

1 **APPROXIMATIONS TO THE SOLUTION OF THE
2 KUSHNER-STRATONOVICH EQUATION FOR THE STOCHASTIC
3 CHEMOSTAT***

4 JOSÉ AUGUSTO FONTENELE MAGALHÃES[†], MUHAMMAD FUADY EMZIR[‡], AND
5 FRANCESCO CORONA[§]

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

20 **Key words.** stochastic chemostat; state estimation, non-linear filtering, Fokker-Planck equa-
21 tion, Kushner-Stratonovich equation

22 **MSC codes.** 92D25, 60G35, 60H15, 62M20, 65C05

23 **1. Introduction.** The aim of filtering is to estimate the current state of an
24 evolving dynamical system, customarily modelled by a stochastic process, denoted
25 by X and called the signal process. The signal process cannot be measured directly,
26 but only via a related process Y , termed the observation process. The aim is to
27 compute π_t , the conditional distribution of the signal X_t , at the current time t given
28 the observation data accumulated up to this time. More precisely, we compute the
29 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 $\pi = (\pi_t, t \geq 0)$
30 describing this conditional distribution. Such process is the solution to a non-linear
31 stochastic partial integral differential equation, the Kushner-Stratonovich equation.
32 The exact form of the process π can be derived for a specific class of problems, for
33 instance, those with a linearly evolving observation process [3, 21]. Even in this
34 scenario, however, the signal is only allowed to have a non-linear drift under some
35 restrictive conditions (see Beneš condition [3]). In general, π is an infinite-dimensional
36 stochastic process and finding it in closed-form is a non-trivial task due to the inherent
37 non-linearities of the system. In this case, the solution must be approximated.

38 Here, our goal is to approximate the conditional distribution π_t for a process
39 performed in a chemostat in order to learn more about the system. The chemostat
40 was the major player in the use of continuous culture systems to characterise certain

*Preprint April 6, 2023.

[†]Department of Chemical and Metallurgical Engineering, Aalto University, Espoo, FI (augusto.magalhaes@aalto.fi).

[‡]Control and Instrumentation Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, SA (muhammad.emzir@kfupm.edu.sa).

[§]Department of Chemical and Metallurgical Engineering, Aalto University, Espoo, FI (francesco.corona@aalto.fi)

42 kinds of microbes in the mid-20th century, prior to when molecular biology and thence
43 genomics took centre stage instead [6, 17]. Being a realistic and thoroughly tested
44 continuous culture system, it has regained some interest in this post-genomic era. For
45 instance, chemostat-grown cultures lead to reproducible and homogeneous results,
46 which are necessary to identify the roles genes play in the biology of organisms [18].

47 We focus on estimating the state of a chemostat to understand its operation and
48 to devise strategies for optimal operation. Here, we consider two substrate-controlled
49 growth models: a basic model (Monod kinetics) and a more complex model, where
50 an excessive level of the substrate can inhibit the growth (Haldane kinetics). We
51 consider a differential description of the substrate and biomass concentrations in a
52 chemostat under environmental fluctuations [19]; in particular, chemostats subjected
53 to geometric Brownian motion. Our best knowledge about the state of the system
54 is given by its conditional distribution in time, given the distribution of the initial
55 state. Such a function solves a deterministic partial differential equation (PDE), the
56 Kolmogorov forward equation. Our focus is to refine our knowledge about the state
57 of the chemostat when additional information about the system is available in the
58 form of measurements. To describe how measurements and states are related, we
59 augment the dynamics with an observation model in which the measurements are
60 also subjected to Brownian motion. When measurements are collected, our estimate
61 of the state is given by the conditional distribution π_t , now given the distribution of
62 the initial state and the collected measurements.

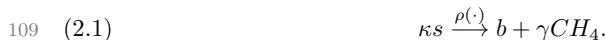
63 For non-linear systems, the classical method for finding π_t is the extended Kalman
64 filter (EKF), which linearises certain non-linearities in the model so that the tradi-
65 tional Kalman-Bucy filter (KBF, [21]) can be applied. A less common approach
66 consists of solving an associated stochastic partial differential equation (SPDE), the
67 Zakai equation. Its solution is an unnormalised density function, given the distribu-
68 tion of the initial state and the collected measurements. In this approach, the solution
69 is also unavailable in closed form for our system, and it must be approximated numer-
70 ically [2]. Despite being an SPDE, classical PDEs methods can still be applied. These
71 methods are successful in low dimensions of the state space but cannot be applied to
72 high-dimensional problems. This is because the number of finite elements required for
73 an accurate approximation goes up exponentially with the dimension of the signal.

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

79 In the context of the stochastic chemostat, the existing state-of-the-art comprises
80 approximations to the first two moments of the distribution π_t [16, 31]. We propose to
81 advance beyond that by presenting results on the approximation to the entire distribu-
82 tion π_t , from which higher-moments could be derived. We focus on the approximation
83 with methods for PDEs. In this work, the available observations are a transformed
84 and noisy version of the signal. In particular, we observe the flow rate of outgoing bio-
85 gas produced inside the chemostat, and not the biomass and substrate concentrations
86 directly. We present alternative scenarios in the supplementary material, in which
87 noisy observations of biomass and/or substrate concentrations are available. After
88 presenting how the solution to π_t looks like, we evaluate the method for PDEs by
89 using the Hellinger distance. More specifically, we quantify the similarity between the
90 resulting approximation and the solution approximated via sequential Monte Carlo.
91 We also present how the EKF 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. In Sections 5 and 6, we present the filtering equations; we explore the problem when observations are corrupted by Gaussian noise, 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 $\rho(\cdot)$:



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

$$(2.2) \quad \frac{db(t)}{dt} = (\mu(\cdot) - D(t))b(t) + D(t)b_{in}(t), \quad \frac{ds(t)}{dt} = D(t)(s_{in}(t) - s(t)) - \kappa\mu(\cdot)b(t),$$

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 Subsection SM1.1 of the supplementary material 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 [17], 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. (2.1) 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 an ordinary differential equation (ODE) on the signal $x(t)$ to describe the dynamics of the minimal system:

$$(2.3) \quad d \begin{bmatrix} b(t) \\ s(t) \end{bmatrix} = \underbrace{\begin{bmatrix} (\mu(s(t)) - D(t))b(t) \\ D(t)(s_{in}(t) - s(t)) - \kappa\mu(s(t))b(t) \end{bmatrix}}_{f(t, x(t), u(t)|\theta)} dt, \quad \underbrace{\begin{bmatrix} b(0) \\ s(0) \end{bmatrix}}_{x(0)} = \underbrace{\begin{bmatrix} b_0 \\ s_0 \end{bmatrix}}_{x_0},$$

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

135 the system has a known starting position, say point x_0 at $t = 0$, and $u(t)$ is given, the
 136 function $x(t)$, also called a solution to Eq. (2.3), is then completely determined in the
 137 future, i.e. we know $x(t)$ for all $t > 0$.

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

141 Case 1 - Monotonic growth μ - Monod type

142 This first case consists of the Monod relation of growth [28], which assumes that
 143 the growth rate is zero when there is no substrate and tends to an upper limit when
 144 the substrate is in great excess:

145 (2.4)
$$\mu(s) = \frac{\mu_{max}s}{K_s + s}.$$

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

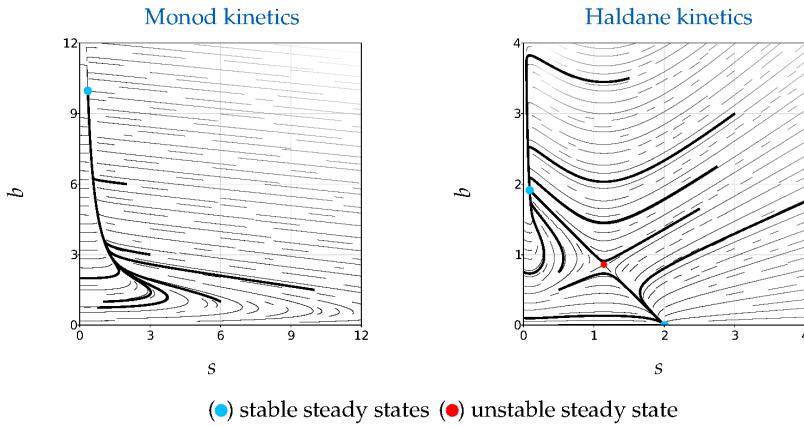
147 Case 2 - Non-monotonic growth μ - Haldane type

148 Some experiments indicate that excessive substrate may inhibit the growth of a
 149 microbial population in the biomass [1]. To consider such scenario, this case works
 150 with the following growth function:

151 (2.5)
$$\mu(s) = \frac{\mu_0 s}{K_s + s + I_s s^2}.$$

152 Here, μ depends on the growth parameter μ_0 , half-saturation coefficient K_s and in-
 153 hibitor coefficient I_s .

154 Let us now look at the solution to Eq. (2.3), given the parameters from Table 1.



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

158 Consider equilibrium pairs (b_e, s_e) for which the right-hand side of Eq. (2.3) is
 159 zero. Case 1 has the equilibrium point $E_1^{(1)} = (9.966, 0.345)$. Furthermore, $E_1^{(1)}$ is a
 160 locally exponentially stable equilibrium [17, Section 2.1.2]. Case 2 has a trivial equilib-
 161 rium point $E_2^{(1)} = (0, s_{in})$. In addition, it has interior equilibria $E_2^{(2)} = (0.859, 1.141)$
 162 and $E_2^{(3)} = (1.912, 0.088)$. It is possible to verify that one of the two interior equilibria
 163 is a stable nodal point and the other is a saddle point [17, Section 2.1.3].

163 **3. Continuous-time dynamic models.**

Table 1: Model parameters and inputs for Eq. (2.3).

	Model parameters		Inputs	
	Specific growth	κ	$D(t) = D(V = 100L)$ (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

3.1. Stochastic differential equations.

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) randomizing 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 $\forall v \geq 0$, 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. (2.3), the following SDE emerges:

$$(3.1) \quad dX_t = f(t, X_t, u(t)|\theta)dt + g(t, X_t, u(t)|\theta)dB_t^x, \quad X_0 = \xi,$$

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 will omit the reference to the vector θ of parameters. We assume a constant input $u(t)$ and autonomous functions f and g , such that Eq. (3.1) can be rewritten as

$$(3.2a) \quad dX_t = f(X_t)dt + g(X_t)dB_t^x, \quad X_0 = \xi,$$

or equivalently in coordinate form

$$(3.2b) \quad dX_t^{d_x} = f_{d_x}(X_t)dt + \sum_{m=1}^M g_{d_x m}(X_t)dB_t^{x,m}, \quad X_0^{d_x} = \xi^{d_x}, \quad d_x = 1, \dots, D_x,$$

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 mapping

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

196 The SDEs in (3.2) are equivalent to the following stochastic integral equation:

197 (3.3)
$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t g(X_s)dB_s^x, \quad t \geq 0,$$

198 where the first integral on the right-hand side is a Riemann integral, and the second
199 one is an Itô integral. Here, f and g satisfy the Lipschitz condition for existence and
200 uniqueness of solutions. In order to demonstrate challenges concerning SDE existence
201 and uniqueness, we now discuss what we mean by solutions of Eq. (3.2) and Eq. (3.3).

202 **3.1.2. Existence and uniqueness.** The goal of this section is to find conditions
203 on the coefficients f and g which guarantee the existence and uniqueness of solutions.

204 To build a notion of time into our probability space, we specify which sub- σ -
205 algebra of information in \mathcal{F} can be gathered by time t . We consider the increasing
206 sequence $\{\mathcal{F}_t\}$ of σ -algebras of \mathcal{F} for which for $t, s \in \mathbb{R}_+$ with $t \leq s$, we have $\mathcal{F}_t \subseteq$
207 $\mathcal{F}_s \subseteq \mathcal{F}$. The quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, \infty)}, \mathbb{P})$ forms our filtered probability space, on
208 which we can now define an M -dimensional \mathcal{F}_t -Brownian motion B_t^x .

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

211 **DEFINITION 3.2** (Strong solutions [10, Section 10.4]). *Let $B^x = (B_t^x, t \geq 0)$ be a
212 Brownian motion and ξ be a random variable on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, \infty)}, \mathbb{P})$ independent of
213 B^x . A stochastic process X is a strong solution of a stochastic differential equation if*

- 214 • X is adapted to the filtration $\{\mathcal{F}_t\}$;
- 215 • \mathbb{P} -almost surely $X_0 = \xi$;
- 216 • \mathbb{P} -almost surely $\int_0^t |f_{d_x}(X_s)| + (g_{dx,m}(X_s))^2 ds < \infty$ holds for all $t \in [0, \infty)$,
- 217 $d_x = 1, \dots, D_x$ and $m = 1, \dots, M$;
- 218 • \mathbb{P} -almost surely X solves Eq. (3.2).

219 As usual, we assume that the sigma-algebra \mathcal{F}_t is completed with all the \mathbb{P} -null
220 sets in Ω [10, Section 10.4]. This definition means that if we are given a probability
221 space which carries B^x and ξ , the solution X must be adapted to the generated
222 filtration $\{\mathcal{F}_t\}$; that is, the path $\{X_t\}_{t \geq 0}$ is completely characterized by the path of
223 the underlying driving process (Brownian motion) and the initial condition.

224 The condition of being adapted to the generated filtration is sometimes too strict
225 (the famous example of Tanaka's equation, found in, for example, [34]). One can find
226 pairs (X, B^x) which are solution to the SDE (3.2) instead of simply seeking X . Such
227 pairs are called *weak solutions*. In what follows, we consider mostly the case of strong
228 solutions of Itô stochastic differential equations.

229 **THEOREM 3.3** (Existence and uniqueness [27, Theorem 2.2]). *Suppose that ξ is a
230 random variable on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, \infty)}, \mathbb{P})$, for which ξ is independent of the Brownian
231 motion B^x also defined on this space. Suppose that*

- 232 • $\xi \in \mathcal{L}^2(\mathbb{P})$, i.e. ξ is square-integrable under the probability measure \mathbb{P} ;
- 233 • f, g satisfy the Lipschitz continuous and linear growth conditions in x .

234 Then there exists a solution $X = (X_t, t \geq 0)$ to the associated stochastic differential
235 equation (3.2) satisfying the initial condition ξ . Moreover, the solution is unique and
236 satisfies $X \in \mathcal{L}^2(\mathbb{T} \times \mathbb{P})$, where \mathbb{T} is the Lebesgue measure on $[0, \infty)$.

237 **EXAMPLE 3.4** (Geometric Brownian motion). *Consider the unidimensional case,
238 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 $\omega_x > 0$:*

239 (3.4)
$$dX_t = \omega_x X_t dB_t^x,$$

240 with initial condition $X_0 = \xi \equiv 1$. Now define a function $V(x) \stackrel{\Delta}{=} \ln(x)$. We apply
 241 Itô's formula [20] on the process $V(X)$, resulting in

$$242 \quad d(\ln(X_t)) = \omega_x dB_t^x - \frac{1}{2}\omega_x^2 t, \quad t \geq 0.$$

243 Since $B_0 = 0$, we further get $X_t = X_0 \exp(\omega_x B_t^x - \frac{1}{2}\omega_x^2 t)$. For this SDE, both exist-
 244 ence and uniqueness properties hold. Moreover, if the initial condition is constrained
 245 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 3.4, assume that this random amount is a function of the current concentration itself. To give an idea of the driving process added to Eq. (2.3), we consider an 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 [19] has the following dynamic behaviour:

$$(3.5) \quad d \begin{bmatrix} b_t \\ s_t \end{bmatrix} = \underbrace{\begin{bmatrix} (\mu(s_t) - D)b_t \\ D(s_{in} - s_t) - \kappa\mu(s_t)b_t \end{bmatrix}}_{f(X_t)} dt + \underbrace{\begin{bmatrix} \omega_b b_t & 0 \\ 0 & \omega_s s_t \end{bmatrix}}_{g(X_t)} d \begin{bmatrix} B_t^b \\ B_t^s \end{bmatrix},$$

with $X_0 \equiv (b_0, s_0)$. Now, we see how the solution to the model in Eq. (3.5) 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.

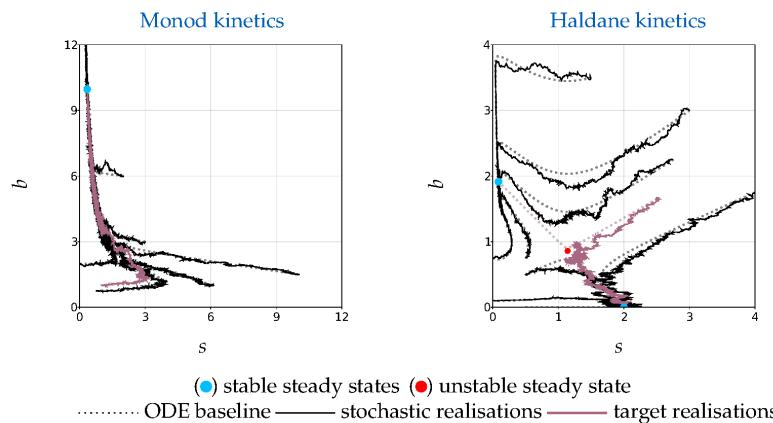


Fig. 2: Trajectories from the stochastic model with growth function of the Monod type (left) and of the Haldane type (right). The dashed lines in the background show the trajectories from the deterministic model, as in Fig. 1. Here, a target realisation denotes a simulated trajectory which will be estimated 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 [17]. 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 X_0 appropriately.

THEOREM 3.5. *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. (3.5) with growth functions from Eq. (2.4) or Eq. (2.5). 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. See Subsection SM1.3 in the supplementary material.

Thm. 3.5 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 was discussed in [33], where sufficient conditions for ergodicity were imposed on s_{in} , D , ω_b and ω_s .

Let ζ be an integrable function with respect to the measure μ , an invariant measure of the process X . If the system is ergodic, then $\mathbb{P}\left(\frac{1}{t} \int_0^t \zeta(X_s) ds \rightarrow \int \zeta(y) \mu(dy)\right) = 1$ [5]. 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 $\langle \zeta \rangle_{\infty,1}^{tim}$ and $\langle \zeta \rangle_{t,\infty}^{ens}$, respectively. Let $X_t^{(n)}$ denote the n -th realisation of X_t , then define

$$(3.6) \quad \langle \zeta \rangle_{\infty,1}^{tim} \triangleq \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \zeta(X_s) ds, \quad \langle \zeta \rangle_{t,\infty}^{ens} \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \zeta(X_t^{(n)}), X_t^{(n)} \sim \mu.$$

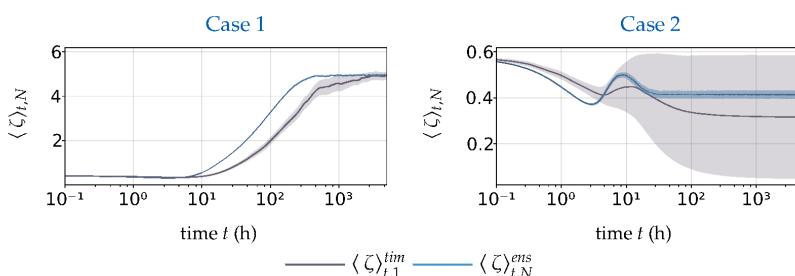


Fig. 3: Estimates of the expectation and the time-average of ζ , where $\zeta(x) = 0.5b - 0.1s$. The estimates are based on 30 experiments. Shaded regions are delimited by two standard deviations above and below the mean. For the Monte Carlo approximation, we used an ensemble of $N = 512$ samples of 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 $\langle \zeta \rangle_{\infty,1}^{tim}$ and $\langle \zeta \rangle_{t,\infty}^{ens}$. In fact, this is confirmed by the fact that the conditions for ergodicity outlined in [33] 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, as confirmed by the non-commutativity in the limits.

248 **3.2. The Markov property and the Kolmogorov forward equation.** In
 249 this section, we use the theory of Markov processes to study the properties of the
 250 SDE (3.5), given that its solution, if it exists, satisfies the Markov property. We will
 251 see that this allows us to apply the Kolmogorov equations (in particular the forward,
 252 or Fokker-Planck, equation) from the theory of Markov process into the SDE setting.
 253 This can then be used to express expectations of functions of the solution to Eq. (3.5)
 254 in terms of certain non-random PDEs.

255 DEFINITION 3.6 (Markov process). *Let the process $X = (X_t, t \geq 0)$ be measurable
 256 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
 257 process if we have*

$$258 \quad (3.7) \quad \mathbb{E}[\phi(X_t)] = \mathbb{E}[\phi(X_t)|X_\tau] =: \phi_{t-\tau}(X_\tau),$$

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

260 THEOREM 3.7 (Markov property and SDEs [12, Theorem 17.2.3]). *Assume existence and uniqueness of the solution to the SDE in Eq. (3.2). The solution X that
 261 satisfies such SDE is an \mathcal{F}_t -Markov process.*

263 The intuitive content behind Thm. 3.7 is clear. Firstly, suppose that the process
 264 X generated by the SDE in Eq. (3.2) begins at time zero, and that we watch it up to
 265 time τ . Based on this information, we want to compute the conditional expectation
 266 $\phi_{t-\tau}(X_\tau)$, where $t \geq \tau$. Then we can pretend that the process is starting at time τ
 267 at its current position, generate the solution to the SDE corresponding to this initial
 268 condition, and compute the expected value of $\phi(X_t)$ generated in this way.

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

273 Let $p(t, \cdot)$ denote the density with respect to Lebesgue measure for the probability
 274 measure \mathbb{P} at time t . We describe the process in terms of a transition kernel
 275 $p(t, x|0, x_0) \triangleq \mathbb{P}(X_t = x | X_0 = x_0), t \geq 0$. This allows us to compute the probability
 276 of having the process X at a point $x \in \mathbb{R}^{D_x}$ at instant t conditioned on starting at
 277 $x_0 \in \mathbb{R}^{D_x}$ at time $t_0 = 0$. We then proceed to use the property shown in Def. 3.6.

$$278 \quad \mathbb{E}[\phi(X_t)] = \mathbb{E}[\mathbb{E}[\phi(X_t)|\mathcal{F}_0]] = \mathbb{E}[\mathbb{E}[\phi(X_t)|X_0]] \\ 279 \quad = \int_{\mathbb{R}^{D_x}} \int_{\mathbb{R}^{D_x}} \phi(x) p(t, x|0, x_0) p(0, x_0) dx_0 dx \\ 280 \quad = \int_{\mathbb{R}^{D_x}} \phi(x) p(t, x) dx. \\ 281 \quad (3.8)$$

282 Now, we find the time evolution of the probability density $p(t, x)$ by differentiating
 283 both sides of equation Eq. (3.8) with respect to time,

$$284 \quad (3.9) \quad \frac{d}{dt} \mathbb{E}[\phi(X_t)] = \int_{\mathbb{R}^{D_x}} \phi(x) \frac{\partial}{\partial t} p(t, x) dx = \int_{\mathbb{R}^{D_x}} \phi(x) \mathcal{A}^* p(t, x) dx,$$

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

287 The following proposition is adapted from [29, Chapter 4].

288 PROPOSITION 3.8 (The Kolmogorov forward equation (KFE)). *Assume that the
 289 \mathcal{F}_t -Markov process $X = (X_t, t \geq 0)$ is solution to the SDE (3.2). Suppose that the*

290 density $p(t, x)$ exists and is C^1 in t and C^2 in x , then $p(t, x)$ satisfies the Kolmogorov
291 forward equation, also called the Fokker-Planck equation:

$$292 \quad (3.10) \quad \frac{\partial}{\partial t} p(t, x) = \mathcal{A}^* p(t, x), \quad p(0, x) = p_0(x),$$

293 where the form of the adjoint operator \mathcal{A}^* is

$$294 \quad (3.11) \quad \mathcal{A}^* p(t, x) = - \sum_{d_x=1}^{D_x} \frac{\partial}{\partial x_{d_x}} [f_{d_x}(x)p(t, x)] + \sum_{d_x=1}^{D_x} \sum_{d'_x=1}^{D_x} \frac{\partial^2}{\partial x_{d_x} \partial x_{d'_x}} [G_{d_x d'_x}(x)p(t, x)].$$

295 Here, $G_{d_x d'_x}(x) = \frac{1}{2} \sum_{m=1}^M g_{d_x m}(x)g_{d'_x m}(x)$ 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. (3.5) through its associated KFE, as in [8]. 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 [8]: we make use of an upwind differencing scheme for the space discretization and of an implicit scheme for the time discretization.

