Model Error in Filtering Random Compressible Flows Utilizing Noisy Lagrangian Tracers

Nan Chen*

Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY.

Andrew J. Majda

Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, USA and Center for Prototype Climate Modeling, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE.

**Corresponding author address:* Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA.

E-mail: chennan@cims.nyu.edu

ABSTRACT

Lagrangian tracers are drifters and floaters that collect real-time information of fluid flows. This paper studies the model error in filtering multiscale random rotating compressible flow fields utilizing noisy Lagrangian tracers. The random flow fields are defined through random amplitudes of Fourier eigenmodes of the rotating shallow water equations that contain both incompressible geostrophically balanced (GB) flows and rotating compressible gravity waves, where filtering the slow-varying GB flows is of primary concern. Despite the inherent nonlinearity in the observations with mixed GB and gravity modes, there are closed analytical formulae for filtering the underlying flows. Besides the full optimal filter, two practical imperfect filters are proposed. An information-theoretic framework is developed for assessing the model error in the imperfect filters, which can apply to a single realization of the observations. All the filters are comparably skillful in a fast rotation regime (Rossby number $\varepsilon = 0.1$). In a moderate rotation regime ($\varepsilon = 1$), significant model errors are found in the reduced filter containing only GB forecast model while the computationally efficient 3D-Var filter with a diagonal covariance matrix remains skillful. First linear then nonlinear coupling of GB and gravity modes is introduced in the random Fourier amplitudes while linear forecast models are retained to ensure the filter estimates have closed analytical expressions. All the filters remain skillful in $\varepsilon = 0.1$ regime. In $\varepsilon = 1$ regime, the full filter with a linear forecast model has an acceptable filtering skill while large model errors are shown in the other two imperfect filters.

1. Introduction

Lagrangian tracers are drifters and floaters that collect real-time information of fluid flows, especially at the center of oceans where Eulerian measurements are inaccessible (Griffa et al. 2007; Gould et al. 2004). An important application of Lagrangian data is to recover the current underlying velocity field. To this end, many approximate filters have been developed for assimilation of Lagrangian data (Molcard et al. 2003; Kuznetsov et al. 2003; Apte et al. 2008) and the properties of these filters are studied through numerical experiments (Salman et al. 2006, 2008; Slivinski et al. 2015).

However, due to the complexity and highly nonlinear nature of Lagrangian data assimilation, there was little systematic analysis of the approximate filters based on rigorous theory. Recently, an analytically tractable nonlinear filtering framework for Lagrangian data assimilation was developed (Chen et al. 2014b, 2015), which allows the study of random incompressible/compressible flow field with full mathematical rigor. In this framework, the turbulent flow field is defined through a finite number of random Fourier modes, which are coupled through the tracer observations in a highly nonlinear way. The key fact is that the resulting signal-observation process forms a conditional Gaussian system conditioned on the observations. Despite the inherent nonlinearity in measuring noisy Lagrangian tracers, it was shown that there are exact closed analytic formulae for the optimal filter in filtering the velocity field involving Riccati equations with random coefficients for the covariance matrix. In (Chen et al. 2014b), this Lagrangian data assimilation framework was applied to random incompressible flows, where a practical information barrier in increasing the number of tracers was revealed. In (Chen et al. 2015), the filtering framework was applied to a realistic multiscale random compressible flow field that is a linear combination of random incompressible geostraphically balanced (GB) flows and random rotating compressible gravity waves. In addition to the full optimal filter, an idealized GB filter, serving as a reference for filtering the slow-varying GB flows, and a practical suboptimal filter with mode reduction in the forecast model were studied. Rigorous theorems through suitable stochastic fast-wave averaging techniques and explicit formulas demonstrated that all these filters have comparably high skill in recovering the slow GB flows in the limit of small Rossby number $\varepsilon \to 0$ for any bounded time interval (Chen et al. 2015).

Since simplifications and approximations are ubiquitous in designing filters, a central practical issue is to understand the model error by utilizing imperfect filters for assimilation of Lagrangian data (Majda 2012; Majda and Harlim 2012). This requires assessing the lack of information in the filter estimate utilizing imperfect filters related to that utilizing perfect one. Yet, despite the application of recursive Bayesian estimation in Lagrangian data assimilation, the filtering skill in the previous works (Salman et al. 2006, 2008; Slivinski et al. 2015; Chen et al. 2014b, 2015) was evaluated mostly based on the path-wise RMS error in the posterior mean estimation, where the uncertainty represented by the posterior covariance was completely ignored. Clearly, a moderate error in the posterior mean estimation utilizing imperfect filters with a tiny posterior covariance is of particular danger since it implies the biased estimation is falsely trusted with high certainty. Likewise, a strongly overestimated posterior covariance utilizing imperfect filters provides little information even if the posterior mean is quite close to that utilizing perfect one. Therefore, it is important to develop a systematic framework for assessing the model error in imperfect filters based on the lack of information in the full posterior distribution.

Below, an information-theoretic framework (Branicki et al. 2013; Majda and Wang 2006; Majda and Branicki 2012; Branicki and Majda 2014) is developed to assess the model error in imperfect filters for filtering the multiscale random rotating compressible flows, which can apply to a single realization of the observations. The lack of information in the posterior distribution utilizing imperfect filters related to that utilizing perfect filter is measured through an information metric, named as the relative entropy (Majda and Wang 2006; Majda et al. 2002), which takes into account not only the error in the mean state estimation but the uncertainty in the filter estimates as well.

Following the general nonlinear filtering framework (Chen et al. 2014b, 2015), the idealized flow fields of the multiscale random rotating compressible flows studied here are defined through random amplitudes of Fourier eigenmodes of the rotating shallow water equations, which involve both the incompressible GB flows and the rotating compressible gravity waves. To ensure the filter estimates of the perfect full filter having closed analytic expressions that facilitates the study of the information model error, linear and independent stochastic dynamics are adopted for the random amplitudes of different modes. These assumptions are often utilized in tests for Lagrangian data assimilation (Kuznetsov et al. 2003; Apte et al. 2008; Slivinski et al. 2015). Despite such decoupling in the true underlying flow fields and thus in the perfect forecast model, the GB and gravity modes are nevertheless coupled in a highly nonlinear way through the tracer observations. Note that many geophysical scenarios involve fast rotating flows, where the Rossby number $\varepsilon \ll 1$ (Vallis 2006). Thus, the random rotating shallow water equations become a slow-fast system and the primary objective in practice is to recover the GB component that dominates the slow-varying geophysical flows (Rossby 1937; Gill 1982; Majda 2003; Cushman-Roisin and Beckers 2011) from the noisy Lagrangian tracer observations.

In addition to the full optimal filter, an idealized GB filter involving only the GB dynamics in the forecast model and artificial noisy observations associated with the GB flow is developed, serving as a reference for filtering the slow-varying GB flows (Chen et al. 2015). Two practical imperfect filters are proposed. First, formally applying the mode reduction (Majda et al. 2003, 1999) to the gravity waves results in a suboptimal filter that contains only the GB dynamics in the forecast model while the noisy observations nevertheless include the coupled GB and gravity modes as in the perfect full filter. This dimension reduction strategy in the forecast model simplifies the filter structure and saves the computational cost. Another practical reduced filter includes the full GB and gravity dynamics in the forecast model but the posterior covariance is assumed to be diagonal. The special structure of such reduced filter leads to a constant diagonal covariance matrix after a short relaxation time and therefore it becomes a 3D-Var type of filter (Navon 2009). Since this diagonal reduced 3D-Var filter only requires the update of the posterior mean, it is computationally efficient. Below, the comparison of the filtering skill and the information model error by utilizing these two reduced imperfect filters will be extensively studied in different dynamical regimes.

Another central issue in this paper involves studying a more complicated and realistic flow field. Recall that the random Fourier amplitudes associated with the GB and gravity modes as discussed above are assumed to evolve independently with each other. Yet, both the mathematical theory of the slow-fast geophysical flows (Embid and Majda 1998; Majda 2003; Gershgorin and Majda 2008) and high resolution of turbulent simulations in slow-fast geophysical regimes (Smith 2001; Smith and Waleffe 2002; Waite and Bartello 2004) indicate the interactive effect between the GB and gravity modes. Therefore, following the theory in (Embid and Majda 1998; Majda 2003; Gershgorin and Majda 2008), a quadratic nonlinear interaction between the GB mode and the two gravity modes with the same Fourier wavenumber is incorporated into the underlying dynamics of the random amplitudes associated with the gravity modes while the GB flow remains evolving independently. However, the perfect filter including the nonlinear forecast model for the random Fourier amplitudes breaks the conditional Gaussian filtering framework in (Chen et al. 2014b, 2015). Thus, the same linear stochastic forecast models where different modes evolve independently as described above are utilized for filtering the nonlinearly coupled flow field, which ensure the filter estimates have closed analytical expressions. Despite this intrinsic model error, such simplification is a common strategy for filtering large dimensional turbulent systems in many practical issues, such as utilizing the extended Kalman filter (Haykin 2004) or adopting the mean stochastic forecast model in filtering (Majda and Harlim 2012; Harlim and Majda 2013). Note that the observational process here remains highly nonlinear and thus the coupled signal-observation system still forms a nonlinear filter. It is of practical importance to understand the effect of model error by dropping the nonlinear coupling between different modes in the forecast models for filtering the random rotating compressible flows with nonlinearly coupled GB and gravity modes in different dynamical regimes.

The remainder of this paper is organized as follows. In Section 2, the multiscale random rotating compressible shallow water flows are described and the analytically tractable nonlinear Lagrangian data assimilation framework is introduced. The description of the four filters is also included in the same section. In Section 3, a general information-theoretic framework for assessing the model error in imperfect filters is developed. Section 4 starts with describing a simple setup of the GB flow field with diverse flow structures varying in time, which is followed by the filtering skill and information model error in filtering the multiscale random rotating compressible flows. Specifically, Section 4c deals with the situation where the GB and gravity modes evolve independently while Section 4d handles the flow field where the underlying dynamics of the random Fourier coefficients contains the nonlinear interaction between GB and gravity modes. Section 5 contains the concluding discussion.

2. Basic set-up

a. Random rotating compressible shallow water flows

The 2-dimensional (2D) random rotating compressible shallow water flows are described in the following way:

$$\begin{bmatrix} \vec{v}(\vec{x},t) \\ h(\vec{x},t) \end{bmatrix} = \sum_{\vec{k}\in K, \alpha\in\{0,\pm\}} \hat{v}_{\vec{k},\alpha}(t) \exp(\mathrm{i}\vec{k}\cdot\vec{x}) \vec{r}_{\vec{k},\alpha}, \tag{1}$$

where $\vec{v}$ is the 2D velocity field and h is the height function. In (1), K is some finite symmetric subset of $\mathbb{Z}^2$, while modes with $\alpha = 0$ represent the geostrophic balanced (GB) part and modes with $\alpha = \pm$ represent the two gravity waves. The vectors $\vec{r}_{\vec{k},\alpha}$ are the eigenvectors associated with different modes, where the projection of $\vec{r}_{\vec{k},0}$ on the velocity components is perpendicular to $\vec{k}$ due to the incompressibility of the GB part (Majda 2003; Embid and Majda 1998; Majda and Embid 1998) and $\vec{r}_{\vec{k},\pm}$ indicate the direction of the compressible gravity waves. The turbulent nature of the underlying flow field is reflected in the wave amplitudes $\hat{v}_{\vec{k},\alpha}(s)$ with stochastic forcing and damping terms (Majda and Harlim 2012; Chen et al. 2015),

$$\mathrm{d}\hat{v}_{\vec{k},0}(t) = \big(-d_B\hat{v}_{\vec{k},0} + f_{\vec{k},0}(t)\big)\mathrm{d}t + \sigma_{\vec{k},0}\mathrm{d}W_{\vec{k},0}(t), \tag{2a}$$

$$\mathrm{d}\hat{v}_{\vec{k},\pm}(t) = \big((-d_g + \mathrm{i}\omega_{\vec{k},\pm} \underline{\pm\mathrm{i}\gamma\hat{v}_{\vec{k},0}})\hat{v}_{\vec{k},\pm}(t) + f_{\vec{k},\pm}(t)\big)\mathrm{d}t + \sigma_{\vec{k},\pm}\mathrm{d}W_{\vec{k},\pm}(t), \tag{2b}$$

where the GB modes $\hat{v}_{\vec{k},0}$ are assumed to be real and the gravity modes $\hat{v}_{\vec{k},\pm}$ are complex. In (2), $\omega_{\vec{k},\pm}$ are the oscillation frequencies of the gravity modes, the details of which will be given in (6), $d_B, d_g > 0$ are damping coefficients, $\sigma_{\vec{k},0}, \sigma_{\vec{k},\pm} > 0$ are stochastic forcing amplitudes and $f_{\vec{k},0}, f_{\vec{k},\pm}$ are deterministic forcing. To guarantee the full flow fields in (1) to be real-valued, we require that $\vec{r}^{\,*}_{\vec{k},\alpha} = \vec{r}_{-\vec{k},-\alpha}$ and $(\hat{v}_{\vec{k},\alpha})^* = \hat{v}_{-\vec{k},-\alpha}$. The equality for the eigenvectors are automatically satisfied which will be discussed below in (5), (7) and (8) and the equality for the Fourier coefficients associated with the gravity modes are enforced by requiring each term in (2b) for the two gravity wave pairs being complex conjugate. For a detailed description of this enforcement, we refer to Appendix A.1 of (Chen et al. 2014b). Note that such a modeling strategy for random turbulence has been widely applied in many other situations (Majda and Harlim 2012). The effect of the slow GB mode on the fast gravity modes is reflected on the nonlinear coupling term with coefficient γ in (2b), which is motivated directly from mathematical theory of the slow-fast geophysical flows (Embid and Majda 1998; Majda 2003; Gershgorin and Majda 2008) and high resolution of turbulent simulations in slow-fast geophysical regimes (Smith 2001; Smith and Waleffe 2002; Waite and Bartello 2004). The situation with $\gamma = 0$ in (2b) implies utilizing linear stochastic model to describe the random Fourier coefficients, where the GB and gravity modes are independently with each other. Despite the linear dynamics associated with each Fourier mode, the stochastic forcing and damping terms compensate the nonlinearity in nature and the full velocity field remains highly turbulent. Some path-wise behaviors of the situation with such uncoupled GB and gravity modes was discussed in (Chen et al. 2015). In this paper, the information model error and the path-wise filtering skill of both linearly independent ($\gamma = 0$) and nonlinearly coupled ($\gamma \neq 0$) GB and gravity modes will be studied.

To provide the motivation of the model (1)–(2) and the choices of the eigenvectors $\vec{r}_{\vec{k},\alpha}$ and rotation frequency $\omega_{\vec{k},\pm}$, we recall the linearized shallow water equations in the non-dimensional form (Section 4.4 of (Majda 2003)):

$$\begin{aligned} \frac{\partial \vec{u}}{\partial t} + \varepsilon^{-1}\vec{u}^{\perp} &= -\varepsilon^{-1}\delta^{1/2}\nabla\eta, \\ \frac{\partial \eta}{\partial t} + \varepsilon^{-1}\delta^{1/2}\nabla\cdot\vec{u} &= 0, \end{aligned} \tag{3}$$

where $\vec{u}$ is a horizontal two dimensional velocity field and η is the height function rescaled by $\delta^{1/2}$ to guarantee the symmetric hyperbolic form in (3). We denote the non-dimensional parameters $\varepsilon = \mathrm{Ro}$ and $\delta = \mathrm{Ro}^2\mathrm{Fr}^{-2}$, where Ro is the Rossby number representing the ratio between the Coriolis term and the advection term and Fr is the Froude number. For most atmosphere-ocean problems, ε is a small number representing fast rotation and δ is either $O(1)$ or $O(\varepsilon)$ (Vallis 2006). Following Section 4.4 of (Majda 2003), the general solution of (3) is given by a superposition of plane waves:

$$\begin{bmatrix} \vec{u}(\vec{x},t) \\ \eta(\vec{x},t) \end{bmatrix} = \sum_{\vec{k}\in\mathbb{Z}^2,\alpha\in\{0,\pm\}} \hat{u}_{\vec{k},\alpha} \exp(\mathrm{i}\vec{k}\cdot\vec{x} - \mathrm{i}\omega_{\vec{k},\alpha}t)\vec{r}_{\vec{k},\alpha}. \tag{4}$$

The modes with $\alpha = 0$ represent the geostrophic balanced (GB) modes, also known as the vortical waves, where the geostrophic balance relation $\vec{u}^{\perp} = -\nabla\eta$ always holds (Majda 2003). The associated rotational speed $\omega_{\vec{k},0} = 0$ and the normalized eigenvector $\vec{r}_{\vec{k},0}$ is given by

$$\vec{r}_{\vec{k},0} = \frac{1}{\sqrt{|\vec{k}|^2+1}} \begin{bmatrix} -\mathrm{i}k_2 \\ \mathrm{i}k_1 \\ 1 \end{bmatrix}. \tag{5}$$

The modes with $\alpha = \pm$ represent the gravity modes also known as the Poincaré waves (Majda 2003). They have a nonzero phase speed:

$$\omega_{\vec{k},\pm} = \pm\varepsilon^{-1}\sqrt{\delta|\vec{k}|^2+1}. \tag{6}$$

The associated normalized eigenvectors $\vec{r}_{\vec{k},\pm}$ are given by

$$\vec{r}_{\vec{k},\pm} = \frac{1}{|\vec{k}|\sqrt{(\delta+\delta^2)|\vec{k}|^2+2}} \begin{bmatrix} \mathrm{i}k_2 \pm k_1\sqrt{\delta|\vec{k}|^2+1} \\ -\mathrm{i}k_1 \pm k_2\sqrt{\delta|\vec{k}|^2+1} \\ \delta|\vec{k}|^2 \end{bmatrix}. \tag{7}$$

