
MULTISCALE NUDGING: FROM MACROSCOPIC OBSERVATIONS TO MICROSCOPIC DYNAMICS

Liyao Lyu

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90024, USA
lyuliyao@math.ucla.edu

Xinyue Yu

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90024, USA
tracy@math.ucla.edu

Hayden Schaeffer

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90024, USA
hayden@math.ucla.edu

ABSTRACT

We introduce a measure-based nudging framework for assimilating macroscopic observations into microscopic mean-field particle dynamics. The central difficulty is a representation mismatch: the forecast is a labeled particle system, while the observations specify only a smoothed, permutation-invariant density. To address this mismatch, we define the forecast-observation discrepancy as a quadratic functional on probability measures after applying the same smoothing operator used by the observation process. The Wasserstein gradient of this functional induces a transport velocity on state space, which yields a particle-level correction without constructing particle-to-particle matching, linearizing the dynamics, or estimating ensemble covariances. For a fixed observation scale, we prove well-posedness of the assimilated McKean–Vlasov dynamics and propagation of chaos for the interacting particle approximation. Under exact smoothed observations and an observability condition at the kernel scale, we establish an L^2 -stability estimate showing exponential decay up to a bias floor controlled by model misspecification. Numerical experiments on linear, bimodal, chaotic, kinetic, and collective-motion systems demonstrate that the method can recover macroscopic structure from incomplete density-level observations.

Keywords mean field approximation · interacting particle models · data assimilation · multiscale dynamics

1 Introduction

Mean-field particle systems are widely used tools in modeling many natural and engineered systems with multiscale phenomena, which arise in fluid dynamics, neuroscience, materials science, and biological systems [1, 2]. These multiscale systems typically involve a large number of degrees of freedom, making direct analysis and simulation at the microscopic level prohibitively expensive. Introduced by [3, 4], mean-field theory has provided a way to connect microscopic interactions with macroscopic dynamics. The mean-field limit replaces the high-dimensional coupled dynamics of N interacting agents with a single representative equation whose drift depends on the law of

the process itself, a formulation known as the McKean-Vlasov stochastic differential equation. On the theoretical side, substantial progress has been made on well-posedness [5, 6], existence and uniqueness of solutions [7], and propagation of chaos [8, 9, 10]. On the applied side, mean-field models have found successful applications in areas ranging from mean-field control [11] and mean-field games [12] to high-dimensional sampling [13, 14] and neural network training [15].

Although mean-field modeling has achieved broad success, several practical challenges remain. In many complex systems, the true mean-field dynamics are unknown, and thus a learned or approximated drift is needed. One approach is to approximate the governing system [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27], which can often introduce bias during forecasting. This issue is often observed in strongly nonlinear, heterogeneous, or transitional regimes, which are difficult to capture using simplified effective dynamics alone [8, 28, 29]. Even when the dominant mechanisms are reasonably well understood, uncertainty in initialization and unresolved interactions can accumulate over time, producing significant forecast drift.

A natural way to reduce such forecast drift is data assimilation [30, 31, 32, 33, 34], which uses observational information to constrain the forecast and correct departures from the reference dynamics. In the mean-field problems considered here, however, the forecast and the observations are represented at different physical scales. The model evolves microscopic particles or agents, such as atoms, molecules, cells, or individual animals, whereas the available data often provide only macroscopic information. In image-based settings, for example, pixel intensities or spatial averages describe a coarse-grained density field rather than the positions and identities of individual particles [35, 36]. Direct access to individual particle positions is therefore unavailable. This creates a cross-scale assimilation problem. The forecast contains a finite particle configuration, but the observation specifies only how mass is distributed at the resolution of the measurement device. Hence, the data do not determine a unique microscopic configuration, nor do they provide a canonical correspondence between observed mass and simulated particles. Even when particle locations can be extracted from measurements, the resulting tracks are often unlabeled, partially observed, or inconsistent across time. The assimilation procedure must therefore compare forecast and data at the observational scale, rather than through direct particle-to-particle matching.

Classical data assimilation has been a powerful tool, particularly in weather prediction and geophysical forecasting. The Ensemble Kalman Filter (EnKF) [37, 38] and its many variants approximate the Bayesian filtering distribution by propagating an ensemble of model states and updating them via a Kalman-type correction. In the present setting, however, the relevant observation is a permutation-invariant functional of the empirical distribution, whereas the standard EnKF update is built from covariances between labeled state coordinates. This can introduce an artificial dependence on particle indexing unless one imposes additional structure, such as reliable particle identities, a canonical ordering, or an explicit matching rule. Variational methods such as 3D-Var and 4D-Var [39, 40] can incorporate nonlinear observation operators, but using coarse density observations would require optimizing over microscopic particle configurations whose macroscopic density matches the data. This lifting from density space back to particle space is highly nonunique. Particle filters [41] provide a fully nonlinear Bayesian alternative, but they suffer from weight degeneracy in high-dimensional systems and do not by themselves resolve the permutation-invariant nature of density observations. More recently, feedback-based approaches, commonly referred to as nudging or synchronization methods [42, 43], have attracted much attention due to their simplicity and suitability to rigorous analysis [44, 45, 46, 47]. In nudging, a relaxation term continuously drives the model state toward consistency with observations, and convergence can often be established under verifiable spectral-gap or dissipativity conditions [48, 49]. Between these classical filters and nudging lies a family of transport- and coupling-based mean-field filters that move particles by a deterministic feedback velocity [50, 51, 52, 53, 54, 55, 56], each of which assumes a likelihood or observation operator acting on the labeled state, rather than on the permutation-invariant density considered here. However, in the present setting, where the state is represented by a finite ensemble of particles while the observations correspond to coarse-grained density fields, a fundamental issue remains: the discrepancy between model and data is naturally defined at the level of measures, rather than labeled states. This mismatch makes a standard L^2 residual unsuitable for comparing particle-based forecasts with coarse-grained density observations and may result in ill-posed or numerically unstable formulations.

This motivates a formulation of the assimilation problem at the level of measures. Rather than constructing a correction in a labeled N -particle state space, we define the forecast-observation discrepancy as a functional of the forecast law. More precisely, we regularize the empirical forecast measure using the same smoothing operator that defines the observed density, and then evaluate the discrepancy between the resulting coarse-grained forecast density and the data. Thus the feedback is driven by quantities that are invariant under permutations of the forecast particles and defined at the same resolution as the measurements. This avoids choosing artificial particle correspondences or lifting a coarse density observation back to a unique microscopic configuration. The central object is therefore an observation-scale, kernel-regularized misfit functional on probability measures.

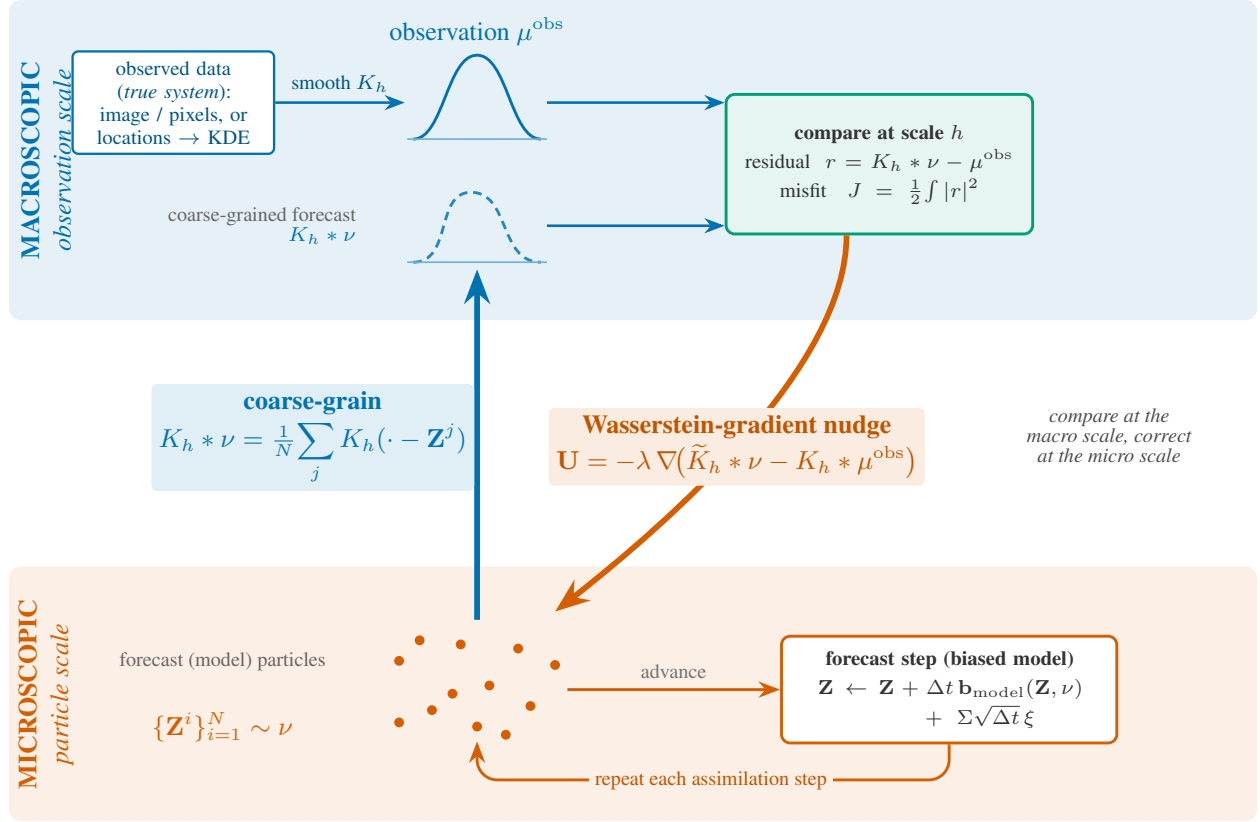


Figure 1: **The Multiscale Nudging pipeline.** Biased-model forecast particles $\{Z^i\} \sim \nu$ (bottom band, *microscopic*) are coarse-grained with the kernel K_h and compared with the observed density μ^{obs} (top band, *macroscopic*). The Wasserstein gradient of the resulting misfit then nudges every particle back at the microscopic scale, and the step repeats. The comparison lives at the observation scale h , but the correction acts on individual particles—without matching, linearization, or ensemble covariances.

Given this measure-level misfit, we seek a mechanism for translating the resulting correction to the particle level. We address this through the Wasserstein gradient-flow method [57, 58, 59, 60]. This theory provides a variational formulation for a broad class of evolution equations including the Fokker–Planck equation, porous medium equations, and aggregation-diffusion models [61, 62]. Within this method, the Wasserstein gradient of a regularized observation-misfit functional defines a transport velocity field on state space. Evaluating this velocity field at the forecast particles yields a microscopic correction that is consistent with the macroscopic density-level discrepancy. The resulting scheme compares forecast and data at the macroscopic density scale, yet implements the correction through microscopic particle dynamics without requiring particle labels or pointwise matching.

In this paper, we develop Multiscale Nudging as a measure-level method for assimilating coarse observations into microscopic mean-field dynamics. The forecast-observation mismatch is defined on probability measures after applying the same smoothing operator used in the observations, so the correction is invariant under particle relabeling and does not require lifting a coarse density to a unique particle configuration. We derive the feedback term as the Wasserstein steepest-descent direction of this misfit and evaluate the resulting velocity at particle locations. For fixed bandwidth and nudging strength, we prove well-posedness of the assimilated McKean–Vlasov dynamics and propagation of chaos for the particle system. With exact smoothed observations and a kernel-scale observability condition, we also obtain an L^2 -error estimate with exponential decay up to a model-error-dependent bias floor. The method is tested on Gaussian, multimodal, chaotic, kinetic, and collective-motion examples. Figure 1 summarizes the overall pipeline: the empirical forecast is coarse-grained and compared with the observation at the macroscopic scale, while the resulting correction is applied to individual particles at the microscopic scale through the Wasserstein gradient.

2 Method

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $(\mathcal{F}_t)_{t \geq 0}$. We consider the reference, or ground-truth, dynamics governed by the following McKean-Vlasov stochastic differential equation:

$$d\mathbf{X}_t = \mathbf{b}_{\text{true}}(\mathbf{X}_t, \mu_t)dt + \Sigma d\mathbf{W}_t, \quad \mathbf{X}_0 \sim \mu_0, \quad (1)$$

where $\mu_t = \text{Law}(\mathbf{X}_t) \in \mathcal{P}_2(\mathbb{R}^d)$ denotes the probability distribution of the true process, $\mathbf{X}_t \in \mathbb{R}^d$ is the state of the reference particle, \mathbf{W}_t is a standard d -dimensional Brownian motion, $\mathbf{b}_{\text{true}} : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is the true drift term, and $\Sigma \in \mathbb{R}^{d \times d}$ is the diffusion coefficient (assumed constant for simplicity). The evolution of μ_t is given by

$$\partial_t \mu_t = -\nabla_{\mathbf{x}} \cdot (\mathbf{b}_{\text{true}}(\mathbf{x}, \mu_t) \mu_t) + \frac{1}{2} \nabla_{\mathbf{x}} \cdot (\Sigma \Sigma^\top \nabla_{\mathbf{x}} \mu_t). \quad (2)$$

Equation (1) can be viewed as the mean-field approximation of an interacting particle system

$$d\mathbf{X}_t^i = \mathbf{b}(\mathbf{X}_t^1, \dots, \mathbf{X}_t^N)dt + \Sigma d\mathbf{W}_t^i, \quad i = 1, \dots, N,$$

with $N \rightarrow \infty$, under appropriate assumptions ensuring propagation of chaos [63, 9]. In practice, the true drift \mathbf{b}_{true} is typically unknown. A considerable body of recent work has focused on learning interaction laws from data [16, 17, 18]. The learned interaction model induces an approximate drift $\mathbf{b}_{\text{model}}$, which generally differs from the true drift with residual

$$\mathbf{R}(\mathbf{x}, \mu) := \mathbf{b}_{\text{model}}(\mathbf{x}, \mu) - \mathbf{b}_{\text{true}}(\mathbf{x}, \mu).$$

The corresponding approximate dynamics are given by

$$d\mathbf{Y}_t = \mathbf{b}_{\text{model}}(\mathbf{Y}_t, \mu_t^{\text{model}})dt + \Sigma d\mathbf{W}_t,$$

where $\mu_t^{\text{model}} = \text{Law}(\mathbf{Y}_t)$, or at the particle level,

$$d\mathbf{Y}_t^{i,N} = \mathbf{b}_{\text{model}}(\mathbf{Y}_t^{i,N}, \mu_t^{\text{model},N})dt + \Sigma d\mathbf{W}_t^{i,N}, \quad (3)$$

where $\mu^{\text{model},N} = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{Y}_t^i}$. Due to initialization error and model misspecification, the discrepancy between the true and approximate systems tends to accumulate over time.

Data assimilation is one strategy to reduce this drift by continuously incorporating observational information into the model evolution. However, in the mean-field setting, one difficulty is that the representative particle \mathbf{X}_t is not itself a physical observable, and even in the interacting particle system $\{\mathbf{X}_t^i\}_{i=1}^N$, individual trajectories are rarely accessible. In many practical applications, the available measurements take the form of aggregate or image-based data, such as density fields reconstructed from pixel observations, from which one cannot establish a correspondence between observed positions and particle indices. For example, in collective motion experiments such as [64, 65, 66], the accessible quantity is a coarse-grained spatial density reconstructed from imaging data, rather than the full microscopic state of each individual. In geophysical forecasting [37], satellite observations similarly provide smoothed integrals of the atmospheric state distribution. In all these settings, the observational quantity takes the form of a smoothed density. For simplicity, we the following two forms of smoothed observations in this paper. Let $K_h(\mathbf{x}) = h^{-d}K(\mathbf{x}/h)$ be a smoothing kernel with bandwidth $h > 0$, where $K : \mathbb{R}^d \rightarrow \mathbb{R}_+$. Concretely, we consider two cases. In the first case, one has access to pixel-level image data, from which the observed density is directly reconstructed through a coarse-graining operator

$$\mu_t^{\text{obs}}(\mathbf{z}) = \mathcal{O}_h(\mu_t)(\mathbf{z}) = \int K_h(\mathbf{z} - \mathbf{x}) \mu_t(d\mathbf{x}).$$

In the second case, only a partial and unlabeled set of particle locations $\{\mathbf{X}_t^{\text{obs},k,M}\}_{k=1}^M$ is available, without consistent index correspondence across time, and the observed density is approximated by the kernel density estimator

$$\mu_t^{\text{obs}}(\mathbf{z}) = \frac{1}{M} \sum_{k=1}^M K_h(\mathbf{z} - \mathbf{X}_t^{\text{obs},k,M}).$$

To reduce discrepancies caused by imperfect initialization and model error, we propose to augment the approximate mean-field dynamics with a nudging term, in the spirit of continuous data assimilation [43]. This term drives the evolving law ν_t toward consistency with the available observations.

