.r.t. c as

$$\mathrm{B}_c(\hat{u}; z) := \int_{\mathscr{X}} c\,(u - \hat{u})\; \mu^z(\mathrm{d}u),$$

and define the *Bayes estimate* $\hat{u}$ as a minimizer of this cost, i.e.,

$$\hat{u} := \operatorname*{argmin}_{u' \in \mathscr{X}} \mathrm{B}_c(u'; z),$$

assuming that such a minimizer exists. The *Bayes estimator* $\hat{\phi} : \mathbb{R}^k \to \mathscr{X}$ is then the mapping which assigns to an observation z the associated Bayes estimate $\hat{u}$, i.e.,

$$\hat{\phi} : z \mapsto \operatorname*{argmin}_{u' \in \mathscr{X}} \mathrm{B}_c(u'; z)\ .$$

We assume measurability of $\hat{\phi}$ in the following. Note that $\hat{\phi}$ is then also the minimizer of the *(prior) Bayes cost*

$$\mathrm{B}_c(\hat{\phi}) := \int_{\mathbb{R}^k} \mathrm{B}_c(\hat{\phi}(z); z)\, \mu_Z(\mathrm{d}z) = \mathbb{E}\left[\mathrm{B}_c(\hat{\phi}(Z); Z)\right],$$

i.e., there holds

$$\mathbb{E}\left[\mathrm{B}_c(\hat{\phi}(Z); Z)\right] \leq \mathbb{E}\left[\mathrm{B}_c(\phi(Z); Z)\right]$$

for any other measurable $\phi : \mathbb{R}^k \to \mathscr{X}$.

Remark 3. Since $\hat{\phi} = \operatorname{argmin}_\phi \mathrm{B}_c(\phi)$ it is possible to determine the estimator $\hat{\phi}$ and thereby also the estimate $\hat{u} = \hat{\phi}(z)$ for a given z without actually computing the posterior measure μ^z, as the integrals in $\mathrm{B}_c(\hat{\phi})$ are w.r.t. the prior measure. Therefore, Bayes estimators are typically easier to approximate than μ^z.

We now introduce two very common Bayes estimators: the *posterior mean estimator* and the *maximum a posteriori estimator*.

2.3.1 Posterior Mean Estimator

For the cost function $c(u) = \|u\|^2$ the posterior Bayes cost

$$\mathrm{B}_c(\hat{u}; z) = \int_{\mathscr{X}} \|u - \hat{u}\|^2\, \mu^z(\mathrm{d}u)$$

is minimized by the posterior mean $\hat{u} = u_{\mathrm{CM}} := \int_{\mathscr{X}} u\, \mu^z(\mathrm{d}u)$. The corresponding Bayes estimator for $c(u) = \|u\|^2$ is then given by

$$\hat{\phi}_{\mathrm{CM}}(z) := \int_{\mathscr{X}} u\, \mu^z(\mathrm{d}u).$$

There holds in particular $\hat{\phi}_{\mathrm{CM}}(Z) = \mathbb{E}[U|Z]$ $\mathbb{P}$-almost surely.

Remark 4. If $\mathscr{X} \simeq \mathbb{R}^n$ and μ^z is unimodal, then the posterior mean minimizes $\int_{\mathscr{X}} c\,(u-\hat{u})\ \mu^z(\mathrm{d}u)$ for any symmetric, convex cost function c, see [28, 40].

Recall that, $\mathbb{E}[U|Z]$ is the best approximation of U in $L^2(\Omega, \sigma(Z), \mathbb{P}; \mathscr{X})$ w.r.t. the norm in $L^2(\Omega, \mathscr{F}, \mathbb{P}; \mathscr{X})$. Hence, the Bayes estimator $\hat{\phi}_{\mathrm{CM}}(Z) = \mathbb{E}[U|Z]$ represents the best L^2-approximation to U w.r.t. the information $\sigma(Z)$ available from the observation process Z.

2.3.2 Maximum a Posteriori Estimator

Another common estimator in Bayesian statistics is the *maximum a posteriori (MAP)* estimator $\hat{\phi}_{\mathrm{MAP}}$. For finite-dimensional $\mathscr{X} \simeq \mathbb{R}^n$ and absolutely continuous prior μ_0, i.e., $\mu_0(\mathrm{d}u) = \pi_0(u)\mathrm{d}u$, the MAP estimate is defined as

$$\hat{\phi}_{\mathrm{MAP}}(z) = \operatorname*{argmin}_{u\in\mathbb{R}^n} \Phi(u,z) - \log \pi_0(u)$$

provided the minimum exists for all $z \in \mathbb{R}^k$. For the definition of the MAP estimate via a cost function and the Bayes cost, we refer to the literature, e.g., [27, Section 16.2]; for MAP estimates in infinite dimensions, we refer to [8].

There is an interesting link between the Bayes estimator $\hat{\phi}_{\mathrm{MAP}}$ and the solution of the associated regularized least-squares problem: If $\mathrm{R} : \mathbb{R}^n \to [0,\infty)$ is a regularizing functional which satisfies $\int_{\mathbb{R}^n} \mathrm{R}(u)\mathrm{d}u < +\infty$, then the solution $\hat{u}_\alpha = \operatorname{argmin} |z - G(u)|^2 + \alpha \mathrm{R}(u)$ corresponds to the MAP estimate $\hat{\phi}_{\mathrm{MAP}}(z)$ for $\varepsilon \sim N(0, \sigma^2 I)$ and $\mu_0(\mathrm{d}u) \propto \exp(-\frac{\alpha}{\sigma^2}\,\mathrm{R}(u))\ \mathrm{d}u$.

2.4 Computational Methods for Bayesian Inverse Problems

We summarize the most common methods for computing the posterior measure and Bayes estimators, referring to the cited literature for details .

In finite dimensions $\mathscr{X} \simeq \mathbb{R}^n$ and in the case of *conjugate priors*, see, e.g., [20], the posterior density is available in closed form since in this case the product of the prior density and the likelihood function belongs to the same class of probability densities as the prior. Therefore only the parameters of the posterior need to be computed, and for these analytical formulas are often available.

