Matthies *et al. Adv. Model. and Simul. in Eng. Sci.(2016)3:24*
DOI 10.1186/s40323-016-0075-7

Advanced Modeling and Simulation in Engineering Sciences

RESEARCH ARTICLE Open Access

Parameter estimation via conditional expectation: a Bayesian inversion

Hermann G. Matthies[1*], Elmar Zander[1], Bojana V. Rosić[1] and Alexander Litvinenko[2]

*Correspondence:
wire@tu-bs.de
[1]Institute of Scientific Computing, Technische Universität Braunschweig, Braunschweig, Germany
Full list of author information is available at the end of the article

Abstract

When a mathematical or computational model is used to analyse some system, it is usual that some parameters resp. functions or fields in the model are not known, and hence uncertain. These parametric quantities are then identified by actual observations of the response of the real system. In a probabilistic setting, Bayes's theory is the proper mathematical background for this identification process. The possibility of being able to compute a conditional expectation turns out to be crucial for this purpose. We show how this theoretical background can be used in an actual numerical procedure, and shortly discuss various numerical approximations.

Keywords: Inverse identification, Uncertainty quantification, Bayesian update, Parameter identification, Conditional expectation, Filters, Functional and spectral approximation

Background

The fitting of parameters resp. functions or fields—these will all be for the sake of brevity be referred to as parameters—in a mathematical computational model is usually denoted as an inverse problem, in contrast to predicting the output or state resp. response of the system given certain inputs, which is called the forward problem. In the inverse problem, the response of the model is compared to the response of the system. The system may be a real world system, or just another computational model—usually a more complex one. One then tries in various ways to match the model response with the system response.

Typical deterministic procedures include such methods as minimising the mean square error (MMSE), leading to optimisation problems in the search of optimal parameters. As the inverse problem is typically ill-posed—the observations do not contain enough information to uniquely determine the parameters—some additional information has to be added to select a unique solution. In the deterministic setting one then typically invokes additional ad-hoc procedures like Tikhonov-regularisation [3,4,28,29].

In a probabilistic setting (e.g. [10,27] and references therein) the ill-posed problem becomes well-posed (e.g. [26]). This is achieved at a cost, though. The unknown parameters are considered as uncertain, and modelled as random variables (RVs). The information added is hence the *prior* probability distribution. This means on one hand that the result of the identification is a probability distribution, and not a single value, and on the other hand the computational work may be increased substantially, as one has to deal with RVs.

Springer Open

That the result is a probability distribution may be seen as additional information though, as it offers an assessment of the residual uncertainty after the identification procedure, something which is not readily available in the deterministic setting. The probabilistic setting thus can be seen as modelling our knowledge about a certain situation—the value of the parameters—in the language of probability theory, and using the observation to update our knowledge, (i.e. the probabilistic description) by *conditioning* on the observation.

The key probabilistic background for this is Bayes's theorem in the formulation of Laplace [10,27]. It is well known that the Bayesian update is theoretically based on the notion of conditional expectation (CE) [1]. Here we take an approach which takes CE not only as a theoretical basis, but also as a basic computational tool. This may be seen as somewhat related to the "Bayes linear" approach [6,13], which has a linear approximation of CE as its basis, as will be explained later.

In many cases, for example when tracking a dynamical system, the updates are performed sequentially step-by-step, and for the next step one needs not only a probability distribution in order to perform the next step, but a random variable which may be evolved through the state equation. Methods on how to transform the prior RV into the one which is conditioned on the observation will be discussed as well [18]. This approach is very different to the very frequently used one which refers to Bayes's theorem in terms of densities and likelihood functions, and typically employs Markov-chain Monte Carlo (MCMC) methods to sample from the posterior (see e.g. [9,16,24]).

Mathematical set-up

Let us start with an example to have a concrete idea of what the whole procedure is about. Imagine a system described by a diffusion equation, e.g. the diffusion of heat through a solid medium, or even the seepage of groundwater through porous rocks and soil:

$$\frac{\partial \tilde{\upsilon}}{\partial t}(x,t) = \dot{\tilde{\upsilon}}(x,t) = \nabla \cdot (\kappa(x,\tilde{\upsilon})\nabla\tilde{\upsilon}(x,t)) + \eta(x,t), \tag{1}$$

$$\tilde{\upsilon}(x,0) = \tilde{\upsilon}_0(x) \quad \text{plus b.c.} \tag{2}$$

Here $x \in \mathcal{G}$ is a spatial coordinate in the domain $\mathcal{G} \subset \mathbb{R}^n$, $t \in [0,T]$ is the time, $\tilde{\upsilon}$ a scalar function describing the diffusing quantity, κ the (possibly non-linear) diffusion tensor, η external sources or sinks, and ∇ the Nabla operator. Additionally assume appropriate boundary conditions so that Eq. (1) is well-posed. Now, as often in such situations, imagine that we do not know the initial conditions $\tilde{\upsilon}_0$ in Eq. (2) precisely, nor the diffusion tensor κ, and maybe not even the driving source η, i.e. there is some uncertainty attached as to what their precise values are.

Data model

A more abstract setting which subsumes Eq. (1) is to view $\tilde{\upsilon}(t) := \tilde{\upsilon}(\cdot,t)$ as an element of a Hilbert-space (for the sake of simplicity) $\mathcal{V}$. In the particular case of Eq. (1) one could take $\mathcal{V} = \mathrm{H}^1_E(\mathcal{G})$, a closed subspace of the Sobolev space $\mathrm{H}^1(\mathcal{G})$ incorporating the essential boundary conditions. Hence we may view Eqs. (1) and (2) as an example of

$$\frac{\mathrm{d}\tilde{\upsilon}}{\mathrm{d}t}(t) = \dot{\tilde{\upsilon}}(t) = A_{\mathcal{V}}(q;\tilde{\upsilon}(t)) + \eta(q;t), \quad \tilde{\upsilon}(0) = \tilde{\upsilon}_0(q) \in \mathcal{V}, \quad t \in [0,T]. \tag{3}$$

Here $A_{\mathcal{V}} : \mathcal{Q} \times \mathcal{V} \to \mathcal{V}$ is a possibly non-linear operator in $\tilde{\upsilon} \in \mathcal{V}$, and $q \in \mathcal{Q}$ are the parameters (like κ, $\tilde{\upsilon}_0$, or η, which more accurately would be described as functions of q),

where we assume for simplicity again that $\mathcal{Q}$ is some Hilbert space. Both $A_{\mathcal{V}}$, $\tilde{v}_0$, and η could involve some noise, so that one may view Eq. (3) as an instance of a stochastic evolution equation. This is our model of the system generating the observed data, which we assume to be well-posed.

Hence assume further that we may observe a function $\hat{Y}(q;\tilde{v}(t))$ of the state $\tilde{v}(t)$ and the parameters q, i.e. $\hat{Y}:\mathcal{Q}\times\mathcal{V}\to\mathcal{Y}$, where we assume that $\mathcal{Y}$ is a Hilbert space. To make things simple, assume additionally that we observe $\hat{Y}(q;\tilde{v}(t))$ at regular time intervals $t_n = n\cdot 1t$, i.e. we observe $y_n = \hat{Y}(q;\tilde{v}_n)$, where $\tilde{v}_n := \tilde{v}(t_n)$. Denote the solution operator Υ of Eq. (3) as

$$\tilde{v}_{n+1} = \Upsilon(t_{n+1}, q, \tilde{v}_n, t_n, \eta), \tag{4}$$

advancing the solution from t_n to t_{n+1}. Hence we are observing

$$\hat{y}_{n+1} = \hat{h}(\hat{Y}(q;\Upsilon(t_{n+1}, q, \tilde{v}_n, t_n, \eta)), \nu_n), \tag{5}$$

where some noise ν_n—inaccuracy of the observation—has been included, and $\hat{h}$ is an appropriate observation operator. A simple example is the often assumed additive noise

$$\hat{h}(y,\nu) := y + S_{\mathcal{V}}(\tilde{v})\nu,$$

where ν is a random vector, and for each $\tilde{v}$, $S_{\mathcal{V}}(\tilde{v})$ is a bounded linear map to $\mathcal{Y}$.

Identification model

Now that the model generating the data has been described, it is the appropriate point to introduce the identification model. Similarly as before in Eq. (3), we have a model

$$\frac{\mathrm{d}u}{\mathrm{d}t}(t) = \dot{u}(t) = A(q;u(t)) + \eta(q;t), \quad u(0) = u_0(q) \in \mathcal{U},\ t\in[0,T], \tag{6}$$

which depends on the same parameters q as in Eq. (3), to be used for the identification, which we shall only write in its abstract from. Hence we assume that we can actually integrate Eq. (6) from t_n to t_{n+1} with its solution operator U

$$u_{n+1} = U(t_{n+1}, q, u_n, t_n, \eta). \tag{7}$$

Observe that the two spaces $\mathcal{V}$ in Eq. (3) and $\mathcal{U}$ in Eq. (6) are not the same, as in general we do not know $\tilde{v}\in\mathcal{V}$, we only have observations $y\in\mathcal{Y}$.

As later not only the state $u\in\mathcal{U}$ in Eq. (6) has to be identified, but also the parameters q, and the identification may happen sequentially, i.e. our estimate of q will change from step n to step $n+1$, we shall introduce an "extended" state vector $x = (u,q)\in\mathcal{X} := \mathcal{Q}\times\mathcal{U}$ and describe the change from n to $n+1$ by

$$x_{n+1} = (u_{n+1}, q_{n+1}) = \hat{f}(x_n) := (U(t_{n+1}, q_n, u_n, t_n, \eta), q_n). \tag{8}$$

Let us explicitly introduce a noise $w\in\mathcal{W}$ to cover the stochastic contribution or modelling errors between Eqs. (6) and (3), so that we set

$$x_{n+1} = f(x_n, w_n), \tag{9}$$

for example

$$f(x,w) = \hat{f}(x) + S_{\mathcal{W}}(x)w,$$

where $w\in\mathcal{W}$ is the random vector, and $S_{\mathcal{W}}(x)\in\mathscr{L}(\mathcal{W},\mathcal{X})$ is for each $x\in\mathcal{X}$ a bounded linear map from $\mathcal{W}$ to $\mathcal{X}$.

