AECSF: ADAPTIVE ENSEMBLE CONDITIONAL SCORE FILTERING FOR HIGH-DIMENSIONAL NONLINEAR DATA ASSIMILATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Bayesian state estimation for high-dimensional nonlinear dynamical systems entails a fundamental tension between statistical fidelity and computational tractability, as particle weights can collapse, while Gaussian ensemble updates can miss non-Gaussian posterior structure. Score-based diffusion filters offer a sampling-based alternative, but existing training-free score filters often rely on heuristic likelihood corrections, which can compromise posterior accuracy by neglecting uncertainty about the system state associated with each noisy reverse particle. To address these issues, we propose AECSF, a training-free adaptive ensemble conditional score filter. AECSF constructs an analytically tractable score estimator from the conditional Tweedie identity, which recasts noisy posterior score estimation as estimating the conditional mean of the system state given a noisy reverse particle and the observation. To estimate these conditional means efficiently, AECSF employs a shared adaptive weighted proposal ensemble, while particle-specific conditional weights yield an estimate for each noisy reverse particle without separate proposal sampling. The proposal ensemble is updated using reverse-particle information within the same reverse-diffusion run to improve conditional-mean estimation. Theoretically, we characterize when a fixed weighted proposal measure yields the exact noisy posterior score. Under stated assumptions, we establish a bound relating conditional-mean estimation errors to reverse-sampling endpoint error. Numerical experiments demonstrate that AECSF improves the accuracy of posterior sampling and nonlinear filtering in high-dimensional problems with limited forecast ensembles.

1 INTRODUCTION

Reliable prediction of complex dynamical systems depends on accurate state estimation and uncertainty quantification. Data assimilation estimates the evolving system state by combining model forecasts with noisy observations (Reich & Cotter, 2015; Sanz-Alonso et al., 2023). In Bayesian filtering, each new observation updates the forecast distribution to a posterior distribution representing the remaining uncertainty about the system state. An ensemble provides a practical representation of these distributions and carries this uncertainty into subsequent forecasts. For high-dimensional nonlinear systems, however, a limited ensemble may poorly represent the posterior, while increasing its size raises the cost of propagating and updating its members. The challenge is therefore to improve posterior inference at a computational cost compatible with repeated assimilation.

Ensemble Kalman and particle methods address this challenge in different ways. Ensemble Kalman methods use sample covariances for efficient updates, but Gaussian approximations and linear regression can miss non-Gaussian posterior structure (Evensen, 1994; Hunt et al., 2007). Particle methods allow more general distributions, but importance weights can concentrate on few members (Liu & Chen, 1998; Doucet et al., 2001), especially in high dimensions (Snyder et al., 2008; Bengtsson et al., 2008). These limitations motivate alternative updates that can represent non-Gaussian posteriors at practical ensemble sizes.

Score-based generative methods offer another way to construct posterior ensembles. They generate samples through reverse diffusion, which progressively transforms random noise into samples

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

of a target distribution (Song et al., 2021). This transformation is driven by score functions, the gradients of log densities at successive noise levels. For Bayesian filtering, these scores must reflect both the forecast distribution and the new observation. Neural approaches learn score models for trajectories or filtering distributions (Rozet & Louppe, 2023; Bao et al., 2024b). The Ensemble Score Filter (EnSF) instead estimates the forecast score directly from the current forecast ensemble and adds an observation-dependent correction (Bao et al., 2024a). EnSF has been applied to surface quasigeostrophic dynamics (Bao et al., 2025), stochastic partial differential equations (Huynh et al., 2026), and data-driven dynamical forecasting (Tang et al., 2026). Although EnSF avoids neural score training, obtaining accurate observation-conditioned scores at each diffusion noise level remains challenging.

The observation likelihood is defined on the system state, whereas reverse diffusion requires the score of the noise-perturbed posterior. At a given diffusion noise level, a noisy reverse particle is generally compatible with multiple system states, which may yield different values of the observation likelihood. The likelihood entering the score correction therefore averages the original observation likelihood over the forecast conditional distribution of the system state given the reverse particle's current noisy state. EnSF uses a pointwise likelihood-gradient correction evaluated at the noisy reverse particle (Bao et al., 2024a), while diffusion posterior sampling approximates this conditional likelihood using a single denoised state estimate (Chung et al., 2023). Such pointwise approximations do not generally perform the required averaging and can introduce errors into the posterior score. The resulting challenge is to account for this conditional uncertainty while keeping score estimation tractable with a limited ensemble.

To address this challenge, we propose the Adaptive Ensemble Conditional Score Filter (AECSF), a training-free method for high-dimensional nonlinear data assimilation. Under Gaussian noising, the conditional Tweedie identity expresses the required score through the mean system state conditioned on the noisy reverse particle and observation (Boys et al., 2024; Peng et al., 2024; Patsenker et al., 2026). AECSF estimates these conditional means from a shared weighted proposal ensemble initialized from the forecast ensemble. Sharing avoids sampling separate proposal ensembles, while each reverse particle applies its own conditional weights. The conditional distributions change as reverse particles evolve, so a small, fixed proposal ensemble may provide inadequate support. AECSF therefore uses the reverse particles to guide proposal updates within a single reverse-diffusion run. The resulting conditional means give an analytically tractable score estimator through the Tweedie identity. The main contributions are:

- We propose AECSF, a training-free filter that constructs an analytically tractable score estimator using the conditional Tweedie identity. AECSF estimates particle-specific conditional means from a shared weighted proposal ensemble, avoiding separate proposal sampling for each reverse particle. It uses the evolving reverse particles to guide proposal updates within the same reverse-diffusion run.
- We characterize when a fixed weighted proposal measure yields the exact noisy posterior score. We establish consistency of adaptive conditional mean estimation and reverse-sampling endpoints relative to a numerical reference, and establish a bound relating conditional-mean estimation errors to reverse-sampling endpoint error.
- AECSF is evaluated on controlled Gaussian sampling problems and high-dimensional nonlinear filtering benchmarks, including 10,000-dimensional Lorenz–96 with 20 members and a 1,024-dimensional Kuramoto–Sivashinsky system. The results demonstrate improved posterior-sampling accuracy over the tested guidance rules, as well as competitive filtering accuracy–cost trade-offs with limited ensembles.

2 Related Work

Nonlinear ensemble filtering. Methods for nonlinear ensemble filtering use tempering, localization, adapted proposals, resampling, and ensemble transforms to mitigate weight degeneracy or sampling error at practical ensemble sizes (Del Moral et al., 2006; Rebeschini & van Handel, 2015; Farchi & Bocquet, 2018; Poterjoy, 2016; Frei & Künsch, 2013; Reich, 2013). These strategies typically modify the ensemble used to approximate the filtering distribution or intermediate assimilation targets. AECSF applies tempering, resampling, and short moves to a shared proposal ensemble used to estimate the conditional score at each reverse-diffusion step.

Score-based and flow-based data assimilation. Ensemble-based generative filters differ in observation incorporation, state representation, and computational cost. The United Filter combines EnSF for state estimation with a direct filter for parameter estimation (Bao et al., 2026) and has been applied to reduced fracture models (Huynh et al., 2025). The Iterative Ensemble Score Filter (IEnSF) approximates an expected likelihood gradient at a conditional mean from a Gaussian reference posterior within a Gaussian-mixture forecast approximation (Zhang et al., 2027). It alternates complete reverse-diffusion runs with ensemble refitting; AECSF adapts proposals within one run. Latent-EnSF learns a shared state–observation space (Si & Chen, 2025), while LD-EnSF learns latent dynamics to reduce forecast cost (Xiao et al., 2026). The Ensemble Flow Filter (EnFF) constructs an observation-conditioned flow using Monte Carlo likelihood weighting or guidance from a local likelihood linearization (Transue et al., 2026).

Conditional diffusion and posterior sampling. Conditional diffusion methods incorporate observations through pointwise guidance (Chung et al., 2023) or measurement-conditioned denoising moments derived from Tweedie identities (Boys et al., 2024; Peng et al., 2024; Patsenker et al., 2026). Monte Carlo and sequential Monte Carlo methods support training-free sampling along diffusion paths (Gleich & Schmidler, 2026; Young et al., 2026). Other sample-based methods construct stochastic maps or conditional scores from state and measurement pairs (Liu et al., 2025; Zhang et al., 2025). Conditional generative models also support bifidelity Bayesian parameter estimation (Tatsuoka et al., 2026). Binder et al. obtain an analytically tractable conditional score from joint kernel density estimation of forecast states and simulated observations (Binder et al., 2026), using fixed forecast-state centers and weights that depend on the observation and noisy state. AECSF instead uses the observation likelihood directly and updates proposal locations during reverse diffusion.

3 Filtering Problem and Conditional-Score Formulation

Let $x_n \in \mathbb{R}^d$ denote the system state at assimilation time n, $y_n \in \mathbb{R}^{m_n}$ the current observation, p_n the transition kernel, and g_n the observation likelihood:

$$x_n \sim p_n(\cdot \,|\, x_{n-1}), \qquad y_n \sim g_n(\cdot \,|\, x_n). \tag{1}$$

The forecast density π_n^f conditions on $y_{1:n-1}$, whereas the analysis density π_n^a additionally conditions on y_n:

$$\begin{aligned} \pi_n^f(x_n) &= \int p_n(x_n \,|\, x_{n-1}) \pi_{n-1}^a(x_{n-1}) \, \mathrm{d}x_{n-1}, \\ \pi_n^a(x_n) &\propto g_n(y_n \,|\, x_n) \pi_n^f(x_n). \end{aligned} \tag{2}$$

The analysis ensemble represents the updated state uncertainty and is propagated to obtain the next forecast ensemble.

Fix one assimilation time and let Z denote the system state before artificial noise is added. For diffusion time $\tau \in [0, 1]$, introduce

$$\begin{aligned} Z \sim \pi_n^f, \qquad Y_n \,|\, Z \sim g_n(\cdot \,|\, Z), \\ X_\tau \,|\, Z \sim \mathcal{N}(a_\tau Z, \sigma_\tau^2 \mathbf{I}_d). \end{aligned} \tag{3}$$

The artificial noise and observation are conditionally independent given Z. The coefficients $a_\tau > 0$ and $\sigma_\tau^2 > 0$ specify the retained signal and noise variance at the score-evaluation levels. Conditional on $Y_n = y_n$, the system state has density π_n^a, and X_τ has the Gaussian-noised posterior density $q_{n,\tau}^a(\cdot \,|\, y_n)$. Reverse diffusion proceeds from high to low noise using

$$s_{n,\tau}^a(x; y_n) = \nabla_x \log q_{n,\tau}^a(x \,|\, y_n). \tag{4}$$

Define the mean system state conditioned on both the noisy state and the observation:

$$D_{n,\tau}(x; y_n) = \mathbb{E}[Z \,|\, X_\tau = x, Y_n = y_n]. \tag{5}$$

The conditional Tweedie identity gives

$$s_{n,\tau}^a(x; y_n) = -\frac{x}{\sigma_\tau^2} + \frac{a_\tau}{\sigma_\tau^2} D_{n,\tau}(x; y_n). \tag{6}$$

Appendix B.1 derives this identity. It reduces score construction to conditional-mean estimation at each noisy reverse state.

4 THE ADAPTIVE ENSEMBLE CONDITIONAL SCORE FILTER

AECSF takes a forecast ensemble $\mathcal{X}_n^f = \{x_n^{f,j}\}_{j=1}^J$, observation y_n, and an evaluable, differentiable likelihood. It maintains J reverse particles and a shared proposal ensemble of M weighted system-state samples. Each reverse particle forms its own conditional weights from these proposals. The reported configuration uses $M = J$, initialized at the forecast members.

The reverse particles guide proposal adaptation; the proposals provide the scores that advance those particles toward the analysis ensemble. Use a decreasing grid $1 = \tau_{N_\tau} > \cdots > \tau_0 = 0$, with $x_i^{(j)} = x_{\tau_i}^{(j)}$.

4.1 SHARED-ENSEMBLE CONDITIONAL-SCORE ESTIMATION

For a reverse particle at x, the conditional mean in (5) is taken over

$$p(z \mid X_\tau = x, Y_n = y_n) \propto \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z)\pi_n^f(z)g_n(y_n \mid z). \tag{7}$$

Here, $\phi_{\sigma_\tau^2 \mathbf{I}_d}$ is the centered Gaussian density with covariance $\sigma_\tau^2 \mathbf{I}_d$. The three factors favor system states compatible with the noisy state, plausible under the forecast, and consistent with the observation.

To support different queries with one proposal ensemble, we introduce a tempered reference distribution independent of the individual noisy query:

$$\pi_{n,\rho}(z) \propto \pi_n^f(z)g_n(y_n \mid z)^\rho, \qquad 0 \le \rho \le 1. \tag{8}$$

It interpolates between the forecast and posterior and gives the factorization

$$p(z \mid X_\tau = x, Y_n = y_n) \propto \pi_{n,\rho}(z)\phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z)g_n(y_n \mid z)^{1-\rho}. \tag{9}$$

The full likelihood is retained through $g_n^\rho g_n^{1-\rho} = g_n$. Increasing ρ places more observation information in the shared reference and less in the remaining particle-specific factor. The reported implementation uses $\rho_i = 1 - \tau_i$.

At step i, proposal locations z_i^m and normalized nonnegative integration weights $W_{i,m}$ define

$$\nu_{n,i}^M = \sum_{m=1}^M W_{i,m}\delta_{z_i^m}. \tag{10}$$

Using this weighted measure to approximate π_{n,ρ_i}, AECSF computes

$$\begin{aligned} F_{i,j}(z) &= \phi_{\sigma_{\tau_i}^2 \mathbf{I}_d}(x_i^{(j)} - a_{\tau_i} z)g_n(y_n \mid z)^{1-\rho_i}, \\ c_{j,m} &= \frac{W_{i,m}F_{i,j}(z_i^m)}{\sum_{\ell=1}^M W_{i,\ell}F_{i,j}(z_i^\ell)}, \\ \widehat{D}_{n,i}^M(x_i^{(j)}; y_n) &= \sum_{m=1}^M c_{j,m} z_i^m. \end{aligned} \tag{11}$$

The conditional weights $c_{j,m}$, with the current step index suppressed, are normalized separately for each reverse particle. Substitution into (6) gives

$$\widehat{s}_{n,\tau_i}^a(x_i^{(j)}; y_n) = -\frac{x_i^{(j)}}{\sigma_{\tau_i}^2} + \frac{a_{\tau_i}}{\sigma_{\tau_i}^2}\widehat{D}_{n,i}^M(x_i^{(j)}; y_n). \tag{12}$$

Weights are evaluated by log-sum-exp normalization.

4.2 REVERSE-PARTICLE-GUIDED PROPOSAL ADAPTATION

Proposal adaptation follows the changing tempered reference and allocates proposals according to the current reverse particles. Let (z_m^-, W_m^-) denote the incoming proposals and weights, with preceding tempering level ρ^-. Incremental weights advance at every adaptation stage, while resampling and movement follow separate schedules.

Incremental target weights. The change from ρ^- to ρ_i gives

$$\omega_{i,m} = \frac{W_m^- g_n(y_n \mid z_m^-)^{\rho_i - \rho^-}}{\sum_\ell W_\ell^- g_n(y_n \mid z_\ell^-)^{\rho_i - \rho^-}}. \tag{13}$$

Since diffusion time decreases, $\rho_i - \rho^- \geq 0$. This increment updates the reference weights before locations are selected or moved.

Guided ancestor selection. Selection based only on ω_i would ignore where conditional estimates are required. Define $K_i(x, z) = \phi_{\sigma^2_{\tau_i} \mathbf{I}_d}(x - a_{\tau_i} z)$ and form

$$r_{i,jm} = \frac{K_i(x_i^{(j)}, z_m^-)\omega_{i,m}}{\sum_\ell K_i(x_i^{(j)}, z_\ell^-)\omega_{i,\ell}}, \qquad s_{i,m} = \frac{1}{J}\sum_{j=1}^{J} r_{i,jm}. \tag{14}$$

Each $r_{i,j}$ allocates ancestors compatible with one reverse particle; s_i averages these allocations across particles. Unlike the conditional weights $c_{j,m}$, these probabilities omit the remaining likelihood factor in (11).

Importance correction. When resampling is scheduled, stratified resampling from s_i gives ancestor indices $A_1, \ldots, A_M$. Because s_i differs from the reference weights ω_i, the descendants receive corrected weights:

$$z_m^+ = z_{A_m}^-, \qquad v_m = \frac{\omega_{i,A_m}}{s_{i,A_m}}, \qquad W_m^+ = \frac{v_m}{\sum_\ell v_\ell}. \tag{15}$$

If $s_{i,m} > 0$ wherever $\omega_{i,m} > 0$, the correction preserves the conditional expectation of the unnormalized weighted sum before movement; normalization can introduce finite-sample bias. Appendix B.3 gives the identity. Without resampling, set $z_m^+ = z_m^-$ and $W_m^+ = \omega_{i,m}$.

Langevin movement. Resampling reallocates existing locations, whereas movement explores new ones. The moves use $\widehat{s}_n^f = \nabla \log \widetilde{\pi}_n^f$, where $\widetilde{\pi}_n^f$ is a fitted forecast-density approximation. In the recursive experiments, it is a diagonal Gaussian with regularized componentwise variances, fixed throughout one analysis update. Starting at $z^{m,0} = z_m^+$, the core unadjusted Langevin step is

$$\begin{aligned} z^{m,\ell+1} = z^{m,\ell} + \frac{h}{2}\left[w_f \widehat{s}_n^f(z^{m,\ell}) + w_g \rho_i \nabla_z \log g_n(y_n \mid z^{m,\ell})\right] \\ + \sqrt{h}\,\xi^{m,\ell}, \qquad \xi^{m,\ell} \sim \mathcal{N}(0, \mathbf{I}_d). \end{aligned} \tag{16}$$

Here h is the step size; the reported configurations use $w_f = w_g = 1$. The moves act on system-state proposals, while (12) determines the noisy reverse-particle update. Moved proposals retain W_m^+, and their likelihood values are refreshed before score estimation. If movement is omitted, locations are retained. The resulting $(z_i^m, W_{i,m})$ define the shared measure in (10).

