Vol. 2(Iss. 1) 2025, pp.240-253, African Scientific Annual Review www.asarev.net ISSN: 3007-0902

A Probabilistic Approach to Solving Unknown Linear Stochastic Differential Equations with Bayesian Uncertainty Quantification and Inference

Collins Musera [1*]
Fredrick Onyango[2]
David Angwenyi[3]

[1*] *collinsmusera@gmail.com*

[1*,2] *Department of Statistics and Actuarial Sciences, Maseno University, Kisumu, Kenya.*
[3] *Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya.*

Original Research Article
DOI: https://doi.org/10.51867/Asarev.Maths.2.1.17

Abstract

This work develops a probabilistic framework for solving unknown linear stochastic differential equations (SDEs) by treating the numerical solution as an inference problem. Starting from classical foundations, including the Fokker–Planck and Kushner–Stratonovich formulations, the study derives the evolution of conditional probability densities and presents both the strong and weak forms of the Kushner–Stratonovich equation. Under linear–Gaussian assumptions, the solution remains Gaussian and is fully characterized by its mean and covariance. Deterministic moment equations are derived for state transition and covariance propagation, providing a computationally tractable alternative to full sampling methods. Embracing probabilistic numerics, the framework interprets numerical solvers as inference procedures that return distributions over trajectories, thereby quantifying both process noise and discretization uncertainty. Connections are established with Bayesian probabilistic numerics, and the results demonstrate that moment-based probabilistic solvers deliver calibrated uncertainty for filtering and forecasting tasks. This approach is particularly valuable in engineering, finance, and climate modeling, where reliable uncertainty quantification is critical for decision-making under stochastic dynamics. The paper concludes by recommending robust numerical schemes for the Kushner–Stratonovich equation, comparative studies against Euler–Maruyama and Milstein methods, extensions to partially observed and real-time systems, and validation of the framework using real-world case studies.

MSC2020 Subject Classification: 60H10, 62F15, 62F35
Keywords: Probabilistic modeling; Stochastic differential equations (SDEs); Linear systems; Uncertainty quantification; State estimation; Bayesian inference

1 Introduction

Stochastic differential equations (SDEs) provide a rigorous mathematical framework for modeling dynamical systems subject to random perturbations, with applications across finance, control theory, physics, and biology [14, 5]. An SDE typically combines deterministic drift dynamics with stochastic diffusion terms driven by Brownian motion. The probability density function (PDF) of the solution evolves according to the Kolmogorov forward equation, also known as the Fokker–Planck equation [15]. While this partial differential equation theoretically captures the full probabilistic structure of the state, exact analytical solutions are only available for a narrow class of scalar linear systems.

When closed-form solutions are not obtainable, numerical schemes such as Euler–Maruyama and Milstein are commonly used [10, 1]. These methods approximate sample paths with known convergence rates but return only point estimates of the trajectories, failing to quantify the numerical uncertainty introduced by time discretization [16]. Similarly, analytical solutions of the Fokker–Planck equation are rare outside basic linear–Gaussian models and often require computationally expensive Monte Carlo techniques.

For systems with noisy observations, Bayesian filtering provides a principled approach to state inference. The Kushner–Stratonovich (K–S) equation describes the continuous-time evolution of the conditional PDF of the state given measurements [12, 17]. In linear–Gaussian settings, this reduces to the Kalman–Bucy filter, which propagates exact mean and covariance dynamics [8]. In nonlinear or non-Gaussian systems, approximate filters such as the extended Kalman filter, unscented Kalman filter, and particle filters are widely used [7, 3]. However, these filtering methods generally assume the underlying SDE can be integrated exactly and do not explicitly account for discretization errors.

Recent advances in *probabilistic numerics* reinterpret numerical solution methods as inference procedures. Rather than producing a single deterministic trajectory, these approaches return a probability distribution over possible solutions, thereby quantifying uncertainty due to discretization [4, 2]. Extensions of this idea to SDEs have produced solvers that incorporate both process noise and solver uncertainty [13] by modeling numerical integration as a Bayesian inference problem, where the solver outputs a probability distribution over trajectories rather than a single path. Such frameworks enable rigorous uncertainty quantification, which is critical in filtering, forecasting, and control applications that demand reliable performance under uncertainty.

