Approximations to the solution of the Kushner-Stratonovich equation for the stochastic chemostat

José Augusto Fontenele Magalhães[a,*], Muhammad Fuady Emzir[b] and Francesco Corona[a]

[a]Department of Chemical and Metallurgical Engineering, Aalto University, Finland
[b]Control and Instrumentation Engineering Department, King Fahd University of Petroleum and Minerals, Saudi Arabia

ARTICLE INFO

Keywords:
stochastic chemostat
state estimation
non-linear filtering
Fokker-Planck equation
Kushner-Stratonovich equation

ABSTRACT

In order to characterise the dynamics of a biochemical system such as the chemostat, we consider a differential description of the evolution of its state under environmental fluctuations. We present solutions to the filtering problem for a chemostat subjected to geometric Brownian motion. Under this modelling assumption, our best knowledge about the state of the system is given by its distribution in time, given the distribution of the initial state. Such a function solves a deterministic partial differential equation, the Kolmogorov forward equation. In this paper, however, we refine our knowledge about the state of the chemostat when additional information about the system is available in the form of measurements. More formally, we are interested in obtaining the distribution of the state conditional on measurements as the solution to a non-linear stochastic partial integral differential equation, the Kushner-Stratonovich equation. For the chemostat, this solution is not available in closed form and it must be approximated. We present approximations to the solution to the Kushner-Stratonovich equation based on methods for partial differential equations. We compare the solution with a linearisation method and with a classical sequential Monte Carlo method known as the Bootstrap particle filter.

1. Introduction

The aim of filtering is to estimate the current state of an evolving dynamical system, customarily modelled by a stochastic process, denoted by X and called the signal process. The signal process cannot be measured directly, but only via a related process Y, termed the observation process. The aim is to compute π_t, the conditional distribution of the signal X_t, at the current time t given the observation data accumulated up to this time. More precisely, we compute the conditional distribution of the signal X_t given the observation filtration $\mathcal{F}_t^Y = \sigma\{Y_s : s \leq t\}$. It is possible to formalise this by defining a stochastic process $\pi = (\pi_t, t \geq 0)$ describing this conditional distribution. Such process is the solution to a non-linear stochastic partial integral differential equation, the Kushner-Stratonovich equation. The exact form of the process π can be derived for a specific class of problems, for instance, those with a linearly evolving observation process (Kálmán and Bucy, 1961; Beneš, 1981). Even in this scenario, however, the signal is only allowed to have a non-linear drift under some restrictive conditions (see Beneš condition (Beneš, 1981)). In general, π is an infinite-dimensional stochastic process and finding it in closed-form is a non-trivial task due to the inherent non-linearities of the system. In this case, the solution must be approximated.

In this work, our goal is to approximate the conditional distribution π_t for a process performed in a chemostat in order to learn more about the system. The chemostat was the major player in the use of continuous culture systems to characterise certain kinds of microbes in the mid-20th century, prior to when molecular biology and thence genomics took centre stage instead (Bull, 2010; Harmand et al., 2017). Being a simple, realistic and thoroughly tested continuous culture system, it has regained some interest in this post-genomic era. For instance, chemostat-grown cultures lead to reproducible and homogeneous results, which are necessary to identify the roles genes play in the biology of organisms (Hoskisson and Hobbs, 2005).

We focus on estimating the state of a chemostat to understand its operation and to devise strategies for optimal operation. Here, we consider a substrate-controlled growth model (Monod kinetics) and a more complex model, where an excessive level of the substrate can inhibit the growth (Haldane kinetics). We consider a differential description of the substrate and biomass concentrations in a chemostat under environmental fluctuations (Imhof and Walcher,

augusto.magalhaes@aalto.fi (J.A.F. Magalhães)
ORCID(s): 0000-0002-0795-665X (J.A.F. Magalhães); 0000-0002-1855-2124 (M.F. Emzir); 0000-0002-3615-1359 (F. Corona)

2005); in particular, chemostats subjected to geometric Brownian motion. Our best knowledge about the state of the system is given by its conditional distribution in time, given the distribution of the initial state. Such a function solves a deterministic partial differential equation, the Kolmogorov forward equation. Our focus is to refine our knowledge about the state of the chemostat when additional information about the system is available in the form of measurements. To describe how measurements and states are related, we augment the dynamics with an observation model in which the measurements are also subjected to Brownian motion. When measurements are collected, our estimate of the state is given by the conditional distribution π_t, now given the distribution of the initial state and the collected measurements.

The most common approach to find π_t for non-linear systems is the extended Kalman filter (EKF), which linearises certain non-linearities in the model so that the traditional linear Kalman-Bucy filter (Kálmán and Bucy, 1961) can be applied. A less common approach consists of solving an associated stochastic partial differential equation, the Zakai equation. Its solution is an unnormalised density function, given the distribution of the initial state and the collected measurements. In this case, the solution is also unavailable in closed form for our system and it must be approximated numerically (see, for example, Chapter 7 in Bain and Crisan (2008)). Although this is a stochastic partial differential equation, classical partial differential equations methods can still be applied. These methods are successful in low dimensions of the state space but cannot be applied to high-dimensional problems. This is because the total number of finite volumes required for an accurate approximation goes up exponentially with the dimension of the signal.

Among other numerical methods for solving the filtering problem, particle filters (also known as sequential Monte Carlo method) are among the most popular ones. In this approximation, particles (also called samples) firstly evolve according to a desirable transition kernel and then, they interact selectively in accordance with their likelihood with respect to the measurements.

In the context of the stochastic chemostat, this paper presents results on the approximation to π_t with a method for partial differential equations. We work with a scenario where the available observations are a transformed and noisy version of the signal. In particular, we observe the flow rate of outgoing biogas produced inside the chemostat, and not the biomass and substrate concentrations directly. We present alternative scenarios in the supplementary material, in which noisy observations of biomass and/or substrate concentrations are available. After presenting how the solution to π_t looks like, we evaluate the method for partial differential equations by using the Hellinger distance. More specifically, we quantify the similarity between the resulting approximation and the solution approximated via sequential Monte Carlo. We also present how the extended Kalman-Bucy filter performs under the same scenario.

This paper is structured as follows: we begin Section 2 by introducing some basic notation and the equations governing the dynamics of a chemostat. The rest of the section is devoted to the exposition of two scenarios, one with Monod kinetics and the other with Haldane kinetics. In Section 3, we present a detailed description of the stochastic chemostat; in particular, our goal is to showcase the prior probabilistic knowledge about the chemostat given some initial condition of the system. In Section 4, we have additional information in the form of measurements, concluding the section with the filtering equations for stochastic observations corrupted by Gaussian noise. In Sections 5 and 6, we introduce some approaches used to solve the filtering problem and we conclude the paper by comparing them.

2. Chemostat equations

The chemostat is a bioreactor in which one or more populations of microorganisms grow in a nutrient medium consisting of a cocktail of various molecules. Here, we call biomass the set of growing microorganisms and substrate the cocktail of various molecules. Let b denote the concentration of biomass and s the concentration of substrate. The following reaction takes place to produce one unit of biomass and γ units of methane (biogas) from κ units of substrate, with reaction rate $\rho(\cdot)$:

$$\kappa s \xrightarrow{\rho(\cdot)} b + \gamma CH_4. \tag{1}$$

Let $F_{in}(t)$ denote the amount of media flowing into a vessel of volume $V(t)$ per hour, and $F_{out}(t)$ denote the outgoing amount departing from the same vessel per hour. Here, the focus is on bioreactor processes performed at continuous mode, where at a specific time t, $F_{in}(t) = F_{out}(t) = F(t)$ and $V(t) = V$. Consequently, the dilution rate is given by $D(t) = F(t)/V$.

The ordinary differential model for the growth of a single species in such a chemostat is given by

$$\frac{db(t)}{dt} = (\mu(\cdot) - D(t))b(t) + D(t)b_{in}(t), \quad \frac{ds(t)}{dt} = D(t)(s_{in}(t) - s(t)) - \kappa\mu(\cdot)b(t), \tag{2}$$

with initial concentrations b_0 and s_0 and incoming concentrations $b_{in}(t)$ and $s_{in}(t)$. Here, μ is the "specific growth velocity", a function which models the growth of an organism population and which guarantees that the reaction rate is zero in the absence of biomass. See Appendix A for the derivation from a mass balance equation.

In this work, we consider the case of no incoming flow of biomass, thus $b_{in}(t) = 0$, $t \geq 0$. We work with the minimal model (Harmand et al., 2017), where we assume that (i) $\mu(\cdot)$ is a function of the substrate only, with $\mu(s) > 0$ for $s > 0$ and $\mu(0) = 0$, and that (ii) the yield coefficient κ is constant. These assumptions are to be interpreted in the following way: (i) the velocity - or kinetics - of the reaction in Eq. (1) does not change as the biomass concentration changes and (ii) temperature and pressure are kept constant, thus κ is also constant.

Now, let $x(t)$ denote the state of the chemostat at instant t, that is $x(t) = (b(t),\ s(t))$. We begin with the ordinary differential equation (ODE) on the signal $x(t)$ to describe the dynamics of the minimal system:

$$d\underbrace{\begin{bmatrix} b(t) \\ s(t) \end{bmatrix}}_{x(t)} = \underbrace{\begin{bmatrix} (\mu(s(t)) - D(t))b(t) \\ D(t)(s_{in}(t) - s(t)) - \kappa\mu(s(t))b(t) \end{bmatrix}}_{f(t,x(t),u(t)|\theta)} dt, \qquad \underbrace{\begin{bmatrix} b(0) \\ s(0) \end{bmatrix}}_{x(0)} = \underbrace{\begin{bmatrix} b_0 \\ s_0 \end{bmatrix}}_{x_0}, \tag{3}$$

where f is a function parametrised by a vector θ of parameters, and $u(t)$ is input at instant t, which for the minimal model refers to $u(t) = (D(t), s_{in}(t))$. If the state of the system has a known starting position, say point x_0 at $t = 0$, and $u(t)$ is given, the function $x(t)$, also called a solution to Eq. (3), is then completely determined in the future, i.e. we know $x(t)$ for all $t > 0$.

In this paper, we consider chemostat cultures with a single growth-limiting substrate. We show two cases for which different types of organisms are being cultivated, each with its own growth function.

Case 1 - Monotonic growth μ - Monod type

This first case consists of the Monod relation of growth (Monod, 1950), which assumes that the growth rate is zero when there is no substrate and tends to an upper limit when the substrate is in great excess.

$$\mu(s) = \frac{\mu_{max} s}{K_s + s}. \tag{4}$$

Here, μ depends on the parameters maximum growth μ_{max} and half-saturation coefficient K_s.

Case 2 - Non-monotonic growth μ - Haldane type

Some experiments indicate that excessive substrate may inhibit the growth of a microbial population in the biomass (Andrews, 1968). To consider such scenario, this case works with the following growth function

$$\mu(s) = \frac{\mu_0 s}{K_s + s + I_s s^2}. \tag{5}$$

Here, μ depends on the growth parameteer μ_0, half-saturation coefficient K_s and inhibitor coefficient I_s.

Let us now see how the solution to Eq. (3) looks like, given the parameters from Table 1.

Table 1: Model parameters and inputs for Eq. (3).