4.3 Reverse sampling and recursive filtering

Let f_τ and r_τ be the drift coefficient and diffusion amplitude of the Gaussian noising process, defined in Appendix B.5. With $\Delta_i = \tau_i - \tau_{i-1}$, the Euler–Maruyama reverse update is

$$\begin{aligned} x_{i-1}^{(j)} = x_i^{(j)} - \Delta_i\left[f_{\tau_i} x_i^{(j)} - r_{\tau_i}^2 \widehat{s}_{n,\tau_i}^a(x_i^{(j)}; y_n)\right] \\ + \sqrt{\Delta_i r_{\tau_i}^2}\,\xi_i^{(j)}, \qquad \xi_i^{(j)} \sim \mathcal{N}(0, \mathbf{I}_d). \end{aligned} \tag{17}$$

Algorithm 1 summarizes the analysis update. Adaptation begins after the first score evaluation. The final reverse particles form $\mathcal{X}_n^a$ and are propagated to the next forecast. The proposal ensemble and fitted forecast density are rebuilt at each assimilation time.

The numerical schedule retains positive endpoint noise and uses a grid uniform in logSNR; Appendix B.5 gives initialization and schedule details. Table 4 lists adaptation settings. Gaussian-kernel and weighted-mean evaluation costs $O(JMN_\tau d)$, implemented through GPU matrix products; likelihoods and proposal moves add model-dependent costs. Figure 3 and Table 8 report costs for increasing $J = M$ and increasing M at fixed J, respectively. Further implementation details appear in Appendix B.6.

Algorithm 1 One AECSF analysis update (reported $M = J$ configuration)

Require: Forecast ensemble $\mathcal{X}_n^f$, observation y_n, likelihood g_n, fitted score $\widehat{s}_n^f$, initial reverse particles, diffusion grid, resampling and move schedules

Initialize proposals at forecast members; set $W_m = 1/M$, $\rho^- = 0$, and cache likelihoods
for $i = N_\tau, \dots, 1$ **do**
  Set $\rho_i = 1 - \tau_i$
  if $i < N_\tau$ **then**
    Compute ω_i by (13); set $W \leftarrow \omega_i$
    if resampling is scheduled **then**
      Draw stratified ancestors from s_i and correct weights by (15)
    end if
    if movement is scheduled **then**
      Apply (16); retain weights and refresh likelihoods
    end if
    Set $\rho^- \leftarrow \rho_i$
  end if
  Compute conditional means and scores by (11)–(12)
  Advance reverse particles by (17)
end for
return $\mathcal{X}_n^a = \{x_0^{(j)}\}_{j=1}^J$

4.4 Exactness and approximation error

For a realized proposal ensemble, define

$$\widehat{\eta}_{n,i}^M(\mathrm{d}z) = \frac{g_n(y_n \mid z)^{1-\rho_i} \nu_{n,i}^M(\mathrm{d}z)}{\nu_{n,i}^M(g_n(y_n \mid \cdot)^{1-\rho_i})}. \tag{18}$$

Holding proposal locations and weights fixed when differentiating in x, the estimator in (12) is the score of the explicit Gaussian mixture

$$\begin{aligned} \widehat{q}_{n,i}^M(x; y_n) &= \int \phi_{\sigma_{\tau_i}^2 \mathbf{I}_d}(x - a_{\tau_i} z) \widehat{\eta}_{n,i}^M(\mathrm{d}z), \\ \widehat{s}_{n,\tau_i}^a(x; y_n) &= \nabla_x \log \widehat{q}_{n,i}^M(x; y_n). \end{aligned} \tag{19}$$

For $a_{\tau_i} \neq 0$ and $\sigma_{\tau_i}^2 > 0$, a fixed proposal measure in this construction yields the target score everywhere if and only if it equals π_{n,ρ_i}, under the positivity and integrability conditions in Appendix C.2.

Fix (τ, x, y_n, ρ), set $\pi = \pi_{n,\rho}$ and $F(z) = \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z) g_n(y_n \mid z)^{1-\rho}$. For a probability measure ν with positive finite denominator and finite weighted first moment, let $D_\nu = \nu(Fz)/\nu(F)$. Then

$$D_\nu - D_{n,\tau}(x; y_n) = \frac{(\nu - \pi)(F[\, z - D_{n,\tau}(x; y_n)\,])}{\nu(F)}, \tag{20}$$

and the associated score satisfies

$$\widehat{s}_{\nu,n,\tau}^a(x; y_n) - s_{n,\tau}^a(x; y_n) = \frac{a_\tau}{\sigma_\tau^2}[D_\nu - D_{n,\tau}(x; y_n)]. \tag{21}$$

These relations identify the weighted moments through which proposal approximation affects conditional-mean and score estimation. Appendix B.3 separates propagated proposal error from the move kernel's invariance defect, and Appendix C.3 controls the normalized moment ratio.

The reverse particles guide proposal updates, while their evolution depends on scores estimated from the same shared proposals. The adaptive analysis accounts for this dependence relative to a numerical proposal reference generated by the initial empirical measure, incremental likelihood weights, and proposal transition kernels. Appendix C.4 establishes consistency of conditional-mean estimation and raw-score reverse-sampling endpoints in probability on a fixed grid, including unbounded proposal states. The comparison density uses the fitted forecast $\widetilde{\pi}_n^f$; Section 3 defines the original filtering target using π_n^f.

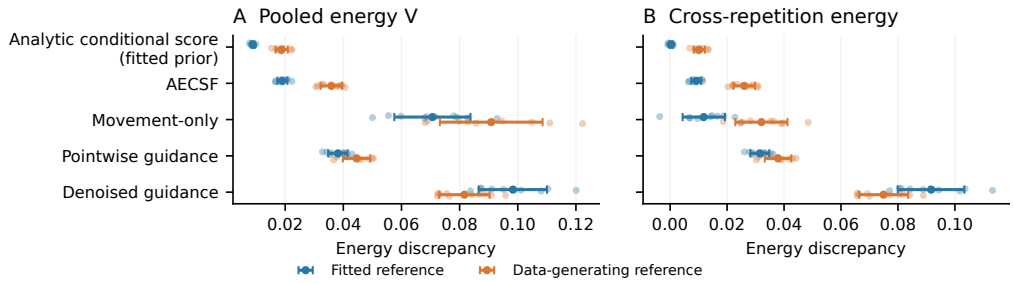


Figure 1: Gaussian endpoint energy discrepancies ($M = J = 20$). A pools four repetitions (80 particles/problem); B excludes same-run particle pairs and reference self-pairs. Colors denote fitted and data-generating retained-noise references. Large points/bars show means/sample SDs across ten problems; small points show individual problems. Definitions: Appendix D.1.

The bound separates propagated initialization error, accumulated score error relative to the numerical reference, and reference-output discrepancy from the original filtering posterior. The latter includes the effects of forecast approximation, proposal-kernel invariance defects, reverse-time discretization, and retained endpoint noise.

5 Experiments

The experiments examine whether AECSF improves posterior sampling, which proposal updates contribute to its performance, and how its accuracy depends on computational budget. Controlled Gaussian problems provide analytic posterior references for comparing sampling endpoints under matched forecast information. Lorenz–96 tests whether proposal adaptation improves recursive filtering with only 20 members in 10,000 dimensions, using component controls at matched Langevin-step budgets and direct posterior ULA at approximately matched analysis time. Kuramoto–Sivashinsky compares AECSF with generative filtering baselines under nonlinear observations and examines how sampling steps and ensemble size affect accuracy and analysis cost.

We evaluate Gaussian sampling endpoints using energy discrepancies and recursive filtering using RMSE, componentwise continuous ranked probability score (CRPS), spread–skill ratio (SSR), and analysis time. Lorenz–96 additionally reports normalized observation misfit. Methods share forecast information in the Gaussian comparisons and truth, observations, and initial ensembles in the recursive comparisons. Appendix D provides metric definitions, configurations, seeds, and aggregation procedures.

5.1 Controlled reverse-sampling endpoints

The Gaussian problem has 128 fully observed coordinates, with $M = J = 20$. All practical methods construct prior approximations from the same forecast ensemble; AECSF and movement-only share the diffusion schedule and 294 Langevin steps. We use ten observation problems with four repetitions each. Both fitted and data-generating posterior references retain endpoint diffusion noise. The analytic-score control uses the fitted Gaussian posterior and the same reverse solver.

AECSF has lower mean pooled and cross-repetition energy discrepancies than movement-only and both guidance rules under either reference (Figure 1). Individual 20-member outputs also improve mean and diagonal-variance errors over movement-only under both references. Full numerical results and per-run timings are reported in Appendix D.1.

5.2 Lorenz–96: nonlinear observations at $d = 10{,}000$

The Lorenz–96 benchmark uses 20 members in 10,000 dimensions with full arctangent observations (noise SD 0.05). We evaluate cycles 50–199 of 200. Appendix D.4.1 records configurations and their selection.

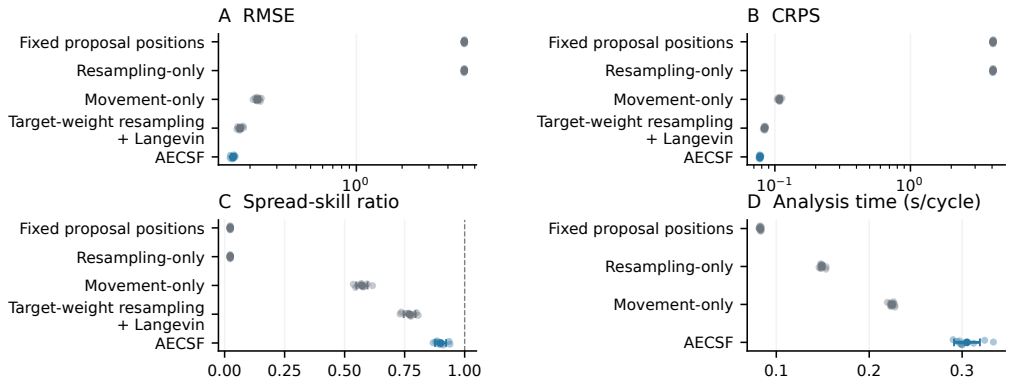


Figure 2: Lorenz–96 proposal-update ablations ($d = 10{,}000$, $M = 20$). Langevin variants share the reverse schedule and 294 steps. A–C summarize cycles 50–199; D gives timings for four variants. Large points/bars show means/sample SDs across ten seeds; small points show paired seed results. Fixed proposal positions retains weight updates; target-weight resampling uses ω.

Table 1: Lorenz–96 results ($d = 10{,}000$, 20 members; $J = M = 20$ for AECSF). Entries are means $\pm$ sample SD across ten seeds, each summarized over cycles 50–199. Timings use serial runs on the same host with GPU analysis for all methods.

Method	RMSE ↓	CRPS ↓	SSR → 1	Norm. obs. misfit	Analysis time/cycle (s)
EnSF	0.2418 ± 0.0006	0.1432 ± 0.0002	1.7366 ± 0.0042	1.8180 ± 0.0028	0.0485 ± 0.0010
LETKF	0.1263 ± 0.0116	0.0578 ± 0.0015	0.6119 ± 0.0604	1.0393 ± 0.0349	9.5530 ± 0.1153
Binder et al.	5.1323 ± 0.0082	4.1428 ± 0.0076	$(24.66 \pm 13.29) \times 10^{-5}$	28.4641 ± 0.0390	0.1507 ± 0.0015
AECSF	0.1545 ± 0.0039	0.0776 ± 0.0007	0.8998 ± 0.0232	0.9209 ± 0.0042	0.3052 ± 0.0139

LETKF has the lowest mean RMSE and CRPS; AECSF improves both over EnSF and has SSR closest to one. Binder shows strong ensemble contraction in all ten runs. Appendix D.4.1 documents its original-protocol check and transfer settings.

Proposal-update ablations. Movement-only tests the addition of corrected resampling; target-weight resampling tests ancestor selection guided by reverse particles with its importance correction. Both controls share AECSF's prior approximation, diffusion schedule, and 294 Langevin steps. Corrected shared resampling reduces mean RMSE by 31.0% and CRPS by 28.1%, with 36.0% more analysis time. Against target-weight resampling at the same stages and Langevin budget, AECSF reduces RMSE by 10.2% and CRPS by 7.5%. Both improvements hold in all ten paired seeds for each control. The two variants without movement have mean RMSE near 5.12, supporting proposal movement alongside corrected, reverse-particle-guided ancestor selection.

Against direct posterior ULA at the same step count and step size, AECSF reduces mean RMSE by 16.2% and CRPS by 1.1%, at 2.23 times the sampling cost (Table 5). At approximately matched analysis time, the reductions are 9.9% and 17.3%, respectively, relative to the development-selected ULA configuration; both metrics improve in all ten runs (Table 6).

5.3 Kuramoto–Sivashinsky: nonlinear observations and ensemble scaling

We compare AECSF with EnSF, EnFF-OT, EnFF-F2P, and IEnSF (Transue et al., 2026; Zhang et al., 2027). The 1,024-dimensional Kuramoto–Sivashinsky system has full arctangent observations with noise SD 0.1. We vary sampling steps with 20 members and ensemble size at 20 steps, using five paired seeds. Appendix D.5 gives the protocol and method settings.

At comparable analysis times near 0.46 seconds, AECSF with 100 steps has lower RMSE and CRPS than IEnSF-L1 with 20 steps (0.0601/0.0431 versus 0.0740/0.0665). Near 2.3 seconds, IEnSF-L1 with 100 steps outperforms AECSF with 500 steps (0.0498/0.0345 versus 0.0579/0.0405; Figure 3). AECSF with 20 steps also has lower RMSE than EnFF-OT with 100 steps (0.0889 versus 0.1237),

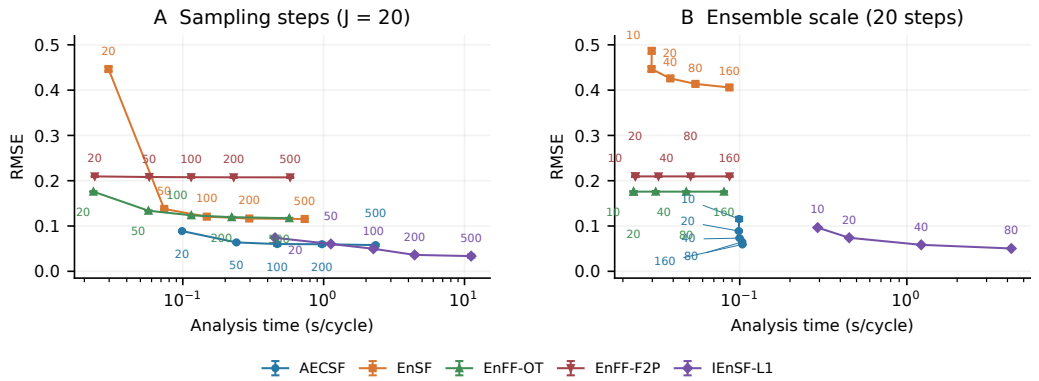


Figure 3: Kuramoto–Sivashinsky accuracy versus analysis time (logarithmic horizontal axes). A: sampling steps at $J = 20$; labels give T or N_t for IEnSF-L1. B: ensemble size at 20 steps; labels give J. AECSF varies forecast, reverse-particle and proposal counts together ($J = M$). IEnSF-L1 extends to $J = 80$ and the other methods to $J = 160$. RMSE uses cycles 950–999 and timing all 1,000 cycles. Points/bars show means/sample SDs over five paired seeds; lines follow increasing steps or ensemble size.

at 0.099 versus 0.115 seconds. Both AECSF and IEnSF-L1 gain less accuracy from additional steps at the largest budgets.

At 20 steps, increasing AECSF's ensemble from 10 to 160 members reduces mean RMSE from 0.1155 to 0.0593, while analysis time stays between 0.099 and 0.104 seconds. IEnSF-L1 improves from 0.0965 to 0.0504 RMSE as its ensemble grows from 10 to 80 members, with time increasing from 0.293 to 4.218 seconds. At each shared ensemble size, IEnSF-L1 has lower RMSE but higher CRPS and cost than AECSF; the EnFF variants improve little in RMSE. AECSF's RMS spread increases slightly as its RMS error falls, raising SSR from 1.60 to 3.33; IEnSF-L1's SSR rises from 2.32 to 4.66.

6 Discussion and Conclusion

AECSF addresses limited-ensemble posterior sampling by adapting the shared proposals used for conditional-score estimation within a single reverse-diffusion run. Each reverse particle uses these proposals to estimate the system-state mean conditioned on its current noisy state and the observation. The analysis connects conditional-mean estimation errors to reverse-sampling endpoint error relative to a numerical reference.

Lorenz–96 ablations support combining proposal movement with corrected ancestor selection guided by reverse particles. AECSF improves on direct posterior ULA on Lorenz–96 and EnFF-OT on Kuramoto–Sivashinsky at comparable analysis costs. The IEnSF-L1 comparison shows that relative accuracy changes with the available budget. Increasing ensemble size gives AECSF a different trade-off: lower RMSE with little additional measured cost, but a larger spread–skill ratio. These results favor considering ensemble size and sampling steps jointly when allocating computation.

Posterior accuracy depends jointly on the forecast approximation and the sampling procedure. The theoretical and empirical findings support AECSF as a practical training-free approach to high-dimensional nonlinear filtering.

AI USE STATEMENT

In this work, we used generative AI tools to assist with implementation and experimental design, mathematical derivations and arguments, and drafting and polishing portions of the manuscript. The authors reviewed all AI-assisted content, independently verified the mathematical arguments and references, and tested the implementation using analytical cases and controlled experiments. The authors take full responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

The main text presents the estimator and principal findings; the appendices provide the derivations, algorithms, model and observation equations, evaluation metrics, configuration-selection history and evaluation seeds, hyperparameters, and supplementary diagnostics. All experiments use simulated models and require no external datasets. The code is available upon request, and it will be made publicly available after acceptance.

REFERENCES

Sergios Agapiou, Omiros Papaspiliopoulos, Daniel Sanz-Alonso, and Andrew M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3):405–431, 2017. doi: 10.1214/17-STS611.

Feng Bao, Zezhong Zhang, and Guannan Zhang. An ensemble score filter for tracking high-dimensional nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 432:117447, 2024a. doi: 10.1016/j.cma.2024.117447.

Feng Bao, Zezhong Zhang, and Guannan Zhang. A score-based filter for nonlinear data assimilation. *Journal of Computational Physics*, 514:113207, 2024b. doi: 10.1016/j.jcp.2024.113207.

Feng Bao, Hristo G. Chipilski, Siming Liang, Guannan Zhang, and Jeffrey S. Whitaker. Nonlinear ensemble filtering with diffusion models: Application to the surface quasigeostrophic dynamics. *Monthly Weather Review*, 153(7):1155–1169, 2025. doi: 10.1175/MWR-D-24-0069.1.