To deal with the extended state, we shall define the identification or predicted observation operator as

$$y_{n+1} = h(x_n, v_n) = H(x_{n+1}, v_n) = H(f(x_n, w_n), v_n), \tag{10}$$

where the noise v_n—the same as in Eq. (5), our model of the inaccuracy of the observation—has been included. A simple example with additive noise is

$$h(x_n, v_n) := Y(q; U(t_{n+1}, q_n, u_n, t_n, \eta)) + S_{\mathcal{V}}(x_n)v_n,$$

where $v \in \mathcal{V}$ is the random vector, and $S_{\mathcal{V}}(x) \in \mathscr{L}(\mathcal{V}, \mathcal{X})$ is for each $x \in \mathcal{X}$ a bounded linear map from $\mathcal{V}$ to $\mathcal{X}$. The mapping $Y : \mathcal{Q}\times\mathcal{U} \to \mathcal{Y}$ is similar to the map $\hat{Y} : \mathcal{Q}\times\mathcal{V} \to \mathcal{Y}$ in the "Data model" section, it predicts the "true" observation without noise v_n. Eq. (9) for the time evolution of the extended state and Eq. (10) for the observation are the basic building blocks for the identification.

Synopsis of Bayesian estimation

There are many accounts of this, and this synopsis is just for the convenience of the reader and to introduce notation. Otherwise we refer to e.g. [6,10,13,27], and in particular for the rôle of conditional expectation (CE) to our work [18,24].

The idea is that the observation $\hat{y}$ from Eq. (5) depends on the unknown parameters q, which ideally should equal y_n from Eq. (10), which in turn both directly and through the state $x = (u(q), q)$ in Eq. (9) depends also on the parameters q, should be equal, and any difference should give an indication on what the "true" value of q should be. The problem in general is—apart from the distracting errors w and v—that the mapping $q \mapsto y = Y(q; u(q))$ is in general not invertible, i.e. y does not contain information to uniquely determine q, or there are many q which give a good fit for $\hat{y}$. Therefore the *inverse* problem of determining q from observing $\hat{y}$ is termed an *ill-posed* problem.

The situation is a bit comparable to Plato's allegory of the cave, where Socrates compares the process of gaining knowledge with looking at the shadows of the real things. The observations $\hat{y}$ are the "shadows" of the "real" things q and $\tilde{v}$, and from observing the "shadows" $\hat{y}$ we want to infer what "reality" is, in a way turning our heads towards it. We hence want to "free" ourselves from just observing the "shadows" and gain some understanding of "reality".

One way to deal with this difficulty is to measure the difference between observed $\hat{y}_n$ and predicted system output y_n and try to find parameters q_n such that this difference is minimised. Frequently it may happen that the parameters which realise the minimum are not unique. In case one wants a unique parameter, a choice has to be made, usually by demanding additionally that some norm or similar functional of the parameters is small as well, i.e. some regularity is enforced. This optimisation approach hence leads to regularisation procedures [3,4,28,29].

Here we take the view that our lack of knowledge or uncertainty of the actual value of the parameters can be described in a *Bayesian* way through a probabilistic model [10,27]. The unknown parameter q is then modelled as a random variable (RV)—also called the *prior* model—and additional information on the system through measurement or observation changes the probabilistic description to the so-called *posterior* model. The second approach is thus a method to update the probabilistic description in such a way as to take account of the additional information, and the updated probabilistic descrip-

tion *is* the parameter estimate, including a probabilistic description of the remaining uncertainty.

It is well-known that such a Bayesian update is in fact closely related to *conditional expectation* [1,6,10,18,24], and this will be the basis of the method presented. For these and other probabilistic notions see for example [22] and the references therein. As the Bayesian update may be numerically very demanding, we show computational procedures to accelerate this update through methods based on *functional approximation* or *spectral representation* of stochastic problems [17,18]. These approximations are in the simplest case known as Wiener's so-called *homogeneous* or *polynomial chaos* expansion, which are polynomials in independent Gaussian RVs—the "chaos"—and which can also be used numerically in a Galerkin procedure [17,18].

Although the Gauss-Markov theorem and its extensions [15] are well-known, as well as its connections to the Kalman filter [7,11]—see also the recent Monte Carlo or *ensemble* version [5]—the connection to Bayes's theorem is not often appreciated, and is sketched here. This turns out to be a linearised version of *conditional expectation*.

Since the parameters of the model to be estimated are uncertain, all relevant information may be obtained via their stochastic description. In order to extract information from the posterior, most estimates take the form of expectations w.r.t. the posterior, i.e. a conditional expectation (CE). These expectations—mathematically integrals, numerically to be evaluated by some quadrature rule—may be computed via asymptotic, deterministic, or sampling methods by typically computing first the posterior density. As we will see, the posterior density does not always exist [23]. Here we follow our recent publications [18,21,24] and introduce a novel approach, namely computing the CE directly and not via the posterior density [18]. This way all relevant information from the conditioning may be computed directly. In addition, we want to change the prior, represented by a random variable (RV), into a new random variable which has the correct posterior distribution. We will discuss how this may be achieved, and what approximations one may employ in the computation.

To be a bit more formal, assume that the uncertain parameters are given by

$$x : \Omega \to \mathcal{X} \text{ as a RV on a probability space } (\Omega, \mathfrak{A}, \mathbb{P}), \tag{11}$$

where the set of elementary events is Ω, $\mathfrak{A}$ a σ-algebra of measurable events, and $\mathbb{P}$ a probability measure. The *expectation* corresponding to $\mathbb{P}$ will be denoted by $\mathbb{E}\,()$, e.g.

$$\bar{\Psi} := \mathbb{E}\,(\Psi) := \int_{\Omega} \Psi(x(\omega))\,\mathbb{P}(\mathrm{d}\omega),$$

for any measurable function Ψ of x.

Modelling our lack-of-knowledge about q in a Bayesian way [6,10,27] by replacing them with random variables (RVs), the problem becomes well-posed [26]. But of course one is looking now at the problem of finding a probability distribution that best fits the data; and one also obtains a probability distribution, not just *one* value q. Here we focus on the use of procedures similar to a linear Bayesian approach [6] in the framework of "white noise" analysis.

As formally q is a RV, so is the state x_n of Eq. (9), reflecting the uncertainty about the parameters and state of Eq. (3). From this follows that also the prediction of the measurement y_n Eq. (10) is a RV; i.e. we have a *probabilistic* description of the measurement.

The theorem of Bayes and Laplace

Bayes original statement of the theorem which today bears his name was only for a very special case. The form which we know today is due to Laplace, and it is a statement about conditional probabilities. A good account of the history may be found in [19].

Bayes's theorem is commonly accepted as a consistent way to incorporate new knowledge into a probabilistic description [10,27]. The elementary textbook statement of the theorem is about conditional probabilities

$$\mathbb{P}(\mathcal{I}_x|\mathcal{M}_y) = \frac{\mathbb{P}(\mathcal{M}_y|\mathcal{I}_x)}{\mathbb{P}(\mathcal{M}_y)}\mathbb{P}(\mathcal{I}_x), \quad \text{if } \mathbb{P}(\mathcal{M}_y) > 0, \tag{12}$$

where $\mathcal{I}_x \subset \mathcal{X}$ is some subset of possible x's on which we would like to gain some information, and $\mathcal{M}_y \subset \mathcal{Y}$ is the information provided by the measurement. The term $\mathbb{P}(\mathcal{I}_x)$ is the so-called *prior*, it is what we know before the observation $\mathcal{M}_y$. The quantity $\mathbb{P}(\mathcal{M}_y|\mathcal{I}_x)$ is the so-called *likelihood*, the conditional probability of $\mathcal{M}_y$ assuming that $\mathcal{I}_x$ is given. The term $\mathbb{P}(\mathcal{M}_y)$ is the so called *evidence*, the probability of observing $\mathcal{M}_y$ in the first place, which sometimes can be expanded with the *law of total probability*, allowing to choose between different models of explanation. It is necessary to make the right hand side of Eq. (12) into a real probability—summing to unity—and hence the term $\mathbb{P}(\mathcal{I}_x|\mathcal{M}_y)$, the *posterior* reflects our knowledge on $\mathcal{I}_x$ *after* observing $\mathcal{M}_y$. The quotient $\mathbb{P}(\mathcal{M}_y|\mathcal{I}_x)/\mathbb{P}(\mathcal{M}_y)$ is sometimes termed the *Bayes* factor, as it reflects the relative change in probability after observing $\mathcal{M}_y$.

This statement Eq. (12) runs into problems if the set observations $\mathcal{M}_y$ has vanishing measure, $\mathbb{P}(\mathcal{M}_y) = 0$, as is the case when we observe *continuous* random variables, and the theorem would have to be formulated in *densities*, or more precisely in probability density functions (pdfs). But the Bayes factor then has the indeterminate form $0/0$, and some form of limiting procedure is needed. As a sign that this is not so simple—there are different and inequivalent forms of doing it—one may just point to the so-called *Borel-Kolmogorov* paradox. See [23] for a thorough discussion.

