Nikolas Nüsken | Sebastian Reich | Paul J. Rozdeba

State and Parameter Estimation from Observed Signal Increments

State and Parameter Estimation from Observed Signal Increments

Nikolas Nüsken[iD]**, Sebastian Reich** *[iD] **and Paul J. Rozdeba**[iD]

Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24/25, D-14476 Potsdam, Germany; nik.nuesken@gmx.de (N.N.); paul.rozdeba@uni-potsdam.de (P.J.R.)

*   Correspondence: sereich@uni-potsdam.de; Tel.: +49-331-977-1859

Abstract: The success of the ensemble Kalman filter has triggered a strong interest in expanding its scope beyond classical state estimation problems. In this paper, we focus on continuous-time data assimilation where the model and measurement errors are correlated and both states and parameters need to be identified. Such scenarios arise from noisy and partial observations of Lagrangian particles which move under a stochastic velocity field involving unknown parameters. We take an appropriate class of McKean–Vlasov equations as the starting point to derive ensemble Kalman–Bucy filter algorithms for combined state and parameter estimation. We demonstrate their performance through a series of increasingly complex multi-scale model systems.

Keywords: parameter estimation; continuous-time data assimilation; ensemble Kalman filter; correlated noise; multi-scale diffusion processes

1. Introduction

The research presented in this paper has been motivated by the state and parameter estimation problem for particles moving under a stochastic velocity field, with the measurements given by partial and noisy observations of their position increments. If the deterministic contributions to the velocity field are stationary, and the position increments of the moving particle are exactly observed, then one is led to a standard parameter estimation problem for stochastic differential equations (SDEs) [1,2]. In [3], this setting was extended to the case where the deterministic contributions to the velocity field themselves undergo a stochastic time evolution. Furthermore, while continuous-time observations of position increments are at the focus of the present study, the assimilation of discrete-time observations of particle positions has been investigated in [4,5] under a so-called Lagrangian data assimilation setting for atmospheric fluid dynamics.

The assumption of exactly and fully observed position increments is not always realistic and the case of partial and noisy observations is at the center of the present study. With access to partial and noisy observations of position increments leads to correlations between the measurement and model errors. The theoretical impact of such correlations on state and parameter estimation problems has been discussed, for example, in [6] in the context of linear systems, and in [7] for nonlinear systems. In particular, one finds that the appropriately adjusted data likelihood involves the gradient of log-densities, which is nontrivial from a computational perspective, and which prevents a straightforward application of standard Markov chain Monte Carlo (MCMC) or sequential Monte Carlo (SMC) methods [8].

In this paper, we instead follow an alternative Monte Carlo approach based on appropriately adjusted McKean–Vlasov filtering equations, an approach pioneered in [9] in the context of the standard state estimation problem for diffusion processes. McKean–Vlasov equations, first studied in [10], are a class of SDEs in which the right-hand side depends on the law of the process itself. We rely

on a particular formulation of McKean–Vlasov filtering equations, the so-called feedback particle filters [11], utilising stochastic innovation processes [12].

Our proposed Monte Carlo formulation avoids the need for estimating log-densities, and can be implemented in a numerically robust manner relying on a generalised ensemble Kalman–Bucy filter approximation applied to an extended state space formulation [13]. The ensemble Kalman–Bucy filter [14,15] has been introduced previously as an extension of the popular ensemble Kalman filter [13,16,17] to continuous-time data assimilation under the assumption of uncorrelated measurement and model errors.

While the McKean-Vlasov formulation is essentially mathematically equivalent to the more conventional one based on the Kushner-Stratonovitch equation [7], these two approaches differ significantly in structure, suggesting different tools for their analysis as well as numerical approximations. More broadly speaking, the McKean–Vlasov approach to filtering is appealing since its Monte Carlo implementations completely avoid the need for resampling characteristic of standard SMC methods. Furthermore, a wide range of approximations are possible within the McKean–Vlasov framework with some of them, such as the ensemble Kalman–Bucy filter, applicable to high-dimensional problems. The McKean–Vlasov approach also arises naturally when analysing sequential Monte Carlo methods [18].

In Section 6, we apply the proposed algorithms to a series of state and parameter estimation problems of increasing complexity. First, we study the state and parameter estimation problem for an Ornstein–Uhlenbeck process [2]. Two further experiments investigate the behaviour of the filters for reduced model equations, with the data being collected from underlying multi-scale models. There we distinguish between the averaging and homogenisation scenarios [19]. Finally, we look at examples of nonparametric drift estimation [3] and parameter estimation for the stochastic heat equation [20].

We finally mention that SMC methods for correlated noise terms in discrete-time have been discussed, for example, in [21] and in the context of the ensemble Kalman filter in [22]. Similar ideas have also been pursued in a more applied context in [23].

2. Mathematical Problem Formulation

We consider the time evolution of a random state variable $X_t \in \mathbb{R}^{N_x}$ in N_x-dimensional state space, $N_x \geq 1$, as prescribed by an SDE of the form

$$\mathrm{d}X_t = f(X_t, a)\,\mathrm{d}t + G\,\mathrm{d}W_t, \tag{1}$$

for time $t \geq 0$, with the drift function $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_a} \to \mathbb{R}^{N_x}$ depending on $N_a \geq 0$ unknown parameters $a = (a^1, \ldots, a^{N_a})^\mathsf{T} \in \mathbb{R}^{N_a}$. Model errors are represented through standard N_w-dimensional Brownian motion W_t, $N_w \geq 1$, and a matrix $G \in \mathbb{R}^{N_x \times N_w}$. We also introduce the associated model error covariance matrix $Q = GG^\mathsf{T}$. We will generally assume that the initial condition X_0 is fixed, that is, $X_0 = x_0$ a.s. for given $x_0 \in \mathbb{R}^{N_x}$. In terms of a more specific example, one can think of X_t denoting the position of a particle at time $t \geq 0$ moving in $N_x = 3$ dimensional space under the influence of a stochastic velocity field, with deterministic contributions given by f and stochastic perturbations by GW_t. In the case $G = 0$, the SDE (1) reduces to an ordinary differential equation with given initial condition x_0.

We assume throughout this paper that (1) possesses unique, strong solutions for all parameter values a. See, for example, [2] (Section 3.3) for sufficient conditions on the drift function f. The distribution of X_t is denoted by π_t, which we also abbreviate by $\pi_t = \mathrm{Law}(X_t)$. We use the same notation for measures and their Lebesgue densities, provided they exist.

Example 1. *A wide class of drift functions can be written in the form*

$$f(x, a) = f_0(x) + B(x)a = f_0(x) + \sum_{i=1}^{N_a} b_i(x)a^i, \tag{2}$$

where $f_0 : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ is a known drift function, the $b_i : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$, $i = 1, \ldots, N_a$, denote appropriate basis functions, and the vector $a = (a^1, \ldots, a^{N_a})^{\mathrm{T}} \in \mathbb{R}^{N_a}$ contains the unknown parameters of the model. The family $\{b_i(x)\}$ of basis functions, which we collect in a matrix-valued function $B(x) = (b_1(x), b_2(x), \ldots, b_{N_a}(x)) \in \mathbb{R}^{N_x \times N_a}$, could arise from a finite-dimensional truncation of some appropriate Hilbert space $\mathcal{H}$. See, for example, [24] for computational approaches to nonparametric drift estimation using a Galerkin approximation in $\mathcal{H}$, where the $b_i(x)$ become finite element basis functions. Furthermore, the expansion coefficients $\{a^i\}$ could be made time-dependent by letting them evolve according to some system of differential equations arising, for example, from the discretisation of an underlying partial differential equation with solutions in $\mathcal{H}$. See [3] for specific examples of such a setting. While the present paper focuses on stationary drift functions, i.e., the parameters $\{a^i\}$ are time-independent, the results from Sections 3 and 5, respectively, can easily be extended to the non-stationary case where the parameters themselves satisfy given evolution equations.

Data and an observation model are required in order to perform state and parameter estimation for SDEs of the form (1). In this paper, we assume that we observe partial and noisy increments $\mathrm{d}Y_t$ of the signal X_t, given by

$$\mathrm{d}Y_t = H\,\mathrm{d}X_t + R^{1/2}\mathrm{d}V_t = Hf(X_t, a)\,\mathrm{d}t + HG\,\mathrm{d}W_t + R^{1/2}\mathrm{d}V_t, \quad Y_0 = X_0 = x_0, \tag{3}$$

for t in the observation interval $[0, T]$, $T > 0$, where $H \in \mathbb{R}^{N_y \times N_x}$ is a given linear operator, V_t denotes standard N_y-dimensional Brownian motion with $N_y \geq 1$ and $R \in \mathbb{R}^{N_y \times N_y}$ is a covariance matrix. We introduce the observation map

$$h(x, a) = Hf(x, a) \tag{4}$$

for later use. Unless $HG = 0$, it is clear that the model error $E_t^{\mathrm{m}} := GW_t$ in (1) and the total observation error

$$E_t^{\mathrm{o}} := HGW_t + R^{1/2}V_t \tag{5}$$

in (3) are correlated. The impact of correlations between the model and measurement errors on the state estimation problem have been discussed by [6,7]. Furthermore, such correlations require adjustments to sequential estimation methods [16,17,25] which are the main focus of this paper. We assume throughout this paper that the covariance matrix

$$C = HGG^{\mathrm{T}}H^{\mathrm{T}} + R = HQH^{\mathrm{T}} + R \tag{6}$$

of the observation error (5) is invertible.

The special case $R = 0$ and $H = I$ leads to a pure parameter estimation problem which has been extensively studied in the literature in the settings of maximum likelihood and Bayesian estimators [1,2]. In Section 3, we provide a reformulation of the Bayesian approach as McKean–Vlasov equations for the parameters, based on the results in [9,11].

If $R \neq 0$, then (1) and (3) lead to a combined state and parameter estimation problem with correlated noise terms. We will first discuss the impact of this correlation on the pure state estimation problem in Section 4 assuming that the parameters of the problem are known. Again, we will derive appropriate McKean–Vlasov equations in the state variables. Our key contribution is a formulation that avoids the need for log-density estimates, and can be put into an appropriately generalised ensemble Kalman–Bucy filter approximation framework [14,15]. We also formally demonstrate that the McKean–Vlasov filter equation reduces to $\mathrm{d}X_t = \mathrm{d}Y_t$ in the limit $R \to 0$ and $H = I$, a property that is less straightforward to demonstrate for filter formulations involving log-densities.

These McKean–Vlasov equations are generalised to the combined state and parameter estimation problem via an augmentation of state space [13] in Section 5. Given the results from Section 4, such an extension is rather straightforward.

The numerical experiments in Section 6 rely exclusively on the generalised ensemble Kalman–Bucy filter approximation to the McKean–Vlasov equations, which are easy to implement and yield robust and accurate numerical results.

3. Parameter Estimation from Noiseless Data

In this section, we treat the simpler Bayesian parameter estimation problem which arises from setting $R = 0$ and $H = I$ in (3), i.e., $N_y = N_x$. This leads to $dX_t = dY_t$ and, furthermore, $X_t = Y_t$ for all $t \in [0, T]$, provided $X_0 = Y_0 = x_0$ which we assume throughout this paper. The requirement that $C = Q$ is invertible requires that G has rank N_x; that is, $N_w \geq N_x$ in (1). The data likelihood

$$l_t(a) = \exp\left(\int_0^t f(Y_s, a)^{\mathrm{T}} Q^{-1} \mathrm{d}Y_s - \frac{1}{2} \int_0^t f(Y_s, a)^{\mathrm{T}} Q^{-1} f(Y_s, a) \, \mathrm{d}s \right) \tag{7}$$

thus follows from the observation model with additive Brownian noise in (3). Given a prior distribution $\Pi_0(a)$ for the parameters, the resulting posterior distribution at any time $t \in (0, T]$ is

$$\Pi_t(a) = \frac{l_t(a)\Pi_0(a)}{\Pi_0[l_t]} \tag{8}$$

according to Bayes' theorem [7]. Here, we have introduced the shorthand

$$\Pi_0[l_t] = \int_{\mathbb{R}^{N_a}} l_t(a)\Pi_0(a) \, \mathrm{d}a \tag{9}$$

for the expectation of l_t with respect to Π_0. It is well-known that the posterior distributions Π_t satisfy the stochastic partial differential equation

$$\mathrm{d}\Pi_t[\phi] = (\Pi_t[\phi\, h_t] - \Pi_t[\phi]\Pi_t[h_t])^{\mathrm{T}} Q^{-1}(\mathrm{d}Y_t - \Pi_t[h_t]\mathrm{d}t) \tag{10}$$

with the time-dependent observation map

$$h_t(a) = f(Y_t, a), \tag{11}$$

where $\phi : \mathbb{R}^{N_a} \to \mathbb{R}$ is a compactly supported smooth test function, and $\Pi_t[\phi]$ again denoting the expectation of ϕ with respect to Π_t. See [7] for a detailed discussion. Equation (10) is a special instance of the well-known Kushner–Stratonovitch equation from time-continuous filtering [7].

