Statistics and Computing (2026) 36:2
https://doi.org/10.1007/s11222-025-10736-1

ORIGINAL PAPER



Importance sampling for rare event tracking within the ensemble Kalman filtering framework

Nadhir Ben Rached[1] · Erik von Schwerin[2] · Gaukhar Shaimerdenova[2,3] · Raúl Tempone[2,4]

Received: 1 August 2024 / Accepted: 21 September 2025
© The Author(s) 2025

Abstract

In this work we employ importance sampling (IS) techniques to track a small over-threshold probability of a running maximum associated with the solution of a stochastic differential equation (SDE) within the framework of ensemble Kalman filtering (EnKF). The proposed method acts as a post-processing step applied to the EnKF output: it uses the ensemble at a given observation time to estimate the probability of a rare event occurring before the next observation, without altering the EnKF itself. Between two observation times of the EnKF, we propose to use IS with respect to the initial condition of the SDE, IS with respect to the Wiener process via a stochastic optimal control formulation, and combined IS with respect to both initial condition and Wiener process. Both IS strategies require the approximation of the solution of Kolmogorov Backward equation (KBE) with boundary conditions. In multidimensional settings, we employ a Markovian projection dimension reduction technique to obtain an approximation of the solution of the KBE by just solving a one dimensional PDE. The proposed ideas are tested on three illustrative examples: Double Well SDE, Langevin dynamics and noisy Charney-deVore model, and showcase a significant variance reduction compared to the standard Monte Carlo method and another sampling-based IS technique, namely, multilevel cross entropy.

Keywords Monte Carlo · ensemble Kalman filter · importance sampling · rare event simulation · stochastic optimal control

Mathematics Subject Classification 35Q93 · 60G35 · 60H35 · 65C05 · 93E20

1 Introduction

Let $\{u_t\}_{t\geq 0} \in \mathbb{R}^d$ be a Markov process which determines the state of the system arising from a stochastic dynamics model. In particlular, we assume the model dynamics is associated with a stochastic differental equation (SDE) driven by the Wiener process. Suppose also we are given partially observed observations $\{y_n\}_{n\in\mathbb{N}}$ arriving sequentially at discrete time instances $\{t_n\}_{n\in\mathbb{N}}$. We wish to approximate the expected value of a quantity of interest $\varphi : \mathbb{R}^d \to \mathbb{R}$ applied to $u_n := u_{t_n}$, taking into account the accumulated observation data $Y_n := (y_1, y_2, ..., y_n)$ up to time t_n, i.e., we seek to approximate $\mathbb{E}[\varphi(u_n)|Y_n]$.

Data assimilation is a powerful methodology to combine the observation data with the model forecast to produce more reliable estimates of true states of the system. There are various data assimilation techniques with their strengths and limitations. Among them, the ensemble Kalman filter (EnKF) has quickly become popular for performing inference online in many applications due to its simplicity in implementation

✉ Gaukhar Shaimerdenova
gaukhar.shaimerdenova@kaust.edu.sa

Nadhir Ben Rached
n.benrached@leeds.ac.uk

Erik von Schwerin
erik.vonschwerin@kaust.edu.sa

Raúl Tempone
raul.tempone@kaust.edu.sa

[1] School of Mathematics, Faculty of Engineering and Physical Sciences, University of Leeds, Leeds, UK

[2] Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia

[3] School of Artificial Intelligence and Data Science, Astana IT University, Astana, Kazakhstan

[4] Chair of Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany

Published online: 26 October 2025

and efficiency in high dimensional problems (Evensen 1994; Kalnay 2003; Aanonsen et al. 2009; Houtekamer et al. 2005; Reich and Cotter 2015). However, the EnKF is impractical when the system exhibits very unlikely, unusual fluctuations, i.e., in failure/rare event occurrences. Suppose we want to estimate the probability of a failure event of a dynamic system between two observation times t_n and t_{n+1} before obtaining the next measurement at time t_{n+1}. For example, we may want to detect a hazard for an aircraft caused by extreme turbulence given radar observations (Gao and Tembine 2016), predict a financial crisis using stock market data (Karmiani et al. 2019), or forecast a severe weather event provided different sources of weather measurements (Jones et al. 2015; Zhang et al. 2015; Maejima and Miyoshi 2020).

Objective: In this work, we are concerned with a particular type of rare event. More precisely, we want to track the probability of the running maximum of the projected one-dimensional stochastic process exceeding a critical threshold $\mathcal{K} > 0$ until the next observation time t_{n+1} within the framework of the EnKF given the data Y_n:

$$\alpha_n := \mathbb{P}\Big(\max_{t_n \le t \le t_{n+1}} P_1 u_t \ge \mathcal{K} \Big| Y_n\Big),$$

where $P_1 \in \mathbb{R}^{1\times d}$ is a pre-defined projection applied to $u_t \in \mathbb{R}^d$.

Since the EnKF is a Monte Carlo (MC)-based filtering method, the estimation of a very small over-threshold probability by EnKF ensembles is computationally expensive for rare event regimes $\alpha_n \ll 1$. Various variance reduction techniques exist to improve the efficiency of the MC simulation. One famous technique to reduce the variance is importance sampling (IS). It is based on the idea of changing the original probability density to another biasing density such that the rare event is more likely to occur (Kroese et al. 2013; Asmussen and Glynn 2007; Rached et al. 2016). To ensure the unbiasedness of the new IS estimator, one must reweight each sample by an appropriate likelihood ratio. However, a key challenge in IS is to select a good choice of biasing density, which results in a significantly smaller variance of the estimator compared to the one provided by the standard MC.

In this work, we incorporate IS techniques into the EnKF framework to enable tracking of rare event probabilities in real time. Specifically, we propose three IS strategies that can be applied at each observation time t_n as a post-processing step, that is, using the output of the EnKF without modifying its internal filtering procedure. These strategies include: IS with respect to (wrt) the initial condition of the SDE (i.e., a change of measure from the original density $\rho_{u_n|Y_n}$ to the biasing IS density $\rho^{IS}_{u_n|Y_n}$); IS wrt the Wiener process via a stochastic optimal control (SOC) problem (Hartmann et al. 2019); and the combination of IS wrt both initial condition and Wiener process. The idea is illustrated in Figure 1,

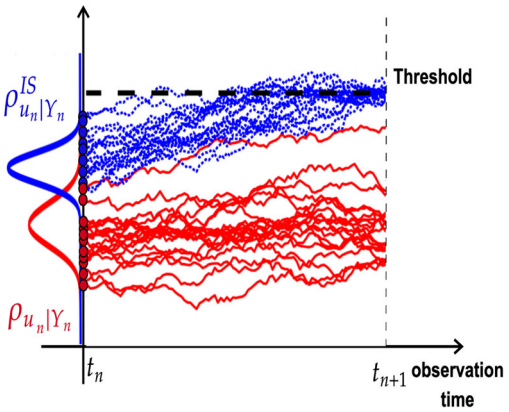


Fig. 1 Illustration of the IS idea for the rare event tracking in the context of the EnKF. The number of blue circles, which depends on the rareness of the estimated event, may differ from the original EnKF ensemble (red circles) number. Rare event paths (blue dashed line) are not reused or propagated in the next observation interval, do not affect the standard EnKF ensembles, and are not involved in data assimilation.

where the red circles form the standard EnKF ensemble and the red trajectories are induced by the Wiener paths, while the blue circles are the auxiliary EnKF ensemble to estimate the rare event probability and the blue trajectories are driven by the controlled SDE obtained via change of measure wrt the Wiener process, thus, allowing to monitor the failure probability between observation times. Here, we note that the auxiliary EnKF ensemble size, which depends on the rareness of the estimated event, is not necessarily the same as the original EnKF ensemble size. We also note that our methodology is designed to be applied independently at each observation time, using only the updated EnKF ensemble available at that step. The generated rare event paths from previous observation time window are not reused or propagated at the next observation time interval, have no influence on the main EnKF ensemble, and are excluded from the data assimilation process. The proposed IS schemes are based on solving the Kolmogorov backward equation (KBE), which is not computationally feasible in high dimensions. To overcome this issue, we employ the Markovian projection technique (Gyöngy 1986; Bayer et al. 2019) that allows us to approximate the multidimensional KBE by a lower-dimensional KBE. We compare our proposed approaches with the multilevel cross entropy (CE) method which is another widely-used technique to estimate rare event probabilities (Kroese et al. 2013; Kroese et al. 2013).

Remark 1 (Aspects beyond the present scope) Note that we operate strictly within the framework of the EnKF. By this we assume that the posterior distribution of the system state at each observation time is approximated by the updated EnKF ensemble and our proposed methodology is applied as a post-processing step. Our goal is to construct and ver-

ify rare event tracking methods between two consecutive observations in practical data assimilation contexts where the EnKF is considered a reasonable and computationally efficient filtering approach. While the EnKF converges to the so-called mean-field Kalman filter under sufficient regularity assumptions and introduces an approximation error relative to the true Bayes filter, addressing or quantifying this discrepancy is beyond the scope of the present study. For more detailed discussions on the accuracy, limitations and convergence properties of the EnKF, we refer the reader to Le Gland et al. (2009); Hoel et al. (2020); Law et al. (2016, 2015).

Literature review. An accurate estimation of rare event probabilities based on dynamic systems is a challenging problem that usually lacks analytical solutions and needs both sophisticated theory and efficient computational tools (Bucklew 2004; James 2016; Sapsis 2021). A considerable amount of literature has been dedicated to analytical and semi-analytical approximation methods from the structural reliability analysis, for instance, the first and second order reliability methods (FORM and SORM) based on first and second order Taylor series expansions, respectively, to find the most likely failure point (Melchers and Beck 2018; Hasofer 1974; Der Kiureghian 2005). However, these methods can be computationally costly in high-dimensional settings since they involve an optimization problem to solve and exhibit large approximation errors of estimates (Valdebenito et al. 2010). An additional perspective to tackle the problem is based on the theory of large deviations which indicates that, in cases of low noise, trajectories may converge towards the most likely path leading to a rare event. The large deviation principle allows to estimate the rare event probability by solving a deterministic optimization problem based simply on the ordinary differential equations. However, this is possible only in the context of the small noise limit. The estimation of extreme events within the framework of large deviations is reviewed in Grafke and Vanden-Eijnden (2019). A combination of large deviations theory with the PDE-constrained optimization is proposed in Tong et al. (2021), with the importance sampling is in Ragone et al. (2018), and with the cross entropy-based importance sampling is in Tong and Stadler (2023).

An alternative approach that has garnered significant attention and is being studied extensively is the use of sampling-based methodologies for rare events in dynamic models. Since a crude Monte Carlo method is unsuitable for the simulation of small probabilities (Rubinstein and Kroese 2016), different advanced variance reduction techniques have been developed to speed up its efficiency while the accuracy is maintained (e.g., subset simulation/splitting algorithms (Siu-Kui and Beck 2001; Botev and Kroese 2012; Cérou et al. 2019; Bréhier and Lelièvre 2019; Webber et al. 2019), sequential MC and IS Heidelberger (1995); Cérou et al. (2012); Papaioannou et al. (2016), line sampling (de Angelis et al. 2015; Straub and Papaioannou 2015), cross-entropy (Kroese et al. 2013; De Boer et al. 2005) methods). In general, sampling methods have demonstrated their strength and efficiency in estimating rare events accurately for high-dimensional settings, as evidenced by their successful implementation in various applications (Siu-Kui and Beck 2003; Beaurepaire et al. 2012; Zio and Pedroni 2009). Another effective sampling method designed for rare events is referred to as genealogical particle analysis. This approach involves selecting and replicating the most favorable realizations of an ensemble of trajectories in an iterative form similar to the idea of MC acceptance-rejection techniques (Del Moral and Garnier 2005).

Recent studies have also explored the relationship between IS schemes and SOC theory when the underlying dynamics is governed by diffusion (Hartmann et al. 2016, 2017, 2018) with the further application for rare event simulation problems (Hartmann et al. 2019). The latter is based on a Gibbs variational principle to find the optimal change of measure, which can be formulated as a SOC problem and solved by the Feynman–Kac representation of the dynamic programming equations (Hartmann et al. 2019). However, these approaches consider a change of measure with respect to the Wiener path only, while keeping the initial condition fixed. The SOC theory is also applied for estimating rare event probabilities in the case of the McKean-Vlasov SDEs (Ben Rached et al. 2022; Rached et al. 2022) and stochastic reaction networks (Ben Hammouda et al. 2023) with further application of the Markovian projection technique in Hammouda et al. (2023).

In the context of data assimilation, the approaches to tracking rare events proposed in this article align with the study presented in Vanden-Eijnden and Weare (2013). The authors combine features of both sampling-based filters and variational methods, resulting in the development of hybrid data assimilation techniques, and test their idea on the Kuroshio model. The study is focused only on the low noise regime and with a small stochastic forcing term, which allows the application of large deviation analysis. However, this work does not consider a change of measure with respect to the initial condition, which is a key component of our proposed importance sampling framework. The work (Wagner et al. 2022) employed the EnKF as a tool to estimate failure probabilities where the rare event problem is reformulated as a Bayesian inverse problem. In particular, the auxiliary EnKF ensembles are used to generate failure samples but not for tracking rare events in real-time, which is the focus of the current study.

Main contributions. This work introduces novel IS strategies for rare event estimation, implemented as a post-processing step within the EnKF framework. While importance sampling and rare event simulation are well-studied in the SDEs literature, particularly via stochastic control and

large deviations techniques, our approach introduces new directions that remain largely unexplored. To the best of our knowledge, IS strategies that involve a change of measure with respect to the initial condition, either independently or in combination with a change of measure with respect to the Wiener process, have not been previously explored or tested numerically in the rare event estimation and SDEs literature. Existing SOC-based approaches typically focus on modifying the path measure while assuming a fixed initial condition (Hartmann et al. 2019, 2017, 2018; Vanden-Eijnden and Weare 2013; Dos Reis et al. 2023). In contrast, our work introduces a principled framework for constructing IS estimators that modify the distribution of the initial condition, as well as a combined approach that jointly alters both the initial distribution and the path measure.

Additionally, the technical approach we introduce for IS wrt the initial condition and IS wrt both initial condition and Wiener process - specifically, the PDE-based approximation of the optimal density combined with Gaussian fitting and the use of the Markovian projection technique for dimensionality reduction - has not, to the best of our knowledge, been previously proposed or evaluated in the filtering context and rare event analysis. We are not aware of any prior work that implements this combination of methods in the manner presented in our manuscript.

Outline. The subsequent sections of the work are structured in the following manner. Section 2 describes a discrete-time filtering problem, followed by a concise summary of the EnKF framework. It also includes a discussion of how the problem relates to the concept of stopping time and the numerical approximation of the quantity of interest (QoI). In Section 3, we propose several IS techniques that can be combined with the EnKF framework. Finally, Section 4 exhibits the numerical part where the proposed IS schemes are applied to three different examples: Double Well SDE, Langevin dynamics and noisy Charney-deVore model.

2 Problem setting

We first describe the discrete-time filtering problem with observations arriving at times $\{t_n\}_{n\in\mathbb{N}}$, and later return to the problem of estimating rare event probabilities between observations.

Let the triple $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete probability space on which we consider the filtering problem. The primary goal in the filtering problem is to track the underlying *signal* u_n governed by the stochastic dynamics

$$u_n(\omega) = \Psi(u_{n-1}, \omega), \quad n = 1, 2, \ldots \tag{1}$$

provided the accumulated data Y_n up to time t_n with *noisy observations* given by

$$y_n(\omega) = Hu_n(\omega) + \eta_n(\omega), \quad n = 1, 2, \ldots \tag{2}$$

where $\omega \in \Omega$, $H \in \mathbb{R}^{m\times d}$, and $\{\eta_k\}_{k\in\mathbb{N}}$ is an independent and identically distributed (i.i.d.) sequence being independent of $\{u_k\}_{k\in\mathbb{N}\cup\{0\}}$ with $\eta_1 \sim N(0, \Gamma)$ for a positive definite $\Gamma \in \mathbb{R}^{m\times m}$. For the sake of simplicity, we further drop the dependence on ω in the notation throughout the paper if there is no confusion.

In other words, the filtering problem is concerned with an approximation of the conditional random variable $u_n|Y_n$ in real-time as soon as the data at time t_n is acquired. Let $\rho_{u_n|Y_n}$ denote the probability density function of u_n given Y_n and $\rho_{u_n|u_{n-1}}$ denote the transition density of u_n given the previous value u_{n-1} typically taken from the model equation in (1). Based on Bayes' theorem, the filtering problem can be expressed formally via two recursive steps of computing the probability density of $u_n|Y_n$ as follows

$$\textbf{prediction step: } \rho_{u_n|Y_{n-1}}(u) \propto \int_{\mathbb{R}^d} \rho_{u_n|u_{n-1}}(u)\rho_{u_{n-1}|Y_{n-1}}(v)dv,$$
$$\textbf{update step: } \rho_{u_n|Y_n}(u) \propto \exp\left(-\left|\Gamma^{-1/2}(y_n - Hu)\right|^2/2\right) \quad \rho_{u_n|Y_{n-1}}(u).$$

If the initial density of $u_0|Y_0$ is assumed to be Gaussian and the dynamics Ψ is linear with additive Gaussian noise, the filtering density becomes Gaussian, which can be entirely described by only two statistics (mean and covariance). In this case, the Kalman filter provides an exact algorithm to track the mean and covariance at each observation timestep. However, in a more general case, when Ψ is nonlinear, an exact solution to the filtering problem is usually not attainable, and we mostly rely on approximation methods. In this work, we focus on an important class of such approximation methods in filtering, namely, the EnKF.