For the special case, $\vec{k} = \vec{0}$, the Poincaré waves have no gravity component and coincide with the inertial waves. The resulting eigenvalues become $\omega_{\vec{0},\pm} = \pm\varepsilon^{-1}$ with the eigenvectors

$$\vec{r}_{\vec{0},+} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathrm{i} \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{r}_{\vec{0},-} = \frac{1}{\sqrt{2}} \begin{bmatrix} -\mathrm{i} \\ 1 \\ 0 \end{bmatrix}. \tag{8}$$

By taking a finite Fourier truncation and replacing the deterministic coefficients $\hat{u}_{\vec{k},\alpha}\exp(-\mathrm{i}\omega_{\vec{k},\alpha}t)$ in (4) with the stochastic processes $\vec{v}_{\vec{k},\alpha}$ modeled by (2), we arrive at the basic rotating compressible random field model in (1)–(2). The additional linear coefficients $\mathrm{i}\omega_{\vec{k},\pm}$ from (6) describe the oscillations of the gravity waves, where a small ε corresponds to a fast rotation. It is worth noticing that ε and δ enter the dynamics only through the gravity waves in $\vec{r}_{\vec{k},\pm}$ and $\omega_{\vec{k},\pm}$. Moreover, $\omega_{\vec{k},\pm}$ is a parameter of order ε^{-1}; its appearance in the linear coefficient for the gravity modes (2) represents the same rotational effect as in the deterministic setting.

b. *Observation process from noisy Lagrangian tracers*

The observations are from the trajectories of L Lagrangian tracers transported by the underlying velocity field with additional noise. The observation process is given by

$$\begin{aligned}\mathrm{d}\vec{X}_l(s) &= \vec{v}(\vec{X}_l(s),s)\mathrm{d}s + \sigma_x \mathrm{d}W_l^x(s)\\ &= \sum_{\vec{k}\in K,\alpha\in\{0,\pm\}} \hat{v}_{\vec{k},\alpha}(t)\exp(\mathrm{i}\vec{k}\cdot\vec{X}_l(s))\mathscr{P}_v\vec{r}_{\vec{k},\alpha}\mathrm{d}s + \sigma_x\mathrm{d}W_l^x(s), \quad l=1,\ldots,L,\end{aligned} \tag{9}$$

where Newton's law is applied in the first row of (9) and the second row is due to (1), where the operator $\mathscr{P}_v$ is the projection of a 3D vector to its first two dimension entries. The noise amplitude σ_x in different tracers is assumed to be the same but the noise W_l^x itself is independent for different l.

c. *Filters for noisy Lagrangian tracers*

Given the observations from the noisy Lagrangian tracers (9), the goal is to filter the underlying flow field $\vec{v}(\vec{x},t)$ in (1), or equivalently the Fourier coefficients $\hat{v}_{\vec{k},\alpha}(t)$ for all $\vec{k}$ and α. For simplicity of notations, we define $\mathbf{k} = \{\vec{k},\alpha\} \in \mathbf{K}$ such that the Fourier coefficient and the eigenvector in (1) can be written as $\hat{v}_{\mathbf{k}}(\vec{x},t)$ and $\vec{r}_{\mathbf{k}}$, respectively. Furthermore, we define $\mathbf{k}_B = \{\vec{k},0\} \in \mathbf{K}_B$ and $\mathbf{k}_g = \{\vec{k},\pm\} \in \mathbf{K}_g$ representing the GB and gravity modes, respectively.

Recall that each trajectory of the noisy Lagrangian tracers is given by (9). Let's group all $\vec{X}_l(s)$ into one $1\times 2L$ column vector

$$\mathbf{X}_s = \begin{bmatrix} \vec{X}_1(s) \\ \vdots \\ \vec{X}_L(s) \end{bmatrix}.$$

Then the abstract form of the observation process for the L noisy Lagrangian tracers follows:

$$\mathrm{d}\mathbf{X}_s = \mathbf{P}_X(\mathbf{X}_s)\mathbf{U}_s\mathrm{d}s + \sigma_x \mathrm{d}W_s^x, \tag{10}$$

where W_s^x is a $2L\times 2L$ diagonal matrix and $\mathbf{P}_X(\mathbf{X}_s)$ is given by, according to (9),

$$\mathbf{P}_X(\mathbf{X}_s) = \begin{bmatrix} P_X(\vec{X}_1(s)) \\ \vdots \\ P_X(\vec{X}_L(s)) \end{bmatrix} = \begin{bmatrix} \cdots & \exp(\mathrm{i}\vec{k}\cdot\vec{X}_1(s))\vec{r}_{\mathbf{k}} & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \exp(\mathrm{i}\vec{k}\cdot\vec{X}_L(s))\vec{r}_{\mathbf{k}} & \cdots \end{bmatrix} := [\mathbf{P}_X^B(\mathbf{X}_s), \mathbf{P}_X^g(\mathbf{X}_s)]. \tag{11}$$

With a slight abuse of the notation, $\vec{r}_{\mathbf{k}}$ in (11) is the eigenvector that has only the first two entries corresponding to the 2D velocity directions.

On the other hand, by formally applying mode reduction over the gravity waves, it is possible to write down the simplified random flow field that contains only GB part of the flow. The corresponding noisy Lagrangian tracers transported by only the GB flow can be formally constructed,

$$\mathrm{d}\vec{X}_l^B(s) = P_X^B(\vec{X}_l^B(s))\mathbf{U}_s^B\mathrm{d}s + \sigma_x\mathrm{d}W_l^B(s), \quad l = 1,\ldots,L.$$

Similar to (10), the abstract form by collecting all L tracers transported by the GB flow is given by

$$\mathrm{d}\mathbf{X}_s^B = \mathbf{P}_X^B(\mathbf{X}^B(s))\mathbf{U}_s^B\mathrm{d}s + \sigma_x\mathrm{d}W_s^B. \tag{12}$$

Note that (12) is an artificial observation process since it is practically impossible to extract the component that corresponds to the random GB signals from the full noisy observations.

With the observation processes in (10) or (12), what remains is to design the forecast models in filters for the velocity field. Recall the dynamics associated with the true velocity field in (2).

In the situation with uncoupled GB and gravity modes, i.e., $\gamma = 0$, the underlying dynamics of the Fourier coefficients for wavenumber $\vec{k}$ associated with the flow field $\vec{v}(\vec{x},t)$ in (1) reduces to a linear stochastic system

$$\mathrm{d}\hat{v}_{\vec{k},0}(t) = \big(-d_B\hat{v}_{\vec{k},0} + f_{\vec{k},0}(t)\big)\mathrm{d}t + \sigma_{\vec{k},0}\mathrm{d}W_{\vec{k},0}(t), \tag{13a}$$

$$\mathrm{d}\hat{v}_{\vec{k},\pm}(t) = \big((-d_g + \mathrm{i}\omega_{\vec{k},\pm})\hat{v}_{\vec{k},\pm}(t) + f_{\vec{k},\pm}(t)\big)\mathrm{d}t + \sigma_{\vec{k},\pm}\mathrm{d}W_{\vec{k},\pm}(t). \tag{13b}$$

As was done for the tracers, the Fourier coefficients for all the GB modes in (13a) and all the gravity modes in (13b) can be grouped into a $1\times|K|$ and a $2\times|K|$ column vector, respectively. Then, the corresponding dynamics of the GB and gravity modes becomes

$$\mathrm{d}\mathbf{U}_s^B = -\Gamma^B\mathbf{U}_s^B\mathrm{d}s + F_s^B\mathrm{d}s + \Sigma_u^B\mathrm{d}W_u^B(s), \tag{14a}$$

$$\mathrm{d}\mathbf{U}_s^g = (-\Gamma^g + \mathrm{i}\Omega_\varepsilon)\mathbf{U}_s^g\mathrm{d}s + F_s^g\mathrm{d}s + \Sigma_u^g\mathrm{d}W_u^g(s), \tag{14b}$$

and jointly:

$$\mathrm{d}\mathbf{U}_s = -\Gamma\mathbf{U}_s\mathrm{d}s + F_s\mathrm{d}s + \Sigma_u\mathrm{d}W_u(s), \tag{15}$$

where Ω_ε in (14b) is a diagonal matrix and its $\mathbf{k}$-th entry is given by, according to (6),

$$\omega_{\mathbf{k}} = \pm\varepsilon^{-1}\sqrt{\delta|\vec{k}|^2+1}, \qquad \mathbf{k}\in\mathbf{K}_g, \tag{16}$$

and Γ in (15) involves both the damping Γ^B, Γ^g and the oscillation frequency $\mathrm{i}\Omega_\varepsilon$.

Utilizing the perfect dynamics of the underlying flow field (15) as the forecast model in the filter, the joint observation-signal system (10) and (15) becomes a conditional Gaussian system given the observations. For such kind of system with Gaussian initial conditions, the conditional distribution of the flow field given the observed noisy Lagrangian tracer trajectories, knowing as posterior distribution, is a Gaussian distribution where the evolutions of the conditional mean and conditional covariance have closed analytic formulae (Liptser and Shiryaev 2001). See Appendix

A for details. This provides an exact and accurate perfect filter for recovering the underlying velocity field.

In the situation where the GB modes affect the gravity modes in a nonlinear way, i.e., $\gamma \neq 0$ in (2), the perfect observation-signal system is no longer a conditional Gaussian system since given the observations the underlying dynamics (2) is a quadratic nonlinear system with non-Gaussian statistics, which breaks the analytically tractable filtering framework in (Chen et al. 2014b, 2015). Due to the high dimensionality of the coupled signal-observation system, it is computationally unaffordable to solve the posterior distribution via direct numerical methods. Thus, in the appearance of the nonlinearly coupled GB and gravity modes ($\gamma \neq 0$) in the true velocity field (2), the linear system in (13) for each Fourier wavenumber is nevertheless utilized as the forecast model in the designed filters to maintain the analytically solvable feature of the filters. Despite this intrinsic model error, such simplification is a common strategy for filtering large dimensional turbulent systems in many practical issues, where the linearized methods such as the extended Kalman filter (Haykin 2004) or the mean stochastic forecast model (Majda and Harlim 2012; Harlim and Majda 2013) are widely adopted. An important practical issue is to understand the effect of model error due to adopting a linear stochastic forecast models with independent GB and gravity components to filter the random rotating compressible flows with nonlinearly coupled GB and gravity modes. Note that, the tracer trajectories in (10) is transported by the true nonlinear dynamics (2) while the linear stochastic turbulent system in (13) is only utilized as the forecast model in the filters.

In the following, four different filters, which all belong to the conditional Gaussian framework, are designed. Their filtering skill will be extensively studied in Section 4.

1) Full filter with linear forecast dynamics

Utilizing the nonlinear observation process (10) and the linear dynamics with independent GB and gravity modes as the forecast model (14), the full filter with linear forecast dynamics is given by

$$
\begin{aligned}
\mathrm{d}\mathbf{X}_s &= \mathbf{P}_X(\mathbf{X}_s)\mathbf{U}_s\mathrm{d}s + \sigma_x \mathrm{d}W_s^x,\\
\mathrm{d}\mathbf{U}_s &= -\Gamma\mathbf{U}_s\mathrm{d}s + F_s\mathrm{d}s + \Sigma_u\mathrm{d}W_u(s).
\end{aligned}
\tag{17}
$$

The filter (17) is a perfect optimal filter if the underlying flow of the truth (2) is also linear, i.e., $\gamma = 0$ in (2b). In such case, we simply name (17) as the full filter. Otherwise ($\gamma \neq 0$), model error comes from ignoring the nonlinear coupling of GB and gravity modes in (2b). The analytic solution of updating the posterior mean and posterior covariance of $\mathbf{U}_t$ given $\mathbf{X}_{s\leq t}$ is shown in Appendix A.

2) Idealized GB filter

In many practical issues, the primary practical objective is to recover the GB component that dominates the slow-varying geophysical flows (Rossby 1937; Gill 1982; Majda 2003; Cushman-Roisin and Beckers 2011). To this end, an idealized GB filter is constructed based on the GB forecast model (14a) and the artificial observations from only GB part of the flow (12),

$$
\begin{aligned}
\mathrm{d}\mathbf{X}_s^B &= \mathbf{P}_X^B(\mathbf{X}_s^B)\mathbf{U}_s^B\mathrm{d}s + \sigma_x^B \mathrm{d}W_s^B,\\
\mathrm{d}\mathbf{U}_s^B &= -\Gamma^B\mathbf{U}_s^B\mathrm{d}s + F_s^B\mathrm{d}s + \Sigma_u^B\mathrm{d}W_u^B(s).
\end{aligned}
\tag{18}
$$

This idealized GB filter (18) is a perfect filter regardless of the coupling coefficient γ in (2) as the nonlinearity in the underlying flow appears only in the gravity modes.

Since the underlying GB flow is incompressible, the properties of this idealized GB filter was well studied in (Chen et al. 2014b). In addition, without being scrambled by the gravity modes, this perfect GB filter indicates the optimality of filtering the GB flow field. Thus, the results from this idealized GB flow are regarded as a reference for checking the filtering skill utilizing other filters. Note that, despite its optimality, the GB filter is not a practical filter because extracting the observations corresponding only to the random GB part of flow from the full noisy observations is impractical in real applications.

3) Reduced filter with only GB forecast model through full observations

Motivated from the idealized GB filter (18), a practical reduced filter for filtering GB part of the flow is formed by adopting only the GB dynamics (14a) as the forecast model while the coupled GB and gravity observations from noisy Lagrangian tracers are utilized as the input in the observation process. This follows the formal application of the mode reduction strategy (Majda et al. 2003, 1999) to the gravity waves in the forecast model. For consistency, the corresponding dynamics of the observation process contains the modes associated with only the GB flow as well, i.e., replacing $\mathbf{P}_X$ in (10) by $\mathbf{P}_X^B$. Therefore, such reduced filter reads,

$$\begin{aligned} \mathrm{d}\mathbf{X}_s &= \mathbf{P}_X^B(\mathbf{X}_s)\mathbf{U}_s^B\mathrm{d}s + \sigma_x^B\mathrm{d}W_s^B, \\ \mathrm{d}\mathbf{U}_s^B &= -\Gamma^B\mathbf{U}_s^B\mathrm{d}s + F_s^B\mathrm{d}s + \Sigma_u^B\mathrm{d}W_u^B(s). \end{aligned} \tag{19}$$

Since the gravity parts of the flow is dropped from the forecast model in (19), the dimension of the flow field $\mathbf{U}_s^B$ in (19) is only $1/3$ compared with $\mathbf{U}_s$ of the full filter in (17) and in turn the number of the entries in the covariance matrix is only $1/9$ of that associated with the perfect filter. Due to the fact that most of the computational cost lies in the update of the posterior covariance, this reduced filter is more computational efficient than the full filter. Yet, the reduced filter (19) is only a suboptimal filter due to the model error from filtering only GB part of the flow through the full mixed observations.

4) Diagonal reduced 3D-Var filter

Another practical reduced filter includes both the GB dynamics and the linearized gravity dynamics in the forecast model (14), which are the same as the full filter (17), but the posterior cross-covariance is assumed to stay zero and thus it reduces to a diagonal filter. Furthermore, as shown in Appendix B, the diagonal entries in the posterior covariance associated with this diagonal reduced filter converge to constant values after a short relaxation time and therefore only the update of the posterior mean is needed afterwards. Due to the same behavior as the 3D-Var with a constant background error covariance (Navon 2009), this filter is named as a diagonal reduced 3D-Var filter,

$$\begin{aligned} d\mathbf{X}_s &= \mathbf{P}_X(\mathbf{X}_s)\mathbf{U}_s ds + \sigma_x dW_s^x, \\ d\mathbf{U}_s &= -\Gamma\mathbf{U}_s ds + F_s ds + \Sigma_u dW_u(s), \end{aligned} \tag{20}$$

Diagonal posterior covariance matrix.

When the true underlying flow field is linear, i.e., $\gamma = 0$ in (2), the only model error in the diagonal reduced 3D-Var filter (20) comes from the ignoring of the off-diagonal entries in the posterior covariance matrix. If the diagonal entries dominate the posterior covariance matrix, then a comparable filtering skill in the diagonal reduced 3D-Var filter (20) is expected as the full filter (17) but (20) is much more efficient. On the other hand, when $\gamma \neq 0$ in the true underlying flow fields (2), an extra model error in the diagonal reduced 3D-Var filter (20) comes from utilizing the linear forecast model for the gravity modes, which is the same as in the full filter (17).

3. An information-theoretic framework in assessing model error

As discussed in Section 1, due to the inevitable approximations and simplifications in real-world Lagrangian data assimilation, it is of practical importance to assess and understand the model error by utilizing imperfect filters with various simplifications. Note that the traditional approach of measuring the filtering skill is based on the path-wise RMS error which takes into account only the point-wise information in the posterior mean while the information in the posterior covariance that represents the uncertainty in the filter estimate is completely ignored. To assess the lack of information in the posterior distribution of imperfect filters, an information-theoretic framework is developed in this section.

Information theory was widely adopted to measure the lack of information in filtering and prediction utilizing imperfect models (Majda and Gershgorin 2010, 2011b,a; Majda et al. 2002; Kleeman 2002). Recently, a systematic information-theoretic approach was developed in (Branicki and Majda 2014) to quantify the statistical accuracy of Kalman filters with model error and the optimality of the imperfect Kalman filters in terms of three information measures was presented. Another application of information theory is illustrated in (Branicki and Majda 2015) for improving imperfect predictions via multi-model ensemble forecasts. Information measures were also adopted for model calibration in predicting the real-time indices of the Madden-Julian oscillation (Chen and Majda 2015d), which shows the significant skill of capturing the extreme events that cannot be assessed by the path-wise RMS error and pattern correlation.