Aside from these special cases μ^z can only be approximated — but how may a probability distribution, possibly on an infinite-dimensional space, be approximated

computationally? Perhaps the simplest and most natural idea is to generate samples distributed according to the posterior measure. A well-known method for this purpose is the *Markov Chain Monte Carlo* method (MCMC). The idea here is to construct a Markov chain with the posterior measure as its stationary resp. limiting distribution. If such a chain is run sufficiently long, it will yield (correlated) samples which are asymptotically distributed according to the posterior measure. For details we refer to [19] and, for the underlying theory of Markov chains, to [30]. The computational efficiency of the chain mainly depends on its transition kernel. Recently, much research has been devoted towards constructing good kernels. We mention [7] for MCMC suited to very high and even infinite dimensions, to [18] for the idea of adapting the kernel to geometrical features of the posterior and to [29], where this idea is realized by a transition kernel derived from the Gauss-Newton-method.

Besides MCMC another common Bayesian method are *particle filters* [24, Section 4.3]. Here samples are generated according to the prior and all samples are assigned initially equal weights. Then, in an updating step, the weights are modified according to the posterior distribution. A further extension, *Gaussian mixture filters* [41], approximate the posterior density by a weighted mean of Gaussian kernels located at samples/particles. Here, in addition to the weights, also the location of the particles are modified according to the posterior.

A further technique for sampling from the posterior is presented in [9]: here a mapping $F : \mathscr{X} \to \mathscr{X}$ is constructed in such a way that $F(U) \sim \mu^z$ for a random variable $U \sim \mu_0$. Given F, which is obtained by solving an optimal transport problem, samples according to μ^z can then easily be generated by evaluating F for samples from the prior.

For the posterior mean, the immediate computational method is numerical integration w.r.t. $\mu^z(\mathrm{d}u)$ or $\mathrm{e}^{-\Phi(u;z)}\mu_0(\mathrm{d}u)$. A Monte Carlo integration is again performed by averaging samples generated by a suitable Markov chain. Recently, sparse quadrature methods based on known quadrature rules for μ_0 have been investigated, see [37, 38]. Due to assumed smoothness of the likelihood $\mathrm{e}^{-\Phi(u;z)}$ w.r.t. u, these methods can yield faster convergence rates than Monte Carlo/MCMC integration and are also suited to infinite dimensions.

Alternatively, the corresponding Bayes estimator ϕ_{CM} could be approximated, e.g., by linear functions, and simply evaluated for the observational data. We return to this approach in Section 3.3 and show that Kalman filters may be viewed as approximation methods of this type.

Computing the MAP estimate is, by construction, a minimization problem for the posterior density and related to classical Tikhonov regularization. Therefore, methods from numerical optimization and computational inverse problems, respectively, can be applied here [10, 44]. Note that in numerical weather prediction the popular methods *3DVar* and *4DVar* are precisely computations of the MAP estimate. The difference between both is that 3DVar treats the typically sequential data recursively, while 4DVar performs the optimization w.r.t. the entire data set at once, see also [28].

3 Analysis of Kalman Filters for Bayesian Inverse Problems

In this section we consider Kalman filters and their application to the nonlinear Bayesian inverse problem (5). We begin with the classical Kalman filter for state estimation in linear dynamics and then consider two generalizations to the nonlinear setting which have been recently proposed for UQ in inverse problems. We show that both methods can be understood as discretizations of the same updating scheme for a certain RV and analyze the properties of this updated variable, thereby characterizing the properties of the approximations provided by the two filtering methods. In particular, we show that Kalman filters do not solve of the nonlinear Bayesian inverse problem, nor can they be justified as approximations to its solution. They are, rather, related to the linear approximation of the Bayes estimator ϕ_{LCM} and its estimation error.

3.1 The Kalman Filter

The Kalman filter [26] is a well-known method for sequential state estimation for incompletely observable, linear discrete-time dynamics, see, e.g., [6, 39] for a broader introduction and discussion. Thus, the Kalman filter may be applied to systems of the form

$$U_n = A_n U_{n-1} + \eta_n, \qquad Z_n = G_n U_n + \varepsilon_n, \qquad n = 1,2,\dots \tag{8}$$

where U_n denotes the unknown, unobservable state and Z_n the observable process at time n, and where U_0, η_n and ε_n are mutually independent RVs. The operators A_n and G_n are linear mappings in state space and from state to observation space, respectively. For the noises η_n and ε_n, zero mean and given covariances Γ_n and Σ_n, respectively, are assumed. Then, given observations $Z_1 = z_1, \dots, Z_n = z_n$ of the process Z, the state U_n is to be inferred. Assume an initial guess $\hat{u}_0$ of the unknown U_0 with minimal variance $\mathrm{trace}(E_0)$ where $E_0 := \mathrm{Cov}(U_0 - \hat{u}_0)$ denotes the error covariance of the estimate $\hat{u}_0$. Then the Kalman filter results in recursive equations for the minimum variance estimates $\hat{u}_n$ of U_n and their error covariances $E_n := \mathrm{Cov}(U_n - \hat{u}_n)$.

Although the main advantage of the Kalman filter is its sequential structure which allows for a significant reduction of computational work (see [42, Section 5.3] for a nice discussion on this topic) we will apply the Kalman filter to our stationary inverse problem

$$Z = GU + \varepsilon, \qquad U \sim N(m_0, C_0), \quad \varepsilon \sim N(0, \Sigma), \tag{9}$$

which is, of course, only a special case of the system (8) in that there are no dynamics, $A_n \equiv I$, $\eta_n \equiv 0$ and only a single update $n = 1$. If we take $\hat{u}_0 = m_0$ as the initial guess this yields $E_0 = C_0$ and the Kalman filter yields the updates

$$\hat{u}_1 = \hat{u}_0 + K(z - G\hat{u}_0), \qquad E_1 = E_0 - KGE_0$$

where $K = E_0 G^*(GE_0 G^* + \Sigma)^{-1}$ is the well-known *Kalman gain*.