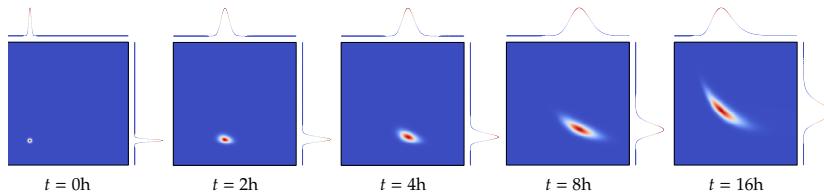


Fig. 4: Case 1 - solution to the KFE for the growth function of Monod type. Initial condition is $X_0 \sim \mathcal{N}([1, 1]^\top, 0.05^2 I_{2 \times 2})$. The discretization in space is in the domain $(0, 5] \times (0, 5]$, with a count of 256^2 finite volumes. Marginal density functions are represented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

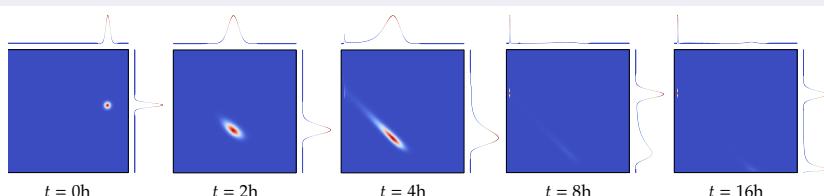


Fig. 5: Case 2 - solution to the KFE for the growth function of Haldane type. Initial condition is $X_0 \sim \mathcal{N}([1.65, 2.5]^\top, 0.05^2 I_{2 \times 2})$. The discretization in space is in the domain $(0, 3] \times (0, 3]$, with a count of 512^2 finite volumes. Again, marginal density functions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

296

297 So far, we have worked with a model describing the dynamic behaviour of the
298 system. The prior probabilistic knowledge about how the state of the system evolves
299 was formally embedded by the adjoint operator \mathcal{A}^* in the Kolmogorov forward equa-
300 tion. Our focus, however, is to refine our knowledge about the state when information
301 about the system is available in the form of measurements. In the following section,
302 we resort to an observation model to describe how the signal gets corrupted. Gener-
303 ally, the state signal cannot be measured directly (e.g. only certain components of a

304 vector-valued process are observed), and the measurements are also subjected to error.
 305 To describe how measurements and states are related, we combine the dynamics
 306 with an observation model in which the error is also stochastic.

307 **4. Observation models.** Let $0 = t_0 < t_1 < \dots < t_n \dots$ be a uniform partition
 308 of the interval $[t_0, \infty)$. We let discrete-time measurements z be available from the
 309 system, in the form of

$$310 \quad (4.1) \quad z_t = h(X_t) + k(t)v_t, \quad t \in \{t_0, t_1 \dots\}.$$

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

314 Instead of proceeding with the model in Eq. (4.1), we will work with an integrated
 315 form of the observations; we use an observation model for a process $Y = (Y_t, t \geq 0)$,
 316 where this process represents the integral of the actually observed process. To explain
 317 this, we begin by defining a D_y -dimensional \mathcal{F}_t -Brownian motion B^y independent of
 318 B^x . We then introduce a process Y which obeys

$$319 \quad (4.2) \quad dY_t = h(X_t)dt + k(t)dB_t^y, \quad Y_0 = 0.$$

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

322 Notice that Y as modelled by Eq. (4.2) is the integral of the observed process,
 323 so that in discrete-time one really observes z rather than Y . We can think of v_t in
 324 Eq. (4.1) as the hypothetical time derivative of the Brownian motion in Eq. (4.2) at
 325 time t , and that v has zero mean and covariance $\mathbb{E}(v_s v_t) = \delta(t - s)$, where $\delta(\cdot)$ is the
 326 Dirac's delta generalized function. By working with the integrated form, we obtain a
 327 mathematical model in which we can disregard the need to work with $\delta(\cdot)$ and instead
 328 use the machinery of stochastic calculus used in the description of the signal model.

In Section 2, we introduced two 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 [4, 30], can be written as $Q(s, b) = \gamma\mu(s)b$, where γ is the yield coefficient for CH_4 in Eq. (2.1). The integrated process considered is

$$(4.3) \quad d[y_t^A] = \underbrace{[Q(s, b)]}_{h_A(x_t)} dt + \underbrace{[\omega_A]}_{k_A(t)} d[B_t^y]^A.$$

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 discretization 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 .

	Observation function	Model parameters			Simulation parameters		
		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	$27.5\mu(s)b$	$0.01 \left(\frac{\delta_y}{\delta_x} \right)^{0.5}$	0.01	2.0			
2A - high-frequency	$2.75\mu(s)b$	0.01	0.005	0.005			
2A - low-frequency	$2.75\mu(s)b$	$0.01 \left(\frac{\delta_y}{\delta_x} \right)^{0.5}$	0.005	2.0			

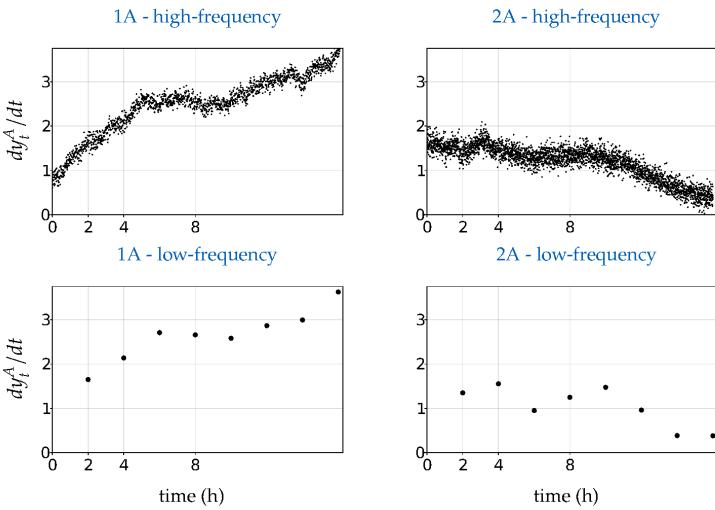


Fig. 6: Measurements simulated from the trajectories highlighted in Fig. 2.

330

331 **5. The filtering problem.** To ensure theoretical developments, we assume that
 332 $X_0 \in \mathcal{L}^2(\mathbb{P})$ (remember that we need this assumption to determine the random process
 333 $X = (X_t, t \geq 0)$ uniquely, according to Thm. 3.3).

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

$$341 \quad (5.1) \quad \pi_t(A) \triangleq \mathbb{P}(X_t \in A | \mathcal{F}_t^Y), \quad \pi_t(\phi) = \mathbb{E}_{\mathbb{P}}[\phi(X_t) | \mathcal{F}_t^Y] = \int_{\mathbb{R}^{D_x}} \phi(x) p_t(x) dx.$$

342 Here, A is an arbitrary set in the σ -algebra $\mathcal{B}([0, \infty) \otimes \mathbb{R}^{D_x})$, ϕ is an arbitrary function
 343 of the state and $\mathbb{E}_{\mathbb{P}}$ denotes expectation under the density p_t . The first equation takes
 344 values in the space of functions from $[0, \infty) \times \mathbb{R}^{D_x}$ into $[0, 1]$. The second equation
 345 allows us to get desirable information about the state through a transformation ϕ ,
 346 assuming that the expectation exists (i.e. $\int |\phi(x)| p_t(x) dx < \infty$). Whether π_t has

347 a density with respect to a reference measure (e.g. Lebesgue measure) is rigorously
 348 discussed in [2]. We proceed by assuming that we can prove the existence of p_t .

349 **5.1. Filtering equations.** We continue working on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, \infty)}, \mathbb{P})$. Our
 350 goal is to compute π_t on the basis of the observations $\{Y_s : s \leq t\}$. We begin by
 351 considering the signal process with evolution in time described by Eq. (3.2), and also
 352 the observation model given by Eq. (4.2). We rewrite here both equations:

$$353 \quad dX_t = f(X_t)dt + g(X_t)dB_t^x, X_0 = \xi; \quad dY_t = h(X_t)dt + k(t)dB_t^y, Y_0 = 0.$$

354 A detailed and formal derivation of the filtering equations is available in [2, 32].
 355 Here, the derivation is briefly outlined, following similar steps as in [23]. Among
 356 the possible ways of deducing the evolution equation for π_t , there are two commonly
 357 used approaches, namely the change of measure method and the innovation process
 358 method. We present the change of probability measure method. For the second
 359 approach, see [2, Section 3.7].

360 **5.1.1. Changes of measure.** If we have two such measures \mathbb{P} and \mathbb{Q} , then \mathbb{P}
 361 is called *absolutely continuous* with respect to \mathbb{Q} if every nullset of \mathbb{Q} is a null set of
 362 \mathbb{P} . Moreover, \mathbb{P} and \mathbb{Q} are called *equivalent* if they have the nullsets. In other words,
 363 if A denotes an event, and $\mathbb{P}(A)$ denotes its probability, then equivalence means that
 364 $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$.

365 Changing the measure allows us to compute expectations of a measurable function
 366 $\phi(x)$ with respect to a measure \mathbb{Q} , which were originally expressed with respect to
 367 measure \mathbb{P} . To see this, consider the two measures \mathbb{P} and \mathbb{Q} for some real-valued
 368 random variable X , and write them in terms of their densities p, q (again, with respect
 369 to Lebesgue measure). We then have

$$370 \quad (5.2) \quad \mathbb{E}_{\mathbb{P}} [\phi(X)] = \int \phi(x)p(x)dx = \int \frac{p(x)}{q(x)}\phi(x)q(x)dx = \mathbb{E}_{\mathbb{Q}} [L(X)\phi(X)],$$

371 where we introduced the likelihood ratio $L(x) \triangleq \frac{p(x)}{q(x)}$. Changing the measure proves
 372 to be useful whenever expectations under \mathbb{Q} are easier to compute than under \mathbb{P} .

373 If \mathbb{P} is a probability measure and we have a collection of processes $(X$ and Y), the
 374 measure \mathbb{P}_t is the restriction of \mathbb{P} to all events that can be described in terms of the
 375 behaviour of $\{X_s\}$ and $\{Y_s\}$ for $0 \leq s \leq t$. In other words, \mathbb{P}_t denotes the restriction
 376 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
 377 and \mathbb{Q}_t are equivalent. By using the Radon–Nikodym theorem [22, Theorem 10.6],
 378 we know that a random variable L_t exists, for which

$$379 \quad (5.3) \quad \mathbb{E}_{\mathbb{P}} [\phi(X_t)] = \mathbb{E}_{\mathbb{Q}} [L_t\phi(X_t)],$$

380 where $L_t \triangleq \frac{d\mathbb{P}_t}{d\mathbb{Q}_t}$ is called the Radon–Nikodym derivative of \mathbb{P}_t with respect to \mathbb{Q}_t .

381 As we previously mentioned, we aim at computing the conditional expectation
 382 for some function ϕ . In terms of a reference probability measure \mathbb{Q} , the conditional
 383 expectation is written as

$$384 \quad (5.4) \quad \mathbb{E}_{\mathbb{P}} [\phi(X_t)|\mathcal{F}_t^Y] = \frac{\mathbb{E}_{\mathbb{Q}} [\phi(X_t)L_t | \mathcal{F}_t^Y]}{\mathbb{E}_{\mathbb{Q}} [L_t | \mathcal{F}_t^Y]} =: \frac{\rho_t(\phi)}{\rho_t(1)}.$$

385 Eq. (5.4) is known as a Bayes' formula for stochastic processes or *Kallianpur–
 386 Striebel formula*. Here, we have defined the estimate $\rho_t(\phi) = \mathbb{E}_{\mathbb{Q}} [\phi(X_t)L_t | \mathcal{F}_t^Y]$.

387 The hope is that we can pick a reference measure \mathbb{Q} such that both L_t and q are
 388 simple enough to make $\mathbb{E}_{\mathbb{P}}[\phi(X_t)|\mathcal{F}_t^Y]$ more tractable to compute via Eq. (5.4) than
 389 via an integration over the measure \mathbb{P} . For instance, some simplification might be
 390 achieved by switching from a model with measure \mathbb{P} in which X and Y are coupled,
 391 i.e. statistically dependent, to a model with measure \mathbb{Q} where they are independent
 392 (while preserving the distribution of X_t). For our filtering problem, we will choose a
 393 reference measure \mathbb{Q} such that the path of the observations $\{Y_{0:t}\}$ (or equivalently the
 394 set of the increments $\{dY_{0:t}\}$) becomes independent of the path of the state process
 395 $\{X_{0:t}\}$, i.e. $q(t, X_{0:t}, dY_{0:t}) = p(t, X_{0:t})q(t, dY_{0:t})$.

396 For the time being, we will concentrate not on $\pi_t(\phi)$, but on its unnormalised
 397 counter-part $\rho_t(\phi)$. Our next step is to find an explicit expression for $\rho_t(\phi)$. We start
 398 by finding the expression for L_t , and proceed by applying Itô's rule to the process
 399 $\phi(X_t)L_t$, then we compute the conditional expectation of this expression. Once we
 400 have accomplished this, the remainder is another application of Itô's rule, which gives
 401 us an expression for the normalized quantity $\pi_t(\phi)$.

402 **5.1.2. Filtering for observations corrupted by Gaussian noise.** For the
 403 derivations in this section, we operate under some technical conditions: (i) $k(t)$ is
 404 invertible for all $t \in [0, \infty)$, and $k(t), k^{-1}(t)$ are locally bounded; (ii) the solution to the
 405 SDE in Eq. (3.2) exists and is unique; (iii) h has linear growth. See Subsection SM1.4
 406 for detailed calculation steps.

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

$$411 \quad (5.5) \quad L_t = \frac{d\mathbb{P}_t}{d\mathbb{Q}_t} = \frac{p(t, X_{0:t}, dY_{0:t})}{q(t, X_{0:t}, dY_{0:t})} = \frac{p(t, dY_{0:t}|X_{0:t})p(t, X_{0:t})}{p(t, X_{0:t})q(t, dY_{0:t})} = \frac{p(t, dY_{0:t}|X_{0:t})}{q(t, dY_{0:t})}.$$

412 The observation model in Eq. (4.2) assumes that increments dY_t are independently
 413 distributed given the state X_t , and we know that $dY_t|X_t \sim \mathcal{N}(dY_t; h(X_t)dt, kk^\top(t)dt)$.
 414 Furthermore, we choose $q(t, dY_t)$ as the density of $\mathcal{N}(dY_t; 0, kk^\top(t)dt)$. Thus, the
 415 Radon-Nikodym derivative L_t can be written as

$$416 \quad (5.6) \quad L_t = \exp \left[\int_0^t (k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_0^t \|k^{-1}(s)h(X_s)\|^2 ds \right].$$

417 *Remark 5.1.* An even more general model arises when we consider that the ob-
 418 servation equation is driven by a signal dependent noise; that is, $k(t, X_t)dB_t^y$ instead
 419 of $k(t)dB_t^y$. The filtering problem with this type of observation model can be then
 420 converted into a classical one (signal independent observation noise) via a suitably
 421 chosen stochastic flow mapping. We refer the interested reader to [13].

422 Eq. (5.6), along with Itô's formula, leads to the following proposition:

423 **PROPOSITION 5.2** (The Zakai equation). *Let ϕ be \mathcal{C}^2 (twice continuously differ-
 424 entiable) and suppose that ϕ and all its derivatives are bounded. Then we can write*

$$425 \quad (5.7) \quad \rho_t(\phi) = \rho_0(\phi) + \int_0^t \rho_s(\mathcal{A}^* \phi)ds + \int_0^t \rho_s(k^{-1}(s)h\phi)^\top d\tilde{Y}_s,$$

426 where $\rho_0(\phi) = \mathbb{E}_{\mathbb{Q}}[\phi(X_0) L_0 | \mathcal{F}_0^Y]$, $h\phi = h(x)\phi(x)$, \mathcal{A}^* is the operator in Eq. (3.9),
 427 and the process $\tilde{Y} = (\tilde{Y}_t, t \geq 0)$ is the rescaled process ($\tilde{Y}_t = k^{-1}(t)Y_t, t \geq 0$).

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

$$429 \quad (5.8) \quad dq_t(x) = \mathcal{A}^* q_t(x) dt + q_t(x)(k^{-1}(t)h(x))^\top d\tilde{Y}_t.$$

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

433 PROPOSITION 5.3 (The Kushner-Stratonovich equation). *Let ϕ be \mathcal{C}^2 and suppose that ϕ and all its derivatives are bounded. Then we can write*

$$435 \quad (5.9) \quad \begin{aligned} \pi_t(\phi) &= \pi_0(\phi) + \int_0^t \pi_s(\mathcal{A}^* \phi) ds + \\ &\int_0^t \pi_s(k^{-1}(s)h\phi) - \pi_s(\phi)\pi_s(k^{-1}(s)h)^\top (d\tilde{Y}_s - \pi_s(k^{-1}(s)h) ds), \end{aligned}$$

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

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

$$438 \quad (5.10) \quad dp_t(x) = \mathcal{A}^* p_t(x) dt + p_t(x) (k^{-1}(t)(h\phi - \pi_t(h)))^\top (d\tilde{Y}_t - \pi_t(k^{-1}(t)h) dt).$$

439 We have now obtained a stochastic integral expression for $\pi_t(\phi)$ (the Kushner-
440 Stratonovich equation) and for the evolution of p_t .

441 **5.2. Approximations to the solution of the Kushner-Stratonovich equa-**
442 **tion.** This section contains an overview of two classes of methods used to approximate
443 the solution to the filtering problem. For each class, we give a brief description of
444 the ideas behind the methods and state some related results. Note that so far we
445 have only presented the filter $\pi_t(\phi)$ on an unspecified set in the Borel σ -algebra of
446 $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{D_y})$ of measure one, as well as its assumed density p_t with respect to Lebesgue
447 measure. We have not yet shown what the filtered estimated $\pi_t(\phi)$ would be for a
448 given observation sample path $z \in \mathcal{C}([0, T], \mathbb{R}^{D_y})$. Typically, the observation data is
449 recorded at discrete times, and only these data are made available and used. Con-
450 sidering the practical implementation of a non-linear filter, we present the notion of
451 pathwise filtering as introduced in [11], and studied by many authors, as in [15] and
452 [14].

453 L_t in Eq. (5.6) is defined in terms of a stochastic integral with the observations
454 as the integrator. However, since the integrand is independent of the observations,
455 we can reverse the role of processes X and Y . Instead of using the observations as
456 the integrator, we compute L_t with an integral for every process sample path $\{X_{0:T}\}$,
457 provided that $h(X_t)$ is a semimartingale. Through an integration by parts:

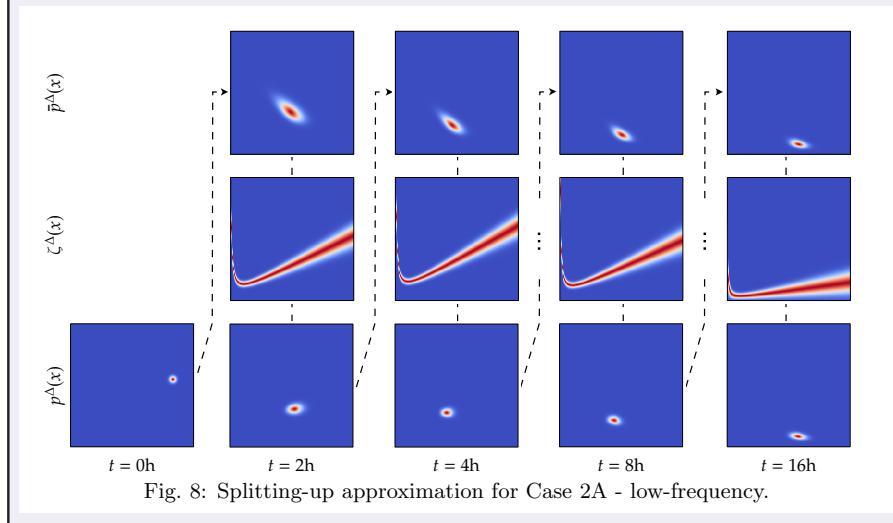
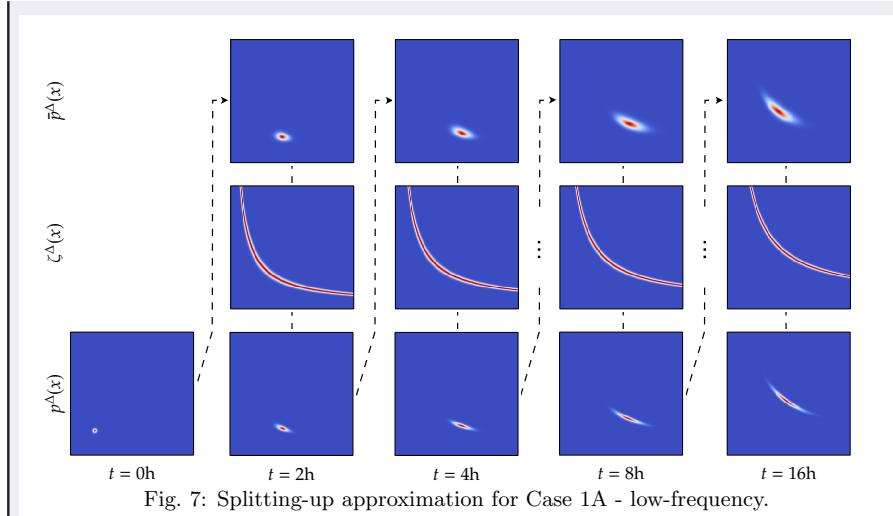
$$458 \quad (5.11) \quad \int_0^T \tilde{h}^\top(X_s) d\tilde{Y}_s = \tilde{h}^\top(X_T) \tilde{Y}_T - \int_0^T \tilde{Y}_s^\top d\tilde{h}(X_s) \quad \text{almost surely.}$$

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

$$461 \quad (5.12) \quad \hat{L}_T = \exp \left[\tilde{h}^\top(X_T) \tilde{Y}_T - \int_0^T \tilde{Y}_s^\top d\tilde{h}(X_s) - \frac{1}{2} \int_0^T \|\tilde{h}(X_s)\|^2 ds \right],$$

462 where \hat{L}_T stands for L_T computed with respect to a fixed observation path. Conse-
463 quently, the SPDEs (5.8) and (5.10) reduce to PDEs with random coefficients.

464 **5.2.1. Finite-differences approach + Splitting-up approximation.** The
 465 following exposition follows closely the description of the method given in [7]. Let
 466 $0 = t_0 < t_1 < \dots < t_n \dots$ be a uniform partition of the interval $[0, \infty)$ with time
 467 step $\Delta = t_n - t_{n-1}$. If the set of observations $\{z_i, 1 \leq i \leq n\}$ is available, the
 468 density p_{t_n} will be approximated by p_n^Δ , where the transition from p_{n-1}^Δ to p_n^Δ follows
 469 a splitting-up approximation [24, 25].



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

473 (5.13)
$$\frac{\partial p_t^n(x)}{\partial t} = \mathcal{A}^* p_t(x), \quad p_{t_{n-1}}^n = p_{n-1}^\Delta,$$

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

475 The second step, called the *correction* step, uses the new observation z_n to update \bar{p}_n^Δ . We use the approximation from pathwise filtering to the Radon-Nikodym
 476 derivative (Eq. (5.12)). We compute $p_n^\Delta(x)$ via $p_n^\Delta(x) \triangleq c_n \zeta_n^\Delta(x) \bar{p}_n^\Delta$, where c_n is a
 477 normalization constant chosen such that $\int_{\mathbb{R}^{D_x}} p_n^\Delta(x) dx = 1$ and
 478

$$479 \quad (5.14) \quad \zeta_n^\Delta(x) \triangleq \exp \left(\tilde{h}^\top(x) \tilde{z}_n - \frac{1}{2} \Delta \|\tilde{h}(x)\|^2 \right),$$

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

481 **5.2.2. Particle methods.** We now switch to particle filtering, a technique used
 482 to approximate the solution to the filtering problem by a finite number of samples
 483 from the conditional distribution π_t . The following exposition is adapted from [2].

484 Let $V_j, j = 1, \dots, N$, be N mutually independent stochastic processes, all in-
 485 dependent of Y , each of them being a solution to Eq. (3.2). The pairs (V_j, Y) are
 486 identically distributed and have the same distribution as the pair (X, Y) under \mathbb{Q} .
 487 Also, let $w_j, j = 1, \dots, N$, be the following exponential martingales

$$488 \quad (5.15) \quad w_j(t) = 1 + \int_0^t w_j(s) (k^{-1}(s) h(V_j(s)))^\top k^{-1}(s) dY_s, \quad t \geq 0,$$

489 that is, for $t \geq 0$

$$490 \quad (5.16) \quad w_j(t) = \exp \left(\int_0^t (k^{-1}(s) h(V_j(s)))^\top k^{-1}(s) dY_s - \frac{1}{2} \int_0^t \|k^{-1}(s) h(V_j(s))\|^2 ds \right).$$

491 As a result, the triples $(V_j, w_j, Y), j = 1, \dots, N$, are identically distributed and have
 492 the same distribution as the triple (X, L, Y) under \mathbb{Q} . Lastly, let $\rho_t^N = \{\rho_t^N, t \geq 0\}$
 493 and $\pi_t^N = \{\pi_t^N, t \geq 0\}$ be the following sequences of measure-valued processes

$$494 \quad (5.17) \quad \rho_t^N \triangleq \frac{1}{N} \sum_{j=1}^N w_j(t) \delta_{V_j(t)}, \quad \pi_t^N \triangleq \frac{\rho_t^N}{\rho_t^N(1)} = \sum_{j=1}^N \bar{w}_j^N(t) \delta_{V_j(t)},$$

495 where the normalised weights $\bar{w}_j(t)$ are given by $\bar{w}_j(t) = \left(\sum_{j'=1}^N w_{j'}(t) \right)^{-1} w_j(t)$, $j' =$
 496 $1, \dots, N$, $t \geq 0$. That is, ρ_t^N is the empirical measure of N (random) particles with
 497 positions $V_j(t)$ and weights $w_j(t), j = 1, \dots, N$, and π_t^N is its normalized version.