Here, the information model error is assessed through the relative entropy (Majda and Wang 2006; Majda et al. 2002),

$$\mathscr{P}(p,q) = \int p \ln \frac{p}{q}, \tag{21}$$

which measures the lack of information in the probability distribution function (PDF) associated with the imperfect model q related to that of the perfect system p. The relative entropy is often interpreted as a 'distance' between the two probability densities but it is not a true metric. It is non-negative with $\mathscr{P} = 0$ only when $p = q$ and it is invariant under nonlinear changes of variables.

Note that when both $p \sim \mathscr{N}(\vec{m}_p, R_p)$ and $q \sim \mathscr{N}(\vec{m}_q, R_q)$ in (21) are Gaussian, the relative entropy has the closed form:

$$\mathscr{P}(p,q) = \left[\tfrac{1}{2}(\vec{m}_p - \vec{m}_q)^T R_q^{-1} (\vec{m}_p - \vec{m}_q)\right] + \tfrac{1}{2}\left[\mathrm{tr}(R_p R_q^{-1}) - N - \ln\det(R_p R_q^{-1})\right], \tag{22}$$

where N is the dimension of both the distributions. The first term in brackets in (22) is called the "*signal*", which measures the lack of information in the mean weighted by model covariance. The second term in brackets is called the "*dispersion*", which involves only the covariance ratio.

Now we develop an information-theoretic framework to measure the model error in imperfect filters. Consider a coupled system with variables $(\mathbf{u_I}, \mathbf{u_{II}})$, where $\mathbf{u_I}$ stands for observations and $\mathbf{u_{II}}$ represents the variables for filtering. Let's denote p and p^M the PDFs of the perfect and the imperfect models, respectively. In a typical scenario, the imperfect model is coarse-grained and thus we assume the distribution p^M is formed only by the conditional moments up to L. Let's further denote p_L the PDF that is reconstructed utilizing the L conditional moments of the perfect model. Then the joint distributions regarding $\mathbf{u_I}$ and $\mathbf{u_{II}}$ can be written as

$$p(\mathbf{u_I}, \mathbf{u_{II}}) = p(\mathbf{u_{II}}|\mathbf{u_I})\pi(\mathbf{u_I})$$

$$p_L(\mathbf{u_I}, \mathbf{u_{II}}) = p_L(\mathbf{u_{II}}|\mathbf{u_I})\pi(\mathbf{u_I})$$

$$p^M(\mathbf{u_I}, \mathbf{u_{II}}) = p_L^M(\mathbf{u_{II}}|\mathbf{u_I})\pi^M(\mathbf{u_I}),$$

According to (Branicki et al. 2013), the lack of information in the imperfect model related to the perfect one is given by

$$\begin{aligned}&\mathscr{P}(p(\mathbf{u_I}, \mathbf{u_{II}}), p_L^M(\mathbf{u_I}, \mathbf{u_{II}}))\\ &\qquad = \mathscr{P}(p(\mathbf{u_I}, \mathbf{u_{II}}), p_L(\mathbf{u_I}, \mathbf{u_{II}})) + \mathscr{P}(p_L(\mathbf{u_I}, \mathbf{u_{II}}), p_L^M(\mathbf{u_I}, \mathbf{u_{II}})),\end{aligned} \tag{23}$$

where the first term on the right hand side of (23) is called the intrinsic barrier that measures the lack of information in the perfect model due to the coarse-grained effect from the insufficient measurement and the second term is the model error where the imperfect model is compared with the perfect model that possesses the same number of moments. Direct calculation (Branicki et al. 2013) shows that

$$\text{Intrinsic barrier} = \int \pi(\mathbf{u_I})\left(\mathscr{S}(p_L(\mathbf{u_{II}})) - \mathscr{S}(p(\mathbf{u_{II}}))\right), \tag{24}$$

$$\text{Model error} = \mathscr{P}(\pi(\mathbf{u_I}), \pi^M(\mathbf{u_I})) + \int \pi^M(\mathbf{u_I})\mathscr{P}(p_L(\mathbf{u_{II}}|\mathbf{u_I}), p_L^M(\mathbf{u_{II}}|\mathbf{u_I}))d\mathbf{u_I}, \tag{25}$$

where $\mathscr{S}$ is the Shannon's entropy (Majda and Wang 2006). In filtering the state variables $\mathbf{u_{II}}$, we assume the observations in the imperfect model $\pi^M(\mathbf{u_I}(s \leq t))$ is the same as those in the perfect model $\pi(\mathbf{u_I}(s \leq t))$. Therefore, the first term in the model error (25) disappears and $\pi^M(\mathbf{u_I})$ in the second term is replaced by $\pi(\mathbf{u_I})$, which simplifies the model error in (25),

$$\text{Model error} = \int \pi(\mathbf{u_I})\mathscr{P}(p_L(\mathbf{u_{II}}|\mathbf{u_I}), p_L^M(\mathbf{u_{II}}|\mathbf{u_I}))d\mathbf{u_I}. \tag{26}$$

Model error for a single realization of the observations.

In filtering the random compressible flow, only one single realization of the observational trajectory associated with each tracer $\mathbf{u}^i_{\mathbf{I}}(s \leq t), i = 1, \ldots, L$ is given. Thus, we simply need to assess the following model error

$$\mathscr{E}(t) = \mathscr{P}(p_L(\mathbf{u_{II}}(t)|\mathbf{u_I}(s)), p_L^M(\mathbf{u_{II}}(t)|\mathbf{u_I}(s))), \qquad 0 \leq s \leq t. \tag{27}$$

In Section 4c, when the underlying flow field is generated from system (2) with decoupled GB and gravity modes, i.e., $\gamma = 0$, the full filter (17) is a perfect filter. Since we have also assumed the observations in the two reduced filters (19) and (20) are the same as those in the full filter, the formula (27) is applied to compute the model error at each time t, where $p_L(\mathbf{u_{II}}(t)|\mathbf{u_I}(s))$ is the posterior distribution of the perfect full filter (17) and $p_L^M(\mathbf{u_{II}}(t)|\mathbf{u_I}(s))$ is that of one of the imperfect filters (19) or (20). Note that all the three filters are conditional Gaussian filters. Thus $L = 2$ in (27) and the model error is splitted into signal and dispersion as described in (22).

On the other hand, in Section 4d, when the underlying flow field is generated from the system with nonlinearly coupled GB and gravity modes, i.e., $\gamma \neq 0$ in (2), the full filter with linear dynamics (17) is no longer a perfect filter. Two alternative approaches are applied to assess the model error in the imperfect filters. In the first method, we assess the model error in the posterior mean estimate of the imperfect filters compared with the true signal. Here, we adopt the general relative entropy formula (21), where p is the time-averaged PDF of the true signal and q is the time-averaged PDF associated with the posterior mean estimation from one of the imperfect filters. Although this model error measures the lack of information based only on the posterior mean, it is nevertheless different from the path-wise RMS error. In fact, this information metric takes into account the spread of both the posterior mean time series and the true signal. Therefore, it is able to quantify the skill of the imperfect filters in capturing the extreme events in the true signal, which is not accessible by the path-wise RMS error and pattern correlation (Chen and Majda 2015d). The second approach involves formally applying the posterior distribution of the idealized GB filter (18) to p_L in (27). Yet, since the observations in the GB filter are different from those in the three imperfect filters, this argument becomes only an approximation in assessing the model error in the filter estimates utilizing the imperfect filters related to that utilizing the perfect one within the information-theoretic framework developed in (27). In Appendix C, we compare the information model error by utilizing either the full filter (17) or the GB filter (18) as the reference distribution p in (27) in the situation with $\gamma = 0$ to justify that the approximation error due to adopting GB filter as the reference filter is acceptable in studying the information model error of the imperfect filters in the dynamics regimes of interest.

4. Numerical experiments

a. Simple GB flow with time-varying flow structures

An interesting and realistic GB flow field involves time-varying flow structures. The simplest setup of such GB flow consists of 5 Fourier wavenumbers, where $k = (0,0), (\pm 1, 0)$ and $(0, \pm 1)$. Since the eigenvector (5) corresponding to the GB mode $k = (0,0)$ has only non-zero entry in h direction, the underlying GB flow is essentially driven by the 4 modes with $|\vec{k}| = 1$, i.e.,

$$\vec{v} = \sum_{|\vec{k}|=1} \hat{v}_{\vec{k}}(t) \exp(i\vec{k} \cdot \vec{x}) \mathscr{P}_v \vec{r}_{\vec{k}}, \tag{28}$$

where for notation simplicity we have dropped the subscript $\cdot_0$ that distinguishes GB flows from gravity waves. To look at the flow structures of the GB flow field (28), we write down the eigenvectors (5) projected on the horizonal and vertical velocity directions,

$$\begin{aligned} \mathscr{P}_v \vec{r}_{(1,0)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ i \end{pmatrix}, \qquad \mathscr{P}_v \vec{r}_{(-1,0)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ -i \end{pmatrix}, \\ \mathscr{P}_v \vec{r}_{(0,1)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 0 \end{pmatrix}, \qquad \mathscr{P}_v \vec{r}_{(0,-1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 0 \end{pmatrix}. \end{aligned} \tag{29}$$

In addition, the four Fourier bases in (28) with $|\vec{k}| = 1$ are given by

$$\begin{aligned} k = (1,0):& \quad \exp(ix) = \cos(x) + i\sin(x), \quad k = (-1,0): \quad \exp(-ix) = \cos(x) - i\sin(x) \\ k = (0,1):& \quad \exp(iy) = \cos(y) + i\sin(y), \quad k = (0,-1): \quad \exp(-iy) = \cos(y) - i\sin(y). \end{aligned} \tag{30}$$

Inserting (29) and (30) into (28), the horizonal and vertical velocities (v_1, v_2) are given by

$$\begin{aligned} \sqrt{2} v_1 &= \hat{v}_{(0,1)} \cdot (-i) \cdot (\cos(y) + i\sin(y)) + \hat{v}_{(0,-1)} \cdot (i) \cdot (\cos(y) - i\sin(y)) \\ &= \hat{v}_{(0,1)}(-i\cos(y) + \sin(y)) + \hat{v}_{(0,-1)}(i\cos(y) + \sin(y)), \\ \sqrt{2} v_2 &= \hat{v}_{(1,0)} \cdot (i) \cdot (\cos(x) + i\sin(x)) + \hat{v}_{(-1,0)} \cdot (-i) \cdot (\cos(x) - i\sin(x)) \\ &= \hat{v}_{(1,0)}(i\cos(x) - \sin(x)) + \hat{v}_{(-1,0)}(-i\cos(x) - \sin(x)). \end{aligned} \tag{31}$$

Since the Fourier coefficients associated with the GB modes are assumed to be real, we have $\hat{v}_{(0,1)} = \hat{v}_{(0,-1)}$ and $\hat{v}_{(1,0)} = \hat{v}_{(-1,0)}$, which simplify (31),

$$\begin{aligned} v_1 &= \sqrt{2}\hat{v}_{(0,1)} \sin(y), \\ v_2 &= -\sqrt{2}\hat{v}_{(1,0)} \sin(x). \end{aligned} \tag{32}$$

and the corresponding stream function is given by

$$\psi = -\sqrt{2}\hat{v}_{(1,0)} \cos(x) + \sqrt{2}\hat{v}_{(0,1)} \cos(y). \tag{33}$$

Thus, we only need to look at the amplitude of the two coefficients $\hat{v}_{(0,1)}$ and $\hat{v}_{(1,0)}$ to obtain the structure of the GB flow. With different choices of $\hat{v}_{(0,1)}$ and $\hat{v}_{(1,0)}$, the streamlines illustrate various profiles that switch between

1. simple shear flow: $\hat{v}_{(0,1)} \ll 1, \hat{v}_{(1,0)} \sim O(1)$ or $\hat{v}_{(1,0)} \ll 1, \hat{v}_{(0,1)} \sim O(1)$,
2. 2D array of swirling eddies: $\hat{v}_{(0,1)} \approx \hat{v}_{(1,0)} \sim O(1)$, and
3. swirling eddies embedded in a shear-flow stream: $\hat{v}_{(0,1)} \sim O(1), \hat{v}_{(1,0)} \sim O(1)$ but $\hat{v}_{(0,1)} \not\approx \hat{v}_{(1,0)}$.

See Chapter 1 of (Majda and Wang 2006) for more detailed description.

b. Filter setup

As in Section 4a, the underlying flow field contains 5 Fourier wavenumbers with $|\vec{k}| \leq 1$. Thus, the total number of GB and gravity modes is $|\mathbf{K}_B| = 5$ and $|\mathbf{K}_g| = 10$, respectively. In most realistic situations, the number of the observations is typically less than the degree of freedom of the underlying system. Thus, we set the number of the tracers $L = 5 < 15 = |\mathbf{K}|$. The observation noise level is set to be $\sigma_x = 0.2$, which is a moderate value, implying that the filters make use of the information in both the forecast models and the observations.

The GB mode at the largest scale $k=(0,0)$ is set to be deterministic while the other four GB modes with $|\vec{k}|=1$ and all the 10 gravity modes are stochastic. The damping and stochastic forcing coefficients are determined in the situation with uncoupled GB and gravity modes, i.e., $\gamma=0$ in (2), and the same values are adopted in the coupled case. The energy in each stochastic GB mode is set to be $E_B=0.3$ and that in each gravity mode is $E_g=0.1$. A relatively small damping $d_B=d_g=0.05$ is utilized for all the stochastic modes, which correspond to a moderately long decorrelation time $\tau=20$ non-dimensional units in the uncoupled flow case. The stochastic forcing in each GB mode is computed by utilizing the formula $\sigma_{\vec{k},0}^2/(2d_B)=E_B$ and similar for that in each gravity mode.

In addition to the typical values mentioned above, the filtering skill dependence on different parameters is of particular interest. Below, the filtering dependence on the number of tracers L, the observation noise σ_x and the energy in the gravity modes E_g will be explored. In each experiment, only one parameter is varied and the others are all set to be their typical values.

The deterministic forcing are chosen in two different ways:

1. Zero deterministic forcing. In this setup, the flow is purely driven by the stochastic forcing, which makes it possible to study the effect of the random forcing in changing the underlying flow structures.

2. Time-periodic deterministic forcing:

$$\begin{aligned} \text{GB modes:} \qquad & f_{\vec{k},0}=a_{\vec{k},0}\cos(\phi t)+b_{\vec{k},0}, \\ \text{Gravity modes:} \qquad & f_{\vec{k},\pm}=a_{\vec{k},\pm}\exp(\mathrm{i}\phi t), \end{aligned} \tag{34}$$

where $a_{\vec{k},0}=\sqrt{3}/10$ and $b_{\vec{k},0}=\sqrt{3}/20$ for mode $(0,0)$; $a_{\vec{k},0}=\sqrt{3}/10$ and $b_{\vec{k},0}=\sqrt{3}/200$ for modes $(\pm 1,0)$; $a_{\vec{k},0}=-\sqrt{3}/10$ and $b_{\vec{k},0}=\sqrt{3}/200$ for modes $(0,\pm 1)$; and $a_{\vec{k},\pm}=1/10$ for all gravity modes. The frequency $\phi=20$. The amplitudes of these large-scale deterministic forcing and stochastic forcing are comparable. This setup implies the flow field has a large-scale background mean flow and a random part. The flow structure is able to switch between nearly straight streamlines and swirling eddies according to (33). Comparing the two situations helps us understand the effect of the deterministic mean flow on the filtering skill.

For the initialization of the filters, the states of all the stochastic modes are set to be consistent with the value at their statistical equilibrium associated with the forecast models, where the initial uncertainty of the stochastic modes is 0.3 and 0.1 for each GB and gravity mode, respectively.

The tracers $\mathbf{X}_s$ utilized in the full filter (17) and the two reduced filters (19) and (20) are identically the same. On the other hand, the tracers $\mathbf{X}_s^B$ in (12) for the GB filter (18) are based only on the GB part of the flow and therefore they are different from those in (10). For the sake of comparing the filtering skill, we impose the same observation noise process in (10) and (12), i.e., $W_s^x = W_s^B$. Furthermore, the initial locations of the tracers utilized in both the full filter and GB filter are the same and are distributed uniformly in the periodic domain $\mathbb{T}^2 = [-\pi,\pi]^2$.

Two dynamical regimes are considered. The first one is a fast rotation regime with small Rossby number $\varepsilon = 0.1$, which mimics the motion in the mid-latitude atmosphere or ocean (Majda 2003). Another dynamical regime involves moderate rotation with $\varepsilon = 1$. Note that the GB flow is kept to be the same in both regimes and the only difference lies in the gravity waves according to the rotation frequency and eigenmodes in Section 2a. The nondimensional number $\delta = 1$ is fixed which implies that the Rossby and the Froude number are equal with each other. Below, the filtering behavior up to a long time $t = 200$ is studied.

c. Results for filtering the random flow fields with uncoupled GB and gravity modes

In this subsection, we study the situation where the random GB and gravity modes evolve independently, i.e., $\gamma = 0$ in (2). Thus, the underlying dynamics of the velocity field for Fourier wavenumber $\vec{k}$ of nature is given by

$$\mathrm{d}\hat{v}_{\vec{k},0}(t) = \big(-d_B\hat{v}_{\vec{k},0} + f_{\vec{k},0}(t)\big)\mathrm{d}t + \sigma_{\vec{k},0}\mathrm{d}W_{\vec{k},0}(t), \tag{35a}$$

$$\mathrm{d}\hat{v}_{\vec{k},\pm}(t) = \big((-d_g + \mathrm{i}\omega_{\vec{k},\pm})\hat{v}_{\vec{k},\pm}(t) + f_{\vec{k},\pm}(t)\big)\mathrm{d}t + \sigma_{\vec{k},\pm}\mathrm{d}W_{\vec{k},\pm}(t). \tag{35b}$$

Recall that the damping and stochastic forcing in (35) compensate the nonlinearity and represent the turbulent behaviors in nature and such strategy for describing random turbulence has been widely applied in many other situations (Majda and Harlim 2012). Since the true dynamics (35) and the forecast model (13) in the full filter (17) are the same for all $\vec{k}$, the full filter becomes a perfect filter.