	Model parameters		Inputs	
	Specific growth	κ	$D(t) = D\,(V = 100\text{L})$ (h^{-1})	$s_{in}(t) = s_{in}$ (mgL^{-1})
Case 1	$\frac{0.3s}{10+s}$	10	0.01	100.0
Case 2	$\frac{5s}{0.5+s+5s^2}$	1	0.7	2.0

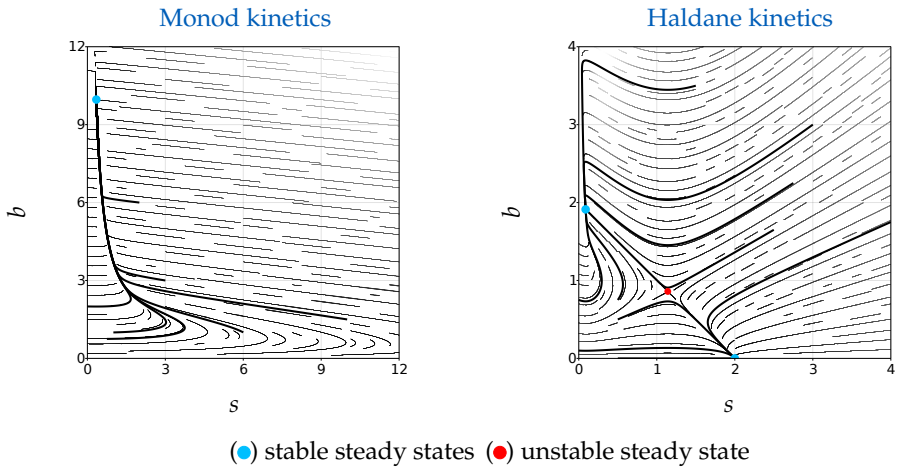

Figure 1: Left: trajectories from the deterministic model with growth function of the Monod type. Right: trajectories from the deterministic model with growth function of the Haldane type.

Consider equilibrium pairs (b_e, s_e) for which the right-hand side of Eq. (3) is zero. Case 1 has the equilibrium point $E_1^{(1)} = (9.966, 0.345)$. Furthermore, $E_1^{(1)}$ is a locally exponentially stable equilibrium (Harmand et al., 2017, Section 2.1.2). Case 2 has a trivial equilibrium point $E_2^{(1)} = (0, s_{in})$. In addition, it has interior equilibria $E_2^{(2)} = (0.859, 1.141)$ and $E_2^{(3)} = (1.912, 0.088)$. It is possible to verify that one of the two interior equilibria is a stable nodal point and the other is a saddle point (Harmand et al., 2017, Section 2.1.3).

3. Continuous-time dynamic models

3.1. Stochastic differential equations

3.1.1. Definition

Now let us consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A solution $x = (x(t), t \geq 0)$ to an ODE becomes a stochastic process $X = (X_t, t \geq 0)$ the moment we introduce some randomness by (i) randomising the initial condition and (ii) introducing a stochastic driving process, e.g. Brownian motion. This procedure results in a class of differential equations known as stochastic differential equations (SDEs), whose solutions involve ordinary and stochastic integrals.

We consider integration with respect to Brownian motion and its properties only. Here, the differential equations are called Itô stochastic differential equations, and the stochastic integrals are called Itô integrals.

Definition 3.1. *A stochastic process $B = (B_t, t \geq 0)$ is called a d-dimensional (standard) Brownian motion if the following conditions are satisfied:*

1. *It starts at zero, i.e. $B_0 = 0$ almost surely;*
2. *$B_t(\omega)_{\omega\in\Omega}$ has independent increments, so that for any $v \geq 0$, the random variable $B_{t+v} - B_t$ is independent of the σ-algebra generated by $(B_s, 0 \leq s \leq t)$;*
3. *For all $0 \leq s < t$, the random variable $B_t - B_s$ is normally distributed with mean 0 and variance $(t - s)$, i.e. $B_t - B_s \sim \mathcal{N}(0, (t - s) \cdot I_{d\times d})$, where $I_{d\times d}$ stands for identity matrix of size d;*
4. *It has almost surely continuous sample paths.*

If we let the Brownian motion $B^x = (B_t^x, t \leq 0)$ drive the dynamics seen in Eq. (3), the following stochastic differential equation emerges:

$$dX_t = f(t, X_t, u(t)|\theta)dt + g(t, X_t, u(t)|\theta)dB_t^x, \quad X_0 = \xi, \tag{6}$$

where $\xi = \xi(\omega)_{\omega\in\Omega}$ is a random variable on Ω, and the deterministic functions f and g are the coefficients of the equation. For notation simplicity, we omit the reference to the vector θ of parameters. We assume the input $u(t)$ to be constant and given for all times t, and f and g to be autonomous functions such that Eq. (6) can be rewritten as

$$dX_t = f(X_t)dt + g(X_t)dB_t^x, \qquad X_0 = \xi, \tag{7a}$$

or equivalently in coordinate form

$$dX_t^{d_x} = f_{d_x}(X_t)dt + \sum_{m=1}^{M} g_{d_x m}(X_t)dB_t^{x,m}, \qquad X_0^{d_x} = \xi^{d_x}, \quad \text{for } d_x = 1, \dots, D_x, \tag{7b}$$

where $B^x = (B^{x,1}, \dots, B^{x,M})$ is an M-dimensional standard Brownian motion ($M \geq 1$). Here, the state process and the coefficients of the equation have the following mapping:

$$X : [0, \infty) \times \Omega \to \mathbb{R}^{D_x}, \quad f : \mathbb{R}^{D_x} \to \mathbb{R}^{D_x}, \quad g : \mathbb{R}^{D_x} \to \mathbb{R}^{D_x} \times \mathbb{R}^{M}.$$

The stochastic differential equations in (7) are equivalent to the following stochastic integral equation:

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t g(X_s)dB_s^x, \quad t \geq 0, \tag{8}$$

where the first integral on the right-hand side is a Riemann integral, and the second one is an Itô integral. Here, f and g satisfy the Lipschitz condition for existence and uniqueness of solutions. In order to demonstrate challenges concerning SDE existence and uniqueness, we now discuss what we mean by solutions of Eq. (7) and Eq. (8).

3.1.2. Existence and uniqueness

The main goal of this section is to find conditions on the coefficients f and g which guarantee the existence and uniqueness of solutions.

To build a notion of time into our probability space, we specify which sub-σ-algebra of information in $\mathcal{F}$ can be gathered by time t. We consider the increasing sequence $\{\mathcal{F}_t\}$ of σ-algebras of $\mathcal{F}$ for which for $t, s \in \mathbb{R}_+$ with $t \leq s$, we have $\mathcal{F}_t \subseteq \mathcal{F}_s \subseteq \mathcal{F}$. The quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$ forms our filtered probability space, on which we can now define an M-dimensional $\mathcal{F}_t$-Brownian motion B_t^x.

The solution of an SDE is usually described in the sense of strong or weak solutions. Firstly, we proceed to detail what we mean by *strong solutions*.

Definition 3.2 (Strong solutions (Chung and Williams, 2014, Section 10.4)). *Let $B^x = (B_t^x, t \geq 0)$ be a Brownian motion and ξ be a random variable independent of B^x on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$. A stochastic process X is a strong solution of a stochastic differential equation if*

- *X is adapted to the filtration $\{\mathcal{F}_t\}$;*
- *$\mathbb{P}$-almost surely $X_0 = \xi$;*
- *$\mathbb{P}$-almost surely $\int_0^t |f_{d_x}(X_s)| + (g_{dx,m}(X_s))^2\, ds < \infty$ holds for all $d_x = 1, 2, \dots, D_x$, $m = 1, 2, \dots, M$ and $t \in [0, \infty)$;*
- *$\mathbb{P}$-almost surely X solves Eq. (7).*

As usual, we assume that the sigma-algebra $\mathcal{F}_t$ is completed with all the $\mathbb{P}$-null sets in Ω (Chung and Williams, 2014, Section 10.4). This definition means that if we are given a probability space which carries B^x and ξ, the solution X must be adapted to the generated filtration $\{\mathcal{F}_t\}$; that is, the path $\{X_t\}_{t\geq 0}$ is completely characterised by the path of the underlying driving process (Brownian motion) and the initial condition.

The condition of being adapted to the generated filtration is sometimes too strict (the famous example of Tanaka's equation, found in, for example, Zvonkin (1974)). One can find pairs (X, B^x) which are solution to the SDE in Eq. (7) instead of simply seeking X. Such pairs are called *weak solutions*. In what follows, we consider mostly the case of strong solutions of Itô stochastic differential equations.

Theorem 3.1 (Existence and uniqueness (Mao, 2007, Theorem 2.2)). *Suppose that ξ is a random variable on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$, for which ξ is independent of the Brownian motion B^x also defined on this space. Suppose that*

- *$\xi \in \mathcal{L}^2(\mathbb{P})$, i.e. ξ is square-integrable under the probability measure $\mathbb{P}$;*

- *f, g satisfy the Lipschitz continuous and linear growth conditions in x.*

Then there exists a solution $X = (X_t,\, t \geq 0)$ to the associated stochastic differential equation (7) satisfying the initial condition ξ. Moreover, the solution is unique and satisfies $X \in \mathcal{L}^2(\mathbb{T} \times \mathbb{P})$, where $\mathbb{T}$ is the Lebesgue measure on $[0, \infty)$.

Example 3.1 (Geometric Brownian motion). *Consider the unidimensional case, for which $X_t : [0, \infty) \times \Omega \mapsto \mathbb{R}$. Let $f(X_t) = 0$ and $g(X_t) = \omega_x X_t$, where ω_x is a positive constant:*

$$dX_t \quad = \quad \omega_x X_t d B_t^x, \tag{9}$$

with initial condition $X_0 = \xi \equiv 1$. Now define a function $V(x) \stackrel{\Delta}{=} \ln(x)$. We apply Itô's formula (Ito, 1944) on the process $V(X)$, resulting in

$$d(\ln(X_t)) \quad = \quad \omega_x d B_t^x - \tfrac{1}{2}\omega_x^2 t, \qquad t \geq 0.$$

Since $B_0 = 0$ by definition, we further get

$$X_t \quad = \quad X_0 \exp\left(\omega_x B_t^x - \tfrac{1}{2}\omega_x^2 t\right).$$

For this SDE, both existence and uniqueness properties hold. Moreover, if the initial condition is constrained to positive real numbers, all the values for paths $\{X_t\}_{t \geq 0}$ will follow the same property.

Now assume that the concentration of an element (e.g. biomass) changes within a time period by a random amount. Similarly to what happens in Example 3.1, assume that this random amount is a function of the current concentration itself. To give an idea of the driving process added to Eq. (3), we consider a SDE with a driving process similar to the geometric Brownian motion. We will see that the conditions for both existence and uniqueness properties have to be slightly modified to prove existence and uniqueness.

The chemostat model considered in Imhof and Walcher (2005) has the following dynamic behaviour:

$$d \underbrace{\begin{bmatrix} b_t \\ s_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} (\mu(s_t) - D) b_t \\ D(s_{in} - s_t) - \kappa \mu(s_t) b_t \end{bmatrix}}_{f(X_t)} dt + \underbrace{\begin{bmatrix} \omega_b b_t & 0 \\ 0 & \omega_s s_t \end{bmatrix}}_{g(X_t)} d \underbrace{\begin{bmatrix} B_t^b \\ B_t^s \end{bmatrix}}_{B_t^x}, \tag{10}$$

with $X_0 \equiv (b_0,\, s_0)$. Now, we see how the solution to the model in Eq. (10) looks like for Case 1 and Case 2. The noise intensities are $\omega_x = (\omega_b = 0.05, \omega_s = 0.05)$ for Case 1 and $\omega_x = (\omega_b = 0.075, \omega_s = 0.075)$ for Case 2.