498 Under certain conditions, it is possible to show both that (i) $\rho_t^N(\phi)$ converges in
 499 expectation to the unnormalised conditional estimator $\rho_t(\phi)$ with rate of convergence
 500 $N^{-\frac{1}{2}}$ and (ii) $\pi_t^N(\phi)$ converges almost surely with a rate slightly lower than $N^{-\frac{1}{2}}$
 501 [2]. Additionally, $\pi_t^N(\phi)$ converges in expectation to $\pi_t(\phi)$, provided the integrability
 502 condition on $\rho_t^{-1}(1)$. Recall the martingale term L introduced in Eq. (5.5). The
 503 validity of these results requires the existence of higher moments of L . In the scenario
 504 where we cannot assure that the higher moments of the martingale L exist, and we
 505 only know that L is \mathbb{Q} -integrable. In this case, $\pi_t^N(\phi)$ still converges almost surely to
 506 $\pi_t(\phi)$ for a fixed observation path $t \mapsto Y_t$ due to the strong law of large numbers [2].

507 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 (Section SM2).

We analyze 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 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.

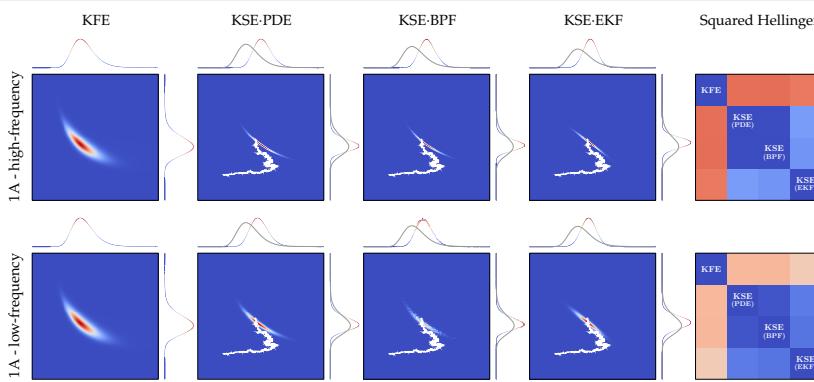


Fig. 9: Case 1A - approximated solution to the Kushner-Stratonovich equation at $t = 16\text{h}$ with the finite differences and the splitting-up approximation for PDEs (KSE-PDE), with the bootstrap particle filter (KSE-BPF) and with the EKF (KSE-EKF). For the PDE method, the discretization in space 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 BPF, we consider $N = 2.5 \times 10^6$. In white, we show the trajectory from which the measurements were simulated. The same trajectory appears in Fig. 2.

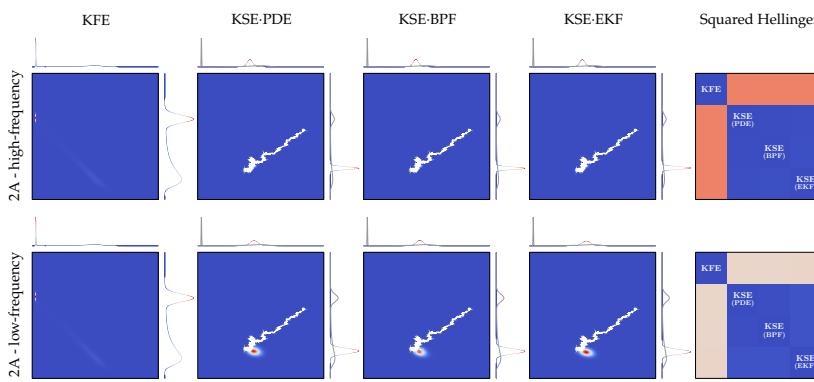


Fig. 10: Case 2A - approximated solution to the Kushner-Stratonovich equation at $t = 8\text{h}$ with the finite differences and the splitting-up approximation for PDEs (KSE-PDE), with the bootstrap particle filter (KSE-BPF) and with the EKF (KSE-EKF). For the PDE method, the discretization in space 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 BPF, $N = 2.5 \times 10^6$. Note how, differently from the solution to the KFE, we move more probability mass towards the washout on account of the measurements.

In Fig. 11, we plot the square of the Hellinger distance between the conditional distribution given by different setups of the method for PDEs and the approximation given by the BPF. It is clear that the discretization error decreases with a finer refinement of the grid. We also include the square of the Hellinger distance between the approximation with the 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 Section SM3 of the supplementary material, where we plot the shape of the conditional distribution for different t 's.

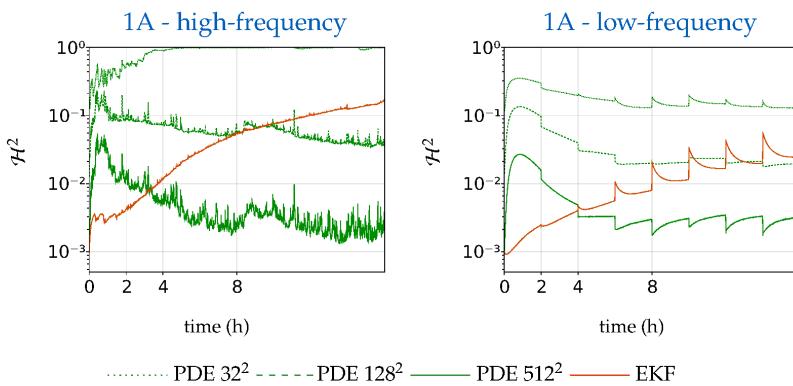


Fig. 11: Case 1A - Squared Hellinger distances between the distributions obtained by approximating the solution to the Kushner-Stratonovich equation with the extended Kalman filter and with methods for partial differential equations using different grid refinements. We consider three scenarios for the 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 $N = 2.5 \times 10^6$.

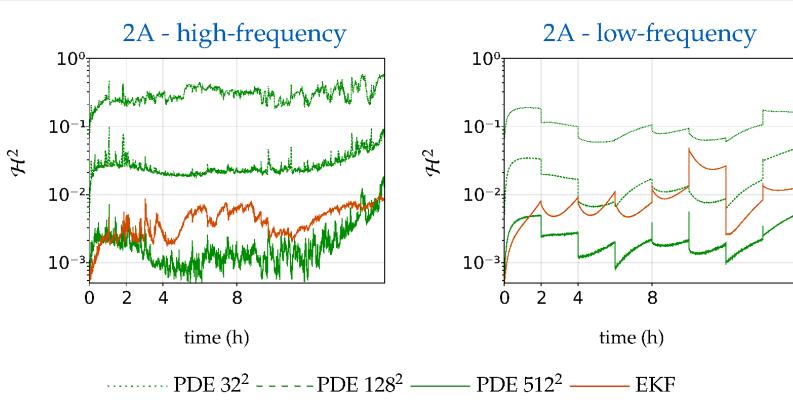


Fig. 12: Case 2A - Squared Hellinger distances between the distributions obtained by approximating the solution to the Kushner-Stratonovich equation with the extended Kalman filter and with methods for partial differential equations using different grid refinements. We consider three scenarios for the 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 con-

ditional distribution given by different setups of the method for PDEs 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, $t > 4h$.

Note that the grid with 512^2 cells fails to give a perfect match between the method for PDEs 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 PDEs or with the EKF 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 BPF, they might need to be tailored to the problem in hand. See [9] and [26] for more methods to prevent sample impoverishment.

510

511 **6. Concluding remarks.** Whether it be an approximation to the Kushner-
512 Stratonovich equation with particle methods or an approximation to the Zakai equa-
513 tion with methods for partial differential equations, no restrictions are imposed on
514 the shape of the distribution. Therefore, these two approximation methods are often
515 more accurate than linearization methods.

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

521 We assumed that measurements of biogas flow rate were available, at either high-
522 or low-frequency. We denoted this setup by scenario A. In the supplementary material,
523 we additionally present results for Cases 1 and 2 under three other scenarios: in B,
524 we are given noisy measurements of the substrate concentration at low-frequency; in
525 C, we are given noisy measurements of the biomass concentration at low-frequency;
526 in D, we are given noisy measurements of both biomass and substrate concentrations
527 at low-frequency. We refer the reader to Section SM3 of the supplementary material
528 for a comparison of the approximations outlined in this work under all four scenarios.

529

REFERENCES

- 530 [1] J. F. ANDREWS, *A mathematical model for the continuous culture of microorganisms utilizing*
531 *inhibitory substrates*, Biotechnology and bioengineering., 10 (1968), <https://doi.org/10.1002/bit.260100602>.
532 [2] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Ap-
533 plied Probability, Springer New York, 2008, <https://doi.org/10.1007/978-0-387-76896-0>.
534 [3] V. E. BENEŠ, *Exact finite-dimensional filters for certain diffusions with nonlinear drift*,
535 *Stochastics*, 5 (1981), pp. 65–92, <https://doi.org/10.1080/17442508108833174>.
536 [4] O. BERNARD, Z. HADJ-SADOK, D. DOCHAIN, A. GENOVESI, AND J.-P. STEYER, *Dynamical*
537 *model development and parameter identification for anaerobic wastewater treatment*
538 *process*, Biotechnology and bioengineering, 75 (2001), <https://doi.org/10.1002/bit.10036>.
539

- 540 [5] G. D. BIRKHOFF, *Proof of the ergodic theorem*, 1931, <https://doi.org/10.1073/pnas.17.2.656>.
541 [6] A. T. BULL, *The renaissance of continuous culture in the post-genomics age*, Journal
542 of Industrial Microbiology and Biotechnology, 37 (2010), [https://doi.org/10.1007/
543 s10295-010-0816-4](https://doi.org/10.1007/s10295-010-0816-4).
544 [7] Z. CAI, F. L. GLAND, AND H. ZHANG, *An adaptive local grid refinement method for nonlinear fil-
545 tering*. Research report 2679, INRIA, 1995, <https://hal.inria.fr/inria-00074012/document>.
546 [8] F. CAMPILLO, M. JOANNIDES, AND I. LARRAMENDY-VALVERDE, *Approximation of the
547 fokker-planck equation of the stochastic chemostat*, Mathematics and Computers in Sim-
548 ulation, 99 (2014), pp. 37–53, <https://doi.org/10.1016/j.matcom.2013.04.012>.
549 [9] J. CARPENTER, P. CLIFFORDY, AND P. FEARNHEAD, *An improved particle filter for nonlinear
550 problems*, IEE Proc. Radar, Sonar Navig., 146 (1999), pp. 2–7, [https://doi.org/10.1049/
551 ip-rsn:19990255](https://doi.org/10.1049/ip-rsn:19990255).
552 [10] K. CHUNG AND R. WILLIAMS, *Introduction to Stochastic Integration*, Modern Birkhäuser Clas-
553 sics, Springer New York, 2014.
554 [11] J. M. C. CLARK, *The design of robust approximations to the stochastic differential equations
555 of nonlinear filtering*, 1978, https://doi.org/10.1007/978-94-011-7577-7_41.
556 [12] S. COHEN AND R. ELLIOTT, *Stochastic Calculus and Applications*, Probability and Its Appli-
557 cations, Springer New York, 2015, <https://doi.org/10.1007/978-1-4939-2867-5>.
558 [13] D. CRISAN, M. KOURITZIN, AND J. XIONG, *Nonlinear filtering with signal dependent obser-
559 vation noise*, Electronic Journal of Probability, 14 (2009), <https://doi.org/10.1214/EJP.v14-687>.
560 [14] D. CRISAN, T. G. KURTZ, AND S. ORTIZ-LATORRE, *Particle representation for the solution of
561 the filtering problem. application to the error expansion of filtering discretizations*, Journal
562 of Stochastic Analysis, 2 (2021), <https://doi.org/10.31390/josa.2.3.15>.
563 [15] M. H. A. DAVIS, *A pathwise solution of the equations of nonlinear filtering*, Theory of Probab-
564 ity & Its Applications, 27 (1982), <https://doi.org/10.1137/1127017>.
565 [16] O. HADJ-ABDELKADEER AND A. HADJ-ABDELKADER, *Estimation of substrate and biomass con-
566 centrations in a chemostat using an extended kalman filter*, International Journal of Bioau-
567 tomation, 23 (2019), pp. 215–232, <https://doi.org/10.7546/ijba.2019.23.2.000551>.
568 [17] J. HARMAND, C. LOBRY, A. RAPAPORT, AND T. SARI, *The Chemostat: Mathematical Theory
569 of Microorganism Cultures*, John Wiley and Sons, ISTE, 2017, <https://doi.org/10.1002/9781119437215>.
570 [18] P. A. HOSKISSON AND G. HOBBS, *Continuous culture-making a comeback?*, Microbiology, 151
571 (2005), <https://doi.org/10.1099/mic.0.27924-0>.
572 [19] L. IMHOF AND S. WALCHER, *Exclusion and persistence in deterministic and stochastic chemo-
573 stat models*, Journal of Differential Equations, 217 (2005), <https://doi.org/10.1016/j.jde.2005.06.017>.
574 [20] K. ITO, *Stochastic integral*, Proceedings of the Imperial Academy, 20 (1944), <https://doi.org/10.3792/pia/1195572786>.
575 [21] R. E. KÁLMÁN AND R. S. BUCY, *New results in linear filtering and prediction theory*, Journal
576 of Basic Engineering, 83 (1961), <https://doi.org/10.1115/1.3658902>.
577 [22] F. KLEBANER, *Introduction to Stochastic Calculus with Applications*, Introduction to Stochastic
578 Calculus with Applications, Imperial College Press, 2005.
579 [23] A. KUTSCHIREITER, S. C. SURACE, AND J.-P. PFISTER, *The hitchhiker’s guide to nonlinear
580 filtering*, Journal of Mathematical Psychology, 94 (2020), <https://doi.org/10.1016/j.jmp.2019.102307>.
581 [24] F. LE GLAND, *Time discretization of nonlinear filtering equations*, Dec. 1989, <https://doi.org/10.1109/CDC.1989.70650>.
582 [25] F. LEGLAND, *Splitting-up approximation for SPDE’s and SDE’s with application to nonlinear
583 filtering*, 1992, <https://doi.org/10.1007/BFb0007332>.
584 [26] T. LI, S. SUN, T. P. SATTAR, AND J. M. CORCHADO, *Fight sample degeneracy and impov-
585 erishment in particle filters: A review of intelligent approaches*, Expert Systems with
586 Applications, 41 (2014), <https://doi.org/10.1016/j.eswa.2013.12.031>.
587 [27] X. MAO, *Stochastic differential equations and applications*, 2nd ed., 2007, <https://doi.org/10.1533/9780857099402>.
588 [28] J. MONOD, *La technique de culture continue: theorie et applications*, 1950, <https://doi.org/10.1016/B978-0-12-460482-7.50023-3>.
589 [29] I. NOBUYUKI AND S. WATANABE, *Stochastic Differential Equations and Diffusion Processes*,
590 North-Holland Pub. Co.; Kodansha; Sole Distributors for the U.S.A. and Canada, 1981,
591 [https://doi.org/10.1016/S0924-6509\(08\)70217-4](https://doi.org/10.1016/S0924-6509(08)70217-4).
592 [30] P. RENARD, D. DOCHAIN, G. BASTIN, H. P. NAVEAU, AND J. NYNS, *Adaptive control of anaer-
593 obic digestion processes—a pilot-scale application*, Biotechnology and Bioengineering, 31
594 (1988), <https://doi.org/10.1002/bit.260310402>.
595 [31] P. RENARD, D. DOCHAIN, G. BASTIN, H. P. NAVEAU, AND J. NYNS, *Adaptive control of anaer-
596 obic digestion processes—a pilot-scale application*, Biotechnology and Bioengineering, 31
597 (1988), <https://doi.org/10.1002/bit.260310402>.
598 [32] P. RENARD, D. DOCHAIN, G. BASTIN, H. P. NAVEAU, AND J. NYNS, *Adaptive control of anaer-
599 obic digestion processes—a pilot-scale application*, Biotechnology and Bioengineering, 31
600 (1988), <https://doi.org/10.1002/bit.260310402>.

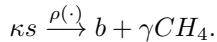
- 602 [31] P. SINNER, M. STIEGLER, O. GOLDBECK, G. M. SEIBOLD, C. HERWIG, AND J. KAGER, *Online*
603 *estimation of changing metabolic capacities in continuous Corynebacterium glutamicum*
604 *cultivations growing on a complex sugar mixture*, Biotechnology and Bioengineering, 119
605 (2021), pp. 575–590, <https://doi.org/10.1002/bit.28001>.
- 606 [32] R. VAN HANDEL, *Filtering, stability, and robustness*, PhD thesis, 2007, <https://doi.org/10.7907/4p53-1h42>.
- 608 [33] L. WANG AND D. JIANG, *A note on the stationary distribution of the stochastic chemostat*
609 *model with general response functions*, Applied Mathematics Letters, 73 (2017), pp. 22–
610 28, <https://doi.org/10.1016/j.aml.2017.04.029>.
- 611 [34] A. K. ZVONKIN, *A transformation of the phase space of a diffusion process that removes the*
612 *drift*, Mathematics of the USSR-Sbornik, 22 (1974), pp. 129–149, <https://doi.org/10.1070/sm1974v02n01abeh001689>.

1 **SUPPLEMENTARY MATERIALS: APPROXIMATIONS TO THE
2 SOLUTION OF THE KUSHNER-STRATONOVICH EQUATION FOR
3 THE STOCHASTIC CHEMOSTAT***

4 JOSÉ AUGUSTO FONTENELE MAGALHÃES[†], MUHAMMAD FUADY EMZIR[‡], AND
5 FRANCESCO CORONA[§]

6 **SM1. Mathematical perspective of stochastic filtering for the chemo-
7 stat: proofs.**

8 **SM1.1. Chemostat equations.** We begin by describing the dynamics in a
9 chemostat. Let b denote the concentration of biomass and s the concentration of
10 substrate. Consider the following reaction to produce one unit of biomass and γ units
11 of methane (biogas) from κ units of substrate, with reaction rate $\rho(\cdot)$:



13 Since biomass here is not a chemically defined compound, we make use of the so-
14 called yield coefficients κ and γ under the assumptions that the biomass composition
15 is constant and that the ratio surface/volume for microbial cells is constant at a
16 population level [SM11]. These are mild assumptions when one is interested not in
17 the full details of intracellular reactions but, rather, in the macroscopic behaviour of
18 the system.

19 Let us now assume that $\rho(\cdot) = \mu(\cdot)b$, where μ is called “specific growth velocity”.
20 With this assumption, we guarantee that the reaction rate is zero in the absence of
21 biomass. At a specific time t , let $F_{in}(t)$ express how much media flows into the vessel
22 per hour. The incoming dilution rate is calculated by dividing the incoming flow rate
23 by the culture volume $V(t)$, that is $D_{in}(t) = F_{in}(t)/V(t)$. Analogously, $F_{out}(t)$ and
24 $D_{out}(t)$ denote the outgoing flow rate and its corresponding dilution rate, respectively.

25 To establish the equations of the chemostat, we apply a mass balance in which, for
26 a period of time dt , the variation in the mass of an element is the net result between
27 four possible terms: the quantity of that element that has been brought into the
28 system, the produced quantity in the vessel, the consumed quantity inside the vessel
29 and the extracted quantity. We start with the equation for the change in volume:

$$30 \quad \frac{dV(t)}{dt} = F_{in}(t) - F_{out}(t),$$

31 and we then apply the chain rule on the derivative of the quantity of substrate with
32 respect to time,

$$33 \quad (\text{SM1.1}) \quad \frac{d(s(t)V(t))}{dt} = s(t)(F_{in}(t) - F_{out}(t)) + \frac{ds(t)}{dt}V(t).$$

34 The variation in the substrate mass, that is $\frac{d(s(t)V(t))}{dt}$, is the balance between the
35 four possible terms outlined above.

*Preprint April 6, 2023.

[†]Department of Chemical and Metallurgical Engineering, Aalto University, Espoo, FI (augusto.magalhaes@aalto.fi).

[‡]Control and Instrumentation Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, SA (muhammad.emzir@kfupm.edu.sa).

[§]Department of Chemical and Metallurgical Engineering, Aalto University, Espoo, FI (francesco.corona@aalto.fi)

36 After rearranging the terms in Eq. (SM1.1), we get

$$\begin{aligned} 37 \quad V(t) \frac{ds(t)}{dt} &= \frac{d(s(t)V(t))}{dt} - s(t)(F_{in}(t) - F_{out}(t)) \\ 38 \quad &= [s_{in}(t)F_{in}(t) - s(t)F_{out}(t) - \kappa V(t)\rho(\cdot)] - s(t)(F_{in}(t) - F_{out}(t)) \\ 39 \quad &= F_{in}(t)(s_{in}(t) - s(t)) - \kappa V(t)\rho(\cdot). \end{aligned}$$

40 Analogously, we apply the chain rule for the biomass with respect to time:

$$\begin{aligned} 41 \quad V(t) \frac{db(t)}{dt} &= \frac{d(b(t)V(t))}{dt} - b(t)(F_{in}(t) - F_{out}(t)) \\ 42 \quad &= [b_{in}(t)F_{in}(t) - b(t)F_{out}(t) + V(t)\rho(\cdot)] - b(t)(F_{in}(t) - F_{out}(t)) \\ 43 \quad &= V(t)\rho(\cdot) - F_{in}(t)(b(t) - b_{in}(t)). \end{aligned}$$

44 We are now left with the following balance equations:

$$\begin{aligned} 45 \quad \frac{dV(t)}{dt} &= F_{in}(t) - F_{out}(t), \\ 46 \quad \frac{db(t)}{dt} &= \rho(\cdot) - D_{in}(t)(b(t) - b_{in}(t)), \\ 47 \quad \frac{ds(t)}{dt} &= D_{in}(s_{in}(t) - s(t)) - \kappa\rho(\cdot). \end{aligned}$$

48 Here, the focus is on bioreactor processes performed at continuous mode, where
49 $F_{in}(t) = F_{out}(t) = F(t)$, $V(t) = V$ and $D_{in}(t) = D_{out}(t) = D(t)$. Under the previous
50 assumption $\rho(\cdot) = \mu(\cdot)b$, we arrive at the ordinary differential model for the growth
51 of a single species in a chemostat:

$$\begin{aligned} 52 \quad \frac{db(t)}{dt} &= (\mu(\cdot) - D(t))b(t) + D(t)b_{in}(t), \\ 53 \quad \frac{ds(t)}{dt} &= D(t)(s_{in}(t) - s(t)) - \kappa\mu(\cdot)b(t), \end{aligned}$$

54 with initial concentrations b_0 and s_0 .

55 **SM1.2. Stochastic differential equations (SDEs).** Let us consider the fol-
56 lowing stochastic differential equation

57 (SM1.2a) $dX_t = f(X_t)dt + g(X_t)dB_t^x, \quad X_0 = \xi,$

58 or equivalently in coordinate form

(SM1.2b)

$$59 \quad dX_t^{d_x} = f_{d_x}(X_t)dt + \sum_{m=1}^M g_{d_x m}(X_t)dB_t^{x,m}, \quad X_0^{d_x} = \xi^{d_x}, \quad \text{for } d_x = 1, \dots, D_x,$$

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

$$\begin{aligned} 62 \quad X : [0, \infty) \times \Omega &\rightarrow \mathbb{R}^{D_x}, \\ 63 \quad f : \mathbb{R}^{D_x} &\rightarrow \mathbb{R}^{D_x}, \\ 64 \quad g : \mathbb{R}^{D_x} &\rightarrow \mathbb{R}^{D_x} \times \mathbb{R}^M. \end{aligned}$$

65 DEFINITION SM1.1 (Itô's formula). Let $\psi : [0, \infty) \times \mathbb{R}^{d_x} \rightarrow \mathbb{R}$ be a function that
 66 is differentiable in the first argument and twice-differentiable in the second argument.
 67 Suppose that $X = (X_t, t \geq 0)$ is a solution to the SDE (SM1.2), then

$$68 \quad (\text{SM1.3a}) \quad d\psi(t, X_t) = \left[\frac{\partial \psi}{\partial t}(t, X_t) + (\nabla_X \psi)^\top f(X_t) + \frac{1}{2} \text{Trace}(\Gamma(X_t) H_X \psi) \right] dt \\ + (\nabla_X \psi)^\top g(X_t) dB_t^x,$$

69 or equivalently in integral form

$$70 \quad (\text{SM1.3b}) \quad \psi(t, X_t) = \psi(0, X_0) + \int_0^t \left[\frac{\partial \psi}{\partial t}(s, X_s) + (\nabla_X \psi)^\top f(X_s) + \frac{1}{2} \text{Trace}(\Gamma(X_s) H_X \psi) \right] ds \\ + \int_0^t (\nabla_X \psi)^\top g(X_s) dB_s^x.$$

71 where $\Gamma(X_s) \triangleq gg^\top(X_s)$, and ∇_X and H_X denote the gradient and Hessian operators,
 72 respectively.

73 EXAMPLE SM1.1 (No global solution).

$$74 \quad (\text{SM1.4}) \quad dX_t = X_t^3 dt + X_t^2 dB_t^x,$$

75 with initial condition $X_0 \equiv 1$. Suppose that $X = (X_t, t \geq 0)$ is a solution to
 76 Eq. (SM1.4). We can find such X by applying Itô's formula (SM1.3) on the process
 77 $\psi(t, X_t) = \frac{1}{X_t}$, yielding $\frac{1}{X_t} = C - B_t^x$, with C a constant. Since $X_0 \equiv 1$ and $B_0^x = 0$,

$$78 \quad X_t = \frac{1}{1 - B_t^x}.$$

79 Note how the solution X grows when the process B^x is close to 1. A solution exists
 80 as long as B^x remains in the interval $(-\infty, 1)$, but no global solution exists.