3.1. Feedback Particle Filter

We now state a McKean–Vlasov reformulation of the Kushner–Stratonovitch Equation (10) as a special instance of the feedback particle filter of [11,12]. The key idea is to formulate a stochastic differential equation in the parameters in which they are treated as time-dependent random variables. We introduce the notation $\widetilde{A}_t$ for these, and require that the law of $\widetilde{A}_t$ coincide with (8) for $t \in [0, T]$, i.e., with the solution to (10).

Lemma 1 (Feedback particle filter). *Consider the McKean–Vlasov equations*

$$\mathrm{d}\widetilde{A}_t = K_t(\widetilde{A}_t) \, \mathrm{d}I_t + \Omega_t(\widetilde{A}_t) \, \mathrm{d}t, \tag{12}$$

where the matrix-valued Kalman gain $K_t \in \mathbb{R}^{N_a \times N_y}$ satisfies

$$\nabla \cdot \left(\widetilde{\Pi}_t \left(K_t Q \right) \right) = -\widetilde{\Pi}_t \left(h_t - \widetilde{\Pi}_t[h_t] \right)^{\mathrm{T}}, \quad \widetilde{\Pi}_t = \mathrm{Law}(\widetilde{A}_t). \tag{13}$$

The innovation process I_t can be chosen to be given by either

$$\mathrm{d}I_t = \mathrm{d}Y_t - \frac{1}{2}\left(h_t(\widetilde{A}_t) + \widetilde{\Pi}_t[h_t]\right)\mathrm{d}t, \tag{14}$$

or

$$\mathrm{d}I_t = \mathrm{d}Y_t - \left\{h_t(\widetilde{A}_t)\,\mathrm{d}t + G\,\mathrm{d}W_t\right\}, \tag{15}$$

and

$$\Omega_t^i = \frac{1}{2}\sum_{j=1}^{N_a}\sum_{k,l=1}^{N_y} Q^{kl}K_t^{jl}\left(\partial_j K_t^{ik}\right), \quad i = 1,\dots,N_a. \tag{16}$$

Then, the distribution $\widetilde{\Pi}_t = \mathrm{Law}(\widetilde{A}_t)$ coincides with the solution to (10), provided that the initial distributions agree. In other words, $\widetilde{\Pi}_t = \Pi_t$ for all $t \in [0, T]$.

Throughout this paper, we write (12) in the more compact Stratonovitch form

$$\mathrm{d}\widetilde{A}_t = K_t(\widetilde{A}_t) \circ \mathrm{d}I_t, \tag{17}$$

where the Stratonovitch interpretation is to be applied only to $\widetilde{A}_t$ in $K_t(\widetilde{A}_t)$, while the explicit time-dependence of K_t remains in its Itô interpretation. It should be noted that the matrix-valued function K_t is not uniquely defined by the PDE (13). Indeed, provided K_t solves (13), $K_t + \beta_t$ is also a solution whenever $\nabla \cdot \left(\widetilde{\Pi}_t \beta_t\right) = 0$. As discussed in [15], the minimiser over all suitable K_t with respect to a kinetic energy-type functional is of the form

$$K_t = \nabla\Psi_t Q^{-1} \tag{18}$$

for a vector of potential functions $\Psi_t = (\psi_t^1, \dots, \psi_t^{N_x})$, $\psi_t^k : \mathbb{R}^{N_a} \to \mathbb{R}$. Inserting (18) into (13) leads to N_x elliptic partial differential equations (often referred to as Poisson equations),

$$\nabla \cdot \left(\widetilde{\Pi}_t \nabla\Psi_t\right) = -\widetilde{\Pi}_t\left(h_t - \widetilde{\Pi}_t[h_t]\right)^{\mathrm{T}}, \quad \widetilde{\Pi}_t[\Psi_t] = 0, \tag{19}$$

understood component wise, where the centring condition $\widetilde{\Pi}_t[\Psi_t] = 0$ makes the solution unique under mild assumptions on $\widetilde{\Pi}_t$ (see [26]). The numerical approximation of (19) in the context of the feedback particle filter has been discussed in [27]. Finally, (15) yields a particularly appealing formulation, since it is based on a direct comparison of $\mathrm{d}Y_t$ with a random realisation of the right hand side of the SDE (1), given a parameter value $a = \widetilde{A}_t(\omega)$ and a realisation of the noise term $\mathrm{d}W_t(\omega)$. This fact will be explored further in Section 4.

Remark 1. *For clarity, let us repeat Equations (13) and (18) in their index forms:*

$$\sum_{i=1}^{N_a}\sum_{j=1}^{N_y} \partial_i\left(\widetilde{\Pi}_t\left(K_t^{ij}Q^{jk}\right)\right) = -\widetilde{\Pi}_t\left(h_t^k - \widetilde{\Pi}_t[h_t^k]\right), \quad k = 1,\dots,N_y, \tag{20}$$

$$\sum_{j=1}^{N_y} K_t^{ij}(a)Q^{jk} = \partial_i\psi_t^k(a), \quad i = 1,\dots,N_a, \quad k = 1,\dots,N_y. \tag{21}$$

3.2. Ensemble Kalman–Bucy Filter

Let us now assume that the initial distribution Π_0 is Gaussian, and that f is linear in the unknown parameters such as in (2). Then, the distributions $\widetilde{\Pi}_t$ remain Gaussian for all times with mean $\bar{a}_t$

and covariance matrix P_t^{aa}. The elliptic PDE (13) is solved by the parameter-independent Kalman gain matrix

$$K_t = P_t^{aa} B(Y_t)^{\mathrm{T}} Q^{-1} \tag{22}$$

and one obtains the McKean–Vlasov formulation

$$\mathrm{d}\widetilde{A}_t = P_t^{aa} B(Y_t)^{\mathrm{T}} Q^{-1} \mathrm{d}I_t \tag{23}$$

of the Kalman–Bucy filter, with the innovation process I_t defined by either

$$\mathrm{d}I_t = \mathrm{d}Y_t - \left(f_0(Y_t) + \frac{1}{2} B(Y_t)(\widetilde{A}_t + \bar{a}_t) \right) \mathrm{d}t \tag{24}$$

or

$$\mathrm{d}I_t = \mathrm{d}Y_t - \left\{ \left(f_0(Y_t) + B(Y_t)\widetilde{A}_t \right) \mathrm{d}t + G \, \mathrm{d}W_t \right\}. \tag{25}$$

Please note that the Stratonovitch formulation (17) reduces to the standard Itô interpretation, since K_t no longer depends explicitly on $\widetilde{A}_t$.

The McKean–Vlasov Equation (23) can be extended to nonlinear, non-Gaussian parameter estimation problems by generalising the parameter-independent Kalman gain matrix (22) to

$$K_t = P_t^{ah} Q^{-1}, \quad P_t^{ah} = \widetilde{\Pi}_t \left[(a - \bar{a}_t)(h_t(a) - \widetilde{\Pi}_t[h_t])^{\mathrm{T}} \right] = \widetilde{\Pi}_t \left[a \, (h_t(a) - \widetilde{\Pi}_t[h_t])^{\mathrm{T}} \right] \tag{26}$$

Clearly, the gain (26) provides only an approximation to the solution of (13). However, such approximations have become popular in nonlinear state estimation in the form of the ensemble Kalman filter [16,17], and we will test its suitability for parameter estimation in Section 6.

Numerical implementations of the proposed McKean–Vlasov approaches rely on Monte–Carlo approximations. More specifically, given M samples $\widetilde{A}_0^i$, $i = 1, \ldots, M$, from the initial distribution Π_0, we introduce the interacting particle system

$$\mathrm{d}\widetilde{A}_t^i = K_t^M(\widetilde{A}_t^i) \circ \mathrm{d}I_t^i, \tag{27}$$

where the innovation processes I_t^i are defined by either

$$\mathrm{d}I_t^i = \mathrm{d}Y_t - \frac{1}{2} \left(h_t(\widetilde{A}_t^i) + \bar{h}_t^M \right) \mathrm{d}t, \quad \bar{h}_t^M = \frac{1}{M} \sum_{i=1}^{M} h_t(\widetilde{A}_t^i), \tag{28}$$

or, alternatively,

$$\mathrm{d}I_t^i = \mathrm{d}Y_t - \left(h_t(\widetilde{A}_t^i) \, \mathrm{d}t + G \, \mathrm{d}W_t^i \right), \tag{29}$$

and W_t^i, $i = 1, \ldots, M$, denote independent N_w-dimensional Brownian motions. For K_t^M, we will use the parameter-independent empirical Kalman gain approximation

$$K_t^M = \widehat{P}_t^{ah} Q^{-1}, \quad \widehat{P}_t^{ah} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{A}_t^i (h_t(\widetilde{A}_t^i) - \bar{h}_t^M)^{\mathrm{T}}, \tag{30}$$

in our numerical experiments, which leads to the so-called ensemble Kalman–Bucy filter [14,15]. Please note that $\widehat{P}_t^{ah}$ provides an unbiased estimator of P_t^{ah}.

Finally, a robust and efficient time-stepping procedure for approximating $\widetilde{A}_{t_n}$, $t_n = n\Delta t$, is provided in [28–30]. Denoting the approximations at time t_n by $\widetilde{A}_n^i$, $i = 1, \ldots, M$, we obtain

$$\widetilde{A}_{n+1}^i = \widetilde{A}_n^i + \Delta t \widehat{P}_n^{ah} \left(Q + \Delta t \widehat{P}_n^{hh} \right)^{-1} \Delta I_n^i \tag{31}$$

with step size $\Delta t > 0$, empirical covariance matrices

$$\widehat{P}_n^{ah} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{A}_n^i (h_n(\widetilde{A}_n^i) - \overline{h}_n^M)^\mathrm{T}, \qquad \widehat{P}_n^{hh} = \frac{1}{M-1} \sum_{i=1}^{M} h_n(\widetilde{A}_n^i)(h_n(\widetilde{A}_n^i) - \overline{h}_n^M)^\mathrm{T}, \tag{32}$$

and innovation increments ΔI_n^i given by either

$$\Delta I_n^i = \Delta Y_n - \frac{1}{2}\left(h_n(\widetilde{A}_n^i) + \overline{h}_n^M\right)\Delta t, \qquad \overline{h}_n^M = \frac{1}{M}\sum_{i=1}^{M} h_n(\widetilde{A}_n^i), \tag{33}$$

or

$$\Delta I_n^i = \Delta Y_n - \left(h_n(\widetilde{A}_n^i)\,\Delta t + \Delta t^{1/2} G \Xi_n^i\right), \qquad \Xi_n^i \sim \mathrm{N}(0, I). \tag{34}$$

Here we have used the abbreviations $h_n(a) = f(Y_n, a)$, $Y_n = Y_{t_n}$, and $\Delta Y_n = Y_{t_{n+1}} - Y_{t_n}$.