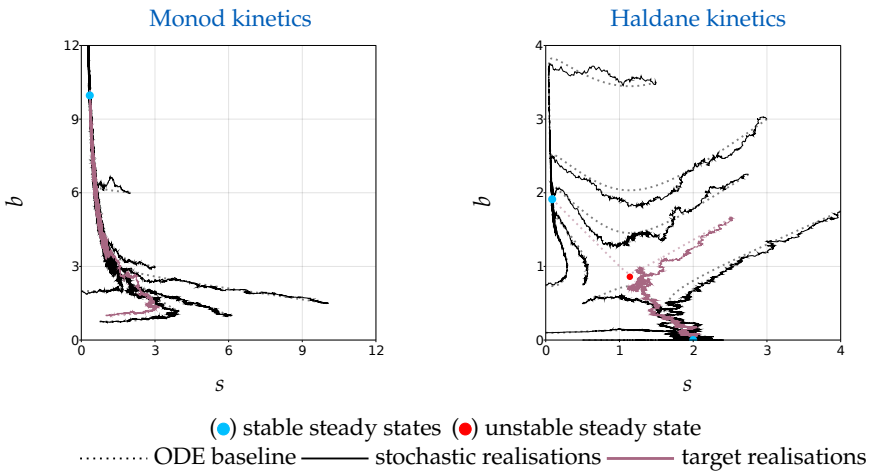

Figure 2: Trajectories from the stochastic model with growth function of the Monod type (left) and of the Haldane type (right). The dashed lines in the background show the trajectories from the deterministic model, as in Fig. 1. Here, a target realisation denotes a simulated trajectory which will be estimated from noisy measurements in Section 5.

Fig. 2 (left) suggests that processes X move around the trajectories of the deterministic chemostat model, with small fluctuations. Here, $D < \mu(s_{in})$. Under the deterministic model, this condition implies that the microbial population forming the biomass survives regardless of the initial concentrations of substrate and biomass (Harmand et al., 2017). In Fig. 2 (right), processes X also move around the trajectories of the deterministic chemostat model, but in order to avoid "near" extinction of the population around $E_2^{(1)} = (0, s_{in})$, we would have to choose appropriate initial concentrations.

Theorem 3.2. *For any initial condition $X_0 \in (\mathbb{R}^2)^*_+$, i.e. any $\{(b_0, s_0) \in \mathbb{R}^2 | b_0 > 0, s_0 > 0\}$, there exists a unique strong solution $X = (X_t, t \geq 0)$ of the system in Eq. (10) with growth functions from Eq. (4) or Eq. (5). Moreover, the solution remains in $(\mathbb{R}^2)^*_+$ with probability one, i.e. $X_t \in (\mathbb{R}^2)^*_+$ for $t \geq 0$ almost surely.*

Proof 3.1. *See Section A(c) in the supplementary material.*

Thm. 3.2 assures the positivity of the solution for all $t \geq 0$. However, we have not yet ensured persistence of the microbial population, i.e. that the system admits a unique stationary distribution which is also ergodic. In the context of the chemostat, persistence has been discussed in Wang and Jiang (2017), where sufficient conditions for ergodicity were imposed on s_{in}, D, ω_b and ω_s.

Let $\zeta(x)$ be an integrable function with respect to the measure μ, an invariant measure of the process X. If the system is ergodic, then $\mathbb{P}\left(\frac{1}{t}\int_0^t \zeta(X_s)ds \to \int \zeta(y)\mu(dy)\right) = 1$ (Birkhoff, 1931). When we let time remove the stochasticity, the time-average of ζ converges almost surely (if the limit exists) to the expectation of ζ at an arbitrary time t. Define the time average limit and the Monte Carlo approximation to the expectation limit by $\langle\zeta\rangle^{tim}_{\infty,1}$ and $\langle\zeta\rangle^{ens}_{t,\infty}$, respectively. Let $X_t^{(n)}$ denote the n-th realisation of X_t, then define

$$\langle\zeta\rangle^{tim}_{\infty,1} \triangleq \lim_{t\to\infty} \frac{1}{t}\int_0^t \zeta(X_s)\,ds, \quad \langle\zeta\rangle^{ens}_{t,\infty} \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \zeta(X_t^n), X_t^{(n)} \sim \mu. \tag{11}$$

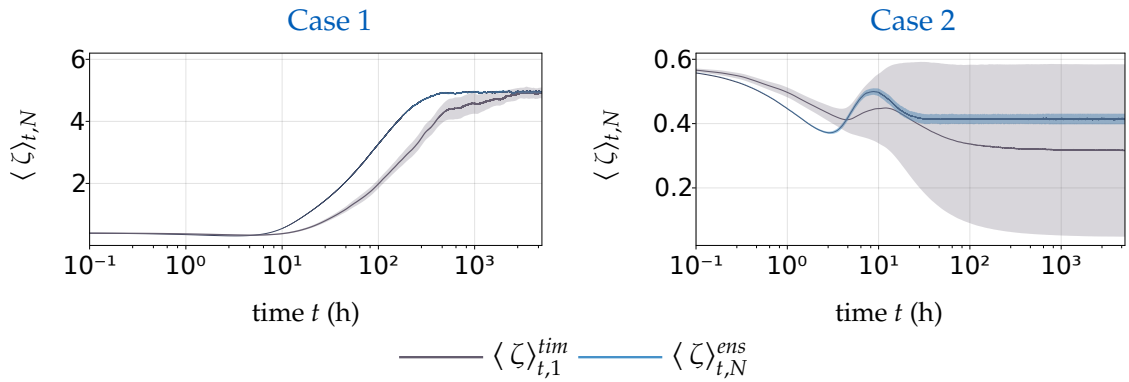

Figure 3: Estimates of the expectation and the time-average of ζ, where $\zeta(x) = 0.5b - 0.1s$. The estimates are based on 30 experiments. Shaded regions are delimited by two standard deviations above and below the mean. For the Monte Carlo approximation, we used an ensemble of $N = 512$ samples of X_t when computing the sample averages (note that $N = 1$ when computing the time averages). The initial concentrations are $(b_0, s_0) = (1.0\,\text{mgL}^{-1}, 1.0\,\text{mgL}^{-1})$ for Case 1 (Monod kinetics) and $(b_0, s_0) = (1.65\,\text{mgL}^{-1}, 2.5\,\text{mgL}^{-1})$ for Case 2 (Haldane kinetics).

Fig. 3 suggests that ergodicity holds for the system in Case 1, as shown by the commutativity in the two limits $\langle\zeta\rangle^{tim}_{\infty,1}$ and $\langle\zeta\rangle^{ens}_{t,\infty}$. In fact, this is confirmed by the fact that the conditions for ergodicity outlined in Wang and Jiang (2017) hold. For Case 2, however, some trajectories of the microbial population go towards the washout node $E_2^{(3)}$, instead of the interior node $E_2^{(1)}$ (as in Fig. 2). Moreover, they do not escape this attractor in finite time. This can be seen by the non-commutativity in the limits.

3.2. The Markov property and the Kolmogorov forward equation

In this section, we use the theory of Markov processes to study the properties of the stochastic differential equation (10), given that its solution, if it exists, satisfies the Markov property. We will see that this allows us to apply the Kolmogorov equations (in particular the forward, or Fokker-Planck, equation) from the theory of Markov process into the SDE setting. This can then be used to express expectations of functions of the solution to Eq. (10) in terms of certain non-random partial differential equations (PDEs).

Definition 3.3 (Markov process). *Let the process $X = (X_t,\, t \geq 0)$ be measurable under the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$. X is called an $\mathcal{F}_t$-Markov process if we have*

$$\mathbb{E}[\phi(X_t)|\mathcal{F}_\tau] \;=\; \mathbb{E}[\phi(X_t)|X_\tau] \;=:\; \phi_{t-\tau}(X_\tau), \tag{12}$$

for all $t \geq \tau$ and all bounded measurable functions ϕ. Here, $\mathbb{E}[\cdot]$ denotes expectation.

Theorem 3.3 (Markov property and SDEs (Cohen and Elliott, 2015, Theorem 17.2.3)). *Assume existence and uniqueness of the solution to the SDE in Eq. (7). The solution X that satisfies such SDE is an $\mathcal{F}_t$-Markov process.*

The intuitive content behind Thm. 3.3 is clear. Firstly, suppose that the process X generated by the SDE in Eq. (7) begins at time zero, and that we watch it up to time τ. Based on this information, we want to compute the conditional expectation $\phi_{t-\tau}(X_\tau)$, where $t \geq \tau$. Then we can pretend that the process is starting at time τ at its current position, generate the solution to the stochastic differential equation corresponding to this initial condition, and compute the expected value of $\phi(X_t)$ generated in this way.

For simplicity, we start by studying the random process $X = (X_t,\, t \geq 0)$ for $\tau = 0$, i.e. given only the $\mathcal{F}_0$-measurable D_x-dimensional random variable ξ. Moreover, rather than studying the path of this random process itself, we can now look at how the non-random function ϕ_t varies with t.

Let $p(t,\cdot)$ denote the density with respect to Lebesgue measure for the probability measure $\mathbb{P}$ at time t. We describe the process in terms of a transition kernel $p(t,x|0,x_0) \overset{\Delta}{=} \mathbb{P}(X_t = x|X_0 = x_0), t \geq 0$. This allows us to compute the probability of having the process X at a point $x \in \mathbb{R}^{D_x}$ at instant t conditioned on starting at a point $x_0 \in \mathbb{R}^{Dx}$ at time $t_0 = 0$. We then proceed to use the property shown in Def. 3.3.

$$\begin{aligned}\mathbb{E}[\phi(X_t)] = \mathbb{E}[\mathbb{E}[\phi(X_t)|\mathcal{F}_0]] = \mathbb{E}[\mathbb{E}[\phi(X_t)|X_0]] &= \int_{\mathbb{R}^{D_x}}\int_{\mathbb{R}^{D_x}} \phi(x)p(t,x|0,x_0)p(0,x_0)dx_0dx \\ &= \int_{\mathbb{R}^{D_x}} \phi(x)p(t,x)dx.\end{aligned} \tag{13}$$

Now, we find the time evolution of the probability density $p(t,x)$ by differentiating both sides of equation Eq. (13) with respect to time,

$$\frac{d}{dt}\mathbb{E}[\phi(X_t)] \;=\; \int_{\mathbb{R}^{D_x}} \phi(x)\frac{\partial}{\partial t}p(t,x)dx \;=\; \int_{\mathbb{R}^{D_x}} \phi(x)\mathcal{A}^*p(t,x)dx, \tag{14}$$

where we introduced the adjoint operator $\mathcal{A}^*$ to describe the time evolution of the probability density.

The proposition that follows is adapted from Nobuyuki and Watanabe (1981, Chapter 4).