There is one special case where something resembling Eq. (12) may be achieved with pdfs, namely if y and x have a *joint* pdf $\pi_{y,x}(y,x)$. As y is essentially a function of x, this is a special case depending on conditions on the error term v. In this case Eq. (12) may be formulated as

$$\pi_{x|y}(x|y) = \frac{\pi_{y,x}(y,x)}{Z_s(y)}, \tag{13}$$

where $\pi_{x|y}(x|y)$ is the *conditional* pdf, and the "evidence" Z_s (from German *Zustandssumme* (sum of states), a term used in physics) is a normalising factor such that the conditional pdf $\pi_{x|y}(\cdot|y)$ integrates to unity

$$Z_s(y) = \int_\Omega \pi_{y,x}(y, x(\omega))\, \mathbb{P}(\mathrm{d}\omega).$$

The joint pdf may be split into the *likelihood density* $\pi_{y|x}(y|x)$ and the *prior* pdf $\pi_x(x)$

$$\pi_{y,x}(y,x) = \pi_{y|x}(y|x)\pi_x(x),$$

so that Eq. (13) has its familiar form ([27] Ch. 1.5)

$$\pi_{x|y}(x|y) = \frac{\pi_{y|x}(y|x)}{Z_s(y)}\pi_x(x). \tag{14}$$

These terms are in direct correspondence with those in Eq. (12) and carry the same names. Once one has the conditional measure $\mathbb{P}(\cdot|\mathcal{M}_y)$ or even a conditional pdf $\pi_{x|y}(\cdot|y)$, the

conditional expectation (CE) $\mathbb{E}\left(\cdot|\mathcal{M}_y\right)$ may be defined as an integral over that conditional measure resp. the conditional pdf. Thus classically, the conditional measure or pdf implies the conditional expectation:

$$\mathbb{E}\left(\Psi|\mathcal{M}_y\right) := \int_{\mathcal{X}} \Psi(x)\,\mathbb{P}(\mathrm{d}x|\mathcal{M}_y)$$

for any measurable function Ψ of x.

Please observe that the model for the RV representing the error $\nu(\omega)$ determines the likelihood functions $\mathbb{P}(\mathcal{M}_y|\mathcal{I}_x)$ resp. the existence and form of the likelihood density $\pi_{y|x}(\cdot|x)$. In computations, it is here that the computational model Eqs. (6) and (10) is needed to predict the measurement RV y given the state and parameters x as a RV.

Most computational approaches determine the pdfs, but we will later argue that it may be advantageous to work directly with RVs, and not with conditional probabilities or pdfs. To this end, the concept of conditional expectation (CE) and its relation to Bayes's theorem is needed [1].

Conditional expectation

To avoid the difficulties with conditional probabilities like in the Borel-Kolmogorov paradox alluded to in the "The theorem of Bayes and Laplace" section, *Kolmogorov* himself—when he was setting up the axioms of the mathematical theory probability—turned the relation between conditional probability or pdf and conditional expectation around, and defined as a first and fundamental notion *conditional expectation* [1,23].

It has to be defined not with respect to measure-zero observations of a RV y, but w.r.t sub-σ-algebras $\mathfrak{B} \subset \mathfrak{A}$ of the underlying σ-algebra $\mathfrak{A}$. The σ-algebra may be loosely seen as the collection of subsets of Ω on which we can make statements about their probability, and for fundamental mathematical reasons in many cases this is *not* the set of *all* subsets of Ω. The sub-σ-algebra $\mathfrak{B}$ may be seen as the sets on which we learn something through the observation.

The simplest—although slightly restricted—way to define the conditional expectation [1] is to just consider RVs with *finite variance*, i.e. the Hilbert-space

$$\mathcal{S} := \mathrm{L}_2(\Omega, \mathfrak{A}, \mathbb{P}) := \{r : \Omega \to \mathbb{R} \ : \ r \text{ measurable w.r.t. } \mathfrak{A}, \mathbb{E}\left(|r|^2\right) < \infty\}.$$

If $\mathfrak{B} \subset \mathfrak{A}$ is a sub-σ-algebra, the space

$$\mathcal{S}_{\mathfrak{B}} := \mathrm{L}_2(\Omega, \mathfrak{B}, \mathbb{P}) := \{r : \Omega \to \mathbb{R} \ : \ r \text{ measurable w.r.t. } \mathfrak{B}, \mathbb{E}\left(|r|^2\right) < \infty\} \subset \mathcal{S}$$

is a *closed* subspace, and hence has a well-defined continuous orthogonal projection $P_{\mathfrak{B}} : \mathcal{S} \to \mathcal{S}_{\mathfrak{B}}$. The *conditional expectation* (CE) of a RV $r \in \mathcal{S}$ w.r.t. a sub-σ-algebra $\mathfrak{B}$ is then defined as that orthogonal projection

$$\mathbb{E}\left(r|\mathfrak{B}\right) := P_{\mathfrak{B}}(r) \in \mathcal{S}_{\mathfrak{B}}. \tag{15}$$

It can be shown [1] to coincide with the classical notion when that one is defined, and the *unconditional* expectation $\mathbb{E}\,()$ is in this view just the CE w.r.t. the minimal σ-algebra $\mathfrak{B} = \{\emptyset, \Omega\}$. As the CE is an orthogonal projection, it minimises the squared error

$$\mathbb{E}\left(|r - \mathbb{E}\left(r|\mathfrak{B}\right)|^2\right) = \min\{\mathbb{E}\left(|r - \tilde{r}|^2\right) \ : \ \tilde{r} \in \mathcal{S}_{\mathfrak{B}}\}, \tag{16}$$

from which one obtains the *variational equation* or orthogonality relation

$$\forall \tilde{r} \in \mathcal{S}_{\mathfrak{B}} : \mathbb{E}\left(\tilde{r}(r - \mathbb{E}\left(r|\mathfrak{B}\right))\right) = 0; \tag{17}$$

and one has a form of *Pythagoras's* theorem

$$\mathbb{E}\left(|r|^2\right) = \mathbb{E}\left(|r - \mathbb{E}\left(r|\mathfrak{B}\right)|^2\right) + \mathbb{E}\left(|\mathbb{E}\left(r|\mathfrak{B}\right)|^2\right).$$

The CE is therefore a form of a *minimum mean square error* (MMSE) estimator.

Given the CE, one may completely characterise the *conditional* probability, e.g. for $A \subset \Omega, A \in \mathfrak{B}$ by

$$\mathbb{P}(A|\mathfrak{B}) := \mathbb{E}\left(\chi_A|\mathfrak{B}\right),$$

where χ_A is the RV which is unity iff $\omega \in A$ and vanishes otherwise—the *usual* characteristic function, sometimes also termed an indicator function. Thus if we know $\mathbb{P}(A|\mathfrak{B})$ for each $A \in \mathfrak{B}$, we know the conditional probability. Hence having the CE $\mathbb{E}\left(\cdot|\mathfrak{B}\right)$ allows one to know everything about the conditional probability; the conditional or posterior density is not needed. If the prior probability was the distribution of some RV r, we know that it is completely characterised by the *prior* characteristic function—in the sense of probability theory—$\varphi_r(s) := \mathbb{E}\left(\exp(\mathrm{i}rs)\right)$. To get the *conditional* characteristic function $\varphi_{r|\mathfrak{B}}(s) = \mathbb{E}\left(\exp(\mathrm{i}rs)|\mathfrak{B}\right)$, all one has to do is use the CE instead of the unconditional expectation. This then completely characterises the conditional distribution.

In our case of an observation of a RV y, the sub-σ-algebra $\mathfrak{B}$ will be the one generated by the *observation* $y = h(x, \nu)$, i.e. $\mathfrak{B} = \sigma(y)$, these are those subsets of Ω on which we may obtain *information* from the observation. According to the *Doob-Dynkin* lemma the subspace $\mathcal{S}_{\sigma(y)}$ is given by

$$\mathcal{S}_{\sigma(y)} := \{r \in \mathcal{S} \ : \ r(\omega) = \phi(y(\omega)), \phi \text{ measurable}\} \subset \mathcal{S}, \tag{18}$$

i.e. functions of the observation. This means intuitively that anything we learn from an observation is a function of the observation, and the subspace $\mathcal{S}_{\sigma(y)} \subset \mathcal{S}$ is where the information from the measurement is lying.

Observe that the CE $\mathbb{E}\left(r|\sigma(y)\right)$ and conditional probability $\mathbb{P}(A|\sigma(y))$—which we will abbreviate to $\mathbb{E}\left(r|y\right)$, and similarly $\mathbb{P}(A|\sigma(y)) = \mathbb{P}(A|y)$—is a RV, as y is a RV. Once an observation has been made, i.e. we observe for the RV y the fixed value $\hat{y} \in \mathcal{Y}$, then—for almost all $\hat{y} \in \mathcal{Y}$—$\mathbb{E}\left(r|\hat{y}\right) \in \mathbb{R}$ is just a number—the *posterior expectation*, and $\mathbb{P}(A|\hat{y}) = \mathbb{E}\left(\chi_A|\hat{y}\right)$ is the *posterior probability*. Often these are also termed conditional expectation and conditional probability, which leads to confusion. We therefore prefer the attribute *posterior* when the actual observation $\hat{y}$ has been observed and inserted in the expressions. Additionally, from Eq. (18) one knows that for some function ϕ_r—for each RV r it is a possibly different function—one has that

$$\mathbb{E}\left(r|y\right) = \phi_r(y) \quad \text{and} \quad \mathbb{E}\left(r|\hat{y}\right) = \phi_r(\hat{y}) \tag{19}$$

In relation to Bayes's theorem, one may conclude that if it is possible to compute the CE w.r.t. an observation y or rather the posterior expectation, then the conditional and especially the posterior probabilities after the observation $\hat{y}$ may as well be computed, regardless whether joint pdfs exist or not. We take this as the starting point to Bayesian estimation.

The conditional expectation has been formulated for scalar RVs, but it is clear that the notion carries through to vector-valued RVs in a straightforward manner, formally by seeing a—let us say—$\mathcal{Y}$-valued RV as an element of the tensor Hilbert space $\mathscr{Y} = \mathcal{Y} \otimes \mathcal{S}$ [8], as

$$\mathscr{Y} = \mathcal{Y} \otimes \mathcal{S} \cong \mathrm{L}_2(\Omega, \mathfrak{A}, \mathbb{P}; \mathcal{Y}),$$

the RVs in $\mathcal{Y}$ with finite *total* variance

$$|||\tilde{y}|||^2_{\mathscr{Y}} = \int_\Omega \|\tilde{y}(\omega)\|^2_{\mathcal{Y}}\, \mathbb{P}(\mathrm{d}\omega) < \infty.$$

Here $\|\tilde{y}(\omega)\|^2_{\mathcal{Y}} = \langle \tilde{y}(\omega), \tilde{y}(\omega)\rangle_{\mathcal{Y}}$ is the norm squared on the deterministic component $\mathcal{Y}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{Y}}$; and the total L_2-norm of an elementary tensor $y \otimes r \in \mathcal{Y} \otimes \mathcal{S}$ with $y \in \mathcal{Y}$ and $r \in \mathcal{S}$ can also be written as

$$|||y \otimes r|||^2_{\mathscr{Y}} = \langle\!\langle y \otimes r, y \otimes r\rangle\!\rangle_{\mathscr{Y}} = \|y\|^2_{\mathcal{Y}} \|r\|^2_{\mathcal{S}} = \langle y, y\rangle_{\mathcal{Y}} \langle r, r\rangle_{\mathcal{S}},$$

where $\langle r, r\rangle_{\mathcal{S}} = \|r\|^2_{\mathcal{S}} := \mathbb{E}\left(|r|^2\right)$ is the usual inner product of scalar RVs.

