Physica D 484 (2025) 135013

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Coupled stochastic-statistical equations for filtering multiscale turbulent systems

Di Qi [a],*, Jian-Guo Liu [b]

[a] Department of Mathematics, Purdue University, 150 North University Street, West Lafayette, 47906, IN, USA
[b] Department of Mathematics and Department of Physics, Duke University, 285 Physics Building, Durham, 27708, NC, USA

ARTICLE INFO

Communicated by Stefan Wiggins

MSC:
60G25
62M20
93E11
76F55
35Q84
60H15

Keywords:
Nonlinear filtering
Statistical modeling
Multiscale system
Ensemble method

ABSTRACT

We present a new strategy for the statistical forecasts of multiscale nonlinear systems involving non-Gaussian probability distributions with the help of observation data from leading-order moments. A stochastic-statistical modeling framework is designed to enable systematic theoretical analysis and support efficient numerical simulations. The nonlinear coupling structures of the explicit stochastic and statistical equations are exploited to develop a new multiscale filtering system using statistical observation data, which is represented by an infinite-dimensional Kalman–Bucy filter satisfying conditional Gaussian dynamics. To facilitate practical implementation, a finite-dimensional stochastic filtering model is proposed that approximates the intractable infinite-dimensional filter solution. We prove that this approximating filter effectively captures key non-Gaussian features, demonstrating consistent statistics with the optimal filter first in its analysis step update, then at the long-time limit guaranteeing stable convergence to the optimal filter. Finally, we build a practical ensemble filter algorithm based on the stochastic filtering model. Robust performance of the modeling and filtering strategies is demonstrated on prototype models, implying wider applications on challenging problems in statistical prediction and uncertainty quantification of multiscale turbulent states.

1. Introduction

Complex turbulent phenomena characterized by nonlinearly coupled spatiotemporal scales and inherent internal instability are widely observed in science and engineering systems [1–4]. A probabilistic formulation is required to quantify uncertainties in the high-dimensional turbulent states [5–7]. Traditional ensemble approaches using a particle system to approximate the probability evolution quickly become computationally prohibitive since a sufficiently large sample size is necessary to capture the extreme non-Gaussian outliers even for relatively low-dimensional systems [8,9]. As a result, rigorous analysis often becomes intractable and direct numerical simulations are likely to be expensive and inaccurate [10,11].

Filtering strategies [12–15] have long been used for finding the optimal probability estimate of a stochastic state based on partial and noisy observation data. Filtering theories [16–19] and corresponding numerical solutions [20–22] for general nonlinear systems have been investigated through different approaches. In predicting nonlinear turbulent signals, ensemble Kalman filters [23,24] as well as the related particle methods [25–27] provide effective tools for state and parameter estimations. Despite wide applications [28–30], difficulties persist for accurate statistical forecast of turbulent states especially when non-Gaussian features are present in the target probability distribution. Conventional ensemble-based approaches often suffer inherent difficulties in estimating the crucial higher-order moment statistics and maintaining stable prediction with finite number of particles [11,31,32]. Instead of using a single trajectory observation of the stochastic signal, observations of low-order statistics, such as the mean and covariance of the large-scale states, can be obtained to improve the prediction and uncertainty quantification of high-order statistical information in filtering. For example, statistical data can be retrieved from coarse-grained statistical observations from local average in a small spatial neighborhood or temporal average during a short period of time given ergodicity of the stochastic systems [33,34], as well as many low-order strategies adopting observation operators from the equilibrium measure [6,35,36]. Therefore, a promising research direction is to propose new filtering models that have skill to recover crucial high-order moments information in non-Gaussian probability distributions using partial observation data from statistical observations from the leading-order mean and covariance [37].

* Corresponding author.
E-mail addresses: qidi@purdue.edu (D. Qi), jian-guo.liu@duke.edu (J.-G. Liu).

https://doi.org/10.1016/j.physd.2025.135013
Received 16 November 2024; Received in revised form 25 October 2025; Accepted 28 October 2025
Available online 31 October 2025

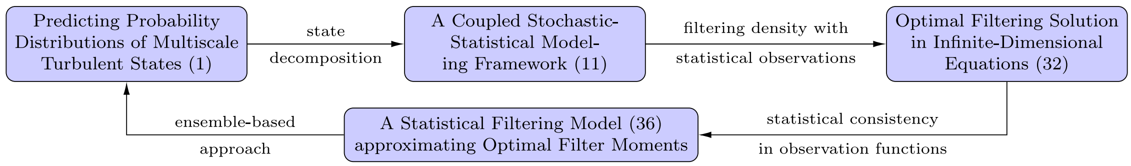

Fig. 1. Diagram illustrating main ideas in constructing the filtering framework.

1.1. General problem setup

We start with a general mathematical formulation [7] modeling a high-dimensional stochastic state $u_t \in \mathbb{R}^d$ involving nonlinear multiscale interactions satisfying the following stochastic differential equation (SDE)

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = \Lambda u_t + B\left(u_t, u_t\right) + F_t + \sigma_t \dot{W}_t. \tag{1}$$

On the right hand side of the above equation, the linear operator, $\Lambda = L - D : \mathbb{R}^d \to \mathbb{R}^d$, represents linear dispersion $L^{\intercal} = -L$ and dissipation $D < 0$ effects. The nonlinear effect in the dynamical system is introduced via a bilinear quadratic operator, $B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$. The system is subject to external forcing effects that are decomposed into a deterministic component, F_t, and a stochastic component represented by a Gaussian white noise $\dot{W}_t \in \mathbb{R}^s$ with coefficient $\sigma_t \in \mathbb{R}^{d\times s}$. The model emphasizes the important role of quadratic interactions through $B(u,u)$. Importantly, the quadratic term is assumed to satisfy an energy conservation law (see next in (3a)) so that the first two moments of u_t remains finite. The structure property in (1) is inherited from a finite-dimensional truncation of the corresponding continuous equation, for example, a spectral projection of the nonlinear advection in the fluid models [38,39]. Many realistic systems with wide applications [40–42] can be categorized in this general dynamical Eq. (1).

The evolution of the model state u_t depends on the sensitivity to the randomness in initial conditions and external stochastic effects, which will be further amplified in time by the inherent internal instability due to the nonlinear coupling term [6]. Then u_t induces a probability measure on $(0,\infty) \times \mathbb{R}^d$. The time evolution of the probability density function (PDF) p_t is governed by the associated Fokker–Planck equation (FPE) starting from an initial distribution $p_{t=0} = p_0$

$$\frac{\partial p_t}{\partial t} = \mathcal{L}_{\mathrm{FP}} p_t := -\nabla_u \cdot \left[\left(\Lambda u + B(u,u) + F_t\right) p_t\right] + \frac{1}{2}\nabla_u \cdot \left[\nabla_u \cdot \left(\sigma_t \sigma_t^{\intercal} p_t\right)\right], \tag{2}$$

where $\mathcal{L}_{\mathrm{FP}}$ represents the Fokker–Planck operator, and $\nabla \cdot (\nabla \cdot A) = \sum_{k,l} \frac{\partial^2 A_{kl}}{\partial u_k \partial u_l}$. The existence and uniqueness of solution p_t to the linear FPE (2) is guaranteed by the uniformly elliptic operator $\mathcal{L}_{\mathrm{FP}}$ [43] and we have that the second moments of the state u_t remain bounded during the time evolution from the finite initial distribution (see Lemma 1 in Section 2.1). However, it remains a challenging task for directly solving the FPE (2) as a high-dimensional PDE. As an alternative approach, ensemble forecast by tracking the Monte-Carlo solutions estimates the essential statistics through empirical averages among a group of samples drawn i.i.d. from the initial distribution $u^{(i)}(0) \sim p_0$ at the starting time $t = 0$. In practice, large errors will still be introduced to the empirical estimations since only a finite sample approximation is available in modeling the non-Gaussian probability distribution and statistics in a high dimensional space.

It is expected that the prediction errors from the finite ensemble estimation of the PDF of stochastic states can be effectively corrected by filtering with the help of the available observation data. In designing new filtering strategies, we propose to use statistical observations from mean and covariance to improve the accuracy and stability in the forecasts of higher-order moments through finite ensemble approximations. However, the general formulation (1) as well as the associated FPE (2) becomes inconvenient to use since all the multiscale stochastic processes are mixed together in the equation as well as all high-order moments of the PDF. The *main goal of this paper* is thus to develop a systematic modeling framework with strategies to accurately capture the (potentially highly non-Gaussian) PDF p_t assisted by statistical measurements in the leading moments.

1.2. Overview of the paper

In this paper, we study nonlinear filtering of the general multiscale turbulent system (1). The new multiscale nonlinear filtering model is constructed under the following step-by-step procedure, which will finally lead a nonlinear ensemble filtering strategy to recover non-Gaussian PDFs:

- First, we propose a coupled stochastic-statistical system (11) demonstrating rigorous statistical consistency with the original system (1): the stochastic dynamics will serve as the signal process in filtering including high-order non-Gaussian features, while the reinforced statistical equations provide the observation process;
- Second, a statistical filtering problem is formulated based on the coupled stochastic-statistical model: optimal filter Eqs. (32) are derived as the precise unbiased least square estimate for the non-Gaussian stochastic state of (11), conditional on the mean and covariance as a natural choice of the observed state;
- Third, a statistical filtering model (36) is developed as an approximation to the optimal filter solution in leading-order statistics: a stochastic McKean–Vlasov equation is adapted from the optimal filter for practical implementation.

The coupled stochastic-statistical model (11) by itself can serve as an effective tool for statistical forecasts and uncertainty quantification [39, 41,44]. Further, combined with the observation data, the resulting filtering McKean–Vlasov SDE (36) is linked to the probability distribution only through the moments, which can be computed directly from the corresponding statistical equations. This enables efficient computational schemes (58) to effectively improve the accuracy and stability in capturing high-order non-Gaussian features based on only observation from the lower moments. The main steps in building effective models for capturing probability distributions is illustrated in Fig. 1.

Still, developing complete and rigorous theories for approximate filtering models involving dominant nonlinear terms remains an inherently challenging task. In this paper, we seek to establish a comprehensive theoretical formulation to help improve the understanding of the complex phenomena arising from nonlinear dynamics, so that practical numerical strategies and applications can follow based on this adaptive general framework. In addition, we present initial results addressing the forecast and analysis steps of filtering separately, demonstrating the potential and validating the predictive capacities within the approximate filtering models: (i) Proposition 3 shows that the highly tractable coupled stochastic-statistical model (11) demonstrates consistent statistics as the original system (1); (ii) the equation for optimal filter solution (32) is given by exploiting the conditional Gaussian structure of the forward equation, and the approximate filter model is found to recover the same key statistics during the analysis step update in Theorem 8; (iii) the long-time convergence in statistics to the optimal filter is demonstrated in Theorem 11 concerning the entire filtering procedure using the statistical filtering model (36). Finally, the effectiveness of the new filtering model is tested on the prototype triad and Lorenz models.

The structure of the paper is organized as follows: In Section 2, we first set up a statistically consistent formulation for the general multiscale system (1) that is suitable for the construction of the statistical filtering models. The main ideas of finding the optimal filter solution and constructing the approximating filter are shown in Section 3. Long-time convergence and stability of the filtering model is then discussed in Section 4. Combining each component of the ideas, the ensemble filtering algorithm is developed in Section 5, followed by preliminary numerical tests in Section 6. A summary of this paper is given in Section 7. More discussions on the filtering formulation and detailed proofs of the results are shown in the Appendices.

2. A statistically consistent modeling framework for multiscale dynamics

We start with a new formulation for the system (1) using an explicit macroscopic and microscopic decomposition of the multiscale state. In particular, we show that the new formulation provides consistent statistics with the original system including higher-order statistics. In addition, the new formulation also enjoys a more tractable dynamical structure to be adapted to the filtering methods.

2.1. The statistical and stochastic equations

In the first place, the well-posedness of the SDE (1) can be established guaranteeing finite statistical moments of u_t with respect to the law p_t from the PDF solution of (2) during the time evolution by the following lemma.

Lemma 1. *Assume that the linear coupling term in the system (1) is negative definite and the quadratic term conserves energy such that*

$$u \cdot \Lambda u < 0, \quad u \cdot B(u,u) = 0, \tag{3a}$$

for all $u \in \mathbb{R}^d$, and the forcing terms are uniformly bounded and satisfy uniformly elliptic condition

$$\left|F_t\right| \leq C, \quad \sigma_t\sigma_t^{\intercal} \geq cI, \tag{3b}$$

with positive constants $C > 0$ and $c > 0$. Then, given finite moments in the initial condition $\int \left(1+|u|^{2k}\right) p_0(u)\,\mathrm{d}u < \infty$ for $k = 1,2,\ldots$, there will exist a unique global probability density function $p_t \in \mathcal{P}\left(\mathbb{R}^d\right) \cap C\left(\mathbb{R}^d\right)$ to the FPE (2) for time $t > 0$, and all moments of the distribution will maintain bounded for all $t > 0$

$$\int_{\mathbb{R}^d} \left(1+|u|^{2k}\right) p_t(u)\,\mathrm{d}u < \infty. \tag{4}$$

The assumptions in (3a) can be implied by the conservation of energy in the nonlinear coupling and the linear term only containing dissipation effect. A statistical energy equation [45] is induced from the conservation law that maintains the finite second moments in the statistical solution. Conditions in (3b) guarantee that the operator is coercive in the FPE. These assumptions are commonly satisfied in many applications of fluid systems considering the physical energy conservation and dissipation laws. Then, (4) implies the existence of a unique probability solution with all finite moments. The proof of this lemma can be found in Appendix B.

In order to identify the detailed multiscale interactions in the general system (1), we decompose the random model state u_t into a statistical mean $\bar{u}_t$ and stochastic fluctuations u_t' in a finite-dimensional representation under an orthonormal basis $\left\{\hat{v}_k\right\}_{k=1}^d$ with $\hat{v}_k \cdot \hat{v}_l = \delta_{kl}$

$$u_t = \bar{u}_t + u_t' = \sum_{k=1}^d \bar{u}_{k,t}\hat{v}_k + \sum_{k=1}^d Z_{k,t}\hat{v}_k, \text{ with } \bar{u}_{k,t} = \hat{v}_k \cdot \bar{u}_t,\ Z_{k,t} = \hat{v}_k \cdot \left(u_t - \bar{u}_t\right). \tag{5}$$

Above, the deterministic $\bar{u}_t \in \mathbb{R}^d$ represents the statistical mean structure (for example, the zonal jets in geophysical turbulence or the coherent radial flow in fusion plasmas), and the stochastic modes $Z_t = \left[Z_{1,t},\ldots,Z_{d,t}\right]^{\intercal} \in \mathbb{R}^d$ are random fluctuations projected on each eigenmode $\hat{v}_k$, whose randomness illustrates the uncertainty in each single scale of u_t'. In particular, we will show that the dynamics of the stochastic modes Z_t contain nonlinear interactions among a large number of multiscale fluctuations, which demonstrate the characterizing feature of strong energy cascades in turbulent systems [6,46,47].

2.1.1. Statistical equations for the macroscopic states

First, we define the leading-order mean and covariance according to the state decomposition (5)

$$\begin{aligned}\bar{u}_t &= \mathbb{E}_{p_t}[u] := \int_{\mathbb{R}^d} u p_t(u)\,\mathrm{d}u,\\ R_{kl,t} &= \mathbb{E}_{p_t}\left[Z_{k,t}Z_{l,t}\right] := \int_{\mathbb{R}^d} \hat{v}_k \cdot \left(u-\bar{u}_t\right)\left(u-\bar{u}_t\right)\cdot \hat{v}_l p_t(u)\,\mathrm{d}u,\ 1 \leq k,l \leq d,\end{aligned} \tag{6}$$

It implies from Lemma 1 that the first two moments in (6) will maintain finite values in time due to the finite total statistical energy conservation, $E = \frac{1}{2}\left(\left|\bar{u}_t\right|^2 + \operatorname{tr} R_t\right) < \infty$. Statistical states of mean $\bar{u}_t$ and covariance R_t represent the *macroscopic* physical quantities that are easiest to achieve from direct measurements. The mean and covariance can be solved by the following statistical equations

$$\frac{\mathrm{d}\bar{u}_{k,t}}{\mathrm{d}t} = \hat{v}_k \cdot \left[\Lambda \bar{u}_t + B\left(\bar{u}_t,\bar{u}_t\right)\right] + \sum_{m,n=1}^d \gamma_{kmn}\mathbb{E}_{p_t}\left[Z_{m,t}Z_{n,t}\right] + \hat{v}_k \cdot F_t, \tag{7a}$$

$$\begin{aligned}\frac{\mathrm{d}R_{kl,t}}{\mathrm{d}t} =& \sum_{m=1}^d \left[L_{km}\left(\bar{u}_t\right)R_{ml,t} + R_{km,t}L_{lm}\left(\bar{u}_t\right)\right] + Q_{t,kl},\\ &+ \sum_{m,n=1}^d \gamma_{kmn}\mathbb{E}_{p_t}\left[Z_{m,t}Z_{n,t}Z_{l,t}\right] + \gamma_{lmn}\mathbb{E}_{p_t}\left[Z_{m,t}Z_{n,t}Z_{k,t}\right].\end{aligned} \tag{7b}$$

The above equations for the mean and covariance can be derived by directly applying Itô's formula to the model states (see Appendix B for the detailed derivation). We define the nonlinear coupling coefficients $\gamma_{kmn} = \hat{v}_k \cdot B\left(\hat{v}_m,\hat{v}_n\right)$, and the white noise coefficient as $\Sigma_t = \left[\left(\hat{v}_1^{\intercal}\sigma_t\right)^{\intercal},\ldots,\left(\hat{v}_d^{\intercal}\sigma_t\right)^{\intercal}\right]^{\intercal}$ with $Q_t = \Sigma_t\Sigma_t^{\intercal} \in \mathbb{R}^{d\times d}$. The operator $L\left(\bar{u}_t\right) \in \mathbb{R}^{d\times d}$ dependent on the statistical mean state $\bar{u}_t$ is defined as

$$L_{kl}(u) = \hat{v}_k \cdot \left[\Lambda\hat{v}_l + B\left(u,\hat{v}_l\right) + B\left(\hat{v}_l,u\right)\right]. \tag{8}$$

Notice that the right hand side of (7b) involves the fluctuation modes Z_t defined from u_t in (5), then the expectations on third moments are taken w.r.t. the PDF p_t. Therefore, the resulting statistical Eqs. (7) are not closed and need to be combined with the FPE (2) to achieve a complete formulation for the leading-order mean and covariance in the nonlinear system.

2.1.2. Stochastic equations for the microscopic processes

Second, we introduce the SDE describing the time evolution of the multiscale stochastic processes Z_t as the *microscopic* state consisting of the many subscale fluctuations

$$\mathrm{d}Z_t = L\left(\bar{u}_t\right)Z_t\mathrm{d}t + \Gamma\left(Z_tZ_t^{\intercal} - R_t\right)\mathrm{d}t + \Sigma_t\mathrm{d}W_t. \tag{9}$$

Above, $L\left(\bar{u}_t\right)$ is the same mean-fluctuation coupling operator defined in (8) involving the statistical mean $\bar{u}_t$. The multiscale feature of the system is also represented by the nonlinear coupling among the different projected modes in (5). We define the quadratic coupling operator $\Gamma : \mathbb{R}^{d\times d} \to \mathbb{R}^d$ as a linear combination of the entries of the input matrix $R \in \mathbb{R}^{d\times d}$ describing the nonlinear coupling involving the covariance R_t

$$\Gamma_k(R) = \sum_{m,n=1}^d \hat{v}_k \cdot B\left(\hat{v}_m,\hat{v}_n\right)R_{mn}. \tag{10}$$

The form of the stochastic dynamics (9) can be found by directly subtracting the mean Eq. (7a) from the original equation for u_t then projecting on the basis $\hat{v}_k$ based on the state decomposition (5). Detailed derivation for (9) and (7a) can be found in Appendix B and in [7]. Similar to the statistical equations (7), the dynamics on the right hand side of (9) is linked to the macroscopic quantities $\bar{u}_t$ and R_t, which in turn requires additional information of p_t from the original model state u_t. This makes the stochastic equation also unclosed requiring additional information from the PDF solution of (2). The derivations of the above statistical and stochastic Eqs. (9) and (7a) from the original general formulation (1) can be found in the proof of Proposition 3.

2.2. A coupled stochastic-statistical model with explicit higher-order feedbacks

Combining the ideas in the stochastic Eq. (9) and the statistical equations (7), we propose a statistically consistent *stochastic-statistical model* based on the following self-consistent coupling of the microscopic stochastic processes Z_t and the macroscopic statistics $\bar{u}, R_t$

$$\begin{aligned} \mathrm{d}Z_t &= L\left(\bar{u}_t\right) Z_t \mathrm{d}t + \Gamma\left(Z_t Z_t^{\intercal} - R_t\right)\mathrm{d}t + \Sigma_t \mathrm{d}W_t, \\ \frac{\mathrm{d}\bar{u}_t}{\mathrm{d}t} &= M\left(\bar{u}_t\right) + Q_m\left(\mathbb{E}\left[Z_t \otimes Z_t\right]\right) + F_t, \\ \frac{\mathrm{d}R_t}{\mathrm{d}t} &= L\left(\bar{u}_t\right) R_t + R_t L\left(\bar{u}_t\right)^{\intercal} + Q_v\left(\mathbb{E}\left[Z_t \otimes Z_t \otimes Z_t\right]\right) + \Sigma_t \Sigma_t^{\intercal}. \end{aligned} \tag{11}$$

Above, the expectations are all w.r.t. the PDF ρ_t of the stochastic states Z_t. In the first moment equation for $\bar{u}_t$, with a bit abuse of notation, we denote $\bar{u}_t = \left[\bar{u}_{1,t}, \ldots, \bar{u}_{d,t}\right]^{\intercal} \in \mathbb{R}^d$ with each component $\bar{u}_{k,t} = \bar{u}_t \cdot \hat{v}_k$, $M = \left[M_1, \ldots, M_d\right]^{\intercal} \in \mathbb{R}^d$ where $M_k\left(\bar{u}_t\right) = \sum_{p,q} \hat{v}_k \cdot \left[\Lambda \hat{v}_p \bar{u}_{p,t} + B\left(\hat{v}_p, \hat{v}_q\right)\bar{u}_{p,t}\bar{u}_{q,t}\right]$ for $1 \le k \le d$ and $F_t = \left[\hat{v}_1 \cdot F_t, \ldots, \hat{v}_d \cdot F_t\right]^{\intercal} \in \mathbb{R}^d$; in the second moment equation for $R_t \in \mathbb{R}^{d\times d}$, the operator $L\left(\bar{u}_t\right) \in \mathbb{R}^{d\times d}$ indicates mean-fluctuation interactions defined in (8); and in the stochastic equation for Z_t, $\Gamma : \mathbb{R}^{d\times d} \to \mathbb{R}^d$ is the quadratic coupling operator defined in (10). The two higher-moment feedbacks for the mean and covariance, Q_m, Q_v, related to the second and third moments of Z_t respectively, are defined as

$$\begin{aligned} Q_{m,k}\left(\mathbb{E}\left[Z_t \otimes Z_t\right]\right) &= \sum_{p,q=1}^{d} \gamma_{kpq} \mathbb{E}\left[Z_{p,t} Z_{q,t}\right], \\ Q_{v,kl}\left(\mathbb{E}\left[Z_t \otimes Z_t \otimes Z_t\right]\right) &= \sum_{p,q=1}^{d} \gamma_{kpq} \mathbb{E}\left[Z_{p,t} Z_{q,t} Z_{l,t}\right] + \gamma_{lpq} \mathbb{E}\left[Z_{p,t} Z_{q,t} Z_{k,t}\right], \end{aligned} \tag{12}$$

for $1 \le k, l \le d$ with coupling coefficients $\gamma_{kpq} = \hat{v}_k \cdot B\left(\hat{v}_p, \hat{v}_q\right)$. Above, Q_m models the feedback in the mean equation due to the second moments $\mathbb{E}\left[Z_t \otimes Z_t\right]$, and $Q_{v,kl}$ is the symmetric feedback in the covariance equation due to all the third moments $\mathbb{E}\left[Z_t \otimes Z_t \otimes Z_t\right]$. Notice that Q_m, Q_v can be both viewed as linear operators w.r.t. ρ_t.

Different from the unclosed stochastic and statistical Eqs. (9) and (7a) inherently dependent on the intractable PDF p_t of the original model state u_t, the new coupled stochastic-statistical model (11) provides a clean self-consistent formulation for tractable theoretical analysis and direct numerical implementations. A new PDF ρ_t of the stochastic process Z_t is introduced to close the system. The statistical states $\bar{u}_t, R_t$ are first treated as new individual processes subject to higher-order moments w.r.t. ρ_t. Then, the microscopic stochastic equation for Z_t models the high-dimensional multiscale process with explicit dependence on the macroscopic states $\bar{u}_t$ and R_t.

In the rest part of this section, we built a precise link between the new model (11) and the coupled Eqs. (9) and (7a) from the original system (1). First, the following lemma provides the self-consistency in the leading moments of the stochastic modes Z_k and statistical states $\bar{u}_t, R_t$ in (11).