While the feedback particle formulation (17) and its ensemble Kalman–Bucy filter approximation (31) are special cases of already available formulations, they provide the starting point for our novel McKean–Vlasov equations and their numerical approximation of the combined state and parameter estimation problem with correlated measurement and model errors, which we develop in the following two sections.

4. State Estimation for Noisy Data

We return to the observation Model (3) with $R \neq 0$ and general H. The pure state estimation problem is considered first; that is, $f(x, a) = f(x)$ in (1).

Using E_t^o, given by (5), and E_t^c defined by

$$E_t^c = G(I - G^\mathrm{T} H^\mathrm{T} C^{-1} HG)W_t - QH^\mathrm{T} C^{-1} R^{1/2} V_t \tag{35}$$

with the total measurement error covariance matrix C given by (6), we find that

$$GW_t = E_t^c + QH^\mathrm{T} C^{-1} E_t^o, \tag{36}$$

and the covariations [2] satisfy

$$\langle E^o, E^c \rangle_t = 0, \quad \langle E^o, E^o \rangle_t = Ct, \quad \langle E^c, E^c \rangle_t = G(I - G^\mathrm{T} H^\mathrm{T} C^{-1} HG)G^\mathrm{T} t. \tag{37}$$

These errors naturally suggest linear combinations of W_t and V_t in (1) and (3) that shift the correlation between measurement and model errors to the signal dynamics, yielding

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + G(I - G^\mathrm{T} H^\mathrm{T} C^{-1} HG)^{1/2}\mathrm{d}\widehat{W}_t + QH^\mathrm{T} C^{-1/2}\mathrm{d}\widehat{V}_t, \tag{38a}$$

$$\mathrm{d}Y_t = Hf(X_t)\,\mathrm{d}t + C^{1/2}\mathrm{d}\widehat{V}_t, \tag{38b}$$

where $\widehat{W}_t$ and $\widehat{V}_t$ denote mutually independent standard Brownian motions of dimension N_w and N_y, respectively. These equations correspond exactly to the correlated noise example from [7] (Section 3.8). Furthermore, $H = I$ and $R = 0$ lead to $E_t^c = 0$, $QH^\mathrm{T} C^{-1/2} = C^{1/2}$, and, hence, $\mathrm{d}X_t = \mathrm{d}Y_t$.

A straightforward application of the results from [7] (Section 3.8) yields the following statement:

Lemma 2 (Generalised Kushner–Stratonovich equation)**.** *The conditional expectations* $\pi_t[\phi] = \mathbb{E}[\phi(X_t)|Y_{[0,t]}]$ *satisfy*

$$\pi_t[\phi] = \pi_0[\phi] + \int_0^t \pi_s[\mathcal{L}\phi]\,\mathrm{d}s + \int_0^t \pi_s\left[\phi h + HQ\nabla\phi - \phi\pi_s[h]\right]^\mathrm{T} C^{-1}\left(\mathrm{d}Y_s - \pi_s[h]\,\mathrm{d}s\right), \tag{39}$$

where We use the notation $Q : \nabla\nabla\phi = \sum_{i,j=1}^{N_x} Q^{ij}\partial_i\partial_j\phi$.

$$\mathcal{L} = f \cdot \nabla + \frac{1}{2}Q : \nabla\nabla \tag{40}$$

is the generator of (1), *$h(x) = Hf(x)$ denotes the observation map, and ϕ is a compactly supported smooth function.*

For the convenience of the reader, we present an independent derivation in Appendix A. We note that (39) also arises as the Kushner–Stratonovitch equations for an SDE Model (1) with observations Y_t satisfying the observation model

$$dY_t = H\left(f(X_t) - Q\nabla \log \pi_t(X_t)\right)dt + C^{1/2}d\widetilde{V}_t, \tag{41}$$

where $\widetilde{V}_t$ denotes N_y-dimensional Brownian motion independent of the Brownian motion W_t in (1). Here we have used that $\pi_t\left[HQ\nabla\pi_t\right] = 0$. This reinterpretation of our state estimation problem in terms of uncorrelated model and observation errors and modified observation map

$$\widetilde{h}_t(x) = H\left(f(x) - Q\nabla \log \pi_t(x)\right) \tag{42}$$

allows one to apply available MCMC and SMC methods for continuous-time filtering and smoothing problems. See, for example, [16]. However, there are two major limitations of such an approach. First, it requires approximating the gradient of the log-density. Second, the modified observation Model (41) is not well-defined in the limit $R \to 0$ and $H = I$, since the density π_t collapses to a Dirac delta function under the given initial condition $X_0 = x_0$ a.s.

In order to circumvent these complications, we develop an alternative approach based on an appropriately modified feedback particle filter formulation in the following subsection.

4.1. Generalised Feedback Particle Filter Formulation

While it is clearly possible to apply the standard feedback particle filter formulations using (41), the following alternative formulation avoids the need for approximating the gradient of the log-density.

Lemma 3 (Feedback particle filter with correlated innovation). *Consider the McKean–Vlasov equation*

$$d\widetilde{X}_t = f(\widetilde{X}_t)\,dt + G\,dW_t + K_t(\widetilde{X}_t) \circ dI_t + \Omega_t(\widetilde{X}_t)\,dt, \tag{43}$$

where the gain $K_t \in \mathbb{R}^{N_x \times N_y}$ solves

$$\nabla \cdot \left(\widetilde{\pi}_t\left(K_t C - QH^{\mathsf{T}}\right)\right) = -\widetilde{\pi}_t\left(h - \widetilde{\pi}_t[h]\right)^{\mathsf{T}}, \quad \widetilde{\pi}_t = \mathrm{Law}(\widetilde{X}_t), \tag{44}$$

with observation map $h(x) = Hf(x)$. The function Ω_t is given by

$$\Omega_t^i = -\frac{1}{2}\sum_{l=1}^{N_x}\sum_{j=1}^{N_y}\partial_l K_t^{ij}(QH^{\mathsf{T}})^{lj}, \quad i = 1,\ldots,N_x, \tag{45}$$

and the innovation process I_t by

$$dI_t = dY_t - \left(h(\widetilde{X}_t)\,dt + HG\,dW_t + R^{1/2}\,dU_t\right). \tag{46}$$

Here, W_t and U_t denote mutually independent N_x-dimensional and N_y-dimensional Brownian motions, respectively. Then, $\widetilde{\pi}_t = \mathrm{Law}(\widetilde{X}_t)$ coincides with the solution to (39), provided that the initial distributions agree.

It should be stressed that W_t in (43) and (46) denote the same Brownian motion, resulting in correlations between the innovation process and model noise.

Proof. In this proof the Einstein summation convention over repeated indices is employed, noting that (44) takes the form

$$\partial_i \left(\tilde{\pi}_t \left(K_t^{ij} C^{jk} - (QH^{\mathrm{T}})^{ik} \right) \right) = -\tilde{\pi}_t \left(h^k - \tilde{\pi}_t[h^k] \right), \quad k = 1, \dots, N_y. \tag{47}$$

We begin by writing (43) in its Itô-form,

$$\mathrm{d}\tilde{X}_t = f(\tilde{X}_t)\,\mathrm{d}t + G\,\mathrm{d}W_t + K_t(\tilde{X}_t)\,\mathrm{d}I_t + \hat{\Omega}_t(\tilde{X}_t)\,\mathrm{d}t, \tag{48}$$

where

$$\begin{aligned}
\hat{\Omega}_t^i &= \Omega_t^i + \frac{1}{2}\left\{ -\left(\partial_l K_t^{ij}\right)(QH^{\mathrm{T}})^{lj} + 2\left(\partial_l K_t^{ij}\right)K_t^{lk}C^{kj} \right\}\\
&= \left(\partial_l K_t^{ij}\right)\left\{ K_t^{lk}C^{kj} - (QH^{\mathrm{T}})^{lj} \right\}
\end{aligned} \tag{49}$$

Here, we have used that the covariation between K_t and I_t satisfies

$$\mathrm{d}\left\langle K^{ij}, I^j \right\rangle_t = \partial_l K_t^{ij}\left(G^{lk}\,\mathrm{d}\left\langle W^k, I \right\rangle_t + K_t^{lk}\,\mathrm{d}\left\langle I^k, I^j \right\rangle_t \right). \tag{50}$$

Furthermore, $\langle GW, I \rangle_t = -QH^{\mathrm{T}}t$ and $\langle I, I \rangle_t = 2Ct$.

For a smooth compactly supported test function ϕ, Itô's formula implies

$$\phi(\tilde{X}_t) = \phi(\tilde{X}_0) + \int_0^t \partial_i \phi(\tilde{X}_s)\,\mathrm{d}\tilde{X}_s^i + \frac{1}{2}\int_0^t \partial_i\partial_j\phi(\tilde{X}_s)\,\mathrm{d}\langle \tilde{X}^i, \tilde{X}^j \rangle_s, \tag{51}$$

where the covariation process is given by

$$\langle \tilde{X}, \tilde{X} \rangle_t = tQ - \int_0^t \left(K_s HQ + QH^{\mathrm{T}}K_s^{\mathrm{T}} \right)\,\mathrm{d}s + 2\int_0^t K_s C K_s^{\mathrm{T}}\,\mathrm{d}s. \tag{52}$$

Our aim is to show that $\tilde{\pi}_t[\phi]$ coincides with $\pi_t[\phi]$ as defined by the Kushner–Stratonovich Equation (39). To this end, we insert (48) and (52) into (51) and take the conditional expectation, arriving at

$$\begin{aligned}
\tilde{\pi}_t[\phi] &= \tilde{\pi}_0[\phi] + \int_0^t \tilde{\pi}_s[\mathcal{L}\phi]\,\mathrm{d}s + \int_0^t \tilde{\pi}_s\left[(\partial_i\phi)K_s^{ij} \right]\,\mathrm{d}Y_s^j - \int_0^t \tilde{\pi}_s\left[(\partial_i\phi)K_s^{ij}h^j \right]\,\mathrm{d}s\\
&+ \int_0^t \tilde{\pi}_s\left[(\partial_i\phi)\hat{\Omega}_s^i \right]\,\mathrm{d}s + \int_0^t \tilde{\pi}_s\left[(\partial_i\partial_j\phi)\left(K_s(CK_s^{\mathrm{T}} - HQ) \right)^{ij} \right]\,\mathrm{d}s,
\end{aligned} \tag{53}$$

recalling that the generator $\mathcal{L}$ has been defined in (40). Under the assumption that K_t satisfies (44), the two Equations (39) and (53) coincide. Indeed,

$$\tilde{\pi}_s\left[(\partial_i\phi)(K_s^{ik}C^{kj} - (QH^{\mathrm{T}})^{ij}) \right] = \tilde{\pi}_s\left[\phi\left(h^j - \tilde{\pi}_s\left[h^j \right] \right) \right] \tag{54}$$

implies

$$\tilde{\pi}_s[\nabla\phi \cdot K_s] = \tilde{\pi}_s\left[\phi h + HQ\nabla\phi - \phi\tilde{\pi}_s[h] \right]^{\mathrm{T}} C^{-1}, \tag{55}$$

and the $\mathrm{d}Y_s$-contributions agree. To verify the same for the $\mathrm{d}s$-contributions, we use (44) to obtain

$$\begin{aligned}
\tilde{\pi}_s\left[(\partial_i\phi)K_s^{ij}(h^j - \tilde{\pi}_t[h^j]) \right] &= -\int_{\mathbb{R}^{N_x}} (\partial_i\phi)K_s^{ij}\partial_l\left(\tilde{\pi}_s\left(K_s^{ln}C^{nj} - (QH^{\mathrm{T}})^{lj} \right) \right)\,\mathrm{d}x\\
&= \tilde{\pi}_s\left[(\partial_i\phi)\hat{\Omega}_s^i \right] + \tilde{\pi}_s\left[(\partial_i\partial_j\phi)\left(K_s(CK_s^{\mathrm{T}} - K_s HQ) \right)^{ij} \right].
\end{aligned} \tag{56}$$

Finally, collecting terms in (53) and (56), and applying (55) to the remaining d*s*-contribution, i.e., $-\widetilde{\pi}_s[\nabla \phi \cdot K_s]\widetilde{\pi}_s[h]$, leads to the desired result. $\square$

We note that the correlation between the innovation process I_t and the model error W_t leads to a correction term Ω_t in (43) which cannot be subsumed into a Stratonovitch correction, in contrast to the standard feedback particle filter formulation (17).