Proposition 3.1 (The Kolmogorov forward equation (KFE)). *Assume that the $\mathcal{F}_t$-Markov process $X = (X_t,\, t \geq 0)$ is solution to the stochastic differential equation (7). Suppose that the density $p(t,x)$ exists and is C^1 in t and C^2 in x, then $p(t,x)$ satisfies the Kolmogorov forward equation, also called the Fokker-Planck equation:*

$$\frac{\partial}{\partial t}p(t,x) = \mathcal{A}^*p(t,x), \tag{15}$$

where the form of the adjoint operator is

$$\mathcal{A}^*p(t,x) = -\sum_{d_x=1}^{D_x}\frac{\partial}{\partial x_{d_x}}[f_{d_x}(x)p(t,x)] \;+\; \sum_{d_x=1}^{D_x}\sum_{d'_x=1}^{D_x}\frac{\partial^2}{\partial x_{d_x}\partial x_{d'_x}}[G_{d_xd'_x}(x)p(t,x)]. \tag{16}$$

Here, $G_{d_xd'_x}(x) = \frac{1}{2}\sum_{m=1}^{M} g_{d_xm}(x)g_{d'_xm}(x)$ and the initial condition $p(0,x)$ is given.

At this point, our best knowledge about the state of the system is given by the distribution whose probability density solves the KFE, given the distribution of its initial state. This motivates us to study the model in Eq. (10) through its associated KFE, as in Campillo et al. (2014). For instance, one can find the solution given the distribution of initial concentrations, and then verify how likely the state of a system is to be within regions around $E_2^{(1)}$ or $E_2^{(3)}$. We present the solution to the KFE for Cases 1 and 2 in Figs. 4 and 5, respectively. We solve the PDE by following similar steps as in Campillo et al. (2014): we make use of an upwind differencing scheme for the space discretisation and of an implicit scheme for the time discretisation.

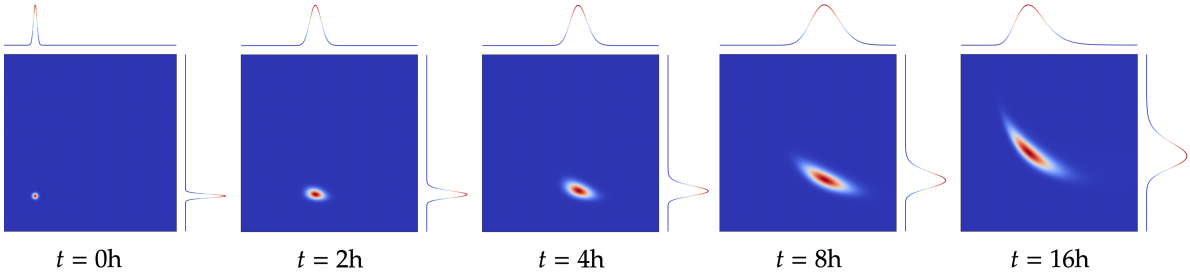

Figure 4: Case 1 - solution to the KFE for the growth function of Monod type. Initial condition is $X_0 \sim \mathcal{N}([1,\ 1]^\top, 0.05^2 I_{2\times 2})$. The discretisation in space is in the domain $(0, 5] \times (0, 5]$, with a count of 256^2 finite volumes. Marginal density functions are represented by marginal curves. Each marginal curve has been normalised by its maximum value for the sake of illustration.

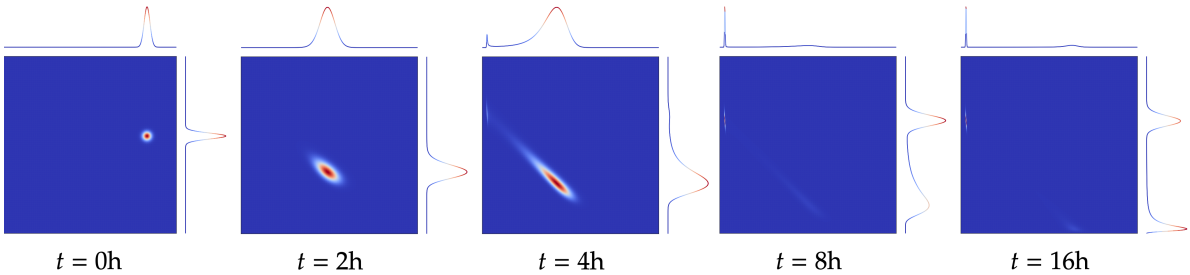

Figure 5: Case 2 - solution to the KFE for the growth function of Haldane type. Initial condition is $X_0 \sim \mathcal{N}([1.65,\ 2.5]^\top, 0.05^2 I_{2\times 2})$. The discretisation in space is in the domain $(0, 3] \times (0, 3]$, with a count of 512^2 finite volumes.

So far, we have worked with a model describing the dynamic behaviour of the system. The prior probabilistic knowledge about how the state of the system evolves was formally embedded by the adjoint operator $\mathcal{A}^*$ in the Kolmogorov forward equation. Our focus, however, is to refine our knowledge about the state when information about the system is available in the form of measurements. In the following section, we resort to an observation model to describe how the signal gets corrupted. Generally, the state signal cannot be measured directly (e.g. only certain components of a vector-valued process are observed), and the measurements are also subjected to error. To describe how measurements and states are related, we combine the dynamics with an observation model in which the error is also stochastic.

4. Observation models

Let $0 = t_0 < t_1 < \cdots < t_n \ldots$ be a uniform partition of the interval $[t_0, \infty)$. We let discrete-time measurements z be available from the system, in the form of

$$z_t = h(X_t) + k(t)v_t, \quad t \in \{t_0, t_1 \ldots\}. \tag{17}$$

Here, $k : [0, \infty) \to \mathbb{R}^{D_y} \times \mathbb{R}^{D_y}$ denotes noise intensity, and $h : \mathbb{R}^{D_x} \to \mathbb{R}^{D_y}$ is the observation function. We simplify again our notation by omitting the dependency on the vector θ of parameters and on the input function $u(t)$.

Instead of proceeding with the model in Eq. (17), we will work with an integrated form of the observations; we use an observation model for a process Y, where this process represents the integral of the actually observed process. To

explain this, we begin by defining a D_y-dimensional $\mathcal{F}_t$-Brownian motion B^y independent of B^x. We then introduce a process $Y = (Y_t, t \geq 0)$ which obeys

$$dY_t = h(X_t)dt + k(t)dB_t^y, \quad Y_0 = 0. \tag{18}$$

Here, Y_0 is considered to be identically zero (this does not mean that measurements at time 0 are zero-valued, but rather that there is no information available initially).

Notice that Y as modelled by Eq. (18) is the integral of the actual observed process, so that in discrete-time one really observes z rather than Y. We can think of the white Gaussian noise v_t in Eq. (17) as the hypothetical time derivative of the Brownian motion in Eq. (18) at time t, and that v has zero mean and covariance $\mathbb{E}(v_s v_t) = \delta(t-s)$, where $\delta(\cdot)$ is the Dirac's delta generalized function. By working with the integrated form, we obtain a mathematical model in which we can disregard the need to work with $\delta(\cdot)$ and instead use the same machinery of stochastic calculus used in the description of the signal model.

In Section 2, we introduced two different settings for the dynamics governing the state of the chemostat: Case 1 with Monod kinetics and Case 2 with Haldane kinetics.

We now additionally explore an observation model in which the sensor transforms the state variables in one-dimensional measurements. We assume that we are given the flow rate of biogas produced by the biomass after substrate consumption. This setting is denoted by scenario A.

Let Q denote the flow of biogas production. The flow Q, being proportional to the microbial activity (Renard et al., 1988; Bernard et al., 2001), can be written as $Q(s,b) = \gamma\mu(s)b$, where γ is the yield coefficient for CH_4 in Eq. (1). The integrated observation process considered is

$$d\left[y_t^A\right] = \underbrace{\left[Q(s,b)\right]}_{h_A(x_t)} dt + \underbrace{\left[\omega_A\right]}_{k_A(t)} d\left[B_t^{y^A}\right]. \tag{19}$$

We combine Cases 1 and 2 with the observation model in scenario A to form the experiments described in Table 2. During a sub-interval of interest $t \in [0, \delta_u]$ for which the input $u(t)$ is held constant, we simulate trajectories from processes X and Y with time discretisation δ_x and δ_y, respectively. We simulate the trajectories for $\delta_u = 16$ hours. We assume that we are able to collect the measurements $\{y_t^A\}$ at high-frequency ($\delta_y = \delta_x$) or at low-frequency ($\delta_y \gg \delta_x$). Again, the initial concentrations are $(b_0, s_0) = (1.0\,\text{mgL}^{-1},\ 1.0\,\text{mgL}^{-1})$ for Case 1 and $(b_0, s_0) = (1.65\,\text{mgL}^{-1},\ 2.5\,\text{mgL}^{-1})$ for Case 2.

Table 2: Additional model parameters and inputs for differential equations of X and Y.

	Model parameters		Simulation parameters	
	Observation function	ω_A	δ_x (h)	δ_y (h)
1A - high-frequency	$27.5\mu(s)b$	0.01	0.01	0.01
1A - low-frequency	$27.5\mu(s)b$	$0.01(\delta_y/\delta_x)^{0.5}$	0.01	2.0
2A - high-frequency	$2.75\mu(s)b$	0.01	0.005	0.005
2A - low-frequency	$2.75\mu(s)b$	$0.01(\delta_y/\delta_x)^{0.5}$	0.005	2.0

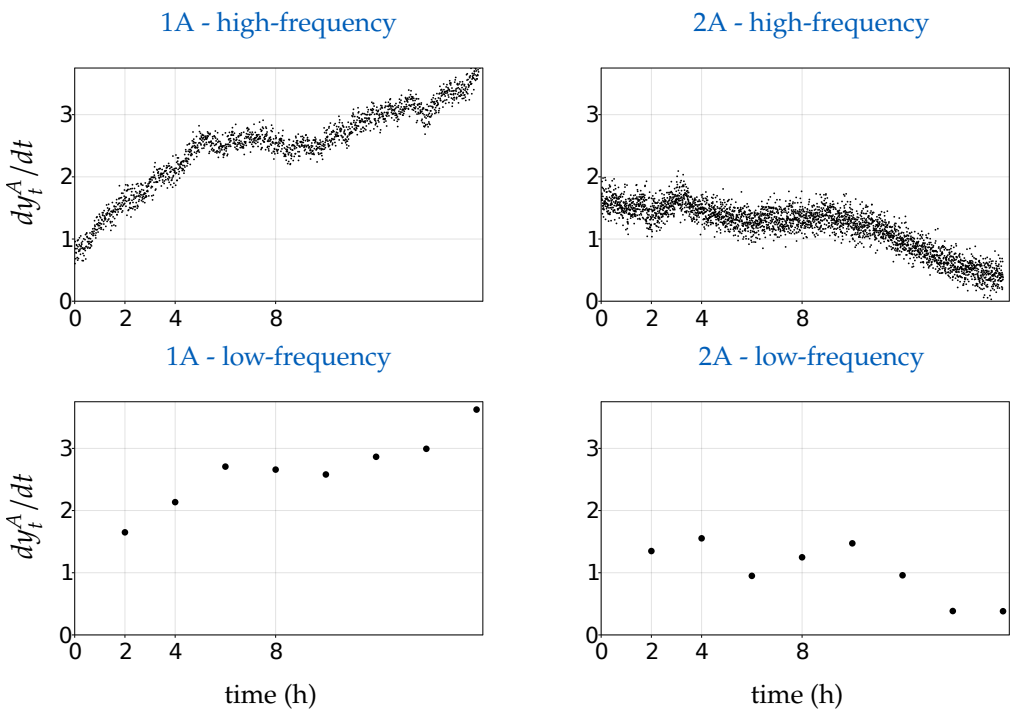

Figure 6: Measurements simulated from the trajectories highlighted in Fig. 2.