The EnKF is an extension of the Kalman filter to nonlinear settings, which produces an ensemble of particles whose empirical measure approximates the true filtering distribution of $u_n|Y_n$. Let P denote the EnKF ensemble size and the pair $\left(v_{n,i}, \hat{v}_{n,i}\right)$ denote the prediction and updated ensemble particles, respectively, corresponding to the sample ω_i at time t_n. Then, the EnKF algorithm with perturbed observations consists of two iterative steps for $n \in \mathbb{N}$, namely

$$\textbf{prediction step: } \begin{cases} v_{n,i} = \Psi^N_{n-1}(\hat{v}_{n-1,i}), \quad i = 1, 2, \ldots, P, \\ m_n = \dfrac{1}{P}\displaystyle\sum_{i=1}^{P} v_{n,i}, \\ C_n = \dfrac{1}{P-1}\displaystyle\sum_{i=1}^{P}\left(v_{n,i} - m_n\right)\left(v_{n,i} - m_n\right)^{\mathbf{T}}, \end{cases} \tag{3}$$

where Ψ^N_{n-1} denotes the numerical discretization of the dynamics Ψ_{n-1} using $N \geq 1$ uniform timesteps, and

$$\textbf{update step:} \begin{cases} \tilde{y}_{n,i} = y_n + \eta_{n,i}, \ i = 1, 2, ..., P, \\ \mathrm{K}_n = C_n H^{\mathbf{T}} (H C_n H^{\mathbf{T}} + \Gamma)^{-1}, \\ \hat{v}_{n,i} = (I - \mathrm{K}_n H) v_{n,i} + \mathrm{K}_n \tilde{y}_{n,i}, \ i = 1, 2, ..., P, \end{cases} \tag{4}$$

where $\eta_{n,i}$ are i.i.d. draws from $N(0, \Gamma)$, the matrix K_n is known as the Kalman gain and H is the observation matrix in (2) (Hoel et al. 2020, 2022).

Let us consider discrete observation time points over the interval $[0, T]$, denoted by $0 = t_0 < t_1 < ... < t_O = T$. In the remainder of this work, we will consider low probability events between two consecutive observations, at times t_n and t_{n+1}. As mentioned earlier, it will be assumed that the underlying signal process is considered as the evolution of the dynamics Ψ associated with an SDE of the form

$$\begin{cases} du_t = a(u_t)dt + b(u_t)dW_t^{\mathbb{P}}, \quad t \in (t_n, t_{n+1}), \\ u_n \sim \rho_0, \end{cases} \tag{5}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d$ is called the drift coefficient, $b : \mathbb{R}^d \to \mathbb{R}^{d \times d_W}$ is called the diffusion coefficient, $W_t^{\mathbb{P}} : \Omega \times [t_n, \infty) \to \mathbb{R}^{d_W \times 1}$ is an d_W-dimensional Wiener process, independent of u_k, $k \leq n$, under the probability measure $\mathbb{P}$, and the initial condition ρ_0 is defined by the updated EnKF density $\hat{\rho}_{u_n|Y_n}$ at time t_n[1]. The drift and diffusion coefficients are generally expected to satisfy the regularity conditions to ensure the existence and uniqueness of the solution to (5) (see, e.g. (Kloeden and Platen 1992, Section 4.5) for sufficient conditions).

Remark 2 (Tail assumptions) Note that this work does not consider more challenging cases such as SDEs resulting in heavy-tailed events. For that purpose, we assume that the drift coefficient a is measurable and has a linear growth in space, and the diffusion coefficient b is bounded and satisfies the usual uniform ellipticity condition. This ensures that the SDE (5) has a unique weak solution and admits the density whose tails decay as Gaussian (see, e.g. Menozzi et al. (2021), Eq. (1.2)-(1.3)).

According to the main objective of the work, we restrict our attention to the observation interval $[t_n, t_{n+1}]$ (although the proposed approaches in this work will be applicable for all observation time intervals). We define the first time when the running maximum process $\{\max_{t_n \leq t \leq t_{n+1}} P_1 u_t\}$ hits the threshold $\mathcal{K}$ given $P_1 u_n < \mathcal{K}$ at the initial time t_n by

$$\tau_{\mathcal{K}} := \inf\{t > t_n : \ P_1 u_t \geq \mathcal{K} \,|\, P_1 u_n < \mathcal{K}\}.$$

In the literature, the random variable $\tau_{\mathcal{K}}$ is referred to as *a first exit/passage time* and $\tau := \min\{\tau_{\mathcal{K}}, t_{n+1}\}$ is *a stopping time*. There is a connection between the events $\{\max_{t_n \leq t \leq t_{n+1}} P_1 u_t \geq \mathcal{K}\}$ and $\{\tau_{\mathcal{K}} \leq t_{n+1}\}$, which allows us to write

$$\begin{aligned} \alpha_n &= \mathbb{E}_{\rho_0 \otimes \mathbb{P}}[\mathbb{1}_{\{\max_{t_n \leq t \leq t_{n+1}} P_1 u_t \geq \mathcal{K}\}} | Y_n] \\ &= \mathbb{E}_{\rho_0 \otimes \mathbb{P}}[\mathbb{1}_{\{P_1 u_\tau \geq \mathcal{K}\}} | Y_n], \end{aligned} \tag{6}$$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function. Furthermore, for convenience, we will omit the notation for dependence on the data Y_n.

A standard procedure to estimate the QoI in (6) is to employ a time-discretization scheme such as Euler-Maruyama and use an MC simulation for computing the mean estimation of the exit events. Let $t_n = t_{n,0} < t_{n,1} < ... < t_{n,k} < ... < t_{n,K} = t_{n+1}$ be a partition of the interval $[t_n, t_{n+1}]$ with $t_{n,k} := t_n + k\Delta t$ and $\Delta t = \frac{t_{n+1} - t_n}{K}$. The discrete forward Euler-Maruyama (EM) approximation of the SDE (5) is

$$\begin{cases} \bar{u}_{n,k+1} = \bar{u}_{n,k} + a(\bar{u}_{n,k})\Delta t + b(\bar{u}_{n,k})\Delta W_{n,k}^{\mathbb{P}}, \\ \bar{u}_{n,0} = u_n, \end{cases} \tag{7}$$

for $k = 0, ..., K-1$, with $\Delta W_{n,k}^{\mathbb{P}} = W_{n,k+1}^{\mathbb{P}} - W_{n,k}^{\mathbb{P}} \sim N(0, \Delta t \mathbb{I}_{d_W})$ where $\mathbb{I}_{d_W}$ is the d_W-dimensional identity matrix.

For sufficiently small Δt, we can approximate the QoI by a function of K discrete EM points, i.e.,

$$\alpha_n \approx \mathbb{E}_{\rho_0 \otimes \mathbb{P}}[\mathbb{1}_{\mathcal{U}}(\bar{u}_{n,0:K})], \tag{8}$$

where $\bar{u}_{n,0:K} := \{\bar{u}_{n,k}\}_{k=0}^K$ and

$$\mathcal{U} := \{x_k \in \mathbb{R}^d, k = 0, ..., K : P_1 x_k \geq \mathcal{K} \text{ for some } k\}.$$

Hereafter, we work with this EM approximation of the QoI to perform IS techniques.

Remark 3 (Stopped diffusions) One of the challenges in approximating the QoI involving stopped diffusions on the boundary is that a continuous path may leave the domain (i.e., hit the threshold) even if a discrete approximate solution does not exit the boundary (see Figure 2). This leads to a slow convergence of the time-discretization error (for more discussions, see (Gobet 2000, 2009; Dzougoutov et al. 2005)). In particular, the EM scheme in (7) achieves a weak convergence, $\mathcal{O}(\sqrt{\Delta t})$, of the relative error, as $\Delta t \to 0$, compared to a standard Euler-type convergence $\mathcal{O}(\Delta t)$ under certain conditions (Kloeden and Platen 1992; Bally and Talay 1996; Higham 2001). One way to recover the first-order convergence of the EM method is by performing a straightforward hitting test after each time step. This involves examining the distribution of the Brownian bridge between two discrete EM nodes to determine if the threshold was crossed during the

[1] refers to a Gaussian density characterized by the empirical mean and covariance matrix computed from the EnKF ensemble at time t_n.

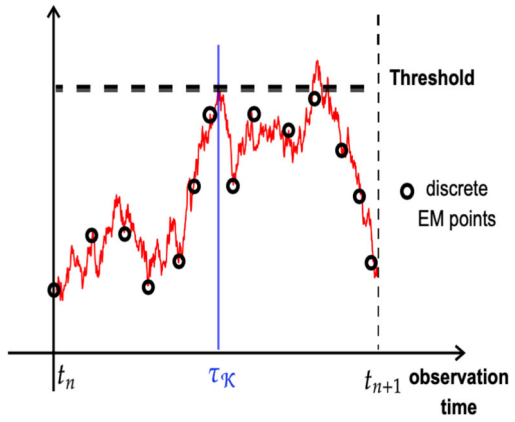


Fig. 2 Illustration of the challenge with stopped diffusions in the boundary discussed in Remark 3. The continuous path (red) hits the threshold at $\tau_{\mathcal{K}}$ although the discrete EM solutions do not exit the boundary (or may exit after $\tau_{\mathcal{K}}$).

given time step. This approach is referred to as *the Brownian bridge technique* (Gobet 2000) and implemented in this work (see Algorithm 2 in Appendix D). However, note that this is not the main challenge considered in this work.

3 Importance Sampling techniques

The primary focus of the work is to reduce the variance effectively via appropriate importance sampling techniques. For that purpose, observe that there are two sources of randomness in (5) coming from the initial condition and also from the Wiener process. Correspondingly, we can perform IS wrt ρ_0, IS wrt $W_t^{\mathbb{P}}$ and IS wrt both ρ_0 and $W_t^{\mathbb{P}}$. In this section, we describe the methodology of each proposal applied to the EM approximation of the QoI.

3.1 IS with respect to initial condition

Let us consider the joint density $\rho_{\bar{u}_{n,0:K}}$ of a random vector $(\bar{u}_{n,0:K})$ with $\rho_{\bar{u}_{n,0}} = \rho_0$. Now we introduce an alternative density function $\tilde{\rho}_0$, such that it is non-negative and $\int \tilde{\rho}_0(x_0)dx_0 = 1$. Using the shorthand $x_{1:K}$ for a K-tuple $(x_1, ..., x_K)$, rewrite the approximation of the QoI as

$$\begin{aligned}
\hat{\alpha}_n &:= \mathbb{E}_{\rho_0\otimes\mathbb{P}}[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)] \\
&= \int \mathbb{1}_{\mathcal{U}}\left(x_{0:K}\right)\rho_{\bar{u}_{n,0:K}}(x_{0:K})dx_{0:K} \\
&= \int \mathbb{1}_{\mathcal{U}}\left(x_{0:K}\right)\rho_0(x_0)\rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0)dx_{0:K} \\
&= \int \mathbb{1}_{\mathcal{U}}\left(x_{0:K}\right)\frac{\rho_0(x_0)}{\tilde{\rho}_0(x_0)}\rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0)\tilde{\rho}_0(x_0)dx_{0:K} \\
&= \mathbb{E}_{\tilde{\rho}_0\otimes\mathbb{P}}\Big[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)L_0(\bar{u}_{n,0})\Big],
\end{aligned} \tag{9}$$

where $L_0(x_0) := \frac{\rho_0(x_0)}{\tilde{\rho}_0(x_0)}$ is *the likelihood ratio*.

In order to select a proper density $\tilde{\rho}_0$ that results in a smaller variance, let us consider the second moment of the QoI approximation,

$$\begin{aligned}
&\mathbb{E}_{\tilde{\rho}_0\otimes\mathbb{P}}\Big[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)L_0^2(\bar{u}_{n,0})\Big] \\
&= \int \mathbb{1}_{\mathcal{U}}\left(x_{0:K}\right)\frac{\rho_0^2(x_0)}{\tilde{\rho}_0^2(x_0)}\rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0)\tilde{\rho}_0(x_0)dx_{0:K} \\
&= \int \mathbb{1}_{\mathcal{U}}\left(x_{0:K}\right)\frac{\rho_0^2(x_0)}{\tilde{\rho}_0(x_0)}\rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0)dx_{0:K} \\
&= \int \mathbb{E}[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)|\bar{u}_{n,0} = x_0]\frac{\rho_0^2(x_0)}{\tilde{\rho}_0(x_0)}dx_0
\end{aligned} \tag{10}$$

where the last equality follows from Fubini's theorem. Now employing a classical inequality from analysis to the last term[2], we have

$$\begin{aligned}
&\int \left(\sqrt{\mathbb{E}[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)|\bar{u}_{n,0} = x_0]}\frac{\rho_0(x_0)}{\sqrt{\tilde{\rho}_0(x_0)}}\right)^2 dx_0 \\
&\underbrace{\int \left(\sqrt{\tilde{\rho}_0(x_0)}\right)^2 dx_0}_{=1} \\
&\geq \left(\int \left|\sqrt{\mathbb{E}[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)|\bar{u}_{n,0} = x_0]}\rho_0(x_0)\right| dx_0\right)^2,
\end{aligned} \tag{11}$$

which shows that the second moment of the QoI approximation is always at least large as right-hand side of (11). We can deduce that (11) becomes an equality if $\tilde{\rho}_0(x_0)$ is proportional to

$$\begin{aligned}
\tilde{\rho}_0(x_0) &\propto \rho_0(x_0)\sqrt{\mathbb{E}[\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)|\bar{u}_{n,0} = x_0]} \\
&= \rho_0(x_0)\sqrt{\mathbb{P}\Big(\max_{0\leq k\leq K} P_1\bar{u}_{n,k} \geq \mathcal{K}\Big|\bar{u}_{n,0} = x_0\Big)},
\end{aligned} \tag{12}$$

which is the strategy for the optimal importance sampling density when we change a measure only with respect to initial condition.

Here, ρ_0 is given and the exit probability needs to be approximated.

3.1.1 PDE-based method

We propose a PDE-based approach to computing the probability in (12). Let us define the function

$$\gamma(x,t) := \mathbb{P}\Big(\max_{t\leq s\leq t_{n+1}} P_1 u_s \geq \mathcal{K}\Big|u_t = x\Big).$$

[2] The Cauchy-Schwartz inequality: $\left(\int |u(x)v(x)|\,dx\right)^2 \leq \int u^2(x)dx \int v^2(x)dx$. Note that the inequality becomes an equality when two functions u and v are proportional to each other.

This $\gamma(x,t)$ solves the Kolmogorov Backward Equation (KBE) associated with the SDE (1). We will first state this problem for $d=1$, where the standard discretization-based numerical PDE approximation methods are efficient, and then propose the use of Markovian projection (Gyöngy 1986; Bayer et al. 2019) as a way of approximating the solution of the d-dimensional PDE by a solution of a lower dimensional one.

One-dimensional setup. For $d=1$, the process $P_1 u_t = u_t$ is a one-dimensional Markov process, and thus, $\gamma(x,t)$ satisfies the KBE with a Dirichlet boundary condition on the domain $(-\infty, \mathcal{K}] \times [t_n, t_{n+1}]$,

$$\begin{cases} \frac{\partial \gamma}{\partial t} = -a(x)\frac{\partial \gamma}{\partial x} - \frac{1}{2}b(x)b(x)^T\frac{\partial^2 \gamma}{\partial x}, \\ \gamma(x,t_{n+1}) = 0, \quad x < \mathcal{K}, \\ \gamma(\mathcal{K},t) = 1, \quad t \in [t_n, t_{n+1}], \\ \lim\limits_{x\to-\infty} \gamma(x,t) = 0, \quad t \in [t_n, t_{n+1}]. \end{cases} \tag{13}$$

Note that there is a discontinuity in the data at $(\mathcal{K}, t_{n+1})$. By numerically approximating (13) backward in time with the technique of freezing the coefficients in a neighborhood of the discontinuity (see Appendix C for more details about the numerical solver), we can obtain a sufficiently accurate approximation for the discrete-time exit probability in (12) by a terminal time solution

$$\mathbb{P}\Big(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} \Big| \bar{u}_{n,0} = x_0\Big) \approx \begin{cases} \gamma(x_0,t_n), & \text{if } x_0 < \mathcal{K}, \\ 1, & \text{if } x_0 \ge \mathcal{K}, \end{cases}$$

and compute for $x_0 \in \mathbb{R}$

$$\tilde{\rho}_0^{PDE,1}(x_0) \propto \rho_0(x_0)\sqrt{\gamma(x_0,t_n)}. \tag{14}$$

For simplicity, due to the Gaussian assumption of EnKF, we fit a Gaussian density with parameters $(\mu_0^{\text{fit}}, \sigma_0^{\text{fit}})$ to the optimal IS initial density via a moment matching approach. Specifically, we first evaluate the product $\rho_0(x_0)\sqrt{\gamma(x_0,t_n)}$ over a sufficiently fine 1D grid covering the relevant domain. We then normalize it using numerical integration via the trapezoidal rule. The Gaussian parameters μ_0^{fit} and σ_0^{fit} are obtained by matching the first and second moments of the normalized optimal IS initial density:

$$\begin{aligned} \mu_0^{\text{fit}} &= \int x_0 \cdot \tilde{\rho}_0^{\text{PDE},1}(x_0)\, dx_0, \\ (\sigma_0^{\text{fit}})^2 &= \int x_0^2 \cdot \tilde{\rho}_0^{\text{PDE},1}(x_0)\, dx_0 - (\mu_0^{\text{fit}})^2. \end{aligned}$$

where the integrals are approximated numerically via the trapezoidal integration rule.

Let us assume $\rho_0(x_0) \sim N(\mu_0, \sigma_0)$. Then, the likelihood L_0 is defined by the ratio of two Gaussian densities

$$L_0(x_0) = \frac{\rho_0(x_0)}{\tilde{\rho}_0^{PDE,1}(x_0)} = \frac{\sigma_0^{\text{fit}}}{\sigma_0} \exp\left(-\frac{(x_0-\mu_0)^2}{2\sigma_0^2} + \frac{\left(x_0-\mu_0^{\text{fit}}\right)^2}{2\left(\sigma_0^{\text{fit}}\right)^2}\right).$$

Note that the methodology is initially developed within the EnKF framework because its Gaussian structure simplifies both sampling and likelihood ratio computations in the IS setting. Without this assumption, identifying and sampling from an optimal initial distribution would require additional techniques-such as nonparametric density estimation or variational inference-which are beyond the scope of this study.

