Boundary-Layer Meteorology (2026) 192:8
https://doi.org/10.1007/s10546-025-00954-5

RESEARCH ARTICLE

An Importance Sampling Method for Lagrangian Stochastic Modeling of Atmospheric Turbulence

Jihoon Shin[1] · **Jong-Jin Baik**[2]

Received: 1 September 2025 / Accepted: 28 November 2025

Abstract
Accurately representing atmospheric turbulence requires capturing its highly non-Gaussian probability density functions (PDFs), which are poorly represented by traditional turbulence closure models. The recently proposed Lagrangian Stochastic Subgrid Turbulence Model (LSSTM) directly simulates turbulence PDFs using particle-based stochastic differential equations (SDEs), but its Monte Carlo formulation is subject to high sampling variance and computational cost. In this study, we introduce an importance sampling (IS) framework for LSSTM that reduces variance by modifying the drift of the governing SDEs to tilt their invariant distribution, combined with adaptive resampling to prevent weight degeneracy. A sampling weight proportional to vertical velocity fluctuations is proposed to improve the estimation of turbulent fluxes while preserving stability in non-turbulent regions. Validation with one-dimensional diffusion processes demonstrates variance reductions of 30–80% across statistical moments and tail probabilities. Applied to dry convective boundary layer simulations, IS improves the accuracy of turbulence statistics, reduces systematic biases in mean profiles, and achieves up to 50% variance reduction in vertical fluxes and up to 60% in rare-event probabilities. These results demonstrate that IS substantially enhances the efficiency and robustness of LSSTM, providing a practical pathway toward affordable PDF-based turbulence modeling.

Keywords Importance sampling · Atmospheric turbulence modeling · Lagrangian stochastic modeling

✉ Jong-Jin Baik
jjbaik@snu.ac.kr

[1] Department of Environmental Atmospheric Sciences, Pukyong National University, Busan, South Korea

[2] School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea

Published online: 11 December 2025

1 Introduction

Accurately simulating atmospheric turbulence requires capturing not only mean flow properties but also the complete spectrum of turbulent fluctuations. Due to thermal stratification and interaction with other physical processes, atmospheric turbulence exhibits complex and often highly non-Gaussian probability density functions (PDFs) (Bogenschutz et al. 2010; Chinita et al. 2018). Traditional turbulence closure models typically provide statistical representations limited to low-order moments. This limitation can lead to significant inaccuracies in representing turbulence PDFs, introducing errors in vertical flux estimates and ultimately degrading the accuracy of atmospheric state predictions (Guo et al. 2015; Fitch 2019). A full PDF-based modeling approach has the potential to resolve subgrid-scale variability more comprehensively within parameterizations.

Efforts to represent turbulence using full PDF-based approaches have a rich history in the broader turbulence modeling community. Early developments (Van Slooten and Jayesh 1998; Minier and Pozorski 1999; Jenny et al. 2001) formulated transport equations for joint PDFs of velocity and scalar quantities using stochastic Lagrangian methods, demonstrating their ability to capture non-Gaussian turbulence statistics in inhomogeneous and reacting flows. Pope (2000) provided a comprehensive theoretical foundation for these so-called "PDF methods", which directly solve transport equations for the joint probability density of turbulent variables, typically represented using ensembles of stochastic particles governed by generalized Langevin equations (Pope 1983). In this approach, notional "particle" tracers are evolved according to stochastic differential equations (SDEs) for velocity and passive scalars, and turbulence statistics are calculated by averaging the particle properties. The methods are distinguished from the assumed PDF method used in numerical atmospheric models (Golaz et al. 2002; Bogenschutz and Krueger 2013), which reconstructs PDFs from prognosed turbulent moments, in that the PDF methods fully resolve turbulence PDFs.

Shin and Baik (2024) introduced a Lagrangian stochastic subgrid turbulence model (LSSTM) that directly simulates the joint PDF of turbulence in a stratified atmospheric boundary layer in a single column framework. The work extends the model of Das and Durbin (2005), originally developed for homogeneous stratified turbulence, to represent inhomogeneous turbulence and wall effect, enabling the simulation of boundary-layer turbulence across a wide range of stability. While Lagrangian stochastic modeling has long been applied to atmospheric dispersion problems (Thomson 1987; Boughton et al. 1987; Stockie 2011), such as tracking air pollutants, LSSTM distinguishes itself by serving as a stand-alone atmospheric turbulence model (Shin and Baik 2024, 2025). While the method has potential for enabling full PDF turbulence modeling, a high computational cost is still a limitation. Calculating converged turbulence statistics requires evolving a large ensemble of particles, placing the computational expense of LSSTM between that of traditional closure models and that of large-eddy simulations.

Since LSSTM is based on Monte Carlo integration, the statistical variances in estimated turbulence statistics must be minimized to reduce computational cost. While many other options are available, a common variance reduction method for diffusion processes is importance sampling. In atmospheric modeling, importance sampling has been widely adopted in dispersion simulations and source attribution to accurately estimate rare events. For instance, the hybrid single-particle Lagrangian integrated trajectory (HYSPLIT) model implements a puff-splitting strategy, which can be viewed as a form of importance sampling

by duplicating particles in regions of interest (Draxler and Hess 1998). Rajaona et al. (2015) introduced an adaptive multiple importance sampling framework for source attribution, which iteratively refines the biasing distribution using a recycling strategy that preserves sample diversity and enhances convergence. Esler (2015) developed an adaptive trajectory algorithm based on the "go-with-the-winners" branching process originally proposed by Grassberger (2002), wherein high-likelihood trajectories are preferentially cloned, while low-likelihood ones are discarded.

The goal of this study is to develop a flexible importance sampling technique particularly appropriate for LSSTM. Many previously developed importance sampling methods for dispersion models are based on a resampling (or splitting) strategy (Draxler and Hess 1998; Scire et al. 2000; Jones et al. 2007; Pisso et al. 2019). However, frequent resampling can lead to sample impoverishment by generating multiple identical particles, reducing the diversity of the particle ensemble. Since the full turbulent fluctuation needs to be resolved in LSSTM, frequent resampling may be a problem. Furthermore, since LSSTM involves the computation of a wide array of statistical measures, a more generalized sampling strategy is required to improve the quality of many statistics, rather than one that is narrowly optimized for a single statistic (e.g., pollutant concentration).

The importance sampling technique introduced in this study combines a change of measure method, which is achieved by modifying the underlying SDEs, with a resampling strategy. This hybrid approach effectively controls both weight degeneracy and sample impoverishment, which are common problems in importance sampling for diffusion processes. Additionally, a sampling weight function suitable for LSSTM is proposed to mitigate statistical uncertainties in vertical fluxes and other statistics. Reducing uncertainties in the estimation of vertical fluxes is critical for enhancing the accuracy and robustness of atmospheric predictions, as these fluxes govern the temporal evolution of mean atmospheric fields. The performance of the proposed method is evaluated using simplified one-dimensional diffusion processes and applied to LSSTM simulations.

2 Theory and Method

2.1 Importance Sampling Method

The proposed importance sampling (IS) method is designed to reduce variance in Monte Carlo estimates of statistics for discretized continuous diffusion processes, especially for the application to atmospheric turbulence modeling. By modifying the drift of the underlying SDE to tilt its invariant measure from the baseline density $G(x)$ to a weighted density $f(x)G(x)$ (where $f(x)$ is the sampling weight), we can generate more samples in the regions that contribute most to target statistics. Subsequently, a resampling technique ensures each particle's importance weight is inversely proportional to $f(x)$, preserving unbiased estimates.

2.1.1 Modified SDE for Weighted Distribution

The standard approach for IS of diffusion processes is to use the change of measure by modifying the drift of the original SDE, based on the theory of h-transforms (Rogers and Williams 2000). Consider the baseline real-valued SDE:

$$dX_t = \mu(x)\,dt + \sigma(x)\,dW_t, \tag{1}$$

where X_t is the random state variable evolving over time, $\mu(x)$ and $\sigma(x)$ are the drift and diffusion coefficients, respectively, and W_t is the standard Brownian motion. We assume that this SDE is constructed so that its invariant density is $G(x)$. This SDE is interpreted in the Itô sense, where the stochastic integral $\int_0^t \sigma(X_s)\,dW_s$ is defined by evaluating the integrand at the beginning of each infinitesimal time interval, so W_t remains a martingale (Øksendal 2003). Under the Itô interpretation, the Fokker–Planck equation for this process is:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\Big(\mu(x)\,p(x,t)\Big) + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(\sigma^2(x)\,p(x,t)\Big), \tag{2}$$

where $p(x, t)$ is the probability density function of the process X_t (Øksendal 2003). At stationarity ($p(x) = G(x)$) and assuming detailed balance (zero probability current), we require that:

$$\mu(x)G(x) - \frac{1}{2}\,\frac{d}{dx}\Big(\sigma^2(x)G(x)\Big) = 0. \tag{3}$$

Then, the drift can be rewritten as:

$$\begin{aligned}\mu(x) &= \frac{1}{2}\,\frac{d}{dx}\Big(\sigma^2(x)G(x)\Big)\Big/G(x)\\ &= \frac{1}{2}\,\sigma^2(x)\,\frac{G'(x)}{G(x)} + \frac{1}{2}\,\sigma^{2\prime}(x),\end{aligned} \tag{4}$$

and the drift ensures that $G(x)$ is the invariant density of the baseline SDE.