5. The filtering problem

To ensure theoretical developments, we assume that $X_0 \in \mathcal{L}^2(\mathbb{P})$ (remember that we need this assumption to determine the random process $X = (X_t,\, t \geq 0)$ uniquely, according to Thm. 3.1).

In state-space models, the Bayesian methodology of computing posterior distributions of a latent state based on a history of noisy measurements is known as filtering and smoothing. The aim of solving a filtering problem is to determine the conditional distribution of the signal X_t given the observation σ-algebra $\mathcal{F}_t^Y = \sigma\{Y_s : s \leq t\}$. We formalise this by defining a stochastic process $\pi = (\pi_t, t \geq 0)$ describing this conditional distribution. If we assume that such distribution has a density p_t with respect to Lebesgue measure, then

$$\pi_t(A) \stackrel{\Delta}{=} \mathbb{P}\left(X_t \in A | \mathcal{F}_t^Y\right), \quad \pi_t(\phi) = \mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right] = \int_{\mathbb{R}^{D_x}} \phi(x) p_t(x) dx.$$

Here, A is an arbitrary set in the σ-algebra $\mathcal{B}([0,\infty) \otimes \mathbb{R}^{D_x})$, ϕ is an arbitrary function of the state and $\mathbb{E}_{\mathbb{P}}$ denotes expectation under the density p_t. The first equation takes values in the space of functions from $[0,\infty) \times \mathbb{R}^{D_x}$ into $[0,1]$ (the probability of an event is always between 0 and 1). The second equation allows us to get desirable information about the state through a transformation ϕ, assuming that the expectation exists (that is, $\int |\phi(x)| p_t(x)\, dx < \infty$). Whether π_t has a density with respect to a reference measure (e.g. Lebesgue measure) is rigorously discussed in Bain and Crisan (2008). We proceed here by assuming that we can prove the existence of p_t.

5.1. Filtering equations

We continue working on the quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$. Our goal is to compute π_t on the basis of the observations $\{Y_s : s \leq t\}$. We begin by considering the signal process with evolution in time described by Eq. (7), and also the observation model given by Eq. (18). We rewrite here both equations:

$$\begin{aligned} dX_t &= f(X_t)dt + g(X_t)dB_t^x, \quad X_0 = \xi, \\ dY_t &= h(X_t)dt + k(t)dB_t^y, \quad Y_0 = 0. \end{aligned}$$

Here, the derivation is briefly outlined. We follow similar steps as in Kutschireiter et al. (2020). For a more detailed and formal derivation, see van Handel (2007); Bain and Crisan (2008). Among the possible ways of deducing the evolution equation for π_t, there are two commonly used approaches, namely the change of measure method and the innovation process method. We present the change of probability measure method. For the second approach, see Bain and Crisan (2008, Section 3.7).

5.1.1. Changes of measure

A probability measure is a function that assigns numbers ("probabilities") to events. If we have two such measures $\mathbb{P}$ and $\mathbb{Q}$, then $\mathbb{P}$ is called *absolutely continuous* with respect to $\mathbb{Q}$ if every nullset of $\mathbb{Q}$ is a null set of $\mathbb{P}$. Moreover, $\mathbb{P}$ and $\mathbb{Q}$ are called *equivalent* if they have the same nullsets. In other words, if A denotes an event, and $\mathbb{P}(A)$ denotes its probability, then equivalence means that $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$.

Changing the measure allows us to compute expectations of a measurable function $\phi(x)$ with respect to a measure $\mathbb{Q}$, which were originally expressed with respect to measure $\mathbb{P}$. To see this, consider the two measures $\mathbb{P}$ and $\mathbb{Q}$ for some real-valued random variable X, and write them in terms of their densities p, q (again, with respect to Lebesgue measure). We then have

$$\mathbb{E}_{\mathbb{P}}\left[\phi(X)\right] = \int \phi(x)p(x)\,dx = \int \frac{p(x)}{q(x)}\phi(x)q(x)\,dx = \mathbb{E}_{\mathbb{Q}}\left[L(X)\phi(X)\right], \tag{20}$$

where we introduced the likelihood ratio $L(x) \stackrel{\Delta}{=} \frac{p(x)}{q(x)}$. Thus, changing the measure proves to be useful whenever expectations under $\mathbb{Q}$ are easier to compute than under $\mathbb{P}$.

If $\mathbb{P}$ is a probability measure and we have a collection of processes (X and Y), the measure $\mathbb{P}_t$ is the restriction of $\mathbb{P}$ to all events that can be described in terms of the behaviour of $\{X_s\}$ and $\{Y_s\}$ for $0 \leq s \leq t$. In other words, $\mathbb{P}_t$ denotes the restriction of the measure $\mathbb{P}$ to $[0, t] \times \mathbb{P}$. If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, also their restrictions $\mathbb{P}_t$ and $\mathbb{Q}_t$ are equivalent. By using the Radon–Nikodym theorem (Klebaner, 2005, Theorem 10.6), we know that a random variable L_t exists, for which

$$\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)\right] = \mathbb{E}_{\mathbb{Q}}\left[L_t\phi(X_t)\right], \tag{21}$$

where $L_t \stackrel{\Delta}{=} \frac{d\mathbb{P}_t}{d\mathbb{Q}_t}$ is called the Radon–Nikodym derivative or density of $\mathbb{P}_t$ with respect to $\mathbb{Q}_t$.

As we previously mentioned, our problem is to compute the conditional expectation (i.e. an expected value under the posterior distribution π_t) for some function ϕ. The conditional expectation can then be rewritten in terms of a reference probability measure $\mathbb{Q}$:

$$\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right] = \frac{\mathbb{E}_{\mathbb{Q}}\left[\phi(X_t)\,L_t|\mathcal{F}_t^Y\right]}{\mathbb{E}_{\mathbb{Q}}\left[L_t|\mathcal{F}_t^Y\right]} =: \frac{\rho_t(\phi)}{\rho_t(1)}. \tag{22}$$

Eq. (22) is known as a Bayes' formula for stochastic processes or *Kallianpur-Striebel formula*. Here, we have defined the unnormalised estimate $\rho_t(\phi) = \mathbb{E}_{\mathbb{Q}}\left[\phi(X_t)\,L_t|\mathcal{F}_t^Y\right]$. The hope is that we can pick a reference measure $\mathbb{Q}$ such that both L_t and q are simple enough to make $\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right]$ more tractable to compute via Eq. (22) than via an integration over the measure $\mathbb{P}$. For instance, some simplification might be achieved by switching from a model with measure $\mathbb{P}$ in which X and Y are coupled, i.e. statistically dependent, to a model with measure $\mathbb{Q}$ where they are independent (while preserving the distribution of X_t). For our filtering problem, we will choose a reference measure $\mathbb{Q}$ such that the path of the observations $\{Y_{0:t}\}$ (or equivalently the set of the increments $\{dY_{0:t}\}$) becomes independent of the path of the state process $\{X_{0:t}\}$, i.e. $q(t, X_{0:t}, dY_{0:t}) = p(t, X_{0:t})q(t, dY_{0:t})$.

For the time being, we will concentrate not on $\pi_t(\phi)$, but on its unnormalised counter-part $\rho_t(\phi)$. Our next step is to find an explicit expression for $\rho_t(\phi)$. We start by finding the expression for L_t, and proceed by applying Itô's rule to the process $\phi(X_t)L_t$, then we compute the conditional expectation of this expression (making use of the independence properties under $\mathbb{Q}_t$). Once we have accomplished this, the remainder is another application of Itô's rule, which gives us an expression for the normalised quantity $\pi_t(\phi)$.

5.1.2. Filtering equations for observations corrupted by Gaussian noise

For the derivations in this section, we operate under some technical conditions: (i) $k(t)$ is invertible for all $t \in [0, \infty)$, and $k(t), k^{-1}(t)$ are locally bounded; (ii) the solution to the SDE in Eq. (7) exists and is unique; (iii) h has linear growth. See Appendix A for detailed calculation steps.

We chose a reference measure $\mathbb{Q}$ such that the path of the observations $\{dY_{0:t}\}$ becomes independent of the path of the state process $\{X_{0:t}\}$, and yet the law of X is the same under $\mathbb{P}$ as it is under $\mathbb{Q}$. By definition, let us see what the likelihood ratio L_t has to look like for this measure change:

$$L_t = \frac{d\mathbb{P}_t}{d\mathbb{Q}_t} = \frac{p(t, X_{0:t}, dY_{0:t})}{q(t, X_{0:t}, dY_{0:t})} = \frac{p(t, dY_{0:t}|X_{0:t})p(t, X_{0:t})}{p(t, X_{0:t})q(t, dY_{0:t})} = \frac{p(t, dY_{0:t}|X_{0:t})}{q(t, dY_{0:t})}. \tag{23}$$

Our observation model in Eq. (18) assumes that increments dY_t are independently distributed given the state X_t, and we know that $dY_t|X_t \sim \mathcal{N}(dY_t;\, h(X_t)dt, kk^\top(t)dt)$. Furthermore, we choose $q(t, dY_t)$ as the density of $\mathcal{N}(dY_t; 0, kk^\top(t)dt)$. Thus, the Radon-Nikodym derivative L_t can be written as

$$L_t \;=\; \exp\left[\int_0^t (k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2}\int_0^t \|k^{-1}(s)h(X_s)\|^2 ds\right]. \tag{24}$$

Remark 5.1. *An even more general model arises when we consider that the observation equation is driven by a signal dependent noise; that is, $k(t, X_t)dB_t^y$ instead of $k(t)dB_t^y$. The filtering problem with this type of observation model can be then converted into a classical one (signal independent observation noise) via a suitably chosen stochastic flow mapping. We refer the interested reader to Crisan et al. (2009).*

With Eq. (24) at hand, an application of Itô's formula leads to the following proposition:

Proposition 5.1 (The Zakai equation). *Let ϕ be C^2 (twice continuously differentiable) and suppose that ϕ and all its derivatives are bounded. Then we can write*

$$\rho_t(\phi) \;=\; \rho_0(\phi) + \int_0^t \rho_s(\mathcal{A}^*\phi)ds + \int_0^t \rho_s(k^{-1}(s)h\phi)^\top d\tilde{Y}_s, \tag{25}$$

where $\rho_0(\phi) = \mathbb{E}_{\mathbb{Q}}\left[\phi(X_0)\, L_0|\mathcal{F}_0^Y\right]$, $\mathcal{A}^*$ is the adjoint operator introduced in Eq. (14), $h\phi = h(x)\phi(x)$ and the process $\tilde{Y} = (\tilde{Y}_t,\, t \geq 0)$ is the rescaled observation process ($\tilde{Y}_t = k^{-1}(t)Y_t,\, t \geq 0$).

If we apply integration by parts to the Zakai equation, we obtain

$$dq_t(x) \;=\; \mathcal{A}^* q_t(x)dt + q_t(x)(k^{-1}(t)h(x))^\top d\tilde{Y}_t. \tag{26}$$

Now that we have an equation for $\rho_t(\phi)$, the equation for $\pi_t(\phi)$ results from another application of Itô's formula, this time on $\rho_t(\phi)/\rho_t(1)$, and of the Kallianpur-Striebel formula introduced previously.

