# FREESIA: COVARIANCE-AWARE POSTERIOR TRANSPORT FOR EXPRESSIVE AND SCALABLE DATA ASSIMILATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Data assimilation aims to infer the state of complex dynamical systems based on observational data. However, accurate inference of the multimodal posteriors induced by nonlinear or non-injective observation operators remains a key challenge under high-dimensional and sparse observation conditions. Ensemble filters scale to high dimensions but are confined by restrictive distributional assumptions, while training-free generative filters (e.g., EnSF, EnFF) alleviate this limitation but may introduce structural errors and hinder information propagation under sparse observations. To address these issues, we propose a training-free, asymptotically exact posterior transport method. Firstly, a covariance-aware posterior transport scheme is designed, which embeds the forecast cross-covariance into flow-based transport and accurately recovers unobserved states while preserving the non-Gaussian posterior structure. Furthermore, the method combines a tractable observation-adaptive proposal with posterior correction, ensuring accurate approximation of the nonlinear posterior distribution. Finally, we establish the corresponding posterior flow theory, from which the asymptotic exactness of the proposed method relative to finite-ensemble surrogates and the Wasserstein error bound are derived. Experiments on Double-Well, Lorenz–96, and Kolmogorov flow show that the proposed method captures complex posterior structure and remains accurate under sparse, nonlinear, and non-injective observations. In the sparse non-injective setting, it reduces RMSE by 56% relative to the best baseline.

## 1 INTRODUCTION

Data assimilation (DA) combines dynamical-model forecasts with observations to estimate the evolving state of a dynamical system and quantify its uncertainty, with important applications in weather forecasting, climate modeling, and other areas of science and engineering (Law et al., 2015). In Bayesian filtering, the state estimate and its uncertainty are represented by the filtering distribution, which characterizes the state conditioned on the available observations and is updated sequentially as new data become available. In high-dimensional nonlinear systems, sparse observations provide only partial information about the state, while nonlinear or non-injective observation operators can produce non-Gaussian or multimodal filtering distributions. Effective filtering must therefore both propagate observational information to unobserved variables and represent complex filtering distributions.

Classical ensemble Kalman methods such as the EnKF and LETKF propagate observational information across the state through forecast covariances, using localization and inflation to remain effective with finite ensembles in high dimensions (Evensen, 2003; Hunt et al., 2007; Gaspari & Cohn, 1999). Iterative and tempered variants further improve robustness to nonlinear observations through repeated updates (Emerick & Reynolds, 2012). However, these methods rely on near-Gaussian approximations of the filtering distribution, which can be restrictive under strongly nonlinear or non-injective observations. By contrast, particle filters can represent non-Gaussian and multimodal filtering distributions, but suffer severe weight degeneracy in high dimensions (Gordon et al., 1993; Snyder et al., 2008).



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.

These limitations have motivated score-based generative approaches for representing complex, high-dimensional, non-Gaussian filtering distributions. Score-based Data Assimilation learns a generative score model of dynamical trajectories and incorporates observations through guidance during assimilation (Rozet & Louppe, 2023), whereas SSLS recursively learns the forecast score at each assimilation step (Ding et al., 2026). Related learned generative filters include flow-based Bayesian filtering, which uses normalizing flows to construct a latent linear-Gaussian state-space model for efficient high-dimensional nonlinear filtering (Wang et al., 2025). These approaches therefore rely on learned generative representations, either trained offline or re-estimated from data. To avoid such training dependence, the Ensemble Score Filter (EnSF) estimates the forecast score directly from the current ensemble and incorporates observational information through likelihood guidance during reverse diffusion (Bao et al., 2024). This training-free score-filtering framework has since been extended to geophysical turbulence, multiphase-flow state estimation, adaptive learning of stochastic PDE solutions, and data-driven dynamical-system forecasting (Bao et al., 2025; Hu et al., 2026; Huynh et al., 2026; Tang et al., 2026).

However, EnSF's heuristic posterior-score construction introduces structural error, while under sparse component-wise observations the likelihood gradient acts directly only on observed variables. Several recent extensions have sought to address these limitations from different directions. IEnSF derives a covariance-aware posterior-score formulation under a Gaussian-mixture prior and iteratively refines its approximation to reduce structural error (Zhang et al., 2027). For partial observations, EnSF with image inpainting reconstructs unobserved states after the score-based update (Liang et al., 2025), while EnSF-LR transfers observed-state analysis increments to unobserved variables through prior-covariance linear regression (Xiong et al., 2026). Latent-EnSF instead addresses sparse observations through learned coupled latent representations, but requires pretrained encoder–decoder models and performs assimilation in the learned latent space (Si & Chen, 2025).

Beyond score-based diffusion, flow matching provides an ODE-based alternative for generative transport (Lipman et al., 2023; Tong et al., 2024; Feng et al., 2025). Its flexible design of probability paths and couplings can yield straighter trajectories, allowing accurate sampling with fewer integration steps. The Ensemble Flow Filter (EnFF) adapts this framework to training-free data assimilation and introduces a filtering-to-predictive flow that exploits the sequential Bayesian structure of DA, improving sampling efficiency and robustness (Transue et al., 2025). EnFF incorporates observations through a guidance mechanism, but its practical localized guidance simplifies the covariance-dependent matrix structure to a tuned scalar multiple of the identity. This simplification limits the explicit use of cross-covariances for propagating sparse observational information to unobserved variables.

Taken together, these developments motivate a covariance-aware, training-free posterior-flow formulation for nonlinear data assimilation. We therefore propose FREESIA (Flow-based Reweighted Ensemble Endpoint Sampling for Inference and Assimilation), which constructs the posterior flow directly from the current forecast ensemble. FREESIA embeds forecast cross-covariances into the transport, allowing sparse observational information to propagate to unobserved variables while retaining non-Gaussian posterior structure. To realize this construction, the forecast distribution is represented by a Gaussian mixture model (GMM) surrogate whose covariance structure is retained in the conditional endpoint distribution. For nonlinear observation operators, local linearization is used only to construct tractable endpoint proposals, while likelihood-ratio reweighting preserves the intended nonlinear Bayesian target under this surrogate.

Our main contributions are:

- We develop covariance-aware posterior transport for training-free nonlinear data assimilation, embedding forecast cross-covariances into the transport to enable accurate recovery of unobserved states while retaining non-Gaussian posterior structure.
- We introduce a proposal-and-correction strategy that combines tractable observation-adaptive proposals with posterior correction, maintaining fidelity to the nonlinear posterior target.
- We establish theoretical guarantees for the resulting posterior flow, proving asymptotic exactness with respect to the finite-ensemble surrogate and deriving a one-step Wasserstein error bound that separates finite-sampling, flow-discretization, and forecast-surrogate errors.

We evaluate FREESIA on Double-Well, Lorenz–96, and Kolmogorov flow, covering multimodal posterior approximation, sparse information transfer, and non-injective observations. We begin in Section 2 by formulating the sequential Bayesian filtering problem and the finite-ensemble surrogate underlying FREESIA. Section 3 then develops the covariance-aware posterior transport and its proposal-and-correction construction, followed by the theoretical analysis in Section 4. We evaluate the resulting method on increasingly challenging nonlinear data-assimilation problems in Section 5 and conclude in Section 6.

## 2 Problem Formulation

This section defines the sequential Bayesian filtering problem and the finite-ensemble surrogate targeted by FREESIA. We retain only the notation needed to distinguish the underlying forecast/analysis laws from the ensemble-based surrogate used to construct the posterior flow in Section 3.

At assimilation times $k = 1, \dots, K$, the latent states $X_k \in \mathbb{R}^d$ and observations $Y_k \in \mathbb{R}^{m_k}$ satisfy

$$X_k = \psi_{k-1}(X_{k-1}) + \omega_k, \qquad Y_k = h_k(X_k) + \eta_k, \tag{2.1}$$

where $X_0 \sim p_0$, $\psi_{k-1}$ is the dynamical model, and $h_k$ is a potentially nonlinear observation operator. The model-error law is left general, whereas the additive observation error satisfies

$$\eta_k \sim \mathcal{N}(0, R_k), \qquad R_k \succ 0. \tag{2.2}$$

The observation noises are assumed independent across assimilation cycles and independent of the state and model-noise processes. Let $p_k(x \mid x') := p(X_k = x \mid X_{k-1} = x')$ denote the Markov transition density, and let $y_{1:k} = (y_1, \dots, y_k)$ denote the realized observations.

Given $y_{1:k-1}$, the forecast density is

$$p_k^f(x) := p(x \mid y_{1:k-1}) = \int p_k(x \mid x') p_{k-1}^a(x')\, dx'. \tag{2.3}$$

Given $y_k$, Bayes' rule gives the analysis density

$$p_k^a(x) := p(x \mid y_{1:k}) = \frac{\ell_k(x; y_k) p_k^f(x)}{\mathcal{Z}_k(y_k)}, \qquad \mathcal{Z}_k(y_k) = \int \ell_k(x'; y_k) p_k^f(x')\, dx', \tag{2.4}$$

where $\ell_k(x; y_k) := p(y_k \mid x)$ is the observation likelihood.

FREESIA does not assume direct access to $p_k^f$. Instead, the current forecast ensemble

$$\mathcal{E}_k^f = \{x_k^{f,(i)}\}_{i=1}^M \tag{2.5}$$

defines a finite-ensemble surrogate $\widetilde{p}_k^f$, and the corresponding surrogate analysis is

$$\widetilde{p}_k^a(x) \propto \ell_k(x; y_k) \widetilde{p}_k^f(x). \tag{2.6}$$

Section 3 constructs a posterior flow for this surrogate target, while Section 4 separates finite-flow error from forecast-surrogate error.

## 3 FREESIA Posterior Flow

This section derives the FREESIA posterior flow for a single assimilation cycle from the current forecast ensemble. We use $k$ for the assimilation-cycle index, $t$ for probability-flow time, $j$ for GMM components, and $s$ for endpoint importance samples. Within one cycle, we suppress $k$, write the forecast ensemble as $\mathcal{E}^f = \{x_j^f\}_{j=1}^M$, and set $h := h_k$, $R := R_k$, and $y := y_k$.

The posterior-flow path targets the surrogate analysis distribution $\widetilde{p}^a$ defined in Section 2, with $Z_0, Z_t, Z_1$ denoting auxiliary flow variables rather than the physical filtering state $X_k$. The observation is associated with the endpoint $Z_1$ through $Y = h(Z_1) + \eta$, where $\eta \sim \mathcal{N}(0, R)$ is independent of $Z_0$ and $Z_1$. FREESIA constructs a localized Gaussian-mixture surrogate of the forecast distribution from the current ensemble. Its shared covariance structure yields an analytically tractable

conditional endpoint distribution that retains the forecast cross-covariances used for sparse information transfer.

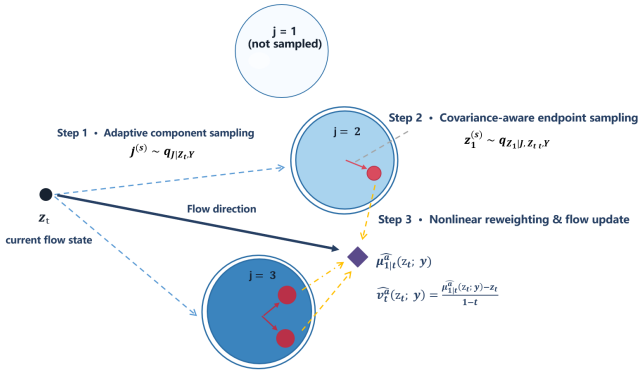


Figure 1: Overview of the FREESIA posterior-flow update. Given $z_t$, Step 1 samples an observation-adaptive GMM component, Step 2 draws a covariance-aware conditional endpoint, and Step 3 applies nonlinear likelihood-ratio correction to estimate the endpoint mean and advance the flow.

### 3.1 Posterior-Flow Construction with a Gaussian-Mixture Surrogate

This subsection derives the conditional endpoint distribution along the straight probability path

$$Z_t = (1-t)Z_0 + tZ_1, \qquad t \in [0,1], \tag{3.1}$$

where $Z_1 \sim \widetilde{p}^f$ follows the forecast surrogate and $Z_0$ is an independent source variable. Conditioning this path on the current observation defines the surrogate posterior path

$$\widetilde{p}^a_t := \mathcal{L}(Z_t \mid Y = y), \qquad t \in [0,1], \qquad \widetilde{p}^a_1 = \widetilde{p}^a. \tag{3.2}$$

Since $Z_0$ is independent of $Y$, the path starts from the source distribution $\mathcal{L}(Z_0)$ and terminates at the surrogate posterior $\widetilde{p}^a$. The velocity is given by conditional endpoint regression.

**Theorem 1.** *For a fixed observation $Y = y$ and $t \in [0,1)$,*

$$v^a_t(z_t; y) = \frac{\mathbb{E}[Z_1 \mid Z_t = z_t, Y = y] - z_t}{1-t}. \tag{3.3}$$

The proof is given in Appendix A.1. We write $v^a_t$ and $\mu^a_{1|t}$ for the velocity and conditional endpoint mean associated with the surrogate posterior path $\widetilde{p}^a_t$. The theorem therefore reduces posterior-flow construction to estimating the conditional endpoint mean

$$\mu^a_{1|t}(z_t; y) := \mathbb{E}[Z_1 \mid Z_t = z_t, Y = y].$$

To evaluate this mean, we specify the ensemble-GMM surrogate. Given $\mathcal{E}^f = \{x^f_j\}_{j=1}^M$, assign equal mixture weights $\pi_j = 1/M$, $j = 1, \ldots, M$. Hence

$$\bar{x}^f = \frac{1}{M}\sum_{j=1}^M x^f_j, \qquad \mu_j = (1-\gamma)\bar{x}^f + \gamma x^f_j, \qquad 0 \le \gamma < 1. \tag{3.4}$$

Let $\Sigma^f \succ 0$ be the localized, inflated covariance described in Appendix A.10. We specify $Z_0$ and $Z_1$ as follows, with the mixture law of $Z_1$ represented hierarchically by $J$ and $Z_1 \mid J$:

$$\begin{aligned} &Z_0 \sim \mathcal{N}(0, \Sigma^f), \qquad Z_1 \sim \frac{1}{M}\sum_{j=1}^M \mathcal{N}(\mu_j, (1-\gamma^2)\Sigma^f), \qquad Z_0 \perp (Z_1, J), \\ &J \sim \mathrm{Cat}\left(\frac{1}{M}, \ldots, \frac{1}{M}\right), \qquad Z_1 \mid J = j \sim \mathcal{N}(\mu_j, (1-\gamma^2)\Sigma^f). \end{aligned} \tag{3.5}$$

Together with the straight path, these Gaussian endpoint distributions yield a closed-form conditional endpoint distribution. It decomposes as follows. Here $r_{J|Z_t}$ denotes the conditional compo-

nent probabilities, and $\phi_{1|t,j}$ denotes the component conditional densities:

$$\begin{aligned} p_{Z_1|Z_t}(z_1 \mid z_t) &= \sum_{j=1}^{M} r_{J|Z_t}(j \mid z_t)\, \phi_{1|t,j}(z_1 \mid z_t), \\ r_{J|Z_t}(j \mid z_t) &= \frac{\pi_j \mathcal{N}\big(z_t; t\mu_j, [(1-t)^2 + (1-\gamma^2)t^2]\Sigma^f\big)}{\sum_\ell \pi_\ell \mathcal{N}(z_t; t\mu_\ell, [(1-t)^2 + (1-\gamma^2)t^2]\Sigma^f)}, \\ \phi_{1|t,j}(z_1 \mid z_t) &:= \mathcal{N}(z_1; \mu_{1|t,j}(z_t), \Sigma_{1|t}), \\ \mu_{1|t,j}(z_t) &= \mu_j + \frac{t(1-\gamma^2)}{(1-t)^2 + (1-\gamma^2)t^2}(z_t - t\mu_j), \\ \Sigma_{1|t} &= \frac{(1-\gamma^2)(1-t)^2}{(1-t)^2 + (1-\gamma^2)t^2}\Sigma^f. \end{aligned} \tag{3.6}$$

The conditional means $\mu_{1|t,j}(z_t)$ determine the component-wise endpoint locations along the path, while $\Sigma_{1|t}$ retains the cross-coordinate covariance structure needed to propagate sparse observational information in the subsequent conditioning step.

### 3.2 Nonlinear Conditioning and Finite-Sample Approximation

Direct Monte Carlo estimation of $\mu^a_{1|t}(z_t; y)$ using samples from the observation-independent bridge can be inefficient when the observation likelihood is highly concentrated relative to the bridge distribution. FREESIA instead constructs an observation-adaptive locally linearized proposal distribution and corrects it using the exact nonlinear likelihood ratio. The resulting computation consists of observation-adaptive component sampling, covariance-aware endpoint sampling, and nonlinear likelihood-ratio correction to estimate the conditional endpoint mean and posterior-flow velocity.

For each component and flow state $z_t$, locally linearize $h$ at $\mu_{1|t,j}(z_t)$, using $H_{t,j}(z_t) := Dh(\mu_{1|t,j}(z_t))$ at differentiable points and a bounded local linearization selection at nondifferentiable points; Appendix A.5 covers both cases. The exact and component-wise linearized likelihoods are

$$\begin{aligned} p(y \mid z_1) &= \mathcal{N}(y; h(z_1), R), \\ p_{\text{lin}}(y \mid z_1, j, z_t) &= \mathcal{N}\big(y; h(\mu_{1|t,j}(z_t)) + H_{t,j}(z_t)(z_1 - \mu_{1|t,j}(z_t)), R\big)\,. \end{aligned} \tag{3.7}$$

The nonlinear likelihood defines the target, whereas the local linearization is used only to construct the proposal distribution.

Conditioned on the current flow state $z_t$, the nonlinear augmented target and its locally linearized proposal are

$$\begin{aligned} p_{J,Z_1|Z_t,Y}(j, z_1 \mid z_t, y) &\propto r_{J|Z_t}(j \mid z_t)\, \phi_{1|t,j}(z_1 \mid z_t)\, p(y \mid z_1), \\ q_{J,Z_1|Z_t,Y}(j, z_1 \mid z_t, y) &= q_{J|Z_t,Y}(j \mid z_t, y)\, q_{Z_1|J,Z_t,Y}(z_1 \mid j, z_t, y) \\ &\propto r_{J|Z_t}(j \mid z_t)\, \phi_{1|t,j}(z_1 \mid z_t)\, p_{\text{lin}}(y \mid z_1, j, z_t). \end{aligned} \tag{3.8}$$

Gaussian marginalization and conditioning give

$$\begin{aligned} S_{t,j}(z_t) &= H_{t,j}(z_t)\Sigma_{1|t}H_{t,j}(z_t)^\top + R, \\ G_{t,j}(z_t) &= \Sigma_{1|t}H_{t,j}(z_t)^\top S_{t,j}(z_t)^{-1}, \\ \mu^q_{1|t,j}(z_t; y) &= \mu_{1|t,j}(z_t) + G_{t,j}(z_t)[y - h(\mu_{1|t,j}(z_t))], \\ \Sigma^q_{1|t,j}(z_t) &= \Sigma_{1|t} - G_{t,j}(z_t)H_{t,j}(z_t)\Sigma_{1|t}. \end{aligned} \tag{3.9}$$

$$\begin{aligned} q_{J|Z_t,Y}(j \mid z_t, y) &\propto r_{J|Z_t}(j \mid z_t)\, \mathcal{N}\big(y; h(\mu_{1|t,j}(z_t)), S_{t,j}(z_t)\big)\,, \\ q_{Z_1|J,Z_t,Y}(z_1 \mid j, z_t, y) &= \mathcal{N}\Big(z_1; \mu^q_{1|t,j}(z_t; y), \Sigma^q_{1|t,j}(z_t)\Big)\,. \end{aligned} \tag{3.10}$$

For brevity, we suppress dependence on $z_t$ and $y$ where no ambiguity arises. The categorical proposal $q_{J|Z_t,Y}$ favors mixture components compatible with the observation, while $G_{t,j}$ propagates observational information through the conditional covariance $\Sigma_{1|t}$. For compactness, write $r_j := r_{J|Z_t}(j \mid z_t)$, $\phi_j(z_1) := \phi_{1|t,j}(z_1 \mid z_t)$, $q(j, z_1) := q_{J,Z_1|Z_t,Y}(j, z_1 \mid z_t, y)$, and $g(j, z_1) := p(y \mid z_1)/p_{\text{lin}}(y \mid z_1, j, z_t)$.

**Theorem 2.** *For the nonlinear target and proposal defined in Eq. 3.8,*

$$\begin{aligned}\mu^a_{1|t}(z_t;y) &= \frac{\sum_j r_j \int z_1 \phi_j(z_1) p(y \mid z_1)\, dz_1}{\sum_j r_j \int \phi_j(z_1) p(y \mid z_1)\, dz_1} \\ &= \frac{\sum_j \int z_1 q(j,z_1) g(j,z_1)\, dz_1}{\sum_j \int q(j,z_1) g(j,z_1)\, dz_1}.\end{aligned} \tag{3.11}$$

At the population level, likelihood-ratio reweighting exactly recovers the nonlinear surrogate target; local linearization affects only the proposal. The proof is in Appendix A.3.