Assume that the smoothing kernel is even, i.e. $K_h(\mathbf{x}) = K_h(-\mathbf{x})$. For a forecast law ν and a fixed observation μ^{obs} , define the observation-scale residual

$$r_{\nu,h} := K_h * \nu - \mu^{\text{obs}}.$$

We measure the forecast–observation mismatch by

$$J_{\text{nud},h}(\nu, \mu^{\text{obs}}) = \frac{1}{2} \int_{\mathbb{R}^d} |r_{\nu,h}(\mathbf{x})|^2 \, d\mathbf{x}.$$

Since convolution with an even kernel is self-adjoint in L^2 , differentiating $J_{\text{nud},h}$ along signed measure perturbations gives the first variation

$$\phi_{\nu,h}(\mathbf{z}) := \frac{\delta J_{\text{nud},h}}{\delta \nu}(\mathbf{z}) = (K_h * r_{\nu,h})(\mathbf{z}) = (\tilde{K}_h * \nu - K_h * \mu^{\text{obs}})(\mathbf{z}), \quad \tilde{K}_h := K_h * K_h.$$

We define the nudging correction as the Wasserstein steepest-descent direction of $J_{\text{nud},h}$, with μ^{obs} held fixed during the correction step. Equivalently, for an artificial nudging time step $\tau > 0$, the corrected law is formally given by the minimizing-movement problem

$$\nu^+ \in \operatorname{argmin}_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \frac{1}{2\tau} W_2^2(\rho, \nu) + \lambda J_{\text{nud},h}(\rho, \mu^{\text{obs}}) \right\}.$$

In infinitesimal form, the corresponding tangent-space problem is

$$\mathbf{u}_{\text{nud},h} = \operatorname{argmin}_{\mathbf{u} \in T_\nu \mathcal{P}_2} \left\{ \int_{\mathbb{R}^d} \nabla \phi_{\nu,h}(\mathbf{z}) \cdot \mathbf{u}(\mathbf{z}) \, d\nu(\mathbf{z}) + \frac{1}{2\lambda} \int_{\mathbb{R}^d} |\mathbf{u}(\mathbf{z})|^2 \, d\nu(\mathbf{z}) \right\}.$$

The Euler–Lagrange condition for this quadratic problem yields

$$\mathbf{u}_{\text{nud},h}(\mathbf{z}, \nu, \mu^{\text{obs}}) = -\lambda \nabla \phi_{\nu,h}(\mathbf{z}).$$

Therefore,

$$\mathbf{u}_{\text{nud},h}(\mathbf{z}, \nu, \mu^{\text{obs}}) = -\lambda \nabla (\tilde{K}_h * \nu - K_h * \mu^{\text{obs}})(\mathbf{z}).$$

Substituting into (3), we obtain the regularized assimilated dynamics

$$d\mathbf{Z}_t = \mathbf{b}_{\text{model}}(\mathbf{Z}_t, \nu_t) \, dt - \lambda \nabla (\tilde{K}_h * \nu - K_h * \mu^{\text{obs}})(\mathbf{Z}_t) \, dt + \Sigma d\mathbf{W}_t. \quad (4)$$

Formally, the Law of \mathbf{Z}_t , denoted by ν_t , satisfies the following nonlinear Fokker-Planck equation:

$$\partial_t \nu_t = -\nabla \cdot (\nu_t \mathbf{b}_{\text{model}}(\cdot, \nu_t)) + \nabla \cdot (A \nabla \nu_t) + \lambda \nabla \cdot [\nu_t \nabla (K_h * r_t)], \quad A = \frac{1}{2} \Sigma \Sigma^\top. \quad (5)$$

Remark 2.1. If we assume that $\mathbf{b}_{\text{model}}$ and Σ take the following form

$$\mathbf{b}_{\text{model}}(\mathbf{z}, \nu) = -\nabla \left(\frac{\delta F_{\text{model}}}{\delta \nu} \right)(\mathbf{z}), \quad \Sigma = \sqrt{2\beta} \mathbf{I},$$

for some functional $F_{\text{model}}(\mu)$, then the assimilated dynamics (4) can be interpreted as a Wasserstein gradient flow of the free energy

$$\mathcal{F}_t(\nu) = F_{\text{model}}(\nu) + \lambda J_{\text{nud},h}(\nu, \mu_t^{\text{obs}}) + \beta \operatorname{Ent}(\nu), \quad (6)$$

where $\operatorname{Ent}(\nu) = \int \nu \log(\nu) \, d\mathbf{z}$.

Algorithm 1 Multiscale Nudging: particle implementation from coarse density observations

Input: Particles $\{\mathbf{Z}_0^i\}_{i=1}^N$, model drift $\mathbf{b}_{\text{model}}$, diffusion Σ , step Δt .
 Grid $\mathcal{G} = \{\mathbf{x}_q\}_{q=1}^G$, weights $\{w_q\}$, kernel K_h , obs. times \mathcal{T}_{obs} .
 Nudging strength λ , substeps L_{nud} , substep size $\Delta\tau = \Delta t/L_{\text{nud}}$.
Output: Assimilated trajectories $\{\mathbf{Z}_n^i\}_{i=1}^N$.

1: **for** $n = 0, \dots, N_t - 1$ **do**
 2: Empirical forecast law $\nu_n^N = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{Z}_n^i}$.
 3: **Forecast.** Push each particle with $\xi_n^i \sim \mathcal{N}(0, I_d)$:
 $\tilde{\mathbf{Z}}_{n+1}^i = \mathbf{Z}_n^i + \Delta t \mathbf{b}_{\text{model}}(\mathbf{Z}_n^i, \nu_n^N) + \Sigma \sqrt{\Delta t} \xi_n^i$.
 4: **Observe.** Receive coarse density $y_{n+1,q} \approx \mu_{t_{n+1}}^{\text{obs}}(\mathbf{x}_q)$ on \mathcal{G} .
 5: Initialize $\mathbf{Z}_{n+1}^{i,(0)} \leftarrow \tilde{\mathbf{Z}}_{n+1}^i$.
 6: **for** $\ell = 0, \dots, L_{\text{nud}} - 1$ **do**
 7: Smoothed density $\rho_q^{(\ell)} = \frac{1}{N} \sum_{j=1}^N K_h(\mathbf{x}_q - \mathbf{Z}_{n+1}^{j,(\ell)})$ for all q .
 8: Residual $r_q^{(\ell)} = \rho_q^{(\ell)} - y_{n+1,q}$.
 9: **Nudge.** Wasserstein-gradient velocity at each particle:

$$\mathbf{U}_i^{(\ell)} = -\lambda \sum_{q=1}^G w_q \nabla_{\mathbf{z}} K_h(\mathbf{z} - \mathbf{x}_q) \Big|_{\mathbf{z}=\mathbf{Z}_{n+1}^{i,(\ell)}} r_q^{(\ell)}.$$

 10: Substep $\mathbf{Z}_{n+1}^{i,(\ell+1)} \leftarrow \mathbf{Z}_{n+1}^{i,(\ell)} + \Delta\tau \mathbf{U}_i^{(\ell)}$.
 11: **end for**
 12: Commit $\mathbf{Z}_{n+1}^i \leftarrow \mathbf{Z}_{n+1}^{i,(L_{\text{nud}})}$.
 13: **end for**

3 Theoretical Analysis

This section establishes three properties of the Multiscale Nudging scheme: the kernel-regularized feedback yields well-posed dynamics, the finite-particle implementation converges to the mean-field model, and the feedback provably reduces the forecast error. The analysis begins with Lemma 3.1, which shows the two key properties of the smoothed kernel \tilde{K}_h : a globally Lipschitz gradient, supplying the regularity for well-posedness, and the H^1 approximation $\tilde{K}_h * v \rightarrow v$ as $h \downarrow 0$, quantifying the accuracy of observing only at scale h . Proposition 3.2 then gives a unique strong solution to the assimilated McKean-Vlasov dynamics (4), and Proposition 3.3 establishes propagation of chaos, justifying the particle discretization of Algorithm 1. With the strict positivity from Proposition 3.4, our main result, Theorem 3.5, shows that under a kernel-scale observability condition and a lower bound on the nudging strength, the L^2 error decays exponentially to a floor of order Δ^2 , vanishing when the model is exact.

Lemma 3.1 (Periodized Gaussian kernel). *Let $\Omega = \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. For $h > 0$, define the periodized Gaussian kernel*

$$K_h^{\mathbb{T}}(\mathbf{x}) = \sum_{\mathbf{m} \in \mathbb{Z}^d} h^{-d} \pi^{-d/2} \exp\left(-\frac{\|\mathbf{x} + \mathbf{m}\|^2}{h^2}\right), \quad \mathbf{x} \in \mathbb{T}^d.$$

Let

$$\tilde{K}_h = K_h^{\mathbb{T}} *_{\mathbb{T}} K_h^{\mathbb{T}},$$

where $*_{\mathbb{T}}$ denotes convolution on the torus. Then:

1. $K_h^{\mathbb{T}}$ and \tilde{K}_h belong to $C^\infty(\mathbb{T}^d)$. In particular, for each fixed $h > 0$, \tilde{K}_h is globally Lipschitz on \mathbb{T}^d .

2. For each fixed $h > 0$,

$$\nabla \tilde{K}_h \in W^{1,\infty}(\mathbb{T}^d).$$

In particular, $\nabla \tilde{K}_h \in L^\infty(\mathbb{T}^d)$, and $\nabla \tilde{K}_h$ is globally Lipschitz.

3. For every $v \in H^1(\mathbb{T}^d)$,

$$\|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2(\mathbb{T}^d)} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

Moreover, if $v \in H^2(\mathbb{T}^d)$, then there exists a constant $C > 0$, independent of h and v , such that for $0 < h \leq 1$,

$$\|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2(\mathbb{T}^d)} \leq Ch \|D^2 v\|_{L^2(\mathbb{T}^d)}.$$

The proof is standard and is given in Appendix A for completeness. Although Lemma 3.1 is stated on the periodic domain \mathbb{T}^d , the periodicity assumption is not essential for these kernel estimates. In the non-periodic whole-space setting, the same conclusions hold on \mathbb{R}^d for any even, normalized $C_c^\infty(\mathbb{R}^d)$ mollifier K , with the usual scaling $K_h(\mathbf{x}) = h^{-d}K(\mathbf{x}/h)$ and $\tilde{K}_h = K_h * K_h$. The proof follows from the standard approximate-identity and translation estimate arguments, and we omit the routine variant for brevity. Based on the properties of the kernel, we have the following well-posedness result.

Assumption 1. Assume that the $\mathbf{b}_{\text{model}}$ and \mathbf{b}_{true} satisfies: (A1) There exists a constant $B_{\text{model}} > 0$ such that for all $\rho \in \mathcal{P}_2(\Omega)$,

$$\|\mathbf{b}_{\text{model}}(\cdot, \rho)\|_{L^\infty(\Omega)} \leq B_{\text{model}}, \quad \|\nabla \cdot \mathbf{b}_{\text{model}}(\cdot, \rho)\|_{L^\infty(\Omega)} \leq B_{\text{model}}.$$

(A2) Assume that \mathbf{b}_{true} and $\mathbf{b}_{\text{model}}$ is globally Lipschitz, i.e. there exists a constant $L_{\text{true}}, L_{\text{model}} > 0$ such that for all $x, y \in \Omega$ and all $\rho, \rho' \in \mathcal{P}_2(\Omega)$,

$$\|\mathbf{b}_{\text{true}}(\mathbf{x}_1, \rho_1) - \mathbf{b}_{\text{true}}(\mathbf{x}_2, \rho_2)\| \leq L_{\text{true}}(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\rho_1, \rho_2)),$$

and

$$\|\mathbf{b}_{\text{model}}(\mathbf{x}_1, \rho_1) - \mathbf{b}_{\text{model}}(\mathbf{x}_2, \rho_2)\| \leq L_{\text{model}}(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\rho_1, \rho_2)).$$

We now turn to the well-posedness of the regularized assimilated dynamics. Although the nudging term depends nonlocally on the law ν_t , convolution against \tilde{K}_h renders it Lipschitz in both the state and the measure, so the standard McKean-Vlasov well-posedness theory applies.

Proposition 3.2 (Well-Posedness of Equation (4)). Under (A2), and also assume that $\nabla \tilde{K}_h$ is globally Lipschitz, i.e. there exists $L > 0$ such that for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$,

$$\|\nabla \tilde{K}_h(\mathbf{x}_1) - \nabla \tilde{K}_h(\mathbf{x}_2)\| \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|.$$

For any $T > 0$ and $\text{Law}(\mathbf{Z}_0) \in \mathcal{P}_2(\mathbb{R}^d)$, the SDE (4) has a unique strong solution on $[0, T]$ and consequently, its law is the unique solution to the Fokker-Planck equation (5).

The proof is given in Appendix B. Next, we approximate the assimilated mean-field process by finite particles as

$$\begin{aligned} d\mathbf{Z}_t^{i,N} = & \mathbf{b}_{\text{model}}(\mathbf{Z}_t^{i,N}, \nu_t^N) dt \\ & - \lambda \nabla \left[\left(\tilde{K}_h * \nu_t^N \right) (\mathbf{Z}_t^{i,N}) - (K_h * \mu_t^{\text{obs}}) (\mathbf{Z}_t^{i,N}) \right] dt + \Sigma d\mathbf{W}_t^i, \end{aligned}$$

where $\nu_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{\mathbf{Z}_t^{j,N}}$ is the empirical measure of the system. Expanding the convolution terms, the explicit interaction dynamics are given by:

$$\begin{aligned} d\mathbf{Z}_t^{i,N} = & \mathbf{b}_{\text{model}}(\mathbf{Z}_t^{i,N}, \nu_t^N) dt \\ & - \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_t^{i,N} - \mathbf{Z}_t^{j,N}) - \nabla K_h * \mu_t^{\text{obs}}(\mathbf{Z}_t^{i,N}) \right) dt + \Sigma d\mathbf{W}_t^i. \end{aligned} \quad (7)$$

If we only observe the unlabeled locations $\{\mathbf{X}_t^{\text{obs},k,M}\}_{k=1}^M$, then

$$\begin{aligned} d\mathbf{Z}_t^{i,N} = & \mathbf{b}_{\text{model}}(\mathbf{Z}_t^{i,N}, \nu_t^N) dt \\ & - \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_t^{i,N} - \mathbf{Z}_t^{j,N}) - \frac{1}{M} \sum_{k=1}^M \nabla \tilde{K}_h(\mathbf{Z}_t^{i,N} - \mathbf{X}_t^{\text{obs},k,M}) \right) dt \\ & + \Sigma d\mathbf{W}_t^i. \end{aligned} \quad (8)$$

Proposition 3.3 (Mean-Field Convergence and Propagation of Chaos for the Nudging Particle System). Let the assumptions of Proposition 3.2 hold. Also, assume that $\nabla \tilde{K}_h$ is bounded, i.e. $\|\nabla \tilde{K}_h\|_\infty < \infty$. Let $(\mathbf{Z}_t^{i,N})_{i=1}^N$ be the N -particle system solving (7) with f_0 -chaotic initial data $\mathbf{Z}_0^{1,N} \sim f_0$. Let f_t be the unique solution to the mean field Fokker-Planck equation (5) with initial condition f_0 . Then, the N -particle system (7) converges to mean field model (4) as $N \rightarrow \infty$. That is, for any $T > 0$, the N -particle distribution $f_t^N = \text{Law}(\mathbf{Z}_t^{1,N}, \dots, \mathbf{Z}_t^{N,N})$ is f_t -chaotic, satisfying:

$$\lim_{N \rightarrow \infty} W_2(f_{[0,T]}^{1,N}, f_{[0,T]}) = 0,$$

where $f_t^{1,N}$ is the first marginal of f_t^N .

The proof appears in Appendix C.

We next show that the solution of equation (4) remains strictly positive on $[0, T]$ for any fixed final time $T > 0$, a property that will be needed in the proof of convergence.

Proposition 3.4 (Positivity on finite time intervals). *Let $T > 0$ be fixed and let $\Omega = \mathbb{T}^d$. Assume that Assumption 1 holds, and let K_h be the Gaussian kernel introduced in Lemma 3.1. Let ν be a classical solution on $[0, T]$ of*

$$\partial_t \nu = \nabla \cdot (A \nabla \nu) - \nabla \cdot (\nu c[\nu]),$$

where

$$c[\nu](x, t) := \mathbf{b}_{\text{model}}(x, \nu_t) - \lambda \nabla \cdot \left(K_h * (K_h * \nu_t - \mu_t^{\text{obs}}) \right)(x).$$

Assume moreover that

$$\sup_{t \in [0, T]} \left(\|K_h * \nu_t\|_{L^\infty(\Omega)} + \|\mu_t^{\text{obs}}\|_{L^\infty(\Omega)} \right) < \infty,$$

and that the initial density satisfies

$$\nu_0(x) \geq \underline{\nu}_0 > 0, \quad \forall x \in \Omega.$$

Then there exists a constant $\underline{\nu}_T > 0$ such that

$$\nu_t(x) \geq \underline{\nu}_T, \quad \forall (x, t) \in \Omega \times [0, T].$$

More precisely,

$$\underline{\nu}_T = \underline{\nu}_0 e^{-CT},$$

where

$$C := B_{\text{model}} + \lambda \|\Delta K_h\|_{L^1(\Omega)} \sup_{t \in [0, T]} \|K_h * \nu_t - \mu_t^{\text{obs}}\|_{L^\infty(\Omega)}.$$

The proof is provided in Appendix D.

The following result should be interpreted as an idealized contraction estimate under exact smoothed observations. Finite-sample and noisy observations add additional residual terms to the bias floor.

Theorem 3.5 (L^2 -error decay under kernel-regularized nudging). *Let $\Omega = \mathbb{T}^d$ and let $A = \frac{1}{2} \Sigma \Sigma^\top$ be a constant, symmetric positive-definite matrix with smallest eigenvalue $\kappa > 0$. Let $(\mu_t)_{t \geq 0}$ and $(\nu_t)_{t \geq 0}$ be classical solutions of the true Fokker–Planck equation (2) and the assimilated equation (5), respectively, with initial conditions satisfying*

$$\int_{\Omega} (\nu_0 - \mu_0) dx = 0. \text{ Define the error } e_t := \nu_t - \mu_t \text{ and the } L^2\text{-error energy}$$

$$V(t) := \frac{1}{2} \|e_t\|_{L^2(\Omega)}^2.$$

Suppose the following conditions hold.

(C1) *The model drift satisfies Assumption 1, and, in addition,*

$$\|\mathbf{b}_{\text{model}}(\cdot, \mu) - \mathbf{b}_{\text{model}}(\cdot, \nu)\|_{L^2(\Omega)} \leq L_\mu \|\mu - \nu\|_{L^2(\Omega)}$$

for all $\mu, \nu \in \mathcal{P}_2(\Omega)$.

(C2) *The true density and the assimilated density are uniformly bounded above,*

$$\|\mu_t\|_{L^\infty}, \|\nu_t\|_{L^\infty} \leq \bar{\rho}$$

, and the assimilated density is uniformly bounded below, $\nu_t(x) \geq \underline{\nu} > 0$, for all $(x, t) \in \Omega \times [0, T]$. (The lower bound is guaranteed by Proposition 3.4 in some special cases.)

(C3) *There exists a function $\delta(h) \rightarrow 0$ as $h \rightarrow 0$ such that*

$$\|\nabla e_t - \nabla(\tilde{K}_h * e_t)\|_{L^2(\Omega)} \leq \delta(h) \|\nabla e_t\|_{L^2(\Omega)}.$$

Define the advection instability coefficient

$$C_{\text{adv}} := B_{\text{model}} + \bar{\rho} L_\mu,$$

the model-error bound

$$\Delta := \sup_{t \in [0, T]} \|\mathbf{b}_{\text{model}}(\cdot, \mu_t) - \mathbf{b}_{\text{true}}(\cdot, \mu_t)\|_{L^\infty(\Omega)},$$

and the Poincaré constant C_P of Ω for zero-mean functions. If the observation resolution h and the nudging intensity λ satisfy

$$\delta(h) \leq \frac{\underline{\nu}}{2\bar{\rho}}, \quad (9)$$

$$\lambda > \frac{1}{\underline{\nu}} \left(\frac{2C_{\text{adv}}^2 C_P}{\kappa} - \kappa \right), \quad (10)$$

then the L^2 error decays exponentially to a model-error floor. Explicitly, with the rate

$$\alpha := \frac{\kappa + \lambda \underline{\nu}}{C_P} - \frac{2C_{\text{adv}}^2}{\kappa} > 0,$$

it holds for all $t \geq 0$ that

$$V(t) \leq e^{-\alpha t} V(0) + \frac{\bar{\rho}^2}{\alpha \kappa} \Delta^2. \quad (11)$$

In particular,

$$\limsup_{t \rightarrow \infty} \|\nu_t - \mu_t\|_{L^2(\Omega)}^2 \leq \frac{2\bar{\rho}^2}{\alpha \kappa} \Delta^2,$$

so the long-time error is controlled by a bias floor proportional to Δ^2 that vanishes when the model is exact ($\Delta = 0$).