Proposition 5.2 (The Kushner-Stratonovich equation). *Let ϕ be C^2 and suppose that ϕ and all its derivatives are bounded. Then we can write*

$$\begin{aligned}\pi_t(\phi) = \pi_0(\phi) + \int_0^t (\pi_s(\mathcal{A}^*\phi)\, ds + \\ \int_0^t \pi_s(k^{-1}(s)h\phi) - \pi_s(\phi)\pi_s(k^{-1}(s)h))^\top (d\tilde{Y}_s - \pi_s(k^{-1}(s)h)\, ds),\end{aligned} \tag{27}$$

where $\pi_0(\phi) = \mathbb{E}_{\mathbb{Q}}\left[\phi(X_0)\, L_0|\mathcal{F}_0^Y\right] = \mathbb{E}_{\mathbb{P}}\left[\phi(X_0)\right]$ and $h\phi = h(x)\phi(x)$.

If we apply integration by parts to the Kushner-Stratonovich equation, we obtain

$$dp_t(x) \;= \mathcal{A}^* p_t(x)\, dt \;+\; p_t(x)\left\{k^{-1}(t)(h\phi - \pi_t(h))\right\}^\top (d\tilde{Y}_t - \pi_t(k^{-1}(t)h)\, dt). \tag{28}$$

We have now obtained both a stochastic integral expression - the Kushner-Stratonovich equation - for $\pi_t(\phi)$ and the evolution of the density p_t of the state conditioned on measurements.

5.2. Approximations to the solution of the Kushner-Stratonovich equation

This section contains an overview of two classes of methods used to approximate the solution to the filtering problem. For each class, we give a brief description of the ideas behind the methods and state some related results. Note that so far we have only presented the filter $\pi_t(\phi)$ on an unspecified set in the Borel σ-algebra of $C(\mathbb{R}_+, \mathbb{R}^{D_y})$ of measure one, as well as its assumed density p_t with respect to the Lebesgue measure. We have not yet shown what the filtered estimated $\pi_t(\phi)$ would be for a given observation sample path $z \in C([0,T], \mathbb{R}^{D_y})$. Typically, the observation data is recorded at discrete times, and only these data are made available and used. Considering the practical implementation of a non-linear filter, we present the notion of pathwise filtering as introduced in Clark (1978), and studied by many authors, as in Davis (1982) and Crisan et al. (2021).

L_t in Eq. (24) is defined in terms of a stochastic integral with the observations as the integrator. However, since the integrand is independent of the observations, we can reverse the role of processes X and Y. Instead of using the observations as the integrator, we compute L_t with an integral for every process sample path $\{X_{0:T}\}$, provided that $h(X_t)$ is a semimartingale. Through an integration by parts:

$$\int_0^T \tilde{h}^\top(X_s)d\tilde{Y}_s \quad = \quad \tilde{h}^\top(X_T)\tilde{Y}_T - \int_0^T \tilde{Y}_s^\top d\tilde{h}(X_s) \quad \text{almost surely.} \tag{29}$$

Here, the right-hand side can be interpreted pathwise with respect to the observations. The pathwise Kallianpur-Striebel formula reads as follows:

$$\hat{L}_T \quad = \quad \exp\left[\tilde{h}^\top(X_T)\tilde{Y}_T - \int_0^T \tilde{Y}_s^\top d\tilde{h}(X_s) - \frac{1}{2}\int_0^T \|\tilde{h}(X_s)\|^2 ds\right], \tag{30}$$

where $\hat{L}_T$ stands for L_T computed with respect to a fixed observation path. Consequently, the stochastic PDEs in Eqs. (26) and (28) reduce to ordinary PDEs with random coefficients.

5.2.1. Finite-differences approach + Splitting-up approximation

The following exposition follows closely the description of the method given in Cai et al. (1995). Let $0 = t_0 < t_1 < \cdots < t_n \ldots$ be a uniform partition of the interval $[0, \infty)$ with time step $\Delta = t_n - t_{n-1}$. If the set of observations $\{z_i, 1 \leq i \leq n\}$ is available, the density p_{t_n} will be approximated by p_n^Δ, where the transition from p_{n-1}^Δ to p_n^Δ follows a splitting-up approximation (Le Gland, 1989; LeGland, 1992).

The first step, called the *prediction* step, consists in solving the Fokker-Planck equation for the time interval $[t_{n-1}, t_n]$. We rewrite the partial differential equation here,

$$\frac{\partial p_t^n(x)}{\partial t} \quad = \quad \mathcal{A}^* p_t(x), \qquad p_{t_{n-1}}^n \; = \; p_{n-1}^\Delta, \tag{31}$$

and we denote the prior estimate by $\bar{p}_n^\Delta \stackrel{\Delta}{=} p_{t_n}$.

The second step, called the *correction* step, uses the new observation z_n to update $\bar{p}_n^\Delta$. We use the approximation from pathwise filtering to the Radon-Nikodym derivative (Eq. (30)). We compute $p_n^\Delta(x)$ via $p_n^\Delta(x) \stackrel{\Delta}{=} c_n \zeta_n^\Delta(x)\bar{p}_n^\Delta$, where c_n is a normalisation constant chosen such that $\int_{\mathbb{R}^{D_x}} p_n^\Delta(x)dx = 1$ and

$$\zeta_n^\Delta(x) \stackrel{\Delta}{=} \exp\left(\tilde{h}^\top(x)\tilde{z}_n - \frac{1}{2}\Delta\|\tilde{h}(x)\|^2\right), \tag{32}$$

Here, $\tilde{z}_n = k^{-1}(t_n)z_n$ and $\tilde{h}(x) = k^{-1}(t_n)h(x)$.

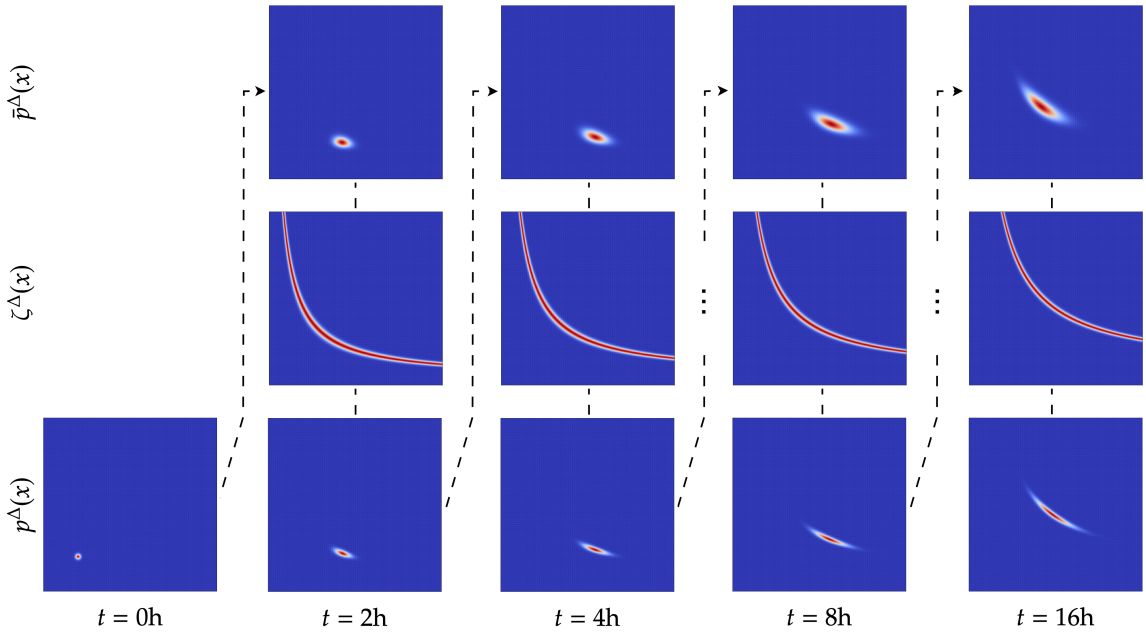

Figure 7: Splitting-up approximation for Case 1A - low-frequency.

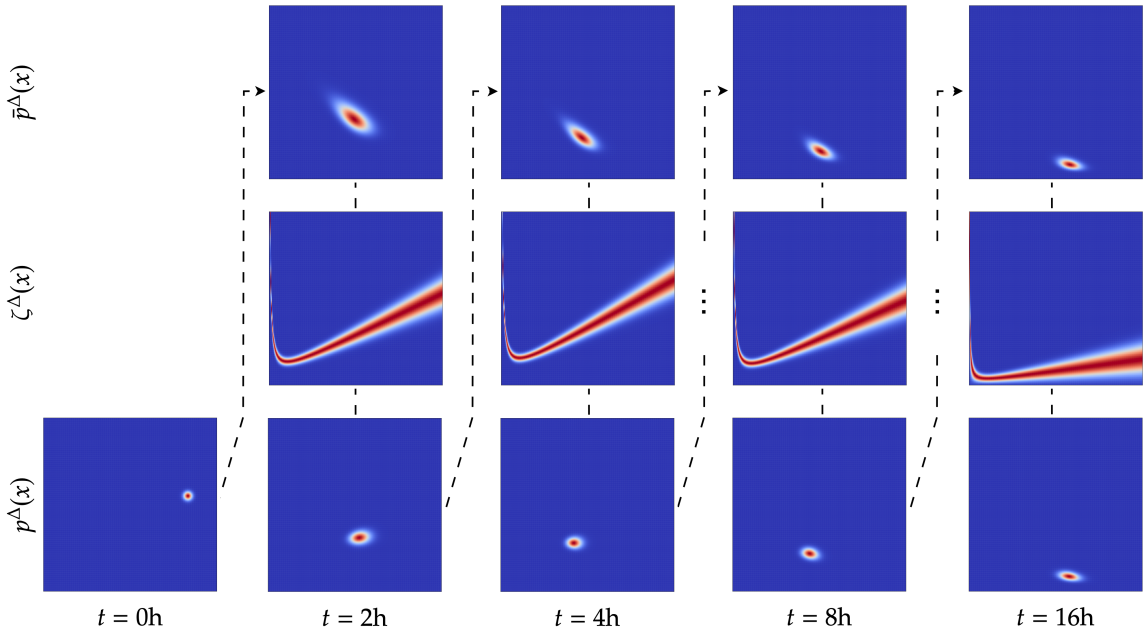

Figure 8: Splitting-up approximation for Case 2A - low-frequency.

5.2.2. Particle methods

We now switch to particle filtering methods, a technique that can be used to approximate the solution to the filtering problem by a finite number of samples, or particles, from the conditional distribution π_t. The following exposition is adapted from Bain and Crisan (2008).