**Corollary 1.** *For $s = 1, \dots, N_{\mathrm{IS}}$, let $J^{(s)} \sim q_{J|Z_t,Y}(\cdot \mid z_t, y)$ and $Z_1^{(s)} \sim q_{Z_1|J,Z_t,Y}(\cdot \mid J^{(s)}, z_t, y)$. Define*

$$\widetilde{w}^{(s)} = g(J^{(s)}, Z_1^{(s)}), \qquad w^{(s)} = \frac{\widetilde{w}^{(s)}}{\sum_{r=1}^{N_{\mathrm{IS}}} \widetilde{w}^{(r)}}. \tag{3.12}$$

*Then*

$$\hat{\mu}^a_{1|t,N_{\mathrm{IS}}}(z_t;y) = \sum_{s=1}^{N_{\mathrm{IS}}} w^{(s)} Z_1^{(s)}, \qquad \hat{v}^a_{t,N_{\mathrm{IS}}}(z_t;y) = \frac{\hat{\mu}^a_{1|t,N_{\mathrm{IS}}}(z_t;y) - z_t}{1-t}. \tag{3.13}$$

Appendix A.5 proves pointwise consistency under first-moment conditions and, under the stronger accumulated self-normalized importance sampling (SNIS) regularity condition, controls the flow-integrated finite-sample contribution at rate $O(N_{\mathrm{IS}}^{-1/2})$. Starting from source particles $Z_0^{(i)} \sim \mathcal{N}(0, \Sigma^f)$, we integrate the posterior-flow ODE using the velocity estimator in Eq. 3.13 and take the terminal particles as the analysis ensemble. The complete finite-ensemble algorithm, structured-covariance implementation, and leading-order computational and memory complexity are detailed in Appendix A.10.

# 4 THEORETICAL ANALYSIS

This section quantifies the one-step error of the finite FREESIA update relative to the Bayesian posterior $p^a$. We separate the error arising from finite posterior-flow computation from the discrepancy introduced by the ensemble-based forecast surrogate.

For one assimilation cycle, the underlying and surrogate posterior densities are

$$p^a(x) \propto p(y \mid x) p^f(x), \qquad \widetilde{p}^a(x) \propto p(y \mid x) \widetilde{p}^f(x).$$

Define the surrogate-posterior discrepancy by

$$\varepsilon_{\mathrm{sur}}(y) := W_1(\widetilde{p}^a, p^a). \tag{4.1}$$

Let $\widehat{p}^a_{N_{\mathrm{IS}},T}$ denote the law produced by the finite FREESIA update using $N_{\mathrm{IS}}$ endpoint proposals per velocity evaluation and $T$ Euler flow steps.

The total one-step error decomposes as

$$W_1\big(\widehat{p}^a_{N_{\mathrm{IS}},T}, p^a\big) \le W_1\big(\widehat{p}^a_{N_{\mathrm{IS}},T}, \widetilde{p}^a\big) + \varepsilon_{\mathrm{sur}}(y). \tag{4.2}$$

Although Eq. 3.3 contains the factor $1/(1-t)$, the conditional Gaussian bridge removes the apparent endpoint singularity. The finite-sampling error can be controlled directly for bounded nonlinear and affine observation maps. Appendix A.5 gives the proofs and retains an accumulated-SNIS formulation for more general unbounded nonlinear observations.

**Theorem 3.** *Suppose the ideal posterior velocity $v_t^a$ is uniformly $L_v$-Lipschitz in the state for $t \in [0,1]$, and its exact flow has a uniform one-step Euler truncation error of order $O((\Delta t)^2)$. For bounded nonlinear observations, assume*

$$B_y := \sup_x \|y - h(x)\|_{R^{-1}} < \infty,$$

*and*

$$\kappa := (1-\gamma^2) \sup_x \lambda_{\max}\Big((\Sigma^f)^{1/2} H(x)^\top R^{-1} H(x) (\Sigma^f)^{1/2}\Big) < \infty.$$

*Here $H(x)$ denotes the local observation linearization used in Section 3.2, and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue. For every $0 < \delta \leq 1$ satisfying $\delta\kappa < 1$,*

$$W_1\big(\widehat{p}^{a}_{N_{\mathrm{IS}},T}, p^a\big) \leq C_\delta N_{\mathrm{IS}}^{-\delta/(1+\delta)} + C_2 T^{-1} + \varepsilon_{\mathrm{sur}}(y). \tag{4.3}$$

*If instead $h(x) = Ax + b$ is affine, its local linearization is exact and the likelihood-ratio correction satisfies $g \equiv 1$. In this case,*

$$W_1\big(\widehat{p}^{a}_{N_{\mathrm{IS}},T}, p^a\big) \leq C_1 N_{\mathrm{IS}}^{-1/2} + C_2 T^{-1} + \varepsilon_{\mathrm{sur}}(y), \tag{4.4}$$

*without requiring bounded observation residuals or an importance-weight overlap condition. The constants are independent of $N_{\mathrm{IS}}$ and $T$.*

**Corollary.** In the bounded nonlinear regime, if $\kappa < 1$, taking $\delta = 1$ in Theorem 3 yields the standard $N_{\mathrm{IS}}^{-1/2}$ finite-sampling rate.

The first term is finite endpoint-sampling error, the second is first-order flow-discretization error, and $\varepsilon_{\mathrm{sur}}(y)$ is the forecast-surrogate discrepancy defined in Eq. 4.1. Thus the full-evidence transport converges to $\widetilde{p}^a$ as $N_{\mathrm{IS}}, T \to \infty$ under either observation regime above. For more general unbounded nonlinear observations, Appendix A.5 gives the accumulated-SNIS alternative. Convergence to the underlying posterior additionally requires $\varepsilon_{\mathrm{sur}}(y) \to 0$. The result remains one-step and law-level. The diagonal component-evidence approximation used in the high-dimensional experiments is analyzed separately in Appendix A.11.

## 5 Experiments

We evaluate FREESIA along three increasingly difficult dimensions: posterior multimodality, sparse-observation information transfer, and high-dimensional nonlinear assimilation. Double-Well provides a grid-resolved posterior, Lorenz–96 separates many-to-one conditioning from sparse transfer at 1000 dimensions, and Kolmogorov flow combines both challenges at 32768 dimensions. Metrics average independent trajectories and analysis times. Appendix B.4 summarizes computational complexity and empirical analysis-time scaling for training-free generative filters on fully observed Lorenz–96 up to $d = 10^6$, including a diagonal specialization of FREESIA for extremely high-dimensional settings. Probabilistic accuracy is evaluated using the continuous ranked probability score (CRPS) (Bröcker, 2012); the Double-Well posterior comparison additionally reports the Jensen–Shannon divergence (Lin, 1991). IEnSF is not included in the numerical comparison because, although we independently reproduced the method, we could not verify that the resulting performance was sufficiently representative of the published method for a fair comparison.

### 5.1 Double-Well: Resolving a Multimodal Filtering Posterior

This experiment isolates posterior multimodality in a setting where the full filtering posterior is numerically available, allowing direct evaluation of posterior shape rather than only pointwise state error.

**Setup.** Following Ding et al. (2026), we use the double-well dynamics and many-to-one observation $Y_k = |X_k - 0.4| + \epsilon_k$. A 10,000-point grid Bayes filter provides the posterior reference, and metrics average $N_{\mathrm{run}} = 10$ runs; complete dynamical and numerical settings are given in Appendix B.1.

**Results.** FREESIA is closest to the grid Bayes posterior across all reported metrics and tracks both posterior modes and their relative masses (Table 1; Appendix Fig. 3). EnKF and EnKF-MDA Gaussianize the analysis, while EnFF-F2P can concentrate on only part of the posterior support, so the gain is distributional rather than only a reduction in posterior-mean error.

**Sensitivity.** Appendix B.1 reports sensitivity to $\gamma$, $T$, and $N_{\mathrm{IS}}$: component separation is important in this multimodal regime, while roughly 20 flow steps capture most of the improvement.

Table 1: Double-well posterior approximation errors averaged over experiments and analysis times. Lower is better; best and second-best results are bold and underlined.

| Method | JS ↓ | $W_1$ ↓ | Mode err. ↓ | RMSE ↓ | CRPS ↓ |
|---|---|---|---|---|---|
| **FREESIA** (ours) | **0.031** | **0.059** | **0.035** | **0.144** | **0.099** |
| EnKF-MDA | 0.185 | 0.320 | 0.259 | 0.433 | 0.374 |
| EnKF | <u>0.184</u> | 0.233 | 0.158 | 0.325 | 0.257 |
| EnSF | 0.315 | 0.666 | 0.359 | 0.699 | 0.379 |
| EnFF-OT | 0.550 | 0.202 | 0.130 | 0.263 | 0.254 |
| EnFF-F2P | 0.358 | <u>0.152</u> | <u>0.085</u> | <u>0.232</u> | <u>0.170</u> |

### 5.2 Lorenz–96: Multimodality and Sparse Information Propagation

Lorenz–96 separates the two main mechanisms in a 1000-dimensional chaotic system: dense many-to-one observations test nonlinear posterior ambiguity, while sparse monotone observations isolate covariance-mediated information transfer.

**Setup.** We use $d = 1000$, 20 particles, and 101 analyses under three regimes: dense $|\arctan(x_i)|$, stride-2 $\arctan(x_{2i})$, and stride-4 $\arctan(x_{4i})$. Results average $N_{\text{run}} = 10$ runs; time stepping, noise levels, baselines, localization, and method-specific settings are given in Appendix B.2. The stride-4 case uses $\sigma = 0.01$ as a high-SNR sparse-observation stress test; Appendix B.2 also reports the $\sigma = 0.05$ setting.

**Dense many-to-one observations.** With every coordinate observed through $|\arctan|$, FREESIA achieves RMSE $0.490$ and CRPS $0.174$, substantially below all competitors in Table 2, extending the multimodal-conditioning advantage to a 1000-dimensional chaotic system.

Table 2: Lorenz–96 filtering under three nonlinear observation regimes. RMSE and CRPS are averaged across repeated experiments and all 101 assimilation cycles. Lower is better; best and second-best values are bold and underlined.

| | Dense $\lvert\arctan\rvert$ | | Stride-2 arctan | | | | Stride-4 arctan | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | RMSE ↓ | CRPS ↓ | RMSE ↓ | Obs. ↓ | Unobs. ↓ | CRPS ↓ | RMSE ↓ | Obs. ↓ | Unobs. ↓ | CRPS ↓ |
| **FREESIA** (ours) | **0.490** | **0.174** | **0.615** | **0.417** | **0.757** | <u>0.241</u> | **1.575** | <u>0.373</u> | **1.805** | **0.584** |
| EnKF-MDA | <u>1.384</u> | <u>0.586</u> | <u>0.987</u> | <u>0.673</u> | <u>1.211</u> | **0.232** | <u>2.056</u> | **0.366** | <u>2.358</u> | <u>0.955</u> |
| EnKF | 3.053 | 1.712 | 1.380 | 1.120 | 1.573 | 0.422 | 2.684 | 1.689 | 2.928 | 1.356 |
| LETKF | 6.161 | 3.921 | 1.533 | 1.243 | 1.753 | 0.494 | 3.185 | 1.853 | 3.513 | 1.648 |
| EnSF | 3.901 | 2.340 | 2.027 | 1.531 | 2.422 | 0.976 | 2.472 | 1.143 | 2.777 | 1.242 |
| EnFF-OT | 4.341 | 2.959 | 2.817 | 2.303 | 3.247 | 1.825 | 3.749 | 2.637 | 4.051 | 2.636 |
| EnFF-F2P | 1.933 | 0.721 | 1.486 | 1.045 | 1.821 | 0.609 | 3.960 | 4.412 | 3.796 | 2.243 |

**Sparse monotone observations.** FREESIA has the lowest total RMSE in both sparse regimes. At stride 4, EnKF-MDA is marginally better on observed coordinates ($0.366$ versus $0.373$), but FREESIA is substantially better on the much larger unobserved subset ($1.805$ versus $2.358$). Because both sparse experiments use $\gamma = 0$, this gain isolates covariance-aware information propagation rather than mixture separation. Stride-4 ablations raise the time-averaged RMSE from $1.575$ to $2.352$ with diagonal covariance and $2.969$ without likelihood-ratio correction (Figure 11); Appendix B.2 gives details and the $\sigma = 0.05$ robustness result.

### 5.3 Kolmogorov Flow: Sparse and Non-Injective PDE Assimilation

This benchmark combines the two challenges studied separately in Lorenz–96 at PDE scale: extreme observation sparsity and non-injective nonlinear measurements.

**Setup.** We use a periodically forced incompressible Navier–Stokes benchmark on a periodic $128 \times 128$ grid, giving $d = 32768$ from the two velocity components, with 20 particles. We observe either $1\%$ of velocity entries through $\arctan(10x)$ or $2\%$ through $|\arctan(10x)|$, averaging $N_{\text{run}} = 10$ trajectories; complete settings are given in Appendix B.3. The block-structured forecast-surrogate covariance is detailed in Appendix A.10, and the reported comparison uses $N_{\text{IS}} = 5$.

**Sparse monotone observations.** With only $1\%$ of velocity entries observed through the monotone $\arctan(10x)$ operator, FREESIA is the only training-free generative method in the comparison that reaches a strongly reduced-error regime. Its time-averaged RMSE is $0.095$ with an unobserved-state RMSE of $0.095$, whereas EnSF, EnFF-OT, and EnFF-F2P remain near order-one error with RMSEs $0.942$, $1.288$, and $0.974$, respectively. Classical covariance-based filters can also reduce the error, but FREESIA remains best overall, improving over the strongest baseline EnKF-MDA from RMSE $0.148$ to $0.095$ and achieving the lowest CRPS ($0.052$). This result extends covariance-mediated

Table 3: Kolmogorov-flow filtering under sparse nonlinear observations. Metrics average ten trajectories and all 51 analyses; lower is better, with best and second-best values bold and underlined.

| | 1% random arctan | | | | 2% random \| arctan \| | | | |
|---|---|---|---|---|---|---|---|---|
| Method | RMSE ↓ | Obs. ↓ | Unobs. ↓ | CRPS ↓ | RMSE ↓ | Obs. ↓ | Unobs. ↓ | CRPS ↓ |
| **FREESIA** (ours) | **0.095** | **0.076** | **0.095** | **0.052** | **0.422** | **0.410** | **0.422** | **0.209** |
| EnKF-MDA | <u>0.148</u> | <u>0.115</u> | <u>0.148</u> | <u>0.081</u> | 1.121 | 1.115 | 1.121 | 0.797 |
| EnKF | 0.377 | 0.372 | 0.377 | 0.200 | 0.987 | 0.984 | 0.987 | 0.681 |
| LETKF | 0.218 | 0.177 | 0.218 | 0.117 | <u>0.970</u> | 1.088 | <u>0.967</u> | 0.572 |
| EnSF | 0.942 | 0.568 | 0.945 | 0.547 | 0.974 | <u>0.967</u> | 0.975 | <u>0.569</u> |
| EnFF-OT | 1.288 | 1.148 | 1.290 | 1.024 | 1.342 | 1.342 | 1.342 | 1.069 |
| EnFF-F2P | 0.974 | 0.776 | 0.976 | 0.567 | 0.980 | 0.977 | 0.980 | 0.570 |

sparse information transfer to a 32768-dimensional nonlinear PDE while retaining non-Gaussian posterior transport.

**Sparse non-injective observations.** In the $2\% \mid \arctan \mid$ regime, FREESIA achieves RMSE $0.422$ and CRPS $0.209$, compared with the lowest evaluated baseline RMSE of $0.970$ and the lowest evaluated baseline CRPS of $0.569$. Figure 2 further shows recovery of the dominant spatial structures and physically relevant velocity directions under sparse sign-ambiguous observations. The time-resolved comparison is reported in Appendix B.3.

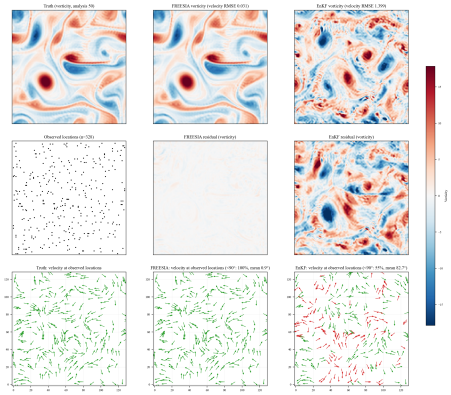

Figure 2: Final-time reconstruction for the 2% absolute-arctangent Kolmogorov-flow experiment. Top: truth, FREESIA, and EnKF vorticity, with analysis-mean velocity RMSE $0.031$ vs. $1.399$. Middle: 328 observed spatial locations, corresponding to 656 observed velocity components (2% of the state), and vorticity residuals. Bottom: observed-point velocity directions; FREESIA/EnKF achieve $100\%/55\%$ within $90°$, with mean angular errors $0.9°/82.7°$.

## 6 Conclusion

We propose FREESIA, a training-free, covariance-aware posterior transport framework for high-dimensional nonlinear data assimilation. By embedding forecast cross-covariances into flow-based transport, FREESIA propagates sparse observational information to unobserved states while retaining non-Gaussian posterior structure. Observation-adaptive proposals and posterior correction preserve the nonlinear surrogate target. We establish surrogate-relative asymptotic exactness and derive a one-step Wasserstein error bound under the stated regularity conditions, separating finite-sampling, flow-discretization, and forecast-surrogate errors.

Experiments on Double-Well, Lorenz–96, and Kolmogorov flow show that these mechanisms translate into more accurate posterior approximation, ensemble sampling, and state estimation under multimodal, sparse, and non-injective observations. FREESIA more accurately recovers posterior modes and their probability mass, improves inference of unobserved states through covariance-mediated information transfer, and preserves spatial and directional structure in challenging high-dimensional regimes. The ablations further isolate the contributions of cross-covariance transfer and nonlinear likelihood correction. At PDE scale, FREESIA reduces RMSE by 56% relative to the best evaluated baseline in the sparse non-injective case. Future work targets efficient sampling, scalable surrogates, and repeated-cycle theory under stability conditions.

## AI USE STATEMENT

Generative AI tools were used to assist with writing and language polishing, drafting and revising parts of the manuscript, literature retrieval and discovery, research ideation and methodological discussion, implementation and debugging of research code, experimental design and analysis, and the development and checking of mathematical claims and proof arguments. AI tools were also used to assist with reference formatting and manuscript consistency checks. All AI-assisted mathematical derivations, proof steps, implementation code, experimental procedures, retrieved references, numerical results, and manuscript text were independently reviewed, tested, or verified by the authors. AI-generated or AI-edited code was inspected and validated against the intended algorithms and experimental configurations before use, and reported experimental results were generated from the authors' computational experiments rather than accepted directly from AI-generated output. The authors take full responsibility for the final methodology, implementation, experiments, mathematical claims, and content of the paper.

## REFERENCES

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. 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, 2024. doi: 10.1016/j.cma.2024.117447.

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.

Jochen Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Quarterly Journal of the Royal Meteorological Society*, 138(667):1611–1617, 2012. doi: 10.1002/qj.1891.

Zhao Ding, Chenguang Duan, Yuling Jiao, Jerry Zhijian Yang, Cheng Yuan, and Pingwen Zhang. Nonlinear assimilation via score-based sequential Langevin sampling. *IEEE Transactions on Information Theory*, 2026. doi: 10.1109/TIT.2026.3715654.

Alexandre A. Emerick and Albert C. Reynolds. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16(3):639–659, 2012. doi: 10.1007/s10596-012-9275-5.

Geir Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. doi: 10.1007/s10236-003-0036-9.

Ruiqi Feng, Chenglei Yu, Wenhao Deng, Peiyan Hu, and Tailin Wu. On the guidance of flow matching. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 16993–17029. PMLR, 2025.

Nicolas Fournier. Convergence of the empirical measure in expected Wasserstein distance: Non-asymptotic explicit bounds in $\mathbb{R}^d$. *ESAIM: Probability and Statistics*, 27:749–775, 2023. doi: 10.1051/ps/2023011.

Gregory Gaspari and Stephen E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999. doi: 10.1002/qj.49712555417.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, 140(2):107–113, 1993. doi: 10.1049/ip-f-2.1993.0015.

Ruoyu Hu, Sanjeeb Poudel, Feng Bao, and Sanghyun Lee. Ensemble score filter for data assimilation of two-phase flow models in porous media. *Journal of Computational Physics*, 544:114416, 2026. doi: 10.1016/j.jcp.2025.114416.

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, 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.

Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62 of *Texts in Applied Mathematics*. Springer International Publishing, Cham, 2015. doi: 10.1007/978-3-319-20325-6.

Siming Liang, Hoang Tran, Feng Bao, Hristo G. Chipilski, Peter Jan van Leeuwen, and Guannan Zhang. Ensemble score filter with image inpainting for data assimilation in tracking surface quasi-geostrophic dynamics with partial observations, 2025. arXiv:2501.12419.

J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Edward N. Lorenz. Predictability – a problem partly solved. In Tim Palmer and Renate Hagedorn (eds.), *Predictability of Weather and Climate*, pp. 40–58. Cambridge University Press, 2006. doi: 10.1017/CBO9780511617652.004.

François Rozet and Gilles Louppe. Score-based data assimilation. In *Advances in Neural Information Processing Systems 36*, pp. 40521–40541. Neural Information Processing Systems Foundation, Inc., 2023. doi: 10.52202/075280-1763.

Phillip Si and Peng Chen. Latent-EnSF: A latent ensemble score filter for high-dimensional data assimilation with sparse observation data. In *International Conference on Learning Representations*, 2025.

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.

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.