Remark 2. *Assuming that there exist potential functions* $\Psi_t = (\psi_t^1, \ldots, \psi_t^{N_y})$, $\psi_t^k : \mathbb{R}^{N_x} \to \mathbb{R}$, *solving the Poisson equation(s)* (19) *(with* $\widetilde{\Pi}_t$ *being replaced by* $\widetilde{\pi}_t$*),* (44) *can be solved by requiring*

$$K_t = (\nabla \Psi_t + QH^{\mathrm{T}})C^{-1}, \tag{57}$$

thus generalising (18).

Remark 3. *If we set* $R = 0$, $H = I$, *and* $K_t = QH^{\mathrm{T}}C^{-1} = I$ *in* (43)*, then one obtains*

$$\mathrm{d}\widetilde{X}_t = \mathrm{d}Y_t \tag{58}$$

since Ω_t *vanishes, and all other terms in* (43) *cancel each other out. If, furthermore,* $Y_0 = \widetilde{X}_0 = x_0$ *a.s., then* $\widetilde{X}_t = Y_t$ *for all* $t \in [0, T]$*, which in turn justifies our assumption that the gain* K_t *is independent of the state variable. Hence, the McKean–Vlasov formulation* (43) *reproduces the exact reference trajectory* Y_t *in the case of no measurement errors and perfectly known initial conditions.*

We develop a simplified version of the feedback particle filter formulation (43) for linear SDEs and Gaussian distributions in the following subsection, which will form the basis of the generalised ensemble Kalman–Bucy filter put forward in the follow-up Section 4.3.

4.2. Generalised Kalman–Bucy Filter

Let us assume that $f(x) = Fx$ with $F \in \mathbb{R}^{N_x \times N_x}$, i.e., Equations (1) and (3) take the form

$$\mathrm{d}X_t = FX_t\,\mathrm{d}t + G\,\mathrm{d}W_t, \tag{59a}$$

$$\mathrm{d}Y_t = HFX_t\,\mathrm{d}t + HG\,\mathrm{d}W_t + R^{1/2}\,\mathrm{d}V_t, \tag{59b}$$

with initial conditions drawn from a Gaussian distribution. In this case π_t stays Gaussian for all $t > 0$, i.e., $\pi_t \sim \mathrm{N}(\overline{x}_t, P_t)$ with $\overline{x}_t \in \mathbb{R}^{N_x}$, $P_t \in \mathbb{R}^{N_x \times N_x}$. Equation (19) can be solved uniquely by $\nabla_x \Psi = P_t F^{\mathrm{T}} H^{\mathrm{T}}$, and thus the McKean–Vlasov equations for the feedback particle filter (43) reduce to

$$\mathrm{d}\widetilde{X}_t = F\widetilde{X}_t\,\mathrm{d}t + G\,\mathrm{d}W_t + \left(P_t F^{\mathrm{T}} H^{\mathrm{T}} + QH^{\mathrm{T}}\right)C^{-1}\mathrm{d}I_t, \tag{60}$$

with the innovation process (46) leading to

$$\mathrm{d}I_t = \mathrm{d}Y_t - HF\widetilde{X}_t\mathrm{d}t - HG\,\mathrm{d}W_t - R^{1/2}\mathrm{d}U_t. \tag{61}$$

We take the expectation in (60) and (61) and end up with

$$\mathrm{d}\overline{x}_t = F\overline{x}_t\,\mathrm{d}t + \left(P_t F^{\mathrm{T}} + Q\right)H^{\mathrm{T}}C^{-1}\left(\mathrm{d}Y_t - HF\overline{x}_t\,\mathrm{d}t\right). \tag{62}$$

Defining $u_t := \widetilde{X}_t - \overline{x}_t$, we see that

$$\mathrm{d}u_t = Fu_t\,\mathrm{d}t + G\,\mathrm{d}\widetilde{W}_t - \left(P_t F^{\mathrm{T}} + Q\right)H^{\mathrm{T}}C^{-1}\left(HFu_t\,\mathrm{d}t + HG\,\mathrm{d}W_t + R^{1/2}\mathrm{d}t\right). \tag{63}$$

Next we use

$$d\left(u_t u_t^\mathsf{T}\right) = du_t u_t^\mathsf{T} + u_t du_t^\mathsf{T} + d\langle u, u^\mathsf{T}\rangle_t \tag{64}$$

and $P_t = \mathbb{E}[u_t u_t^\mathsf{T}]$ to obtain, after some calculations,

$$dP_t = (FP_t + P_t F^\mathsf{T})\, dt - \left(P_t F^\mathsf{T} + Q\right) H^\mathsf{T} C^{-1} H\, (FP_t + Q)\, dt + Q\, dt. \tag{65}$$

Hence we have shown that our McKean–Vlasov formulation (60) agrees with the standard Kalman–Bucy filter equations for the mean and the covariance matrix in the correlated noise case [6].

4.3. Ensemble Kalman–Bucy Filter

The McKean–Vlasov Equation (60) for linear systems, along with Gaussian prior and posterior distributions, suggest approximating the feedback particle filter formulation (43) for nonlinear systems by

$$d\widetilde{X}_t = f(\widetilde{X}_t)\, dt + G\, dW_t + \left(P_t^{xh} + QH^\mathsf{T}\right) C^{-1} dI_t, \tag{66}$$

where the innovation process I_t given by (46) as before. In other words, we approximate the gain matrix K_t in (43) by the state independent term $\left(P_t^{xh} + QH^\mathsf{T}\right) C^{-1}$ with the covariance matrix P_t^{xh} defined by

$$P_t^{xh} = \widetilde{\pi}_t \left[(x - \bar{x}_t)(h(x) - \widetilde{\pi}_t[h])^\mathsf{T}\right] = \widetilde{\pi}_t \left[x\,(h(x) - \widetilde{\pi}_t[h])^\mathsf{T}\right] \tag{67}$$

where $\widetilde{\pi}_t$ denotes the law of $\widetilde{X}_t$.

We can now generalise the ensemble Kalman–Bucy filter formulation (31) for the pure parameter estimation problem to the state estimation problem with correlated noise. We assume that M initial state values $\widetilde{X}_0^i$ have been sampled from an initial distribution π_0 or, alternatively, $X_0^i = x_0$ for all $i = 1, \ldots, M$ in case the initial condition is known exactly. These state values are then propagated under the time-stepping procedure

$$\widetilde{X}_{n+1}^i = \widetilde{X}_n^i + \Delta t f(\widetilde{X}_n^i) + \Delta t^{1/2} G\Theta_n^i + \left(\widehat{P}_n^{xh} + QH^\mathsf{T}\right) \left(C + \Delta t \widehat{P}_n^{hh}\right)^{-1} \Delta I_n^i \tag{68}$$

with $\Theta_n^i \sim \mathrm{N}(0, I)$, step size $\Delta t > 0$, empirical covariance matrices

$$\widehat{P}_n^{xh} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{X}_n^i (h(\widetilde{X}_n^i) - \bar{h}_n^M)^\mathsf{T}, \qquad \bar{h}_n^M = \frac{1}{M} \sum_{i=1}^{M} h(\widetilde{X}_n^i), \tag{69a}$$

$$\widehat{P}_n^{hh} = \frac{1}{M-1} \sum_{i=1}^{M} h(\widetilde{X}_n^i)(h(\widetilde{X}_n^i) - \bar{h}_n^M)^\mathsf{T}, \tag{69b}$$

and innovation increments ΔI_n^i given by

$$\Delta I_n^i = \Delta Y_n - \Delta t h(\widetilde{X}_n^i) - \Delta t^{1/2} HG\Theta_n^i - \Delta t^{1/2} R^{1/2} \Xi_n^i, \qquad \Xi_n^i \sim \mathrm{N}(0, I). \tag{70}$$

The McKean–Vlasov equations of this section form the basis for the methods proposed for the combined state and parameter estimation problem to be considered next.

5. Combined State and Parameter Estimation

We now return to the combined state and parameter estimation problem, and consider the augmented dynamics

$$dX_t = f(X_t, A_t)\, dt + G\, dW_t, \tag{71a}$$

$$dA_t = 0, \tag{71b}$$

with observations (3) as before. The initial conditions satisfy $X_0 = x_0$ a.s., and $A_0 \sim \Pi_0$. Let us introduce the extended state space variable $Z_t = (X_t^\mathsf{T}, A_t^\mathsf{T})^\mathsf{T}$. In terms of Z_t, the Equations (3) and (71) take the form

$$dZ_t = \bar{f}(Z)\, dt + \bar{G}\, dW_t, \tag{72a}$$

$$dY_t = \bar{H}\, dZ_t + R^{1/2}\, dV_t, \tag{72b}$$

with

$$\bar{f}(z) = \begin{pmatrix} f(x,a) \\ 0 \end{pmatrix}, \quad \bar{G} = \begin{pmatrix} G & 0 \\ 0 & 0 \end{pmatrix}, \quad \bar{H} = \begin{pmatrix} H & 0 \end{pmatrix}. \tag{73}$$

Thus we end up with an augmented state estimation problem of the general structure considered in detail in Section 4 already. Below we provide details on some of the necessary modifications.

5.1. Feedback Particle Filter Formulation

The appropriately extended feedback particle filter Equation (43) leads to

$$d\widetilde{X}_t = f(\widetilde{X}, \widetilde{A}_t)\, dt + G\, dW_t + (\nabla_x \Psi_t(\widetilde{X}_t, \widetilde{A}_t) + QH^\mathsf{T})C^{-1} \circ dI_t + \Omega_t(\widetilde{X}_t, \widetilde{A}_t), \tag{74a}$$

$$d\widetilde{A}_t = \nabla_a \Psi_t(\widetilde{X}_t, \widetilde{A}_t)C^{-1} \circ dI_t, \tag{74b}$$

where (46) takes the form

$$dI_t = dY_t - \left(h(\widetilde{X}_t, \widetilde{A}_t)\, dt + HG\, dW_t + R^{1/2}dU_t \right) \tag{75}$$

with observation map (4) and correction Ω_t given by (45), with Q replaced by $\bar{Q} = \bar{G}\bar{G}^T$ and H by $\bar{H}$. In the Poisson equation(s) (19), $\widetilde{\Pi}_t$ is replaced by $\widetilde{\pi}_t$ denoting the joint density of $(\widetilde{X}_t, \widetilde{A}_t)$. We also stress that Ψ_t becomes a function of x and a, and we distinguish between gradients with respect to x and a using the notation ∇_x and ∇_a, respectively.

Numerical implementations of the extended feedback particle filter are demanding due to the need for solving the Poisson equation(s) (19). Instead, we again rely on the ensemble Kalman–Bucy filter approximation, which we describe next.

5.2. Ensemble Kalman–Bucy Filter

We approximate the joint density $\widetilde{\pi}_t$ of $\widetilde{Z}_t$ by an ensemble of particles

$$\widetilde{Z}_t^i = \begin{pmatrix} \widetilde{X}_t^i \\ \widetilde{A}_t^i \end{pmatrix}, \tag{76}$$

that is,

$$\widetilde{\pi}_t \approx \frac{1}{M} \sum_{i=1}^M \delta_{\widetilde{Z}_t^i}, \tag{77}$$

where $\delta_{z'}$ denotes the Dirac delta function centred at z'. The initial ensemble satisfies $X_0^i = x_0$ for all $i = 1, \dots, M$, and the initial parameter values A_0^i are independent draws from the prior distribution Π_0.