81 EXAMPLE SM1.2 (No unique solution).

$$82 \quad (\text{SM1.5}) \quad dX_t = 3X_t^{1/3} dt + 3X_t^{2/3} dB_t^x,$$

83 with initial condition $X_0 \equiv 0$. Define a function $\eta_a(v) = (v-a)^3 \mathbb{1}_{\{v>a\}}$, with constant
 84 $a > 0$ fixed. One can show that for any $a > 0$, the SDE has the solution $\eta_a(B^x)$. To
 85 show this, we apply Itô's formula (SM1.3) on the process $\psi(t, X_t) = \eta_a(B_t^x)$. Since
 86 $B_0^x = 0$ by definition, we get that

$$87 \quad d\eta_a(B_t^x) = 3\eta_a(B_t^x)^{1/3} dt + 3\eta_a(B_t^x)^{2/3} dB_t^x,$$

88 which is exactly in the form of the SDE (SM1.5). Therefore, under the same initial
 89 condition, one finds infinitely many solutions.

90 Eqs. (SM1.4) and (SM1.5) have a property in common: their respective coefficients
 91 f and g do not satisfy the conditions for existence and uniqueness of solutions.

92 **SM1.3. Existence, uniqueness and positivity of the solution to SDEs
 93 for the chemostat.** For a given initial condition $X_0 \in (\mathbb{R}^2)_+^*$, there is a unique
 94 solution $X = (X_t, t \geq 0)$ of

$$95 \quad (\text{SM1.6}) \quad d \begin{bmatrix} b_t \\ s_t \end{bmatrix} = \underbrace{\begin{bmatrix} (\mu(s_t) - D)b_t \\ D(s_{in} - s_t) - \kappa\mu(s_t)b_t \end{bmatrix}}_{f(X_t)} dt + \underbrace{\begin{bmatrix} \omega_b b_t & 0 \\ 0 & \omega_s s_t \end{bmatrix}}_{g(X_t)} d \begin{bmatrix} B_t^b \\ B_t^s \end{bmatrix},$$

96 with growth functions $\mu(s) = \frac{\mu_{max}s}{K_s+s}$ or $\mu(s) = \frac{\mu_0 s}{K_s+s+I_s s^2}$. Moreover, the solution
 97 remains in $(\mathbb{R}^2)_+^*$ with probability one, i.e. $X_t \in (\mathbb{R}^2)_+^*$ for $t \geq 0$ almost surely.

98 CONDITION SM1.1 (Local Lipschitz continuous). *f(x)* and *g(x)* are local Lip-
 99 schitz continuous in *x*, that is, for all $C > 0$, there exist constants $K_C^f > 0$ and
 100 $K_C^g > 0$ such that

$$101 \quad \|f(x) - f(x')\| \leq K_C^f \|x - x'\|, \|g(x) - g(x')\| \leq K_C^g \|x - x'\|, \text{ for all } \|x\|, \|x'\| \leq C,$$

102 where the matrix norm is the Hilbert–Schmidt and the vector norm is the Euclidean.

103 Under condition SM1.1, there exists a unique solution $X = (X_t, t \geq 0)$ to the
 104 state equation (SM1.6), on $t \in [0, \tau_e]$, where τ_e is the explosion time, namely when
 105 X_t leaves $(\mathbb{R}^2)_+^*$. See, for example, [SM7] or [SM2].

106 To show that this solution is global, we need to show that $\tau_e = \infty$ almost surely.
 107 Let m_0 be a sufficient large integer such that $\frac{1}{m_0} < X_0^{d_x} < m_0$, $d_x \in \{1, 2\}$. For each
 108 integer $m \geq m_0$, define the stopping time

$$109 \quad \tau_m \stackrel{\Delta}{=} \inf\{t \in [0, \tau_e) : X_t^{d_x} \notin (1/m, m)\},$$

110 where throughout this derivation we set $\inf \emptyset = \infty$, where \emptyset denotes the empty set.
 111 Let $\tau_\infty \stackrel{\Delta}{=} \lim_{m \rightarrow \infty} \tau_m$, we then have $\tau_\infty \leq \tau_e$ almost surely. The idea is to show that
 112 no explosion occurs, and consequently that $X_t \in (\mathbb{R}^2)_+^*$ for all $t \geq 0$. To proof this, we
 113 first assume that the opposite result is true. We show that such assumption results
 114 in a contradictory conclusion. The same approach can be found in [SM12] and [SM5].

115 Assume that X will explode almost surely. Under this assumption, the random
 116 variable τ_e and, consequently, the sequence of random variables τ_m are bounded in
 117 probability; for any $\varepsilon \geq 0$, there exists a constant $T(\varepsilon)$ and a value m' such that

$$118 \quad (\text{SM1.7}) \quad \mathbb{P}(|\tau_m| \leq T) \geq 1 - \varepsilon, \quad \text{for all } m \geq m'.$$

119 In accordance with the Khasminskii test for verifying non-explosion on solutions
 120 of SDEs, we construct a suitable function to prove the existence of a global solution.
 121 Define a \mathcal{C}^2 -function $V : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ by

$$122 \quad V(x) = \kappa(x_1 - 1 - \log x_1) + x_2 - C_2 \left(1 + \log \frac{x_2}{C_2} \right),$$

123 with $x = (x_1, x_2)$ and constant C_2 . The non-negativity of this function comes from the
 124 inequality $x \geq 1 + \log(x)$. Under the initial condition $X_0 \equiv x_0$, and with $V(x_0) < \infty$,
 125 we apply Itô's formula (SM1.3) to the process $\psi(t, X_t) = V(X_t)$, $t \geq 0$:

$$\begin{aligned}
126 \quad dV(X_t) &= \underbrace{\left[\frac{\kappa b_t - \kappa}{b_t} \quad \frac{s_t - C_2}{s_t} \right] \left[\begin{array}{c} (\mu(s_t) - D)b_t \\ D(s_{in} - s_t) - \kappa\mu(s_t)b_t \end{array} \right]}_{(\nabla_x V)^\top f(X_t)} dt \\
&\quad + \frac{1}{2} \text{Trace} \left(\underbrace{\left[\begin{array}{cc} \omega_b^2 b_t^2 & 0 \\ 0 & \omega_s^2 s_t^2 \end{array} \right] \left[\begin{array}{cc} \frac{\kappa}{b_t^2} & 0 \\ 0 & \frac{C_2}{s_t^2} \end{array} \right]}_{\Gamma(X_t) H_X V} \right) dt \\
&\quad + \underbrace{\left[\frac{\kappa b_t - \kappa}{b_t} \quad \frac{s_t - C_2}{s_t} \right] \left[\begin{array}{cc} \omega_b b_t & 0 \\ 0 & \omega_s s_t \end{array} \right]}_{(\nabla_x V)^\top g(X_t)} d \underbrace{\left[\begin{array}{c} B_t^b \\ B_t^s \end{array} \right]}_{B_t^x} \\
&= \left((\kappa b_t - \kappa) \left(\frac{\mu_0 s_t}{K_s + s_t + I_s s_t^2} - D \right) \right. \\
&\quad \left. + \frac{s_t - C_2}{s_t} \left((D s_{in} - D s_t) - \kappa \frac{\mu_0 s_t b_t}{K_s + s_t + I_s s_t^2} \right) + \frac{1}{2} (\kappa w_b^2 + C_2 w_s^2) \right) dt \\
&\quad + w_b (\kappa b_t - \kappa) dB_t^b + w_s (s_t - C_2) dB_t^s \\
&\leq \left(\kappa (\mu_0 b_t - D b_t + D) + D s_{in} + C_2 D + \kappa C_2 \frac{\mu_0}{K_s} b_t + \frac{1}{2} (\kappa w_b^2 + C_2 w_s^2) \right) dt \\
&\quad + w_b \kappa (b_t - 1) dB_t^b + w_s (s_t - C_2) dB_t^s.
\end{aligned}$$

127 Let $C_2 \triangleq \frac{K_s(D-\mu_0)}{\mu_0}$, then the terms dependent on b_t cancel out. Therefore,

$$\begin{aligned}
128 \quad dV(X_t) &\leq (D(\kappa + s_{in} + C_2) + \frac{1}{2} (\kappa w_b^2 + C_2 w_s^2)) dt + w_b (b_t - \kappa) dB_t^b + w_s (s_t - C_2) dB_t^s \\
129 \quad (\text{SM1.8}) &\leq \tilde{C}_2 dt + w_b (b_t - \kappa) dB_t^b + w_s (s_t - C_2) dB_t^s,
\end{aligned}$$

130 where $\tilde{C}_2 > 0$ is a constant.

131 Since for any $t \geq 0$, $X_{t \wedge \tau_m} \in \mathbb{R}_+^2$, we can integrate both sides of Eq. (SM1.8):

$$132 \quad \int_0^{\tau_m \wedge T} dV(X_t) \leq \int_0^{\tau_m \wedge T} \tilde{C}_2 dt + \int_0^{\tau_m \wedge T} (w_b (b_t - \kappa) dB_t^b + w_s (s_t - C_2) dB_t^s).$$

133 Moreover, we can take expectations on both sides, which leads to

$$\begin{aligned}
134 \quad \mathbb{E}_X[V(X_{\tau_m \wedge T})] &\leq V(x_0) + \tilde{C}_2 \mathbb{E}_X[\tau_m \wedge T] \\
135 \quad (\text{SM1.9}) \quad &= V(x_0) + \tilde{C}_2 (\tau_m \wedge T).
\end{aligned}$$

136 Define the set $\Omega_m = \{\tau_m \leq T\}, \forall m > m'$. Then, $\mathbb{P}(\Omega_m) \geq 1 - \varepsilon$. Note that for
137 every $\omega \in \Omega_m$, there is some d_x such that $X_{\tau_m}^{d_x}(\omega)$ equals either m or $1/m$, and hence

$$V(X_{\tau_m}(\omega)) \geq (\kappa(m - 1 - \log m)) \wedge (m - C_2 - C_2 \log \frac{m}{C_2})$$

138 or

$$139 \quad V(X_{\tau_m}(\omega)) \geq \left(\kappa \left(\frac{1}{m} - 1 - \log \frac{1}{m} \right) \right) \wedge \left(\frac{1}{m} - C_2 - C_2 \log \frac{1/m}{C_2} \right).$$

140 Combining both inequalities,

$$\begin{aligned} 141 \quad V(X_{\tau_m}(\omega)) &\geq \left[(\kappa(m - 1 - \log m)) \wedge (m - C_2 - C_2 \log \frac{m}{C_2}) \right] \wedge \\ 142 \quad &\left[\left(\kappa \left(\frac{1}{m} - 1 + \log m \right) \right) \wedge \left(\frac{1}{m} - C_2 + C_2 \log (mC_2) \right) \right]. \\ 143 \end{aligned}$$

144 Under the bounded in probability property (SM1.7), the inequality (SM1.9) can
145 be further developed into

$$\begin{aligned} 146 \quad V(x_0) + \tilde{C}T &\geq \mathbb{E}_X[V(X_{\tau_m \wedge T})] \\ &\geq \mathbb{E}_X[\mathbb{1}_{\Omega_m}(\omega)V(X_{\tau_m})] \\ &\geq \mathbb{P}(\Omega_m) \left[(\kappa(m - 1 - \log m)) \wedge (m - C_2 - C_2 \log \frac{m}{C_2}) \right] \wedge \\ &\quad \left[\left(\kappa \left(\frac{1}{m} - 1 + \log m \right) \right) \wedge \left(\frac{1}{m} - C_2 + C_2 \log (mC_2) \right) \right] \\ &\geq (1 - \varepsilon) \left[(\kappa(m - 1 - \log m)) \wedge (m - C_2 - C_2 \log \frac{m}{C_2}) \right] \wedge \\ &\quad \left[\left(\kappa \left(\frac{1}{m} - 1 + \log m \right) \right) \wedge \left(\frac{1}{m} - C_2 + C_2 \log (mC_2) \right) \right], \end{aligned}$$

147 where the second and third inequalities result from the Markov's inequality.

148 We are interested in the limiting behavior of τ_m when m goes to infinity. Taking
149 the limit on both sides of Eq. (SM1.3) results in

$$\begin{aligned} 150 \quad \lim_{m \rightarrow \infty} (V(x_0) + \tilde{C}_2 T) &\geq \lim_{m \rightarrow \infty} (1 - \varepsilon) \left[(\kappa(m - 1 - \log m)) \wedge \left(m - C_2 - C_2 \log \frac{m}{C} \right) \right] \wedge \\ 151 \quad &\left[\left(\kappa \left(\frac{1}{m} - 1 + \log m \right) \right) \wedge \left(\frac{1}{m} - C_2 + C_2 \log \frac{m}{C_2} \right) \right]. \end{aligned}$$

152 That is, $V(x_0) + \tilde{C}_2 T \geq \infty$, which is a contradiction, since $0 < V(x_0) < \infty$ and
153 $0 \leq \tilde{C}_2 T < \infty$. Therefore, $\tau_m \rightarrow \infty$ almost surely.

154 What we have shown is that for locally Lipschitz coefficients the solution is unique
155 up to τ_e , the exit time from $(\mathbb{R}^2)_+^*$. With $\tau_m \rightarrow \infty$, we consequently have $\tau_e = \infty$,
156 and we get a unique solution remaining in $(\mathbb{R}^2)_+^*$ for all time t .

157 **SM1.4. The Kushner-Stratonovich equation.**

158 PROPOSITION SM1.2 (SDE for the Radon-Nikodym derivative). *The derivative*
159 L_t *can be written as*

$$160 \quad L_t = \exp \left[\int_0^t (k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_0^t \|k^{-1}(s)h(X_s)\|^2 ds \right],$$

161 and it evolves according to the following SDE:

$$162 \quad dL_t = L_t \tilde{h}^\top(X_t) d\tilde{Y}_t,$$

163 where $\tilde{Y} = (\tilde{Y}_t, t \geq 0)$ is a rescaled observation process (i.e. $\tilde{Y}_t = k^{-1}(t)Y_t$, $t \geq 0$),
164 and \tilde{h} is the rescaled observation function $\tilde{h}(X_t) = k^{-1}(t)h(X_t)$.

165 *Proof.* Let $q(t, dY_t)$ be the density of $\mathcal{N}(dY_t; 0, kk^\top(t)dt)$. The Radon-Nikodym

166 derivative L_t can be written as □

$$\begin{aligned}
 L_t &= \frac{p(t, dY_{0:t} | X_{0:t})}{q(t, dY_{0:t})} \\
 &= \frac{\prod_{s=0}^t p(t, dY_s | X_s)}{\prod_{s=0}^t q(t, dY_s)} \\
 &= \prod_{s=0}^t \frac{\text{density of } \mathcal{N}(dY_s; h(X_s)ds, kk^\top(s))ds}{\text{density of } \mathcal{N}(dY_s; 0, kk^\top(s))ds} \\
 &= \prod_{s=0}^t \exp \left[(k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2} \|k^{-1}(s)h(X_s)\|^2 ds \right] \\
 &\stackrel{\lim_{dt \rightarrow 0}}{=} \exp \left[\int_0^t (k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_0^t \|k^{-1}(s)h(X_s)\|^2 ds \right],
 \end{aligned}$$

168 where in the last step we took the continuum limit $\lim_{dt \rightarrow 0}$.

169 Let

$$\begin{aligned}
 170 \quad \Lambda_t &\triangleq \log L_t = \left[\int_0^t (k^{-1}(s)h(X_s))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_0^t \|k^{-1}(s)h(X_s)\|^2 ds \right], \\
 171 \quad d\Lambda_t &= \left[(k^{-1}(t)h(X_t))^\top k^{-1}(t)dY_t - \frac{1}{2} \|k^{-1}(t)h(X_t)\|^2 dt \right].
 \end{aligned}$$

172 Then, exploit the fact that $dL_t = d(\exp \Lambda_t)$ and apply Itô's formula (SM1.3) to
173 the process $\exp \Lambda$:

$$\begin{aligned}
 174 \quad dL_t &= d(\exp \Lambda_t) \\
 175 &= \exp \Lambda_t d\Lambda_t + \frac{1}{2} \exp \Lambda_t (d\Lambda_t)^2 \\
 176 &= \exp \Lambda_t d\Lambda_t + \frac{1}{2} \exp \Lambda_t \|k^{-1}(t)h(X_t)\|^2 dt \\
 177 &= L_t \left[(k^{-1}(t)h(X_t))^\top k^{-1}(t)dY_t - \frac{1}{2} \|k^{-1}(t)h(X_t)\|^2 dt \right] + \frac{1}{2} L_t \|k^{-1}(t)h(X_t)\|^2 dt \\
 178 &= L_t \tilde{h}^\top(X_t) d\tilde{Y}_t.
 \end{aligned}$$

179 PROPOSITION SM1.3 (The Kushner-Stratonovich equation - Proposition 5.3 re-
180 visited). Let ϕ be C^2 and suppose that ϕ and all its derivatives are bounded. Then we
181 can write

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

185 Here, $\pi_0(\phi) = \mathbb{E}_{\mathbb{P}} [\phi(X_0) | \mathcal{F}_0^Y] = \mathbb{E}_{\mathbb{P}} [\phi(X_0)]$, and the adjoint operator \mathcal{A}^* is
(SM1.10)

$$186 \quad \mathcal{A}^* p(t, x) = - \sum_{d_x=1}^{D_x} \frac{\partial}{\partial x_{d_x}} [f_{d_x}(x)p(t, x)] + \sum_{d_x=1}^{D_x} \sum_{d'_x=1}^{D_x} \frac{\partial^2}{\partial x_{d_x} \partial x_{d'_x}} [G_{d_x d'_x}(x)p(t, x)],$$

187 where

$$188 \quad G_{d_x d'_x}(t, x) = \frac{1}{2} \sum_{m=1}^M g_{d_x m}(x) g_{d'_x m}(x).$$

189 *Proof.* Our problem is to compute a conditional expectation (i.e. an expected
190 value under the posterior distribution π_t) for some function ϕ . The conditional ex-
191 pectation can be rewritten in terms of a reference probability measure \mathbb{Q} :

$$192 \quad (\text{SM1.11}) \quad \mathbb{E}_{\mathbb{P}} [\phi(X_t) | \mathcal{F}_t^Y] = \frac{\mathbb{E}_{\mathbb{Q}} [\phi(X_t) L_t | \mathcal{F}_t^Y]}{\mathbb{E}_{\mathbb{Q}} [L_t | \mathcal{F}_t^Y]} =: \frac{\rho_t(\phi)}{\rho_t(1)}.$$

193 Here, we have defined the unnormalized estimate $\rho_t(\phi) \triangleq \mathbb{E}_{\mathbb{Q}} [\phi(X_t) L_t | \mathcal{F}_t^Y]$.
194 Eq. (SM1.11) is known as the Bayes' formula for stochastic processes or *Kallianpur-*
195 *Striebel formula*. Under \mathbb{Q} , the stochastic differential can be taken inside the expec-
196 tation (see [SM14, Ch. 7, Lemma 7.2.7]), and we obtain using Itô's formula¹(SM1.3):

$$\begin{aligned} 197 \quad d\mathbb{E}_{\mathbb{P}} [\phi(X_t) | \mathcal{F}_t^Y] &= d\left(\frac{\rho_t(\phi)}{\rho_t(1)}\right) \\ 198 \quad &= \frac{1}{\rho_t(1)} d\rho_t(\phi) + \rho_t(\phi) d\left(\frac{1}{\rho_t(1)}\right) + d\rho_t(\phi) d\left(\frac{1}{\rho_t(1)}\right) \\ 199 \quad &= \frac{1}{\rho_t(1)} \mathbb{E}_{\mathbb{Q}} [d(\phi(X_t) L_t) | \mathcal{F}_t^Y] + \mathbb{E}_{\mathbb{Q}} [\phi(X_t) L_t | \mathcal{F}_t^Y] d\left(\frac{1}{\rho_t(1)}\right) + \\ 200 \quad (\text{SM1.12}) \quad &\quad \mathbb{E}_{\mathbb{Q}} [d(\phi(X_t) L_t) | \mathcal{F}_t^Y] d\left(\frac{1}{\rho_t(1)}\right). \\ 201 \end{aligned}$$

202 Let us evaluate these terms separately. The first term can be expressed as

$$\begin{aligned} 203 \quad \frac{1}{\rho_t(1)} \mathbb{E}_{\mathbb{Q}} [d(\phi(X_t) L_t) | \mathcal{F}_t^Y] &= \frac{1}{\rho_t(1)} \mathbb{E}_{\mathbb{Q}} \left[L_t d\phi_t + \phi_t dL_t + \underbrace{d\phi_t dL_t}_{=0} \right] \\ 204 \quad &= \frac{1}{\rho_t(1)} \left(\mathbb{E}_{\mathbb{Q}} [L_t \mathcal{A}^* \phi_t(X_t) | \mathcal{F}_t^Y] dt + \mathbb{E}_{\mathbb{Q}} [\phi_t(X_t) L_t \tilde{h}^\top(X_t) | \mathcal{F}_t^Y] d\tilde{Y}_t \right) \\ 205 \quad &= \mathbb{E}_{\mathbb{P}} [\mathcal{A}^* \phi_t(X_t) | \mathcal{F}_t^Y] dt + \mathbb{E}_{\mathbb{P}} [\phi_t(X_t) \tilde{h}^\top(X_t) | \mathcal{F}_t^Y] d\tilde{Y}_t, \end{aligned}$$

206 where we used the Itô formula for products. The term $d\phi_t(X_t) dL_t$ equals zero because
207 their noise components are independent.

208 For the second term, we first need to obtain $d\left(\frac{1}{\rho_t(1)}\right)$. For this, we apply Itô
209 formula to the inverse of the process $\rho(1) = (\rho_t(1), t \geq 0)$, where L_t solves (SM1.10).

$$\begin{aligned} 210 \quad d\rho(1) &= d(\mathbb{E}_{\mathbb{Q}} [L_t | \mathcal{F}_t^Y]) \\ 211 \quad &= \mathbb{E}_{\mathbb{Q}} [L_t \tilde{h}^\top(X_t) d\tilde{Y}_t | \mathcal{F}_t^Y]. \end{aligned}$$

¹Recall that we have to consider the product of differentials, since the Itô's formula corresponds to a Taylor expansion up to second order for diffusion processes.

212 Using Itô formula, we find:

$$\begin{aligned}
 213 \quad d\left(\frac{1}{\rho_t(1)}\right) &= -\rho_t(1)^{-2}d\rho_t(1) + \rho_t(1)^{-3}(d\rho_t(1))^2 \\
 214 \quad &= -\rho_t(1)^{-2}\mathbb{E}_{\mathbb{Q}}\left[L_t\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right]d\bar{Y}_t + \rho_t(1)^{-3}\left(\mathbb{E}_{\mathbb{Q}}\left[L_t\left\|\tilde{h}(X_t)\right\||\mathcal{F}_t^Y\right]\right)^2dt \\
 215 \quad &= -\rho_t(1)^{-1}\mathbb{E}_{\mathbb{P}}\left[\left(\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right)d\bar{Y}_t + \rho_t(1)^{-1}\left(\mathbb{E}_{\mathbb{P}}\left[\left\|\tilde{h}(X_t)\right\||\mathcal{F}_t^Y\right]\right)^2dt\right. \\
 216 \quad (\text{SM1.13}) \quad &= \left.-\rho_t(1)^{-1}\left(\mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]\right)^{\top}\left(d\tilde{Y} - \mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt\right).\right]
 \end{aligned}$$

217 Thus, the second term in Eq. (SM1.12) reads:

$$218 \quad \mathbb{E}_{\mathbb{Q}}\left[\phi(X_t)L_t|\mathcal{F}_t^Y\right]d\left(\frac{1}{\rho_t(1)}\right) = -\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right]\left(\mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]\right)^{\top}\left(d\tilde{Y} - \mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt\right). \blacksquare$$

219 Finally, the third term uses the result in Eq. (SM1.13) together with the following
220 unnormalized posterior expectation:

$$\begin{aligned}
 221 \quad \mathbb{E}_{\mathbb{Q}}\left[d(\phi(X_t)L_t)|\mathcal{F}_t^Y\right] &= \mathbb{E}_{\mathbb{Q}}\left[d(\phi(X_t))L_t + \phi(X_t)d(L_t) + d(\phi(X_t))d(L_t)|\mathcal{F}_t^Y\right] \\
 222 \quad (\text{SM1.14}) \quad &= \mathbb{E}_{\mathbb{Q}}\left[L_t\mathcal{A}^*\phi(X_t)|\mathcal{F}_t^Y\right]dt + \mathbb{E}_{\mathbb{Q}}\left[\phi(X_t)L_t\tilde{h}^{\top}(X_t)d\tilde{Y}_t|\mathcal{F}_t^Y\right].
 \end{aligned}$$

223 Here, we again used $\mathbb{E}_{\mathbb{Q}}\left[d(\phi_t(X_t))d(L_t)|\mathcal{F}_t^Y\right] = 0$ because their noise components are
224 independent.