High-dimensional setup. For $d > 1$, the process $S_t := P_1 u_t$ with $u_t \in \mathbb{R}^d$ is in general no longer Markovian, and thus, we apply a *Markovian projection technique* to be able to employ the proposed PDE-based method above (Gyöngy 1986; Bayer et al. 2019). We introduce a Markovian surrogate process $\check{S}_t \in \mathbb{R}$ which follows the SDE

$$\begin{cases} d\check{S}_t = \check{a}(\check{S}_t,t)dt + \check{b}(\check{S}_t,t)d\check{W}_t^{\mathbb{P}}, \quad t \in (t_n, t_{n+1}), \\ \check{S}_{t_n} = P_1 u_n, \quad u_n \sim \rho_0, \end{cases} \tag{15}$$

where $\check{W}_t^{\mathbb{P}} : \Omega \times [0,\infty) \to \mathbb{R}$ is an one-dimensional Wiener process independent of the initial condition $\check{S}_{t_n}$, and $\check{a}, \check{b} : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ are non-random coefficients defined by

$$\begin{cases} \check{a}(y,t) = \mathbb{E}[P_1 a(u_t) | P_1 u_t = y], \\ \check{b}^2(y,t) = \mathbb{E}[(P_1 b b^T P_1^T)(u_t) | P_1 u_t = y]. \end{cases} \tag{16}$$

It is proved that, in the case of a fixed initial condition, if $P_1 a(x)$ and $P_1 b(x)$ are bounded measurable functions such that $(P_1 b b^T P_1^T)(x)$ is uniformly positive definite, then the SDE (15) admits a weak solution $\check{S}_t$ which has the same one-dimensional probability distribution as S_t for every t (Gyöngy 1986). Since we have a random initial condition, it is assumed that the conditional expected values in (16) exist and are well-defined too. One can use a discrete L^2 regression to approximate (16). We provide the details of the approximation method in Appendix A.

For simplicity, we assume that the coordinate system is oriented so that one component is orthogonal to the hyperplane that defines the rare event threshold, e.g., $P_1 u_t = u_t^i$ for a given $i = 1, ..., d$. For a general form of the projection P_1, we can use an orthogonal coordinate transformation such that the rare event is defined by a single component in the transformed coordinates with the condition that the first row of the rotation matrix and P_1^T are collinear

Following the same idea as in one dimension, we solve the KBE corresponding to the surrogate process $\check{S}_t$ defined by

$$\begin{cases} \frac{\partial \check{\gamma}}{\partial t} = -\check{a}(y,t)\frac{\partial \check{\gamma}}{\partial y} - \frac{1}{2}\check{b}(y,t)\check{b}(y,t)^T\frac{\partial^2 \check{\gamma}}{\partial y}, \\ \check{\gamma}(y,t_{n+1}) = 0, \quad y < \mathcal{K}, \\ \check{\gamma}(\mathcal{K},t) = 1, \quad t \in [t_n, t_{n+1}], \\ \lim_{s\to-\infty} \check{\gamma}(y,t) = 0, \quad t \in [t_n, t_{n+1}]. \end{cases} \tag{17}$$

and approximate the high-dimensional optimal IS initial density for $x_0 \in \mathbb{R}^d$ by

$$\tilde{\rho}_0^{PDE,1}(x_0) \propto \rho_0(x_0)\sqrt{\gamma(x_0,t_n)} \approx \rho_0(x_0)\sqrt{\check{\gamma}(P_1 x_0, t_n)}. \tag{18}$$

Note that $\rho_0(x_0)$ is here a multivariate Gaussian density. For simplicity, let us provide an example of how we fit a Gaussian density to the product $\rho_0(x_0)\sqrt{\check{\gamma}(P_1x_0,t_n)}$ in the two-dimensional case.

Example 1 Let $x_0 = (x_0^1, x_0^2) \in \mathbb{R}^2$ and the projection $P_1 x_0 = x_0^1$. Let $\rho_0(x_0)$ be a bivariate Gaussian density with mean $\begin{bmatrix}\mu_0^1 \\ \mu_0^2\end{bmatrix}$ and covariance $\begin{bmatrix}(\sigma_0^1)^2 & \varrho\sigma_0^1\sigma_0^2 \\ \varrho\sigma_0^1\sigma_0^2 & (\sigma_0^2)^2\end{bmatrix}$ where ϱ is a correlation coefficient between x_0^1 and x_0^2. Then, the optimal IS initial density $\tilde{\rho}_0(x_0^1, x_0^2) \propto \rho_0(x_0^2|x_0^1)\rho_0(x_0^1)\sqrt{\bar{\gamma}(x_0^1, x_0^2; t_n)}$ can be approximated by the probability density

$$\check{\tilde{\rho}}_0(x_0^1, x_0^2) \propto \rho_0(x_0^2|x_0^1)\rho_0(x_0^1)\sqrt{\check{\gamma}(x_0^1,t_n)}.$$

We can fit the univariate Gaussian density to the product $\rho_0(x_0^1)\sqrt{\check{\gamma}(x_0^1,t_n)}$ with parameters μ_0^{fit} and σ_0^{fit} via the moment matching approach as described in one-dimensional setting. Note that $\check{\tilde{\rho}}_0(x_0^2|x_0^1) = \rho_0(x_0^2|x_0^1)$ and

$$\begin{aligned} \check{\tilde{\rho}}_0(x_0^1) &\propto \int \rho_0(x_0^2|x_0^1)\rho_0(x_0^1)\sqrt{\check{\gamma}(x_0^1,t_n)}dx_0^2 \\ &= \rho_0(x_0^1)\sqrt{\check{\gamma}(x_0^1,t_n)}. \end{aligned}$$

Applying a definition of the conditional bivariate Gaussian distribution, we have

$$\check{\tilde{\rho}}_0(x_0^2|x_0^1) \sim N\Big(\mu_0^2 + \varrho\frac{\sigma_0^2}{\sigma_0^1}(x_0^1 - \mu_0^1), (1-\varrho^2)(\sigma_0^2)^2\Big),$$

where $x_0^1 \sim N(\mu_0^{\text{fit}}, \sigma_0^{\text{fit}})$. The likelihood ratio, in this case, is defined by

$$L_0(x_0) = \frac{\rho_0(x_0^1)}{\check{\tilde{\rho}}_0(x_0^1)} = \frac{\sigma_0^{\text{fit}}}{\sigma_0^1}\exp\left(-\frac{(x_0^1-\mu_0^1)^2}{2(\sigma_0^1)^2} + \frac{(x_0^1-\mu_0^{\text{fit}})^2}{2(\sigma_0^{\text{fit}})^2}\right)$$

A similar approach is used in the numerical section 4.2.

3.1.2 Multilevel Cross-Entropy method

For the purpose of comparison, we consider a simple, efficient iterative procedure known as the multilevel cross-entropy (CE) method to estimate the optimal IS density (12). The core idea of the CE is to assume $\tilde{\rho}_0(x_0)$ to be from a particular family of densities $\{\rho_0(x_0;\boldsymbol{\nu})\}$ (in the framework of EnKF - Gaussian family) and select the parameter vector $\boldsymbol{\nu}$ such that the distance between $\rho_0(x_0;\boldsymbol{\nu})$ and $\tilde{\rho}_0(x_0)$ in (12) is minimal (De Boer et al. 2005; Kroese et al. 2013). A methodology and basic algorithm of the multilevel CE adjusted to the EnKF setting are briefly provided in Appendix B. We denote the optimal IS density obtained via the multilevel CE method by $\tilde{\rho}_0^{CE}$.

3.2 IS with respect to Wiener processes using SOC

Here, we base the analysis on the continuous time problem (5). We apply IS wrt Wiener processes obtained by stochastic optimal control for the SDEs of the form (5) with a fixed initial condition x_0.

One-dimensional setup. Let us introduce a process

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} - \int_{t_n}^t \xi(u_s, s)ds$$

where $\xi : \mathbb{R} \times [t_n, t_{n+1}] \to \mathbb{R}^{d_W \times 1}$ and substitute it in (5) for $t \in (t_n, t_{n+1})$ to get

$$\begin{cases} du_t = (a(u_t) + b(u_t)\xi(u_t,t))dt + b(u_t)dW_t^{\mathbb{Q}}, \\ u_n = x_0. \end{cases} \tag{19}$$

Girsanov's theorem (Karatzas and Shreve 2012) states under certain conditions that $(W^{\mathbb{Q}}(t))_{t\geq 0}$ is a Wiener process under the probabilty measure $\mathbb{Q}$ with the likelihood ratio

$$\begin{aligned} &\frac{d\mathbb{P}}{d\mathbb{Q}}|_{[t_n,t]} \\ &= \exp\left(-\int_{t_n}^t \xi(u_s,s)^T dW_s^{\mathbb{Q}} - \frac{1}{2}\int_{t_n}^t \xi(u_s,s)^T\xi(u_s,s)ds\right). \end{aligned} \tag{20}$$

Then, a change of measure wrt the Wiener process is as follows

$$\mathbb{E}_{\rho_0\otimes\mathbb{P}}[\mathbb{1}_{\{u_\tau\geq\mathcal{K}\}}] = \mathbb{E}_{\rho_0\otimes\mathbb{Q}}\Big[\mathbb{1}_{\{u_\tau^\xi\geq\mathcal{K}\}}\frac{d\mathbb{P}}{d\mathbb{Q}}|_{[t_n,\tau]}\Big],$$

where u_t^ξ follows the controlled SDE

$$\begin{cases} du_t^\xi = (a(u_t^\xi) + b(u_t^\xi)\xi(u_t^\xi,t))dt + b(u_t^\xi)dW_t^{\mathbb{Q}}, \\ u_n^\xi = x_0. \end{cases}$$

Now, the aim is to find a control ξ that minimizes the variance of the QoI. Note that it is sufficient to minimize the second moment of the estimator that is

$$\min_{\xi} \mathbb{E}_{\rho_0 \otimes \mathbb{Q}}\Big[\mathbb{1}_{\{u_\tau^\xi \geq \mathcal{K}\}} e^{-2\int_{t_n}^{\tau} \xi(u_s^\xi,s)^T dW_s^{\mathbb{Q}} - \int_{t_n}^{\tau} \xi(u_s^\xi,s)^T \xi(u_s^\xi,s) ds} \Big| u_n^\xi = x_0\Big].$$

Let us define the value function that minimizes the second moment of the MC estimator:

$$\theta(x,t) = \min_{\xi} \mathbb{E}_{\rho_0 \times \mathbb{Q}}\Big[\mathbb{1}_{\{u_\tau^\xi \geq \mathcal{K}\}} e^{-2\int_{t}^{\tau} \xi(u_s^\xi,s)^T dW_s^{\mathbb{Q}} - \int_{t}^{\tau} \xi(u_s^\xi,s)^T \xi(u_s^\xi,s) ds} \Big| u_t^\xi = x\Big].$$

Then, $\theta(x,t)$ satisfies the following terminal-boundary problem (Fleming and Soner 2006; Hartmann and Schütte 2012; Hartmann et al. 2017, 2019; Nüsken and Richter 2021)

$$\begin{cases} \frac{\partial \theta}{\partial t} = -a(x)\frac{\partial \theta}{\partial x} - \frac{1}{2} b(x)b(x)^T \frac{\partial^2 \theta}{\partial x^2} + \frac{1}{2\theta} b(x)b(x)^T \left(\frac{\partial \theta}{\partial x}\right)^2, \\ \theta(x, t_{n+1}) = 0, \quad x < \mathcal{K}, \\ \theta(\mathcal{K}, t) = 1, \quad t \in [t_n, t_{n+1}] \\ \lim_{x \to -\infty} \theta(x,t) = 0, \ t \in [t_n, t_{n+1}], \end{cases} \tag{21}$$

with the optimal control

$$\xi^\star(x,t) = \frac{1}{2} b(x)^T \frac{\partial \log \theta(x,t)}{\partial x}.$$

Alternatively, we can consider $\theta(x,t) = \gamma^2(x,t)$ which leads to the KBE defined in (13) with the optimal control (Fleming and Soner 2006; Hartmann and Schütte 2012; Hartmann et al. 2017, 2019; Nüsken and Richter 2021; Ben Rached et al. 2022)

$$\xi^*(x,t) = b(x)^T \frac{\partial \log \gamma(x,t)}{\partial x}. \tag{22}$$

High-dimensional setup. In more than one dimension, similarly to Section 3.1.1, we apply the Markovian projection technique and apply a change of measure wrt the Wiener process to the one-dimensional surrogate process $\check{S}_t$. Then, the value function satisfies the above-mentioned PDEs since the Markov property for $\check{S}_t$ holds. The high-dimensional optimal control with $x \in \mathbb{R}^d$ can be approximated via a solution of the one-dimensional PDE by

$$\xi^*(x,t) \approx b(x)^T P_1^T \frac{\partial \log \check{\gamma}(P_1 x, t)}{\partial y}, \tag{23}$$

where the partial derivative is taken wrt the component on which we project.

Correspondingly, the high-dimensional controlled SDE takes the form

$$\begin{cases} du_t^\xi = (a(u_t^\xi) + b(u_t^\xi)\xi(u_t^\xi, t))dt + b(u_t^\xi) dW_t^{\mathbb{Q}}, \\ u_n^\xi = x_0, \end{cases}$$

where $\xi : \mathbb{R}^d \times [t_n, t_{n+1}] \to \mathbb{R}^{d_W \times 1}$ and the likelihood ratio is

$$\frac{d\mathbb{P}}{d\mathbb{Q}}|_{[t_n,t]} = \exp\left(-\int_{t_n}^{t} \xi(u_s,s)^T dW_s^{\mathbb{Q}} - \frac{1}{2}\int_{t_n}^{t} \xi(u_s,s)^T \xi(u_s,s) ds\right). \tag{24}$$

3.3 IS with respect to both initial condition and Wiener processes

Recall that $\rho_{\bar{u}_{n,0:K}}$ denotes the joint density of a random vector $(\bar{u}_{n,0:K})$ with $\rho_{\bar{u}_{n,0}} = \rho_0$. Now we introduce an alternative joint density function $\tilde{\rho}_{\bar{u}^\xi_{n,0:K}}$ and consider

$$\begin{aligned} &\mathbb{E}_{\rho_0 \otimes \mathbb{P}}[\mathbb{1}_{\mathcal{U}}(\bar{u}_{n,0:K})] \\ &= \int \mathbb{1}_{\mathcal{U}}(x_{0:K})\, \rho_{\bar{u}_{n,0:K}}(x_{0:K}) dx_{0:K} \\ &= \int \mathbb{1}_{\mathcal{U}}(x_{0:K}) \frac{\rho_{\bar{u}_{n,0:K}}(x_{0:K})}{\tilde{\rho}_{\bar{u}^\xi_{n,0:K}}(x_{0:K})} \tilde{\rho}_{\bar{u}^\xi_{n,0:K}}(x_{0:K}) dx_{0:K} \\ &= \mathbb{E}_{\tilde{\rho}_0 \otimes \mathbb{Q}}\Big[\mathbb{1}_{\mathcal{U}}(\bar{u}^\xi_{n,0:K}) \frac{\rho_{\bar{u}_{n,0:K}}(\bar{u}^\xi_{n,0:K})}{\tilde{\rho}_{\bar{u}^\xi_{n,0:K}}(\bar{u}^\xi_{n,0:K})}\Big] \\ &= \mathbb{E}_{\tilde{\rho}_0 \otimes \mathbb{Q}}\Big[\mathbb{1}_{\mathcal{U}}(\bar{u}^\xi_{n,0:K}) L_0(\bar{u}^\xi_{n,0}) L_W(\bar{u}^\xi_{n,1:K})\Big], \end{aligned} \tag{25}$$

where $L_0(\cdot)$ is a likelihood ratio of the IS wrt the initial condition and $L_W(\cdot)$ is a likelihood ratio of the IS wrt the Wiener processes.

Using the Cauchy-Schwartz inequality, similarly to (11), the optimal importance joint density that reduces the variance of the estimator is proportional to

$$\tilde{\rho}_{\bar{u}^\xi_{n,0:K}}(x_{0:K}) \propto \mathbb{1}_{\mathcal{U}}(x_{0:K})\, \rho_{\bar{u}_{n,0:K}}(x_{0:K}).$$

Furthermore, we can deduce that

$$\begin{aligned} &\tilde{\rho}_{\bar{u}^\xi_{n,0}}(x_0) = \int \tilde{\rho}_{\bar{u}^\xi_{n,0:K}}(x_{0:K}) dx_{1:K} \propto \\ &\int \mathbb{1}_{\mathcal{U}}(x_{0:K})\, \rho_{\bar{u}_{n,0:K}}(x_{0:K}) dx_{1:K} \\ &= \int \mathbb{1}_{\mathcal{U}}(x_{0:K})\, \rho_0(x_0) \rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0) dx_{1:K} \\ &= \rho_0(x_0) \int \big(\mathbb{1}_{\{P_1 x_0 \geq \mathcal{K}\}}(x_0) + \mathbb{1}_{\{P_1 x_0 < \mathcal{K}\}}(x_0) \\ &\mathbb{1}_{\mathcal{U}}(x_{1:K})\big) \rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}}(x_{1:K}|x_0) dx_{1:K} \\ &= \rho_0(x_0) \begin{cases} \mathbb{P}(\max_{1 \leq k \leq K} P_1 \bar{u}_{n,k} \geq \mathcal{K} | \bar{u}_{n,0} = x_0), & P_1 x_0 < \mathcal{K}, \\ 1, & P_1 x_0 \geq \mathcal{K}. \end{cases} \end{aligned} \tag{26}$$

One-dimensional setup. Similarly to Section 3.1, for $d = 1$, the optimal importance initial density in the IS wrt both initial condition and Wiener process is given by

$$\tilde{\rho}_0^{PDE,2}(x_0) \propto \rho_0(x_0)\gamma(x_0, t_n), \quad x_0 < \mathcal{K},$$

where $\gamma(x_0, t_n)$ is an approximation for the terminal time solution of the KBE (13). Notice that (12) differs from (26) by having a square root in the exit probability.

The optimal control is defined similar to (22) and the likelihood ratios are correspondingly given by

$$\begin{aligned} L_0(\bar{u}^{\xi}_{n,0}) &= \frac{\rho_0(\bar{u}^{\xi}_{n,0})}{\tilde{\rho}_0^{PDE,2}(\bar{u}^{\xi}_{n,0})} \\ &= \frac{\sigma_0^{\text{fit}}}{\sigma_0} \exp\left(-\frac{(x_0-\mu_0)^2}{2\sigma_0^2} + \frac{\left(x_0-\mu_0^{\text{fit}}\right)^2}{2\left(\sigma_0^{\text{fit}}\right)^2}\right), \\ L_W(\bar{u}^{\xi}_{n,1:K}) &= \prod_{k=1}^{K-1} \exp\Big(-\xi(t_{n,k}, \bar{u}^{\xi}_{n,k})^T \Delta W^{\mathbb{Q}}_{n,k} \\ &\quad -\frac{1}{2}\Delta t_{n,k}\xi(t_{n,k}, \bar{u}^{\xi}_{n,k})^T \xi(t_{n,k}, \bar{u}^{\xi}_{n,k})\Big), \end{aligned} \tag{27}$$

where the control ξ is the zero vector after a stopping time. Note that L_W here is the discrete form of the likelihood defined in (20).