Feng Bao, Zezhong Zhang, and Guannan Zhang. United filter for jointly estimating state and parameters of stochastic dynamical systems. *Communications in Computational Physics*, 39(3): 747–774, 2026. doi: 10.4208/cicp.OA-2024-0009.

Thomas Bengtsson, Peter Bickel, and Bo Li. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Institute of Mathematical Statistics Collections*, 2:316–334, 2008. doi: 10.1214/193940307000000518.

Alexandros Beskos, Dan Crisan, and Ajay Jasra. On the stability of sequential monte carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445, 2014. doi: 10.1214/13-AAP951.

Brianna Binder, Agnimitra Dasgupta, and Assad Oberai. Closed-form conditional diffusion models for data assimilation, 2026. URL https://arxiv.org/abs/2603.21291.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O. Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=4unJi0qrTE.

Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018. doi: 10.1214/17-AAP1326.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. International Conference on Learning Representations, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006. doi: 10.1111/j.1467-9868.2006.00553.x.

Arnaud Doucet, Nando de Freitas, and Neil Gordon (eds.). *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York, 2001. ISBN 978-0-387-95146-1. doi: 10.1007/978-1-4757-3437-9.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Geir Evensen. 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. doi: 10.1029/94JC00572.

Alban Farchi and Marc Bocquet. Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes in Geophysics*, 25(4):765–807, 2018. doi: 10.5194/npg-25-765-2018.

Marco Frei and Hans R. Künsch. Bridging the ensemble kalman and particle filters. *Biometrika*, 100(4):781–800, 2013. doi: 10.1093/biomet/ast020.

Aidan Gleich and Scott C. Schmidler. Multilevel and sequential monte carlo for training-free diffusion guidance, 2026. URL https://arxiv.org/abs/2601.21104.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Brian R. Hunt, Eric J. Kostelich, and Istvan Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, 230(1–2): 112–126, 2007. doi: 10.1016/j.physd.2006.11.008.

Phuoc-Toan Huynh, Thi-Thao-Phuong Hoang, Guannan Zhang, and Feng Bao. Joint state-parameter estimation for the reduced fracture model via the united filter. *Journal of Computational Physics*, 538:114159, 2025. doi: 10.1016/j.jcp.2025.114159.

Phuoc-Toan Huynh, Ruth Y. López Fajardo, Guannan Zhang, Lili Ju, and Feng Bao. A score-based diffusion model approach for adaptive learning of stochastic partial differential equation solutions. *Journal of Computational Physics*, 556:114814, 2026. doi: 10.1016/j.jcp.2026.114814.

Jun S. Liu and Rong Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998. doi: 10.1080/01621459.1998.10473765.

Yanfang Liu, Yuan Chen, Dongbin Xiu, and Guannan Zhang. A training-free conditional diffusion model for learning stochastic dynamical systems. *SIAM Journal on Scientific Computing*, 47(5): C1144–C1171, 2025. doi: 10.1137/24M1699589.

Edward N. Lorenz. Predictability: A problem partly solved. In *Seminar on Predictability*, volume 1, pp. 1–18, Reading, United Kingdom, 1996. European Centre for Medium-Range Weather Forecasts. URL https://www.ecmwf.int/en/elibrary/75462-predictability-problem-partly-solved.

Radford M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001. doi: 10.1023/A:1008923215028.

Jonathan Patsenker, Henry Li, Myeongseob Ko, Ruoxi Jia, and Yuval Kluger. Injecting measurement information yields a fast and noise-robust diffusion-based inverse problem solver. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 2323–2331. PMLR, 2026. URL https://proceedings.mlr.press/v300/patsenker26a.html.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40347–40370. PMLR, 2024. URL https://proceedings.mlr.press/v235/peng24h.html.

Jonathan Poterjoy. A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144(1):59–76, 2016. doi: 10.1175/MWR-D-15-0163.1.

Patrick Rebeschini and Ramon van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015. doi: 10.1214/14-AAP1061.

Sebastian Reich. A nonparametric ensemble transform method for bayesian inference. *SIAM Journal on Scientific Computing*, 35(4):A2013–A2024, 2013. doi: 10.1137/130907367.

Sebastian Reich and Colin Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, 2015. doi: 10.1017/CBO9781107706804.

François Rozet and Gilles Louppe. Score-based data assimilation. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-1763. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/7f7fa581cc8a1970a4332920cdf87395-Abstract-Conference.html.

Daniel Sanz-Alonso, Andrew M. Stuart, and Armeen Taeb. *Inverse Problems and Data Assimilation*. Cambridge University Press, 2023. doi: 10.1017/9781009414319.

Phillip Si and Peng Chen. Latent-EnSF: A latent ensemble score filter for high-dimensional data assimilation with sparse observation data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=urcEYsZOBz.

G. I. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames–i. derivation of basic equations. *Acta Astronautica*, 4(11–12):1177–1206, 1977. doi: 10.1016/0094-5765(77)90096-0.

Chris Snyder, Thomas Bengtsson, Peter Bickel, and Jeff Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008. doi: 10.1175/2008MWR2529.1.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. International Conference on Learning Representations, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Jingqiao Tang, Ryan Bausback, Feng Bao, Guannan Zhang, and Phuoc-Toan Huynh. A score-filter-enhanced data assimilation framework for data-driven dynamical systems. *Numerical Methods for Partial Differential Equations*, 42(5):e70138, 2026. doi: 10.1002/num.70138.

Caroline Tatsuoka, Minglei Yang, Dongbin Xiu, and Guannan Zhang. Bifidelity parameter estimation using conditional diffusion models. *SIAM/ASA Journal on Uncertainty Quantification*, 14(3):743–778, 2026. doi: 10.1137/25M1734567.

Taos Transue, Bohan Chen, So Takao, and Bao Wang. Flow matching for efficient and scalable data assimilation. *SIAM/ASA Journal on Uncertainty Quantification*, 2026. URL https://arxiv.org/abs/2508.13313. Accepted for publication.

Pengpeng Xiao, Phillip Si, and Peng Chen. LD-EnSF: Synergizing latent dynamics with ensemble score filters for fast data assimilation with sparse observations. International Conference on Learning Representations, 2026. URL https://openreview.net/forum?id=AWSVzzhbr7.

James Matthew Young, Paula Cordero-Encinar, Sebastian Reich, Andrew Duncan, and O. Deniz Akyildiz. Diffusion path samplers via sequential monte carlo, 2026. URL https://arxiv.org/abs/2601.21951.

Zezhong Zhang, Caroline Tatsuoka, Dongbin Xiu, and Guannan Zhang. Exact conditional score-guided generative modeling for amortized inference in uncertainty quantification. *SIAM Journal on Scientific Computing*, 2025. URL https://arxiv.org/abs/2506.18227. Accepted for publication.

Zezhong Zhang, Feng Bao, and Guannan Zhang. IEnSF: Iterative ensemble score filter for reducing error in posterior score estimation in nonlinear data assimilation. *Journal of Computational Physics*, 568:115357, 2027. doi: 10.1016/j.jcp.2026.115357.

A FILTERING MODEL AND NOTATION

We distinguish assimilation time n from diffusion time τ. The system state $x_n \in \mathbb{R}^d$ evolves according to the transition kernel p_n, and the observation $y_n \in \mathbb{R}^{m_n}$ has likelihood $g_n(y_n \mid x_n)$. The forecast density π_n^f conditions on $y_{1:n-1}$, whereas the posterior density π_n^a also conditions on y_n, as in (2). Likelihood values determine the normalized conditional kernel coefficients, and likelihood gradients enter the Langevin steps.

Within one assimilation step, Z denotes the system state before artificial noise is added. The variable z denotes a possible value of the system state; z_i^m denotes a member of the proposal ensemble. The joint model in (3) defines the noisy state X_τ, whose conditional density given y_n is $q_{n,\tau}^a$. Its score $s_{n,\tau}^a$ is determined by the conditional mean $D_{n,\tau}$ through (6).

The forecast ensemble $\mathcal{X}_n^f$ initializes the M-member proposal ensemble $\mathcal{Z}_{n,\tau}$. The reported initialization uses $M = J$, with one proposal at each forecast member. The J reverse particles $x_\tau^{(j)}$ share the proposal ensemble but assign its members different normalized conditional weights. Their final states form the analysis ensemble $\mathcal{X}_n^a$. The tempered density $\pi_{n,\rho}$ specifies the sampling distribution assumed in the importance-sampling derivation. Appendix B gives the derivation and implementation details; Appendix C states the assumptions, consistency under independent sampling, and conditional-mean error bounds.

B CONDITIONAL-SCORE DERIVATION AND IMPLEMENTATION DETAILS

B.1 EXACT CONDITIONAL TWEEDIE IDENTITY

We first derive the score when the forecast distribution is known exactly. The result is a conditional form of Tweedie's identity (Efron, 2011; Boys et al., 2024; Peng et al., 2024; Patsenker et al., 2026); the next subsection constructs the conditional-mean estimator from the proposal ensemble.

Consider the joint model

$$Z \sim \pi_n^f, \tag{22}$$

$$Y_n \mid Z \sim g_n(\cdot \mid Z), \tag{23}$$

$$X_\tau \mid Z \sim \mathcal{N}(a_\tau Z, \sigma_\tau^2 \mathbf{I}_d). \tag{24}$$

The artificial diffusion noise is independent of the observation conditional on the system state Z. Conditioning on $Y_n = y_n$ changes the law of Z from the forecast density π_n^f to the analysis density π_n^a. The corresponding conditional density of X_τ is therefore the posterior after Gaussian noising, $q_{n,\tau}^a(\cdot \mid y_n)$.

Proposition 1 (conditional Tweedie identity). Under the conditions in Appendix C.1, for any $\tau \in (0, 1]$ with $a_\tau > 0$ and $\sigma_\tau > 0$, the conditional score satisfies

$$s_{n,\tau}^a(x; y_n) = -\frac{x}{\sigma_\tau^2} + \frac{a_\tau}{\sigma_\tau^2} D_{n,\tau}(x; y_n), \tag{25}$$

where

$$D_{n,\tau}(x; y_n) := \mathbb{E}[Z \mid X_\tau = x, Y_n = y_n]. \tag{26}$$

To see this, write the Gaussian-smoothed analysis density as

$$q_{n,\tau}^a(x \mid y_n) = \frac{1}{C_n(y_n)} \int \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z) g_n(y_n \mid z) \pi_n^f(z) \, \mathrm{d}z, \tag{27}$$

where $C_n(y_n)$ is independent of x. The integral gives the density of the noisy state X_τ conditional on $Y_n = y_n$. For a fixed noisy state x, the conditional density of the system state is

$$p(z \mid X_\tau = x, Y_n = y_n) = \frac{\phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z) g_n(y_n \mid z) \pi_n^f(z)}{\int \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau u) g_n(y_n \mid u) \pi_n^f(u) \, \mathrm{d}u}. \tag{28}$$

The conditional mean $D_{n,\tau}(x; y_n)$ in (26) is the first moment of (28). Differentiating (27) under the integral and dividing by $q^a_{n,\tau}(x \mid y_n)$ yields

$$\nabla_x \log q^a_{n,\tau}(x \mid y_n) = -\frac{x}{\sigma_\tau^2} + \frac{a_\tau}{\sigma_\tau^2} \frac{\int z\, \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z) g_n(y_n \mid z) \pi_n^f(z)\, \mathrm{d}z}{\int \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z) g_n(y_n \mid z) \pi_n^f(z)\, \mathrm{d}z}, \tag{29}$$

which is exactly (25).

B.2 Conditional-mean estimation from the proposal ensemble

At reverse step i, the analytical estimator in (12) is defined by the realized empirical measure $\nu^M_{n,i}$ in (10). For a fixed auxiliary query x, let

$$F_{i,x}(z) = \phi_{\sigma^2_{\tau_i} \mathbf{I}_d}(x - a_{\tau_i} z) g_n(y_n \mid z)^{1-\rho_i}. \tag{30}$$

Whenever $\nu^M_{n,i}(F_{i,x}) > 0$, its conditional-mean estimate is the weighted moment

$$D_{\nu^M_{n,i}}(x; y_n) = \frac{\nu^M_{n,i}(F_{i,x} z)}{\nu^M_{n,i}(F_{i,x})} = \sum_{m=1}^{M} c_m(x) z_i^m. \tag{31}$$

For the reverse-particle queries, this is $\widehat{D}^M_{n,i}$ in (11). The corresponding log coefficients are

$$b_{j,m}(\tau_i) = \log W_{i,m} - \frac{\left\| x_i^{(j)} - a_{\tau_i} z_i^m \right\|^2}{2\sigma^2_{\tau_i}} + (1 - \rho_i) \log g_n(y_n \mid z_i^m), \tag{32}$$

followed by log-sum-exp normalization over m. The resulting score estimate is

$$\widehat{s}^a_{n,\tau_i}(x; y_n) = -\frac{x}{\sigma^2_{\tau_i}} + \frac{a_{\tau_i}}{\sigma^2_{\tau_i}} D_{\nu^M_{n,i}}(x; y_n). \tag{33}$$

The likelihood-tilted empirical measure $\widehat{\eta}^M_{n,i}$ in (18) defines the Gaussian mixture

$$\widehat{q}^M_{n,i}(x; y_n) = \int \phi_{\sigma^2_{\tau_i} \mathbf{I}_d}(x - a_{\tau_i} z) \widehat{\eta}^M_{n,i}(\mathrm{d}z). \tag{34}$$

Holding the proposal ensemble fixed while differentiating with respect to x gives

$$\nabla_x \log \widehat{q}^M_{n,i}(x; y_n) = -\frac{x}{\sigma^2_{\tau_i}} + \frac{a_{\tau_i}}{\sigma^2_{\tau_i}} D_{\nu^M_{n,i}}(x; y_n). \tag{35}$$

B.3 Resampling and Langevin steps

Stratified ancestor sampling uses independent offsets $U_\ell \sim \mathrm{Unif}((\ell - 1)/M, \ell/M)$, followed by $A_\ell = \min\{m : \sum_{r=1}^m s_{i,r} \geq U_\ell\}$.

For the incoming proposal ensemble in Algorithm 1, let $r_j = (r_{i,j1}, \ldots, r_{i,jM})$ be the ancestor-selection distribution for reverse particle j in (14). Its average $s_i = J^{-1} \sum_j r_j$ minimizes the mean relative entropy:

$$s_i = \arg\min_{p \in \Delta_M} \frac{1}{J} \sum_{j=1}^{J} \mathrm{KL}(r_j \| p), \qquad \Delta_M = \left\{ p \in [0,1]^M : \sum_m p_m = 1 \right\}. \tag{36}$$

Indeed,

$$\frac{1}{J} \sum_{j=1}^{J} \mathrm{KL}(r_j \| p) = C + \mathrm{KL}(s_i \| p), \tag{37}$$

where C is independent of p. This identity characterizes ancestor allocation among the current proposal indices. The conditional-score coefficients $c_{j,m}$ in (11) have a different role and include the residual likelihood.

Conditional expectation under ancestor correction. Condition on the incoming positions, target weights ω_i, and reverse queries, and let $\mathcal{F}_i$ denote this information. Assume $s_{i,m} > 0$ whenever $\omega_{i,m} > 0$, and that the test function φ is integrable under the target empirical measure. If the offspring counts N_m satisfy $\mathbb{E}[N_m \mid \mathcal{F}_i] = M s_{i,m}$, then

$$\mathbb{E}\left[\frac{1}{M}\sum_{\ell=1}^{M}\frac{\omega_{i,A_\ell}}{s_{i,A_\ell}}\varphi(z^-_{A_\ell})\,\middle|\,\mathcal{F}_i\right] = \sum_m \omega_{i,m}\varphi(z^-_m). \tag{38}$$

To verify this, group descendants by their ancestor. The left side becomes

$$\frac{1}{M}\sum_m \mathbb{E}[N_m \mid \mathcal{F}_i]\frac{\omega_{i,m}}{s_{i,m}}\varphi(z^-_m),$$

which gives the stated sum. Stratified resampling has the required expected counts, using an independent uniform draw within each stratum. Taking $\varphi = 1$ gives expected total mass one. Normalizing the realized weights yields a finite-sample ratio estimator of the discrete target integral.

Effect of the move kernel. Let $\mu_i = \sum_m \omega_{i,m}\delta_{z^-_m}$. Conditional on $\mathcal{F}_i$, assume that the parameters of the Markov kernel Q_i are fixed and that a descendant of $z^-_{A_\ell}$ has transition law $Q_i(z^-_{A_\ell}, \cdot)$. Retain its unnormalized weight $v_\ell = \omega_{i,A_\ell}/s_{i,A_\ell}$, and define $\widetilde{\nu}_i^+(\varphi) = M^{-1}\sum_\ell v_\ell\varphi(Z^+_\ell)$. Under the preceding support and integrability conditions, conditional expectation first over the moves and then over offspring counts gives

$$\mathbb{E}[\widetilde{\nu}_i^+(\varphi) \mid \mathcal{F}_i] = \mu_i(Q_i\varphi), \tag{39}$$

$$\mathbb{E}[\widetilde{\nu}_i^+(\varphi) \mid \mathcal{F}_i] - \pi_{n,\rho_i}(\varphi) = (\mu_i - \pi_{n,\rho_i})(Q_i\varphi) + \pi_{n,\rho_i}(Q_i\varphi - \varphi). \tag{40}$$

The first term propagates the incoming proposal error through Q_i; the second measures the kernel's invariance defect relative to the reference target. A kernel that preserves the target makes the second term zero after any finite number of steps. Forecast approximation and ULA discretization can introduce a defect. The fitted Gaussian and move parameters are fixed from the forecast ensemble before ancestor selection. Taking $\varphi = F_{i,j}$ and its coordinatewise products with z gives the expected denominator and numerator errors for conditional mean estimation. The normalized estimator also depends on the random denominator (Appendix C.3).

The reference Langevin steps in (16) use the forecast-score approximation and likelihood gradient. If $\widehat{s}_n^f = \nabla \log \widetilde{\pi}_n^f$, their drift direction is the gradient of

$$\log\left[(\widetilde{\pi}_n^f)^{w_f} g_n(y_n \mid \cdot)^{w_g\rho_i}\right]. \tag{41}$$

The diagonal approximation supplies coordinatewise prior gradients without estimating a full covariance matrix. Short Langevin moves adjust the proposals within the sampling budget; their contribution to filtering performance is evaluated by the component controls.