For linear SDEs driven by Gaussian noise, the first two moments of the state distribution—the mean and covariance—evolve according to deterministic differential equations [8]. This analytical structure makes linear systems an ideal setting for developing and validating probabilistic solution methods. Nevertheless, existing approaches typically compute these moments using deterministic solvers, thereby ignoring the uncertainty introduced by time discretization and failing to deliver a fully probabilistic characterization of the solution, even in this analytically tractable linear–Gaussian setting [9].

2 Problem Statement

Existing numerical and filtering methods for stochastic differential equations (SDEs) are fundamentally limited. They rely on the implicit assumption that the underlying solutions can be computed exactly or with negligible error. In realistic applications, this assumption breaks down because system parameters are frequently unknown, uncertain, or time-varying. As a result, no existing framework simultaneously delivers a complete probabilistic characterization of solutions to unknown linear SDEs, coherently incorporates both process noise and discretization uncertainty in a Bayesian setting, and remains computationally tractable by leveraging the linear–Gaussian structure. Addressing these deficiencies is critical for reliable modeling, filtering, and control in complex stochastic systems. This research directly confronts these challenges by developing a novel probabilistic approach to solving unknown linear SDEs. Unlike previous methods, this framework is the first to unify solution approximation, parameter estimation, and uncertainty quantification within a single Bayesian methodology that is both rigorous and computationally efficient, marking a significant advance over classical deterministic and approximate techniques.

3 Methods

3.1 Model Formulation

We consider a linear SDE of the form:

$$df_t = h(f_t, t)\,dt + u(f_t)\,dt + g(f_t, t)\,d\eta_t, \quad t_0 \le t, \tag{1}$$

where: f_t is the latent state process representing the true but unobserved system trajectory. Here, $h(f_t, t)$ models the underlying system dynamics, $u(f_t)$ accounts for model discrepancy or discretization error, $g(f_t, t)$ determines the strength of stochastic fluctuations, and η_t is a standard Wiener process representing intrinsic noise or unresolved subscale effects. This equation specifies a probability model for the state trajectories by combining the deterministic evolution term with a stochastic forcing term. In a Bayesian setting, prior probability distributions are placed on unknown parameters within h, u, and g, and potentially on the initial state. with observations given by:

$$dy_t = s(f_t, t)\,dt + Q^{1/2}(t)\,d\beta_t. \tag{2}$$

where: y_t denotes the observation vector at time t, $s(f_t, t)$ is the *observation operator*, mapping the latent state f_t into the observation space, possibly involving spatial projection, subsampling, or transformation, $Q^{1/2}(t)$ is the *square-root noise covariance matrix*, defining the amplitude and correlation structure of the measurement noise and β_t is another independent Wiener process, representing observational uncertainty, sensor noise, or external measurement error. The measurement model formalizes the link between the unobserved true dynamics and the available data, acknowledging that observations are *noisy, indirect, and partial*.

The state f_t follows a linear SDE that is linear in f_t with time-varying coefficients, where $h(t)$ and $g(t)$ may vary over time, though the exact solution remains unknown or inaccessible. Our objective

is to build a probability model over the unknown infinite-dimensional solution, reconstructing it from finite-dimensional evaluations and quantifying uncertainty via Bayesian inference. Embracing the approach from probabilistic numerics, we treat the unknown solution itself as a random process, placing a probability measure—typically a Gaussian process prior—over function space. To facilitate tractability, we represent the prior over solutions and potentially latent functions in a state space enabling efficient inference.

3.2 Inference via Bayesian Filtering

Applying Bayesian recursive filtering yields the conditional density $\pi_t(f \mid Y_t)$ of the state given observations up to time t. From this, we derive the evolution via the Kushner–Stratonovich equation, combining dynamics with measurement updates in continuous time. The filtering update incorporates measurement corrections, while the prior propagation mirrors a Fokker–Planck-like evolution for the unobserved diffusion.

3.3 Moments Computation

For linear–Gaussian assumptions, our probabilistic solution of the unknown linear stochastic differential equation yields explicit evolution equations for the mean and covariance. The mean evolves according to a linear ODE governed by the system dynamics, while the covariance satisfies a Riccati-type ODE describing the propagation of process noise. Together, these equations characterize the full Gaussian state posterior.