Lemma 2. *With consistent initial conditions $\mathbb{E}\left[Z_0\right] = 0$ and $\mathbb{E}\left[Z_0 Z_0^{\intercal}\right] = R_0$, the leading moments of the stochastic modes Z_t of the coupled model (11) satisfy*

$$\mathbb{E}\left[Z_t\right] = 0, \quad \mathbb{E}\left[Z_t Z_t^{\intercal}\right] = R_t, \tag{13}$$

for all $t > 0$ where the expectation is taken w.r.t. the PDF ρ_t of Z_t, and R_t is the solution of the second-order moment equation in (11).

The identities (13) can be found through the direct application of Itô's formula and we put the proof in Appendix B. Lemma 2 demonstrates that the mean zero coefficients Z_t maintain the same covariance with the statistical equation of R_t, while it also contains more information of the higher-order statistics. Furthermore, we show that the coupled stochastic-statistical model generates the same statistical solution as the original system (1). The following proposition describes the statistical consistency between the coupled model (11) and the original system (1).

Proposition 3. *Assume that p_t is the PDF that solves the FPE (2) of the system (1) for u_t, and the solution $\left\{\bar{u}_t, R_t; Z_t\right\}$ of the stochastic-statistical model (11) has the PDF ρ_t for Z_t together with the deterministic solutions for $\bar{u}_t$ and R_t. Then from the same initial conditions, the two models give the same statistical solution, that is, for all $t > 0$*

$$\mathbb{E}_{p_t}\left[u_t\right] = \bar{u}_t, \quad \mathbb{E}_{p_t}\left[u_t' u_t'^{\intercal}\right] = R_t, \tag{14}$$

where $u_t' = u_t - \mathbb{E}_{p_t}\left[u_t\right]$. Furthermore, for any function $\varphi \in C_b^2\left(\mathbb{R}^d\right)$ we have

$$\mathbb{E}_{p_t}\left[\left(1 + \left|u_t'\right|^2\right)\varphi\left(u_t'\right)\right] = \mathbb{E}_{\rho_t}\left[\left(1 + \left|\tilde{u}_t\right|^2\right)\varphi\left(\tilde{u}_t\right)\right], \tag{15}$$

where $\tilde{u}_t = \sum_{k=1}^d Z_{k,t}\hat{v}_k$ is the fluctuation component in the coupled model (11).

Notice that the left hand sides of (14) and (15) consist of the statistics requiring solving the PDF p_t of the original system, while the right hand sides are purely w.r.t. the PDF ρ_t from the coupled model. The proof of Proposition 3 can be found in Appendix B through detailed computation of each moments of (1) compared with that of (11). Together with Lemma 1, (15) shows that all higher-order statistics with each high-order of moments are also preserved in finite amplitude in the new model formulation (11), thus the new PDF solution ρ_t can be effectively used to represent the statistics in the original system under p_t.

Remark. 1. Proposition 3 confirms the direct link between the new stochastic-statistical model (11) to the statistics in the original fully coupled multiscale system. The consistency in the statistical solution guarantees the existence of solution for the proposed new system given the original system has a unique statistical solution that starts from the same initial condition. Still, more care may be required for the ergodic properties of the original SDE (1). In our case of high but fixed dimensional stochastic problems, ergodicity holds as long as there is a quadratic trap potential due to (3a), $\frac{1}{2}\langle u, Du\rangle > 0$, at far field. The existence of equilibrium invariant measures and ergodicity of SPDEs become a more delicate problem. There have been a series of studies demonstrating ergodic behaviors for the Lorenz equation [48], the 2D Navier–Stokes equations with degenerate random forcing [33], and the Rayleigh–Bénard convection with an additive noise [49] as typical examples of the general stochastic model.

2. We can also propose a first-order coupled model involving only the statistical mean equation coupled with the McKean–Vlasov SDE

$$\begin{aligned} \mathrm{d}Z_t &= L\left(\bar{u}_t\right) Z_t \mathrm{d}t + \Gamma\left(Z_t Z_t^{\intercal} - \mathbb{E}\left[Z_t Z_t^{\intercal}\right]\right)\mathrm{d}t + \Sigma_t \mathrm{d}W_t, \\ \frac{\mathrm{d}\bar{u}_t}{\mathrm{d}t} &= M\left(\bar{u}_t\right) + \sum_{p,q} \mathbb{E}\left[Z_{p,t} Z_{q,t}\right] B\left(\hat{v}_p, \hat{v}_q\right) + F_t. \end{aligned} \tag{16}$$

(16) can serve as an intermediate model for uncertainty quantification and filtering schemes. However, the above first-order equations can only rely on the stochastic model to compute the second moments and the dynamics of the SDE for Z_t will directly involve expectation w.r.t. its law ρ_t. Thus this model will often suffer larger numerical errors and instability in practical applications using finite sample sizes [41].

2.3. The stochastic closure equation as a multiscale interacting system

From the stochastic-statistical formulation (11), the SDE for Z_t is given by a stochastic McKean–Vlasov equation depending on its own probability distribution ρ_t. In particular, the resulting McKean–Vlasov SDE can be viewed naturally as the mean-field limit of the ensemble approximation of N individual trajectories

$$\mathrm{d}Z_t^{(i)} = L\left(\bar{u}_t^N\right) Z_t \mathrm{d}t + \Gamma\left(Z_t^{(i)} Z_t^{(i)\intercal} - R_t^N\right) \mathrm{d}t + \Sigma_t \mathrm{d}W_t^{(i)}, \quad i = 1, \ldots, N, \tag{17}$$

where $\left\{W_t^{(i)}\right\}_{i=1}^N$ are independent white noise processes, and the initial samples $\left\{Z_0^{(i)}\right\}_{i=1}^N$ are drawn from the same initial distribution ρ_0 of Z_t. Notice that the ensemble members $Z_t^{(i)}$ as interacting particles are not evolving independently with each other, but are coupled through feedbacks of the leading-order statistics $\bar{u}_t^N$ and R_t^N according to the statistical equations

$$\begin{aligned}\frac{\mathrm{d}\bar{u}_t^N}{\mathrm{d}t} &= M\left(\bar{u}_t^N\right) + Q_m\left(\mathbb{E}^N\left[\mathbf{Z}_t \otimes \mathbf{Z}_t\right]\right) + F_t,\\ \frac{\mathrm{d}R_t^N}{\mathrm{d}t} &= L\left(\bar{u}_t^N\right) R_t^N + R_t^N L^\intercal\left(\bar{u}_t^N\right) + Q_v\left(\mathbb{E}^N\left[\mathbf{Z}_t \otimes \mathbf{Z}_t \otimes \mathbf{Z}_t\right]\right)\\ &\quad + \Sigma_t \Sigma_t^\intercal + \epsilon^{-1}\left(\mathbb{E}^N\left[\mathbf{Z}_t \mathbf{Z}_t^\intercal\right] - R_t^N\right).\end{aligned} \tag{18}$$

The statistical equations (18) involving high-order moments computed directly by the law of the stochastic state Z_t from the ensemble estimate in (17) can be viewed as the *closed model*, and no additional information is needed for the PDF p_t of the original system. The expectations are computed through the empirical average of the interacting particles $\mathbf{Z}_t = \left\{Z_t^{(i)}\right\}_{i=1}^N$ from the ensemble simulation

$$\mathbb{E}^N\left[\varphi\left(\mathbf{Z}_t\right)\right] = \frac{1}{N}\sum_{i=1}^N \varphi\left(Z_t^{(i)}\right). \tag{19}$$

In general, large errors will be introduced by computing the statistics using (18) from a finite ensemble approximation of the stochastic process (17). In the second moment equation for R_t^N, one major difficulty found in many realistic situations for accurate prediction [7,47] is the inherent instability represented by the positive eigenvalues in the coefficient $L\left(\bar{u}_t^N\right)$. This will induce positive growth rate in the unstable modes, while this unbounded growth can be only balanced by the third-moment feedback in Q_v. An additional relaxation term with a parameter $\epsilon > 0$ is thus introduced. This term will not modify the original statistical dynamics given the consistent second-order moment from Proposition 3, and is playing a crucial role as a 'reinforcement' term in maintaining stable performance by introducing an 'effective damping' effect with the negative sign in R_t^N (see [44] for examples in numerical tests) especially with high instability induced by the strong mean-fluctuation coupling from $L\left(\bar{u}_t^N\right)$ while the finite sample approximation becomes not sufficient to balance the strong linear instability.

The coupled ensemble approximation Eqs. (17) and (18) have advantages in practical applications. Unlike the general McKean–Vlasov SDEs [50], (17) avoids the direct inclusion of the PDF of Z_t, which is very difficult to approximate accurately from finite particles. Instead, the mean and covariance equations are used to link the contributions from higher moments, enabling the effective computational algorithm shown in Section 5. Effective computational algorithms with consistent statistics then can be proposed (such as using the efficient random batch methods [41,44]) for the straightforward ensemble model approximation. Besides in practical computation, the relaxation term in R_t^N provides additional restoring forcing as a correction term to numerical errors with finite sample approximation to reinforce stable dynamics and consistent statistics especially in the case where internal instability is involved.

In particular, it is well-known [51] that the empirical measure converges weakly to the true distribution, ρ_t, as well as the leading-order statistics in (17), $\bar{u}^N \to \bar{u}, R_t^N \to R_t$, as $N \to \infty$ under relatively weak assumptions. The dynamical equation for the continuous density function ρ_t of Z_t is given by the corresponding equation

$$\begin{aligned}\frac{\partial \rho_t}{\partial t} &= \mathcal{L}_t^*\left(\bar{u}_t, R_t\right)\rho_t := -\nabla_z \cdot \left[L\left(\bar{u}_t\right) z\rho_t(z) + \Gamma\left(zz^\intercal - R_t\right)\rho_t(z)\right]\\ &\quad + \frac{1}{2}\nabla_z \cdot \left[\nabla_z \cdot \left(\Sigma_t \Sigma_t^\intercal \rho_t(z)\right)\right],\end{aligned} \tag{20}$$

where $\mathcal{L}_t^*$ is the adjoint of the generator $\mathcal{L}_t$ that is also dependent on the law of Z_t shown in the statistics of the mean $\bar{u}_t$ and covariance R_t. In general, the probability density ρ_t will demonstrate non-Gaussian features due to the nonlinear stochastic coupling effects. On the other hand, quantification of non-Gaussian statistics in Z_t relies on the accurate estimation of the leading-order mean and covariance, which can be assisted from the observation data. These desirable structure inspires the construction of effective filtering methods next in Section 3 to include leading-order statistical observations to improve forecast of highly non-Gaussian statistics.

3. Filtering models using observations in mean and covariance

In this section, we propose a new strategy of predicting the probability distribution in the coupled stochastic-statistical model (11) involving highly non-Gaussian statistics. An optimal filter for the ensemble estimate of ρ_t from the coupled equations (17) and (18) is developed by combining the stochastic forecast model describing unobserved microscopic states and leading-order statistics introduced as macroscopic observations. We start with a precise description of the optimal filter equations satisfying a functional conditional Gaussian process, then a new statistical filtering model is proposed approximating the optimal filtering solution showing equivalent statistics.

3.1. Filtering probability distributions using statistical observations

We first formulate the filtering problem for predicting probability distributions based on observations from the leading-order statistics. From the stochastic-statistical Eqs. (11), we can reformulate the general multiscale system (1) for u_t as a composition of the macroscopic state from the first two moments $\bar{u}_t, R_t$ and the microscopic stochastic processes Z_t. In practice, statistical observations are often available through measurements of the macroscopic states. The macroscopic mean state could be achieved from taking the local average in coarse-grained grids and variances estimated by the local fluctuations from the mean state. Another situation includes uncertainty quantification for low-order model predictions. Especially, there has been a large group of data-driven models [52,53] that can be used to produce low-order statistical data while lacking the skill to predict high-order moments. Other reduced-order approaches such as the statistical response theory [47] can also be adopted to exact the low-order statistics in certain problems using the available equilibrium distribution. Therefore, it is natural to incorporate the statistical observation data to improve the estimation of the unobserved microscopic processes, especially to recover the unobserved higher-order statistics (such as the deviation from the normal distribution indicating the occurrence of high impact extreme events).

We start with the target process of the original model state u_t with the associated PDF $p_t \in \mathcal{P}\left(\mathbb{R}^d\right)$ belonging to the space of continuous probability density functions on $\mathbb{R}^d$ with all finite moments (from Lemma 1). Using the statistical consistency in Proposition 3, we can track the dynamical evolution of the equivalent PDF solution ρ_t with the coupled stochastic-statistical Eqs. (11). Furthermore, in practical numerical implementations, the signal process in the forecast model is generated by the ensemble simulation with particles in (17) satisfying the law $Z_t^{(i)} \sim \rho_t^N$. The observation process is generated by the higher-order statistical moments computed from the finite ensemble approximation (18) denoted as $y_t^N = \left(\bar{u}_t^N, R_t^N\right) \in \mathbb{R}^p$. This leads to the following *infinite-dimensional filtering system with statistical observations*

$$\mathrm{d}\rho_t^N = \mathcal{L}_t^*\left(y_t^N\right)\rho_t^N \mathrm{d}t, \qquad \rho_{t=0}^N \sim \mu_0, \tag{21a}$$

$$\mathrm{d}y_t^N = \left[\mathcal{H}\rho_t^N + h_t\left(y_t^N\right)\right]\mathrm{d}t + \Gamma_t \mathrm{d}B_t, \quad y_{t=0}^N = y_0, \tag{21b}$$

where μ_0 is a probability measure of the $\mathcal{P}\left(\mathbb{R}^d\right)$-valued random field. Above, $\mathcal{L}_t$ is the infinitesimal generator of the corresponding SDE for $Z_t^{(i)}$ given by the explicit form in (20); the general observation process y_t^N satisfies the dynamical equation subject to a linear observation operator $\mathcal{H}: \mathcal{P}\left(\mathbb{R}^d\right) \to \mathbb{R}^p$ (with the explicit forms shown next in (24)) acting on the continuous PDF ρ_t^N, as well as the deterministic function $h_t: \mathbb{R}^p \to \mathbb{R}^p$. The additional noise term $\Gamma_t \mathrm{d}B_t$ is introduced to account for the errors from finite ensemble approximation (further explanation will be given next in the explicit model (23)). Notice here ρ_t^N becomes a random field due to the randomness in the finite ensemble approximation for the dynamics of y_t^N as well as its initial uncertainty. Given the observation process $y_s^N, s \le t$, ρ_t^N satisfies a conditional linear dynamics. Thus we can solve the solution for $\rho_t^N \in \mathcal{P}\left(\mathbb{R}^d\right)$ satisfying a conditional Gaussian probability measure. Next, we give a detailed formulation for the observation data y_t as in (23) based on the first two moments.

Remark. In practice, it is common that only partial observation data in the mean and covariance are available for the filtering update. In this case, the forecast Eq. (21a) can be further decomposed into the part only dependent on the observation process, and the residual process containing all the unobserved states (possibly also using proper closure models). This introduces an imperfect model approximation that allows models errors and maintains the conditional Gaussian dynamics [54].

3.1.1. Statistical observations from leading-order moments

Let p_t be the (unknown) PDF of the state u_t in (1), that is, the deterministic solution of the FPE (2). Then, we can assume that observations are drawn from the mean and covariance of the state u_t as

$$\bar{u}_{k,t} = \mathbb{E}_{p_t}\left[u_t \cdot \hat{v}_k\right], \quad R_{kl,t} = \mathbb{E}_{p_t}\left[\left(\hat{v}_k \cdot u_t'\right)\left(u_t' \cdot \hat{v}_l\right)\right], \tag{22}$$

projected to the observed large-scale modes $\hat{v}_k, k \le d'$ in (5). We refer it as the full observation case with $d' = d$, and partial observation case with $d' < d$. For simplicity, we may always consider the full observation case $d' = d$ (that is, $y_t = \left(\bar{u}_t, R_t\right) \in \mathbb{R}^p$ with $p = d + d^2$) in this paper without confusion. According to Proposition 3, the statistical equations for $\bar{u}_t, R_t$ in (11) provide statistical solutions consistent with the law p_t of the state u_t. Thus, the dynamical equations for $\bar{u}_t, R_t$ can be introduced according to the coupled model (11).

On the other hand, model errors always exist in the numerical schemes due to the finite ensemble approximation in the practical implementation model (17) and (18). The detailed equations for the observation process (21b) can be rewritten according to the finite particle estimate (18) as the following SDEs for $\bar{u}_t^N \in \mathbb{R}^d$ and $R_t^N \in \mathbb{R}^{d^2}$

$$\begin{aligned} \mathrm{d}\bar{u}_t^N &= \left[\mathcal{H}_m \rho_t^N + h_{m,t}\left(\bar{u}_t^N\right)\right]\mathrm{d}t + \Gamma_m \mathrm{d}B_{m,t}, \\ \mathrm{d}R_t^N &= \left[\mathcal{H}_v \rho_t^N + h_{v,t}\left(\bar{u}_t^N, R_t^N\right)\right]\mathrm{d}t + \Gamma_v \mathrm{d}B_{v,t}. \end{aligned} \tag{23}$$

where $h_{m,t}(\bar{u}) = M(\bar{u}) + F_t$ and $h_{v,t}(\bar{u}, R) = L(\bar{u})R + RL(\bar{u})^\intercal + \Sigma_t \Sigma_t^\intercal$ are deterministic functions, while the linear *observation operators*, $\mathcal{H}_m, \mathcal{H}_v$, are defined by the high-order statistical feedback functions (12)

$$\begin{aligned} \mathcal{H}_{m,k}\rho &= \int_{\mathbb{R}^d} H_k^m(z)\rho(z)\,\mathrm{d}z, \quad H_k^m = \sum_{p,q=1}^d \gamma_{kpq} z_p z_q, \\ \mathcal{H}_{v,kl}\rho &= \int_{\mathbb{R}^d} H_{kl}^v(z)\rho(z)\,\mathrm{d}z, \quad H_{kl}^v = \sum_{p,q=1}^d \left(\gamma_{kpq} z_p z_q z_l + \gamma_{lpq} z_p z_q z_k\right). \end{aligned} \tag{24}$$

Importantly, additional correction terms, modeled by independent white noises, $B_{m,t}$ and $B_{v,t}$, are added to the statistical equations accounting for errors from the finite ensemble approximations. In fact, the empirical averages in the mean and covariance equations (18) can be both decomposed into the expectation w.r.t. the continuous ρ^N and the additional noises are used to represent the fluctuating errors from the finite N sample estimation, that is, $\mathbb{E}^N H(Z)\,\mathrm{d}t \approx \mathbb{E}H(Z)\,\mathrm{d}t + \Gamma \mathrm{d}B$. The left hand side of the identity contains randomness from the empirical average (19), while the right hand side uses the white noise to model the uncertainty according to the central limit theorem. Therefore, the observation data in (22) can be viewed as a special realization of the observation processes with noises (23).

We assume that $\rho^N \in \mathcal{P}\left(\mathbb{R}^d\right)$ is the probability density with finite moments in all orders, thus we have $\mathcal{H}_{m,k}\rho^N < \infty$ and $\mathcal{H}_{v,kl}\rho^N < \infty$ for all modes k, l. Then (23) fits into the general observation equation (21b) by setting $y_t^N = \left(\bar{u}_t^N, R_t^N\right) \in \mathbb{R}^p$ with $p = d + d^2$ in a column vector, and letting $\mathcal{H} = \left(\mathcal{H}_m, \mathcal{H}_v\right)^\intercal$, $h_t = \left(h_{m,t}, h_{v,t}\right)^\intercal$, and $\Gamma_t = \operatorname{diag}\left(\Gamma_m, \Gamma_v\right)$. With this explicit setup of the signal and observation processes for filtering, we will consider the optimal filtering solution for the probability density ρ_t of Z_t based on the statistical observation data $Y_t = \left\{\left(\bar{u}_s, R_s\right), s \le t\right\}$ we abuse the notation a bit by neglecting the superscript N in the following sections for the theoretical development.

3.2. Optimal filter with conditional Gaussian structure

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the complete probability space, and denote $\mathcal{P}_2\left(\mathbb{R}^d\right)$ as the space of probability density functions with bounded second moments. We first define the $\mathcal{P}_2\left(\mathbb{R}^d\right)$-valued stochastic process ρ_t (denoting $\omega \in \Omega$ as the random event and will be dropped in the following notation) as

$$\rho_t: \mathbb{R}^d \times \Omega \to \mathbb{R}^+, \ (z, \omega) \mapsto \rho_t(z; \omega), \text{ with } \rho_t(\cdot; \omega) \in \mathcal{P}_2\left(\mathbb{R}^d\right), \tag{25}$$

which is thereafter referred to as a *random field*. In contrast to the standard filtering problem concerning the nonlinear SDE of the random model states Z_t, for derivation purpose of the exact optimal equations, we lift the problem into filtering the random field ρ_t based on the observation information $y_s, s \le t$ as in (21). A stochastic model (36) on $\mathbb{R}^d$ will follow for practical implementations next in Section 3.4. Let $\mathcal{G}_t = \sigma\left\{\omega : y_s, s \le t\right\}$ be the σ-algebra generated by the observations. We define the space $\mathcal{V}_t$ as the collection of $\mathcal{G}_t$-measurable square-integrable random fields

$$\mathcal{V}_t := L^2\left(\Omega, \mathcal{G}_t, \mathbb{P}; \mathcal{P}_2\left(\mathbb{R}^d\right) \cap L^2\left(\mathbb{R}^d\right)\right), \tag{26}$$

satisfying $\int \|v(\cdot;\omega)\|_{L^2(\mathbb{R}^d)}^2 \mathrm{d}\mathbb{P}(\omega) < \infty$ and $v(\cdot;\omega) \in \mathcal{P}_2\left(\mathbb{R}^d\right) \cap L^2\left(\mathbb{R}^d\right)$ for $v \in \mathcal{V}_t$. In this infinite-dimensional filtering problem, we aim to find the optimal approximation of ρ_t in the space $\mathcal{V}_t$. The *optimal filtering solution* $\hat{\rho}_t$ is then introduced as the least-square estimate with the minimum variance as

$$\hat{\rho}_t := \underset{v \in \mathcal{V}_t}{\arg\min}\, \mathbb{E}\left[\left\|\rho_t - v\right\|_{L^2(\mathbb{R}^d)}^2\right] = \mathsf{P}_{\mathcal{V}_t}\left[\rho_t\right], \tag{27}$$

where the optimal solution $\hat{\rho}_t$ can be viewed as the unbiased projection of ρ_t onto the space $\mathcal{V}_t$. (27) indicates that $\hat{\rho}_t$ gives the estimation closest to the true distribution ρ_t in the mean square sense in agreement with the observations $\mathcal{G}_t$.

Accordingly, we define the *optimal filter distribution* $\mu_t : \mathcal{P}_2\left(\mathbb{R}^d\right) \times \Omega \to [0,1]$ as the regular conditional measure of the stochastic process ρ_t given $\mathcal{G}_t$. That is, for any Borel set $A \in \mathcal{B}\left(\mathcal{P}_2\left(\mathbb{R}^d\right)\right)$, μ_t is given by the conditional probability of ρ_t given $\mathcal{G}_t$ such that

$$\mu_t(A; \cdot) := \mathbb{P}\left(\rho_t \in A \mid \mathcal{G}_t\right), \quad \mathbb{P} - \text{a.s.} \tag{28}$$

Notice that $\mu_t(A; \cdot) \in \mathcal{G}_t$ is still a stochastic process. For any functional $F \in C\left(\mathcal{P}_2\left(\mathbb{R}^d\right)\right)$ and $t > 0$, we can introduce the conditional expectation w.r.t. the measure μ_t given $\mathcal{G}_t$ as

$$\mathbb{E}\left[F\left(\rho_t\right) \mid \mathcal{G}_t\right] := \int_{\mathcal{P}(\mathbb{R}^d)} F(\rho)\,\mu_t(\mathrm{d}\rho).$$

Therefore, the optimal filter solution (27) is the following random field

$$\hat{\rho}_t = \mathbb{E}\left[\rho_t \mid \mathcal{G}_t\right] \tag{29}$$

given by the conditional expectation of ρ_t w.r.t. μ_t. Furthermore, for any linear operator $\mathcal{M}: L^2(\mathbb{R}^d) \to \mathbb{R}^p$, we have $\mathcal{M}\hat{\rho}_t = \mathbb{E}\left[\mathcal{M}\rho_t \mid \mathcal{G}_t\right]$. And the second moment of $\mathcal{M}\rho_t$ is given by

$$\mathbb{E}\left[\left(\mathcal{M}\rho_t - \mathcal{M}\hat{\rho}_t\right)\left(\mathcal{M}\rho_t - \mathcal{M}\hat{\rho}_t\right)^{\intercal} \mid \mathcal{G}_t\right] = \mathcal{M}\hat{C}_t\mathcal{M}^*,$$

where $\hat{C}_t(\omega): L^2(\mathbb{R}^d;\mathbb{R}^p) \to L^2(\mathbb{R}^d;\mathbb{R}^p)$ with $\hat{C}_t^* = \hat{C}_t$ is the self-adjoint covariance operator for any $f \in L^2(\mathbb{R}^d;\mathbb{R}^p)$, such that

$$\hat{C}_t f = \mathbb{E}\left[\left(\rho - \hat{\rho}_t\right)\int_{\mathbb{R}^d}\left(\rho - \hat{\rho}_t\right)(z) f(z)\,\mathrm{d}z \mid \mathcal{G}_t\right]. \tag{30}$$

Notice again $\hat{C}_t f(z;\cdot)$ is also a random field conditional on $\mathcal{G}_t$. For clarification of notations, we will call $\hat{\rho}_t$ and $\hat{C}_t$ the optimal filter solution and its covariance, and μ_t the optimal filter distribution in the rest part of the paper.