At the same time, we make the approximation $\widetilde{Z}_t \sim \mathrm{N}(\bar{z}_t^M, \widehat{P}_t^{zz})$ when dealing with the Kalman gain of the feedback particle filter. Here the empirical mean $\bar{z}_t^M$ has components

$$\bar{x}_t^M = \frac{1}{M} \sum_{i=1}^M \widetilde{X}_t^i, \quad \bar{a}_t^M = \frac{1}{M} \sum_{i=1}^M \widetilde{A}_t^i, \tag{78}$$

and the joint empirical covariance matrix is given by

$$\widehat{P}_t^{zz} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{Z}_t^i (\widetilde{Z}_t - \bar{z}_t^M)^{\mathrm{T}} = \begin{pmatrix} \widehat{P}_t^{xx} & \widehat{P}_t^{xa} \\ (\widehat{P}_t^{xa})^{\mathrm{T}} & \widehat{P}_t^{aa} \end{pmatrix}. \tag{79}$$

As in Section 4.3, the solution to (19) can be approximated by

$$\nabla_x \Psi_t = P_t^{xh}, \quad \nabla_a \Psi_t = P_t^{ah}, \tag{80}$$

where finally, the covariance matrices P_t^{xh} and P_t^{ah} are estimated by their empirical counterparts

$$\widehat{P}_t^{xh} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{X}_t^i (h(\widetilde{X}_t^i, \widetilde{A}_t^i) - \bar{h}_t^M)^{\mathrm{T}}, \tag{81a}$$

$$\widehat{P}_t^{ah} = \frac{1}{M-1} \sum_{i=1}^{M} \widetilde{A}_t^i (h(\widetilde{X}_t^i, \widetilde{A}_t^i) - \bar{h}_t^M)^{\mathrm{T}}, \tag{81b}$$

with $\bar{h}_t^M$ defined by

$$\bar{h}_t^M = \frac{1}{M} \sum_{i=1}^{M} h(\widetilde{X}_t^i, \widetilde{A}_t^i). \tag{82}$$

Summing everything up, we obtain the following generalised ensemble Kalman–Bucy filter equations

$$d\widetilde{X}_t^i = f(\widetilde{X}_t^i, \widetilde{A}_t^i)\, dt + G\, dW_t^i + (\widehat{P}_t^{xh} + QH^{\mathrm{T}})C^{-1}\, dI_t^i, \tag{83a}$$

$$d\widetilde{A}_t^i = \widehat{P}_t^{ah} C^{-1}\, dI_t^i, \tag{83b}$$

where the innovations are given by

$$dI_t^i = dY_t - \left(h(\widetilde{X}_t^i, \widetilde{A}_t^i)\, dt + HG\, dW_t^i + R^{1/2} dU_t^i \right), \tag{84}$$

and W_t^i and U_t^i denote independent N_x-dimensional and N_y-dimensional Brownian motions, respectively, for $i = 1, \ldots, M$.

The interacting particle Equation (83) can be time-stepped along the lines discussed in Section 4.3 for the pure state estimation formulation of the ensemble Kalman–Bucy filter.

6. Numerical Results

We now apply the generalised ensemble Kalman–Bucy filter formulation (83) with innovation (84) to five different model scenarios.

6.1. Parameter Estimation for the Ornstein–Uhlenbeck Process

Our first example is provided by the Ornstein–Uhlenbeck process

$$dX_t = a\, X_t\, dt + Q^{1/2} dW_t \tag{85}$$

with unknown parameter $a \in \mathbb{R}$, and known initial condition $X_0 = 1/2$. We assume an observation model of the form (3) with $H = 1$, and a measurement error taking values $R = 0.01$, $R = 0.0001$, and $R = 0$. The model error variance is set to either $Q = 0.5$ or $Q = 0.005$. Except for the case $R = 0$ a combined state and parameter estimation problem is to be solved. We implement the ensemble Kalman–Bucy filter (83) with innovation (84), step size $\Delta t = 0.005$, and ensemble size $M = 1000$. The data is generated using the Euler–Maruyama method applied to (85), with $a = -1/2$ and integrated over a time-interval $[0, 500]$ with the same step size. The prior distribution Π_0 for the

parameter is Gaussian with mean $\bar{a} = -1/2$ and variance $\sigma_a^2 = 2$. The results can be found in Figure 1. We find that the ensemble Kalman–Bucy filter is able to successfully identify the unknown parameter under all tested experimental settings, except for the largest measurement error case where $R = 0.01$. There, a small systematic offset of the estimated parameter value can be observed. One can also see that the variance in the parameter estimate monotonically decreases in time in all cases, while the variance in the state estimates approximately reaches a steady state.
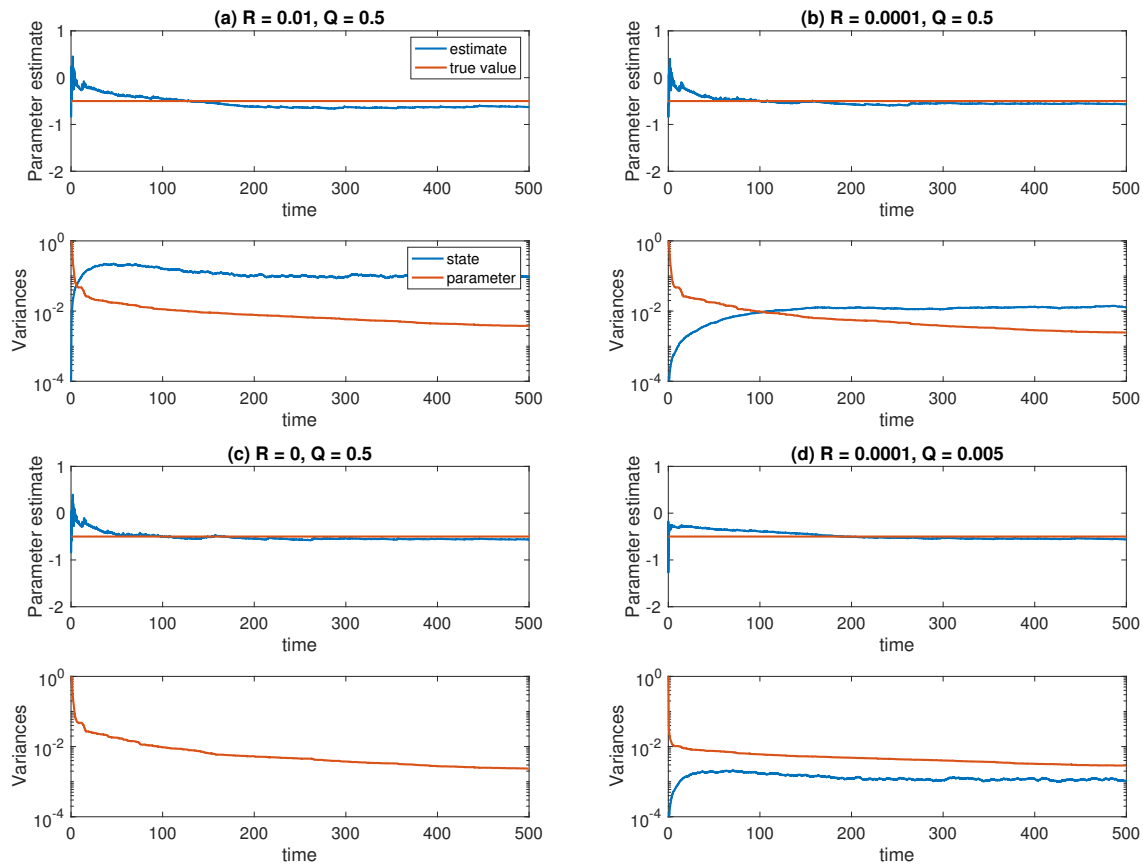
Figure 1. Results for the Ornstein–Uhlenbeck state and parameter estimation problem under different experimental settings: (**a**) $Q = 1/2$, $R = 0.01$; (**b**) $Q = 1/2$, $R = 0.0001$; (**c**) $Q = 1/2$, $R = 0$ (pure parameter estimation); (**d**) $Q = 0.005$, $R = 0.0001$. The ensemble size is set to $M = 1000$ in all cases. Displayed are the ensemble mean $\bar{a}_n$ and the ensemble variance in $\widetilde{A}_n$ and $\widetilde{X}_n$. The variance of $\widetilde{X}_n$ is zero when $R = 0$ in case (**b**).

6.2. Averaging

Consider the equations

$$\mathrm{d}Y_t = \left(1 - Z_t^2\right) Y_t \, \mathrm{d}t + Q^{1/2} \mathrm{d}W_t^y, \tag{86a}$$

$$\mathrm{d}Z_t = -\frac{\alpha}{\epsilon} Z_t \, \mathrm{d}t + \sqrt{\frac{2\lambda}{\epsilon}} \mathrm{d}W_t^z \tag{86b}$$

from [19] for $\lambda, \alpha, \gamma, \epsilon > 0$, and initial condition $Y_0 = 1/2$, $Z_0 = 0$. The reduced equations in the limit $\epsilon \to 0$ are given by (85), with parameter value

$$a = 1 - \frac{\lambda}{\alpha} \tag{87}$$

and initial condition $X_0 = 1/2$. The reduced dynamics corresponds to a (stable) Ornstein–Uhlenbeck process for $\lambda/\alpha > 1$. We wish to estimate the parameter a from observed increments

$$\Delta Y_n = Y_{n+1} - Y_n + \Delta t^{1/2} R^{1/2} \Xi_n, \qquad \Xi_n \sim N(0,1), \tag{88}$$

where the sequence of $\{Y_n\}_{n \geq 0}$ is obtained by time-stepping (86) using the Euler–Maruyama method with a step size Δt. We set $\lambda = 3$, $\alpha = 2$ (so that $a = -1/2$), $Q = 0.5$, and $\epsilon \in \{0.1, 0.01\}$ in our experiments. The measurement noise is set to $R = 0.01$ or $R = 0$ (pure parameter estimation).

We implement the ensemble Kalman–Bucy filter (83) with innovation (84), step size $\Delta t = \epsilon/50$, and ensemble size $M = 1000$ for the reduced Equation (87). The data is generated from an Euler–Maruyama discretization of (86) with the same step size. We also investigate the effect of subsampling the observations for $\epsilon = 0.01$ by solving (86) with step size $\Delta t = \epsilon/50$ and storing only every tenth solution Y_n, while the reduced equations and the ensemble Kalman–Bucy filter equations are integrated with $\Delta t = \epsilon/5$. The results are shown in Figure 2. Figure 3 shows the results for the same experiments repeated with a smaller ensemble size of $M = 10$. We find that the smaller ensemble size leads to more noisy estimates for the variance in $\widetilde{X}_n$ and a faster decay of the variance in $\widetilde{A}_n$, but the estimated parameter values are equally well converged. Subsampling does not lead to significant changes in the estimated parameter values. This is in contrast to the example considered next.
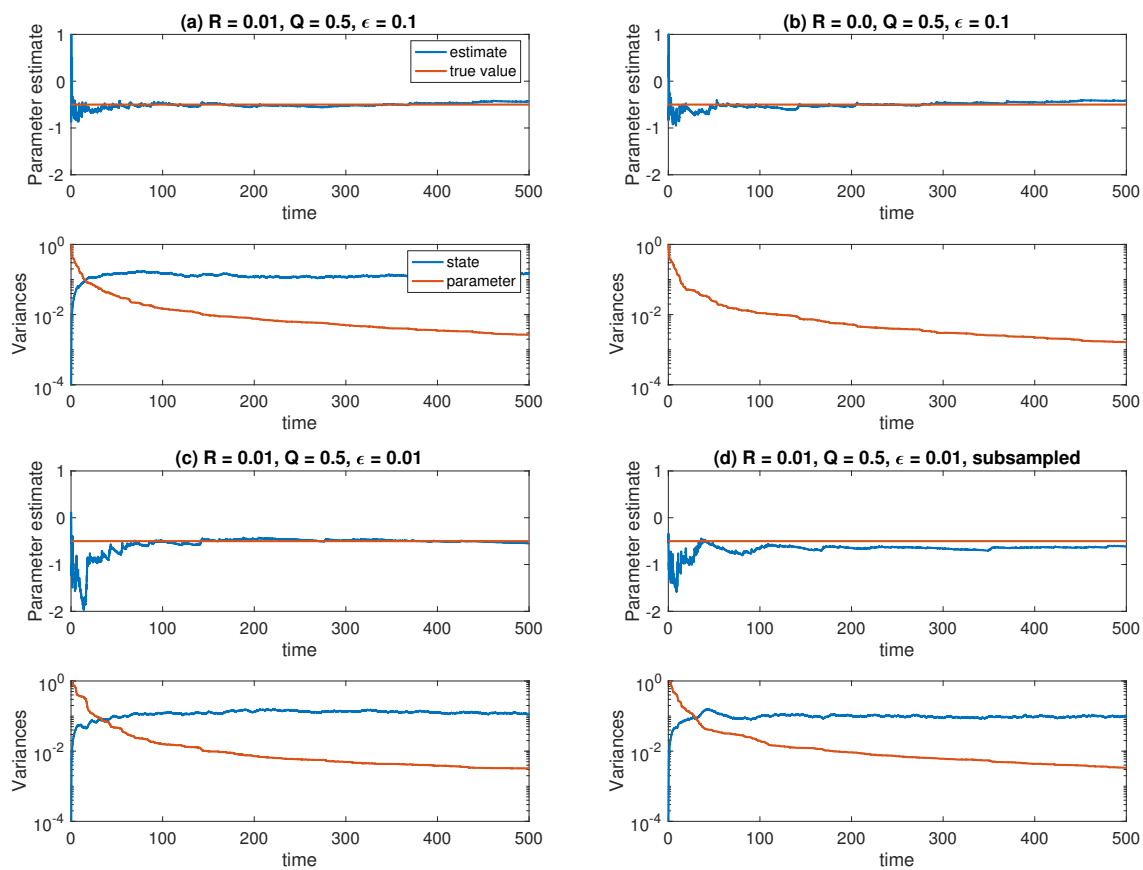