4 Results and Discussion

We want to obtain the current best estimate of the state f_t given measurements $Y_t = y_{[t_0,t]}$ using the techniques for integrating noisy measurements with noisy dynamics via the Bayesian approach. The development of the conditional density is given in the Kushner–Stratonovich equation.

4.1 Derivation of the Kushner–Stratonovich Equation

The Kushner–Stratonovich equation is an additive perturbation of the Fokker–Planck equation with corrections resulting from measurement knowledge. Bayes' rule is used to quantify the additional information emerging from the measurements. To derive the Kushner–Stratonovich equation we begin by deriving the Fokker–Planck equation.

The Fokker–Planck equation is a partial differential equation (PDE), which describes the time evolution of the PDF associated with the underlying SDE. Consider the scalar Itô stochastic differential equation (1). It can be shown [6] that the process $\{f_t, t \in [t_0, T]\}$ generated by equation (1) is a Markov process. Of interest in Markov processes is the density function:

$$\pi_t(f) = \pi(f, t), \quad \forall t \in [t_0, T], \tag{3}$$

together with the transition probability density function:

$$\pi_{t|\tau}(f|z) = \pi\big((f,t) \mid z,\tau\big), \quad \forall \tau < t \in [t_0, T], \tag{4}$$

which both characterize $\{f_t, t \in [t_0, T]\}$. For all $t > \tau \in [t_0, T]$, the Fokker–Planck equation is the evolution equation of the density function $\pi_t(f)$ and the conditional density $\pi_{t|\tau}(f|z)$. Because the process $\{f_t, t > 0\}$ generated by equation (1) is a Markov process, then, given $t_1 < t_2 < t_3$,

$$\pi_{t_3|t_1,t_2}(f|z,y) = \pi_{t_3|t_2}(f|y), \tag{5}$$

and the following equation holds true:

$$\pi_{t_3|t_1}(f|z) = \int \pi_{t_3|t_2}(f|y)\, \pi_{t_2|t_1}(y|z)\, dy. \tag{6}$$

Equation (6) is referred to as the Chapman–Kolmogorov equation, and it applies to all Markov processes. Using this equation and a Taylor expansion, an equation for the development of the transition probability density function may be obtained (see [6] for derivation):

$$\frac{\partial \pi_{t|\tau}(f|y)}{\partial t} = -\frac{\partial}{\partial f}\Big(\pi_{t|\tau}(f|y)\, h(f,t)\Big) + \frac{1}{2}\frac{\partial^2}{\partial f^2}\Big(\pi_{t|\tau}(f|y)\, g^2(f,t)\Big) \;\pm\; \frac{\partial}{\partial f}\Big(\pi_{t|\tau}(f|y)\, u(f)\Big), \tag{7}$$

Taking expectation in (7) with respect to $\pi_\tau(y)$ and observing that

$$\mathbb{E}_\tau\big[\pi_{t|\tau}(f|y)\big] = \int \pi_{t|\tau}(f|y)\, \pi_\tau(y)\, dy = \pi_t(f), \tag{8}$$

we obtain the evolution equation for $\pi_t(f)$:

$$\frac{\partial \pi_t(f)}{\partial t} = -\frac{\partial}{\partial f}\Big(\pi_t(f)\, h(f,t)\Big) - \frac{\partial}{\partial f}\Big(\pi_t(f)\, u(f)\Big) + \frac{1}{2}\frac{\partial^2}{\partial f^2}\Big(\pi_t(f)\, g^2(f,t)\Big). \tag{9}$$

For the vector case, evolving each component yields

$$\begin{aligned}\frac{\partial \pi_t(f,t)}{\partial t} = &-\sum_{k=1}^{n} \frac{\partial}{\partial f_k}\Big(\pi_t(f,t)\, h_k(f,t)\Big) - \sum_{k=1}^{n} \frac{\partial}{\partial f_k}\Big(\pi_t(f,t)\, u_k(f)\Big) \\ &+ \frac{1}{2}\sum_{k,l=1}^{n} \frac{\partial^2}{\partial f_k\, \partial f_l}\Big(\pi_t(f,t)\, \big(g(f,t) g^\top(f,t)\big)_{kl}\Big) =: \xi\, \pi_t(f,t),\end{aligned} \tag{10}$$

where

$$\xi\,\pi_t(f,t) = -\sum_{k=1}^{n}\frac{\partial}{\partial f_k}\Big(\pi_t(f,t)\,h_k(f,t)\Big) - \sum_{k=1}^{n}\frac{\partial}{\partial f_k}\Big(\pi_t(f,t)\,u_k(f)\Big) + \frac{1}{2}\sum_{k,l=1}^{n}\frac{\partial^2}{\partial f_k\,\partial f_l}\Big(\pi_t(f,t)\,(gg^\top)_{kl}\Big). \tag{11}$$