We can characterize the optimal filter solution $\hat{\rho}_t$ as the best estimate in each order of moments. The following result describes the accuracy of the filter approximations in any finite-dimensional projections with the proof in Appendix B.

Proposition 4. *Let ρ_t be the random field from the system (21) and $\hat{\rho}_t = \mathbb{E}\left[\rho_t \mid \mathcal{G}_t\right]$ the optimal filter solution given the observations $\mathcal{G}_t$. For any linear operator $\mathcal{M}: \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^p$ defined by $\mathcal{M}\rho = \int M(z)\rho(\mathrm{d}z)$ and $M \in C(\mathbb{R}^d;\mathbb{R}^p)$, $\mathcal{M}\hat{\rho}_t = \mathbb{E}\left[\mathcal{M}\rho_t \mid \mathcal{G}_t\right]$ gives the best unbiased estimate of $\mathcal{M}\rho_t$ in the sense of minimum mean square error, that is,*

$$\mathbb{E}\left[\left|\mathcal{M}\rho_t - \mathcal{M}\hat{\rho}_t\right|^2\right] = \min_{v\in\mathcal{V}_t}\mathbb{E}\left[\left|\mathcal{M}\rho_t - \mathcal{M}v\right|^2\right], \quad \text{with} \quad \mathbb{E}\left[\mathcal{M}\hat{\rho}_t\right] = \mathbb{E}\left[\mathcal{M}\rho_t\right]. \tag{31}$$

By taking the operator $\mathcal{M}$ as the expectation on $M\left(Z_t\right) = \left|Z_t\right|^m$ with any integer m, $\mathcal{M}\rho_t$ and $\mathcal{M}\hat{\rho}_t$ give the mth order moments of Z_t under the random field ρ_t in (25) and the optimal filter approximation $\hat{\rho}_t$ respectively. A direct implication from (31) shows that we have the unbiased statistics in all finite-dimensional moments $\mathbb{E}\left[\mathcal{M}\hat{\rho}_t\right] = \mathbb{E}\left[\mathcal{M}\rho_t\right]$ with the minimum error $\mathbb{E}\left[\left|\mathcal{M}\rho_t - \mathcal{M}\hat{\rho}_t\right|^2\right]$ from the infinite-dimensional optimal filter solution.

Importantly, the model equations (21) satisfy the desirable conditional Gaussian process [55], that is, given the observations of $Y_t = \left\{y_s = \left(\bar{u}_s, R_s\right), s \le t\right\}$ and Gaussian initial distribution for ρ_0, the random field ρ_t follows a Gaussian distribution at each time t. Let ρ_t be the signal state satisfying linear dynamics (21a), and y_t the observed statistical process subject to linear observation operators in (21b). The optimal filter distribution μ_t (28) conditional on Y_t then becomes an infinite-dimensional Gaussian distribution, $\mu_t = \mathcal{N}\left(\hat{\rho}_t, \hat{C}_t\right)$, where the mean $\hat{\rho}_t$ and covariance $\hat{C}_t$ give the solution to (29) and (30) respectively. Therefore, the equations for the mean and covariance are given by the generalized version of Kalman–Bucy (KB) filter [56] for the infinite-dimensional conditional Gaussian process

$$\begin{aligned}
\mathrm{d}\hat{\rho}_t &= \mathcal{L}_t^*\left(\bar{u}_t, R_t\right)\hat{\rho}_t\mathrm{d}t + \hat{C}_t\mathcal{H}_m^*\Gamma_m^{-2}\left\{\mathrm{d}\bar{u}_t - \left[\mathcal{H}_m\hat{\rho}_t + h_{m,t}\left(\bar{u}_t\right)\right]\mathrm{d}t\right\}\\
&\quad + \hat{C}_t\mathcal{H}_v^*\Gamma_v^{-2}\left\{\mathrm{d}R_t - \left[\mathcal{H}_v\hat{\rho}_t + h_{v,t}\left(\bar{u}, R\right)\right]\mathrm{d}t\right\},\\
\mathrm{d}\hat{C}_t &= \left[\mathcal{L}_t^*\left(\bar{u}_t, R_t\right)\hat{C}_t + \hat{C}_t\mathcal{L}_t\left(\bar{u}_t, R_t\right)\right]\mathrm{d}t - \hat{C}_t\left(\mathcal{H}_m^*\Gamma_m^{-2}\mathcal{H}_m + \mathcal{H}_v^*\Gamma_v^{-2}\mathcal{H}_v\right)\hat{C}_t\mathrm{d}t.
\end{aligned} \tag{32}$$

The existence and uniqueness of solution to the equations of the conditional Gaussian processes (32) are shown in Chapter 12 of [55] for finite-dimensional systems. The results are then generalized to infinite-dimensional Hilbert space [57,58] (see a summary of the results in Appendix A). The system (32) gives a closed set of coupled SPDEs (due to the randomness in $\bar{u}_t, R_t$) enabling more detailed analysis and development of practical methods for computing the optimal solution.

Remark. A similar filtering problem using statistical observations is introduced and analyzed as the ensemble Fokker–Planck filter by [37]. Inspired by the idea, we propose the filtering equations (32) for the more general nonlinearly coupled conditional processes. Very different from the linear setup in [37], we propose a new nonlinear filtering model that incorporates the general stochastic-statistical modeling framework (11) that is suitable for effective statistical forecasts of non-Gaussian statistics.

3.3. A surrogate filtering model for approximating the optimal filter solution

The resulting optimal filtering problem from (21) requires solving the infinite-dimensional system (32) concerning the function $\hat{\rho}_t$ and the operator $\hat{C}_t$. It becomes intractable in finding such infinite-dimensional solutions from direct methods. In developing practical strategies to realize the optimal filter solution, it is more useful to find a surrogate model for the stochastic process $\tilde{Z}_t$, based on which effective ensemble-based approaches can be built. Therefore, we aim to construct an approximating filter from designing a new dynamical equation for $\tilde{Z}_t$, whose PDF $\tilde{\rho}_t$ constrained in the probability space $\mathcal{P}_\infty\left(\mathbb{R}^d\right)$ with all finite moments to effectively represent that of the optimal filter solution $\hat{\rho}_t$.

3.3.1. Filtering updating cycle in a split two-step procedure

For a clear characterization of the filtering process, we follow the general procedure in [27] to first describe the filtering process by concatenated iterations of transporting maps on the corresponding probability distribution. We propose a new stochastic process $\tilde{Z}_t$, whose law $\tilde{\rho}_t \in \mathcal{V}_t$ is a $\mathcal{P}_\infty\left(\mathbb{R}^d\right)$-valued random field dependent on the same statistical observation Y_t as in the optimal filter satisfying (32). Thus, the filtering updating cycle during the time interval $[t, t+\tau]$ can be characterized by the transport of the probability density $\tilde{\rho}_t$ of $\tilde{Z}_t$ in a split two-step procedure.

First, the *forecast step* can be viewed as the push-forward operator acting on the probability density at time instant t with time step $\tau > 0$

$$\tilde{\rho}_t \to \tilde{\rho}_{t+\tau}^- := \mathcal{F}_t^\tau\tilde{\rho}_t = e^{\int_t^{t+\tau}\mathcal{L}_s^*\left(y_s\right)\mathrm{d}s}\tilde{\rho}_t, \tag{33}$$

where $\mathcal{F}_t^\tau$ represents the forecast updating operator with forward time step τ, and $\mathcal{L}_t\left(y_t\right)$ is the same generator as in (20). Second, the *analysis step* updates the prior distribution $\tilde{\rho}^-$ to the posterior distribution $\tilde{\rho}^+$ by incorporating the observation data up to $Y_{t+\tau} = \left\{y_s, s \le t+\tau\right\}$, that is

$$\tilde{\rho}_{t+\tau}^- \to \tilde{\rho}_{t+\tau}^+ := \mathcal{A}_t^\tau\left(\tilde{\rho}_{t+\tau}^-; Y_{t+\tau}\right), \tag{34}$$

where $\mathcal{A}_t^\tau$ represents the analysis updating operator. Therefore, the full filtering cycle from t to $t+\tau$ can be summarized as the composition of the two maps

$$\tilde{\rho}_{t+\tau} = \mathcal{A}_t^\tau\left(\mathcal{F}_t^\tau\tilde{\rho}_t; Y_{t+\tau}\right). \tag{35}$$

Notice that $\mathcal{F}_t^\tau$ is a linear operator on $\tilde{\rho}_t$, while $\mathcal{A}_t^\tau$ could contain nonlinear actions due to the normalization of the probability distribution. The continuous equation for $\partial_t\tilde{\rho}_t = \lim_{\tau\to 0}\frac{1}{\tau}\left(\tilde{\rho}_{t+\tau} - \tilde{\rho}_t\right)$ is then achieved by letting the discrete time step $\tau \to 0$. Next, we first propose the general new filtering model for $\tilde{Z}_t \sim \tilde{\rho}_t$ as a combination of the above two-step procedure, then detailed analysis can be done according to the design of the forecast and analysis step operators $\mathcal{F}_t^\tau$ and $\mathcal{A}_t^\tau$ accordingly.

3.3.2. Construction of the statistically equivalent approximating filter

For simplicity of notations, we still use the general statistical observation processes (23) for $y_t = \left(\bar{u}_t, R_t\right)$ taking the compact formulation

$$\mathrm{d}y_t = \left[\mathcal{H}\rho_t + h_t\left(y_t\right)\right]\mathrm{d}t + \Gamma_t\mathrm{d}B_t,$$

where the general observation operator (24), $\mathcal{H}\rho_t = \int H(z)\rho_t(z)\,\mathrm{d}z$, is defined with the general observation function $H \in C\left(\mathbb{R}^d;\mathbb{R}^p\right)$ acting on the density function ρ_t. Following the general construction in [17,19], we seek the approximating filtering model adopting the following McKean–Vlasov representation with functionals a_t, K_t

$$\begin{aligned}\mathrm{d}\tilde{Z}_t &= L\left(\bar{u}_t\right)\tilde{Z}_t\mathrm{d}t + \Gamma\left(\tilde{Z}_t\tilde{Z}_t^{\intercal} - R_t\right)\mathrm{d}t + \Sigma_t\mathrm{d}\tilde{W}_t\\ &+ a_t\left(\tilde{Z}_t;\tilde{\rho}_t\right)\mathrm{d}t + K_t\left(\tilde{Z}_t;\tilde{\rho}_t\right)\left\{\mathrm{d}y_t - \left[H\left(\tilde{Z}_t\right) + h_t\left(y_t\right)\right]\mathrm{d}t - \Gamma_t\mathrm{d}\tilde{B}_t\right\},\end{aligned} \tag{36}$$

where $\tilde{W}_t, \tilde{B}_t$ are white noise processes independent of W_t, B_t. We use the name 'statistical filtering' to refer the above new approximating filtering model emphasizing our main goal of filtering statistical moments different from the common filtering case. The first row of the above equation models the forecast step of the filtering process, while the second row is the analysis step. The forecast step accepts the same dynamical model of (21a) dependent on the mean and covariance $\left(\bar{u}_t, R_t\right)$. On the other hand, the analysis step serves as an additional correction as an control over statistical observations y_t. New functionals known as the drift $a_t : \mathbb{R}^d \times \mathcal{P}_\infty\left(\mathbb{R}^d\right) \to \mathbb{R}^d$ and the gain operator $K_t : \mathbb{R}^d \times \mathcal{P}_\infty\left(\mathbb{R}^d\right) \to \mathbb{R}^{d\times p}$ are introduced, resulting in an approximating filtering model about the process $\tilde{Z}_t$. Most importantly, as will be shown next in Theorem 8 under proper condition, a_t and K_t are only implicitly dependent on the PDF $\tilde{\rho}_t$ through its leading moments, without the need to compute the (potentially highly non-Gaussian) density function $\tilde{\rho}_t$ explicitly.

In a more clear identification of the filtering updates involving several levels of approximations, we take the split-step strategy to analyze the coupled forecast step and analysis step of the filtering equation (36) separately. In particular, the forecast step in the first row of (36) is given by the exactly same form as the stochastic-statistical equations (11) developed in Section 2. Thus in practice, the updating step with the forecast operator can be implemented adopting the efficient uncertainty prediction methods such as [41,44]. Then, the remaining task is to propose proper analysis step update in the second line of (36) concerning consistent statistics with the optimal solution $\hat{\rho}_t$ in (32).

3.4. Statistical consistency in analysis step update of the approximating filter

Now, we focus on updating posterior PDF $\tilde{\rho}_t$ in (34) of the proposed approximating filter (36) based on the statistical observation y_t satisfying (23). Concentrating on the analysis step, the resulting optimal filter Eqs. (32) for the mean and covariance $\left(\hat{\rho}_t, \hat{C}_t\right)$ become

$$\begin{aligned}\mathrm{d}\hat{\rho}_t &= \hat{C}_t\mathcal{H}^*\Gamma_t^{-2}\left\{\mathrm{d}y_t - \left[\mathcal{H}\hat{\rho}_t + h_t\left(y_t\right)\right]\mathrm{d}t\right\},\\ \mathrm{d}\hat{C}_t &= -\hat{C}_t\mathcal{H}^*\Gamma_t^{-2}\mathcal{H}\hat{C}_t\mathrm{d}t.\end{aligned} \tag{37}$$

Correspondingly, the approximating statistical filtering model for $\tilde{Z}_t$ satisfies the second line of the SDE (36) as

$$\mathrm{d}\tilde{Z}_t = a_t\left(\tilde{Z}_t\right)\mathrm{d}t + K_t\left(\tilde{Z}_t\right)\left\{\mathrm{d}y_t - \left[H\left(\tilde{Z}_t\right) + h_t\left(y_t\right)\right]\mathrm{d}t - \Gamma_t\mathrm{d}\tilde{B}_t\right\}. \tag{38}$$

Following the similar idea in the McKean–Vlasov representation of the filtering equation [19,59], we expect the PDF $\tilde{\rho}_t$ of $\tilde{Z}_t$ to satisfy the following Kushner–Stratonovich-type equation (with requirements on a_t, K_t given next in (41))

$$\frac{\partial\tilde{\rho}_t}{\partial t} = \left[H\left(z\right) - \mathcal{H}\tilde{\rho}_t\right]^{\intercal}\Gamma_t^{-2}\left[\frac{\mathrm{d}y_t}{\mathrm{d}t} - \mathcal{H}\tilde{\rho}_t - h_t\left(y_t\right)\right]\tilde{\rho}_t. \tag{39}$$

Again, the goal here is to approximate the optimal filter mean $\hat{\rho}_t$ in (37) by $\tilde{\rho}_t$ generated by the surrogate SDE model (38) in the sense of consistent statistics.

Unfortunately, the approximation (39) and the optimal filtering Eq. (37) will in general have different continuous solutions for $\tilde{\rho}_t$ and $\hat{\rho}_t$ due to their distinctive dynamics. In order to compare key statistics of the two distributions, we apply the linear operator $\mathcal{H}$ to the optimal Eqs. (37) as a finite-dimensional projection on leading moments based on observations. The resulting optimal mean and covariance equations become finite dimensional as

$$\begin{aligned}\mathrm{d}\left(\mathcal{H}\hat{\rho}_t\right) &= \left(\mathcal{H}\hat{C}_t\mathcal{H}^*\right)\Gamma_t^{-2}\left\{\mathrm{d}y_t - \left[\mathcal{H}\hat{\rho}_t + h_t\left(y_t\right)\right]\mathrm{d}t\right\},\\ \mathrm{d}\left(\mathcal{H}\hat{C}_t\mathcal{H}^*\right) &= -\left(\mathcal{H}\hat{C}_t\mathcal{H}^*\right)\Gamma_t^{-2}\left(\mathcal{H}\hat{C}_t\mathcal{H}^*\right)\mathrm{d}t.\end{aligned} \tag{40}$$

Above, remind that the observation operator $\mathcal{H} : L^2\left(\mathbb{R}^d\right) \to \mathbb{R}^p$ and its adjoint $\mathcal{H}^* : \mathbb{R}^p \to L^2\left(\mathbb{R}^d\right)$ are defined based on the observation function $H \in C^2\left(\mathbb{R}^d;\mathbb{R}^p\right)$ as

$$\mathcal{H}\rho = \int H\left(z\right)\rho\left(z\right)\mathrm{d}z,\quad \left[\mathcal{H}^*u\right]\left(z\right) = u\cdot H\left(z\right),$$

and the covariance operator $\hat{C}_t : L^2\left(\mathbb{R}^d\right) \to L^2\left(\mathbb{R}^d\right)$ is defined in (30). Therefore, (40) gives the equations for the finite-dimensional quantities $\mathcal{H}\hat{\rho}_t \in \mathbb{R}^p$ and $\mathcal{H}\hat{C}_t\mathcal{H}^* \in \mathbb{R}^{p\times p}$ as the first two moments of H w.r.t. $\hat{\rho}_t$. The idea then is to design the analysis step operator $\mathcal{A}_t^{\tau}$ according to the approximating filter process $\tilde{Z}_t$ in (38), so that consistency in the first and second-order moments (40) can be achieved.

Denote the expectation, $\tilde{\mathbb{E}}[\cdot] := \mathbb{E}\left[\cdot \mid \mathcal{G}_t\right] = \mathbb{E}_{\tilde{\rho}_t}[\cdot]$, w.r.t.$\tilde{\rho}_t$ in (39) conditional on the observation $\mathcal{G}_t$. We define $\bar{H}_t = \tilde{\mathbb{E}}H\left(\tilde{Z}_t\right) = \int H\left(z\right)\tilde{\rho}_t\left(z\right)\mathrm{d}z$ and $C_t^H = \tilde{\mathbb{E}}\left[H_t\left(\tilde{Z}_t\right) - \bar{H}_t\right]\left[H_t\left(\tilde{Z}_t\right) - \bar{H}_t\right]^{\intercal}$ as the first and second-order moments of H w.r.t. $\tilde{\rho}_t$. Assume that the drift a_t and gain K_t in the SDE approximation (38) satisfy the following identities

$$a_t = \nabla\cdot\left(K_t\Gamma_t^2K_t^{\intercal}\right) - K_t\Gamma_t^2\nabla\cdot K_t^{\intercal},\quad -\nabla\cdot\left(K_t^{\intercal}\tilde{\rho}_t\right) = \tilde{\rho}_t\Gamma_t^{-2}\left(H\left(z\right) - \tilde{\mathbb{E}}H\right), \tag{41}$$

where the divergence on a matrix is defined columnwise as $(\nabla\cdot A)_i = \sum_j \partial_{z_j}A_{ij}$. We first have the following result concerning the evolution equations of $\bar{H}_t$ and C_t^H given the realization $Y_t = \left\{y_s, s\le t\right\}$.

Lemma 5. *Given that $\Gamma_t \succ 0$ in (21) and the identities in (41) are satisfied, the evolution equations for the mean and covariance of the observation function $H\left(\tilde{Z}_t\right)$ associated with the SDE (38) are given by*

$$\begin{aligned}\mathrm{d}\bar{H}_t &= C_t^H\Gamma_t^{-2}\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right]\mathrm{d}t\right\},\\ \mathrm{d}C_t^H &= Q_t^H\Gamma_t^{-2}\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right]\mathrm{d}t\right\} - C_t^H\Gamma_t^{-2}C_t^H\mathrm{d}t,\end{aligned} \tag{42}$$

where $Q_t^H : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ is defined as

$$Q_t^H = \tilde{\mathbb{E}}\left[\left(H_t'H_t'^{\intercal}\right)\otimes H_t'^{\intercal}\right],$$

containing third moments of $H_t' = H\left(\tilde{Z}_t\right) - \bar{H}_t$.

We put the detailed derivation of (42) in Appendix B. Notice that (42) goes back to the Kalman–Bucy filter if we set linear observation $H\left(\tilde{Z}_t\right) = \tilde{Z}_t$ satisfying a normal distribution as in [37]. However, here we are considering the more general nonlinear dynamics and quadratic and cubic observation functions from (24).

Comparing (40) and (42) implies that the same statistical solution can be reached in $\left(\bar{H}_t, C_t^H\right)$ and $\left(\mathcal{H}\hat{\rho}_t, \mathcal{H}\hat{C}_t\mathcal{H}^*\right)$ if we have $Q_t^H = 0$. In order to achieve this, we further introduce the projection operator on the space of probability distributions using the Kullback–Leibler (KL) divergence as an unbiased metric.

Definition 6. Define the operator $\mathcal{S}_H$ making symmetric projection on the probability density $\rho \in \mathcal{P}_\infty\left(\mathbb{R}^d\right)$ with finite moments

$$\mathcal{S}_H\rho = \underset{\nu\in\mathcal{V}_H}{\arg\min}\, d_{\mathrm{KL}}\left(\nu \parallel \rho\right), \tag{43}$$

where d_{KL} is the KL-divergence between two probability measures. The minimization is among the probability measures in the following set

$$\begin{aligned}&\mathcal{V}_H\left(\bar{H}, C^H\right)\\ &= \left\{\nu\in\mathcal{P}_\infty\left(\mathbb{R}^d\right) : \mathbb{E}_\nu H = \bar{H},\ \mathbb{E}_\nu\left[H'H'^{\intercal}\right] = C^H,\ \text{and}\ \mathbb{E}_\nu\left[H_l'H_m'H_n'\right] = 0\right\},\end{aligned}$$

for all $1\le l,m,n\le d$ and $H'(Z) = H(Z) - \mathbb{E}_\nu[H(Z)]$.

In Definition 6, $\mathcal{S}_H$ acts as a symmetric approximation of probability measures with vanishing third-order moments of the observation function H, while maintains consistent first two leading moments of H. It is clear that given $\bar{H}, C^H$, the set $\mathcal{V}_H$ is closed with respect to weak convergence of measures. From Proposition 2.1 of [60,61], we have for any ρ and weakly convergent sequence $\left\{\nu_n\right\}$ to ν_*

$$\liminf_{n\to\infty} d_{\mathrm{KL}}\left(\nu_n \parallel \rho\right) \ge d_{\mathrm{KL}}\left(\nu_* \parallel \rho\right).$$

It follows immediately that there exists $v_* \in \mathcal{V}_H$ that reaches the minimum. Therefore, we have the following lemma guaranteeing the existence of the minimizer in the proposed projection (43).

Lemma 7. *Assume that there exists one $v \in \mathcal{V}_H$ such that the KL-divergence $d_{\mathrm{KL}}(v \parallel \rho) < \infty$ for given $\rho \in \mathcal{P}_\infty(\mathbb{R}^d)$. Then a minimizer exists in (43).*

Lemma 7 makes sure that we can always find one minimizer given the same mean and covariance $\bar{H} = \tilde{\mathbb{E}}[H(Z)], C^H = \tilde{\mathbb{E}}[H'(Z)H'(Z)^\intercal]$. This provides the density function satisfying the required symmetric statistics about $H(Z_t)$. Denote the push-forward operator for the new SDE (38) with the structure functions (41) as $\tilde{\rho}_{t+s} = \mathcal{Q}_t^s(\tilde{\rho}_t; Y_{t+s})$ for any $s \geq 0$. Under the above construction, we can finally propose the forward operator in analysis step (34) as

$$\mathcal{A}_t^s(\tilde{\rho}_t) := \mathcal{Q}_t^s(\mathcal{S}_H\tilde{\rho}_t; Y_{t+s}), \tag{44}$$

where the projection $\mathcal{S}_H$ in (43) is a linear operator acting on the random fields in space $\mathcal{V}_t$ in (26). In a similar fashion as in the proof of Lemma 5, by applying the addition projection $\mathcal{S}_H\tilde{\rho}_t$ in the expectations on the SDEs for $H(\tilde{Z}_t)$ and $H_t'(\tilde{Z}_t)H_t'(\tilde{Z}_t)^\intercal$ (see also (B.10) and (B.11) in Appendix B), $Q_t^H = 0$ is automatically guaranteed in (42) w.r.t. the new projected density $\mathcal{S}_H\tilde{\rho}_t$. In addition, the first two moments $\bar{H}_t$ and C_t^H in (42) will stay the same w.r.t. $\mathcal{S}_H\tilde{\rho}_t$. The continuous filtering process by letting $s \to 0$ will satisfy the Eqs. (42) with $Q_t^H \equiv 0$. Therefore, under the same initial condition and the uniqueness of the solution, the same solution will be reached in both (37) and (42). This leads to the main result of this section concerning the analysis step update in the approximating filter solution.

Theorem 8. *Consider the analysis step update (44) of the statistical filtering model (38). Assume that a_t, K_t in the statistical filtering SDE are designed to satisfy (41) and the probability set $\mathcal{V}_H$ according to H defined in (43) is not empty. Under the same statistical observations $y_t, t \in [0, T]$ and the same initial conditions, the following identities hold for $t \in [0, T]$*

$$\mathcal{H}\hat{\rho}_t = \tilde{\mathbb{E}}[H(\tilde{Z}_t)], \quad \mathcal{H}\hat{C}_t\mathcal{H}^* = \tilde{\mathbb{E}}[H'(\tilde{Z}_t)H'(\tilde{Z}_t)^\intercal], \tag{45}$$

where $(\hat{\rho}_t, \hat{C}_t)$ is the solution of (37), and $\tilde{\mathbb{E}}$ is w.r.t. $\tilde{\rho}_t$ given by the solution of (39) with $H' = H - \mathcal{H}\hat{\rho}_t$.