In the Gaussian case (9), for which (U,Z) is a jointly Gaussian RV, the posterior measure μ^z is again Gaussian, i.e., $\mu^z = N(m^z, C^z)$. Moreover, the posterior mean m^z and the posterior covariance C^z are given by

$$m^z = m_0 + K(z - Gm_0), \qquad C^z = C_0 - KGC_0,$$

where $K = C_0 G^*(GC_0 G^* + \Sigma)^{-1}$. Thus, for (9) the Kalman filter is seen to yield the solution of the Bayesian inverse problem by providing the posterior mean and covariance. However, we emphasize that the Kalman filter does not directly approximate the posterior measure. The filter provides estimates and error covariances which, in the Gaussian case, coincide with the posterior mean and covariance which, in turn, uniquely determine a Gaussian posterior measure. Whenever the linearity of G or Gaussianity of the prior $U \sim \mu_0$ or noise $\varepsilon \sim N(0,\Sigma)$ do not hold, then neither does the Kalman filter yield the first two posterior moments nor is the posterior measure necessarily Gaussian. We will return to the interpretation of the Kalman filter for linear G but non-Gaussian U or ε in Section 3.3.

3.2 Kalman Filter Extensions for Nonlinear Inverse Problems

Besides the extended Kalman filter (EKF), which is based on linearizations of the nonlinear forward map G but which we shall not consider here, a widely used method for nonlinear systems is the Ensemble Kalman Filter (EnKF) introduced by Evensen [13]. In addition, a more recent development, the Polynomial Chaos Expansion Kalman Filter (PCE-KF) developed by Matthies et al. [32, 34, 35] can also be applied to the nonlinear inverse problem (5).

3.2.1 The Ensemble Kalman Filter

Since its introduction in 1994, the EnKF has been investigated and evaluated in many publications [14, 5, 15, 16, 31]. However, the focus is usually on its application to state or parameter estimation rather than solving Bayesian inverse problems. Recently, the interest in the EnKF for UQ in inverse problems has increased, see, e.g., [21, 22, 27].

If we consider $Z = G(U) + \varepsilon$ with $U \sim \mu_0$ and $\varepsilon \sim \mu_\varepsilon$ and given observations $z \in \mathbb{R}^k$, the EnKF algorithm proceeds as follows:

1. **Initial ensemble:** Generate samples $u_1, \ldots, u_M$ of U according to μ_0.
2. **Forecast:** Generate samples $z_1, \ldots, z_M$ of Z by

$$z_j = G(u_j) + \varepsilon_j, \qquad j = 1, \ldots, M,$$

where $\varepsilon_1,\dots,\varepsilon_M$ are samples of ε according to μ_ε.

3. **Analysis:** Update the inital ensemble $\mathbf{u}=(u_1,\dots,u_M)$ member by member via

$$u_j^a = u_j + \tilde{K}(z-z_j), \qquad j=1,\dots,M, \tag{10}$$

where $\tilde{K}=\mathrm{Cov}(\mathbf{u},\mathbf{z})\mathrm{Cov}(\mathbf{z})^{-1}$ and $\mathrm{Cov}(\mathbf{u},\mathbf{z})$ and $\mathrm{Cov}(\mathbf{z})=\mathrm{Cov}(\mathbf{z},\mathbf{z})$ are the empirical covariances of the samples $\mathbf{u}$ and $\mathbf{z}=(z_1,\dots,z_M)$. This yields an *analysis ensemble* $\mathbf{u}^a=(u_1^a,\dots,u_M^a)$.
The empirical mean of $\mathbf{u}^a$ serves as estimate $\hat{u}$ for the unknown u and the empirical covariance of $\mathbf{u}^a$ as an indicator for the accuracy of the estimate.

Note that for dynamical systems such as (8), the analysis ensemble $A_n(\mathbf{u}^a)$ serves as the initial ensemble for the next step n.

3.2.2 The Polynomial Chaos Expansion Kalman Filter

In [32, 34, 35] the authors propose a sampling-free Kalman filtering scheme for nonlinear systems. Rather than updating samples of the unknown, this is carried out for the coefficient vector of a polynomial chaos expansion of the unknown. This necessitates the construction of a polynomial chaos expansion distributed according to the prior measure μ_0: we assume there exist countably many independent real-valued random variables $\xi=(\xi_m)_{m\in\mathbb{N}}$, and *chaos coefficients* $u_\alpha\in\mathscr{X}$, $\varepsilon_\alpha\in\mathbb{R}^k$ for each

$$\alpha\in\mathbb{J}:=\{\alpha\in\mathbb{N}_0^{\mathbb{N}} : \alpha_j\neq 0 \text{ for only finitely many } j\},$$

such that

$$\sum_{\alpha\in\mathbb{J}}\|u_\alpha\|^2<+\infty \quad \text{and} \quad \sum_{\alpha\in\mathbb{J}}|\varepsilon_\alpha|^2<+\infty,$$

and

$$\Big(\sum_{\alpha\in\mathbb{J}} u_\alpha P_\alpha(\xi),\quad \sum_{\alpha\in\mathbb{J}}\varepsilon_\alpha P_\alpha(\xi)\Big)\sim\mu_0\otimes\mu_\varepsilon.$$

Here, $P_\alpha(\xi)=\prod_{m\geq 1}P_{\alpha_m}^{(m)}(\xi_m)$ denotes the product of univariate orthogonal polynomials $P_{\alpha_m}^{(m)}$ where we require $\{P_\alpha^{(m)}\}_{\alpha\in\mathbb{N}}$ to be a CONS in $L^2(\Gamma_m,\mathscr{B}(\Gamma_m),\mathbb{P}_{\xi_m})$, $\Gamma_m=\xi_m(\Omega)\subseteq\mathbb{R}$. Note, that the completeness of orthogonal polynomials will depend in general on properties of the measure $\mathbb{P}_{\xi_m}$, see [12] for a complete characterization.

We then define $U:=\sum_{\alpha\in\mathbb{J}}u_\alpha P_\alpha(\xi)$ and $\varepsilon:=\sum_{\alpha\in\mathbb{J}}\varepsilon_\alpha P_\alpha(\xi)$, denoting their PCE vectors $(u_\alpha)_{\alpha\in\mathbb{J}}$ and $(\varepsilon_\alpha)_{\alpha\in\mathbb{J}}$ by $[U]$ and $[\varepsilon]$. For the same problem considered for the EnKF, the PCE-KF algorithm is as follows.