Our objective is to sample from a proposal density:

$$\pi(x) \propto f(x)\,G(x), \tag{5}$$

where $f(x) > 0$ is an unnormalized sampling weight. Note that $f(x)$ itself does not need to be normalizable, but $\pi(x)$ and $G(x)$ must be properly normalizable to define valid probability densities. The proposal density can be written in Gibbs form:

$$\pi(x) \propto e^{-V(x)}, \tag{6}$$

where the potential is:

$$V(x) = -\ln\big(f(x)G(x)\big) = -\ln f(x) - \ln G(x). \tag{7}$$

For SDEs with state-dependent diffusion, the Langevin (or reversible) dynamics that yield a specified invariant density take the form:

$$dX_t = -\frac{1}{2}\sigma^2(x)V'(x)\,dt + \frac{1}{2}\,\sigma^{2\prime}(x)\,dt + \sigma(x)\,dW_t. \tag{8}$$

Taking the derivative of the potential gives:

$$V'(x) = -\frac{f'(x)}{f(x)} - \frac{G'(x)}{G(x)}. \tag{9}$$

Substituting this into the drift term yields the modified drift $\mu^*(x)$ where:

$$\begin{aligned} \mu^*(x) &= -\frac{1}{2}\sigma^2(x)V'(x) + \frac{1}{2}\sigma^{2\prime}(x) \\ &= \frac{1}{2}\sigma^2(x)\left(\frac{f'(x)}{f(x)} + \frac{G'(x)}{G(x)}\right) + \frac{1}{2}\sigma^{2\prime}(x). \end{aligned} \tag{10}$$

Substituting Equation 4, the modified SDE (Equation 8) then becomes:

$$\begin{aligned} dX_t &= \mu^*(x)\,dt + \sigma(x)\,dW_t \\ &= \left[\mu(x) + \frac{1}{2}\,\sigma^2(x)\frac{f'(x)}{f(x)}\right]dt + \sigma(x)\,dW_t, \end{aligned} \tag{11}$$

where the additional drift $\frac{1}{2}\sigma^2(x)\frac{f'(x)}{f(x)}$ tilts the invariant measure from $G(x)$ to $\pi(x)$.

2.1.2 Calculation of Importance Weights

When samples from the modified SDE exactly follow the invariant density, the expected importance weight is proportional to $1/f(x)$. However, for general cases where SDEs do not have invariant densities, the importance weight of each particle should be traced. The diffusion process can be simulated by integrating a discretized SDE for multiple samples, and the following provides a procedure to keep track of the importance weights. Let $t_0 < t_1 < \cdots < t_n$ be an arbitrary time grid (n is the number of time steps), and suppose the modified SDE (Equation 11) is integrated with any numerical scheme. At time t_k, the one-step proposal transition density $q_k\left(X_{t_{k+1}} \mid X_{t_k}\right)$ is defined as the conditional density predicted by the modified SDE, and the one-step target transition density $p_k\left(X_{t_{k+1}} \mid X_{t_k}\right)$ is defined as the conditional density predicted by the original SDE. Then, the importance weight at each step k is:

$$W_k \;=\; \frac{p_k\left(X_{t_{k+1}} \mid X_{t_k}\right)}{q_k\left(X_{t_{k+1}} \mid X_{t_k}\right)}. \tag{12}$$

For a full trajectory $\{X_{t_k}\}_{k=0}^{n}$, the overall importance weight is:

$$\widetilde{W} \;=\; \prod_{k=0}^{n-1} W_k \;=\; \prod_{k=0}^{n-1} \frac{p_k\left(X_{t_{k+1}} \mid X_{t_k}\right)}{q_k\left(X_{t_{k+1}} \mid X_{t_k}\right)}, \tag{13}$$

in the context of sequential IS. Equivalently, one may accumulate log-weights for numerical stability:

$$\ln \widetilde{W} \;=\; \sum_{k=0}^{n-1} \Big[\ln p_k\big(X_{t_{k+1}} \mid X_{t_k}\big) - \ln q_k\big(X_{t_{k+1}} \mid X_{t_k}\big) \Big]. \tag{14}$$

For the example of the Euler-Maruyama (EM) scheme, the transitions are approximated by Gaussian distributions:

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + \mu(X_{t_k})\Delta t + \sigma(X_{t_k})\sqrt{\Delta t}\xi_k, \quad \text{(original)} \\ X_{t_{k+1}} &= X_{t_k} + \mu^*(X_{t_k})\Delta t + \sigma(X_{t_k})\sqrt{\Delta t}\xi_k, \quad \text{(modified)} \end{aligned} \tag{15}$$

where $\xi_k \sim \mathcal{N}(0,1)$. Then, p_k and q_k are Gaussian densities with respective means and identical variances. The importance weight becomes:

$$W_k = \exp\left(-\frac{1}{2\sigma^2 \Delta t} \left[(X_{t_{k+1}} - X_{t_k} - \mu \Delta t)^2 - (X_{t_{k+1}} - X_{t_k} - \mu^* \Delta t)^2 \right] \right). \tag{16}$$

The discretized single-step transition density is approximated as Gaussian, since any finite-variance random process tends toward Gaussian behavior as $\Delta t \to 0$ by the central limit theorem. The importance weight is computed consistently with this transition density; however, discretization errors in importance weight may accumulate when the distribution of X_t becomes highly non-Gaussian. Therefore, for highly non-Gaussian PDFs, Δt should be chosen sufficiently small ($\Delta t \ll \sigma^2/\mu^2$) to ensure the validity of the central limit theorem and to keep discretization errors in the importance weights minimal.

2.1.3 Resampling to Control Weight Degeneracy

In long-time integrations of the modified SDE with sample number N, the importance weights $\{\widetilde{W}^{(i)}\}_{i=1}^N$ may become highly imbalanced and deviate from the expected importance weight $1/f(x)$, leading to weight degeneracy. The resampling is done to control the weight degeneracy issue and force samples to have importance weights proportional to $1/f(x)$. Here, the resampling weight is calculated as:

$$\widetilde{W}_{\text{resample}} \propto \frac{\widetilde{W}}{1/f(X)}, \tag{17}$$

and the effective sample size is calculated as:

$$\text{ESS} \;=\; \frac{\left(\sum_{i=1}^N \widetilde{W}^{(i)}_{\text{resample}}\right)^2}{\sum_{i=1}^N \left(\widetilde{W}^{(i)}_{\text{resample}}\right)^2}. \tag{18}$$

At each time step, the resampling weight and effective sample size are calculated, and resampling is triggered whenever $\text{ESS} < \text{ESS}_{\min}$, where $\text{ESS}_{\min}$ is a specified threshold.

At resampling, we replace the weighted ensemble by N particles drawn according to $\{\widetilde{W}^{(i)}_{\text{resample}}\}$, using any standard resampling scheme (e.g. residual, stratified, or systematic

resampling). Kuptametee and Aunsri (2022) provides a comprehensive review of the resampling techniques. After resampling, each particle $X^{(i)}$ is assigned the reset weight:

$$\widetilde{W}^{(i)} \propto \frac{1}{f\left(X^{(i)}\right)} \quad \text{(after resampling)}, \tag{19}$$

so that the important weight remains unbiased with respect to the expected value. This procedure restores diversity of samples while preserving consistency under the tilted invariant measure. Finally, any statistics can be calculated as an importance-weighted ensemble mean; the expectation from the tilted distribution is approximated as:

$$\mathbb{E}_G[h(X)] \approx \frac{\sum_{i=1}^{N} \widetilde{W}^{(i)}\, h\left(X_{t_n}^{(i)}\right)}{\sum_{i=1}^{N} \widetilde{W}^{(i)}}, \tag{20}$$

where $h(x)$ is any observable whose expectation is desired.

The optimal value of $\mathrm{ESS}_{\min}$ depends on the variance of importance weights, which increases with the mismatch between the target and proposal transition densities (Doucet et al. 2001). In highly complex or non-Gaussian PDFs, the ESS typically decreases rapidly, and a smaller $\mathrm{ESS}_{\min}$ may be preferable to avoid excessive resampling. Although IS itself does not bias the statistical estimates, frequent resampling may eliminate particles carrying valuable Lagrangian history and duplicate identical samples, leading to the loss of spatio-temporal correlation along particle trajectories (sample impoverishment). Therefore, it is advisable to maintain a high ESS while performing resampling as infrequently as possible to preserve particle diversity.

2.1.4 Proposed Sampling Weight for Atmospheric Turbulence Modeling

While any positive function can be used as a sampling weight in this IS technique, a proper sampling weight is necessary to reduce variances in vertical fluxes and other turbulence statistics effectively. A simple and straightforward proposal would be to sample more particles that exhibit high vertical velocity fluctuations, w'. By doing so, the number of samples that have a high impact on vertical fluxes increases. Therefore, it is natural to set the sampling weight proportional to $|w'|$, and more generally $|w'|^p$ where p is a positive exponent.

However, using the sampling weight of $|w'|^p$ creates a weight degeneracy issue near $w' = 0$, since accumulated importance weights diverge near this point. Therefore, we propose a sampling weight as:

$$f(x) \propto (\sigma_w(z))^p \left[1 + \frac{1}{p}\left(\frac{w'}{s_f\, \sigma_w(z)}\right)^2\right]^{\frac{p}{2}} \tag{21}$$

where s_f is a scale factor and $\sigma_w(z)$ is the standard deviation of w at the height z where the sample is located. For large $|w'|$, this distribution increases as $|w'|^p$, but it remains well-behaved near zero. Figure 1 illustrates $1/f(x)$, which is the expected importance weight, for various parameter values (when $\sigma_w = 1$). As the exponent p increases, more samples with

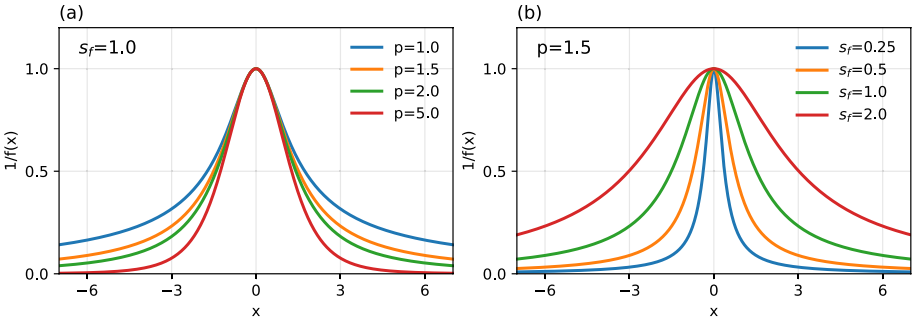

Fig. 1 Plots of the reciprocal of $f(x)$: **a** varying p with fixed $s_f = 1.0$; **b** varying s_f with fixed $p = 1.5$

high vertical velocity are drawn, and the corresponding importance weights for large $|w'|$ decrease. The parameter s_f controls the sharpness of the function near $w' = 0$. A smaller s_f results in a steep gradient in importance weights around zero, which can cause the accumulated weights to diverge rapidly. Therefore, choosing an appropriate value of s_f is crucial to prevent weight degeneracy.