Theorem 8 validates the use of the statistical filtering model density $\tilde{\rho}_t$ by solving (36) to approximate the optimal filter $\hat{\rho}_t$ from (32). Though restricted only on the first two moments of the observation function H, the resulting consistent statistics during analysis step play a key role in accurate statistical forecast. Notice that based on the statistical model in (11), accurate prediction of the important leading statistics, $\bar{u}_t, R_t$, is determined by key higher-order feedbacks in the related functional $\mathcal{H}\rho_t$ (more specifically, the terms $\mathcal{H}_m\rho_t$ and $\mathcal{H}_v\rho_t$ in (24)). According to Proposition 4, the optimal $\mathcal{H}\hat{\rho}_t$ gives the least mean square estimate of the random variable $\mathcal{H}\rho_t$ given the statistical observations. Thus, consistent filtering approximation $\tilde{\mathbb{E}}H(\tilde{Z}_t)$ for $\mathcal{H}\hat{\rho}_t$ as well as its error estimate guarantees accurate recovery of key model statistics.

For example, applying the explicit forms of the observation function (24), the quadratic observation operator $\mathcal{H}_m$ in the mean equation gives

$$\tilde{\mathbb{E}}H^m(\tilde{Z}_t) = \mathcal{H}_m\hat{\rho}_t \;\Leftrightarrow\; \sum_{p,q}\gamma_{kpq}\tilde{\mathbb{E}}[\tilde{Z}_{p,t}\tilde{Z}_{q,t}] = \sum_{p,q}\gamma_{kpq}\int z_p z_q \hat{\rho}_t(z)\,\mathrm{d}z,$$

which implies consistent statistical feedbacks in the mean Eq. (7a) from the statistical filtering model $\tilde{\rho}_t$ and the optimal filter solution $\hat{\rho}_t$. This demonstrates that the new approximating filter maintains the accuracy in the statistical mean prediction $\bar{u}_t$. In addition, the covariance operator characterizes the essential uncertainty in the optimal filter $\hat{\rho}_t$

$$\mathcal{H}_m\hat{C}_t\mathcal{H}_m^* = \mathbb{E}_{\mu_t}\left[(\mathcal{H}_m\rho_t - \mathcal{H}_m\hat{\rho}_t)(\mathcal{H}_m\rho_t - \mathcal{H}_m\hat{\rho}_t)^\intercal\right],$$

which is also linked to the approximation by $C_t^{H^m} = \tilde{\mathbb{E}}\left[(H_t^m - \bar{H}_t^m)(H_t^m - \bar{H}_t^m)^\intercal\right] = \mathcal{H}_m\hat{C}_t\mathcal{H}_m^*$, demonstrating a consistent error estimate in the statistical filtering model. Similar conclusion can be reached for the accurate prediction in the model covariance prediction for R_t based on the cubic observation operator $\mathcal{H}_v$.

Remark. Still, the statistical consistency in the analysis step does not guarantee the consistency in the entire two-step updating procedure in (35). In particular, the forecast models of (32) and (36) satisfy the following forecast equations

$$\begin{aligned}\partial_t\hat{\rho}_t &= \mathcal{L}_t^*\hat{\rho}_t \\ \partial_t\tilde{\rho}_t &= \mathcal{L}_t^*\tilde{\rho}_t\end{aligned} \;\Rightarrow\; \begin{aligned}\partial_t(\mathcal{H}\hat{\rho}_t) &= \mathcal{H}\mathcal{L}_t^*\hat{\rho}_t = \int(\mathcal{L}_tH)(z)\hat{\rho}_t(z)\,\mathrm{d}z \\ \partial_t(\tilde{\mathbb{E}}H) &= \tilde{\mathbb{E}}\mathcal{L}_tH = \int(\mathcal{L}_tH)(z)\tilde{\rho}_t(z)\,\mathrm{d}z\end{aligned}$$

where the generator $\mathcal{L}_t$ is defined as in (20). The analysis step update only gives consistent first two moments of H, while higher moments may be included in $\mathcal{L}_tH$. The forecast model may require additional consistent condition between $\hat{\rho}_t$ and $\tilde{\rho}_t$ as well as their covariances. More work is still needed for the complete analysis combining approximations in both the forecast and analysis step of the full filtering model.

4. Stability and convergence of the statistical filtering model

Following the new approximating filtering model, we discuss the long-time performance of the filtering problem for finding the optimal filter PDF $\hat{\rho}_t$ of the statistical filtering system (21) based on statistical observations $Y_t = \{y_s, s \leq t\}$. Section 3 shows that the statistical filtering model (36) constitutes the approximating filter $\tilde{\rho}_t$ with consistent mean and covariance, $\bar{H}_t$ and C_t^H, during the analysis step update. Here, we show further that the full filter approximation of the observation function $\bar{H}_t$ will approach the optimal filter $\mathcal{H}\hat{\rho}_t$ at the long-time limit as $t \to \infty$. This guarantees the stable performance of the proposed new filtering strategy.

4.1. Closed statistical filtering equations based on the observation operator

We consider the optimal filter solution based on the conditional Gaussian processes (21). The finite-dimensional statistical states $\mathcal{H}\hat{\rho}_t \in \mathbb{R}^p$ and $\hat{C}_t^{\mathcal{H}} = \mathcal{H}\hat{C}_t\mathcal{H}^* \in \mathbb{R}^{p\times p}$ under the observation operator can be solved by the Kalman–Bucy equations (32) as

$$\begin{aligned}\mathrm{d}(\mathcal{H}\hat{\rho}_t) &= \langle\mathcal{L}_t(y_t)H\hat{\rho}_t\rangle\,\mathrm{d}t + \hat{C}_t^{\mathcal{H}}\Gamma^{-2}\{\mathrm{d}y_t - [\mathcal{H}\hat{\rho}_t + h_t(y_t)]\,\mathrm{d}t\}, \\ \mathrm{d}\hat{C}_t^{\mathcal{H}} &= [\langle\mathcal{L}_t(y_t)H(\hat{C}_tH^\intercal)\rangle + \langle(\hat{C}_tH)\mathcal{L}_t(y_t)H^\intercal\rangle]\,\mathrm{d}t - \hat{C}_t^{\mathcal{H}}\Gamma^{-2}\hat{C}_t^{\mathcal{H}}\mathrm{d}t.\end{aligned} \tag{46}$$

Above, $\langle F\rangle = \int F(z)\,\mathrm{d}z$ denotes the componentwise integration of the matrix-valued functions F. For simplicity, we use constant observation noise, $\Gamma_t \equiv \Gamma$. In the first equation for the mean, we rewrite the forecast step dynamics as

$$\mathcal{H}\mathcal{L}_t^*(y_t)\hat{\rho}_t = \int H(z)\mathcal{L}_t^*(y_t)\hat{\rho}_t(z)\,\mathrm{d}z = \langle\mathcal{L}_t(y_t)H\hat{\rho}_t\rangle.$$

Similarly, we rewrite the covariance equation using the definition of $\hat{C}_t$ in (30) under the conditional measure $\mu_t(\cdot) = \mathbb{P}(\rho \in \cdot \mid \mathcal{G}_t)$

$$\begin{aligned}\mathcal{H}\mathcal{L}_t^*\hat{C}_t\mathcal{H}^* &= \mathbb{E}\int H(x)\mathcal{L}_t^*(\rho - \hat{\rho}_t)(x)\,\mathrm{d}x\int(\rho - \hat{\rho}_t)(z)H(z)^\intercal\,\mathrm{d}z \\ &= \langle\mathcal{L}_tH(\hat{C}_tH^\intercal)\rangle, \\ \mathcal{H}\hat{C}_t\mathcal{L}_t\mathcal{H}^* &= \mathbb{E}\int H(z)(\rho - \hat{\rho}_t)(z)\,\mathrm{d}z\int(\rho - \hat{\rho}_t)(x)\mathcal{L}_tH(x)^\intercal\,\mathrm{d}x \\ &= \langle(\hat{C}_tH)\mathcal{L}_tH^\intercal\rangle,\end{aligned}$$

where $\hat{C}_tH(x) = \mathbb{E}\left[(\rho - \hat{\rho}_t)(x)\int(\rho - \hat{\rho}_t)(z)H(z)\,\mathrm{d}z\right] \in \mathbb{R}^d$ can be viewed as an unnormalized density that satisfies $\int\hat{C}_tH(x)\,\mathrm{d}x = 0$.

Correspondingly, consider the approximating filtering model (36) with $\Sigma_t \equiv 0$. Let $\tilde{\rho}_t$ be the PDF after the symmetric projection (43), that is, satisfying vanishing third-order moments $\tilde{\mathbb{E}}\left[H_n'H_p'H_q'\right] = 0$ for all n, p, q. The first and second-order moments $\bar{H}_t = \tilde{\mathbb{E}}[H(\tilde{Z}_t)]$ and $C_t^H = \tilde{\mathbb{E}}[H_t'(\tilde{Z}_t)H_t'(\tilde{Z}_t)^\intercal]$ according to the observation function H satisfy the following equations (by combining the forecast step update with generator $\mathcal{L}_t$ and the analysis step dynamics (42))

$$\begin{aligned}\mathrm{d}\bar{H}_t &= \left\langle \mathcal{L}_t\left(y_t\right) H\tilde{\rho}_t\right\rangle \mathrm{d}t + C_t^H \Gamma^{-2}\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right]\mathrm{d}t\right\},\\ \mathrm{d}C_t^H &= \left[\left\langle \mathcal{L}_t\left(y_t\right) H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle + \left\langle\left(\tilde{\rho}_t H_t'\right)\mathcal{L}_t\left(y_t\right) H^{\intercal}\right\rangle\right]\mathrm{d}t - C_t^H \Gamma^{-2} C_t^H \mathrm{d}t,\end{aligned} \tag{47}$$

where $H_t'(z) = H(z) - \bar{H}_t$.

First, we introduce the following assumptions for the approximating filter PDF $\tilde{\rho}_t$ approximating the optimal filter solution $\hat{\rho}_t$ as random fields according to the same observation process $Y = \left\{y_t, t \geq 0\right\}$:

Assumption 9. Assume that the optimal filter (32) and the approximating filtering model (36) have probability density functions $\hat{\rho}_t$ and $\tilde{\rho}_t$ respectively that satisfy the following conditions:

• Unique equilibrium solutions $\left(\mathcal{H}\hat{\rho}_\infty, \hat{C}_\infty^{\mathcal{H}}\right)$ and $\left(\mathcal{H}\tilde{\rho}_\infty, C_\infty^H\right)$ exist to the systems (46) and (47) respectively.

• There exist deterministic matrices $L_\infty^m, L_\infty^v \in \mathbb{R}^{p\times p}$, such that the generator $\mathcal{L}_t$ (20) reaches the same statistical limit under H w.r.t. both $\hat{\rho}_t$ and $\tilde{\rho}_t$, that is,

$$\begin{aligned}&\left\langle \mathcal{L}_t\left(y_t\right) H\hat{\rho}_t\right\rangle \to L_\infty^m \mathcal{H}\hat{\rho}_\infty,\ \left\langle \mathcal{L}_t\left(y_t\right) H\left(\hat{C}_t H^{\intercal}\right)\right\rangle \to L_\infty^v \hat{C}_\infty^{\mathcal{H}},\\ &\left\langle \mathcal{L}_t\left(y_t\right) H\tilde{\rho}_t\right\rangle \to L_\infty^m \mathcal{H}\tilde{\rho}_\infty,\ \left\langle \mathcal{L}_t\left(y_t\right) H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle \to L_\infty^v C_\infty^H,\end{aligned} \tag{48}$$

a.s. as $t \to \infty$ conditional on the observation data Y.

• Further, the real parts of eigenvalues of the limit matrices $L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}$ and $L_\infty^v - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}$ are all negative.

The first condition in Assumption 9 guarantees that both filter equations will finally converge to finite steady-state statistical solutions. The second condition requires consistent first and second-order moments under H w.r.t. the optimal and approximating model distributions, such as for the covariance as $t \to \infty$

$$\begin{aligned}\int \mathcal{L}_t\left(y_t\right) H(z)\hat{C}_t H(z)^{\intercal}\mathrm{d}z &\to L_\infty^v \int H(z)\left(\hat{C}_\infty H^{\intercal}\right)(z)\,\mathrm{d}z,\\ \int \mathcal{L}_t\left(y_t\right) H_t'(z) H_t'(z)^{\intercal}\tilde{\rho}_t(z)\,\mathrm{d}z &\to L_\infty^v \int H_\infty'(z) H_\infty'(z)^{\intercal}\tilde{\rho}_\infty(z)\,\mathrm{d}z.\end{aligned}$$

In addition, we may further introduce the convergence rate, that is, there exist constants $\bar{\lambda} > 0$ and $K > 0$ such that a.s.

$$\begin{aligned}&\left|\left\langle \mathcal{L}_t\left(y_t\right) H\hat{\rho}_t\right\rangle - L_\infty^m \mathcal{H}\hat{\rho}_\infty\right| \leq K e^{-\bar{\lambda}t}, \quad \left\|\left\langle \mathcal{L}_t\left(y_t\right) H\left(\hat{C}_t H^{\intercal}\right)\right\rangle - L_\infty^v \hat{C}_\infty^{\mathcal{H}}\right\| \leq K e^{-\bar{\lambda}t},\\ &\left|\left\langle \mathcal{L}_t\left(y_t\right) H\tilde{\rho}_t\right\rangle - L_\infty^m \mathcal{H}\tilde{\rho}_\infty\right| \leq K e^{-\bar{\lambda}t}, \quad \left\|\left\langle \mathcal{L}_t\left(y_t\right) H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle - L_\infty^v C_\infty^H\right\| \leq K e^{-\bar{\lambda}t},\end{aligned} \tag{49}$$

where $\|\cdot\|$ is the matrix norm. And the third condition requires that the filter solutions will be stabilized at the long-time limit. These assumptions are based on the observation that in practice a final equilibrium state will be reached and maintain stable dynamics. Next, we ask the limiting behaviors in the mean and covariance $\left(\bar{H}_t, C_t^H\right)$ from the approximating model (47) compared with the optimal filter solution $\left(\mathcal{H}\hat{\rho}_t, \hat{C}_t^{\mathcal{H}}\right)$ from (46).

4.2. Asymptotic stability of the equilibrium covariance matrix

With the assumption (48), we first have consistent equilibrium covariances as $t \to \infty$ in the two model solutions, that is,

$$\begin{aligned}L_\infty^v \hat{C}_\infty^{\mathcal{H}} + \hat{C}_\infty^{\mathcal{H}} L_\infty^{v\intercal} - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\hat{C}_\infty^{\mathcal{H}} = 0,\\ L_\infty^v C_\infty^H + C_\infty^H L_\infty^{v\intercal} - C_\infty^H \Gamma^{-2} C_\infty^H = 0.\end{aligned}$$

Uniqueness of the solution directly implies that the final equilibrium covariances satisfy $\hat{C}_\infty^{\mathcal{H}} = C_\infty^H$ with no randomness. Further, we can find that the covariance C_t^H in the approximating filter will approach the optimal equilibrium covariance $\hat{C}_\infty^{\mathcal{H}}$ as described in the following lemma.

Lemma 10. *Suppose that Assumption 9 is satisfied, C_t^H is the covariance solution to the statistical filtering model (47) and $\hat{C}_\infty^{\mathcal{H}}$ is the unique equilibrium solution to the optimal model (46). Then there is*

$$\left\|C_t^H - \hat{C}_\infty^{\mathcal{H}}\right\| \to 0, \quad \text{a.s. as } t \to \infty. \tag{50}$$

Further, the convergence rate will be exponential if (49) is also satisfied.

Proof. Combining the covariance equation in (47) and the equilibrium equation of (46), we have

$$\begin{aligned}\mathrm{d}\left(C_t^H - \hat{C}_\infty^{\mathcal{H}}\right) = &\left[L_\infty^v - \frac{1}{2}\left(C_t^H + \hat{C}_\infty^{\mathcal{H}}\right)\Gamma^{-2}\right]\left(C_t^H - \hat{C}_\infty^{\mathcal{H}}\right)\mathrm{d}t\\ &+ \left(C_t^H - \hat{C}_\infty^{\mathcal{H}}\right)\left[L_\infty^{v\intercal} - \frac{1}{2}\Gamma^{-2}\left(C_t^H + \hat{C}_\infty^{\mathcal{H}}\right)\right]\mathrm{d}t\\ &+ \left[\left\langle \mathcal{L}_t H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle - L_\infty^v C_t^H\right]\mathrm{d}t\\ &+ \left[\left\langle\left(\tilde{\rho}_t H_t'\right)\mathcal{L}_t H^{\intercal}\right\rangle - C_t^H L_\infty^{v\intercal}\right]\mathrm{d}t.\end{aligned}$$

For the last row of the above equation, using the uniqueness of the solution we have $C_t^H \to C_\infty^H$ as $t \to \infty$. Then denote

$$F_t = \left\langle \mathcal{L}_t H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle - L_\infty^v C_t^H.$$

We get $\left\|F_t\right\| \to 0$ as $t \to \infty$ by using

$$\begin{aligned}&\left\|\left\langle \mathcal{L}_t H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle - L_\infty^v C_t^H\right\|\\ &\leq \left\|\left\langle \mathcal{L}_t H\left(\tilde{\rho}_t H_t'^{\intercal}\right)\right\rangle - L_\infty^v C_\infty^H\right\| + \left\|L_\infty^v\left(C_\infty^H - C_t^H\right)\right\| \to 0.\end{aligned}$$

Next, by taking $\lambda_v = \min\left\{\mathrm{Re}\lambda : \lambda \text{ is the eighenvalue of } -L_\infty^v + \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right\}$, we have $\lambda_v > 0$ from Assumption 9. This implies $\left\|e^{\int_s^t\left[L_\infty^v - \frac{1}{2}\left(C_\tau^H + \hat{C}_\infty^{\mathcal{H}}\right)\Gamma^{-2}\right]\mathrm{d}\tau}\right\| \leq K e^{-\lambda_v(t-s)}$. Together with F_t vanishing as $t \to \infty$, we have for any $t > T$

$$\begin{aligned}\left\|C_t^H - \hat{C}_\infty^{\mathcal{H}}\right\| &\leq K\left\|C_T^H - \hat{C}_\infty^{\mathcal{H}}\right\| e^{-2\lambda_v(t-T)} + 2K\int_T^t e^{-2\lambda_v(t-s)}\left\|F_s\right\|\mathrm{d}s\\ &\leq K_1 e^{-2\lambda_v t} + K_2 \sup_{s\geq T}\left\|F_s\right\|.\end{aligned}$$

Therefore, by first letting $t \to \infty$ then letting $T \to \infty$, we reach a.s. $\left\|C_t^H - \hat{C}_\infty^{\mathcal{H}}\right\| \to 0$.

Further, if we assume exponential convergence rate $\bar{\lambda}$ in the covariances under the generator $\mathcal{L}_t$ as in (49), we can have exponential convergence rate in both $\hat{C}_t$ and C_t^H as $t \to \infty$ a.s.

$$\left\|\hat{C}_t^{\mathcal{H}} - \hat{C}_\infty^{\mathcal{H}}\right\| \leq K_H e^{-\min\left\{\lambda_v, \bar{\lambda}\right\}t}, \quad \left\|C_t^H - \hat{C}_\infty^{\mathcal{H}}\right\| \leq K_H e^{-\min\left\{\lambda_v, \bar{\lambda}\right\}t}. \quad \square \tag{51}$$

In addition, from the definition of the optimal filter solution (31), the covariance is defined as

$$\begin{aligned}\mathrm{tr}\mathbb{E}\left[\hat{C}_t^{\mathcal{H}}\right] &= \mathbb{E}\mathrm{tr}\left[\left(\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right)\left(\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right)^{\intercal}\right]\\ &= \mathbb{E}\mathrm{tr}\left[\left(\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right)^{\intercal}\left(\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right)\right]\\ &= \mathbb{E}\left[\left|\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right|^2\right].\end{aligned}$$

Notice that $\mathcal{H}\hat{\rho}_t$ and $\hat{C}_t^{\mathcal{H}}$ are still $\mathcal{G}_t$-measurable random field. Under Assumption 9 $\hat{C}_t^{\mathcal{H}} \to \hat{C}_\infty^{\mathcal{H}}$ a.s. conditional on the observations $\mathcal{G}_t$, we have as $t \to \infty$

$$\mathbb{E}\left[\left|\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right|^2\right] = \mathrm{tr}\mathbb{E}\left[\hat{C}_t^{\mathcal{H}}\right] \to \mathrm{tr}\hat{C}_\infty^{\mathcal{H}}. \tag{52}$$

This confirms that the total uncertainty at equilibrium in the optimal filter solution $\mathcal{H}\hat{\rho}_t$ is estimated by the total variance $\mathrm{tr}\hat{C}_\infty^{\mathcal{H}}$.

4.3. Convergence of the statistical state under the observation operator

Next, we consider the convergence of the statistical observation function $\bar{H}_t$ from the approximating filter to the optimal filter solution $\mathcal{H}\hat{\rho}_t$. We have the long-term stability in the statistical solution in the following theorem.

Theorem 11. *Suppose that Assumption 9 holds and the covariances go to the same deterministic limit*

$$\hat{C}_t^{\mathcal{H}} \to \hat{C}_\infty^{\mathcal{H}}, \quad C_t^H \to \hat{C}_\infty^{\mathcal{H}},$$

a.s. as $t \to \infty$. Then there is

$$\mathbb{E}\left[\left|\mathcal{H}\hat{\rho}_t - \bar{H}_t\right|^2\right] \to 0, \tag{53}$$

as $t \to \infty$. Furthermore, assume exponential convergence rate in $\hat{C}_t^{\mathcal{H}}, C_t^H$ as in (51), and let $\lambda = \min\left\{\lambda_m, \lambda_v\right\}$ where λ_m and λ_v are the minimums of the real parts of the eigenvalues of $-L_\infty^m + \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}$ and $-L_\infty^v + \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}$ respectively. There is also exponential convergence as

$$\mathbb{E}\left[\left|\mathcal{H}\hat{\rho}_t - \bar{H}_t\right|^2\right] \le K_H e^{-\min\{\lambda,\bar{\lambda}\}t}, \tag{54}$$

with K_H a constant only dependent on the observation function H.