1. **Initialization:** Compute a PCE with coefficient vector $[U]$ such that $U\sim\mu_0$.
2. **Forecast:** Compute the PC vector $[G(U)]$ of $G(U)$ and set

$$[Z]:=[G(U)]+[\varepsilon],$$

where $[\varepsilon]$ is a PC vector such that then $\varepsilon \sim \mu_\varepsilon$. Note that, by linearity, $[Z]$ is the PC vector of the RV defined by $Z := G(U) + \varepsilon$.

3. **Analysis:** Update the inital PC vector by

$$[U]^a = [U] + K \otimes I_{\mathbb{J}} \left([z] - [Z]\right), \tag{11}$$

where $[z] = (z, 0, \ldots)$ is the PC vector of the observed data $z \in \mathbb{R}^k$ and $K := \mathrm{Cov}(U,Z)\mathrm{Cov}(Z)^{-1}$. The action of the covariances as operators can be described, e.g. in the case of $\mathrm{Cov}(U,Z) : \mathbb{R}^k \to \mathscr{X}$, by

$$\mathrm{Cov}(U,Z)z = \sum_{\alpha \in \mathbb{J}} \sum_{\beta \in \mathbb{J}} z_\beta^\top z \, u_\alpha.$$

The result of one step of the PCE-KF algorithm is an *analysis PC vector* $[U]^a$.

Remark 5. Neither the independence of the $\{\xi_m\}_{m\in\mathbb{N}}$ nor an expansion in polynomials $\{P_\alpha(\xi)\}$ is crucial for the PCE-KF. In principle, only a countable CONS $\{\Psi_\alpha\}_{\alpha\in\mathbb{N}}$ for the space $L^2(\Gamma, \mathscr{B}(\Gamma), \mathbb{P}_\xi)$, $\Gamma = \xi(\Omega) \subseteq \mathbb{R}^{\mathbb{N}}$, is required such that $\left(\sum_\alpha u_\alpha \Psi_\alpha(\xi), \sum_\alpha \varepsilon_\alpha \Psi_\alpha(\xi)\right) \sim \mu_0 \otimes \mu_\varepsilon$. However, the independence structure of $\mu_0 \otimes \mu_\varepsilon$ requires at least two independent random vectors $\eta = (\eta_1, \ldots, \eta_M)$, $\zeta = (\zeta_1, \ldots, \zeta_N)$, $\xi = (\eta, \zeta)$, and expansions of the form $\sum_\alpha u_\alpha \Psi_\alpha(\eta_1, \ldots, \eta_M)$ and $\sum_\alpha \varepsilon_\alpha \Psi_\alpha(\zeta_1, \ldots, \zeta_N)$.

3.2.3 The Analysis Variable

Note that the analysis PC vector $[U]^a$ defines an *analysis variable* $U^a := \sum_{\alpha\in\mathbb{J}} u_\alpha^a P_\alpha(\xi)$. Indeed, both EnKF and PCE-KF perform discretized versions of an update for RVs, namely,

$$U^a = U + K(z - Z), \qquad K = \mathrm{Cov}(U,Z)\mathrm{Cov}(Z)^{-1},$$

where $Z := G(U) + \varepsilon$, and $(U, \varepsilon) \sim \mu_0 \otimes \mu_\varepsilon$, providing samples $\mathbf{u}^a$ and PCE vectors $[U]^a = [U^a]$ of U^a, respectively. This raises the question of how the analysis variable U^a is to be understood in context of Bayesian inverse problems?

3.3 The Linear Conditional Mean

To relate the results produced by the EnKF or PCE-KF to the Bayesian setting, we introduce a new Bayes estimator, or, more precisely, a linear approximation to the Bayes estimator $\hat{\phi}_{\mathrm{CM}}$ resp. the conditional mean $\mathbb{E}[U|Z]$. The *linear posterior mean estimator* $\hat{\phi}_{\mathrm{LCM}}$ is given by

$$\hat{\phi}_{\mathrm{LCM}} = \operatorname*{argmin}_{\phi \in \mathrm{span}\{1,z\}} \mathbb{E}\left[\|U - \phi(Z)\|^2\right], \tag{12}$$

here $\operatorname{span}\{1,z\} = \{\phi : \phi(z) = b + Az \text{ with } b \in \mathscr{X}, A : \mathbb{R}^k \to \mathscr{X} \text{ linear and bounded}\}$.

Moreover, we denote the RV $\hat{\phi}_{\mathrm{LCM}}(Z)$ as the *linear conditional mean*. Thus, $\hat{\phi}_{\mathrm{LCM}}(Z)$ is the best $L^2(\Omega,\mathscr{F},\mathbb{P};\mathscr{X})$-approximation to $U \sim \mu_0$ in the subspace $\operatorname{span}\{1,Z\} \subset L^2(\Omega,\sigma(Z),\mathbb{P};\mathscr{X})$. Or, alternatively, ϕ_{LCM} is the linear estimator with minimal prior Bayes cost for $c(u) = \|u\|^2$. Furthermore, there holds

$$\hat{\phi}_{\mathrm{LCM}}(z) = \mathbb{E}[U] + K(z - \mathbb{E}[Z]),$$

with the usual Kalman gain $K = \operatorname{Cov}(U,Z)\operatorname{Cov}(Z)^{-1}$, and we immediately obtain the following result.