Figure 2. Results for the averaged Ornstein–Uhlenbeck state and parameter estimation problem under different experimental settings: (**a**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.1$; (**b**) $Q = 1/2$, $R = 0$, $\epsilon = 0.1$ (pure parameter estimation); (**c**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.01$; (**d**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.01$ and subsampling by a factor of ten. The ensemble size is set to $M = 1000$ in all cases. Displayed are the ensemble mean and the ensemble variance in $\widetilde{A}_n$ and $\widetilde{X}_n$. The variance of $\widetilde{X}_n$ is zero when $R = 0$ in case (**b**).
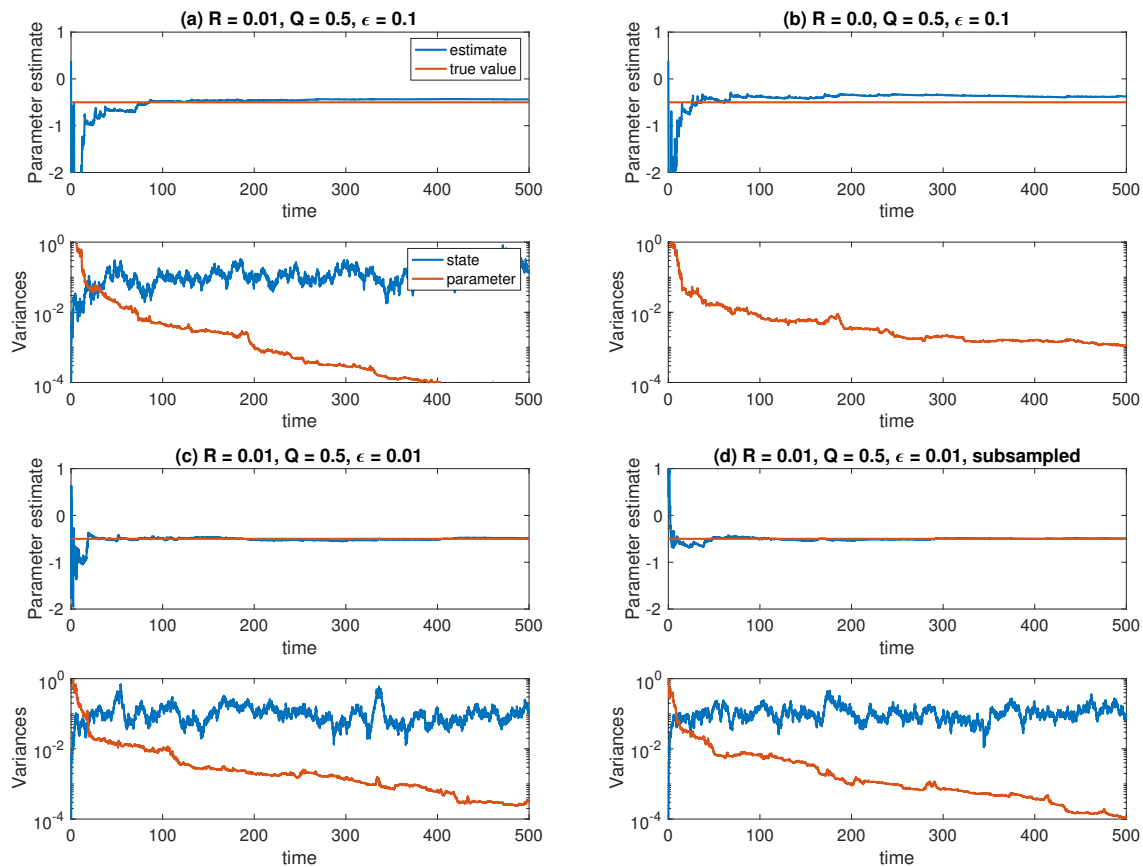
Figure 3. Results for the averaged Ornstein-Uhlenbeck process, now with a smaller ensemble size $M = 10$. Otherwise, panels (**a**–**d**) correspond to the same experimental settings as in Figure 2.

We finally mention [31] for alternative approaches to sequential estimation in the context of averaging using however different assumptions on the data.

6.3. Homogenisation

In this example, the data is produced by integrating the multi-scale SDE

$$dY_t = \left(\frac{\sqrt{\sigma/2}}{\epsilon} Z_t + a\, Y_t \right) dt, \tag{89a}$$

$$dZ_t = -\frac{1}{\epsilon^2} Z_t\, dt + \frac{\sqrt{2}}{\epsilon} dW_t^z \tag{89b}$$

with parameter values $\epsilon = 0.1$, $a = -1/2$, $\sigma = 1/2$, and initial condition $Y_0 = 1/2$, $Z_0 = 0$. Here, W_t^z denotes standard Brownian motion. The equations are discretised with step size $\Delta\tau = \epsilon^2/50 = 0.0002$, and the resulting increments (88) are stored over a time interval $[0, 500]$. See [32] for more details.

According to homogenisation theory, the reduced model is given by (85) with $Q = \sigma$, and we wish to estimate the parameter a from the data $\{\Delta Y_n\}$ produced according to (88). It is known that a standard maximum likelihood estimator (MLE) given by

$$a_{\mathrm{ML}} = \frac{\sum_n Y_{t_n} (Y_{t_{n+1}} - Y_{t_n})}{\sum_n Y_{t_n}^2 \Delta\tau} \tag{90}$$

leads to $a_{\mathrm{ML}} = 0$ in the limit $\Delta\tau \to 0$ and the observation interval $T \to \infty$. This MLE corresponds to $H = I$ and $R = 0$ in our extended state space formulation of the problem. Subsampling can be achieved by choosing an appropriate time-step $\Delta t > \Delta\tau$ in the ensemble Kalman–Bucy filter equations and a corresponding subsampling of the data points Y_n in (88). We used $\Delta t = 50\Delta\tau = 0.01$ and $\Delta t = 500\Delta\tau = 0.1$, respectively. The results can be found in Figure 4. It can be seen that only the larger subsampling leads to a correct estimate of the parameter a. This is in line with known results for the maximum likelihood estimator (90). See [32] and references therein.
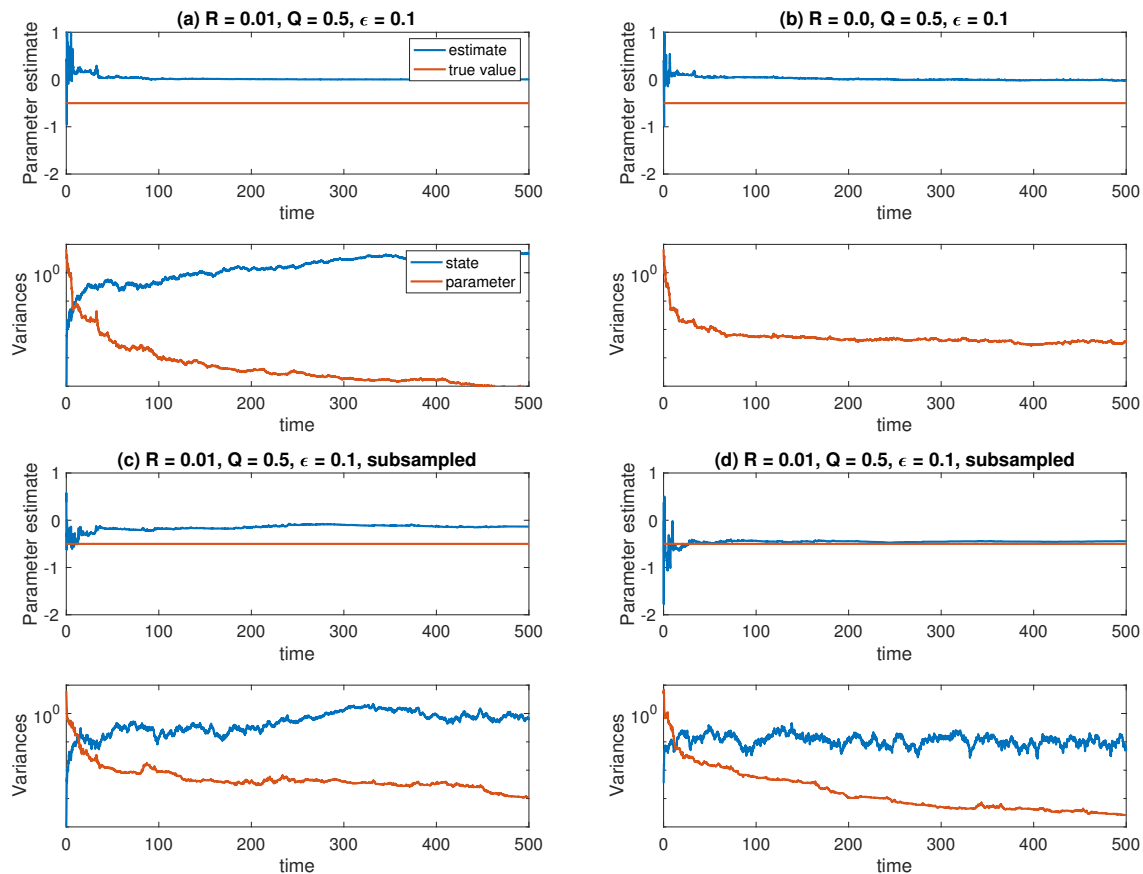
Figure 4. Results for the homoginsation Ornstein–Uhlenbeck state and parameter estimation problem under different experimental settings: (**a**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.1$; (**b**) $Q = 1/2$, $R = 0$, $\epsilon = 0.1$ (pure parameter estimation); (**c**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.1$ and subsampling by a factor of fifty; (**d**) $Q = 1/2$, $R = 0.01$, $\epsilon = 0.1$ and subsampling by a factor of five hundred. The ensemble size is set to $M = 10$ in all cases. Displayed are the ensemble mean and the ensemble variance in $\widetilde{A}_n$ and $\widetilde{X}_n$. The variance of $\widetilde{X}_n$ is zero under (**c**).