Roger Temam. *Navier–Stokes Equations*. American Mathematical Society, Providence, Rhode Island, 2001. doi: 10.1090/chel/343.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Taos Transue, Bohan Chen, So Takao, and Bao Wang. Flow matching for efficient and scalable data assimilation, 2025. arXiv:2508.13313.

Xintong Wang, Xiaofei Guan, Ling Guo, and Hao Wu. Flow-based Bayesian filtering for high-dimensional nonlinear stochastic dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 446:118285, 2025. doi: 10.1016/j.cma.2025.118285.

Zixiang Xiong, Feng Bao, Hristo G. Chipilski, Siming Liang, Jingqiao Tang, and Guannan Zhang. A two-step ensemble score filter for data assimilation in partially observed systems, 2026. arXiv:2606.28264.

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 THEORETICAL DERIVATIONS AND PROOFS

Throughout this appendix, we fix one assimilation cycle and suppress its index.

## A.1 PROOF OF THEOREM 1

For fixed $y$, define the posterior-reweighted endpoint coupling

$$\mathbb{P}_y^a := \mathcal{L}(Z_0, Z_1 \mid Y = y). \tag{A.1}$$

Since $Z_t = (1-t)Z_0 + tZ_1$, the surrogate posterior path satisfies

$$\widetilde{p}_t^a = (Z_t)_{\#}\mathbb{P}_y^a. \tag{A.2}$$

For any $\varphi \in C_c^1(\mathbb{R}^d)$, under the required first-moment condition,

$$\frac{d}{dt}\mathbb{E}_{\mathbb{P}_y^a}[\varphi(Z_t)] = \mathbb{E}_{\mathbb{P}_y^a}\left[\nabla\varphi(Z_t)^\top (Z_1 - Z_0)\right]. \tag{A.3}$$

Conditioning on $Z_t$ shows that its marginal law satisfies the continuity equation in the weak sense with

$$v_t^a(z_t; y) = \mathbb{E}[Z_1 - Z_0 \mid Z_t = z_t, Y = y]. \tag{A.4}$$

Since $Z_t = (1-t)Z_0 + tZ_1$, for $t < 1$,

$$Z_1 - Z_0 = \frac{Z_1 - Z_t}{1-t}, \tag{A.5}$$

and therefore

$$v_t^a(z_t; y) = \frac{\mathbb{E}[Z_1 \mid Z_t = z_t, Y = y] - z_t}{1-t}, \tag{A.6}$$

which proves Theorem 1.

## A.2 ANALYTIC GMM BRIDGE DERIVATION

For FREESIA, $Z_0 \sim \mathcal{N}(0, \Sigma^f)$, $Z_1 \mid J = j \sim \mathcal{N}(\mu_j, (1-\gamma^2)\Sigma^f)$, and $Z_0 \perp (Z_1, J)$. Therefore

$$\begin{aligned} \mathrm{Cov}(Z_t \mid J = j) &= [(1-t)^2 + (1-\gamma^2)t^2]\Sigma^f, \\ \mathrm{Cov}(Z_1, Z_t \mid J = j) &= t(1-\gamma^2)\Sigma^f. \end{aligned} \tag{A.7}$$

Applying the Gaussian conditioning formula gives

$$\begin{aligned} \mu_{1|t,j}(z_t) &= \mu_j + \frac{t(1-\gamma^2)}{(1-t)^2 + (1-\gamma^2)t^2}(z_t - t\mu_j), \\ \Sigma_{1|t} &= \frac{(1-\gamma^2)(1-t)^2}{(1-t)^2 + (1-\gamma^2)t^2}\Sigma^f. \end{aligned} \tag{A.8}$$

Bayes' rule for $J$ yields the normalized responsibility

$$r_{J|Z_t}(j \mid z_t) = \frac{\pi_j \mathcal{N}\big(z_t; t\mu_j, [(1-t)^2 + (1-\gamma^2)t^2]\Sigma^f\big)}{\sum_\ell \pi_\ell \mathcal{N}(z_t; t\mu_\ell, [(1-t)^2 + (1-\gamma^2)t^2]\Sigma^f)}. \tag{A.9}$$

These are the bridge quantities used in Section 3.1.

## A.3 PROOF OF THEOREM 2 AND AFFINE CLOSURE

For fixed $z_t, y$, write $r_j = r_{J|Z_t}(j \mid z_t)$ and $\phi_j(z_1) = \phi_{1|t,j}(z_1 \mid z_t)$. The nonlinear augmented target satisfies

$$p(j, z_1 \mid z_t, y) \propto r_j \phi_j(z_1) p(y \mid z_1), \tag{A.10}$$

whereas the proposal satisfies

$$q(j, z_1 \mid z_t, y) \propto r_j \phi_j(z_1) p_{\mathrm{lin}}(y \mid z_1, j, z_t). \tag{A.11}$$

Consequently,

$$\frac{p(j, z_1 \mid z_t, y)}{q(j, z_1 \mid z_t, y)} \propto \frac{p(y \mid z_1)}{p_{\mathrm{lin}}(y \mid z_1, j, z_t)}. \tag{A.12}$$

Taking the $z_1$-expectation under the normalized target gives the corrected endpoint-regression identity in Theorem 2.

If $h$ is affine, then $p_{\mathrm{lin}} = p$, so $q = p$ and the correction weights in Corollary 1 are constant.

### A.4 STOCHASTIC KALMAN REALIZATION

For the component-wise sampling rule in Eq. A.67 of Appendix A.10, the mean and covariance are

$$\begin{aligned}\mathbb{E}[Z_1^{(s)}] &= \mu_{1|t,j} + G_{t,j}[y - h(\mu_{1|t,j})] = \mu^q_{1|t,j},\\ \mathrm{Cov}(Z_1^{(s)}) &= (I - G_{t,j}H_{t,j})\Sigma_{1|t}(I - G_{t,j}H_{t,j})^\top + G_{t,j}RG_{t,j}^\top\\ &= \Sigma_{1|t} - G_{t,j}H_{t,j}\Sigma_{1|t} = \Sigma^q_{1|t,j}.\end{aligned} \tag{A.13}$$

Thus $Z_1^{(s)} \sim q_{Z_1|J,Z_t,Y}(\cdot \mid j, z_t, y)$.

### A.5 SNIS CONSISTENCY AND FINITE-SAMPLE VELOCITY ERROR

Fix $t < 1$, a flow state $z_t$, and the observation $y$. Let $P$ and $Q$ denote the exact augmented conditional target and the locally linearized proposal from Eq. 3.8. As in Section 3, write

$$g(j, z_1) = \frac{p(y \mid z_1)}{p_{\mathrm{lin}}(y \mid z_1, j, z_t)}.$$

Then

$$\frac{dP}{dQ}(j, z_1) = \frac{g(j, z_1)}{\mathbb{E}_Q[g(J, Z_1)]}.$$

For the finite-sample analysis, we estimate the velocity directly. Define the local function

$$u_t(z_1; z_t) := \frac{z_1 - z_t}{1 - t}, \qquad t < 1. \tag{A.14}$$

Along the straight path $Z_t = (1 - t)Z_0 + tZ_1$,

$$u_t(Z_1; Z_t) = Z_1 - Z_0.$$

Consequently, Theorem 1 gives

$$v_t^a(z_t; y) = \frac{\mathbb{E}_Q[g(J, Z_1)u_t(Z_1; z_t)]}{\mathbb{E}_Q[g(J, Z_1)]}. \tag{A.15}$$

Moreover, the estimator in Eq. 3.13 is exactly

$$\widehat{v}^a_{t,N_{\mathrm{IS}}}(z_t; y) = \frac{\sum_{s=1}^{N_{\mathrm{IS}}} g(J^{(s)}, Z_1^{(s)})u_t(Z_1^{(s)}; z_t)}{\sum_{s=1}^{N_{\mathrm{IS}}} g(J^{(s)}, Z_1^{(s)})}. \tag{A.16}$$

Thus the practical FREESIA velocity is a self-normalized importance-sampling estimator of the conditional velocity.

**Proposition 1** (Automatic pointwise consistency of the corrected velocity estimator)**.** *For fixed $t < 1$, $z_t$, and $y$, assume $R \succ 0$, $0 \le \gamma < 1$, $\Sigma^f \succ 0$, and finite local linearization matrices $H_{t,j}$. Then*

$$0 < \mathbb{E}_Q[g(J, Z_1)] < \infty, \qquad \mathbb{E}_Q[g(J, Z_1)\|u_t(Z_1; z_t)\|] < \infty.$$

*Consequently,*

$$\widehat{v}^a_{t,N_{\mathrm{IS}}}(z_t; y) \longrightarrow v_t^a(z_t; y) \quad \textit{almost surely as } N_{\mathrm{IS}} \to \infty,$$

*or, equivalently, $\widehat{\mu}^a_{1|t,N_{\mathrm{IS}}}(z_t; y) \to \mu^a_{1|t}(z_t; y)$ almost surely.*

*Proof.* Writing $\mathcal{Z}_q(t, z_t) > 0$ for the proposal normalizer, the cancellation between the proposal and likelihood ratio gives

$$\mathbb{E}_Q[g] = \frac{1}{\mathcal{Z}_q}\sum_j r_j \int \phi_j(z_1)p(y \mid z_1)\, dz_1.$$

The Gaussian observation likelihood satisfies

$$0 < p(y \mid z_1) \le C_R, \qquad C_R := (2\pi)^{-m/2}|R|^{-1/2},$$

so the first moment is finite and strictly positive. Similarly,

$$\mathbb{E}_Q[g\|u_t\|] = \frac{1}{\mathcal{Z}_q}\sum_j r_j \int \phi_j(z_1)p(y \mid z_1)\|u_t(z_1; z_t)\|\, dz_1$$

is bounded above by

$$\frac{C_R}{\mathcal{Z}_q}\sum_j r_j \int \phi_j(z_1)\|u_t(z_1; z_t)\|\, dz_1,$$

which is finite for every fixed $t < 1$ because $\phi_j$ is Gaussian. The strong law applies separately to the numerator and denominator of Eq. A.16; ratio convergence proves the velocity claim, and Eq. 3.13 gives the endpoint-mean claim. □

**Pointwise bridge regularity.** Set, only in this appendix,

$$c_\gamma := 1 - \gamma^2 > 0, \qquad D_t := (1-t)^2 + c_\gamma t^2.$$

From Eq. 3.6,

$$\Sigma_{1|t} = \frac{c_\gamma(1-t)^2}{D_t}\Sigma^f.$$

Since

$$\min_{t\in[0,1]} D_t = \frac{c_\gamma}{1+c_\gamma},$$

we have

$$\frac{\Sigma_{1|t}}{(1-t)^2} = \frac{c_\gamma}{D_t}\Sigma^f \preceq (1+c_\gamma)\Sigma^f \preceq 2\Sigma^f. \tag{A.17}$$

The bridge mean satisfies

$$\mu_{1|t,j}(z_t) = \mu_j + \frac{tc_\gamma}{D_t}(z_t - t\mu_j),$$

and direct algebra gives

$$\frac{\mu_{1|t,j}(z_t) - z_t}{1-t} = \frac{(1-t)\mu_j + [t(1+c_\gamma) - 1]z_t}{D_t}. \tag{A.18}$$

For every fixed $0 \le \gamma < 1$, there is a constant $C_\gamma < \infty$ such that

$$\left\|\frac{\mu_{1|t,j}(z_t) - z_t}{1-t}\right\| \le C_\gamma(1 + \|\mu_j\| + \|z_t\|), \qquad 0 \le t < 1. \tag{A.19}$$

Thus, for every fixed $z_t \in \mathbb{R}^d$ and fixed $0 \le \gamma < 1$, the rescaled bridge mean remains finite for $t < 1$ and converges to $z_t$ as $t \uparrow 1$, while the rescaled bridge covariance remains uniformly bounded in flow time. This regularity is pointwise in $z_t$, with at most linear state dependence. No uniformity is claimed as $\gamma \uparrow 1$.

**Endpoint behavior of the proposal.** Assume $R \succeq r_0 I$ for some $r_0 > 0$. We consider two sufficient regularity regimes for the local observation model. In the smooth regime, $h$ is $C^2$ on the relevant component neighborhoods, with finite bounds on $\|Dh\|$ and $\|D^2h\|$. In the Lipschitz regime, $h$ is Lipschitz on the proposal states under consideration,

$$\|h(x) - h(x')\| \le L_h\|x - x'\|,$$

and the local linearization matrix is a measurable selection $H(x)$ satisfying

$$\|H(x)\| \le L_H.$$

At differentiable points, $H(x) = Dh(x)$; at nondifferentiable points, the implementation may use any bounded local linearization selection. State-dependent versions of $L_h$ and $L_H$, as well as polynomial-growth derivative bounds in the smooth regime, are admissible when the resulting constants have the flow-law moments required below. The smooth regime gives a quadratic local-linearization remainder, whereas the Lipschitz regime gives a linear remainder. Both are sufficient for the endpoint likelihood-ratio mismatch to vanish along the contracting bridge. For fixed $z_t$, let

$L_H(z_t)$ denote the corresponding finite bound on $\|H\|$ over the relevant component neighborhoods. In the smooth regime, let $L_{h,2}(z_t)$ denote the corresponding finite bound on $\|D^2 h\|$.

The component-wise proposal in Eq. 3.10 has mean and covariance

$$\mu^q_{1|t,j} = \mu_{1|t,j} + G_{t,j}[y - h(\mu_{1|t,j})], \qquad \Sigma^q_{1|t,j} = \Sigma_{1|t} - G_{t,j} H_{t,j} \Sigma_{1|t}.$$

Because

$$S_{t,j} = H_{t,j} \Sigma_{1|t} H_{t,j}^\top + R \succeq R,$$

we have $\|S_{t,j}^{-1}\| \le r_0^{-1}$ and hence

$$\|G_{t,j}\| = \|\Sigma_{1|t} H_{t,j}^\top S_{t,j}^{-1}\| \tag{A.20}$$

$$\le \frac{c_\gamma (1-t)^2}{D_t} \|\Sigma^f\| L_H(z_t) r_0^{-1} \le C_G(z_t)(1-t)^2, \tag{A.21}$$

where $C_G(z_t)$ is finite for each fixed current state and independent of $t$. Moreover,

$$0 \preceq \Sigma^q_{1|t,j} \preceq \Sigma_{1|t},$$

so Eq. A.17 implies

$$0 \preceq \frac{\Sigma^q_{1|t,j}}{(1-t)^2} \preceq (1 + c_\gamma)\Sigma^f \preceq 2\Sigma^f. \tag{A.22}$$

Finally,

$$\frac{\mu^q_{1|t,j} - z_t}{1-t} = \frac{\mu_{1|t,j} - z_t}{1-t} + \frac{G_{t,j}}{1-t}[y - h(\mu_{1|t,j})]. \tag{A.23}$$

By Eq. A.21, the second term is $O(1-t)$ for each fixed $z_t$ whenever the observation residual is finite. Hence, for every fixed finite $p$ for which the conditional residual moments are finite,

$$\sup_{t_0(z_t) \le t < 1} \mathbb{E}_Q[\|u_t(Z_1; z_t)\|^p] < \infty$$

for some $t_0(z_t) < 1$.

For

$$m := \mu_{1|t,j}(z_t), \qquad H := H_{t,j}, \qquad \Delta z := Z_1 - m,$$

the proposal satisfies

$$\mathbb{E}_Q[\Delta z \mid J = j] = G_{t,j}[y - h(m)] = O((1-t)^2),$$

and

$$\mathrm{Cov}_Q(\Delta z \mid J = j) = \Sigma^q_{1|t,j} = O((1-t)^2).$$

Gaussian moment bounds therefore give $\Delta z = O_{L^p}(1-t)$ for every fixed finite $p$ and fixed state $z_t$ for which the corresponding moments are finite.

In the smooth regime, Taylor's theorem gives

$$h(m + \Delta z) = h(m) + H\Delta z + \rho_h(\Delta z), \qquad \|\rho_h(\Delta z)\| \le \frac{L_{h,2}(z_t)}{2} \|\Delta z\|^2,$$

and hence $\rho_h(\Delta z) = O_{L^p}((1-t)^2)$. In the Lipschitz regime, define the same local remainder by

$$\rho_h(\Delta z) := h(m + \Delta z) - h(m) - H\Delta z.$$

The Lipschitz and bounded-linearization assumptions give

$$\|\rho_h(\Delta z)\| \le \|h(m + \Delta z) - h(m)\| + \|H\Delta z\| \le (L_h + L_H)\|\Delta z\|.$$

Thus, $\rho_h(\Delta z) = O_{L^p}(1-t)$. The remainder vanishes in both regimes, quadratically in the smooth case and linearly in the Lipschitz case.

Writing

$$a := y - h(m) - H\Delta z,$$

the Gaussian likelihoods give

$$\log g(J, Z_1) = a^\top R^{-1}\rho_h(\Delta z) - \frac{1}{2}\rho_h(\Delta z)^\top R^{-1}\rho_h(\Delta z). \tag{A.24}$$

Under either regularity regime, $a = O_{L^p}(1)$ for every required finite $p$. The smooth regime therefore gives $\log g(J, Z_1) = O_{L^p}((1-t)^2)$, whereas the Lipschitz regime gives $\log g(J, Z_1) = O_{L^p}(1-t)$. Hence, for each such fixed $z_t$,

$$g(J, Z_1) \to 1 \quad \text{in } Q\text{-probability as } t \uparrow 1. \tag{A.25}$$

The local-linearization mismatch therefore vanishes along the contracting bridge. The nonsmooth observation maps used in the experiments fall within the Lipschitz regime: $x \mapsto |x - c|$ and $x \mapsto |\arctan x|$ are globally 1-Lipschitz, while $x \mapsto |\arctan(10x)|$ is globally 10-Lipschitz. The branch derivative away from each kink and the bounded selection used at the kink satisfy the required bound on $H$.

**Endpoint importance moments at a fixed state.** For fixed $t$ and $z_t$, define the normalizing constant of the locally linearized proposal by

$$\mathcal{Z}_q(t, z_t) := \sum_{j=1}^{M} r_j \int \phi_j(z_1) p_{\text{lin}}(y \mid z_1, j, z_t)\, dz_1.$$

Gaussian marginalization gives

$$\mathcal{Z}_q(t, z_t) = \sum_{j=1}^{M} r_j \mathcal{N}\big(y; h(\mu_{1|t,j}(z_t)), S_{t,j}(z_t)\big)\,.$$

For each fixed $z_t$ and fixed $0 \le \gamma < 1$, continuity and positivity of the Gaussian evidence, together with $S_{t,j} \to R$, imply that there are $t_0(z_t) < 1$ and $c_q(z_t) > 0$ such that

$$\inf_{t_0(z_t) \le t < 1} \mathcal{Z}_q(t, z_t) \ge c_q(z_t) > 0. \tag{A.26}$$

Fix finite $r \ge 1$ and $s \ge 0$. The likelihood-ratio definition yields

$$\begin{aligned}
&\mathbb{E}_Q[g(J, Z_1)^r\,(1 + \|u_t(Z_1; z_t)\|^s)] \\
&= \frac{1}{\mathcal{Z}_q(t, z_t)} \sum_{j=1}^{M} r_j \int \phi_j(z_1) p(y \mid z_1)^r p_{\text{lin}}(y \mid z_1, j, z_t)^{1-r} \\
&\qquad\qquad \times \left(1 + \left\| \frac{z_1 - z_t}{1-t} \right\|^s\right) dz_1.
\end{aligned} \tag{A.27}$$

For bounded nonlinear likelihoods, a representative sufficient condition for the Gaussian tail in component $j$ to have a finite $r$th weight moment is

$$(r-1)\lambda_{\max}\Big(\Sigma_{1|t}^{1/2} H_{t,j}^\top R^{-1} H_{t,j} \Sigma_{1|t}^{1/2}\Big) < 1. \tag{A.28}$$

The inverse linearized likelihood contributes the positive quadratic form $(r-1)H_{t,j}^\top R^{-1} H_{t,j}$, whereas the Gaussian bridge contributes $-\Sigma_{1|t}^{-1}$, both with the same factor $1/2$. This condition is sufficient rather than necessary. The same condition applies in the Lipschitz regime because it depends only on the bounded local linearization matrix $H$, the bridge covariance, and $R$, rather than on second derivatives of $h$. Since $\Sigma_{1|t} = O((1-t)^2)$, it becomes easier to satisfy near the endpoint for every fixed finite $r$; small $R$, broad forecast uncertainty, or early flow times can make complete-flow control more restrictive. The exact Gaussian likelihood is uniformly bounded above. The inverse power $p_{\text{lin}}^{1-r}$ contributes an exponent that grows at most quadratically in $z_1 - \mu_{1|t,j}$. With the change of variables

$$z_1 = \mu_{1|t,j} + (1-t)\zeta,$$

the bridge density becomes a Gaussian density in $\zeta$ with covariance

$$\frac{\Sigma_{1|t}}{(1-t)^2} = \frac{c_\gamma}{D_t}\Sigma^f,$$

which is uniformly bounded for fixed $\gamma < 1$. The positive quadratic coefficient contributed by $p_{\text{lin}}^{1-r}$ is multiplied by $(1-t)^2$. For fixed $z_t$, the relevant derivative and residual constants are finite, so this perturbation is dominated by the negative Gaussian quadratic form for all $t$ sufficiently close to 1. The rescaled increment becomes

$$u_t(z_1; z_t) = \frac{\mu_{1|t,j} - z_t}{1-t} + \zeta,$$

and its first term is uniformly bounded by Eq. A.19. Equation A.26 then gives, after increasing $t_0(z_t)$ if necessary,

$$\boxed{\sup_{t_0(z_t) \le t < 1} \mathbb{E}_Q[g(J, Z_1)^r \, (1 + \|u_t(Z_1; z_t)\|^s)] < \infty.} \tag{A.29}$$

Thus, for each fixed state, the contracting bridge controls every fixed finite importance-weight and rescaled-endpoint moment required below. Together with Eq. A.25, the same bound at a slightly higher moment order gives uniform integrability in flow time and hence $g(J, Z_1) \to 1$ in $L^r(Q)$ for each fixed finite admissible $r$. In particular, after increasing $t_0(z_t)$ if necessary,

$$\inf_{t_0(z_t) \le t < 1} \mathbb{E}_Q[g(J, Z_1)] > 0.$$

Under either regularity regime, the bridge introduces no additional endpoint singularity in the importance weights. Complete-flow behavior away from the endpoint is handled below either directly from the observation model or through the general accumulated-SNIS alternative.