Proof. By taking the difference of the mean equations in (46) and (47), we have

$$\begin{aligned}\mathrm{d}\left(\mathcal{H}\hat{\rho}_t - \bar{H}_t\right) = &\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)\left(\mathcal{H}\hat{\rho}_t - \bar{H}_t\right)\mathrm{d}t + \left(\hat{C}_t^{\mathcal{H}} - C_t^H\right)\Gamma^{-2}\left[\mathrm{d}y_t - h_t\left(y_t\right)\mathrm{d}t\right] \\ &+ \left[\left(\hat{C}_\infty^{\mathcal{H}} - \hat{C}_t^{\mathcal{H}}\right)\Gamma^{-2}\mathcal{H}\hat{\rho}_t - \left(\hat{C}_\infty^{\mathcal{H}} - C_t^H\right)\Gamma^{-2}\bar{H}_t\right]\mathrm{d}t \\ &+ \left[\left\langle \mathcal{L}_t\left(y_t\right)H\hat{\rho}_t\right\rangle - L_\infty^m\mathcal{H}\hat{\rho}_t\right]\mathrm{d}t + \left[\left\langle \mathcal{L}_t\left(y_t\right)H\tilde{\rho}_t\right\rangle - L_\infty^m\bar{H}_t\right]\mathrm{d}t.\end{aligned}$$

By applying Itô's formula to the above equation, there is

$$\begin{aligned}&\mathrm{d}\left[e^{-t\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\mathcal{H}\hat{\rho}_t - \bar{H}_t\right)\right] \\ &= e^{-t\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left\{\left(\hat{C}_t^{\mathcal{H}} - C_t^H\right)\Gamma^{-2}\left[\mathrm{d}y_t - h_t\left(y_t\right)\mathrm{d}t\right] + G_t\left(y_t\right)\mathrm{d}t\right\} \\ &+ e^{-t\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left[\left(\hat{C}_\infty^{\mathcal{H}} - \hat{C}_t^{\mathcal{H}}\right)\Gamma^{-2}\mathcal{H}\hat{\rho}_t - \left(\hat{C}_\infty^{\mathcal{H}} - C_t^H\right)\Gamma^{-2}\bar{H}_t\right]\mathrm{d}t.\end{aligned}$$

Above, we denote the residual term as

$$G_t\left(y_t\right) = \left[\left\langle \mathcal{L}_t\left(y_t\right)H\hat{\rho}_t\right\rangle - L_\infty^m\mathcal{H}\hat{\rho}_t\right] + \left[\left\langle \mathcal{L}_t\left(y_t\right)H\tilde{\rho}_t\right\rangle - L_\infty^m\bar{H}_t\right].$$

By similar estimates as in Lemma 10 according to the assumption (48), there is

$$\left|G_t\right| \le \begin{matrix}\left|\left\langle \mathcal{L}_t\left(y_t\right)H\hat{\rho}_t\right\rangle - L_\infty^m\mathcal{H}\hat{\rho}_\infty\right| + \left|L_\infty^m\left(\mathcal{H}\hat{\rho}_\infty - \mathcal{H}\hat{\rho}_t\right)\right| \\ + \left|\left\langle \mathcal{L}_t\left(y_t\right)H\tilde{\rho}_t\right\rangle - L_\infty^m\bar{H}_\infty\right| + \left|L_\infty^m\left(\bar{H}_\infty - \bar{H}_t\right)\right|\end{matrix} \le Ke^{-\bar{\lambda}t} \to 0,$$

a.s. as $t \to \infty$ using the uniqueness of the solutions $\mathcal{H}\hat{\rho}_t \to \mathcal{H}\hat{\rho}_\infty$ and $\bar{H}_t \to \bar{H}_\infty$ (with exponential decay in the stronger convergence case (51)). Equivalently, the above SDE can be written as (ignoring initial condition by assuming $\mathcal{H}\hat{\rho}_0 = \bar{H}_0$)

$$\begin{aligned}\mathcal{H}\hat{\rho}_t - \bar{H}_t = &\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left\{\left(\hat{C}_s^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\left[\mathrm{d}y_s - h_s\left(y_s\right)\mathrm{d}s\right] + G_s\left(y_s\right)\mathrm{d}s\right\} \\ &+ \int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left[\left(\hat{C}_\infty^{\mathcal{H}} - \hat{C}_s^{\mathcal{H}}\right)\Gamma^{-2}\mathcal{H}\hat{\rho}_s - \left(\hat{C}_\infty^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\bar{H}_s\right]\mathrm{d}s.\end{aligned}$$

Therefore, by first taking the square on both sides of the above identity and then taking the expectation, we get

$$\begin{aligned}\mathbb{E}\left[\left|\mathcal{H}\hat{\rho}_t - \bar{H}_t\right|^2\right] \le &5\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_s^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\mathrm{d}y_s\right|^2 \\ &+ 5\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}G_s\left(y_s\right)\mathrm{d}s\right|^2 \\ &+ 5\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_s^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}h_s\left(y_s\right)\mathrm{d}s\right|^2 \\ &+ 5\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_\infty^{\mathcal{H}} - \hat{C}_s^{\mathcal{H}}\right)\Gamma^{-2}\mathcal{H}\hat{\rho}_s\mathrm{d}s\right|^2 \\ &+ 5\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_\infty^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\bar{H}_s\mathrm{d}s\right|^2.\end{aligned}$$

The second line above follows the same argument as in Lemma 10 and using Cauchy–Schwarz inequality

$$\begin{aligned}&\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}G_s\left(y_s\right)\mathrm{d}s\right|^2 \\ &\le \int_0^t \left\|e^{\frac{1}{2}(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\right\|^2\mathrm{d}s\,\mathbb{E}\int_0^t\left|e^{\frac{1}{2}(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}G_s\left(y_s\right)\right|^2\mathrm{d}s. \\ &\le \lambda_m^{-1}\mathbb{E}\int_0^t e^{-\lambda_m(t-s)}\left|G_s\left(y_s\right)\right|^2\mathrm{d}s \le Ke^{-\lambda_1 t} \to 0.\end{aligned}$$

Similar results can be achieved for line three to five following the convergence (or exponential convergence) of the integrants. Finally, for the first line using the observation equation in (21), that is, $\mathrm{d}y_t = \left[\mathcal{H}\rho_t + h_t\left(y_t\right)\right]\mathrm{d}t + \Gamma\mathrm{d}B_t$, there is

$$\begin{aligned}&\mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_s^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\mathrm{d}y_s\right|^2 \\ &\le \mathbb{E}\left|\int_0^t e^{(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left(\hat{C}_s^{\mathcal{H}} - C_s^H\right)\Gamma^{-2}\left[\mathcal{H}\rho_t + h_t\left(y_t\right)\right]\mathrm{d}s\right|^2 \\ &+\mathbb{E}\int_0^t e^{2(t-s)\left(L_\infty^m - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2}\right)}\left\|\hat{C}_s^{\mathcal{H}} - C_s^H\right\|^2\mathrm{d}s \\ &\le K_1\int_0^t e^{-\lambda_m(t-s)}\left(\mathbb{E}\left\|\hat{C}_s^{\mathcal{H}} - \hat{C}_\infty^{\mathcal{H}}\right\|^2 + \mathbb{E}\left\|C_s^H - \hat{C}_\infty^{\mathcal{H}}\right\|^2\right)\mathrm{d}s \\ &\le Ke^{-\lambda_1 t} \to 0.\end{aligned}$$

Above in the second to the last inequality, we use the uniform boundedness of $\mathbb{E}\left|\mathcal{H}\rho_t\right|$ and $\mathbb{E}\left|h_t\left(y_t\right)\right|$ as $t \to \infty$, and the last line uses the convergence (or exponential convergence) of the covariance (50) or (51). Combining all the above bounds, we finally get the convergence in (53) and (54). □

Together with the equilibrium estimate in (52) combining the result in Theorem 11, we can also get the same error estimate compared with the target field as $t \to \infty$

$$\mathbb{E}\left[\left|\mathcal{H}\rho_t - \bar{H}_t\right|^2\right] \le \mathbb{E}\left[\left|\mathcal{H}\rho_t - \mathcal{H}\hat{\rho}_t\right|^2\right] + \mathbb{E}\left[\left|\mathcal{H}\hat{\rho}_t - \bar{H}_t\right|^2\right] \to \mathrm{tr}\hat{C}_\infty^{\mathcal{H}}.$$

Thus, we show that the statistical filtering model solution $\bar{H}_t$ converges to the optimal filter solution with the same mean square error, and exponential convergence is reached if the forecast model has exponential convergence to the equilibrium.

As a final comment, we can further relax Assumption 9 as there exist deterministic uniformly continuous functions, $L_\infty^m(y)$ and $L_\infty^v(y)$, so that for any $y_t \to y_\infty$, there are

$$\begin{aligned}&\left\langle \mathcal{L}_t\left(y_t\right)H\hat{\rho}_t\right\rangle \to L_\infty^m\left(y_\infty\right)\mathcal{H}\hat{\rho}_\infty, \ \left\langle \mathcal{L}_tH\left(\hat{C}_tH^\intercal\right)\right\rangle \to L_\infty^v\left(y_\infty\right)\hat{C}_\infty^{\mathcal{H}}, \\ &\left\langle \mathcal{L}_t\left(y_t\right)H\tilde{\rho}_t\right\rangle \to L_\infty^m\left(y_\infty\right)\mathcal{H}\tilde{\rho}_\infty, \ \left\langle \mathcal{L}_tH\left(\tilde{\rho}_tH'^\intercal\right)\right\rangle \to L_\infty^v\left(y_\infty\right)C_\infty^H,\end{aligned} \tag{55}$$

a.s. as $t \to \infty$. And the limiting matrices are uniformly bounded by negative-definite matrices A_m, A_v

$$L_\infty^m(y) - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2} \le A_m \prec 0, \quad L_\infty^v(y) - \hat{C}_\infty^{\mathcal{H}}\Gamma^{-2} \le A_v \prec 0. \tag{56}$$

In addition, non-zero noise in (36) can be included satisfying $\Sigma_t \to 0$ as $t \to \infty$. Then, the same result applies for the convergence of the observation mean and covariance at the long-time limit as in Lemma 10 and Theorem 11.

5. Ensemble approximation of the statistical filtering model

Finally, we discuss numerical implementations of ensemble methods for solving the statistical filtering model (36) with discrete observations and explicit filtering operators in analysis update.

5.1. Numerical algorithm for implementing the approximating filter

Assume that the observation data comes at discrete times $t_n = n\delta$ with a constant observation frequency δ. We can linearly interpolate $y_n = \left(\bar{u}_{t_n}, R_{t_n}\right) \in \mathbb{R}^p$ as

$$\frac{\mathrm{d}y_t^\delta}{\mathrm{d}t} = \frac{y_{n+1} - y_n}{t_{n+1} - t_n} = \frac{\Delta y_{n+1}}{\delta}, \tag{57}$$

during time interval $t \in [t_n, t_{n+1}]$. We propose an ensemble algorithm to approximate the filtering distribution $\tilde{\rho}_{t_n} \sim \tilde{Z}_{t_n}$ conditional on the statistical observations $Y_t^\delta = \{y_s^\delta, s \le t\}$ based on the statistical filter Eq. (36). First, N independent particles, $\tilde{\mathbf{Z}}_t = \{\tilde{Z}_t^{(i)}\}_{i=1}^N$, are drawn to sample the initial distribution of the stochastic state. Then, the particles are evolved according to the following SDE with drift terms a_t^m, a_t^v and control gains K_t^m, K_t^v

$$\begin{aligned}
\mathrm{d}\tilde{Z}_t^{(i)} =\ & L(\bar{u}_t^N)\tilde{Z}_t^{(i)}\mathrm{d}t + \Gamma\left(\tilde{Z}_t^{(i)}\tilde{Z}_t^{(i)\intercal} - R_t^N\right)\mathrm{d}t + \Sigma_t \mathrm{d}\tilde{W}_t^{(i)} \\
&+ a_t^m\left(\tilde{Z}_t^{(i)}\right)\mathrm{d}t + K_t^m\left(\tilde{Z}_t^{(i)}\right)\left\{\mathrm{d}\bar{u}_t^\delta - \left[H^m\left(\tilde{Z}_t^{(i)}\right) + h_{m,t}(\bar{u}_t^\delta)\right]\mathrm{d}t - \Gamma_{m,t}\mathrm{d}\tilde{B}_{m,t}^{(i)}\right\} \\
&+ a_t^v\left(\tilde{Z}_t^{(i)}\right)\mathrm{d}t + K_t^v\left(\tilde{Z}_t^{(i)}\right)\left\{\mathrm{d}R_t^\delta - \left[H^v\left(\tilde{Z}_t^{(i)}\right) + h_{v,t}(\bar{u}_t^\delta, R_t^\delta)\right]\mathrm{d}t - \Gamma_{v,t}\mathrm{d}\tilde{B}_{v,t}^{(i)}\right\},
\end{aligned} \tag{58}$$

where the expressions for h_m, h_v are defined in (23), H^m, H^v are defined in (24), and $\tilde{B}_{m,t}^{(i)}, \tilde{B}_{v,t}^{(i)}$ are independent white noises. Above, in the first line of (58) for the forecast step of the filter, the first two moments $(\bar{u}_t^N, R_t^N)$ can be explicitly solved by the statistical equations according to the stochastic-statistical model (11)

$$\begin{aligned}
\frac{\mathrm{d}\bar{u}_t^N}{\mathrm{d}t} &= M(\bar{u}_t^N) + F_t + \mathbb{E}^N\left[H^m(\tilde{\mathbf{Z}}_t)\right], \\
\frac{\mathrm{d}R_t^N}{\mathrm{d}t} &= L(\bar{u}_t^N)R_t^N + R_t^N L(\bar{u}_t^N)^\intercal + \Sigma_t\Sigma_t^\intercal \\
&\quad + \mathbb{E}^N\left[H^v(\tilde{\mathbf{Z}}_t)\right] + \epsilon^{-1}\left(\mathbb{E}^N\left[\tilde{\mathbf{Z}}_t\tilde{\mathbf{Z}}_t^\intercal\right] - R_t^N\right).
\end{aligned} \tag{59}$$

The expectation is computed by the empirical average, $\mathbb{E}^N f(\tilde{\mathbf{Z}}) = \frac{1}{N}\sum_{i=1}^N f(\tilde{Z}^{(i)})$. In this way, the particle simulation of (58) can be carried out easily for each individual sample $\tilde{Z}_t^{(i)}$, and the dependence on the distribution of the whole interacting particles is only introduced through the empirical average in the statistical Eqs. (59). We summarize the ensemble filtering strategy in Algorithm 1.

Algorithm 1 Ensemble statistical filter with observations in mean and covariance

Model Setup: Get the interpolated sequence of statistical observations Eq. (57) $y_t^\delta = \{\bar{u}_t^\delta, R_t^\delta\}, t \in [0, T]$; determine time integration step τ, and the initial state distribution ρ_0.

Initial condition: Draw an ensemble of samples $\{\tilde{Z}_0^{(i)}\}_{i=1}^N$ from the initial distribution ρ_0, and compute the initial mean and covariance $\{\bar{u}_0^N, R_0^N\}$ w.r.t. ρ_0.

1: **for** $n = 0$ while $n < \lfloor T/\tau \rfloor$, during the time interval $t \in [t_n, t_{n+1}]$ with $t_n = n\tau$ **do**
2: Compute the gain functions $K_{t_n}^m$ and $K_{t_n}^v$ using Eq. (62) and the associated drift functions $a_{t_n}^m$ and $a_{t_n}^v$.
3: Update the samples $\{\tilde{Z}_{t_{n+1}}^{(i)}\}_{i=1}^N$ using Eq. (58) with the statistical states $\{\bar{u}_{t_n}^N, R_{t_n}^N\}$ and observation data y_t^δ.
4: Update the statistical mean and covariance $\{\bar{u}_{t_{n+1}}^N, R_{t_{n+1}}^N\}$ by integrating Eq. (59) to the next time step using the empirical average of all samples.
5: **end for**

Remark. Solving the Eqs. (58) may still demand high computational cost for resolving the multiple nonlinear coupling terms in high dimension $d \gg 1$. One potential approach to address the computational challenge is to adopt the efficient random batch approach [41,44] developed for the coupled models (11). A detailed investigation of efficient numerical methods will be performed in the follow-up research.

5.2. Construction of explicit model operators in the analysis step

In the second and third lines of (58) for the analysis step update of filtering, we still need to propose explicit expressions for model parameters a_t^m, K_t^m and a_t^v, K_t^v according to the observations of the mean and covariance respectively. According to Theorem 8, the gain function K_t needs to be solved from Eq. (41), that is,

$$-\nabla\cdot\left(K_t^\intercal\tilde{\rho}_t\right) = \tilde{\rho}_t\Gamma_t^{-2}\left(H(z) - \bar{H}_t\right).$$

Then the drift function a_t can be directly computed from the solution of K_t as

$$a_t = \nabla\cdot\left(K_t\Gamma_t^2K_t^\intercal\right) - K_t\Gamma_t^2\nabla\cdot K_t^\intercal.$$

In general, it is still difficult to find solutions of the above equations. By multiplying H on both sides and integrating about z, the identity for K_t implies a necessary condition

$$\tilde{\mathbb{E}}\left[K_t^\intercal\nabla H\right] = \Gamma_t^{-2}C_t^H, \tag{60}$$

where $C_t^H = \tilde{\mathbb{E}}\left[\left(H(\tilde{Z}_t) - \bar{H}_t\right)\left(H(\tilde{Z}_t) - \bar{H}_t\right)^\intercal\right]$ is the covariance of H. Therefore, we can first design proper gain functions K_t by solving (60) according to the specific structures of $H^m : \mathbb{R}^d \to \mathbb{R}^d$ and $H^v : \mathbb{R}^d \to \mathbb{R}^{d^2}$ required in our problem in (24)

$$\begin{aligned}
H_k^m(z) &= \sum_{m,n}\gamma_{kmn}z_mz_n = z^\intercal A_kz, \\
H_{kl}^v(z) &= \sum_{m,n}\gamma_{kmn}z_mz_nz_l + \gamma_{lmn}z_mz_nz_k = \left(z^\intercal A_kz\right)z_l + z_k\left(z^\intercal A_lz\right),
\end{aligned} \tag{61}$$

for all $1 \le k, l \le d$ where we rewrite the quadratic and cubic functions using the symmetric coefficient matrix, $A_k^\intercal = A_k \in \mathbb{R}^{d\times d}$, satisfying the assumed structural symmetry in the coupling coefficient γ_{kmn}. The resulting gain functions are then constructed with the following specific expressions.

Proposition 12. *The matrix-valued gain functions $K_t^m = \tilde{K}^m\Gamma_{m,t}^{-2}$ and $K_t^v = \tilde{K}^v\Gamma_{v,t}^{-2}$ with $\tilde{K}^m(z) \in \mathbb{R}^{d\times d}$ and $\tilde{K}^v(z) \in \mathbb{R}^{d\times d^2}$ in the following expressions*

$$\begin{aligned}
\tilde{K}_{j,k}^m(z) &= \frac{1}{2}z_j\left[H_k^m(z) - \bar{H}_k^m\right], \\
\tilde{K}_{j,kl}^v(z) &= \frac{1}{3}z_j\left[H_{kl}^v(z) - \bar{H}_{kl}^v\right].
\end{aligned} \tag{62}$$

for $1 \le k, l \le d$ and $1 \le j \le d$ satisfy Eq. (60) according to the structures of the functions H^m and H^v in the forms (61) respectively, and $\bar{H}^m = \tilde{\mathbb{E}}\left[H^m(\tilde{Z})\right]$ and $\bar{H}^v = \tilde{\mathbb{E}}\left[H^v(\tilde{Z})\right]$.

The proof of Proposition 12 is put in Appendix B. The average terms, $\bar{H}^m, \bar{H}^v$, are already computed in the statistical Eqs. (59) thus no additional computational cost is needed. On the other hand, it is noticed that (62) can only give a necessary condition for the gain operators and may not guarantee the original identity for K_t in general. However, in the proof of Theorem 8, it shows that (60) is the main relation needed to derive the consistent analysis statistics on the mean of $H(\tilde{Z}_t)$. Therefore, (62) provides a desirable candidate for practical implementations of the algorithm concerning the consistency in the leading moments.

6. Numerical tests on prototype models

In this section, we test the performance of the proposed filtering algorithm on simple but nevertheless instructive prototype models. Though relatively low-dimensional, these models can demonstrate a wide variety of different statistical regimes, making desirable first experiments for confirming the skill of new filtering strategies.

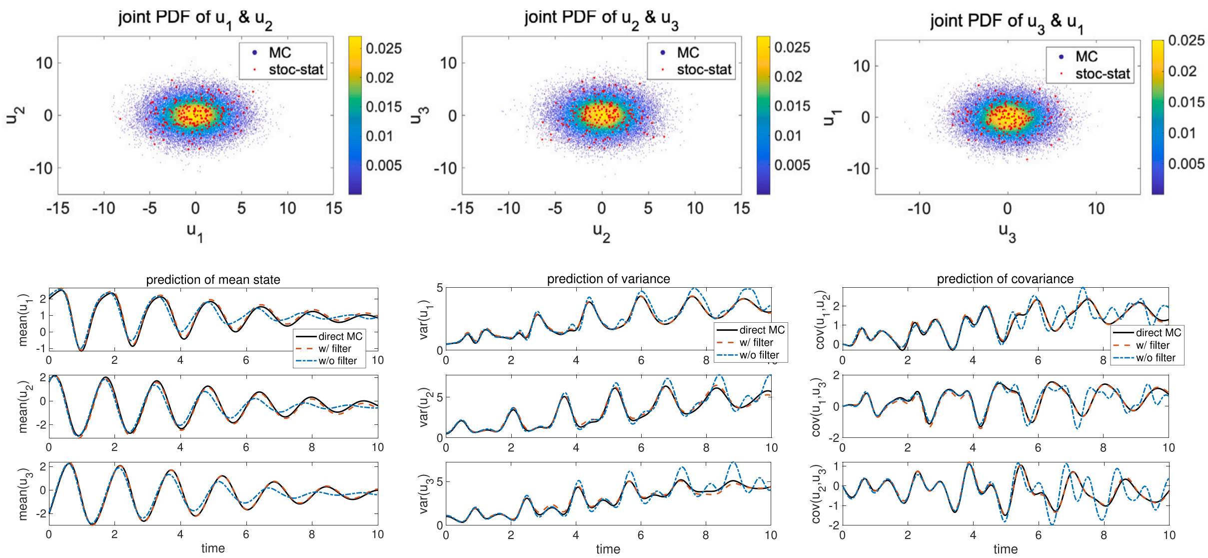

Fig. 2. Filtering performance in the near-Gaussian region of the triad model with the sample size $N_1 = 100$.

6.1. Numerical setup of the triad model

The triad model [7,36] is given by a three-dimensional ODE system about $\mathbf{u} = (u_1, u_2, u_3)^{\intercal}$ with a quadratic nonlinear coupling term

$$\begin{aligned}
\frac{du_1}{dt} &= L_2 u_3 - L_3 u_2 - d_1 u_1 + B_1 u_2 u_3 + \sigma_1 \dot{W}_1,\\
\frac{du_2}{dt} &= L_3 u_1 - L_1 u_3 - d_2 u_2 + B_2 u_3 u_1 + \sigma_2 \dot{W}_2,\\
\frac{du_3}{dt} &= L_1 u_2 - L_2 u_1 - d_3 u_3 + B_3 u_1 u_2 + \sigma_3 \dot{W}_3.
\end{aligned} \tag{63}$$

The above system (63) fits into the general formulation (1) and the nonlinear coupling term is energy preserving if $B_1 + B_2 + B_3 = 0$. The triad system constitutes an elementary building block for more general turbulent systems as a three-mode Galerkin projection of high-dimensional dynamics with energy conserving nonlinear interactions [6].

Though simple in appearance, the triad system (63) has representative statistical features including energy cascade between modes and internal instabilities that can be created by choosing the model parameters. In particularly, we can generate distinct statistical features from Gaussian to highly non-Gaussian PDFs in two typical dynamical regimes:

- *Region I: equipartition of energy*. Gaussian equilibrium distribution, $p_{\mathrm{eq}} \sim \exp\left(-\frac{1}{2}\sigma_{\mathrm{eq}}^{-2}\mathbf{u}\cdot\mathbf{u}\right)$, will be reached under this set-up by setting the equipartition of energy in model parameters, $\frac{\sigma_1^2}{2d_1} = \frac{\sigma_2^2}{2d_2} = \frac{\sigma_3^2}{2d_3} = \sigma_{\mathrm{eq}}^2$.
- *Region II: cascade of energy*. In this case, energy is injected in the strongly forced first mode u_1 while the other two strongly damped modes u_2, u_3 are less energetic, inducing energy cascades through nonlinear coupling.

Region I is the simplest but nevertheless representative with near-Gaussian statistics. The higher-order moment effects are relatively small in the final equilibrium state, while the nonlinear dynamics still produces dominantly non-Gaussian statistics as the system evolves from initial state. Region II contains important third-order interactions, and large errors will be introduced if the unobserved cross-covariances are ignored without care. The model parameters of the two test regions are listed in Table 1.

Table 1
Model parameters of the triad model in two test regions.

Region	(L_1, L_2, L_3)	(d_1, d_2, d_3)	(B_1, B_2, B_3)	$(\sigma_1, \sigma_2, \sigma_3)$
I	$(3, -2, -1)$	$(0.02, 0.01, 0.01)$	$(1, -0.5, -0.5)$	$2.5 \times \left(\sqrt{2d_1}, \sqrt{2d_2}, \sqrt{2d_3}\right)$
II	$(0, 0, 0)$	$(0.1, 0.2, 0.2)$	$(1, -0.5, -0.5)$	$\left(\sqrt{10}, 0.1, 0.1\right)$

The true statistical features of the triad system in the above dynamical regimes are resolved through direct Monte-Carlo simulations. We run an ensemble of $N = 5 \times 10^4$ particles, which shall be enough for capturing the essential non-Gaussian statistics in a three-dimensional phase space. A fourth-order Runge–Kutta scheme with time step $\Delta t = 1\times10^{-3}$ is used to integrate the system in time, and the stochastic forcing is simulated through the standard *Euler–Maruyama* scheme. The initial ensemble is chosen from a standard Gaussian random sampling.

6.2. Filtering performance in different statistical regions of the triad model

In the statistical forecast problem, we aim to capture the probability evolution of the model state (u_1, u_2, u_3) using a small ensemble size $N_1 = 100$. The statistical filtering scheme can be directly applied to the triad model according to the numerical formulation (58) and (59). Observation data is drawn from the leading moments of the mean $(\bar{u}_1, \bar{u}_2, \bar{u}_3)$ and variance (r_1, r_2, r_3). Notice that due to the nonlinear coupling structure in the triad model (63), the accurate prediction of the mean and covariance in (18) relies on the estimation of the cross-covariance and third-moments which are all not included in the observation data.