First, we look at the tracer behaviors. Row (a) of Figure 1 includes the comparison of the tracer trajectories utilizing the full filter (17) and the GB filter (18) at an initial period from $t = 0$ to $t = 10$ in the two dynamics regimes with different ε, where the large-scale deterministic forcing is set to be zero. For conciseness, only one of the five tracers associated with each filter is shown. The two trajectories starting at the same location almost overlap with each other during this short initial period in $\varepsilon = 0.1$ regime while the two trajectories diverge quickly in $\varepsilon = 1$ regime. Comparing the snapshot of the GB flow (column III) with the full flow (column I and II) at $t = 10$, it is clear that the gravity waves have non-negligible contributions to the total flow at each time instant. Fortunately, due to the fast oscillation nature of the gravity waves in $\varepsilon = 0.1$ regime, the effect of the gravity waves is averaged out and therefore the two trajectories align with each other. Row (b) is similar to row (a) but the time-periodic deterministic forcing in the underlying flow (35) is nonzero as described in (34) in Section 4b and the initial period shown is shortened up to $t = 7$.

The same phenomenon is found in row (b) in the two different dynamics regimes, despite that the tracers move faster due to the deterministic background flow velocity. We have also found that the deterministic forcing has no effect on the RMS error and the uncertainty in the filter estimates. In row (c) and (d), we compare the posterior variance for GB mode $(1,0)$ as a function of time up to $t = 25$. The difference by adopting different deterministic forcing is insignificant. Yet, it is obvious that the relaxation time of the posterior variance towards the statistical equilibrium state is longer in $\varepsilon = 1$ regime. Since the large-scale deterministic forcing only affects the tracer speed while it has little influence on the filtering skill, below we focus on the situation with no large-scale deterministic forcing.

Next, we study the long-term behavior of tracers' distribution. In Figure 2 we show the distributions of the tracers utilizing the full filter (17) and the GB filter (18) in the two dynamical regimes at $t = 199$. In addition to showing the distribution with $L = 5$ tracers, the results with $L = 20$ are included to provide a more clear vision. Since the GB filter deals with only the incompressible GB flow, it has been proved (Chen et al. 2014b) that the tracers are uniformly distributed at the statistical equilibrium state. With the interference of the gravity modes, the distribution of tracers at $t = 199$ remains nearly uniform in $\varepsilon = 0.1$ regime since the fast oscillation averages out the effect from the random compressible gravity waves. On the other hand, pronounced clustering of tracers is found in $\varepsilon = 1$ regime due to the compressible nature of the underlying flow. In addition, it is clear that with $L = 5$ tracers, the underlying GB flow can be filtered with high accuracy in both dynamical regimes utilizing both the full and the GB filter.

We now focus on the filtering skill utilizing different filters. As stated in (33), the structure of GB flow is controlled by the two Fourier coefficients $\hat{v}_{(0,1)}$ and $\hat{v}_{(1,0)}$. To this end, we show in Figure 3 the truth and the posterior mean estimates of these two coefficients in the two dynamical regimes. In $\varepsilon = 0.1$ regime, the filtered solutions of $\hat{v}_{(0,1)}$ and $\hat{v}_{(1,0)}$ utilizing all the four filters are quite close to the truth while in $\varepsilon = 1$ regime a significant error with many unexpected oscillations is found (row d) in the filter estimate utilizing the reduced filter with only GB forecast model (19). To provide an intuitive illustration of the error in the filter estimates, the recovered streamlines of the GB flow is demonstrated in Figure 4 at two time instants, where the true GB streamline at $t = 142.3$ is 2D array of swirling eddies and at $t = 161.4$ it becomes a shear-flow stream. Consistent with Figure 3, the filtered streamlines utilizing all the filters are nearly the same as the truth in $\varepsilon = 0.1$ regime. On the other hand, despite the skillful filter estimates utilizing both the GB (18) and full filter (17), the reduced filter with only GB forecast model (19) leads to a large disparity in the recovered the streamlines, where the swirling eddies at $t = 142.3$ are falsely recovered by shear flows (row c) and the weak shear-flow stream at $t = 161.4$ becomes strong swirling eddies in the filtered solution (row d). In addition, although some inaccuracy is also found in the filter estimate utilizing the diagonal reduced 3D-Var filter (20), the recovered streamlines are qualitatively similar to the truth.

To understand the dependence of the filters' behavior on different parameters, we show in Figure 5–7 the filtering skill as a function of the tracer numbers L, the observation noise σ_x and the energy in the gravity modes E_g, respectively. Both the RMS error in the posterior mean estimate and averaged posterior variance are computed over time interval $t \in [20, 200]$, where only the statistics of mode $(1,0)$ is shown for simplicity. The information model error in filtering the GB flows utilizing the two imperfect reduced filters (19) and (20) compared with the perfect full filter (17) through the relative entropy (27) is computed, where the information model error is splitted into the signal and dispersion parts utilizing the formula in (22). The model error averaged over time interval $t \in [20, 200]$ in filtering the GB flow field is shown in these figures.

First, we look at the RMS error in filtering the GB mode $(1,0)$. The RMS error decreases in the filter estimates utilizing the GB filter (18) with the increase of L and the decrease of σ_x and E_g. In

$\varepsilon = 0.1$ regime, all the filters have comparably high filtering skill, despite that the reduced filter with only GB forecast model (19) leads to a slightly larger RMS error with a small observation noise σ_x or a large increase in the energy associated with the gravity modes E_g. In $\varepsilon = 1$ regime, the filtering skill utilizing both the full filter (17) and the diagonal reduced 3D-Var filter (20) remains close to that utilizing the idealized GB filter (18). However, the RMS error utilizing the reduced filter with only GB forecast model (20) is much larger than that utilizing the other three filters. Note that the RMS error in the reduced filter with only GB forecast model (19) shoots up with a decrease of σ_x when σ_x is small, which is a different trend compared with the other filters. Clearly, a small σ_x means the filter trusts more towards the observations, which however implies the filter (19) falsely regards the scrambled GB and gravity observations as the observations associated the GB modes in (19).

Now we focus on the information model error (27). As shown in row (c) of Figure 5–7, the model error in the reduced filter with only GB forecast model (19) is significant in $\varepsilon = 1$ regime, where the signal part has a dominant portion. In contrast to (19), the model error in the diagonal reduced 3D-Var filter (20) shown in row (d) is much smaller, implying an insignificant lack of information in its posterior distribution related to that of the perfect filter. In addition, the model error in $\varepsilon = 0.1$ regime utilizing both the imperfect filters is smaller than that in $\varepsilon = 1$ regime. Note that different trends in large L and small σ_x are found in the RMS error and information model error utilizing the diagonal reduced 3D-Var filter (20). This is because the signal part of the information model error (22) is proportional to the inverse of the covariance of the imperfect model. With a slowly-varying gap in the mean estimates, a smaller covariance implies a more certain estimate of the incorrect state and thus a larger information model error. It is worthwhile pointing out that the information model error has no upper bound and thus it is very sensitive when the model covariance becomes extremely small. A bounded measurement for checking the model error in the posterior distribution is the Hellinger distance (Beran 1977; Branicki and Majda 2014), which is however not able to be explained as a measure of information gain. The definition of the Hellinger distance and its comparison with information model error (27) is included in Appendix D.

Finally, to provide a deeper understanding of the two imperfect filters, we include in panel (a)-(d) of Figure 8 some time series of the filtered solutions for mode $(1,0)$. Panel (a) and (b) show the absolute error in the posterior mean estimate of GB mode $(1,0)$ utilizing the reduced filter with only GB forecast model (19), where the y-axis limit is the same as that in Figure 3 of the truth. In $\varepsilon = 0.1$ regime, the error amplitude remains significantly smaller than the true signal. On the other hand, except a small error at the initial period for $t \leq 20$ in $\varepsilon = 1$ regime, the amplitude of the error is comparable with that of the true signal, which leads to a large RMS error and a significant lack of information in the signal part. In panel (c) and (d), the posterior covariance for mode $(1,0)$ utilizing both the full filter (17) and the diagonal reduced 3D-Var filter (20) is shown. Clearly, the diagonal components of the covariance matrix of the full filter, i.e., both the variance of the GB mode (blue) and that of the gravity mode (black), have much larger amplitudes than the cross-covariance between them (magenta). The negligible cross-covariance is possibly due to the orthogonality of the eigenvectors associated with GB and gravity modes. We have also checked the cross-covariance between different GB and different gravity modes and they are small as well. These are evident proofs for the skillful behavior of the reduced 3D-Var filter (20). It is also noticeable that the posterior variance of the diagonal reduced 3D-Var filter (20) becomes a constant after a short initial relaxation time, which is justified in Appendix B. Note that the variance of the GB modes utilizing both filters (blue and green) are close to each other in $\varepsilon = 0.1$ regime while the reduced 3D-Var filter results in a smaller variance than the full filter in $\varepsilon = 1$ regime, which leads to the increase of the information model error. A natural improvement for the diagonal reduced 3D-Var filter is to inflate its diagonal covariance matrix by a factor r with $r \cdot R_t$. In panel (e), we show the information model error as a function of the inflation factor r. When $r = 1.6$, which is around the ratio of the averaged variance utilizing the full filter over the variance at the statistical equilibrium utilizing the diagonal reduced filter, the information model error is reduced by 40%. The lack of information in the dispersion part is nearly zero as expected and that in the signal part is also reduced since the signal part is proportional to the inverse of the model covariance.

d. Results for filtering the random flow fields with coupled GB and gravity modes

From now on, we study the skill of filtering the multiscale random rotating compressible flow in the situation that each GB mode affects the underlying dynamics of the two corresponding gravity modes through quadratic nonlinear interactions, which is motivated directly from mathematical theory of the slow-fast geophysical flows (Embid and Majda 1998; Majda 2003; Gershgorin and Majda 2008) and high resolution of turbulent simulations in slow-fast geophysical regimes (Smith 2001; Smith and Waleffe 2002; Waite and Bartello 2004). Let's recall the governing equations of the underlying flow field for Fourier wavenumber k,

$$\mathrm{d}\hat{v}_{\vec{k},0}(t) = \big(-d_B\hat{v}_{\vec{k},0} + f_{\vec{k},0}(t)\big)\mathrm{d}t + \sigma_{\vec{k},0}\mathrm{d}W_{\vec{k},0}(t), \tag{36a}$$

$$\mathrm{d}\hat{v}_{\vec{k},\pm}(t) = \big((-d_g + \mathrm{i}\omega_{\vec{k},\pm}\underline{\pm \mathrm{i}\gamma\hat{v}_{\vec{k},0}})\hat{v}_{\vec{k},\pm}(t) + f_{\vec{k},\pm}(t)\big)\mathrm{d}t + \sigma_{\vec{k},\pm}\mathrm{d}W_{\vec{k},\pm}(t), \tag{36b}$$

where the coupling coefficient γ is non-zero. On the other hand, such nonlinear coupling between the GB and gravity modes is dropped in the forecast models of both the full filter (17) and the diagonal reduced 3D-Var filter (20) and therefore these forecast models become linear independent stochastic model (35) as discussed in Section 2c. Due to this model error, the full filter is no longer a perfect filter. Note that despite the linear independent forecast model for the random

Fourier amplitudes, the observational processes in (17), (19) and (20) remain highly nonlinear with coupled GB and gravity modes.

We first look at the intrinsic change in the coupled flow fields with the coupling effect. In Figure 9, the sample trajectories and the associated power spectrums of the gravity mode $(1,0)$ are demonstrated, and those of the GB mode are also shown as comparison. The spectrum of the gravity mode becomes more and more flat with the increase of the coupling coefficient γ in both regimes. In $\varepsilon = 0.1$ regime, the spectrums of the GB and gravity modes remain having almost no overlapped band even with $\gamma = 5$, which implies a clear scale separation between them and therefore skillful filtering results of the GB flow are expected. On the other hand, the spectrum bands of the GB and the gravity modes in $\varepsilon = 1$ regime become completely overlapped with each other for $\gamma > 1$, which indicates that the GB and gravity flows are hard to be distinguished from the mixed observations. Therefore, the filtering skill in $\varepsilon = 1$ regime is expected to be deteriorated.

We show in Figure 10 the filtered GB modes $(1,0)$ and $(0,1)$ and the reconstructed streamlines with $\gamma = 2$ in $\varepsilon = 1$ regime. Here the true GB flow is adopted to be the same as that in Section 4c and therefore the two Fourier modes in Figure 10 remain the same as those in Figure 3. The filter estimate of the GB filter (18) has very little change due to the randomness in the observation noise. However, the filter estimates utilizing all the three imperfect filters contain evident errors, where the bias utilizing reduced filter with only GB forecast model (19) is the most significant. This is reflected in the recovered streamlines at five different time instants. The reduced filter with only GB forecast model (19) leads to completely wrong flow structures while the full filter with linear forecast model (17) at least has some skill at $t = 8.5$ and $t = 105.5$ and the diagonal reduced 3D-Var filter (20) is skillful for recovering the shear flow at time $t = 8.5$ as well.

In Figure 11, we show the RMS error in the posterior mean estimation and the averaged posterior variance for mode $(1,0)$, where the filtering skill in both the GB and one of the gravity modes is included. As motivated from Figure 9, the nonlinear coupling up to $\gamma = 5$ has little effect on the filtering skill of GB mode utilizing all the filters in $\varepsilon = 0.1$ regime due to the apparent scale separation. The error in the filtered solution of the gravity mode is also almost unchanged with different γ, which is possibly due to the fact that its intrinsic oscillation in this fast oscillation regime dominates the stochastic oscillation from the interaction with the GB mode and therefore the stochastic oscillation behaves as insignificant random noise. On the other hand, the filtering skill of the GB mode utilizing all the three imperfect filters deteriorates with a gradual increase of γ in $\varepsilon = 1$ regime. Among the three imperfect filters, the largest RMS error remains in the reduced filter with only GB forecast model (19). In addition, unlike the uncoupled situation where the full filter and the diagonal reduced 3D-Var filter always have comparable filtering skill, with a nonzero γ the error utilizing the diagonal reduced 3D-Var filter (20) becomes more significant than the full filter with linear forecast model (17). Furthermore, filtering the gravity waves becomes less skillful utilizing both the full filter with linear forecast model and the diagonal reduced 3D-Var filter with the increase of γ in $\varepsilon = 1$ regime.

Finally, we study the information model error in the imperfect filters. Note that the full filter with linear forecast model (17) is no longer a perfect filter and therefore the model error in both (17) and the two reduced filters (19) and (20) are assessed following the discussion at the end of Section 3.

Panel (a) and (b) of Figure 12 show the model error in the time-averaged PDF of the posterior mean estimation utilizing the imperfect filters related to that of the true signal over time interval $t \in [20, 200]$ for GB mode $(1, 0)$. In $\varepsilon = 0.1$ regime, the model error remains small for all the filters. In $\varepsilon = 1$ regime, the model error of the three imperfect filters becomes large for $\gamma \geq 1$, where the largest model error is found in the diagonal reduced 3D-Var filter. In panel (c), we compare the time series of the posterior mean estimate and the true signal with $\gamma = 2$ in $\varepsilon = 1$ regime and the associated PDFs are shown in panel (d). Clearly, the difference in the PDF of the posterior mean estimates compared with the truth, reflecting the lack of information, is obvious utilizing all the three imperfect filters. Particularly, the large information model error in the diagonal reduced 3D-Var filter (20) is due to the fact that its PDF is more concentrated than that of the truth. This implies the posterior mean estimation of (20) misses many extreme events, such as those around $t = 140$. Note that with a nonzero coupling coefficient γ, a non-negligible cross-covariance between the GB and gravity modes appears and therefore a large model error is expected by dropping the off-diagonal entries in the posterior covariance matrix. It is worthwhile pointing out that the RMS error and the information model error provide different views in assessing the filtering skill in the posterior mean estimation. Despite a smaller RMS error compared with the reduced filter with only GB forecast model (19), a larger information model error in the diagonal reduced 3D-Var filter (20) implies the potential danger in utilizing (20) with a moderate or large γ due to its failure in capturing the important extreme events.

Figure 13 shows the information model error in the posterior distribution p^M of the GB flow utilizing the three imperfect filters (17), (19) and (20) compared to p utilizing the idealized GB filter (18). The statistics shown is averaged over time $t \in [20, 200]$. It is clear that the information model error in all the three imperfect filters remain small in $\varepsilon = 0.1$ regime while it becomes significant larger in $\varepsilon = 1$ regime and increases as a function of γ. Again, the signal part dominates the model error. As expected, the full filter with linear forecast model (17) has the smallest lack of information. Among the two reduced filters, the computational efficient diagonal reduced 3D-Var filter (20) has smaller model error than the reduced filter by completely dropping the forecast models associated with the gravity waves (19). Yet, the lack of information in the diagonal reduced 3D-Var filter increases much more significantly with γ than the full filter with linear forecast model.