The probability density function $\pi_t(f)$ is obtained by solving the Fokker–Planck equation (10). Analytical solutions exist for basic scalar systems [17], but most SDEs, particularly those including nonlinear drift factors, are difficult to solve analytically; numerical solutions are therefore preferred. With these results, we now derive the Kushner–Stratonovich equation, which serves as the foundation for our approach.

Theorem 1 (Kushner–Stratonovich evolution)**.** *Suppose that h, g and s are real-valued, uniformly Lipschitz in f on $[t_0,T]$, and bounded on bounded sets; suppose also that f_{t_0} is finite in the mean-square sense and independent of $\{d\eta_t\}$. For the system (1)–(2), the conditional density evolves as*

$$\pi_t(f \mid Y_t) = \pi_{t_0}(f) + \int_{t_0}^{t}\xi\big(\pi_\tau(f \mid Y_\tau)\big)\,d\tau + \int_{t_0}^{t}\pi_\tau(f \mid Y_\tau)\,\big(s(f,\tau)-\hat{s}_\tau\big)^\top Q^{-1}(\tau)\,\big(dy_\tau - \hat{s}_\tau\,d\tau\big), \tag{12}$$

where the adjoint operator is $\xi^* = h^\top\nabla(\cdot) + \frac{1}{2}\operatorname{tr}\big(gg^\top\Delta(\cdot)\big)$, and

$$\nabla = \begin{bmatrix}\frac{\partial}{\partial f_1}\\ \frac{\partial}{\partial f_2}\\ \vdots\\ \frac{\partial}{\partial f_n}\end{bmatrix}, \qquad \Delta = \begin{bmatrix}\frac{\partial^2}{\partial f_1^2} & \frac{\partial^2}{\partial f_1\partial f_2} & \cdots & \frac{\partial^2}{\partial f_1\partial f_n}\\ \frac{\partial^2}{\partial f_2\partial f_1} & \frac{\partial^2}{\partial f_2^2} & \cdots & \frac{\partial^2}{\partial f_2\partial f_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2}{\partial f_n\partial f_1} & \frac{\partial^2}{\partial f_n\partial f_2} & \cdots & \frac{\partial^2}{\partial f_n^2}\end{bmatrix}, \qquad \hat{s}_t = \int s(f,t)\,\pi_t(f \mid Y_t)\,df. \tag{13}$$

Equation (12) is called Kushner-Stratonovich equation. In the absence of observations, that is, when $Q^{-1}(t) \equiv 0$, we obtain the Fokker-Planck equation.

Proof. We provide a sketch of the proof. To make things easier, we rewrite equation (12) in differential form:

$$d\pi_t = \xi(\pi_t)dt + \pi_t(s-\hat{s}_t)^T Q^{-1}(t)(dy_t - \hat{s}_t dt) \tag{14}$$

and remember that the conditional density, π_t, is a notationally an easy way of expressing $\pi_t(f \mid Y_t)$- that is. the conditional probability density function of the state,f_t , at time, t, given all measurements from the start time, t_0, to the time of interest, t. Following an infinitesimal change in time, δt, an infinitesimal measurement, δY_t, is obtained using the measurements equation. equation (2); similarly. an infinitesimal change in state, δf_t is realized. The proof then takes into account the infinitesimal change in the conditional density function, $\delta\pi_t$, caused by fresh information on both the signal and

the measurements, which may be stated as follows:

$$\begin{aligned}\delta\pi_t =& \mid \delta\pi_t \mid_{dynamic} + \mid \delta\pi_t \mid_{measurements},\\ &= [\pi_{t+\delta t}(f \mid Y_t + \delta y_t) - \pi_t(f \mid Y_t + \delta y_t)] + [\pi_t(f \mid Y_t + \delta y_t) - \pi_t(f \mid Y_t)]\end{aligned} \tag{15}$$