6.2.1. Filtering with near-Gaussian statistics

First, we consider the filter performance in Region I with near-Gaussian statistics. Scatter plots of joint two-dimensional distributions from direct MC simulations with a large ensemble are shown in the first row of Fig. 2 with different colors indicating the density of the samples. The joint distributions of states show the near-Gaussian PDFs. Nevertheless, large errors will be gradually developed by running the statistical equations with a very small sample size due to the insufficient estimation of the expectations in (59) which will be amplified in time. This will lead to the gradual increase of errors in the prediction of mean, variance, and cross-covariance as shown in the second panel of Fig. 2. In contrast, stability and accuracy are maintained using the

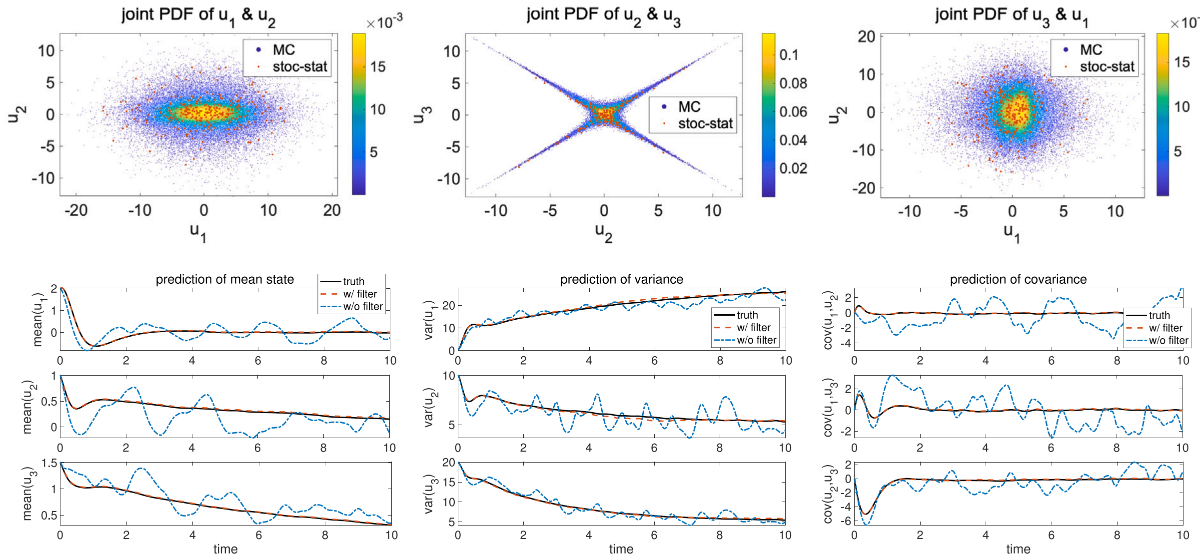

Fig. 3. Filtering performance in the non-Gaussian region of the triad model with the sample size $N_1 = 100$.

filter scheme during the entire evolution of the solutions. Especially, the unobserved cross-covariances between the modes play an important role in the accurate prediction of the mean and covariance. The divergence of the direct method without filter is induced by the accumulated errors in the cross-covariances, while the filter effectively stabilizes the unobserved statistics and guarantees accurate prediction.

6.2.2. Filtering with non-Gaussian statistics

Next, in Region II with nonlinear energy cascade, non-Gaussian distributions with extreme events can be observed in the PDFs especially with a star-shaped joint-distribution as illustrated in the upper panel of Fig. 3. This demonstrates strong nonlinear effects between the modes and a more challenging case for accurate statistical forecasts. Note that the high non-Gaussianity in u_2, u_3 can affect the final structure in the dominant mode u_1 despite its closer to Gaussian marginal distribution. This illustrates important contributions of third-order moments in this case for accurate predictions. Due to this strong reliance on the accurate estimation of higher moments, the small ensemble prediction without the filter will quickly degenerate from the initial time as observed in the lower panel of Fig. 3. The growth of large errors is more obvious in the crucial cross-covariances that govern the dynamics of the mean equation. Again, the filter scheme maintains robust accurate prediction against the large errors induced by the small ensemble estimates of crucial higher-order moments. The accurate prediction and long-time stability shown in the tests involving strong nonlinearity and non-Gaussian statistics demonstrate high skill of the statistical filter enabling efficient long-time probabilistic forecasts of non-Gaussian statistics.

6.3. Filtering performance in the Lorenz 96 model

As a further test on higher dimensional systems, we apply the filtering algorithm on the Lorenz 96 (L-96) model [62] as another prototype model to examine data assimilation schemes

$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = -u_j + \left(u_{j+1} - u_{j-2}\right) u_{j-1} + F, \; j = 1, \ldots, J, \tag{64}$$

with periodic boundary condition $u_{J+1} = u_1$ and constant uniform forcing F. Various representative statistical features can be found in the L-96 solution [63]. Notice that by taking the dimension of the system $J = 3$, the L-96 Eq. (64) shares similar dynamical structures as the triad system (63) with homogeneous linear terms and energy-conserving quadratic nonlinear coupling.

In the numerical tests here, we adopt the constant forcing $F = 6$ that demonstrates strong non-Gaussian statistics in the state solution u_j, and a higher dimension $J = 10$ to test the filter performance. The true statistical solution of the L-96 model (64) is resolved using a sufficiently large ensemble size $N = 5 \times 10^4$. The model state starts with an initial distribution with independent Gaussian distribution with small variances, while the internal instability will rapidly amplify the uncertainty among the modes. By applying the filtering method on the L-96 model containing a large number of internal unstable modes, we aim to capture the key model statistics in a high dimension $J = 10$ using a small number of samples $N_1 = 100$. The prediction results for the mean and variances captured by the particles are plotted in Fig. 4. Similar to the triad model results, the direct numerical prediction using a small ensemble size fails to capture the statistics showing large fluctuation errors and diverging solution. On the other hand, the predictions of mean and variances in the filtering model stay accurate during the entire evolution time against the strong inherent instability in the Lorenz system.

7. Summarizing discussions

We developed a systematic statistical filtering strategy that enables effective ensemble approximation of non-Gaussian probability distributions of multiscale turbulent states using observations in the leading-order moments. The filtering model is based on a closed stochastic-statistical formulation established for modeling general turbulent dynamical systems involving nonlinear coupling. Statistical observations in the first two moments are used to improve the accuracy in capturing crucial non-Gaussian statistics in filtering updates. In practical implementation of the framework, white noise corrections are introduced to represent model errors from the finite ensemble approximation and incomplete observation data such as only statistics in large scales is allowed. There are several potential approaches to obtain the statistical observations depending on the problems to be applied, including: (i) taking local spatial averages to get large-scale statistics; (ii) using time averages in a short time window in the trajectories; (iii) coarse-grained simulations from multiple imperfect low-order approximation models; and (iv) approaches using linear response theory and reduced-order methods to extract leading-order statistics from equilibrium measures. The non-Gaussian features then

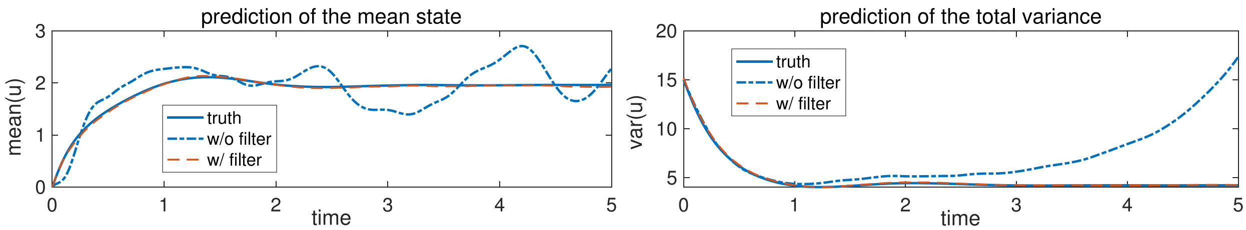

Fig. 4. Filtering performance in the 10-dimensional L-96 model with the sample size $N_1 = 100$.

can be characterized by a McKean–Vlasov SDE taking into account both the stochastic forecast equation and corrections according to the observed first two moments. Importantly, the SDE model for the finite-dimensional stochastic state does not require the explicit computation of the infinite-dimensional probability distribution, but just relies on the feedbacks from the leading moments that can be computed from the associated statistical equations. This leads to straightforward numerical algorithms using ensemble approximations. The new filtering model offers a flexible approach to recover essential statistics, making it applicable to a wide range of problems in uncertainty quantification and data assimilation.

Limitations of the statistical filtering model and future research directions. Still, many interesting problems remain open in both rigorous mathematical analysis of the approximate filtering model and the practical computational strategies for realistic applications. In this paper, we are only able to show the statistical consistency of the approximate filter in the first two moments and using full observation function under restricted assumptions. Further explorations exploiting specific model structures such as the conservation properties will be used to provide a thorough understanding of the approximation skill of the filter predictions. In the immediate applications of this research, the performance of the new filtering strategy need to be tested on typical nonlinear systems with close realistic relevance. We plan to perform systematic numerical experiments on a series of turbulent systems, starting from prototype models to realistic applications in really high-dimensional systems. Using the simple prototype models [54], we are able to carry out a systematic numerical investigation of the filtering scheme for different non-Gaussian features generated by the nonlinear dynamics. On the other hand, to deal with the computational challenge in high-dimensional systems, the current filtering scheme needs to be combined with additional model reduction strategies such as the random batch approximations [41,44] that have been shown a promising way to compute high dimensional problems. Furthermore, we would like to point out that even though the results in this paper are mostly focusing on continuous observations, the statistical filtering scheme described in Algorithm 1 can be adapted for discrete observation data y_t and can only include the moments in the leading modes. This is particularly suitable for many practical situations where only large-scale observations in discrete times are available. We plan to perform detailed numerical performance comparison based on different types of observations in the follow-up research [54].

CRediT authorship contribution statement

Di Qi: Validation, Methodology, Investigation, Conceptualization, Writing – original draft. **Jian-Guo Liu:** Investigation, Writing – original draft, Validation, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under Grant No. DMS-2424139, while J.-G. L. was in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the Fall 2025 semester. J.-G. L. is partially supported by the NSF Grant DMS-2106988. D. Q. is partially supported by ONR Grant N00014-24-1-2192 and NSF Grant DMS-2407361.

Appendix A. General backgrounds about filtering

We summarize useful results needed in the main text of the paper mostly following the Refs. [55,58,64].

A.1. Filtering equations for general stochastic systems

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the complete probability space. The *signal process* $u_t \in \mathbb{R}^d$ and the *observation process* $y_t \in \mathbb{R}^p$ are defined on the probability space satisfying the following SDEs

$$\mathrm{d}u_t = F\left(u_t\right)\mathrm{d}t + \Sigma \mathrm{d}W_t, \quad u_{t=0} = u_0, \tag{A.1a}$$

$$\mathrm{d}y_t = H\left(u_t\right)\mathrm{d}t + \mathrm{d}B_t, \quad y_{t=0} = y_0, \tag{A.1b}$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ and $H : \mathbb{R}^d \to \mathbb{R}^p$ are bounded and globally Lipschitz continuous functions, and $W_t \in \mathbb{R}^s, B_t \in \mathbb{R}^p$ are independent standard Wiener processes with matrix coefficient $\Sigma \in \mathbb{R}^{d\times s}$. The aim of the general filtering problem is to determine the conditional probability distribution μ_t of the signal process, u_t, given the accumulated observation process, $Y_t = \left\{y_s, s \le t\right\}$.

Define the observation filtration $\mathcal{G}_t = \sigma\left\{y_s, s \le t\right\}$. The random conditional distribution $\mu_t : \mathbb{R}^d \times \Omega \to [0,1]$ is defined as the $\mathcal{P}\left(\mathbb{R}^d\right)$-valued stochastic process which is measurable w.r.t. $\mathcal{G}_t$, so that for any function $\varphi \in C_b^2\left(\mathbb{R}^d\right)$ a.s.

$$\mathbb{E}\left[\varphi\left(u_t\right) \mid \mathcal{G}_t\right] = \mu_t(\varphi) := \int_{\mathbb{R}^d} \varphi(u)\, \mu_t(\mathrm{d}u).$$

In particular, the optimal filter solution, $\hat{u}_t = \mathbb{E}\left[u_t \mid \mathcal{G}_t\right]$, can be defined based on μ_t. It shows that $\hat{u}_t$ is the minimizer $\mathbb{E}\left[\left|\hat{u}_t - u_t\right|^2\right] = \min_v \mathbb{E}\left[\left|v - u_t\right|^2\right]$ among all $v \in L^2\left(\Omega, \mathcal{G}_t, \mathbb{P}\right)$ in the set of $\mathcal{G}_t$-measurable square-integrable random variables for any fixed t. The filtering equation for the conditional probability distribution μ_t is verified to satisfy the Kushner–Stratonovich equation

$$\mathrm{d}\mu_t(\varphi) = \mu_t(\mathcal{L}\varphi)\,\mathrm{d}t + \sigma\left(H, \varphi; \mu_t\right)\mathrm{d}\nu_t. \tag{A.2}$$

On the right hand side of the above equation, the first term is the drift due to the infinitesimal generator $\mathcal{L} = F \cdot \nabla + \frac{1}{2}\Sigma\Sigma^\intercal : \nabla\nabla$ of the signal process (A.1a); the second term represents the correction from the observation process (A.1b). The innovation process, $\mathrm{d}\nu_t = \mathrm{d}y_t - \mu_t(H)\,\mathrm{d}t \in \mathbb{R}^p$, is a $\mathcal{G}_t$-Brownian motion under the probability measure $\mathbb{P}$, and $\sigma\left(H, \varphi; \mu_t\right) = \mu_t\left(\varphi H^\intercal\right) - \mu_t(\varphi)\,\mu_t\left(H^\intercal\right) \in \mathbb{R}^{1\times p}$ gives the coefficient with finite quadratic variation, where $\mu_t(H)$ is the componentwise measure of the vector-valued function H. In addition, the filtering Eq. (A.2) is shown to have a unique solution under proper conditions (Theorem 3.30 and 4.19 in [64] and Theorem 7.7 in [12]) that is also stable

(Theorem 2.7 in [65]). Therefore, this guarantees that the solution of the Kushner–Stratonovich Eq. (A.2) uniquely characterizes the filter distribution μ_t as a $\mathcal{P}\left(\mathbb{R}^d\right)$-valued stochastic process.

Next, assume that the conditional probability μ_t possesses a square integrable density, $\mu_t(\mathrm{d}x) = \rho_t(x)\,\mathrm{d}x$, with respect to the Lebesgue measure. It can be shown under proper conditions (Corollary 7.18 in [20,64]), the conditional probability solution μ_t of (A.1) has a probability density $\rho_t \in W_k^2\left(\mathbb{R}^d\right)$. According to (A.2) for the conditional probability, ρ_t can be found to be the unique solution to the following SPDE

$$\partial_t \rho_t = \mathcal{L}^* \rho_t \mathrm{d}t + \rho_t \left(H - \bar{H}_t\right)^{\intercal} \left(\mathrm{d}y_t - \bar{H}_t \mathrm{d}t\right), \quad \rho_{t=0} = \rho_0, \tag{A.3}$$

where $\bar{H}_t = \int_{\mathbb{R}^d} H(x)\,\rho_t(x)\,\mathrm{d}x$ and $\rho_0 \in L^2\left(\mathbb{R}^d\right)$ is the absolute continuous density of μ_0. The randomness of the above SPDE only comes from the innovation process $\mathrm{d}\nu_t = \mathrm{d}y_t - \mu_t(H)\,\mathrm{d}t$ as a finite-dimensional white noise in time.

At last, if the functions on the right hand sides of (A.1) satisfy the Ornstein–Uhlenbeck processes with matrix coefficients, that is, $F(u) = Fu + f_t$ and $H(u) = Hu + h_t$. With Gaussian initial condition, $u_0 \sim \mathcal{N}\left(\hat{u}_0, C_0\right)$, the conditional distribution $\mu_t = \mathcal{N}\left(\hat{u}_t, C_t\right)$ given $\mathcal{G}_t$ in (A.2) becomes a multivariate normal distribution, where $\hat{u}_t = \mathbb{E}\left[u_t \mid \mathcal{G}_t\right]$ and $C_t = \mathbb{E}\left[\left(u_t - \hat{u}_t\right)\left(u_t - \hat{u}_t\right)^{\intercal} \mid \mathcal{G}_t\right]$. The filtering equations for $\hat{u}_t \in \mathbb{R}^d$ and $C_t \in \mathbb{R}^{d\times d}$ are given by the Kalman–Bucy filter [56]

$$\mathrm{d}\hat{u}_t = \left(F\hat{u}_t + f_t\right)\mathrm{d}t + K_t\left[\mathrm{d}y_t - \left(H\hat{u}_t + h_t\right)\mathrm{d}t\right], \tag{A.4a}$$

$$\dot{C}_t = FC_t + C_t F^{\intercal} - K_t K_t^{\intercal} + \Sigma\Sigma^{\intercal}. \tag{A.4b}$$

with the Kalman gain matrix $K_t = C_t H^{\intercal}$. Above, (A.4a) is an SDE coupled with the deterministic Riccati equation (A.4b).

A.2. Infinite dimensional filtering in Hilbert space

It is shown that the linear Kalman–Bucy filter can be generalized to linear stochastic equations on a Hilbert space [57,58]. Let H be a Hilbert space. Denote $L^2(\Omega,\mathcal{G},\mathbb{P};H)$ as the collection of all H-valued $\mathcal{G}$-measurable square-integrable random variables. The expectation of $u \in L^2(\Omega,\mathcal{G},\mathbb{P};H)$ is denoted by

$$\mathbb{E}[u] = \int_\Omega u(\omega)\,\mathrm{d}\mathbb{P}(\omega). \tag{A.5}$$

The inner product for states $u,v \in L^2(\Omega,\mathcal{G},\mathbb{P};H)$ can be defined as $\langle u,v\rangle_2 = \mathbb{E}\left[\langle u,v\rangle_H\right] = \int_\Omega \langle u(\omega), v(\omega)\rangle_H \,\mathrm{d}\mathbb{P}(\omega)$. With the above notations, the *covariance operator* C can be introduced as an element in the linear transformations $\mathcal{L}(H;H)$.

Definition 13. Let $u,v \in L^2(\Omega,\mathcal{G},\mathbb{P};H)$ be two H-valued random variables. Then the covariance of u and v is given by

$$C(u,v) = \mathbb{E}[u\otimes v] - \mathbb{E}[u]\otimes\mathbb{E}[v], \tag{A.6}$$

where $u(\omega)\otimes v(\omega) \in \mathcal{L}(H;H)$ is a linear transformation of H into H defined for any $f \in H$ as

$$(u\otimes v)\,f = u\,\langle v,f\rangle_H\,.$$

It is easy to check that the adjoint $C(u,v)^* = C(v,u)$ and $C(u,u)^* = C(u,u)$ is self-adjoint since

$$\langle f,(u\otimes v)\,g\rangle_H = \langle u,f\rangle_H\,\langle v,g\rangle_H$$

Notice that if $H = \mathbb{R}^d$ is finite-dimensional, for any $x,y \in L^2(\Omega,\mathcal{G},\mathbb{P};\mathbb{R}^d)$, $x\otimes y = xy^{\intercal} \in \mathbb{R}^{d\times d}$, then the covariance $C(x,y) \in \mathbb{R}^{d\times d}$ becomes the $d\times d$ matrix

$$C(x,y) = \mathbb{E}\left[xy^{\intercal}\right] - \mathbb{E}[x]\,\mathbb{E}[y]^{\intercal} = \mathbb{E}\left[(x-\mathbb{E}[x])\,(y-\mathbb{E}[y])^{\intercal}\right].$$

Then, we call the $u_t(\omega)$ from $[0,T]\times\Omega$ to H an H-valued stochastic process. An infinite-dimensional H-valued Wiener process W_t can be defined accordingly and the Itô integral can be generalized to infinite-dimensional Hilbert space accordingly (see Chapter 2 of [5] with precise validations). Therefore, the signal process u_t of filtering can be given by the following H-valued SDE

$$\mathrm{d}u_t = \mathcal{A}_t u_t \mathrm{d}t + \mathcal{Q}_t \mathrm{d}W_t, \quad u_{t=0} = u_0, \tag{A.7}$$

where $\mathcal{A}_t \in L^\infty([0,T];\mathcal{L}(H;H))$ is a regulated mapping of $[0,T]$ into $\mathcal{L}(H;H)$ (which is further generalized to unbounded operators in [66]), $\mathcal{Q}_t \in L^2([0,T];\mathcal{L}(H;H))$, and W_t is the H-valued Wiener process. It can be shown (Theorem 2.13 in [5] and Theorem 5.1 in [58]) that (A.7) has unique solution in $L^2(\Omega,\mathcal{G},\mathbb{P};H)$ with initial value $\mathbb{E}\left[\left|u_0\right|_H^2\right] < \infty$. The observation stochastic process $y_t \in \mathbb{R}^p$ can be generated by the SDE

$$\mathrm{d}y_t = \mathcal{H}_t u_t \mathrm{d}t + \mathrm{d}B_t, \quad y_{t=0} = y_0, \tag{A.8}$$

where $\mathcal{H}_t$ is a linear mapping of $[0,T]$ into $\mathcal{L}(H;\mathbb{R}^p)$, and B_t is the Wiener process in $\mathbb{R}^p$ independent of W_t. The infinite-dimensional filtering problem can be then described as: given $y_s, s\le t$, determine the optimal estimate $\hat{u}_t$ of u_t that minimizes

$$\mathbb{E}\left[\left|u_t - v\right|_H^2\right], \quad v \in L^2\left(\Omega,\mathcal{G}_t,\mathbb{P};H\right), \tag{A.9}$$

where $\mathcal{G}_t = \sigma\left\{y_s, s\le t\right\}$ is generated by the observations up to time t.

Finally, in parallel to the Kalman–Bucy filter (A.4) in finite-dimensional space, similar result can be extended to the above infinite-dimensional filtering problem (A.7) and (A.8). Below, we summarize the main results in Theorem 7.10, 7.14 of [58].

Theorem 14. *The optimal filter solution $\hat{u}_t$ of (A.9) exists and is unique, which satisfies the following infinite-dimensional SDE*

$$\mathrm{d}\hat{u}_t = \mathcal{A}_t\hat{u}_t\mathrm{d}t + \mathcal{K}_t\left(\mathrm{d}y_t - \mathcal{H}_t\hat{u}_t\mathrm{d}t\right), \quad \hat{u}_{t=0} = u_0, \tag{A.10}$$

where $\mathcal{K}_t = C_t\mathcal{H}_t^* \in \mathcal{L}(\mathbb{R}^p;H)$ with $\mathcal{H}_t^*$ the adjoint of $\mathcal{H}_t$. And the covariance operator $C_t = \mathbb{E}\left[\left(u_t-\hat{u}_t\right)\otimes\left(u_t-\hat{u}_t\right)\right]$ satisfies the following Riccati equation

$$\dot{C}_t = \mathcal{A}_t C_t + C_t\mathcal{A}_t^* - \mathcal{K}_t\mathcal{K}_t^* + \mathcal{Q}_t\mathcal{Q}_t^*, \quad C_{t=0} = C\left(u_0,u_0\right). \tag{A.11}$$

A.3. Kalman–Bucy filter with conditional Gaussian processes

The linear Kalman–Bucy filter can be generalized to nonlinear filtering accepting the conditional Gaussian processes [55]. The conditional Gaussian process $\left(v_t,y_t\right), 0\le t\le T$ is given by the solution of the following coupled equations

$$\mathrm{d}v_t = \left[F_t\left(y_t\right)v_t + f_t\left(y_t\right)\right]\mathrm{d}t + \Sigma_t\mathrm{d}W_t, \quad v_{t=0} = v_0, \tag{A.12a}$$

$$\mathrm{d}y_t = \left[H_t\left(y_t\right)v_t + h_t\left(y_t\right)\right]\mathrm{d}t + \Gamma_t\mathrm{d}B_t, \quad y_{t=0} = y_0, \tag{A.12b}$$

where W_t, B_t are mutually independent standard Gaussian white noise processes, and the initial states $\left(v_0,y_0\right)$ are random variables independent of W_t, B_t. In general, $v_t \in \mathbb{R}^d$ represents the signal process and $y_t \in \mathbb{R}^p$ represents the observation process. The functions f_t, F_t and h_t, H_t are globally Lipschitz continuous and uniformly bounded on the observed state y_t over the time interval $0\le t\le T$. Assume that the sequence $\left(v_t,y_t\right)$ is obtained from a realization ω and the initial condition, $\left\{v_0,y_0\right\}$. We can then define the observation sequence $Y_t = \left\{y_s(\omega), s\le t\right\}$ as well as the unobserved signal $v_t = v_t(\omega)$. The above system (A.12) is called the conditional Gaussian process since the conditional distribution $\mu_t = \mathbb{P}\left(v_t\in\cdot\mid Y_t\right)$ given Y_t becomes a Gaussian distribution a.s. if $\mu_0 = \mathbb{P}\left(v_0\in\cdot\mid y_0\right)$ is Gaussian (Theorem 12.6 of [55]).

Next, let $\mathcal{G}_t = \sigma\left\{y_s, s\le t\right\}$, and define the mean $\hat{v}_t = \mathbb{E}\left[v_t\mid\mathcal{G}_t\right]$ and covariance matrix $\hat{C}_t = \mathbb{E}\left[\left(v_t-\hat{v}_t\right)\left(v_t-\hat{v}_t\right)^{\intercal}\mid\mathcal{G}_t\right]$ w.r.t. the conditional Gaussian distribution $\mu_t = \mathcal{N}\left(\hat{v}_t,\hat{C}_t\right)$. Then, it shows that the explicit dynamical equations for $\left(\hat{v}_t,\hat{C}_t\right)$ can be derived based on the conditional Gaussian process (A.12). As a result, filtering equations from the linear Kalman–Bucy filter (A.4) can be directly applied to the conditional linear system regardless of its essentially nonlinear dynamics. The equations of the conditional Gaussian processes and the uniqueness of the solutions are proved under suitable conditions for the model coefficients in Chapter 12 of [55]. We summarize the results according to Theorem 12.7 of [55] for the nonlinear conditional Gaussian filter.