Proof. Subtracting the true equation (2) from the assimilated equation (5), the error $e_t = \nu_t - \mu_t$ satisfies

$$\partial_t e_t = \nabla \cdot (A \nabla e_t) - \nabla \cdot (\nu_t \mathbf{b}_{\text{model}}(\cdot, \nu_t)) + \nabla \cdot (\mu_t \mathbf{b}_{\text{true}}(\cdot, \mu_t)) + \lambda \nabla \cdot (\nu_t \nabla (K_h * r_t)), \quad (12)$$

where $r_t = K_h * \nu_t - \mu_t^{\text{obs}} = K_h * e_t$. Since (12) is in divergence form, the zero-mean condition is preserved: $\int_{\Omega} e_t dx = 0$ for all $t \geq 0$. Hence the Poincaré inequality applies:

$$\|e_t\|_{L^2(\Omega)}^2 \leq C_P \|\nabla e_t\|_{L^2(\Omega)}^2. \quad (13)$$

Testing (12) against e_t and integrating by parts yields

$$\begin{aligned} V'(t) = & \underbrace{-\langle \nabla e_t, A \nabla e_t \rangle}_{I_{\text{diff}}} + \underbrace{\langle \nabla e_t, \nu_t \mathbf{b}_{\text{model}}(\cdot, \nu_t) - \mu_t \mathbf{b}_{\text{true}}(\cdot, \mu_t) \rangle}_{I_{\text{adv}}} \\ & - \underbrace{\lambda \langle \nabla e_t, \nu_t \nabla (K_h * r_t) \rangle}_{I_{\text{nud}}}. \end{aligned} \quad (14)$$

For the diffusion term, by the coercivity of A ,

$$I_{\text{diff}} = -\langle \nabla e_t, A \nabla e_t \rangle \leq -\kappa \|\nabla e_t\|^2.$$

Next, for the advection term, we decompose the flux difference as

$$\begin{aligned} \nu_t \mathbf{b}_{\text{model}}(\cdot, \nu_t) - \mu_t \mathbf{b}_{\text{true}}(\cdot, \mu_t) = & e_t \mathbf{b}_{\text{model}}(\cdot, \nu_t) + \mu_t (\mathbf{b}_{\text{model}}(\cdot, \nu_t) - \mathbf{b}_{\text{model}}(\cdot, \mu_t)) \\ & + \mu_t (\mathbf{b}_{\text{model}}(\cdot, \mu_t) - \mathbf{b}_{\text{true}}(\cdot, \mu_t)). \end{aligned}$$

Using (C1), (C2), and Cauchy–Schwarz, each inner product with ∇e_t is bounded by

$$I_{\text{adv}} \leq B_{\text{model}} \|\nabla e_t\| \|e_t\| + \bar{\rho} L_{\mu} \|\nabla e_t\| \|e_t\| + \bar{\rho} \|\nabla e_t\| \Delta = C_{\text{adv}} \|\nabla e_t\| \|e_t\| + \bar{\rho} \|\nabla e_t\| \Delta.$$

Applying Young's inequality $ab \leq \frac{\epsilon}{2} a^2 + \frac{1}{2\epsilon} b^2$ to each product with a parameter $\epsilon > 0$ gives

$$I_{\text{adv}} \leq \epsilon \|\nabla e_t\|^2 + \frac{C_{\text{adv}}^2}{2\epsilon} \|e_t\|^2 + \frac{\bar{\rho}^2}{2\epsilon} \Delta^2. \quad (15)$$

Lastly, we bound the nudging term. Since $K_h * r_t = \tilde{K}_h * e_t$, we add and subtract ∇e_t inside the inner product:

$$\begin{aligned} I_{\text{nud}} = & -\lambda \langle \nu_t \nabla e_t, \nabla (\tilde{K}_h * e_t) \rangle \\ = & -\lambda \langle \nu_t \nabla e_t, \nabla e_t \rangle - \lambda \langle \nu_t \nabla e_t, \nabla (\tilde{K}_h * e_t) - \nabla e_t \rangle. \end{aligned}$$

The first term is bounded using (C2): $-\lambda \langle \nu_t \nabla e_t, \nabla e_t \rangle \leq -\lambda \underline{\nu} \|\nabla e_t\|^2$. For the second, Cauchy–Schwarz and the kernel-approximation condition (C3) yield

$$\left| \lambda \langle \nu_t \nabla e_t, \nabla (\tilde{K}_h * e_t) - \nabla e_t \rangle \right| \leq \lambda \bar{\rho} \|\nabla e_t\| \delta(h) \|\nabla e_t\| = \lambda \bar{\rho} \delta(h) \|\nabla e_t\|^2.$$

Combining,

$$I_{\text{nud}} \leq -\lambda(\underline{\nu} - \bar{\rho}\delta(h))\|\nabla e_t\|^2. \quad (16)$$

Substituting (15) and (16) into (14) gives

$$V'(t) \leq (-\kappa - \lambda(\underline{\nu} - \bar{\rho}\delta(h)) + \epsilon)\|\nabla e_t\|^2 + \frac{C_{\text{adv}}^2}{2\epsilon}\|e_t\|^2 + \frac{\bar{\rho}^2}{2\epsilon}\Delta^2.$$

We now make two parameter choices. First, condition (9) gives $\underline{\nu} - \bar{\rho}\delta(h) \geq \underline{\nu}/2$, so

$$-\kappa - \lambda(\underline{\nu} - \bar{\rho}\delta(h)) + \epsilon \leq -\kappa - \frac{\lambda\underline{\nu}}{2} + \epsilon.$$

Second, choosing $\epsilon = \kappa/2$ to retain half of the diffusive dissipation yields

$$V'(t) \leq -\left(\frac{\kappa}{2} + \frac{\lambda\underline{\nu}}{2}\right)\|\nabla e_t\|^2 + \frac{C_{\text{adv}}^2}{\kappa}\|e_t\|^2 + \frac{\bar{\rho}^2}{\kappa}\Delta^2.$$

Applying the Poincaré inequality (13) to the gradient term,

$$V'(t) \leq -\underbrace{\left[\frac{1}{C_P}\left(\frac{\kappa}{2} + \frac{\lambda\underline{\nu}}{2}\right) - \frac{C_{\text{adv}}^2}{\kappa}\right]}_{=: \frac{1}{2}\alpha}\|e_t\|^2 + \frac{\bar{\rho}^2}{\kappa}\Delta^2.$$

Condition (10) ensures $\alpha > 0$. Since $\|e_t\|^2 = 2V(t)$, we obtain

$$V'(t) \leq -\alpha V(t) + \frac{\bar{\rho}^2}{\kappa}\Delta^2.$$

The bound (11) follows by Gronwall's inequality. \square

Condition (9) requires the observation bandwidth h to be small enough so that the smoothing error does not overwhelm the density lower bound. Condition (10) requires the nudging intensity λ to be large enough to overcome the advection instability caused by model error and nonlinear transport. Together, they guarantee exponential convergence of the full L^2 -error $\|\nu_t - \mu_t\|_{L^2}$ to a neighborhood of zero whose radius is controlled by the model bias Δ .

4 Numerical experiments

We evaluate the proposed multiscale nudging method on five examples of increasing complexity: a linear Gaussian benchmark, a bistable bimodal system, a mean-field Lorenz model, Vlasov–Poisson, and a real collective-motion dataset. In each case, we compare three evolutions: the reference dynamics, a biased forecast model without assimilation, and the assimilated forecast obtained by adding the nudging term.

4.1 Simple linear system

As a first test, we consider the one-dimensional linear interacting particle system

$$dX_t^i = -a(X_t^i - m_t) dt + dW_t^i, \quad (17)$$

where

$$m_t = \frac{1}{N} \sum_{i=1}^N X_t^i$$

is the empirical mean. As $N \rightarrow \infty$, the associated mean-field limit is given by

$$d\bar{X}_t = -a(\bar{X}_t - \bar{m}_t) dt + dW_t, \quad \bar{m}_t = \int x \mu_t(dx),$$

where $\mu_t = \text{Law}(\bar{X}_t)$. The corresponding Fokker–Planck equation reads

$$\partial_t \mu_t = \partial_x \left(a \left(x - \int x \mu_t(dx) \right) \mu_t \right) + \frac{1}{2} \partial_{xx} \mu_t.$$

This example is useful as a first test since the true law remains close to Gaussian and its variance provides an example of the distributional error. We generate the reference data using the true interaction coefficient $a = 1$ with initial condition

$X_0^i \sim \mathcal{N}(0, 0.5)$. The particle system is integrated by the Euler–Maruyama scheme with time step 0.01 up to $T = 5$. Observations are produced from the smoothed density $K_h * \mu_t$ with bandwidth $h = 0.5$. To test robustness with respect to model bias, the forecast model uses the same dynamics but with $a \in \{0.5, 2, 5\}$.

Figure 2 reports the variance of the trajectories for the under-interacting case $a = 0.5$. The four panels correspond to different numbers of nudging substeps. Without assimilation, the forecast systematically underestimates the growth of the variance and drifts away from the reference solution. When $\lambda = 1$, the feedback is too weak to noticeably change the forecast. Increasing the nudging strength to 10 and 100 reduces the variance gap. The strongest correction, $\lambda = 1000$, gives the closest match once enough nudging substeps are applied, but with only one nudging update, it produces a sharp overshoot, illustrating the expected stability-accuracy trade-off. The same qualitative behavior persists for the over-interacting cases $a = 2$ and $a = 5$; see Appendix Figures 14 and 15.

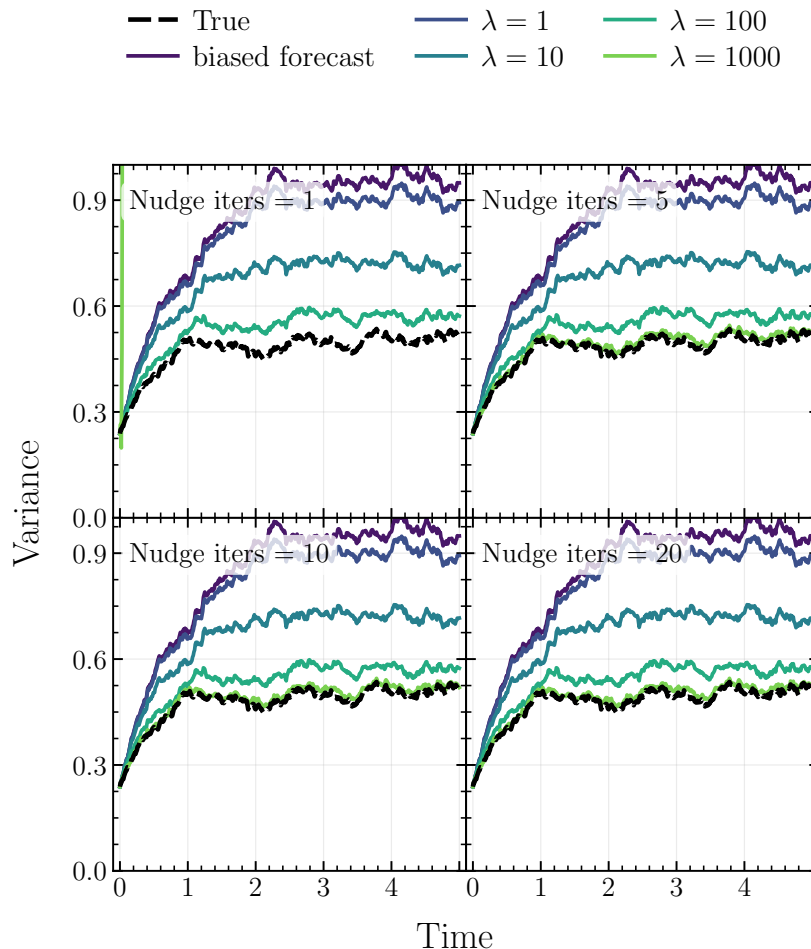


Figure 2: **Variance dynamics in the one-dimensional linear benchmark** ($a = 0.5$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. The four panels correspond to different numbers of nudging substeps. Increasing λ improves tracking accuracy, while excessively large nudging can introduce temporary numerical instability when the correction is applied too aggressively.

To move beyond a single moment diagnostic, Figure 3 compares the full space-time density for the true dynamics, the biased forecast, and two assimilated solutions. The biased forecast is overly diffuse and fails to reproduce the concentration near $x = 0$. A moderate correction ($\lambda = 10$) partially restores the correct density profile, while the stronger correction ($\lambda = 1000$) yields a space-time distribution much closer to the reference one. This confirms that the nudging term corrects the full law.

Figure 4 summarizes the final-time and time-averaged W_2 errors for all three biased models as functions of λ . Both error metrics decrease nearly monotonically as the nudging strength increases, with the most visible gains occurring between 10 and 100. The reduction is consistent across both under-interacting and over-interacting forecast models,

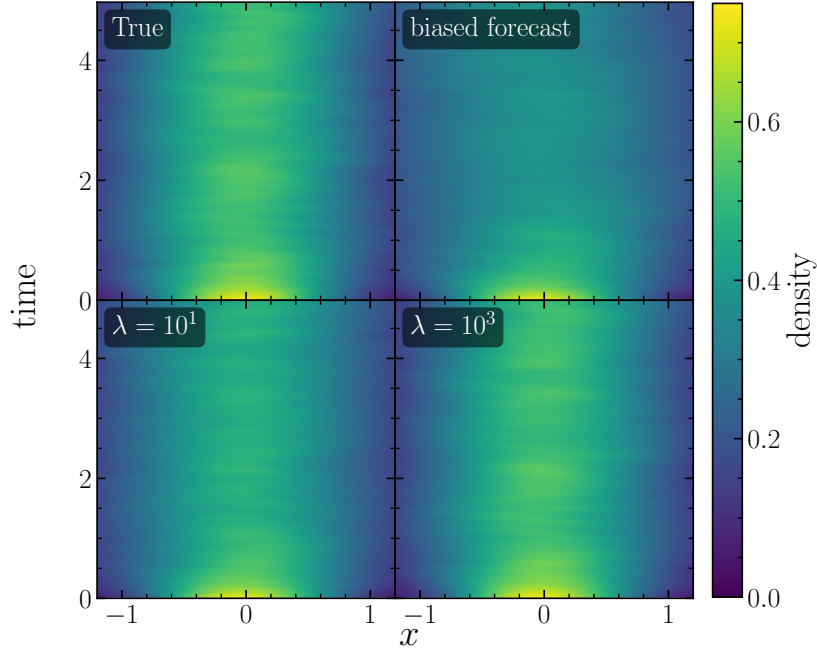


Figure 3: **Space–time density evolution in the linear benchmark** ($a = 0.5$). Top-left: reference density. Top-right: biased forecast. Bottom-left: assimilated density with $\lambda = 10$. Bottom-right: assimilated density with $\lambda = 1000$. Larger nudging strength restores both the location and the spread of the true law.

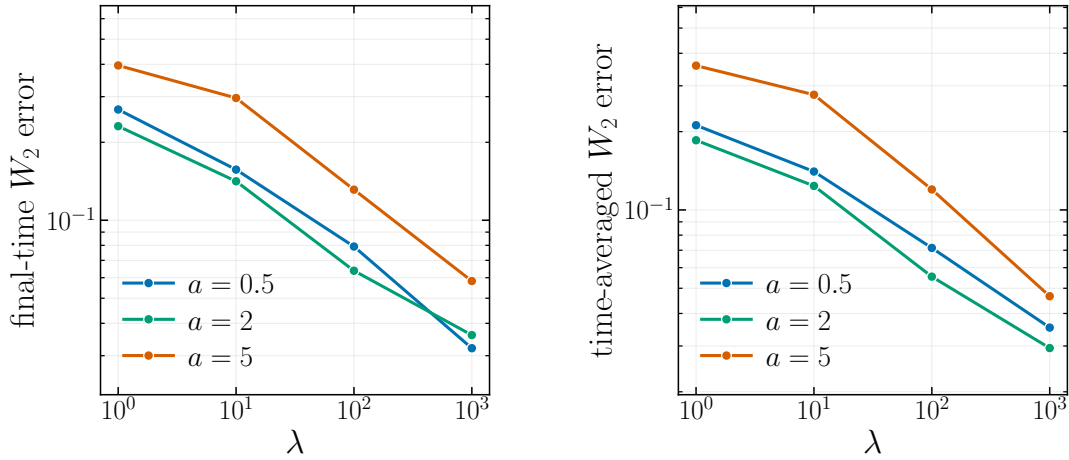


Figure 4: **Distributional error versus nudging strength in the linear benchmark**. Left: final-time W_2 error. Right: time-averaged W_2 error. In all three biased models $a \in \{0.5, 2, 5\}$, stronger nudging produces a smaller distributional discrepancy with the reference dynamics.

which supports the theoretical conclusion of Theorem 3.5: sufficiently strong observation feedback suppresses the error induced by model misspecification.