5. Summary conclusions

In this paper, the filtering skill and the multiscale information model error of filtering the random rotating compressible flows utilizing noisy Lagrangian tracers are extensively studied. The random flow fields are defined through random amplitudes of Fourier eigenmodes of the rotating shallow water equations, which involve both the random incompressible GB flows and the random rotating compressible gravity waves (Section 2a). The GB and gravity modes are coupled in a highly nonlinear way in the tracer observations (Section 2b). An information-theoretic framework (Section 3) is developed to assess the lack of information and model error in imperfect filters, which applies to a single realization of the observations. Two scenarios of the underlying dynamics of the flow fields are taken into consideration.

First, linear stochastic equations with extra damping and stochastic forcing that represent the turbulent nature are utilized to model the underlying dynamics of the random amplitudes of Fourier modes, where the GB and gravity modes are assumed to be independent with each other (Equation (13)). The joint signal-observation system then becomes a conditional Gaussian system given the observations. Despite the high nonlinearity in the observations, such system allows analytical solutions for the update of the posterior states in the optimal filter (Appendix A). In addition to the full optimal filter, an idealized GB filter (18) is proposed as a reference for filtering the slow-varying GB flow which is of primary concern in practice and two practical imperfect filters with different simplifications (19) and (20) are developed. The truth of the GB flow field is designed based on a simple setup with time-varying flow structures (Section a). Shown in Section 4c, in the dynamical regime with fast rotation (Rossby number $\varepsilon = 0.1$), all the four filters have comparably high filtering skill and the lack of information in the two imperfect filters related to the perfect full filter remains small. In a moderate rotation regime ($\varepsilon = 1$), a significant information model error in the posterior distribution is found in the filtered solutions utilizing the reduced filter with only GB forecast model through the full observations (19). On the other hand, the diagonal reduced 3D-Var filter (20) is not only computationally efficient but nearly as skillful as the optimal filters in filtering the GB modes as well.

In the second part of this paper, a more realistic situation with coupled GB and gravity modes in the underlying dynamics is considered, where each GB mode affects the two gravity modes with the same Fourier wavenumber through a quadratic nonlinear interaction (36) following the mathematical theory of the slow-fast atmosphere flows (Embid and Majda 1998; Majda 2003; Gershgorin and Majda 2008). Since the full filter with nonlinear forecast model no longer belongs to the conditional Gaussian filtering framework, the same linear forecast model as in the situation with uncoupled Fourier modes is adopted in both the full and the diagonal reduced 3D-Var filters, which follows the common practical strategy for filtering high dimensional turbulent systems (Majda and Harlim 2012; Harlim and Majda 2013). Again, as shown in Section 4d, all the four filters are comparably skillful in $\varepsilon = 0.1$ regime even in the appearance of a strong nonlinear coupling in the true underlying flow. In $\varepsilon = 1$ regime, the three imperfect filters, i.e., the full filter with linear dynamics and the two reduced filters, lose their filtering skill as the increase of the nonlinear coupling. The filtering skill of the full filter with linear forecast model remains acceptable. On the other hand, information theory shows that the diagonal reduced 3D-Var filter fails to recover the extreme events while the reduced filter with only GB forecast model suffers from a significant lack of information in the posterior distribution compared to that of the idealized GB filter.

It is worthwhile pointing out that the conditional Gaussian filtering framework adopted here has many other desirable applications. Examples of this framework includes filtering the stochastic skeleton model of the Madden-Julian oscillation (MJO) (Chen and Majda 2015b), initialization of the unobserved variables in predicting the MJO/Monsoon indices (Chen et al. 2014a; Chen and

Majda 2015d,c), exploring the model error in dyad and triad models and analyzing the parameter estimations skill for a wide class of models (Chen and Majda 2015a).

In addition, utilizing information-theoretic framework for the assessment of filter performance is an important topic in filtering turbulent systems. A systematic description of quantifying the statistical accuracy of Kalman filters with model error and the optimality of the imperfect Kalman filters in terms of different information measures is presented in (Branicki and Majda 2014).

Acknowledgments. This research of A.J.M is partially supported by the Office of Naval Research grant ONR MURI N00014-12-1-0912. N.C. is supported as a graduate research assistant on this grant. A.J.M also gratefully acknowledges the generous support of the Center for Prototype Climate Modeling of NYU Abu Dhabi Research Institute grant.

APPENDIX A

Filtering formulae for the conditional Gaussian systems

Here, we show the formulae for updating the posterior mean and posterior covariance of the full linearized filter (17). It is straightforward to derive formulae for those of the idealized GB filter (18) and the reduced filter with only GB forecast model (19).

Recall the full filter with linearized dynamics

$$\begin{aligned} d\mathbf{X}_s &= \mathbf{P}_X(\mathbf{X}_s)\mathbf{U}_s ds + \sigma_x dW_s^x, \\ d\mathbf{U}_s &= -\Gamma \mathbf{U}_s ds + F_s ds + \Sigma_u dW_u(s). \end{aligned} \tag{A1}$$

where $\mathbf{X}_s$ is the observed tracer trajectories and $\mathbf{U}_s$ is the linearized forecast model for the flow field. Despite the conditional Gaussianity, the full system (A1) remains highly nonlinear due to the nonlinear observation process. Following theorem 12.7 in (Liptser and Shiryaev 2001), given bounded $\Gamma_t, F_t, \mathbf{P}_X$ processes being functions of $\mathbf{X}_t$, if $\mathbb{P}(\mathbf{U}_0 \in \cdot | \mathbf{X}_0)$ is $\mathcal{N}(m_0, R_0)$, then conditioned on $\mathbf{X}_{s\le t}$, $\mathbb{P}(\mathbf{U}_t \in \cdot|\mathbf{X}_{s\le t})$ is Gaussian $\mathcal{N}(m_t, R_t)$, with m_t, R_t being solutions to the following with initial value m_0, R_0:

$$\mathrm{d}m_t = [-\Gamma m_t + F_t]\mathrm{d}t + \sigma_x^{-1} R_t \mathbf{P}_X^*(\mathbf{X}_t)[\mathrm{d}\mathbf{X}_t - \mathbf{P}_X(\mathbf{X}_t)m_t\mathrm{d}t], \tag{A2}$$

$$\mathrm{d}R_t = [-\Gamma R_t - R_t\Gamma^* + \Sigma_u\Sigma_u^* - \sigma_x^{-2} R_t \mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t)R_t]\mathrm{d}t. \tag{A3}$$

APPENDIX B

Filtering formulae for the diagonal reduced 3D-Var filter

In the diagonal reduced 3D-Var filter, the posterior covariance is set to be diagonal. The formulae of updating the posterior mean are the same as that in (A2). To see the update of the posterior covariance, we denote $R_{t,i}$ to be the (i,i)-th entry of R_t. Then the update of the posterior covariance R_t in (A3) becomes $|\mathbf{K}|$ independent 1-D equations

$$\mathrm{d}R_{t,i} = [-\Gamma_{ii}R_{t,i} - R_{t,i}\Gamma_{ii}^* + (\Sigma_u\Sigma_u^*)_{ii} - \sigma_x^{-2} R_{t,i}^2 (\mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t))_{ii}]\mathrm{d}t, \tag{B1}$$

where $(\cdot)_{ii}$ means the (i,i)-th entry of the matrix. In each time step, after solving each $R_{t,i}$, we insert R_t into the posterior mean update (A2). It is worthwhile noticing that the (i,i)-th component of $\mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t)$ is simply $|\vec{r}_\mathbf{k}|^2$, as is seen in (11) due to the fact that $\exp(-\mathrm{i}\vec{k}\cdot\vec{X}_\mathbf{k}(s))\cdot\exp(\mathrm{i}\vec{k}\cdot\vec{X}_\mathbf{k}(s)) = 1$ for $\mathbf{k}\in\mathbf{K}$. Thus, the equation (B1) is deterministic. In addition, the diagonal entry $R_{t,i}$ converges to a constant equilibrium value after a short relaxation time.

On the other hand, the $\mathbf{P}_X^*(\mathbf{X}_t)\mathbf{P}_X(\mathbf{X}_t)$ matrix in the full filter (A1) is not a constant matrix because the tracer locations play important roles in the off-diagonal components. Due to the non-linearity in R_t, the diagonal entries affected by the off-diagonal one also becomes time-dependent in each update.

APPENDIX C

Approximation error by utilizing the GB filter as the reference in assessing the information model error

Here, we compare the information model error (27) in the imperfect filters (19) and (20) in the situation with $\gamma = 0$ by choosing different reference perfect filters. Note that when $\gamma = 0$, both the full filter (17) and the GB filter (18) are perfect filters. The posterior distribution associated with either (17) or (18) is chosen as p in while that associated with the two reduced filters (19) and (20) is chosen as p^M. The goal is to see the approximation error in (27) by choosing the GB filter (18) as the reference filter in assessing the information model error. We show the results as a function of the energy in the gravity modes in Figure 14.

It is clear that the information model error in both the imperfect filters by utilizing the posterior distribution associated with GB filter (18) as the reference distribution p in (27) is slightly larger than utilizing that associated with the full filter (17) due to the extra lack of information in the observations. Fortunately, the qualitative conclusions by utilizing different reference distribution p remain the same. The lack of information by utilizing the reduced filter with only GB forecast model (19) is significantly larger than that utilizing the diagonal reduced 3D-Var filter (20) in $\varepsilon = 1$ regime while the lack of information in both imperfect filters remains small in $\varepsilon = 0.1$ regime. In addition, the large model error utilizing the reduced filter with only GB forecast model (19) in $\varepsilon = 1$ regime dominates the approximation error due to the idealized artificial observations in the GB filter (18). These results imply the justification of utilizing the GB filter as the reference in assessing the model error in the more complicated situation with $\gamma \neq 0$.

APPENDIX D

Hellinger distance

As discussed in Section d, the relative entropy is unbounded from above and it is very sensitive when the model uncertainty R_q in (22) becomes small. To provide a bounded measurement, we introduce the Hellinger distance (Beran 1977; Branicki and Majda 2014),

$$d_H(p,q) = \frac{1}{2}\int (\sqrt{p}-\sqrt{q})^2 = 1 - \int \sqrt{pq}, \tag{D1}$$

where p and q are the distribution associated with the perfect and imperfect model, respectively. In Gaussian framework $p \sim \mathcal{N}(\vec{m}_p, R_p)$ and $q \sim \mathcal{N}(\vec{m}_q, R_q)$, the Hellinger distance becomes

$$d_H(p,q) = 1 - \frac{|R_p|^{\frac{1}{4}}|R_q|^{\frac{1}{4}}}{\left|\frac{1}{2}R_p + \frac{1}{2}R_q\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{4}(\vec{m}_p - \vec{m}_q)^T \left(R_p + R_q\right)^{-1}(\vec{m}_p - \vec{m}_q)\right). \tag{D2}$$

It is clear that the Hellinger distance is bounded $0 \le d_H(p,q) \le 1$. Yet, the drawback of Hellinger distance is that it cannot be explained as a measure of information gain.

In Figure 15, we show the Hellinger distance between the posterior distribution utilizing the two reduced filters and that of perfect full filter (17) as a function of L, σ_x and E_g in the situation that $\gamma = 0$ in the true underlying flow field, which can be compared with the information model error in Figure 5, 6 and 7. Same trends of the filtering dependence are found by utilizing Hellinger distance and the information model error but the Hellinger distance is clearly bounded from above.

References

Apte, A., C. K. Jones, and A. Stuart, 2008: A Bayesian approach to Lagrangian data assimilation. *Tellus A*, **60 (2)**, 336–347.

Beran, R., 1977: Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, 445–463.

Branicki, M., N. Chen, and A. J. Majda, 2013: Non-Gaussian test models for prediction and state estimation with model errors. *Chinese Annals of Mathematics, Series B*, **34 (1)**, 29–64.

Branicki, M., and A. Majda, 2014: Quantifying Bayesian filter performance for turbulent dynamical systems through information theory. *Comm. Math. Sci*, **12 (5)**, 901–978.

Branicki, M., and A. J. Majda, 2015: An information-theoretic framework for improving imperfect dynamical predictions via Multi-Model Ensemble forecasts. *Journal of Nonlinear Science*, **25 (3)**, 489–538.

Chen, N., and A. J. Majda, 2015a: Filtering nonlinear turbulent dynamical system through conditional Gaussian statistics. *Monthly Weather Review*, submitted.

Chen, N., and A. J. Majda, 2015b: Filtering the stochastic skeleton model for the Madden-Julian oscillation. *Monthly Weather Review*, accepted.

Chen, N., and A. J. Majda, 2015c: Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Mathematics of Climate and Weather Forecasting*, **1 (1)**, 1–20.

Chen, N., and A. J. Majda, 2015d: Predicting the real-time multivariate Madden-Julian oscillation index through a low-order nonlinear stochastic model. *Monthly Weather Review*, **143 (6)**, 2148–2169.

Chen, N., A. J. Majda, and D. Giannakis, 2014a: Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model. *Geophysical Research Letters*, **41 (15)**, 5612–5619.

Chen, N., A. J. Majda, and X. T. Tong, 2014b: Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. *Nonlinearity*, **27 (9)**, 2133–2163.

Chen, N., A. J. Majda, and X. T. Tong, 2015: Noisy Lagrangian tracers for filtering random rotating compressible flows. *Journal of Nonlinear Science*, **25 (3)**, 451–488.

Cushman-Roisin, B., and J.-M. Beckers, 2011: *Introduction to geophysical fluid dynamics: physical and numerical aspects*, Vol. 101. Academic Press.

Embid, P. F., and A. J. Majda, 1998: Low Froude number limiting dynamics for stably stratified flow with small or finite Rossby numbers. *Geophysical & Astrophysical Fluid Dynamics*, **87 (1-2)**, 1–50.

Gershgorin, B., and A. Majda, 2008: A nonlinear test model for filtering slow-fast systems. *Communications in Mathematical Sciences*, **6 (3)**, 611–649.

Gill, A. E., 1982: *Atmosphere-ocean dynamics*, Vol. 30. Academic press.

Gould, J., and Coauthors, 2004: Argo profiling floats bring new era of in situ ocean observations. *EOS, Transactions American Geophysical Union*, **85 (19)**, 185–191.

Griffa, A., A. Kirwan Jr, A. J. Mariano, T. Özgökmen, and H. T. Rossby, 2007: *Lagrangian analysis and prediction of coastal and ocean dynamics*. Cambridge University Press.

Harlim, J., and A. J. Majda, 2013: Test models for filtering and prediction of moisture-coupled tropical waves. *Quarterly Journal of the Royal Meteorological Society*, **139 (670)**, 119–136.

Haykin, S., 2004: *Kalman filtering and neural networks*, Vol. 47. John Wiley & Sons.

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *Journal of the atmospheric sciences*, **59 (13)**, 2057–2072.

Kuznetsov, L., K. Ide, and C. Jones, 2003: A method for assimilation of Lagrangian data. *Monthly Weather Review*, **131 (10)**, 2247–2260.

Liptser, R. S., and A. N. Shiryaev, 2001: *Statistics of Random Processes II: II. Applications*, Vol. 2. Springer.

Majda, A., 2003: *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Vol. 9. American Mathematical Soc.

Majda, A., R. Kleeman, and D. Cai, 2002: A mathematical framework for quantifying predictability through relative entropy. *Methods and Applications of Analysis*, **9 (3)**, 425–444.

Majda, A., and X. Wang, 2006: *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press.

Majda, A. J., 2012: Challenges in climate science and contemporary applied mathematics. *Communications on Pure and Applied Mathematics*, **65 (7)**, 920–948.

Majda, A. J., and M. Branicki, 2012: Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete Cont. Dyn. Systems*, **32 (9)**.

Majda, A. J., and P. Embid, 1998: Averaging over fast gravity waves for geophysical flows with unbalanced initial data. *Theoretical and computational fluid dynamics*, **11 (3-4)**, 155–169.

Majda, A. J., and B. Gershgorin, 2010: Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences*, **107 (34)**, 14 958–14 963.

Majda, A. J., and B. Gershgorin, 2011a: Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proceedings of the National Academy of Sciences*, **108 (25)**, 10 044–10 049.

Majda, A. J., and B. Gershgorin, 2011b: Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proceedings of the National Academy of Sciences*, **108 (31)**, 12 599–12 604.

Majda, A. J., and J. Harlim, 2012: *Filtering Complex Turbulent Systems*. Cambridge University Press.

Majda, A. J., I. Timofeyev, and E. V. Eijnden, 1999: Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences*, **96 (26)**, 14 687–14 691.

Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 2003: Systematic strategies for stochastic mode reduction in climate. *Journal of the Atmospheric Sciences*, **60 (14)**, 1705–1722.

Molcard, A., L. I. Piterbarg, A. Griffa, T. M. Özgökmen, and A. J. Mariano, 2003: Assimilation of drifter observations for the reconstruction of the Eulerian circulation field. *Journal of Geophysical Research: Oceans (1978–2012)*, **108 (C3)**.

Navon, I. M., 2009: Data assimilation for numerical weather prediction: a review. *Data assimilation for atmospheric, oceanic and hydrologic applications*, Springer, 21–65.

Rossby, C., 1937: On the mutual adjustment of pressure and velocity distributions in certain simple current systems. *J. Mar. Res*, **1 (19)**, 280.

Salman, H., K. Ide, and C. K. Jones, 2008: Using flow geometry for drifter deployment in Lagrangian data assimilation. *Tellus A*, **60 (2)**, 321–335.