High-dimensional setup. For $d > 1$, similarly to the previous sections, as a result of the Markovian projection technique, the optimal control is given by (23), the optimal initial density is approximated by

$$\tilde{\rho}_0^{PDE,2}(x_0) \propto \rho_0(x_0)\gamma(x_0, t_n) \approx \rho_0(x_0)\check{\gamma}(P_1 x_0, t_n), \quad P_1 x_0 < \mathcal{K},$$

and the likelihood ratios are respectively given by

$$\begin{aligned} L_0(\bar{u}^{\xi}_{n,0}) &= \frac{\rho_0(\bar{u}^{\xi}_{n,0})}{\tilde{\rho}_0^{PDE,2}(\bar{u}^{\xi}_{n,0})}, \\ L_W(\bar{u}^{\xi}_{n,1:K}) &= \prod_{k=1}^{K-1} \Big(-\xi(t_{n,k}, \bar{u}^{\xi}_{n,k})^T \Delta W^{\mathbb{Q}}_{n,k} \\ &\quad -\frac{1}{2}\xi(t_{n,k}, \bar{u}^{\xi}_{n,k})^T \xi(t_{n,k}, \bar{u}^{\xi}_{n,k})\Delta t_{n,k}\Big), \end{aligned} \tag{28}$$

where the control ξ is the zero vector after a stopping time. Note that L_W here is the discrete form of the likelihood defined in (24).

4 Numerical examples

In this section, we provide numerical examples that showcase the application of the IS techniques presented in Section 3. We test the proposed IS approaches in a range of problems: a Double Well SDE, Langevin dynamics and a noisy Charney-deVore model (Hoel et al. 2020, 2022; Grafke and Vanden-Eijnden 2019). Experimental support for the variance reduction of the proposed methods is provided in comparison with the standard MC and the multilevel CE methods.

The standard MC estimator to approximate the probability $\hat{\alpha}_n$ in (9) of the discretized dynamics is

$$\hat{\alpha}_n^{MC} := \frac{1}{J}\sum_{i=1}^{J} \mathbb{1}_{\mathcal{U}}\left(\bar{u}^{[i]}_{n,0:K}\right)$$

where $\left\{\mathbb{1}_{\mathcal{U}}\left(\bar{u}^{[i]}_{n,0:K}\right)\right\}_{i=1}^{J}$ are i.i.d realizations of $\left\{\mathbb{1}_{\mathcal{U}}\left(\bar{u}_{n,0:K}\right)\right\}$. The estimator is unbiased, which means $\mathbb{E}[\hat{\alpha}_n^{MC}] = \hat{\alpha}_n$, and the variance is

$$\mathbb{V}[\hat{\alpha}_n^{MC}] = \frac{\hat{\alpha}_n - \hat{\alpha}_n^2}{J}.$$

The asymptotic result in the CLT motivates the confidence interval (CI) for $\hat{\alpha}_n$ as

$$\hat{\alpha}_n^{MC} \pm z_c\sqrt{\mathbb{V}[\hat{\alpha}_n^{MC}]},$$

where z_c is a constant corresponding to a given level of confidence. We consider 95% CI in our numerical simulations by setting $z_c = 1.96$.

In the regime of rare events, that is, $\hat{\alpha}_n$ tends to be very small, it is appropriate to measure the accuracy of the estimation as a relative statistical error, which we estimate by the ratio of the standard deviation and the mean

$$\epsilon_{st}^{MC} = z_c\frac{\sqrt{\mathbb{V}[\hat{\alpha}_n^{MC}]}}{\hat{\alpha}_n^{MC}}.$$

The method of numerical approximation to (13) is desribed in Appendix C. A numerical approximation of the corresponding optimal control is obtained by a central difference with linear interpolation throughout the domain (see (Ben Rached et al. 2022, Appendix E) for more details).

4.1 Double-Well problem

We consider nonlinear dynamics with a constant diffusion and a drift arising for a double well (DW) potential

$$du_t = -V'(u_t)dt + bdW_t,$$

where $V(u) = \frac{1}{2+4u^2} + \frac{u^2}{4}$. Figure 3 serves an illustrative purpose, visualizing the metastable behaviour of the DW dynamics between two wells over $T = 100$ observation

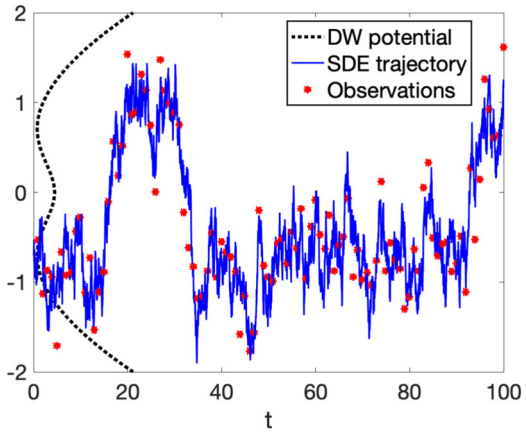


Fig. 3 A trajectory of the DW SDE from Section 4.1 over $T = 100$ observation times with given observations and the double-well potential. The model parameters: $b = 0.5$, $u_0 \sim N(-0.7, 0.1)$. The EnKF parameters: $H = 1, \Gamma = 0.1$.

times. Notice that the trajectory tends to remain in one well for a considerably longer time horizon and transit to another well.

In Figure 4, we illustrate a comparison of the approximations to the optimal IS initial density $\tilde{\rho}_{u_0}$ obtained via the PDE-based and multilevel CE-based methods when the original initial density ρ_{u_0} is given in one well and the threshold $\mathcal{K}$ is set in the other well. Three subplots demonstrate how the IS densities change from the original one in the cases of different initial standard deviation σ_0. With a relatively small σ_0, all IS densities shift towards the threshold slightly, whereas with larger σ_0 values, they shift significantly to the rare event region. In particular, the CE-based IS density exhibits a considerable shift towards the threshold, however, it has the same scale as the original density by design. The PDE-based IS density $\tilde{\rho}_{u_0}^{PDE,2}$ wrt both initial condition and Wiener process

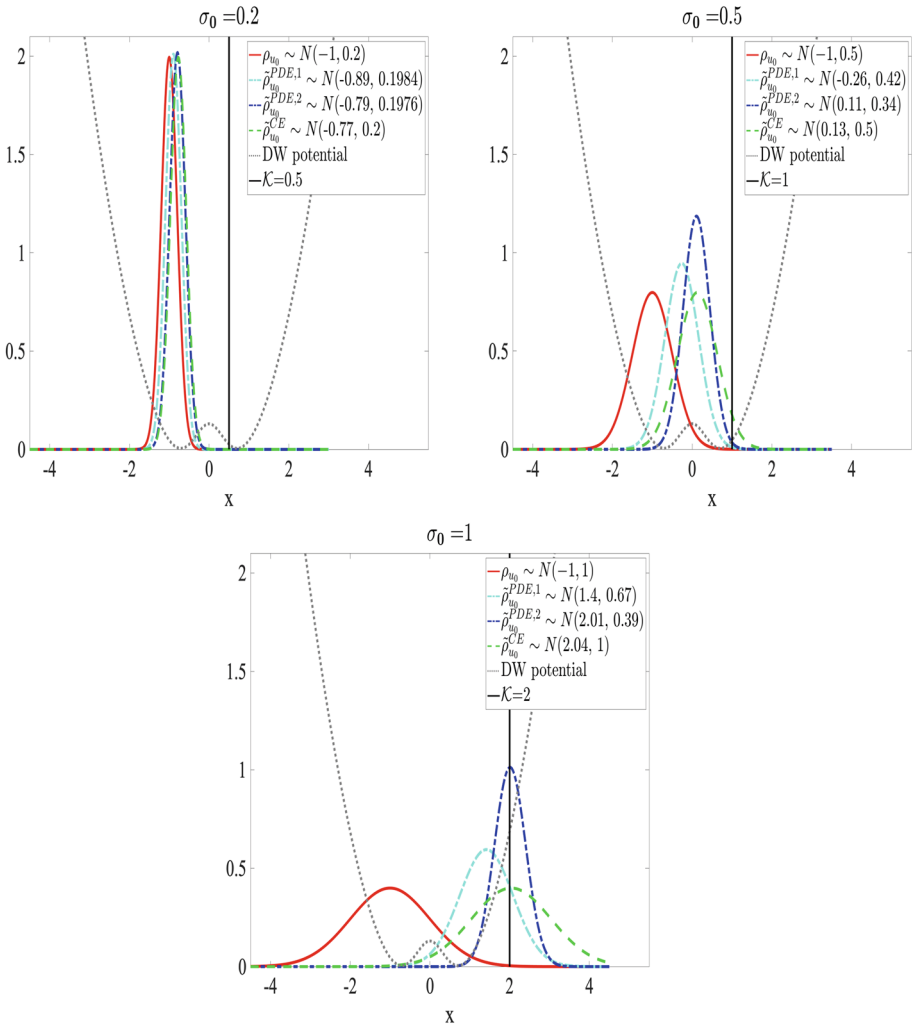


Fig. 4 Double Well example. The original initial density ($\rho_{u_0} \sim N(-1, \sigma_0)$ in red solid line) is centered in one well with the threshold $\mathcal{K}$ in the other well. For all three values of σ_0, $\mathcal{K}$ is such that the rare event probability is around 10^{-3}. A comparison of the approximated optimal initial densities based on: IS wrt ρ_0 ($\tilde{\rho}_{u_0}^{PDE,1}$ in cyan dash-dotted line), IS wrt both ρ_0 and $W(t)$ ($\tilde{\rho}_{u_0}^{PDE,2}$ in blue dash-dotted line), and CE-based IS ($\tilde{\rho}_{u_0}^{CE}$ in green dashed line)

Table 1 Double Well example. Model parameter: $b = 0.5$. Simulation parameters: $T = 1$, $\Delta t = 0.01$, $J = 10^6$, $u_0 \sim N(\mu_0, \sigma_0)$ with $\mu_0 = -1$, $\sigma_0 = 0.2$. Numerical results for the last two lines of the CE method are missing due to the impracticability of the algorithm for the given parameter setting and a limit of computer capacity to accomplish the simulation

$\mathcal{K}$	**Estimator with 95% CI**				
	$\hat{\alpha}^{MC}$	$\hat{\alpha}^{CE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,W_t}$	$\hat{\alpha}^{PDE,\text{both}}$
0	$\mathbf{4.65 \times 10^{-2}}$	$\mathbf{4.66 \times 10^{-2}}$	$\mathbf{4.65 \times 10^{-2}}$	$\mathbf{4.63 \times 10^{-2}}$	$\mathbf{4.63 \times 10^{-2}}$
	[0.0461, 0.0469]	[0.0462, 0.0470]	[0.0461, 0.0469]	[0.0462, 0.0464]	[0.04626, 0.04632]
0.5	$\mathbf{4.83 \times 10^{-3}}$	$\mathbf{4.76 \times 10^{-3}}$	$\mathbf{4.84 \times 10^{-3}}$	$\mathbf{4.81 \times 10^{-3}}$	$\mathbf{4.82 \times 10^{-3}}$
	[0.0047, 0.0050]	[0.0046, 0.0049]	[0.0047, 0.0050]	[0.00480, 0.00483]	[0.004815, 0.004822]
1	$\mathbf{1.81 \times 10^{-4}}$	-	$\mathbf{1.90 \times 10^{-4}}$	$\mathbf{1.89 \times 10^{-4}}$	$\mathbf{1.89 \times 10^{-4}}$
	[1.6e-04, 2.1e-04]		[1.7e-04, 2.1e-04]	[1.88e-04, 1.90e-04]	[1.891e-04, 1.895e-04]
1.2	$\mathbf{3.70 \times 10^{-5}}$	-	$\mathbf{3.18 \times 10^{-5}}$	$\mathbf{3.12 \times 10^{-5}}$	$\mathbf{3.13 \times 10^{-5}}$
	[2.5e-05, 4.9e-05]		[2.4e-05, 4.0e-05]	[3.1e-05, 3.2e-05]	[3.125e-05, 3.132e-05]

$\mathcal{K}$	**Relative statistical error**					**Variance reduction**			
	ϵ_{st}^{MC}	$\epsilon_{st}^{CE,\tilde{\rho}_0}$	$\epsilon_{st}^{PDE,\tilde{\rho}_0}$	ϵ_{st}^{PDE,W_t}	$\epsilon_{st}^{PDE,\text{both}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{CE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,W_t}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\text{both}}}$
0	**0.9%**	**0.8%**	**0.9%**	**0.2%**	**0.06%**	**1.00**	**1.14**	**28**	**208**
0.5	**2.8%**	**3.0%**	**2.5%**	**0.3%**	**0.07%**	**0.90**	**1.29**	**96**	**1534**
1	**14.6%**	-	**10.4%**	**0.5%**	**0.09%**	-	**1.78**	**691**	**20992**
1.2	**32.2%**	-	**26.2%**	**0.7%**	**0.11%**	-	**2.06**	**2975**	**124939**

Table 2 Double Well example. Model parameter: $b = 0.5$. Simulation parameters: $T = 1$, $\Delta t = 0.01$, $J = 10^6$, $u_0 \sim N(\mu_0, \sigma_0)$ with $\mu_0 = -1$, $\sigma_0 = 1$

$\mathcal{K}$	**Estimator with 95% CI**				
	$\hat{\alpha}^{MC}$	$\hat{\alpha}^{CE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,W_t}$	$\hat{\alpha}^{PDE,\text{both}}$
1.5	$\mathbf{1.48 \times 10^{-2}}$	$\mathbf{1.47 \times 10^{-2}}$	$\mathbf{1.48 \times 10^{-2}}$	$\mathbf{1.47 \times 10^{-2}}$	$\mathbf{1.48 \times 10^{-2}}$
	[0.0145, 0.0150]	[0.0147, 0.0148]	[0.0147, 0.0149]	[0.0145, 0.0149]	[0.01479, 0.01480]
2	$\mathbf{2.5 \times 10^{-3}}$	$\mathbf{2.58 \times 10^{-3}}$	$\mathbf{2.59 \times 10^{-3}}$	$\mathbf{2.53 \times 10^{-3}}$	$\mathbf{2.58 \times 10^{-3}}$
	[0.0024, 0.0026]	[0.00256, 0.00260]	[0.00257, 0.00260]	[0.0025, 0.0026]	[0.002578, 0.002584]
2.5	$\mathbf{3.8 \times 10^{-4}}$	$\mathbf{3.98 \times 10^{-4}}$	$\mathbf{3.98 \times 10^{-4}}$	$\mathbf{3.81 \times 10^{-4}}$	$\mathbf{3.97 \times 10^{-4}}$
	[3.4e-04, 4.2e-04]	[3.96e-04, 4.00e-04]	[3.96e-04, 3.99e-04]	[3.5e-04, 4.1e-04]	[3.96e-04, 3.97e-04]
3	$\mathbf{5.8 \times 10^{-5}}$	$\mathbf{5.06 \times 10^{-5}}$	$\mathbf{5.06 \times 10^{-5}}$	$\mathbf{5.98 \times 10^{-5}}$	$\mathbf{5.07 \times 10^{-5}}$
	[4.3e-05, 7.3e-05]	[5.03e-05, 5.09e-05]	[5.04e-05, 5.09e-05]	[4.6e-05, 7.3e-05]	[5.06e-05, 5.07e-05]

$\mathcal{K}$	**Relative statistical error**					**Variance reduction**			
	ϵ_{st}^{MC}	$\epsilon_{st}^{CE,\tilde{\rho}_0}$	$\epsilon_{st}^{PDE,\tilde{\rho}_0}$	ϵ_{st}^{PDE,W_t}	$\epsilon_{st}^{PDE,\text{both}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{CE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,W_t}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\text{both}}}$
1.5	**1.6**%	**0.5**%	**0.4**%	**1.3**%	**0.06**%	**9**	**15**	**1.6**	**677**
2	**3.9**%	**0.6**%	**0.5**%	**3.2**%	**0.10**%	**40**	**61**	**1.4**	**1327**
2.5	**10.1**%	**0.6**%	**0.5**%	**8.7**%	**0.10**%	**261**	**360**	**1.3**	**6150**
3	**25.7**%	**0.6**%	**0.4**%	**22.4**%	**0.10**%	**2641**	**4580**	**1.2**	**46372**

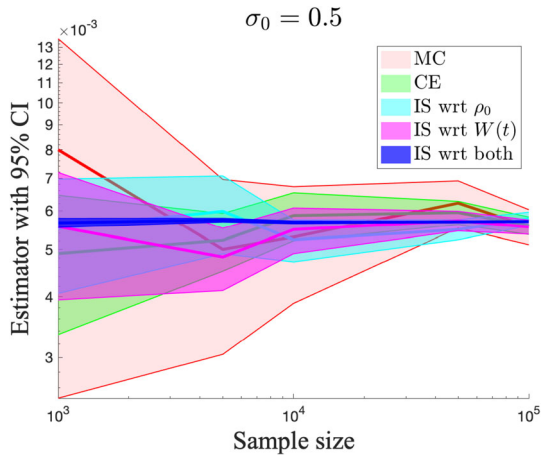


Fig. 5 Double Well example Model parameter: $b = 0.5$. 95% CI of the estimator with the simulation parameters: $T = 1$, $\Delta t = 0.01$, $\mathcal{K} = 1$, $u_0 \sim N(\mu_0, \sigma_0)$ with $\mu_0 = -1$, $\sigma_0 = 0.5$.

has a larger shift and a higher concentration in the rare event region compared to the initial density $\tilde{\rho}_{u_0}^{PDE,1}$ obtained by IS wrt the initial condition. This is because of the difference in the optimal IS initial density formulas of each approach presented in (12) and (26).