B.4 Importance sampling with a tempered proposal density

For comparison, the conditional distribution of the system state is

$$\mu^{x,y_n}_{n,\tau}(z) := p(z \mid X_\tau = x, Y_n = y_n) \propto \phi_{\sigma_\tau^2\mathbf{I}_d}(x - a_\tau z)\pi_n^f(z)g_n(y_n \mid z). \tag{42}$$

For normalized diffusion time $\tau \in [0,1]$, the experiments use

$$\rho_\tau = 1 - \tau. \tag{43}$$

For any fixed $\rho \in [0,1]$, define the tempered proposal density

$$\pi_{n,\rho}(z) \propto \pi_n^f(z)g_n(y_n \mid z)^\rho. \tag{44}$$

The same conditional distribution can be written as

$$\mu^{x,y_n}_{n,\tau}(z) \propto \pi_{n,\rho}(z)\underbrace{\phi_{\sigma_\tau^2\mathbf{I}_d}(x - a_\tau z)}_{\text{Gaussian noising}}\underbrace{g_n(y_n \mid z)^{1-\rho}}_{\text{remaining likelihood}}. \tag{45}$$

Because $g_n^\rho g_n^{1-\rho} = g_n$, changing ρ leaves the conditional distribution unchanged. If z^m are independent samples from $\pi_{n,\rho}$, the conditional mean $D_{n,\tau}(x; y_n)$ can be estimated by

$$\widehat{D}^{\mathrm{IS}}_{n,\tau,\rho}(x; y_n) = \sum_{m=1}^{M} \omega_m^{(\rho)}(\tau, x) z^m, \tag{46}$$

where

$$b_m^{(\rho)}(\tau, x) = -\frac{\|x - a_\tau z^m\|^2}{2\sigma_\tau^2} + (1-\rho)\log g_n(y_n \mid z^m), \tag{47}$$

$$\omega_m^{(\rho)}(\tau, x) = \frac{\exp(b_m^{(\rho)}(\tau, x) - c)}{\sum_{\ell=1}^{M} \exp(b_\ell^{(\rho)}(\tau, x) - c)}, \qquad c = \max_\ell b_\ell^{(\rho)}(\tau, x). \tag{48}$$

Here the coefficients are self-normalized importance weights because the samples come from $\pi_{n,\rho}$. Larger ρ places more likelihood information in that proposal and leaves less in the residual weight. Such likelihood-tempering paths are also used in sequential Monte Carlo and annealed importance sampling (Del Moral et al., 2006; Beskos et al., 2014; Neal, 2001).

B.5 Reverse-diffusion analysis update and recursive filtering

Algorithm 1 gives the analysis update in the main text. For the differentiable Gaussian noising schedule, define

$$f_\tau = \partial_\tau \log a_\tau, \qquad r_\tau^2 = \partial_\tau \sigma_\tau^2 - 2 f_\tau \sigma_\tau^2, \qquad \Delta_i = \tau_i - \tau_{i-1}. \tag{49}$$

The schedule is chosen so that $r_\tau^2 \geq 0$. The reported recursive experiments use

$$a_\tau = 1 - (1-\epsilon_a)\tau, \qquad \sigma_\tau^2 = \epsilon_b + (1-\epsilon_b)\tau, \tag{50}$$

with $0 < \epsilon_a, \epsilon_b < 1$. Thus $a_0 = 1$ and $\sigma_0^2 = \epsilon_b > 0$, and the diffusion rate in (49) is nonnegative. The grid is uniform in logSNR:

$$\lambda(\tau) = \log \frac{a_\tau^2}{\sigma_\tau^2}, \qquad \lambda(\tau_i) = \lambda(1) + \frac{N_\tau - i}{N_\tau}[\lambda(0) - \lambda(1)]. \tag{51}$$

Inverting this relation gives τ_i, with $\rho_i = 1 - \tau_i$ and $\Delta_i = \tau_i - \tau_{i-1}$. The full logSNR convention differs by a constant factor of two from $\log(a_\tau/\sigma_\tau)$ and generates the same uniform grid.

The reverse particles are advanced by (17). At each assimilation time, independent standard-normal entries are drawn and then standardized componentwise across the J particles to zero mean and unit sample standard deviation. This standardization makes the initialized members dependent; Proposition C.3 quantifies its member-marginal discrepancy from the terminal noised posterior. The conditional-score estimator can also be evaluated at prescribed query states without running the reverse solver.

Algorithm 2 Recursive filtering with AECSF

Require: Initial analysis ensemble $\mathcal{X}_0^a$, observations $\{y_n\}_{n=1}^{N_a}$, forecast model $\mathcal{M}_n$, differentiable likelihoods g_n, reverse-particle initialization rule, and the settings in Algorithm 1

1: **for** $n = 1, \ldots, N_a$ **do**
2: Forecast: propagate $\mathcal{X}_{n-1}^a$ through $\mathcal{M}_n$ to obtain $\mathcal{X}_n^f$
3: Analysis: initialize the reverse particles at $\tau = 1$ and apply Algorithm 1 to $\mathcal{X}_n^f$ and y_n
4: **end for**
5: **return** analysis ensembles $\{\mathcal{X}_n^a\}_{n=1}^{N_a}$

B.6 Numerical stability and computational cost

The implementation computes (32) in the log domain and normalizes the conditional kernel coefficients after subtracting their largest log value:

$$c = \max_m b_m, \tag{52}$$

$$\widetilde{c}_m = \exp(b_m - c), \tag{53}$$

$$c_m = \widetilde{c}_m / \sum_\ell \widetilde{c}_\ell. \tag{54}$$

The same log-sum-exp operation normalizes the incremental target weights, ancestor selection probabilities, and corrected descendant weights separately. For finite log coefficients, subtracting their maximum preserves the normalized coefficients and prevents exponential overflow. At least one unnormalized coefficient equals one.

With J reverse particles, M proposal members, N_τ reverse steps, and state dimension d, evaluating the Gaussian factors and weighted conditional means requires

$$O(JN_\tau Md) \tag{55}$$

arithmetic work. At fixed J, this work grows linearly with M; when $M = J$, it grows quadratically with ensemble size. For the Gaussian kernel, we use

$$\|x_j - a_\tau z_m\|^2 = \|x_j\|^2 + a_\tau^2\|z_m\|^2 - 2a_\tau x_j^T z_m.$$

The implementation applies this identity after centering the states and computes the pairwise inner products and weighted conditional means by matrix multiplication on the GPU. This avoids storing a $J\times M\times d$ difference tensor and requires $O(JM+(J+M)d)$ working storage for these calculations. If each proposal receives L Langevin steps and one gradient evaluation costs $C_\nabla(d)$, their gradient work is $O(LMC_\nabla(d))$; likelihood evaluation contributes its observation-model cost separately. At $T = 20$, the measured analysis time for $J = M \in \{10, 20, 40, 80, 160\}$ remains within 0.099–0.104 seconds on the RTX 4090 (Figure 3). The arithmetic and storage costs still grow with ensemble size.

C Exact identities and approximation error

C.1 Conditions for the conditional-score identity

For the fixed observation, assume that its marginal density is positive and finite:

$$C_n(y_n) = \int g_n(y_n \mid z)\pi_n^f(z)\,\mathrm{d}z, \qquad 0 < C_n(y_n) < \infty. \tag{56}$$

This condition ensures that $\pi_n^a(z) = g_n(y_n \mid z)\pi_n^f(z)/C_n(y_n)$ is a probability density. For fixed τ with $a_\tau > 0$ and $\sigma_\tau > 0$, the Gaussian kernel and its first derivative in x are bounded as functions of z. For fixed x, the function $\|z\|\,\phi_{\sigma_\tau^2\mathbf{I}_d}(x - a_\tau z)$ is also bounded. These bounds justify differentiation under the integral and ensure that the conditional first moment is finite. The derivation in Appendix B therefore yields (25). The likelihood factorization in (45) leaves this conditional distribution unchanged for every $\rho \in [0, 1]$.

C.2 Consistency under independent sampling

Fix (τ, x, y_n, ρ), and let $z^1, \dots, z^M$ be independent samples from the tempered density $\pi_{n,\rho}$. Define the unnormalized importance weight for this fixed query by

$$W(z) = \phi_{\sigma_\tau^2\mathbf{I}_d}(x - a_\tau z)g_n(y_n \mid z)^{1-\rho}. \tag{57}$$

Assume $0 < \mathbb{E}_{\pi_{n,\rho}}[W(Z)] < \infty$ and $\mathbb{E}_{\pi_{n,\rho}}[W(Z)\,\|Z\|] < \infty$. The importance-sampling estimator in (46) then converges to the population conditional mean:

$$\widehat{D}^{\mathrm{IS}}_{n,\tau,\rho}(x; y_n) \xrightarrow[M\to\infty]{p} D_{n,\tau}(x; y_n). \tag{58}$$

To establish this result, write the estimator as the ratio of sample averages:

$$\begin{aligned}\widehat{D}^{\mathrm{IS}}_{n,\tau,\rho}(x; y_n) &= \frac{M^{-1}\sum_{m=1}^M W(z^m)z^m}{M^{-1}\sum_{m=1}^M W(z^m)} \\ &\xrightarrow[M\to\infty]{p} \frac{\mathbb{E}_{\pi_{n,\rho}}[W(Z)Z]}{\mathbb{E}_{\pi_{n,\rho}}[W(Z)]} = D_{n,\tau}(x; y_n).\end{aligned} \tag{59}$$

The law of large numbers applies separately to the numerator and denominator. Their ratio converges because the limiting denominator is positive and finite. The final equality follows from the conditional-density factorization in (45). Importance sampling becomes more demanding as the target and proposal separate, particularly in high intrinsic dimension (Agapiou et al., 2017; Chatterjee & Diaconis, 2018).

For a different proposal density q whose support covers the conditional target, the corresponding unnormalized importance weight is instead

$$W_q(z) = \frac{\phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z)\pi_n^f(z)g_n(y_n \mid z)}{q(z)}. \tag{60}$$

Setting $q = \pi_{n,\rho}$ reduces (60) to (57), up to a constant that cancels under normalization.

Target condition for a fixed proposal measure. Write $p = \pi_n^f$, $g(z) = g_n(y_n \mid z)$, and $\pi_\rho(\mathrm{d}z) \propto p(z)g(z)^\rho \mathrm{d}z$. Let ν_τ be a probability measure held fixed during score-field evaluation. Assume g is positive and finite on the supports of p and ν_τ, and all normalizing constants below and for π_ρ, π_1 are positive and finite. Define

$$\eta_\tau(\mathrm{d}z) = \frac{g(z)^{1-\rho}\nu_\tau(\mathrm{d}z)}{\nu_\tau(g^{1-\rho})}, \qquad q_\tau^\nu(x) = \int \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z)\eta_\tau(\mathrm{d}z). \tag{61}$$

Proposition. For $a_\tau \neq 0$ and $\sigma_\tau^2 > 0$, q_τ^ν equals the noised posterior everywhere if and only if $\nu_\tau = \pi_\rho$. The same equivalence holds for their score fields.

Proof. Gaussian noising maps a measure η to a density with Fourier transform

$$\widehat{q}(\xi) = e^{-\sigma_\tau^2 \|\xi\|^2/2}\widehat{\eta}(a_\tau \xi).$$

The Gaussian multiplier is nonzero and the scaling invertible, so equality of noised densities is equivalent to $\eta_\tau = \pi_1$. Inverting the likelihood tilt gives

$$\nu_\tau(\mathrm{d}z) = \frac{g(z)^{\rho-1}\pi_1(\mathrm{d}z)}{\pi_1(g^{\rho-1})} = \pi_\rho(\mathrm{d}z).$$

Conversely, this substitution gives $\eta_\tau = \pi_1$. Positive normalized noised densities have identical scores exactly when they are equal.

At initialization, $\rho = 0$ and the population choice $\nu_1 = p$ gives $\eta_1 = \pi_1$. Subsequent stages advance the empirical target weights even when resampling is skipped. The weighted-moment analysis below characterizes the discrepancy between the updated proposal measure and the tempered target.

C.3 Conditional-mean error in terms of weighted moments

Fix (τ, x, y_n, ρ), write $\pi = \pi_{n,\rho}$, and set

$$F(z) = \phi_{\sigma_\tau^2 \mathbf{I}_d}(x - a_\tau z)g_n(y_n \mid z)^{1-\rho}. \tag{62}$$

For any probability measure ν with $0 < \nu(F) < \infty$ and $\nu(F\,\|z\|) < \infty$, define $D_\nu = \nu(Fz)/\nu(F)$. Since $D_{n,\tau}(x; y_n) = \pi(Fz)/\pi(F)$, the conditional-mean error obeys the exact identity

$$D_\nu - D_{n,\tau}(x; y_n) = \frac{(\nu - \pi)(F[\, z - D_{n,\tau}(x; y_n)\,])}{\nu(F)}. \tag{63}$$

Let

$$\delta_0 = |(\nu - \pi)(F)|\,, \qquad \delta_1 = \|(\nu - \pi)(Fz)\|\,. \tag{64}$$

If $\delta_0 < \pi(F)$, then

$$\|D_\nu - D_{n,\tau}(x; y_n)\| \leq \frac{\delta_1 + \|D_{n,\tau}(x; y_n)\|\,\delta_0}{\pi(F) - \delta_0}. \tag{65}$$

Under the stated moment conditions and $\delta_0 < \pi(F)$, this bound also applies to the empirical measure $\nu_{n,i}^M$ of the proposal ensemble. Thus, convergence of both weighted moments is sufficient for convergence of the conditional-mean estimate at the fixed query.

Let $\widehat{s}_{\nu,n,\tau}^a$ denote the corresponding field. At a fixed query state, its error follows directly from (25):

$$\widehat{s}_{\nu,n,\tau}^a(x; y_n) - s_{n,\tau}^a(x; y_n) = \frac{a_\tau}{\sigma_\tau^2}\left[D_\nu - D_{n,\tau}(x; y_n)\right]. \tag{66}$$

Resampling and Langevin steps affect the weighted moments in (63). The recursive filter also introduces time discretization, terminal initialization, and a retained final noise variance, as specified in Appendix B.5.

C.4 ADAPTIVE PROPOSALS AND REVERSE-SAMPLING ERROR

For this subsection, fix one assimilation cycle. Use the execution index $k = 1, \dots, K$, where $K = N_\tau$, $\theta_k = \tau_{N_\tau - k + 1}$, and $\delta_k = \tau_{N_\tau - k + 1} - \tau_{N_\tau - k}$. Here X_k^j denotes a reverse particle indexed by execution stage; the unnoised system state in the conditional Tweedie identity is denoted by Z. Thus X_{k-1}^j corresponds to the input $x_{N_\tau - k + 1}^{(j)}$ in Algorithm 1, and X_k^j to its output $x_{N_\tau - k}^{(j)}$. Write $\varrho_k = 1 - \theta_k$ and $\varrho_0 = 0$. The first score evaluation uses the initial proposal ensemble, so $G_1 = 1$ and $Q_1 = I$; X_K^j is the numerical endpoint.

We condition on the forecast ensemble, observation, and initial proposal ensemble. Let $p = \widetilde{\pi}_n^f$ be the fitted forecast density, g the likelihood, and $\pi_k = p g^{\varrho_k} / Z_k$, where $Z_k = \int p g^{\varrho_k}$. Let Q_k be the transition kernel for one proposal member at stage k, including the identity when movement is omitted. Its parameters are fixed within each reverse run. Across proposal counts, Q_k abbreviates the corresponding deterministic kernel Q_k^M; each reference uses that run's forecast fit and initial proposals. Define

$$G_k = g^{\varrho_k - \varrho_{k-1}}, \qquad T_k f = G_k Q_k f, \qquad \Gamma_k = \eta_0 T_1 \cdots T_k, \qquad \eta_k = \Gamma_k / \Gamma_k(1).$$

The initial measure η_0 is the empirical proposal measure. Thus η_k is the numerical reference generated by the initial proposal ensemble, incremental weighting, and proposal transitions.

Assumptions. The number of stages and the dimension are fixed. Conditional on the inputs, G_k and Q_k are deterministic, $0 < G_k \le b_k < \infty$, and all displayed integrals exist. At stage k, the current queries and incoming proposal ensemble determine the ancestor probabilities before new sampling noise is drawn. Selection covers every positive target weight and uses the correction $v = \omega / s$. Stratified offsets and descendant transition noises are conditionally independent. Ratios and variance sums are taken on the positive-weight support, with zero-weight terms contributing zero.

Proposition C.1 (corrected adaptive sampling). Let ν_{k-1} be the normalized incoming proposal measure, $c_k = \nu_{k-1}(G_k)$, and

$$Y_k(f) = \frac{1}{M} \sum_{\ell=1}^{M} \frac{\omega_{k,A_\ell}}{s_{k,A_\ell}} f(Z_\ell), \quad R_k = Y_k(1), \quad m_k = m_{k-1} c_k R_k, \quad \widehat{\Gamma}_k = m_k \nu_k, \quad m_0 = 1,$$

where Z_ℓ is the moved descendant and $\nu_k = Y_k / Y_k(1)$. Without resampling, let $Z'_m \sim Q_k(z_m, \cdot)$ independently conditional on the incoming ensemble and define

$$Y_k(f) = \sum_m \omega_{k,m} f(Z'_m), \qquad R_k = Y_k(1) = 1.$$

The identity kernel covers stages with no movement. The same definition also covers a deterministic one-member transition. If $\mathcal{F}_k^-$ contains the incoming proposal ensemble, queries, and past randomness, then

$$\mathbb{E}[\widehat{\Gamma}_k(f) \mid \mathcal{F}_k^-] = \widehat{\Gamma}_{k-1}(T_k f), \qquad \mathbb{E}\widehat{\Gamma}_k(f) = \Gamma_k(f).$$

With resampling, the first equality follows by grouping descendants by ancestor and using the expected count $M s_{k,m}$, which cancels the selection probability in $\omega_{k,m}/s_{k,m}$. Without resampling, $\mathbb{E}[Y_k(f) \mid \mathcal{F}_k^-] = \sum_m \omega_{k,m} Q_k f(z_m)$ gives the same equality. These statements concern unnormalized weighted sums and allow dependence between the queries and incoming proposals.