225 Keeping only terms up to $\mathcal{O}(dt)$, the third term in Eq. (SM1.12) reads:

$$226 \quad \mathbb{E}_{\mathbb{Q}}\left[d(\phi(X_t)L_t)|\mathcal{F}_t^Y\right]d\left(\frac{1}{\rho_t(1)}\right) = -\mathbb{E}_{\mathbb{P}}\left[\phi_t(X_t)\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right]\mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt.$$

227 Adding up and rearranging the terms, we end up with

$$\begin{aligned}
 228 \quad d\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right] &= \mathbb{E}_{\mathbb{P}}\left[\mathcal{A}^*\phi_t(X_t)|\mathcal{F}_t^Y\right]dt + \mathbb{E}_{\mathbb{P}}\left[\phi_t(X_t)\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right]d\tilde{Y}_t \\
 229 \quad &\quad - \mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right]\left(\mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]\right)^{\top}\left(d\tilde{Y} - \mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt\right) \\
 230 \quad &\quad - \mathbb{E}_{\mathbb{P}}\left[\phi_t(X_t)\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right]\mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt \\
 231 \quad &= \mathbb{E}_{\mathbb{P}}\left[\mathcal{A}^*\phi_t(X_t)|\mathcal{F}_t^Y\right]dt + \\
 232 \quad &\quad \left(\mathbb{E}_{\mathbb{P}}\left[\phi(X_t)\tilde{h}^{\top}(X_t) - \mathbb{E}_{\mathbb{P}}\left[\phi(X_t)|\mathcal{F}_t^Y\right]\mathbb{E}_{\mathbb{P}}\left[\tilde{h}^{\top}(X_t)|\mathcal{F}_t^Y\right]\middle|\mathcal{F}_t^Y\right]\right) \times \\
 233 \quad (\text{SM1.15}) \quad &\quad \left(d\tilde{Y} - \mathbb{E}_{\mathbb{P}}\left[\tilde{h}(X_t)|\mathcal{F}_t^Y\right]dt\right). \\
 234
 \end{aligned}$$

235 Note that we can also write Eq. (SM1.15) in the integrated form:

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

239 where $\pi_0(\phi) = \mathbb{E}_{\mathbb{P}}\left[\phi(X_0)L_0|\mathcal{F}_0^Y\right] = \mathbb{E}_{\mathbb{P}}\left[\phi(X_0)\right]$. \square

240 COROLLARY SM1.4 (The Zakai equation - Proposition 5.2 revisited). *Let ϕ be \mathcal{C}^2
 241 and suppose that ϕ and all its derivatives are bounded. Then we can write*

$$242 \quad \rho_t(\phi) = \rho_0(\phi) + \int_0^t \rho_s(\mathcal{A}^* \phi) ds + \int_0^t \rho_s(k^{-1}(s) h\phi)^\top d\tilde{Y}_s,$$

243 where $\rho_0(\phi) = \mathbb{E}_{\mathbb{Q}} [\phi(X_0) L_0 | \mathcal{F}_0^Y] = \mathbb{E}_{\mathbb{Q}} [\phi(X_0)].$

244 *Proof.* The result follows from Eq. (SM1.14). □

245 **SM2. Mathematical perspective of stochastic filtering for the chemo-**
 246 **stat: algorithms.**

247 **SM2.1. Solving the Kushner-Stratonovich equation with methods for**
 248 **partial differential equations (PDEs).** Consider a twice-differentiable continuous
 249 function $v : [0, \infty) \times \mathbb{R}^{D_x} \rightarrow \mathbb{R}_+$ and its evolution law

250 (SM2.1)
$$\frac{\partial v(t, x)}{\partial t} = \mathcal{A}^* v(t, x), \quad v(0, x) = v_0(x),$$

251 where the operator \mathcal{A}^* is the second-order differential operator in Eq. (SM1.10), i.e.

252
$$\mathcal{A}^* v(t, x) = - \sum_{d_x=1}^{D_x} \frac{\partial}{\partial x_{d_x}} [f_{d_x}(x)v(t, x)] + \sum_{d_x=1}^{D_x} \sum_{d'_x=1}^{D_x} \frac{\partial^2}{\partial x_{d_x} \partial x_{d'_x}} [G_{d_x d'_x}(x)v(t, x)].$$

253 We approximate the solution to equations in the form of (SM2.1) by using a finite
 254 difference scheme on a given D_x -dimensional regular grid Ω^l with mesh $l = (l_1, \dots, l_{d_x})$
 255 in order to approximate the differential operator \mathcal{A}^* . Let ϵ_{d_x} be a unit vector in the
 256 d_x -th coordinate. The scheme approximates first-order derivatives of functions $z(\cdot)$
 257 evaluated at x according to the following up-wind scheme:

258
$$\left. \frac{\partial z}{\partial x_{d_x}} \right|_x \simeq \begin{cases} \frac{z(x + \epsilon_{d_x} l_{d_x}) - z(x)}{l_{d_x}} & \text{if } f_{d_x} \geq 0, \text{ i.e. forward difference} \\ \frac{z(x) - z(x - \epsilon_{d_x} l_{d_x})}{l_{d_x}} & \text{if } f_{d_x} < 0, \text{ i.e. backward difference,} \end{cases}$$

260 and the second-order derivates as

261
$$\left. \frac{\partial^2 z}{\partial x_{d_x}^2} \right|_{x_x} \simeq \frac{z(x + \epsilon_{d_x} l_{d_x}) - 2z(x) + z(x - \epsilon_{d_x} l_{d_x})}{l_{d_x}^2}.$$

263 The cross term $\left. \frac{\partial^2 z}{\partial x_{d_x} \partial x_{d'_x}} \right|_x$ is zero valued for the dynamics discussed in this paper.
 264 For general cases, see [SM1, Chapter 8].

265 Let $0 = t_0 < t_1 < \dots < t_n \dots$ be a uniform partition of the interval $[0, \infty)$ with
 266 time step $\Delta = t_n - t_{n-1}$. If the set of observations $\{z_i, 1 \leq i \leq n\}$ is available, the
 267 density p_{t_n} will be approximated by p_n^Δ , with the transition from p_{n-1}^Δ to p_n^Δ divided
 268 into two steps:

- 269 • The first step, called the *prediction* step, consists in solving the Fokker-Planck
 270 equation for the time interval $[t_{n-1}, t_n]$,

271
$$\frac{\partial p_t^n(x)}{\partial t} = \mathcal{A}^* p_t(x), \quad p_{t_{n-1}}^n = p_{n-1}^\Delta.$$

272 We then let the prior estimate be $\bar{p}_n^\Delta \stackrel{\Delta}{=} p_{t_n}$.

- 273 • The second step, called the *correction* step, uses the new observation z_n
 274 to update \bar{p}_n^Δ . For this, we use the approximation to the Radon-Nikodym
 275 derivative from the theory of pathwise filtering, described by Eq. (SM2.2)
 276 below. We compute $p_n^\Delta(x)$ for $x \in \mathbb{R}^{D_x}$ via

277
$$p_n^\Delta(x) \stackrel{\Delta}{=} c_n \zeta_n^\Delta(x) \bar{p}_n^\Delta,$$

where

$$(SM2.2) \quad \zeta_n^\Delta(x) \triangleq \exp\left(\tilde{h}^\top(x)\tilde{z}_n - \frac{1}{2}\Delta\|\tilde{h}(x)\|^2\right),$$

c_n is a normalization constant chosen such that $\int_{\mathbb{R}^{D_x}} p_n^\Delta(x)dx = 1$, $\tilde{z}_n = k^{-1}(t_n)z_n$ and $\tilde{h}(x) = k^{-1}(t_n)h(x)$.

Algorithm SM2.1 Splitting algorithm for the Kushner-Stratonovich Equation with PDE methods

Result: $p^\Delta = (p_t^\Delta, t \geq 0)$

- 1 Initialise p_0^Δ with a desirable density and let $\{t_n\}, n = 1, 2, \dots$ denote the times to estimate $\pi_t(\phi)$ for a test function ϕ ;
- 2 **for** $n = 1, 2, \dots$ **do**
- 3 For $t \in [t_{n-1}, t_n]$, solve the following partial differential equation

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

- 4 Define $\bar{p}_n^\Delta := p_{t_n}^n(x)$;
- 5 **if** t_n is an observation time **then**
- 6 Update the density with information from observation z_n , that is, compute $p_n^\Delta(x)$ for $x \in \mathbb{R}^{D_x}$ via

$$p_n^\Delta(x) = c_n \zeta_n^\Delta(x) \bar{p}_n^\Delta,$$

where

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

\tilde{z} and \tilde{h} stands for rescaled versions of z and h , respectively, and c_n is a normalisation constant chosen such that

$$\int_{\mathbb{R}^{D_x}} p_n^\Delta(x)dx = 1;$$

- 7 **end**
 - 8 **else**
 - 9 | Set $p_n^\Delta \leftarrow p_{t_n}^n$.
 - 10 **end**
 - 11 **end**
-

Algorithm SM2.2 Continuous-Discrete Bootstrap Particle Filter

Result: $\pi^N(\phi) = (\pi_{t_n}^N(\phi), t_n = t_0, t_1, \dots)$

```

12 Initialize  $p_0^N$  with a desirable density and let  $\{t_n\}, n = 1, 2, \dots$ , denote the times to
   estimate  $\pi_t(\phi)$  for a test function  $\phi$ ;
13 Sample  $V_j(0), j = 1, \dots, N$  mutually independent stochastic processes from  $p_0^\Delta$ , each
   with initial weight  $1/N$ ;
14 for  $n = 1, 2, \dots$  do
    /* Use a resampling step to eliminate particles with low weights
       and multiply those with high weights. */
15 Resample  $\{V_j(t_{n-1}), w_j(t_{n-1})\}_{j=1}^N$ , resulting in equally weighed particles
    $\{\tilde{V}_j(t_{n-1}), 1/N\}_{j=1}^N$ ;
16 for  $j = 1, 2, \dots, N$  do
    /* Evolve the particles from  $t = t_{n-1}$  to  $t = t_n$  */
17
    
$$V_j(t_n) = \tilde{V}_j(t_{n-1}) + \int_{t_{n-1}}^{t_n} f(\tilde{V}_j(t))dt + \int_{t_{n-1}}^{t_n} g(\tilde{V}_j(t))dB_t^x,$$

    which can be solved by any numerical solver for SDEs;
18 end
19 if  $t_n$  is an observation time then
20   Update the density with information from observation  $z_n$ , that is, update the
      weights  $w_j(t)$  via
      
$$w_j(t_n) = e^{\left(\int_0^{t_n} (k^{-1}(s)h(V_j(s)))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_0^{t_n} \|k^{-1}(s)h(V_j(s))\|^2 ds\right)}$$

      
$$\propto w_j(t_{n-1}) e^{\left(\int_{t_{n-1}}^{t_n} (k^{-1}(s)h(V_j(s)))^\top k^{-1}(s)dY_s - \frac{1}{2} \int_{t_{n-1}}^{t_n} \|k^{-1}(s)h(V_j(s))\|^2 ds\right)}$$

      
$$\approx w_j(t_{n-1}) e^{\left((k^{-1}(t_n)h(V_j(t_n)))^\top \left(\frac{k(t_n)}{\Delta}\right)^{-1} z_n - \frac{1}{2} \|k^{-1}(t_n)h(V_j(t_n))\|^2 \Delta\right)},$$

      where  $e^a \equiv \exp a$ . Normalize the weights with
      
$$\bar{w}_j(t) = \frac{w_j(t)}{\sum_{j'=1}^N w_{j'}(t)}, \quad j = 1, \dots, N;$$

      Set  $w_j(t_n) \leftarrow \bar{w}_j(t_n)$ ;
21 end
22 Use the approximation
23

$$\pi_{t_n}^N(\phi) = \sum_{j=1}^N w_j(t_n) \phi(V_j(t_n)).$$

24 end

```

282 **SM2.2. Solving the Kushner-Stratonovich equation with particle meth-**
 283 **ods.**

284 **SM2.3. Solving the Kushner-Stratonovich equation with linearization**
 285 **methods.**

286 **SM2.3.1. The Kalman-Bucy Filter.** Before we present more approximation
 287 methods, let us first examine a case for which the Kushner-Stratonovich equation has
 288 a closed-form solution. Consider the signal process with evolution in time described
 289 by Eq. (SM1.2) and the measurement model written as

290 $dY_t = h(X_t)dt + k(t)dB_t^y, \quad Y_0 = 0.$

291 As a starting point, let us assume that f and h are linear functions of the state
 292 and that g is only function of t . It is possible to rewrite our state-space model as:

293 $dX_t = a(t)X_t + g(t)dB_t^x, \quad X_0 = \xi; \quad dY_t = c(t)X_t + k(t)dB_t^y, \quad Y_0 = 0.$

294 Now suppose that

- 295 • ξ is a Gaussian random variable with mean m_0 and covariance P_0 ;
- 296 • $k(t)$ is invertible for all $t \in [0, \infty)$;
- 297 • $a(t), g(t), c(t), k(t), k^{-1}(t)$ are continuous.

298 THEOREM SM2.1 (Kalman-Bucy filter). *The estimates for the conditional mean*
 299 $m_t \stackrel{\Delta}{=} \mathbb{E}[X_t | \mathcal{F}_t^Y]$ *and error covariance* $P \stackrel{\Delta}{=} \mathbb{E}[(X_t - m_t)(X_t - m_t)^\top]$ *form the solu-*
 300 *tion to:*

301 $dm_t = a(t)m_t dt + P_t (k^{-1}(t)c(t))^\top d\tilde{B}_t^x,$
 302 $\frac{dP_t}{dt} = a(t)P_t + P_t a^\top(t) - P_t c^\top(t) (kk^\top(t))^{-1} c(t)P_t + kk^\top(t),$

303 with initial conditions m_0 and P_0 . Here, we made use of the so-called innovation
 304 process, where $d\tilde{B}_t^x = k^{-1}(t)(dY_t - c(t)m_t dt)$.

305 It is known that the solution p_t to this filtering problem refers to the density of
 306 a Gaussian distribution with mean m_t and covariance matrix P_t . One can verify this
 307 claim by noting that

308 $q_t \propto \exp\left(-\frac{1}{2}(x - m_t)^\top P_t^{-1}(x - m_t)\right)$

309 satisfies the Zakai equation for the setup in Eq. (SM2.3.1),

310 $dq_t(x) = q_t(x)(k^{-1}(t)c_t^u)^\top d\tilde{Y}_t - \left[\sum_{d_x=1}^{D_x} \frac{\partial}{\partial x_{d_x}} [(a(t)x)_{d_x} p(t, x)] \right] dt$
 311 $+ \left[\frac{1}{2} \sum_{d_x=1}^{D_x} \sum_{d'_x=1}^{D_x} \frac{\partial^2}{\partial x_{d_x} \partial x_{d'_x}} [(gg^\top(t))_{d_x d'_x} p(t, x)] \right] dt$

313 Note that showing that the unnormalised conditional distribution ρ_t has a density
 314 q_t of a Gaussian distribution automatically implies that π_t has the same property.

315 As presented in the main manuscript, once we fix the initial condition $X_0 \sim$
 316 $\mathcal{N}(m_0, P_0)$, the exact solution to the linear Gaussian filtering problem is given by the
 317 Kalman-Bucy filter [SM6]. Starting at iteration $n = 1$, we propagate the estimates
 318 via the procedure below.

- 319 • The *prediction* step consists of finding the estimate for the predictive mean
 320 m_t^- satisfying

321

$$\frac{dm_t^-}{dt} = a(t)m_t^-, \quad m_{t_{n-1}}^- = m_{t_{n-1}},$$

322 and the estimate for the predictive covariance P_t^- satisfies the equation

323

$$\frac{dP_t^-}{dt} = a(t)P_t^- + P_t^-a^\top(t) + gg^\top(t) \quad P_{t_{n-1}}^- = P_{t_{n-1}}.$$

- 324 • The *correction* step consists of using the new observation z_n at time step t_n .
 325 The state mean m_{t_n} and covariance P_{t_n} are updated according to

326

$$K_n = P_{t_n}^{-1}c^\top(t_n) (c(t_n)P_{t_n}^{-1}c^\top(t_n) + kk^\top(t_n))^{-1},$$

327

$$m_{t_n} = m_{t_n} + K_n(z_n - c(t_n)m_{t_n}^-),$$

328

$$P_{t_n} = (I - K_n c(t_n)) P_{t_n}^- (I - K_n c(t_n))^\top + K_n k(t_n)(K_n k(t_n))^\top.$$

Algorithm SM2.3 Continuous-Discrete Kalman-Bucy Filter

Result: $m = \{m_t\}, P = \{P_t\}, t \geq 0$

25 Initialize m_0 and P_0 with desirable values and let $\{t_n\}, n = 1, 2, \dots$, denote the times
 to estimate the state variable X_t ;

26 **for** $n = 1, 2, \dots$ **do**

27 Compute the predicted state mean and covariance $(m^-(t), P^-(t))$ as solutions to

$$dm_t^- = a(t)m_t^- dt$$

$$\frac{dP_t^-}{dt} = a(t)P_t^- + P_t^-a^\top(t) + gg^\top(t),$$

28 for $t \in [t_{n-1}, t_n]$, and with initial conditions $m_{n-1}^- = m_{n-1}$ and $P_{n-1}^- = P_{n-1}$;

29 **if** t_n is an observation time **then**

30 Update state estimates with information from observation z_n :

$$e_n = z_n - c(t_n)m_n^-$$

$$S_n = c(t_n)P_{t_n}^-c^\top(t_n) + kk^\top(t_n)$$

$$K_n = P_{t_n}^{-1}(c(t_n)m_n^-)^\top S_n^{-1}$$

$$m_n = m_n^- + K_n e_n$$

$$P_n = P_{t_n}^- - K_n S_n^{-1} K_n^\top;$$

31 **end**

32 **else**

33 | Set $m_n \leftarrow m_n^-$ and $P_n \leftarrow P_n^-$.

34 **end**

35 **end**

329 **SM2.3.2. The Extended Kalman Filter.** Now we proceed to detail a lineariza-
 330 tion method to cope with non-linear models, for which f and/or h are non-linear

functions of X . We present the most common approach, the extended Kalman filter. The method assumes that the conditional density is nearly Gaussian, so that third and higher order odd central moments are essentially zero, and the fourth and higher order even central moments can be expressed in terms of the covariance.

We apply the assumption above to truncate high-order terms in a Taylor series expansion to functions f and h . Let \bar{x}_t be a solution of the ODE $\frac{d\bar{x}}{dt} = f(\bar{x}_t)$, $\bar{x}_0 = m_0$.

Consider a small-time frame for which the contribution of the stochastic terms $g(X_t)dB_t^x$ and $k(t)dB_t^y$ remains small, such that a trajectory $t \mapsto X_t$ may be viewed as a perturbation from the deterministic trajectory $t \mapsto \bar{x}_t$. Instead of solving the filtering problem for the original system of differential equations, we work with the following Taylor-like expansion:

$$\begin{aligned} dX_t &\simeq (F(\bar{x}_t)(X_t - \bar{x}_t) + f(\bar{x}_t))dt + g(\bar{x}_t)dB_t^x, \\ dY_t &\simeq (H(\bar{x}_t)(X_t - \bar{x}_t) + h(\bar{x}_t))dt + k(t)dB_t^y, \end{aligned}$$

where, \simeq means approximately equal without a rigorous mathematical meaning to it, F and H are the derivatives of f and h , that is

$$F(\bar{x}_t) \triangleq \left. \frac{\partial f}{\partial x} \right|_{x_t=\bar{x}_t}, \quad H(\bar{x}_t) \triangleq \left. \frac{\partial h}{\partial x} \right|_{x_t=\bar{x}_t}.$$

If we let $\bar{x}_t \equiv m_t$, the resulting equations form the extended Kalman filter.

- For the time interval $[t_{n-1}, t_n]$, the *prediction* step consists in finding the predictive estimate m_t^- for the mean satisfying

$$\frac{dm_t^-}{dt} = f(m_t^-), \quad m_{t_{n-1}}^- = m_{t_{n-1}},$$

and the predictive estimate P_t^- for the covariance satisfying the equation

$$\begin{aligned} \frac{dP_t^-}{dt} &= F(m_t^-)P_t^- + R_t^- F^\top(m_t^-) + gg^\top(m_t), \\ P_{t_{n-1}}^- &= P_{t_{n-1}}, \\ F(m_t^-) &\triangleq \left. \frac{\partial f}{\partial x} \right|_{x_t=m_t^-}. \end{aligned}$$

- The *correction* step consists of using the new observation z_n at time step t_n . The state mean m_{t_n} and covariance P_{t_n} are updated according to

$$\begin{aligned} H(m_{t_n}) &\triangleq \left. \frac{\partial h}{\partial x} \right|_{x_t=m_{t_n}}, \\ K_n &= P_{t_n}^- H^\top(m_n) (H(m_{t_n})P_{t_n}^- H^\top(m_{t_n}) + kk^\top(t_n))^{-1}, \\ m_{t_n} &= m_{t_n}^- + K_n(z_n - h(m_{t_n})), \\ P_{t_n} &= (I - K_n H^\top(m_{t_n})) P_{t_n}^- (I - K_n H^\top(m_{t_n})) + K_n k(t_n) (K_n k(t_n))^\top. \end{aligned}$$

To cope with numerical errors, Kalman filter equations are often implemented such that the matrix square roots of covariance matrices are used in computations instead of their actual values. Following the derivation in [SM9] and [SM13], we can rewrite the extended Kalman filter equations in terms of generalized or modified

SUPPLEMENTARY MATERIALS: KSE FOR THE STOCHASTIC CHEMOSTAT SM17

365 Cholesky factors. Let $S_t S_t^\top = P_t$, then we solve the differential equation for either
366 $\frac{dS_t}{dt}$ or $\frac{dS_t^\top}{dt}$. If one still chooses to implement the standard solution, the problem
367 requires ODE solvers with variable step size for their accurate and efficient numerical
368 integration [SM3, SM8, SM10]. Lastly, it may be beneficial to maintain symmetry
369 of P by evaluating the expression $P = 0.5 * (P + P')$ on every cycle of the Riccati
370 equation [SM4].

371 Pseudocode for the squared-root extended Kalman filter is as follows:

Algorithm SM2.4 Squared-root Extended Kalman Filter

Result: $m = \{m_t\}, S = \{S_t\}, t \geq 0$

- 35 Initialise m_0 and S_0 with desirable values, such that $S_0 S_0^\top = P_0$, and let $\{t_n\}, n = 1, \dots$, denote the times to estimate the state variable X_t ;
 36 Define the matrix-valued function $\Phi(\cdot)$ that returns the lower triangular part of argument matrix M , as follows:

$$\Phi_{ij}(M) = \begin{cases} M_{ij} & \text{if } i > j \\ 0.5M_{ij} & \text{if } i = j \\ 0 & \text{if } i < j. \end{cases}$$

- 37 **for** $n = 1, 2, \dots$ **do**
 38 Compute the predicted state estimators (m_t^-, S_t^-) as solutions to

$$dm_t^- = f(m_t^-)dt, \quad \frac{dS_t^-}{dt} = S_t \Phi(A_t + A_t^\top + B_t),$$

where

$$A_t = (S_t)^{-1} F_t S_t, \quad B_t = (S_t)^{-1} g g^\top (m_t^-)((S_t)^{-1})^\top, \quad F_t = \left. \frac{\partial f}{\partial x} \right|_{m_t^-},$$

for $t \in [t_{n-1}, t_n]$, and with initial conditions $m_{t_{n-1}}^- = m_{t_{n-1}}$ and $S_{t_{n-1}}^- = S_{t_{n-1}}$;

- 39 **if** t_n is an observation time **then**
 40 Update state estimates. Start with a QR decomposition of the matrix on the LHS of the following equation

$$\begin{pmatrix} k(t_n)(t_n - t_{n-1})^{1/2} & H_x(m_{t_n}^-)S_{t_n}^- \\ 0^\top & S_{t_n}^- \end{pmatrix}^\top = Q_n \begin{pmatrix} R_n^{1/2} & 0 \\ \tilde{K}_n & S_{t_n} \end{pmatrix}^\top,$$

where $H_x(m) \equiv \left. \frac{\partial h}{\partial x} \right|_{X_t=m}$. You can then extract Q_n as the Q -factor and let the last matrix on the RHS be the R -factor. Directly extract S_{t_n} from the latter;

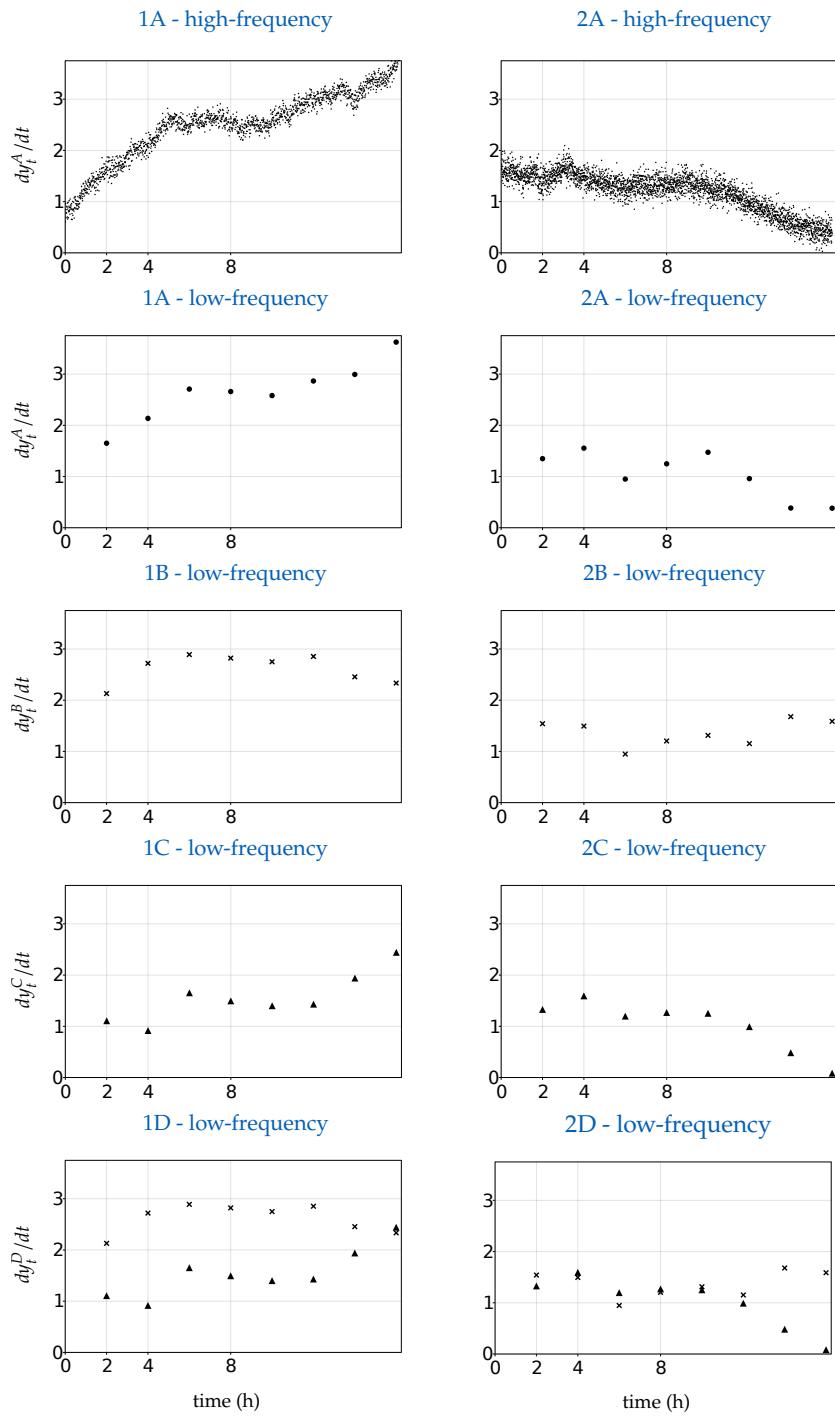
- 41 For the mean estimate:

$$\begin{aligned} e_n &= y_n - h(m_{t_n}^-), \\ K_n &= \tilde{K}_n((R_n^{1/2})^{-1})^\top, \\ m_{t_n} &= m_{t_n}^- + K_n e_n; \end{aligned}$$

- 42 **end**
 43 **else**
 44 | Set $m_{t_n} \leftarrow m_{t_n}^-$ and $S_{t_n} \leftarrow S_{t_n}^-$.
 45 **end**
 46 **end**
-

SM3. Numerical experiments.**Table SM1** Additional model parameters and inputs for differential equations of X and Y (Extension of Table 2).