Theorem 15. *The conditional distribution μ_t of the stochastic processes υ_t given Y_t from (A.12) is Gaussian, $\mathcal{N}\left(\hat{\upsilon}_t, \hat{C}_t\right)$. Then, with $\Gamma_t\Gamma_t^{\intercal} > 0$ and the initial mean and covariance $\hat{\upsilon}_0, \hat{C}_0$, the solutions for the mean $\hat{\upsilon}_t$ and covariance matrix $\hat{C}_t$ are uniquely given by the following closed equations*

$$\begin{aligned} d\hat{\upsilon}_t &= \left[F_t\left(y_t\right)\hat{\upsilon}_t + f_t\left(y_t\right)\right]dt \\ &\quad + K_t\left(y_t\right)\left(\Gamma_t\Gamma_t^{\intercal}\right)^{-1}\left\{dy_t - \left[H_t\left(y_t\right)\hat{\upsilon}_t + h_t\left(y_t\right)\right]dt\right\}, \qquad \text{(A.13a)} \\ d\hat{C}_t &= \left[F_t\left(y_t\right)\hat{C}_t + \hat{C}_tF_t\left(y_t\right)^{\intercal} + \Sigma_t\Sigma_t^{\intercal}\right]dt \\ &\quad - K_t\left(y_t\right)\left(\Gamma_t\Gamma_t^{\intercal}\right)^{-1}K_t^{\intercal}\left(y_t\right)dt, \qquad \text{(A.13b)} \end{aligned}$$

where $K_t\left(y_t\right) = \hat{C}_t\left(y_t\right)H_t\left(y_t\right)^{\intercal}$. The matrix $\hat{C}_t$ will remain positive-definite for all $0 \le t \le T$ if $\hat{C}_0 \succ 0$.

Appendix B. Detailed proofs of theorems in the main text

Proof of Lemma 1. Denote the energy $E_t = \int |u|^2 p_t(u)\,du$. Multiplying the factor $|u|^2$ on both sides of the FPE (2) and integrating in space $u \in \mathbb{R}^d$ yield that

$$\begin{aligned} \frac{dE_t}{dt} &= \int 2u \cdot \left[\Lambda u + B(u,u) + F_t\right] p_t du + \operatorname{tr}\left(\sigma_t\sigma_t^{\intercal}\right) \\ &\le -2\lambda_0 E_t + \bar{u}_t \cdot F_t + \operatorname{tr}\left(\sigma_t\sigma_t^{\intercal}\right) \\ &\le -\lambda_0 E_t + \frac{\left|F_t\right|^2}{\lambda_0} + \operatorname{tr}\left(\sigma_t\sigma_t^{\intercal}\right). \end{aligned}$$

Above, the second line follows from the conservation relation (3a) in the quadratic coupling, $u \cdot B(u,u) = 0$ and $-\lambda_0$ is the largest eigenvalue of the negative-definite coefficient. The third line uses the estimate $\bar{u}_t \cdot F_t \le \lambda_0 \bar{u}_t^2 + \frac{|F_t|^2}{\lambda_0} \le \lambda_0 E_t + \frac{|F_t|^2}{\lambda_0}$. Then using the uniform boundedness of the forcing coefficients, we have the finite solution E_t as the direct conclusion from Grönwall's inequality starting from finite initial state $E_0 < \infty$. This proves that p_t has finite second moments during the time evolution, that is, $p_t \in \mathcal{P}_2\left(\mathbb{R}^d\right)$. Next, the regularity of the solution for $t > 0$ follows from the coercive operator from the dissipation coefficient (3b) following the standard proof as in [43].

Finally, for the finiteness of higher order moments, consider the high-order energy $E_t^{2k} = \int |u|^{2k} p_t(u)\,du$. By multiplying $|u|^{2k-2}u$ on both sides of Eq. (1) and using the finite moments in the lower order, the boundedness of each higher order moments will be reached following the same line of arguments. □

Proof of Lemma 2. First, applying Itô's formula for Z_t with any test function $\varphi \in C_b^2\left(\mathbb{R}^d\right)$ gives

$$\begin{aligned} d\varphi\left(Z_t\right) &= \mathcal{L}\left(\bar{u}_t, R_t\right)\varphi\left(Z_t\right)dt + \nabla\varphi\left(Z_t\right)^{\intercal}\Sigma_t dW_t \\ &= \nabla\varphi\left(Z_t\right)^{\intercal}\left[L\left(\bar{u}_t\right)Z_t + \Gamma\left(Z_tZ_t^{\intercal} - R_t\right)\right]dt \qquad \text{(B.1)} \\ &\quad + \frac{1}{2}\Sigma_t\Sigma_t^{\intercal} : \nabla\nabla\varphi\left(Z_t\right)dt + \nabla\varphi\left(Z_t\right)^{\intercal}\Sigma_t dW_t, \qquad \text{(B.2)} \end{aligned}$$

where $\mathcal{L}$ is the generator of Z_t. Given any statistical solution $\left(\bar{u}_t, R_t\right)$ and taking expectation using $\varphi(Z) = Z$ for $|Z| \le C$, the equation for the first moment of Z_τ with $\tau = t \wedge \sigma$ and $\sigma = \inf\{t : |Z_t| \ge C\}$ can be found as

$$\frac{d}{dt}\mathbb{E}\left[Z_\tau\right] = \left[L\left(\bar{u}_\tau\right)\mathbb{E}\left[Z_\tau\right] + \Gamma\left(\mathbb{E}\left[Z_\tau Z_\tau^{\intercal}\right] - R_\tau\right)\right]. \qquad \text{(B.3)}$$

Then notice that $\sigma \uparrow \infty$ and the first two moments are finite, we get the equation for Z_t. Next by taking $\varphi(Z) = Z_kZ_l$ for $|Z| \le C$, in a similarly way we have

$$\begin{aligned} d\left(Z_{k,t}Z_{l,t}\right) &= \sum_m \left[L_{km}\left(\bar{u}_t\right)Z_{m,t}Z_{l,t} + Z_{k,t}Z_{m,t}L_{lm}\left(\bar{u}_t\right)\right]dt + \Sigma_{km,t}\Sigma_{lm,t}dt \\ &\quad + \sum_{m,n}\gamma_{mnk}\left(Z_{m,t}Z_{n,t}Z_{l,t} - R_{mn,t}Z_{l,t}\right)dt \\ &\quad + \gamma_{mnl}\left(Z_{m,t}Z_{n,t}Z_{k,t} - R_{mn,t}Z_{k,t}\right)dt \\ &\quad + \sum_m \Sigma_{km,t}Z_{l,t}dW_{m,t} + \Sigma_{lm,t}Z_{k,t}dW_{m,t}. \end{aligned}$$

This implies the second moment equation of Z_t as

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[Z_tZ_t^{\intercal}\right] &= \sum_m \left[L_{km}\left(\bar{u}_t\right)\mathbb{E}\left[Z_{m,t}Z_{l,t}\right] + \mathbb{E}\left[Z_{k,t}Z_{m,t}\right]L_{lm}\left(\bar{u}_t\right)\right]dt + \Sigma_t\Sigma_t^{\intercal} \\ &\quad + \sum_{m,n}\gamma_{mnk}\left(\mathbb{E}\left[Z_{m,t}Z_{n,t}Z_{l,t}\right] - R_{mn,t}\mathbb{E}\left[Z_{l,t}\right]\right)dt \\ &\quad + \gamma_{mnl}\left(\mathbb{E}\left[Z_{m,t}Z_{n,t}Z_{k,t}\right] - R_{mn,t}\mathbb{E}\left[Z_{k,t}\right]\right)dt. \end{aligned} \qquad \text{(B.4)}$$

Assuming $\mathbb{E}\left[Z_t\right] = 0$ and $\mathbb{E}\left[Z_tZ_t^{\intercal}\right] = R_t$ for any time instant t, first we see that the right hand side of (B.3) will always stay zero. Then, with the same statistics in the third moments of Z_t, the right hand side of (B.4) becomes equal to the right hand of the statistical equation of R_t in (11) with (12). Uniqueness of solution in the statistical equations given the same initial values implies that the leading two moments of Z_t satisfy $\mathbb{E}\left[Z_t\right] = 0$ and $\mathbb{E}\left[Z_tZ_t^{\intercal}\right] = R_t$ for all $t > 0$. □

Proof of Proposition 3. First, consider $\mathbb{E}\varphi\left(u_t\right)$ with test function $\varphi \in C_b^2\left(\mathbb{R}^d\right)$ w.r.t. the PDF p_t for the state of the original system (1). Itô's lemma shows that

$$\frac{d\mathbb{E}\varphi\left(u_t\right)}{dt} = \mathbb{E}\left[\left(\Lambda u_t + B\left(u_t,u_t\right) + F_t\right)\cdot\nabla\varphi\left(u_t\right)\right] + \frac{1}{2}\mathbb{E}\left[\sigma_t\sigma_t^{\intercal} : \nabla\nabla\varphi\left(u_t\right)\right],$$

where $A : \nabla\nabla\varphi = \sum_{mn} a_{mn}\partial_{u_m}\partial_{u_n}\varphi$. In addition, the above equation can be generalized to all $\varphi \in C^2\left(\mathbb{R}^d\right)$ by applying Dynkin's formula for a strong Markov process. In fact, consider the test function φ defined within $|u| \le C$ and introduce the stopping time $\tau = t \wedge \sigma$ with $\sigma = \inf\{t : |u_t| \ge C\}$. We have the equation for u_τ by Dynkin's formula. Since all the moments of u_t remain bounded from Lemma 1, we get the same equation for $\varphi\left(u_t\right)$ by letting $C \uparrow \infty$ and $\sigma \uparrow \infty$.

Using the decomposition $u' = u - \mathbb{E}u = \sum_k u'_{k,t}\hat{\upsilon}_k$ with $u'_k = \hat{\upsilon}_k \cdot u'$, we have first for $\varphi = u$

$$\begin{aligned} \frac{d\mathbb{E}u_t}{dt} &= \mathbb{E}\left[\Lambda u_t + B\left(u_t,u_t\right) + F_t\right] \\ &= \Lambda\mathbb{E}u_t + B\left(\mathbb{E}u_t,\mathbb{E}u_t\right) + \sum_{k,l}\mathbb{E}\left[u'_{k,t}u'_{l,t}\right]B\left(\hat{\upsilon}_k,\hat{\upsilon}_l\right) + F_t. \end{aligned} \qquad \text{(B.5)}$$

Above, we use the bilinearity of the operator B and $\mathbb{E}u' = 0$, such that

$$\begin{aligned} \mathbb{E}B(u,u) &= \mathbb{E}B\left(\mathbb{E}u + u', \mathbb{E}u + u'\right) \\ &= \mathbb{E}\left[B(\mathbb{E}u,\mathbb{E}u) + B\left(\mathbb{E}u,u'\right) + B\left(u',\mathbb{E}u\right) + B\left(u',u'\right)\right] \\ &= B(\mathbb{E}u,\mathbb{E}u) + \mathbb{E}\sum_{k,l}B\left(u'_{k,t}\hat{\upsilon}_k, u'_{l,t}\hat{\upsilon}_l\right) \\ &= B(\mathbb{E}u,\mathbb{E}u) + \sum_{k,l}\mathbb{E}\left[u'_{k,t}u'_{l,t}\right]B\left(\hat{\upsilon}_k,\hat{\upsilon}_l\right). \end{aligned}$$

Similarly, by taking $\varphi = \left(\hat{\upsilon}_k \cdot u'_t\right)\left(u'_t \cdot \hat{\upsilon}_l\right) = u_t'^{\intercal}A_{kl}u'_t$ with $A_{kl} = \hat{\upsilon}_l\hat{\upsilon}_k^{\intercal}$, we find

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[u'_{k,t}u'_{l,t}\right] &= \mathbb{E}\left[\left(\Lambda u_t + B\left(u_t,u_t\right) + F_t\right)\cdot\left(A_{kl} + A_{kl}^{\intercal}\right)u'_t\right] + \frac{1}{2}\sigma_t\sigma_t^{\intercal} : \left(A_{kl} + A_{kl}^{\intercal}\right) \\ &= \mathbb{E}\left[\left(u'_t\cdot\hat{\upsilon}_l\right)\left(\hat{\upsilon}_k^{\intercal}\Lambda u'_t\right) + \left(u_t'^{\intercal}\Lambda^{\intercal}\hat{\upsilon}_l\right)\left(\hat{\upsilon}_k\cdot u'_t\right)\right] \\ &\quad + \mathbb{E}\left[\hat{\upsilon}_l\cdot B\left(\mathbb{E}u,u'\right)u'_{k,t} + \hat{\upsilon}_k\cdot B\left(\mathbb{E}u,u'\right)u'_{l,t}\right] \\ &\quad + \mathbb{E}\left[\hat{\upsilon}_l\cdot B\left(u',\mathbb{E}u\right)u'_{k,t} + \hat{\upsilon}_k\cdot B\left(u',\mathbb{E}u\right)u'_{l,t}\right] \\ &\quad + \mathbb{E}\left[B\left(u'_t,u'_t\right)\cdot\left(\hat{\upsilon}_lu'_{k,t} + \hat{\upsilon}_ku'_{l,t}\right)\right] + \left(\sigma_t\cdot\hat{\upsilon}_k\right)\left(\hat{\upsilon}_l\cdot\sigma_t\right) \\ &= \sum_m\left(\hat{\upsilon}_k^{\intercal}\Lambda\hat{\upsilon}_m\right)\mathbb{E}\left[u'_{m,t}u'_{l,t}\right] + \mathbb{E}\left[u'_{k,t}u'_{m,t}\right]\left(\hat{\upsilon}_l^{\intercal}\Lambda^{\intercal}\hat{\upsilon}_m\right) + \left(\sigma_t\cdot\hat{\upsilon}_k\right)\left(\hat{\upsilon}_l\cdot\sigma_t\right) \\ &\quad + \sum_m\left[\hat{\upsilon}_k^{\intercal}B\left(\mathbb{E}u,\hat{\upsilon}_m\right)\mathbb{E}\left[u'_{m,t}u'_{l,t}\right] + \mathbb{E}\left[u'_{k,t}u'_{m,t}\right]\hat{\upsilon}_l^{\intercal}B\left(\mathbb{E}u,\hat{\upsilon}_m\right)\right] \\ &\quad + \sum_m\left[\hat{\upsilon}_k^{\intercal}B\left(\hat{\upsilon}_m,\mathbb{E}u\right)\mathbb{E}\left[u'_{m,t}u'_{l,t}\right] + \mathbb{E}\left[u'_{k,t}u'_{m,t}\right]\hat{\upsilon}_l^{\intercal}B\left(\hat{\upsilon}_m,\mathbb{E}u\right)\right] \end{aligned} \qquad \text{(B.6)}$$

$$+\sum_{m,n}\left[\hat{v}_k^{\intercal}B\left(\hat{v}_m,\hat{v}_n\right)\mathbb{E}\left[u'_{l,t}u'_{m,t}u'_{n,t}\right]+\hat{v}_l^{\intercal}B\left(\hat{v}_m,\hat{v}_n\right)\mathbb{E}\left[u'_{k,t}u'_{m,t}u'_{n,t}\right]\right].$$

Above in the second equality, we use the projection on the two modes $\left(A_{kl}+A_{kl}^{\intercal}\right)u'_t=\left(\hat{v}_l\hat{v}_k^{\intercal}+\hat{v}_k\hat{v}_l^{\intercal}\right)u'_t=\hat{v}_lu'_{k,t}+\hat{v}_ku'_{k,t}$, and the bilinearity of the quadratic operator B; and in the third equality, we use the decomposition of fluctuation modes, $u'_t=\sum_k u'_{k,t}\hat{v}_k$. We find the coupling operator $L_{km}=\left(\hat{v}_k^{\intercal}\Lambda\hat{v}_m\right)+\hat{v}_k^{\intercal}B\left(\hat{v}_m,\mathbb{E}u\right)+\hat{v}_k^{\intercal}B\left(\hat{v}_m,\mathbb{E}u\right)$, the third-order coupling coefficients $\gamma_{mnk}=\hat{v}_k^{\intercal}B\left(\hat{v}_m,\hat{v}_n\right)$, as well as the noise term $\left(\sigma_t\cdot\hat{v}_k\right)\left(\hat{v}_l\cdot\sigma_t\right)$.

In addition, by subtracting the mean Eq. (B.5) from the original system (1), we find the SDE for the stochastic state

$$\begin{aligned}\frac{\mathrm{d}u'_t}{\mathrm{d}t}=\ &\frac{\mathrm{d}}{\mathrm{d}t}\left(u_t-\mathbb{E}u_t\right)=\Lambda u_t+B\left(u_t,u_t\right)+\sigma_t\dot{W}_t\\&-\Lambda\mathbb{E}u_t-B\left(\mathbb{E}u,\mathbb{E}u\right)-\sum_{m,n}\mathbb{E}\left[u'_{m,t}u'_{n,t}\right]B\left(\hat{v}_m,\hat{v}_n\right)\\&=\sum_m u'_{m,t}\left[\Lambda\hat{v}_m+B\left(\mathbb{E}u,\hat{v}_m\right)+B\left(\hat{v}_m,\mathbb{E}u\right)\right]\\&+\sum_{m,n}\left[u'_{m,t}u'_{n,t}-\mathbb{E}\left[u'_{m,t}u'_{n,t}\right]\right]B\left(\hat{v}_m,\hat{v}_n\right)+\sigma_t\dot{W}_t.\end{aligned}$$

Again in the second equality, we use the spectral decomposition of the fluctuation state $u'_{k,t}=u'_t\cdot\hat{v}_k$. By projecting the state u'_t on the basis $\hat{v}_k$, we have

$$\frac{\mathrm{d}u'_{k,t}}{\mathrm{d}t}=\sum_k L_{km}\left(\mathbb{E}u_t\right)u'_{m,t}+\sum_{m,n}\gamma_{mnk}\left[u'_{m,t}u'_{n,t}-\mathbb{E}\left[u'_{m,t}u'_{n,t}\right]\right]+\hat{v}_k^{\intercal}\sigma_t\dot{W}_t,\quad\text{(B.7)}$$

with the same parameters L_{km} and γ_{mnk} defined before. The generator $\mathcal{L}_t^u$ of u'_t can be written as

$$\begin{aligned}\mathcal{L}_t^u\left(p_t\right)=&\left[\sum_k L_{km}\left(\mathbb{E}u_t\right)u'_{m,t}+\sum_{m,n}\gamma_{mnk}\left(u'_{m,t}u'_{n,t}-\mathbb{E}\left[u'_{m,t}u'_{n,t}\right]\right)\right]\cdot\nabla_{u'}\\&+\frac{1}{2}\Sigma_t\Sigma_t^{\intercal}:\nabla_{u'}\nabla_{u'},\end{aligned}$$

where $\mathcal{L}_t^u$ is dependent on p_t in computing the expectations.

Next, we consider the closure model (11) without the relaxation term

$$\begin{aligned}\frac{\mathrm{d}\bar{u}_t}{\mathrm{d}t}&=\Lambda\bar{u}_t+B\left(\bar{u}_t,\bar{u}_t\right)+Q_m\left(\mathbb{E}\left[Z_t\otimes Z_t\right]\right)+F_t,\\\frac{\mathrm{d}R_t}{\mathrm{d}t}&=L\left(\bar{u}_t\right)R_t+R_tL\left(\bar{u}_t\right)+Q_v\left(\mathbb{E}\left[Z_t\otimes Z_t\otimes Z_t\right]\right)+\Sigma_t\Sigma_t^{\intercal},\qquad\text{(B.8)}\\\frac{\mathrm{d}}{\mathrm{d}t}&\mathbb{E}\left[\varphi\left(Z_t\right)\right]=\mathbb{E}\left[\mathcal{L}_t\left(\bar{u}_t,R_t\right)\varphi\left(Z_t\right)\right].\end{aligned}$$

where $\mathcal{L}_t$ is the generator from (20) defined from the McKean–Vlasov SDE of Z_t

$$\mathcal{L}_t\left(\bar{u}_t,R_t\right)=\left[L\left(\bar{u}_t\right)Z_t+\Gamma\left(Z_tZ_t^{\intercal}-R_t\right)\right]\cdot\nabla_z+\frac{1}{2}\Sigma_t\Sigma_t^{\intercal}:\nabla_z\nabla_z.$$

The closure terms Q_m and Q_v in (12) have exactly the same structure as that in the original system derived in (B.5) and (B.6). In addition, by comparing the above SDEs (9) and (B.7), it is realized that their generators, $\mathcal{L}_t^u\left(p_t\right)$ and $\mathcal{L}_t\left(\bar{u}_t,R_t\right)$, share the same dynamical structure with the dependence on the first two moments w.r.t. p_t and the statistical solutions $\bar{u}_t,R_t$ in (B.8). Therefore, at any time instant t if we assume consistent statistics

$$\mathbb{E}_{p_t}\left[u_t\right]=\bar{u}_t,\quad\mathbb{E}_{p_t}\left[\left(u'_t\cdot\hat{v}_k\right)\left(u'_t\cdot\hat{v}_l\right)\right]=R_{kl,t},$$

as well as

$$\mathbb{E}_{p_t}\left[\varphi\left(u'_t\right)\right]=\mathbb{E}\left[\varphi\left(\sum_{k=1}^d Z_{k,t}\hat{v}_k\right)\right],$$

the right hand sides of the original model (B.5), (B.6), and (B.7) become the same as that of the closure model (B.8). Starting from the same initial condition with uniqueness of the solution, it directly implies that the statistical solutions of the two systems (1) and (B.8) will remain the same during the entire time evolution. Finally, adding the additional coefficient $\left(1+\left|u_t\right|^2\right)$ in front of the test function φ will follow by repeating the same argument given that the second moments w.r.t. p_t are finite. □

Proof of Proposition 4. Given any $\mathcal{G}_t$-measurable square-integrable stochastic process v, we have from direct computation

$$\begin{aligned}&\mathbb{E}\left[\left|\mathcal{M}\rho_t-\mathcal{M}v\right|^2\right]-\mathbb{E}\left[\left|\mathcal{M}\rho_t-\mathcal{M}\hat{\rho}_t\right|^2\right]\\&=\mathbb{E}\left[\left(\mathcal{M}\hat{\rho}_t-\mathcal{M}v\right)\cdot\left(2\mathcal{M}\rho_t-\mathcal{M}v-\mathcal{M}\hat{\rho}_t\right)\right]\\&=\mathbb{E}\left\{\mathbb{E}\left[\left(\mathcal{M}\hat{\rho}_t-\mathcal{M}v\right)\cdot\left(2\mathcal{M}\rho_t-\mathcal{M}v-\mathcal{M}\hat{\rho}_t\right)\mid\mathcal{G}_t\right]\right\}\\&=\mathbb{E}\left\{\left(\mathcal{M}\hat{\rho}_t-\mathcal{M}v\right)\cdot\left[\mathbb{E}\left[2\mathcal{M}\rho_t\mid\mathcal{G}_t\right]-\left(\mathcal{M}v+\mathcal{M}\hat{\rho}_t\right)\right]\right\}\\&=\mathbb{E}\left[\left|\mathcal{M}\hat{\rho}_t-\mathcal{M}v\right|^2\right]\geq 0.\end{aligned}$$

Above, the third equality uses the fact $\mathcal{M}\hat{\rho}_t-\mathcal{M}v=\mathbb{E}\left[\mathcal{M}\rho_t\mid\mathcal{G}_t\right]-\mathcal{M}v$ is $\mathcal{G}_t$-measurable. Thus, we get $\mathcal{M}\hat{\rho}_t$ minimizes the mean square error. The consistency under the expectation in $\mathcal{M}\hat{\rho}_t$ and $\mathcal{M}\rho_t$ can be directly implied by definition. □

Proof of Lemma 5. We rewrite the filter model (38) for $\tilde{Z}_t$ by substituting the explicit equation for the observation process, $\mathrm{d}y_t=\left[\mathcal{H}\rho_t+h_t\left(y_t\right)\right]\mathrm{d}t+\Gamma_t\mathrm{d}B_t$, in (23)

$$\mathrm{d}\tilde{Z}_t=a_t\left(\tilde{Z}_t\right)\mathrm{d}t+K_t\left(\tilde{Z}_t\right)\left\{\left[\mathcal{H}\rho_t-H\left(\tilde{Z}_t\right)\right]\mathrm{d}t+\Gamma_t\mathrm{d}B_t-\Gamma_t\mathrm{d}\tilde{B}_t\right\}.$$

By applying Itô's formula on the above SDE, we have for $\varphi\in C_b^2\left(\mathbb{R}^d\right)$

$$\begin{aligned}\mathrm{d}\varphi\left(\tilde{Z}_t\right)=&\ \nabla\varphi\cdot\left[\left(a_t-K_t\left(H\left(\tilde{Z}_t\right)+h_t\left(y_t\right)\right)\right)\right]\mathrm{d}t\\&-\nabla\varphi\cdot K_t\Gamma_t\mathrm{d}\tilde{B}_t+\nabla\varphi\cdot K_t\mathrm{d}y_t+K_t\Gamma_t^2K_t^{\intercal}:\nabla\nabla\varphi\mathrm{d}t,\qquad\text{(B.9)}\end{aligned}$$

where we define $A:\nabla\nabla\varphi=\sum_{i,j=1}^d A_{ij}\partial_{z_iz_j}\varphi$ and take the convention $\left(\nabla f\right)_{ij}=\partial_{z_i}f_j$ for the gradient of vector-valued functions $f\in C^1\left(\mathbb{R}^d;\mathbb{R}^p\right)$. Notice that above the coefficient in the last term is 1 considering the additional contributions from the independent white noise process $\Gamma_t\mathrm{d}B_t=\mathrm{d}y_t-\left[\mathcal{H}\rho_t+h_t\left(y_t\right)\right]\mathrm{d}t$ in the observation process besides the original $\mathrm{d}\tilde{B}_t$, that is,

$$\frac{1}{2}\nabla\nabla\varphi:\mathrm{d}\left\langle K\Gamma\tilde{B},K\Gamma\tilde{B}\right\rangle_t+\frac{1}{2}\nabla\nabla\varphi:\mathrm{d}\left\langle K\Gamma B,K\Gamma B\right\rangle_t=\nabla\nabla\varphi:K_t\Gamma_t^2K_t^{\intercal}\mathrm{d}t,$$

where we denote $\langle M,N\rangle_t$ as the Meyer's process of two martingales M_t and N_t.