The CE on $\mathscr{Y}$ is then formally given by $\mathbb{E}_{\mathscr{Y}}(\cdot|\mathfrak{B}) := I_{\mathcal{Y}} \otimes \mathbb{E}(\cdot|\mathfrak{B})$, where $I_{\mathcal{Y}}$ is the identity operator on $\mathcal{Y}$. This means that for an elementary tensor $y \otimes r \in \mathcal{Y} \otimes \mathcal{S}$ one has

$$\mathbb{E}_{\mathscr{Y}}(y \otimes r|\mathfrak{B}) = y \otimes \mathbb{E}(r|\mathfrak{B}).$$

The vector valued conditional expectation

$$\mathbb{E}_{\mathscr{Y}}(\cdot|\mathfrak{B}) = I_{\mathcal{Y}} \otimes \mathbb{E}(\cdot|\mathfrak{B}) : \mathscr{Y} = \mathcal{Y} \otimes \mathcal{S} \to \mathcal{Y}$$

is also an orthogonal projection, but in $\mathscr{Y}$, for simplicity also denoted by $\mathbb{E}(\cdot|\mathfrak{B}) = P_{\mathfrak{B}}$ when there is no possibility of confusion.

Constructing a posterior random variable

We recall the equations governing our model Eqs. (9) and (10), and interpret them now as equations acting on RVs, i.e. for $\omega \in \Omega$:

$$\hat{x}_{n+1}(\omega) = f(x_n(\omega), w_n(\omega)), \tag{20}$$

$$y_{n+1}(\omega) = h(x_n(\omega), v_n(\omega)), \tag{21}$$

where one may now see the mappings $f : \mathscr{X} \times \mathscr{W} \to \mathscr{X}$ and $h : \mathscr{X} \times \mathscr{V} \to \mathscr{Y}$ acting on the tensor Hilbert spaces of RVs with finite variance, e.g. $\mathscr{Y} := \mathcal{Y} \otimes \mathcal{S}$ with the inner product as explained in "Conditional expectation" section; and similarly for $\mathscr{X} := \mathcal{X} \otimes \mathcal{S}$ resp. $\mathscr{W}$ and $\mathscr{V}$.

Updating random variables

We now focus on the step from time t_n to t_{n+1}. Knowing the RV $x_n \in \mathscr{X}$, one predicts the new state $\hat{x}_{n+1} \in \mathscr{X}$ and the measurement $y_{n+1} \in \mathscr{Y}$. With the CE operator from the measurement prediction y_{n+1} in Eq. (21)

$$\mathbb{E}(\Psi(x_{n+1})|\sigma(y_{n+1})) = \phi_\Psi(y_{n+1}), \tag{22}$$

and the actual observation $\hat{y}_{n+1}$ one may then compute the *posterior* expectation operator

$$\mathbb{E}(\Psi(x_{n+1})|\hat{y}_{n+1}) = \phi_\Psi(\hat{y}_{n+1}). \tag{23}$$

This has all the information about the posterior probability.

But to then go on from t_{n+1} to t_{n+2} with the Eqs. (20) and (21), one needs a new RV x_{n+2} which has the posterior distribution described by the mappings $\phi_\Psi(\hat{y}_{n+1})$ in Eq. (23). Bayes's theorem only specifies this probabilistic content. There are many RVs which have this posterior distribution, and we have to pick a particular representative to continue the computation. We will show a method which in the simplest case comes back to MMSE.

Here it is proposed to construct this new RV x_{n+1} from the predicted $\hat{x}_{n+1}$ in Eq. (20) with a mapping, starting from very simple ones and getting ever more complex. For the sake of brevity of notation, the forecast RV will be called $x_f = \hat{x}_{n+1}$, and the forecast measurement $y_f = y_{n+1}$, and we will denote the measurement just by $\hat{y} = \hat{y}_{n+1}$. The RV x_{n+1} we want to construct will be called the *assimilated* RV $x_a = x_{n+1}$—it has assimilated the new observation $\hat{y} = \hat{y}_{n+1}$. Hence what we want is a new RV which is an *update* of the forecast RV x_f

$$x_a = B(x_f, y_f, \hat{y}) = x_f + \Xi(x_f, y_f, \hat{y}), \tag{24}$$

with a Bayesian update map B resp. a change given by the *innovation* map Ξ. Such a transformation is often called a *filter*—the measurement $\hat{y}$ is filtered to produce the update.

Correcting the mean

We take first the task to give the new RV the correct posterior *mean* $\bar{x}_a = \mathbb{E}(x_a|\hat{y})$, i.e. we take $\Psi(x) = x$ in Eq. (23). Remember that according to Eq. (15) $\mathbb{E}\left(x_a|\sigma(y_f)\right) = \phi_{x_f}(y_f) =: \phi_x(y_f)$ is an orthogonal projection $P_{\sigma(y_f)}(x_f)$ from $\mathscr{X} = \mathcal{X} \otimes \mathcal{S}$ onto $\mathscr{X}_\infty := \mathcal{X} \otimes \mathcal{S}_\infty$, where $\mathcal{S}_\infty := \mathcal{S}_{\sigma(y)} = \mathrm{L}_2(\Omega, \sigma(y_f), \mathbb{P})$. Hence there is an orthogonal decomposition

$$\mathscr{X} = \mathcal{X} \otimes \mathcal{S} = \mathscr{X}_\infty \oplus \mathscr{X}_\infty^\perp = (\mathcal{X} \otimes \mathcal{S}_\infty) \oplus (\mathcal{X} \otimes \mathcal{S}_\infty^\perp), \tag{25}$$

$$x_f = P_{\sigma(y_f)}(x_f) + (I_{\mathscr{X}} - P_{\sigma(y_f)})(x_f) = \phi_x(y_f) + (x_f - \phi_x(y_f)). \tag{26}$$

As $P_{\sigma(y_f)} = \mathbb{E}\left(\cdot|\sigma(y_f)\right)$ is a projection, one sees from Eq. (26) that the second term has vanishing CE for any measurement $\hat{y}$:

$$\mathbb{E}\left(x_f - \phi_x(y_f)|\sigma(y_f)\right) = P_{\sigma(y_f)}(I_{\mathscr{X}} - P_{\sigma(y_f)})(x_f) = 0. \tag{27}$$

One may view this also in the following way: From the measurement y_a resp. $\hat{y}$ we only learn something about the subspace $\mathscr{X}_\infty$. Hence when the measurement comes, we change the decomposition Eq. (26) by only fixing the component $\phi_x(y_f) \in \mathscr{X}_\infty$, and leaving the orthogonal rest unchanged:

$$x_{a,1} = \phi_x(\hat{y}) + (x_f - \phi_x(y_f)) = x_f + (\phi_x(\hat{y}) - \phi_x(y_f)). \tag{28}$$

Observe that this is just a linear *translation* of the RV x_f, i.e. a very simple map B in Eq. (24). From Eq. (27) follows that

$$\bar{x}_{a,1} = \mathbb{E}(x_{a,1}|\hat{y}) = \phi_x(\hat{y}) = \mathbb{E}(x_a|\hat{y}),$$

hence the RV $x_{a,1}$ from Eq. (28) has the *correct* posterior mean.

Observe that according to Eq. (27) the term $x_\perp := (x_f - \phi_x(y_f))$ in Eq. (28) is a zero mean RV, hence the covariance and total variance of $x_{a,1}$ is given by

$$\mathrm{cov}(x_{a,1}) = \mathbb{E}(x_\perp \otimes x_\perp) = \mathbb{E}\left(x_\perp^{\otimes 2}\right) =: C_1, \tag{29}$$

$$\mathrm{var}(x_{a,1}) = \mathbb{E}\left(\|x_\perp(\omega)\|_{\mathcal{X}}^2\right) = \mathrm{tr}(\mathrm{cov}(x_{a,1})). \tag{30}$$

Correcting higher moments

Here let us just describe two small additional steps: we take $\Psi(x) = \|x - \phi_x(\hat{y})\|_{\mathcal{X}}^2$ in Eq. (23), and hence obtain the total posterior variance as

$$\mathrm{var}(x_a) = \mathbb{E}\left(\|x_f - \phi_x(y_f)\|_{\mathcal{X}}^2|\hat{y}\right) = \phi_{x-\bar{x}}(\hat{y}). \tag{31}$$

Now it is easy to correct Eq. (28) to obtain

$$x_{a,t} = \phi_x(\hat{y}) + \frac{\mathrm{var}(x_a)}{\mathrm{var}(x_{a,1})}(x_f - \phi_x(y_f)), \tag{32}$$

a RV which has the correct posterior mean *and* the correct posterior total variance

$$\mathrm{var}(x_{a,t}) = \mathrm{var}(x_a).$$

Observe that this is just a linear translation and partial scaling of the RV x_f, i.e. still a very simple map B in Eq. (24).