The IS approach wrt the initial condition gives a substantial variance reduction only for comparatively large standard deviation values (see Table 2 for the example $\sigma_0 = 1$). In the regime of small σ_0, the IS approach wrt Wiener processes via the SOC theory proposed in Section 3.2 works well (see Table 1 for the example $\sigma_0 = 0.2$. The combination of both approaches (i.e., IS wrt both initial condition and Wiener processes) provides considerably better variance reduction over a crude MC method for all three cases of σ_0 (see Table 1 and 2). Figure 5 illustrates the consistency of the rare event probability estimates produced by different methods for the case of $\sigma_0 = 0.5$, where the 95% confidence intervals (CI) overlap and converge as the MC sample size increases, and even with relatively small sample sizes, the IS wrt both ρ_0 and W_t demonstrates superior improvement in reducing variance. A robustness of the numerical results is verified via bootstrapping. In particular, Figure 6 shows 95% CI of two independent runs, which are overlapping, for

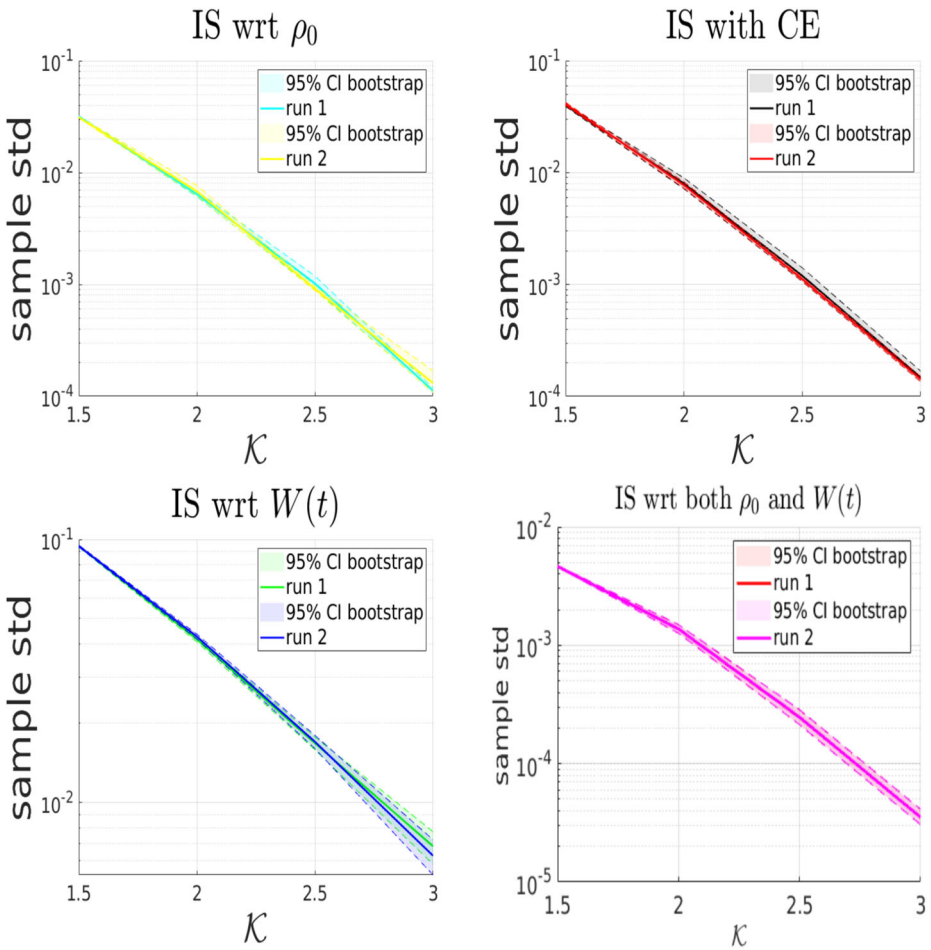


Fig. 6 Double Well example. Two independent simulation results of the bootstrapping for a sample standard deviation produced by the CE- and PDE-based methods with parameters of the Table 2 and with the bootstrap number $B = 10^4$.

the sample standard deviation (std) produced by the CE- and the PDE-based methods via bootstrap sampling with a bootstrap number $B = 10^4$ over $J = 10^6$ i.i.d. samples for the parameter setting of Table 2.

4.2 Langevin dynamics

We consider two-dimensional Langevin dynamics

$$\begin{cases} du_t = v_t dt, \\ dv_t = -V'(u_t)dt - \kappa v_t dt + (2\kappa\mathcal{T})^{1/2} dW_t, \end{cases} \tag{29}$$

where V is the same double-well potential as in Section 4.1. Figure 7 illustrates trajectories of the Langevin dynamics in different temperature $\mathcal{T}$ settings. The rare event threshold is defined by the velocity, so that the projection is $P_1 = [0\ 1]$. In Figure 8, a change of the initial density towards the importance region with $\mathcal{K} = 3$ is shown for, respectively, PDE- and CE-based IS techniques wrt ρ_0.

Similarly to the previous DW example, from Table 3, we can observe that the combination of the PDE-based method wrt the initial distribution and Wiener paths reduces the variance significantly compared to its single approaches and the multilevel CE approach. Figure 9 shows the consistency of the estimates produced by different methods for the case of correlated initial condition $\begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$, where the 95% confidence intervals (CI) overlap and converge as the MC sample size increases, and the IS wrt both ρ_0 and W_t provides the highest level of variance reduction when compared to alternative methods. Once again, Figure 10 exhibits a robustness of methods by displaying 95% CI of two independent pilot runs for the parameter setting indicated in Table 3.

4.3 A noisy Charney-deVore model

We consider the six-dimensional Charney–deVore (CdV) model polluted with additive Gaussian noise (Grafke and Vanden-Eijnden 2019)

$$\begin{cases} du_1 = (\tilde{\gamma}_1 u_3 - C(u_1 - u_1^*))dt + \sqrt{2b}dW_1, \\ du_2 = (-(\zeta_1 u_1 - \beta_1)u_3 - Cu_2 - \delta_1 u_4 u_6)dt + \sqrt{2b}dW_2, \\ du_3 = ((\zeta_1 u_1 - \beta_1)u_2 - \gamma_1 u_1 - Cu_3 + \delta_1 u_4 u_5)dt + \sqrt{2b}dW_3, \\ du_4 = (\tilde{\gamma}_2 u_6 - C(u_4 - u_4^*) + \eta(u_2 u_6 - u_3 u_5))dt + \sqrt{2b}dW_4, \\ du_5 = (-(\zeta_2 u_1 - \beta_2)u_6 - Cu_5 - \delta_2 u_3 u_4)dt + \sqrt{2b}dW_5, \\ du_6 = ((\zeta_2 u_1 - \beta_2)u_5 - \gamma_2 u_4 - Cu_6 + \delta_2 u_2 u_4)dt + \sqrt{2b}dW_6, \end{cases} \tag{30}$$

where for $m \in 1, 2$,

$$\zeta_m = \frac{8\sqrt{2}}{\pi}\frac{m^2}{4m^2-1}\frac{q^2+m^2-1}{b^2+m^2}, \quad \beta_m = \frac{\beta q^2}{q^2+m^2},$$
$$\gamma_m = \gamma\frac{\sqrt{2}q}{\pi}\frac{4m^3}{(4m^2-1)(q^2+m^2)}, \quad \tilde{\gamma}_m = \gamma\frac{\sqrt{2}q}{\pi}\frac{4m}{4m^2-1},$$
$$\delta_m = \frac{64\sqrt{2}}{15\pi}\frac{q^2-m^2+1}{q^2+m^2}, \quad \eta = \frac{16\sqrt{2}}{5\pi}.$$

The CdV was among the pioneering atmospheric models that demonstrated multiple invariant measures. The model (30) features two distinct metastable modes: one is the "zonal" mode and the other is the "blocked" mode. These modes correspond to atmospheric blocking phenomena commonly observed in meteorology (Charney and DeVore 1979; Crommelin et al. 2004; Majda and Harlim 2012).

In all numerical tests, we use the parameters provided in Crommelin et al. (2004) which result in a chaotic behavior of dynamics: zonally symmetric forcing terms $u_1^* = 0.95$ and $u_4^* = -0.76095$; the thermal relaxation damping timescale $C = 0.1$ corresponding to 10 days; the orographic height $\gamma = 0.2$ corresponding to a 200m amplitude; the Coriolis parameter $\beta = 1.25$ defining a central latitude of 45^o; the channel width–length ratio $q = 0.5$ corresponding to a channel of 6300km $\times$ 1600km. For simplicity, we consider

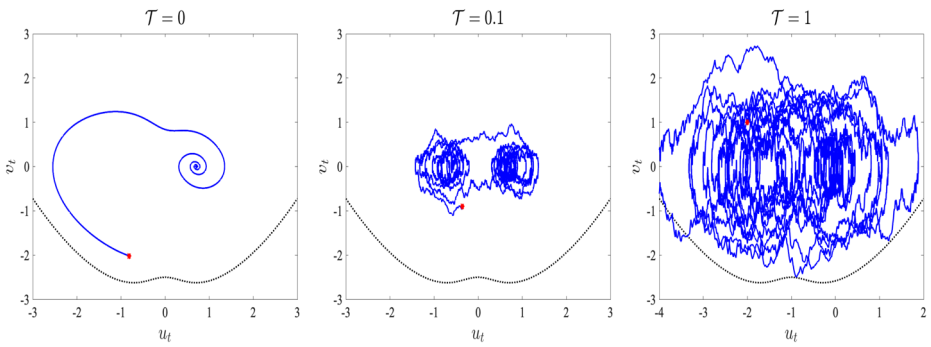


Fig. 7 Example: trajectories of the Langevin dynamics with different temperature $\mathcal{T}$ values up to time $T = 100$. The red star represents the initial value and the dashed black line represents the double well potential.

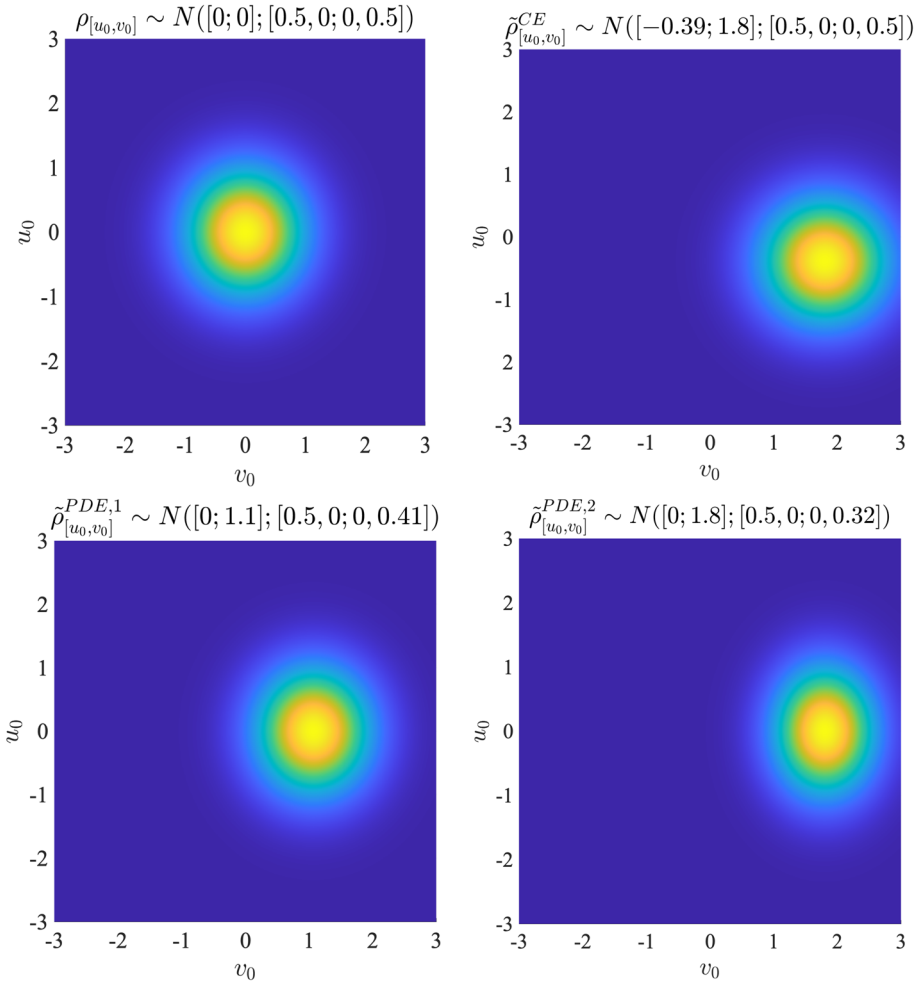


Fig. 8 Langevin dynamics. A comparison of the original initial density (top left) with the approximately optimal IS initial densities based on the multilevel CE method (top right), the PDE approaches wrt only initial condition (bottom left) and wrt both initial condition and Wiener process (bottom right).

the rare event defined by the first component, i.e., the projection $P_1 = [1\ 0\ 0\ 0\ 0\ 0]$. We set the initial condition $\boldsymbol{u}_0 := [u_1(0)\ u_2(0)\ u_3(0)\ u_4(0)\ u_5(0)\ u_6(0)]$ such that the first component is sampled from the blocked region

$$\boldsymbol{u}_0 \sim N([0.7650, 0.2288, -0.2990, -0.3657, -0.1636, 0.3108], \sigma_0^2 I)$$

where I is a 6-dimensional identity matrix and the rare event threshold is defined from the zonal region. The metastable feature of the dynamics can be observed in Figure 11, which illustrates a sample trajectory of the tracked component u_1 for a given different values of the diffusion parameter $b = 0,\ 0.0001,\ 0.001,\ 0.01$ over time $T = 1000$.

In a similar fashion to previous examples, Table 4 provides numerical results of the CdV model with the diffusion parameter $b = 0.01$. In this case, we observe a considerable variance reduction in the PDE-based IS technique wrt the Wiener process and a much more significant reduction in the IS approach wrt both the initial condition and the Wiener process. Similar to the Double Well problem in Section 4.1, it is expected to get a variance reduction in IS wrt the initial condition if much larger values of σ_0^2 are considered in the initial condition. Furthermore, we provide the bootstrapping results for the sample standard deviation in Figure 12 for the parameter setting indicated in Table 4 as numerical evidence of the robustness of the considered methods. We have to note that when a diffusion coefficient is very small, the simple numerical PDE solvers with uniform discretization (see Appendix C) may not provide accurate solutions, and special adaptive mesh refinement strategies may be needed to solve the PDEs numerically. Instead, the large deviation principle-based methods might be more efficient in a very small noise regime (Grafke and Vanden-Eijnden 2019; Vanden-Eijnden and Weare 2013).

Table 3 **Langevin dynamics.** The model parameters: $\kappa = 2^{-5}\pi^2$, $\mathcal{T} = 1$. Simulation parameters: $T = 1$, $\Delta t = 0.01$, $J = 10^6$, $[u_0, v_0] \sim N([0\ 0], [0.5\ 0; 0\ 0.5])$. PDE-based methods uses $\Delta x_{\mathbf{PDE}} = 0.006$, $\Delta t_{\mathbf{PDE}} = \Delta x_{\mathbf{PDE}}/5$.

$\mathcal{K}$	**Estimator with 95% CI**				
	$\hat{\alpha}^{MC}$	$\hat{\alpha}^{CE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,W_t}$	$\hat{\alpha}^{PDE,\text{both}}$
2	$\mathbf{2.53 \times 10^{-2}}$	$\mathbf{2.55 \times 10^{-2}}$	$\mathbf{2.53 \times 10^{-2}}$	$\mathbf{2.54 \times 10^{-2}}$	$\mathbf{2.56 \times 10^{-2}}$
	[0.025, 0.026]	[0.0252, 0.0256]	[0.0251, 0.0255]	[0.0252, 0.0256]	[0.02559, 0.02569]
2.5	$\mathbf{4.2 \times 10^{-3}}$	$\mathbf{4.2 \times 10^{-3}}$	$\mathbf{4.25 \times 10^{-3}}$	$\mathbf{4.17 \times 10^{-3}}$	$\mathbf{4.27 \times 10^{-3}}$
	[0.0041, 0.0043]	[0.0041, 0.0043]	[0.0042, 0.0043]	[0.0041, 0.0042]	[0.00426, 0.00428]
3	$\mathbf{4.7 \times 10^{-4}}$	$\mathbf{4.6 \times 10^{-4}}$	$\mathbf{4.7 \times 10^{-4}}$	$\mathbf{4.8 \times 10^{-4}}$	$\mathbf{4.68 \times 10^{-4}}$
	[4.3e-04, 5.1e-04]	[4.5e-04, 4.8e-04]	[4.5e-04, 4.8e-04]	[4.6e-04, 4.9e-04]	[4.67e-04, 4.69e-04]
3.5	$\mathbf{3.8 \times 10^{-5}}$	$\mathbf{3.3 \times 10^{-5}}$	$\mathbf{3.4 \times 10^{-5}}$	$\mathbf{3.3 \times 10^{-5}}$	$\mathbf{3.43 \times 10^{-5}}$
	[2.6e-05, 5.0e-05]	[3.0e-05, 3.6e-05]	[3.1e-05, 3.7e-05]	[3.1e-05, 3.6e-05]	[3.42e-05, 3.44e-05]

$\mathcal{K}$	**Relative statistical error**					**Variance reduction**			
	ϵ_{st}^{MC}	$\epsilon_{st}^{CE,\tilde{\rho}_0}$	$\epsilon_{st}^{PDE,\tilde{\rho}_0}$	ϵ_{st}^{PDE,W_t}	$\epsilon_{st}^{PDE,\text{both}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{CE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,W_t}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\text{both}}}$
2	**1.21%**	**1.18%**	**0.8%**	**0.7%**	**0.2%**	**1.1**	**2.6**	**3.5**	**35**
2.5	**3.0%**	**2.5%**	**1.5%**	**1.5%**	**0.2%**	**1.3**	**3.9**	**5.8**	**180**
3	**9.1%**	**3.8%**	**3.5%**	**2.9%**	**0.2%**	**5.7**	**6.6**	**9.3**	**2000**
3.5	**31.8%**	**8.9%**	**9.1%**	**7.3%**	**0.3%**	**15.2**	**15.4**	**24.7**	**12156**

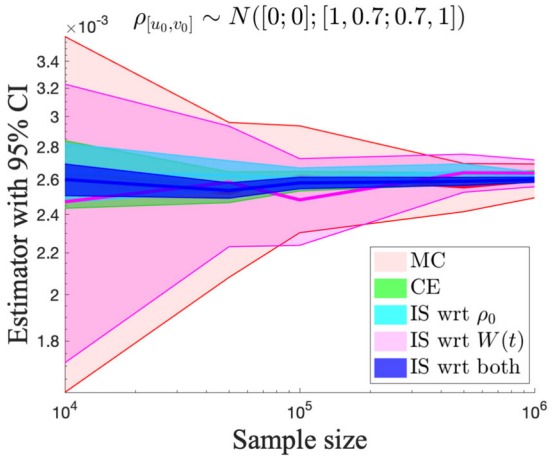


Fig. 9 Langevin dynamics. The model parameters: $\kappa = 2^{-5}\pi^2$, $\mathcal{T} = 1$. Simulation parameters: $T = 1$, $\Delta t = 0.01$, $\mathcal{K} = 2.5$, $[u_0, v_0] \sim N([0\ 0], [1\ 0.7; 0.7\ 1])$. PDE-based methods uses $\Delta x_{\mathbf{PDE}} = 0.006$, $\Delta t_{\mathbf{PDE}} = \Delta x_{\mathbf{PDE}}/5$.