Theorem 2. *Consider* (5) *and let Assumptions 1 – 3 be satisfied. Then for any $z \in \mathbb{R}^k$ the* analysis variable $U^a = U + K(z - Z)$, $K = \operatorname{Cov}(U,Z)\operatorname{Cov}(Z)^{-1}$, *coincides with*

$$U^a = \hat{\phi}_{\mathrm{LCM}}(z) + (U - \hat{\phi}_{\mathrm{LCM}}(Z)).$$

In particular, there holds

$$\mathbb{E}[U^a] = \hat{\phi}_{\mathrm{LCM}}(z) \quad \text{and} \quad \operatorname{Cov}(U^a) = \operatorname{Cov}(U) - K\operatorname{Cov}(Z,U).$$

We summarize the consequences of Theorem 2 as follows:

- The analysis variable U^a, to which the EnKF and the PCE-KF provide approximations, is the sum of a Bayes estimate $\hat{\phi}_{\mathrm{LCM}}(z)$ and the prior error $U - \hat{\phi}_{\mathrm{LCM}}(Z)$ of the corresponding Bayes estimator $\hat{\phi}_{\mathrm{LCM}}$.
- The resulting mean of the EnKF analysis ensemble or the PCE-KF analysis vector corresponds to the linear posterior mean estimate and therefore provides an approximation to the true posterior mean.
- The covariance approximated by the empirical covariance of the EnKF analysis ensemble, as well as that of the PCE-KF analysis vector, is independent of the actual observational data $z \in \mathbb{R}^k$. It therefore constitutes a prior rather than a posterior measure of uncertainty.
- In particular, the randomness in U^a is entirely determined by the prior measures μ_0 and μ_ε. Only the location, i.e., the mean, of U^a is influenced by the observation data z; the randomness of U^a is independent of z and determined only by the projection error $U - \hat{\phi}_{\mathrm{LCM}}(Z)$ w.r.t. the prior measures.
- By to the last two items, the analysis variable U^a, and therefore the EnKF analysis ensemble or the result of the PCE-KF, are in general not distributed according to the posterior measure μ^z. Moreover, the difference between μ^z and the distribution of U^a depends on the data z and can become quite large for nonlinear problems, see Example 3.

Remark 6. Note that in particular the second and third item above explain the observations made in [27], i.e., that "[...] (i) with appropriate parameter choices, approximate filters can perform well in reproducing the mean of the desired probability distribution, (ii) they do not perform as well in reproducing the covariance [...] ".

We illustrate the conceptual difference between the distribution of the analysis variable U^a and the posterior measure μ^z with a simple yet striking example.

Example 3. We consider $U \sim N(0,1)$, $\varepsilon \sim N(0,\sigma^2)$ and $G(u) \equiv u^2$. Given data $z \in \mathbb{R}$, the posterior measure, obtained from Bayes' rule for the densities, is

$$\mu^z(\mathrm{d}u) = C\exp\left(-\frac{\sigma^2u^2 + (z-u^2)^2}{2\sigma^2}\right)\mathrm{d}u.$$

Due to the symmetry of μ^z we have $\hat{u}_{\mathrm{CM}} = \int_{\mathscr{X}} u\,\mu^z(\mathrm{d}u) = 0$ for any $z \in \mathbb{R}^k$. Thus, $\mathbb{E}[U|Z] \equiv 0$ and $\hat{\phi}_{\mathrm{LCM}} \equiv \hat{\phi}_{\mathrm{CM}}$. In particular, we have $K = 0$ due to

$$\mathrm{Cov}(U,Z) = \mathrm{Cov}(U,U^2) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} u(u^2-1)\mathrm{e}^{-u^2/2}\mathrm{d}u = 0,$$

which in turn yields $U^a = U \sim N(0,1)$. Hence the analysis variable is distributed according to the prior measure. This is not surprising as, by definition, its mean is the best linear approximation to the posterior mean according to μ^z and its fluctuation is simply the prior estimation error $U - \hat{\phi}_{\mathrm{LCM}}(Z) = U - 0 = U$. This illustrates that U^a is suited for approximating the posterior mean, but not appropriate as a method for uncertainty quantification in nonlinear inverse problems. As displayed in Figure 1, the distribution of U^a can be markedly different from the true posterior distribution.

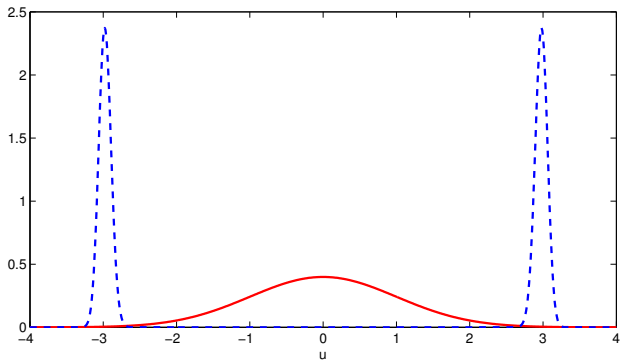

Fig. 1 Density of the posterior μ^z (*dashed, blue line*) and the probability density of the analysis variable U^a (*solid, red line*) for $z = 9$ and $\sigma = 0.5$.

4 Numerical Example: 1D Elliptic Boundary Value Problem

To illustrate the application of the EnKF and PCE-KF to a simple Bayesian inverse problems, we consider the following PDE model on $D = [0,1]$:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\exp(u_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x) = f(x), \qquad p(0) = p_0, \quad p(1) = u_2. \tag{13}$$

Here $u = (u_1, u_2)$ are the unknown parameters to be identified. The solution of (13) is given by

$$p(x) = p_0 + (u_2 - p_0)x + \exp(-u_1)\left(S_x(F) - S_1(F)x\right), \tag{14}$$

where $S_x(g) := \int_0^x g(y)\,\mathrm{d}y$ and $F(x) = S_x(f) = \int_0^x f(y)\,\mathrm{d}y$. For simplicity we choose $f \equiv 1$ and $p_0 = 0$ in the following.

Assume now that noisy measurements of p are available at $x_1 = 0.25$ and $x_2 = 0.75$, namely $z = (27.5, 79.7)$. We wish to infer u based on this data and on a priori information modelled by the prior distributions of the independent random variables

$$u_1 \sim N(0,1), \quad \text{and} \quad u_2 \sim \mathrm{Uni}(90,110).$$

Here $\mathrm{Uni}(90,110)$ denotes the uniform distribution on the interval $[90,110]$. Thus, the forward map here is $G(u) = (p(x_1), p(x_2))$ with p according to (14) for $f \equiv 1$, and the model for the measurement noise is $\varepsilon \sim N(0, 0.01 \cdot I_2)$.