For stratified resampling define $\Lambda_k = \sum_m \omega_{k,m}^2 / s_{k,m}$. The query-dependent selection rule admits a direct bound on this factor. Let

$$\zeta_{k,j} = \sum_m \omega_{k,m} K_k(X_{k-1}^j, z_m), \qquad H_{k,m} = \frac{1}{J} \sum_{j=1}^{J} \frac{K_k(X_{k-1}^j, z_m)}{\zeta_{k,j}}.$$

Then $s_{k,m} = \omega_{k,m} H_{k,m}$ and

$$\Lambda_k = \sum_m \frac{\omega_{k,m}}{H_{k,m}} = 1 + \chi^2(\omega_k \| s_k).$$

If $\|z_m\| \le B$, the Gaussian noising kernel satisfies

$$\frac{\max_m K_k(X_{k-1}^j, z_m)}{\min_m K_k(X_{k-1}^j, z_m)} \le e^{D_{k,j}}, \qquad D_{k,j} = \frac{2|a_{\theta_k}|B\|X_{k-1}^j\| + a_{\theta_k}^2 B^2/2}{\sigma_{\theta_k}^2}.$$

Consequently,

$$H_{k,m} \ge \frac{1}{J}\sum_{j=1}^{J} e^{-D_{k,j}}, \qquad \Lambda_k \le \left(\frac{1}{J}\sum_{j=1}^{J} e^{-D_{k,j}}\right)^{-1} \le \exp\left(\frac{1}{J}\sum_{j=1}^{J} D_{k,j}\right).$$

This connects query–proposal mismatch to the correction variance in the resampling branch. In particular, $\mathbb{E}[m_{k-1}^2 \exp(J^{-1}\sum_j D_{k,j})] < \infty$ controls its weighted local-variance factor.

Branch-wise local variance. For a bounded test function f, independent stratified offsets and independent transition noises give

$$\mathrm{Var}(Y_k(f) \mid \mathcal{F}_k^-) \le \frac{1}{M}\sum_m \frac{\omega_{k,m}^2}{s_{k,m}} Q_k(f^2)(z_m) \le \frac{\Lambda_k}{M}\|f\|_\infty^2$$

when resampling is used, by summing second moments across strata. Without resampling, conditional independence of the one-member moves gives

$$\begin{aligned}\mathrm{Var}(Y_k(f) \mid \mathcal{F}_k^-) &= \sum_m \omega_{k,m}^2 \left[Q_k(f^2)(z_m) - (Q_k f(z_m))^2\right] \\ &\le \|f\|_\infty^2 \sum_m \omega_{k,m}^2.\end{aligned} \tag{67}$$

This variance is zero for a deterministic transition. Define

$$\Lambda_k^{\mathrm{loc}} = \begin{cases} \sum_m \omega_{k,m}^2/s_{k,m}, & \text{with resampling,} \\ M\sum_m \omega_{k,m}^2, & \text{without resampling, with a random move,} \\ 0, & \text{without resampling, with a deterministic transition.} \end{cases} \tag{68}$$

Then all branches obey $\mathrm{Var}(Y_k(f) \mid \mathcal{F}_k^-) \le \Lambda_k^{\mathrm{loc}}\|f\|_\infty^2/M$. The following quantitative bounds assume $C_k := \mathbb{E}[m_{k-1}^2 \Lambda_k^{\mathrm{loc}}] < \infty$. In the unresampled random-move branch, this explicitly controls $M\sum_m \omega_{k,m}^2 = M/\,\mathrm{ESS}(\omega_k)$, where $\mathrm{ESS}(\omega_k) = (\sum_m \omega_{k,m}^2)^{-1}$.

Set

$$V_k = \frac{1}{M}\sum_{r=1}^{k} C_r \prod_{\ell=r}^{k} b_\ell^2, \qquad E_k = 32V_k/\Gamma_k(1)^2.$$

For every fixed bounded f,

$$\mathbb{E}|\widehat{\Gamma}_k(f) - \Gamma_k(f)|^2 \le V_k\|f\|_\infty^2, \qquad \mathbb{E}|\nu_k(f) - \eta_k(f)|^2 \le \min(4, E_k)\|f\|_\infty^2.$$

To see the first bound, apply the local-variance inequality to the deterministic backward functions $T_{r+1}\cdots T_k f$. Their suprema are at most $\|f\|_\infty \prod_{\ell=r+1}^k b_\ell$, while $c_r = \nu_{r-1}(G_r) \le b_r$. The martingale increment $\widehat{\Gamma}_r(T_{r+1}\cdots T_k f) - \widehat{\Gamma}_{r-1}(T_r \cdots T_k f)$ therefore has second moment at most $C_r\|f\|_\infty^2 \prod_{\ell=r}^k b_\ell^2/M$. Summing the orthogonal increments proves the bound. For the normalized bound, on $m_k \ge \Gamma_k(1)/2$ use

$$\nu_k(f) - \eta_k(f) = \frac{\widehat{\Gamma}_k(f) - \Gamma_k(f) - \eta_k(f)[m_k - \Gamma_k(1)]}{m_k}.$$

The two unnormalized error bounds contribute at most $16V_k\|f\|_\infty^2/\Gamma_k(1)^2$ on this event. On its complement, use $|\nu_k(f) - \eta_k(f)| \le 2\|f\|_\infty$ and the mass-error bound to obtain the same contribution. The trivial $4\|f\|_\infty^2$ bound yields the displayed minimum.

Retaining the test function in the local variance gives a more specific bound. For $h_r = T_{r+1} \cdots T_k f$, define

$$B_r(h) = \begin{cases} M^{-1} \sum_m \omega_{r,m}^2 Q_r(h^2)(z_m)/s_{r,m}, & \text{with resampling,} \\ \sum_m \omega_{r,m}^2 [Q_r(h^2)(z_m) - (Q_r h(z_m))^2], & \text{without resampling.} \end{cases}$$

The same martingale argument yields

$$\mathbb{E}|\widehat{\Gamma}_k(f) - \Gamma_k(f)|^2 \le \sum_{r=1}^{k} \mathbb{E}[(m_{r-1}c_r)^2 B_r(h_r)].$$

Proposition C.1a (full-chain consistency). Suppose the initial proposal measures have uniformly bounded rth moments for some $r > 0$, and their largest initial weights tend to zero. Let each G_k be continuous and let the one-member kernels satisfy

$$Q_k(\|\cdot\|^r)(z) \le A_k(1 + \|z\|^r)$$

with A_k uniform in M. Assume average tightness of the queries:

$$q_{k,R} := \sup_M \frac{1}{J_M} \sum_{j=1}^{J_M} \Pr(\|X_{k-1}^{j,M}\| > R) \longrightarrow 0 \quad \text{as } R \to \infty.$$

For the Gaussian affinity with fixed $a_{\theta_k} \neq 0$ and $\sigma_{\theta_k} > 0$, the corrected proposal recursion satisfies

$$\sup_{\|f\|_\infty \le 1} \mathbb{E}|\nu_k(f) - \eta_k^M(f)|^2 \longrightarrow 0 \qquad (k \le K).$$

The supremum is over deterministic bounded measurable functions and is outside the expectation. This includes $J_M = M$ and queries dependent on the proposal ensemble.

Proof. The reference moment bound follows by induction. An incoming moment bound places at least half the reference mass in a fixed ball, where G_k has a positive minimum. Hence $\eta_{k-1}^M(G_k)$ is uniformly positive and

$$\eta_k^M(\|z\|^r) \le \frac{b_k A_k [1 + \eta_{k-1}^M(\|z\|^r)]}{\eta_{k-1}^M(G_k)}.$$

Thus $\gamma_k^M := \Gamma_k^M(1)$ has a uniform positive lower bound, $\gamma_k^M \le \bar{\gamma}_k := \prod_{r=1}^k b_r$, and $t_{k,B} := \sup_M \Gamma_k^M(1_{\{\|z\|>B\}}) \to 0$ as $B \to \infty$. Proposition C.1 gives $\mathbb{E}m_k = \gamma_k^M$, so $\Pr(m_k > L) \le \bar{\gamma}_k/L$.

Let $E_{k,R}$ be the event that at least half the queries lie in the ball of radius R. Its complement has probability at most $2q_{k,R}$. Cancel the common Gaussian prefactor in $H_{k,m}$, so $\zeta_{k,j} \le 1$; on this event every parent with $\|z_m\| \le B$ satisfies

$$H_{k,m} \ge c_{k,R,B} := \tfrac{1}{2} \exp\left[-\frac{(R + |a_{\theta_k}|B)^2}{2\sigma_{\theta_k}^2} \right] > 0.$$

Write $u_{k,m}$ for the unnormalized member weights, so $\widehat{\Gamma}_k = \sum_m u_{k,m} \delta_{z_{k,m}}$. On $E_{k,R} \cap \{m_{k-1} \le L\}$, descendants of these parents have $u_{k,m} \le L b_k/(M c_{k,R,B})$. The expected total weight of descendants whose parents lie outside the ball is at most $b_k t_{k-1,B}$, by the conditional expectation identity in Proposition C.1. Markov's inequality therefore gives $\max_m u_{k,m} \to 0$ in probability: fix L, R, B, let $M \to \infty$, then enlarge the three bounds. At a stage without resampling, $\max_m u_{k,m} \le b_k \max_m u_{k-1,m}$, which completes the induction from the initial weights.

For deterministic $\|h\|_\infty \le 1$, define the innovation $\mathcal{E}_k(h) = \widehat{\Gamma}_k(h) - \widehat{\Gamma}_{k-1}(T_k h)$. At a resampling stage, split it according to whether the parent lies inside or outside the ball of radius B. On the same event as above, the conditional variance of the inside contribution is at most $L^2 b_k^2/(M c_{k,R,B})$. The outside contribution has expected absolute value at most $2b_k t_{k-1,B}$. Conditional Chebyshev and Markov inequalities give

$$\sup_{\|h\|_\infty \le 1} \Pr(|\mathcal{E}_k(h)| > \epsilon) \le \bar{\gamma}_{k-1}/L + 2q_{k,R} + \frac{4L^2 b_k^2}{M c_{k,R,B} \epsilon^2} + \frac{4 b_k t_{k-1,B}}{\epsilon}.$$

The same order of limits makes this probability vanish. Without resampling, the conditional variance is bounded by $b_k^2(\max_m u_{k-1,m})m_{k-1}$, which tends to zero in probability. Taking expectations of the conditional probability bound truncated at one gives the same conclusion. Deterministic transitions have zero innovation at these stages.

Finally, telescope with the deterministic backward functions $h_r = T_{r+1} \cdots T_k f$:

$$\widehat{\Gamma}_k(f) - \Gamma_k^M(f) = \sum_{r=1}^{k} \mathcal{E}_r(h_r), \qquad \|h_r\|_\infty \le \|f\|_\infty \prod_{\ell=r+1}^{k} b_\ell.$$

The uniform innovation bounds give convergence in probability uniformly over $\|f\|_\infty \le 1$. Taking $f = 1$ controls $m_k - \gamma_k^M$. The positive reference mass bound then permits normalization. Since $|\nu_k(f) - \eta_k^M(f)| \le 2$, this uniform probability convergence also gives the stated mean-square convergence.

Moment growth for ordinary Langevin moves. Suppose the drift satisfies $\|b_k^M(z)\| \le L_k\|z\| + B_k$, with constants uniform in M. For $Z' = z + (h/2)b_k^M(z) + \sqrt{h}\,\xi$, where $\xi \sim \mathcal{N}(0, I_d)$,

$$\mathbb{E}[\|Z'\|^2 \mid z] \le 2(1 + hL_k/2)^2\|z\|^2 + h^2B_k^2/2 + hd.$$

Iteration over a fixed number of moves establishes the moment-growth condition in Proposition C.1a with $r = 2$. For the diagonal Gaussian forecast, a uniform variance lower bound and bounded forecast means give this linear-growth bound for the prior gradient. With fixed observation y, the arctangent observation map and its derivative are bounded, so the likelihood gradient is bounded. A linear Gaussian observation gives an affine likelihood gradient and also satisfies the same growth condition.

Proposition C.2 (adaptive conditional means). For this proposition, suppose proposal measures and numerical references are supported in a common ball of radius B, uniformly in M. The residual likelihood is continuous and positive there, the noising variance is positive, and the adaptive queries $X = X_M$ are uniformly tight. Cancel the noising factor independent of z and write

$$F_x(z) = \exp\left(\frac{a_{\theta_k}x^Tz}{\sigma_{\theta_k}^2} - \frac{a_{\theta_k}^2\|z\|^2}{2\sigma_{\theta_k}^2}\right) g(z)^{1-\varrho_k}, \qquad D_\mu(x) = \frac{\mu(zF_x)}{\mu(F_x)}.$$

On $\|x\| \le R$, let $U_R = \sup F_x(z)$ and $l_R = \inf F_x(z) > 0$ over the proposal ball. The weighted-moment ratio identity and Proposition C.1 give the fixed-query squared-error bound $A_R = 4dB^2(U_R/l_R)^2E_k$. Moreover, $\nabla_x D_\mu = (a_{\theta_k}/\sigma_{\theta_k}^2)\,\mathrm{Cov}_{\mu,x}(Z)$, so the difference between two such conditional means is $2|a_{\theta_k}|B^2/\sigma_{\theta_k}^2$-Lipschitz. An ϵ-net argument therefore gives, for an adaptive query X,

$$\mathbb{E}\|D_{\nu_k}(X) - D_{\eta_k}(X)\|^2 \le \min\left\{4B^2, 2(1+2R/\epsilon)^dA_R + \frac{8a_{\theta_k}^2B^4\epsilon^2}{\sigma_{\theta_k}^4} + 4B^2\Pr(\|X\| > R)\right\}.$$

The bound allows dependence between X and ν_k. If the C_k are uniformly bounded, $\Gamma_k(1)$ is uniformly positive, and the query tightness condition holds, taking $M \to \infty$, then $\epsilon \to 0$ and $R \to \infty$, yields mean-square convergence of the conditional means.

Proposition C.2a (conditional means on an unbounded state space). Under Proposition C.1a, suppose $\ell(z) = g(z)^{1-\varrho_k}$ is positive, continuous and bounded by L_ℓ. Then, for an independent uniform member index I_M,

$$D_{\nu_k}(X_{k-1}^{I_M}) - D_{\eta_k^M}(X_{k-1}^{I_M}) \longrightarrow 0 \quad \text{in probability.}$$

If the squared errors are uniformly integrable, convergence also holds in mean square.

Proof. Write $a = a_{\theta_k}$, $\sigma = \sigma_{\theta_k}$ and retain the full Gaussian factor, $\widetilde{F}_x(z) = \exp[-\|x - az\|^2/(2\sigma^2)]\ell(z)$. It gives the same ratio $D_\mu = \mu(z\widetilde{F}_x)/\mu(\widetilde{F}_x)$. For $\|x\| \le R$, uniformly over $z \in \mathbb{R}^d$,

$$|\widetilde{F}_x| \le L_\ell, \qquad \|z\widetilde{F}_x\| \le \frac{L_\ell}{|a|}(R + \sigma/\sqrt{e}),$$

$$\|\nabla_x\widetilde{F}_x\| \le \frac{L_\ell}{\sigma\sqrt{e}}, \quad \|\nabla_x(z\widetilde{F}_x)\|_{\mathrm{op}} \le \frac{L_\ell}{|a|}\left(\frac{R}{\sigma\sqrt{e}} + \frac{2}{e}\right).$$

These bounds follow from $\|z\| \le (R+\|x-az\|)/|a|$ and the first two polynomial Gaussian maxima. Reference tightness places a fixed positive mass in a ball where ℓ has a positive minimum, giving $\inf_M \inf_{\|x\|\le R} \eta_k^M(\widetilde{F}_x) > 0$. Apply Proposition C.1a to the numerator coordinates and denominator on a finite query net. The derivative bounds extend convergence to the whole query ball, and the positive denominator bound gives uniform convergence in probability of the ratio. Average query tightness handles the probability outside that ball. Uniform integrability gives the additional mean-square conclusion.

Proposition C.3 (standardized terminal initialization). For one coordinate, let $G_1, \dots, G_J$ be independent standard Gaussians, $\bar{G} = J^{-1}\sum_j G_j$, $S^2 = (J-1)^{-1}\sum_j (G_j - \bar{G})^2$, and $\widetilde{S} = \max(S, \varepsilon)$. The recursive implementation initializes $X_0^j = (G_j - \bar{G})/\widetilde{S}$ componentwise. If I is uniform on $\{1, \dots, J\}$ and independent of the ensemble, then

$$\mathbb{E}(X_0^I - G_I)^2 = \kappa_{J,\varepsilon} := \frac{1}{J} + \frac{J-1}{J}\mathbb{E}\Big[S^2(\widetilde{S}^{-1} - 1)^2\Big].$$

For $\varepsilon = 0$ this becomes

$$\kappa_{J,0} = \frac{1}{J} + \frac{2(J-1)}{J}(1 - c_J), \qquad c_J = \sqrt{\frac{2}{J-1}}\,\frac{\Gamma(J/2)}{\Gamma((J-1)/2)}.$$

Across d independent coordinates, $\mathbb{E}\|X_0^I - G_I\|^2 = d\kappa_{J,\varepsilon}$. Since the exact terminal noised posterior has the form $\epsilon_a Z + G$ for the reported schedule, where Z has the analysis distribution and $G \sim \mathcal{N}(0, I_d)$ is independent of Z,

$$W_2(\mathcal{L}(X_0^I), \mathcal{L}(\epsilon_a Z + G)) \le \sqrt{d\kappa_{J,\varepsilon}} + \epsilon_a(\mathbb{E}\|Z\|^2)^{1/2}.$$

This is a member-marginal bound; standardization makes the J initialized members dependent.

Theorem C.4 (reverse endpoint control). Let s_k^R be the score obtained from the numerical proposal reference and let $\Psi_k(x) = (1 - \delta_k f_{\theta_k})x + \alpha_k s_k^R(x)$, where $\alpha_k = \delta_k r_{\theta_k}^2$. Assume Ψ_k is β_k-Lipschitz. Couple the actual reverse particles X_k^j and reference particles Y_k^j with the same Euler noise, and define

$$q_k^2 = \mathbb{E}\frac{1}{J}\sum_j \|\widehat{s}_k^j(X_{k-1}^j) - s_k^R(X_{k-1}^j)\|^2, \qquad w_k = \alpha_k \prod_{\ell=k+1}^{K} \beta_\ell.$$

For a uniform member index I and any target $q_\star$ with a finite second moment,

$$W_2(\mathcal{L}(X_K^I), q_\star) \le \left(\prod_{k=1}^{K}\beta_k\right)\left(\mathbb{E}\|X_0^I - Y_0^I\|^2\right)^{1/2} + \sum_{k=1}^{K} w_k q_k + W_2(\mathcal{L}(Y_K^I), q_\star).$$

The first term can be bounded by Proposition C.3 in the recursive setting; it is zero in the controlled Gaussian experiment, which samples the analytic terminal reference. Under Proposition C.2, the conditional Tweedie identity gives $q_k \to 0$ for raw scores and for scores transformed by the same nonexpansive map in both chains.