Obs. function	Model parameters		Simul. parameters	
	$\omega_{l, l \in \{A, B, C, D\}}$	Initial condition (mgL^{-1} , mgL^{-1})	δ_x (h)	δ_y (h)
1A - high-freq.	$27.5\mu(s)b$	0.01	(1.0, 1.0)	0.01
1A - low-freq.	$27.5\mu(s)b$	$0.01(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.0, 1.0)	0.01
1B - low-freq.	s	$0.02(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.0, 1.0)	0.01
1C - low-freq.	b	$0.02(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.0, 1.0)	0.01
1D - low-freq.	(b, s)	$0.02(\delta_y/\delta_x)^{\frac{1}{2}} I_{2 \times 2}$	(1.0, 1.0)	0.01
2A - high-freq.	$2.75\mu(s)b$	0.01	(1.65, 2.5)	0.005
2A - low-freq.	$2.75\mu(s)b$	$0.01(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.65, 2.5)	0.005
2B - low-freq.	s	$0.02(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.65, 2.5)	0.005
2C - low-freq.	b	$0.02(\delta_y/\delta_x)^{\frac{1}{2}}$	(1.65, 2.5)	0.005
2D - low-freq.	(b, s)	$0.02(\delta_y/\delta_x)^{\frac{1}{2}} I_{2 \times 2}$	(1.65, 2.5)	0.005

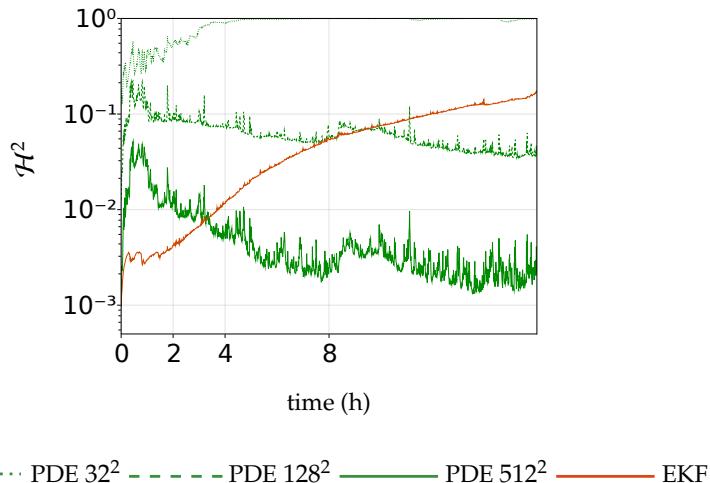
**Fig. SM1** Measurements simulated from trajectories highlighted in Figure 2.

375 We now present the approximation to the solution of the Kushner-Stratonovich
 376 equation for the growth function of the Monod type with different measurement sce-
 377 narios. The initial condition is $X_0 \sim \mathcal{N}([1, 1]^\top, 0.05^2 I_2)$.

378
 379 For each scenario, we firstly show the squared Hellinger distances between the
 380 densities obtained by approximating the solution to the Kushner-Stratonovich equa-
 381 tion with particle filters and with (i) methods for PDEs or (ii) the EKF. For the
 382 PDE methods, the discretisation in space is in the domain $(0, 5.0] \times (0, 5.0]$, and we
 383 consider three scenarios for the refinement of the grid, each containing a total count
 384 of $32^2, 128^2, 512^2$ finite volumes. The squared Hellinger distance is computed with
 385 respect to the approximation from the BPF with $N = 2.5 \times 10^6$. Secondly, we present
 386 snapshots of the approximations at selected times t . The prior knowledge from the
 387 Kolmogorov forward equation is represented by marginal density functions in gray.
 388 The updated marginal density function is represented by coloured curves.

389
 390 For scenarios with sparse measurements, we additionally present the splitting-up
 391 approximation.

392 **1A - high-frequency (Monod kinetics, measurements of biogas flow
 393 rate).**



394
 395 **Fig. SM2** Squared Hellinger distance for 1A - high-frequency. BPF: $N = 2.5 \times 10^6$.

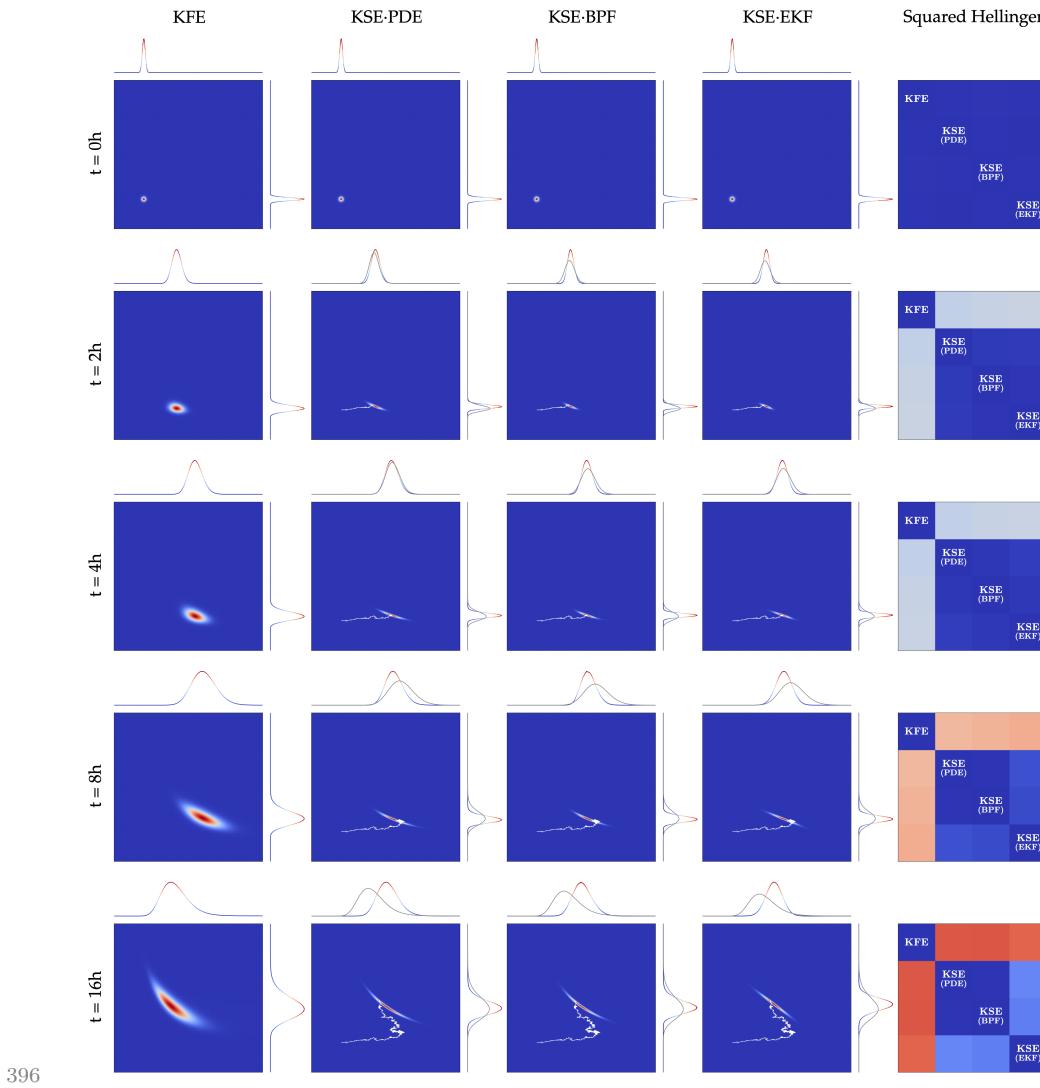


Fig. SM3 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Monod type with continuous stream of measurements of biogas flow rate. 256^2 grid. $N = 2.5 \times 10^6$.

398 **1A - low-frequency (Monod kinetics, sparse measurements of biogas
399 flow rate).**

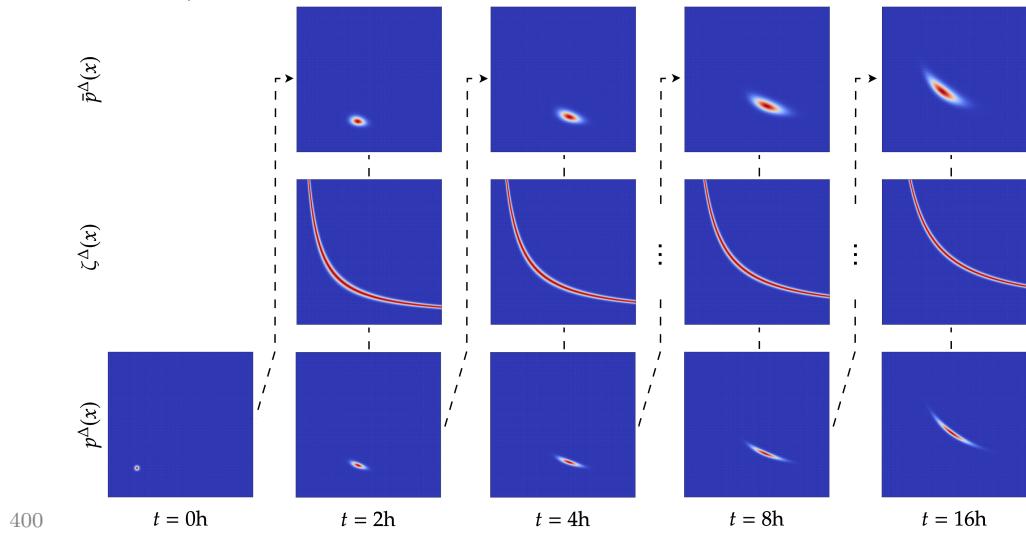


Fig. SM4 Splitting-up approximation for the Kushner-Stratonovich Equation using
401 a 256^2 grid.

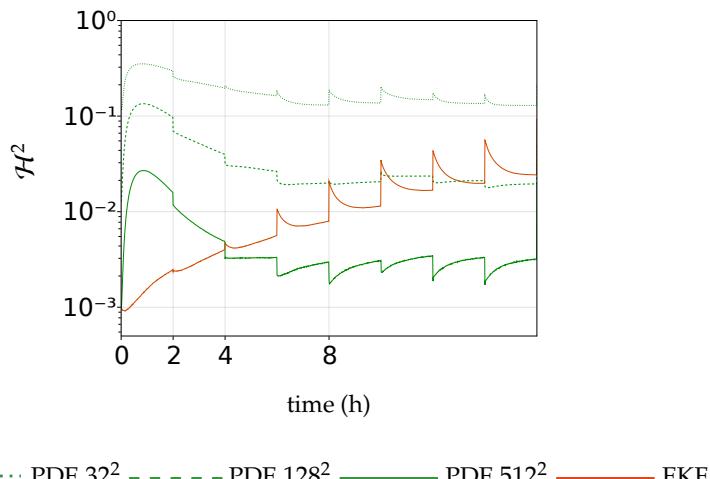


Fig. SM5 Squared Hellinger distance for 1A - low-frequency. BPF: $N = 2.5 \times 10^6$.
403

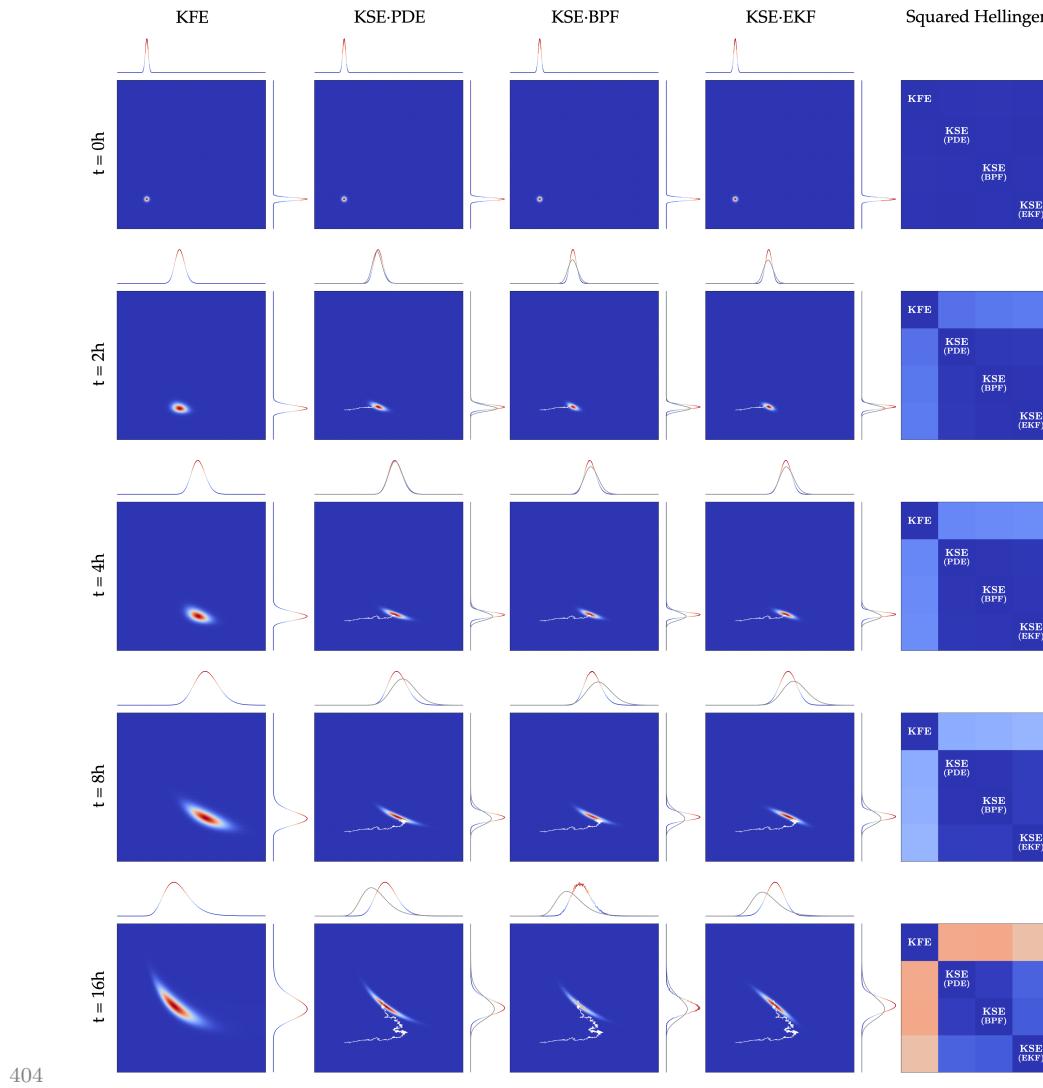
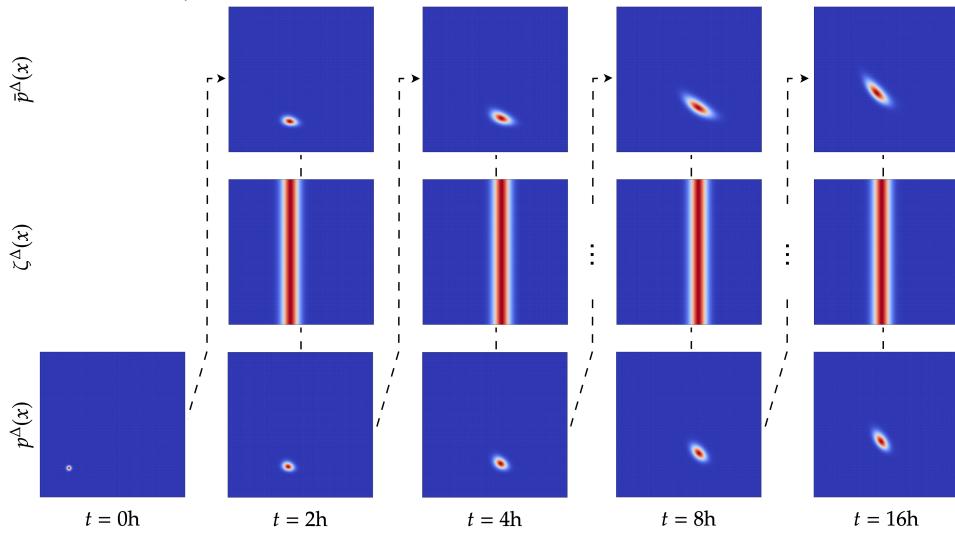
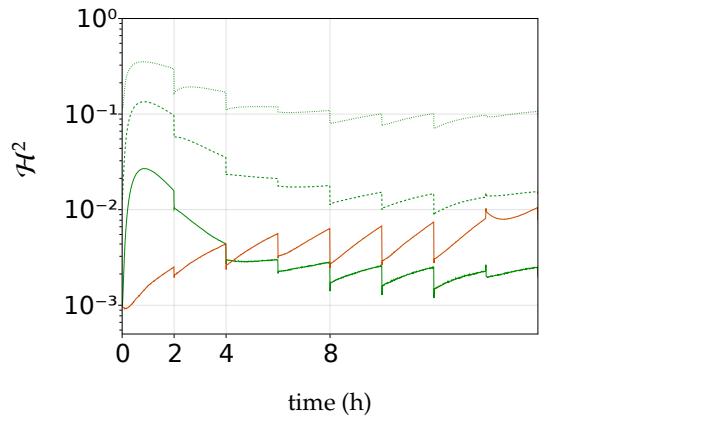


Fig. SM6 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Monod type with sparse measurements of biogas flow rate.
256² grid. $N = 2.5 \times 10^6$.

406 **1B - low-frequency (Monod kinetics, sparse measurements of substrate**
 407 **concentration).**



409 **Fig. SM7** Splitting-up approximation for the Kushner-Stratonovich Equation using
 a 256^2 grid.



410
 411 **Fig. SM8** Squared Hellinger distance for 1B - low-frequency. BPF: $N = 2.5 \times 10^6$.

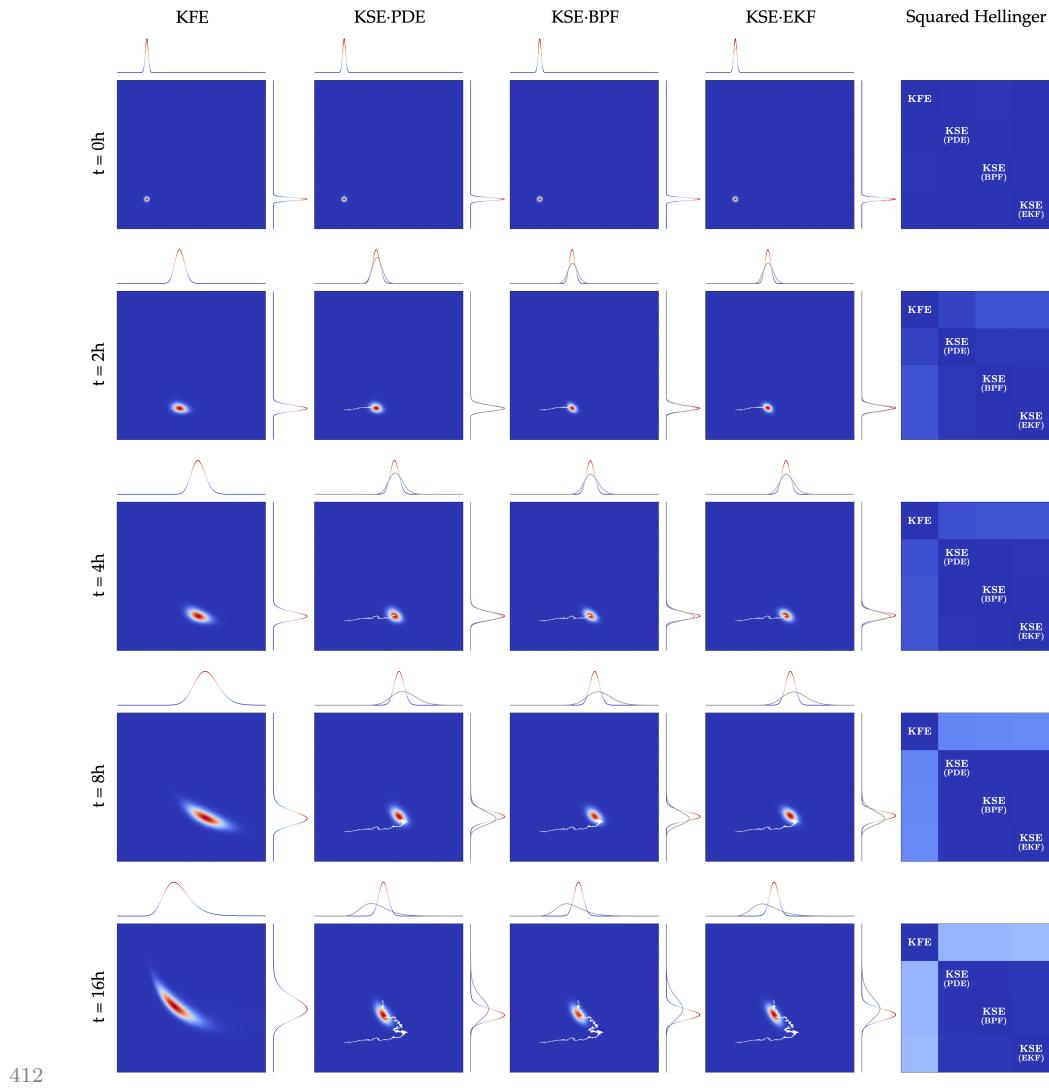
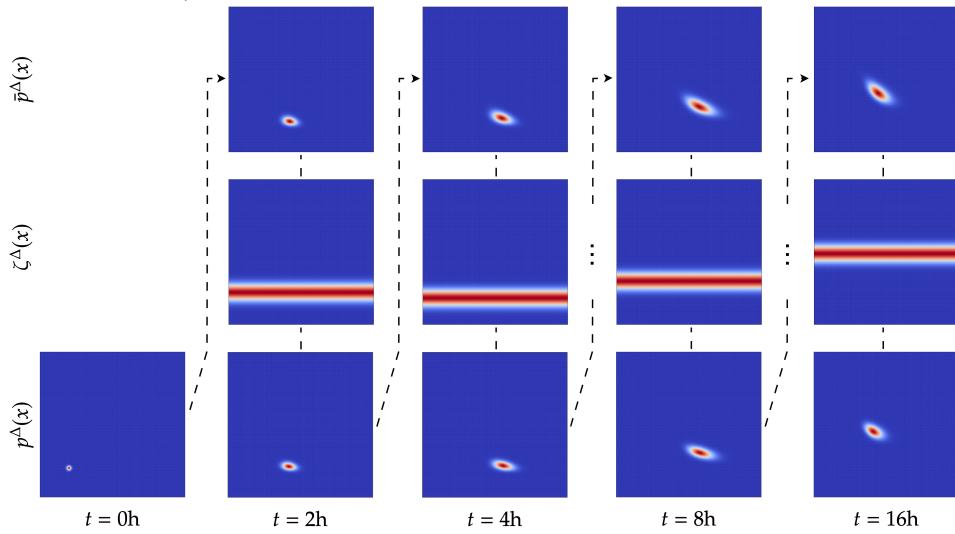
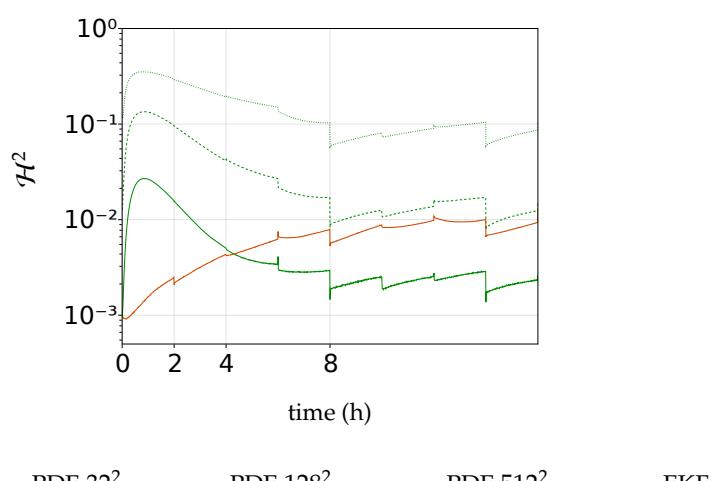


Fig. SM9 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Monod type with sparse measurements of substrate concentration. 256² grid. $N = 2.5 \times 10^6$.