The term $\sigma_w(z)^p$ governs the sampling probability with height, such that more samples are drawn in regions of high vertical velocity variance. However, we found that the performance of IS deteriorates near the boundary layer top, as $\sigma_w(z)^p$ becomes small in this region. To address this, $\sigma_w(z)^p$ is smoothed using a running average with a 500 m vertical window, thereby preventing $f(x)$ from becoming excessively small near the boundary layer top. Note that the proposed sampling weight is not an optimal function, since the optimal sampling weight depends on the target statistic. In the Lagrangian stochastic approach, many statistics are calculated simultaneously from a particle ensemble, so there is no universally optimal sampling weight.

2.2 Lagrangian Stochastic Subgrid Turbulence Model (LSSTM)

LSSTM proposed by Shin and Baik (2024) directly solves turbulence PDF transport by tracking the stochastic turbulent motion of each air parcel. The equations for the velocity component u_i and passive scalar ϕ are adopted from Das and Durbin (2005) with the inclusion of a pressure transport:

$$dx_i = u_i dt, \tag{22}$$

$$\begin{aligned} du_i = &-\frac{c_1}{2}\frac{\epsilon}{k}u_i' dt + c_{2\phi}u_k'\frac{\partial \overline{u_i}}{\partial x_k}dt + c_{3\phi}u_k'\frac{\partial \overline{u_k}}{\partial x_i}dt - (1-c_{5\phi})\, g_i \frac{\theta_v'}{\theta_{v0}}dt \\ &+ C_{pt}\left(\frac{u_k'u_k'}{2k} - 1\right)\frac{\partial k}{\partial x_i}dt + \sqrt{c_0\epsilon}dW_i, \end{aligned} \tag{23}$$

$$d\phi = -\left(c_{1\phi} - \frac{c_1}{2}\right)\frac{\epsilon}{k}\phi' dt + c_{4\phi}u_k'\frac{\partial \overline{\phi}}{\partial x_k}dt + \sqrt{c_\phi}dW_\phi, \tag{24}$$

where the overbar denotes Reynolds-averaging, x_i is the particle position, θ_v and θ_{v0} are the virtual potential temperature and basic-state virtual potential temperature, respectively, g_i is the component of gravitational acceleration, dW is the increment of the Wiener process, $k = (1/2)\overline{u_k'^2}$ is the turbulent kinetic energy (TKE), and ϵ is the Reynolds-averaged dissipation rate. c_1, $c_{1\phi}$, $c_{2\phi}$, $c_{3\phi}$, $c_{4\phi}$, $c_{5\phi}$, c_0, and c_ϕ are model constants, and their values are set as in Shin and Baik (2024).

The SDE for the dissipation rate is adopted from Van Slooten and Jayesh (1998) and is expressed in terms of turbulence frequency $\omega = \epsilon^*/k$, where ϵ^* is the local dissipation rate:

$$\begin{aligned} d\omega &= -C_3(\omega - \overline{\omega})\Omega dt - \Omega\omega S_\omega dt + \sqrt{2C_3C_4\overline{\omega}\Omega\omega}dW, \\ S_\omega &= (C_{\epsilon 2} - 1) - (C_{\epsilon 1} - 1)\frac{\mathcal{P}}{\epsilon} - (C_{\epsilon 3} - 1)\frac{\mathcal{G}}{\epsilon}, \end{aligned} \tag{25}$$

where $\mathcal{P}$ and $\mathcal{G}$ are shear and buoyancy productions, respectively, and Ω is the conditional-mean turbulence frequency. The Reynolds-averaged dissipation is determined as $\epsilon = \Omega k$. C_3, C_4, $C_{\epsilon 1}$, $C_{\epsilon 2}$, and $C_{\epsilon 3}$ are model constants, and their values are set as in Shin and Baik (2024).

At each time step, particle properties and positions are updated using a second-order time integration scheme. Turbulence statistics are obtained by averaging particle values at each vertical layer, employing a particle-in-cell approach. The mean fields evolve according to the convergence of the computed vertical fluxes. Further details of the model formulation and implementation are given in Shin and Baik (2024). LSSTM adopts a single-column model framework that employs the Reynolds-averaged Navier–Stokes (RANS) equations, where the Reynolds operator represents a horizontal average under the assumption of horizontal homogeneity. Unlike the classical RANS approach, which relies on temporal or ensemble averaging, this framework does not apply temporal averaging.

3 Results

3.1 Tests with Simple 1D Diffusion Processes

3.1.1 Description of Diffusion Processes

To validate the IS framework in a controlled setting, we examine two one-dimensional diffusion processes whose invariant densities correspond to the normal and log-gamma distributions, respectively. First, the SDE whose invariant density is a normal distribution, $G(x) \propto \exp\left(-\frac{\theta}{\sigma^2}x^2\right)$, is associated with the well-known Ornstein–Uhlenbeck process, given by:

$$\mathrm{d}X_t = -\theta\, X_t\, \mathrm{d}t + \sigma\, \mathrm{d}W_t. \tag{26}$$

Next, we consider a diffusion process whose invariant density follows a log-gamma distribution, $G(x) \propto e^{\beta x}e^{-e^x/\alpha}$, characterized by scale and shape parameters, α and β, respectively. The corresponding SDE is:

$$dX_t = \frac{\sigma^2}{2}\left(\beta - \frac{e^{X_t+m}}{\alpha}\right)dt + \sigma dW_t. \tag{27}$$

The term $m = \psi(\beta) + \ln\alpha$, where $\psi(\cdot)$ denotes the digamma function (i.e., the logarithmic derivative of the gamma function), shifts the distribution to ensure a zero mean. The log-gamma distribution is used to evaluate the robustness of IS under skewed distributions, which are often observed in convective boundary layers. In the following experiments, the parameters for SDEs are set as $\theta = 0.5$, $\sigma = 1$, $\alpha = 0.5$, and $\beta = 1$.

The 1D diffusion process simulations are conducted using a specified number of samples over 100 time steps with a time increment of $\Delta t = 0.1$, employing the EM scheme for time integration. For the resampling method, we found that the performance of IS is insensitive to the specific choice of method. Accordingly, we adopted systematic resampling, which is both simple and computationally efficient. Particles are initially drawn from a standard normal distribution and assigned equal weights. Statistical estimates are computed at the final time step, where the sample distribution has reached statistical stationarity.

3.1.2 Determining Parameters

To enhance the performance of IS, the controllable parameters p, s_f, and $\mathrm{ESS}_{\min}$ need to be optimized. For simplicity, we fix $p = 1.5$, based on the following rationale. Consider reducing variance in the Monte Carlo estimation of vertical fluxes ($\overline{w'\phi'}$). Given that ϕ' is correlated with w', the optimal exponent for sampling weight $|w'|^p$ generally falls within the range $1 < p < 2$. Consequently, we focus on optimizing s_f and $\mathrm{ESS}_{\min}$ while keeping p fixed at 1.5, which is simply chosen as the midpoint of this range. The optimal choice of p depends on the specific target statistics and simulation case, so there is no single optimal value that is universally adoptable. In practice, it is advisable to tune p according to the particular application.

We ran a two-dimensional grid search across various values of $\mathrm{ESS}_{\min}$ and s_f to find optimal values. At each grid point, 10^5 independent simulations with IS are conducted using a sample size of $N = 100$, and then the $(s_f, \mathrm{ESS}_{\min})$ pair that minimized the statistical variance of the second moment is selected. The second moment is set as the target statistics because statistical robustness is mostly required for the second moments in turbulence models.

Figure 2 summarizes the grid search optimization result. The normal and log-gamma density processes show slightly different dependency on s_f and $\mathrm{ESS}_{\min}$, while the log-gamma shows more robust dependency over s_f and $\mathrm{ESS}_{\min}$. The optimized parameters for the normal density process are $s_f = 1.0$ and $\mathrm{ESS}_{\min} = 0.9N$, and the optimized parameters for the log-gamma density process are $s_f = 0.8$ and $\mathrm{ESS}_{\min} = 0.85N$. However, slightly changing these parameters does not noticeably degrade the performance of IS. Therefore, we selected $s_f = 1.0$ and $\mathrm{ESS}_{\min} = 0.9N$, since these values work well in both skewed and non-skewed distributions.

When $\mathrm{ESS}_{\min} = 1N$, resampling occurs almost every time step, leading to a significant increase in the standard deviation due to deteriorating sample quality caused by frequent duplication. Conversely, when $\mathrm{ESS}_{\min} < 0.6N$, resampling rarely occurs, resulting in severe weight divergence. An intermediate value, $\mathrm{ESS}_{\min} \approx 0.9N$, provides a balanced

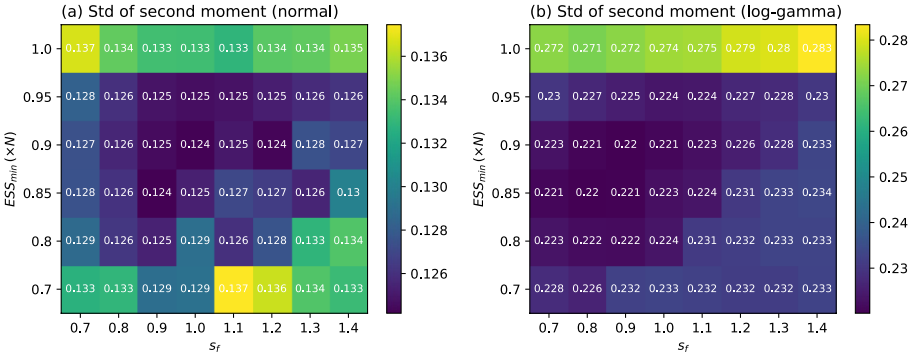

Fig. 2 Heatmaps showing standard deviations of second-moment estimation from diffusion process simulations using IS with different $\mathrm{ESS}_{\min}$ and s_f values, which have invariant density of **a** normal and **b** log-gamma distribution. The standard deviations are calculated from independent 10^5 simulations with a sample size $N = 100$

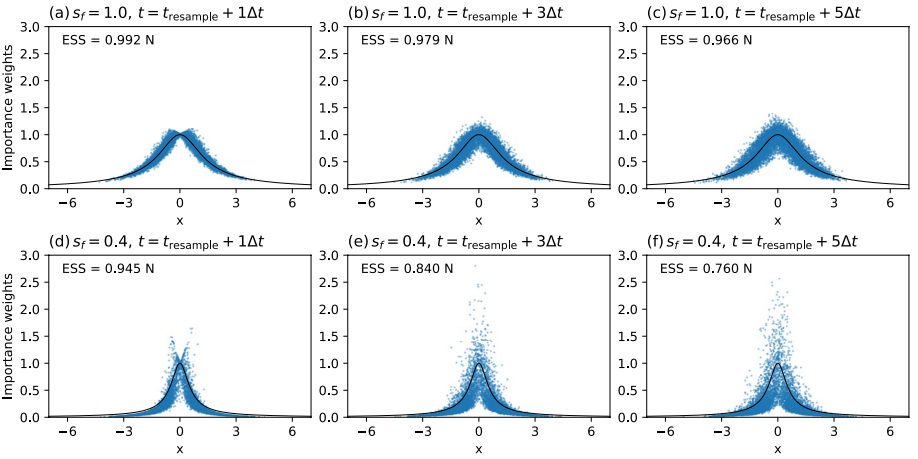

Fig. 3 Distribution of unnormalized importance weights of the normal invariant density process at **a**, **d**$+1\Delta t$, **b**, **e**$+3\Delta t$, and **c**, **f**$+5\Delta t$ after resampling, where the top row corresponds to the simulation with $s_f = 1.0$, and the bottom row to $s_f = 0.4$. The black solid line indicates the expected importance weight $1/f(x) = [1 + \frac{1}{p}(w'/s_f)^2]^{-p/2}$ where σ_w is set as 1. The effective sample size (ESS) is shown in the top-left corner of each panel

resampling frequency (approximately 6% of all time steps) and helps maintain sample quality while avoiding excessive duplication.