Corollary C.4a (numerical endpoint consistency). Use the proposal system of Proposition C.2a. Let P_k be a common nonexpansive score map with $\sup_v \|P_k(v)\| \le C_k^S$, and define

$$\widehat{S}_k(x) = P_k\left[-\frac{x}{\sigma_{\theta_k}^2} + \frac{a_{\theta_k}}{\sigma_{\theta_k}^2} D_{\nu_k}(x)\right], \qquad S_k^R(x) = P_k\left[-\frac{x}{\sigma_{\theta_k}^2} + \frac{a_{\theta_k}}{\sigma_{\theta_k}^2} D_{\eta_k^M}(x)\right].$$

Couple the Euler updates using the same noise for each paired member. Assume uniformly bounded initial average second moments and

$$\sup_M \frac{1}{J_M}\sum_j \mathbb{E}\|\xi_k^{j,M}\|^2 < \infty \qquad (k \le K).$$

If $\mathbb{E}\|X_0^{I_M} - Y_0^{I_M}\|^2 \to 0$, then

$$\mathbb{E}\|X_K^{I_M} - Y_K^{I_M}\|^2 \to 0, \qquad W_2(\mathcal{L}(X_K^{I_M}), \mathcal{L}(Y_K^{I_M})) \to 0.$$

Here the independent uniform index I_M is kept along the entire trajectory.

Proof. Set $A_k^E = 1 - \delta_k f_{\theta_k}$ and $\alpha_k = \delta_k r_{\theta_k}^2$. The bounded scores and uniform noise moments give a uniform average second-moment bound at each stage, using

$$\mathbb{E}\|X_k^{I_M}\|^2 \le 3|A_k^E|^2\mathbb{E}\|X_{k-1}^{I_M}\|^2 + 3\alpha_k^2(C_k^S)^2 + 3\mathbb{E}\|\xi_k^{I_M}\|^2,$$

and the same inequality for Y. This establishes average query tightness directly from the reverse recursion. Proposition C.2a then gives $\widehat{S}_k(X_{k-1}^{I_M}) - S_k^R(X_{k-1}^{I_M}) \to 0$ in probability and in mean square, since its norm is bounded by $2C_k^S$. The Gaussian derivative and denominator bounds in its proof make S_k^R uniformly locally Lipschitz. Tightness and a previous-stage coupling error tending to zero in probability therefore imply $S_k^R(X_{k-1}^{I_M}) - S_k^R(Y_{k-1}^{I_M}) \to 0$ in probability. The full score difference is bounded by $2C_k^S$, so it converges in mean square. The common noise cancels, leaving

$$\|X_k^{I_M} - Y_k^{I_M}\|_{L^2} \le |A_k^E|\|X_{k-1}^{I_M} - Y_{k-1}^{I_M}\|_{L^2} + |\alpha_k|\|\widehat{S}_k(X_{k-1}^{I_M}) - S_k^R(Y_{k-1}^{I_M})\|_{L^2}.$$

Finite-stage induction proves the endpoint result; this coupling also bounds W_2. A common nonexpansive output map preserves the bound.

Proposition C.4b (endpoint consistency with raw scores). Use the corrected proposal recursion of Proposition C.1a, retaining its initial-measure, potential and one-member kernel conditions and the residual-likelihood conditions of Proposition C.2a. Suppose the initial reverse queries and the Euler noise at each stage are uniformly average-tight. Couple the actual and numerical reference chains with the same Euler noise, using their raw conditional scores. For an independent uniform member index I_M, kept along the trajectory, if $X_0^{I_M} - Y_0^{I_M} \to 0$ in probability, then

$$X_k^{I_M} - Y_k^{I_M} \longrightarrow 0 \quad \text{in probability} \qquad (k \le K).$$

The reverse queries remain average-tight at every stage, and the endpoint laws approach each other in bounded-Lipschitz distance.

Proof. The reference moment recursion in Proposition C.1a depends only on the initial proposals, potentials and move kernels. It gives tight reference measures before considering the reverse queries. The Gaussian envelopes and positive denominator bound in Proposition C.2a then make the reference scores $s_k^{R,M}$ uniformly bounded and Lipschitz on every fixed query ball, with constants uniform in M.

Proceed jointly through the proposal and reverse stages. Tight initial queries and the initial coupling give tight reference queries. Suppose both chains are tight through stage $k-1$ and their coupled difference tends to zero in probability. The localization proof of Proposition C.1a applies to the proposal prefix through stage k, whose input queries are all tight by induction. Proposition C.2a therefore gives

$$e_k^M := \widehat{s}_k(X_{k-1}^{I_M}) - s_k^{R,M}(X_{k-1}^{I_M}) \longrightarrow 0 \quad \text{in probability.}$$

Local boundedness of the reference score and tightness of the query imply tightness of this reference score evaluation. Adding e_k^M gives tightness of the actual score evaluation. The Euler update and tight noise thus give tight queries at stage k in both chains.

Set $A_k = 1 - \delta_k f_{\theta_k}$. On the event that both previous queries lie in a ball of radius R, the common noise cancels and

$$\begin{aligned}\|X_k^{I_M} - Y_k^{I_M}\| \le (|A_k| + |\alpha_k|L_{k,R})\|X_{k-1}^{I_M} - Y_{k-1}^{I_M}\| \\ + |\alpha_k|\|e_k^M\|,\end{aligned}$$

where $L_{k,R}$ is a uniform local Lipschitz constant of $s_k^{R,M}$. The right side tends to zero in probability for fixed R. The probability outside this event tends uniformly to zero as $R \to \infty$, completing the finite-stage induction. For any function with sup norm and Lipschitz constant at most one,

$$|\mathbb{E}f(X_K^{I_M}) - \mathbb{E}f(Y_K^{I_M})| \le \mathbb{E}\min\{2, \|X_K^{I_M} - Y_K^{I_M}\|\} \to 0.$$

This proves the endpoint law comparison.

The reported recursive experiments use $M = J$, with one proposal initialized at each forecast member. Along $M = J \to \infty$, Proposition C.3 controls the terminal-member marginal and Propositions C.1a–C.2a control adaptive proposal integration.

Reference bias relative to the target. Let $f_{r:k} = T_{r+1} \cdots T_k f$. The numerical proposal reference obeys the exact identity

$$\Gamma_k(f) - Z_k \pi_k(f) = (\eta_0 - p)(T_1 \cdots T_k f) + \sum_{r=1}^{k} Z_r \pi_r[(Q_r - I) f_{r:k}].$$

It follows by iterating $\Gamma_k - \gamma_k^* = (\Gamma_{k-1} - \gamma_{k-1}^*)T_k + \gamma_k^*(Q_k - I)$, where $\gamma_k^* = Z_k \pi_k$. Put $D_k^*(x) = \pi_k(zF_x)/\pi_k(F_x)$ and $\varphi_x = (z - D_k^*(x))F_x$. Then

$$D_{\eta_k}(x) - D_k^*(x) = \frac{(\eta_0 - p)T_1 \cdots T_k \varphi_x + \sum_{r=1}^{k} Z_r \pi_r[(Q_r - I)T_{r+1} \cdots T_k \varphi_x]}{\Gamma_k(F_x)}.$$

If the displayed backward test functions are Lipschitz, define $L_0 = \mathrm{Lip}(T_1 \cdots T_k \varphi_x)$, $L_r = \mathrm{Lip}(T_{r+1} \cdots T_k \varphi_x)$, and $\Delta_r = W_1(\pi_r Q_r, \pi_r)$. The preceding identity gives

$$\|D_{\eta_k}(x) - D_k^*(x)\| \le \frac{L_0 W_1(\eta_0, p) + \sum_{r=1}^{k} Z_r L_r \Delta_r}{\Gamma_k(F_x)}.$$

Thus the target discrepancy separates the initial forecast representation from the invariance defect of each proposal transition.

For an unprocessed ULA step, write $\widetilde{P}_h(x, \cdot) = \mathcal{L}(x + h\widetilde{b}(x) + \sqrt{h}\,\xi)$, where $\widetilde{b}$ is its drift approximation, and let P_h use the ideal drift $b = \frac{1}{2}\nabla \log \pi$. A synchronous coupling gives

$$W_1(\pi \widetilde{P}_h, \pi) \le h\, \mathbb{E}_\pi \|\widetilde{b} - b\| + W_1(\pi P_h, \pi).$$

If b is globally L_b-Lipschitz, π is stationary for the continuous Langevin diffusion, and $B_b = \mathbb{E}_\pi \|b\| < \infty$, then

$$W_1(\pi P_h, \pi) \le L_b \left(\tfrac{1}{2} B_b h^2 + \tfrac{2}{3}\sqrt{d}\, h^{3/2} \right).$$

The first term measures drift approximation and the second ULA discretization. For several steps, the corresponding one-step defects propagate through the Lipschitz constants of the intervening kernels. These target-bias terms, the reverse-time discretization, and the retained endpoint noise do not generally vanish when only the proposal count grows.

C.5 An exactly computable Gaussian proposal reference

For a Gaussian fitted forecast and linear Gaussian observations, the numerical proposal reference has a finite Gaussian-mixture representation. This gives a computable example of the preceding reference-bias decomposition. Here we use the ordinary linear ULA transition; this example is separate from the compact-support assumptions of Proposition C.2.

Let $p = \mathcal{N}(m_f, C_f)$, $C_f \succ 0$, and let $y = Lz + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, R)$ and $R \succ 0$. Using the execution index k defined above, start from $\eta_0 = \sum_{m=1}^{M} u_{0,m} \delta_{z_m}$ and write

$$\eta_{k-1} = \sum_{m=1}^{M} u_{k-1,m} \mathcal{N}(b_{k-1,m}, C_{k-1,m}),$$

where zero covariance denotes a point mass. For $\Delta\varrho_k = \varrho_k - \varrho_{k-1} > 0$, set $V_k^{\mathrm{obs}} = R/\Delta\varrho_k$ and define

$$S_{k,m} = L C_{k-1,m} L^T + V_k^{\mathrm{obs}}, \qquad K_{k,m} = C_{k-1,m} L^T S_{k,m}^{-1}.$$

Incremental likelihood weighting gives

$$\bar{b}_{k,m} = b_{k-1,m} + K_{k,m}(y - L b_{k-1,m}), \qquad \bar{C}_{k,m} = C_{k-1,m} - K_{k,m} L C_{k-1,m},$$
$$\bar{u}_{k,m} \propto u_{k-1,m} \phi_{S_{k,m}}(y - L b_{k-1,m}).$$

The omitted likelihood normalization factor is common to all components and cancels when their masses are normalized. When $\Delta\varrho_k = 0$, weighting is the identity.

For a ULA step of size $h > 0$, define

$$\mathcal{H}_k = C_f^{-1} + \varrho_k L^T R^{-1} L, \qquad v_k = C_f^{-1} m_f + \varrho_k L^T R^{-1} y, \qquad A_k = I - \tfrac{h}{2}\mathcal{H}_k.$$

Each step maps a component's moments as

$$b \longmapsto A_k b + \tfrac{h}{2} v_k, \qquad C \longmapsto A_k C A_k^T + hI,$$

with unchanged component mass. Applying this map for the scheduled number of Langevin steps computes η_k with M components at every stage; if movement is omitted, the weighted mixture is retained. At the initial score evaluation, both weighting and movement are identities. These formulas follow by completing the square for likelihood weighting and applying an affine Gaussian transition.

For the tempered target $\pi_k = \mathcal{N}(\mathcal{H}_k^{-1} v_k, \mathcal{H}_k^{-1})$, one ULA step preserves its mean and increases its covariance by $h^2 \mathcal{H}_k / 4$. After ℓ steps from π_k, the covariance excess is

$$\frac{h^2}{4} \sum_{r=0}^{\ell-1} A_k^r \mathcal{H}_k (A_k^r)^T.$$

If $0 < h\lambda_{\max}(\mathcal{H}_k) < 4$, the stationary covariance is $[\mathcal{H}_k(I - h\mathcal{H}_k/4)]^{-1}$. These identities quantify the transition defect at a fixed step size.

To compute $D_{\eta_k}(x)$, update each component first with the remaining likelihood $g^{1-\varrho_k}$ and then with the Gaussian observation $x = a_{\theta_k} z + \sigma_{\theta_k} \xi$. The first update uses observation covariance $R/(1-\varrho_k)$ when $\varrho_k < 1$ and is omitted when $\varrho_k = 1$. The second uses observation matrix $a_{\theta_k} I$ and covariance $\sigma_{\theta_k}^2 I$. Normalize the component masses after multiplying by both marginal likelihoods and average their conditional means. The corresponding reference score is

$$s_k^R(x) = \frac{a_{\theta_k} D_{\eta_k}(x) - x}{\sigma_{\theta_k}^2}.$$

Thus the proposal reference and its conditional score can be evaluated without sampling additional proposals.

D Experimental protocols and supplementary analyses

D.1 Controlled Gaussian evaluation

The data-generating prior is $Z \sim \mathcal{N}(0, \mathbf{I}_{128})$, and the observation is $Y = Z + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, 0.25\mathbf{I}_{128})$. For each of ten observation problems (seeds 10–19), an independent forecast ensemble of 20 prior samples defines the common diagonal Gaussian approximation $\widetilde{\pi}^f = \mathcal{N}(m_f, \mathrm{diag}(v_f))$. The implementation regularizes its variances with jitter 10^{-5} and a variance floor of 10^{-3}. The observation, forecast ensemble, fitted approximation, and reference samples are fixed across the four sampling repetitions and all methods within a problem. Repetitions vary algorithmic random streams; the ten observation problems are the units of statistical aggregation.

The fitted posterior has componentwise variance $v_a = (v_f^{-1} + 4)^{-1}$ and mean $m_a = v_a(m_f/v_f + 4y)$. The data-generating posterior has mean $0.8y$ and variance $0.2\mathbf{I}$. Each endpoint reference includes the retained noising variance $\epsilon_b = 0.025$. We draw 4096 independent samples from each reference per problem. All methods start reverse diffusion from the same fitted noisy-posterior distribution at $\tau = 1$. This common analytic initialization isolates the subsequent sampling procedure; recursive filtering instead uses the standardized Gaussian initialization described in the main text. The diffusion schedule uses $\epsilon_a = 0.001$, 100 steps uniform in logSNR, and the actual nonuniform time increments.

AECSF and movement-only use the same fitted prior gradient, 294 Langevin steps with $h = 0.02$, and the same separate streams for reverse diffusion and proposal movement. Movement-only retains incremental proposal weights but omits resampling. The other controls use the same fitted Gaussian information. Writing $b_\tau = \sigma_\tau^2$ and interpreting vector products componentwise, define

$$\begin{aligned}
s_\tau^f(x) &= \frac{a_\tau m_f - x}{a_\tau^2 v_f + b_\tau}, \\
D_\tau^f(x) &= m_f + \frac{a_\tau v_f}{a_\tau^2 v_f + b_\tau}(x - a_\tau m_f), \\
s_\tau^{\mathrm{point}}(x; y) &= s_\tau^f(x) + (1-\tau)\nabla \log g(y \,|\, x), \\
s_\tau^{\mathrm{denoise}}(x; y) &= s_\tau^f(x) + J_{D_\tau^f}(x)^\top \nabla \log g(y \,|\, D_\tau^f(x)).
\end{aligned} \tag{69}$$

Here $D_\tau^f(x)$ is the conditional mean of the system state given $X_\tau = x$ under the fitted forecast, without conditioning on the current observation. Its Jacobian $J_{D_\tau^f}$ is diagonal with entries $a_\tau v_f/(a_\tau^2 v_f + b_\tau)$. The analytic conditional-score control uses the exact score of the noised posterior under the fitted prior $(a_\tau m_a - x)/(a_\tau^2 v_a + b_\tau)$, integrated with the same numerical reverse solver and finite output ensemble.

Endpoint energy statistics. Let X_{rm} be output member m of repetition r, where $R = 4$, $M = 20$, and $n = RM = 80$; write X_p for the pooled samples. In this subsection, $M = J$ counts output samples and n is the pooled sample count, not assimilation time. Let Y_k, $k = 1, \dots, K$, be the reference samples, with $K = 4096$, and set $d(u, v) = \|u - v\|/\sqrt{128}$. The two reported discrepancies are

$$E_V = \frac{2}{nK}\sum_{p,k} d(X_p, Y_k) - \frac{1}{n^2}\sum_{p,q} d(X_p, X_q) - \frac{1}{K^2}\sum_{k,l} d(Y_k, Y_l), \tag{70}$$

$$E_{\text{cross}} = \frac{2}{nK}\sum_{p,k} d(X_p, Y_k) - \frac{\sum_{r\neq s}\sum_{m,l} d(X_{rm}, X_{sl})}{R(R-1)M^2} - \frac{\sum_{k\neq l} d(Y_k, Y_l)}{K(K-1)}. \tag{71}$$

The V-statistic evaluates the pooled empirical distribution of 80 particles. The cross-repetition statistic pairs particles from different runs and excludes reference self-pairs. Conditional on each problem, independent repetitions and reference samples make it an estimator of energy discrepancy for the marginal output distribution averaged over member indices. Negative estimates are retained. Figures show means and sample standard deviations across ten problems, with individual problem values overlaid.

Table 2: Gaussian endpoint summaries across ten problems. Each problem uses four repetitions of 20 particles. The two energy statistics are defined in Eqs. (70)–(71).

Method	Fitted E_V	True E_V	Fitted E_{cross}	True E_{cross}	Seconds/run
Analytic conditional score (fitted prior)	0.008850	0.018799	0.000284	0.010230	0.069935
AECSF	0.019040	0.035886	0.009168	0.026011	0.300461
Movement-only	0.070636	0.090871	0.011818	0.032051	0.216578
Pointwise guidance	0.038129	0.044543	0.031485	0.037897	0.062395
Denoised guidance	0.098303	0.081641	0.091629	0.074965	0.066121

To evaluate individual 20-member outputs, let $\bar{X}_r = M^{-1}\sum_m X_{rm}$ and $\widehat{v}_{r,q} = (M-1)^{-1}\sum_m (X_{rm,q} - \bar{X}_{r,q})^2$. For each repetition, we compute

$$e_{\mu,r} = d^{-1}\|\bar{X}_r - \mu_\star\|^2, \qquad e_{v,r} = d^{-1}\|\widehat{v}_r - v_\star\|^2,$$

using the analytic mean and diagonal variance of the corresponding retained-noise reference. We first average the four repetitions within each problem, then summarize the ten problems. The diagonal-variance error includes finite-ensemble sampling variability.