4.2 Multimodal distribution

We next consider a one-dimensional interacting particle system with a bistable confining potential,

$$dX_t^i = -\nabla U(X_t^i) dt - a(X_t^i - m_t) dt + dW_t^i, \tag{18}$$

where

$$U(x) = \frac{x^4}{4} - \frac{x^2}{2},$$

and where the empirical mean is

$$m_t = \frac{1}{N} \sum_{i=1}^N X_t^i.$$

As $N \rightarrow \infty$, the associated mean-field limit is given by

$$d\bar{X}_t = -(\bar{X}_t^3 - \bar{X}_t) dt - a(\bar{X}_t - \bar{m}_t) dt + dW_t, \quad \bar{m}_t = \int x \mu_t(dx),$$

where $\mu_t = \text{Law}(\bar{X}_t)$. The corresponding Fokker–Planck equation reads

$$\partial_t \mu_t = \partial_x ((x^3 - x)\mu_t) + \partial_x \left(a \left(x - \int x \mu_t(dx) \right) \mu_t \right) + \frac{1}{2} \partial_{xx} \mu_t.$$

This benchmark is more challenging than the linear Gaussian case since the double-well potential generates a bimodal law with metastable transitions between the two wells. In this regime, matching only the mean or the variance is insufficient; two distributions can have similar low-order moments while placing mass in the wrong well. We therefore use this example to test whether the kernel-based nudging term can transfer probability mass across the barrier at $x = 0$, recover the correct modal locations, and preserve the relative weights of the two modes from smoothed macroscopic observations. In the experiments, we generate the reference data with the true interaction coefficient $a = 0.25$ and initial position $X_0^i \sim \mathcal{N}(0, 0.5)$. We integrate the particle system using Euler–Maruyama with time step $\Delta t = 0.01$ up to final time $T = 5$. As before, observations are generated from the smoothed density $K_h * \mu_t$ with $h = 0.5$, and the forecast model is misspecified by taking the interaction coefficient $a \in \{0.1, 0.5, 1.5\}$. The quality of assimilation is evaluated through density snapshots together with the corresponding Wasserstein error. This example highlights that the proposed correction mechanism is not restricted to unimodal or near-Gaussian laws.

Figure 5 shows the variance dynamics for $a = 1.5$ for four different nudging substeps. The biased forecast model performs worse than all assimilated forecast models. After enough nudging substeps, the assimilated model with nudging intensity $\lambda = 10^3$ produces the closest approximation to the reference solution, but with just one nudging iteration, the model overestimates the variance and generates large oscillations, as expected due to the trade-off between accuracy and stability.

In Figure 6, the space-time densities for the true reference solution, biased forecast, and two assimilated forecasts with $a = 1.5$ are shown. The prediction from the biased model is overly concentrated around $x = 0$ and does not capture the density profile of the true solution, which has two high probability regions. The assimilated model with $\lambda = 10$ moderately diffuses the concentration near $x = 0$ and partially resolves the two modes, whereas the model with $\lambda = 10^3$ achieves the closest match to the reference density and clearly shows the two stable regions of high probability.

Figure 7 displays the final-time and time-averaged W_2 errors for the forecast models with $a \in \{0.1, 0.5, 1.5\}$ as functions of the nudging intensity λ . As λ increases, both errors decrease monotonically across all three biased models, which is consistent with the conclusion of Theorem 3.5. Additional experiments with $a = 0.1$ and $a = 0.5$ show similar qualitative behavior; see Appendix Figures 16, 17, 18, and 19.

4.3 Lorenz dynamics

We further consider a mean-field Lorenz system

$$\begin{aligned} dX_t^i &= \sigma_L(m_{t,y} - X_t^i) dt + \sigma dW_t^{x,1}, \\ dY_t^i &= [m_{t,x}(\rho - m_{t,z}) - Y_t^i] dt + \sigma dW_t^{y,1}, \\ dZ_t^i &= [m_{t,x}m_{t,y} - \beta Z_t^i] dt + \sigma dW_t^{z,1}, \end{aligned}$$

where $m_{t,x}$, $m_{t,y}$, and $m_{t,z}$ denote the empirical means of the three coordinates. As a biased forecast model, we use N individual Lorenz systems without the mean-field coupling,

$$\begin{aligned} dX_t^i &= \sigma_L(Y_t^i - X_t^i) dt + \sigma dW_t^{x,1}, \\ dY_t^i &= [X_t^i(\rho - Z_t^i) - Y_t^i] dt + \sigma dW_t^{y,1}, \\ dZ_t^i &= [X_t^i Y_t^i - \beta Z_t^i] dt + \sigma dW_t^{z,1}. \end{aligned}$$

We use the classical Lorenz parameters

$$\sigma_L = 10, \quad \rho = 28, \quad \beta = 8/3,$$

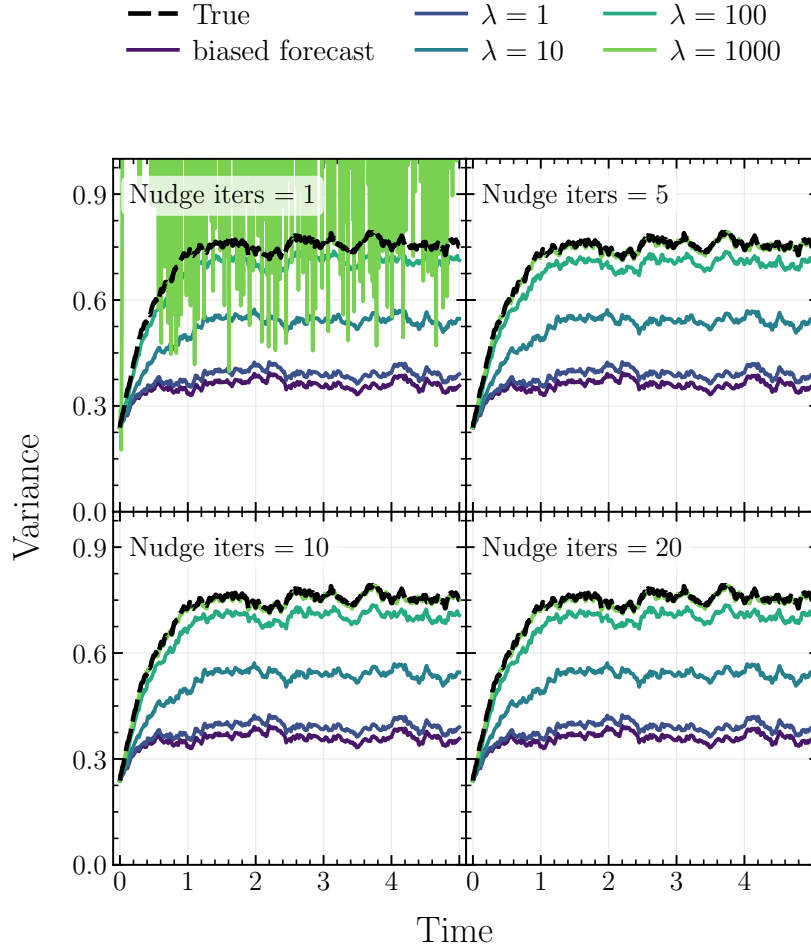


Figure 5: **Variance dynamics in the multimodal benchmark** ($a = 1.5$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. The four panels correspond to different numbers of nudging substeps. The same stability-accuracy trade-off observed in Figure 2 persists in this regime.

with noise amplitude $\sigma_{\text{noise}} = 1$, time step $\Delta t = 0.01$, particle number $N = 1000$, and kernel bandwidth $h = 0.5$. This example probes a strongly nonlinear and chaotic regime in which small modeling errors amplify rapidly. Figure 8 compares the mean trajectory $(m_{t,x}, m_{t,y}, m_{t,z})$ of the assimilated system with the reference attractor for $\lambda \in \{1000, 100, 10\}$. When $\lambda = 1000$, the corrected trajectory remains close to the true attractor and the mean error stays uniformly small. For $\lambda = 100$, the dynamics still capture the correct global structure, but a visible phase discrepancy remains. When $\lambda = 10$, the correction is too weak and the bottom panel shows repeated spikes in the mean-trajectory error, corresponding to intermittent departures from the correct lobe of the attractor. These results show that density-level feedback can stabilize a biased microscopic forecast even in a chaotic setting.

4.4 Vlasov–Poisson equation

We next consider the one-dimensional Vlasov–Poisson system

$$\begin{aligned} dX_t^i &= V_t^i dt, \\ dV_t^i &= E(X_t^i, t) dt, \end{aligned}$$

where the self-consistent electric field is given by

$$E(x, t) = -\nabla_x \phi(x, t),$$

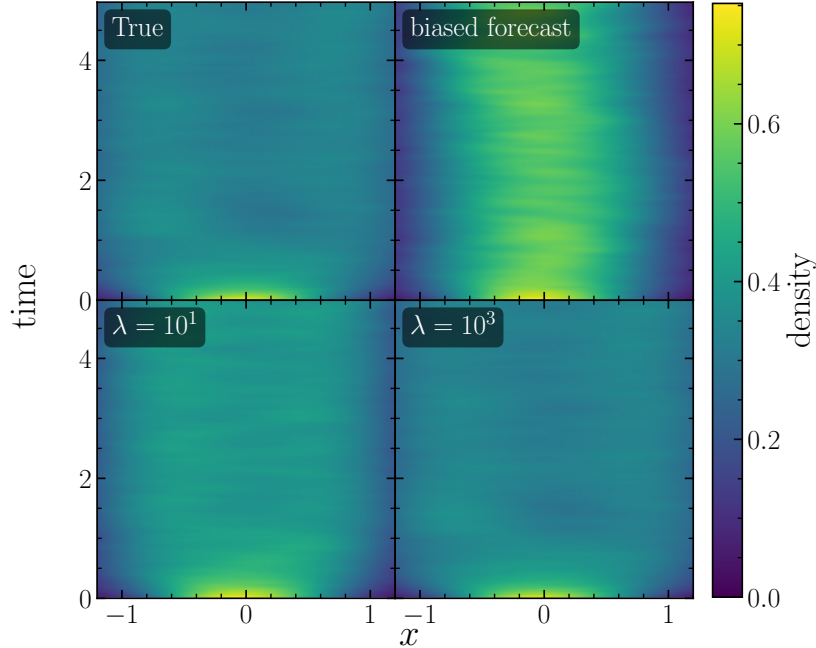


Figure 6: **Space-time density evolution in the multimodal benchmark** ($a = 1.5$). Top-left: reference density. Top-right: biased forecast. Bottom-left: assimilated density with $\lambda = 10$. Bottom-right: assimilated density with $\lambda = 1000$. Larger nudging strength restores both the location and the spread of the true law.

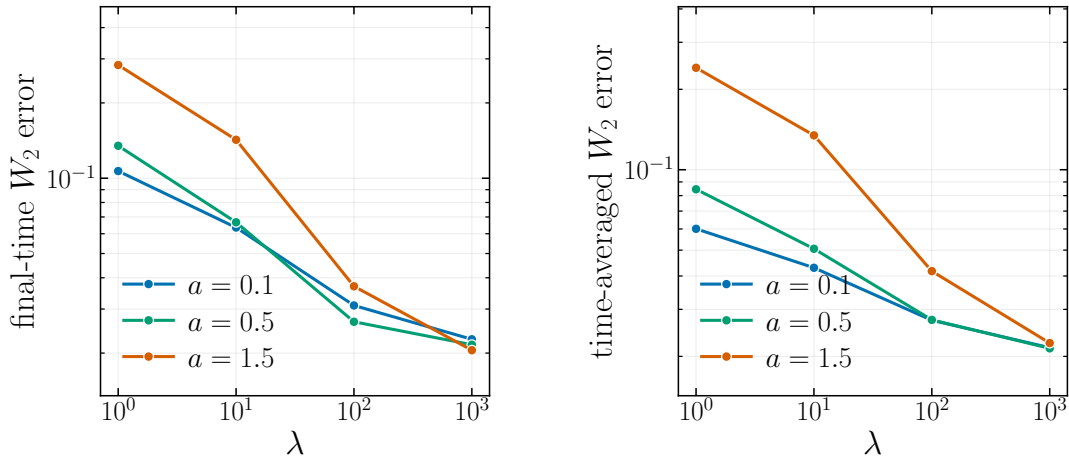


Figure 7: **Distributional error versus nudging strength in the multimodal benchmark**. Left: final-time W_2 error. Right: time-averaged W_2 error. In all three biased models $a \in \{0.1, 0.5, 1.5\}$, stronger nudging produces a smaller distributional discrepancy with the reference dynamics.

and the electrostatic potential ϕ solves the Poisson equation

$$-\Delta\phi = \rho - 1, \quad \rho(x, t) = \int f(x, v, t) dv.$$

In these Vlasov-Poisson experiments, observations are taken on the full phase space. Namely, for $z = (x, v)$, we use the smoothed phase-space density

$$y_t(z) = (K_h * f_t)(z),$$

and compute the nudging residual between this observation and the smoothed forecast density. Although ρ_t determines the electric field through Poisson's equation, it does not uniquely determine the full kinetic distribution $f(x, v, t)$. The spatial-marginal observation problem is therefore left to future work.

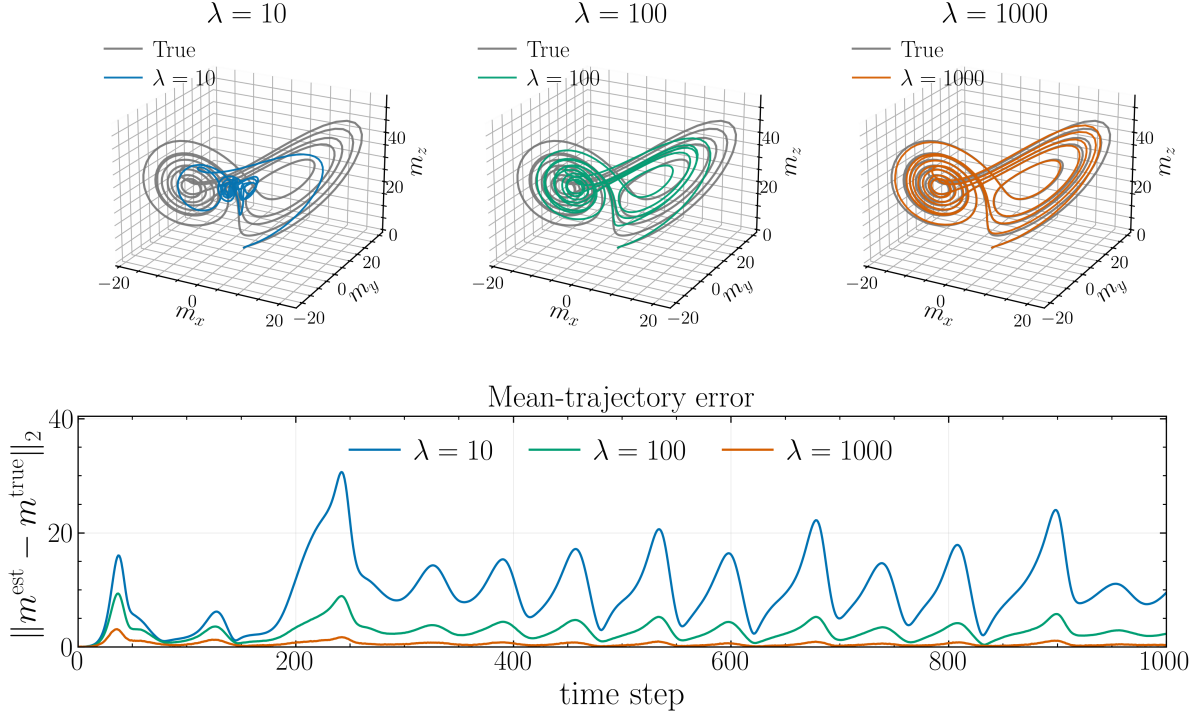


Figure 8: **Mean-field Lorenz test.** Top: trajectories of the mean state $(m_{t,x}, m_{t,y}, m_{t,z})$ for $\lambda = 1000, 100, 10$ compared with the reference attractor. Bottom: time series of the mean-trajectory error. Stronger nudging keeps the corrected dynamics close to the true attractor and suppresses intermittent excursions.

Landau damping We first consider the classical Landau damping problem. The ground-truth dynamics are initialized with

$$f(x, v, 0) \propto f_{M,1}(v)(1 + \varepsilon \cos(\kappa x)), \quad f_{M,\sigma}(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{v^2}{2\sigma^2}\right),$$

which corresponds to a Maxwellian velocity distribution perturbed by a spatial density modulation. In our experiments, we take $\varepsilon = 0.5$ and $\kappa = \frac{1}{2}$. The spatial domain is periodic with period 4π , i.e., $x \in [0, 4\pi]$, while the velocity domain is truncated to $v \in [-6, 6]$ for numerical implementation.

To quantify the damping, we monitor the amplitude of the fundamental Fourier mode of the electric field,

$$E_k(t) = \left| \hat{E}(k, t) \right|, \quad \hat{E}(k, t) = \frac{1}{L} \int_0^L E(x, t) e^{-ikx} dx,$$

with $k = \kappa$ and $L = 4\pi$. In the linear regime, E_k decays exponentially at the Landau damping rate.

The forecast model uses the same dynamical system but is initialized from an unperturbed Maxwellian ($\varepsilon = 0$), resulting in a biased prediction that lacks the initial density modulation. Figure 9 compares the reference solution, the biased forecast, and the assimilated dynamics. The results show that the nudging term successfully reconstructs the damping behavior and recovers the correct exponential decay of E_k .

Two-stream instability We also consider the two-stream instability, which exhibits strongly nonlinear behavior. The ground-truth initial condition is given by

$$f(x, v, 0) \propto (0.5f_{M,\sigma}(v - u_0) + 0.5f_{M,\sigma}(v + u_0))(1 + \varepsilon \cos(\kappa x)),$$

with parameters $\sigma = 0.2$, $u_0 = 1$, $\varepsilon = 0.01$, and $\kappa = 0.5$. This corresponds to a bimodal velocity distribution with a small spatial perturbation. In contrast, the forecast model is initialized from a single Maxwellian profile,

$$f(x, v, 0) \propto f_{M,1}(v)(1 + \varepsilon \cos(\kappa x)),$$

which fails to capture the underlying two-stream structure.

Figure 10 shows the evolution of the system at different times. The biased model is unable to reproduce the filamentation and phase-space structures characteristic of the instability. In contrast, the assimilated dynamics recover the correct

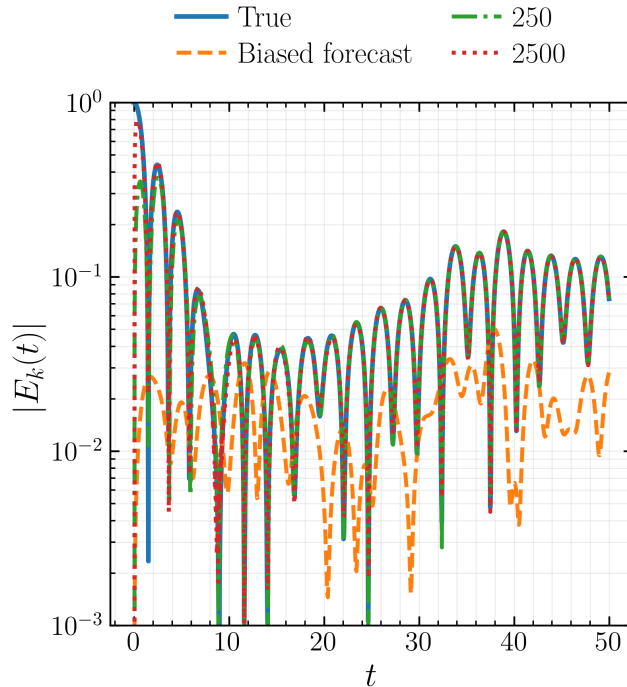


Figure 9: **Landau damping.** Time evolution of the fundamental electric-field mode $|E_k(t)|$ for the reference dynamics, the biased forecast, and the assimilated forecasts. The nudged systems recover the damping profile from density-level observations.