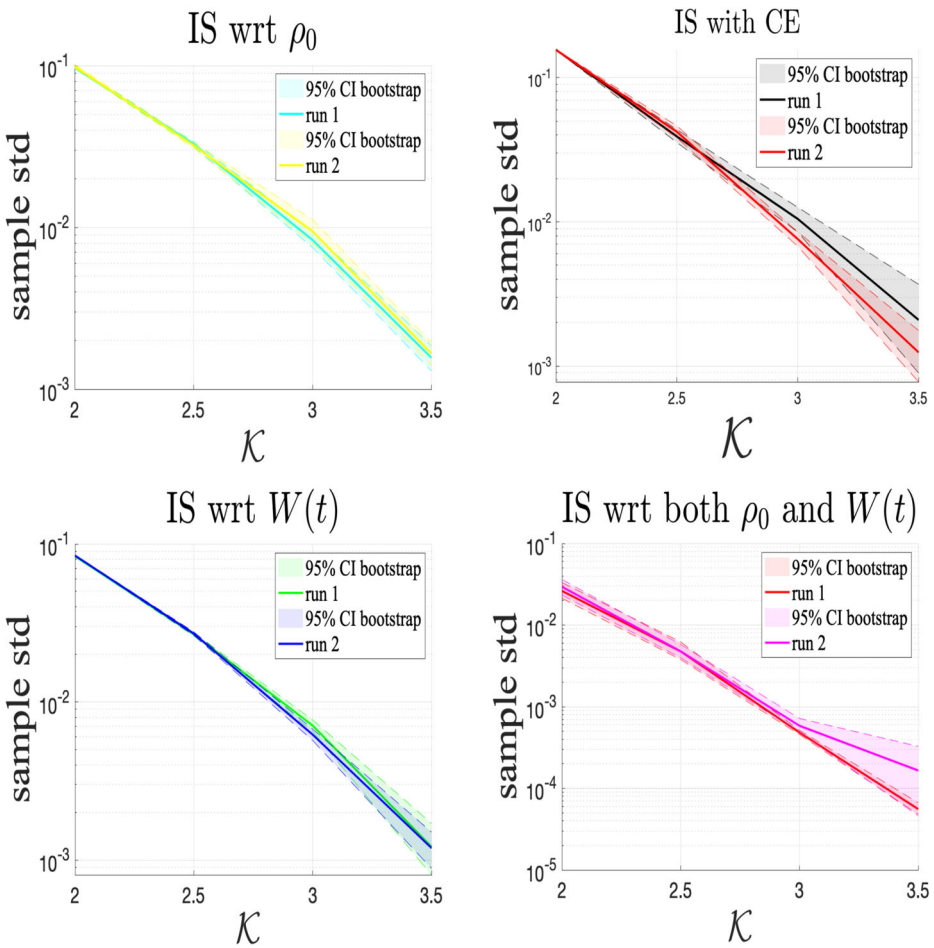


Fig. 10 Langevin dynamics. Two independent simulation results of the bootstrapping for a sample standard deviation produced by the CE- and PDE-based methods with parameters of the Table 3 and with the bootstrap number $B = 10^4$

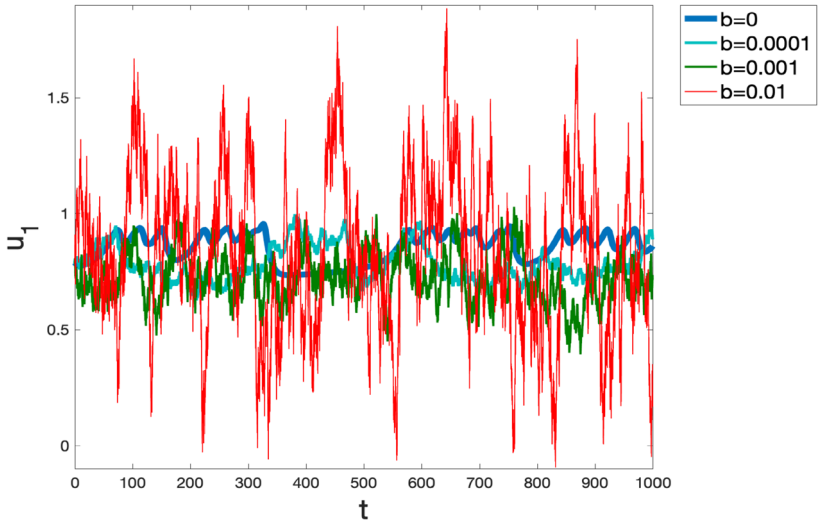


Fig. 11 Charney-deVore model. Sample trajectories of the first component corresponding to different values of the diffusion parameter b over $T = 1000$ time with $\Delta t = 0.01$.

5 Conclusion

In this study, we made use of IS techniques to monitor a rare event probability exceeding a critical threshold in relation to the running maximum associated with the solution of an SDE. We conducted this investigation within the framework of EnKF. In the interval between two observations in the EnKF method, we put forward three IS strategies: one that changes a measure wrt the SDE's initial condition, another that involves a change of measure wrt the Wiener process using a SOC formulation, and a third that combines these IS approaches for both the initial condition and the Wiener process. To our knowledge, the use of IS with respect to the initial condition and its combination with path-space IS has not been previously explored or tested in the filtering and SDEs literature. All of these IS methods hinge on approximating the solution of the KBE while considering specific boundary conditions. In more than one-dimensional setting, we employed a dimension reduction technique through the Markovian projection. This technique allows us to approximate the KBE solution by solving a simpler one-dimensional PDE.

A key feature of our methodology is that these rare event estimations are carried out in a post-processing step, independently at each observation time, using only the EnKF output. This design ensures flexibility and modularity in practical data assimilation workflows. It was demonstrated through numerical examples that the proposed estimation methods significantly reduced relative variance compared to the standard MC and the multilevel CE techniques.

The future research trajectory will involve extending the proposed IS methodology to particle filtering frameworks (Crisan and Doucet 2002), which can provide more accurate filtering distributions in highly nonlinear or non-Gaussian systems. This extension would enable broader applicability and facilitate a more comprehensive comparison between filtering methods in the context of rare event estimation.

Appendix A A discrete L^2 regression

We want to approximate $\breve{a}$ and $\breve{b}$, in (16), by linear combinations of bivariate polynomials defined as

$$\psi_p(t,s) = t^{p_1} s^{p_2}, \qquad p = (p_1, p_2) \in \mathbb{N}^2.$$

Restrict p to a tensor product index set $\Lambda(\varpi) \subset \mathbb{N}^2$ such that $\max_{p\in\Lambda} \max_{k=1,2} p_k \leq \varpi$ and let $\mathcal{V}_{\Lambda(\varpi)} = \text{span}\{\psi_p, p \in \Lambda(\varpi)\}$. We approximate $\breve{a}$ and $\breve{b}$ by $\hat{\breve{a}}$ and $\hat{\breve{b}}$ respectively, obtained by computing

$$\begin{cases} \hat{\breve{a}} = \arg\min_{h_a\in\mathcal{V}_\Lambda} \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=1}^{M} |P_1 a(u^{(i)}(t_n)) \\ -h_a(t_n, P_1 u^{(i)}(t_n))|^2, \\ \hat{\breve{b}}^2 = \arg\min_{h_b\in\mathcal{V}_\Lambda} \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=1}^{M} |(P_1 b b^T P_1^T)(u^{(i)}(t_n)) \\ -h_b(t_n, P_1 u^{(i)}(t_n))|^2. \end{cases} \tag{31}$$

First, in order to ensure the stabilty, we consider *the orthonormalised basis expansion* $\bar{\psi}_p$ via orthonormalising ψ_p by the modified Gram-Schmidt algorithm wrt a discrete scalar product for $\forall p, q \in \Lambda(\varpi)$:

$$< \psi_p, \psi_q >= \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{i=1}^{M} \psi_p(t_n, P_1 u^{(i)}(t_n)) \psi_q(t_n, P_1 u^{(i)}(t_n)).$$

Table 4 A noisy Charney-deVore model with $b = 0.01$. Simulation parameters: $\sigma_0^2 = 0.0025$, $T = 1$, $\Delta t = 0.01$, $J = 10^6$. Numerical results for the last line of the CE method is missing due to the impracticability of the algorithm for the given parameter setting and a limit of computer capacity to accomplish the simulation

$\mathcal{K}$	**Estimator with 95% CI**				
	$\hat{\alpha}^{MC}$	$\hat{\alpha}^{CE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,\tilde{\rho}_0}$	$\hat{\alpha}^{PDE,W_t}$	$\hat{\alpha}^{PDE,\text{both}}$
1.1	$\mathbf{3.1 \times 10^{-2}}$	$\mathbf{3.1 \times 10^{-2}}$	$\mathbf{3.1 \times 10^{-2}}$	$\mathbf{3.08 \times 10^{-2}}$	$\mathbf{3.08 \times 10^{-2}}$
	[0.0308, 0.0314]	[0.0308, 0.0315]	[0.0308, 0.0314]	[0.0307, 0.0309]	[0.0308, 0.0309]
1.2	$\mathbf{4.2 \times 10^{-3}}$	$\mathbf{4.2 \times 10^{-3}}$	$\mathbf{4.2 \times 10^{-3}}$	$\mathbf{4.1 \times 10^{-3}}$	$\mathbf{4.11 \times 10^{-3}}$
	[0.0041, 0.0043]	[0.0040, 0.0043]	[0.0041, 0.0043]	[0.00409, 0.00412]	[0.00411, 0.00412]
1.3	$\mathbf{3.3 \times 10^{-4}}$	$\mathbf{3.4 \times 10^{-4}}$	$\mathbf{3.44 \times 10^{-4}}$	$\mathbf{3.5 \times 10^{-4}}$	$\mathbf{3.45 \times 10^{-4}}$
	[2.9e-04, 3.7e-04]	[3.0e-04, 3.8e-04]	[3.2e-04, 3.8e-04]	[3.42e-04, 3.46e-04]	[3.44e-04, 3.46e-04]
1.4	$\mathbf{2.1 \times 10^{-5}}$	–	$\mathbf{1.5 \times 10^{-5}}$	$\mathbf{1.8 \times 10^{-5}}$	$\mathbf{1.81 \times 10^{-5}}$
	[1.2e-05, 3.0e-05]		[8.9e-06, 2.2e-05]	[1.78e-05, 1.81e-05]	[1.80e-05, 1.81e-05]

$\mathcal{K}$	**Relative statistical error**					**Variance reduction**			
	ϵ_{st}^{MC}	$\epsilon_{st}^{CE,\tilde{\rho}_0}$	$\epsilon_{st}^{PDE,\tilde{\rho}_0}$	ϵ_{st}^{PDE,W_t}	$\epsilon_{st}^{PDE,\text{both}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{CE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\tilde{\rho}_0}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,W_t}}$	$\frac{\mathbb{V}_{MC}}{\mathbb{V}_{PDE,\text{both}}}$
1.1	**1.1**%	**1.1**%	**1.0**%	**0.2**%	**0.1**%	**1.0**	**1.1**	**21**	**78**
1.2	**3.0**%	**3.1**%	**2.6**%	**0.4**%	**0.2**%	**1.0**	**1.3**	**75**	**296**
1.3	**11**%	**12**%	**8.5**%	**0.5**%	**0.2**%	**0.8**	**1.4**	**394**	**1985**
1.4	**43**%	–%	**42**%	**0.7**%	**0.3**%	–	**2.0**	**5395**	**38467**

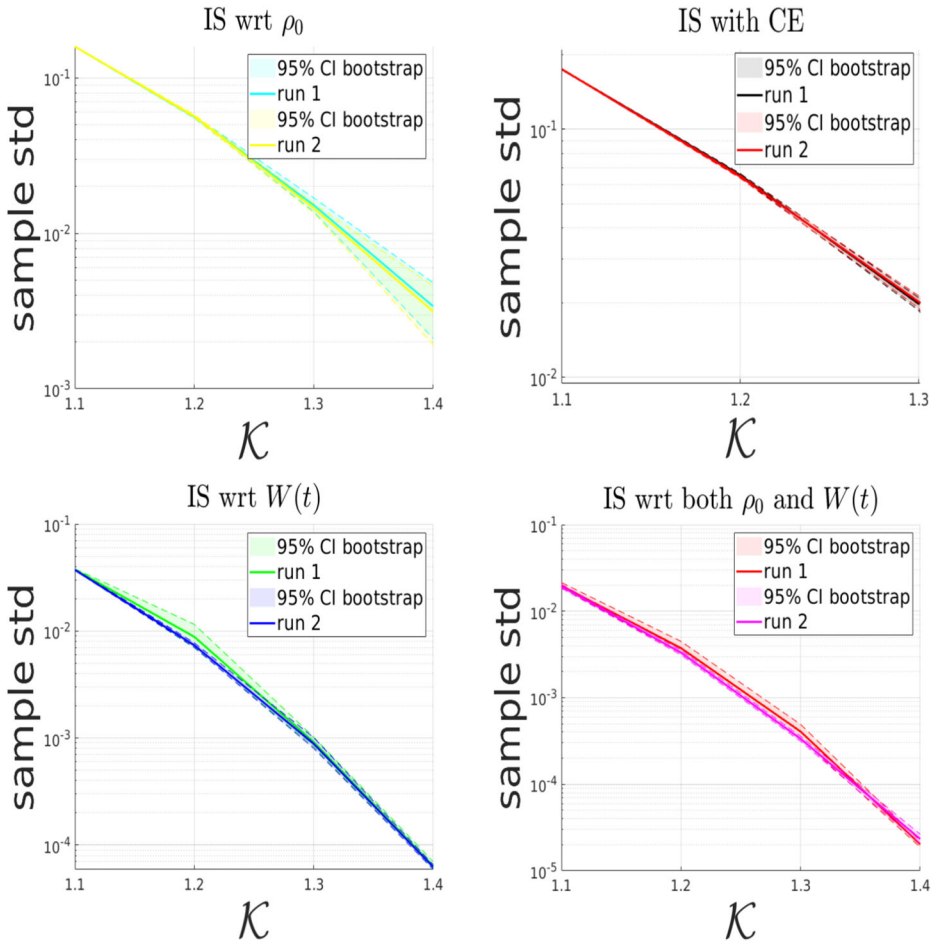


Fig. 12 Charney-deVore model. Two independent simulation results of the bootstrapping for a sample standard deviation produced by the CE- and PDE-based methods with parameters of the Table 4 and with the bootstrap number $B = 10^4$.

The modified Gram-Schmidt algorithm produces a QR decomposition for non-orthonormalised matrix Ψ as follows:

$$\underbrace{\left[\psi_p^1 \middle| \psi_p^2 \middle| \dots \middle| \psi_p^{\#\Lambda}\right]}_{\Psi} = \underbrace{\left[\bar{\psi}_p^1 \middle| \bar{\psi}_p^2 \middle| \dots \middle| \bar{\psi}_p^{\#\Lambda}\right]}_{Q} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1\#\Lambda} \\ & r_{22} & \dots & r_{2\#\Lambda} \\ & & \dots & \\ & & & r_{\#\Lambda\#\Lambda} \end{bmatrix}}_{R}$$

where $Q^T Q = I$ and R is the upper-triangular matrix. Note that since $Q = \Psi R^{-1}$, we can use the entries of R^{-1} to construct the orthonormalised function $\bar{\psi}_p(t, s)$.

Then, solving (31) is equivalent to solving the least squares problem for the expansion coefficients a_p and b_p:

$$\begin{cases} \hat{h}_a(t,s) = \sum_{p\in\Lambda} a_p \bar{\psi}_p(t,s), \\ \hat{h}_b(t,s) = \sum_{p\in\Lambda} b_p \bar{\psi}_p(t,s). \end{cases} \tag{32}$$

The expansion coefficients in (32) are computed via solving the respective normal equations for $\underline{a} = \{a_p\}_{p\in\Lambda}$ and $\underline{b} = \{b_p\}_{p\in\Lambda}$:

$$\begin{cases} Q^T Q \underline{a} = Q^T \underline{f} \\ Q^T Q \underline{b} = Q^T \underline{g}, \\ (\underline{f})_j = P_1 a(t_n, u^{(i)}(t_n)), \\ (\underline{g})_j = P_1 b b^T P_1^T (t_n, u^{(i)}(t_n)), \\ Q_{jp} = \bar{\psi}_p(t_n, P_1 u^{(i)}(t_n)) \text{ for } p \in \Lambda, \ j = 1, ..., NM. \end{cases} \tag{33}$$

Appendix B Multilevel Cross-Entropy method

We mainly follow the description of the method provided in De Boer et al. (2005); Kroese et al. (2013). In the framework of the EnKF, the optimal original initial density is assumed to be Gaussian. Let $\rho_0(x_0) = \phi(x_0; \boldsymbol{v})$ belong to a family of Gaussian densities with a parameter vector $\boldsymbol{v}$ which is approximated by the EnKF ensembles. Then, similarly,

to (12), the optimal importance sampling density is

$$\tilde{\rho}_0(x_0) \propto \phi(x_0; \boldsymbol{v}) \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = x_0)}. \quad (34)$$

We want to approximate $\tilde{\rho}_0(x_0)$ by a member of the Gaussian family, $\phi(x_0; \boldsymbol{\zeta})$, for some parameters $\boldsymbol{\zeta}$. We consider the *Kullback-Leibler divergence* defined as

$$\mathcal{D}(\tilde{\rho}_0(\cdot), \phi(\cdot; \boldsymbol{\zeta})) = \int (\log \tilde{\rho}_0(x_0) - \log \rho_0(x_0; \boldsymbol{\zeta})) \tilde{\rho}_0(x_0) dx_0. \quad (35)$$

This is also known as the cross-entropy between the given densities.