414 **1C - low-frequency (Monod kinetics, sparse measurements of biomass**
 415 **concentration).**



416 **Fig. SM10** Splitting-up approximation for the Kushner-Stratonovich Equation using
 417 a 256^2 grid.



418 **Fig. SM11** Squared Hellinger distance for 1C - low-frequency. BPF: $N = 2.5 \times 10^6$.
 419

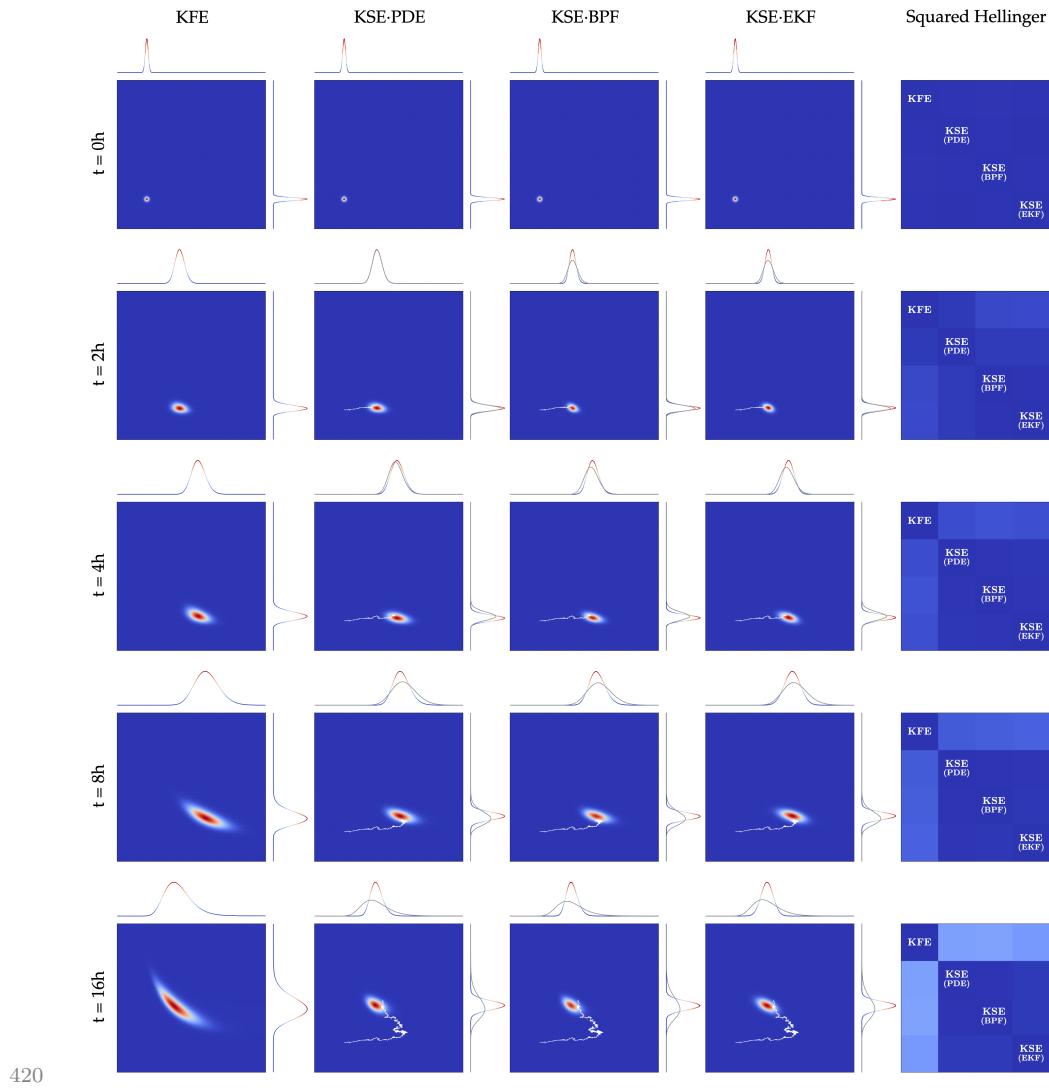
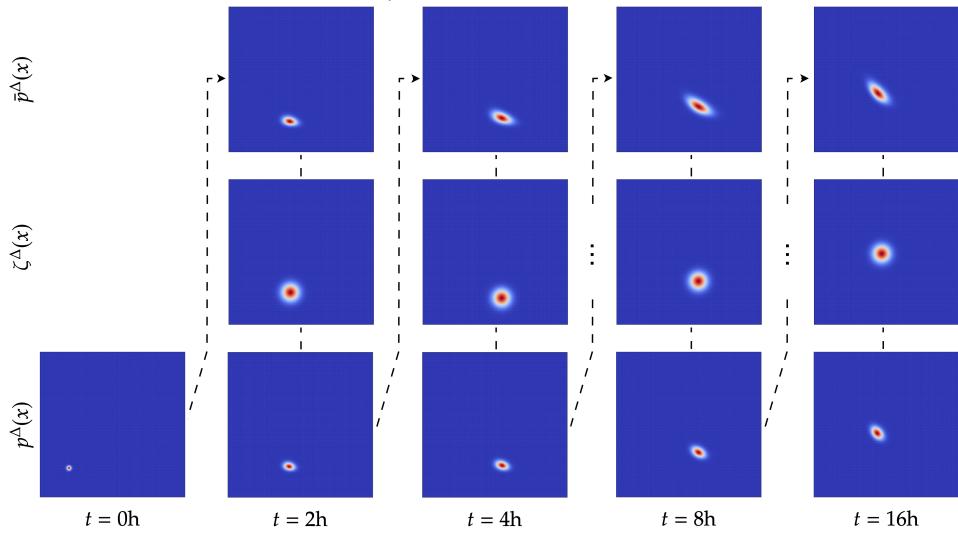
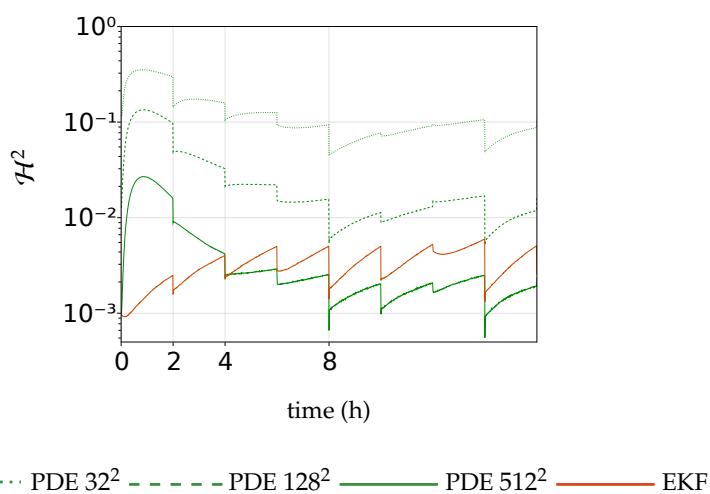


Fig. SM12 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Monod type with sparse measurements of biomass concentration. 256^2 grid. $N = 2.5 \times 10^6$.

422 **1D - low-frequency (Monod kinetics, sparse measurements of biomass**
 423 **and substrate concentrations).**



424 **Fig. SM13** Splitting-up approximation for the Kushner-Stratonovich Equation using
 425 a 256^2 grid.



426 **Fig. SM14** Squared Hellinger distance for 1D - low-frequency. BPF: $N = 2.5 \times 10^6$.
 427

428

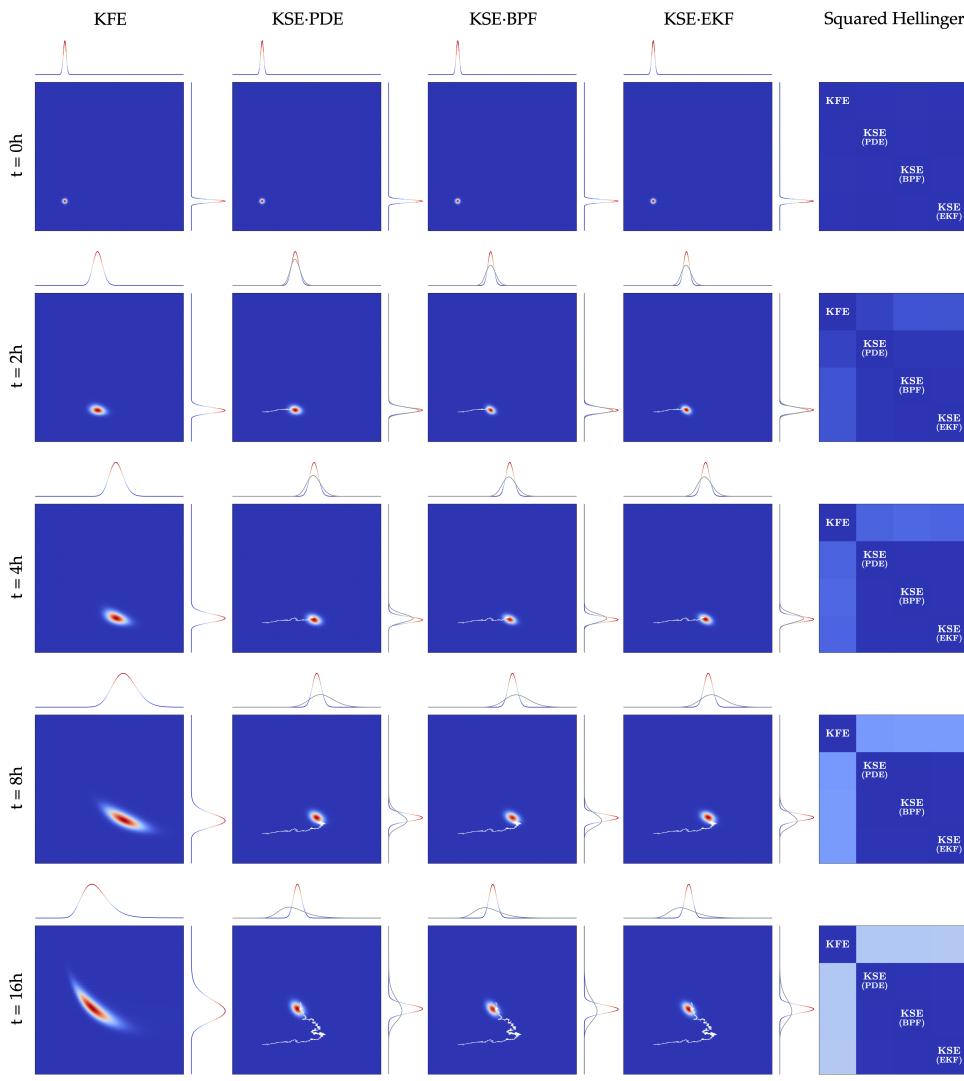


Fig. SM15 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Monod type with sparse measurements of biomass and substrate concentrations. 256^2 grid. $N = 2.5 \times 10^6$.

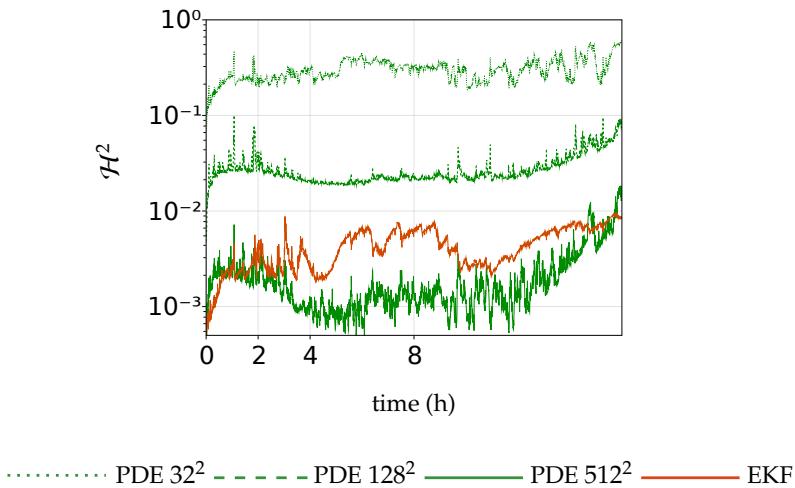
429

430 We now present the approximation to the solution of the Kushner-Stratonovich
 431 equation for the growth function of the Haldane type with different measurement sce-
 432 narios. The initial condition is $X_0 \sim \mathcal{N}([1.65, 2.5]^\top, 0.05^2 I_2)$. The prior knowledge
 433 from the Kolmogorov forward equation is represented by marginal density functions
 434 in gray. The updated marginal density function is represented by coloured curves.
 435

436 For each scenario, we firstly show the squared Hellinger distances between the
 437 densities obtained by approximating the solution to the Kushner-Stratonovich equa-
 438 tion with particle filters and with (i) methods for partial differential equations or (ii)
 439 the extended Kalman filter. For the PDE methods, the discretisation in space is in
 440 the domain $(0, 3.0] \times (0, 3.0]$, and we consider three scenarios for the refinement of
 441 the grid, each containing a total count of $32^2, 128^2, 512^2$ finite volumes. The squared
 442 Hellinger distance is computed with respect to the approximation from the bootstrap
 443 particle filter with $N = 2.5 \times 10^6$. Secondly, we present snapshots of the approxima-
 444 tions at selected times t . The prior knowledge from the Kolmogorov forward equation
 445 is represented by marginal density functions in gray. The updated marginal density
 446 function is represented by coloured curves.

447
 448 For scenarios with sparse measurements, we additionally present the splitting-up
 449 approximation.

450 **2A - high-frequency (Haldane kinetics, measurements of biogas flow
 451 rate).**



452

453 **Fig. SM16** Squared Hellinger distance for 2A - high-frequency. BPF: $N = 2.5 \times 10^6$.

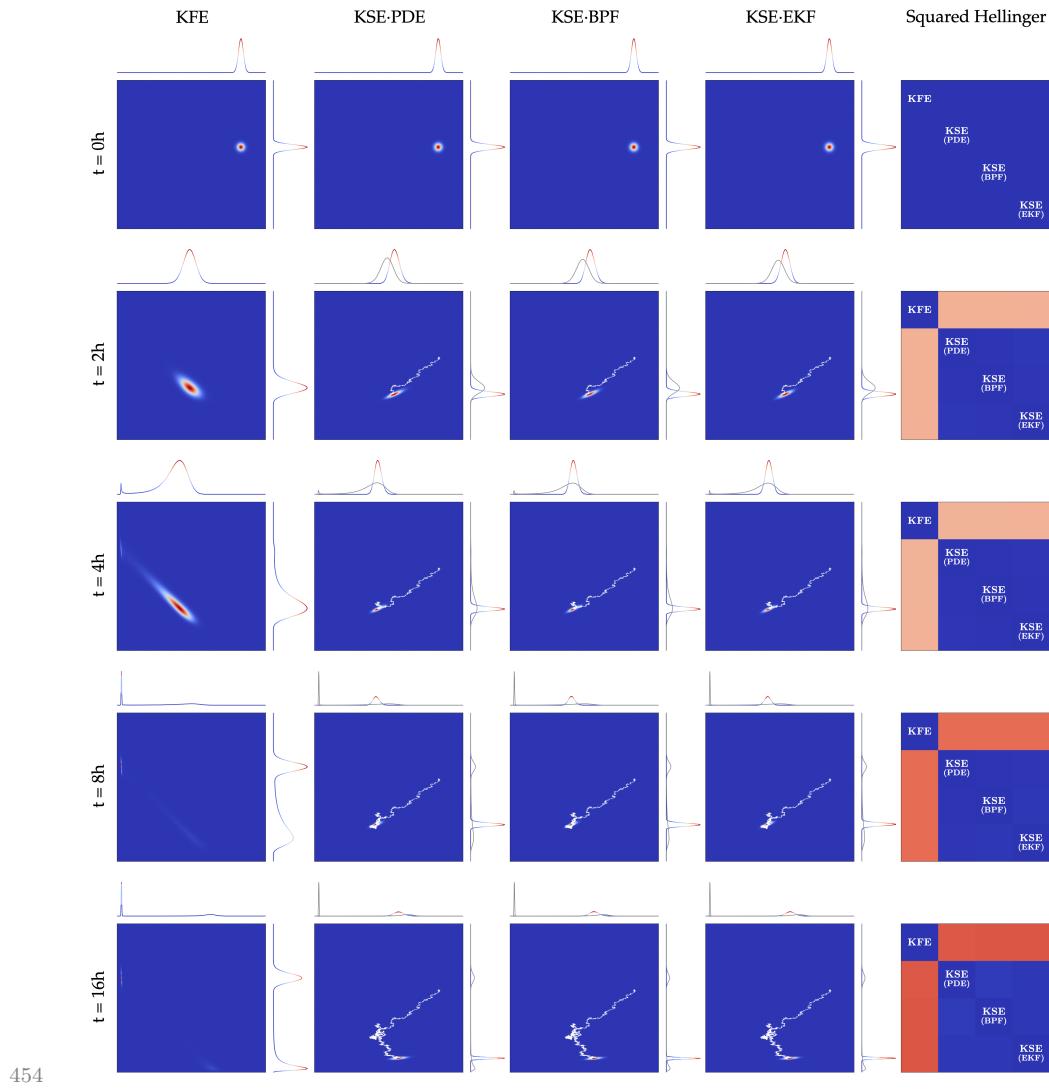
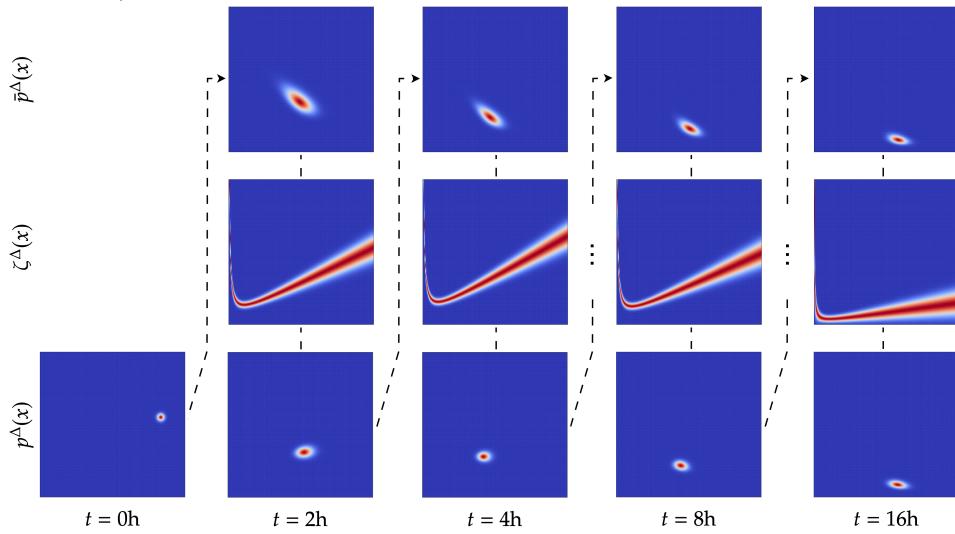


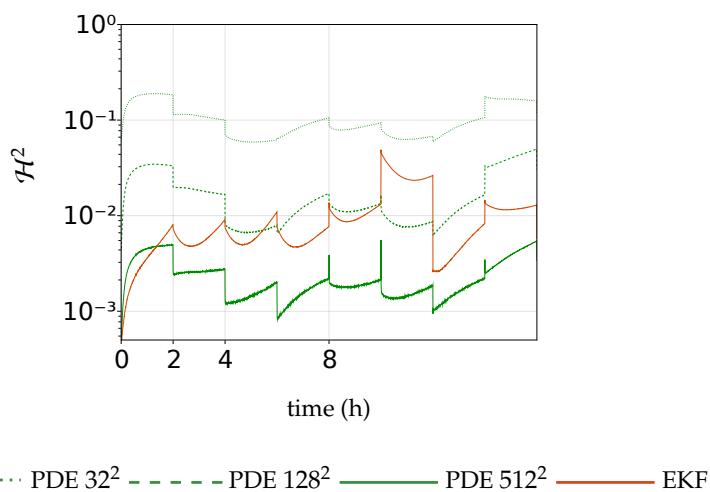
Fig. SM17 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Haldane type with continuous stream of measurements of biogas flow rate. 512^2 grid. $N = 2.5 \times 10^6$.

455

456 **2A - low-frequency (Haldane kinetics, sparse measurement of biogas
457 flow rate).**



458 **Fig. SM18** Splitting-up approximation for the Kushner-Stratonovich Equation using
459 a 256^2 grid.



460 **Fig. SM19** Squared Hellinger distance for 2A - low-frequency. BPF: $N = 2.5 \times 10^6$.
461

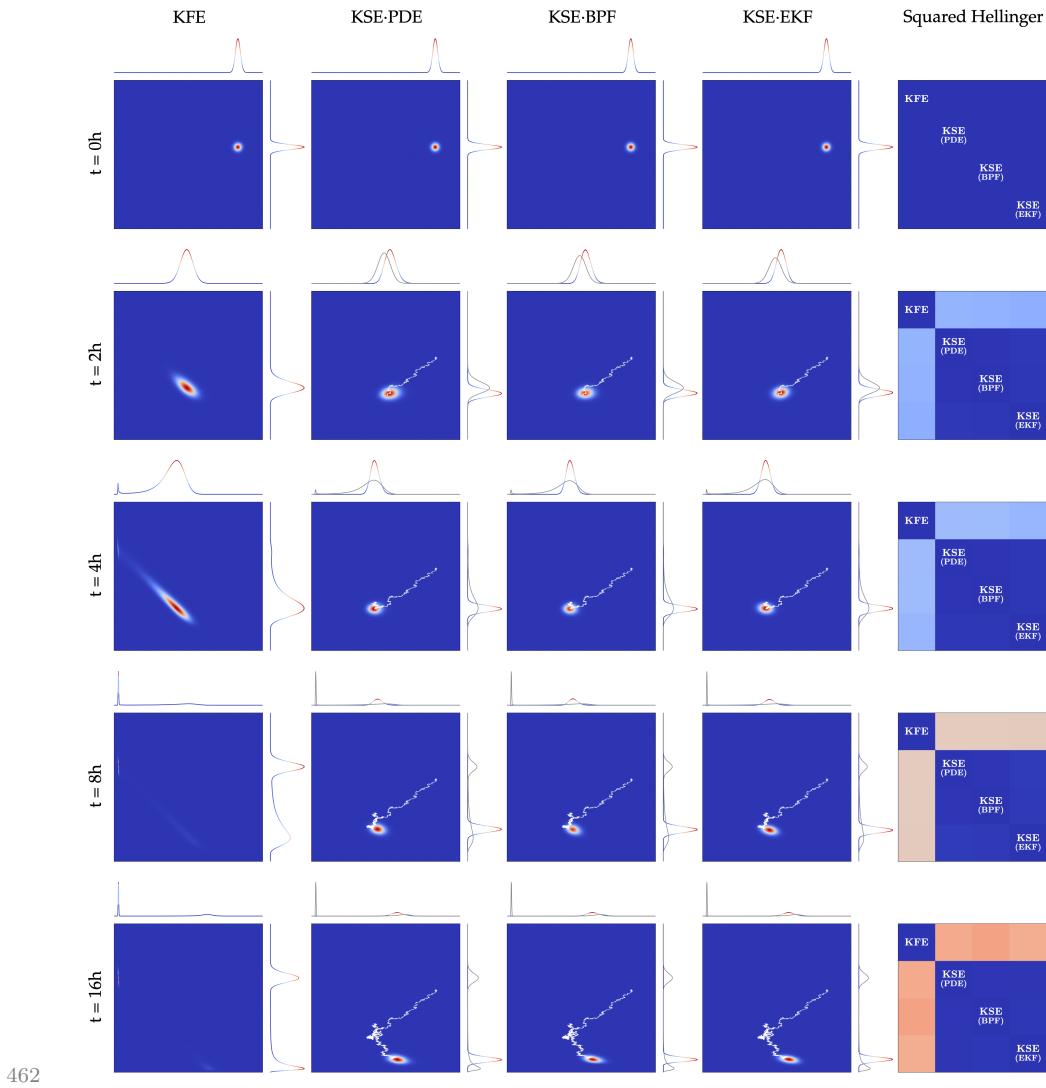
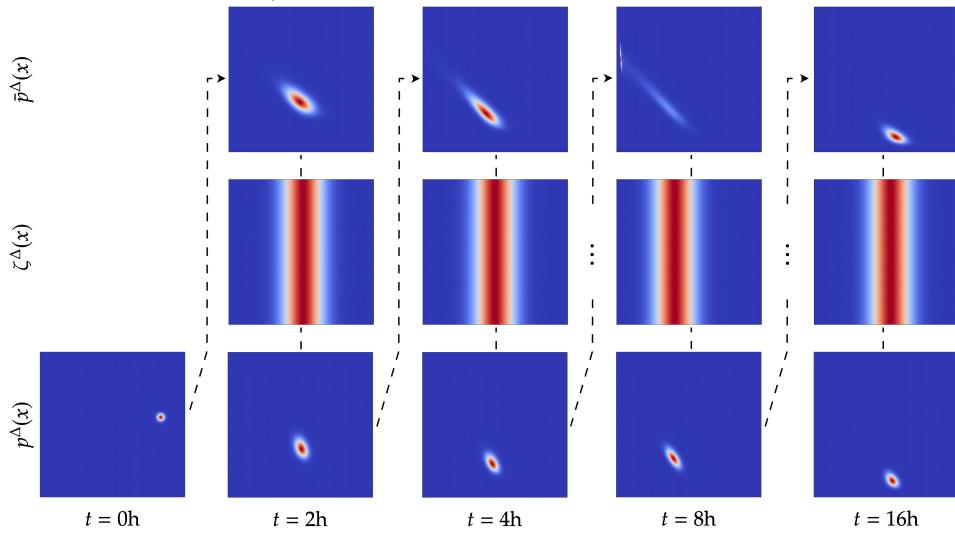


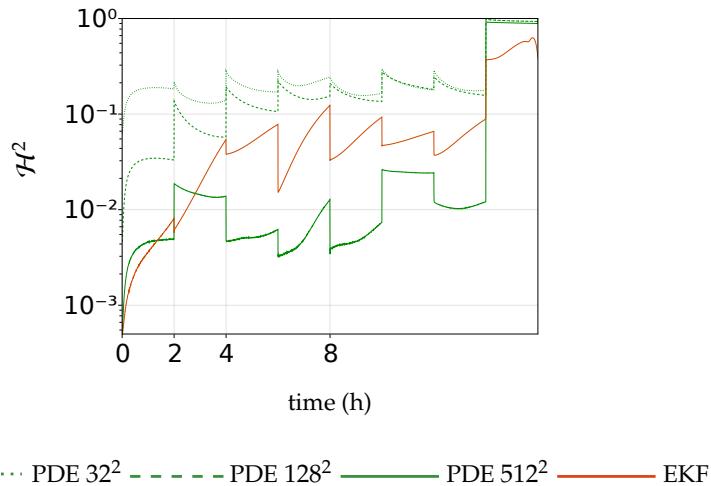
Fig. SM20 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Haldane type with sparse measurements of biogas flow rate. 512^2 grid. $N = 2.5 \times 10^6$.