Let $V_j, j = 1, \dots, N$ be N mutually independent stochastic processes, all independent of Y, each of them being a solution to Eq. (7). The pairs (V_i, Y) are identically distributed and have the same distribution as the pair (X, Y) under $\mathbb{Q}$. Also, let $w_j, j = 1, \dots, N$ be the following exponential martingales

$$w_j(t) = 1 + \int_0^t w_j(s)(k^{-1}(s)h(V_j(s)))^\top k^{-1}(s)dY_s, \quad t \geq 0, \tag{33}$$

that is, for $t \geq 0$

$$w_j(t) = \exp\left(\int_0^t (k^{-1}(s)h(V_j(s)))^\top k^{-1}(s)dY_s - \frac{1}{2}\int_0^t \|k^{-1}(s)h(V_j(s))\|^2 ds\right). \tag{34}$$

As a result, the triples $(V_j, w_j, Y), j = 1, \dots, N$ are identically distributed and have the same distribution as the triple (X, L, Y) under $\mathbb{Q}$.

Let $\rho^N = \{\rho_t^N, t \geq 0\}$ and $\pi^N = \{\pi_t^N, t \geq 0\}$ be the following sequences of measure-valued processes

$$\rho_t^N \stackrel{\Delta}{=} \frac{1}{N}\sum_{j=1}^N w_j(t)\delta_{V_j(t)}, \quad \pi_t^N \stackrel{\Delta}{=} \frac{\rho_t^N}{\rho_t^N(1)} = \sum_{j=1}^N \bar{w}_j^N(t)\delta_{V_j(t)}, \tag{35}$$

where the normalised weights $\bar{w}_j(t)$ are given by

$$\bar{w}_j(t) = \frac{w_j(t)}{\sum_{i=1}^N w_{j'}(t)}, j' = 1, \dots, N, \; t \geq 0. \tag{36}$$

That is, ρ_t^N is the empirical measure of N (random) particles with positions $V_j(t), j = 1, \dots, N$ and weights $w_j(t), j = 1, \dots, N$, and π_t^N is its normalised version.

Under certain conditions, it is possible to show both that (i) $\rho_t^N(\phi)$ converges in expectation to the unnormalised conditional estimator $\rho_t(\phi)$ with rate of convergence $1/\sqrt{N}$ and (ii) $\pi_t^N(\phi)$ converges almost surely with a rate slightly lower than $1/\sqrt{N}$ (Bain and Crisan, 2008). Additionally, $\pi_t^N(\phi)$ converges in expectation to $\pi_t(\phi)$, provided the integrability condition on $\rho_t^{-1}(1)$. Recall the martingale term L introduced in Eq. (23). The validity of these results requires the existence of higher moments of the martingale L.

Consider the scenario in which we cannot assure that the higher moments of the martingale L exist, and we only know that L is $\mathbb{Q}$-integrable. In this case, $\pi_t^N(\phi)$ still converges almost surely to $\pi_t(\phi)$ for a fixed observation path $t \mapsto Y_t$ as a consequence of the strong law of large numbers (Bain and Crisan, 2008).

Let us revisit the case outlined in Table 2 and Fig. 6. Without observations, the probabilities of the hidden state variables evolved according to the Fokker-Planck equation, as illustrated in Fig. 4 and Fig. 5. We then considered an observation model for which the solution to each of the corresponding Kushner-Stratonovich equation had to be approximated. Now, we firstly find the solution via finite-differences and a splitting-up approximation. We then apply a sequential Monte Carlo method (the continuous time bootstrap particle filter) and a linearisation method (the extended Kalman filter). Fig. 9 and 10 show the results for a fixed instant t. Algorithms are available in the supplementary material (Appendix B).

We analyse the similarity among the approximated solutions by computing the Hellinger distance $\mathcal{H}$ between two probability distributions, say F_1 and F_2. With respect to the Lebesgue measure, let f_1 and f_2 denote the density functions of the distributions F_1 and F_2, respectively. The square of the Hellinger distance is then defined as $\mathcal{H}^2(f_1, f_2) = \frac{1}{2}\int\left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 dx$. Note how this distance function is symmetric and is zero-valued when the compared densities are equal.

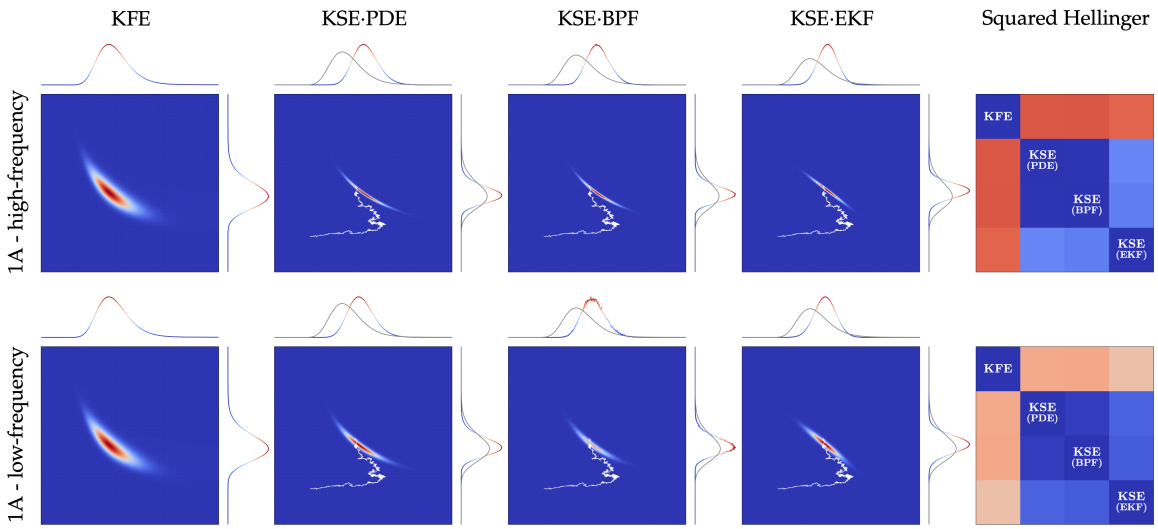

Figure 9: Case 1A - approximated solution to the Kushner-Stratonovich equation at the instant $t = 16$h with the finite differences and the splitting-up approximation for partial differential equations (KSE·PDE), with the bootstrap particle filter (KSE·BPF) and with the extended Kalman filter (KSE·EKF). For the PDE method, the discretisation in space is in the domain $(0, 5.0] \times (0, 5.0]$. We consider a refinement of the grid of a total count of 256^2 finite volumes. As for the particle filter, we consider $N = 2.5 \times 10^6$. In white, we show the trajectory from which measurements were simulated. The same trajectory is highlighted in Fig. 2.

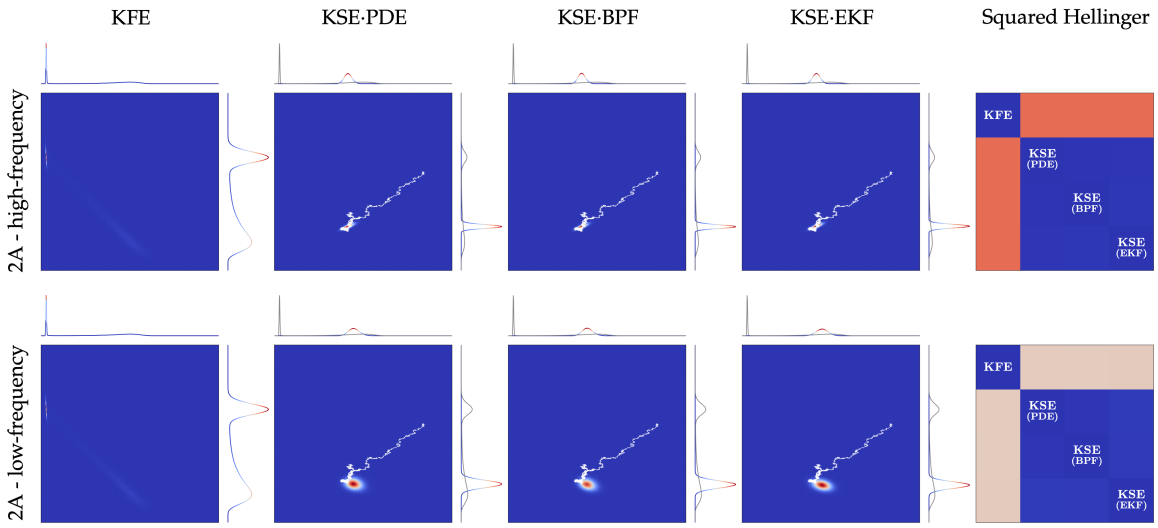

Figure 10: Case 2A - approximated solution to the Kushner-Stratonovich equation at the instant $t = 8$h with the finite differences and the splitting-up approximation for partial differential equations (KSE·PDE), with the bootstrap particle filter (KSE·BPF) and with the extended Kalman filter (KSE·EKF). For the PDE method, the discretisation in space is in the domain $(0, 3.0] \times (0, 3.0]$. We consider a refinement of the grid of a total count of 256^2 finite volumes. As for the particle filter, we consider $N = 2.5 \times 10^6$. Note how, differently from the solution to the KFE, we move more probability mass towards the washout on account of the additional information from measurements.

In Fig. 11, we plot the square of the Hellinger distance between the conditional distribution given by different setups of the method for partial differential equations and the approximation given by the bootstrap particle filter (BPF). It is clear that the discretisation error decreases with a finer refinement of the grid. We also include the

square of the Hellinger distance between the approximation with the extended Kalman filter (EKF) and with the BPF. The EKF fails to capture higher moments of the distribution, hence its divergence from a sequential Monte Carlo. Non-zero skewness and excess kurtosis result from having both a non-linear dynamic process and a non-linear measurement process with respect to the state variables. This can be seen in the Appendix C of the supplementary material, where we plot the shape of the conditional distribution for different t's.

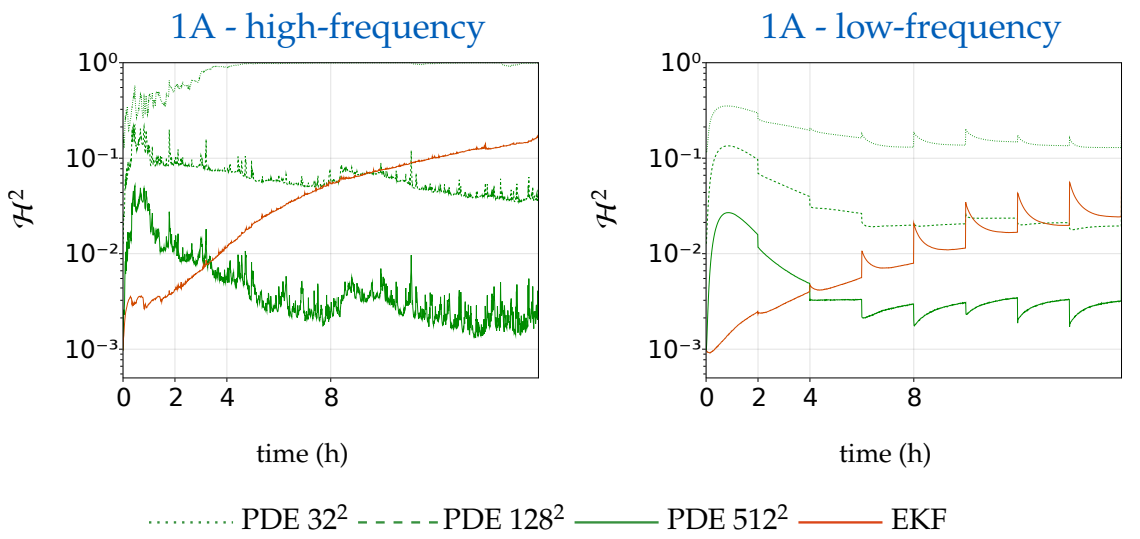

Figure 11: Case 1A - Squared Hellinger distances between the distributions obtained by approximating the solution to the Kushner-Stratonovich equation with the extended Kalman filter and with methods for partial differential equations using different grid refinements. For the PDE methods, we consider three scenarios for the refinement of the grid, each containing a total count of $32^2, 128^2, 512^2$ finite volumes, respectively. As for the particle filter, we consider $N = 2.5 \times 10^6$.