With more computational effort, one may choose $\Psi(x) = (x - \phi_x(\hat{y}))^{\otimes 2}$ in Eq. (23), to obtain the covariance of x_a:

$$\mathrm{cov}(x_a) = \mathbb{E}\left((x - \phi_x(\hat{y}))^{\otimes 2}|\hat{y}\right) = \phi_{\otimes 2}(\hat{y}) =: C_a. \tag{33}$$

Instead of just scaling the RV as in Eq. (32), one may now take

$$x_{a,2} = \phi_x(\hat{y}) + B_a {B_1}^{-1}(x_f - \phi_x(y_f)), \tag{34}$$

where B_1 is any operator "square root" that satisfies $B_1 {B_1}^* = C_1$ in Eq. (29), and similarly $B_a {B_a}^* = C_a$ in Eq. (33). One possibility is the real square root—as C_1 and C_a are positive definite—$B_1 = C_1^{1/2}$, but computationally a Cholesky factor is usually cheaper. In any case, no matter which "square root" is chosen, the RV $x_{a,2}$ in Eq. (34) has the correct posterior mean *and* the correct posterior covariance. Observe that this is just an affine transformation of the RV x_f, i.e. still a fairly simple map B in Eq. (24).

By combining further transport maps [20] it seems possible to construct a RV x_a which has the desired posterior distribution to any accuracy. This is beyond the scope of the present paper, and is ongoing work on how to do it in the simplest way. For the following, we shall be content with the update Eq. (28) in "Correcting the mean" section.

The Gauss-Markov-Kalman filter (GMKF)

It turned out that practical computations in the context of Bayesian estimation can be extremely demanding, see [19] for an account of the history of Bayesian theory, and the break-throughs required in computational procedures to make Bayesian estimation possible at all for practical purposes. This involves both the Monte Carlo (MC) method and the Markov chain Monte Carlo (MCMC) sampling procedure. One may have gleaned this also already from the "Constructing a posterior random variable" section.

To arrive at computationally feasible procedures for computationally demanding models Eqs. (20) and (21), where MCMC methods are not feasible, approximations are necessary. This means in some way not using all information but having a simpler computation. Incidentally, this connects with the Gauss-Markov theorem [15] and the Kalman filter (KF) [7,11]. These were initially stated and developed without any reference to Bayes's theorem. The Monte Carlo (MC) computational implementation of this is the *ensemble* KF (EnKF) [5]. We will in contrast use a white noise or polynomial chaos approximation [18,21,24]. But the initial ideas leading to the abstract Gauss-Markov-Kalman filter (GMKF) are independent of any computational implementation and are presented first. It is in an abstract way just *orthogonal projection*, based on the update Eq. (28) in "Correcting the mean" section.

Building the filter

Recalling Eqs. (20) and (21) together with Eq. (28), the algorithm for forecasting and assimilating with just the posterior mean looks like

$$\begin{aligned}
\hat{x}_{n+1}(\omega) &= f(x_n(\omega), w_n(\omega)),\\
y_{n+1}(\omega) &= H(f(x_n(\omega), w_n(\omega)), v_n(\omega)),\\
x_{n+1}(\omega) &= \hat{x}_{n+1}(\omega) + (\phi_x(\hat{y}_{n+1}) - \phi_x(y_{n+1}(\omega))).
\end{aligned}$$

For simplicity of notation the argument ω will be suppressed. Also it will turn out that the mapping ϕ_x representing the CE can in most cases only be computed approximately, so we want to look at update algorithms with a general map $g : \mathcal{Y} \to \mathcal{X}$ to possibly approximate ϕ_x:

$$\begin{aligned}
x_{n+1} &= f(x_n, w_n) + (g(\hat{y}_{n+1}) - g(H(f(x_n, w_n), v_n)))\\
&= f(x_n, w_n) - g(H(f(x_n, w_n), v_n)) + g(\hat{y}_{n+1}),
\end{aligned} \tag{35}$$

where the first two equations have been inserted into the last. This is the filter equation for tracking and identifying the extended state of Eq. (20). One may observe that the normal evolution model Eq. (20) is corrected by the innovation term. This is the *best unbiased* filter, with $\phi(\hat{y})$ a MMSE estimate. It is clear that the *stability* of the solution to Eq. (35) will depend on the contraction properties or otherwise of the map $f - g \circ H \circ f = (I - g \circ H) \circ f$ as applied to x_n, but that is not completely worked out yet and beyond the scope of this paper.

By combining the minimisation property Eq. (16) and the Doob-Dynkin lemma Eq. (18), we see that the map ϕ_Ψ is defined by

$$\|\Psi(x) - \phi_\Psi(y)\|^2_{\mathcal{X}} = \min_{\varpi} \|\Psi(x) - \varpi(y)\|^2_{\mathcal{X}} = \min_{z \in \mathcal{X}_\infty} \|\Psi(x) - z\|^2_{\mathcal{X}}, \tag{36}$$

where ϖ ranges over all measurable maps $\varpi : \mathcal{Y} \to \mathcal{X}$. As $\mathcal{X}_{\sigma(y)} = \mathcal{X}_\infty$ is $\mathcal{L}$-closed [2,18], it is characterised similarly to Eq. (17), but by orthogonality in the $\mathcal{L}$-invariant sense

$$\forall z \in \mathcal{X}_\infty : \quad \mathbb{E}\,(z \otimes (\Psi(x) - \phi_\Psi(y))) = 0, \tag{37}$$

i.e. the RV $(\Psi(x) - \varpi(y))$ is orthogonal in the $\mathcal{L}$-invariant sense to all RVs $z \in \mathcal{X}_\infty$, which means its correlation operator vanishes. Although the CE $\mathbb{E}\,(x|y) = P_{\sigma(y)}(x)$ is an orthogonal projection, as the measurement operator Y, resp. h or H, which evaluates y, is not necessarily linear in x, hence the optimal map $\phi_x(y)$ is also not necessarily linear in y. In some sense it has to be the opposite of Y.

The linear filter

The minimisation in Eq. (36) over all measurable maps is still a formidable task, and typically only feasible in an approximate way. One problem of course is that the space $\mathcal{X}_\infty$ is in general infinite-dimensional. The standard Galerkin approach is then to approximate it by finite-dimensional subspaces, see [18] for a general description and analysis of the Galerkin convergence.

Here we replace $\mathcal{X}_\infty$ by much smaller subspace; and we choose in some way the simplest possible one

$$\mathcal{X}_1 = \{z \,:\, z = \Phi(y) = L(y(\omega)) + b,\ L \in \mathscr{L}(\mathcal{Y}, \mathcal{X}),\ b \in \mathcal{X}\} \subset \mathcal{X}_\infty \subset \mathcal{X}, \tag{38}$$

where the Φ are just *affine* maps; they are certainly measurable. Note that $\mathscr{X}_1$ is also an $\mathcal{L}$-invariant subspace of $\mathscr{X}_\infty \subset \mathscr{X}$.

Note that also other, possibly larger, $\mathcal{L}$-invariant subspaces of $\mathscr{X}_\infty$ can be used, but this seems to be smallest useful one. Now the minimisation Eq. (36) may be replaced by

$$\|x - (K(y) + a)\|^2_{\mathscr{X}} = \min_{L,b} \|x - (L(y) + b)\|^2_{\mathscr{X}}, \tag{39}$$

and the optimal affine map is defined by $K \in \mathscr{L}(\mathcal{Y}, \mathcal{X})$ and $a \in \mathcal{X}$.

Using this $g(y) = K(y) + a$, one disregards some information as $\mathscr{X}_1 \subset \mathscr{X}_\infty$ is usually a true subspace—observe that the subspace represents the information we may learn from the measurement—but the computation is easier, and one arrives in lieu of Eq. (28) at

$$x_{a,1L} = x_f + (K(\hat{y}) - K(y)) = x_f + K(\hat{y} - y). \tag{40}$$

This is the *best linear* filter, with the linear MMSE $K(\hat{y})$. One may note that the constant term a in Eq. (39) drops out in the filter equation.

The algorithm corresponding to Eq. (35) is then

$$\begin{aligned} x_{n+1} &= f(x_n, w_n) + K((\hat{y}_{n+1}) - H(f(x_n, w_n), v_n)) \\ &= f(x_n, w_n) - K(H(f(x_n, w_n), v_n)) + K(\hat{y}_{n+1}). \end{aligned} \tag{41}$$

The Gauss-Markov theorem and the Kalman filter

The optimisation described in Eq. (39) is a familiar one, it is easily solved, and the solution is given by an extension of the *Gauss-Markov* theorem [15]. The same idea of a linear MMSE is behind the *Kalman* filter [5–7,11,22]. In our context it reads

Theorem 1 *The solution to Eq.* (39)*, minimising*

$$\|x - (K(y) + a)\|^2_{\mathscr{X}} = \min_{L,b} \|x - (L(y) + b)\|^2_{\mathscr{X}}$$

is given by $K := cov(x, y)cov(y)^{-1}$ *and* $a := \bar{x} - K(\bar{y})$*, where* $cov(x, y)$ *is the covariance of* x *and* y*, and* $cov(y)$ *is the auto-covariance of* y*. In case* $cov(y)$ *is singular or nearly singular, the pseudo-inverse can be taken instead of the inverse.*

The operator $K \in \mathscr{L}(\mathcal{Y}, \mathcal{X})$ is also called the *Kalman* gain, and has the familiar form known from least squares projections. It is interesting to note that initially the connection between MMSE and Bayesian estimation was not seen [19], although it is one of the simplest approximations.

The resulting filter Eq. (40) is therefore called the **Gauss-Markov-Kalman** filter (GMKF). The original Kalman filter has Eq. (40) just for the means

$$\bar{x}_{a,1L} = \bar{x}_f + K(\hat{y} - \bar{z}).$$

It easy to compute that

Theorem 2 *The covariance operator corresponding to Eq.* (29) $\mathrm{cov}(x_{a,1L})$ *of* $x_{a,1L}$ *is given by*

$$\mathrm{cov}(x_{a,1L}) = \mathrm{cov}(x_f) - K\mathrm{cov}(x_f, y)^T = \mathrm{cov}(x_f) - \mathrm{cov}(x_f, y)\mathrm{cov}(z)^{-1}\mathrm{cov}(x_f, y)^T,$$

which is Kalman's formula for the covariance.

This shows that Eq. (40) is a true extension of the classical Kalman filter (KF). Rewriting Eq. (40) explicitly in less symbolic notation

$$x_a(\omega) = x_f(\omega) + \mathrm{cov}(x_f, y)\mathrm{cov}(z)^{-1}(\hat{y} - y(\omega)), \tag{42}$$

one may see that it is a relation between RVs, and hence some further *stochastic* discretisation is needed to be numerically implementable.