**Observation-model control in the bounded-residual regime.** Assume

$$B_y := \sup_{x \in \mathbb{R}^d} \|y - h(x)\|_{R^{-1}} < \infty, \qquad \Lambda_H := \sup_x \lambda_{\max}\Big((\Sigma^f)^{1/2} H(x)^\top R^{-1} H(x) (\Sigma^f)^{1/2}\Big) < \infty, \tag{A.30}$$

and define $\kappa := (1-\gamma^2)\Lambda_H$. If $\|H(x)\| \le L_H$, a simple sufficient bound is

$$\Lambda_H \le \frac{L_H^2 \|\Sigma^f\|}{\lambda_{\min}(R)}, \qquad \kappa \le \frac{(1-\gamma^2) L_H^2 \|\Sigma^f\|}{\lambda_{\min}(R)}.$$

These are coarse upper bounds, not identities.

The exact likelihood and its linearized counterpart satisfy

$$p(y \mid z_1) = C_R \exp\left[-\frac{1}{2}\|y - h(z_1)\|_{R^{-1}}^2\right] \ge C_R e^{-B_y^2/2}, \qquad p_{\text{lin}}(y \mid z_1, j, z_t) \le C_R.$$

Consequently,

$$g(j, z_1) \ge g_\star := e^{-B_y^2/2} > 0. \tag{A.31}$$

Moreover, writing

$$\Sigma_{1|t} = a_t \Sigma^f, \qquad a_t := \frac{(1-\gamma^2)(1-t)^2}{(1-t)^2 + (1-\gamma^2)t^2} \le 1 - \gamma^2,$$

the definition of $\kappa$ and $H_{t,j} = H(\mu_{1|t,j})$ give

$$R^{-1/2} H_{t,j} \Sigma_{1|t} H_{t,j}^\top R^{-1/2} \preceq \kappa I.$$

Thus

$$R \preceq S_{t,j} \preceq R^{1/2}(I + \kappa I) R^{1/2}, \qquad |S_{t,j}| \le |R|(1+\kappa)^m,$$

while $S_{t,j}^{-1} \preceq R^{-1}$ implies

$$(y - h(\mu_{1|t,j}))^\top S_{t,j}^{-1} (y - h(\mu_{1|t,j})) \le B_y^2.$$

Every component evidence is therefore bounded below by $C_R(1+\kappa)^{-m/2} e^{-B_y^2/2}$, and hence

$$\mathcal{Z}_q(t, z_t) \ge Z_\star := C_R(1+\kappa)^{-m/2} e^{-B_y^2/2} > 0. \tag{A.32}$$

The constant $Z_\star$ may depend strongly on the observation dimension $m$; no dimension-free claim is made.

Let $p := 1 + \delta$ with $0 < \delta \le 1$ and $\delta\kappa < 1$. In Eq. A.27, set

$$m_j := \mu_{1|t,j}(z_t), \qquad \Delta := z_1 - m_j, \qquad b_j := y - h(m_j).$$

The likelihood bounds give

$$p(y \mid z_1)^p \le C_R^p, \qquad p_{\text{lin}}^{1-p} = C_R^{1-p} \exp\left[\frac{\delta}{2}(b_j - H_{t,j}\Delta)^\top R^{-1}(b_j - H_{t,j}\Delta)\right].$$

The resulting Gaussian integral is finite whenever

$$I - \delta\Sigma_{1|t}^{1/2} H_{t,j}^\top R^{-1} H_{t,j} \Sigma_{1|t}^{1/2} \succ 0.$$

Its largest quadratic eigenvalue is at most $\delta\kappa < 1$, uniformly in $t$ and $j$. Completing the square, using $\|b_j\|_{R^{-1}} \le B_y$, and applying the rescaling $\Delta = (1-t)\xi$ together with Eqs. A.17 and A.19 shows that

$$\mathbb{E}_Q\big[g^{1+\delta}\big(1 + \|u_t\|^{1+\delta}\big)\big] \le C_\delta\big(1 + \|z_t\|^{1+\delta}\big), \qquad 0 \le t < 1. \tag{A.33}$$

Here the finite forecast ensemble ensures $\max_j \|\mu_j\| < \infty$ within the assimilation cycle. The constant may depend on the fixed surrogate, $y, R, \gamma, \Sigma^f, H, d, m$, and $\delta$, but not on $N_{\text{IS}}$, $T$, or $t$.

Set $N := N_{\text{IS}}$ and, for i.i.d. proposal samples, write

$$g_s := g(J^{(s)}, Z_1^{(s)}), \qquad u_s := u_t(Z_1^{(s)}; z_t), \qquad v := v_t^a(z_t; y), \qquad X_s := g_s(u_s - v).$$

Then $\mathbb{E}_Q[X_s] = 0$ and

$$\widehat{v}_{t,N}^a - v = \frac{N^{-1}\sum_{s=1}^N X_s}{N^{-1}\sum_{s=1}^N g_s}.$$

Equation A.31 makes the denominator at least $g_\star$. For $p = 1 + \delta \in (1, 2]$, a componentwise von Bahr–Esseen or Marcinkiewicz–Zygmund inequality gives

$$\mathbb{E}\left\| N^{-1}\sum_{s=1}^N X_s \right\|^p \le C_{p,d} N^{1-p} \mathbb{E}\|X_1\|^p.$$

Equations A.31 and A.33 first imply $\|v_t^a(z_t; y)\| \le C(1 + \|z_t\|)$ and then $\mathbb{E}_Q\|X_1\|^p \le C_p(1 + \|z_t\|^p)$. Hence

$$\mathbb{E}\big[\|\widehat{v}_{t,N}^a(z_t; y) - v_t^a(z_t; y)\| \,\big|\, z_t\big] \le C_\delta N^{-\beta_\delta}(1 + \|z_t\|), \qquad \beta_\delta := \frac{\delta}{1+\delta}. \tag{A.34}$$

For $\delta = 1$, this is the usual root-$N$ rate.

It remains to control evaluation at the random numerical states. Since the normalized weights are nonnegative and sum to one,

$$\|\widehat{v}\|^p \le \sum_{s=1}^N w_s \|u_s\|^p.$$

Using $\sum_s g_s \ge N g_\star$ and Eq. A.33 yields

$$\mathbb{E}[\|\widehat{v}_{t,N}(z_t)\|^p \mid z_t] \le C_p(1 + \|z_t\|^p). \tag{A.35}$$

For the Euler recursion $\widehat{Z}_{n+1} = \widehat{Z}_n + \Delta t\, \widehat{v}_{t_n,N}(\widehat{Z}_n)$, convexity and $\|x+y\|^p \le 2^{p-1}(\|x\|^p + \|y\|^p)$ give

$$\mathbb{E}\|\widehat{Z}_{n+1}\|^p \le (1 + C\Delta t)\mathbb{E}\|\widehat{Z}_n\|^p + C\Delta t.$$

Because $Z_0$ is Gaussian, discrete Grönwall therefore gives

$$\sup_{N\ge1,\, T\ge1} \max_{0\le n\le T} \mathbb{E}\|\widehat{Z}_n\|^p \le C_Z < \infty. \tag{A.36}$$

Evaluating Eq. A.34 at $z_t = \widehat{Z}_n$ now gives

$$a_n := \mathbb{E}\|\widehat{v}_{t_n,N}(\widehat{Z}_n) - v_{t_n}(\widehat{Z}_n)\| \le C_\delta N^{-\beta_\delta}, \tag{A.37}$$

and consequently

$$\Delta t \sum_{n=0}^{T-1} a_n \le C_\delta N^{-\beta_\delta}. \tag{A.38}$$

**Affine observation operators.** Suppose $h(x) = Ax + b$, with fixed $A$ and $b$. Then $H_{t,j} = A$ for every component and flow state. Writing $m = \mu_{1|t,j}(z_t)$, we have

$$h(m) + A(z_1 - m) = Am + b + A(z_1 - m) = Az_1 + b = h(z_1).$$

Thus the locally linearized likelihood is exact, even when $h$ is unbounded:

$$p_{\mathrm{lin}}(y \mid z_1, j, z_t) = p(y \mid z_1), \qquad g(j, z_1) = 1. \tag{A.39}$$

The proposal $Q$ therefore equals the surrogate conditional target $P$, and its independent endpoint samples give ordinary Monte Carlo estimators,

$$\widehat{\mu}^a_{1|t,N}(z_t; y) = \frac{1}{N}\sum_{s=1}^{N} Z_1^{(s)}, \qquad \widehat{v}^a_{t,N}(z_t; y) = \frac{1}{N}\sum_{s=1}^{N} u_t(Z_1^{(s)}; z_t). \tag{A.40}$$

The bridge bounds in Eqs. A.17 and A.19 give $\Sigma_{1|t}/(1-t)^2 \preceq 2\Sigma^f$ and $\|(\mu_{1|t,j}(z_t) - z_t)/(1-t)\| \le C_\gamma(1 + \|\mu_j\| + \|z_t\|)$. For the affine Gaussian proposal, $0 \preceq \Sigma^q_{1|t,j} \preceq \Sigma_{1|t}$ and $G_{t,j} = \Sigma_{1|t}A^\top S^{-1}_{t,j}$, with $S_{t,j} \succeq R$. Hence $\|G_{t,j}/(1-t)\| \le C(1-t)$. Since the finite component means are fixed and $\|y - Am - b\| \le C(1 + \|z_t\|)$, the proposal mean shift divided by $1 - t$ is also at most linear in $\|z_t\|$. Averaging over the component probabilities therefore yields, uniformly for $t < 1$,

$$\mathbb{E}_Q\big[\|u_t(Z_1; z_t)\|^2 \mid z_t\big] \le C(1 + \|z_t\|^2). \tag{A.41}$$

Because $v^a_t(z_t; y) = \mathbb{E}_Q[u_t(Z_1; z_t) \mid z_t]$, standard Monte Carlo variance and Jensen's inequality give

$$\mathbb{E}\big[\|\widehat{v}^a_{t,N}(z_t; y) - v^a_t(z_t; y)\| \,\big|\, z_t\big] \le CN^{-1/2}(1 + \|z_t\|). \tag{A.42}$$

Equation A.41 also bounds the second moment of the sample-mean velocity. The random-state moment recursion used in Eq. A.36, now with $p = 2$, gives $\sup_{N,T} \max_n \mathbb{E}\|\widehat{Z}_n\|^2 < \infty$. Evaluating Eq. A.42 at $\widehat{Z}_n$ and summing the Euler steps yields

$$\Delta t \sum_{n=0}^{T-1} \mathbb{E}\Big[\|\widehat{v}^a_{t_n,N}(\widehat{Z}_n) - v^a_{t_n}(\widehat{Z}_n)\|\Big] \le CN^{-1/2}. \tag{A.43}$$

This estimate does not require a bounded residual or an importance-weight overlap condition.

**General accumulated-SNIS alternative.** The bounded nonlinear result uses the residual assumption $B_y < \infty$. For more general unbounded nonlinear observation maps, we retain the following accumulated condition on the standard pointwise SNIS constants. For each fixed $(t, z_t)$ satisfying the standard SNIS moment conditions, apply Theorem 2.3 of Agapiou et al. (2017) with the unbounded test function

$$\varphi_\ell(j, z_1) := u_t(z_1; z_t)_\ell.$$

Let $C_{\mathrm{MSE},\ell}(t, z_t)$ denote the resulting finite non-asymptotic MSE constant. Applying the theorem coordinatewise to Eq. A.16 gives the pointwise conditional estimate

$$\mathbb{E}\Big[\big|\widehat{v}^a_{t,N_{\mathrm{IS}},\ell}(z_t; y) - v^a_{t,\ell}(z_t; y)\big|^2 \,\Big|\, z_t\Big] \le \frac{C_{\mathrm{MSE},\ell}(t, z_t)}{N_{\mathrm{IS}}}. \tag{A.44}$$

The pointwise estimate above does not require a uniform-in-time bound on the SNIS constants. For the global finite-computation analysis, it is sufficient to control their accumulated contribution along the numerical posterior flow. Define

$$\kappa_{\mathrm{IS}}(t, z) := \left(\sum_{\ell=1}^{d} C_{\mathrm{MSE},\ell}(t, z)\right)^{1/2}.$$

We assume that there exists $C_{\mathrm{IS}} < \infty$, independent of $N_{\mathrm{IS}}$ and $T$, such that

$$\sup_{N_{\mathrm{IS}} \ge 1,\, T \ge 1} \Delta t \sum_{n=0}^{T-1} \mathbb{E}\Big[\kappa_{\mathrm{IS}}(t_n, \widehat{Z}_n)\Big] \le C_{\mathrm{IS}}, \qquad \Delta t = \frac{1}{T}. \tag{A.45}$$

This condition controls the accumulated first-moment contribution of the standard pointwise SNIS MSE constants along the numerical flow. It is weaker than a uniform bound at every flow time: the latter implies Eq. A.45 by Jensen's inequality, whereas the accumulated condition allows the importance-sampling difficulty to vary with $t$ provided that its total contribution remains finite.

Taking conditional expectations in Eq. A.44, summing over coordinates, and applying Jensen's inequality gives, at each numerical state,

$$\mathbb{E}\Big[\Big\|\widehat{v}^a_{t_n,N_{\mathrm{IS}}}(\widehat{Z}_n;y)-v^a_{t_n}(\widehat{Z}_n;y)\Big\|\Big]\le N_{\mathrm{IS}}^{-1/2}\mathbb{E}\Big[\kappa_{\mathrm{IS}}(t_n,\widehat{Z}_n)\Big]. \tag{A.46}$$

Therefore Eq. A.45 implies

$$\Delta t\sum_{n=0}^{T-1}\mathbb{E}\Big[\Big\|\widehat{v}^a_{t_n,N_{\mathrm{IS}}}(\widehat{Z}_n;y)-v^a_{t_n}(\widehat{Z}_n;y)\Big\|\Big]\le C_{\mathrm{IS}}N_{\mathrm{IS}}^{-1/2}. \tag{A.47}$$

For any fixed $(t,z_t)$ satisfying the pointwise SNIS conditions, Eqs. 3.3 and 3.13 give

$$\widehat{\mu}^a_{1|t,N_{\mathrm{IS}}}(z_t;y)-\mu^a_{1|t}(z_t;y)=(1-t)\left(\widehat{v}^a_{t,N_{\mathrm{IS}}}(z_t;y)-v^a_t(z_t;y)\right).$$

Hence Eq. A.44 implies

$$\mathbb{E}\Big[\Big\|\widehat{\mu}^a_{1|t,N_{\mathrm{IS}}}(z_t;y)-\mu^a_{1|t}(z_t;y)\Big\|\,\Big|\,z_t\Big]\le(1-t)N_{\mathrm{IS}}^{-1/2}\kappa_{\mathrm{IS}}(t,z_t). \tag{A.48}$$

Thus the endpoint-regression error retains the explicit $(1-t)$ contraction factor without requiring a uniform-in-time SNIS constant. A continuous-time analogue,

$$\int_0^1\mathbb{E}[\kappa_{\mathrm{IS}}(t,Z_t)]\,dt<\infty,$$

similarly controls the accumulated SNIS error along the exact flow, but it is not required for the finite Euler error bound below.

The conditional-velocity identity also gives the terminal value:

$$v^a_t(z_t;y)=\mathbb{E}[Z_1-Z_0\mid Z_t=z_t,Y=y].$$

At $t=1$, $Z_t=Z_1$. Since $Z_0$ is independent of $(Z_1,Y)$ and $\mathbb{E}[Z_0]=0$,

$$v^a_1(z;y):=\mathbb{E}[Z_1-Z_0\mid Z_1=z,Y=y]=z. \tag{A.49}$$

Equations A.18–A.29 give $v^a_t(z;y)\to z$ as $t\uparrow 1$ for every fixed $z$ covered by the pointwise assumptions. The apparent endpoint singularity in Eq. 3.3 is therefore removable. The numerical method evaluates at $t_n=n/T<1$ and advances its final Euler step to $t=1$.

Appendix B.2 reports finite-sample normalized-ESS diagnostics for the sparse, small-noise Lorenz–96 regime. They show substantial weight concentration over the early and intermediate flow and rapid recovery near the endpoint, consistent with the endpoint contraction analysis. For general unbounded nonlinear observations, complete-flow moment control remains nontrivial and is supplied by the accumulated-SNIS alternative above.

### A.6 Finite-computation error

Let

$$\Delta t=\frac{1}{T},\qquad t_n=n\Delta t=\frac{n}{T},\qquad n=0,\dots,T.$$

Assume that the continuously extended ideal velocity admits a deterministic Lipschitz constant $L_v<\infty$ on the supports relevant to the coupled exact and numerical flow laws:

$$\|v^a_t(x;y)-v^a_t(x';y)\|\le L_v\|x-x'\|$$

for $t\in[0,1]$. Assume also that one Euler step for the exact ODE has mean local truncation error bounded by $C_E(\Delta t)^2$, uniformly in $n$.

**Theorem 4.** *Under these flow regularity conditions, let $Z_t$ solve*

$$\dot{Z}_t = v_t^a(Z_t), \qquad Z_0 \sim \mathcal{L}(Z_0),$$

*and let*

$$\widehat{Z}_{n+1} = \widehat{Z}_n + \Delta t\, \widehat{v}^a_{t_n, N_{\mathrm{IS}}}(\widehat{Z}_n), \qquad \widehat{Z}_0 = Z_0. \tag{A.50}$$

*In the bounded nonlinear regime of Appendix A.5, for every $0 < \delta \le 1$ with $\delta\kappa < 1$,*

$$W_1\big(\widehat{p}^a_{N_{\mathrm{IS}},T}, \widetilde{p}^a\big) \le e^{L_v} C_\delta N_{\mathrm{IS}}^{-\delta/(1+\delta)} + B(L_v) C_E T^{-1}. \tag{A.51}$$

*In the affine regime,*

$$W_1\big(\widehat{p}^a_{N_{\mathrm{IS}},T}, \widetilde{p}^a\big) \le e^{L_v} C_1 N_{\mathrm{IS}}^{-1/2} + B(L_v) C_E T^{-1}. \tag{A.52}$$

*Here*

$$B(L_v) := \begin{cases} (e^{L_v} - 1)/L_v, & L_v > 0, \\ 1, & L_v = 0. \end{cases}$$

*Proof.* Couple the exact and numerical flows through the same initial variable $Z_0$, and let

$$e_n := \mathbb{E}\|\widehat{Z}_n - Z_{t_n}\|.$$

One exact ODE step satisfies

$$Z_{t_{n+1}} = Z_{t_n} + \Delta t\, v^a_{t_n}(Z_{t_n}) + \rho_n, \qquad \mathbb{E}\|\rho_n\| \le C_E(\Delta t)^2.$$

Subtracting the numerical update, adding and subtracting $v^a_{t_n}(\widehat{Z}_n)$, and using the $L_v$-Lipschitz property gives

$$e_{n+1} \le (1 + L_v \Delta t) e_n + \Delta t\, a_n + C_E(\Delta t)^2, \tag{A.53}$$

where

$$a_n := \mathbb{E}\left\| \widehat{v}^a_{t_n, N_{\mathrm{IS}}}(\widehat{Z}_n) - v^a_{t_n}(\widehat{Z}_n) \right\|.$$

Equation A.37 bounds $a_n$ in the bounded nonlinear regime. In the affine regime, Eq. A.42 and the random-state second-moment bound give the corresponding root-$N$ estimate. Thus $a_n \le r_N$, where

$$r_N := \begin{cases} C_\delta N_{\mathrm{IS}}^{-\delta/(1+\delta)}, & \text{bounded nonlinear}, \\ C_1 N_{\mathrm{IS}}^{-1/2}, & \text{affine}. \end{cases}$$

Writing $q = 1 + L_v \Delta t$ and iterating the recursion from $e_0 = 0$,

$$e_T \le \Delta t \sum_{n=0}^{T-1} q^{T-1-n} a_n + C_E(\Delta t)^2 \sum_{n=0}^{T-1} q^{T-1-n}.$$

Since $q^T \le e^{L_v}$,

$$\Delta t \sum_{n=0}^{T-1} q^{T-1-n} a_n \le e^{L_v} r_N.$$

Moreover,

$$C_E(\Delta t)^2 \sum_{k=0}^{T-1} q^k = C_E \Delta t\, \frac{q^T - 1}{L_v} \le B(L_v) C_E T^{-1},$$

with the continuous interpretation $B(0) = 1$. Therefore

$$e_T \le e^{L_v} r_N + B(L_v) C_E T^{-1}.$$

The coupling has terminal marginals $\widehat{p}^a_{N_{\mathrm{IS}},T}$ and $\widetilde{p}^a$, hence

$$W_1\big(\widehat{p}^a_{N_{\mathrm{IS}},T}, \widetilde{p}^a\big) \le \mathbb{E}\|\widehat{Z}_T - Z_1\| = e_T,$$

which proves Eqs. A.51 and A.52. □

For more general unbounded nonlinear observation maps, the same recursion combined with the accumulated-SNIS condition of Appendix A.5 recovers the corresponding general finite-computation bound.