In Fig. 12, we again plot the square of the Hellinger distance between the conditional distribution given by different setups of the method for partial differential equations and the approximation given by the BPF, now for Case 2A. We again include the square of the Hellinger distance between the approximations with the EKF and the BPF. Here, the EKF suffers in the highly non-linear region of the state-space, namely when the state is near the saddle point, $t > 4$h.

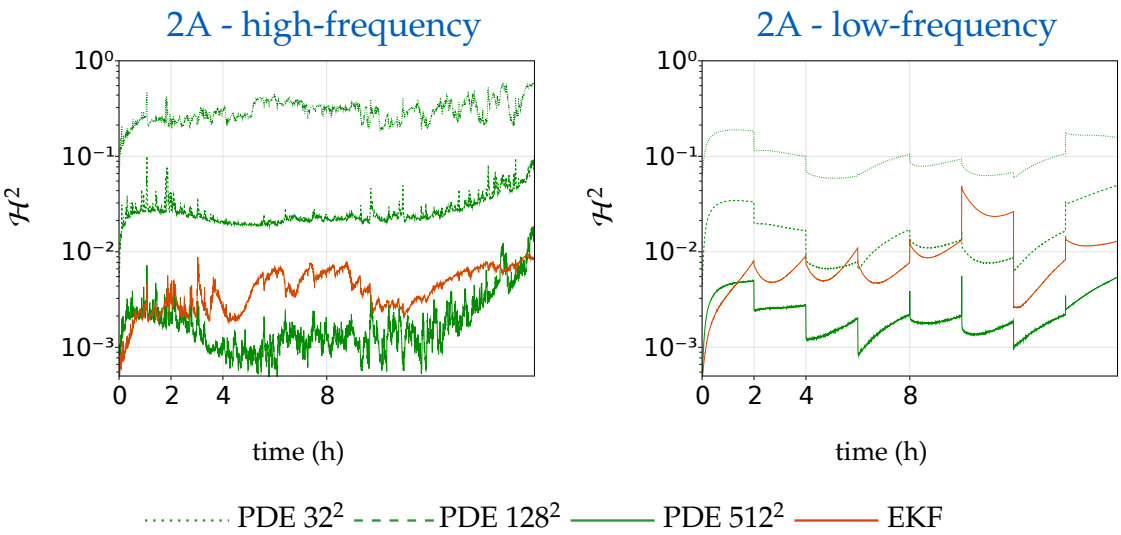

Figure 12: Case 2A - Squared Hellinger distances between the distributions obtained by approximating the solution to the Kushner-Stratonovich equation with the extended Kalman filter and with methods for partial differential equations using different grid refinements. For the PDE methods, we consider three scenarios for the refinement of the grid, each containing a total count of $32^2, 128^2, 512^2$ finite volumes, respectively. As for the particle filter, we consider $N = 2.5 \times 10^6$.

Note that the grid with 512^2 cells fails to give a perfect match between the method for partial differential equations and the particle method. At $t = 8h$, for instance, the Hellinger distance between both distributions increases, even though both methods approximate the same updating equation. The approximations with methods for partial differential equations or with the extended Kalman filter are naturally smooth, whereas for such fine grid, estimates based on the particles' paths degenerate. This issue is similar to the so-called sample impoverishment, which is a reduction in the number of distinct sample values when the process model noise variance is very low, such that samples will hardly diversify and after a few iterations all particles will collapse into a single point in the state space. There are different solutions for this problem, the simplest one to implement being an increase in the number of particles. This, however, leads to unreasonable computational load. Other solutions are more complicated to implement and, unlike the bootstrap particle filter, they might need to be tailored to the problem in hand. See Carpenter et al. (1999) and Li et al. (2014) for more methods to prevent sample impoverishment.

6. Concluding remarks

Whether it be an approximation to the Kushner-Stratonovich equation with particle methods or an approximation to the Zakai equation with methods for partial differential equations, no restrictions are imposed on the shape of the distribution. Therefore, these two approximation methods are often more accurate than linearisation methods.

In this paper, we presented a scenario where persistence of the microorganisms forming the biomass is ensured under the deterministic model, but not under the stochastic one (Case 2). The corresponding filtering distribution π_t helped to identify the actual state of the system. π_t can be used, for instance, to design a control strategy on the input variables $D(t)$ and $s_{in}(t)$ in order to avoid extinction of the microorganisms.

We assumed that measurements of biogas flow rate were available, at either high- or low-frequency. We denoted this setup by scenario A. In the supplementary material, we additionally present results for Cases 1 and 2 under three other scenarios: in B, we are given noisy measurements of the substrate concentration at low-frequency; in C, we are given noisy measurements of the biomass concentration at low-frequency; in D, we are given noisy measurements of both biomass and substrate concentrations at low-frequency. We refer the reader to Appendix C of the supplementary material for a detailed comparison of the approximations outlined in this work under all four scenarios.

Acknowledgments: this work was supported by starting grant from Aalto University (Project 96310033).
Declarations of interest: none.

References

Andrews, J.F., 1968. A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnology and bioengineering. 10. doi:10.1002/bit.260100602.

Bain, A., Crisan, D., 2008. Fundamentals of Stochastic Filtering. Stochastic Modelling and Applied Probability, Springer New York. doi:10.1007/978-0-387-76896-0.

Beneš, V.E., 1981. Exact finite-dimensional filters for certain diffusions with nonlinear drift. Stochastics 5, 65–92. doi:10.1080/17442508108833174.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., Steyer, J.P., 2001. Dynamical model development and parameter identification for anaerobic wastewater treatment process. Biotechnology and bioengineering 75. doi:10.1002/bit.10036.

Birkhoff, G.D., 1931. Proof of the ergodic theorem, pp. 656–660. doi:10.1073/pnas.17.2.656.

Bull, A.T., 2010. The renaissance of continuous culture in the post-genomics age. Journal of Industrial Microbiology and Biotechnology 37. doi:10.1007/s10295-010-0816-4.

Cai, Z., Gland, F.L., Zhang, H., 1995. An adaptive local grid refinement method for nonlinear filtering. Research report 2679, INRIA. URL: https://hal.inria.fr/inria-00074012/document.

Campillo, F., Joannides, M., Larramendy-Valverde, I., 2014. Approximation of the fokker–planck equation of the stochastic chemostat. Mathematics and Computers in Simulation 99, 37–53. doi:10.1016/j.matcom.2013.04.012.

Carpenter, J., Cliffordy, P., Fearnhead, P., 1999. An improved particle filter for nonlinear problems. IEE Proc. Radar, Sonar Navig. 146, 2–7. doi:10.1049/ip-rsn:19990255.

Chung, K., Williams, R., 2014. Introduction to Stochastic Integration. Modern Birkhäuser Classics, Springer New York.

Clark, J.M.C., 1978. The design of robust approximations to the stochastic differential equations of nonlinear filtering, in: Communication Systems and Random Process Theory. Springer Netherlands. doi:10.1007/978-94-011-7577-7_41.

Cohen, S., Elliott, R., 2015. Stochastic Calculus and Applications. Probability and Its Applications, Springer New York. doi:10.1007/978-1-4939-2867-5.

Crisan, D., Kouritzin, M., Xiong, J., 2009. Nonlinear filtering with signal dependent observation noise. Electronic Journal of Probability 14. doi:10.1214/EJP.v14-687.

Crisan, D., Kurtz, T.G., Ortiz-Latorre, S., 2021. Particle representation for the solution of the filtering problem. application to the error expansion of filtering discretizations. Journal of Stochastic Analysis 2. doi:10.31390/josa.2.3.15.

Davis, M.H.A., 1982. A pathwise solution of the equations of nonlinear filtering. Theory of Probability & Its Applications 27. doi:10.1137/1127017.

van Handel, R., 2007. Filtering, stability, and robustness. Ph.D. thesis. doi:10.7907/4p53-1h42.

Harmand, J., Lobry, C., Rapaport, A., Sari, T., 2017. The Chemostat: Mathematical Theory of Microorganism Cultures. John Wiley and Sons, ISTE. doi:10.1002/9781119437215.

Hoskisson, P.A., Hobbs, G., 2005. Continuous culture–making a comeback? Microbiology 151. doi:10.1099/mic.0.27924-0.

Imhof, L., Walcher, S., 2005. Exclusion and persistence in deterministic and stochastic chemostat models. Journal of Differential Equations 217. doi:10.1016/j.jde.2005.06.017.

Ito, K., 1944. Stochastic integral. Proceedings of the Imperial Academy 20. doi:10.3792/pia/1195572786.

Kálmán, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. Journal of Basic Engineering 83. doi:10.1115/1.3658902.

Klebaner, F., 2005. Introduction to Stochastic Calculus with Applications. Introduction to Stochastic Calculus with Applications, Imperial College Press.

Kutschireiter, A., Surace, S.C., Pfister, J.P., 2020. The hitchhiker's guide to nonlinear filtering. Journal of Mathematical Psychology 94. doi:10.1016/j.jmp.2019.102307.

Le Gland, F., 1989. Time discretization of nonlinear filtering equations, in: Proceedings of the 28th IEEE Conference on Decision and Control, Tampa 1989, IEEE–CSS. pp. 2601–2606. doi:10.1109/CDC.1989.70650.

LeGland, F., 1992. Splitting-up approximation for SPDE's and SDE's with application to nonlinear filtering, in: Rozovskii, B.L., Sowers, R.B. (Eds.), Stochastic Partial Differential Equations and Their Applications, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 177–187. doi:10.1007/BFb0007332.

Li, T., Sun, S., Sattar, T.P., Corchado, J.M., 2014. Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. Expert Systems with Applications 41. doi:10.1016/j.eswa.2013.12.031.

Mao, X., 2007. Stochastic differential equations and applications. 2nd ed. doi:10.1533/9780857099402.

Monod, J., 1950. La technique de culture continue: theorie et applications, Academic Press New York. doi:10.1016/B978-0-12-460482-7.50023-3.

Nobuyuki, I., Watanabe, S., 1981. Stochastic Differential Equations and Diffusion Processes. North-Holland Pub. Co.; Kodansha; Sole Distributors for the U.S.A. and Canada. doi:10.1016/S0924-6509(08)70217-4.

Renard, P., Dochain, D., Bastin, G., Naveau, H.P., Nyns, J., 1988. Adaptive control of anaerobic digestion processes—a pilot-scale application. Biotechnology and Bioengineering 31. doi:10.1002/bit.260310402.

Wang, L., Jiang, D., 2017. A note on the stationary distribution of the stochastic chemostat model with general response functions. Applied Mathematics Letters 73, 22–28. doi:10.1016/j.aml.2017.04.029.

Zvonkin, A.K., 1974. A transformation of the phase space of a diffusion process that removes the drift. Mathematics of the USSR-Sbornik 22, 129–149. doi:10.1070/sm1974v022n01abeh001689.