Nonlinear filters

The derivation of nonlinear but polynomial filters is given in [18]. It has the advantage of showing the connection to the "Bayes linear" approach [6], to the Gauss-Markov theorem [15], and to the *Kalman* filter [11,22]. Correcting higher moments of the posterior RV has been touched on in the "Correcting higher moments" section, and is not the topic here. Now the focus is on computing better than linear (see "The linear filter" section) approximations to the CE operator, which is the basic tool for the whole updating and identification process.

We follow [18] for a more general approach not limited to polynomials, and assume a set of linearly independent measurable functions, not necessarily orthonormal,

$$\mathcal{B} := \{\psi_\alpha \mid \alpha \in \mathcal{A},\ \psi_\alpha(y(\omega)) \in \mathcal{S}\} \subseteq \mathcal{S}_\infty \tag{43}$$

where $\mathcal{A}$ is some countable index set. Galerkin convergence [18] will require that

$$\mathcal{S}_\infty = \overline{\operatorname{span}}\ \mathcal{B},$$

i.e. that $\mathcal{B}$ is a Hilbert basis of $\mathcal{S}_\infty$.

Let us consider a general function $\Psi : \mathcal{X} \to \mathcal{R}$ of x, where $\mathcal{R}$ is some Hilbert space, of which we want to compute the conditional expectation $\mathbb{E}(\Psi(x)|y)$. Denote by $\mathcal{A}_k$ a finite part of $\mathcal{A}$ of cardinality k, such that $\mathcal{A}_k \subset \mathcal{A}_\ell$ for $k < \ell$ and $\bigcup_k \mathcal{A}_k = \mathcal{A}$, and set

$$\mathscr{R}_k := \mathcal{R} \otimes \mathcal{S}_k \subseteq \mathscr{R}_\infty := \mathcal{R} \otimes \mathcal{S}_\infty, \tag{44}$$

where the finite dimensional and hence closed subspaces $\mathcal{S}_k$ are given by

$$\mathcal{S}_k := \operatorname{span}\{\psi_\alpha \mid \alpha \in \mathcal{A}_k,\ \psi_\alpha \in \mathcal{B}\} \subseteq \mathcal{S}. \tag{45}$$

Observe that the spaces $\mathscr{R}_k$ from Eq. (44) are $\mathcal{L}$-closed, see [18]. In practice, also a "spatial" discretisation of the spaces $\mathcal{X}$ resp. $\mathcal{R}$ has to be carried out; but this is a standard process and will be neglected here for the sake of brevity and clarity.

For a RV $\Psi(x) \in \mathscr{R} = \mathcal{R} \otimes \mathcal{S}$ we make the following 'ansatz' for the optimal map $\phi_{\Psi,k}$ such that $P_{\mathscr{R}_k}(\Psi(x)) = \phi_{\Psi,k}(y)$:

$$\Phi_{\Psi,k}(y) = \sum_{\alpha \in \mathcal{A}_k} \nu_\alpha \psi_\alpha(y), \tag{46}$$

with as yet unknown coefficients $\nu_\alpha \in \mathcal{R}$. This is a normal *Galerkin*-ansatz, and the Galerkin orthogonality Eq. (37) can be used to determine these coefficients.

Take $\mathcal{Z}_k := \mathbb{R}^{\mathcal{A}_k}$ with canonical basis $\{\boldsymbol{e}_\alpha \mid \alpha \in \mathcal{A}_k\}$, and let

$$\boldsymbol{G}_k := (\langle \psi_\alpha(y(x)), \psi_\beta(y(x))\rangle_{\mathcal{S}})_{\alpha,\beta \in \mathcal{A}_k} \in \mathscr{L}(\mathcal{Z}_k)$$

be the symmetric positive definite Gram matrix of the basis of $\mathcal{S}_k$; also set

$$\boldsymbol{v} := \sum_{\alpha \in \mathcal{A}_k} \boldsymbol{e}_\alpha \otimes \nu_\alpha \in \mathcal{Z}_k \otimes \mathcal{R},$$
$$\boldsymbol{r} := \sum_{\alpha \in \mathcal{A}_k} \boldsymbol{e}_\alpha \otimes \mathbb{E}(\psi_\alpha(y(x))R(x)) \in \mathcal{Z}_k \otimes \mathcal{R}.$$

Theorem 3 *For any $k \in \mathbb{N}$, the coefficients $\{v_\alpha\}_{\alpha\in\mathcal{A}_k}$ of the optimal map $\phi_{\Psi,k}$ in Eq. (46) are given by the unique solution of the Galerkin equation*

$$(\boldsymbol{G}_k \otimes I_{\mathcal{R}})\boldsymbol{v} = \boldsymbol{r}. \tag{47}$$

It has the formal solution

$$\boldsymbol{v} = (\boldsymbol{G}_k \otimes I_{\mathcal{R}})^{-1}\boldsymbol{r} = (\boldsymbol{G}_k^{-1} \otimes I_{\mathcal{R}})\boldsymbol{r} \in \mathcal{Z}_k \otimes \mathcal{R}.$$

Proof The Galerkin Eq. (47) is a simple consequence of the Galerkin orthogonality Eq. (37). As the Gram matrix $\boldsymbol{G}_k$ and the identity $I_{\mathcal{R}}$ on $\mathcal{R}$ are positive definite, so is the tensor operator $(\boldsymbol{G}_k \otimes I_{\mathcal{R}})$, with inverse $(\boldsymbol{G}_k^{-1} \otimes I_{\mathcal{R}})$. □

The block structure of the equations is clearly visible. Hence, to solve Eq. (47), one only has to deal with the 'small' matrix $\boldsymbol{G}_k$.

The update corresponding to Eq. (35), using again $\Psi(x) = x$, one obtains a possibly nonlinear filter based on the basis $\mathcal{B}$:

$$x_a \approx x_{a,k} = x_f + \big(\phi_{x,k}(\hat{y}) - \phi_{x,k}(y(x_f))\big) = x_f + x_{\infty,k}. \tag{48}$$

In case that $\mathcal{Y}^* \subseteq \operatorname{span}\{\psi_\alpha\}_{\alpha\in\mathcal{A}_k}$, i.e. the functions with indices in $\mathcal{A}_k$ generate all the linear functions on $\mathcal{Y}$, this is a true extension of the Kalman filter.

Observe that this allows one to compute the map in Eq. (19) or rather Eq. (23) to any desired accuracy. Then, using this tool, one may construct a new random variable which has the desired posterior expectations; as was started in the "Correcting the mean" and "Correcting higher moments" sections. This is then a truly nonlinear extension of the linear filters described in "The Gauss-Markov-Kalman filter (GMKF)" section, and one may expect better tracking properties than even for the best linear filters. This could for example allow for less frequent observations of a dynamical system.

Numerical realisation

This is only going to be a rough overview on possibilities of numerical realisations. Only the simplest case of the linear filter will be considered, all other approximations can be dealt with in an analogous manner. Essentially we will look at two different kinds of approximations, *sampling* and *functional* or *spectral* approximations.

Sampling

As starting point take Eq. (42). As it is a relation between RVs, it certainly also holds for *samples* of the RVs. Thus it is possible to take an *ensemble* of sampling points $\omega_1, \ldots, \omega_N$ and require

$$\forall \ell = 1, \ldots, N: \quad \boldsymbol{x}_a(\omega_\ell) = \boldsymbol{x}_f(\omega_\ell) + \boldsymbol{C}_{x_f y}\boldsymbol{C}_y^{-1}(\check{y} - y(\omega_\ell)), \tag{49}$$

and this is the basis of the *ensemble* Kalman filter, the EnKF [5]; the points $\boldsymbol{x}_f(\omega_\ell)$ and $\boldsymbol{x}_a(\omega_\ell)$ are sometimes also denoted as *particles*, and Eq. (49) is a simple version of a *particle filter*. In Eq. (49), $\boldsymbol{C}_{x_f y} = \operatorname{cov}(x_f, y)$ and $\boldsymbol{C}_y = \operatorname{cov}(y)$

Some of the main work for the EnKF consists in obtaining good estimates of $\boldsymbol{C}_{x_f y}$ and $\boldsymbol{C}_y$, as they have to be computed from the samples. Further approximations are possible, for example such as *assuming* a particular form for $\boldsymbol{C}_{x_f y}$ and $\boldsymbol{C}_y$. This is the basis for methods like *kriging* and *3DVAR* resp. *4DVAR*, where one works with an approximate Kalman gain $\tilde{\boldsymbol{K}} \approx \boldsymbol{K}$. For a recent account see [12].

Functional approximation

Here we want to pursue a different tack, and want to discretise RVs not through their samples, but by *functional* resp. *spectral approximations* [14,17,30]. This means that all RVs, say $\boldsymbol{\nu}(\omega)$, are described as functions of *known* RVs $\{\xi_1(\omega), \ldots, \xi_\ell(\omega), \ldots\}$. Often, when for example stochastic processes or random fields are involved, one has to deal here with *infinitely* many RVs, which for an actual computation have to be truncated to a finite vector $\boldsymbol{\xi}(\omega) = [\xi_1(\omega), \ldots, \xi_n(\omega)]$ of significant RVs. We shall assume that these have been chosen such as to be independent. As we want to approximate later $\boldsymbol{x} = [x_1, \ldots, x_n]$, we do not need more than n RVs $\boldsymbol{\xi}$.

One further chooses a finite set of linearly independent functions $\{\psi_\alpha\}_{\alpha\in\mathcal{J}_M}$ of the variables $\boldsymbol{\xi}(\omega)$, where the index α often is a *multi-index*, and the set $\mathcal{J}_M$ is a finite set with cardinality (size) M. Many different systems of functions can be used, classical choices are [14,17,30] multivariate polynomials—leading to the *polynomial chaos expansion* (PCE), as well as trigonometric functions, kernel functions as in kriging, radial basis functions, sigmoidal functions as in artificial neural networks (ANNs), or functions derived from fuzzy sets. The particular choice is immaterial for the further development. But to obtain results which match the above theory as regards $\mathcal{L}$-invariant subspaces, we shall assume that the set $\{\psi_\alpha\}_{\alpha\in\mathcal{J}_M}$ includes all the *linear* functions of $\boldsymbol{\xi}$. This is easy to achieve with polynomials, and w.r.t kriging it corresponds to *universal* kriging. All other function systems can also be augmented by a linear trend.