We note that assuming limitations in equation (15) as $\delta t \to 0$ results in

$$\mid \delta\pi_t \mid_{dynamic} = \xi(\pi_t)dt \tag{16}$$

and

$$\begin{aligned}\mid \delta\pi_t \mid_{measurements} &= [\pi_t(f \mid Y_t + \delta y_t) - \pi_t(f \mid Y_t)]\\ &= \pi_t(f)(s(f,t) - \hat{s_t})^T Q^{-1}(t)(dy_t - \hat{s_t}dt)\end{aligned} \tag{17}$$

Because the Fokker-Planck equation is already provided in equation (16), all that is required to be established is equation (17). We may get an expression for an infinitesimal measurement from the measurement's equation, equation (2).

$$\delta y_t = s(f_t)\delta t + Q^{1/2}(t)\delta\beta_t, \quad t_0 \le t \tag{18}$$

and take note of

$$\mathbb{E}[\delta y_t \delta y_t^T] = Q(t)\delta t + o(\delta t) \tag{19}$$

When infinitesimal measurement is available, δy_t, we obtain a new value for the transition density $\pi_t(f \mid Y_t + \delta y_t)$, which is provided by [11].

$$\pi_(f \mid Y_t + \delta y_t) = \frac{\pi_t(\delta y \mid f_t)\pi_t(f \mid Y_t)}{\int \pi_t(\delta y \mid f_t)\pi_t(f \mid Y_t)df} \tag{20}$$

Where

$$\pi_t(\delta y \mid f_t) \sim \mathcal{N}(s(f_t)\delta t, Q(t)\delta t) \tag{21}$$

We then define a function $p(\delta t, \delta y_t)$ as follows

$$p(\delta t, \delta y_t) = \frac{\pi_t(f \mid Y_t + \delta y_t)}{\pi_t(f \mid Y_t)} \tag{22}$$

Then, using Taylor series expansion. we estimate $p(\delta t, \delta y_t)$ about (0,0) up to the first derivative with respect to δt and the second derivative with respect to δy_t, while ignoring terms containing mixed derivatives. The terms $\delta y_t \delta y_t^T$ are substituted with their expected values, yielding.

$$p(\delta t, \delta y_t) = 1 + (s - \hat{s_t})^T Q^{-1}(\delta y_t - \hat{s_t}\delta t) + o(\delta t) \tag{23}$$

From equations (22) and (23).

$$\pi_t(f \mid Y_t + \delta y_t) = \pi_t(f \mid Y_t) + (s - \hat{s_t})^T Q^{-1}(\delta y_t - \hat{s_t}\delta t)\pi_t(f \mid Y_t) + o(\delta t) \tag{24}$$

By going to the formal limit as $\delta t \to 0$, we have the required equation (17). □

4.2 The weak form of the Kushner-Stratonovich Equation

If $\phi(f)$ satisfies the conditions of the weak form of the Fokker-Planck equation, the weak form of the Kushner-Stratonovich equation can be obtained by integration by parts.

$$\begin{aligned}\int \partial\pi_t\phi(f)\partial f = &-\int \frac{\partial\pi_t h(f,t)}{\partial f}\phi(f)\partial t\partial f - \int \frac{\partial\pi_t u(f_t)}{\partial f}\phi(f)\partial t\partial f + \frac{1}{2}\frac{\partial^2(\pi_t g^2(f,t)}{\partial f^2}\phi(f)\partial t\partial f \\ &+ \int \phi(f)\pi_t(s-\hat{s}_t)Q^{-1}(t)(\partial y_t - \hat{s}_t\partial t)\partial f\end{aligned} \tag{25}$$

By integrating equation (25) by parts, we get

$$\begin{aligned}\int \partial\pi_t\phi(f)\partial f = &\frac{\partial\phi(f)}{\partial f}\pi_t h(f,t)\partial t\partial f + \frac{\partial\phi(f)}{\partial f}\pi_t u(f_t)\partial t\partial f + \frac{1}{2}\int \frac{\partial^2\phi(f)}{\partial f}\pi_t g^2(f,t)\partial t\partial f \\ &+ \int \pi_t(s-\hat{s}_t)Q^{-1}(t)(\delta y_t - \hat{s}_t\partial t)\phi(f)\partial f\end{aligned} \tag{26}$$