Salman, H., L. Kuznetsov, C. Jones, and K. Ide, 2006: A method for assimilating Lagrangian data into a shallow-water-equation ocean model. *Monthly Weather Review*, **134 (4)**, 1081–1101.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle–ensemble Kalman filter for Lagrangian data assimilation. *Monthly Weather Review*, **143 (1)**, 195–211.

Smith, L. M., 2001: Numerical study of two-dimensional stratified turbulence. *Contemporary Mathematics*, **283**, 91–106.

Smith, L. M., and F. Waleffe, 2002: Generation of slow large scales in forced rotating stratified turbulence. *Journal of Fluid Mechanics*, **451**, 145–168.

Vallis, G. K., 2006: *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.

Waite, M. L., and P. Bartello, 2004: Stratified turbulence dominated by vortical motion. *Journal of Fluid Mechanics*, **517**, 281–308.

LIST OF FIGURES

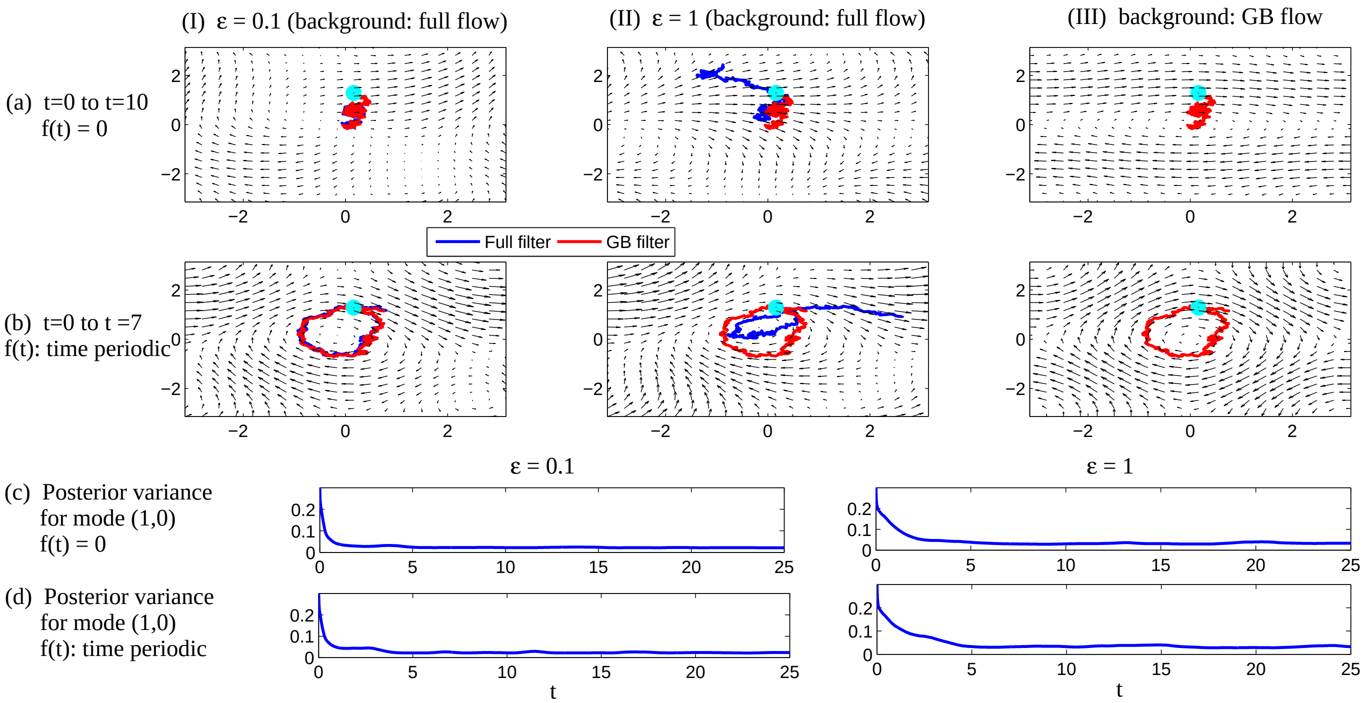

FIG. 1. Row (a) and (b): Tracer trajectories utilizing the full filter (17) (blue) and the GB filter (18) (red) at an initial period in two situations with different large-scale deterministic forcing as described in Section 4b. The total number of tracers in filtering is $L = 5$ but only one tracer trajectory of each filter is shown for conciseness. The cyan dot indicates the initial location of the tracers. The flow field is shown at time $t = 10$ and $t = 7$ for the two situations, respectively, where the full flow is shown in column (I) and (II) while only the GB flow is shown in column (III). Panel (c) and (d) show the posterior variance associated with GB mode $(1,0)$ with different large-scale deterministic forcing.

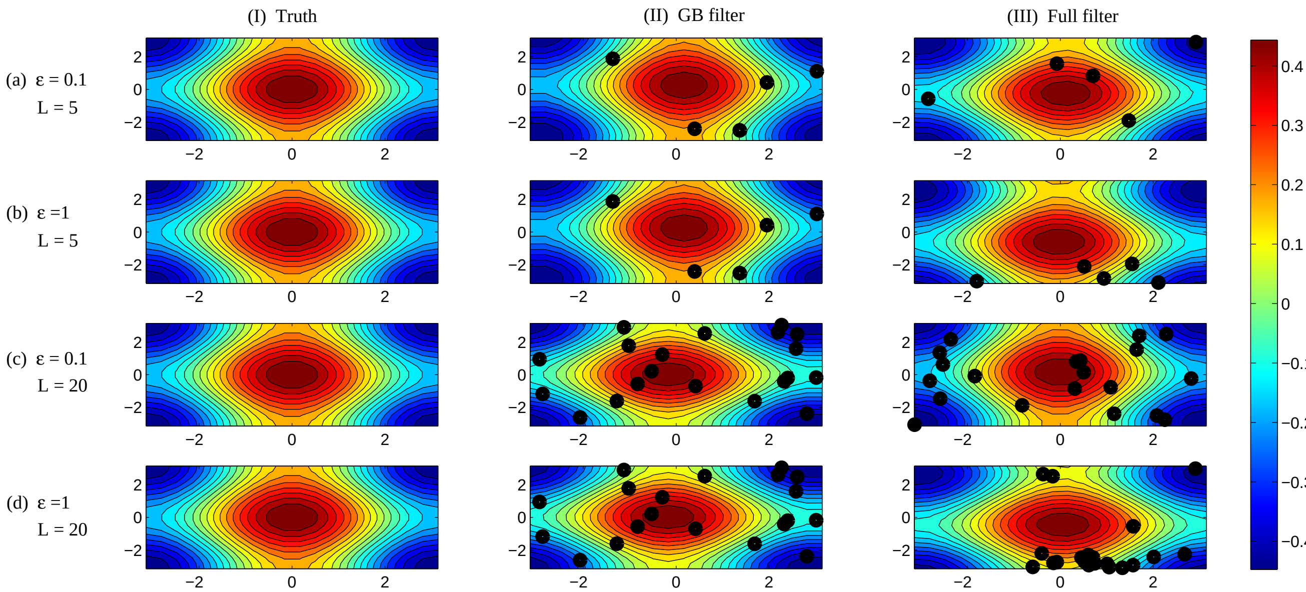

FIG. 2. Distribution of the tracer locations at time $t = 199$. Note that the distribution of GB filter (column II) is always uniform. The streamline in each panel corresponds to the GB flow.

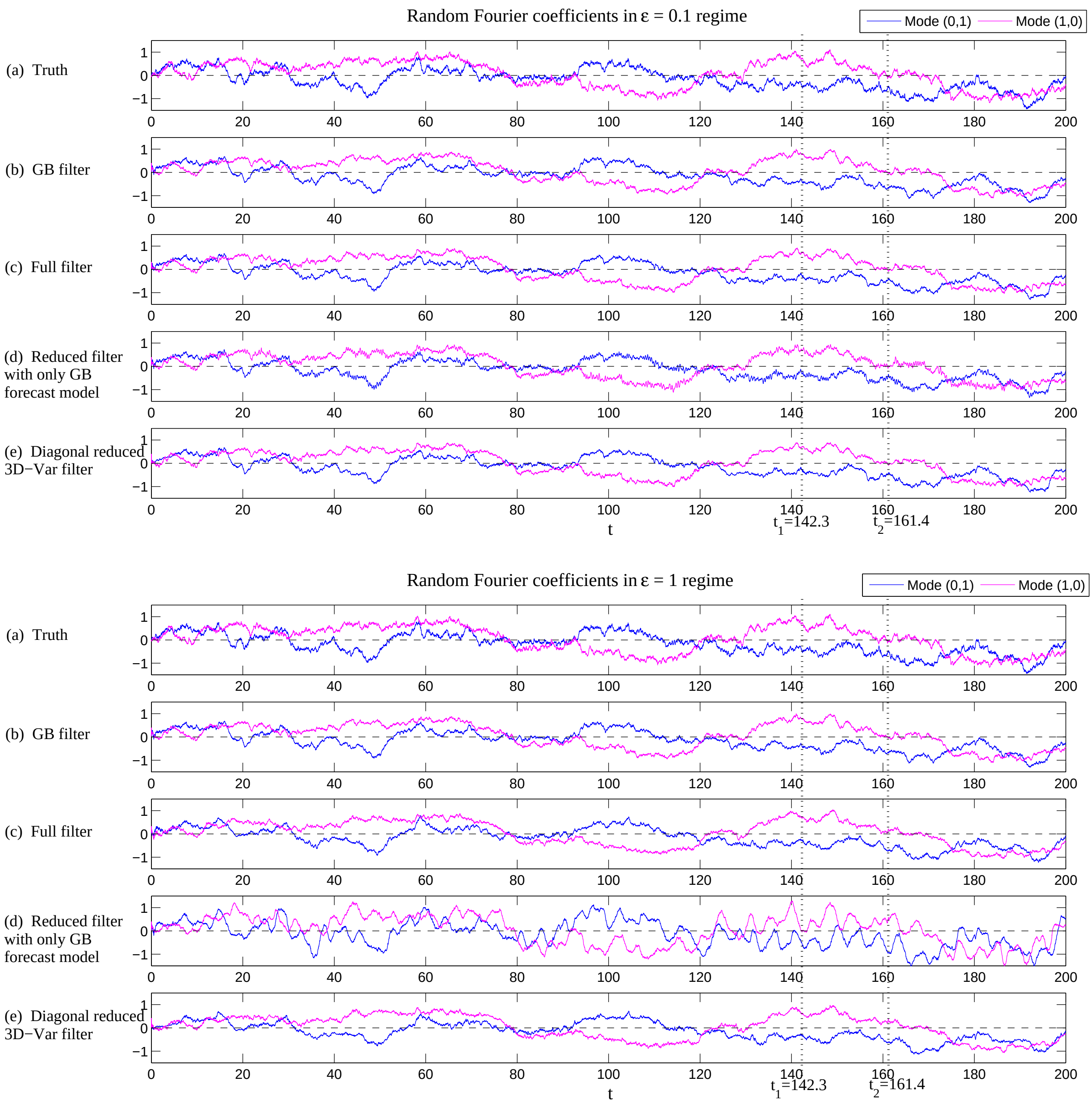

FIG. 3. Truth and the filtered solutions (posterior mean estimates) of $\hat{v}_{(0,1)}$ and $\hat{v}_{(1,0)}$ of the GB flow in (33). These two Fourier coefficients determine the flow structure. Top: $\varepsilon = 0.1$ regime. Bottom: $\varepsilon = 1$ regime.

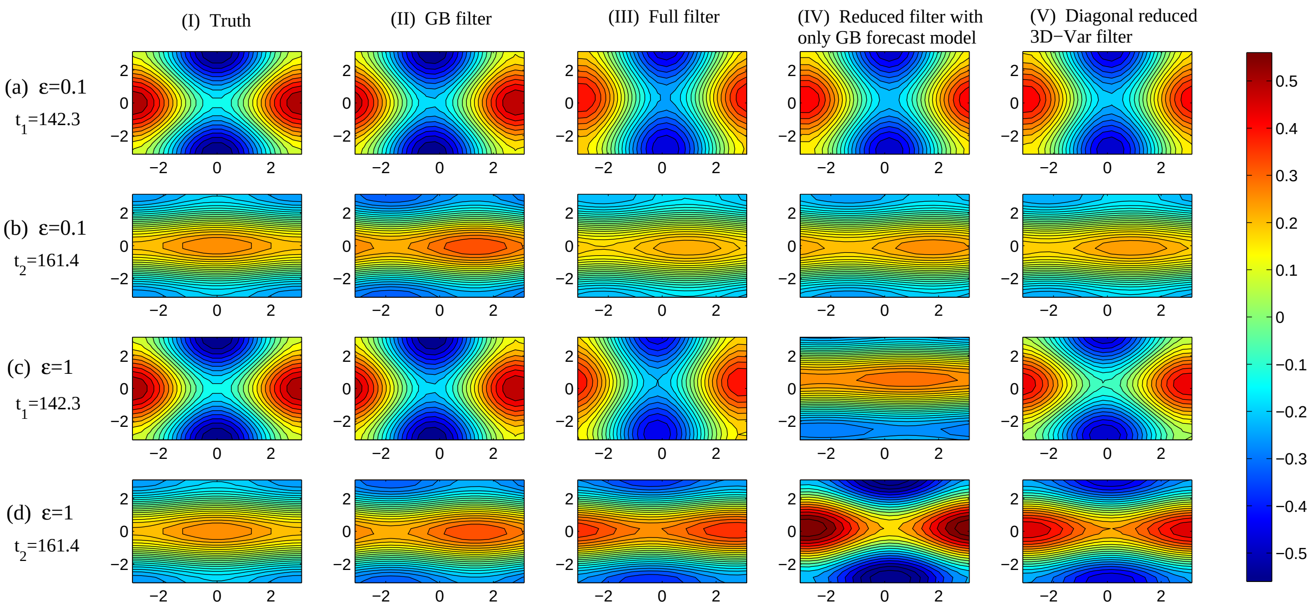

FIG. 4. Truth and the filtered streamlines of the GB flow in (33) utilizing different filters at time instants $t = 142.3$ and $t = 161.4$ as shown in Figure 3, where the underlying GB flow forms 2D array of swirling eddies at $t = 142.3$ and it becomes a shear-flow stream at $t = 161.4$.

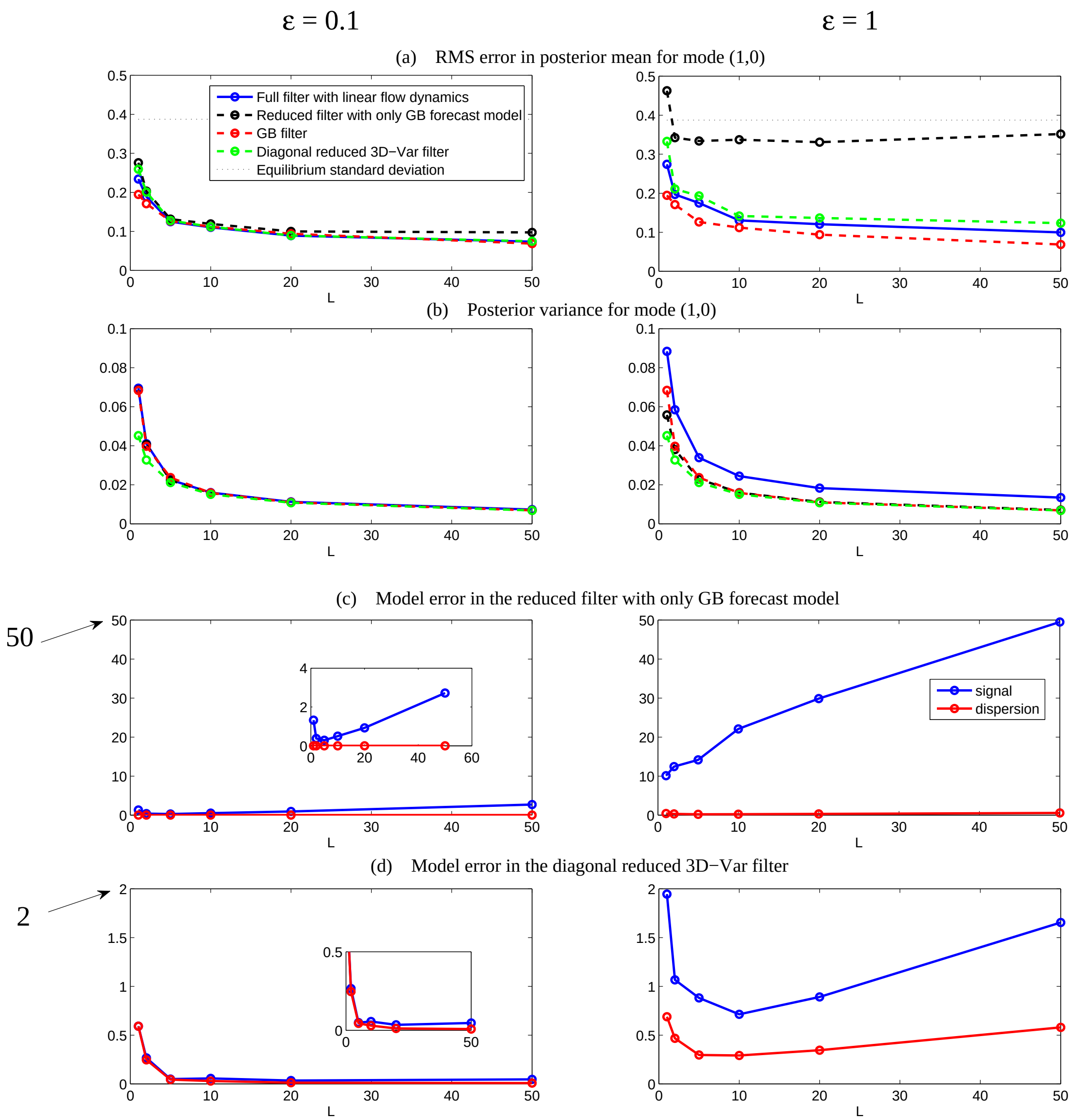

FIG. 5. Panel (a) and (b): The RMS error in the posterior mean estimate and the averaged posterior covariance over time $t \in [20,200]$ for GB mode $(1,0)$ utilizing different filters. Panel (c) and (d): the information model error in filtering the GB flow utilizing the two reduced filters as a function of the number of tracers L. Here the information model error is computed utilizing formula in (27) and then averaged in time for $t \in [20,200]$. In (27), p is the posterior distribution of the perfect full filter (17) and p^M is that of the reduced filter with only GB forecast model (19) (row c) and the diagonal reduced filter (20) (row d), respectively.