### A.7 Stability of Bayesian reweighting

For this proof, write $\ell_y(x) = p(y \mid x)$ and define the Bayesian reweighting operator

$$\mathcal{B}_y(\mu)(dx) := \frac{\ell_y(x)\mu(dx)}{\mu(\ell_y)}. \tag{A.54}$$

**Theorem 5.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Assume $0 \le \ell_y \le M_L$, $\ell_y$ is $G_L$-Lipschitz, and define*

$$\begin{aligned} &\mathcal{Z}_\mu := \mu(\ell_y) > 0, \qquad \mathcal{Z}_\nu := \nu(\ell_y) > 0, \\ &\mathcal{Z}_\star := \min\{\mathcal{Z}_\mu, \mathcal{Z}_\nu\}, \qquad B_2 := \max_{\rho \in \{\mu,\nu\}} \left( \int \|x\|^2 \rho(dx) \right)^{1/2}. \end{aligned} \tag{A.55}$$

*Then*

$$W_1(\mathcal{B}_y(\mu), \mathcal{B}_y(\nu)) \le C_{\mathrm{Bayes}}(y) W_2(\mu, \nu), \tag{A.56}$$

*where*

$$C_{\mathrm{Bayes}}(y) = \frac{M_L + G_L B_2}{\mathcal{Z}_\star} + \frac{M_L G_L B_2}{\mathcal{Z}_\star^2}. \tag{A.57}$$

*Proof.* By Kantorovich–Rubinstein duality, consider a 1-Lipschitz $f$ with $f(0) = 0$, so $|f(x)| \le \|x\|$. Decompose

$$\mathcal{B}_y(\mu)(f) - \mathcal{B}_y(\nu)(f) = \frac{\mu(f\ell_y) - \nu(f\ell_y)}{\mathcal{Z}_\mu} + \nu(f\ell_y) \left( \frac{1}{\mathcal{Z}_\mu} - \frac{1}{\mathcal{Z}_\nu} \right). \tag{A.58}$$

For any $\Gamma \in \Pi(\mu, \nu)$ and $(U, V) \sim \Gamma$, Cauchy–Schwarz gives

$$|\mu(f\ell_y) - \nu(f\ell_y)| \le (M_L + G_L B_2) \left( \mathbb{E}_\Gamma \|U - V\|^2 \right)^{1/2}. \tag{A.59}$$

Moreover, $|\mathcal{Z}_\mu - \mathcal{Z}_\nu| \le G_L W_2(\mu, \nu)$ and $|\nu(f\ell_y)| \le M_L B_2$. Taking the infimum over $\Gamma \in \Pi(\mu, \nu)$, then the supremum over $f$, yields the stated constant. □

### A.8 Forecast and posterior surrogate error

Let

$$\nu_M^f = \frac{1}{M} \sum_{j=1}^{M} \delta_{x_j^f}$$

denote the equal-weight empirical forecast measure. Its mean is

$$\bar{x}^f = \frac{1}{M} \sum_{j=1}^{M} x_j^f,$$

and its covariance is

$$C_{\nu_M}^f := \mathrm{Cov}(\nu_M^f) = \frac{1}{M} \sum_{j=1}^{M} (x_j^f - \bar{x}^f)(x_j^f - \bar{x}^f)^\top. \tag{A.60}$$

**Lemma 1.** *Let $U \sim \nu_M^f$ and $\xi \sim \mathcal{N}(0, I)$ independently, with $\Sigma^f \succeq 0$. The FREESIA forecast-GMM surrogate is the law of*

$$\widetilde{U} = \bar{x}^f + \gamma(U - \bar{x}^f) + \sqrt{1 - \gamma^2} (\Sigma^f)^{1/2} \xi. \tag{A.61}$$

*Then*

$$\begin{aligned} W_2(\widetilde{p}^f, p^f) \le W_2(\nu_M^f, p^f) &+ \sqrt{2(1-\gamma)\,\mathrm{tr}\, C_{\nu_M}^f} \\ &+ \sqrt{1-\gamma^2} \left\| (\Sigma^f)^{1/2} - (C_{\nu_M}^f)^{1/2} \right\|_F. \end{aligned} \tag{A.62}$$

*Proof.* Using the same $(U, \xi)$, compare $\widetilde{U}$ with

$$U_C = \bar{x}^f + \gamma(U - \bar{x}^f) + \sqrt{1-\gamma^2}(C^f_{\nu_M})^{1/2}\xi. \tag{A.63}$$

The two couplings satisfy

$$\mathbb{E}\|U_C - U\|^2 = 2(1-\gamma)\,\mathrm{tr}\, C^f_{\nu_M}, \qquad \mathbb{E}\|\widetilde{U} - U_C\|^2 = (1-\gamma^2)\left\|(\Sigma^f)^{1/2} - (C^f_{\nu_M})^{1/2}\right\|_F^2. \tag{A.64}$$

The triangle inequality with $W_2(\nu^f_M, p^f)$ proves the bound. Independence also gives

$$\mathbb{E}\|\widetilde{U}\|^2 = \|\bar{x}^f\|^2 + \gamma^2\,\mathrm{tr}\, C^f_{\nu_M} + (1-\gamma^2)\,\mathrm{tr}\,\Sigma^f, \tag{A.65}$$

so $\widetilde{p}^f \in \mathcal{P}_2(\mathbb{R}^d)$. □

We leave $W_2(\nu^f_M, p^f)$ explicit rather than impose a dimension-free empirical-measure rate, since generic Wasserstein empirical convergence depends on dimension and moment assumptions (Fournier, 2023).

Recall that $\varepsilon_{\mathrm{sur}}(y) = W_1(\widetilde{p}^a, p^a)$ from Eq. 4.1.

**Corollary 2.** *Under the conditions of Theorem 5 for $(p^f, \widetilde{p}^f)$,*

$$\begin{aligned}\varepsilon_{\mathrm{sur}}(y) \le C_{\mathrm{Bayes}}(y)\Bigg[&W_2(\nu^f_M, p^f) + \sqrt{2(1-\gamma)\,\mathrm{tr}\, C^f_{\nu_M}} \\ &+ \sqrt{1-\gamma^2}\left\|(\Sigma^f)^{1/2} - (C^f_{\nu_M})^{1/2}\right\|_F\Bigg].\end{aligned} \tag{A.66}$$

*Proof.* Apply Theorem 5 to $(p^f, \widetilde{p}^f)$ and substitute Lemma 1. □

The three contributions in this bound are finite-ensemble discrepancy, GMM smoothing, and covariance modification, respectively.

### A.9 Proof of Theorem 3

By the triangle inequality and Eq. 4.2,

$$W_1(\widehat{p}^a_{N_{\mathrm{IS}},T}, p^a) \le W_1(\widehat{p}^a_{N_{\mathrm{IS}},T}, \widetilde{p}^a) + \varepsilon_{\mathrm{sur}}(y).$$

Applying the corresponding case of Theorem 4 gives

$$W_1(\widehat{p}^a_{N_{\mathrm{IS}},T}, p^a) \le \begin{cases} C_\delta N_{\mathrm{IS}}^{-\delta/(1+\delta)} + C_2 T^{-1} + \varepsilon_{\mathrm{sur}}(y), & \text{bounded nonlinear,} \\ C_1 N_{\mathrm{IS}}^{-1/2} + C_2 T^{-1} + \varepsilon_{\mathrm{sur}}(y), & \text{affine,} \end{cases}$$

with constants relabeled as in Theorem 3. When $\kappa < 1$, $\delta = 1$ is admissible in the bounded nonlinear case and yields the stated corollary.

### A.10 Covariance and high-dimensional implementation

**Complete finite-ensemble update.** Algorithm 1 gives the full-evidence FREESIA update; the high-dimensional experiments use the diagonal categorical-evidence approximation described below. It implements the three stages derived in Section 3: observation-adaptive component sampling, covariance-aware endpoint sampling, and nonlinear correction with a flow update. In Algorithm 1, $z_t^{(i)}$ denotes a deterministic numerical flow particle, whereas $\zeta_1^{(i,s)}$ denotes its $s$th stochastic endpoint proposal.

**Algorithm 1** FREESIA: Training-Free Ensemble Posterior Flow

**Require:** Forecast ensemble $\mathcal{E}^f = \{x_j^f\}_{j=1}^M$, observation $y$, observation operator $h$, observation-noise covariance $R$, GMM separation $\gamma$, flow steps $T$, and endpoint importance samples $N_{\mathrm{IS}}$

Construct the GMM component means $\{\mu_j\}_{j=1}^M$ and the localized/inflated covariance $\Sigma^f = L_f L_f^\top$
Set $\Delta t = 1/T$
Draw $z_0^{(i)} = L_f \xi_0^{(i)}$, $\xi_0^{(i)} \sim \mathcal{N}(0, I)$, for $i = 1, \dots, M$
**for** $n = 0, \dots, T-1$ **do**
  $t \leftarrow n\Delta t$
  **for** $i = 1, \dots, M$ **do**
    Compute $\{r_{J|Z_t}(j \mid z_t^{(i)}), \mu_{1|t,j}(z_t^{(i)})\}_{j=1}^M$ and $\Sigma_{1|t}$ using Eq. 3.6
    Compute $\{H_{t,j}^{(i)}, S_{t,j}^{(i)}, q_{J|Z_t,Y}(j \mid z_t^{(i)}, y), G_{t,j}^{(i)}\}_{j=1}^M$ using Eqs. 3.9–3.10
    **for** $s = 1, \dots, N_{\mathrm{IS}}$ **do**
      **Step 1: Observation-adaptive component sampling.** Draw $j^{(i,s)} \sim q_{J|Z_t,Y}(\cdot \mid z_t^{(i)}, y)$
      **Step 2: Covariance-aware endpoint sampling.** Draw $\zeta_1^{(i,s)} \sim q_{Z_1|J,Z_t,Y}(\cdot \mid j^{(i,s)}, z_t^{(i)}, y)$ using Eq. A.67 in the sampling paragraph below
    **end for**
    **Step 3: Nonlinear correction and flow update.** Compute $\widetilde{w}^{(i,s)} = p(y \mid \zeta_1^{(i,s)}) / p_{\mathrm{lin}}(y \mid \zeta_1^{(i,s)}, j^{(i,s)}, z_t^{(i)})$ for $s = 1, \dots, N_{\mathrm{IS}}$, and normalize $w^{(i,s)} = \widetilde{w}^{(i,s)} / \sum_{r=1}^{N_{\mathrm{IS}}} \widetilde{w}^{(i,r)}$
    Set $\widehat{\mu}_{1|t}^a = \sum_{s=1}^{N_{\mathrm{IS}}} w^{(i,s)} \zeta_1^{(i,s)}$ and $\widehat{v}_t^a = (\widehat{\mu}_{1|t}^a - z_t^{(i)})/(1-t)$
    Update $z_{t+\Delta t}^{(i)} = z_t^{(i)} + \Delta t\, \widehat{v}_t^a$
  **end for**
**end for**
Set $x^{a,(i)} \leftarrow z_1^{(i)}$ for $i = 1, \dots, M$
**return** $\mathcal{E}^a = \{x^{a,(i)}\}_{i=1}^M$

**Sampling and numerical integration.** Factor $\Sigma^f = L_f L_f^\top$ once per assimilation cycle. For a sampled component $J^{(s)} = j$, draw $\xi^{(s)} \sim \mathcal{N}(0, I)$ and $\epsilon^{(s)} \sim \mathcal{N}(0, R)$ independently and use

$$\begin{aligned} \widetilde{Z}_1^{(s)} &= \mu_{1|t,j} + \sqrt{\frac{(1-\gamma^2)(1-t)^2}{(1-t)^2 + (1-\gamma^2)t^2}}\, L_f \xi^{(s)}, \\ Z_1^{(s)} &= \widetilde{Z}_1^{(s)} + G_{t,j}\Big[y + \epsilon^{(s)} - h(\mu_{1|t,j}) - H_{t,j}(\widetilde{Z}_1^{(s)} - \mu_{1|t,j})\Big]. \end{aligned} \tag{A.67}$$

Conditional on $J^{(s)} = j$, this samples exactly from $q_{Z_1|J,Z_t,Y}(\cdot \mid j, z_t, y)$; see Appendix A.4.

All likelihood-ratio weights are evaluated in the log domain,

$$\log \widetilde{w}^{(s)} = \log p(y \mid Z_1^{(s)}) - \log p_{\mathrm{lin}}(y \mid Z_1^{(s)}, J^{(s)}, z_t), \tag{A.68}$$

and normalized by subtracting a log-sum-exp term (implemented as a stable softmax) to avoid numerical underflow or overflow in high-dimensional observation spaces.

With $t_n = n/T$ and $\Delta t = 1/T$, each flow particle is advanced by $z_{t_{n+1}}^{(i)} = z_{t_n}^{(i)} + \Delta t\, \hat{v}_{t_n,N_{\mathrm{IS}}}^a(z_{t_n}^{(i)}; y)$. Although the endpoint-regression form in Eq. 3.13 is written for $t < 1$, Appendix A.5 shows that its apparent endpoint singularity is removable under the stated regularity conditions. Algorithm 1 evaluates the velocity only at $t_n = n/T < 1$ and advances the final Euler step directly to $t = 1$, exactly as analyzed in Appendix A.6.

After $T$ flow steps, set $x^{a,(i)} := z_1^{(i)}$ and $\mathcal{E}^a := \{x^{a,(i)}\}_{i=1}^M = \{z_1^{(i)}\}_{i=1}^M$.

Two covariance matrices play distinct roles. Recall from Appendix A.8 that $C_{\nu_M}^f$ denotes the covariance of the equal-weight empirical forecast measure. The standard uniform-ensemble sample covariance is

$$\widehat{C}_M^f := \frac{1}{M-1} \sum_{j=1}^M (x_j^f - \bar{x}^f)(x_j^f - \bar{x}^f)^\top. \tag{A.69}$$

The factor $1/M$ in $C_{\nu_M}^f$ arises because it is the covariance of the equal-weight empirical probability measure, whereas $\widehat{C}_M^f$ uses the standard unbiased $1/(M-1)$ ensemble covariance employed in the practical localized/inflated update. The practical covariance used by FREESIA is

$$\Sigma^f = \lambda_{\mathrm{infl}}\big(\Lambda \odot \widehat{C}_M^f\big). \tag{A.70}$$

The covariance-modification term in Corollary 2 measures the discrepancy between these choices. The density and responsibility formulas assume $\Sigma^f \succ 0$, whereas the surrogate Wasserstein bound in Lemma 1 only requires $\Sigma^f \succeq 0$. In implementation, a small numerical jitter is added when necessary to obtain a positive-definite factorization.

The same factor $L_f$ generates source and endpoint noise and evaluates bridge responsibilities. Because the determinant is shared across components,

$$\log r_j = c(z_t, t) + \log \pi_j - \frac{\|L_f^{-1}(z_t - t\mu_j)\|^2}{2[(1-t)^2 + (1-\gamma^2)t^2]}, \tag{A.71}$$

where $c(z_t, t)$ is independent of the component label $j$.

**Diagonal evidence approximation.** For large observation dimension, component-label adaptation may use $S_{t,j} \mapsto \mathrm{Diag}(\mathrm{diag}\, S_{t,j})$. With

$$e_j = \mathcal{N}(y; h(\mu_{1|t,j}), S_{t,j}), \qquad \widetilde{e}_j = \mathcal{N}\big(y; h(\mu_{1|t,j}), \mathrm{Diag}(\mathrm{diag}\, S_{t,j})\big)\,, \tag{A.72}$$

let $q_j = r_j e_j / A$, $\widetilde{q}_j = r_j \widetilde{e}_j / \widetilde{A}$, $A = \sum_j r_j e_j$, and $\widetilde{A} = \sum_j r_j \widetilde{e}_j$. Then

$$\|q - \widetilde{q}\|_1 \le \frac{2}{A} \sum_{j=1}^{M} r_j |e_j - \widetilde{e}_j|. \tag{A.73}$$

This diagonal component-evidence approximation is used in all reported Lorenz–96 and Kolmogorov-flow FREESIA experiments. It changes only the categorical component-adaptation probabilities; the conditional endpoint proposal in Eqs. 3.9–3.10 is otherwise unchanged. The nonlinear likelihood-ratio correction alone does not remove the categorical approximation introduced by replacing $e_j$ with $\widetilde{e}_j$. Its induced population-flow error is analyzed in Appendix A.11.

**Block-structured forecast surrogate for Kolmogorov flow.** For the $128 \times 128$ Kolmogorov-flow experiments, the practical forecast-surrogate covariance is represented by a block-structured approximation over the full $d = 32768$ state. The spatial domain is partitioned into 16 non-overlapping blocks indexed by $b = 1, \dots, 16$, and the localized covariance is factorized independently within each block:

$$\Sigma^f_{\mathrm{blk}} = \mathrm{blkdiag}\Big(\Sigma^f_1, \dots, \Sigma^f_{16}\Big)\,, \quad L_{\mathrm{blk}} = \mathrm{blkdiag}(L_1, \dots, L_{16})\,, \quad \Sigma^f_b = L_b L_b^\top. \tag{A.74}$$

For this benchmark, $\Sigma^f_{\mathrm{blk}}$ is the covariance defining the practical forecast surrogate, rather than an auxiliary covariance used only for noise generation. The source law, bridge covariance, component responsibilities, covariance–observation products, $S_{t,j}$, $G_{t,j}$, and conditional endpoint sampling all use this same covariance. Accordingly, the equations in Section 3 and Algorithm 1 are interpreted with $\Sigma^f = \Sigma^f_{\mathrm{blk}}$ and $L_f = L_{\mathrm{blk}}$ for this experiment. Matrix-free ensemble products evaluate actions of this covariance without introducing a second covariance model. The discrepancy between $\Sigma^f_{\mathrm{blk}}$ and the empirical forecast covariance is included in the covariance-modification contribution in Corollary 2. The nonlinear likelihood-ratio correction is evaluated globally over the complete observed vector for each endpoint proposal, rather than normalized independently by block.

**Computational complexity.** Let $M$, $T$, $d$, and $m$ denote the ensemble size, number of generative steps, state dimension, and observation dimension, respectively. We separate the cost of one component-conditioned proposal into $C_{\mathrm{cond}}$ for constructing the observation-space conditional quantities, including $S_{t,j} = H_{t,j}\Sigma_{1|t}H_{t,j}^\top + R$ and its linear-solve factorization; $C_{\mathrm{end}}$ for drawing one conditional endpoint with Eq. A.67 after these quantities are available; and $C_{\mathrm{like}}$ for evaluating the nonlinear likelihood-ratio correction. The leading analysis work is therefore

$$O\big(TM^2 C_{\mathrm{cond}} + TMN_{\mathrm{IS}}(C_{\mathrm{end}} + C_{\mathrm{like}})\big)\,. \tag{A.75}$$

For structured covariance and sampling-factor actions that cost $O(d)$, component-wise or row-sparse observation Jacobians, and a direct dense solve in observation space, $m$ covariance–observation products and factorization of the $m \times m$ matrix $S_{t,j}$ give $C_{\mathrm{cond}} = O(dm + m^3)$. Once this factorization is available, a structured noise draw, observation-space triangular solve, and covariance

back-projection give $C_{\text{end}} = O(d+m^2)$. For component-wise nonlinear observations and diagonal $R$, as in the reported high-dimensional experiments, $C_{\text{like}} = O(m)$. With $m \le d$, a representative structured realization thus has

$$O\big[TM^2(dm+m^3) + TMN_{\text{IS}}(d+m^2)\big] \tag{A.76}$$

analysis work, up to lower-order terms and method-dependent Jacobian costs. A matrix-free realization avoids storing the full $d\times m$ state–observation covariance and uses $O(Md+S_\Sigma+m^2)$ memory, where $S_\Sigma$ is structured-covariance storage; explicitly storing $\Sigma H^\top$ adds $O(dm)$ memory. Dense observation Jacobians or alternative linear solvers change these $m$-dependent costs. Appendix B.4 studies the diagonal specialization that removes the dense observation-space algebra.

### A.11 Error analysis for the diagonal component-evidence approximation

Fix $t<1$, a flow state $z_t$, and an observation $y$. Write $r_j := r_{J|Z_t}(j \mid z_t)$ and $\phi_j(z_1) := \phi_{1|t,j}(z_1 \mid z_t)$, and abbreviate $S_j := S_{t,j}$. Define the diagonal evidence covariance and the corresponding component probabilities by

$$\begin{aligned} D_j &:= \operatorname{Diag}(\operatorname{diag} S_j),\\ e_j &:= \int \phi_j(z_1)p_{\text{lin}}(y \mid z_1, j, z_t)\,dz_1 = \mathcal{N}\big(y; h(\mu_{1|t,j}), S_j\big)\,,\\ \widetilde{e}_j &:= \mathcal{N}\big(y; h(\mu_{1|t,j}), D_j\big)\,,\\ q_j &:= \frac{r_j e_j}{\sum_k r_k e_k}, \qquad \widetilde{q}_j := \frac{r_j \widetilde{e}_j}{\sum_k r_k \widetilde{e}_k}. \end{aligned} \tag{A.77}$$

The conditional endpoint density $q(z_1 \mid j)$ remains the Gaussian conditional in Eq. 3.10; only the categorical component probabilities are approximated.