In Figure 2 we show the prior and the posterior densities as well as 1000 ensemble members of the initial and analysis ensemble obtained by the EnKF. A total ensemble size of $M = 10^5$ was chosen in order to reduce the sampling error to a negligible level. It can be seen, however, that the analysis EnKF-ensemble does not follow the posterior distribution, although its mean $(-2.92, 105.14)$ is quite close to the true posterior mean $(-2.65, 104.5)$ (computed by quadrature). To illustrate

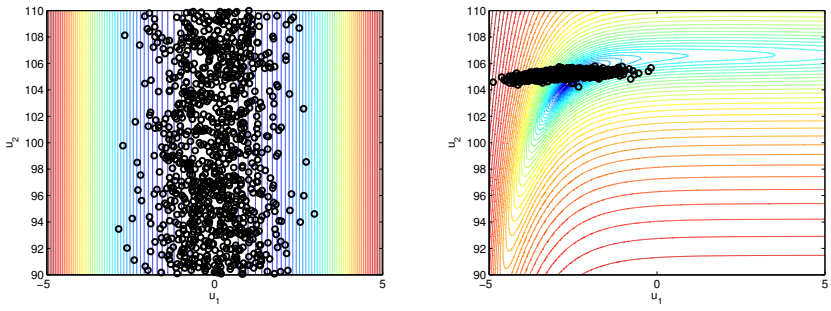

Fig. 2 Left: Contour plot of the negative logarithm of the prior density and the locations of 1000 ensemble members of the initial EnKF-ensemble.
Right: Contour plot of the logarithm of the negative logarithm of the posterior density and the locations of the updated 1000 ensemble members in the analysis EnKF-ensemble.

the difference between the distribution of the analysis ensemble resp. variable and the true posterior distribution, we present the marginal posterior distributions of u_1 and u_2 in Figure 3. For the posterior the marginals were evaluated by quadrature, whereas for the analysis ensemble we show a relative frequency plot.

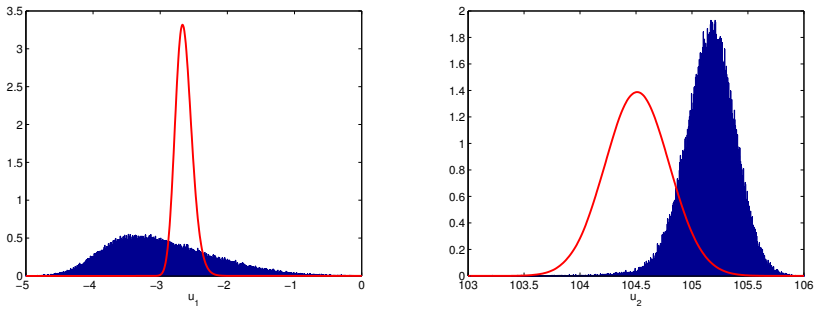

Fig. 3 Left: Posterior marginal and relative frequencies in the analysis ensemble for u_1.
Right: The same for u_2.

We note that slightly changing the observational data to $\tilde{z} = (23.8, 71.3)$ moves the analysis ensemble resp. variable much closer to the true posterior, see Figure 4. Also, the mean of the analysis ensemble $(0.33, 94.94)$ provides a better fit to the true posterior mean $(0.33, 94.94)$ here.

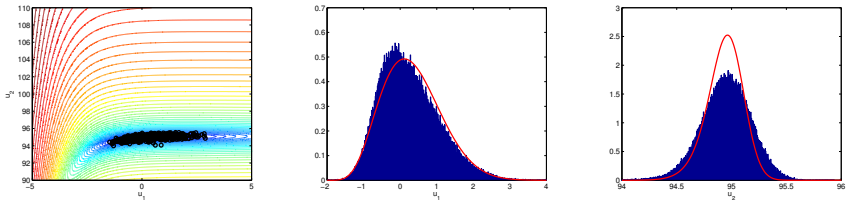

Fig. 4 Left: Contours of the logarithm of the negative log posterior density and locations of 1000 members of the analysis EnKF-ensemble.
Middle: Posterior marginal and relative frequencies in the analysis ensemble for u_1.
Right: The same for u_2.

To reaffirm the fact that only the mean of analysis variable U^a depends on the actual data, we show density estimates for the marginals of u_1 and u_2 of U^a in Figure 5. Here we have used once the data $z = (27.5, 79.7)$ (blue lines) and once $\tilde{z} = (23.8, 71.3)$ (green lines). The density estimates were obtained by normal kernel density estimation (KDE, in this case MATLAB's `ksdensity` routine) based on the resulting analysis ensembles $(\mathbf{u}_1^a, \mathbf{u}_2^a)$ and $(\tilde{\mathbf{u}}_1^a, \tilde{\mathbf{u}}_2^a)$, respectively, of the EnKF for these two data sets $z, \tilde{z}$. In the left picture we show the KDE for $\mathbf{u}_1^a$ and $\tilde{\mathbf{u}}_1^a$ and in the middle picture we display the KDE for the corresponding centered ensembles

$\mathbf{u}_1^a - \mathbb{E}[\mathbf{u}_1^a]$ and $\tilde{\mathbf{u}}_1^a - \mathbb{E}[\tilde{\mathbf{u}}_1^a]$. In the right picture we provide the KDEs for the centered ensembles of u_2. Note that the marginal distributions of the centered ensembles coincide, in agreement with Theorem 2.

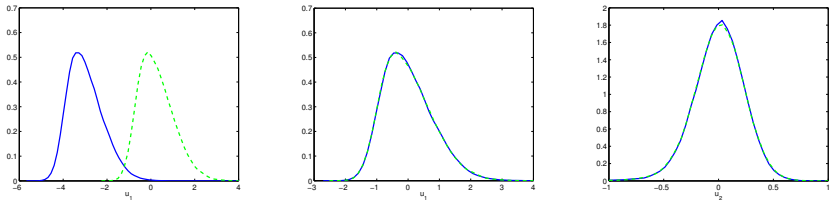

Fig. 5 Left: Kernel density estimates for $\mathbf{u}_1^a$ (*blue, solid line*) and $\tilde{\mathbf{u}}_1^a$ (*green, dashed line*). Middle: Kernel density estimates for $\mathbf{u}_1^a - \mathbb{E}[\mathbf{u}_1^a]$ (*blue, solid*) and $\tilde{\mathbf{u}}_1^a - \mathbb{E}[\tilde{\mathbf{u}}_1^a]$ (*green, dashed*). Right: Kernel density estimates for $\mathbf{u}_2^a - \mathbb{E}[\mathbf{u}_2^a]$ (*blue, solid*) and $\tilde{\mathbf{u}}_2^a - \mathbb{E}[\tilde{\mathbf{u}}_2^a]$ (*green, dashed*).