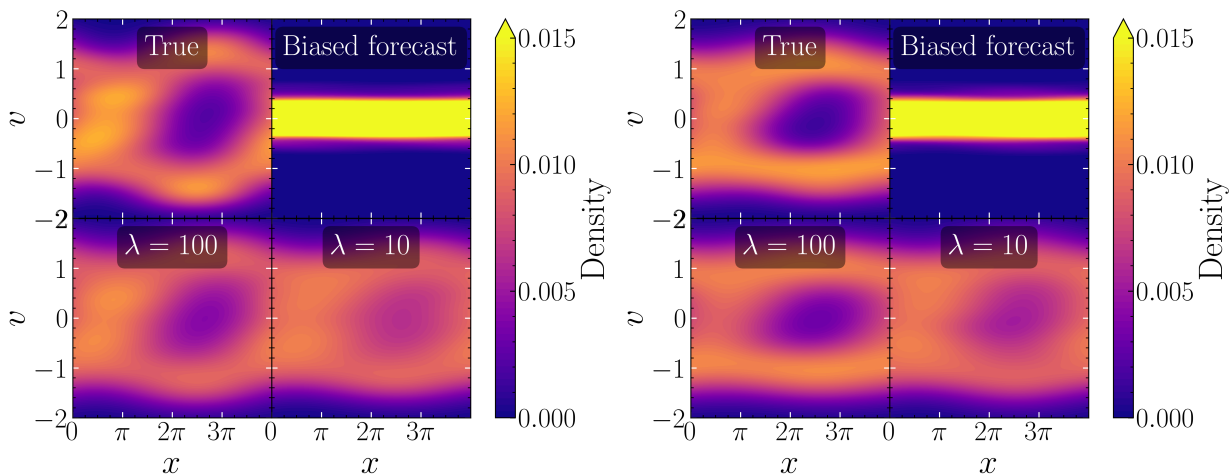


Figure 10: **Two-stream instability.** Phase-space density at intermediate (left, step 1000) and late (right, step 2000) times. Each panel compares the reference two-stream dynamics (top), the biased single-Maxwellian forecast (center), and the assimilated solution (bottom). The nudging recovers the filamentation and vortex-merging structures absent from the biased model.

qualitative features, demonstrating that the proposed nudging mechanism can reconstruct complex multimodal structures from macroscopic observations.

4.5 Collective-motion data

We apply Multiscale Nudging to trajectory data obtained from experiments with approximately $N \approx 1126$ fish swimming in a quasi-two-dimensional tank, processed by the TREX tracker [67]. The dataset consists of discrete-time position measurements $\{\mathbf{X}_i(t_n)\}_{n=0}^T$, which are spatially and temporally incomplete due to visual occlusions and

tracking dropouts; at every frame, a substantial fraction of trajectories carry a missing flag, and identities are not consistent across frames [67]. Exact particle-to-particle correspondence is therefore unavailable, which makes a permutation-invariant, density-level assimilation strategy the natural choice.

Since the underlying physical interactions among the fish are not known in closed form, we follow [18] and use a learned mean-field drift

$$\mathbf{b}_{\text{model}}(\mathbf{x}, \nu) = \text{MLP}_{\theta_2}(\mathbf{x}, \langle \text{MLP}_{\theta_1}(\cdot) \rangle_{\nu}) \quad (19)$$

trained on a held-out portion of the same dataset. Two modest MLPs ([12, 24, 48] and [128, 128, 128, 2]) are composed as a feature network and a descriptor network, with the inner empirical mean providing the mean-field coupling. The trained drift captures the average circulating motion of the school of fish but, as we show below, accumulates errors when rolled out without observational feedback.

The observation operator \mathcal{O}_h is the same Gaussian KDE used throughout this paper; at every observed frame we evaluate the smoothed density $\mu_t^{\text{obs}} = K_h * \hat{\mu}_t^{\text{TREX}}$ on a 125×125 grid covering the $[0, L]^2$ tank ($L = 128$ px) with bandwidth $h = 2$ px and exclude particles flagged as missing from the empirical measure used to form μ_t^{obs} . The assimilated forecast is integrated by Euler–Maruyama with $\Delta t = 0.025$ s and $L_{\text{nud}} = 100$ inner nudging substeps per outer step; the nudging strength is $\lambda = 1$. We compare three trajectories sharing the same initial configuration $\mathbf{Z}_0 = \mathbf{X}(t_0)$: the experimental reference, a biased forecast obtained by rolling out (19) without any observation feedback, and the Multiscale Nudging forecast obtained from the smoothed-density observations.

Figure 11 compares the raw image, the reference ensemble, the biased forecast, and the Multiscale Nudging forecast at four snapshots $t \in \{0, 1.25, 3.75, 6.25\}$ s. Each particle is rendered as a small line segment oriented along its instantaneous velocity, so that the ring-shaped circulation pattern of the school is directly visible. The reference school maintains a coherent counterclockwise circulation along the tank boundary throughout the window; the biased forecast captures the initial pattern but loses coherence quickly. In particular, by $t = 3.75$ s, the rotational structure is fragmented, and spurious particles concentrate near the upper-left corner. The Multiscale Nudging forecast tracks both the boundary circulation and the central low-density core for the full window. The same comparison at the level of the smoothed density is shown in Figure 12. The reference field exhibits a characteristic ring of high occupancy along the tank boundary together with a structured interior. The biased forecast develops a spurious concentration in the upper-left corner at $t = 1.25$ s and never recovers the boundary ring, whereas the assimilated forecast matches the reference density throughout.

To quantify the assimilation, we compute the L^2 error of the density:

$$\text{Err}_t = \|K_h * \hat{\mu}_t^{\text{forecast}} - K_h * \hat{\mu}_t^{\text{ref}}\|_{L^2(\Omega)},$$

where the empirical measures use only the unmasked particles at each frame. Figure 13 reports Err_t for the biased and assimilated forecasts on a logarithmic scale. The biased forecast saturates near $\text{Err} \approx 30$ within roughly one second, whereas the Multiscale Nudging forecast remains close to $\text{Err} \approx 2$ for the entire window; the assimilated error stays approximately one order of magnitude below the biased baseline.

5 Conclusion

We introduced Multiscale Nudging, a measure-based method for assimilating coarse observations into microscopic mean-field particle dynamics. The method addresses a representation mismatch that arises when the forecast is a labeled particle ensemble, but the data are available only as smoothed, permutation-invariant densities. By defining the observation mismatch on probability measures and applying its Wasserstein gradient at the particle level, the method produces a practical feedback correction without particle matching, model linearization, or ensemble covariance estimation. For a fixed observation scale, we established well-posedness of the assimilated McKean–Vlasov dynamics and propagation of chaos for the particle approximation. We also proved an L^2 -stability estimate under exact smoothed observations and a kernel-scale observability condition, showing exponential convergence up to a model-error-dependent bias floor. Numerical results across Gaussian, multimodal, chaotic, kinetic, and experimental collective-motion systems show that the approach can correct biased microscopic forecasts using only density-level information.

Several limitations remain. The present stability result assumes exact observations and requires the error to be observable at the kernel scale. In problems with hidden variables, such as kinetic systems observed only through spatial density, some phase-space structures may be only partially identifiable. Extending the method to noisy observations, nonlinear observation operators, adaptive bandwidths, and partial observability is a natural direction for future work.

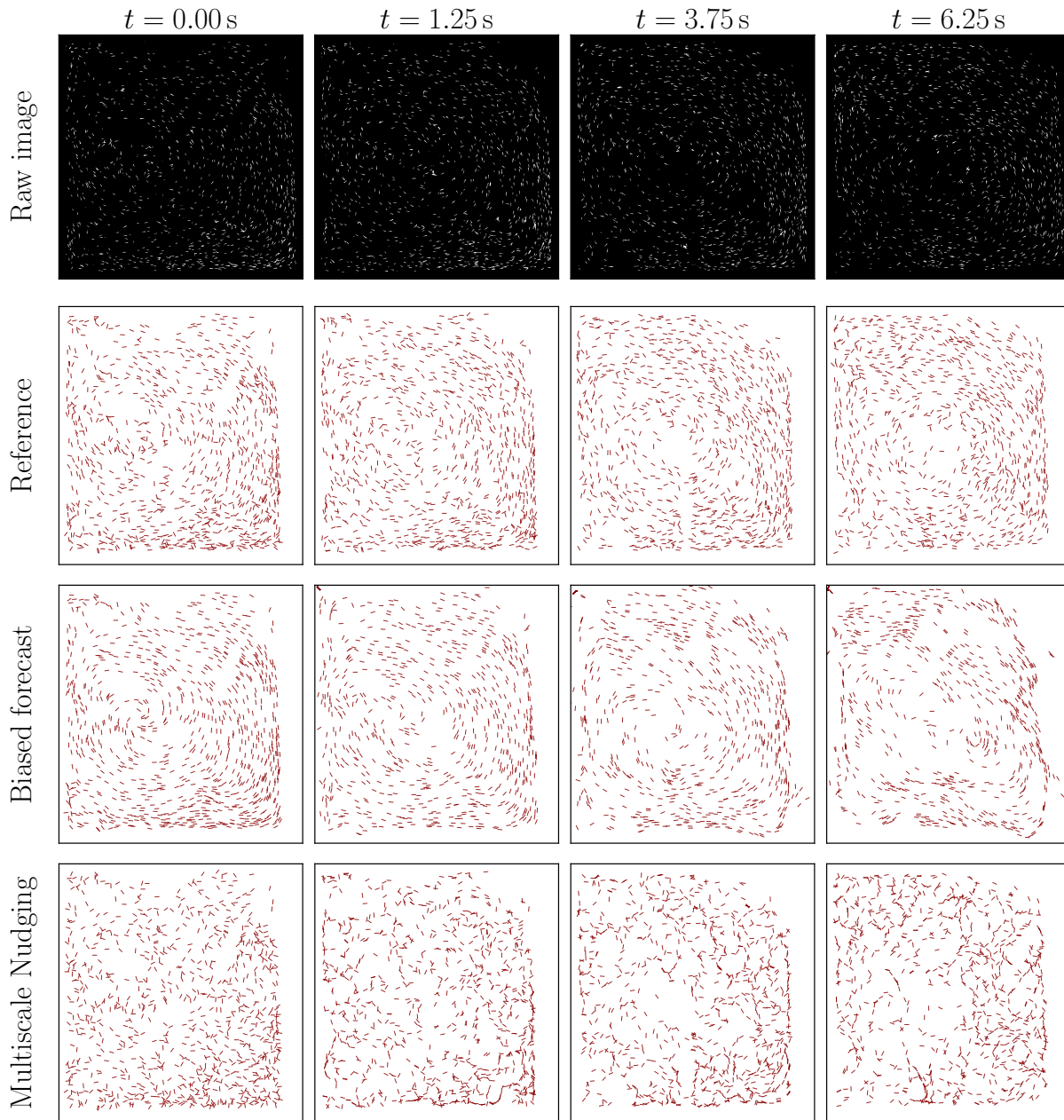


Figure 11: **Particle-level comparison on the fish dataset.** Rows, top to bottom: raw image frame, reference ensemble, biased forecast (learned drift, no observation feedback), and Multiscale Nudging. Columns: $t = 0, 1.25, 3.75, 6.25$ s after the common initial condition. Each particle is shown as a short segment oriented along its instantaneous velocity. The biased forecast loses the boundary ring and accumulates spurious mass in the upper-left corner; the Multiscale Nudging forecast preserves the circulation pattern of the school.

Acknowledgments

This work used Anvil at Purdue through allocation MTH260007 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program[68], which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

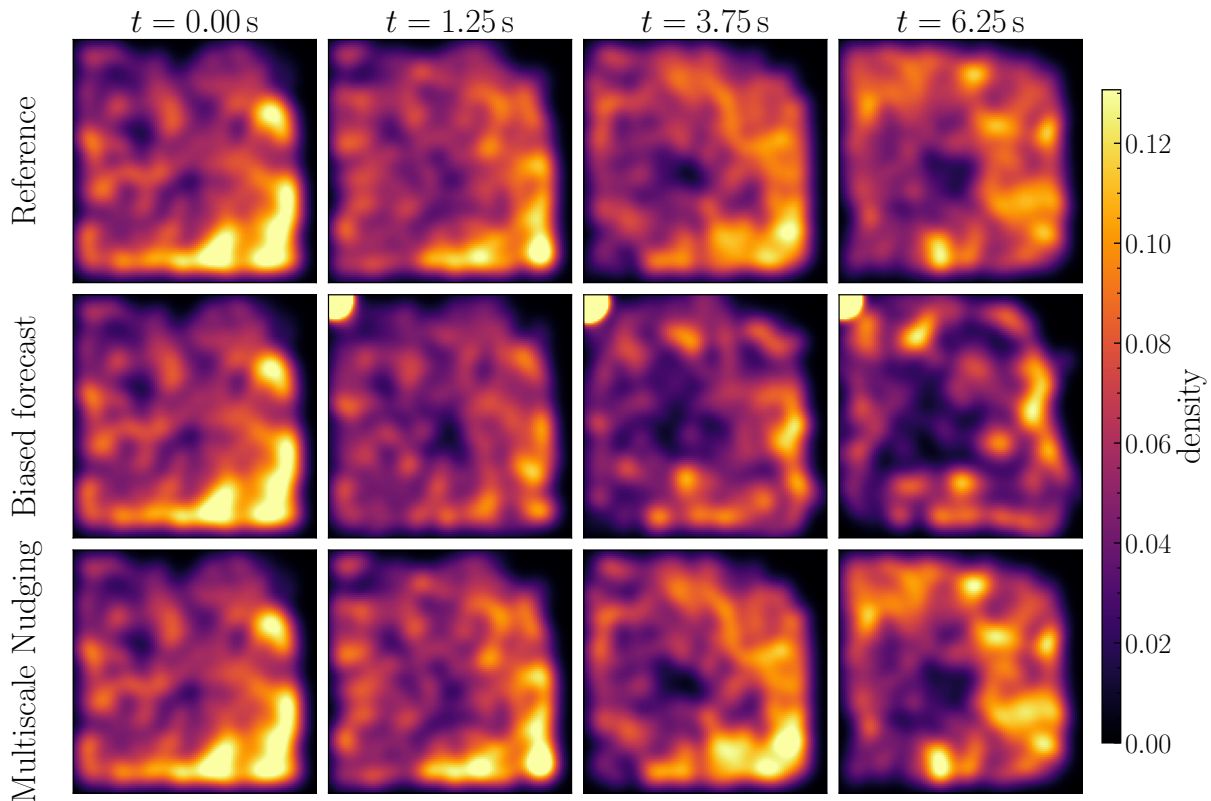


Figure 12: **Smoothed-density comparison on the fish dataset.** Rows: reference density, biased forecast density, Multiscale Nudging density. Columns: $t = 0, 1.25, 3.75, 6.25$ s. Densities are evaluated by Gaussian KDE with bandwidth $h = 2$ px on a 125×125 grid. The biased forecast develops a singular concentration in the upper-left corner; the assimilated forecast tracks the reference throughout.

References

- [1] Jürgen Jost. *Dynamical systems: examples of complex behaviour*. Springer, 2005.
- [2] Stephen Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*. Springer, 2003.
- [3] Mark Kac. Foundations of kinetic theory. In *Proc. Third Berkeley Symp. on Math. Stat. and Prob.*, volume 3, pages 171–197, 1956.
- [4] Henry P McKean Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences*, 56(6):1907–1911, 1966.
- [5] PE Chaudru De Raynal. Strong well posedness of Mckean–Vlasov stochastic differential equations with Hölder drift. *Stochastic Processes and their Applications*, 130(1):79–107, 2020.
- [6] Xing Huang and Feng-Yu Wang. Distribution dependent SDEs with singular coefficients. *Stochastic Processes and their Applications*, 129(11):4747–4770, 2019.
- [7] Yuliya Mishura and Alexander Veretennikov. Existence and uniqueness theorems for solutions of Mckean–Vlasov stochastic equations. *Theory of Probability and Mathematical Statistics*, 103:59–101, 2020.
- [8] Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d’été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 2006.
- [9] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. ii. applications. *arXiv preprint arXiv:2106.14812*, 2021.
- [10] Pierre-Emmanuel Jabin and Zhenfu Wang. Mean field limit for stochastic particle systems. In *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, pages 379–402. Springer, 2017.
- [11] Rainer Buckdahn, Juan Li, and Jin Ma. A mean-field stochastic control problem with partial observations. *The Annals of Applied Probability*, 27(5):3201–3245, 2017.

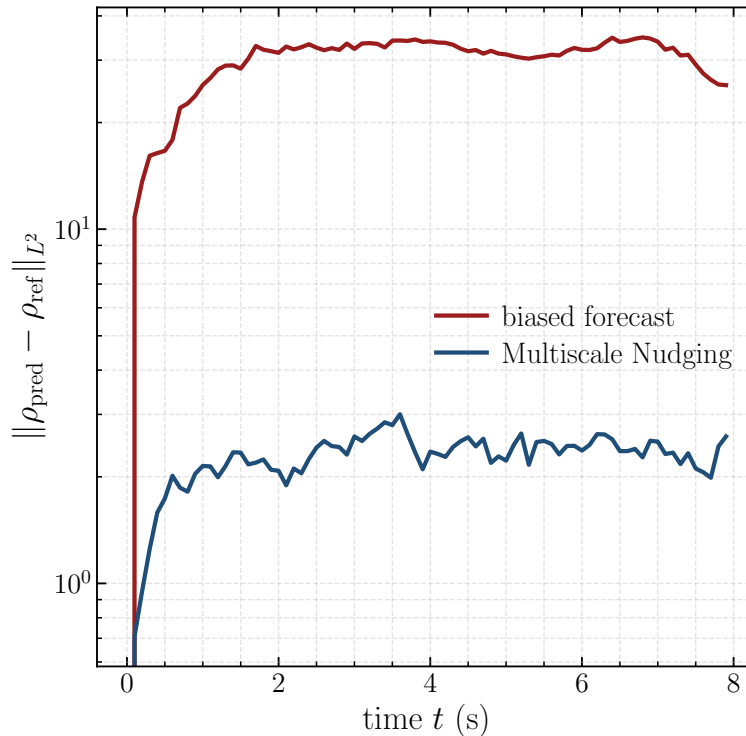


Figure 13: **Density L^2 error on the fish dataset.** Dynamics of $\|K_h * \hat{\mu}_t^{\text{forecast}} - K_h * \hat{\mu}_t^{\text{ref}}\|_{L^2}$ for the biased forecast (red) and Multiscale Nudging (blue). The biased forecast saturates near 30 within ~ 1 s; the assimilated forecast stays approximately one order of magnitude lower.