Finding a minimizer of (35) is equivalent to finding

$$\begin{aligned} \boldsymbol{\zeta}^* &= \operatorname*{argmax}_{\boldsymbol{\zeta}} -\mathcal{D}(\tilde{\rho}_0(\cdot), \phi(\cdot; \boldsymbol{\zeta})) = \operatorname*{argmax}_{\boldsymbol{\zeta}} \\ &\int \tilde{\rho}_0(x_0) \log \phi(x_0; \boldsymbol{\zeta}) dx_0 \\ &= \operatorname*{argmax}_{\boldsymbol{\zeta}} \int \phi(x_0, \boldsymbol{v}) \\ &\sqrt{\mathbb{P}\Big(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = x_0\Big)} \log \phi(x_0; \boldsymbol{\zeta}) dx_0 \\ &= \operatorname*{argmax}_{\boldsymbol{\zeta}} \mathbb{E}_{\boldsymbol{w}}\Big[\sqrt{\mathbb{P}\Big(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = X_0\Big)} \\ &\log \phi(X_0; \boldsymbol{\zeta}) L(X_0; \boldsymbol{v}, \boldsymbol{w})\Big] \end{aligned} \quad (36)$$

where the likelihood ratio is defined by

$$L(x_0; \boldsymbol{v}, \boldsymbol{w}) = \frac{\rho_0(x_0; \boldsymbol{v})}{\phi(x_0; \boldsymbol{w})}.$$

$\mathbb{E}_{\boldsymbol{w}}[\cdot]$ means the expectaction is taken wrt the density $\phi(\cdot; \boldsymbol{w})$, which is, in our setting, another family of Gaussian densities with an arbitrary parameter vector $\boldsymbol{w}$. The reason of introducing it is to estimate $\boldsymbol{\zeta}^*$ using random samples $\{X_0^{(j)}\}_{j=1}^J$ from $\phi(\cdot; \boldsymbol{w})$. Then, the approximate optimization becomes

$$\max_{\boldsymbol{\zeta}} \underbrace{\frac{1}{J}\sum_{j=1}^{J} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = X_0^{(j)})} \log \rho_0(X_0^{(j)}; \boldsymbol{\zeta}) L(X_0^{(j)}; \boldsymbol{v}, \boldsymbol{w})}_{-\hat{D}}. \quad (37)$$

Assuming that $\hat{D}$ is convex and differentiable with respect to $\boldsymbol{\zeta}$, we solve the system of equations

$$\frac{1}{J}\sum_{j=1}^{J} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = X_0^{(j)})} \nabla \log \rho_0(X_0^{(j)}; \boldsymbol{\zeta}) L(X_0^{(j)}; \boldsymbol{v}, \boldsymbol{w}) = \mathbf{0}$$

where the gradient is with respect to $\boldsymbol{\zeta}$ and $\mathbf{0}$ denotes a zero vector in size of $\boldsymbol{v}$. We can often compute the analytical solution for the above problem if the random variable distributions belong to a natural exponential family, which is the case in the EnKF.

It is crucial to note that the optimization problem (37) can be difficult to carry out due to the rareness of the events $\{\max_{0\le k\le K} P_1 \bar{u}_{n,k} \ge \mathcal{K} | \bar{u}_{n,0} = X_0^{(j)}\}$. We can employ a so-called *multilevel* algorithm to overcome the difficulty with the rare event probability (De Boer et al. 2005). The main idea is to construct a sequence of reference parameters $\{\boldsymbol{\zeta}_\ell, \ell \ge 0\}$ and a sequence of threshold levels $\{\mathcal{K}_\ell, \ell \ge 1\}$ (see Algorithm 1).

Algorithm 1: Multilevel Cross-Entropy method for Rare-Event Probability

1 Define $\hat{\boldsymbol{\zeta}}_0 = \boldsymbol{v}$. Set $\ell = 1$.

2 Generate $\{\bar{u}_{n,0}^{(i)}\}_{i=1}^{J_1} \sim \phi(x_0; \hat{\boldsymbol{\zeta}}_{\ell-1})$ and $\{\bar{u}_{n,1:K}^{(i)}\}_{i=1}^{J_1} \sim \rho_{\bar{u}_{n,1:K}|\bar{u}_{n,0}^{(i)}}(x_{1:K}|x_0)$. Compute the sample $(1-\beta)-$quantile $\hat{\mathcal{K}}_\ell$ by calculating $\bar{M}_{\Delta t,(i)}^n = \max_{0\le k\le K} P_1 \bar{u}_{n,k}^{(i)}$ and arranging them in ascending order

$$\bar{M}_{\Delta t,(1)}^n \le \ldots \le \bar{M}_{\Delta t,(J_1)}^n$$

where β is an usually not small number, say $\beta \simeq 10^{-2}$. Evaluate $\hat{\mathcal{K}}_\ell = \bar{M}_{\Delta t,(\lceil (1-\beta)J_1 \rceil)}^n$ if $\hat{\mathcal{K}}_\ell < \mathcal{K}$, otherwise $\hat{\mathcal{K}}_\ell = \mathcal{K}$.

3 Given $\boldsymbol{\zeta}_{\ell-1}$, $\hat{\mathcal{K}}_\ell$ and the **same** sample $\bar{u}_{n,0:K}^{(1)}, \ldots, \bar{u}_{n,0:K}^{(J_1)}$, find

$$\hat{\boldsymbol{\zeta}}_\ell = \operatorname*{argmax}_{\boldsymbol{\zeta}} \frac{1}{J_1}\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k}^{(i)} \ge \mathcal{K} | \bar{u}_{n,0}^{(i)})} L(\bar{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\zeta}}_{\ell-1}) \log \phi(\bar{u}_{n,0}^{(i)}; \boldsymbol{\zeta})$$

4 If $\hat{\mathcal{K}}_\ell < \mathcal{K}$, set $\ell = \ell + 1$ and go to **Step 2**, otherwise, go to **Step 5**.

5 Estimate the rare event probability α_n by

$$\hat{\alpha}_n = \frac{1}{J_2}\sum_{i=1}^{J_2} \mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k}^{(i)} \ge \mathcal{K} | \bar{u}_{n,0}^{(i)}) L(\bar{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\zeta}}_{\mathcal{L}})$$

where $\mathcal{L}$ denotes the final number of levels/iterations.

For example, let $\boldsymbol{\zeta} = [\tilde{\mu}_0, \tilde{\sigma}_0]$ and $\phi(x_0, \boldsymbol{v}) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(x_0-\mu_0)^2/2\sigma_0^2}$, where $\boldsymbol{v} = [\mu_0, \sigma_0]$. Then, in **Step 3** of the Algorithm 1, we have to solve the following system of equations

$$\begin{cases} \frac{1}{J_1}\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k}^{(i)} \ge \mathcal{K} | \bar{u}_{n,0}^{(i)})} L(\bar{u}_{n,0}^{(i)}; \\ \boldsymbol{v}, \hat{\boldsymbol{\zeta}}_{\ell-1}) \Big(\frac{\bar{u}_{n,0}^{(i)} - \tilde{\mu}_0}{\tilde{\sigma}_0^2}\Big) = \mathbf{0} \\ \frac{1}{J_1}\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \bar{u}_{n,k}^{(i)} \ge \mathcal{K} | \bar{u}_{n,0}^{(i)})} L(\bar{u}_{n,0}^{(i)}; \\ \boldsymbol{v}, \hat{\boldsymbol{\zeta}}_{\ell-1}) \Big(\frac{(\bar{u}_{n,0}^{(i)} - \tilde{\mu}_0)^2}{\tilde{\sigma}_0^3} - \frac{1}{\tilde{\sigma}_0}\Big) = \mathbf{0} \end{cases} \quad (38)$$

which have the following analytical solutions

$$\begin{cases} \tilde{\mu}_{0,\ell} = \dfrac{\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \tilde{u}_{n,k}^{(i)} \ge \mathcal{K} | \tilde{u}_{n,0}^{(i)})} L(\tilde{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\xi}}_{\ell-1}) \tilde{u}_{n,0}^{(i)}}{\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \tilde{u}_{n,k}^{(i)} \ge \mathcal{K} | \tilde{u}_{n,0}^{(i)})} L(\tilde{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\xi}}_{\ell-1})}, \\ \tilde{\sigma}_{0,\ell}^2 = \dfrac{\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \tilde{u}_{n,k}^{(i)} \ge \mathcal{K} | \tilde{u}_{n,0}^{(i)})} L(\tilde{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\xi}}_{\ell-1}) \left(\tilde{u}_{n,0}^{(i)} - \tilde{\mu}_0\right)^2}{\sum_{i=1}^{J_1} \sqrt{\mathbb{P}(\max_{0\le k\le K} P_1 \tilde{u}_{n,k}^{(i)} \ge \mathcal{K} | \tilde{u}_{n,0}^{(i)})} L(\tilde{u}_{n,0}^{(i)}; \boldsymbol{v}, \hat{\boldsymbol{\xi}}_{\ell-1})}. \end{cases} \tag{39}$$

Note that the CE method with a Gaussian assumption for the initial condition may lead to an infinite variance of the estimator. Therefore, we observe certain non-robustness of the estimator variance produced by the CE method. To resolve this issue, we propose to fix $\tilde{\sigma}_0 = \sigma_0$ and shift only $\tilde{\mu}_0$ by (39). $\tilde{\rho}_0^{CE}$ corresponds to the optimal IS density obtained via the multilevel CE method with the shifted mean and fixed variance.

Appendix C Numerical PDE solver for the problem with a discontinuity at the boundary

Here, we briefly describe the numerical method used to approximate the solution to the auxiliary terminal-boundary value problem of the type

$$\frac{\partial \gamma}{\partial t} = -a(x,t)\frac{\partial \gamma}{\partial x} - \frac{1}{2} b(x,t) b(x,t)^T \frac{\partial^2 \gamma}{\partial x}, \quad (x,t) \in (-\infty, \mathcal{K}) \times (0,T) \tag{40a}$$

$$\begin{cases} \gamma(x,T) = 0, & x < \mathcal{K}, \\ \gamma(\mathcal{K},t) = 1, & t \le T, \\ \lim_{x\to-\infty} \gamma(x,t) = 0, & t \le T. \end{cases} \tag{40b}$$

Compare (13) and (17).

Since the accuracy of the approximation to (40) does not directly affect the accuracy of the estimate of the rare event probability α_n, only the efficiency of the proposed IS algorithm, and since even a qualitatively correct approximation can give us a useful IS we may use a simple method here. In our numerical tests, we used the Crank-Nickolson scheme with the uniform discretization of a truncated domain $(x_{\min}, \mathcal{K}) \times (0,T)$ with an artificial boundary condition $\frac{\partial^2 \gamma(x,t)}{\partial x^2} = 0$ at $x = x_{\min}$ approximated from known neighboring values by a linear extrapolation. However, the jump discontinuity in the corner $(\mathcal{K}, T)$ requires special care with this very basic approach. We outline our treatment here.

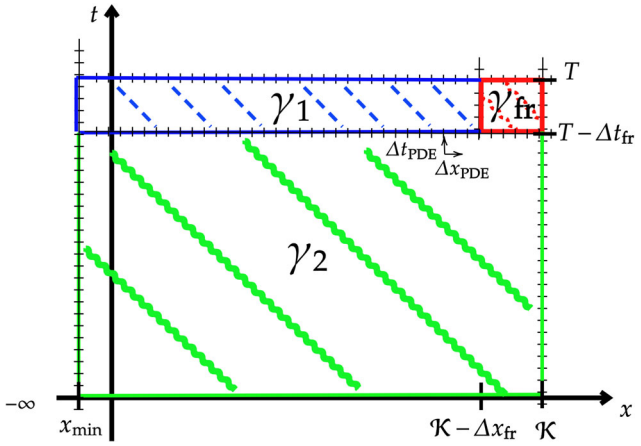


Fig. 13 Division of the problem domain for the PDE solver.

The constant coefficient version, $a(x,t) \equiv \mathbf{a}$, $b(x,t) \equiv \mathbf{b}$ of (40) has the closed-form solution

$$\gamma_{\mathbf{const}}(x,t) = \frac{1}{2}\operatorname{erfc}\left(\frac{\mathcal{K} - x - \mathbf{a}(T-t)}{2\sqrt{\frac{\mathbf{b}^2}{2}(T-t)}}\right) + \frac{1}{2} e^{\frac{\mathbf{a}(\mathcal{K}-x)}{\mathbf{b}^2/2}} \operatorname{erfc}\left(\frac{\mathcal{K} - x + \mathbf{a}(T-t)}{2\sqrt{\frac{\mathbf{b}^2}{2}(T-t)}}\right), \tag{41}$$

where $\operatorname{erfc}(z) := \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt$ is a complementary error function; see e.g., Harleman and Rumer (1963). Let $\gamma_{\mathbf{fr}}(x,t)$ be the solution to the constant coefficient problem with the frozen coefficients $\mathbf{a} = a(\mathcal{K},T)$ and $\mathbf{b} = b(\mathcal{K},T)$, i.e.,

$$\gamma_{\mathbf{fr}}(x,t) = \frac{1}{2}\operatorname{erfc}\left(\frac{\mathcal{K} - x - a(\mathcal{K},T)(T-t)}{|b(\mathcal{K},T)|\sqrt{2(T-t)}}\right) + \frac{1}{2} e^{\frac{a(\mathcal{K},T)(\mathcal{K}-x)}{b^2(\mathcal{K},T)/2}} \operatorname{erfc}\left(\frac{\mathcal{K} - x + a(\mathcal{K},T)(T-t)}{|b(\mathcal{K},T)|\sqrt{2(T-t)}}\right). \tag{42}$$

We divide the computational domain into three regions, see Figure 13, and for appropriately chosen $(\Delta x_{\mathbf{PDE}}, \Delta t_{\mathbf{PDE}})$ and $(\Delta x_{\mathbf{fr}}, \Delta t_{\mathbf{fr}})$ approximate γ as follows:

I. In the rectangle $[\mathcal{K} - \Delta x_{\mathbf{fr}}, \mathcal{K}] \times [T - \Delta t_{\mathbf{fr}}, T]$, we approximate $\gamma(x,t)$ by $\gamma_{\mathbf{fr}}(x,t)$. Observe that $\gamma_{\mathbf{fr}}$ decays very fast toward 0 as x decreases.

II. In $[x_{\min}, \mathcal{K} - \Delta x_{\mathbf{fr}}) \times [T - \Delta t_{\mathbf{fr}}, T]$, we use the finite difference approximation of (40a) with the terminal-boundary conditions

$$\begin{cases} \gamma(x,T) = 0, & x \in (x_{\min}, \mathcal{K} - \Delta x_{\mathbf{fr}}), \\ \gamma(\mathcal{K} - \Delta x_{\mathbf{fr}}, t) = \gamma_{\mathbf{fr}}(\mathcal{K} - \Delta x_{\mathbf{fr}}, t), & t \in [T - \Delta t_{\mathbf{fr}}, T], \\ \frac{\partial^2 \gamma(x_{\min},t)}{\partial x^2} = 0, & t \in [T - \Delta t_{\mathbf{fr}}, T]. \end{cases}$$

Denote this approximation γ_1.

Table 5 The choice for $(\Delta x_{\mathbf{fr}}, \Delta t_{\mathbf{fr}})$ used in the numerical PDE solver for the problems considered in Section 4.

$\mathcal{K}$	**Double Well**		$\mathcal{K}$	**Langevin**	
	$\Delta x_{\mathbf{fr}}$	$\Delta t_{\mathbf{fr}}$		$\Delta x_{\mathbf{fr}}$	$\Delta t_{\mathbf{fr}}$
0	0.185	0.0225	**2**	0.170	0.00625
0.5	0.195	0.0250	**2.5**	0.121	0.0025
1	0.185	0.0225	**3**	0.120	0.0025
1.2	0.185	0.0225	**3.5**	0.1365	0.00375
1.5	0.185	0.0225			
2	0.185	0.0225			
2.5	0.200	0.0275			
3	0.195	0.0250			

III. In $[x_{\min}, \mathcal{K}] \times [0, T - \Delta t_{\mathbf{fr}}]$, we use the same finite difference approximation of (40a) with the terminal-boundary conditions

$$\gamma(x, T - \Delta t_{\mathbf{fr}}) = \begin{cases} \gamma_1(x, T - \Delta t_{\mathbf{fr}}), & x \in (x_{\min}, \mathcal{K} - \Delta x_{\mathbf{fr}}), \\ \gamma_{\mathbf{fr}}(x, T - \Delta t_{\mathbf{fr}}), & x \in [\mathcal{K} - \Delta x_{\mathbf{fr}}, \mathcal{K}), \end{cases}$$
$$\gamma(\mathcal{K}, t) = 1, \qquad t \in [0, T - \Delta t_{\mathbf{fr}}],$$
$$\frac{\partial^2 \gamma(x_{\min}, t)}{\partial x^2} = 0, \quad t \in [0, T - \Delta t_{\mathbf{fr}}].$$

Denote this approximation γ_2.

We patch together the numerical approximation from the three subdomains and denote the PDE solution obtained via this strategy by $\gamma_{\mathbf{fr}}^{\mathbf{PDE}}(x, t)$ and the corresponding control by $\xi_{\mathbf{fr}}^{\mathbf{PDE}}(x, t)$.

The appropriate choice of the pair $(\Delta x_{\mathbf{fr}}, \Delta t_{\mathbf{fr}})$ depends on the discretization parameters $(\Delta x_{\mathbf{PDE}}, \Delta t_{\mathbf{PDE}})$ of the finite-difference scheme for γ_1 and γ_2. One can choose $(\Delta x_{\mathbf{fr}}, \Delta t_{\mathbf{fr}})$ such that the gradient of the closed-form solution at $(\mathcal{K} - \Delta x_{\mathbf{fr}}, T - \Delta t_{\mathbf{fr}})$ should closely match with $(\Delta x_{\mathbf{PDE}}, \Delta t_{\mathbf{PDE}})$. In the numerical tests in Section 4, we use $\Delta x_{\mathbf{PDE}} = 0.005$, $\Delta t_{\mathbf{PDE}} = \frac{\Delta x_{\mathbf{PDE}}}{2}$ with $x_{\min} = -5$ in the Double Well problem, and $\Delta x_{\mathbf{PDE}} = 0.006$, $\Delta t_{\mathbf{PDE}} = \frac{\Delta x_{\mathbf{PDE}}}{5}$ with $x_{\min} = -3$ in the Langevin problem. Correspondingly, the choice $(\Delta x_{\mathbf{fr}}, \Delta t_{\mathbf{fr}})$ used to define the red region in Figure 13 is provided in Table 5 for both problems given the different thresholds $\mathcal{K}$.