The value of s_f also plays a critical role in the performance of IS. To examine its effect, we test the normal invariant density process using $s_f = 1.0$ and $s_f = 0.4$ and analyze the evolution of importance weights after resampling (Fig. 3). In these simulations, $\mathrm{ESS}_{\min}$ is set to $0.7N$, with a total sample size of $N = 10^4$. When $s_f = 1.0$, the importance weights diverge from the expected weights relatively slowly. For instance, five time steps after resampling, ESS remains high at $\mathrm{ESS} = 0.966N$. In contrast, with $s_f = 0.4$, the importance weights diverge from the expected value rapidly, particularly near $x = 0$. The ESS drops to $0.760N$ at five time steps after resampling. This behavior arises because the one-step

importance weight W_k (see Equation 12 and Equation 16 for the Euler–Maruyama scheme) becomes extremely small or large when the gradient of $f(x)$ is large, given that the additional drift term is $\frac{1}{2}\sigma^2 \frac{f'(x)}{f(x)}$. This result explains the tendency that the optimal $\mathrm{ESS}_{\min}$ decreases as s_f decreases in Figure 1, where a smaller $\mathrm{ESS}_{\min}$ helps prevent overly frequent resampling when s_f is small.

3.1.3 Variance Reduction in Estimated Moments

To evaluate the effectiveness of the proposed IS with the optimized parameters, $p = 1.5$, $s_f = 1.0$, and $\mathrm{ESS}_{\min} = 0.9$, we compare statistical estimates of moments and tail probabilities computed with and without IS for the normal and log-gamma invariant density processes. The results are summarized in Table 1, showing mean estimates along with their standard deviations from repeated simulations.

For the normal distribution, IS achieves noticeable variance reduction across all evaluated statistics. The variance of the second moment shows a reduction of 28.1%, while higher-order moments benefit more substantially: the third and fourth moments exhibit variance reductions of 44.5% and 54.3%, respectively. IS is most effective for estimating the tail probability $P(|x| > 3\sigma)$, where variance is reduced by 63.0%. The log-gamma distribution, which represents a skewed and heavy-tailed density, demonstrates even greater improvements. IS reduces variance by 50.1% for the second moment and by 68.9% and 78.4% for the third and fourth moments, respectively. The variance reduction for the tail probability is somewhat similar to the normal distribution, where 64.5% is reduced. Overall, the results demonstrate that IS significantly improves the precision of statistical estimates, particularly for higher-order moments and rare-event probabilities.

The means of the estimated statistics show slight deviations from the analytical means, with a bias of order $O(\Delta t)$, arising from time discretization by the EM scheme. For the normal density process under the EM scheme, the expected second, third, and fourth moments (when the sample size is infinite) can be computed analytically as 1.02564, 0, and 3.15582, respectively. The expected moments of the log-gamma density process cannot be obtained in closed form. The simulations with and without IS both yield means close to the expected values, but there is a tendency for the IS method to slightly amplify the numerical bias in

Table 1 Statistical estimates in the repeated simulations of the normal and log-gamma invariant density diffusion processes with and without IS

Invariant density	Statistics	Without IS	With IS	Variance reduction (%)
Normal	2nd moment	1.0257 ± 0.0458	1.0261 ± 0.0388	28.1
	3rd moment	− 0.0005 ± 0.1266	0.0002 ± 0.0943	44.5
	4th moment	3.1559 ± 0.3259	3.1585 ± 0.2204	54.3
	$P(\|x\| > 3\sigma)$	0.0027 ± 0.0015	0.0027 ± 0.0009	63.0
Log-gamma	2nd moment	1.5892 ± 0.0977	1.5902 ± 0.0690	50.1
	3rd moment	− 1.9880 ± 0.4972	− 1.9927 ± 0.2775	68.9
	4th moment	12.0384 ± 3.0349	12.0648 ± 1.4094	78.4
	$P(\|x\| > 3\sigma)$	0.0108 ± 0.0025	0.0109 ± 0.0015	64.5

The values are reported as the mean ± standard deviation of each statistic at the final time step, computed from 10^5 simulations with the sample size $N = 10^3$. The final column shows the percentage reduction in variance achieved by IS. Statistics include second, third, fourth moments, and the tail probability $P(|x| > 3\sigma)$

the means of estimated statistics. Nevertheless, these biases are much smaller than the variances, and the variance reduction achieved through IS offers a substantially greater benefit.

Figure 4 shows the simulated stationary distributions obtained from unweighted and weighted samples in diffusion process simulations with IS. As displayed in the figure, the additional drift in SDEs effectively tilts the samples to the proposal density $f(x)G(x)$, and the weighted samples follow the target invariant density $G(x)$ well. Due to the tilting, the sample number in the tails is substantially increased, and this explains the preferential enhancement of higher-order and tail statistics. The effect of the tilting is more pronounced in the tail of the log-gamma distribution. The variance of the weighted density is relatively high near $x = 0$, especially in the normal distribution, since importance weights deviate strongly near this point (Fig. 3). However, the samples near $x = 0$ contribute less to the statistics, so the deviation has little impact on the variance of the estimated statistics.

The variance of the estimated statistics of Monte Carlo integration is expected to scale with the sample size as N^{-1}, and our results follow this slope closely (Fig. 5). The percentage reduction in variance achieved by IS is found to be nearly independent of sample size, implying that computational efficiency gains can be expected for any N. Since computational cost is proportional to N, the reduction in cost required to achieve a given level of statistical uncertainty is equivalent to the variance reduction. For example, employing IS with the log-gamma distribution reduces the computational cost by approximately 50% for second-moment estimation and by 78% for fourth-moment estimation.

To understand better how IS performance responds to model parameters and PDF shape, sensitivity tests on the exponent of the sampling weight function p, and the shape parameter of the log-gamma distribution, β, are conducted (Fig. 6). In general, the variances of the tested statistics decrease as p increases for both the normal and log-gamma density processes (Fig. 6a). This behavior occurs because a larger p increases the number of samples in the distribution tails, which strongly influences the statistical variance. However, the sensitivities to p differ among the statistics. The variance reduction of the second moment peaks near $p = 2$ and then slightly decreases for $p > 2$. The variance reductions of the third and fourth moments tend to saturate around $p = 3$ and $p = 4$, respectively. These results suggest that the optimal value of p is approximately equal to the effective order of the target statistic. Although Fig. 6a indicates that increasing p reduces the variances of statistics overall, excessively large p values are not recommended; greater mismatches between the target and

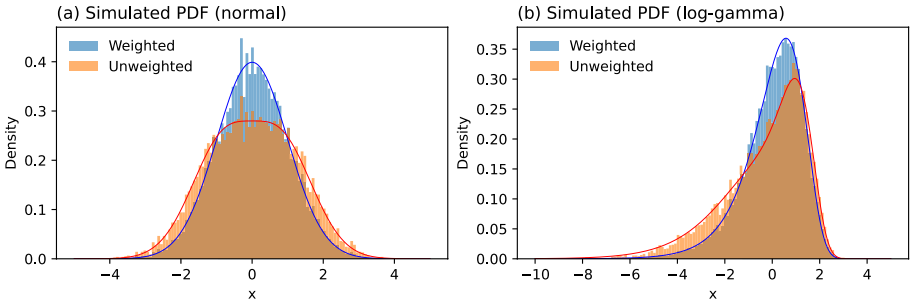

Fig. 4 Histogram showing simulated distributions of unweighted and weighted samples from the final time step of diffusion process simulations with IS, which have invariant density of **a** normal and **b** log-gamma distribution. The sample number for each simulation is $N = 10^4$. The blue solid line indicates the target invariant density $G(x)$, and the red solid line indicates the proposal density $f(x)G(x)$