- [12] Lisang Ding, Wuchen Li, Stanley Osher, and Wotao Yin. A mean field game inverse problem. *Journal of Scientific Computing*, 92(1):7, 2022.
- [13] Qiang Liu. Stein variational gradient descent as gradient flow. *Advances in neural information processing systems*, 30, 2017.
- [14] José A Carrillo, Yifan Chen, Daniel Zhengyu Huang, Jiaoyang Huang, and Dongyi Wei. Fisher-Rao gradient flow: geodesic convexity and functional inequalities. *arXiv preprint arXiv:2407.15693*, 2024.
- [15] Grant Rotskoff and Eric Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022.
- [16] Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proceedings of the National Academy of Sciences*, 116(29):14424–14433, 2019.
- [17] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. *Journal of Machine Learning Research*, 22(32):1–67, 2021.
- [18] Liyao Lyu, Xinyue Yu, and Hayden Schaeffer. MVNN: A measure-valued neural network for learning McKean-Vlasov dynamics from particle data. *arXiv preprint arXiv:2604.00333*, 2026.
- [19] Renyi Chen and Molei Tao. Data-driven prediction of general Hamiltonian dynamics via learning exactly-symplectic maps. In *International conference on machine learning*, pages 1717–1727. PMLR, 2021.
- [20] Qiang Du, Yiqi Gu, Haizhao Yang, and Chao Zhou. The discovery of dynamics via linear multistep methods and deep learning: Error estimation. *SIAM Journal on Numerical Analysis*, 60(4):2014–2045, 2022.
- [21] Louis Sharrock, Nikolas Kantas, Panos Parpas, and Grigorios A Pavliotis. Parameter estimation for the McKean-Vlasov stochastic differential equation. *arXiv preprint arXiv:2106.13751*, 2021.
- [22] Zecheng Zhang, Christian Moya, Wing Tat Leung, Guang Lin, and Hayden Schaeffer. Bayesian deep operator learning for homogenized to fine-scale maps for multiscale PDE. *Multiscale Modeling & Simulation*, 22(3):956–972, 2024.
- [23] Hayden Schaeffer, Giang Tran, Rachel Ward, and Linan Zhang. Extracting structured dynamical systems using sparse optimization with very few samples. *Multiscale Modeling & Simulation*, 18(4):1435–1461, 2020.

- [24] Yuxuan Liu, Scott G McCalla, and Hayden Schaeffer. Random feature models for learning interacting dynamical systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 479(2275), 2023.
- [25] Yubin Lu, Xiaofan Li, Chun Liu, Qi Tang, and Yiwei Wang. Learning generalized diffusions using an energetic variational approach. *arXiv preprint arXiv:2412.04480*, 2024.
- [26] Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2197), 2017.
- [27] Hayden Schaeffer and Scott G McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.
- [28] Pierre-Emmanuel Jabin and Zhenfu Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones mathematicae*, 214(1):523–591, 2018.
- [29] Patrick Cattiaux, Arnaud Guillin, and Florent Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140(1):19–40, 2008.
- [30] Yoonsang Lee and Andrew J Majda. Multiscale methods for data assimilation in turbulent systems. *Multiscale Modeling & Simulation*, 13(2):691–713, 2015.
- [31] Quanling Deng, Nan Chen, Samuel N Stechmann, and Jiuhua Hu. LEMDA: A Lagrangian-Eulerian multiscale data assimilation framework. *Journal of Advances in Modeling Earth Systems*, 17(2):e2024MS004259, 2025.
- [32] John Harlim and Andrew J Majda. Test models for filtering with superparameterization. *Multiscale Modeling & Simulation*, 11(1):282–308, 2013.
- [33] Quanling Deng, Samuel N Stechmann, and Nan Chen. Particle-continuum multiscale modeling of sea ice floes. *Multiscale Modeling & Simulation*, 22(1):230–255, 2024.
- [34] Dmitry Burov, Dimitrios Giannakis, Krithika Manohar, and Andrew Stuart. Kernel analog forecasting: Multiscale test problems. *Multiscale Modeling & Simulation*, 19(2):1011–1040, 2021.
- [35] José A Carrillo and Young-Pil Choi. Mean-field limits: from particle descriptions to macroscopic equations. *Archive for Rational Mechanics and Analysis*, 241(3):1529–1573, 2021.
- [36] Thierry Paul and Emmanuel Trélat. From microscopic to macroscopic scale equations: mean field, hydrodynamic and graph limits. *arXiv preprint arXiv:2209.08832*, 2022.
- [37] Geir Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.
- [38] Geir Evensen. *Data assimilation: the ensemble Kalman filter*. Springer, 2009.
- [39] Andrew C Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.
- [40] PHILIPPE Courtier, J-N Thépaut, and Anthony Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.
- [41] Nicolas Chopin, Omiros Papaspiliopoulos, et al. *An introduction to sequential Monte Carlo*, volume 4. Springer, 2020.
- [42] Abderrahim Azouani, Eric Olson, and Edriss S Titi. Continuous data assimilation using general interpolant observables. *Journal of Nonlinear Science*, 24(2):277–304, 2014.
- [43] Hakima Bessaih, Eric Olson, and Edriss S Titi. Continuous data assimilation with stochastically noisy data. *Nonlinearity*, 28(3):729–753, 2015.
- [44] Joshua Newey, Jared P Whitehead, and Elizabeth Carlson. Model discovery on the fly using continuous data assimilation. *Journal of Computational Physics*, 537:114121, 2025.
- [45] Jingcheng Lu, Kunlun Qi, Li Wang, and Jeff Calder. Continuous data assimilation for hydrodynamics: consistent discretization and application to moment recovery. *Journal of Computational Physics*, 538:114199, 2025.
- [46] Patricio Clark Di Leoni, Andrea Mazzino, and Luca Biferale. Inferring flow parameters and turbulent configuration with physics-informed data assimilation and spectral nudging. *Physical Review Fluids*, 3(10):104604, 2018.
- [47] Aseel Farhat, Nathan E Glatt-Holtz, Vincent R Martinez, Shane A McQuarrie, and Jared P Whitehead. Data assimilation in large Prandtl Rayleigh–Bénard convection from thermal measurements. *SIAM Journal on Applied Dynamical Systems*, 19(1):510–540, 2020.
- [48] Débora AF Albanez, Helena J Nussenzveig Lopes, and Edriss S Titi. Continuous data assimilation for the three-dimensional Navier–Stokes- α model. *Asymptotic Analysis*, 97(1-2):139–164, 2016.

- [49] Michael S Jolly, Vincent R Martinez, and Edriss S Titi. A data assimilation algorithm for the subcritical surface quasi-geostrophic equation. *Advanced Nonlinear Studies*, 17(1):167–192, 2017.
- [50] Tao Yang, Prashant G Mehta, and Sean P Meyn. Feedback particle filter with mean-field coupling. In *2011 50th IEEE conference on decision and control and European control conference*, pages 7909–7916. IEEE, 2011.
- [51] Alastair Gregory, Colin J Cotter, and Sebastian Reich. Multilevel ensemble transform particle filtering. *SIAM Journal on Scientific Computing*, 38(3):A1317–A1338, 2016.
- [52] Mr Fred Daum, Jim Huang, and Arjang Noushin. Exact particle flow for nonlinear filters. In *Signal processing, sensor fusion, and target recognition XIX*, volume 7697, page 92. SPIE, 2010.
- [53] Manuel Pulido and Peter Jan van Leeuwen. Sequential Monte Carlo with kernel embedded mappings: The mapping particle filter. *Journal of Computational Physics*, 396:400–415, 2019.
- [54] Edoardo Calvello, Sebastian Reich, and Andrew M Stuart. Ensemble Kalman methods: A mean-field perspective. *Acta Numerica*, 34:123–291, 2025.
- [55] 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.
- [56] Feng Bao, Zezhong Zhang, and Guannan Zhang. A score-based filter for nonlinear data assimilation. *Journal of Computational Physics*, 514:113207, 2024.
- [57] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.
- [58] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2005.
- [59] Yifan Chen, Daniel Zhengyu Huang, Jiaoyang Huang, Sebastian Reich, and Andrew M Stuart. Gradient flows for sampling: Mean-field models, Gaussian approximations and affine invariance. *arXiv preprint arXiv:2302.11024*, 2023.
- [60] Yifan Chen, Daniel Zhengyu Huang, Jiaoyang Huang, Sebastian Reich, and Andrew M Stuart. Sampling via gradient flows in the space of probability measures. *arXiv preprint arXiv:2310.03597*, 2023.
- [61] José A Carrillo, Robert J McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana*, 19(3):971–1018, 2003.
- [62] José A Carrillo, Katy Craig, and Yao Yao. Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In *Active Particles, Volume 2: Advances in Theory, Models, and Applications*, pages 65–108. Springer, 2019.
- [63] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. models and methods. *arXiv preprint arXiv:2203.00446*, 2022.
- [64] Michele Ballerini, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Vivien Lecomte, Alberto Orlandi, Giorgio Parisi, Andrea Procaccini, et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the national academy of sciences*, 105(4):1232–1237, 2008.
- [65] Andrea Cavagna, Alessio Cimarrelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.
- [66] He-Peng Zhang, Avraham Be’er, E-L Florin, and Harry L Swinney. Collective motion and density fluctuations in bacterial colonies. *Proceedings of the National Academy of Sciences*, 107(31):13626–13630, 2010.
- [67] Tristan Walter and Iain D Couzin. TRex, a fast multi-animal tracking system with markerless identification, and 2d estimation of posture and visual fields. *Elife*, 10:e64000, 2021.
- [68] Timothy J Boerner, Stephen Deems, Thomas R Furlani, Shelley L Knuth, and John Towns. ACCESS: Advancing innovation: NSF’s advanced cyberinfrastructure coordination ecosystem: Services & support. In *Practice and experience in advanced research computing 2023: Computing for the common good*, pages 173–176. Association for Computing Machinery, Inc, 2023.
- [69] René Carmona. *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*. SIAM, 2016.
- [70] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26, 1958.

[71] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

A Proof of Lemma 3.1

Proof. We first prove the regularity statements. The periodized Gaussian $K_h^\mathbb{T}$ is C^∞ , periodic, even, and has unit mass on \mathbb{T}^d . Moreover, all of its derivatives are obtained by periodizing derivatives of the Euclidean Gaussian, and the corresponding series converge uniformly for each fixed $h > 0$. Hence $K_h^\mathbb{T} \in C^\infty(\mathbb{T}^d)$. Since convolution preserves smoothness on the torus,

$$\tilde{K}_h = K_h^\mathbb{T} *_{\mathbb{T}} K_h^\mathbb{T} \in C^\infty(\mathbb{T}^d).$$

Because \mathbb{T}^d is compact, all derivatives of \tilde{K}_h are bounded. In particular,

$$\nabla \tilde{K}_h \in L^\infty(\mathbb{T}^d), \quad D^2 \tilde{K}_h \in L^\infty(\mathbb{T}^d).$$

Thus \tilde{K}_h is globally Lipschitz, and $\nabla \tilde{K}_h$ is globally Lipschitz. This proves (1) and (2).

We now prove the approximation statement. Let

$$v(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{v}_{\mathbf{k}} e^{2\pi i \mathbf{k} \cdot \mathbf{x}}$$

be the Fourier series of v . The Fourier coefficients of the periodized Gaussian are

$$\widehat{K}_h^\mathbb{T}(\mathbf{k}) = e^{-\pi^2 h^2 |\mathbf{k}|^2},$$

therefore,

$$\widehat{\tilde{K}_h}(\mathbf{k}) = \widehat{K}_h^\mathbb{T}(\mathbf{k})^2 = e^{-2\pi^2 h^2 |\mathbf{k}|^2}.$$

Since differentiation commutes with convolution on \mathbb{T}^d ,

$$\widehat{(\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v))}(\mathbf{k}) = 2\pi i \mathbf{k} \left(1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2}\right) \hat{v}_{\mathbf{k}}.$$

By Parseval's identity,

$$\|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2}^2 = \sum_{\mathbf{k} \in \mathbb{Z}^d} (2\pi)^2 |\mathbf{k}|^2 \left|1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2}\right|^2 |\hat{v}_{\mathbf{k}}|^2.$$

If $v \in H^1(\mathbb{T}^d)$, then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} (2\pi)^2 |\mathbf{k}|^2 |\hat{v}_{\mathbf{k}}|^2 < \infty,$$

so for each fixed \mathbf{k} , we have the following limit

$$1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

The coefficients are bounded above and below by:

$$0 \leq 1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2} \leq 1.$$

And applying the dominated convergence theorem therefore yields the limit

$$\|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2(\mathbb{T}^d)} \rightarrow 0 \quad \text{as } h \downarrow 0.$$

It remains to prove the quantitative estimate for $v \in H^2(\mathbb{T}^d)$. Using the elementary inequality

$$1 - e^{-t} \leq C\sqrt{t}, \quad t \geq 0,$$

we obtain

$$1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2} \leq Ch|\mathbf{k}|.$$

Hence

$$\begin{aligned} \|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2}^2 &= \sum_{\mathbf{k} \in \mathbb{Z}^d} (2\pi)^2 |\mathbf{k}|^2 \left|1 - e^{-2\pi^2 h^2 |\mathbf{k}|^2}\right|^2 |\hat{v}_{\mathbf{k}}|^2 \\ &\leq Ch^2 \sum_{\mathbf{k} \in \mathbb{Z}^d} |\mathbf{k}|^4 |\hat{v}_{\mathbf{k}}|^2. \end{aligned}$$

By the Fourier characterization of the Sobolev norm, we have

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} |\mathbf{k}|^4 |\widehat{v}_{\mathbf{k}}|^2 \sim \|D^2 v\|_{L^2(\mathbb{T}^d)}^2,$$

and thus we conclude that

$$\|\nabla v - \nabla(\tilde{K}_h *_{\mathbb{T}} v)\|_{L^2(\mathbb{T}^d)} \leq Ch \|D^2 v\|_{L^2(\mathbb{T}^d)},$$

which completes the proof. \square

B Proof of Proposition 3.2

Proof. The proof is based on Theorem 1.7 in [69]. We only need to show that for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$ and $\nu_1, \nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned} & \|\mathbf{b}_{\text{model}}(\mathbf{x}_1, \nu_1) - \lambda \nabla(\tilde{K}_h * \nu_1 - K_h * \mu^{\text{obs}})(\mathbf{x}_1) \\ & \quad - (\mathbf{b}_{\text{model}}(\mathbf{x}_2, \nu_2) - \lambda \nabla(\tilde{K}_h * \nu_2 - K_h * \mu^{\text{obs}})(\mathbf{x}_2))\| \\ & \leq C(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\nu_1, \nu_2)). \end{aligned}$$

By the Lipschitz property of $\mathbf{b}_{\text{model}}$ and triangle inequality, we have:

$$\begin{aligned} & \|\mathbf{b}_{\text{model}}(\mathbf{x}_1, \nu_1) - \lambda \nabla(\tilde{K}_h * \nu_1 - K_h * \mu^{\text{obs}})(\mathbf{x}_1) \\ & \quad - (\mathbf{b}_{\text{model}}(\mathbf{x}_2, \nu_2) - \lambda \nabla(\tilde{K}_h * \nu_2 - K_h * \mu^{\text{obs}})(\mathbf{x}_2))\| \\ & \leq c(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\nu_1, \nu_2)) + \lambda \|\nabla(\tilde{K}_h * \nu_1)(\mathbf{x}_1) - \nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_1)\| + \\ & \quad \lambda \|\nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_1) - \nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_2)\| \\ & \quad + \lambda \|\nabla(K_h * \mu^{\text{obs}})(\mathbf{x}_1) - \nabla(K_h * \mu^{\text{obs}})(\mathbf{x}_2)\| \end{aligned} \tag{20}$$

The difference is bounded by

$$\begin{aligned} & \|\nabla(\tilde{K}_h * \nu_1)(\mathbf{x}_1) - \nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_1)\| \\ & = \left\| \int \nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) d\nu_1(\mathbf{y}) - \int \nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) d\nu_2(\mathbf{y}) \right\| \\ & = \left\| \int \nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) d(\nu_1 - \nu_2)(\mathbf{y}) \right\| \\ & \leq LW_1(\nu_1, \nu_2) \\ & \leq LW_2(\nu_1, \nu_2) \end{aligned}$$

using Kantorovich-Rubinstein duality and the Lipschitz continuity of $\nabla \tilde{K}_h$.