**Diagonal evidence and corrected component weights.** Let

$$a_j := \int \phi_j(z_1)p(y \mid z_1)\,dz_1, \qquad \pi_j^a(dz_1) := \frac{\phi_j(z_1)p(y \mid z_1)}{a_j}\,dz_1. \tag{A.78}$$

The full-evidence corrected endpoint law can be written as

$$P^a(dz_1) = \sum_{j=1}^M \alpha_j \pi_j^a(dz_1), \qquad \alpha_j := \frac{r_j a_j}{\sum_k r_k a_k}. \tag{A.79}$$

For the diagonal categorical proposal, the conditional density and nonlinear correction are

$$q(z_1 \mid j) = \frac{\phi_j(z_1)p_{\text{lin}}(y \mid z_1, j, z_t)}{e_j}, \qquad g(j, z_1) = \frac{p(y \mid z_1)}{p_{\text{lin}}(y \mid z_1, j, z_t)}.$$

Consequently,

$$\begin{aligned} \widetilde{q}_j q(z_1 \mid j) g(j, z_1) &\propto r_j \widetilde{e}_j \frac{\phi_j(z_1)p_{\text{lin}}(y \mid z_1, j, z_t)}{e_j} \frac{p(y \mid z_1)}{p_{\text{lin}}(y \mid z_1, j, z_t)}\\ &= r_j \frac{\widetilde{e}_j}{e_j}\phi_j(z_1)p(y \mid z_1). \end{aligned} \tag{A.80}$$

With $c_j := \widetilde{e}_j / e_j$, the resulting population law is

$$P^a_{\text{diag}}(dz_1) = \sum_{j=1}^M \widetilde{\alpha}_j \pi_j^a(dz_1), \qquad \widetilde{\alpha}_j := \frac{r_j a_j c_j}{\sum_k r_k a_k c_k} = \frac{\alpha_j c_j}{\sum_k \alpha_k c_k}. \tag{A.81}$$

Thus diagonal evidence leaves the corrected within-component posterior laws $\pi_j^a$ unchanged and perturbs only their mixture weights.

**Evidence and component-weight error.** Let $d_j := y - h(\mu_{1|t,j})$ and $\eta_j := \log(\widetilde{e}_j/e_j)$. Directly from the two Gaussian evidences,

$$\eta_j = \frac{1}{2}\left[\log\frac{|S_j|}{|D_j|} + d_j^\top (S_j^{-1} - D_j^{-1}) d_j\right]. \tag{A.82}$$

Define the whitened off-diagonal perturbation and residual by

$$F_j := D_j^{-1/2}(S_j - D_j)D_j^{-1/2}, \qquad \xi_j := D_j^{-1/2} d_j. \tag{A.83}$$

Since $S_j = D_j^{1/2}(I + F_j)D_j^{1/2}$, Eq. A.82 is equivalently

$$\eta_j = \frac{1}{2}\left[\log\det(I + F_j) + \xi_j^\top\big((I + F_j)^{-1} - I\big)\xi_j\right]. \tag{A.84}$$

Thus the error is determined by omitted observation-space correlations through a log-determinant term and a correlated Mahalanobis-residual term. If $\rho_j := \|F_j\|_2 < 1$, then $F_j$ has zero diagonal and $\mathrm{tr}(F_j) = 0$. The eigenvalue expansion of $\log\det(I + F_j)$ and the resolvent identity give

$$|\log\det(I + F_j)| \le \frac{\|F_j\|_F^2}{2(1-\rho_j)}, \qquad \big\|(I + F_j)^{-1} - I\big\|_2 \le \frac{\rho_j}{1-\rho_j}. \tag{A.85}$$

Therefore

$$|\eta_j| \le \frac{\|F_j\|_F^2}{4(1-\rho_j)} + \frac{\rho_j}{2(1-\rho_j)}\|\xi_j\|^2. \tag{A.86}$$

The relevant perturbation measure is the oscillation

$$\omega_e(t, z_t) := \max_j \eta_j - \min_j \eta_j, \tag{A.87}$$

because a common additive shift in the log evidences cancels under categorical normalization. Centering the $\eta_j$ values within their range and applying the bounded density-ratio inequality to the normalized exponential tilt in Eq. A.81 yields

$$\|\widetilde{\alpha} - \alpha\|_1 \le 2\tanh\Big(\frac{\omega_e}{2}\Big). \tag{A.88}$$

**Endpoint, velocity, and terminal-law error.** Define

$$m_j^a(t, z_t; y) := \int z_1 \pi_j^a(dz_1), \qquad M_a(t, z_t) := \max_j \|m_j^a - \mu_{1|t}^a\|. \tag{A.89}$$

Then $\mu_{1|t}^a = \sum_j \alpha_j m_j^a$ and $\mu_{1|t}^{a,\mathrm{diag}} = \sum_j \widetilde{\alpha}_j m_j^a$. Using $\sum_j(\widetilde{\alpha}_j - \alpha_j) = 0$ together with Eq. A.88 gives

$$\|\mu_{1|t}^{a,\mathrm{diag}} - \mu_{1|t}^a\| \le 2M_a(t, z_t)\tanh\left(\frac{\omega_e(t, z_t)}{2}\right). \tag{A.90}$$

Thus the corresponding population velocities satisfy

$$\|v_t^{a,\mathrm{diag}} - v_t^a\| \le \Delta_{\mathrm{diag}}(t, z_t) := \frac{2M_a(t, z_t)}{1-t}\tanh\left(\frac{\omega_e(t, z_t)}{2}\right). \tag{A.91}$$

For the high-dimensional experiments, $R$ is diagonal. The off-diagonal part of $S_{t,j} = H_{t,j}\Sigma_{1|t}H_{t,j}^\top + R$ is therefore generated entirely by the bridge covariance term. Here the Kolmogorov-flow bridge uses the same block-structured forecast-surrogate covariance defined in Eq. A.74. Since $\Sigma_{1|t} = O((1-t)^2)$, bounded $H_{t,j}$ gives $S_j - D_j = O((1-t)^2)$. Moreover, $D_j \succeq R \succeq r_0 I$, so $\rho_j = O((1-t)^2)$. Under bounded normalized-residual moments, Eqs. A.86–A.87 give $\omega_e = O_{L^p}((1-t)^2)$. If $M_a(t, z_t)$ remains finite near the endpoint, then

$$\Delta_{\mathrm{diag}}(t, z_t) = O_{L^p}(1-t). \tag{A.92}$$

The diagonal component-evidence approximation therefore introduces no additional $1/(1-t)$-type endpoint singularity in the population velocity.

Let $Z_t$ and $Z_t^{\mathrm{diag}}$ solve the full- and diagonal-evidence population ODEs from the same source variable, and let $p^a_{\mathrm{diag}}$ denote the terminal law of the latter. If the full velocity is $L_v$-Lipschitz in the state, Grönwall's inequality under this synchronous coupling gives

$$W_1(p^a_{\mathrm{diag}}, \widetilde{p}^a) \le \varepsilon_{\mathrm{diag}}(y),$$
$$\varepsilon_{\mathrm{diag}}(y) := \int_0^1 e^{L_v(1-t)} \mathbb{E}\left[\frac{2M_a(t, Z_t^{\mathrm{diag}})}{1-t} \tanh\left(\frac{\omega_e(t, Z_t^{\mathrm{diag}})}{2}\right)\right] dt. \tag{A.93}$$

Under the same moment conditions, the endpoint contraction in Eq. A.92 makes the integrand locally integrable near $t = 1$.

Let $\widehat{p}^a_{\mathrm{diag},N_{\mathrm{IS}},T}$ denote the law produced by the corresponding finite diagonal-evidence update with $N_{\mathrm{IS}}$ endpoint proposals per velocity evaluation and $T$ Euler steps.

Under the bounded-residual nonlinear conditions of Appendix A.5 and the corresponding numerical-flow regularity conditions, the same finite-computation analysis applies to the diagonal-evidence population flow. For any $0 < \delta \le 1$ satisfying $\delta\kappa < 1$,

$$W_1\big(\widehat{p}^a_{\mathrm{diag},N_{\mathrm{IS}},T}, p^a\big) \le C_\delta^{\mathrm{diag}} N_{\mathrm{IS}}^{-\delta/(1+\delta)} + C_2^{\mathrm{diag}} T^{-1} + \varepsilon_{\mathrm{diag}}(y) + \varepsilon_{\mathrm{sur}}(y). \tag{A.94}$$

If $\kappa < 1$, the finite-sampling term is $O(N_{\mathrm{IS}}^{-1/2})$. For affine observations, the local likelihood correction has $g \equiv 1$ by Eq. A.39. The diagonal categorical evidence can still differ from the full evidence, but the same ordinary Monte Carlo argument gives

$$W_1\big(\widehat{p}^a_{\mathrm{diag},N_{\mathrm{IS}},T}, p^a\big) \le C_1^{\mathrm{diag}} N_{\mathrm{IS}}^{-1/2} + C_2^{\mathrm{diag}} T^{-1} + \varepsilon_{\mathrm{diag}}(y) + \varepsilon_{\mathrm{sur}}(y). \tag{A.95}$$

For more general unbounded nonlinear maps, the accumulated-SNIS alternative applies under its stated condition. The four terms are, respectively, finite endpoint sampling, flow discretization, diagonal component-evidence approximation, and forecast-surrogate error. Equation A.73 quantifies the local source of the diagonal approximation, but by itself does not imply a dimension-independent terminal bound on $\varepsilon_{\mathrm{diag}}(y)$.

**Collapsed-Gaussian special case.** When $\gamma = 0$, all component means and bridge distributions coincide. The full and diagonal evidences are then each independent of the component label, so $c_j = \widetilde{e}_j / e_j$ is common to all components and cancels under categorical normalization. Consequently,

$$\gamma = 0 \quad \Longrightarrow \quad \widetilde{\alpha}_j = \alpha_j, \quad \omega_e = 0, \quad \varepsilon_{\mathrm{diag}}(y) = 0. \tag{A.96}$$

This exact special case applies to the stride-2 and stride-4 Lorenz–96 regimes, which use $\gamma = 0$. The dense Lorenz–96 and Kolmogorov-flow experiments use $\gamma = 0.3$ and $\gamma = 0.2$, respectively, so no zero-error conclusion is made for those regimes.

# B Experimental Details

This appendix summarizes the dynamical systems, initialization protocols, and numerical settings used in the three benchmarks. Within each repetition, all compared methods use the same truth trajectory and observation locations. The Double-Well, Lorenz–96, and Kolmogorov-flow comparisons average $N_{\mathrm{run}} = 10$ independent trajectories and all analysis times.

**Baseline configuration.** The numerical comparisons include EnKF-MDA (Emerick & Reynolds, 2012), EnKF (Evensen, 2003), LETKF (Hunt et al., 2007), EnSF (Bao et al., 2024), and the EnFF variants (Transue et al., 2025). Hyperparameters were selected using reduced-cost pilot runs on trajectories generated independently from those used in the final reported evaluation. Within the final evaluation set, all compared methods use matched truth trajectories and observation locations. Tables 4 and 6 report covariance localization and inflation settings only for FREESIA and the Kalman-type baselines. EnSF and EnFF use their method-specific sampling/guidance hyperparameters, which are reported separately where relevant. Method-specific sampling budgets are specified in Sections B.1–B.3.

EnSF-LR is not included in the numerical comparison. Its observed-to-unobserved update relies on ensemble-covariance linear regression (Xiong et al., 2026), while the released implementation

does not include localization for this step. In our high-dimensional small-ensemble settings, the observed-state sample covariance is necessarily rank deficient, so extending EnSF-LR would require additional localization or regularization choices beyond the released formulation. We therefore discuss it as closely related work without introducing a separately modified baseline.

IEnSF is a closely related covariance-aware extension of EnSF. We implemented its published formulation but could not independently validate the resulting performance against a reference implementation; we therefore omit numerical results and restrict the comparison to methodological differences supported by the published formulation. The experiments reported here should not be interpreted as an empirical comparison with IEnSF.

### B.1 Double-Well

Following Ding et al. (2026), the double-well benchmark is

$$X_{k+1} = X_k - \delta t\, U'(X_k) + \beta\sqrt{\delta t}\, V_k, \qquad U(x) = x^4 - 2x^2,$$

where $V_k \sim \mathcal{N}(0,1)$, $\delta t = 0.1$, and $\beta = 0.8$. The initial filtering distribution is $X_0 \sim \mathcal{N}(0, 0.5^2)$. Observations are generated by

$$Y_k = |X_k - 0.4| + \epsilon_k, \qquad \epsilon_k \sim \mathcal{N}(0, 0.1^2).$$

The experiment uses 100 analysis steps and an ensemble size of 100. A 10,000-point grid on $[-2.5, 2.5]$ provides the numerical Bayes reference. FREESIA uses $\gamma = 0.9$, 20 flow steps, and $N_{\mathrm{IS}} = 10$ endpoint importance samples.

For a density $p$ and observation kink $x_0 = 0.4$, the implemented left and right mode masses are

$$P_L(p) = \int_{x<x_0} p(x)\, dx, \qquad P_R(p) = \int_{x\geq x_0} p(x)\, dx.$$

With $p_{\mathrm{ref}}$ denoting the grid Bayes posterior, the reported mode-mass error is

$$E_{\mathrm{mode}}(p, p_{\mathrm{ref}}) = \frac{1}{2}\left(|P_L(p) - P_L(p_{\mathrm{ref}})| + |P_R(p) - P_R(p_{\mathrm{ref}})|\right). \tag{B.1}$$

This posterior-shape metric is distinct from RMSE and CRPS, which evaluate state or predictive accuracy.

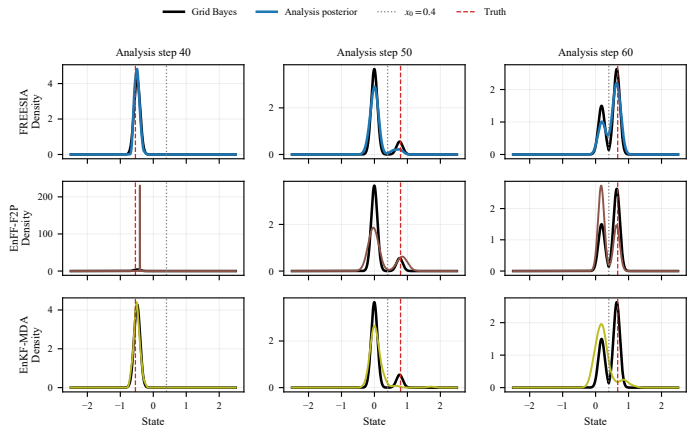


Figure 3: Double-well posterior densities at analysis steps 40, 50, and 60. Rows show FREESIA, EnFF-F2P, and EnKF-MDA against the grid Bayes posterior.

Figure 4 shows that separating the GMM components is important in this multimodal regime: performance improves markedly from the collapsed Gaussian limit, while about 20 flow steps already capture most of the gain.

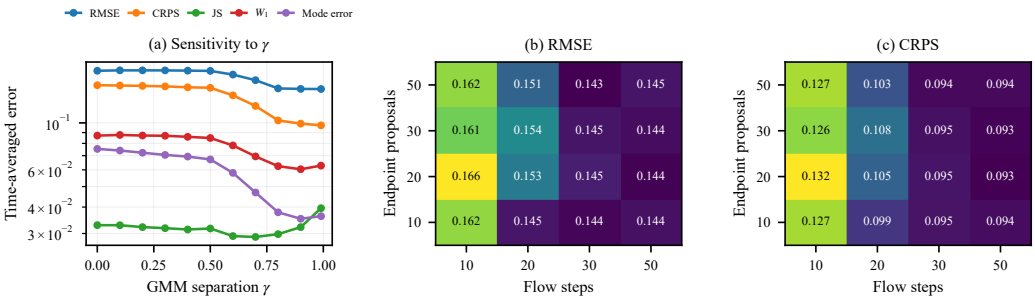


Figure 4: Double-well hyperparameter search with $N_{\text{run}} = 3$ per setting. Left: all five reported errors versus the GMM separation parameter $\gamma$, with a logarithmic vertical axis; the rightmost point is $\gamma = 0.99$. Center and right: RMSE and CRPS over flow steps and endpoint importance samples $N_{\text{IS}}$. Lower is better throughout; values average all search runs and analysis times.

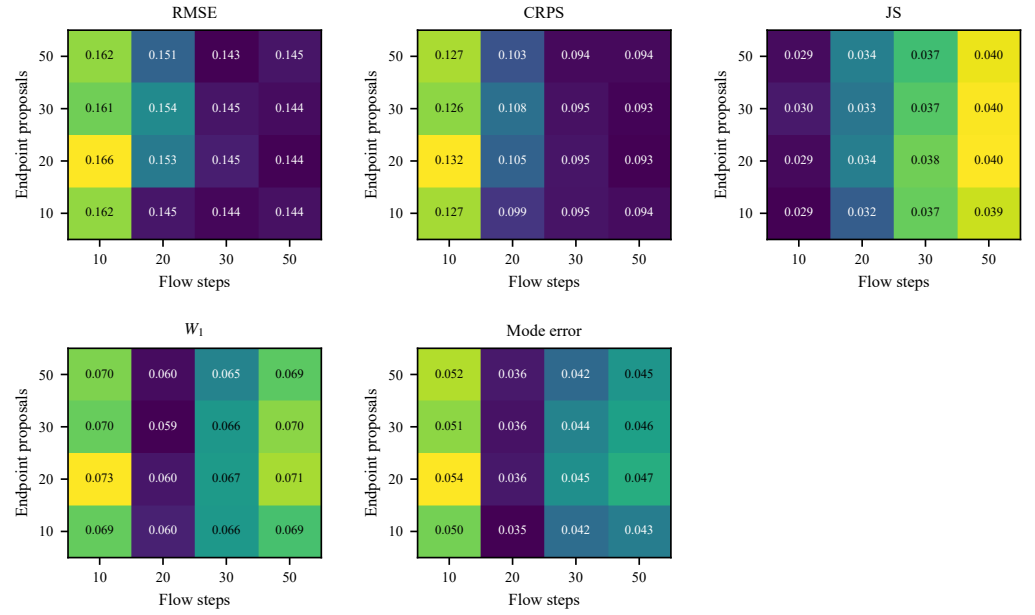


Figure 5: Complete double-well hyperparameter search over flow steps and endpoint importance samples $N_{\text{IS}}$. The panels report RMSE, CRPS, Jensen–Shannon divergence, Wasserstein-1 distance, and mode-mass error; lower is better for every metric. Cell values are averaged over the hyperparameter-search repetitions and analysis times.

### B.2 LORENZ–96

We use the Lorenz–96 system (Lorenz, 2006), whose state $X = (x_1, \ldots, x_d)^\top$ evolves according to

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \qquad i = 1, \ldots, d,$$

with cyclic indexing, $F = 8$, and $d = 1000$. Before assimilation, the system is integrated for 10 physical time units so that initialization is taken from the chaotic attractor. Truth trajectories and initial forecast ensembles are drawn from this post-spin-up regime rather than formed by small perturbations of the truth. The spin-up interval is excluded from all reported filtering statistics.

The truth model uses $\Delta t_{\mathrm{truth}} = 0.01$, while all forecast models use $\Delta t_{\mathrm{model}} = 0.02$. Observations arrive every 0.2 time units, with an ensemble size of 20 and 101 analysis times. We consider three observation regimes: dense $|\arctan(x_i)|$ observations with noise standard deviation 0.05, stride-2 $\arctan(x_{2i})$ observations with noise standard deviation 0.05, and stride-4 $\arctan(x_{4i})$ observations with noise standard deviation 0.01. The stride-4 experiment is a high-SNR sparse-observation stress test. Its directly observed coordinates are strongly constrained, making the experiment diagnostic of information transfer to the substantially larger unobserved subset; the smaller observation covariance also produces a sharper likelihood for guidance and posterior correction.

All FREESIA configurations in the main Lorenz–96 comparison use 20 flow steps and $N_{\mathrm{IS}} = 5$ endpoint importance samples. EnKF-MDA uses 20 analysis substeps, EnSF uses 500 reverse-SDE steps, and each EnFF variant uses 20 path grid points. These are the computational budgets used in the comparison rather than matched-cost settings. The dense FREESIA experiment uses $\gamma = 0.3$, whereas both sparse regimes use the collapsed single-Gaussian limit $\gamma = 0$.

**Localization and inflation.** Table 4 reports the localization radius and inflation used in the three Lorenz–96 regimes. Radius is measured in cyclic Lorenz–96 grid points. FREESIA, EnKF, EnKF-MDA, and LETKF use cyclic Gaspari–Cohn localization through their respective covariance or ensemble updates.