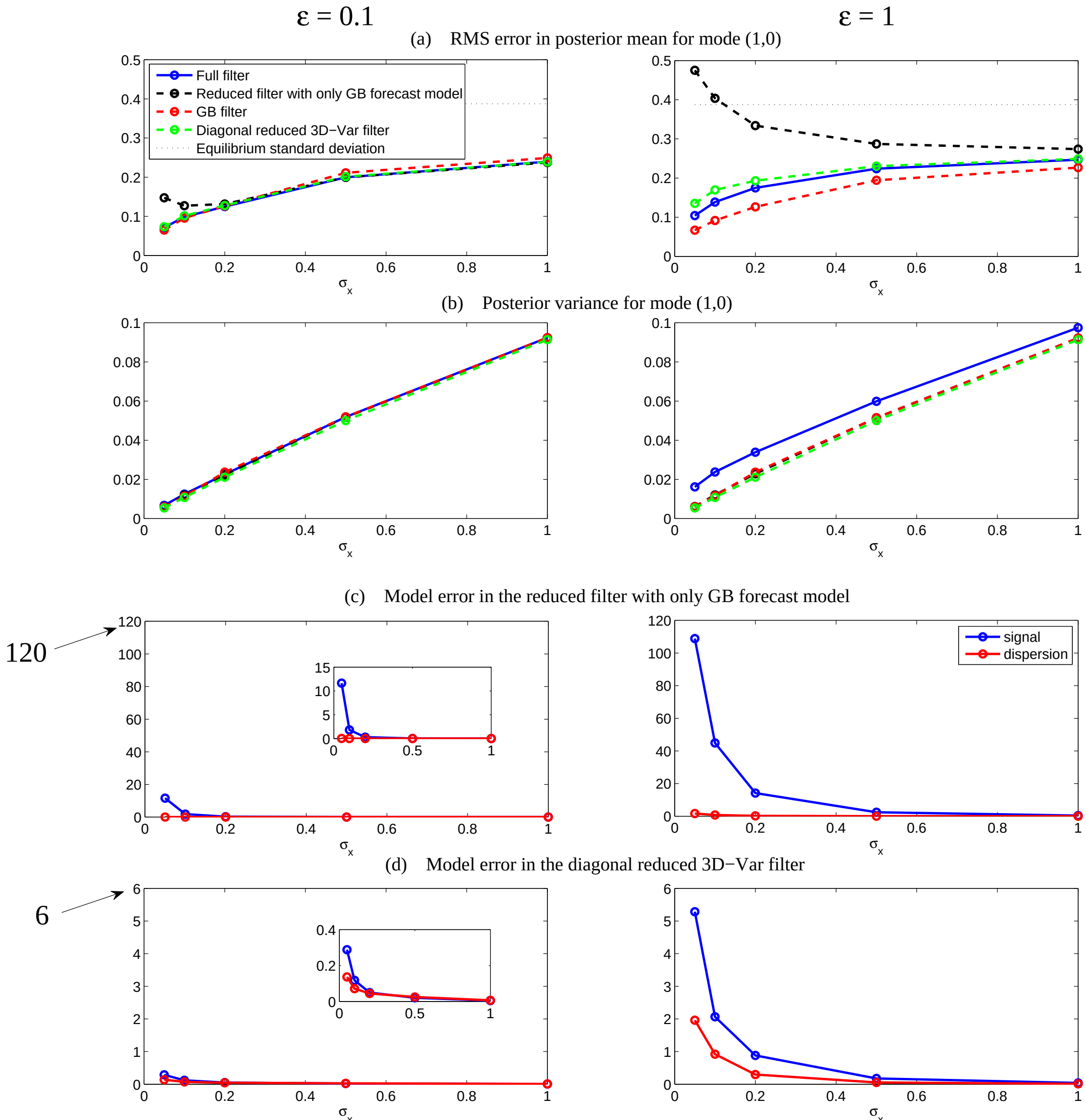

FIG. 6. Same as Figure 5 but as a function of the observational noise σ_x.

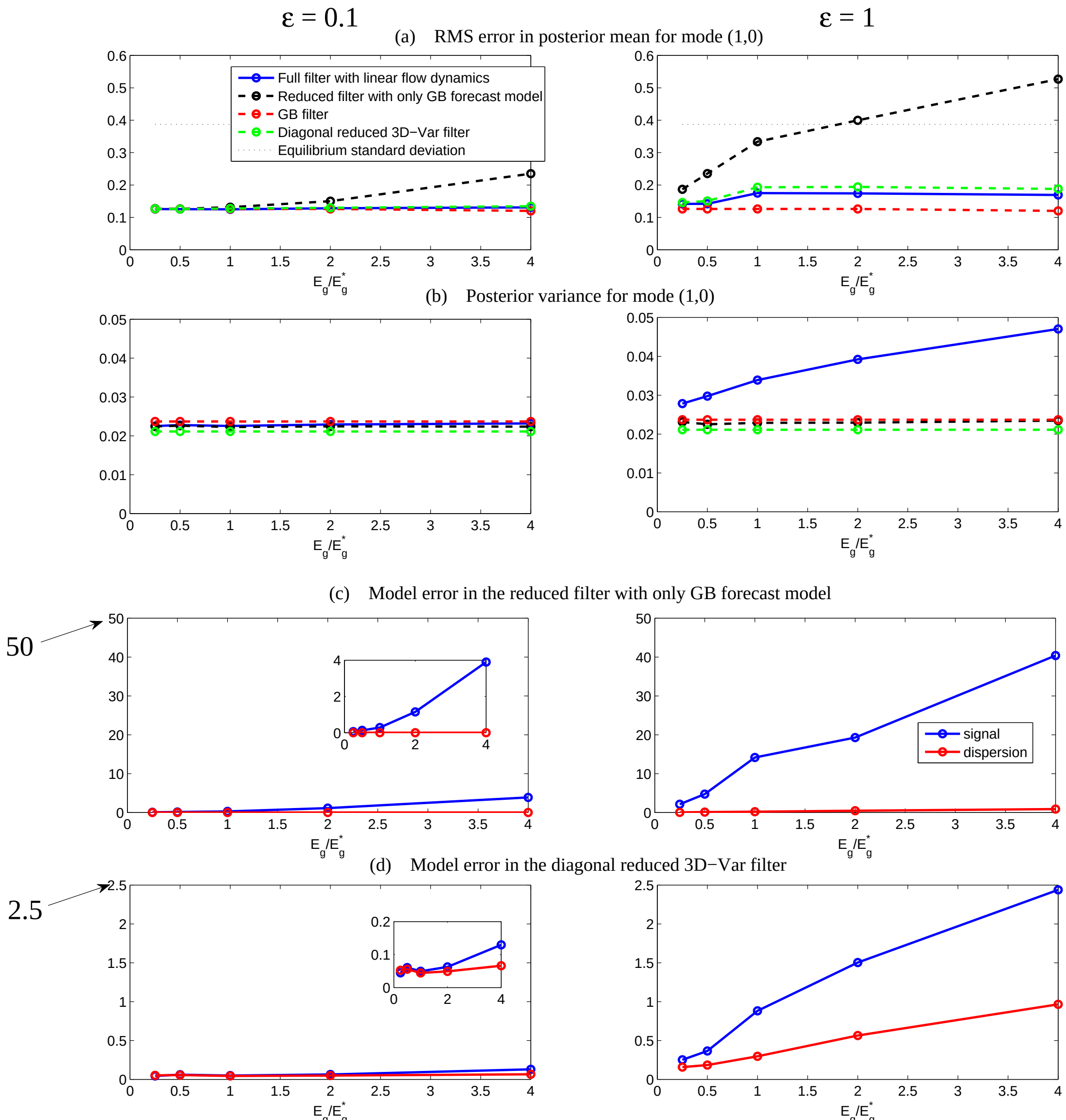

FIG. 7. Same as Figure 5 but as a function of the energy $E_g$in gravity modes, where $E_g^* = 0.1$ is the standard value in each gravity mode.

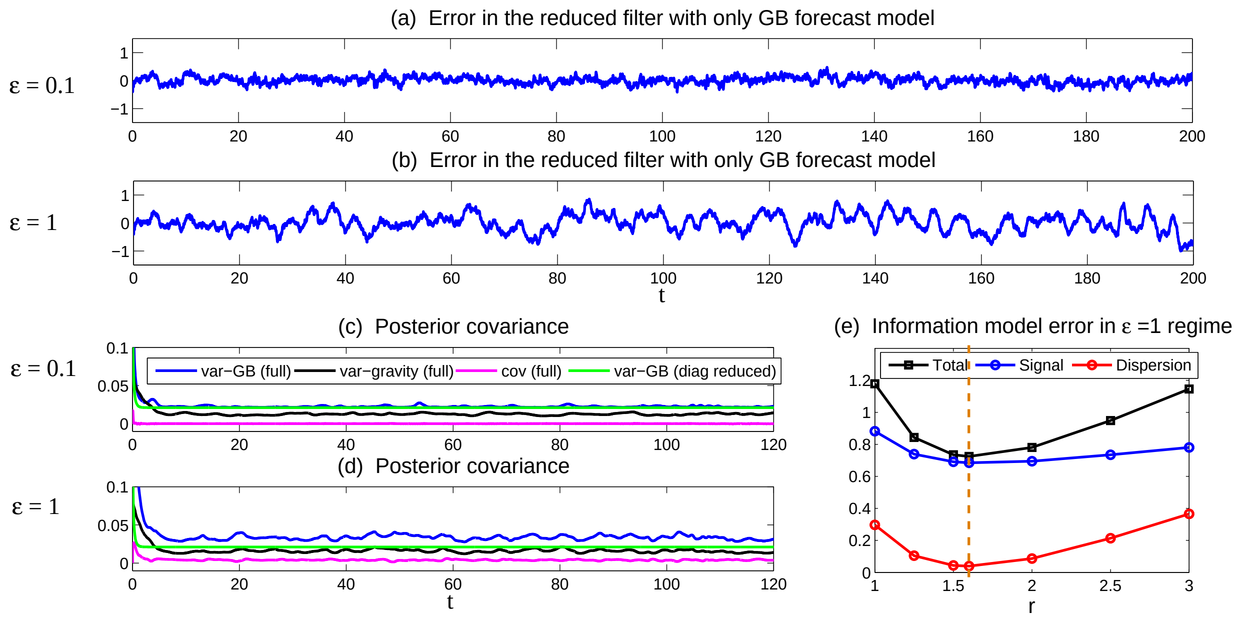

FIG. 8. Panel (a) and (b): The absolute error utilizing the reduced filter with only GB forecast model (19) in GB mode $(1,0)$ as a function of time, where the y-axis limit is the same as that in Figure 3 of the truth. Panel (c) and (d): The posterior variance of GB mode $(1,0)$ (blue), gravity mode $(1,0)$ (black) and the cross-covariance between GB and gravity modes $(1,0)$ (magenta) utilizing the full filter (17), and the posterior variance of mode $(1,0)$ (green) utilizing the diagonal reduced 3D-Var filter (20). Panel (e): the information model error as a function of the inflation factor r of the covariance in $\varepsilon = 1$ regime.

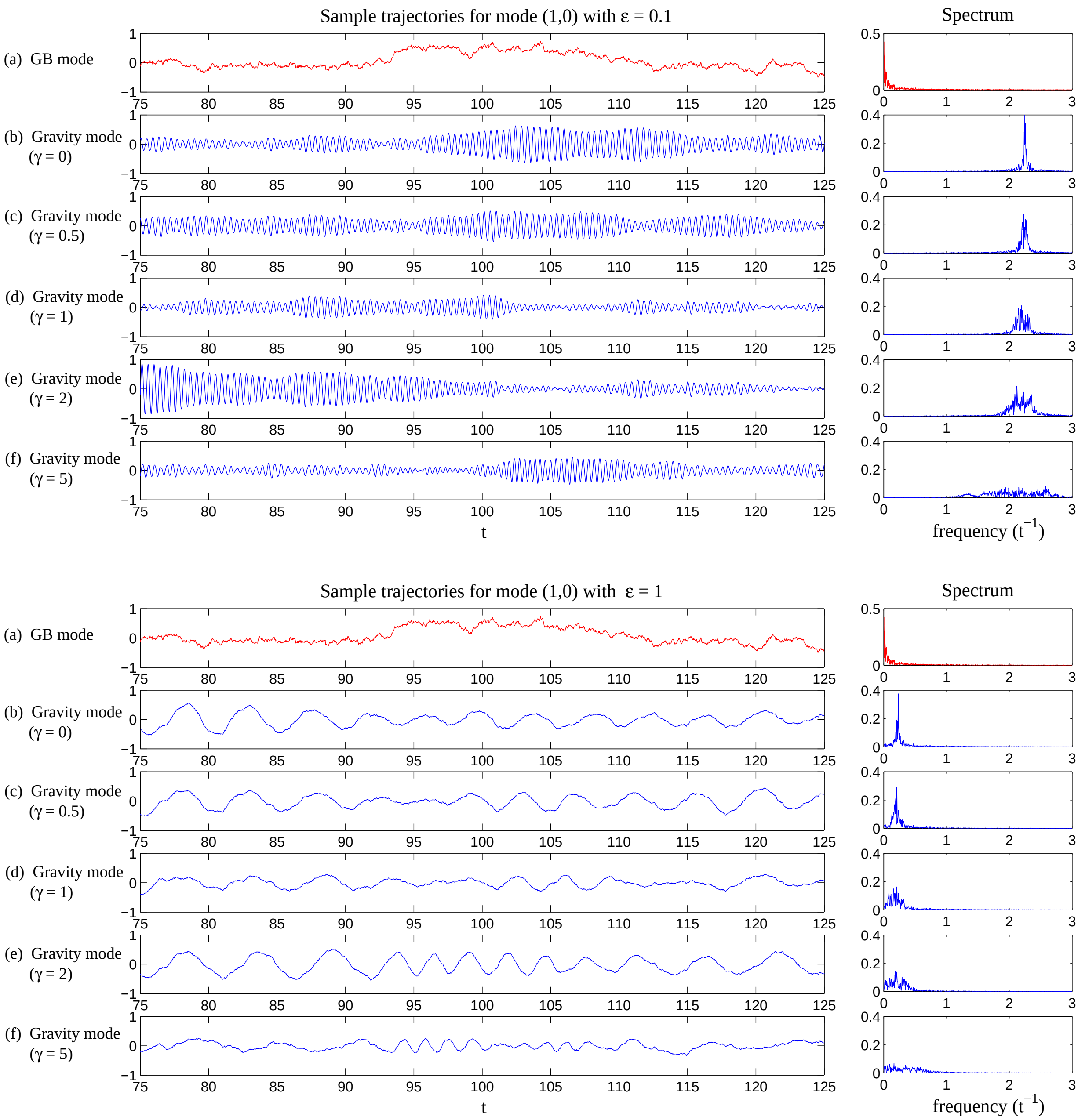

FIG. 9. Trajectory and power spectrum of mode $(1,0)$ for GB flow (panel (a)) and gravity wave with different values of coupling coefficient γ in the true underlying dynamics (36). Top: $\varepsilon = 0.1$; Bottom: $\varepsilon = 1$.