Appendix D Adaptive time-stepping scheme for IS simulation

The optimal control given by (22) blows up as $t \to T$. For this reason, it is not appropriate to use uniform time discretizations in the Euler-Maruyama approximation of the controlled SDE. Our numerical experiments used a simple adaptive time-stepping scheme based on the solution $\gamma_{\mathbf{fr}}$ to the frozen coefficient problem. The control corresponding to $\gamma_{\mathbf{fr}}$ is

$$\begin{aligned} \xi_{\mathbf{fr}}(x, t) =& b(\mathcal{K}, T) \frac{\partial(\log(\gamma_{\mathbf{fr}}(x, t)))}{\partial x} \\ =& \frac{b(\mathcal{K}, T)}{\gamma_{\mathbf{fr}}(x, t)} \frac{\partial(\gamma_{\mathbf{fr}}(x, t))}{\partial x} \\ =& \frac{b(\mathcal{K}, T)}{\gamma_{\mathbf{fr}}(x, t)} \\ & \Bigg[\frac{1}{|b(\mathcal{K}, T)| \sqrt{2\pi(T - t)}} \\ & \left(e^{-(R^-(x))^2} + e^{-(R^+(x))^2 + \frac{2a(\mathcal{K}, T)(\mathcal{K} - x)}{b(\mathcal{K}, T)^2}} \right) \\ & - \frac{a(\mathcal{K}, T)}{b(\mathcal{K}, T)^2} e^{\frac{2a(\mathcal{K}, T)(\mathcal{K} - x)}{b(\mathcal{K}, T)^2}} \mathbf{erfc}(R^+(x)) \Bigg], \end{aligned} \tag{43}$$

where $R^{\pm}(x) := \frac{\mathcal{K} - x \pm a(\mathcal{K}, T)(T - t)}{|b(\mathcal{K}, T)| \sqrt{2(T - t)}}$.

Asymptotically, as $t \to T$, for fixed x

$$\xi_{\mathbf{fr}}(x, t) \cong \xi_{\mathbf{fr}}^{\infty}(x, t) := \frac{1}{b(\mathcal{K}, T)} \left(\frac{\mathcal{K} - x}{T - t} \right) - a(\mathcal{K}, T).$$

We extend this asymptotic formula for the control to non-constant coefficients as

$$\xi^{\infty}(x, t) = \frac{1}{b(x, t)} \left(\frac{K - x}{T - t} \right) - a(x, t).$$

and use the control $\xi^{\infty}(x, t)$ for $t \in (T - \delta t, T]$, while we use $\xi_{\mathbf{fr}}^{\mathbf{PDE}}(x, t)$, based on the numerical approximation $\gamma_{\mathbf{fr}}^{\mathbf{PDE}}(x, t)$, for $t \in [0, T - \delta t]$.

In Algorithm 2, below, we provide a sketch of the adaptive time-stepping scheme with the Brownian bridge technique used in Monte Carlo simulations of all methods (Crude MC, IS wrt ρ_0, IS wrt $W(t)$, IS wrt both ρ_0 and $W(t)$ and IS with CE) considered in this work.

* $\xi(x, t) = 0$ in the crude MC, IS wrt ρ_0 and CE simulations.

In both numerical tests in Section 4, we set $\delta t = 10\Delta t$ and $\varepsilon = 10^{-6}$ as this choice was sufficient to ensure a high level of variance reduction for each method.

Algorithm 2: Adaptive time-stepping scheme with Brownian bridge technique

Input: T, K, $\mathcal{K}$, model parameters, δt and ε.

Define the uniform Euler-Maruyama timestep: $\Delta t = \frac{T}{K}$;

Set the optimal control such that

$$\xi(x,t) = \begin{cases} \xi_{\mathbf{fr}}^{\mathbf{PDE}}(x,t), & \text{if } t \in [0, T-\delta t], \\ \xi^{\infty}(x,t), & \text{if } t \in (T-\delta t, T]; \end{cases} \tag{44}$$

$k = 0$, $t_{n,k} = 0$, $\mathrm{h} = \Delta t$;

Set the initial condition $\bar{u}_{n,0}$ according to each method;

while $(t_{n,k} \leq T)$, $(P_1 \bar{u}_{n,k} < \mathcal{K})$ *and* $(\mathrm{h} > \varepsilon)$ **do**

  Adaptive rule: $\mathrm{h} = \min(\Delta t, \frac{T - t_{n,k}}{2})$;

  $\Delta W_{n,k}^{\mathbb{P}} \sim N(0, \mathrm{h}\mathbb{I}_{d_W})$;

  Simulate the trajectory $\bar{u}_{n,k+1} = \bar{u}_{n,k} + a(\bar{u}_{n,k})\mathrm{h} + b(\bar{u}_{n,k})\xi(\bar{u}_{n,k}, t_{n,k})\mathrm{h} + b(\bar{u}_{n,k})\Delta W_{n,k}^{\mathbb{P}}$ *;

  Compute the corresponding likelihood according to each method;

  Hitting test via Brownian bridge technique:

  compute the exiting probability $q = \exp\left(-\frac{2\max(\mathcal{K} - P_1\bar{u}_{n,k}, 0)\max(\mathcal{K} - P_1\bar{u}_{n,k+1}, 0)}{(P_1 b(\bar{u}_{n,k}) b(\bar{u}_{n,k})^T P_1^T)\mathrm{h}}\right)$;

  sample a uniform random variable $r \sim U(0,1)$;

  if $r < q$ **then**

    set $P_1\bar{u}_{n,k+1} = \mathcal{K}$ and the stopping time $\tau = t_{n,k} + 0.5\mathrm{h}$;

  else

    $t_{n,k} = t_{n,k} + \mathrm{h}$;

  $k = k + 1$;

Acknowledgements The authors would like to thank the anonymous reviewer and the external examiner of G. Shaimerdenova's Ph.D. dissertation for their constructive and insightful feedback, which greatly helped to improve the clarity and presentation of this manuscript.

Author Contributions All authors contributed to the conceptualisation of the manuscript. G.S. wrote the main manuscript text, obtained the numerical results, and created the figures. N.B, E.V and R.T provided continuous feedback during the development process. All authors reviewed the manuscript.

Funding Open access publishing provided by King Abdullah University of Science and Technology (KAUST). This work was supported by the KAUST Office of Sponsored Research (OSR) under Award No. URF/1/2584-01-01 and the Alexander von Humboldt Foundation. E. von Schwerin, G. Shaimerdenova and R. Tempone are members of the KAUST SRI Center for Uncertainty Quantification in Computational Science and Engineering.

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B.: The ensemble Kalman filter in reservoir engineering-a review. SPE Journal **14**(03), 393–412 (2009)

Asmussen, S., Glynn, P.W.: Stochastic simulation: algorithms and analysis. SMAP, vol. 57. Springer, New York (2007)

Siu-Kui, A., Beck, J.L.: Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic engineering mechanics **16**(4), 263–277 (2001)

Siu-Kui, A., Beck, J.L.: Subset simulation and its application to seismic risk based on dynamic analysis. J. Eng. Mech. **129**(8), 901–917 (2003)

Bally, V., Talay, D.: The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. Probability theory and related fields **104**, 43–60 (1996)

Bayer, C., Häppölä, J., Tempone, R.: Implied stopping rules for american basket options from markovian projection. Quantitative Finance **19**(3), 371–390 (2019)

Beaurepaire, P., Valdebenito, M.A., Schuëller, G.I., Jensen, H.A.: Reliability-based optimization of maintenance scheduling of mechanical components under fatigue. Comput. Methods Appl. Mech. Eng. **221**, 24–40 (2012)

James, L.: Beck and Konstantin M. Zuev. Rare Event Simulation. Springer International Publishing, Cham (2016)

Ben Hammouda, C., Ben Rached, N., Tempone, R., Wiechert, S.: Learning-based importance sampling via stochastic optimal control for stochastic reaction networks. Statistics and Computing, 33(3):58, (2023)

Ben Rached, N., Haji-Ali, A., Mohan, S., Tempone, R.: Double loop Monte Carlo estimator with importance sampling for McKean-Vlasov stochastic differential equation. arXiv e-prints, pages arXiv:2207.06926, (2022)

Rached, N.B., Kammoun, A., Alouini, M.S., Tempone, R.: Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels. IEEE Journal of Selected Topics in Signal Processing **10**(2), 376–388 (2016)

Botev, Z.I., Kroese, D.P.: Efficient monte carlo simulation via the generalized splitting method. Stat. Comput. **22**, 1–16 (2012)

Bréhier, C.E., Lelièvre, T.: On a new class of score functions to estimate tail probabilities of some stochastic processes with adaptive multilevel splitting. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(3):033126, (2019)

Bucklew, J.A.: Introduction to rare event simulation, volume 5 of SSS. Springer New York, 1 edition, (2004)

Cérou, F., Del Moral, P., Furon, T., Guyader, A.: Sequential monte carlo for rare event estimation. Stat. Comput. **22**(3), 795–808 (2012)

Cérou, F., Guyader, A., Rousset, M.: Adaptive multilevel splitting: Historical perspective and recent results. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(4):043108, (2019)

Charney, J.G., DeVore, J.G.: Multiple flow equilibria in the atmosphere and blocking. Journal of Atmospheric Sciences, 36(7):1205–1216, (1979)

Crisan, D., Doucet, A.: A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Signal Process. **50**(3), 736–746 (2002)

Crommelin, D.T., Opsteegh, JD., Verhulst, F.: A mechanism for atmospheric regime behavior. Journal of the atmospheric sciences, 61(12):1406–1419, (2004)

de Angelis, M., Patelli, E., Beer, M.: Advanced line sampling for efficient robust reliability analysis. Struct. Saf. **52**, 170–182 (2015)

De Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y.: A tutorial on the cross-entropy method. Ann. Oper. Res. **134**, 19–67 (2005)

Del Moral, P., Garnier, J.: Genealogical particle analysis of rare events. Ann. Appl. Probab. **15**(4), 2496–2534 (2005)

Der Kiureghian, A.: First-and second-order reliability methods. Engineering design reliability handbook, 14, (2005)

Dos Reis, G., Smith, G., Tankov, P.: Importance sampling for mckean-vlasov sdes. Appl. Math. Comput. **453**, 128078 (2023)

Dzougoutov, A., Moon, K.S., Schwerin, E.v., Szepessy, A., Tempone, R.: Adaptive Monte Carlo algorithms for stopped diffusion. In Multiscale methods in science and engineering, pages 59–88. Springer, (2005)

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans **99**(C5), 10143–10162 (1994)

Fleming, W.H., Soner, H.M.: Controlled Markov processes and viscosity solutions, volume 25. Springer Science & Business Media, (2006)

Gao, J., Tembine, H.: Distributed mean-field-type filters for big data assimilation. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1446–1453. IEEE, (2016)

Gobet, E.: Weak approximation of killed diffusion using euler schemes. Stochastic processes and their applications **87**(2), 167–197 (2000)

Gobet, E.: Advanced Monte Carlo methods for barrier and related exotic options. In Handbook of Numerical Analysis, volume 15, pages 497–528. Elsevier, (2009)

Grafke, T., Vanden-Eijnden, E.: Numerical computation of rare events via large deviation theory. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(6):063118, (2019)

Gyöngy, I.: Mimicking the one-dimensional marginal distributions of processes having an itô differential. Probab. Theory Relat. Fields **71**(4), 501–516 (1986)

Hammouda, C.B., Rached, N.B., Tempone, R., Wiechert, S.: Learning-based importance sampling via stochastic optimal control for stochastic reaction networks. Statistics and Computing, 33(58), (2023)

Harleman, D.R.F., Rumer, R.R.: Longitudinal and lateral dispersion in an isotropic porous medium. J. Fluid Mech. **16**(3), 385–394 (1963)

Hartmann, C., Kebiri, O., Neureither, L., Richter, L.: Variational approach to rare event simulation using least-squares regression. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(6):063107, (2019)

Hartmann, C., Richter, L., Schütte, C., Zhang, W.: Variational characterization of free energy: theory and algorithms. Entropy **19**(11), 626 (2017)

Hartmann, C., Schütte, C.: Efficient rare event simulation by optimal nonequilibrium forcing. J. Stat. Mech: Theory Exp. **2012**(11), P11004 (2012)

Hartmann, C., Schütte, C., Weber, M., Zhang, W.: Importance sampling in path space for diffusion processes with slow-fast variables. Probab. Theory Relat. Fields **170**, 177–228 (2018)

Hartmann, C., Schütte, C., Zhang, W.: Model reduction algorithms for optimal control and importance sampling of diffusions. Nonlinearity **29**(8), 2298 (2016)

Hasofer, A.M.: An exact and invarient first order reliability format. J. Eng. Mech. Div., Proc. ASCE, 100(1):111–121, (1974)

Heidelberger, P.: Fast simulation of rare events in queueing and reliability models. ACM Transactions on Modeling and Computer Simulation (TOMACS) **5**(1), 43–85 (1995)

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

Hoel, H., Shaimerdenova, G., Tempone, R.: Multilevel ensemble kalman filtering based on a sample average of independent enkf estimators. Foundations of Data Science **2**(4), 351 (2020)

Hoel, H., Shaimerdenova, G., Tempone, R.: Multi-index ensemble kalman filtering. J. Comput. Phys. **470**, 111561 (2022)

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B.: Atmospheric data assimilation with an ensemble kalman filter: results with real observations. Mon. Weather Rev. **133**(3), 604–620 (2005)

Jones, T.A., Stensrud, D., Wicker, L., Minnis, P., Palikonda, R.: Simultaneous radar and satellite data storm-scale assimilation using an ensemble kalman filter approach for 24 may 2011. Mon. Weather Rev. **143**(1), 165–194 (2015)

Kalnay, E.: Atmospheric modeling, data assimilation and predictability. Cambridge university press, (2003)

Karatzas, I., Shreve, S.: Brownian motion and stochastic calculus, volume 113 of GTM. Springer Science & Business Media, (2012)

Karmiani, D., Kazi, R., Nambisan, A., Shah, A., Kamble, V.: Comparison of predictive algorithms: backpropagation, svm, lstm and kalman filter for stock market. In 2019 amity international conference on artificial intelligence (AICAI), pages 228–234. IEEE, (2019)

Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations, vol. 23. Springer, Berlin, Heidelberg (1992)

Kroese, D.P., Rubinstein, R.Y., Glynn, P.W.: The cross-entropy method for estimation. In Handbook of statistics, volume 31, pages 19–34. Elsevier, (2013)

Kroese, D.P., Taimre, T., Botev, Z.I.: Handbook of Monte Carlo methods. John Wiley & Sons, (2013)

Law, K., Stuart, A., Zygalakis, K.: Data assimilation. Cham, Switzerland: Springer, 214, (2015)

Law, K.JH., Tembine, H., Tempone, R.: Deterministic mean-field ensemble kalman filtering. SIAM Journal on Scientific Computing, 38(3):A1251–A1279, (2016)

Le Gland, F., Monbet, V., Tran, V.D.: Large sample asymptotics for the ensemble Kalman filter. PhD thesis, INRIA, (2009)

Maejima, Y., Miyoshi, T.: Impact of the window length of four-dimensional local ensemble transform kalman filter: a case of convective rain event. SOLA **16**, 37–42 (2020)

Majda, A.J., Harlim, J.: Filtering complex turbulent systems. Cambridge University Press, (2012)

Melchers, R.E., Beck, A.T.: Structural reliability analysis and prediction. John wiley & sons, (2018)

Menozzi, S., Pesce, A., Zhang, X.: Density and gradient estimates for non degenerate brownian sdes with unbounded measurable drift. J. Differential Equations **272**, 330–369 (2021)

Nüsken, N., Richter, L.: Solving high-dimensional hamilton-jacobi-bellman pdes using neural networks: perspectives from the theory of controlled diffusions and measures on path space. Partial differential equations and applications **2**(4), 48 (2021)

Papaioannou, I., Papadimitriou, C., Straub, D.: Sequential importance sampling for structural reliability analysis. Struct. Saf. **62**, 66–75 (2016)

Rached, N.B., Haji-Ali, A.L., Pillai, S.M.S., Tempone, R.: Multilevel importance sampling for McKean-Vlasov stochastic differential equation. arXiv preprint arXiv:2208.03225, (2022)

Ragone, F., Wouters, J., Bouchet, F.: Computation of extreme heat waves in climate models using a large deviation algorithm. Proc. Natl. Acad. Sci. **115**(1), 24–29 (2018)

Reich, S., Cotter, C.: Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press, (2015)

Rubinstein, R.Y., Kroese, D.P.: Simulation and the Monte Carlo method. John Wiley & Sons, (2016)

Sapsis, T.P.: Statistics of extreme events in fluid flows and waves. Annual Review of Fluid Mechanics, 53(1):85–111, (2021)

Straub, D., Papaioannou, I.: Bayesian updating with structural reliability methods. J. Eng. Mech. **141**(3), 04014134 (2015)

Tong, S., Stadler, G.: Large deviation theory-based adaptive importance sampling for rare events in high dimensions. SIAM/ASA Journal on Uncertainty Quantification **11**(3), 788–813 (2023)

Tong, S., Vanden-Eijnden, E., Stadler, G.: Extreme event probability estimation using pde-constrained optimization and large deviation theory, with application to tsunamis. Communications in Applied Mathematics and Computational Science **16**(2), 181–225 (2021)

Valdebenito, M.A., Pradlwarter, H.J., Schuëller, G.I.: The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities. Structural Safety, 32(2):101–111, (2010)

Vanden-Eijnden, E., Weare, J.: Data assimilation in the low noise regime with application to the kuroshio. Mon. Weather Rev. **141**(6), 1822–1841 (2013)

Wagner, F., Papaioannou, I., Ullmann, E.: The ensemble kalman filter for rare event estimation. SIAM/ASA Journal on Uncertainty Quantification **10**(1), 317–349 (2022)

Webber, R.J., Plotkin, D.A., O'Neill, M.E., Abbot, D.S., Weare, J.: Practical rare event sampling for extreme mesoscale weather. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(5), (2019)

Zhang, K., Manning, T., Suqin, W., Rohm, W., Silcock, D., Choy, S.: Capturing the signature of severe weather events in australia using gps measurements. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **8**(4), 1839–1847 (2015)

Zio, E., Pedroni, N.: Estimation of the functional failure probability of a thermal-hydraulic passive system by subset simulation. Nucl. Eng. Des. **239**(3), 580–599 (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.