Table 4: Localization radius $r_{\mathrm{loc}}$ and inflation parameter $\lambda_{\mathrm{infl}}$ in the three Lorenz–96 experiments. GC denotes cyclic Gaspari–Cohn localization.

| Method | Dense abs-atan | | Stride-2 atan | | Stride-4 atan | |
|---|---|---|---|---|---|---|
| | $r_{\mathrm{loc}}$ | $\lambda_{\mathrm{infl}}$ | $r_{\mathrm{loc}}$ | $\lambda_{\mathrm{infl}}$ | $r_{\mathrm{loc}}$ | $\lambda_{\mathrm{infl}}$ |
| FREESIA | 4 | 1.8 | 4 | 1.2 | 2 | 1.0 |
| EnKF | 2 | 1.1 | 4 | 1.2 | 3 | 1.1 |
| EnKF-MDA | 3 | 1.6 | 3 | 1.1 | 2 | 1.2 |
| LETKF | 2 | 1.2 | 2 | 1.5 | 2 | 1.3 |

**Supplementary diagnostics.** The observed/unobserved RMSE decomposition provides additional diagnostics for the two sparse-observation regimes. Representative coordinate-wise trajectories are reported for the dense and sparse regimes.

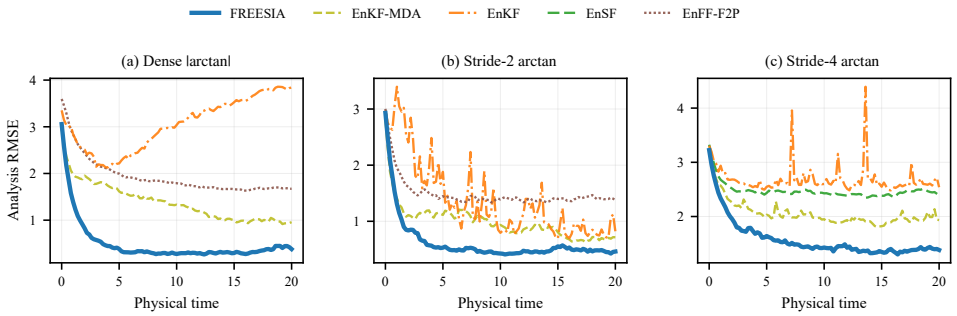


Figure 6: Total analysis RMSE for dense absolute-arctangent, stride-2 arctangent, and stride-4 arctangent observations. Each panel shows the four methods with the lowest time-averaged RMSE in that regime; the shared legend lists the union of methods shown across panels.

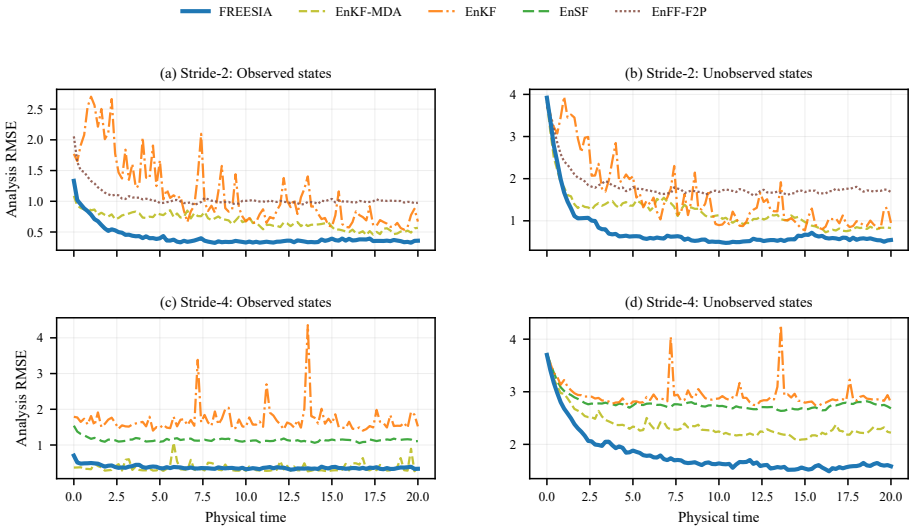


Figure 7: Observed- and unobserved-coordinate RMSE for stride-2 (top) and stride-4 (bottom) observations. The right column directly evaluates information transfer to unobserved state coordinates.

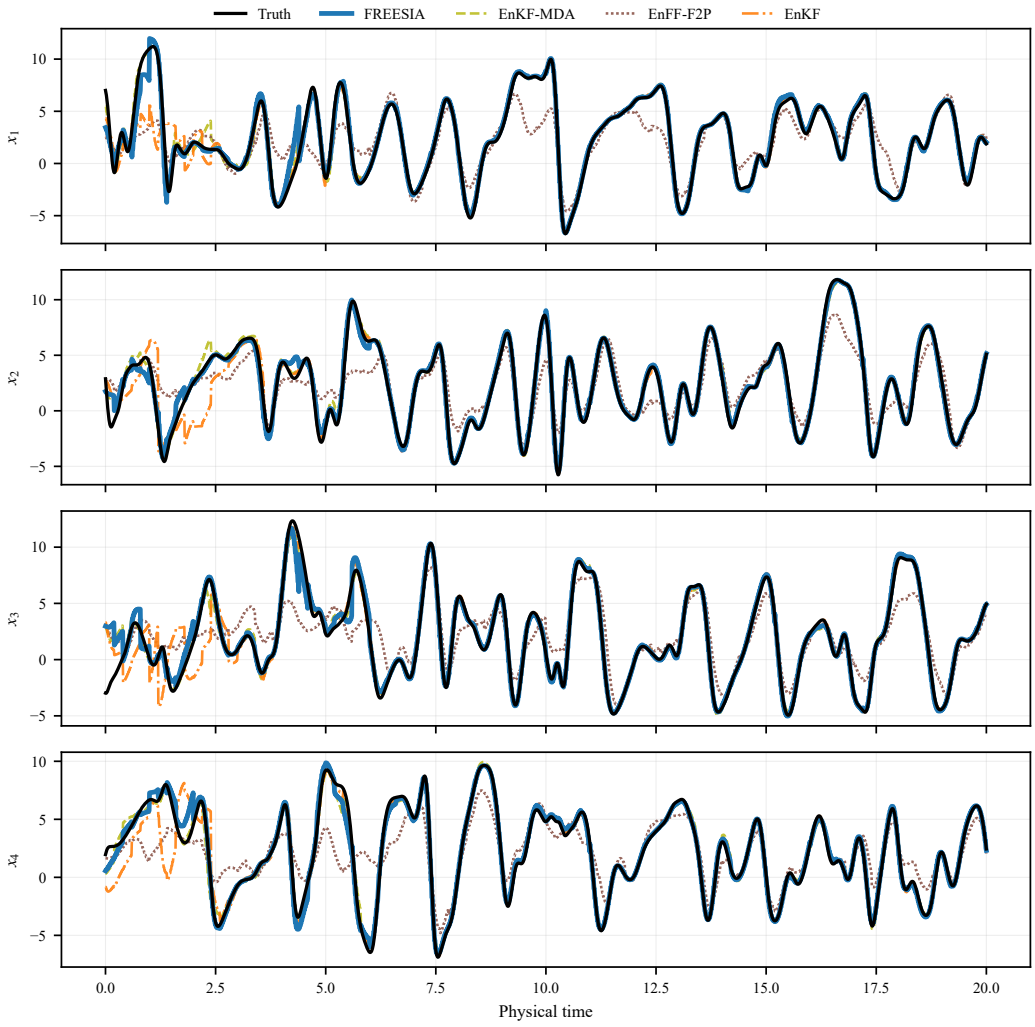


Figure 8: Representative trajectories for the first four Lorenz–96 state coordinates under dense absolute-arctangent observations. The panels show the truth and analysis trajectories for the methods selected by the corresponding RMSE comparison.

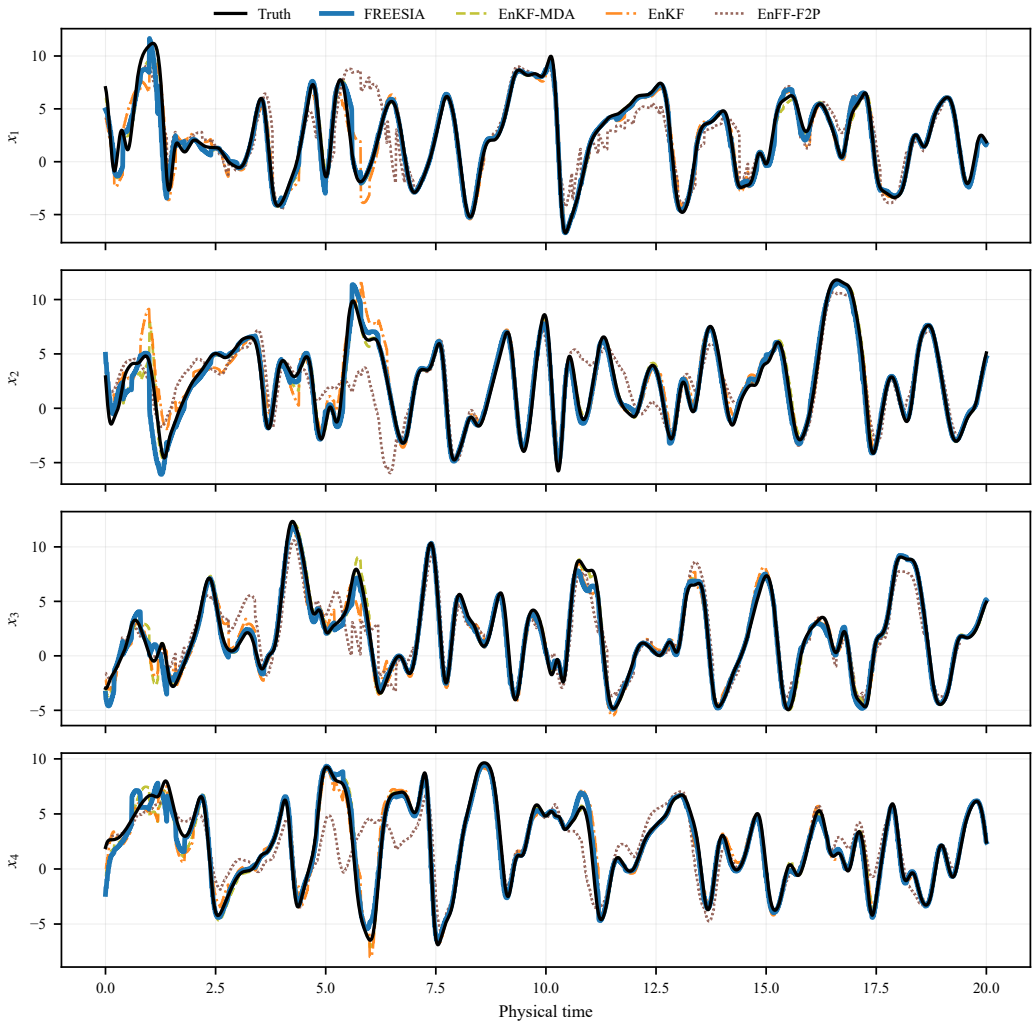


Figure 9: Representative trajectories for the first four Lorenz–96 state coordinates under stride-2 arctangent observations. The panels show the truth and analysis trajectories for the methods selected by the corresponding RMSE comparison.

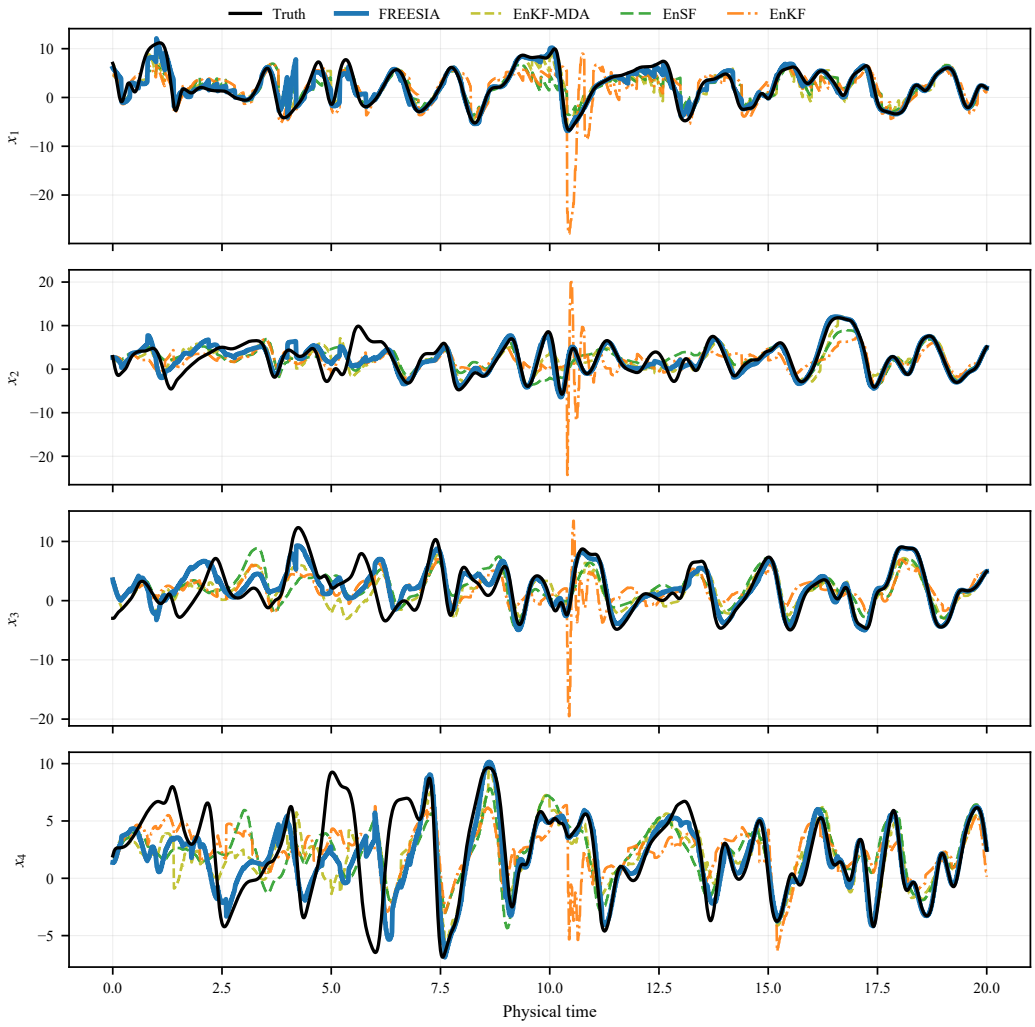


Figure 10: Representative trajectories for the first four Lorenz–96 state coordinates under stride-4 arctangent observations. The panels show the truth and analysis trajectories for the methods selected by the corresponding RMSE comparison.

**Stride-4 ablation.** We repeat the stride-4 experiment with two targeted variants. The first replaces the localized forecast covariance by its diagonal, removing cross-coordinate covariances while retaining marginal variances. The second replaces the nonlinear likelihood-ratio weights by uniform endpoint weights. Across the same 10 initializations and 101 analyses, the time-averaged RMSE increases from 1.575 for FREESIA to 2.352 with diagonal covariance and 2.969 without likelihood correction. The diagonal variant mainly degrades the unobserved coordinates (2.702 versus 1.805), while the no-correction variant decreases initially and then accumulates error, reaching final-time RMSE 3.728 versus 1.382 for FREESIA. These changes separately support the roles of cross-covariance transfer and nonlinear proposal correction in repeated sparse assimilation.

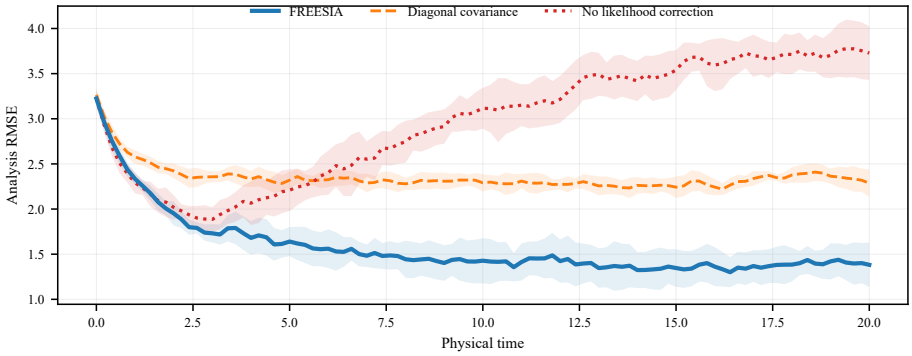


Figure 11: Lorenz–96 stride-4 ablation. Curves show mean analysis RMSE over 10 matched runs, and shading denotes one standard deviation across runs. Diagonal covariance removes cross-coordinate covariance terms. The no-likelihood-correction variant replaces the nonlinear likelihood-ratio weights with uniform endpoint weights.

**Noise robustness.** As a robustness check, we repeat the stride-4 experiment with observation-noise standard deviation $\sigma = 0.05$. FREESIA remains the lowest-RMSE method, showing that the sparse-transfer advantage is not specific to the low-noise setting. The performance gaps among several baselines become smaller, and some methods improve relative to the $\sigma = 0.01$ regime, suggesting that the sharper likelihood in the high-SNR experiment contributes to the difficulty of the assimilation problem.

Table 5: Time-averaged total, observed-coordinate, and unobserved-coordinate RMSE and CRPS for Lorenz–96 stride-4 observations under two noise levels. Lower is better.

| | $\sigma = 0.01$ | | | | $\sigma = 0.05$ | | | |
|---|---|---|---|---|---|---|---|---|
| Method | RMSE | Obs. | Unobs. | CRPS | RMSE | Obs. | Unobs. | CRPS |
| FREESIA | **1.575** | 0.373 | **1.805** | **0.584** | **2.076** | 0.938 | **2.335** | **1.013** |
| EnKF-MDA | 2.056 | **0.366** | 2.358 | 0.955 | 2.544 | **0.894** | 2.891 | 1.342 |
| EnKF | 2.684 | 1.689 | 2.928 | 1.356 | 2.642 | 1.603 | 2.903 | 1.407 |
| EnSF | 2.472 | 1.143 | 2.777 | 1.242 | 2.865 | 1.937 | 3.113 | 1.582 |
| EnFF-F2P | 3.960 | 4.412 | 3.796 | 2.243 | 2.893 | 2.263 | 3.074 | 1.565 |
| LETKF | 3.185 | 1.853 | 3.513 | 1.648 | 3.069 | 1.756 | 3.391 | 1.654 |
| EnFF-OT | 3.749 | 2.637 | 4.051 | 2.636 | 4.751 | 4.573 | 4.807 | 3.447 |

**Accuracy–cost trade-off.** Figure 12 fixes $N_{\mathrm{IS}} = 5$ and varies only $T \in \{5, 10, 20, 50\}$, while baseline points use their reported configurations. Increasing the number of flow steps from 5 to 20 progressively reduces filtering error, whereas further increasing the flow resolution to 50 yields negligible additional RMSE reduction despite substantially higher analysis cost. The default $T = 20$ setting therefore lies near the empirical accuracy–cost knee. At a comparable analysis time, FREESIA substantially improves over EnSF, providing evidence that the observed accuracy gain is not solely attributable to a larger computational budget.

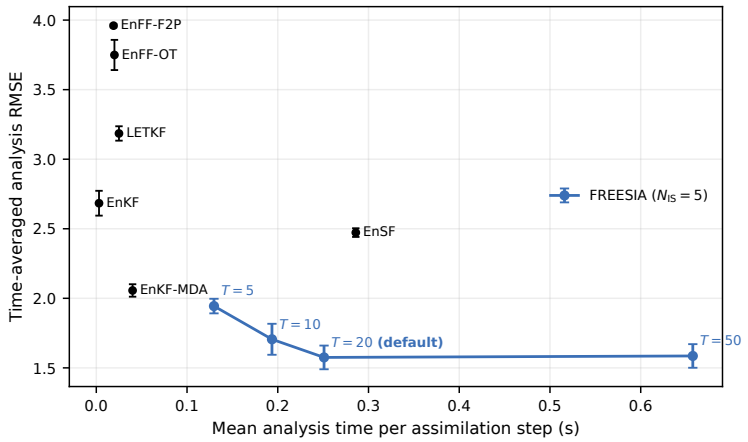


Figure 12: Accuracy–cost trade-off for Lorenz–96 stride-4 assimilation. FREESIA fixes $N_{\mathrm{IS}} = 5$ and varies only the number of flow steps $T$; error bars denote one standard deviation across repeated runs. Baseline points use the configurations reported in the main comparison.

**Importance-weight diagnostics.** For normalized endpoint weights $w^{(s)}$, define

$$\mathrm{ESS} = \frac{1}{\sum_{s=1}^{N_{\mathrm{IS}}} (w^{(s)})^2}, \qquad \mathrm{nESS} = \frac{\mathrm{ESS}}{N_{\mathrm{IS}}} = \frac{1}{N_{\mathrm{IS}} \sum_s (w^{(s)})^2},$$

so $1/N_{\mathrm{IS}} \le \mathrm{nESS} \le 1$. The higher-resolution diagnostic uses $T = 50$ and $N_{\mathrm{IS}} = 20$, a finer flow discretization and more endpoint proposals, only to resolve the near-endpoint behavior of the importance weights; the default configurations used for the main accuracy comparisons are unchanged. In the stride-4 small-noise regime, importance weights are strongly concentrated over the early and intermediate portions of the flow. The concentration decreases rapidly near $t = 1$, consistent with the endpoint contraction analysis in Appendix A.5. Across the filtering sequence, we do not observe systematic deterioration of the weight behavior at fixed flow times. These diagnostics are qualitative and do not verify the accumulated-SNIS condition in Eq. A.45.