However, note that, particularly in this example where the prior, and thus posterior, support for u_2 is bounded, the EnKF may yield members in the analysis ensemble which are outside this support. This is a further consequence of Theorem 2: Since the analysis ensemble of the EnKF follows the distribution of the analysis variable rather than that of the true posterior distribution, ensemble members lying outside the posterior support can always occur whenever the support of the analysis variable is not a subset of the support of the posterior.

In addition, we would like to stress that, whether or not the distribution of the analysis variable is a good fit to the true posterior distribution depends entirely on the observed data — which can neither be controlled nor are known a priori.

Applying the PCE-KF to this simple example problem can be done analytically. We require four basic independent random variables $\xi_1 \sim N(0,1)$, $\xi_2 \sim \mathrm{Uni}(0,1)$, $\xi_3 \sim N(0,1)$ and $\xi_4 \sim N(0,1)$ to define PCEs which yield random variables distributed according to the prior and error distributions:

$$U := (\xi_1,\, 90 + 20\xi_2)^\top \sim \mu_0, \qquad \varepsilon := (0.1\xi_3,\, 0.1\xi_4)^\top \sim \mu_\varepsilon.$$

Moreover, due to (14), $G(U)$ is also available in closed form as

$$G(U) = \begin{pmatrix} c_{11}(90+20\xi_2) + c_{12}\sum_{n=0}^{\infty}(-1)^n \frac{\sqrt{e}}{\sqrt{n!}} H_n(\xi_1) \\ c_{21}(90+20\xi_2) + c_{22}\sum_{n=0}^{\infty}(-1)^n \frac{\sqrt{e}}{\sqrt{n!}} H_n(\xi_1) \end{pmatrix},$$

where H_n denotes the nth normalized Hermite polynomial and $c_{11}, c_{12}, c_{21}, c_{22}$ can be deduced from inserting $x = 0.25$ and $x = 0.75$ into (14). Here, we have used the expansion of $\exp(-\xi)$ in Hermite polynomials, see also [43, Example 2.2.7]. Thus, the PCE coefficient vectors $[U]$ and $[G(U) + \varepsilon]$ w.r.t. the polynomials

$$P_\alpha(\xi) = H_{\alpha_1}(\xi_1) L_{\alpha_2}(\xi_2) H_{\alpha_3}(\xi_3) H_{\alpha_4}(\xi_4), \qquad \alpha \in \mathbb{N}_0^4,$$

can be obtained explicitly. Here H_α and L_α denote the αth normalized Hermite and Legendre polynomials, respectively. In particular, the nonvanishing chaos coefficients involve only the basis polynomials

$$P_0(\xi) \equiv 1, \quad P_1(\xi) = L_1(\xi_2), \quad P_2(\xi) = H_1(\xi_3), \quad P_3(\xi) = H_1(\xi_4)$$

and $P_\alpha(\xi) = H_{\alpha-3}(\xi_1)$ for $\alpha \geq 4$. Arranging the two-dimensional chaos coefficients of U and $G(U)$ as the column vectors $[U], [G(U) + \varepsilon] \in \mathbb{R}^{2\times\mathbb{N}_0}$, and denoting by $[\dot{U}]$ the matrix $(u_1, u_2, \ldots) \in \mathbb{R}^{2\times\mathbb{N}}$ we get

$$K = [\dot{U}][G(\dot{U})]^\top \left([G(\dot{U})][G(\dot{U})]^\top + 0.01 I_2 \right)^{-1}.$$

Thus, the only numerical error for applying the PCE-KF to the example is the truncation of the PCE. We have carried out this calculation using a truncated PCE of length $J = 4 + 50$ according to the reduced basis above, evaluated the approximation to K by using the truncated vector $[G(U)]$ in the formula above and then performed the update of the PCE vectors according to (11). We then sampled the resulting random variable U^a again $M = 10^5$ times. The resulting empirical distributions were essentially indistinguishable from the results obtained by the EnKF described previously and are therefore omitted.

Remark 7. Although a detailed complexity analysis of these methods is beyond the scope of this contribution, we would like to mention that the EnKF calls for M evaluations of the forward map $G(u_j)$, $j = 1, \ldots, M$, whereas the PCE-KF requires computing the chaos coefficients of $G(U)$ by, e.g., the Galerkin method. Thus the former yields, in general, many small systems to solve, whereas the latter typically requires the solution of a large coupled system. Moreover, we emphasize the computational savings by applying Kalman filters compared to a "full Bayesian update", i.e., sampling from the posterior measure by MCMC methods. In particular, each MCMC run one may require calculating many hundreds of thousands forward maps $G(u)$, e.g., for each iteration u_j of the Markov chain as in the case of Metropolis-Hastings MCMC. Hence, if one is interested in only the posterior mean as a Bayes estimate, then EnKF and PCE-KF provide substantially less expensive alternatives to MCMC for its approximation by means of the linear posterior mean.

5 Conclusions

We have constrasted the deterministic and Bayesian formulations of nonlinear inverse problems such as arise in parameter estimation and data assimilation settings. An important distinction lies in the objectives of the two approaches: the identification of a particular value of the unknown quantity in the deterministic case versus the updating of a prior to a posterior probability measure encoding the uncertainty associated with the unknown quantity due to new observations. Moreover, we have also pointed out the relation beween regularized least-squares solutions and the concept

of Bayesian (point) estimators. Among the computational methods for Bayesian inverse problems we have focused on Kalman filters such as the EnKF and PCE-KF and presented a precise characterization of these methods in the Bayesian setting. A summary of the contrasting features of Bayesian inversion, Bayes estimators and Kalman filter-based methods is given in Table 1.