Thus a RV $\boldsymbol{\nu}(\omega)$ would be replaced by a functional approximation

$$\boldsymbol{\nu}(\omega) = \sum_{\alpha\in\mathcal{J}_M} \boldsymbol{\nu}_\alpha \psi_\alpha(\boldsymbol{\xi}(\omega)) = \sum_{\alpha\in\mathcal{J}_M} \boldsymbol{\nu}_\alpha \psi_\alpha(\boldsymbol{\xi}) = \boldsymbol{\nu}(\boldsymbol{\xi}). \tag{50}$$

The argument ω will be omitted from here on, as we transport the probability measure $\mathbb{P}$ on Ω to $\boldsymbol{\Xi} = \Xi_1 \times \cdots \times \Xi_n$, the range of $\boldsymbol{\xi}$, giving $\mathbb{P}_\xi = \mathbb{P}_1 \times \cdots \times \mathbb{P}_n$ as a product measure, where $\mathbb{P}_\ell = (\xi_\ell)_*\mathbb{P}$ is the distribution measure of the RV ξ_ℓ, as the RVs ξ_ℓ are independent. All computations from here on are performed on $\boldsymbol{\Xi}$, typically some subset of $\mathbb{R}^n$. Hence n is the dimension of our problem, and if n is large, one faces a high-dimensional problem. It is here that low-rank tensor approximations [8] become practically important.

It is not too difficult to see that the linear filter, when applied to the spectral approximation, has exactly the same form as shown in Eq. (42). Hence the basic formula Eq. (42) looks formally the same in both cases, once it is applied to samples or "particles", in the other case to the functional approximation of RVs, i.e. to the coefficients in Eq. (50).

In both of the cases described here in the "Sampling" and "Functional approximation" sections, the question as how to compute the covariance matrices in Eq. (42) arises. In the EnKF in "Sampling" section they have to be computed from the samples [5], and in the case of functional resp. spectral approximations they can be computed from the coefficients in Eq. (50), see [21,24].

In the sampling context, the samples or particles may be seen as δ-measures, and one generally obtains weak-$*$ convergence of convex combinations of these δ-measures to the continuous limit as the number of particles increases. In the case of functional resp. spectral approximation one can bring the whole theory of Galerkin-approximations to bear on the problem, and one may obtain convergence of the involved RVs in appropriate norms [18]. We leave this topic with this pointer to the literature, as this is too extensive to be discussed any further and hence is beyond the scope of the present work.

Examples

The first example is a dynamic system considered in [21], it is the well-known Lorenz-84 chaotic model, a system of three nonlinear ordinary differential equations operating in the chaotic regime. This is an example along the description of Eqs. (3) and (5) in the "Data model" section. Remember that this was originally a model to describe the evolution of some amplitudes of a spherical harmonic expansion of variables describing world climate. As the original scaling of the variables has been kept, the time axis in Fig. 1 is in *days*. Every 10 days a noisy measurement is performed and the state description is updated. In between the state description evolves according to the chaotic dynamic of the system. One may observe from Fig. 1 how the uncertainty—the width of the distribution as given by the quantile lines—shrinks every time a measurement is performed, and then increases again due to the chaotic and hence noisy dynamics. Of course, we did not really measure the world climate, but rather simulated the "truth" as well, i.e. a *virtual* experiment, like the others to follow. More details may be found in [21] and the references therein. All computations are performed in a functional approximation with polynomial chaos expansions as alluded to in the "Functional approximation" section, and the filter is linear according to Eq. (42).

To introduce the nonlinear filter as sketched in "Nonlinear filters" section, where for the nonlinear filter the functions in Eq. (46) included polynomials up to quadratic terms, one may look shortly at a very simplified example, identifying a value, where only the third power of the value plus a Gaussian error RV is observed. All updates follow Eq. (28), but the update map is computed with different accuracy.

Shown are the pdfs produced by the linear filter according to Eq. (42)—Linear polynomial chaos Bayesian update (Linear PCBU)—a special form of Eq. (28), also with an iterated linear filter—iterative LPCBU—using Newton iterations, i.e. an iterated version of Eq. (42), and using polynomials up to order two, the quadratic polynomial chaos Bayesian update

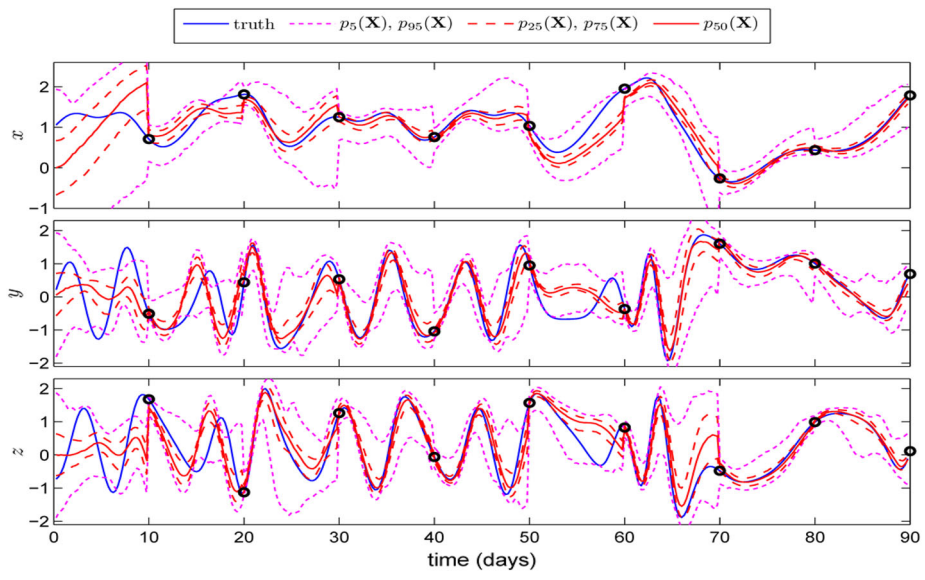

Fig. 1 Time evolution of the Lorenz-84 model with state identification with the LBU, from [21]. For the estimated state uncertainty the 50 (*full line*), ±25, and ±45 % quantiles are shown

(QPCBU). One may observe that due to the nonlinear observation, the differences between the linear filters and the quadratic one are already significant, the QPCBU yielding a better update.

We go back to the example shown in Fig. 1, but now consider only for one step a nonlinear filter like in Fig. 2, see [18].

As a first set of experiments we take the measurement operator to be linear in the state variable to be identified, i.e. we can observe the *whole* state directly. At the moment we consider updates after each day—whereas in Fig. 1 the updates were performed every 10 days. The update is done once with the linear Bayesian update (LBU), and again with a *quadratic* nonlinear BU (QBU). The results for the posterior pdfs are given in Fig. 3, where the linear update is dotted in blue and labelled $z1$, and the full red line is the quadratic QBU labelled $z2$; there is hardly any difference between the two except for the z-component of the state, most probably indicating that the LBU is already very accurate.

Now the same experiment, but the measurement operator is cubic:

These differences in posterior pdfs after one update may be gleaned from Fig. 4, and they are indeed larger than in the linear case Fig. 3, due to the strongly nonlinear measurement operator, showing that the QBU may provide a much more accurate tracking of the state, especially for non-linear observation operators.

As a last example we follow [18] and take a strongly nonlinear and also non-smooth situation, namely elasto-plasticity with linear hardening and large deformations and a *Kirchhoff-St. Venant* elastic material law [24,25]. This example is known as *Cook's mem-*

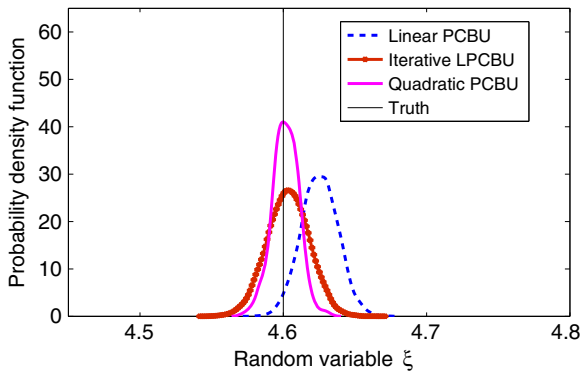

Fig. 2 Perturbed observations of the cube of a RV, different updates—linear, iterative linear, and quadratic update

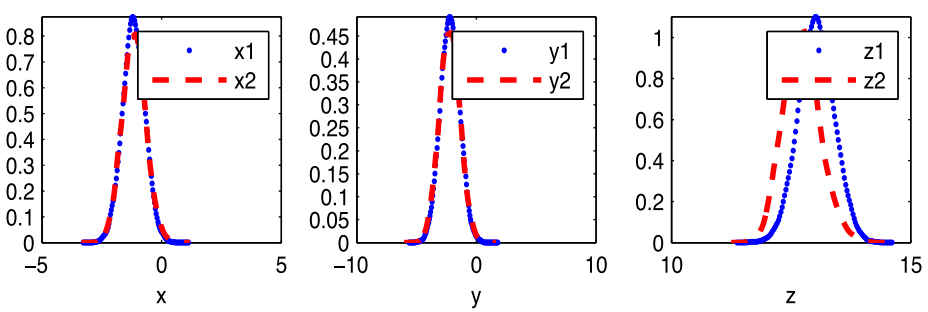

Fig. 3 Lorenz-84 model, perturbed linear observations of the state: posterior for LBU and QBU after one update, from [18]

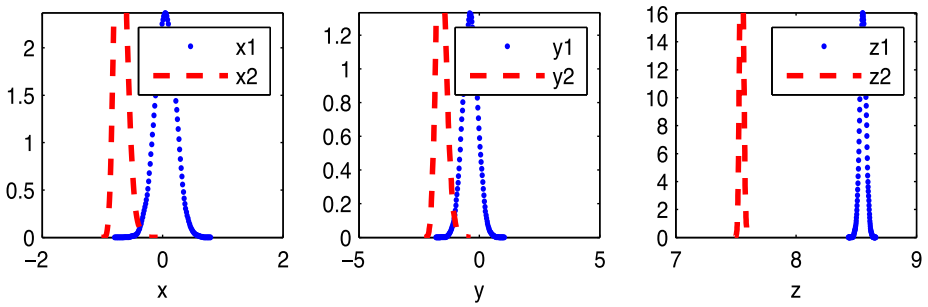

Fig. 4 Lorenz-84 model, perturbed cubic observations of the state: posterior for LBU and QBU after one update, from [18]

brane, and is shown in Fig. 5 with the undeformed mesh (initial), the deformed one obtained by computing with average values of the elasticity and plasticity material constants (deterministic), and finally the average result from a stochastic forward calculation of the probabilistic model (stochastic), which is described by a variational inequality [25].