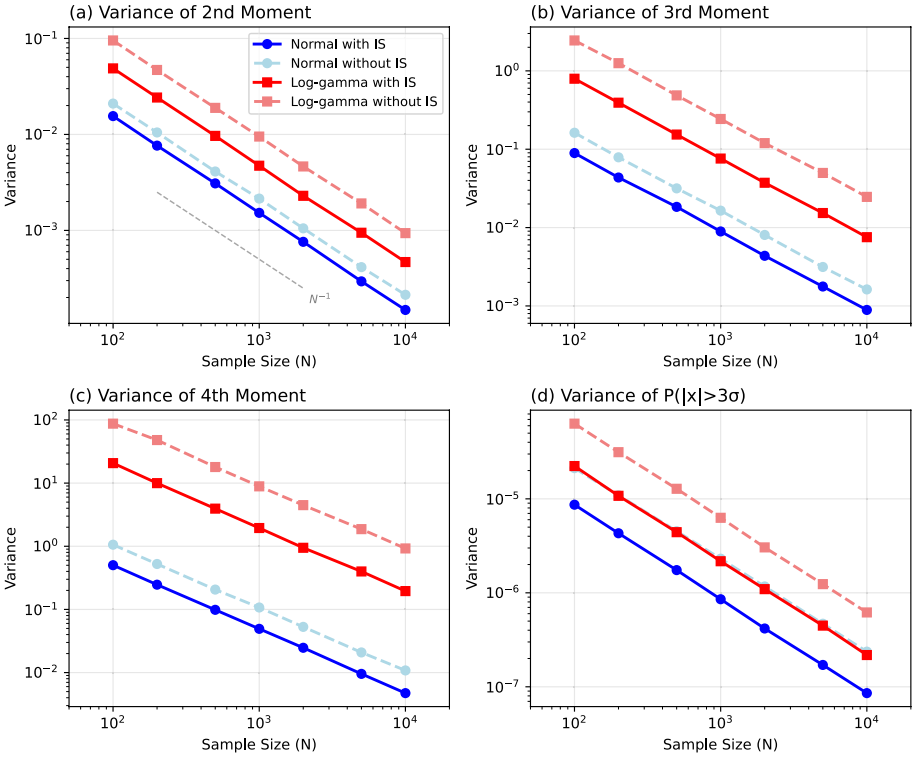

Fig. 5 Variances of the estimated **a** second moment, **b** third moment, **c** fourth moment, and **d** $P(|x| > 3\sigma)$ as functions of sample size N, computed from 10^5 simulations for each case. Results from simulations of normal and log-gamma invariant density diffusion processes, with and without IS, are compared. The gray dashed line in panel **a** denotes a reference slope of N^{-1}

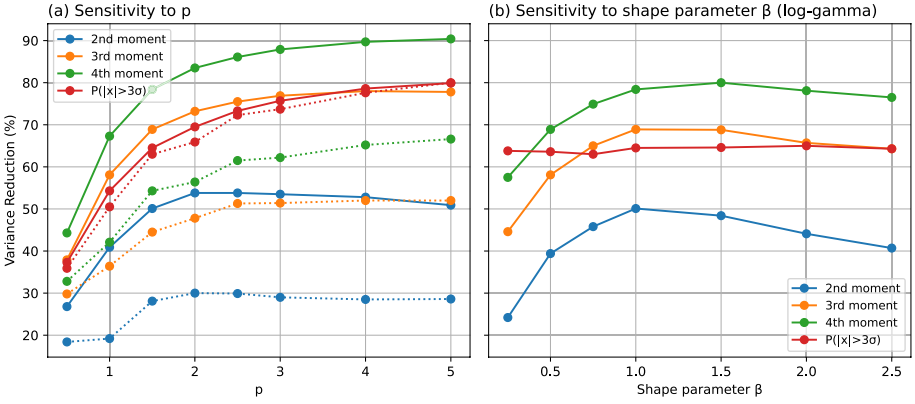

Fig. 6 Percentage reductions in variance achieved using IS for the second, third, and fourth moments, as well as for $P(|x| > 3\sigma)$, shown as functions of **a** p and **b** the shape parameter β. Solid and dashed lines represent results for the normal and log-gamma density diffusion process, respectively. Each data point is obtained using an ensemble-averaged statistic at the final time step, computed from 10^5 simulations with the sample size $N = 10^3$

proposal transition densities can lead to overly frequent resampling, thereby disrupting the preservation of Lagrangian sample histories.

Looking into the sensitivity to β, the performance of IS exhibits a non-monotonic trend. The variance reduction increases with β for $\beta < 1$ and decreases for $\beta > 1.5$. The skewness of the log-gamma distribution decreases with β. Therefore, the decrease in IS performance with increasing β when $\beta > 1.5$ is expected, since a thinner left tail limits the contribution of extreme events. When $\beta < 1$, however, the variance reduction decreases even as skewness grows, which appears counterintuitive. Our analysis reveals that this behavior arises from weight divergence at the right tail: the right tail of the log-gamma distribution decays super-exponentially, becoming extremely thin when $\beta < 1$, and mismatches between the target and proposal transition densities become severe. Fortunately, atmospheric turbulence typically exhibits exponential tail behavior, so such degradation is unlikely in practical applications. To further mitigate potential weight divergence at the distribution tails, a limiter can be introduced on the additional drift term in the SDEs, and/or adaptive time-stepping can be applied to ensure bounded particle motion and prevent excessive weight accumulation in extreme regions. The performance of IS for $P(|x| > 3\sigma)$ remains nearly independent of β, implying that IS performance for tail probabilities is primarily governed by overall sampling efficiency rather than by the detailed shape of the baseline PDF.

3.2 Simulation of Boundary Layer Turbulence

A dry convective boundary layer (DCBL) case using LSSTM is simulated to evaluate the performance of IS in a full boundary layer simulation. The DCBL case is forced with a constant sensible heat flux of $0.1\,\mathrm{K\,m\,s^{-1}}$, and the initial potential temperature profile is specified with a constant gradient of $\partial\bar{\theta}/\partial z = 3\,\mathrm{K\,km^{-1}}$ and a surface potential temperature of 300 K. The surface pressure and roughness length are set to 1015 hPa and 0.01 m, respectively. The vertical domain extends to 2 km, with a uniform vertical grid spacing of 20 m. A fixed time step of $\Delta t = 5\,\mathrm{s}$ is used, and the case is integrated for 4 h. Turbulence statistics (e.g., $\overline{w'\theta'}$), defined as horizontally Reynolds-averaged quantities, are computed at each model time step. The temporal mean and temporal variance of these statistics are then calculated from the turbulence statistics obtained at all time steps during the time window from 3.75 to 4.0 simulation hours. The reason for the specific time window is explained later (Fig. 10). To enhance statistical robustness, the temporal means and variances are ensemble-averaged across 10 ensemble members. The ensembles are generated using different random seeds for pseudo-random number generation.

We conducted simulations with and without IS, using total particle counts of $N = 10^4$ and $N = 10^5$. According to Shin and Baik (2024), the mean profiles simulated by LSSTM are nearly converged when $N > 5 \times 10^4$. Therefore, the case with $N = 10^4$ represents a simulation that has not yet reached statistical convergence, while $N = 10^5$ corresponds to a nearly converged simulation. In addition, a reference (ground-truth) simulation with a very large particle count of $N = 10^8$ is performed without IS, as it is expected to exhibit negligible bias error. The parameters for IS are set as the values obtained from the optimization using 1D diffusion, where $p = 1.5$, $s_f = 1.0$, and $\mathrm{ESS}_{\min} = 0.9N$. Systematic resampling is employed for its simplicity, and the simulation results are found to be insensitive to the choice of resampling method.

For the simulations with IS, the resampling step is modified to ensure consistency between the particle mass density and the base-state density profile. In LSSTM, the algorithm of Wild (2013) is employed to correct particle mass density by adding a small correction velocity to the particle vertical velocity. However, the method does not eliminate particle density biases over larger vertical extents, because it applies corrections locally. To address this, the resampling is modified to increase particles where particle mass density is low, multiplying the ratio of the particle mass density to the base-state density by the resampling weight. The resulting resampling weight takes the form:

$$\widetilde{W}_{\text{resample}} \propto \frac{\rho_0(z)}{\overline{q}(z)} \frac{\widetilde{W}}{1/f(X)}, \tag{28}$$

where $\rho_0(z)$ and $\overline{q}(z)$ denote the base-state density and the particle mass density at height z, respectively. The new IS technique thereby enables a more robust consistency criterion at negligible additional computational cost.

Figure 7 presents the ensemble-averaged vertical profiles of turbulence statistics simulated by LSSTM. Although LSSTM employs a Monte Carlo approach, which is, in principle, unbiased for a single time integration, a systematic bias can develop over time due to the nonlinear interaction between fluctuations in turbulence statistics and particle distribu-

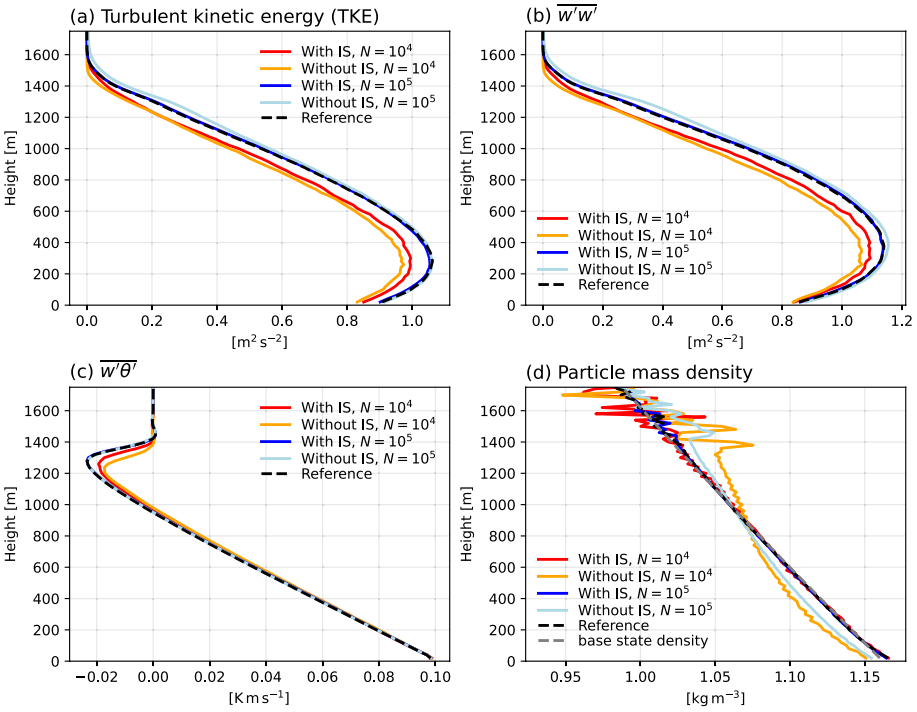

Fig. 7 Ensemble-averaged vertical profiles of temporal means of **a** turbulent kinetic energy (TKE), **b** $\overline{w'w'}$, **c** $\overline{w'\theta'}$, and **d** particle mass density from the LSSTM simulations of the DCBL case, comparing results with and without IS and with total particle counts of $N = 10^4$ and $N = 10^5$. Each profile is averaged over 10 ensemble members over 3.75–4.0 simulation hours. The black dashed lines denote the profiles from the reference simulation, and the gray dashed line in **d** denotes the base state density

tions. Specifically, Pope (1995) identified that the statistical fluctuations in the coefficients of the SDEs are the source of this bias, where the bias scales as N^{-1}. Shin and Baik (2024) demonstrated that LSSTM exhibits a systematic bias in which entrainment is weakened in the entrainment zone when the sample size is small. Consistently, the simulations with $N = 10^4$ show weaker entrainment compared to the reference simulation (Fig. 7c). The simulations with $N = 10^4$ also produce weaker TKE and vertical velocity variance. The use of IS slightly mitigates this bias, enhancing both TKE and entrainment strength, and the profiles are closer to the reference simulation.