6.4. Nonparametric Drift and State Estimation

We consider nonparametric drift estimation for one-dimensional SDEs over a periodic domain $[0, 2\pi)$ in the setting considered from a theoretical perspective in [33]. There, a zero-mean Gaussian process prior $\mathcal{GP}(0, \mathcal{D}^{-1})$ is placed on the unknown drift function, with inverse covariance operator

$$\mathcal{D} := \eta[(-\Delta)^p + \kappa I]. \tag{91}$$

The integer parameter p sets the regularity of the process, whereas $\eta, \kappa \in \mathbb{R}^+$ control its characteristic correlation length and stationary variance.

Spatial discretization of the problem is carried out by first defining a grid of N_d evenly spaced points on the domain, at locations $x_i = i\Delta x$, $\Delta x = 2\pi/N_d$. The drift function is projected onto compactly supported functions centred at these points, which are piecewise linear with

$$b_i(x_j) = \delta_{ij} \tag{92}$$

and linear interpolation is used to define a drift function $f(x, a)$ for all $x \in [0, 2\pi)$, that is, it is of the form (2) with $f_0(x) \equiv 0$. In this example, we set $N_d = 200$. Sample realisations, as well as the reference drift f^*, can be found in Figure 5a.
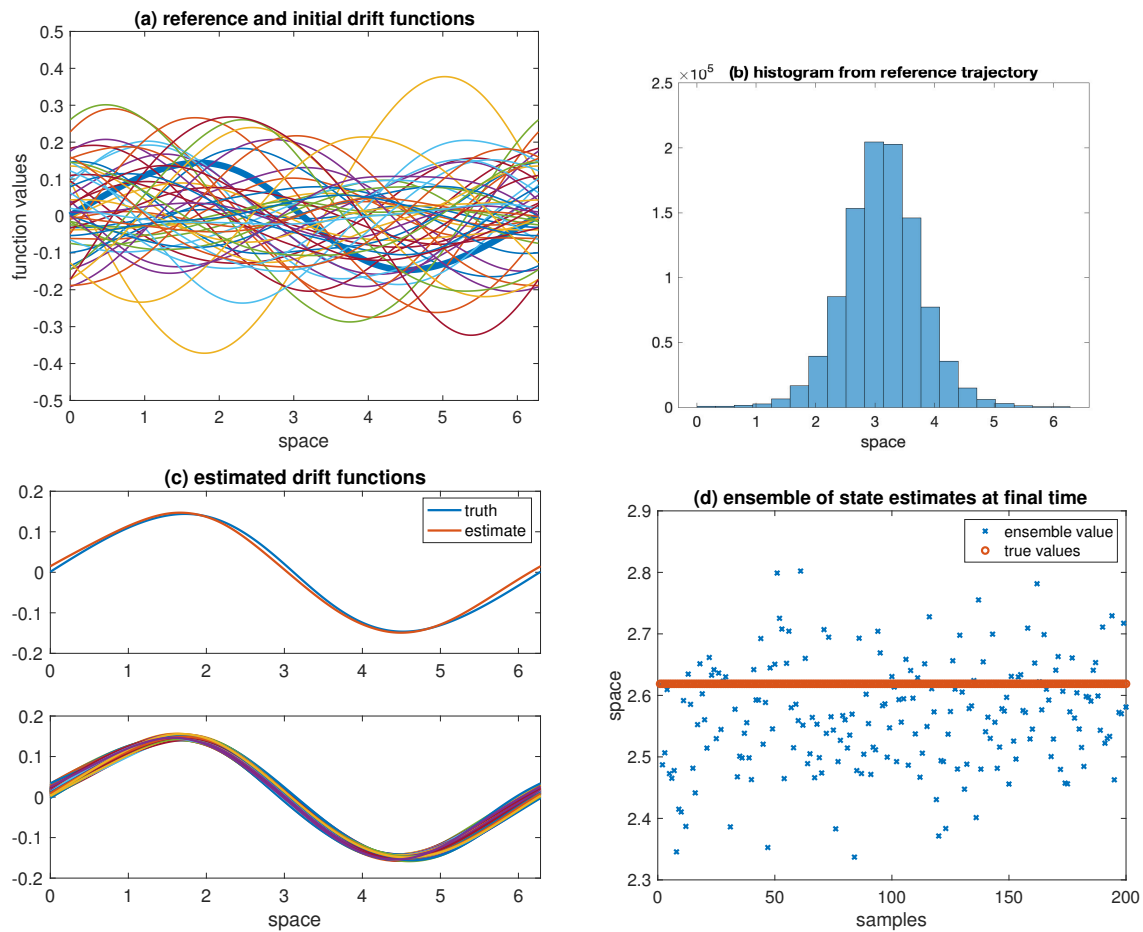
Figure 5. Results for the nonparametric drift and state estimation problem: (**a**) reference drift function (thick line) and ensemble of drift functions drawn from the prior distribution; (**b**) histogram of samples from the reference trajectory; (**c**) reference drift function and its estimate (**top**) and ensemble of drift functions (**bottom**) at final time; (**d**) ensemble of states and the true value at final time.

Data is generated by integrating the SDE (1) with drift f^* forward in time from initial condition $X_0 = \pi$ and with noise level $Q = 0.1$, using the Euler–Maruyama discretisation with step size $\Delta t = 0.1$ over one million time-steps. The spatial distribution of the solutions X_n is plotted in Figure 5b. The data is then given by

$$\Delta Y_n = X_{n+1} - X_n + \Delta t^{1/2} R^{1/2} \Xi_n \tag{93}$$

with $R = 0.00001$. Data assimilation is performed using the time-discretised ensemble Kalman–Bucy filter Equation (83) with innovation (84), ensemble size $M = 200$, and step size $\Delta t = 0.1$.

The final estimate of the drift function (ensemble mean) and the ensemble of drift functions can be found in Figure 5c. Figure 5d displays the ensemble of state estimates and the value of the reference

solution at the final time. We find that the ensemble Kalman–Bucy filter is able to successfully estimate the drift function and the model states. Further experiments reveal that the drift function can only be identified for sufficiently small measurement errors.

6.5. Spde Parameter Estimation

Consider the stochastic heat equation on the periodic domain $x \in [0, 2\pi)$, given in conservative form by the stochastic partial differential equation (SPDE)

$$\mathrm{d}u(x,t) = \nabla \cdot (\theta(x)\nabla u(x,t))\,\mathrm{d}t + \sigma^{1/2}\,\mathrm{d}W(x,t), \tag{94}$$

where $W(x,t)$ is space-time white noise. With constant $\theta(x) = \theta$, this SPDE reduces to

$$\mathrm{d}u(x,t) = \theta\Delta u(x,t)\mathrm{d}t + \sigma^{1/2}\,\mathrm{d}W(x,t). \tag{95}$$

In this example, we examine the estimation of θ from incremental measurements of a locally averaged quantity $q(x,t)$ that arises naturally in a standard finite volume discretisation of (95).

To discretise the system, one first defines $q_t^i = q(x_i, t)$ around $N_d = 200$ grid points x_i on a regular grid, separated by distances Δx, as

$$q_t^i = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} u(x,t)\mathrm{d}x. \tag{96}$$

The conservative (drift) term in (94) reduces to

$$\int_{x_i - \Delta x/2}^{x_i + \Delta x/2} \nabla \cdot (\theta(x)\nabla u(x,t))\,\mathrm{d}x = \theta_{i+1/2}\nabla u_t^{i+1/2} - \theta_{i-1/2}\nabla u_t^{i-1/2}, \tag{97}$$

where $\theta_{i\pm1/2} \equiv \theta(x_i + \Delta x/2)$, etc. The following standard finite difference approximations

$$\nabla u_t^{i+1/2} \simeq \frac{u_t^{i+1} - u_t^i}{\Delta x}, \quad u_t^i \simeq \Delta x^{-1}q_t^i \tag{98}$$

yield the N_d-dimensional SDE

$$\mathrm{d}q_t^i = \theta \left(\frac{q_t^{i+1} - 2q_t^i + q_t^{i-1}}{\Delta x^2} \right) \mathrm{d}t + \sigma^{1/2}\Delta x^{1/2}\mathrm{d}W_t^i \tag{99}$$

for constant θ, where W_t^i are independent one-dimensional Brownian motions in time.

Following recent results from [20] we consider the case of estimation of a constant $a = \theta$ value from measurements $\mathrm{d}q_t^*$ at a fixed location/index $j^* \in \{1, \ldots, N_d\}$. The data trajectory is thus given by

$$\mathrm{d}Y_t = \mathrm{d}q_t^* + R^{1/2}\mathrm{d}V_t \tag{100}$$

where $R^{1/2}$ is a scalar and V_t is a standard Brownian motion in one dimension. We perform numerical experiments in which the initial state q_0^i is set to zero for all indices i and the prior on the unknown parameter $a = \theta$ is uniform over the interval $[0.2, 1.8]$.

The increment data is generated by first integrating (95) forward in time from the known initial condition $q_i(0) = 0$ for all i. The equation is discretised in time using the Euler-Maruyama method. It is known that $\Delta t < \theta\Delta x^2/2$ is required for stability of the Euler–Maruyama discretisation; we use the much smaller time step $\Delta t = \Delta x^2/80$. The solution is sampled with this same time step, and increment

measurements are approximated at time t_n by setting the measurement noise level R to zero in (100), resulting in

$$\Delta Y_n = q_{n+1}^* - q_n^*. \tag{101}$$

Please note that the associated model error in (1) is given by $G = \sigma^{1/2} \Delta x^{1/2} I$ and the matrix H in (3) projects the vector of state increments onto a single component with index $j^* = N_d/2$. Simulations are performed over the time-interval $[0, 20]$. The results can be found in Figure 6a. We also compute the model evidence for a sequence of parameter values $\theta \in \{0.2, 0.3, \ldots, 1.8\}$ based on a standard Kalman–Bucy filter [6] for the associated linear state estimation problem. See Figure 6b. Both approaches agree with the reference value $\theta = 1$.
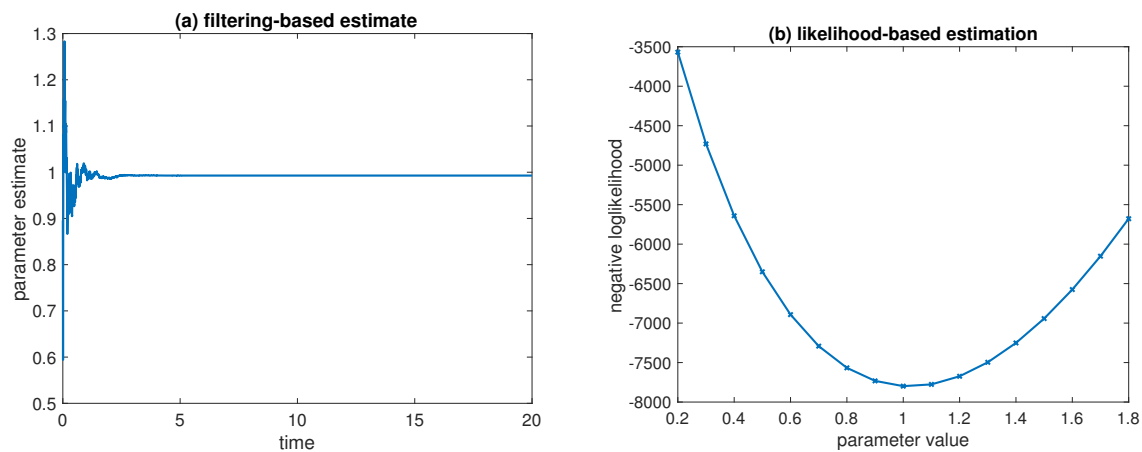

Figure 6. Results for SPDE parameter estimation: (**a**) estimate of θ as a function of time as obtained by the ensemble Kalman–Bucy filter; (**b**) evidence based on a Kalman–Bucy filter for state estimation applied to a sequence of parameter values $\theta \in \{0.2, 0.3, \ldots, 1.8\}$.