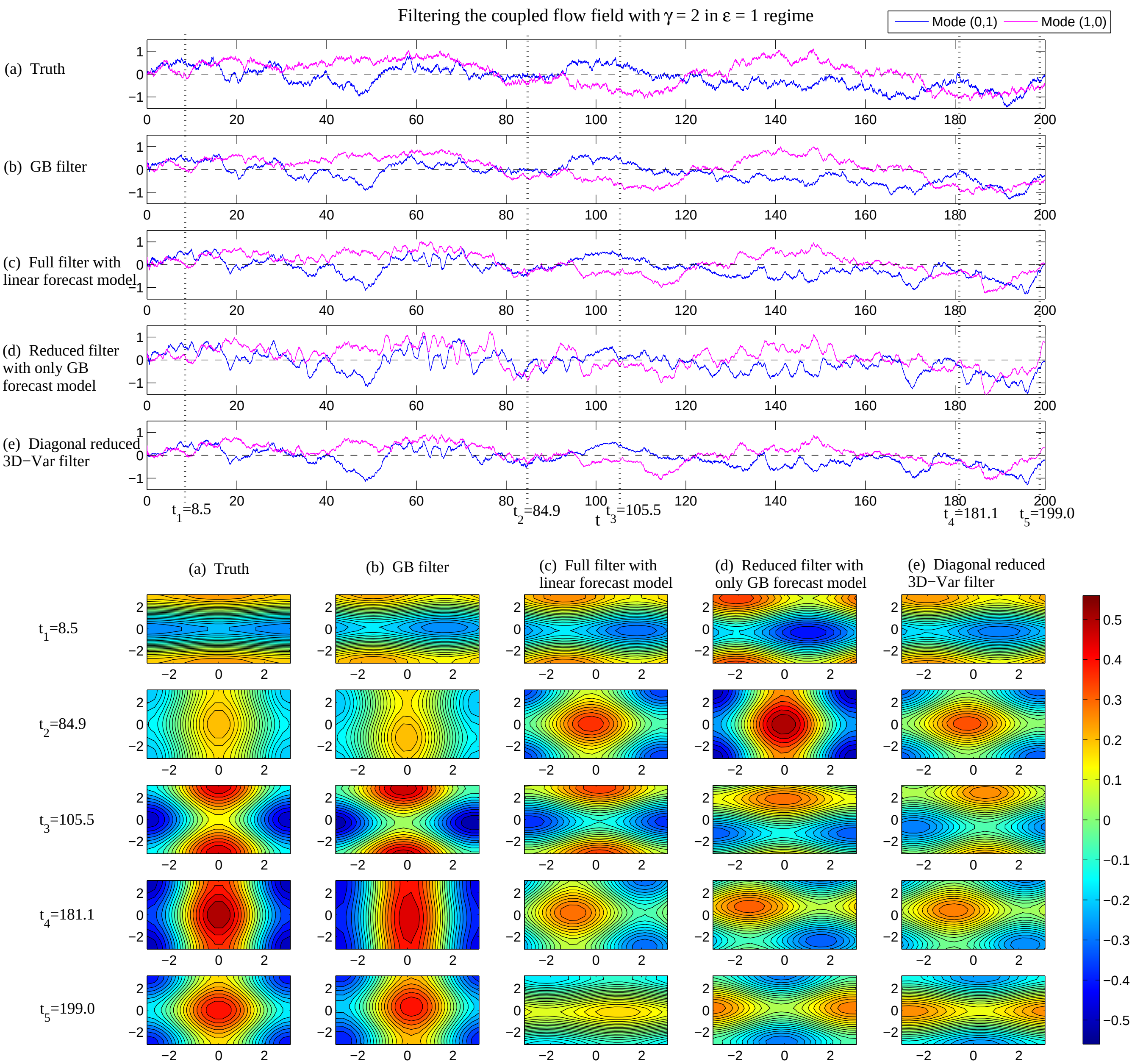

FIG. 10. Top: Truth and the filtered solutions of GB mode $(1,0)$ and $(0,1)$ utilizing different filters. Bottom: The true and recovered streamlines at five time instant in $\varepsilon = 1$ regime. Here, no large-scale forcing is imposed in the underlying flow and the coupling coefficient $\gamma = 2$.

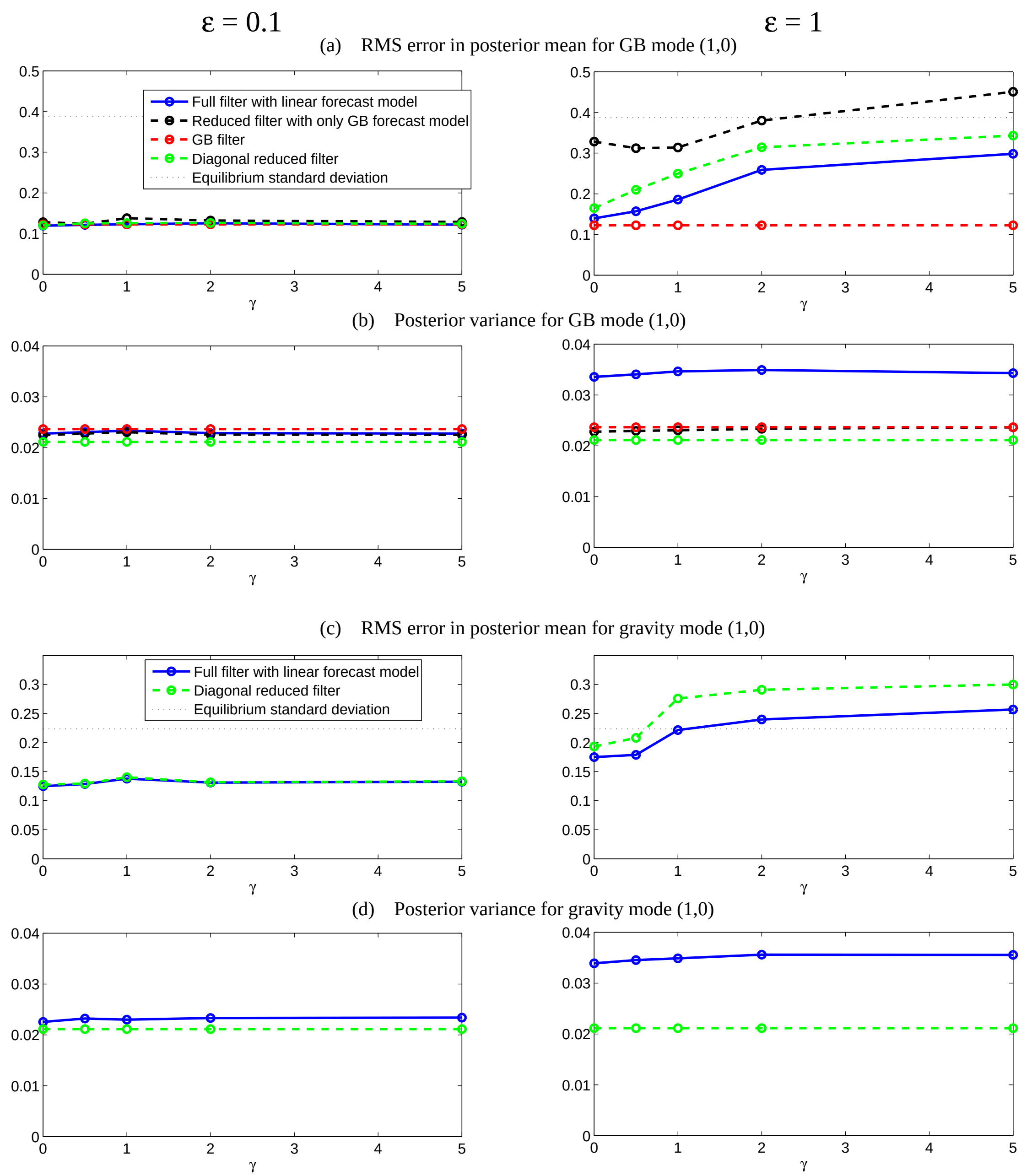

FIG. 11. The RMS error in posterior mean estimate and the averaged posterior variance over time $t \in [20,200]$ for Fourier mode $(1,0)$ as a function nonlinear coupling coefficient γ. Panel (a) and (b): GB mode; Panel (c) and (d): Gravity mode.

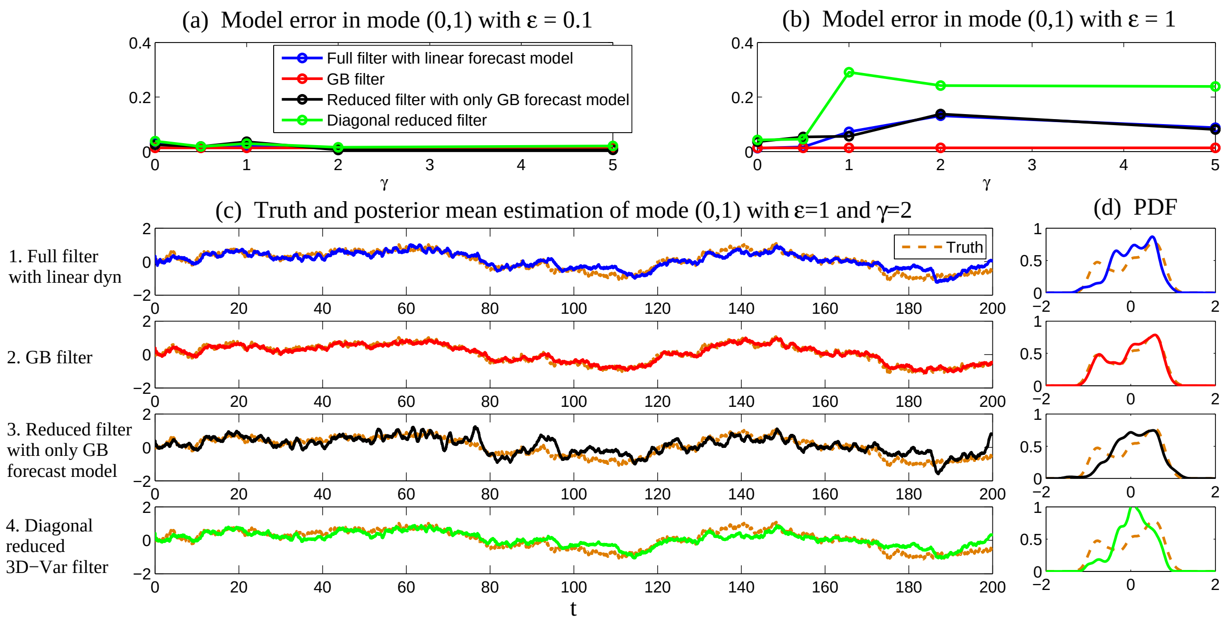

FIG. 12. Panel (a) and (b): Information model error in the time-averaged PDF of the posterior mean estimate for GB mode $(0,1)$ related to that of the truth as a function of coupling coefficient γ. Panel (c) and (d): Time series and the associated time-averaged PDFs of the filtered solution of GB mode $(0,1)$ with $\gamma = 2$ and $\varepsilon = 1$ compared with those of the truth.

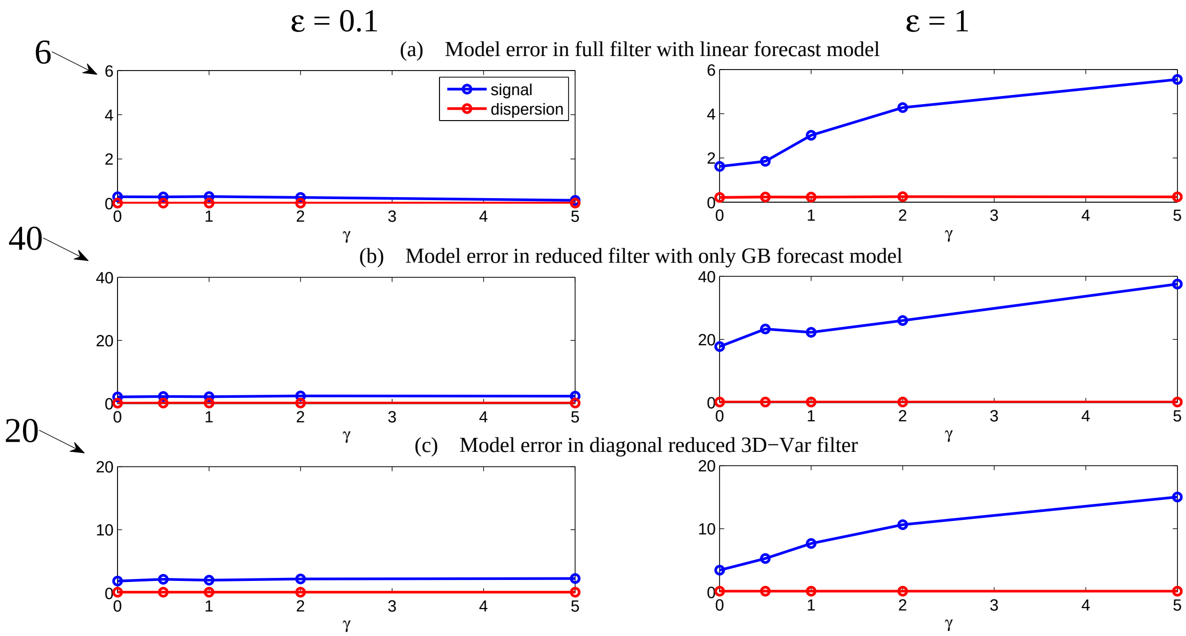

FIG. 13. Information model error (27) in the posterior distribution utilizing the three imperfect filters related to that utilizing the perfect GB filter in filtering the GB part of the flow, where p^M is the posterior distribution of (a) the full filter with linear dynamics (17), (b) the reduced filter with only GB forecast model (19), and (c) the diagonal reduced 3D-Var filter (20), and p is that of the GB filter. The statistics shown are the averaged value over time interval $t \in [20, 200]$ for the information model error computed at each time instant.

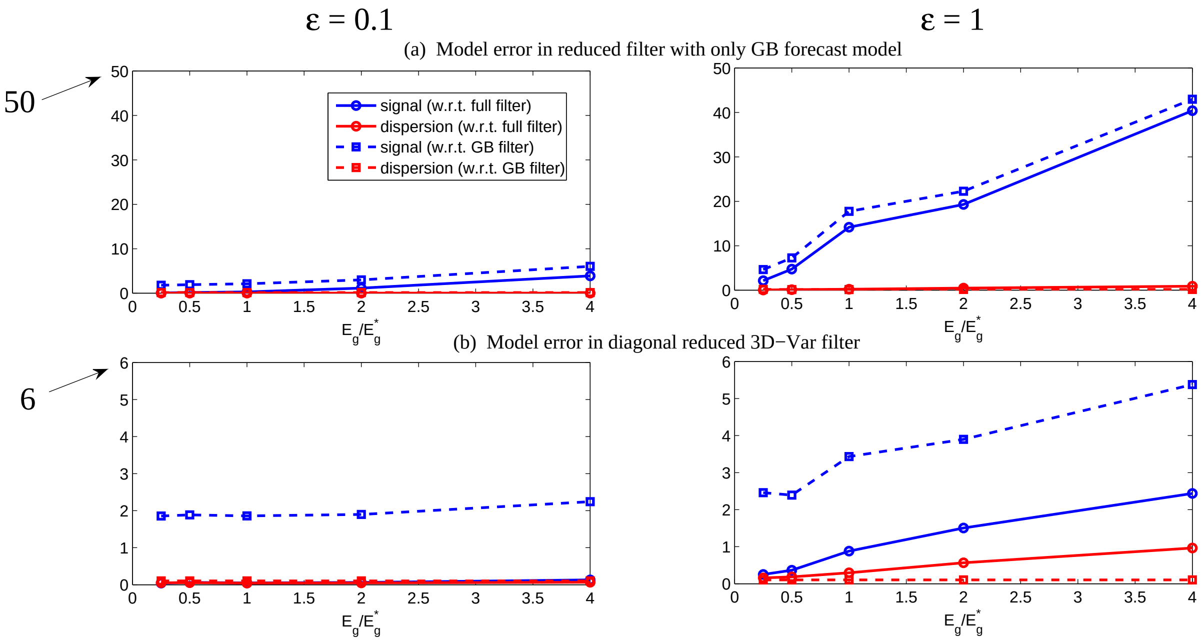

FIG. 14. Comparison of the information model error (27) in the posterior distribution p^M of the two imperfect filters related to that p of the perfect filters. Here p is associated with either the full filter (17) (solid lines) or the idealized GB filter in (18) (dashed lines) in different dynamical regimes.

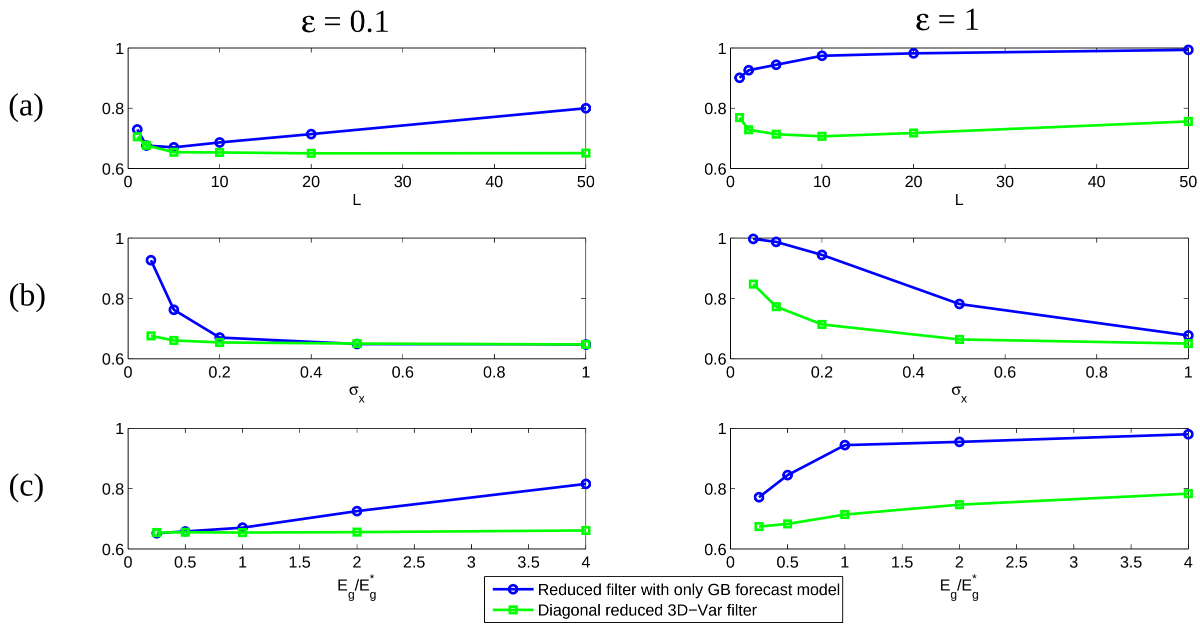

FIG. 15. Hellinger distance as a function of L, σ_x and E_g in the situation that $\gamma = 0$ in the true underlying flow field. Compare row (a), (b) and (c) with the information model error in Figure 5, 6 and 7.