For the simulations with $N = 10^5$, the difference between cases with and without IS is small since the mean profiles approximately reach statistical convergence when $N > 5 \times 10^4$, and both simulations produce mean profiles close to the reference. However, the simulations with IS exhibit slightly lower TKE and $\overline{w'w'}$ in the entrainment zone (height range of 1000–1400 m), yielding better agreement with the reference profiles. This difference likely arises from discrepancies in particle mass density. Simulations with IS produce particle mass densities that closely match the base-state density, whereas simulations without IS underestimate particle mass densities below 1000 m and overestimate them above (Fig. 7d). The bias in particle mass density is more pronounced for $N = 10^4$ than for $N = 10^5$. Accumulation of particles in the upper layer means enhanced upward transport, which in turn increases turbulence strength in the upper boundary layer. The simulations with IS effectively reduce this bias through the modified resampling procedure.

The reduction of the bias in the mean profile arises from the decrease in statistical fluctuations achieved by IS. Figure 8 compares the temporal variances of simulated second-order statistics. Here, the temporal variances are computed as the variance of statistical estimates collected across all time steps within the time window. For both $N = 10^4$ and $N = 10^5$, the temporal variances are substantially lower in the IS simulations than in the non-IS simulations. Variance reduction occurs throughout the boundary layer, although its magnitude varies with height. The fractional reduction is greatest in the middle of the boundary layer and somewhat smaller near the boundary layer top. The variance is relatively high for $\overline{w'\theta'}$ near the boundary layer top, since particles that penetrate the entrainment layer, which correspond to strong thermal convection, have high vertical velocity and temperature perturbation. The IS also reduces deviations between ensembles, which is denoted in the shading in the figure.

The magnitude of variance reduction depends on the specific statistic considered: the largest reduction, approximately 47%, is observed in $\overline{w'\theta'}$, while the smallest reduction is found in TKE (Table 2). The effectiveness of IS is determined by both the heaviness of the distribution tails and the functional dependence of the observable associated with the target statistic. Since the proposed sampling weight is proportional to $|w'|^{1.5}$, IS is most efficient when the observable exhibits strong dependence on $|w'|$. This dependence was quantified from the simulated data by fitting each observable to an effective exponent p in $|w'|^p$. The estimated exponent for $\overline{w'\theta'}$ is 2.3, which is larger than those of the other statistics: 1.0 for $\overline{w'w'}$, 2.0 for $\overline{w'w'}$, and 1.3 for TKE. The particularly strong variance reduction in $\overline{w'\theta'}$ arises from buoyancy-driven acceleration, whereby w' increases nonlinearly with θ', amplifying its dependence on vertical velocity.

As in the 1D diffusion process tests, the variance reduction is larger for third-order and tail statistics than for second-order statistics in the boundary layer simulation (Fig. 9). For third moments, variances are reduced by roughly half throughout the boundary layer, with

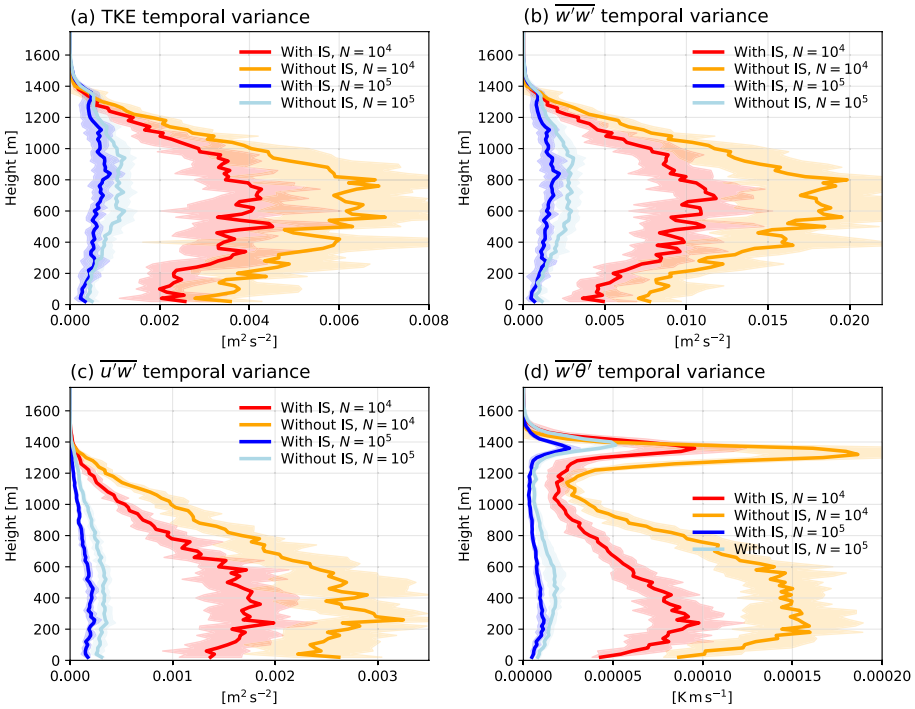

Fig. 8 Ensemble-averaged vertical profiles of temporal variances of **a** turbulent kinetic energy (TKE), **b** $\overline{w'w'}$, **c** $\overline{u'w'}$, and **d** $\overline{w'\theta'}$ from the LSSTM simulations of the DCBL case over 3.75–4.0 simulation hours, comparing results with and without IS and with total particle counts of $N = 10^4$ and $N = 10^5$. Each profile is an average of 10 ensemble members, with solid lines indicating ensemble means and shading denoting the ± 1 standard deviation range

Table 2 Boundary-layer-averaged temporal variances and variance reduction with IS from LSSTM simulations of the DCBL case over 3.75–4.0 simulation hours

Variable	LSSTM $N = 10^4$			LSSTM $N = 10^5$		
	Without IS	With IS	Reduction (%)	Without IS	With IS	Reduction (%)
TKE [m^{2}s^{-2}]	0.003915	0.002559	34.6	0.000747	0.000518	30.7
$\overline{w'w'}$[m^{2}s^{-2}]	0.010729	0.006525	39.2	0.001875	0.001201	35.9
$\overline{u'w'}$[m^{2}s^{-2}]	0.001587	0.000970	38.9	0.000211	0.000116	45.3
$\overline{w'\theta'}$[K m s^{-1}]	0.000098	0.000051	47.5	0.000014	0.000007	47.3
$\overline{w'w'w'}$[m^3 s^{-3}]	0.047979	0.024488	49.0	0.008584	0.003928	54.2
$\overline{w'w'\theta'}$[K m^2 s^{-2}]	0.000346	0.000158	54.3	0.000057	0.000026	54.1
$P(w' > 2\,\mathrm{m\,s^{-1}})$	0.000141	0.000066	53.0	0.000020	0.000010	50.7
$P(w' > 3\,\mathrm{m\,s^{-1}})$	0.000018	0.000008	57.5	0.000003	0.000001	62.9

The temporal variances are averaged across the vertical range $z = 0$–1600 m and over 10 ensemble members

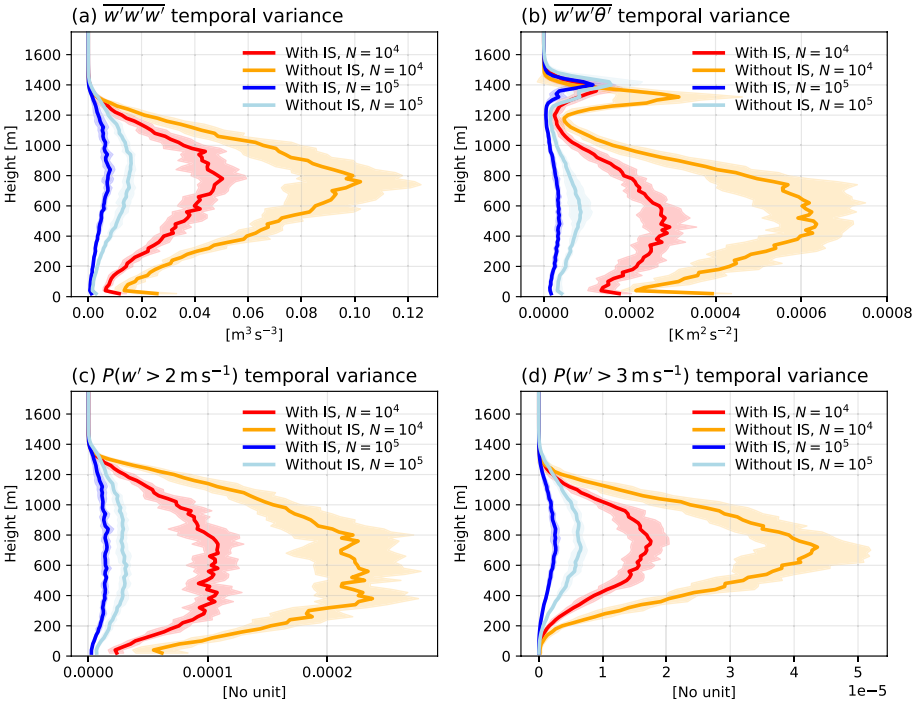

Fig. 9 Ensemble-averaged vertical profiles of temporal variances of **a** $\overline{w'w'w'}$, **b** $\overline{w'w'\theta'}$, **c** $P(w' > 2\,\mathrm{m\,s^{-1}})$, and **d** $P(w' > 3\,\mathrm{m\,s^{-1}})$ from the LSSTM simulations of the DCBL case over 3.75–4.0 simulation hours, comparing results with and without IS and with total particle counts of $N = 10^4$ and $N = 10^5$. Each profile is averaged over 10 ensemble members, with solid lines indicating ensemble means and shading denoting the ± 1 standard deviation range

the reduction in $\overline{w'w'\theta'}$ slightly greater than in $\overline{w'w'w'}$ (Table 2). The tail probabilities $P(w' > 2\,\mathrm{m\,s^{-1}})$ and $P(w' > 3\,\mathrm{m\,s^{-1}})$ represent the fractions of high vertical velocity, which can be interpreted as the convective updraft fractional area in the boundary layer. The variances of these tail probabilities are reduced by 53% (50.7%) for $P(w' > 2\,\mathrm{m\,s^{-1}})$ and by 57.5% (62.9%) for $P(w' > 3\,\mathrm{m\,s^{-1}})$ with $N = 10^4$ ($N = 10^5$). These results clearly demonstrate the advantage of IS in sampling the tails of the distribution. In particular, convective updrafts characterized by large w' values are critical for entrainment and vertical transport in the boundary layer, yet they are rarely captured in standard Monte Carlo approaches due to their low occurrence. IS substantially improves the statistical representation of these rare but dynamically important convective events. We also perform simulations with different values of N, and the percentage of variance reduction achieved by IS is found to be largely insensitive to N, similar to what was observed in the 1D diffusion process tests.