6.6. Discussion

The results presented here demonstrate that the proposed methodology can be applied to a broad range of continuous-time state and parameter estimation problems with correlated measurement and model errors. Alternatively, one could have employed standard SMC or MCMC methods utilising the modified observation Model (41) as implied by the Kushner–Stratonovitch formulation (39) of the filtering problem. However, such implementations require the approximation of the additional $Q\nabla \log \pi_t$ term which is nontrivial if only samples from π_t are available. Furthermore, the limiting behaviour of such implementations in the limit $R \to 0$ and $H = I$ (pure parameter estimation problem) is unclear since π_t degenerates into a Dirac delta distribution, potentially leading to numerical difficulties in this singular regime. The proposed generalised feedback particle filter formulation avoids these issues through the use of stochastic innovations which are correlated with the model noise. In other words, the distribution π_t does not appear explicitly in the innovation process (46), and the correlated noise terms cancel each other out as discussed in Remark 3 for $R = 0$ and $H = I$. The main computational challenge of the feedback particle filter approach is given by the need for finding the Kalman gain matrix (57). However, the constant gain ensemble Kalman–Bucy approximation

$$K_t \approx \left(P^{xh} + QH^{\mathrm{T}} \right) C^{-1} \tag{102}$$

is easy to implement. In fact, the only differences with the standard ensemble Kalman–Bucy filter formulation of [14] are in the additional QH^{T} term in the Kalman gain, and a correlation between the stochastic innovation process and the model error. While the ensemble Kalman–Bucy filter gave rather satisfactory results for the numerical experiments displayed in Section 6, strongly non-Gaussian

distributions might require more accurate approximations to the Kalman gain matrix (57). In that case, one could rely on the particle-based diffusion map approximation considered in [27].

7. Conclusions

In this paper, we have derived McKean–Vlasov equations for combined state and parameter estimation from continuously observed state increments. An approximate and robust implementation of these McKean–Vlasov equations in the form of a generalised ensemble Kalman–Bucy filter has been provided and applied to a range of increasingly complex model systems. Future work will address the treatment of temporally-correlated measurement and model errors, as well as a rigorous analysis of these McKean–Vlasov equations in the contexts of multi-scale dynamics and nonparametric drift estimation.

Author Contributions: Methodology, N.N. and S.R.; software, S.R. and P.J.R.; validation, N.N., S.R. and P.J.R.; writing—original draft preparation, N.N., S.R.; writing—review and editing, N.N., S.R. and P.J.R.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. The Filtering Equations for Correlated Noise

In this appendix we outline a derivation of the Kushner-Stratonovich Equation (39) for the signal-observation dynamics given by (38). In fact, we only compute the evolution equation (termed *modified Zakai equation*) for the unnormalised filtering distribution $\rho_t[\phi] = \mathbb{E}\left[l_t\phi(X_t)|Y_{[0,t]}\right]$, where the likelihood l_t is given by

$$l_t \equiv l(Y_{[0,t]}|X_{[0,t]}) = \exp\left(\int_0^t f(X_s)^\mathsf{T} H^\mathsf{T} C^{-1}\,\mathrm{d}Y_s - \frac{1}{2}\int_0^t f(X_s)^\mathsf{T} H^\mathsf{T} C^{-1} Hf(X_s)\,\mathrm{d}s\right). \qquad (A1)$$

Obtaining the Kushner-Stratonovich formulation is then standard, applying Itô's formula to the Kallianpur-Striebel formula $\pi[\phi] = \rho_t[\phi]/\rho_t[\mathbf{1}]$, see ([7], Chapter 3). The following result is in agreement with the corollaries 3.39 and 3.40 in [7].

Lemma A1. *The modified Zakai equation is given by*

$$\rho_t[\phi] = \rho_0[\phi] + \int_0^t \rho_s[\mathcal{L}\phi]\,\mathrm{d}s + \int_0^t \rho_s\left[\phi f^\mathsf{T} H^\mathsf{T} C^{-1}\right]\mathrm{d}Y_s + \int_0^t \rho_s\left[\nabla\phi\right] QH^\mathsf{T} C^{-1}\,\mathrm{d}Y_s, \qquad (A2)$$

where the generator $\mathcal{L}$ has been defined in (40).

Proof. For convenience, let us define the process

$$M_t = \int_0^t f(X_s)^\mathsf{T} H^\mathsf{T} C^{-1}\mathrm{d}Y_s, \qquad (A3)$$

where Y_s satisfies (38b). From $\langle Y\rangle_t = Ct$ we see that

$$\langle M\rangle_t = \int_0^t f(X_s)^\mathsf{T} H^\mathsf{T} C^{-1} Hf(X_s)\,\mathrm{d}s. \qquad (A4)$$

Hence, the likelihood takes the form

$$l_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right), \qquad (A5)$$

satisfying the SDE

$$\mathrm{d}l_t = l_t \, \mathrm{d}M_t. \tag{A6}$$

For an arbitrary smooth compactly supported test function ϕ, Itô's formula implies

$$l_t \phi(X_t) = \phi(X_0) + \int_0^t \phi(X_s) \, \mathrm{d}l_s + \int_0^t l_s \nabla \phi(X_s) \cdot \mathrm{d}X_s \tag{A7a}$$

$$+ \frac{1}{2} \sum_{i,j=1}^{N_x} \int_0^t l_s \partial_i \partial_j \phi(X_s) \, \mathrm{d}\langle X^i, X^j \rangle_s + \sum_{i=1}^{N_x} \int_0^t \partial_i \phi(X_s) \, \mathrm{d}\langle l, X^i \rangle_s, \tag{A7b}$$

where X_s satisfies (38a). For the covariation process $\langle l, X \rangle_t$ we obtain

$$\langle l, X \rangle_t = l_t \langle M, X \rangle_t = l_t f(X_t)^{\mathrm{T}} H^{\mathrm{T}} C^{-1} H Q t, \tag{A8}$$

using $\langle Y, X \rangle_t = HQt$. Furthermore, $\langle X, X \rangle_t = Qt$, which follows from the definition of the stochastic contributions in (38a).

We now apply the conditional expectation to (A7). Noticing that

$$\int_0^t \phi(X_s) \, \mathrm{d}l_s = \int_0^t l_s \phi(X_s) f(X_s)^{\mathrm{T}} H^{\mathrm{T}} C^{-1} \, \mathrm{d}Y_s, \tag{A9}$$

the result follows from (A6). $\quad\square$

References

1. Kutoyants, Y. *Statistical Inference for Ergodic Diffusion Processes*; Springer: New York, NY, USA, 2004.
2. Pavliotis, G. *Stochastic Processes and Applications*; Springer: New York, NY, USA, 2014.
3. Apte, A.; Hairer, M.; Stuart, A.; Voss, J. Sampling the posterior: An approach to non-Gaussian data assimilation. *Phys. D Nonlinear Phenom.* **2007**, *230*, 50–64. [CrossRef]
4. Salman, H.; Kuznetsov, L.; Jones, C.; Ide, K. A method for assimilating Lagrangian data into a shallow-water-equation ocean model. *Mon. Weather Rev.* **2006**, *134*, 1081–1101. [CrossRef]
5. Apte, A.; Jones, C.; Stuart, A. A Bayesian approach to Lagrangian data assimilation. *Tellus A* **2008**, *60*, 336–347. [CrossRef]
6. Simon, D. *Optimal State Estimation*; Wiley: Hoboken, NJ, USA, 2006.
7. Bain, A.; Crisan, D. *Fundamentals of Stochastic Filtering*; Springer: New York, NY, USA, 2009.
8. Liu, J. *Monte Carlo Strategies in Scientific Computing*; Springer: New York, NY, USA, 2001.
9. Crisan, D.; Xiong, J. Approximate McKean-Vlasov representation for a class of SPDEs. *Stochastics* **2010**, *82*, 53–68. [CrossRef]
10. McKean, H. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1907–1911. [CrossRef] [PubMed]
11. Yang, T.; Mehta, P.; Meyn, S. Feedback particle filter. *IEEE Trans. Autom. Control* **2013**, *58*, 2465–2480. [CrossRef]
12. Reich, S. Data assimilation: The Schrödinger perspective. *Acta Numer.* **2019**, *28*, 635–710.
13. Majda, A.; Harlim, J. *Filtering Complex Turbulent Systems*; Cambridge University Press: Cambridge, UK, 2012.
14. Bergemann, K.; Reich, S. An ensemble Kalman–Bucy filter for continuous data assimilation. *Meteorol. Z.* **2012**, *21*, 213–219. [CrossRef]
15. Taghvaei, A.; de Wiljes, J.; Mehta, P.; Reich, S. Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. *ASME. J. Dyn. Syst. Meas. Control* **2017**, *140*. [CrossRef]
16. Law, K.; Stuart, A.; Zygalakis, K. *Data Assimilation: A Mathematical Introduction*; Springer: New York, NY, USA, 2015.
17. Reich, S.; Cotter, C. *Probabilistic Forecasting and Bayesian Data Assimilation*; Cambridge University Press: Cambridge, UK, 2015.
18. Moral, P.D. *Mean Field Simulation for Monte Carlo Integration*; Chapman and Hall/CRC: London, UK, 2013.
19. Pavliotis, G.; Stuart, A. *Multiscale Methods*; Springer: New York, NY, USA, 2008.

20. Altmeyer, R.; Reiß, M. *Nonparametric Estimation for Linear SPDEs from Local Measurements*; Technical Report; Humboldt University Berlin: Berlin, Germany, 2019.

21. Saha, S.; Gustafsson, F. Particle filtering with dependent noise processes. *IEEE Trans. Signal Process.* **2012**, *60*, 4497–4508. [CrossRef]

22. Berry, T.; Sauer, T. Correlations between systems and observation errors in data assimilation. *Mon. Weather Rev.* **2018**, *146*, 2913–2931. [CrossRef]

23. Mitchell, H.L.; Daley, R. Discretization error and signal/error correlation in atmospheric data assimilation: (I). All scales resolved. *Tellus A* **1997**, *49*, 32–53. [CrossRef]

24. Papaspiliopoulos, O.; Pokern, Y.; Roberts, G.; Stuart, A. Nonparametric estimation of diffusion: A differential equation approach. *Biometrika* **2012**, *99*, 511–531. [CrossRef]

25. Särkkä, S. *Bayesian Filtering and Smoothing*; Cambridge University Press: Cambridge, UK, 2013.

26. Laugesen, R.S.; Mehta, P.G.; Meyn, S.P.; Raginsky, M. Poisson's equation in nonlinear filtering. *SIAM J. Control Optim.* **2015**, *53*, 501–525. [CrossRef]

27. Taghvaei, A.; Mehta, P.; Meyn, S. *Gain Function Approximation in the Feedback Particle Filter*; Technical Report; University of Illinois at Urbana-Champaign: Champaign, IL, USA, 2019.

28. Amezcua, J.; Kalnay, E.; Ide, K.; Reich, S. Ensemble transform Kalman-Bucy filters. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 995–1004. [CrossRef]

29. De Wiljes, J.; Reich, S.; Stannat, W. Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise. *SIAM J. Appl. Dyn. Syst.* **2018**, *17*, 1152–1181. [CrossRef]

30. Blömker, D.; Schillings, C.; Wacker, P. A strongly convergent numerical scheme for ensemble Kalman inversion. *SIAM J. Numer. Anal.* **2018**, *56*, 2537–2562. [CrossRef]

31. Harlim, J. Model error in data assimilation. In *Nonlinear and Stochastic Climate Dynamics*; Franzke, C., Kane, T.O., Eds.; Cambridge University Press: Cambridge, UK, 2017; pp. 276–317.

32. Krumscheid, S.; Pavliotis, G.; Kalliadasis, S. Semi-parametric drift and diffusion estimation for multiscale diffusions. *SIAM J. Multiscale Model. Simul.* **2011**, *11*, 442–473. [CrossRef]

33. Van Waaij, J.; van Zanten, H. Gaussian process methods for one-dimensional diffusion: Optimal rates and adaptation. *Electron. J. Stat.* **2016**, *10*, 628–645. [CrossRef]