When equation (26) is expressed in terms of expectations the result is

$$\begin{aligned}d\mathbb{E}[\phi(f)] = &\mathbb{E}\left[\frac{\partial\phi(f)}{\partial f}h(f,t)\right]dt + \mathbb{E}\left[\frac{\partial\phi(f)}{\partial f}u(f_t)\right]dt + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2\phi(f)}{\partial f}g^2(f,t)\right]dt \\ &+ ((\mathbb{E}[\phi(f)s] - \hat{\phi}\hat{s}_t)Q^{-1}(t)(\delta y_t - \hat{s}_t dt))\end{aligned} \tag{27}$$

The vector version of equation (27) can be constructed as,

$$d\pi_t[\phi] = \pi_t[[\xi^*]dt + (\pi_t[\phi(f)s] - \hat{\phi}\hat{s}_t)^T Q^{-1}(t)(\delta y_t - \hat{s}_t)] \tag{28}$$

this is the weak form of K-S equation.

4.3 Moments of the Probabilistic Solution

In the previous section, we derived the Fokker–Planck Equation (FPE), which, in principle, offers a complete probabilistic characterization of the state of a stochastic system. This partial differential equation governs the time evolution of the probability density function (PDF) of the system's state. From the solution of the FPE, one can compute various statistical properties of the system, such as the mean, covariance, and higher-order moments.

In particular, the first and second statistical moments of the state distribution namely the mean and covariance are of special interest. These moments summarize the central tendency and the spread of the state distribution, respectively, and are often sufficient for characterizing the behavior of systems governed by Gaussian processes or for linear systems with additive Gaussian noise.

A moment of a random variable refers to the expected value of some function of that variable. If $\phi(f)$,

is a function of the state variable f, the moment associated with ϕ is defined as:

$$\mathbb{E}[\phi(f)] = \int \phi(f)p(f,t)df \tag{29}$$

where $p(f,t)$ is the probability density function of the state at time t, as determined by the solution to the FPE. To illustrate this, consider a linear stochastic differential equation (SDE) describing the evolution of the state $f(t)$:

$$df_t = h(f_t,t)dt + u(f_t)dt + g(f_t,t)d\eta_t \tag{30}$$

where the initial conditions are $h(t_0) \sim N(m_0, C_0)$, $h(f_t,t)$ and $g(f_t,t)$ are matrix valued functions of time, $u(f_t)$ is a vector valued function of time and $d\eta_t$ is a Brownian motion with diffusion matrix Q. The mean and covariance can be solved from equations (31) and (32)

$$\frac{dm}{dt} = E[h(f_t,t)] \tag{31}$$

$$\frac{dC}{dt} = E[h(f_t,t)(f-m)^T] + E[(f-m)h^T(f_t,t)] + E[g(f_t,t)Qg^T(f_t,t)] \tag{32}$$

which are the differential equations for mean and covariance of the state, the equations provide a useful starting point for forming Gaussian approximations to stochastic differential equations. Using equation (30) the mean of the state is given by

$$\frac{dm(t)}{dt} = h(t)m(t) + u(t) \tag{33}$$

and using equation (31) the covariance of the state is given by

$$\frac{dC(t)}{dt} = h(t)C(t) + C(t)h^T(t) + g(t)Qg^T(t) \tag{34}$$

with the initial conditions $m(t_0) = m_0$ and $C(t_0) = C_0$. The general solutions to these differential equations are

$$\begin{aligned} m(t) &= \psi(t,t_0)m(t_0) + \int_{t_0}^{t} \psi(t,\tau)u(\tau)d\tau \\ C(t) &= \psi(t,t_0)C(t_0)\psi^T(t,t_0) + \int_{t_0}^{t} \psi(t,\tau)g(\tau)Qg^T(\tau)(t,\tau)d\tau \end{aligned} \tag{35}$$

Because the solution is linear transformation of the Brownian motion, which is a Gaussian process, the solution is also Gaussian

$$p(f,t) = N(f(t)m(t), C(t)) \tag{36}$$

This can be verified by checking that the probability distribution solves the corresponding Fokker-Planck-Kolmogorov equation. The evolution of the mean characterizes the deterministic dynamics of the system, while the covariance captures the growth of uncertainty over time due to process noise.

These moment equations are linear and deterministic, evolving independently of the specific sample paths of the stochastic process. They form the foundation of many filtering and control algorithms most notably the Kalman filter where they are used to propagate uncertainty in the system state in the absence of measurements.