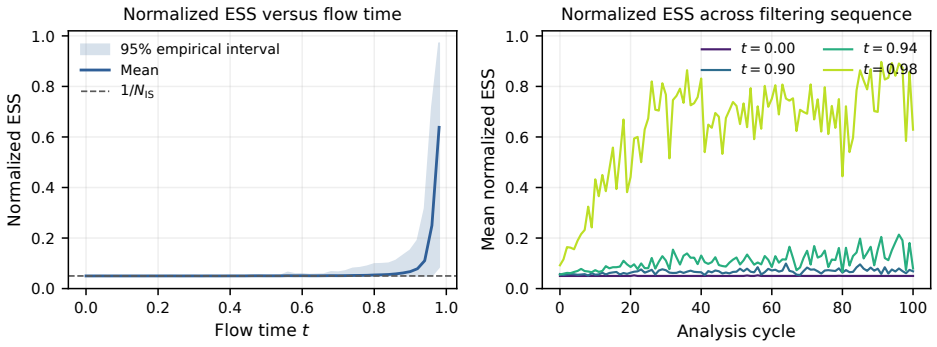


Figure 13: Importance-weight diagnostics for the Lorenz–96 stride-4 experiment. Left: normalized effective sample size (nESS) as a function of flow time, aggregated over filtering states and flow particles. Right: nESS across the filtering sequence at selected flow times. Importance weights are strongly concentrated over the early and intermediate portions of the flow, while nESS rises rapidly near the endpoint as the conditional bridge contracts. The concentration is primarily flow-time dependent rather than showing systematic deterioration across the filtering sequence. The higher-resolution configuration is used only to resolve the near-endpoint weight behavior.

### B.3 KOLMOGOROV FLOW

The high-dimensional benchmark uses the periodically forced incompressible Navier–Stokes equations (Temam, 2001),

$$\partial_t u + (u \cdot \nabla)u = -\nabla p + \mathrm{Re}^{-1}\Delta u - \alpha u + f, \qquad \nabla \cdot u = 0,$$

on a periodic $128 \times 128$ grid over $[0, 2\pi]^2$. The two velocity components give $d = 32768$, and the Reynolds number is $\mathrm{Re} = 1000$. The linear damping coefficient is $\alpha = 0.1$, and the velocity-form Kolmogorov forcing is $f(x, y) = (\sin(4y), 0)^\top$, with unit amplitude and forcing wavenumber four. Equivalently, the forcing contributes $-4\cos(4y)$ to the vorticity equation. The numerical time step is determined automatically from a Courant stability condition.

Before assimilation, the flow is integrated for 10 physical time units so that both truth trajectories and initial forecast ensembles are initialized from the developed-flow regime. The spin-up interval is excluded from the reported filtering statistics.

Both observation regimes use an ensemble size of 20, an observation interval of approximately 0.2, 50 forecast–analysis intervals, and ten independent trajectories. The arctangent experiment observes 328 randomly selected velocity entries (approximately 1%), whereas the absolute-arctangent experiment observes 656 entries (approximately 2%). All compared methods use the same truth trajectories and observation locations. For the selected velocity entries, the observation noise is independent Gaussian noise, $\epsilon_k \sim \mathcal{N}(0, 0.05^2 I)$.

FREESIA uses 20 flow steps, $N_{\mathrm{IS}} = 5$ endpoint importance samples, $\gamma = 0.2$, localization radius $r_{\mathrm{loc}} = 16$, and covariance inflation 1.1 in both reported observation regimes. The multi-endpoint construction permits nonuniform likelihood-ratio weights while retaining the covariance-aware observation-adaptive proposal. The matched Lorenz–96 stride-4 ablation in Appendix B.2 isolates the contribution of nonlinear likelihood correction.

Table 6 reports the covariance settings used by FREESIA and the Kalman-type baselines. All seven methods complete the ten repeated assimilation sequences in both observation regimes.

**Time-resolved RMSE.** Figure 14 reports the analysis trajectories for the available methods. Each panel retains the four methods with the lowest time-averaged RMSE in that regime.

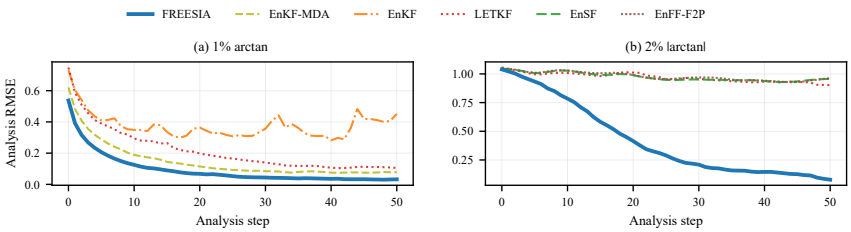


Figure 14: Analysis RMSE for 1% arctangent observations (left) and 2% absolute-arctangent observations (right). Each panel shows the four methods with the lowest time-averaged RMSE across the ten trajectories.

**Accuracy–cost trade-off.** Figure 15 reports an accuracy–cost sweep for the 1% arctangent Kolmogorov-flow regime. The sweep fixes $N_{\mathrm{IS}} = 1$, $r_{\mathrm{loc}} = 16$, and the remaining FREESIA hyperparameters, and varies only the number of flow steps $T \in \{5, 10, 20\}$. The default benchmark configuration ($T = 20, N_{\mathrm{IS}} = 5$), which uses the same localization radius, is included as the rightmost reference point. Filtering error decreases rapidly over the low-to-moderate budget range and then saturates; the default configuration incurs additional cost for comparatively modest further improvement in this monotone-observation regime.

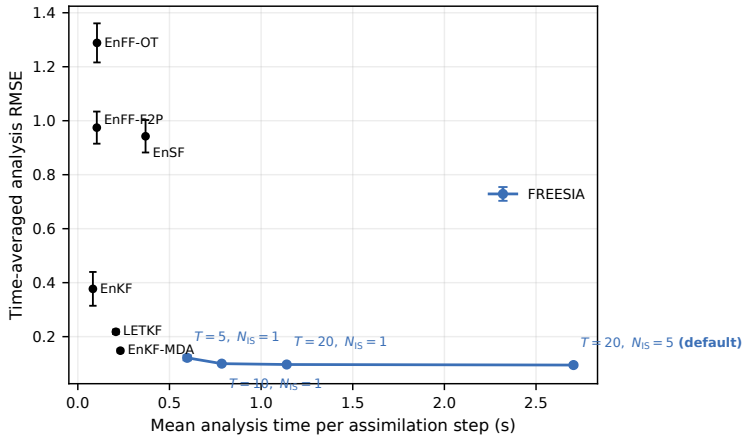


Figure 15: Accuracy–cost trade-off on 2-D Kolmogorov flow with 1% random arctangent observations. The sweep fixes $N_{\mathrm{IS}} = 1$ and $r_{\mathrm{loc}} = 16$, and varies only the number of flow steps $T$. The rightmost reference point is the default benchmark configuration ($T = 20, N_{\mathrm{IS}} = 5$), with the same localization radius. Error bars denote one standard deviation across ten trajectories. Baseline points use their reported configurations.

Table 6: Localization radius $r_{\mathrm{loc}}$ and covariance inflation $\lambda_{\mathrm{infl}}$ for the Kolmogorov-flow comparisons.

| | 1% atan | | 2% abs-atan | |
|---|---|---|---|---|
| Method | $r_{\mathrm{loc}}$ | $\lambda_{\mathrm{infl}}$ | $r_{\mathrm{loc}}$ | $\lambda_{\mathrm{infl}}$ |
| FREESIA | 16 | 1.1 | 16 | 1.1 |
| EnKF | 8 | 1.1 | 6 | 1.1 |
| EnKF-MDA | 10 | 1.2 | 6 | 1.1 |
| LETKF | 3 | 1.1 | 2 | 1.1 |

**Statistical uncertainty.** Tables 7–9 report means and standard deviations of per-run time-averaged metrics. Standard deviations describe variability across independent runs and are not confidence intervals.

Table 7: Double-Well metrics (mean $\pm$ standard deviation over 10 runs).

| Method | JS | $W_1$ | Mode error | RMSE | CRPS |
|---|---|---|---|---|---|
| FREESIA | $0.03134 \pm 0.00047$ | $0.05936 \pm 0.00053$ | $0.03503 \pm 0.00009$ | $0.14417 \pm 0.00035$ | $0.09915 \pm 0.00002$ |
| EnKF-MDA | $0.18521 \pm 0.00001$ | $0.31992 \pm 0.00004$ | $0.25878 \pm 0.00001$ | $0.43262 \pm 0.00003$ | $0.37445 \pm 0.00001$ |
| EnKF | $0.18365 \pm 0.00035$ | $0.23281 \pm 0.00050$ | $0.15807 \pm 0.00010$ | $0.32489 \pm 0.00051$ | $0.25673 \pm 0.00017$ |
| EnSF | $0.31497 \pm 0.00161$ | $0.66648 \pm 0.00404$ | $0.35924 \pm 0.00398$ | $0.69883 \pm 0.00261$ | $0.37946 \pm 0.00186$ |
| EnFF-OT | $0.54998 \pm 0.00007$ | $0.20219 \pm 0.00003$ | $0.12996 \pm 0.00001$ | $0.26335 \pm 0.00004$ | $0.25411 \pm 0.00004$ |
| EnFF-F2P | $0.35776 \pm 0.00115$ | $0.15151 \pm 0.00243$ | $0.08547 \pm 0.00072$ | $0.23215 \pm 0.00183$ | $0.16993 \pm 0.00080$ |

Table 8: Lorenz–96 RMSE and CRPS (mean $\pm$ standard deviation over 10 runs). For the dense observation regime, all coordinates are observed, so only total RMSE and CRPS are reported. For the sparse regimes, total, observed-coordinate, and unobserved-coordinate RMSE are reported together with CRPS.

| | Dense abs-atan | | Stride-2 atan | | | | Stride-4 atan | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | RMSE | CRPS | RMSE | Obs. | Unobs. | CRPS | RMSE | Obs. | Unobs. | CRPS |
| FREESIA | $0.490 \pm 0.038$ | $0.174 \pm 0.007$ | $0.615 \pm 0.060$ | $0.417 \pm 0.025$ | $0.757 \pm 0.082$ | $0.241 \pm 0.014$ | $1.575 \pm 0.085$ | $0.373 \pm 0.017$ | $1.805 \pm 0.098$ | $0.584 \pm 0.041$ |
| EnKF-MDA | $1.384 \pm 0.255$ | $0.586 \pm 0.107$ | $0.987 \pm 0.243$ | $0.673 \pm 0.153$ | $1.211 \pm 0.307$ | $0.232 \pm 0.006$ | $2.056 \pm 0.045$ | $0.366 \pm 0.021$ | $2.358 \pm 0.051$ | $0.955 \pm 0.024$ |
| EnKF | $3.053 \pm 0.155$ | $1.712 \pm 0.148$ | $1.380 \pm 0.153$ | $1.120 \pm 0.141$ | $1.573 \pm 0.168$ | $0.422 \pm 0.026$ | $2.684 \pm 0.089$ | $1.689 \pm 0.130$ | $2.928 \pm 0.078$ | $1.356 \pm 0.020$ |
| LETKF | $6.161 \pm 1.161$ | $3.921 \pm 0.329$ | $1.533 \pm 0.225$ | $1.243 \pm 0.192$ | $1.753 \pm 0.253$ | $0.494 \pm 0.051$ | $3.185 \pm 0.052$ | $1.853 \pm 0.069$ | $3.513 \pm 0.052$ | $1.648 \pm 0.022$ |
| EnSF | $3.901 \pm 0.012$ | $2.340 \pm 0.008$ | $2.027 \pm 0.051$ | $1.531 \pm 0.036$ | $2.422 \pm 0.064$ | $0.976 \pm 0.035$ | $2.472 \pm 0.031$ | $1.143 \pm 0.017$ | $2.777 \pm 0.035$ | $1.242 \pm 0.020$ |
| EnFF-OT | $4.341 \pm 0.206$ | $2.959 \pm 0.213$ | $2.817 \pm 0.085$ | $2.303 \pm 0.069$ | $3.247 \pm 0.100$ | $1.825 \pm 0.076$ | $3.749 \pm 0.109$ | $2.637 \pm 0.073$ | $4.051 \pm 0.119$ | $2.636 \pm 0.101$ |
| EnFF-F2P | $1.933 \pm 0.054$ | $0.721 \pm 0.029$ | $1.486 \pm 0.042$ | $1.045 \pm 0.029$ | $1.821 \pm 0.053$ | $0.609 \pm 0.024$ | $3.960 \pm 0.013$ | $4.412 \pm 0.012$ | $3.796 \pm 0.017$ | $2.243 \pm 0.008$ |

Table 9: Kolmogorov-flow total, observed-coordinate, and unobserved-coordinate RMSE and CRPS (mean $\pm$ standard deviation over ten trajectories).

| | 1% atan | | | | 2% abs-atan | | | |
|---|---|---|---|---|---|---|---|---|
| Method | All | Obs. | Unobs. | CRPS | All | Obs. | Unobs. | CRPS |
| FREESIA | $0.095 \pm 0.005$ | $0.076 \pm 0.007$ | $0.095 \pm 0.005$ | $0.052 \pm 0.003$ | $0.422 \pm 0.101$ | $0.410 \pm 0.105$ | $0.422 \pm 0.101$ | $0.209 \pm 0.036$ |
| EnKF-MDA | $0.148 \pm 0.007$ | $0.115 \pm 0.012$ | $0.148 \pm 0.007$ | $0.081 \pm 0.005$ | $1.121 \pm 0.088$ | $1.115 \pm 0.105$ | $1.121 \pm 0.088$ | $0.797 \pm 0.070$ |
| EnKF | $0.377 \pm 0.063$ | $0.372 \pm 0.070$ | $0.377 \pm 0.063$ | $0.200 \pm 0.051$ | $0.987 \pm 0.123$ | $0.984 \pm 0.123$ | $0.987 \pm 0.123$ | $0.681 \pm 0.100$ |
| LETKF | $0.218 \pm 0.008$ | $0.177 \pm 0.012$ | $0.218 \pm 0.008$ | $0.117 \pm 0.006$ | $0.970 \pm 0.062$ | $1.088 \pm 0.072$ | $0.967 \pm 0.062$ | $0.572 \pm 0.031$ |
| EnSF | $0.942 \pm 0.060$ | $0.568 \pm 0.067$ | $0.945 \pm 0.060$ | $0.547 \pm 0.035$ | $0.974 \pm 0.053$ | $0.967 \pm 0.085$ | $0.975 \pm 0.052$ | $0.569 \pm 0.030$ |
| EnFF-OT | $1.288 \pm 0.072$ | $1.148 \pm 0.081$ | $1.290 \pm 0.072$ | $1.024 \pm 0.060$ | $1.342 \pm 0.041$ | $1.342 \pm 0.052$ | $1.342 \pm 0.041$ | $1.069 \pm 0.030$ |
| EnFF-F2P | $0.974 \pm 0.060$ | $0.776 \pm 0.081$ | $0.976 \pm 0.059$ | $0.567 \pm 0.034$ | $0.980 \pm 0.051$ | $0.977 \pm 0.069$ | $0.980 \pm 0.051$ | $0.570 \pm 0.028$ |

**Runtime.** Table 10 reports mean analysis time per assimilation step for the reported Lorenz–96 and Kolmogorov-flow configurations. The Kolmogorov-flow entries use the main $N_{\mathrm{IS}} = 5$ configuration. Timings are measured on a laptop equipped with an NVIDIA GeForce RTX 5070 Laptop GPU.

Table 10: Mean analysis time per assimilation step in seconds.

| Method | L96 dense | L96 stride-2 | L96 stride-4 | KF $\arctan$ | KF $\lvert\arctan\rvert$ |
|---|---|---|---|---|---|
| **FREESIA** (ours) | 1.377 | 0.478 | 0.251 | 2.704 | 3.397 |
| EnKF-MDA | 0.216 | 0.083 | 0.040 | 0.232 | 0.394 |
| EnKF | 0.012 | 0.005 | 0.003 | 0.082 | 0.194 |
| LETKF | 0.009 | 0.021 | 0.025 | 0.208 | 0.200 |
| EnSF | 0.343 | 0.301 | 0.286 | 0.369 | 0.385 |
| EnFF-OT | 0.021 | 0.022 | 0.020 | 0.105 | 0.102 |
| EnFF-F2P | 0.020 | 0.020 | 0.019 | 0.103 | 0.101 |

### B.4 Computational Complexity and High-Dimensional Scaling

We further examine computational scaling using fully observed Lorenz–96,

$$y_i = \arctan(x_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, 0.05^2),$$

with $d \in \{10^4, 10^5, 10^6\}$ and ensemble size $M = 20$. For each dimension, three independent runs use matched initializations and observation-noise sequences across methods, with 51 analysis times per run. We compare EnSF and EnFF-F2P with **FREESIA-Diag**, a lightweight specialization of FREESIA for extremely high-dimensional fully observed settings. FREESIA-Diag uses a diagonal forecast covariance, $\gamma = 0$, $N_{\mathrm{IS}} = 1$, and $T = 20$; EnFF-F2P uses 20 flow points, and EnSF uses 50 reverse-diffusion steps.

Table 11: Leading analysis complexity of the scaling-study implementations. $M$, $T$, $d$, and $m$ denote ensemble size, generative steps, state dimension, and observation dimension, respectively. For structured FREESIA, $C_{\mathrm{cond}}$, $C_{\mathrm{end}}$, and $C_{\mathrm{like}}$ are defined in Appendix A.10; $S_\Sigma$ denotes structured-covariance storage.

| Method | Computation | Memory |
|---|---|---|
| EnSF (diagonal score) | $O(MTd)$ | $O(Md)$ |
| EnFF-F2P | $O(M^2Td)$ | $O(Md + M^2)$ |
| FREESIA (structured) | $O(TM^2C_{\mathrm{cond}} + TMN_{\mathrm{IS}}(C_{\mathrm{end}} + C_{\mathrm{like}}))$ | $O(Md + S_\Sigma + m^2)$ |
| FREESIA-Diag | $O(MTd)$ | $O(Md)$ |

Table 11 separates the general structured implementation from the diagonal specialization used in this scaling study. As detailed in Appendix A.10, general FREESIA retains explicit dependence on observation dimension through the $m \times m$ innovation covariance $S_{t,j}$. Under structured covariance actions, row-sparse observations, and a direct dense observation-space solve, $C_{\mathrm{cond}} = O(dm+m^3)$, $C_{\mathrm{end}} = O(d + m^2)$, and $C_{\mathrm{like}} = O(m)$ for diagonal $R$. In FREESIA-Diag, the forecast covariance and component-wise observation Jacobian $H = \mathrm{Diag}(h'_1, \ldots, h'_d)$ are diagonal, as is $R$; hence $S_{t,j} = H\Sigma_{1|t}H^\top + R$ is diagonal and its solve reduces to coordinatewise scalar operations. Together with $\gamma = 0$, which makes the mixture components coincide, this gives $O(MTd)$ computation and $O(Md)$ memory for fixed $N_{\mathrm{IS}}$, even when $m = d$.

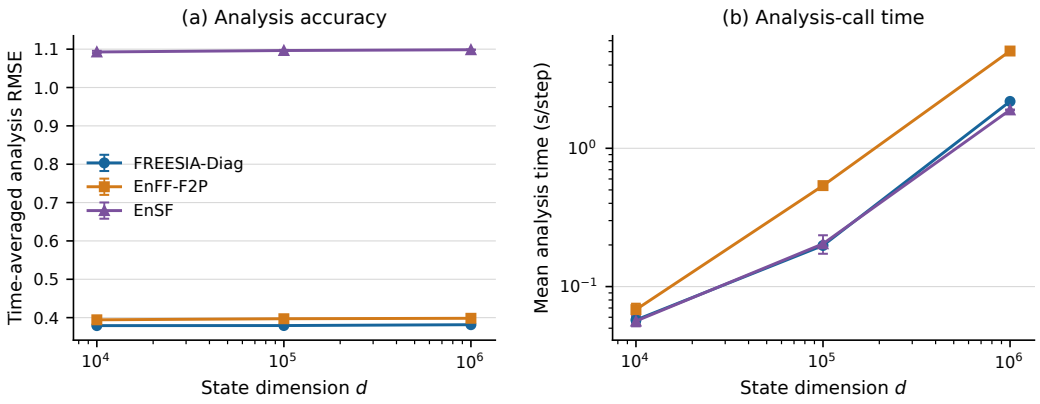


Figure 16: Fully observed Lorenz–96 scaling with $M = 20$. Left: time-averaged analysis RMSE. Right: mean analysis time per assimilation step. Each point averages 51 analyses over three independent runs; error bars denote one standard deviation across run means.

Figure 16 reports the corresponding accuracy and analysis-time scaling. All three implementations scale to $d = 10^6$ in the tested configuration. FREESIA-Diag maintains a nearly dimension-independent time-averaged RMSE of 0.379–0.382; at $d = 10^6$, the RMSEs are 0.382, 0.398, and 1.098 for FREESIA-Diag, EnFF-F2P, and EnSF, respectively. Its analysis time grows approximately linearly with state dimension over the tested range, consistent with the $O(MTd)$ complexity in Table 11. These results show that the diagonal specialization preserves useful filtering accuracy while extending FREESIA to million-dimensional state spaces with linear state-dimensional storage and computation.