	Bayesian Inversion	Bayes Estimators	Kalman Filters
Goal	Merge prior belief with new observational data	Compute best guess w.r.t. posterior belief	Compute best linear guess and associated error
Result	measure μ^z on $\mathscr{X}$	estimate $\hat{u} \in \mathscr{X}$	estimate $\hat{u} \in \mathscr{X}$ and estimation error $U - \hat{\phi}_{\mathrm{LCM}}(Z)$
Allows for	rigorous UQ in post-processing	deterministic post-processing with $\hat{u}$	deterministic post-processing with $\hat{u}$ and certain UQ

Table 1 Distinguishing features of Bayesian inverse problems, Bayes estimators and Kalman filters.

Most important, the RVs approximated by the Kalman filter-based methods, will not, in general, be distributed according to the posterior distribution in the Bayes' sense. They are rather related to a common Bayes estimator – the linear conditional mean – and its estimation error RV, and therefore represent a different uncertainty model than the posterior measure. Some carefully chosen numerical examples were given to illustrate these basic differences.

References

1. S. Anzengruber, B. Hofmann, P. Mathé, doi: 10.1080/00036811.2013.833326
2. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. (Springer, New York, 1985)
3. J. M. Bernardo, in *Probability and Statistics*, ed. by R. Viertl. Encyclopedia of Life Support Systems (UNESCO, Oxford, 2003)
4. M. Burger, F. Lucka, *Computational and theoretical aspects of sparsity-constraints in Bayesian inversion (working title)*, in preparation
5. G. Burgers, P. J. van Leeuwen, G. Evensen, Monthly Weather Review **126**
6. D. E. Catlin, *Estimation, Control, and the Discrete Kalman Filter* (Springer, New York, 1989)
7. S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, Statistical Science **28**
8. M. Dashti, K. J. H. Law, A. Stuart, J. Voss, arXiv:1303.4795v3
9. T. El Moselhy, Y. Marzouk, Journal of Computational Physics **231**(23)
10. H. W. Engl, M. Hanke, A. Neubauer. *Regularization of inverse problems* (Kluwer Academic Publishers, Dordrecht, 2000)
11. H. W. Engl, K. Kunisch, A. Neubauer, Inverse Problems **5**
12. O. G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, ESAIM: Mathematical Modelling and Numerical Analysis **46**(2)
13. G. Evensen, Journal of Geophysical Research **99**(C5)

14. G. Evensen Ocean Dynamics **53**
15. G. Evensen, Control systems magazine **29**(3)
16. G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*, 2nd edition edition (Springer, New York, 2009)
17. A. L. Gibbs, F. E. Su, International Statistical Review **70**(3)
18. M. Girolami, B. Calderhead J. R. Statist. Soc. B **73**(2)
19. C. J. Geyer, in *Handbook of Markov Chain Monte Carlo*, ed. by S. Brooks, A. Gelman, G. J. Jones, X.-L. Meng. Handbooks of Modern Statistical Methods (CRC Press, Boca Raton, 2011), pp. 3-48
20. P. D. Hoff, *A First Course in Bayesian Statistical Methods* (Springer, New York, 2009)
21. M. A. Iglesias, K. J. H. Law, A. M. Stuart, Inverse Problems **29**(4)
22. M. A. Iglesias, K. J. H. Law, A. M. Stuart, Computational Geosciences **17**
23. B. Hofmann, B. Kaltenbacher, C. Pöschl, O. Scherzer, Inverse Problems **23**(3)
24. J. Kaipio, E. Somersalo, *Statistical and Computational Inverse Problems*, (Springer, New York, 2005)
25. O. Kallenberg, *Foundations of Modern Probaility*, (Springer, New York, 2002)
26. R. E. Kalman, Transactions of the AMSE, Journal of Basic Engineering **82**
27. K. J. H. Law, A. M. Stuart, Monthly Weather Review **140**
28. J. M. Lewis, S. Lakshmivarahan, S. Dhall *Dynamic Data Assimilation – A Least Squares Appoach* (Cambridge University Press, Cambridge, 2006)
29. J. Martin, L. C. Wilcox, C. Burstedde, O. Ghattas, SIAM Journal on Scientific Computing **34**(3)
30. S. Meyn, R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd edn. (Cambridge Univeristy Press, Cambridge, 2009)
31. I. Myrseth, H. Omre, in *Large-Scale Inverse Problems and Quantification of Uncertainty*, ed. by L. Biegler. Wiley Series in Computational Statistics (Wiley, Chichester, 2011), pp. 217-246
32. O. Pajonk, B. V. Rosić, A. Litvinenko, H. G. Matthies, Physica D **241**(7)
33. M. M. Rao, *Conditional measures and applications*, 2nd edn. (Chapman and Hall/CRC, 2010)
34. B. V. Rosić, A. Kučerová, J. Sýkora, O. Pajonk, A. Litvinenko, H. G. Matthies, Engineering Structures **50**
35. B. V. Rosić, A. Litvinenko, O. Pajonk, H. G. Matthies, Journal of Computational Physics **231**(17)
36. O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, *Variational Methods in Imaging*, (Springer, New York, 2009)
37. C. Schillings, Ch. Schwab, Inverse Problems **29**(6)
38. Ch. Schwab, A. M. Stuart, Inverse Problems **28**(4)
39. D. Simon, *Optimal state estimation: Kalman, H_∞, and nonlinear approaches*, (Wiley, Hoboken, 2006)
40. J. L. Speyer, W. H. Chung, *Stochastic processes, estimation, and control* (SIAM, Philadelphia, 2008)
41. A. S. Stordal, H. A. Karlsen, G. Nærvdal, H. J. Skaug, B. Vallès, Computational Geosciences **15**
42. A. M. Stuart, Acta Numerica, **19**
43. E. Ullmann, *Solution Strategies for Stochastic Finite Element Discretizations*, PhD thesis (TU Bergakademie Freiberg, 2008)
44. C. R. Vogel, *Computational Methods for Inverse Problems* (SIAM, Philadelphia, 2002)