4.4 Application

To illustrate the practical value of the probabilistic solution, consider a bank pricing a 5-year interest rate swap where the short rate follows a Vasicek SDE with uncertain parameters. The Vasicek model describes the evolution of the short-term interest rate r_t as a mean-reverting stochastic differential equation (SDE):

$$dr_t = a(b - r_t)\,dt + \sigma\,dW_t,$$

where $a > 0$ is the speed of mean reversion, b is the long-term equilibrium rate, $\sigma > 0$ is the volatility, and W_t is a standard Wiener process. This model captures the tendency of interest rates to revert toward a long-term mean while allowing for random fluctuations due to market uncertainty.

Standard solvers such as Euler–Maruyama produce a single simulated rate path, yielding a swap valuation that appears precise but conceals discretization error. Using our probabilistic approach, the short rate is represented as a Gaussian distribution rather than a single trajectory. The mean captures the expected interest rate dynamics, while the covariance quantifies confidence in this forecast, incorporating both market volatility and solver uncertainty. As a result, the bank obtains swap prices with calibrated uncertainty bounds enabling robust bid/ask spreads, improved hedging strategies, and more reliable risk assessment, all without relying on large-scale Monte Carlo simulations.

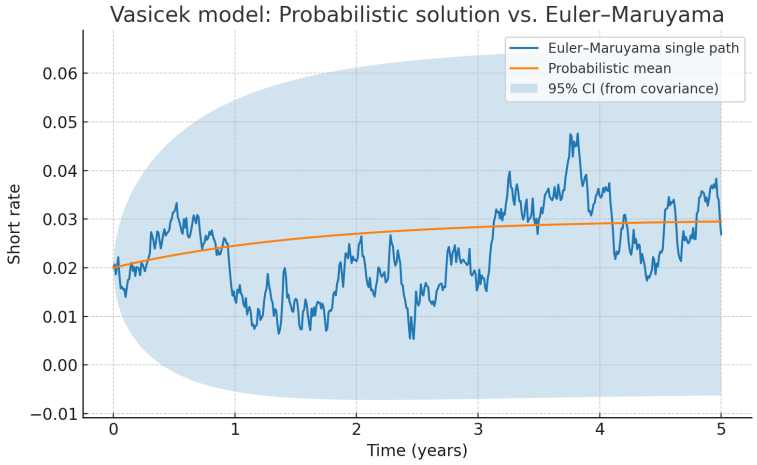

Figure 1: Comparing a single Euler–Maruyama path with the probabilistic solution's for a Vasicek short-rate model.

Figure 1 compares a single Euler–Maruyama trajectory with the probabilistic solution of the Vasicek interest rate model. The Euler–Maruyama method produces a single, noisy path of the short rate,

while the probabilistic solver represents the rate as a Gaussian process whose mean follows the nominal dynamics and whose covariance explicitly tracks uncertainty over time. The shaded 95 percent confidence band shows that this approach captures both process noise and numerical uncertainty in a principled way, rather than providing just one realization of the dynamics.

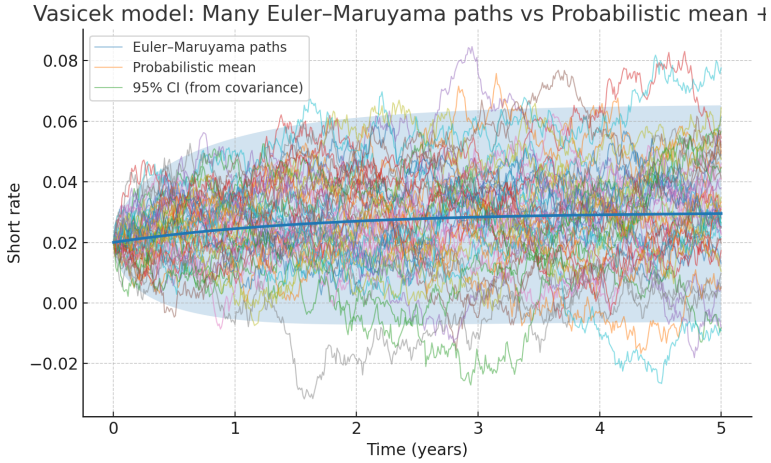

Figure 2: Comparing many Euler–Maruyama paths with the probabilistic solution's for a Vasicek short-rate model.