Here, we examine mathematically how the temporal variance of estimated statistics depends on turbulent fluctuations and on the number of Lagrangian particles. Let the target statistic be $m = \mathbb{E}[h(X)]$, and let its Monte Carlo estimator at a given time step be $\hat{m} = \frac{1}{N}\sum_{i=1}^{N} h(X_i)$. In the absence of IS, the ensemble variance of this estimator is:

$$\mathrm{Var}(\hat{m}) = \frac{1}{N}\left(\mathbb{E}[h(X)^2] - m^2\right). \tag{29}$$

When IS is employed, the variance of the corresponding self-normalized IS estimator is approximately:

$$\mathrm{Var}(\hat{m}) \approx \frac{1}{N}\frac{\mathbb{E}_\pi\left[\widetilde{W}(X)^2\,(h(X)-m)^2\right]}{\left(\mathbb{E}_\pi[\widetilde{W}(X)]\right)^2}, \tag{30}$$

where $\mathbb{E}_\pi[\cdot]$ denotes expectation with respect to the proposal density π. This approximation follows from a central limit theorem for the numerator and denominator of the self-normalized IS estimator. In both the non-IS and IS cases, $\mathrm{Var}(\hat{m})$ scales as N^{-1} and m^2 for fixed PDF shape. In particular, if h represents a variance of turbulent fluctuations such as u'^2, v'^2, or w'^2, then $\mathrm{Var}(\hat{m})$ is proportional to the square of the corresponding turbulence variance.

$\mathrm{Var}(\hat{m})$ represents the ensemble variance of the target statistic at a single time step. In practice, however, this variance is estimated from the LSSTM results by evaluating the temporal variability of $\hat{m}(t)$ over a finite time window (3.75–4.0 simulation hours). Figure 10 shows the boundary-layer–averaged temporal variances of $\overline{w'w'}$ as a function of time-window length from LSSTM simulations. The averaging window is positioned near the end of the simulation period, when boundary-layer growth has slowed. When the time window is shorter than the turbulence decorrelation time scale (approximately 10 min in the DCBL case), the measured temporal variance is underestimated because the process has not fully decorrelated. Conversely, if the window is too long, systematic trends in the statistics may dominate over random fluctuations, leading to overestimation of the variance. Based on this behavior, a time-window length of 0.25 h, for which the temporal variance is nearly converged, was adopted for the analysis.

Figure 11 presents the particle number density as a function of vertical velocity and height (first and second columns) and as a function of height alone (third column). Without IS, particles are primarily concentrated in regions with small negative w', reflecting compensating subsidence, and the probability of convective updrafts is relatively small compared to that of

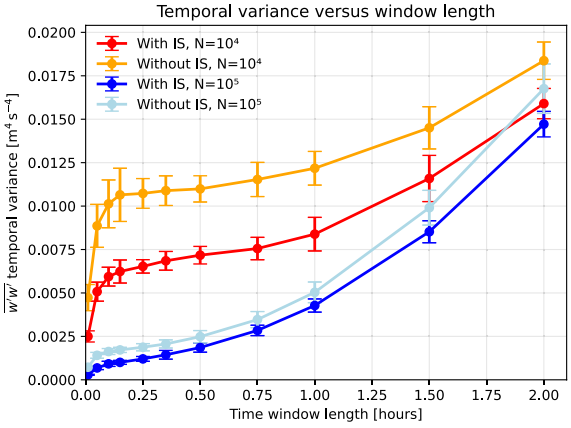

Fig. 10 Boundary-layer-averaged temporal variances of $\overline{w'w'}$ as a function of time-window length from LSSTM simulations of the DCBL case, comparing results with and without IS and with total particle counts of $N = 10^4$ and $N = 10^5$. The end time of each time window is fixed at $t = 4$ h. The temporal variances are averaged over the vertical range $z = 0$–1600 m and across 10 ensemble members, and the error bars indicate standard deviations among the 10 members

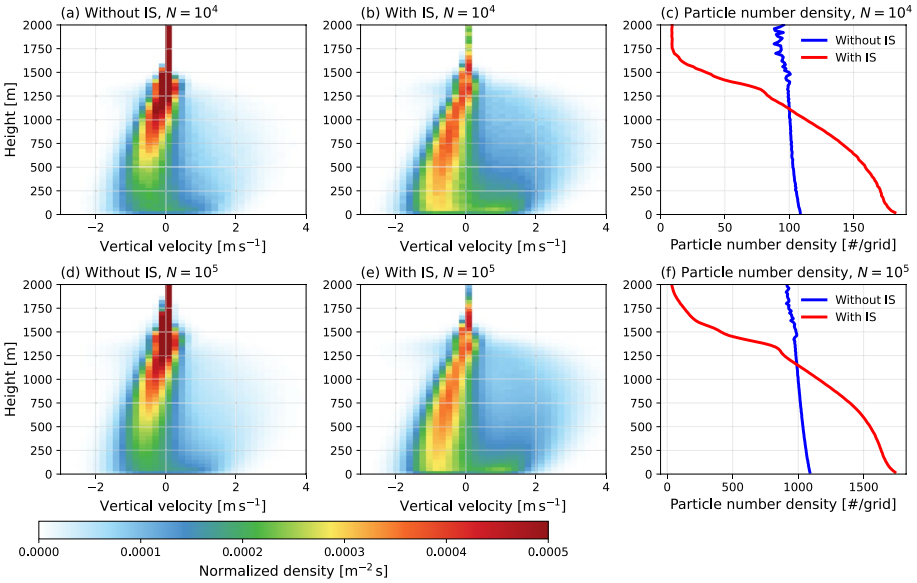

Fig. 11 **a**, **b**, **d**, **e** Two-dimensional density plots of the particle number density as a function of vertical velocity and height and **c**, **f** one-dimensional profiles of particle number density, using samples from the 3.75–4.0 simulation hour period of 10 ensemble LSSTM simulations of the DCBL case, comparing results with and without IS and with total particle counts of $N = 10^4$ and $N = 10^5$

non-convective regions. With IS, the probability of convective updrafts increases substantially, while the number of non-convective samples decreases. As the subsidence descends into the lower boundary layer, the descending speed increases due to negative buoyancy. In the IS simulations, the probability of strong convective downdrafts is significantly enhanced in the lower boundary layer. Near the surface, the magnitudes of updrafts and downdrafts become comparable, and the distribution approaches Gaussian. Without IS, the probability of near-zero vertical velocity is high; however, IS reduces these samples and enhances the occurrence of initial convective updrafts near the surface.

Another advantage of IS is its ability to reduce sampling in non-turbulent regions adaptively. Without IS, the particle number per grid cell is nearly uniform across the vertical domain (Fig. 11c, f). However, particles above the boundary layer add considerable computational cost while contributing little to turbulent fluxes. With IS, the particle number density in each layer is approximately proportional to $\sigma_w(z)^p$, such that particles are preferentially resampled in layers where vertical velocity variance is high. Above the boundary layer, the particle number is greatly reduced but not eliminated, since the $\sigma_w(z)^p$ term in the sampling weight function is smoothed using a running average over height. This ensures that convective particles entering a non-turbulent layer are retained, allowing entrainment processes to be properly represented. As shown in Fig. 11b and e, downdrafts driven by strong negative buoyancy in the entrainment zone are well captured in the IS simulations.

Figure 12 presents the unweighted and weighted distributions of w and θ, along with mean weights as functions of w and θ, at $z = 750\,\mathrm{m}$ in the IS simulations. The distributions are normalized by their integrals, so the effect of increased sample numbers within the layer is not reflected here. IS substantially increases the number of high-speed updraft samples,

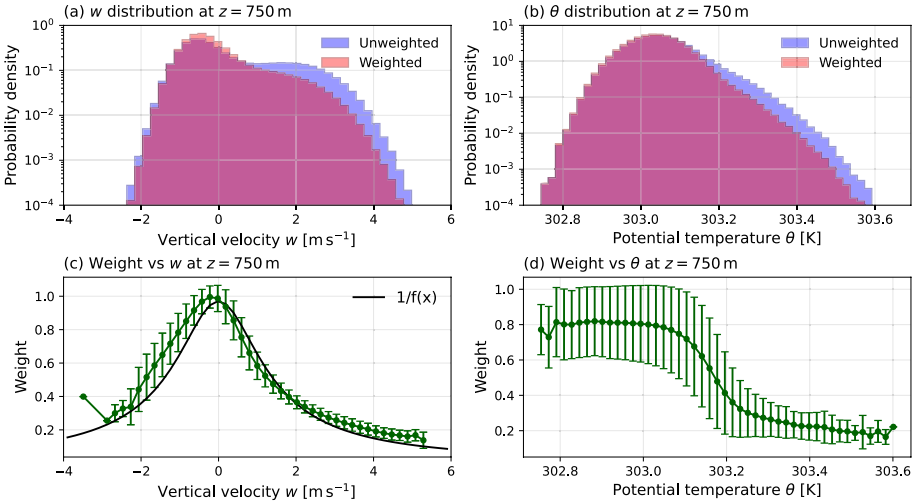

Fig. 12 Particle distribution analysis at the height range $z = 750 \pm 50\,\mathrm{m}$, using the samples from the 3.75–4.0 simulation hour period of 10 ensemble LSSTM simulations of the DCBL case, which is simulated with IS and a total particle count of $N = 10^5$. The first row presents the unweighted and weighted distributions of **a** vertical velocity and **b** potential temperature, while the second row shows the mean particle weights as functions of **c** vertical velocity and **d** potential temperature, with error bars representing the standard deviation. In **c**, the specified importance weight ($1/f(x)$) is plotted as a black solid line

nearly doubling the occurrence of samples with $2\,\mathrm{m\,s^{-1}}$, while reducing the number of low vertical velocity samples. Downdraft samples are also increased, though to a lesser extent than updrafts. This asymmetric sampling is also evident in the mean weight distribution (Fig. 12c). For positive velocities, the mean weights approximately follow the specified importance weight, $1/f(x)$. In contrast, samples with negative vertical velocity have weights larger than $1/f(x)$ and show considerable deviations. These increased weights for downdrafts arise because they originate from the upper boundary layer, where particle number density is low. It should be noted that weights deviate from the specified importance weight due to the multiplication of stepwise importance weights in sequential IS, and they are reset to $1/f(x)$ at each resampling step.