Table 3: Mean and diagonal-variance errors of individual 20-member Gaussian outputs. Each entry averages four repetitions within each of ten problems and then averages across problems. Both references include the retained endpoint noise.

Method	Fitted mean MSE	Fitted variance MSE	True mean MSE	True variance MSE
Analytic conditional score (fitted prior)	0.011516	0.005614	0.018309	0.005799
AECSF	0.015528	0.008179	0.025495	0.009191
Movement-only	0.093729	0.028066	0.105807	0.029772
Pointwise guidance	0.020602	0.009073	0.024058	0.010208
Denoised guidance	0.058519	0.009065	0.047680	0.009776

D.2 Evaluation metrics and aggregation

State RMSE is the primary accuracy measure and mean componentwise CRPS measures probabilistic accuracy. Normalized observation misfit is reported for Lorenz–96. SSR compares ensemble

spread with the error of the ensemble mean. Computational cost is reported as analysis time per assimilation cycle.

The per-cycle RMSE, CRPS, and normalized observation misfit are

$$\mathrm{RMSE}_n = \left(\frac{1}{d} \left\| \bar{x}_n^a - x_n^{\mathrm{true}} \right\|^2 \right)^{1/2}, \tag{72}$$

$$\mathrm{CRPS}_n = \frac{1}{Jd} \sum_{j=1}^{J} \left\| x_n^{a,j} - x_n^{\mathrm{true}} \right\|_1 - \frac{1}{2J^2 d} \sum_{j=1}^{J} \sum_{k=1}^{J} \left\| x_n^{a,j} - x_n^{a,k} \right\|_1, \tag{73}$$

$$\mathrm{NOM}_n = \frac{1}{\sigma_y} \left(\frac{1}{m_n} \left\| \mathcal{H}_n(\bar{x}_n^a) - y_n \right\|^2 \right)^{1/2}. \tag{74}$$

The CRPS expression in (73) applies ensemble CRPS to each state component and then averages over components (Gneiting & Raftery, 2007). Here, σ_y is the observation-noise standard deviation in the reported homoscedastic protocols. Normalized observation misfit measures agreement between the analysis mean and the assimilated observation in units of observation noise. A value of one provides a reference noise scale, not an exact calibration criterion.

For each seed, RMSE and CRPS are averaged over the evaluation window $\mathcal{W}$: cycles 50–199 for Lorenz–96 and 950–999 for Kuramoto–Sivashinsky. Define the ensemble variance $v_{n,k} = J^{-1} \sum_j (x_{n,k}^{a,j} - \bar{x}_{n,k}^a)^2$ and ensemble-mean error $e_{n,k} = \bar{x}_{n,k}^a - x_{n,k}^{\mathrm{true}}$. The spread–skill ratio pools their second moments over the evaluation window:

$$\mathrm{SSR} = \sqrt{\frac{\sum_{n \in \mathcal{W}} \sum_k v_{n,k}}{\sum_{n \in \mathcal{W}} \sum_k e_{n,k}^2}}.$$

The tables report the mean and sample standard deviation of the seed-level summaries. All prescribed seeds are included in each reported comparison.

The recursive benchmarks load identical saved input bundles across methods within each evaluation seed. AECSF uses independent deterministic random streams for resampling, Langevin movement, and reverse diffusion.

D.3 Shared implementation settings and timing

Table 4: AECSF configurations. M counts proposal members and J counts reverse particles; the forecast ensemble has J members. The main benchmark configurations use $M = J$; Table 8 varies M at fixed J. All configurations use stratified resampling and unit forecast and likelihood gains.

Setting	Gaussian	Lorenz–96	Kuramoto–Sivashinsky
M/J	20/20	20/20	J/J
Reverse steps	100	100	T
Weight updates / resampling / move calls	99/50/98	99/50/98	$(T-1)/(T/2)/(T-2)$
Langevin steps per move	3	3	3
Langevin step size h	0.02	0.006	0.002
ϵ_a	0.001	0.0005	0.001
ϵ_b	0.025	0.0125	0.015

Here the forecast ensemble has $J = M$ members. For coordinate q, the diagonal forecast approximation uses $m_q = M^{-1} \sum_m X_{mq}^f$ and

$$v_q = \max\left\{ \max\left[\frac{1}{M} \sum_m (X_{mq}^f - m_q)^2, 10^{-12} \right] + 10^{-5}, 10^{-3} \right\}.$$

The fitted mean and variance remain fixed throughout the analysis update.

The grid is uniform in logSNR. For T score evaluations, the first evaluation at $\tau = 1$ uses the initial proposal ensemble. Numbering the remaining $T-1$ stages from zero, resampling occurs at even-numbered stages and movement at stages $0, \ldots, T-3$. For the even values of T used here, each

analysis makes $T/2$ resampling calls and $T-2$ move calls, with $3(T-2)$ Langevin steps. Weight updates continue when resampling is omitted; the final score stage resamples but does not move. Gaussian and Lorenz–96 use $T = 100$; the Kuramoto–Sivashinsky values of T and J are given in Appendix D.5.

Analysis timings exclude forecast propagation, metric evaluation, and file output, with CUDA synchronization for GPU operations. The Kuramoto–Sivashinsky timings include per-cycle preparation and sampling from device-resident forecasts and observations; one-time model construction is recorded separately. Lorenz–96 times the sampling update, while the controlled Gaussian timings also include per-run model construction. AECSF, EnSF, EnFF, IEnSF, LETKF, and Binder use RTX 4090 GPU analysis implementations.

D.4 Lorenz–96 nonlinear filtering

The Lorenz–96 system provides a scalable chaotic benchmark for recursive data assimilation (Lorenz, 1996):

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \qquad i = 1, \ldots, d, \tag{75}$$

with periodic indexing and forcing $F = 8$. The observation operator is nonlinear and saturating:

$$y_n = \arctan(Px_n) + \varepsilon_n, \qquad \varepsilon_n \sim \mathcal{N}(0, R), \tag{76}$$

where P selects observed components. This setting tests whether the method can assimilate observations that become weakly informative when state magnitudes are large. The integration time step is 0.01, with observations every five steps, corresponding to 0.05 physical time units. AECSF propagates its forecast ensemble with the classical fourth-order Runge–Kutta method in double precision.

Methods and configuration selection. AECSF, EnSF, LETKF, and the method of Binder et al. share physical inputs and the initial ensemble on seeds 10–19. Subsequent forecasts depend on each method's recursive analyses. The EnSF and LETKF configurations were first selected from complete 10×10 grids at $d = 1000$ on seeds 0–9, ranked by mean post-burn-in RMSE. We additionally checked five nearby EnSF configurations at $d = 10{,}000$ on development seeds 0–9. The original configuration, $(\epsilon_a, \epsilon_b) = (0.6, 0.025)$, retained the lowest mean RMSE. LETKF uses cyclic localization radius 2, neighbor size 5, and inflation 1.1. The development results appear below.

AECSF uses $h = 0.006$ and three Langevin steps per scheduled move, with the weighted update and logSNR grid specified in Table 4. The $d = 10{,}000$ experiment has 200 cycles, full arctangent observations, noise standard deviation 0.05, and evaluation cycles 50–199.

Component controls. All five variants use the same prior approximation, diffusion grid, and observation model. Movement-only retains persistent target weights and all Langevin steps but omits resampling. Resampling-only omits the moves. Fixed proposal positions omits both resampling and movement but still advances weights; its integration coefficients are proportional to $K_i g_n$ after combining the accumulated and residual likelihood factors. Target-weight resampling replaces s_i with ω_i at the same resampling times; its ancestor correction factors are one. The comparison with AECSF tests the combined contribution of reverse-particle-guided selection and its associated importance correction. SSR is closer to one for AECSF in 10 of the ten seeds than for target-weight resampling, whereas normalized observation misfit is closer to one for the target-weight control (means 0.9497 versus 0.9209).

Direct posterior sampling control. The direct posterior ULA control uses 20 forecast members, 294 steps with step size 0.006, and the full posterior drift at every step, without resampling or incremental weights. Its physical inputs are paired with AECSF on seeds 10–19. AECSF has lower RMSE in all ten comparisons and lower CRPS in eight. The step size is set to match AECSF's.

We calibrated direct ULA runtime at 294, 588, and 882 steps, then set a budget of 1,350 steps to approximately match AECSF's analysis time. Development runs on seeds 0–9 tested step sizes 0.003, 0.006, and 0.012 at both 294 and 1,350 steps. At 1,350 steps, the lowest mean development RMSE occurred at step size 0.012. The evaluation compares this configuration and AECSF using common truth, observations, and initial ensembles. Each method then propagated its own analysis

Table 5: Matched-step Lorenz–96 comparison with 20 members at dimension 10,000. Both methods use 294 Langevin steps per analysis with step size 0.006. Entries are means and sample standard deviations across seeds 10–19, using cycles 50–199 within each seed. Timing brackets the sampling update and excludes model construction and final analysis processing.

Method	RMSE ↓	CRPS ↓	SSR → 1	Norm. obs. misfit	Sampling-update time (s/cycle)
AECSF	0.154482 ± 0.003892	0.077631 ± 0.000740	0.899823 ± 0.023152	0.920890 ± 0.004196	0.305238 ± 0.013899
Direct posterior ULA	0.184311 ± 0.007514	0.078474 ± 0.001215	0.596951 ± 0.023914	1.042469 ± 0.013634	0.136624 ± 0.003211

ensemble. On the same RTX 4090, timing covered one complete reverse run for AECSF or all 1,350 ULA steps per cycle. CUDA synchronization bracketed each call; forecast propagation and fitting of the forecast density were outside the timed interval.

Table 6: Approximately time-matched posterior sampling on Lorenz–96 with 20 members in 10,000 dimensions. AECSF uses 294 Langevin steps; direct ULA uses 1,350 steps with step size 0.012. Entries are means and sample standard deviations across ten seeds, using cycles 50–199 within each seed.

Method	RMSE ↓	CRPS ↓	SSR → 1	Norm. obs. misfit	Update time (s/cycle)
AECSF	0.154968 ± 0.004083	0.077842 ± 0.000623	0.896359 ± 0.024766	0.921380 ± 0.005849	0.289720 ± 0.001397
Direct posterior ULA	0.171986 ± 0.001801	0.094169 ± 0.000456	1.116732 ± 0.010192	1.149319 ± 0.002589	0.296331 ± 0.000446

Sensitivity to proposal movement. After selecting the main configuration, we evaluated sensitivity to the allocation and step size of Langevin moves using five configurations in the 10,000-dimensional Lorenz–96 benchmark with 20 members on seeds 10–19, keeping the remaining settings fixed. Every configuration uses 294 Langevin steps across 98 movement stages. Besides three steps at every stage, we assign four steps to the first 49 stages and two to the last 49, or reverse this allocation. We also vary the step size by 25% around 0.006 while retaining three steps per stage. The main configuration has the lowest mean RMSE among these five settings; allocating more steps later lowers CRPS in all ten paired seeds, while its mean RMSE increases by 0.20% and its SSR moves further below one. Relative to the main configuration, the largest changes in mean RMSE and CRPS are approximately 1.44% and 2.56%, respectively (Table 7).

Table 7: Local movement sensitivity on Lorenz–96. Entries are means across seeds 10–19 over cycles 50–199. Every setting uses 294 Langevin steps; varying the step size also changes the total Langevin integration time.

Steps in first/last 49 stages	h	RMSE ↓	CRPS ↓	SSR → 1
3/3	0.0060	0.154482	0.077631	0.899823
4/2	0.0060	0.156707	0.079617	0.939406
2/4	0.0060	0.154794	0.077192	0.872296
3/3	0.0045	0.154682	0.078449	0.931029
3/3	0.0075	0.155021	0.077493	0.879071

Proposal capacity at fixed output size. We vary the shared proposal ensemble size in a 1,000-dimensional Lorenz–96 experiment, keeping the forecast ensemble, reverse particles, and output ensemble at 20 members. We use the observation model and remaining settings of the Lorenz–96 benchmark. The proposals are initialized by evenly replicating the forecast members with uniform weights and then updated using the same resampling and movement schedule. Each proposal receives 294 Langevin steps across 98 stages. Table 8 reports ten-seed results. All larger proposal ensembles lower mean RMSE and CRPS relative to $M = 20$, although the improvement is not monotonic in M. At $M = 400$, mean RMSE and CRPS decrease by 4.08% and 2.82%, respectively, with a 1.19% increase in measured sampling time; SSR changes from 0.8935 to 0.9598. Increasing proposal capacity thus improves the estimates while retaining a 20-member output ensemble.

Table 8: Proposal capacity in Lorenz–96 with $d = 1000$ and $J = 20$. Entries are means over seeds 10–19, each summarized over cycles 50–199. All configurations run on one RTX 4090. The per-proposal Langevin budget is fixed, so total gradient work increases with M. Time measures the sampling update.

M	M/J	RMSE ↓	CRPS ↓	SSR → 1	s/update
20	1	0.153809	0.079206	0.8935	0.2995
40	2	0.148537	0.077347	0.9335	0.3018
80	4	0.148571	0.077442	0.9395	0.2998
160	8	0.148918	0.077408	0.9428	0.2969
240	12	0.150529	0.077595	0.9352	0.2985
320	16	0.147748	0.077230	0.9571	0.2986
400	20	0.147535	0.076975	0.9598	0.3031

D.4.1 TUNING GRIDS

Figures 4 and 5 show the baseline Lorenz–96 tuning grids at $d = 1000$. Cells aggregate post-burn-in RMSE over seeds 0–9. Red rectangles mark the three lowest-RMSE configurations, with a thicker rectangle for rank 1.

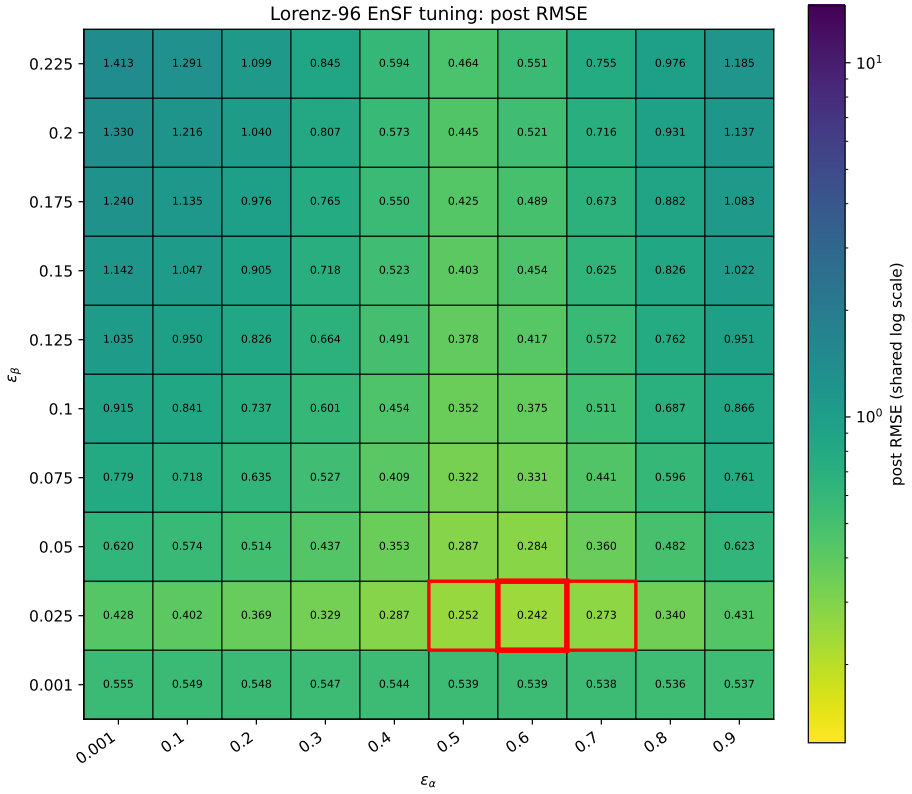


Figure 4: Lorenz–96 EnSF $d = 1000$ tuning grid over seeds $0, \ldots, 9$. The rank-1 configuration transferred to Table 1 is $\epsilon_a = 0.6, \epsilon_b = 0.025$.

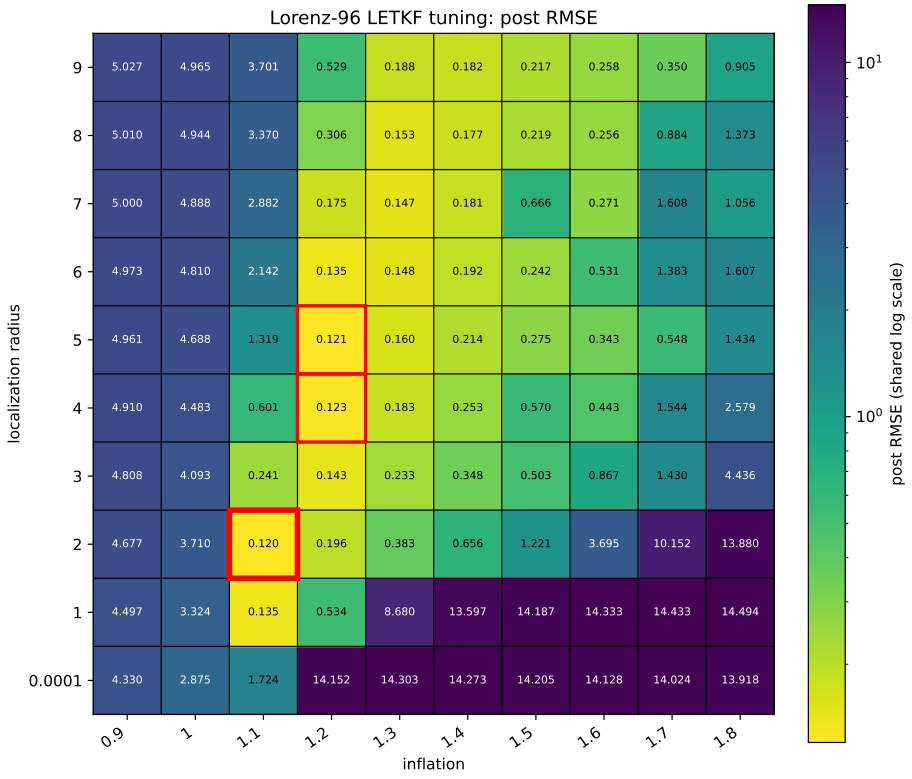


Figure 5: Lorenz–96 LETKF $d = 1000$ tuning grid over seeds $0, \ldots, 9$. The rank-1 configuration transferred to Table 1 uses localization radius 2, neighbor size 5, and inflation 1.1.