Figure 2 extends this comparison by plotting many Euler–Maruyama sample paths against the same probabilistic solution. The ensemble of Monte Carlo trajectories spreads out over time, reflecting stochastic variability, but their empirical mean and dispersion are well-matched by the deterministic mean and covariance produced by the probabilistic method. The confidence interval generated by solving the moment equations accurately envelopes nearly all simulated paths, confirming that the solver provides calibrated uncertainty without resorting to large-scale Monte Carlo simulation.

Together, these results demonstrate that the probabilistic solution not only recovers the correct dynamics but also provides a computationally efficient and principled way to quantify uncertainty. Unlike classical solvers that require many samples to approximate variability, this method outputs full distributional information directly, making it valuable for applications such as risk management, derivative pricing, and hedging in fixed-income markets.

5 Conclusion and Recommendations

5.1 Conclusion

This study formulated a probabilistic framework for solving unknown linear stochastic differential equations (SDEs) by integrating Bayesian inference with continuous-time filtering. Using the Fokker–Planck equation and its extension to the Kushner–Stratonovich equation, we established a coherent approach to characterize the conditional density of the system state, quantify uncertainty, and compute key

statistical moments such as the mean and covariance. The derived moment equations demonstrate that the state evolution under linear–Gaussian assumptions remains analytically tractable and leads naturally to Gaussian approximations for the probabilistic solution. This unified approach provides a foundation for simultaneously solving linear SDEs, estimating unknown parameters, and propagating uncertainty.

5.2 Recommendations

Based on the findings of this study, it is recommended to adopt moment-based probabilistic solvers for linear stochastic differential equations, as the evolution of the mean and covariance can be captured by deterministic equations, enabling efficient uncertainty propagation. Furthermore, the Kushner–Stratonovich equation, which fully characterizes the conditional density of the system state, should be implemented using robust numerical schemes that preserve its probabilistic interpretation. Comparative studies against classical deterministic solvers such as Euler–Maruyama and Milstein methods are necessary to assess the additional benefits of the probabilistic approach, particularly in terms of uncertainty quantification. The methodology should also be extended to partially observed systems, where only noisy measurements are available, to demonstrate its utility in real-time filtering applications. Finally, applying the proposed framework to real-world case studies in engineering, finance, or climate modeling would validate its practical relevance and highlight its potential for quantifying uncertainty in complex stochastic systems.

References

[1] Burrage, K., & Burrage, P. M. (1996). High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations. *Applied Numerical Mathematics*, *22*, 81–101.

[2] Cockayne, J., Oates, C., Sullivan, T., & Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM Review*, *61*(4), 756–789.

[3] Doucet, A., de Freitas, N., & Gordon, N. J. (Eds.). (2001). *Sequential Monte Carlo methods in practice*. Springer.

[4] Hennig, P., Osborne, M., & Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A*, *471*(2179), 20150142.

[5] Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, *43*(3), 525–546.

[6] Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.

[7] Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, *92*(3), 401–422.

[8] Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108.

[9] Kersting, H., & Hennig, P. (2016). Active uncertainty calibration in Bayesian ODE solvers. In *Advances in Neural Information Processing Systems*.

[10] Kloeden, P. E., & Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer.

[11] Kushner, H. J. (1962). On the differential equations satisfied by conditional probability densities of Markov processes, with applications. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, *2*(1), 106–119.

[12] Kushner, H. J. (1967). Dynamical equations for optimal nonlinear filtering. *Journal of Differential Equations*, *3*(2), 179–190.

[13] Le Fay, Y., Särkkä, S., & Corenflos, A. (2025). Modelling pathwise uncertainty of stochastic differential equations samplers via probabilistic numerics. *Bayesian Analysis*. Advance online publication. https://doi.org/10.1214/25-ba1522

[14] Øksendal, B. (2013). *Stochastic differential equations: An introduction with applications* (6th ed.). Springer.

[15] Risken, H. (1996). *The Fokker–Planck equation: Methods of solution and applications* (2nd ed.). Springer.

[16] Särkkä, S., & Solin, A. (2019). *Applied stochastic differential equations*. Cambridge University Press.

[17] Stratonovich, R. L. (1968). Conditional Markov processes. *Theory of Probability and Its Applications*, *13*(1), 156–178.