The shear modulus G, a random field and not a deterministic value in this case, has to be identified, which is made more difficult by the non-smooth non-linearity. In Fig. 6 one may see the 'true' distribution at one point in the domain in an unbroken black line, with the mode—the maximum of the pdf—marked by a black cross on the abscissa, whereas the prior is shown in a dotted blue line. The pdf of the LBU is shown in an unbroken red line, with its mode marked by a red cross, and the pdf of the QBU is shown in a broken purple line with its mode marked by an asterisk. Again we see a difference between the LBU and the QBU. But here a curious thing happens; the mode of the LBU-posterior is actually closer to the mode of the 'truth' than the mode of the QBU-posterior. This means that somehow the QBU takes the prior more into account than the LBU, which is a kind

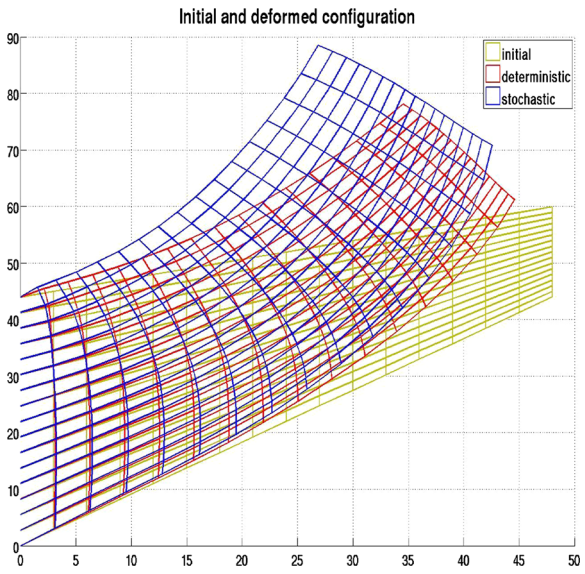

Fig. 5 Cook's membrane—large strain elasto-plasticity, undeformed grid [initial], deformations with mean properties [deterministic], and mean of the deformation with stochastic properties [stochastic], from [18,24,25]

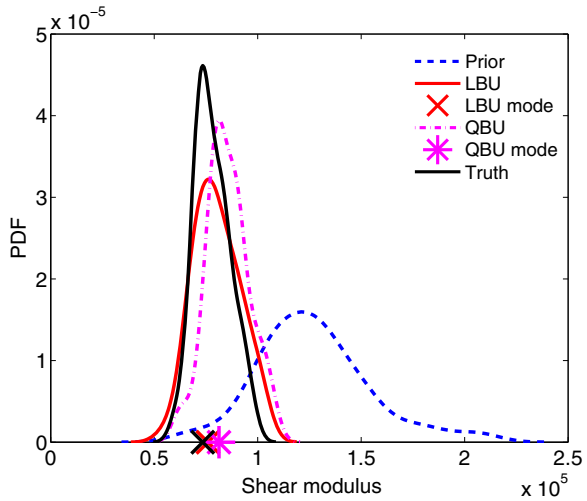

Fig. 6 Cook's membrane—large strain elasto-plasticity, perturbed linear observations of the deformation, LBU and QBU for the shear modulus, from [18]

of overshooting which has been observed at other occasions. On the other hand the pdf of the QBU is narrower—has less uncertainty—than the pdf of the LBU.

Conclusion

A general approach for state and parameter estimation has been presented in a Bayesian framework. The Bayesian approach is based here on the conditional expectation (CE) operator, and different approximations were discussed, where the linear approximation leads to a generalisation of the well-known Kalman filter (KF), and is here termed the Gauss-Markov-Kalman filter (GMKF), as it is based on the classical Gauss-Markov theorem. Based on the CE operator, various approximations to construct a RV with the proper posterior distribution were shown, where just correcting for the mean is certainly the simplest type of filter, and also the basis of the GMKF.

Actual numerical computations typically require a discretisation of both the spatial variables—something which is practically independent of the considerations here—and the stochastic variables. Classical are sampling methods, but here the use of spectral resp. functional approximations is alluded to, and all computations in the examples shown are carried out with functional approximations.

Authors' contributions
HGM provided ideas and wrote draft. EZ and BVR helped improve the research idea, BVR and AL conducted the numerical implementation and computation and the results parts. All authors read and approved the final manuscript.

Author details
[1]Institute of Scientific Computing, Technische Universität Braunschweig, Braunschweig, Germany , [2]Center for Uncertainty Quantification, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.

Acknowledgements
Partly supported by the Deutsche Forschungsgemeinschaft (DFG) through SFB 880.
Dedicated to Pierre Ladevèze on the occasion of his 70th birthday.

Competing interests
The authors declare that they have no competing interests.

Received: 12 March 2016 Accepted: 21 June 2016
Published online: 11 August 2016

References

1. Bobrowski A. Functional analysis for probability and stochastic processes. Cambridge: Cambridge University Press; 2005.
2. Bosq D. Linear processes in function spaces. Theory and applications. In: Lecture notes in statistics, vol. 149. Contains definition of strong or *L*-orthogonality for vector valued random variables. Berlin: Springer; 2000.
3. Engl HW, Groetsch CW. Inverse and ill-posed problems. New York: Academic Press; 1987.
4. Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Dordrecht: Kluwer; 2000.
5. Evensen G. Data assimilation—the ensemble Kalman filter. Berlin: Springer; 2009.
6. Goldstein M, Wooff D. Bayes linear statistics—theory and methods, Wiley series in probability and statistics. Chichester: Wiley; 2007.
7. Grewal MS, Andrews AP. Kalman filtering: theory and practice using MATLAB. Chichester: Wiley; 2008.
8. Hackbusch W. Tensor spaces and numerical tensor calculus. Berlin: Springer; 2012.
9. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57(1):97–109. doi:10.1093/biomet/57.1.97.
10. Jaynes ET. Probability theory, the logic of science. Cambridge: Cambridge University Press; 2003.
11. Kálmán RE. A new approach to linear filtering and prediction problems. J Basic Eng. 1960;82:35–45.
12. Kelly DTB, Law KJH, Stuart AM. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. Nonlinearity. 2014;27:2579–603. doi:10.1088/0951-7715/27/10/2579.
13. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J Royal Stat Soc Series B. 2001;63(3):425–64.
14. Le Maître OP, Knio OM. Spectral methods for uncertainty quantification. Scientific computation. Berlin: Springer; 2010. doi:10.1007/978-90-481-3520-2.
15. Luenberger DG. Optimization by vector space methods. Chichester: Wiley; 1969.
16. Marzouk YM, Najm HN, Rahn LA. Stochastic spectral methods for efficient Bayesian solution of inverse problems. J Comput Phys. 2007;224(2):560–86. doi:10.1016/j.jcp.2006.10.010.
17. Matthies HG. Uncertainty quantification with stochastic finite elements. In: Stein E, de Borst R, Hughes TJR, editors. Encyclopaedia of computational mechanics. Chichester: Wiley; 2007. doi:10.1002/0470091355.ecm071.
18. Matthies HG, Zander E, Rosić BV, Litvinenko A, Pajonk O. Inverse problems in a Bayesian setting. arXiv: 1511.00524 [math.PR]. 2015.
19. McGrayne SB. The theory that would not die. New Haven: Yale University Press; 2011.
20. Moselhy TA, Marzouk YM. Bayesian inference with optimal maps. J Comput Phys. 2012;231:7815–50. doi:10.1016/j.jcp.2012.07.022.
21. Pajonk O, Rosić BV, Litvinenko A, Matthies HG. A deterministic filter for non-Gaussian Bayesian estimation—applications to dynamical system estimation with noisy measurements. Physica D Nonlinear Phenom. 2012;241:775–88. doi:10.1016/j.physd.2012.01.001.
22. Papoulis A. Probability, random variables, and stochastic processes. 3rd ed. New York: McGraw-Hill Series in Electrical Engineering, McGraw-Hill; 1991.
23. Rao MM. Conditional measures and applications. Boca Raton: CRC Press; 2005.
24. Rosić BV, Kučerová A, Sýkora J, Pajonk O, Litvinenko A, Matthies HG. Parameter identification in a probabilistic setting. Eng Struct. 2013;50:179–96. doi:10.1016/j.engstruct.2012.12.029.
25. Rosić BV, Matthies HG. Identification of properties of stochastic elastoplastic systems. In: Papadrakakis M, Stefanou G, Papadopoulos V, editors. Computational methods in stochastic dynamics. Berlin: Springer; 2013. p. 237–53. doi:10.1007/978-94-007-5134-7_14.
26. Stuart AM. Inverse problems: a Bayesian perspective. Acta Numerica. 2010;19:451–559. doi:10.1017/S0962492910000061.
27. Tarantola A. Inverse problem theory and methods for model parameter estimation. Philadelphia: SIAM; 2004.
28. Tikhonov AN, Goncharsky AV, Stepanov VV, Yagola AG. Numerical methods for the solution of ill-posed problems. Berlin: Springer; 1995.
29. Tikhonov AN, Arsenin VY. Solutions of ill-posed problems. Chichester: Wiley; 1977.
30. Xiu D. Numerical methods for stochastic computations: a spectral method approach. Princeton: Princeton University Press; 2010.