Using the global Lipschitz properties and that these measures are unit mass

$$\begin{aligned} & \|\nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_1) - \nabla(\tilde{K}_h * \nu_2)(\mathbf{x}_2)\| \\ & = \left\| \int \nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) d\nu_2(\mathbf{y}) - \int \nabla \tilde{K}_h(\mathbf{x}_2 - \mathbf{y}) d\nu_2(\mathbf{y}) \right\| \\ & \leq \int \|\nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) - \nabla \tilde{K}_h(\mathbf{x}_2 - \mathbf{y})\| d\nu_2(\mathbf{y}) \\ & \leq L\|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned}$$

We have that

$$\begin{aligned} & (K_h * \mu^{\text{obs}})(\mathbf{x}) \\ & = \int K_h(\mathbf{x} - \mathbf{z}) \mu^{\text{obs}}(\mathbf{z}) d\mathbf{z} \\ & = \int \int K_h(\mathbf{x} - \mathbf{z}) K_h(\mathbf{z} - \mathbf{y}) \mu_t(d\mathbf{y}) d\mathbf{z} \\ & = (\tilde{K}_h * \mu_t)(\mathbf{x}). \end{aligned}$$

So for the last term on the right-hand side of (20),

$$\begin{aligned}
& \|\nabla(K_h * \mu^{\text{obs}})(\mathbf{x}_1) - \nabla(K_h * \mu^{\text{obs}})(\mathbf{x}_2)\| \\
&= \|\nabla(\tilde{K}_h * \mu_t)(\mathbf{x}_1) - \nabla(\tilde{K}_h * \mu_t)(\mathbf{x}_2)\| \\
&= \left\| \int \nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) d\mu_t(\mathbf{y}) - \int \nabla \tilde{K}_h(\mathbf{x}_2 - \mathbf{y}) d\mu_t(\mathbf{y}) \right\| \\
&\leq \int \|\nabla \tilde{K}_h(\mathbf{x}_1 - \mathbf{y}) - \nabla \tilde{K}_h(\mathbf{x}_2 - \mathbf{y})\| d\mu_t(\mathbf{y}) \\
&\leq L \|\mathbf{x}_1 - \mathbf{x}_2\|
\end{aligned}$$

Simplifying (20), we obtain the desired result:

$$\begin{aligned}
& \|\mathbf{b}_{\text{model}}(\mathbf{x}_1, \nu_1) - \lambda \nabla(\tilde{K}_h * \nu_1 - K_h * \mu^{\text{obs}})(\mathbf{x}_1) - \\
& \quad (\mathbf{b}_{\text{model}}(\mathbf{x}_2, \nu_2) - \lambda \nabla(\tilde{K}_h * \nu_2 - K_h * \mu^{\text{obs}})(\mathbf{x}_2))\| \\
&\leq c(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\nu_1, \nu_2)) + \lambda L W_2(\nu_1, \nu_2) + \lambda L \|\mathbf{x}_1 - \mathbf{x}_2\| + \lambda L \|\mathbf{x}_1 - \mathbf{x}_2\| \\
&\leq C(\|\mathbf{x}_1 - \mathbf{x}_2\| + W_2(\nu_1, \nu_2)),
\end{aligned}$$

where $C := 2\lambda L + c$. Thus the modified drift inherits the required global Lipschitz continuity in both its spatial and measure arguments, which is the key condition needed for the well-posedness. \square

C Proof of Proposition 3.3

Proof. We use a synchronous coupling argument, as in [8] and Theorem 3.1 of [9]. We construct the N -particle system $(\mathbf{Z}_t^{i,N})_{i=1}^N$ and a set of N independent mean-field processes with the same Brownian motions and show that their L^2 distance vanishes as $N \rightarrow \infty$. We define N independent processes $(\bar{\mathbf{Z}}_t^{1,N}, \dots, \bar{\mathbf{Z}}_t^{N,N})$ defined as the solutions of N SDEs:

$$d\bar{\mathbf{Z}}_t^{i,N} = \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_t^{i,N}, f_t) dt - \lambda \left(\nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_t^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_t^{i,N}) \right) dt + \Sigma d\mathbf{W}_t^i,$$

for $i \in \{1, \dots, N\}$, where (\mathbf{W}_t^i) is the same Brownian motion as in (7) and $f_t = \text{Law}(\bar{\mathbf{Z}}_t^{i,N})$. We will show that:

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\sup_{t \leq T} \left| \mathbf{Z}_t^{i,N} - \bar{\mathbf{Z}}_t^{i,N} \right|^2 \right] \leq \epsilon(N, T). \quad (21)$$

We fix $i = 1$ and define the path-space coupling

$$\pi_N := \text{Law} \left((\mathbf{Z}_t^{1,N})_{t \in [0, T]}, (\bar{\mathbf{Z}}_t^{1,N})_{t \in [0, T]} \right).$$

Its marginals are $f_{[0, T]}^{1, N}$ and $f_{[0, T]}$, respectively. By the definition of the 2-Wasserstein distance on path-space,

$$\begin{aligned}
W_2^2 \left(f_{[0, T]}^{1, N}, f_{[0, T]} \right) &\leq \int \sup_{0 \leq t \leq T} |\mathbf{x}_t - \mathbf{y}_t|^2 \pi_N(d\mathbf{x}, d\mathbf{y}) \\
&= \mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \mathbf{Z}_t^{1, N} - \bar{\mathbf{Z}}_t^{1, N} \right|^2 \right].
\end{aligned}$$

Since the right-hand side of Equation (21) will tend to 0 as $N \rightarrow \infty$, we can conclude that

$$\lim_{N \rightarrow \infty} W_2 \left(f_{[0, T]}^{1, N}, f_{[0, T]} \right) = 0.$$

Furthermore, by exchangeability, the same estimate holds for any fixed finite collection of particles, which yields $f_{[0, T]}$ -chaoticity of the nudged particle system.

By Ito's formula, and since the stochastic terms cancel due to the synchronous coupling, we have:

$$\begin{aligned}
\left| \mathbf{Z}_t^{i, N} - \bar{\mathbf{Z}}_t^{i, N} \right|^2 &= 2 \int_0^t \left\langle \mathbf{Z}_s^{i, N} - \bar{\mathbf{Z}}_s^{i, N}, \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i, N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i, N}, f_s) - \right. \\
& \quad \left. \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i, N} - \mathbf{Z}_s^{j, N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i, N}) \right) + \right. \\
& \quad \left. \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i, N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i, N}) \right) \right\rangle ds.
\end{aligned}$$

We take the supremum and then the expectation:

$$\begin{aligned}
& \mathbb{E} \left[\sup_{t \leq T} \left| \mathbf{Z}_t^{i,N} - \bar{\mathbf{Z}}_t^{i,N} \right|^2 \right] \\
&= \mathbb{E} \left[\sup_{t \leq T} \left| 2 \int_0^t \left\langle \mathbf{Z}_s^{i,N} - \bar{\mathbf{Z}}_s^{i,N}, \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) - \right. \right. \right. \\
&\quad \left. \left. \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \right. \\
&\quad \left. \left. + \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right\rangle ds \right] \\
&\leq 2 \int_0^T \mathbb{E} \left[\left| \left\langle \mathbf{Z}_s^{i,N} - \bar{\mathbf{Z}}_s^{i,N}, \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) - \right. \right. \right. \\
&\quad \left. \left. \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \right. \\
&\quad \left. \left. + \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right\rangle \right] ds \tag{22} \\
&\leq \int_0^T \mathbb{E} \left[\left| \mathbf{Z}_s^{i,N} - \bar{\mathbf{Z}}_s^{i,N} \right|^2 \right] ds + \int_0^T \mathbb{E} \left[\left| \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) - \right. \right. \\
&\quad \left. \left. \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \right. \\
&\quad \left. \left. + \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] ds \\
&\leq \int_0^T \mathbb{E} \left[\sup_{r \leq s} \left| \mathbf{Z}_r^{i,N} - \bar{\mathbf{Z}}_r^{i,N} \right|^2 \right] ds + \int_0^T \mathbb{E} \left[\left| \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) - \right. \right. \\
&\quad \left. \left. \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \right. \\
&\quad \left. \left. + \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] ds.
\end{aligned}$$

In the second integral, we have that:

$$\begin{aligned}
& \mathbb{E} \left[\left| \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) \right. \right. \\
& \quad \left. \left. - \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) + \right. \right. \\
& \quad \left. \left. \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] \\
& \leq 3 \mathbb{E} \left[\left| \mathbf{b}_{\text{model}}(\mathbf{Z}_s^{i,N}, \nu_s^N) - \mathbf{b}_{\text{model}}(\bar{\mathbf{Z}}_s^{i,N}, f_s) \right|^2 \right] \\
& \quad + 3 \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] + \\
& \quad 3 \mathbb{E} \left[\left| \lambda \left(\nabla K_h * \mu^{\text{obs}}(\mathbf{Z}_s^{i,N}) - \nabla K_h * \mu^{\text{obs}}(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right].
\end{aligned} \tag{23}$$

For the first term on the right-hand side of (23), we define $\bar{\nu}_s^N = \frac{1}{N} \sum_{j=1}^N \delta_{\bar{\mathbf{Z}}_s^{j,N}}$ as the empirical measure of the system $(\bar{\mathbf{Z}}_s^{1,N}, \dots, \bar{\mathbf{Z}}_s^{N,N})$ and use triangle inequality and the Lipschitz condition of $\mathbf{b}_{\text{model}}$ to get: The second term on the right-hand side of (24) becomes: We show that $\mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \right] \rightarrow 0$. By Theorem 3 in [70], the empirical measure $\bar{\nu}_s^N$ converges weakly to f_s . Furthermore, by Theorem 1.7 in [69], f_s has bounded second moment. We have convergence of the second moment by the strong law of large numbers. Hence, by Theorem 6.9 in [71], we have that $W_2(\bar{\nu}_s^N, f_s) \rightarrow 0$. If $(W_2(\bar{\nu}_s^N, f_s))^2$ is also uniformly integrable, then by Vitali Convergence Theorem, $\mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \right] \rightarrow 0$. $(W_2(\bar{\nu}_s^N, f_s))^2$ is uniformly integrable if $\sup_N \mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right] \rightarrow 0$ as $R \rightarrow \infty$. We have

$$\begin{aligned}
(W_2(\bar{\nu}_s^N, f_s))^2 &= \inf_{\gamma \in \Gamma(\bar{\nu}_s^N, f_s)} \int |\mathbf{x} - \mathbf{y}|^2 d\gamma(\mathbf{x}, \mathbf{y}) \\
&\leq 2 \int |\mathbf{x}|^2 d\bar{\nu}_s^N + 2 \int |\mathbf{y}|^2 df_s,
\end{aligned}$$

so that

$$\begin{aligned}
& \sup_N \mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right] \\
& \leq 2 \sup_N \int |\mathbf{x}|^2 d\bar{\nu}_s^N \mathbb{E} \left[\mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right] + 2 \int |\mathbf{y}|^2 df_s \sup_N \mathbb{E} \left[\mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right].
\end{aligned}$$

Since $\bar{\nu}_s^N$ and f_s both have bounded second moments and $\sup_N \mathbb{E} \left[\mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right] \rightarrow 0$ because $W_2(\bar{\nu}_s^N, f_s) \rightarrow 0$, we have that $\sup_N \mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \mathbb{1}_{(W_2(\bar{\nu}_s^N, f_s))^2 \geq R} \right] \rightarrow 0$ as $R \rightarrow \infty$.

Hence, we can conclude by Vitali Convergence Theorem that $\mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \right] \rightarrow 0$.

For the second term on the right-hand side of (23), we use triangle inequality to get:

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{z}_s^{i,N} - \mathbf{z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{z}}_s^{i,N}) \right) \right|^2 \right] \\
\leq & \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{z}_s^{i,N} - \mathbf{z}_s^{j,N}) - \frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \bar{\mathbf{z}}_s^{j,N}) \right) \right|^2 \right] + \\
& \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \bar{\mathbf{z}}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{z}}_s^{i,N}) \right) \right|^2 \right]
\end{aligned} \tag{24}$$

The first term on the right-hand side of (24) is simplified using triangle inequality and the Lipschitz property of $\nabla \tilde{K}_h$:

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{z}_s^{i,N} - \mathbf{z}_s^{j,N}) - \frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \bar{\mathbf{z}}_s^{j,N}) \right) \right|^2 \right] \\
\leq & 2\lambda^2 \mathbb{E} \left[\left| \frac{1}{N} \sum_{j=1}^N \left(\nabla \tilde{K}_h(\mathbf{z}_s^{i,N} - \mathbf{z}_s^{j,N}) - \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \mathbf{z}_s^{j,N}) \right) \right|^2 \right] + \\
& 2\lambda^2 \mathbb{E} \left[\left| \frac{1}{N} \sum_{j=1}^N \left(\nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \mathbf{z}_s^{j,N}) - \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \bar{\mathbf{z}}_s^{j,N}) \right) \right|^2 \right] \\
\leq & 2\lambda^2 \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N \left| \nabla \tilde{K}_h(\mathbf{z}_s^{i,N} - \mathbf{z}_s^{j,N}) - \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \mathbf{z}_s^{j,N}) \right|^2 \right] + \\
& 2\lambda^2 \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N \left| \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \mathbf{z}_s^{j,N}) - \nabla \tilde{K}_h(\bar{\mathbf{z}}_s^{i,N} - \bar{\mathbf{z}}_s^{j,N}) \right|^2 \right] \\
\leq & 2\lambda^2 \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N L^2 |\mathbf{z}_s^{i,N} - \bar{\mathbf{z}}_s^{i,N}|^2 \right] + 2\lambda^2 \mathbb{E} \left[\frac{1}{N} \sum_{j=1}^N L^2 |\mathbf{z}_s^{j,N} - \bar{\mathbf{z}}_s^{j,N}|^2 \right] \\
\leq & 4\lambda^2 L^2 \mathbb{E} \left[|\mathbf{z}_s^{i,N} - \bar{\mathbf{z}}_s^{i,N}|^2 \right].
\end{aligned}$$

The second term on the right-hand side of (24) becomes:

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] \\
&= \frac{\lambda^2}{N^2} \sum_{k,l=1}^N \mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{k,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \\
&\quad \left. \left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{l,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right] \\
&\leq \frac{\lambda^2}{N^2} \sum_{k \neq l} \mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{k,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \\
&\quad \left. \left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{l,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right] + 4 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2 \\
&= \frac{\lambda^2}{N^2} \sum_{k \neq l} \mathbb{E} \left[\mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{k,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right. \right. \\
&\quad \left. \left. \left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{l,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \middle| \bar{\mathbf{Z}}_s^{i,N} \right] \right] + \\
&\quad 4 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2 \\
&= \frac{\lambda^2}{N^2} \sum_{k \neq l} \mathbb{E} \left[\mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{k,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \middle| \bar{\mathbf{Z}}_s^{i,N} \right] \right. \\
&\quad \left. \mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{l,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \middle| \bar{\mathbf{Z}}_s^{i,N} \right] \right] + \\
&\quad 4 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2 \\
&= 4 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2,
\end{aligned}$$

where we used the law of total expectation and the fact that $\bar{\mathbf{Z}}_s^{k,N}, \bar{\mathbf{Z}}_s^{l,N}$, are independent of $\bar{\mathbf{Z}}_s^{i,N}$. The last equality is obtained by observing that at least one of k, l is not equal to i ; without loss of generality, assume $k \neq i$. Then since $\text{Law}(\bar{\mathbf{Z}}_s^{k,N}) = f_t$,

$$\begin{aligned}
& \mathbb{E} \left[\left(\nabla \tilde{K}_h(\bar{\mathbf{Z}}_s^{i,N} - \bar{\mathbf{Z}}_s^{k,N}) \right. \right. \\
&\quad \left. \left. - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \middle| \bar{\mathbf{Z}}_s^{i,N} = \mathbf{z} \right] \\
&= \mathbb{E} \left[\nabla \tilde{K}_h(\mathbf{z} - \bar{\mathbf{Z}}_s^{k,N}) \right] - \nabla \tilde{K}_h * f_t(\mathbf{z}) \\
&= \int \nabla \tilde{K}_h(\mathbf{z} - \mathbf{y}) d f_t(\mathbf{y}) - \int \nabla \tilde{K}_h(\mathbf{z} - \mathbf{y}) d f_t(\mathbf{y}) \\
&= 0
\end{aligned}$$

Hence, Equation (24) simplifies to:

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda \left(\frac{1}{N} \sum_{j=1}^N \nabla \tilde{K}_h(\mathbf{Z}_s^{i,N} - \mathbf{Z}_s^{j,N}) - \nabla \tilde{K}_h * f_t(\bar{\mathbf{Z}}_s^{i,N}) \right) \right|^2 \right] \\
&\leq 4 \lambda^2 L^2 \mathbb{E} \left[|\mathbf{Z}_s^{i,N} - \bar{\mathbf{Z}}_s^{i,N}|^2 \right] + 4 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2.
\end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E} \left[\sup_{t \leq T} \left| \mathbf{Z}_t^{i,N} - \bar{\mathbf{Z}}_t^{i,N} \right|^2 \right] &\leq (1 + 24c^2 + 15\lambda^2 L^2) \int_0^T \mathbb{E} \left[\sup_{t \leq s} \left| \mathbf{Z}_t^{i,N} - \bar{\mathbf{Z}}_t^{i,N} \right|^2 \right] ds \\ &\quad + 6Tc^2 \mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \right] + 12 \frac{\lambda^2}{N} \|\nabla \tilde{K}_h\|_\infty^2. \end{aligned}$$

The conclusion follows by Gronwall's lemma and the fact that $\mathbb{E} \left[(W_2(\bar{\nu}_s^N, f_s))^2 \right] \rightarrow 0$. \square

D Proof of Proposition 3.4

Proof. By Assumption 1,

$$\|\nabla \cdot \mathbf{b}_{\text{model}}(\cdot, \nu_t)\|_{L^\infty(\Omega)} \leq B_{\text{model}}, \quad \forall t \in [0, T].$$

For the nudging term, using the identity

$$\Delta(K_h * f) = (\Delta K_h) * f,$$

we obtain

$$\Delta \left(K_h * (K_h * \nu_t - \mu_t^{\text{obs}}) \right) = (\Delta K_h) * (K_h * \nu_t - \mu_t^{\text{obs}}).$$

Since K_h is Gaussian, Lemma 3.1 implies that K_h is smooth, and in particular $\Delta K_h \in L^1(\Omega)$. By Young's inequality,

$$\left\| \Delta \left(K_h * (K_h * \nu_t - \mu_t^{\text{obs}}) \right) \right\|_{L^\infty(\Omega)} \leq \|\Delta K_h\|_{L^1(\Omega)} \|K_h * \nu_t - \mu_t^{\text{obs}}\|_{L^\infty(\Omega)}.$$

Therefore,

$$\|\nabla \cdot c[\nu](\cdot, t)\|_{L^\infty(\Omega)} \leq C, \quad \forall t \in [0, T].$$

We now rewrite the equation as

$$\partial_t \nu - \nabla \cdot (A \nabla \nu) + c[\nu] \cdot \nabla \nu + (\nabla \cdot c[\nu]) \nu = 0.$$

Define $m(t) := \underline{\nu}_0 e^{-Ct}$. Since m is independent of x , we have $\nabla m = 0$ and $\nabla \cdot (A \nabla m) = 0$, and thus

$$\partial_t m - \nabla \cdot (A \nabla m) + c[\nu] \cdot \nabla m + (\nabla \cdot c[\nu]) m = m'(t) + (\nabla \cdot c[\nu]) m \leq (-C + C)m = 0.$$

Hence m is a subsolution of the same linear parabolic equation satisfied by ν . Since

$$\nu(x, 0) = \nu_0(x) \geq \underline{\nu}_0 = m(0), \quad \forall x \in \Omega,$$

the parabolic comparison principle yields

$$\nu(x, t) \geq m(t) = \underline{\nu}_0 e^{-Ct}, \quad \forall (x, t) \in \Omega \times [0, T].$$

The conclusion follows. \square

E Additional results for the one-dimensional linear benchmark

We include two additional linear benchmark cases to test the behavior of the nudging correction under stronger mean-reversion bias. In both experiments, the reference dynamics use the same coefficient as in the main text, $a_{\text{true}} = 1$, while the biased forecast uses $a = 2$ or $a = 5$. Since these values are larger than the reference value, the forecast model pulls particles too strongly toward the empirical mean and therefore produces an overly concentrated distribution with a smaller variance than the true system.