First, by taking $\varphi(z)=H(z)$ and taking expectation $\tilde{\mathbb{E}}$ w.r.t. $\tilde{\rho}_t$ conditional on $Y_t=\left\{y_s,s\leq t\right\}\in\mathcal{G}_t$, we have

$$\begin{aligned}\mathrm{d}\tilde{\mathbb{E}}H\left(\tilde{Z}_t\right)=&\ \tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}a_t\right]\mathrm{d}t-\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}K_t\left(\bar{H}_t+H'_t+h_t\left(y_t\right)\right)\right]\mathrm{d}t\\&+\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}K_t\right]\mathrm{d}y_t+\tilde{\mathbb{E}}\left[K_t\Gamma_t^2K_t^{\intercal}:\nabla\nabla H\left(\tilde{Z}_t\right)\right]\mathrm{d}t\\=&\ \tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}a_t\right]\mathrm{d}t+\tilde{\mathbb{E}}\left[K_t\Gamma_t^2K_t^{\intercal}:\nabla\nabla H\left(\tilde{Z}_t\right)\right]\mathrm{d}t\\&-\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}K_tH'_t\right]\mathrm{d}t+\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}K_t\right]\left\{\mathrm{d}y_t-\left[\bar{H}_t+h_t\left(y_t\right)\right]\mathrm{d}t\right\}.\end{aligned}\qquad\text{(B.10)}$$

In the first line above, we split $H\left(\tilde{Z}_t\right)=\bar{H}_t+H'_t$. Notice that the observation process $y_t\in\mathcal{G}_t$ can be brought out of the expectation $\tilde{\mathbb{E}}[\cdot]=\mathbb{E}\left[\cdot\mid\mathcal{G}_t\right]$. Using the first identity in (41) for a_t, there is

$$\begin{aligned}\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal}a_t\right]&=\int\nabla H(z)^{\intercal}\left[\nabla\cdot\left(K_t\Gamma_t^2K_t^{\intercal}\right)-K_t\Gamma_t^2\nabla\cdot K_t^{\intercal}\right]\tilde{\rho}_t(z)\,\mathrm{d}z\\&=\int\nabla H(z)^{\intercal}\left[\nabla\cdot\left(\tilde{\rho}_tK_t\Gamma_t^2K_t^{\intercal}\right)-K_t\Gamma_t^2K_t^{\intercal}\nabla\tilde{\rho}_t\right.\\&\qquad\left.-\ K_t\Gamma_t^2\nabla\cdot K_t^{\intercal}\tilde{\rho}_t\right]\mathrm{d}z\\&=-\int\nabla\nabla H(z):\left(K_t\Gamma_t^2K_t^{\intercal}\right)\tilde{\rho}_t\mathrm{d}z\\&\quad-\int\nabla H(z)^{\intercal}K_t\Gamma_t^2\nabla\cdot\left(\tilde{\rho}_tK_t^{\intercal}\right)\mathrm{d}z\\&=-\tilde{\mathbb{E}}\left[\left(K_t\Gamma_t^2K_t^{\intercal}\right):\nabla\nabla H\left(\tilde{Z}_t\right)\right]\\&\quad-\int\nabla H(z)^{\intercal}K_t\Gamma_t^2\nabla\cdot\left(\tilde{\rho}_tK_t^{\intercal}\right)\mathrm{d}z.\end{aligned}$$

Then using the second identity in (41) for K_t and denoting $H'_t = H - \bar{H}_t$, the last term above gets simplified to

$$-\int \nabla H^{\intercal} K_t \Gamma_t^2 \nabla \cdot \left(\tilde{\rho}_t K_t^{\intercal}\right) \mathrm{d}z = \int \nabla H_t'^{\intercal} K_t \tilde{\rho}_t H_t' \mathrm{d}z = \tilde{\mathbb{E}}\left[\nabla H_t'^{\intercal} K_t H_t'\right].$$

With the above identities, first line of (B.10) becomes zero. Further with the second identity in (41) for K_t, there is

$$\tilde{\mathbb{E}}\left(K_t^{\intercal} \nabla \psi\right) = \Gamma_t^{-2} \tilde{\mathbb{E}}\left[\left(H\left(\tilde{Z}_t\right) - \mathbb{E}H\right) \psi\left(\tilde{Z}_t\right)^{\intercal}\right],$$

for any regular function ψ with $\mathbb{E}\psi = 0$. By taking $\psi = H - \bar{H}_t$, there is

$$\tilde{\mathbb{E}}\left[\nabla H\left(\tilde{Z}_t\right)^{\intercal} K_t\right] = \tilde{\mathbb{E}}\left[\left(H\left(\tilde{Z}_t\right) - \bar{H}_t\right)\left(H\left(\tilde{Z}_t\right) - \bar{H}_t\right)^{\intercal}\right] \Gamma_t^{-2} = C_t^H \Gamma_t^{-2}.$$

This gives the equation for $\bar{H}_t = \tilde{\mathbb{E}} H\left(\tilde{Z}_t\right)$.

Next, we take $\varphi(z) = H_k(z) H_l(z)$. For the convenience of computation, we separate the mean state $\bar{H}_t$ as

$$\begin{aligned}\varphi(z) &= \left[\bar{H}_{k,t} + H_k'(z)\right]\left[\bar{H}_{l,t} + H_l'(z)\right] \\ &= H_{k,t}'(z) H_{l,t}'(z) + \bar{H}_{k,t} H_l'(z) + H_k'(z) \bar{H}_{l,t} + \bar{H}_{k,t} \bar{H}_{l,t}.\end{aligned}$$

The last term above is independent of z, thus will vanish after applying Itô's formula (B.9). We have for the first term on the right hand side

$$\begin{aligned}\mathrm{d}\tilde{\mathbb{E}}\left[H_{k,t}'\left(\tilde{Z}_t\right) H_{l,t}'\left(\tilde{Z}_t\right)\right] &= \tilde{\mathbb{E}}\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal}\left[\left(a_t - K_t\left(\bar{H}_t + H_t' + h_t\left(y_t\right)\right)\right)\right] \mathrm{d}t \\ &+ \tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t\right] \mathrm{d}y_t + \tilde{\mathbb{E}}\left[K_t \Gamma_t^2 K_t^{\intercal} : \nabla\nabla\left(H_{k,t}' H_{l,t}'\right)\right] \mathrm{d}t \\ &= \tilde{\mathbb{E}}\left[\nabla\left(H_k' H_l'\right)^{\intercal} a_t\right] \mathrm{d}t + \tilde{\mathbb{E}}\left[K_t \Gamma_t^2 K_t^{\intercal} : \nabla\nabla\left(H_{k,t}' H_{l,t}'\right)\right] \mathrm{d}t \\ &+ \tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t\right]\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right] \mathrm{d}t\right\} \\ &- \tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t H_t'\right].\end{aligned} \quad \text{(B.11)}$$

Using the identifies (41) for a_t and K_t, again we can find

$$\begin{aligned}\tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} a_t\right] &= -\tilde{\mathbb{E}}\left[K_t \Gamma_t^2 K_t^{\intercal} : \nabla\nabla\left(H_{k,t}' H_{l,t}'\right)\right] \\ &- \int \nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t \Gamma_t^2 \nabla \cdot \left(\tilde{\rho}_t K_t^{\intercal}\right) \mathrm{d}z \\ &= -\tilde{\mathbb{E}}\left[K_t \Gamma_t^2 K_t^{\intercal} : \nabla\nabla\left(H_{k,t}' H_{l,t}'\right)\right] \\ &+ \tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t H_t'\right].\end{aligned}$$

Therefore, we have $\mathrm{d}\tilde{\mathbb{E}}\left[H_{k,t}' H_{l,t}'\right] = \tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t\right]\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right] \mathrm{d}t\right\}$. Further, using the identity for K_t, the coefficient becomes third moments of H_t'

$$\begin{aligned}\tilde{\mathbb{E}}\left[\nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t\right] &= \int \nabla\left(H_{k,t}' H_{l,t}'\right)^{\intercal} K_t \tilde{\rho}_t \mathrm{d}z \\ &= -\int\left(H_{k,t}' H_{l,t}'\right)\left[\nabla \cdot \left(\tilde{\rho}_t K_t^{\intercal}\right)\right]^{\intercal} \mathrm{d}z \\ &= \int\left(H_{k,t}' H_{l,t}'\right)\left[\tilde{\rho}_t \Gamma_t^{-2}\left(H(z) - \bar{H}_t\right)\right]^{\intercal} \mathrm{d}z \\ &= \tilde{\mathbb{E}}\left[H_{k,t}' H_{l,t}' H'^{\intercal}\right] \Gamma_t^{-2}.\end{aligned}$$

Similarly, by repeating the same procedure for $\bar{H}_{k,t} H_l'(z)$, we have

$$\begin{aligned}\mathrm{d}\tilde{\mathbb{E}}\left[\bar{H}_{k,t} H_l'\left(\tilde{Z}_t\right)\right] &= \tilde{\mathbb{E}}\left[\nabla\left(\bar{H}_{k,t} H_{l,t}'\right)^{\intercal} K_t\right]\left\{\mathrm{d}y_t - \left[\bar{H}_t + h_t\left(y_t\right)\right] \mathrm{d}t\right\} \\ &= \tilde{\mathbb{E}}\left[H_{l,t}' H_t'^{\intercal}\right] \Gamma_t^{-2}\left[\mathrm{d}y_t - \left(\bar{H}_t + h_t\right) \mathrm{d}t\right] \bar{H}_{k,t}.\end{aligned}$$

And similar result can be achieved for $\tilde{\mathbb{E}}\left[H_k'\left(\tilde{Z}_t\right) \bar{H}_{l,t}\right]$.

Finally, applying Itô's formula for $\bar{H}_t \bar{H}_t^{\intercal}$ where $\mathrm{d}\bar{H}_t = C_t^H \Gamma_t^{-2}\left(\mathcal{H}\rho_t - \bar{H}_t\right) \mathrm{d}t + C_t^H \Gamma_t^{-1} \mathrm{d}B_t$ as we have derived, there is,

$$\begin{aligned}\mathrm{d}\left(\bar{H}_t \bar{H}_t^{\intercal}\right) &= \left(\mathrm{d}\bar{H}_t\right) \bar{H}_t^{\intercal} + \bar{H}_t\left(\mathrm{d}\bar{H}_t^{\intercal}\right) + \mathrm{d}\left\langle C^H \Gamma^{-1} B, C^H \Gamma^{-1} B\right\rangle_t \\ &= C_t^H \Gamma_t^{-2}\left[\mathrm{d}y_t - \left(\bar{H}_t + h_t\right) \mathrm{d}t\right] \bar{H}_t^{\intercal} \\ &+ \bar{H}_t\left[\mathrm{d}y_t^{\intercal} - \left(\bar{H}_t + h_t\right)^{\intercal} \mathrm{d}t\right] \Gamma_t^{-2} C_t^H \\ &+ C_t^H \Gamma_t^{-2} C_t^H \mathrm{d}t.\end{aligned}$$

Notice again that the white noise process, $C_t^H \Gamma_t^{-1} \mathrm{d}B_t$, gives the last term in the first equality above. Putting all the above equations together, we get the equation for $\mathrm{d}C_t^H = \mathrm{d}\tilde{\mathbb{E}}\left[H\left(\tilde{Z}_t\right) H\left(\tilde{Z}_t\right)^{\intercal}\right] - \mathrm{d}\left(\bar{H}_t \bar{H}_t^{\intercal}\right)$ where $C_{kl,t}^H = \tilde{\mathbb{E}}\left[H_{k,t}' H_{l,t}'\right]$. □

Proof of Proposition 12. According to (60) with $K = \tilde{K}\Gamma^{-2}$, we need to show

$$-\nabla \cdot \left(\tilde{K}^{\intercal} \tilde{\rho}\right) H^{\intercal} = \tilde{\rho} H' H^{\intercal} \Rightarrow \tilde{\mathbb{E}}\left[\tilde{K}^{\intercal} \nabla H\right] = \tilde{\mathbb{E}}\left[H'\left(\bar{H} + H'\right)^{\intercal}\right] = C^H,$$

according to the specific expressions $H = H^m$ and $H = H^v$. First, we can compute

$$\nabla_z H_l^m = 2A_l z, \quad \nabla_z H_{pq}^v = 2z_q A_p z + \left(z^{\intercal} A_p z\right) \delta_{qj} \hat{e}_j.$$

Above in H^v for simplicity, we only compute half of the symmetric function and $\hat{e}_j$ is the unit vector with value 1 in the jth entry.

From direct computations for H^m and using $H_k^m = z^{\intercal} A_k z$, we have

$$\begin{aligned}\sum_j \tilde{K}_{j,k}^m \frac{\partial H_l^m}{\partial z_j} &= \frac{1}{2}\left[\left(z^{\intercal} A_k z\right) - \bar{H}_k^m\right] \sum_j z_j 2\left(A_l z\right)_j \\ &= \left[\left(z^{\intercal} A_k z\right) - \bar{H}_k^m\right]\left(z^{\intercal} A_l z\right) = H_k^{m\prime} H_l^m.\end{aligned}$$

Similarly for H^v, we can compute

$$\begin{aligned}\sum_j \tilde{K}_{j,kl}^v \frac{\partial H_{pq}^v}{\partial z_j} &= \frac{1}{3}\left[\left(z^{\intercal} A_k z\right) z_l - \bar{H}_{kl}^v\right] \sum_j z_j\left[2z_q\left(A_p z\right)_j + \left(z^{\intercal} A_p z\right) \delta_{qj}\right] \\ &= \frac{2}{3}\left[\left(z^{\intercal} A_k z\right) z_l - \bar{H}_{kl}^v\right]\left(z^{\intercal} A_p z\right) z_q \\ &+ \frac{1}{3}\left[\left(z^{\intercal} A_k z\right) z_l - \bar{H}_{kl}^v\right] z_q\left(z^{\intercal} A_p z\right) \\ &= \left[\left(z^{\intercal} A_k z\right) z_l - \bar{H}_{kl}^v\right]\left(z^{\intercal} A_p z\right) z_q = H_{kl}^{v\prime} H_{pq}^v.\end{aligned}$$

This finishes the proof. □

Data availability

Data will be made available on request.

References

[1] R. Salmon, Lectures on Geophysical Fluid Dynamics, Oxford University Press, USA, 1998.

[2] P.H. Diamond, S.-I. Itoh, K. Itoh, Modern Plasma Physics: Volume 1, Physical Kinetics of Turbulent Plasmas, Cambridge University Press, 2010.

[3] S. Nazarenko, S. Lukaschuk, Wave turbulence on water surface, Annu. Rev. Condens. Matter Phys. 7 (2016) 61–88.

[4] M.-H. Giga, A. Kirshtein, C. Liu, Variational modeling and complex fluids, in: Handbook of Mathematical Analysis in Mechanics of Viscous Fluids, 2017, pp. 1–41.

[5] É. Pardoux, Stochastic partial differential equations, in: Fudan Lecture Notes, vol. 87, 2007.

[6] A.J. Majda, Introduction To Turbulent Dynamical Systems in Complex Systems, Springer, 2016.

[7] A.J. Majda, D. Qi, Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems, SIAM Rev. 60 (3) (2018) 491–549.

[8] M. Leutbecher, T.N. Palmer, Ensemble forecasting, J. Comput. Phys. 227 (7) (2008) 3515–3539.

[9] S. Tong, E. Vanden-Eijnden, G. Stadler, Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis, Commun. Appl. Math. Comput. Sci. 16 (2) (2021) 181–225.

[10] A.J. Majda, X.T. Tong, Performance of ensemble Kalman filters in large dimensions, Comm. Pure Appl. Math. 71 (5) (2018) 892–937.

[11] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, Mon. Weather Rev. 136 (12) (2008) 4629–4640.

[12] J. Xiong, An Introduction to Stochastic Filtering Theory, vol. 18, OUP Oxford, 2008.

[13] S. Reich, C. Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.

[14] K. Law, A. Stuart, K. Zygalakis, Data Assimilation, vol. 214, Springer, Cham, Switzerland, 2015, p. 52.

[15] D. Sanz-Alonso, A. Stuart, A. Taeb, Inverse Problems and Data Assimilation, vol. 107, Cambridge University Press, 2023.

[16] S.-T. Yau, S.S.-T. Yau, Finite dimensional filters with nonlinear drift, xi: Explicit solution of the generalized Kolmogorov equation in Brockett–Mitter program, Adv. Math. 140 (2) (1998) 156–189.
[17] D. Crisan, J. Xiong, Approximate McKean–Vlasov representations for a class of SPDEs, Stochastics an Int. J. Probab. Stochastics Process. 82 (1) (2010) 53–68.
[18] K. Hayden, E. Olson, E.S. Titi, Discrete data assimilation in the Lorenz and 2D Navier–Stokes equations, Phys. D: Nonlinear Phenom. 240 (18) (2011) 1416–1425.
[19] S. Pathiraja, S. Reich, W. Stannat, Mckean–Vlasov SDEs in nonlinear filtering, SIAM J. Control Optim. 59 (6) (2021) 4188–4215.
[20] T.G. Kurtz, J. Xiong, Particle representations for a class of nonlinear SPDEs, Stochastic Process. Appl. 83 (1) (1999) 103–126.
[21] T.G. Kurtz, J. Xiong, Numerical solutions for a class of SPDEs with application to filtering, in: Stochastics in Finite and Infinite Dimensions: In Honor of Gopinath Kallianpur, 2001, pp. 233–258.
[22] K. Bergemann, S. Reich, An ensemble Kalman-Bucy filter for continuous data assimilation, Meteorol. Z. 21 (3) (2012) 213.
[23] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res.: Ocean. 99 (C5) (1994) 10143–10162.
[24] P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev. 126 (3) (1998) 796–811.
[25] P. Del Moral, Nonlinear filtering: Interacting particle resolution, C. R. l'Acad. Sci. -Series I-Mathematics 325 (6) (1997) 653–658.
[26] A. Doucet, N. De Freitas, N.J. Gordon, et al., Sequential Monte Carlo Methods in Practice, vol. 1, Springer, 2001.
[27] E. Calvello, S. Reich, A.M. Stuart, Ensemble Kalman methods: A mean field perspective, Acta Numer. (2025).
[28] J. de Wiljes, X.T. Tong, Analysis of a localised nonlinear ensemble Kalman Bucy filter with complete and accurate observations, Nonlinearity 33 (9) (2020) 4752.
[29] A. Balakrishna, A. Biswas, Determining map, data assimilation and an observable regularity criterion for the three-dimensional Boussinesq system, Appl. Math. Optim. 86 (3) (2022) 28.
[30] A. Biswas, M. Branicki, A unified framework for the analysis of accuracy and stability of a class of approximate Gaussian filters for the Navier-Stokes equations, 2024, arXiv preprint arXiv:2402.14078.
[31] G.A. Gottwald, A.J. Majda, A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks, Nonlinear Process. Geophys. 20 (5) (2013) 705–712.
[32] C. González-Tokman, B.R. Hunt, Ensemble data assimilation for hyperbolic systems, Phys. D: Nonlinear Phenom. 243 (1) (2013) 128–142.
[33] M. Hairer, J.C. Mattingly, Ergodicity of the 2D Navier–Stokes equations with degenerate stochastic forcing, Ann. Math. (2006) 993–1032.
[34] J.C. Mattingly, A.M. Stuart, D.J. Higham, Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise, Stochastic Process. Appl. 101 (2) (2002) 185–232.
[35] S. Dirren, G.J. Hakim, Toward the assimilation of time-averaged observations, Geophys. Res. Lett. 32 (4) (2005).
[36] A. Majda, I. Timofeyev, E. Vanden-Eijnden, A priori tests of a stochastic mode reduction strategy, Phys. D: Nonlinear Phenom. 170 (3–4) (2002) 206–252.
[37] E. Bach, T. Colonius, I. Scherl, A. Stuart, Filtering dynamical systems using observations of statistics, Chaos: An Interdiscip. J. Nonlinear Sci. 34 (3) (2024).
[38] M. Lesieur, Turbulence in Fluids: Stochastic and Numerical Modelling, vol. 488, Nijhoff Boston, MA, 1987.
[39] D. Qi, A.J. Majda, Low-dimensional reduced-order models for statistical response and uncertainty quantification: Barotropic turbulence with topography, Phys. D: Nonlinear Phenom. 343 (2017) 7–27.
[40] R. Hu, T.K. Edwards, L.M. Smith, S.N. Stechmann, Initial investigations of precipitating quasi-geostrophic turbulence with phase changes, Res. Math. Sci. 8 (2021) 1–25.
[41] D. Qi, J.-G. Liu, A random batch method for efficient ensemble forecasts of multiscale turbulent systems, Chaos: An Interdiscip. J. Nonlinear Sci. 33 (2) (2023).
[42] Y. Gao, D. Qi, Mean field games for controlling coherent structures in nonlinear fluid systems, SIAM/ASA J. Uncertain. Quantif. 13 (4) (2025) 1681–1708.
[43] L. Koralov, Y.G. Sinai, Theory of Probability and Random Processes, Springer Science & Business Media, 2007.
[44] D. Qi, J.-G. Liu, High-order moment closure models with random batch method for efficient computation of multiscale turbulent systems, Chaos: An Interdiscip. J. Nonlinear Sci. 33 (10) (2023).
[45] A.J. Majda, Statistical energy conservation principle for inhomogeneous turbulent dynamical systems, Proc. Natl. Acad. Sci. 112 (29) (2015) 8937–8941.
[46] U. Frisch, Turbulence: The Legacy of an Kolmogorov, Cambridge University Press, 1995.
[47] A.J. Majda, D. Qi, Linear and nonlinear statistical response theories with prototype applications to sensitivity analysis and statistical control of complex turbulent dynamical systems, Chaos: An Interdiscip. J. Nonlinear Sci. 29 (10) (2019).
[48] J. Földes, N.E. Glatt-Holtz, D.P. Herzog, Sensitivity of steady states in a degenerately damped stochastic Lorenz system, Stoch. Dyn. 21 (08) (2021) 2150055.
[49] J. Földes, N. Glatt-Holtz, G. Richards, J. Whitehead, Ergodicity in randomly forced Rayleigh–Bénard convection, Nonlinearity 29 (11) (2016) 3309.
[50] H.P. McKean, Propagation of chaos for a class of non-linear parabolic equations, in: Stochastic Differential Equations, in: Lecture Series in Differential Equations, Session 7, Catholic Univ. 1967, 1967, pp. 41–57.
[51] C. Graham, T.G. Kurtz, S. Méléard, P.E. Protter, M. Pulvirenti, D. Talay, S. Méléard, Asymptotic behaviour of some interacting particle systems; mckean-vlasov and Boltzmann models, in: Probabilistic Models for Nonlinear Partial Differential Equations: Lectures Given At the 1st Session of the Centro Internazionale Matematico Estivo (CIME) Held in Montecatini Terme, Italy, May (1995) 22–30, 1996, pp. 42–95.
[52] D. Qi, J. Harlim, Machine learning-based statistical closure models for turbulent dynamical systems, Phil. Trans. R. Soc. A 380 (2229) (2022) 20210205.
[53] N. Chen, D. Qi, A physics-informed data-driven algorithm for ensemble forecast of complex turbulent systems, Appl. Math. Comput. 466 (2024) 128480.
[54] D. Qi, J.-G. Liu, Data assimilation models for computing probability distributions of complex multiscale systems, J. Comput. Phys. 545 (2026) 114465.
[55] R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes II: Applications, vol. 6, Springer Science & Business Media, 2013.
[56] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, 1961.
[57] R. Curtain, A survey of infinite-dimensional filtering, SIAM Rev. 17 (3) (1975) 395–411.
[58] P.L. Falb, Infinite-dimensional filtering: The Kalman-Bucy filter in Hilbert space, 1967.
[59] T. Yang, H.A. Blom, P.G. Mehta, The continuous-discrete time feedback particle filter, in: 2014 American Control Conference, IEEE, 2014, pp. 648–653.
[60] F.J. Pinski, G. Simpson, A.M. Stuart, H. Weber, Kullback–Leibler approximation for probability measures on infinite dimensional spaces, SIAM J. Math. Anal. 47 (6) (2015) 4091–4122.
[61] P. Dupuis, R.S. Ellis, A Weak Convergence Approach To the Theory of Large Deviations, John Wiley & Sons, 2011.
[62] E.N. Lorenz, Predictability: A problem partly solved, in: Proc. Seminar on predictability, vol. 1, Reading, 1996, pp. 1–18.
[63] D. Qi, J.-G. Liu, Oscillatory solutions at the continuum limit of Lorenz 96 systems, Commun. Inf. Syst. 25 (3) (2024) 551–587.
[64] A. Bain, D. Crisan, Fundamentals of Stochastic Filtering, vol. 3, Springer, 2009.
[65] A. Budhiraja, Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter, in: Annales de l'IHP Probabilités et statistiques, vol. 39, 2003, pp. 919–941.
[66] R.F. Curtain, Infinite-dimensional filtering, SIAM J. Control 13 (1) (1975) 89–104.