463

464 **2B - low-frequency (Haldane kinetics, sparse measurements of sub-**
 465 **strate concentration).**



467 **Fig. SM21** Splitting-up approximation for the Kushner-Stratonovich Equation with
 PDE using a 256^2 grid.



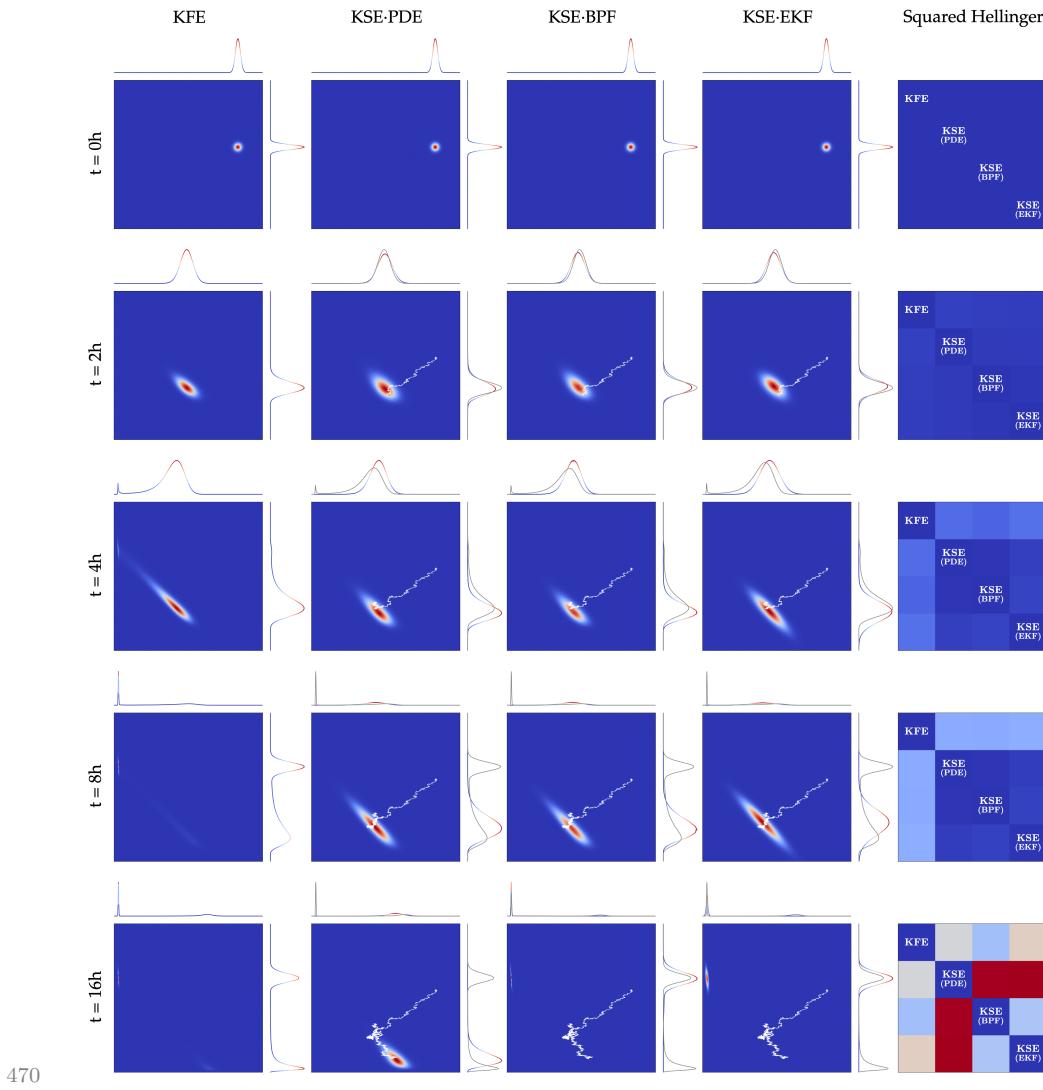
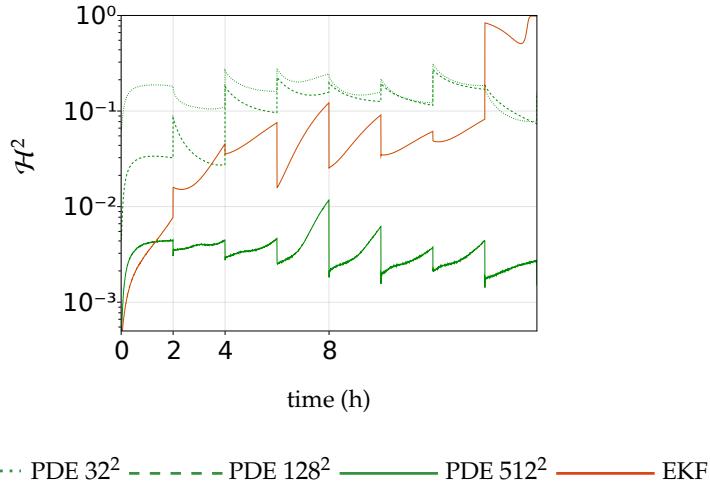


Fig. SM23 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Haldane type with sparse measurements of substrate concentration. 512^2 grid. $N = 2.5 \times 10^6$.

471

472 The distribution of the particles $\{V^j(t)\}_{j=1}^N$ given approximately by the evolution
 473 step is “far” from $\pi_{t_n}^N$ in the sense that the ratio (i.e., the Radon-Nykodym derivative)
 474 of these two distributions generates importance weights with a high variance. One
 475 of the reasons for why this happens comes from our sparse measurements setup: the
 476 dynamics allow particles to disperse far away between measurement times; if N is not
 477 large enough, we will not be able to get a good approximation to Eq. (SM1.11) (the
 478 Kallianpur-Striebel formula).

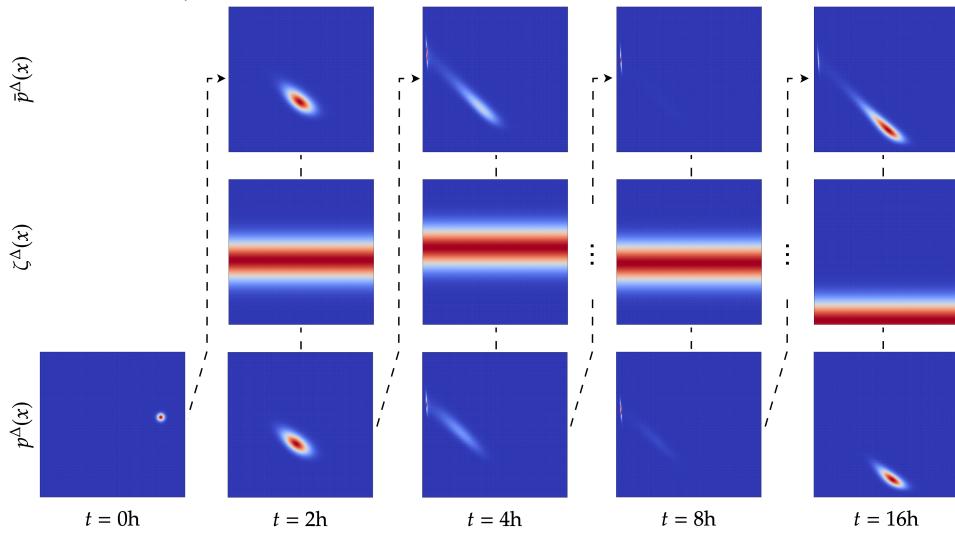
479 To show that a larger number of particles can improve the approximation, we
 480 now compute the Hellinger distance against the BPF with $N = 3.5 \times 10^6$.



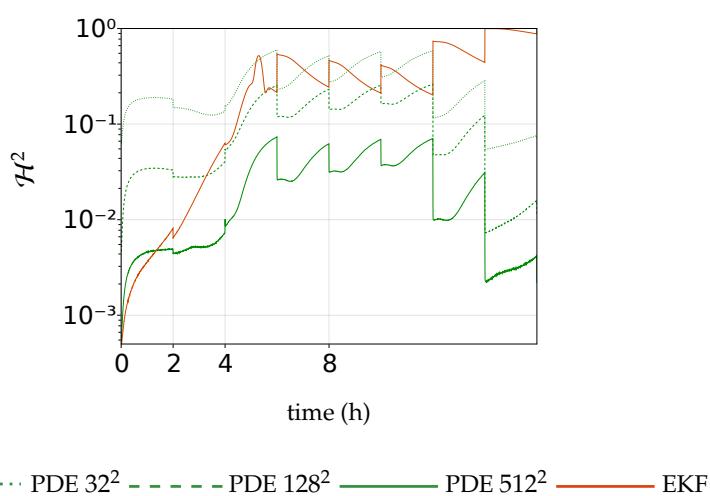
481

Fig. SM24 Squared Hellinger distance for 2B - low-frequency. BPF with $N = 3.5 \times 10^6$. It is clear that the approximation from the particle filter is improved. The resulting approximation is closer to the approximation with methods for PDE with a fine mesh refinement, and the mass of the distribution will be concentrated around
 482 the washout steady state as $t \rightarrow \infty$.

483 **2C - low-frequency (Haldane kinetics, sparse measurements of biomass**
 484 **concentration).**

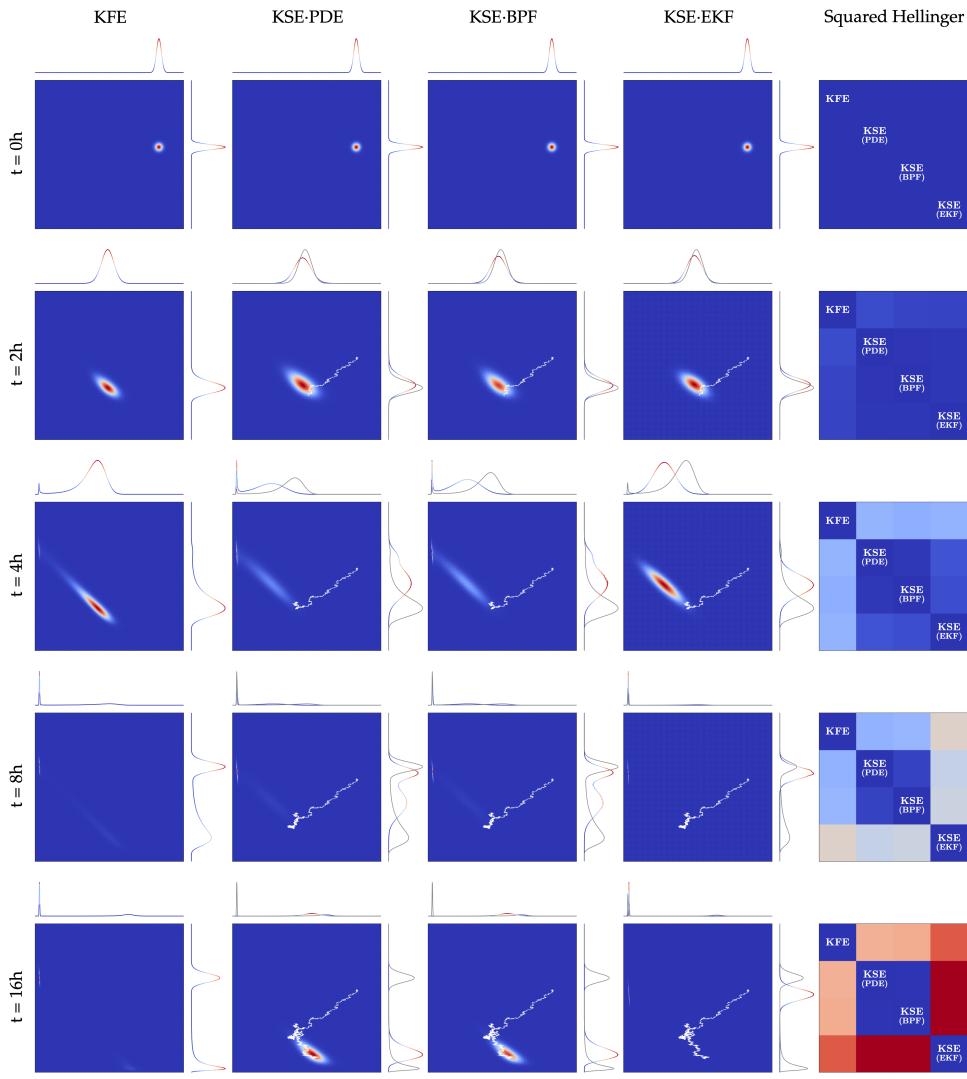


485 **Fig. SM25** Splitting-up approximation for the Kushner-Stratonovich Equation using
 486 a 256^2 grid.



487 **Fig. SM26** Squared Hellinger distance for 2C - low-frequency. BPF: $N = 2.5 \times 10^6$.
 488

SUPPLEMENTARY MATERIALS: KSE FOR THE STOCHASTIC CHEMOSTAT SM39

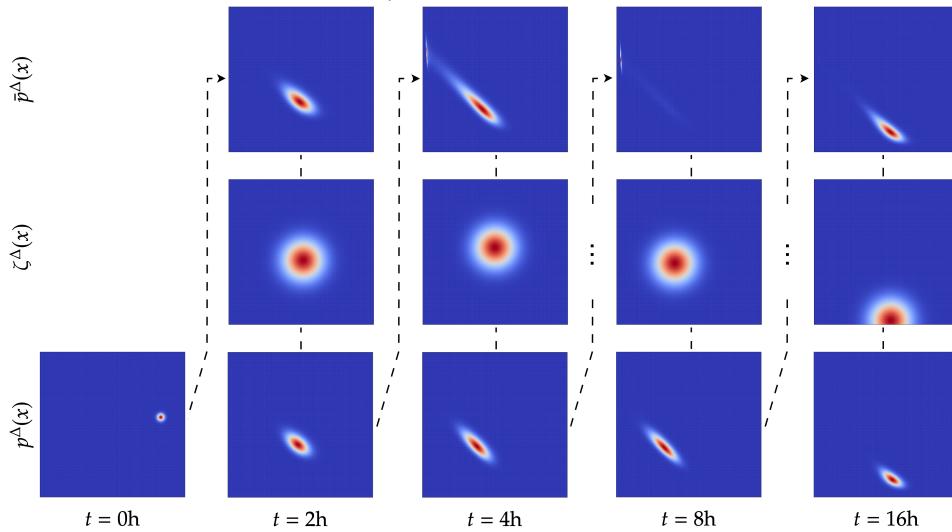


489

Fig. SM27 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Haldane type with sparse measurements of biomass concentration. 512^2 grid. $N = 2.5 \times 10^6$.

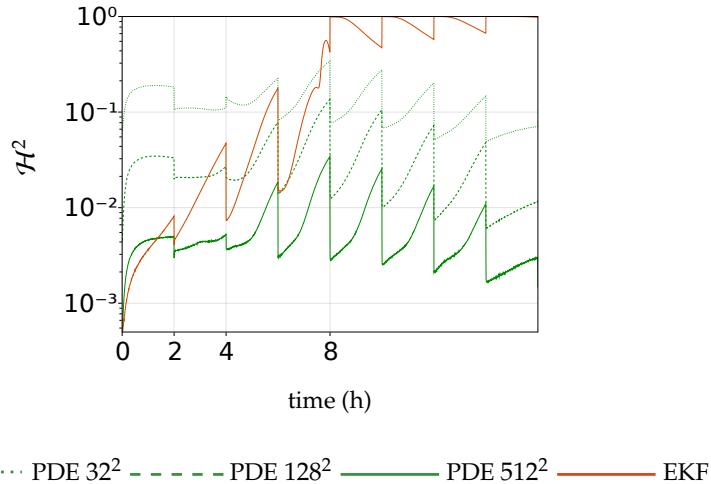
490

491 **2D - low-frequency (Haldane kinetics, sparse measurements of biomass
492 and substrate concentrations).**



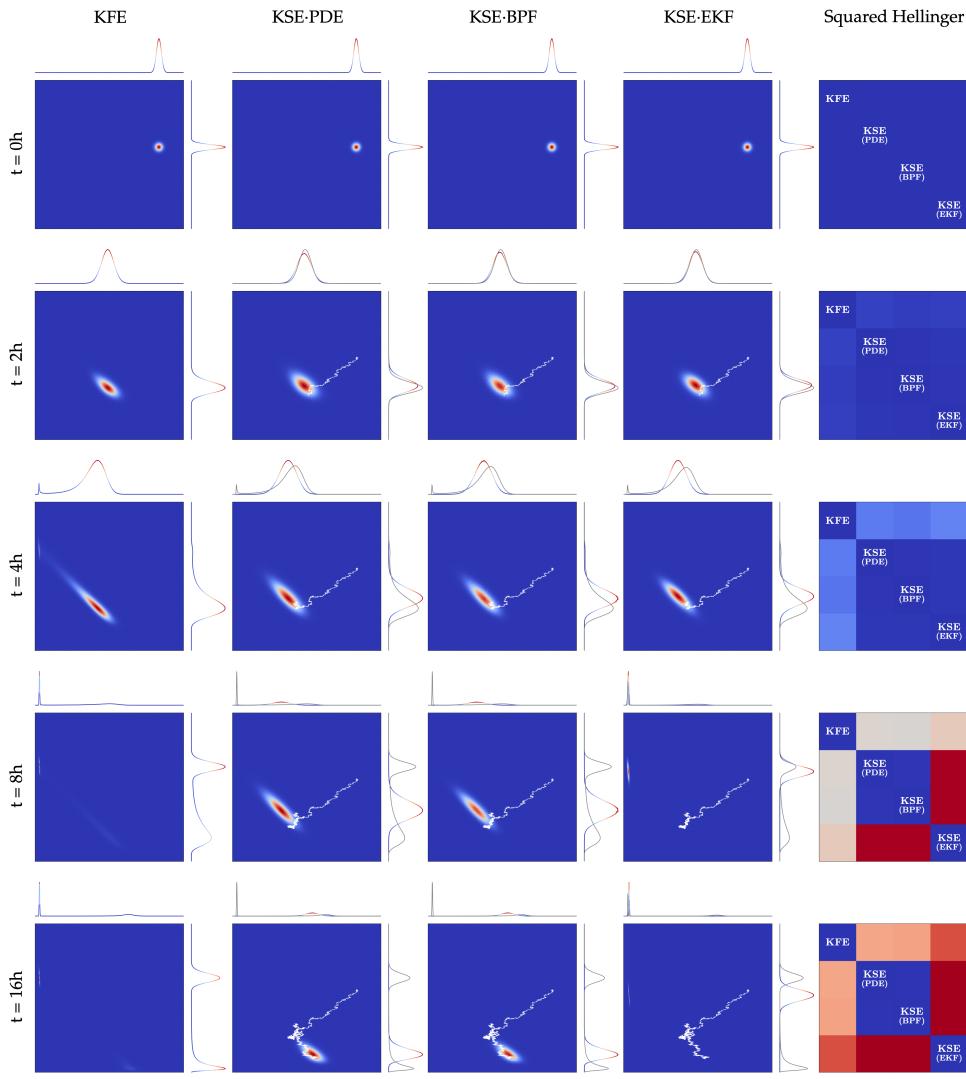
493

Fig. SM28 Splitting-up approximation for the Kushner-Stratonovich Equation using
494 a 256^2 grid.



495

Fig. SM29 Squared Hellinger distance for 2D - low-frequency. BPF: $N = 2.5 \times 10^6$.



497

Fig. SM30 Approximation to the solution of the Kushner-Stratonovich equation for the growth function of the Haldane type with sparse measurements of biomass and substrate concentrations. 512^2 grid. $N = 2.5 \times 10^6$.

498

- 500 [SM1] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability, Springer New York, 2008, <https://doi.org/10.1007/978-0-387-76896-0>.
- 501 [SM2] R. DURRETT, *Stochastic Calculus: A Practical Introduction*, Probability and Stochastics Series, Taylor and Francis, 1996, <https://doi.org/10.1201/9780203738283>.
- 502 [SM3] P. FROGERAIS, J.-J. BELLANGER, AND L. SENHADJI, *Various Ways to Compute the Continuous-Discrete Extended Kalman Filter*, IEEE Transactions on Automatic Control, 57 (2012), pp. 1000 – 1004, <https://doi.org/10.1109/TAC.2011.2168129>.
- 503 [SM4] M. GREWAL AND A. ANDREWS, *Kalman filtering: Theory and practice using matlab*, John Wiley and Sons, (2015), <https://doi.org/10.1002/9781118984987>.
- 504 [SM5] L. IMHOFF AND S. WALCHER, *Exclusion and persistence in deterministic and stochastic chemostat models*, Journal of Differential Equations, 217 (2005), <https://doi.org/10.1016/j.jde.2005.06.017>.
- 505 [SM6] R. E. KÁLMÁN AND R. S. BUCY, *New results in linear filtering and prediction theory*, Journal of Basic Engineering, 83 (1961), <https://doi.org/10.1115/1.3658902>.
- 506 [SM7] R. KHASMINSKII, *Stochastic Stability of Differential Equations*, Stochastic Modelling and Applied Probability, Springer Berlin, Heidelberg, 2012, <https://doi.org/10.1007/978-3-642-23280-0>.
- 507 [SM8] G. KULIKOV AND M. KULIKOVA, *Accurate continuous-discrete unscented kalman filtering for estimation of nonlinear continuous-time stochastic models in radar tracking*, Signal Processing, 139 (2017), pp. 25–35, <https://doi.org/10.1016/j.sigpro.2017.04.002>.
- 508 [SM9] G. KULIKOV AND M. KULIKOVA, *Square-root kalman-like filters for estimation of stiff continuous-time stochastic systems with ill-conditioned measurements*, IET Control Theory & Applications, 11 (2017).
- 509 [SM10] G. Y. KULIKOV AND M. V. KULIKOVA, *The continuous-discrete extended kalman filter revisited*, Russian Journal of Numerical Analysis and Mathematical Modelling, 32 (2017), pp. 27–38, <https://doi.org/DOI:10.1515/rnam-2017-0003>.
- 510 [SM11] J. LOBRY, J. FLANDROIS, G. CARRET, AND A. PAVE, *Monod's bacterial growth model revisited*, Bulletin of Mathematical Biology, 54 (1992), pp. 117–122, [https://doi.org/10.1016/S0092-8240\(05\)80179-X](https://doi.org/10.1016/S0092-8240(05)80179-X).
- 511 [SM12] X. MAO, G. MARION, AND E. RENSHAW, *Environmental brownian noise suppresses explosions in population dynamics*, Stochastic Processes and their Applications, 97 (2002), pp. 95–110, [https://doi.org/10.1016/S0304-4149\(01\)00126-0](https://doi.org/10.1016/S0304-4149(01)00126-0).
- 512 [SM13] S. SÄRKÄÄ, *On unscented kalman filtering for state estimation of continuous-time nonlinear systems*, Automatic Control, IEEE Transactions on, 52 (2007), pp. 1631 – 1641, <https://doi.org/10.1109/TAC.2007.904453>.
- 513 [SM14] R. VAN HANDEL, *Stochastic calculus, filtering, and stochastic control: lecture notes*, 2007. Caltech.
- 514