For the θ distributions at $z = 750\,\mathrm{m}$, IS preferentially increases samples in the right tail, while samples with $\theta < 303.1\,\mathrm{K}$ are reduced (Fig. 12b). Consistently, the low-weight samples are concentrated at high θ (Fig. 12d). This occurs because both convective updrafts and downdrafts exhibit higher temperatures than the surrounding environment in the DCBL case: updrafts are heated by surface sensible heat fluxes, whereas downdrafts are heated by entrainment from the high-θ free atmosphere. These results demonstrate that the proposed IS technique generally enhances the statistical robustness of convection, regardless of the variable considered.

Regarding computational cost, the mean LSSTM integration time of the 4-hour DCBL case for 10 ensemble members is measured. The code of LSSTM is written in the Julia programming language (Bezanson et al. 2017), and the simulations are executed on a single core of a 2.9 GHz Intel Xeon CPU. The simulations without IS take 253 s for $N = 10^5$ and 23.2 s for $N = 10^4$, and those with IS take 260 s for $N = 10^5$ and 23.8 s for $N = 10^4$ on average. Therefore, the new IS technique introduces only a small overhead of about 2–3%

in LSSTM simulations. IS operates efficiently, with the primary additional cost arising from tracking the weights of individual particles. Resampling contributes little to the overall cost, as it is required in only 5–6% of time steps and is implemented using the highly efficient systematic resampling method. Overall, the added computational expense is negligible compared to the substantial benefits gained from statistical variance reduction.

4 Summary and Conclusions

This study presented and evaluated an importance sampling framework for Monte Carlo simulations of diffusion processes, with a particular focus on its application to atmospheric turbulence modeling in the Lagrangian stochastic framework. The central idea of the method is to modify the drift term of the governing stochastic differential equations so that the invariant distribution of the process is tilted from a baseline density to a weighted density. By doing so, the method preferentially samples regions in state space that contribute most strongly to the statistics of interest. A resampling procedure is then applied to maintain unbiased estimates and to prevent weight degeneracy over long integrations.

The theoretical foundation of the method is established by deriving the modified SDE using the theory of h-transforms. Importance weights are tracked stepwise through sequential importance sampling, and resampling is triggered adaptively when the effective sample size falls below a threshold. A specific sampling weight is proposed for atmospheric turbulence modeling, scaling as $|w'|^p$ for large vertical velocities while remaining bounded when $|w'|$ is near zero. This form enhances the representation of rare, high-velocity events without suffering from weight degeneracy.

The method was first validated using one-dimensional diffusion processes with known invariant densities. Tests on the Ornstein–Uhlenbeck process, which corresponds to a normal distribution, and on a log-gamma process, which represents a skewed distribution, both demonstrated significant variance reduction across statistical estimates. The variance reduction was especially strong for higher-order moments and tail probabilities, which are typically the most difficult to capture with standard Monte Carlo approaches. The efficiency gains achieved by importance sampling were nearly independent of sample size, implying that the method reduces computational cost consistently across different resolutions.

The framework was then applied to simulations of a dry convective boundary layer using the Lagrangian Stochastic Subgrid Turbulence Model (LSSTM). Here, importance sampling again proved effective in reducing statistical fluctuations and in mitigating systematic biases that appear when particle counts are modest. The systematic biases are reduced because the statistical fluctuations in the coefficients of the SDEs are mitigated. Variance reductions were most pronounced for statistics whose observables scale strongly with vertical velocity, such as the vertical heat flux, where reductions of nearly 50% were achieved. Even greater improvements were seen in third-order moments and in the representation of tail probabilities, which are related to convective transport and the occurrence of convection, respectively. Furthermore, a modification of the resampling procedure improved the consistency of particle mass density with the prescribed base-state density, reducing structural biases in the mean fields.

Overall, the results demonstrate that the proposed importance sampling method offers statistical benefits, enhancing computational efficiency and mitigating systematic biases in

the model. It enhances the representation of extreme events, reduces variance in estimated statistics, and improves the efficiency of Monte Carlo simulations of turbulence. Although the specific sampling weight proposed here is not universally optimal, the framework offers flexibility to design weights tailored to different statistics or flow regimes. Nevertheless, the method has a limitation that its parameters must be tuned for specific applications. Moreover, LSSTM still exhibits noticeable statistical fluctuations even with importance sampling, leading to computational costs that remain significantly higher than those of conventional turbulence closure models. Future work may extend the approach to more complex scenarios, such as moist convection, and further optimize the sampling weights and parameters for boundary layer simulations. In addition, the development of complementary variance reduction techniques will help make PDF-based turbulence modeling more computationally affordable for full three-dimensional atmospheric models.

Acknowledgements This work was supported by a Research Grant of Pukyong National University (2023).

Author Contributions Jihoon Shin did the main analysis, and Jihoon Shin and Jong-Jin Baik wrote the main manuscript text. All authors reviewed the manuscript.

Data Availability The simulation data and analysis code are available from the authors upon request.

Declarations

Competing interests The authors declare no competing interests.

References

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59(1):65–98

Bogenschutz PA, Krueger SK (2013) A simplified PDF parameterization of subgrid-scale clouds and turbulence for cloud-resolving models. J Adv Model Earth Syst 5(2):195–211

Bogenschutz PA, Krueger SK, Khairoutdinov M (2010) Assumed probability density functions for shallow and deep convection. J Adv Model Earth Syst 2(4)

Boughton BA, Delaurentis JM, Dunn WE (1987) A stochastic model of particle dispersion in the atmosphere. Bound Layer Meteorol 40(1):147–163

Chinita MJ, Matheou G, Teixeira J (2018) A joint probability density-based decomposition of turbulence in the atmospheric boundary layer. Mon Weather Rev 146(2):503–523

Das SK, Durbin PA (2005) A Lagrangian stochastic model for dispersion in stratified turbulence. Phys Fluids 17(2):025109

Doucet A, De Freitas N, Gordon NJ et al (2001) Sequential Monte Carlo methods in practice. Springer, Berlin

Draxler RR, Hess G (1998) An overview of the HYSPLIT_4 modelling system for trajectories. Aust Meteorol Mag 47(4):295–308

Esler JG (2015) Adaptive stochastic trajectory modelling in the chaotic advection regime. J Fluid Mech 769:1–25

Fitch AC (2019) An improved double-Gaussian closure for the subgrid vertical velocity probability distribution function. J Atmos Sci 76(1):285–304

Golaz JC, Larson VE, Cotton WR (2002) A pdf-based model for boundary layer clouds. Part I: method and model description. J Atmos Sci 59(24):3540–3551

Grassberger P (2002) Go with the winners: a general Monte Carlo strategy. Comput Phys Commun 147(1–2):64–70

Guo H, Golaz JC, Donner LJ, Wyman B, Zhao M, Ginoux P (2015) CLUBB as a unified cloud parameterization: opportunities and challenges. Geophys Res Lett 42(11):4540–4547

Jenny P, Pope SB, Muradoglu M, Caughey DA (2001) A hybrid algorithm for the joint PDF equation of turbulent reactive flows. J Comput Phys 166(2):218–252

Jones A, Thomson D, Hort M, Devenish B (2007) The UK Met Office's next-generation atmospheric dispersion model, NAME III. In: Air pollution modeling and its application XVII. Springer, Berlin, pp 580–589

Kuptametee C, Aunsri N (2022) A review of resampling techniques in particle filtering framework. Measurement 193:110836

Minier JP, Pozorski J (1999) Wall-boundary conditions in probability density function methods and application to a turbulent channel flow. Phys Fluids 11(9):2632–2644

Øksendal B (2003) Stochastic differential equations: an introduction with applications. Springer, Berlin

Pisso I, Sollum E, Grythe H, Kristiansen NI, Cassiani M, Eckhardt S, Arnold D, Morton D, Thompson RL, Groot-Zwaaftink CD et al (2019) The Lagrangian particle dispersion model FLEXPART version 10.4. Geosci Model Dev 12(12):4955–4997

Pope SB (1983) A Lagrangian two-time probability density function equation for inhomogeneous turbulent flows. Phys Fluids 26(12):3448–3450

Pope SB (1995) Particle method for turbulent flows: integration of stochastic model equations. J Comput Phys 117(2):332–349

Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge

Rajaona H, Septier F, Armand P, Delignon Y, Olry C, Albergel A, Moussafir J (2015) An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release. Atmos Env 122:748–762

Rogers LCG, Williams D (2000) Diffusions, Markov processes, and martingales, vol 2. Cambridge University Press, Cambridge

Scire JS, Strimaitis DG, Yamartino RJ et al (2000) A user's guide for the CALPUFF dispersion model. Earth Tech Inc, Concord

Shin J, Baik JJ (2024) Lagrangian stochastic modeling of stratified atmospheric boundary layer. Bound Layer Meteorol 190(4):1–27

Shin J, Baik JJ (2025) Lagrangian stochastic modeling of unstable atmospheric surface layer. Bound Layer Meteorol 191:7

Stockie JM (2011) The mathematics of atmospheric dispersion modeling. SIAM Rev 53(2):349–372

Thomson DJ (1987) Criteria for the selection of stochastic models of particle trajectories in turbulent flows. J Fluid Mech 180:529–556

Van Slooten PR, Jayesh PSB (1998) Advances in PDF modeling for inhomogeneous turbulent flows. Phys Fluids 10(1):246–265

Wild MA (2013) General purpose PDF solution algorithm for reactive flow simulations in OpenFOAM. PhD Thesis, ETH Zurich