Binder reproduction and transfer protocol. We implemented the conditional score from the joint KDE of Binder et al. (2026) and checked its numerical evaluation against analytic references. Under the paper's protocol with $d = 10$, $M = 20$, and $(h_x, h_y) = (0.2, 1)$, one 500-cycle run yielded RMSE 3.0198 without ensemble collapse.

For the common benchmark in 10,000 dimensions, bandwidths $(h_x, h_y) = (0.2, 33.5)$ were selected during development and fixed before evaluation on seeds $10, \ldots, 19$. Forecast states and simulated observations were centered coordinatewise and divided by the maximum absolute centered deviation. Coordinates with zero deviation used unit scale. The implementation uses float64 GPU state updates with the Dormand–Prince RK45 coefficients and adaptive step-size rules of SciPy, at relative and absolute tolerances 10^{-6} and 10^{-8}. Analysis timing includes simulated observations, normalization, integration, and required device transfers. The common benchmark differs from the original protocol in observation noise, initialization, assimilation interval, and process noise. Earlier runs of the common protocol in low dimensions also showed ensemble contraction.

D.5 Kuramoto–Sivashinsky filtering with nonlinear observations

We use the periodic Kuramoto–Sivashinsky equation (Sivashinsky, 1977)

$$u_t = -uu_x - u_{xx} - u_{xxxx}, \qquad x \in [0, 128\pi), \tag{77}$$

discretized on 1,024 spatial coordinates with a Fourier pseudospectral ETDRK4 solver and time step 0.25. The nonlinear term is evaluated by transforming the squared physical-space field back to Fourier space, as in the reference implementation. Observations are

$$y_{n,k} = \arctan(u_{n,k}) + \varepsilon_{n,k}, \qquad \varepsilon_n \sim \mathcal{N}(0, 0.1^2 I_{1024}). \tag{78}$$

Four model steps separate successive observations, giving an assimilation interval of one time unit. The first observation is assimilated at the initial time. The model propagation is deterministic.

Inputs and ensemble initialization. The reference initial profile is evolved for 600 steps with the DAPPER initialization routine. For each seed, independent standard Gaussian noise is added to this state, followed by 2,000 model steps before assimilation begins. Initial ensemble members are obtained by adding independent standard Gaussian perturbations to the truth at this point. The first J members of a saved 160-member ensemble are used for each ensemble-size setting. Methods share the saved truth, observations, and initial ensemble at seeds 30–34.

Compared methods. The Kuramoto–Sivashinsky experiment protocol, together with the EnSF, EnFF-OT, and EnFF-F2P implementations and their parameter settings, follows the public EnFF repository (Transue et al., 2026). We extend this setup with the sampling-step and ensemble-size comparisons specified below. The EnSF path parameters are $(\epsilon_a, \epsilon_b) = (0.5, 0.025)$. EnFF-OT uses $\sigma_{\min} = 0.01$ and guidance strength 1; EnFF-F2P uses $\sigma_{\min} = 0.01$ and guidance strength 0.2. EnFF-F2P receives the preceding analysis ensemble when constructing its probability path. The original four methods operate in float32 on an RTX 4090. Computation is divided into blocks of at most 20 queries in the public implementations.

AECSF uses $(\epsilon_a, \epsilon_b) = (0.001, 0.015)$, Langevin step size $h = 0.002$, three steps per move, Stratified ancestor selection, and a uniform-logSNR grid. At each analysis, AECSF initializes reverse particles from Gaussian noise at $\tau = 1$ and standardizes each coordinate to zero sample mean and unit sample SD. Baselines retain their own terminal-noise initialization.

IEnSF-L1 transfer conventions. We implement IEnSF based on the formulation and implementation details provided by its authors. For each output coordinate, we use an overlapping periodic patch of 13 coordinates and retain its centre score. The forecast covariance, normalized by J, is multiplied entrywise by a Gaspari–Cohn taper with scale 3 and support 6. Writing this covariance as C_{GC} and the local forecast mean as m, mixture components use

$$m_j = m + 0.8\sqrt{1-\gamma^2}(x_j - m), \qquad \Sigma = \gamma^2 C_{\mathrm{GC}}, \qquad \gamma = 0.25.$$

The guidance matrix is $\alpha_t \Sigma(\alpha_t^2 \Sigma + \beta_t^2 I)^{-1}$. The Gaussian reference uses mean m and covariance C_{GC}; it is reinitialized from each cycle's forecast. We use one reverse pass ($L = 1$) and invert the

covariance matrices within each local patch. The patch construction, reference localization, centre factor, and nonlinear covariance blending specify our adaptation to this benchmark.

For mixture likelihood weights, the arctangent observation is linearized at each component's conditional clean-state mean. If $C_{0|t}$ is its conditional covariance and H_j the diagonal Jacobian, we use $R + 0.7H_jC_{0|t}H_j^\top + 0.3\,\mathrm{diag}(H_jC_{0|t}H_j^\top)$. Reverse sampling uses Euler–Maruyama with $\alpha_t = 1 - 0.999t$ and $\beta_t^2 = 0.001 + 0.999t$. There are N_t updates evaluated at $t_k = 1 - k/N_t$ for $k = 0, \ldots, N_t - 1$; the score is not evaluated at zero. The total score is clipped componentwise at 1,000. The terminal samples and diffusion increments use separate deterministic streams. IEnSF-L1 starts from independent standard Gaussian samples, without the coordinatewise sample standardization used by AECSF. All configurations retain the same physical input bundles for seeds 30–34, with the first J members of the saved initial ensemble.

Budgets and evaluation. At $J = 20$, all five methods use step counts $\{20, 50, 100, 200, 500\}$. At 20 steps, ensemble sizes are $\{10, 20, 40, 80, 160\}$ for AECSF, EnSF and EnFF, and $\{10, 20, 40, 80\}$ for IEnSF-L1. We limit the IEnSF-L1 sweep to 80 members to keep its computational cost within the study budget. Its particle–mixture evaluations scale quadratically with ensemble size, and measured analysis time rises from 1.219 to 4.218 seconds when increasing from 40 to 80 members. AECSF uses $M = J$; IEnSF-L1 keeps $L = 1$. The shared 20-member, 20-step configuration is evaluated once per method. The public EnSF/EnFF implementations perform $T-1$ sampling updates; AECSF uses T score evaluations and $3(T-2)$ Langevin steps; IEnSF-L1 uses N_t updates.

Each run uses 1,000 assimilation cycles. Quality metrics use cycles 950–999; analysis time uses all cycles. Tables 9–11 report means and sample SDs across five seeds. Analysis timing follows Appendix D and includes IEnSF-L1's terminal preparation, local covariance construction and reverse sampling.

Table 9: Representative Kuramoto–Sivashinsky results with 20 members. Entries are means ± sample standard deviations across seeds 30–34. Quality metrics use cycles 950–999; analysis time uses all 1,000 cycles. AECSF uses $\epsilon_b = 0.015$. Steps denote T for AECSF, EnSF and EnFF, and N_t for IEnSF-L1.

Method	J	Steps	RMSE	CRPS	Energy	SSR	s/cycle
AECSF	20	20	0.0889 ± 0.0018	0.0625 ± 0.0005	0.0792 ± 0.0007	2.1560 ± 0.0417	0.0988 ± 0.0040
AECSF	20	100	0.0601 ± 0.0040	0.0431 ± 0.0012	0.0545 ± 0.0017	2.2261 ± 0.1376	0.4681 ± 0.0188
EnSF	20	20	0.4466 ± 0.0047	0.2722 ± 0.0027	0.3673 ± 0.0038	1.8207 ± 0.0106	0.0297 ± 0.0006
EnSF	20	100	0.1206 ± 0.0006	0.0684 ± 0.0003	0.0876 ± 0.0004	1.0333 ± 0.0075	0.1480 ± 0.0024
EnFF-F2P	20	20	0.2096 ± 0.0014	0.1569 ± 0.0013	0.2076 ± 0.0014	0.0135 ± 0.0006	0.0237 ± 0.0004
EnFF-F2P	20	100	0.2077 ± 0.0014	0.1557 ± 0.0013	0.2058 ± 0.0014	0.0138 ± 0.0006	0.1147 ± 0.0034
IEnSF-L1	20	20	0.0740 ± 0.0006	0.0665 ± 0.0001	0.0833 ± 0.0002	3.1173 ± 0.0258	0.4519 ± 0.0115
IEnSF-L1	20	100	0.0498 ± 0.0026	0.0345 ± 0.0009	0.0433 ± 0.0011	2.0622 ± 0.1020	2.2554 ± 0.0393
EnFF-OT	20	20	0.1759 ± 0.0012	0.1346 ± 0.0009	0.1732 ± 0.0012	0.0228 ± 0.0008	0.0232 ± 0.0005
EnFF-OT	20	100	0.1237 ± 0.0009	0.0955 ± 0.0008	0.1205 ± 0.0009	0.0388 ± 0.0011	0.1150 ± 0.0033

Energy is reported as a supplementary metric, using

$$d^{-1/2}\left[\frac{1}{J}\sum_j \|x_j - x^\star\|_2 - \frac{1}{2J^2}\sum_{j,k}\|x_j - x_k\|_2\right], \qquad d = 1024.$$

The pair sum includes its zero diagonal entries.

Table 10: Complete Kuramoto–Sivashinsky sampling-budget comparison at $J = 20$. Parameter settings, statistical summaries and step notation follow Table 9.

Method	J	Steps	RMSE	CRPS	Energy	SSR	s/cycle
AECSF	20	20	0.0889 ± 0.0018	0.0625 ± 0.0005	0.0792 ± 0.0007	2.1560 ± 0.0417	0.0988 ± 0.0040
AECSF	20	50	0.0638 ± 0.0016	0.0467 ± 0.0005	0.0591 ± 0.0006	2.3116 ± 0.0612	0.2402 ± 0.0086
AECSF	20	100	0.0601 ± 0.0040	0.0431 ± 0.0012	0.0545 ± 0.0017	2.2261 ± 0.1376	0.4681 ± 0.0188
AECSF	20	200	0.0598 ± 0.0035	0.0418 ± 0.0010	0.0530 ± 0.0015	2.1292 ± 0.1276	0.9723 ± 0.0179
AECSF	20	500	0.0579 ± 0.0027	0.0405 ± 0.0008	0.0514 ± 0.0011	2.1330 ± 0.0982	2.3375 ± 0.0765
EnSF	20	20	0.4466 ± 0.0047	0.2722 ± 0.0027	0.3673 ± 0.0038	1.8207 ± 0.0106	0.0297 ± 0.0006
EnSF	20	50	0.1377 ± 0.0010	0.0869 ± 0.0003	0.1057 ± 0.0004	1.4840 ± 0.0200	0.0738 ± 0.0022
EnSF	20	100	0.1206 ± 0.0006	0.0684 ± 0.0003	0.0876 ± 0.0004	1.0333 ± 0.0075	0.1480 ± 0.0024
EnSF	20	200	0.1165 ± 0.0014	0.0660 ± 0.0008	0.0845 ± 0.0010	0.9588 ± 0.0079	0.2964 ± 0.0054
EnSF	20	500	0.1155 ± 0.0006	0.0655 ± 0.0003	0.0837 ± 0.0005	0.9188 ± 0.0139	0.7333 ± 0.0134
EnFF-F2P	20	20	0.2096 ± 0.0014	0.1569 ± 0.0013	0.2076 ± 0.0014	0.0135 ± 0.0006	0.0237 ± 0.0004
EnFF-F2P	20	50	0.2082 ± 0.0014	0.1560 ± 0.0013	0.2062 ± 0.0014	0.0137 ± 0.0006	0.0578 ± 0.0012
EnFF-F2P	20	100	0.2077 ± 0.0014	0.1557 ± 0.0013	0.2058 ± 0.0014	0.0138 ± 0.0006	0.1147 ± 0.0034
EnFF-F2P	20	200	0.2075 ± 0.0014	0.1556 ± 0.0013	0.2056 ± 0.0014	0.0139 ± 0.0006	0.2299 ± 0.0060
EnFF-F2P	20	500	0.2074 ± 0.0014	0.1555 ± 0.0013	0.2054 ± 0.0014	0.0139 ± 0.0006	0.5785 ± 0.0123
IEnSF-L1	20	20	0.0740 ± 0.0006	0.0665 ± 0.0001	0.0833 ± 0.0002	3.1173 ± 0.0258	0.4519 ± 0.0115
IEnSF-L1	20	50	0.0604 ± 0.0014	0.0453 ± 0.0005	0.0567 ± 0.0005	2.3726 ± 0.0541	1.1257 ± 0.0239
IEnSF-L1	20	100	0.0498 ± 0.0026	0.0345 ± 0.0009	0.0433 ± 0.0011	2.0622 ± 0.1020	2.2554 ± 0.0393
IEnSF-L1	20	200	0.0362 ± 0.0023	0.0250 ± 0.0008	0.0317 ± 0.0010	2.0777 ± 0.1289	4.4251 ± 0.1107
IEnSF-L1	20	500	0.0336 ± 0.0037	0.0205 ± 0.0017	0.0264 ± 0.0020	1.6052 ± 0.1626	11.1443 ± 0.2650
EnFF-OT	20	20	0.1759 ± 0.0012	0.1346 ± 0.0009	0.1732 ± 0.0012	0.0228 ± 0.0008	0.0232 ± 0.0005
EnFF-OT	20	50	0.1338 ± 0.0009	0.1034 ± 0.0008	0.1307 ± 0.0010	0.0340 ± 0.0011	0.0571 ± 0.0004
EnFF-OT	20	100	0.1237 ± 0.0009	0.0955 ± 0.0008	0.1205 ± 0.0009	0.0388 ± 0.0011	0.1150 ± 0.0033
EnFF-OT	20	200	0.1195 ± 0.0009	0.0922 ± 0.0008	0.1162 ± 0.0009	0.0411 ± 0.0012	0.2225 ± 0.0058
EnFF-OT	20	500	0.1173 ± 0.0009	0.0904 ± 0.0008	0.1140 ± 0.0009	0.0425 ± 0.0012	0.5704 ± 0.0182

Table 11: Kuramoto–Sivashinsky ensemble-size comparison at 20 sampling steps. The shared $J = 20$ settings are reported in Table 10. Parameter settings, statistical summaries and step notation follow Table 9.

Method	J	Steps	RMSE	CRPS	Energy	SSR	s/cycle
AECSF	10	20	0.1155 ± 0.0047	0.0738 ± 0.0017	0.0942 ± 0.0025	1.6032 ± 0.0687	0.0992 ± 0.0032
AECSF	40	20	0.0732 ± 0.0012	0.0571 ± 0.0003	0.0722 ± 0.0004	2.6609 ± 0.0429	0.0998 ± 0.0033
AECSF	80	20	0.0650 ± 0.0012	0.0546 ± 0.0003	0.0691 ± 0.0004	3.0259 ± 0.0558	0.1034 ± 0.0033
AECSF	160	20	0.0593 ± 0.0010	0.0531 ± 0.0002	0.0671 ± 0.0003	3.3327 ± 0.0530	0.1038 ± 0.0031
EnSF	10	20	0.4867 ± 0.0056	0.2943 ± 0.0035	0.3982 ± 0.0036	1.6259 ± 0.0182	0.0297 ± 0.0006
EnSF	40	20	0.4259 ± 0.0042	0.2625 ± 0.0030	0.3525 ± 0.0032	1.9366 ± 0.0116	0.0384 ± 0.0007
EnSF	80	20	0.4138 ± 0.0031	0.2559 ± 0.0016	0.3443 ± 0.0023	2.0062 ± 0.0240	0.0543 ± 0.0010
EnSF	160	20	0.4059 ± 0.0039	0.2518 ± 0.0020	0.3397 ± 0.0030	2.0536 ± 0.0177	0.0867 ± 0.0020
EnFF-F2P	10	20	0.2096 ± 0.0014	0.1569 ± 0.0013	0.2077 ± 0.0014	0.0131 ± 0.0006	0.0238 ± 0.0007
EnFF-F2P	40	20	0.2095 ± 0.0014	0.1569 ± 0.0013	0.2076 ± 0.0014	0.0137 ± 0.0007	0.0326 ± 0.0003
EnFF-F2P	80	20	0.2095 ± 0.0014	0.1568 ± 0.0013	0.2075 ± 0.0014	0.0139 ± 0.0007	0.0508 ± 0.0007
EnFF-F2P	160	20	0.2095 ± 0.0014	0.1568 ± 0.0013	0.2075 ± 0.0014	0.0139 ± 0.0007	0.0870 ± 0.0015
IEnSF-L1	10	20	0.0965 ± 0.0008	0.0749 ± 0.0002	0.0938 ± 0.0003	2.3223 ± 0.0183	0.2926 ± 0.0065
IEnSF-L1	40	20	0.0585 ± 0.0002	0.0619 ± 0.0000	0.0777 ± 0.0001	3.9905 ± 0.0171	1.2187 ± 0.0101
IEnSF-L1	80	20	0.0504 ± 0.0004	0.0600 ± 0.0001	0.0752 ± 0.0001	4.6569 ± 0.0339	4.2181 ± 0.0229
EnFF-OT	10	20	0.1759 ± 0.0012	0.1347 ± 0.0009	0.1733 ± 0.0012	0.0222 ± 0.0007	0.0232 ± 0.0005
EnFF-OT	40	20	0.1759 ± 0.0012	0.1345 ± 0.0009	0.1731 ± 0.0012	0.0231 ± 0.0008	0.0314 ± 0.0005
EnFF-OT	80	20	0.1759 ± 0.0012	0.1345 ± 0.0009	0.1731 ± 0.0012	0.0233 ± 0.0008	0.0479 ± 0.0017
EnFF-OT	160	20	0.1759 ± 0.0012	0.1345 ± 0.0009	0.1730 ± 0.0012	0.0234 ± 0.0008	0.0805 ± 0.0013