Figure 14 shows the moderately over-interacting case $a = 2$. The biased forecast remains below the reference variance, reflecting excessive contraction around the mean. The nudging correction increases the variance toward the reference curve, with stronger corrections giving better agreement. The improvement is already visible for intermediate values of λ , while the largest value can produce a short initial overshoot when the correction is applied with too few nudging substeps.

Figure 15 considers a more strongly biased forecast model. In this case, the unassimilated trajectory is substantially over-concentrated, so a larger nudging strength is needed to recover the correct variance level. The qualitative trend is consistent with the $a = 2$ case; increasing λ reduces the variance mismatch, but aggressive nudging can introduce temporary numerical instability. These additional experiments support the robustness of the proposed correction mechanism for both moderate and severe over-interaction bias.

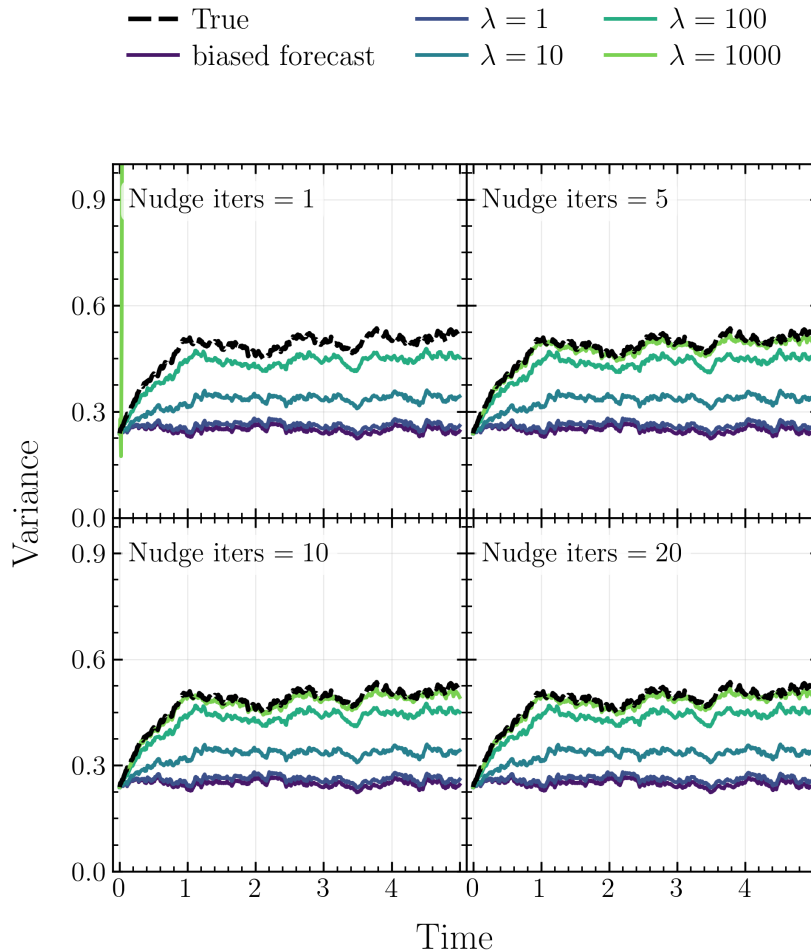


Figure 14: **Variance dynamics in the one-dimensional linear benchmark** ($a = 2$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. Increasing λ generally improves tracking accuracy, but excessively large nudging may reduce numerical stability.

F Additional results for the multimodal benchmark

We include additional experiments for the multimodal benchmark with biased interaction coefficients $a = 0.1$ and $a = 0.5$. These two cases complement the main experiment with $a = 1.5$ and test whether the proposed nudging correction remains effective when the forecast model is only moderately misspecified. The reference system uses $a_{\text{true}} = 0.25$, so the case $a = 0.1$ corresponds to a weaker interaction than the reference dynamics, while $a = 0.5$ corresponds to a stronger interaction.

Figure 16 shows the case where the forecast interaction strength is smaller than the reference value. Since this bias is relatively mild, the unassimilated forecast already follows the reference variance more closely than in the more strongly biased case shown in the main text. Nevertheless, the nudging correction still yields better agreement with the reference trajectory. For moderate and large values of λ , the assimilated variance remains close to the true variance after a short adjustment period. As in the previous experiments, using only one nudging iteration together with a large λ can create visible oscillations, reflecting the stability limitation of an explicit correction step.

Figure 17 considers a forecast model with stronger interaction than the reference system. In this case, the biased model tends to contract the distribution too strongly toward the empirical mean, which leads to a variance mismatch. The nudging term corrects this error by pushing the forecast law toward the observed coarse density. Increasing λ generally reduces the variance gap, while the largest value again requires enough nudging substeps to avoid temporary numerical instability.

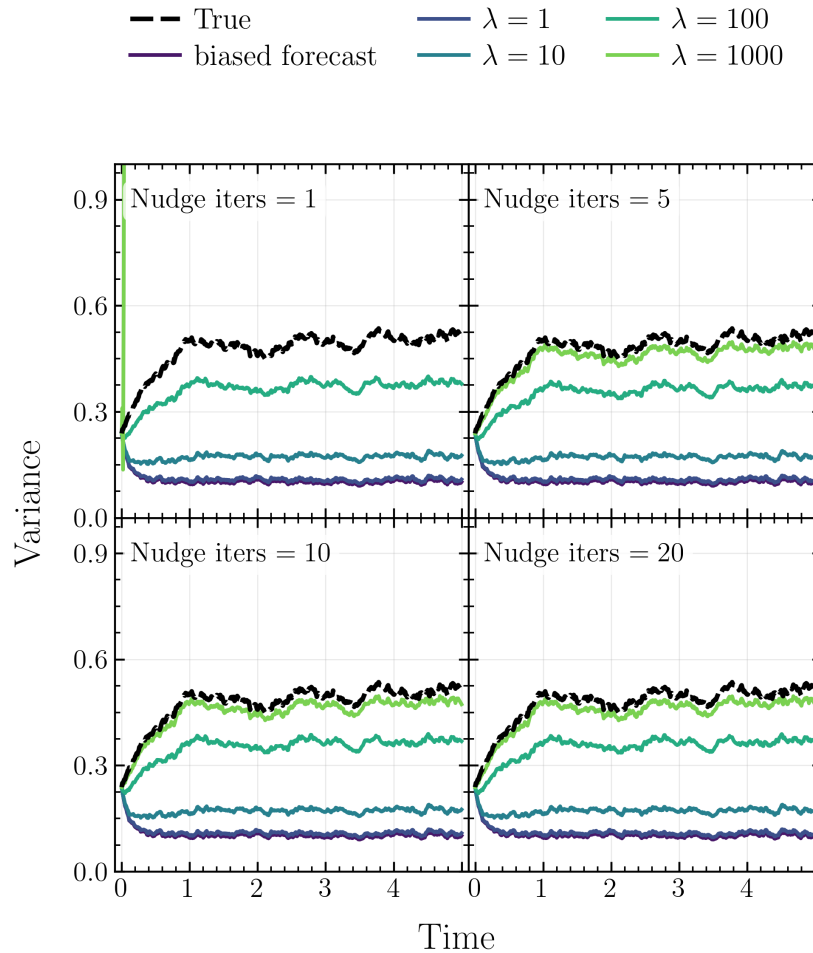


Figure 15: **Variance dynamics in the one-dimensional linear benchmark** ($a = 5$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. Increasing λ generally improves tracking accuracy, but excessively large nudging may reduce numerical stability.

The corresponding space-time density in Figure 18 confirms that the improvement is not limited to the variance. The biased forecast captures the overall bimodal structure but exhibits visible discrepancies in the spread and relative concentration of the density. The assimilated solutions reduce these discrepancies, with the stronger correction giving a density evolution closer to the reference law. This shows that the measure-based feedback can correct the distribution at the level of the full density, not only at the level of a low-order statistic.

Figure 19 shows a similar trend for the stronger-interaction forecast model. The biased prediction has an incorrect density profile, because the interaction term disrupts the balance between concentration near the center and spreading toward the two wells. Nudging with $\lambda = 10$ partially corrects this mismatch, while $\lambda = 1000$ gives a closer reconstruction of the reference space-time density. Together, Figures 16–19 indicate that the proposed correction remains effective for both weaker and stronger interaction bias in the multimodal setting.

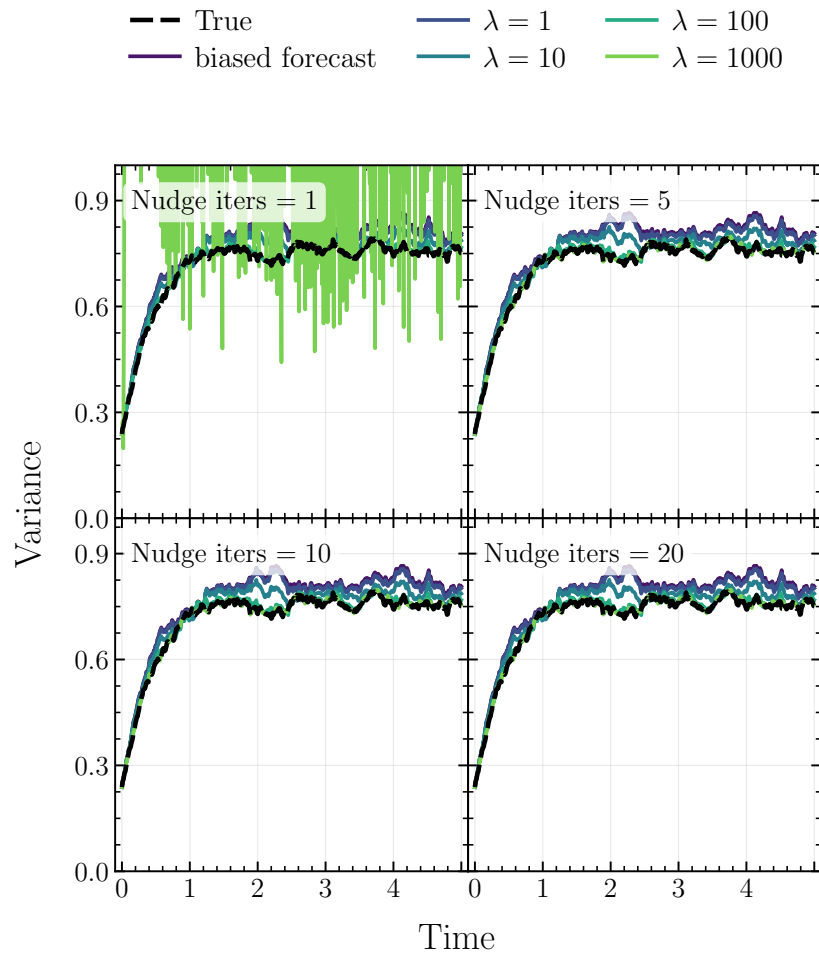


Figure 16: **Variance dynamics in the multimodal benchmark** ($a = 0.1$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. The four panels correspond to different numbers of nudging iterations. Increasing λ improves tracking accuracy, while excessively large nudging can introduce temporary numerical instability when the correction is applied too aggressively.

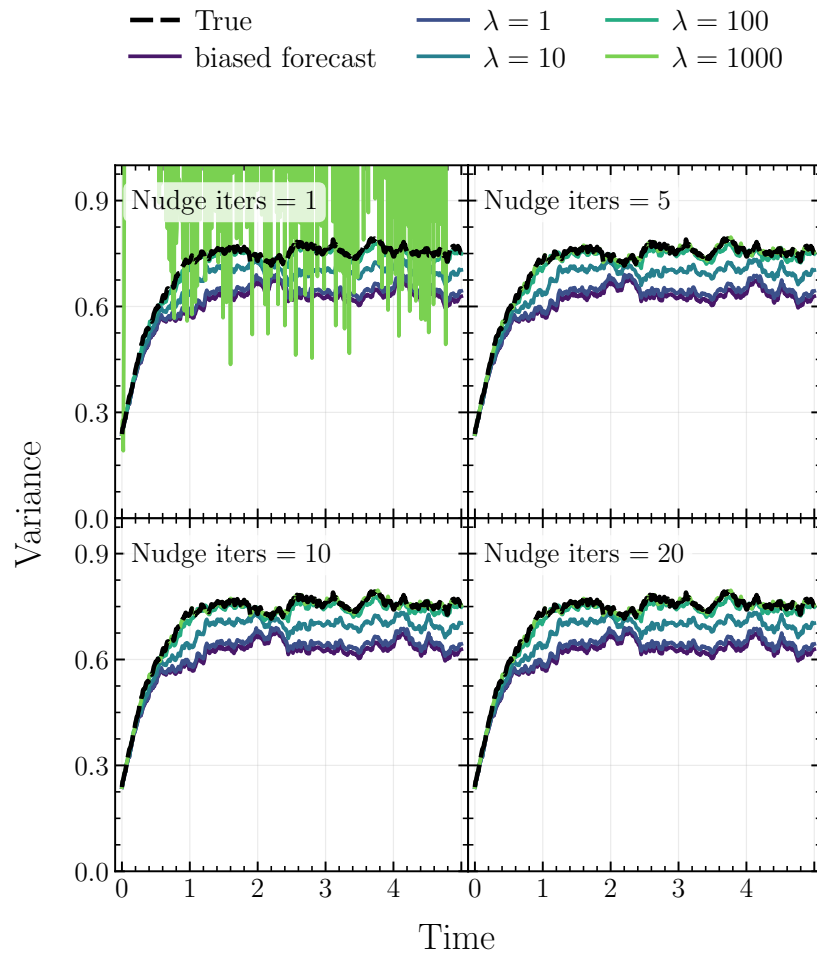


Figure 17: **Variance dynamics in the multimodal benchmark** ($a = 0.5$). We compare the reference system, the biased forecast model, and assimilated (nudged) trajectories with $\lambda \in \{1, 10, 100, 1000\}$. The four panels correspond to different numbers of nudging iterations. Increasing λ improves tracking accuracy, while excessively large nudging can introduce temporary numerical instability when the correction is applied too aggressively.

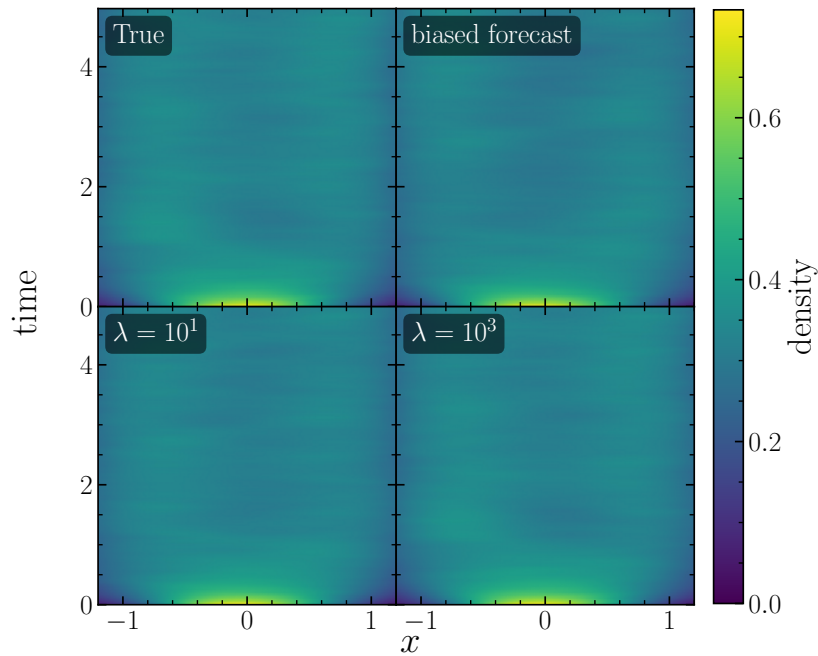


Figure 18: **Space-time density evolution in the multimodal benchmark** ($a = 0.1$). Top-left: reference density. Top-right: biased forecast. Bottom-left: assimilated density with $\lambda = 10$. Bottom-right: assimilated density with $\lambda = 1000$. Larger nudging strength restores both the location and the spread of the true law.

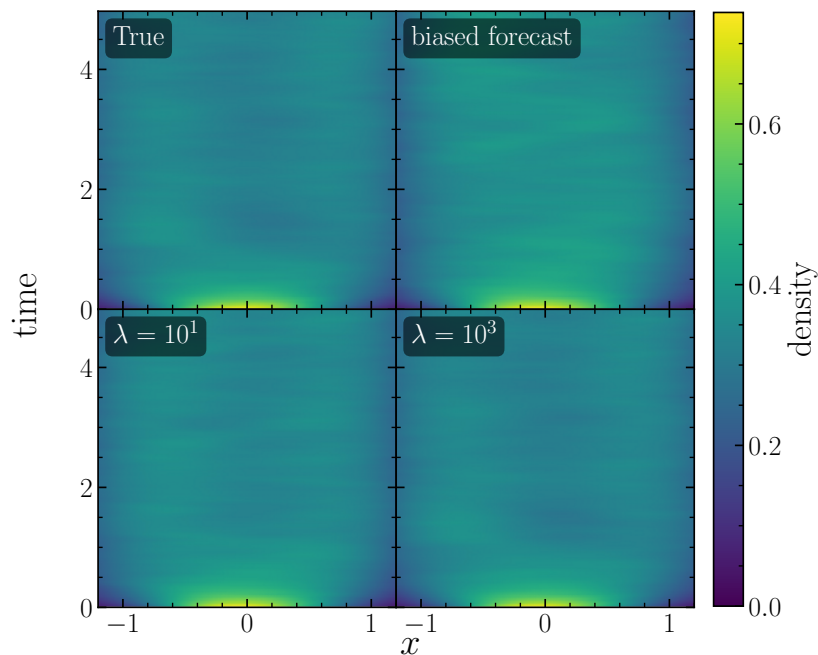


Figure 19: **Space-time density evolution in the multimodal benchmark** ($a = 0.5$). Top-left: reference density. Top-right: biased forecast. Bottom-left: assimilated density with $\lambda = 10$. Bottom-right: assimilated density with $\lambda = 1000$. Larger nudging strength restores both the location and the spread of the true law.